The beginnings of proto-metabolism at the origin of life in alkaline hydrothermal vents

By

Eloi Camprubí Casas

The department of Genetics, Evolution and Environment

University College London

UK

Supervised by

Professor Nick Lane

A thesis submitted for the degree of

Doctor of Philosophy

Division of Biosciences

Declaration

I, Eloi Camprubí Casas, confirm that the work presented in this thesis is my own, except where the work that has formed part of a jointly-authored publication has been included. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The origin of life is still one of the most exciting scientific quests, one that has deep implications that go far beyond science itself. This thesis explores the possible origins of carbon and energy proto-metabolism in Hadean alkaline hydrothermal vents. I first explore the catalytic properties of Fe(Ni)S minerals in the presence of geologically sustained proton gradients, which lower the kinetic barrier to the reaction between H_2 and CO_2. This could have promoted non-enzymatic pathways analogous to the acetyl CoA pathway and reverse incomplete Krebs cycle, which drive carbon metabolism in arguably some of the most ancient microorganisms (methanogenic archaea). One of the predicted prebiotic products is thioacetate, which we have shown can be non-enzymatically phosphorolysed to acetyl phosphate (AcP). AcP is widely used in bacteria and archaea as a key intermediary between acetyl CoA and ATP, driving substrate-level phosphorylation of ADP. Its simple structure and high phosphorylating potential suggest it could have acted as a primordial ATP analogue, driving energy metabolism at the beginnings of proto-metabolism. The thesis then explores the phosphorylating capabilities of AcP on ribose, adenosine and ADP, followed by a study on the condensing capabilities of AcP with glycine and AMP. Finally, I explore how carbon and energy metabolism could have interacted during early abiogenesis. The formose reaction non-enzymatically synthesises sugars (carbon metabolism) from formaldehyde, one of the first products of CO_2 reduction. AcP could hypothetically direct the formose reaction towards biologically interesting sugars such as ribose using AcP (energy metabolism), rather than the much larger and ramified tars often claimed to be its main products. I show how Hadean alkaline hydrothermal vents could have promoted non-enzymatic pathways analogous to the acetyl CoA pathway and reverse incomplete Krebs cycle. I report that AcP can non-enzymatically phosphorylate a variety of biomolecules, but does not promote the polymerisation of glycine or AMP. Most strikingly, I show that AcP can indeed direct the formose reaction towards the synthesis of pentoses such as ribose.

Acknowledgments

Firstly, I would like to express the most profound gratitude to Nick Lane for giving me the opportunity to pursue my passion; without his very early support and encouragement I would not have had such an opportunity to start a career in this fascinating topic of research. His passion for solving some of the most profound scientific questions and his way of igniting an interest in others has been a true inspiration; needless to say, his impact on me goes well beyond the production of this thesis. I would also like to thank my second supervisor John Ward, and the chair of the thesis committee Finn Werner; their input has been crucial and I could not thank them enough for providing a fresh perspective on the issues I faced. I would like to thank my fellow Lane Lab members Alex Whicher, Sean Jordan, Rafaela Vasiliadou, Silvana Pinna, Stuart Harrison, Hanadi Rammu, Sylvia Lim, Hebe Wildi, Jeanne Bonnel, Barry Herschy and Victor Sojo for all their scientific and personal support throughout my PhD. The work presented here would not have been possible without UCL's support staff, facility technicians and postgraduate students who helped me countless times with the myriad of instrumentation used; particularly I would like to thank Kersti Karu for her unparalleled patience and her willingness to always try 'one more time'. I thank 'la Caixa' Foundation and UCL Impact Awards for funding my PhD and allowing this curiosity-driven basic research to not be pushed aside.

Finally, I would like to give special thanks to my friends and family for all their excitement and encouragement. Particularly to my partner Tobías for his love, sacrifice and unending support, which has always gone way beyond the call of duty.

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two-dimensional
ACN	Acetonitrile
AcP	Acetyl phosphate
ADP	Adenosine diphosphate
AEC	3-amino 9-ethylcarbazole
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
cAMP	Cyclic adenosine triphosphate
CoA	Coenzyme A
CODH/ACS	Carbon monoxide dehydrogenase/acetyl CoA synthase
DCM	Dichloromethane
DHA	Dihydroxyacetone
DHAP	Dihydroxyacetone phosphate
DKP	2,5-diketopiperazine (or glycine anhydride)
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E4P	Erythrose 4-phosphate
Ech	Energy-converting hydrogenase
EI	Electron ionisation
ES	External standard
ESI	Electrospray ionisation
e.g.	*exempli gratia* (for example)
F6P	Fructose 6-phosphate

GC-MS	Gas Chromatography – Mass Spectrometry
G3P	Glyceraldehyde 3-phosphate
HMBC	Heteronuclear multiple bond correlation
HPLC	High Performance Liquid Chromatography
IS	Internal standard
i.e.	*id est* (in other words)
KHP	Hydrogen phthalate
LC-MS	Liquid Chromatography – Mass Spectrometry
LTQ	Linear trap quadrupole
LUCA	Last universal common ancestor
MALDI	Matrix-assisted laser desorption/ionisation
MS	Mass spectrometry
MS-MS	Tandem mass spectrometry
Mtr	Methyltransferase
NADH	Reduced nicotinamide adenine dinucleotide
NAD^+	Oxidised nicotinamide adenine dinucleotide
NAG	N-acetyl glycine
NASA	National Aeronautics and Space Administration
NMR	Nuclear magnetic resonance ($^1H/^{13}C/^{31}P$)
PFBOA	o-(2,3,4,5,6-pentafluorobenzyl) hydroxylamine
PFOR	Pyruvate:ferredoxin oxidoreductase
PFTE	Polytetrafluoroethylene (Teflon)
pH	Measure of how acidic or basic a fluid is
Pi	Orthophosphate
PPi	Pyrophosphate
Q-TOF	Quadrupole time of flight

R5P	Ribose 5-phosphate
RNA	Ribonucleic acid
Rnf	Ferredoxin:NAD^+ oxidoreductase
Ru5P	Ribulose 5-phosphate
SD	Standard deviation
SPE	Solid phase extraction
Su7P	Sedoheptulose 7-phosphate
TEAA	Triethylammonium acetate
THAP	2,4,6-trihydroxyacetophenone
tRNA	Transfer ribonucleic acid
UDP	Uridine diphosphate
UPLC	Ultra performance liquid chromatography
UV	Ultraviolet light

Chapter 1

Introduction

1.1 A universal enquiry

The origin of life has been one of humanity's most compelling questions since the cradle of civilization, where innumerable creation myths have tried to shed light onto this essential question without the limitations that a scientific approach to the issue would entail. Until the 19th century, the theory of spontaneous generation was widely accepted, since it was the most comprehensive way to conceive how maggots in rotting meat or mice in grain could appear from apparently thin air. Using the scientific method, Louis Pasteur disproved this theory by showing that small organisms (later known as microorganisms) were everywhere, and that these could not emerge in strictly isolated sterile organic media (Ligon, 2002; Pasteur, 1862). By demonstrating that living organisms could only originate from other living organisms, Pasteur had ultimately unveiled the question of how the first living entity could have emerged.

'How and where did life originate?' remains one of the most fundamental questions for humanity to date. Unfortunately, these enquiries are confronted by the harsh reality that the phenomenon of life has not been fully comprehended yet. One of best examples to visualise this is the lack of a common definition of life (Cleland & Chyba, 2002). The definition most commonly accepted has been proposed by NASA: life is 'a self-sustaining chemical system capable of Darwinian evolution'. This definition (like any other) has problems, arguably the most important being the term 'self-sustaining'. Life is most definitely not self-sustained, just the opposite: it is an open system that exchanges matter and energy with its environment in order to temporarily maintain its far-from-equilibrium state. In Peter Mitchell's words, life and its environment "may be regarded as equivalent phases between which dynamic contact is maintained by the membranes that separate and link them" (Mitchell, 1959).

In turn, most modern definitions avoid the term ‘life’ (a noun) and instead use ‘living’ (an adjective), pointing to the fact that life is more of a transient state which some matter experiences: ‘a living being is any autonomous system with open-ended evolutionary capacities’ (Ruiz-Mirazo et al., 2004). An analogous phenomenon occurs when trying to define entities such as the ‘truth’. If the adjective ‘true’ is used instead (or ‘living’ instead of ‘life’), its study becomes more approachable, but the difficulty of defining the qualities of the overarching phenomenon still remains. These controversies with the definition of life are probably just symptoms of a deeper problem, and unfortunately these have been conducive to arguably the larger issue of the classical divide on the approaches to the origin of life. Each definition of life has classically put its emphasis on a trait(s) expressed by living entities, such as their replicative capabilities, their far-from-equilibrium energetic state, their compartmentalisation, or their evolutionary potential, etc. Each trait has classically been associated with different scientific disciplines (e.g. replication with RNA molecular biology, compartmentalisation with lipidic molecular biophysics, etc.), so each sprung independent lines of research. Mostly due to limitations associated with each highly specialised discipline, but also due to often-antagonistic philosophical positions, these disciplines have often remained strikingly isolated from each other. Needless to say, research on the origin of life is a remarkably broad area of enquiry, requiring expertise encompassing geochemistry, organic chemistry, bioenergetics and molecular biology amongst others; due to this, it cannot be overstated how counterproductive this classical divide has been for the research field.

Despite still being far from the objective, it is undeniable that during the last few decades, considerable experimental progress has been made and, luckily, some of the old controversies have slowly faded away. Hopefully, a better understanding ‘how and where’ life started on Earth will serve as tools in order to comprehend ‘why’ life, as a phenomenon, occurs in the Universe; a goal mankind has always striven to achieve.

1.2 Origin of life research, a historical perspective

Charles Darwin's proposition of a "warm little pond" acting as the cradle of life on Earth was one of the earliest and, due to his scientific pre-eminence, most famous scientific propositions on this issue. Darwin's proposal was rather vague since he avoided the subject publicly and his few comments were in a brief personal note written to Joseph Dalton Hooker, but it sparkled the scientific interest for the study of the origin of life. Darwin's ideas were later developed independently by Haldane (1929) and Oparin (1924). They proposed that life started from simple organic molecules, deriving from CH_4, NH_3, H_2O and H_2 (from a highly reducing atmosphere), which aggregated and in time became the complex biomolecules of life. The scientific community embraced this idea and expanded on it; the Miller-Urey experiment (Miller, 1953) being a crucial contribution which is considered to be where modern experimentation on the origins of life started. Just as Haldane and Oparin suggested, in this experiment CH_4, H_2, H_2O and NH_3 were mixed in an anoxic environment with electrical discharges as a source of energy. These reactions generated a mixture of organic compounds (initially mostly amino acids, then considered to be the building blocks of life) that were proposed to accumulate in the ocean ultimately creating the now well-known "primordial soup". Another postulated source of organic compounds has been the delivery from outer space through meteorites and other space debris; at least some of the organic compounds trapped within would accumulate in the early ocean as well. Regardless of the source of organic material, all these hypotheses imply a heterotrophic origin of life wherein the first living entities would have thrived by fermenting previously formed organics. Historically, this hypothesis has enjoyed a wide, yet often uncritical, acceptance from the scientific community.

The aforementioned hypothesis is experimentally supported by the discovery of several chemical pathways that generate other, more complex, molecules of life today. These proposed chemical reactions require UV radiation, radioactive decay, lightning or shockwaves as the energy sources. Despite this, most are not sustained energy sources and some, including UV radiation, would tend to degrade organics as fast as they promote their synthesis. Many of these

processes also rely on a highly reducing early atmosphere (Oro, 1960; Saladino et al., 2009). There is no consensus on the chemical composition of Hadean (ancient) atmosphere and hydrosphere, particularly regarding nitrogenated compounds, which makes some of the proposed chemistry somewhat tenuous. Furthermore, these chemical systems would readily reach equilibrium (state of maximum entropy) due to the accumulation of reaction products inhibiting the continuing forward reaction. In contrast, living entities are open systems that operate far-from-equilibrium via a continuous influx of reactants and efflux of waste products. Some more dynamic environments, such as tidal beaches, have been postulated which could arguably comprise both a mechanism for the concentration of substrates and for the removal of waste substances. In all of the hypotheses mentioned so far, it is worth pointing that the chemical reactions that synthesised the monomers of life (e.g. amino acids) were not considered to be initially part of it, as, once formed, life supposedly would have invented its metabolism (the pathways to keep synthesising those monomers) basically from scratch. This raises several important questions, which have not been addressed by research into heterotrophic hypothesis: How did life survive the discontinuity between heterotrophic origins and autotrophy, necessary as soon as the primordial soup had been consumed (presumably very early)? Why do all phylogenetic trees place autotrophs as the most deeply branching organisms, and heterotrophs as more recent branches? Why do the bacteria and archaea, the two major groups of prokaryotes, have evolutionarily distinct pathways of glycolysis (Embden-Mayerhoff pathway), a markedly heterotrophic metabolic pathway, but share the same pathway of gluconeogenesis up to C3 sugars, also implying that early life was autotrophic (Berg et al., 2010)? How and why did early life adopt the universally conserved mechanism for conserving energy (chemiosmosis) if it first arose through abiotic mechanisms completely detached from it? All these considerations are far more compatible with an autotrophic origin of life; the ideas for which actually predate those of Haldane and Oparin, going back to Leduc (1911), and Goldschmidt (1952).

All living organisms conserve energy through an electrochemical gradient across membranes, the proton-motive force (Mitchell, 1961), where protons actively pumped across a membrane flux back through the ATP-synthase to

generate ATP. This feature, named chemiosmotic coupling which is common to all life and as conserved as the genetic code, has been proposed to be as crucial as the latter for explaining how, where and why life started (Lane & Martin, 2012).

It is at least feasible that the ubiquity of chemiosmotic coupling suggests a very early origin, and possibly some congruence with an abiotic environment. The question of congruence was addressed by Christian de Duve (2005): when considering the evolution of metabolism, de Duve argues that we must consider how biological catalysts, whether ribozymes or enzymes, first came into existence. His “only reasonable answer” is that they were selected. But enzymes (or ribozymes) could only be selected if they offered some improvement over their protometabolic counterparts, implying that the very first biological catalysts must have enhanced abiotic (geochemical) processes that already existed. This principle argues that biochemistry derived from geochemistry and was gradually improved through the incorporation of more refined enzymes. This implies that to understand how life arose, we need to look at the universally conserved biochemistry within all cells. In this way, it seems reasonable to think that the earliest living entities conserved energy and maintained a carbon metabolism in a way not remarkably different to what the last universal common ancestor (LUCA) did. This, together with the Occam’s razor principle, is the basis of the autotrophic hypothesis for the origins of life.

This hypothesis, as discussed earlier dating back to Leduc (1911) and pioneered more recently by Wächtershäuser (1988b; 1988a) and independently by Russell and colleagues (Russell et al., 1988; Russell et al., 1989), involves a scenario where the continuous flux of small inorganic molecules such as H_2 and CO_2 triggered the appearance of life. Wächtershäuser proposed an “iron-sulphur world” (in contraposition to the well-known “RNA world”) where proto-metabolism was driven by “pyrites pulling”. As initially conceived, from H_2 and CO_2, pyrites pulling has never been demonstrated (Heinen & Lauwers, 1996), and indeed is not found in any living cell either, a glaring lack of congruence. Wächtershäuser was able to reduce CO under ‘black smoker’ conditions (extremely hot and acidic submarine hydrothermal vents). Despite this

experimental success, the actual CO concentrations in these environments are now known to be much lower than of CO_2, which makes this carbon fixating mechanism remarkably implausible. Due to this and the fact that no living organism is known to have a pyrite-driven metabolism, his theory lost some scientific interest but, nevertheless, opened the field for further research on autotrophic origins.

In contrast, Michael Russell proposed alkaline hydrothermal vents (Figure 1.1) as the hatcheries of life (Russell et al., 1988) as these show remarkable congruence with current life, especially in their vectorial chemistry, with both proton gradients and redox gradients across thin inorganic barriers in microporous labyrinths. The discovery of the first deep marine alkaline hydrothermal vent (Kelley et al., 2001) nearly a decade after Russell had predicted their existence greatly supported Russell's views and established an empirically measured set of experimental conditions (pressure, pH, temperature, ion composition, etc.) for further tests. However, it should be noted that all modern vents, including Lost City, are not equivalent to Hadean systems because the ancient oceans were anoxic and rich in CO_2. The anoxic conditions meant that the Hadean oceans were rich in dissolved ferrous iron and other transition metals derived from volcanic vents, notably Ni; and the high CO_2 levels probably rendered the early oceans mildly acidic (pH 5-7) unlike modern oceans, which are mildly alkaline (pH 8). That would have strengthened natural proton gradients in such vent systems and also increased the local concentration of bicarbonate/CO_2 for carbon fixation.

Furthermore, the internal structure of these vents resembles a porous labyrinth, which has caught scientific interest since these porous environments containing steep temperature gradients (up to 90 °C in the inner layers, and down to 20 °C in the outermost ones) are conducive to the physical phenomenon of thermophoresis. Thermophoresis is a process by which solutes accumulate in the colder pockets of a solid porous matrix due to thermal convection of the solution generated by a thermal gradient that crosses it. This phenomenon could have concentrated organics synthesised in the vents by 5000-fold

(Herschy et al., 2014), and even extreme accumulations of ~10^8-fold have been reported (Baaske et al., 2007).

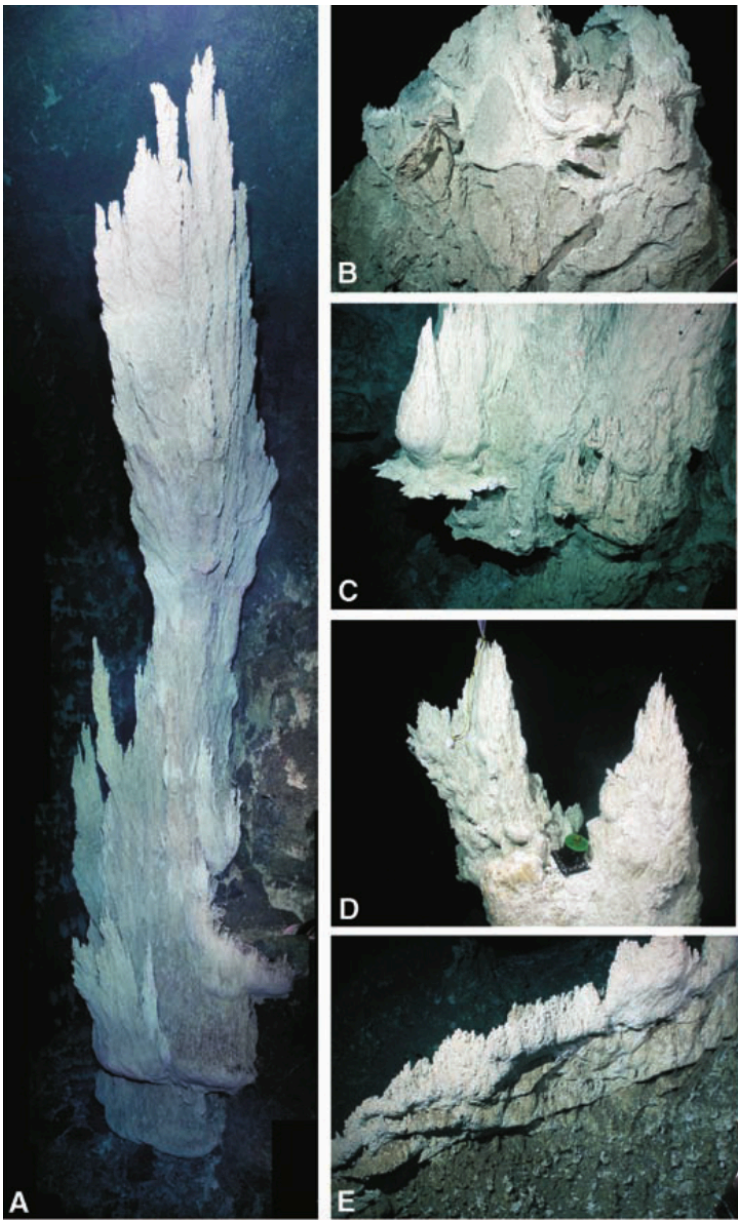


Figure 1.1. Alkaline hydrothermal vents at Lost City (Kelley et al., 2005). (**a**) A 10 m tall actively venting carbonate chimney. (**b**-**e**) Actively venting sections of Lost City at 50 °C.

1.3 From geochemistry to biochemistry

From a biological point of view, the reduction of CO_2 using electrons from H_2 via the acetyl CoA pathway is used by some of the most putatively ancient organisms: methanogenic archaea and acetogenic bacteria (Buckel & Thauer, 2013), which are strict anaerobes. This metabolic pathway is arguably ancient,

as it is the only carbon-fixation pathway conserved in both archaea and bacteria phyla (albeit there are significant differences in the enzymes catalysing the conserved chemistry of transition metal-catalyzed methyl synthesis (Martin, 2012)); this strongly suggests that LUCA conserved energy this way. The acetyl CoA pathway is unique in that it presents the following characteristics: it is the only exergonic pathway of carbon fixation, it is short and linear, it is replete with numerous Fe(Ni)S proteins, and it is obligately chemiosmotic (Fuchs, 2011). Following de Duve's congruence principle, if LUCA used this type of metabolism, then the first protocells did too and, at the same time, they arose in an environment where that type of geochemistry was naturally already occurring. And indeed there is abiotic production of methane in alkaline hydrothermal vents (Proskurowski et al., 2008), which adds to this argument.

Chemical and electrical gradients occur naturally in alkaline hydrothermal vents due to a reaction called serpentinization, which occurs below the seafloor. In this reaction, minerals such as olivine [$(Mg/Fe)_2SiO_4$] react with sea water to produce serpentinite and H_2 dissolved in alkaline fluids. Serpentinization is an exothermic process, releasing heat: the warm, buoyant hydrothermal fluids rise back up to the seafloor and components precipitate into large microporous vent systems. Modern vents such as Lost City consist of spires composed of a mixture of precipitated carbonate (aragonite) and brucite. On the other hand, the Hadean oceans (i.e. circa 4000 million years ago) was acidic due to high atmospheric CO_2 concentrations (Pinti, 2005). The composition of early vents are unknown, but are likely to have contained silicates, brucite, some carbonate, and importantly, Fe(Ni)S minerals such as mackinawite (FeS) and green rust, $Fe(OH)_2$. The structure of the spires is a complex microporous labyrinth with chambers separated by thin mineral walls. This semi-permeable environment created electrochemical H^+ gradients with a magnitude and polarity (acid outside) that was remarkably similar to what can bee seen today across prokaryote membranes (~3 pH units).

The walls of Hadean alkaline vents most probably contained Fe and Ni ions (which, as today, are exhaled from hotter acidic black smokers). The source of Fe and Ni ions in the Hadean alkaline vents is the ocean, which (we know from

the early deposition of banded iron formations) contained high concentrations of these ions due to its increased acidity and absence of dissolved O_2. It has been proposed that these natural proton gradients, together with the Fe(Ni)S clusters acting as catalysts, could have driven abiotic electron flux from H_2 to CO_2 generating organic molecules in an analogous way to modern anaerobes living in equivalent environments (Lane, 2014; Martin & Russell, 2007).

In summary, research based on a heterotrophic origin of life focuses on unveiling chemical routes to generating the components of cells (RNA, proteins, etc.). This approach has commonly been referred as 'bottom-up', since it is centred first in synthesising the molecules of life, and then assessing how these would have later developed life's defining traits (replication, heredity, etc.); this upward directionality (from geology/chemistry up into the tree of life) explains the naming of the approach. These chemical routes are expected to meet a series of criteria (Wong & Lazcano, 2009), achieving high reaction yields and explaining the origin of homochirality being crucial ones. On the other hand, the autotrophic origins hypotheses focus on the metabolism by which current life (and its last universal common ancestor, LUCA) conserves energy, and investigates how these mechanisms could have derived from pre-existing geochemical reactions. This approach has a downward directionality (from biology down into geology/chemistry), and therefore it is commonly known as a 'top-down' approach. The research presented in this thesis, based around alkaline hydrothermal vents, offers a perspective which hopes to balance both approaches and, by doing so, aims to close the current disconnection between prebiotic chemistry and ancient biochemistry.

1.4 Bifurcation between carbon and energy proto-metabolisms

1.4.1 Carbon proto-metabolism

Despite many pieces of evidence supporting the possibility that life could have started in alkaline hydrothermal vents, many questions remain unanswered.

How exactly does CO_2 get reduced to organic compounds such as formaldehyde (HCHO) and more complex organics in these environments? Although the reaction of H_2 and CO_2 to form organics is exergonic overall, the first steps to produce formaldehyde are strongly endergonic, producing a major kinetic barrier to their reaction. So far, experimental efforts to force the reaction of H_2 and CO_2 under plausible abiotic conditions have been somewhat unsuccessful.

Lane has argued that proton gradients across thin semi-conducting barriers could drive the reduction of CO_2 to formaldehyde directly in a manner analogous to modern autotrophic cells such as methanogens (Lane, 2014). Modern cells use the proton gradient to reduce ferredoxin and thence CO_2, using H_2 via the membrane protein Ech (energy-converting hydrogenase). A similar abiotic mechanism is feasible, based on differences in reduction potential depending on pH. Specifically, the reduction potential of all reactions involving protons falls by 59 mV for each pH unit rise in pH. Thus while it is essentially 'impossible' for H_2 to reduce CO_2 at pH 7 (because the reduction potential is not low enough) in vents it could be possible: the reduction potential of CO_2 is higher at pH 6, whereas the reduction potential of H_2 is lower at pH 10, sufficiently low to reduce CO_2 to formaldehyde (Figure 1.2).

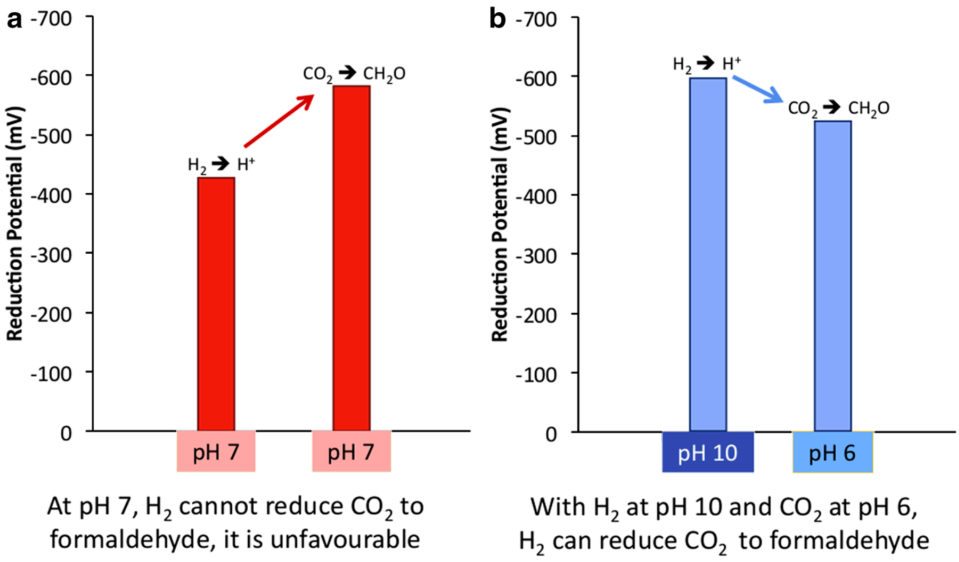


Figure 1.2. Redox reaction between H_2 and CO_2 mediated by pH gradients (Herschy et al., 2014)**.** (**a**) Standard reduction potentials of H_2 and CO_2 at pH 7. Transfer of electrons from H_2 to CO_2 is unfavourable as the reduction potential for CO_2 at this pH is lower (more negative) than H_2. (**b**) With H_2 dissolved in waters at pH 10 and dissolved CO_2 in waters at pH 6 however, the reduction potential for CO_2 becomes higher (more positive) than that of H_2 making the reduction of CO_2 favourable. This would theoretically allow for the reduction of CO_2 to form organic compounds such as formate, formaldehyde, methanol and methane.

The FeS catalytic membrane can also be protonated (Wolthers et al., 2005), so its reduction potential should vary with pH. It should be more easily reduced on the acidic face, and more reducing on the alkaline one; hence electrons should cross the barrier from the alkaline side to the acidic side. This way, the FeS membrane, acting as a semiconductor, should promote the transfer of electrons from H_2 to CO_2 (Figure 1.3). The term semi-conducting used here and thereafter is to emphasize the unidirectional flow of electrons, which is paramount to this hypothesis; in this type of vectorial electrochemistry electrons flow from the highly reducing alkaline fluid to the moderately oxidising acidic oceanic waters.

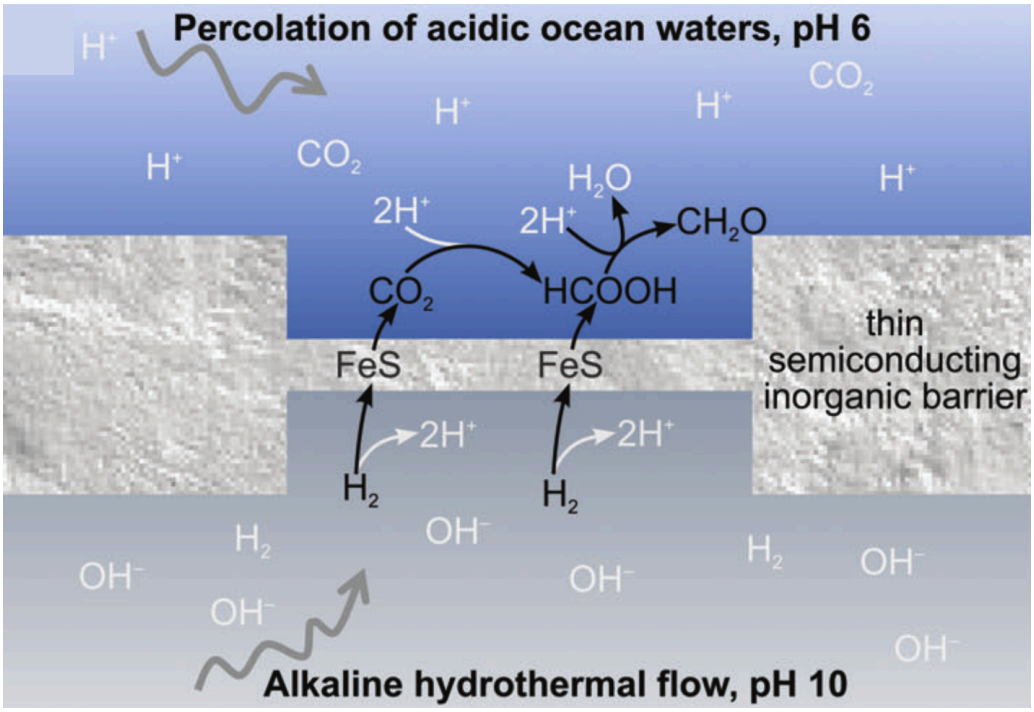


Figure 1.3. Reduction of CO_2 by H_2 across the thin inorganic barrier containing FeS minerals, which mimics a Hadean alkaline hydrothermal vent (Sojo et al., 2016). The reduction potential of the H^+/H_2 couple at pH 10 is lower (–590 mV) than of the CO_2/HCOOH (–370 mV) and HCOOH/HCHO (–520 mV) couples, both at pH 6.

This hypothesis that CO_2 can be reduced by vectorial electron transfer has been tested in the lab (Herschy et al., 2014), and it has been partially successful. Formaldehyde has repeatedly been detected above background noise levels, but it seems that the usage of a large 1.5 L reactor seems to be contributing to the large variation in the data obtained. The lab is currently working on a microfluidics device that should help constrain the reaction coordinates in order to generate more reproducible data. As discussed previously, one of the products of the reaction of CO_2 with H_2 is formaldehyde, which under the conditions found in alkaline hydrothermal vents undergoes the formose reaction. In 1861 Alexander Butlerov discovered the “formose reaction” where formaldehyde undergoes a series of aldol condensation and isomerisation reactions to form sugars, including ribose, in alkaline medium (Figure 1.4). Ribose molecules are the structural backbone of RNA and many other key organic compounds. The formose reaction has since been of interest as an abiotic source of sugar backbones, although it fell out of fashion as feasible abiotic chemistry, as the required strongly alkaline conditions seemed

implausible on the early Earth. Not surprisingly, the discovery of alkaline hydrothermal vents, whose pH of around 11 is perfectly compatible with the formose reaction, changed that perception. Another factor that has contributed to the current interest in the formose reaction is that it has also been proposed to take place in interstellar dust, since its products have been detected remotely (Meinert et al., 2016). It has been demonstrated that the formose reaction serves as a supply of sugars, particularly ribose, under simulated alkaline hydrothermal vents (Herschy et al., 2014, Kopetzki & Antonietti, 2011); even so in a low yield, this reaction, coupled to the concentration-enhancing phenomenon of thermophoresis, constitutes a plausible abiotic source of ribose.

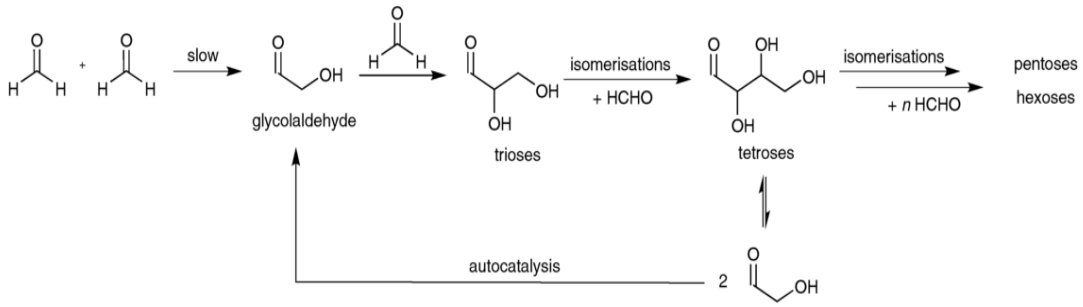


Figure 1.4. Schematic representation of the formose reaction starting with the condensation of two formaldehyde molecules up to pentoses/hexoses (Kopetzki & Antonietti, 2011). Note that the reaction continues further and creates longer chain and branched sugars. A positive feedback cycle by autocatalysis is depicted where a tetrose gets hydrolysed and generates two molecules of glycolaldehyde speeding up the reaction. Many side reactions take place alongside those depicted here but have been omitted for clarity.

Fe(Ni)S minerals at alkaline hydrothermal vents could conceivably promote the synthesis of other organic molecules apart from formaldehyde. As mentioned earlier, the acetyl CoA pathway is arguably an ancient metabolic pathway found both in methanogenic archaea and acetogenic bacteria. These organisms rely on the reaction between CO_2 and H_2, which is exergonic, to grow. The enzymes that catalyse the step-wise reduction of CO_2 in this pathway, such as the acetyl CoA synthase, are replete with Fe(Ni)S clusters. It is indeed conceivable that, perhaps at lower yields, the surface of Fe(Ni)S minerals at alkaline hydrothermal vents could promote (i) the reduction of mineral-bound CO_2 molecules to bound-methyl groups (via hydrogenation), which in turn could be (ii) carbonylated by bound-CO moieties (originally from CO_2), yielding mineral-

bound acetyl groups in a manner analogous to the mechanism of acetyl CoA synthase present in modern methanogens (Figure 1.5). At alkaline hydrothermal vents, a molecule that could elute the bound acetyl group from the mineral surface could be SH^-, yielding thioacetate (or CH_3S^-, yielding methyl thioacetate), which could have acted as a simpler primordial precursor of acetyl CoA. Further carbonylation and hydrogenation reactions could yield Krebs cycle intermediates starting from thioacetate/methyl thioacetate. This kind of reactivity could have constituted the core of carbon proto-metabolism since the acetyl CoA analogous pathway would have been necessary to feed the reverse incomplete Krebs cycle, which in turn would have supplied the carbon skeletons necessary for further amino acid and nucleotide synthesis.

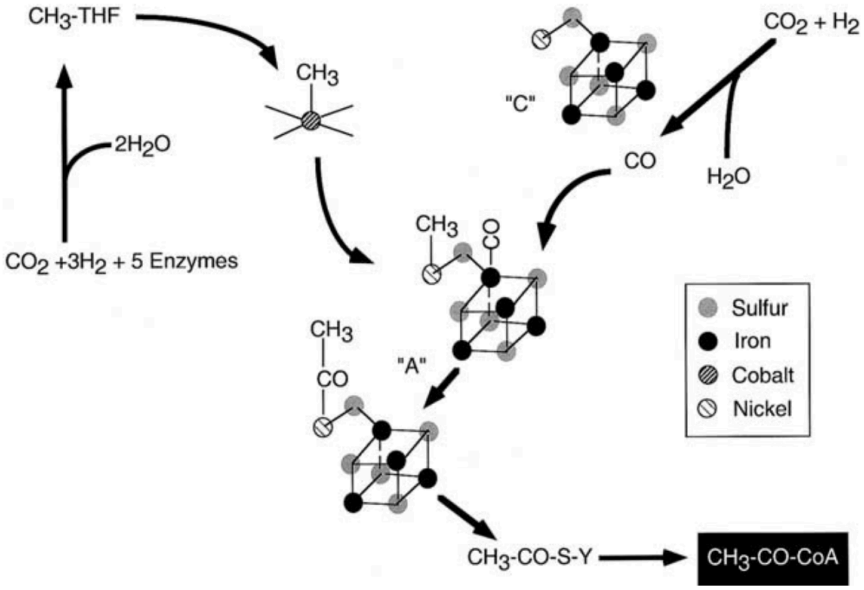


Figure 1.5. Schematic representation of the active sites of acetyl CoA synthase (Cody et al., 2004). The left reaction branch sequentially reduces CO_2 to a tetra-hydrofolate (THF) bound methyl group. This methyl group is transferred to a cobalt cobalamin cofactor and then to a Ni_p(proximal)-Ni_d(distal)-Fe_4S_4 cluster ("A"). The right reaction branch promotes the water-gas shift reaction reducing CO_2 to CO by utilizing a second Ni_p-Ni_d-Fe_4S_4 cluster ("C"). The CO carbonylates cluster "A". Carbonyl migration and insertion (carbonylation) between the Ni-CH_3 bond yields the acetyl group. Transfer of the acetyl group first to a neighbouring cysteine and then to the thiol end of coenzyme A yields acetyl CoA.

In summary, thioacetate or methyl thioacetate could have acted as a primordial precursor of acetyl CoA (Figure 1.6), linking the putative earliest source of energy, the exergonic reaction between CO_2 and H_2. This could have happened

via a prebiotic analogous to the acetyl CoA pathway, with both the incomplete reverse Krebs cycle and gluconeogenesis (carbon proto-metabolism), and the synthesis of acetyl phosphate, which could have acted as a prebiotic precursor to ATP (energy metabolism, see below). This kind of FeS-promoted catalysis has been assessed in this project and is discussed in detail in Chapter 3.

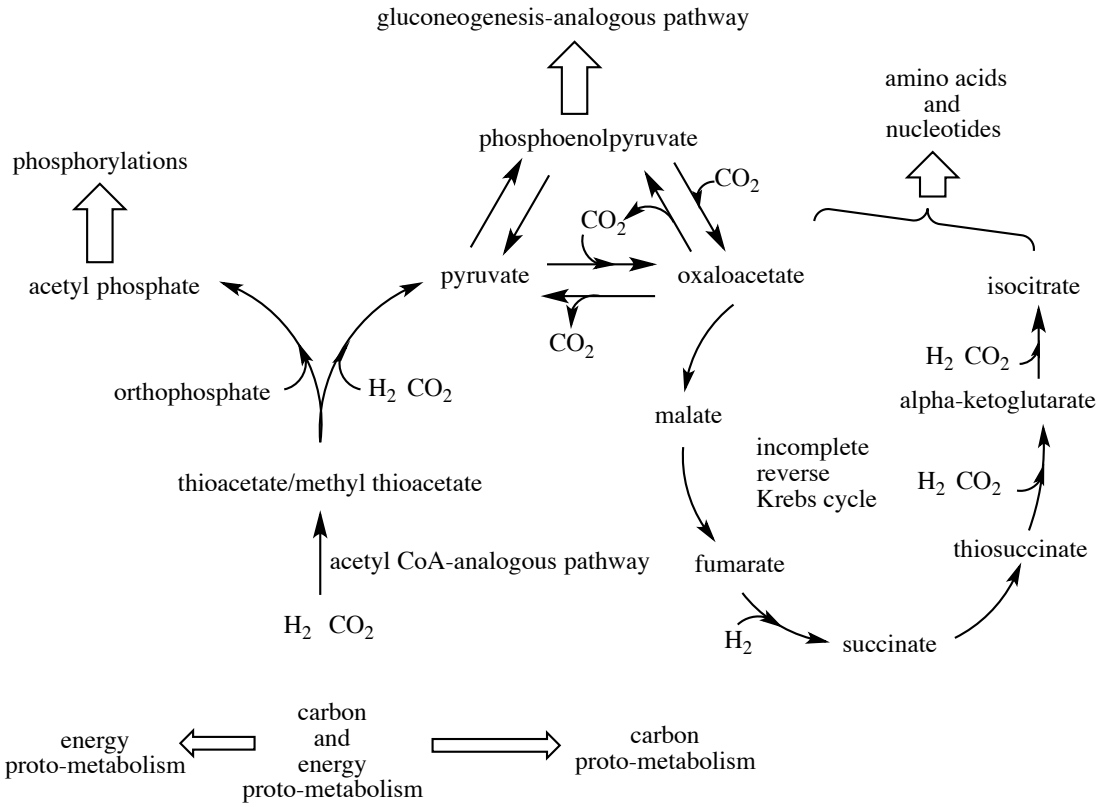


Figure 1.6. Schematic summary of the reactions of carbon and energy metabolism relevant to modern biochemistry and possibly relevant to prebiotic proto-metabolism (amended from Martin & Russell, 2007).

1.4.2 Energy proto-metabolism

From an origins of life point of view, one of the crucial next steps after obtaining simple organics would be to phosphorylate them. Due to its ionising properties, phosphorylation could have helped molecules to be retained either inside protocells (organic membranes are impermeable to large ions), or onto the positively charged mineral surfaces (Westheimer, 1987). Abiotic phosphate

metabolism is of crucial interest as phosphorylated molecules (e.g. ATP) serve as the universal energetic currency of modern cells. Acetyl phosphate (AcP) has been proposed as a primordial phosphorylating molecule (de Duve, 1995; Martin & Russell, 2007; Pratt, 2011; Whicher et al., 2018).

Furthermore, prebiotic phosphorylation is of interest since phosphorylated molecules can, theoretically, condense more easily in aqueous solution, mostly due to the phosphates acting as good leaving groups due to their strong resonance stabilisation. In modern cells, condensation reactions are achieved by the phosphorylation of substrates; and some limited phosphorylation can be achieved by phosphate minerals and other compounds, including ATP, in the absence of enzymes (Burcar et al., 2016; Gull et al., 2015; Costanzo et al., 2007; Saygin, 1983; Schieven & Martin, 1988; Vasant et al., 2005). AcP has a higher phosphorylating potential than ATP due to the more negative free energy of hydrolysis of its 'energy-rich' phosphoester bond (−43 kJ/mol versus −31 kJ/mol in standard conditions) (Martin & Russell, 2007); this allows for a wider variety of synthetic coupled reactions with high energy requirements to take place when using AcP rather than with ATP. AcP is much simpler than ATP and could theoretically be synthesized from the reaction of thiolated compounds (such as acetyl CoA biotically or simpler thioesters abiotically) with inorganic phosphate (de Duve, 1991; Whicher et al., 2018) (Figure 1.5). As explained in the previous section, thiolated compounds, such as thioacetate or methyl thioacetate, could have been synthesized from CO_2, H_2 and SH^-/CH_3S^- using Fe(Ni)S minerals as catalysts in alkaline hydrothermal vents (Camprubi et al., 2017), or from CO and CH_3SH (Huber & Wächtershäuser, 1997). AcP and its phosphorylating properties on sugars, nucleosides and nucleotides have been studied in detail during this project and are discussed in detail in Chapter 4.

Another as to yet unanswered question for the hydrothermal vents hypothesis point of view is how monomers such as nucleotides and amino acids would polymerise (via condensation) to generate oligomers like nucleic acids and peptides respectively in aqueous conditions. A dehydration reaction (which yields water) is a subtype of condensation and is thermodynamically hard to achieve when water is the solvent, due to product inhibition (Figure 1.7a). The

severity of this problem has even led some to suggest that life must have originated on Mars, where there was less water, and was transferred to the Earth on planetary meteorites (Benner, 2013). However, this argument does not stand up to much scrutiny. Other attempts to answer this question propose several Earth-based dehydrating mechanisms (e.g. tidal beaches or low water activity environments (Burcar et al., 2015)), whilst others opt for primordial solvents other than water (e.g. formamide) (Saladino et al., 2009), which eliminate the product inhibition problem of having a dehydration in water altogether. Whilst this approach has generated some measurable results, it also generates a rather large gap between prebiotic chemistry and ancient biochemistry according to de Duve's congruence principle. The proposed solutions to this problem from the perspective of autotrophic origins generally consists of using previously activated monomers, since this is what life does (Figure 1.7b), (e.g. imidazole-activated nucleotides (Burcar et al., 2015) or cyclic nucleotides (Costanzo et al., 2009)) which are in effect pre-condensed versions of the monomer and thus, facilitate further condensations to form the polymer. The main problem with this approach is that the way of obtaining such condensed monomers remains unclear. The capacity of AcP for promoting the condensation of glycine and AMP in water under simulated alkaline hydrothermal vents has been studied in this project and is discussed in Chapter 5. We expected that AcP would help condense nucleotides in aqueous conditions, albeit producing small oligomers in low yields. Higher yields would be expected in future experimentation when adding chelated ions (e.g. Mg^{2+} chelated by aspartate oligomers) or mineral surfaces (e.g. ground brucite mineral), as these conditions would roughly mimic the predicted natural evolution steps occurred in vents: from unspecific and low-yielding reactions in an aqueous medium, to highly productive and stereospecific reactions occurring at the active centres of enzymes or ribozymes.

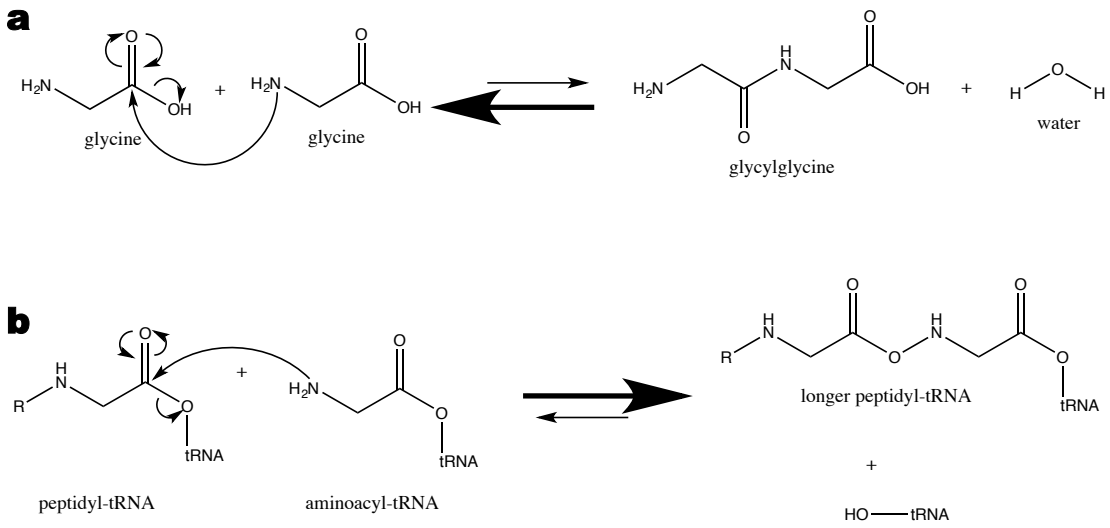


Figure 1.7. Glycine polymerisation mechanisms. (**a**) Dehydration reaction between two unactivated glycine molecules, yielding a water molecule which inhibits the forward reaction in water. (**b**) Condensation reaction between the aminoacyl-tRNA of glycine (acting as a nucleophile) with an activated peptidyl-tRNA glycine molecule, yielding a longer oligopeptide and a free tRNA molecule; no product inhibition occurs here since no water is produced.

AcP might be able not only to drive condensations in water through phosphorylation, but also possibly direct prebiotic pathways such as the formose reaction towards biological products. Even under ideal alkaline hydrothermal conditions, the formose reaction is frequently claimed to yield little more than tars and complex mixtures of biologically irrelevant sugars (Gollihar et al., 2014). Because of this, a way of funnelling the formose reaction towards its pentose products (particularly ribose) has become one of the main focuses of this research field. Several different mechanisms have been proposed to this goal, for example ultraviolet radiations in interstellar dust (Meinert et al., 2016), co-incubation of the formose reaction with borate minerals (Furukawa et al., 2013; Pepi et al., 2010; Ricardo et al., 2004), and phosphorylation-mediated funnelling of the formose reaction (Mellersh & Smith, 2010). While Mellersh & Smith (2010) suggested that phosphorylation could drive the synthesis of ribose phosphate after it has been synthesised by the formose reaction, there is as yet no experimental verification. Chapter 6 of this document explores in depth the formose reaction at alkaline hydrothermal vent conditions, and how the addition of a phosphorylating agent (acetyl phosphate) could funnel its products towards the synthesis of pentoses, such as ribose. In conclusion, Chapter 6 focuses on

assessing the potential early interactions between carbon and energy proto-metabolisms.

1.5 Objectives

This project aims to have a better understanding of the reactivity landscape afforded by Hadean alkaline hydrothermal vents, which could have arguably been conducive to the emergence of proto-metabolism in a congruent step-wise fashion. Accordingly, the main four aims of this study were to:

1. Assess whether Fe(Ni)S minerals could have catalysed chemical reactions analogous to those of the acetyl CoA pathway and the incomplete reverse Krebs cycle, by acting as semiconducting barriers promoting redox reactions between reactants, whose reduction potentials could have been attuned by the steep natural pH gradients present at Hadean alkaline hydrothermal vents. The synthesised intermediate molecules would include thioacetate, which is conducive to the synthesis of AcP.

2. Study whether AcP could have acted as a primordial non-enzymatic ATP analogue by promoting substrate-level phosphorylations on simple organics, such as ribose, adenosine and ADP, under simulated alkaline hydrothermal vent conditions.

3. Evaluate whether similar non-enzymatic phosphorylations on small organics (AMP and glycine) would have been conducive to their non-enzymatic condensation yielding oligomers under simulated alkaline hydrothermal vent conditions.

4. Finally, assess whether the formose reaction could have acted as a credible source of biologically important sugars, such as ribose, at Hadean alkaline hydrothermal vents; and study whether the addition of a

phosphorylating agent (such as AcP) could have funnelled it towards the synthesis of pentoses, as suggested in previous literature.

Chapter 2

Materials and methods

2.1 AcP-mediated phosphorylation experiments

2.1.1 Acetyl phosphate production

A method developed by Crans & Whitesides (1983) was used to synthesize disodium acetyl phosphate. A volume of 45 mL of phosphoric acid (85 wt. %, Sigma-Aldrich) was dissolved in 400 mL of ethyl acetate (Sigma-Aldrich, ≥99.7%). In order to boost the stability of the produced AcP, the solution was cooled to 0 °C using an ice bath and 125 mL of precooled (0 °C) acetic anhydride (Fluka, ≥99%) was slowly added over 15 min. The mixture was stirred for 6 h at 0 °C and added to a suspension of 330 mL of HPLC water, 170 g of ice and 56 g of sodium bicarbonate (Sigma-Aldrich, 99.5-100.5%). This mixture was stirred at 0 °C until no more CO_2 evolved. The mixture was transferred to a separation funnel and mixed vigorously; the organic layer was then separated and discarded. The resulting liquid was washed with 600 mL of ethyl acetate followed by 330 mL of ethyl acetate again to remove as much acetic acid as possible. The resulting volume was neutralized using NaOH, which caused the appearance of a thin ethyl acetate layer that was discarded using a separation funnel. The final product was divided into aliquots and stored at −80 °C.

The usage of commercial AcP (lithium-potassium salt, ≥85%, Sigma) was later adopted for experimentation in order to exclude the potential adverse effects from contaminants (notably acetate and inorganic phosphate) abundantly present in the synthesised stock. A 1.1 M solution of commercial AcP was prepared, divided into aliquots and stored at −80 °C.

2.1.2 Colorimetric quantification of acetyl phosphate

A method from Fowler et al. (2011) was used to determine the concentration of AcP in the prepared stock solution. According to the literature, the prepared solution should contain 1.1 M AcP; this assay was performed to ensure this, so that accurate quantities of phosphorylating agent would be used across all experiments. A 0.1 M of commercial lithium-potassium acetyl phosphate stock solution was prepared adding 0.092 g in 5 mL HPLC water and divided into micro-centrifuge tubes (0.25 mL) and frozen at −20 °C. An AcP dilution series was prepared from the stock AcP solution in HPLC water at a concentration range from 10 mM to 100 µM to act as a calibration curve. A 2 M hydroxylamine hydrochloride (Sigma-Aldrich, 99%) solution was prepared by dissolving 1.39 g in 5 mL HPLC water; the pH was adjusted to 7 using 10 M NaOH, yielding a final volume of 10 mL. This solution could be stored for up to 90 days at 4 °C. A 0.5 M ferric chloride solution was prepared by dissolving 1.35 g of ferric chloride (Sigma-Aldrich, ≥99.99%) in 5 mL of HPLC water and adding 4.13 mL of 12 M HCl. The volume was brought to 10 mL using HPLC water; this solution could be stored at room temperature. A desired volume of development solution was prepared freshly by mixing equal volumes of the ferric chloride solution and HPLC water.

Aliquots of 600 µL of the prepared AcP solution or standard curve were added to 1.5 mL Eppendorf tubes. Hydroxylamine hydrochloride solution (100 µL) were added and left for 5 min at 60 °C. After the incubation time, 200 µL of development solution were added and mixed. Then, each vial was immediately transferred to 1.5 mL polystyrene cuvettes and measured on a spectrophotometer at 500 nm. An aliquot with HPLC water instead of the 600 µL of sample was used as a spectrophotometric blank. The spectrophotometer used was a NanoDrop 2000c (Thermo Scientific) in cuvette mode with NanoDrop 2000/2000c (1.4.2 version) software.

2.1.3 Derivatisation of sugars (phosphates) for LC-MS analysis

Some molecules need to be chemically modified for chromatographic assay; this is the case for both sugars and sugar phosphates for LC-MS analysis. Sugar phosphates should not be analysed using other chromatographic techniques, such as gas chromatography, as the phosphate moiety is too polar for most columns. This polarity also lowers the ionisation efficiency and dramatically reduces the signal intensity. Alternatively, negative ion mode LC-MS could be used for the analysis of un-derivatised sugar phosphates but (i) usually the resulting signal intensity is poor, and (ii) the sugars in the sample can react further, yielding unreliable information. LC-MS analysis using positive ion mode of derivatised sugars is the best option, as it affords a high sensibility assay and the derivatised analytes in the sample are chemically unreactive. The method developed by Han et al. (2013) was employed (Figure 2.1); this method chemically modifies the sugars by attaching a molecule of 3-amino 9-ethylcarbazole (AEC) through a reductive amination of the 1' carbon of the sugar. This modification adds molecular weight and hydrophobicity, which allows for an optimal separation of the sugars during the liquid chromatography, as well as providing good positive ionization sites (the nitrogen atoms of AEC) for the mass spectrometry analysis.

Initially, a 25 mM stock solution of AEC (Sigma-Aldrich, ≥95%) was prepared by diluting 0.0263 g in methanol. A 50mM stock solution of $NaCNBH_3$ (Acros Organics, 95%) was also prepared diluting 7.83 mg in 10 mL water. 10 mL of a solution made of 2:1 DCM-hexane (Sigma-Aldrich, 95%) was created. A disodium D-ribose 5-phosphate (Sigma-Aldrich, ≥98%) 1 mM solution was prepared by diluting in HPLC water to act as a standard.

A 50 µL aliquot of standard solution or sample was introduced into 1.5 mL Eppendorf tubes followed by addition of 100 µL of AEC solution in methanol, 50 µL of $NaCNBH_3$ solution and 20 µL of acetic acid (which acts as a reaction catalyst). The reaction mixtures were incubated at 70 °C for 60 min. After the reaction, the tubes were cooled on ice for 1 min. Then, in order to remove excess derivatisation reagents a liquid-liquid extraction step was performed by

adding 300 µL of HPLC water and 300 µL of the DCM-hexane solution; each tube was vortexed and centrifuged at 10,000 rpm for 5 min in a microcentrifuge. A volume of 150 µL of the upper aqueous phase was transferred to an LC autosampler vial and mixed with 850 µL of HPLC water. The resulting 1 mL was finally manually filtered using 0.22 um filters (13 mm PFTE, Gilson Scientific) in order to preclude any blockage of the chromatography column.

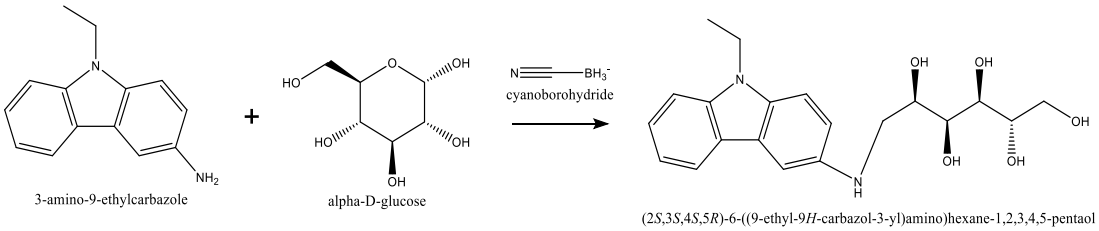


Figure 2.1. Derivatisation reaction of glucose using AEC. Other sugars and sugar phosphates are derivatised in the same way by the reductive amination of the carbonyl group of the sugar.

Later experimental results suggested that the concentration of derivatisation reagents initially used was not enough to fully derivatise high concentrations (> 100 mM) of sugars present in some samples (see Section 4.2.2). Therefore the concentration of derivatisation agents was increased to 238 mM of AEC in methanol (to saturation), and 476 mM of $NaCNBH_3$ in water. Using the same 50:100:50 µL ratios of derivatisation reagent volumes, this new protocol allowed for the full derivatisation of up to 500 mM of total sugars present in the 50 µL of sample. In order to compensate for the higher concentrations of derivatisation reagents used, and therefore to not saturate and damage the chromatography column, the post-derivatisation purification step was expanded into three liquid-liquid extractions. First (i) 300 µL of HPLC water plus 300 µL of 2:1 DCM-hexane solution were added to the 220 µL of derivatisation volume, followed by vortexing and microcentrifuging at 10,000 rpm for 5 min; then 450 µL of the aqueous upper phase were collected in a separate vial. Then (ii) to the remaining derivatisation reaction volume, another 300 µL of HPLC water plus 300 µL of 2:1 DCM-hexane solution were added, followed by the usual mixing and microcentrifuging; this time 250 µL of the aqueous upper phase were collected and added to the previously collected 450 µL. Finally (iii) to the remaining derivatisation reaction volume, another 300 µL of HPLC water plus

300 µL of 2:1 DCM-hexane solution were added, followed by the usual mixing and microcentrifuging; this time 300 µL of the aqueous upper phase were collected and added to the previously collected 700 µL, making a total of 1 mL. In order to wash the final 1 mL of any remnant AEC, 500 µL of 2:1 DCM-hexane solution were added, followed by the usual mixing and microcentrifuging. The resulting 750 µL of upper phase contained all the derivatised analyte whilst containing only trace amounts of AEC. This final volume was then either (i) manually filtered using 200 µm filters (in the case of commercial standards) and directly injected on the LC-MS instrument, or (ii) put through an extra purification step using solid phase extraction (SPE) cartridges in order to eliminate any polar ions (notably phosphates or Ca^{2+}) which would damage the instrument.

2.1.4 Solid phase extraction (SPE)

The presence of high concentrations of ions in samples being analysed by LC-MS is harmful to the instrument, as these tend to precipitate inside the ionisation chamber, reducing ionisation efficiency over time, as well as clogging the internal metallic tubing. Due to this, samples that contained such ions (phosphates and Ca^{2+} mainly) had to be purified using a solid phase extraction (SPE) protocol. This purification protocol was partially developed by undergraduate student Jeanne Bonnel under my supervision. It was decided to use a non-polar C18 stationary phase because sugars, once derivatised using AEC (Section 2.1.3), become rather non-polar and thus would be adsorbed to the SPE cartridge matrix, whilst polar ions would be washed away. Another possibility was to SPE-purify the sugars before derivatisation but this way, due to the similar polarity of analytes (sugars) and contaminants (small ions), it would have been necessary to alter the pH of the samples and that could have easily contributed to generate analytical artefacts, therefore this was discarded. The cartridges used were Thermo Scientific's Hypersep 500 mg/2.8 mL C18-Hypersil. A vacuum manifold was used in order to maximise the time efficiency of the purification process, as this allowed for 12 samples to be purified

simultaneously. The initial SPE purification process was designed using the following scheme:

1. SPE cartridge was conditioned using 1 mL of pure methanol (volume discarded to waste).
2. SPE cartridge was conditioned using 1 mL of HPLC water (to waste).
3. 200 µL of derivatised sample was passed though the matrix and the resulting liquid would not be recovered (to waste).
4. The analytes-containing matrix was washed with 1 mL of HPLC water (to waste).
5. The analytes was eluted using 1 mL of a given % of methanol in water, and this volume would be recovered and analysed further using LC-MS.

In order to assess what percentage of methanol should be used to elute the derivatised sugars from the SPE matrix, a series of tests were carried out calculating the % recovery of an initial concentration of 1 mM of commercial ribose and ribose 5-phosphate when eluted with increasing % of methanol. Methanol was tested first as it is commonly available and provides a good intermediate point in polarity (note that derivatised sugars are only mildly non-polar and the eluting solvent must match that). A concentration of 80% methanol was shown to display the best polarity level for eluting the adsorbed derivatised sugars (Figure 2.2) in the case of ribose; for simplicity, this concentration was also used to elute all the other sugars.

Despite a concentration of 80% methanol being optimal, using a volume of only 1 mL was not enough as only a 50% recovery was achieved. In order to improve this, a test using increasing volumes of 80% methanol during the last eluting step was performed. The results showed that using 3 mL of 80% methanol in water offered the best recoveries with 99.43% in the case of ribose and 86.87% in the case of ribose 5-phosphate; these results were obtained by averaging the results of 10 independent tests with an initial concentration of 500 µM of each sugar. These percentage recovery values were used as a correcting factor when experimentally quantifying sugars; the % recovery value for ribose

was used for the quantification of every non-phosphorylated sugar, and the value for ribose 5-phosphate was used for all phosphorylated ones.

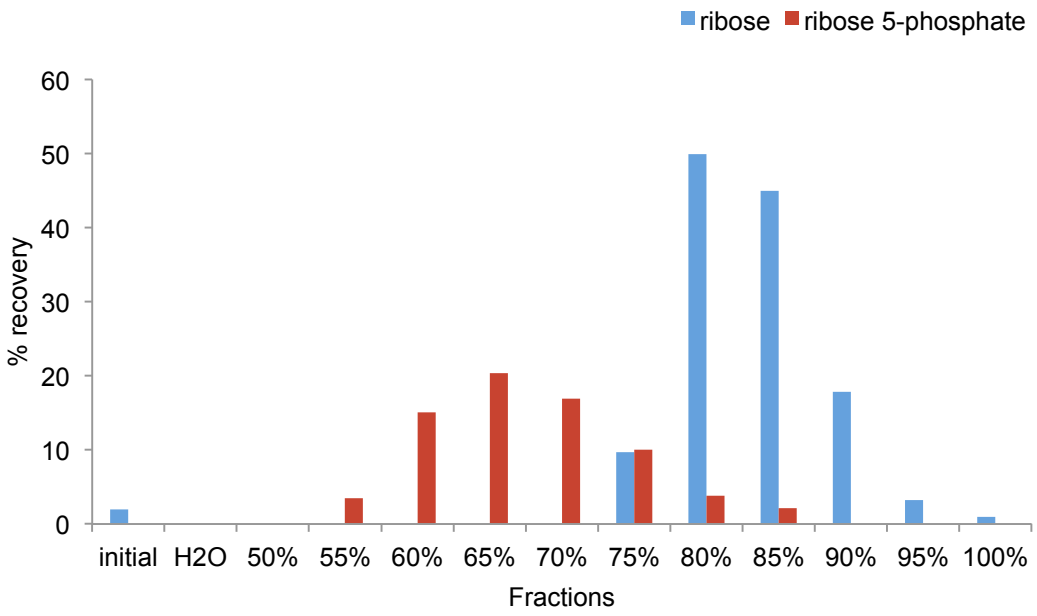


Figure 2.2. Percentage recoveries of commercial ribose (blue) and ribose 5-phosphate (red) at increasing % of methanol (x axis) during the last 1 mL elution step. The initial effluent ('initial') that would normally go to waste, as well as the water wash ('H2O'), were recovered and analysed to ensure that no analyte was being lost during these steps. Data acquired by Jeanne Bonnel under my supervision.

This SPE protocol was applied both to the samples from the late ribose phosphorylation experiments, as well as to those of the formose reaction tests.

2.1.5 Liquid chromatography – mass spectrometry

Once the samples had been derivatised according to the previous method, they were analysed using LC-MS, which is a good method for analysing polar analytes such as sugars and sugar phosphates (even after derivatisation). LC-MS analysis was carried out at the mass spectrometry laboratory at UCL's Christopher Ingold building. The first LC-MS system used was an Acquity Ultra Performance (Waters) LC plus Acquity Sample Manager equipped with a tuneable UV-detector and a Single Quadrupole (Waters) MS in electrospray

ionization (ESI) mode. HPLC water with 0.1% formic acid (solvent A) (Fisher Chemical, ≥98%) and acetonitrile (solvent B) (Sigma-Aldrich, ≥99.9%) were used as solvents. This system was on an open-access regime, so not ideal for long-term data acquisition due to a large number of different labs using it, but was ideal for the initial reaction conditions screening. Data analysis performed in MassLynx (4.1 version) software. Pure methanol blanks were always injected between samples to avoid cross contamination and cluttering of the column. The analytical conditions used are summarised in Table 2.1.

Table 2.1. Operational conditions of the Acquity UPLC-MS.

LC-MS	
Column	Acquity UPLC (Ultra Performance LC) BEH C18 1.7 µm, 2.1 mm x 50 mm
Column temperature	50 °C
Mobile phase	90% (A)/10% (B) (up for 8 min) →72% (A)/28% (B) (for 0.1 min) → 0% (A)/100% (B) (for 2.4 min)
Column flow rate	0.6 mL/min
Ion source	ESI (+)
Ion source temperature	120 °C
Atm. pressure ionisation gas	Nitrogen
Desolvation Temperature	350 °C
Injection volume	10 µL
Scan range	100-600 m/z

Another LC-MS instrument was used at later dates to improve sensitivity and, due to its high mass accuracy capabilities, was used for an isotopic assay with 7-deuterated glucose (97% atom D, Sigma-Aldrich). These analyses were performed using an Agilent 1200 Series LC system connected to an Agilent 6510 Q-TOF mass spectrometer. The mobile phases were identical as in the first LC-MS setup. For the first 0.5 minutes, the effluent from the column was sent to the waste and after to the MS. The analytical conditions used are summarised in Table 2.2.

Table 2.2. Operational conditions of the Agilent LC-MS (Q-TOF).

LC-MS	
Column	Thermo Hypersil Gold C18 1.9 µm, 150 mm x 2.1 mm
Column temperature	30 °C
Mobile phase	Initial: 85% (A)/15% (B) (up during 18 min) →0% (A)/100% (B) (for 1 min) → 85% (A)/15% (B) (for 3 min)
Column flow rate	0.25 mL/min
Ion source	ESI (+)
Injection volume	10 µL
Scan range	150-2000 m/z
Acquisition rate	0.57 spectra/second
Calibration references	121.0509 and 922.0098 m/z
Centroid transformation	Mass Hunter
Electrospray voltage	4000 V
Fragmentor voltage	175 V
Gas temperature	320 °C

A different LC-MS setup was used at later dates to improve the quantification reproducibility, as this instrument was more commonly available. These analyses were performed using a Thermo HPLC Accela 600 pump and autosampler LC system connected to a Thermo Finnigan (linear trap quadrupole) LTQ mass spectrometer. The mobile phases were identical as in the two previous LC-MS setups. The analytical conditions used are summarised in Table 2.3.

Table 2.3. Operational conditions of the Thermo LC-MS (LTQ).

LC-MS	
Column	Thermo Hypersil Gold C18 1.9 µm, 150 mm x 2.1 mm
Column temperature	35 °C
Mobile phase	Initial: 98% (A)/2% (B) (up during 18 min) → 5% (A)/95% (B) (for 1 min) → 98% (A)/2% (B) (for 5.5 min)
Column flow rate	0.2 mL/min
Ion source	ESI (+)
Injection volume	10 µL

2.1.6 Ribose phosphorylation

Solutions of D-ribose (0.3 M) and AcP (0.3 M) were prepared and, when applicable, $MgCl_2$ (hexahydrate, ≥98%, Alpha Aesar) $CaCl_2$ (dihydrate, ≥99%, Sigma-Aldrich) or boric acid (H_3BO_3) (≥99.5%, Sigma-Aldrich) were added (0.15 M). The final reaction volumes of 4 mL were contained in 15 mL Falcon tubes. Experiments were performed at pH 7, 9 and 11 (titrated using negligible volumes of 5 M HCl and 10 M NaOH solutions as needed) at 20 °C, 50 °C and 60 °C and samples taken at 2, 30, 60, 120 and 300 minutes, and 1, 2 and 5 days. After acquisition, all individual samples were immediately frozen at −80 °C, and thawed as batches right before derivatisation. Chemical derivatisation and analysis was performed using the LC-MS analytical protocol described at Sections 2.1.3, 2.1.4 and 2.1.5.

Ribose phosphate was identified by the comparison with a commercial standard of its (i) chromatographic retention time, (ii) accurate mass of the precursor ion, and (iii) unique MS-MS secondary fragmentation pattern. Ribose phosphate was quantified using calibration curves of the integrated chromatographic areas of commercial standards for each experimental condition (pH and ion concentration). D-fucose (Sigma-Aldrich, ≥98%) was used as an external standard in order to minimise data dispersion across months of analytical assays, as well as to monitor instrumental and derivatisation efficiencies. A 250 µM solution of fucose in HPLC water was prepared, divided into 50 µL aliquots and stored at −80 °C. One of the aliquoted fucose solution vials was thawed every time an analytical assay was to be performed and was derivatised together with the rest of the samples. Chromatographic ribose phosphate signal (from a calibration curve or experimental samples) was always divided by the signal of the external standard (fucose), which had been derivatised on the same day and ran first on the sample sequence (in triplicate to maximise reproducibility).

2.1.7 Adenosine diphosphate phosphorylation

Equimolar solutions (0.5 M) of ADP (≥90%, disodium salt, Sigma-Aldrich) and AcP were prepared and mixed. The pH was adjusted to 5.5 with aqueous HCl or NaOH (1M) and samples (500 µL) were taken at time-points (1, 2, 3, 4 and 5 hours) and immediately frozen at –80 °C. Experiments were carried out at 20 and 50 °C. These samples were analysed using the HPLC-UV protocol described in Section 2.1.10.

2.1.8 Nuclear magnetic resonance

For the ribose phosphorylation experiments ^{31}P-NMR was used to try to establish the phosphorylation sites of experimentally phosphorylated ribose, and adenosine diphosphate (ADP): ^{1}H-NMR was used to quantify of acetyl phosphate. Both ^{31}P-NMR and ^{1}H-NMR were carried out in a Bruker-Spectrospin 300 MHz equipped with an automated 60 sample changer and Topspin (2.1) software; a ^{1}H/^{13}C/^{31}P/^{19}F probe with z-gradients was routinely used. Sealed glass capillary tubes (SAMCO) with deuterated water (D_2O) (Sigma-Aldrich, ≥99.9% D atom) were introduced in each NMR tube. D_2O is used for locking purposes, which allows the software to accurately establish its ppm scale; this way, the software sets the frequency of the D_2O peak and then references all other experimentally found frequencies to that point.

For the ADP phosphorylation experiments ^{31}P-NMR (^{1}H-decoupled, Bruker Avance 400 MHz, 152 scans) was used; these experiments were performed by Silvana Pinna under my supervision. Peaks were identified by comparison with the chemical shift of pure standards, and by performing spike tests (addition of commercial standard directly into an already analysed sample in order to see if the putative peak grew). The quantification and statistical analysis were carried out by extrapolating the absolute intensity of the peaks of interest (in the case of ATP, the average absolute intensity of the three peaks formed by the β-phosphate, P_β) and using standard calibration curves.

2.1.9 Acetyl phosphate stability assay

^{1}H-NMR was used to monitor the degradation of AcP in water. The study was performed using the same concentration of AcP used in the ribose phosphorylation experiments and the same reaction volume (4 mL). Reaction conditions comprised pH 7, 9 and 11; 20 °C, 50 °C and 60 °C; and the presence/absence of Mg^{2+} ions and Ca^{2+} ions (20 mM Mg^{2+} and Ca^{2+} respectively). Samples were taken after 2, 30, 60, 120 and 300 minutes. Samples (500 µL) were immediately frozen at –80 °C. Samples were later thawed and analysed as described in Section 2.1.8 with the addition of 5.5 µL of a 1 mM potassium hydrogen phthalate (KHP) solution in water (≥99.95%, Sigma-Aldrich) as an internal standard. The usage of an internal standard allowed correcting for variations in machine sensitivity across different days of analysis. The internal standard was also employed for quantification purposes; analyte concentrations were calculated using the integrated ratio of protons in the internal standard compared to AcP. KHP was detected between 7.4 – 7.6 ppm and AcP at 2.1 ppm. To account for time taken from thawing to start of analysis (which took place in another building), 15 minutes were added to every sample time point.

2.1.10 High performance liquid chromatography (HPLC) – UV spectroscopy

HPLC-UV was used to study the phosphorylation of ADP by AcP. An equivalent experiment as in Section 2.1.7 was carried out, but this time with lower concentration of reagents to prevent chromatographic saturation, by mixing equimolar solutions (250 µM) of ADP and AcP at pH 5.5 at 50 °C. Samples (1 mL) were taken at time-points (1, 2, 3, 4 and 5 hours) and frozen at –80 °C. Samples were thawed and analysed using the following mobile phases A: 0.1 M ammonium acetate, B: 100% acetonitrile. Peaks were identified using pure standards and a spike test, and quantified by comparison with standard calibration curves. The analytical conditions used are summarised in Table 2.4.

Table 2.4. Operational conditions of the Dionex HPLC.

HPLC-UV	
Column	Acclaim 120 C18 2.2 µm, 2.1 x 250 mm
Column temperature	20 °C
Mobile phase	Initial: 100% (A)/0% (B) (up during 5 min) → 95% (A)/5% (B) (up during 6.3 min) → 75% (A)/25% (B) (for 8.7 min)
Column flow rate	0.1 mL/min
Injection volume	20 µL

2.2 Nucleotide polymerization

2.2.1 Sample desalting protocol

In order to reduce the adduct formation, which confounds the peak assignation and therefore the analysis of the results, a desalting protocol was applied to the samples before the mixing with the MALDI matrix. The protocol was adapted from Burcar et al. (2013) with advice from Kristin M. Coari (personal communication). The protocol consists of the following steps:

Solvents used:
ACN solution: 50% acetonitrile (MS grade, FLUKA) with HPLC water
TEAA solution: 0.1 M triethylammonium acetate (TEAA) in HPLC water (1 M in H_2O, FLUKA)

Procedure:
1. A volume of 4 µL of the ACN solution was deposited into a small Eppendorf microcentrifuge tube, closed and set aside.
2. With a Millipore C18 zip tip (Sigma), 10 µL of ACN solution were aspirated up and discarded. This was repeated twice more.
3. The previous three rinses were repeated with 10 µL of the TEAA solution.
4. To allow for the retention of the analyte by the zip tip matrix, 10 µL of sample were aspirated up and down eight times and then discarded.
5. A volume of 10 µL of HPLC water were aspirated up and dispensed.
6. A volume of 10 µL of the TEAA solution were aspirated up and dispensed.
7. Step 5 was repeated.

8. The 4 µL of the microcentrifuge tube with ACN from step 1 were slowly aspirated up and down three times. The analyte was in the final 4 µL.

2.2.2 MALDI-TOF analysis

Matrix-assisted laser desorption/ionisation (MALDI) – time of flight (TOF) mass spectrometry was used for the analysis of potential AMP polymers. A MALDI matrix consisting of 2,4,6-trihydroxyacetophenone monohydrate (THAP) plus ammonium citrate was selected on the basis of previous work showing THAP to be less prone to generating a dirty baseline when compared with other common MALDI matrices; also ammonium citrate reduces matrix-ion-sample adducts (Burcar et al., 2013) which simplifies the analysis. Other matrices such as 3-hydroxypicolinic acid were also tried but discarded as they showed much higher and dirtier baseline when compared to THAP, mostly in the range of 500 – 1000 m/z which is exactly where a strong sample signal (dimer and trimer molecular weights) was expected.

Two solutions were prepared in preparation of the MALDI matrix: 0.0372 g of 2,4,6-trihydroxyacetophenone monohydrate (Sigma-Aldrich, ≥99.5%) were added to 1 mL of acetonitrile (ACN); 0.045 g of ammonium citrate dibasic (Sigma-Aldrich, ≥98%) were added to 1 mL of HPLC water. Each solution was maintained at 4 °C for a maximum of a week. The required equal volumes of each solution were mixed to obtain the MALDI matrix. A volume of 1 µL of freshly prepared matrix solution were mixed with 1 µL of sample and deposited onto a clean steel MALDI-TOF plate. The liquid volume was allowed to evaporate for 30 min before the introduction of the steel plate into the MALDI-TOF machine.

The MALDI-TOF MS analysis was performed using a Waters micro MX mass spectrometer. The analytical conditions used are summarised in Table 2.5.

Table 2.5. Operational conditions of MALDI-TOF MS.

MALDI-TOF MS	
Mode	Reflectron and negative ion
Laser power	280 au
Pulse	2000 V
Detector	2500 V
Flight tube	12000 V
Reflector	5200 V
Negative anode	3738 V
Scan range	500-5000 amu

Before analysis, the mass spectrometer was calibrated using a low-molecular-weight oligonucleotide standard (comprising of a DNA tetramer, pentamer, 7-mer, 9-mer, and 11-mer (Bruker Daltonics)) as external calibrants. Each oligonucleotide standard was initially dissolved in 100 µL of HPLC water, divided in 6 µL aliquots and frozen at –80 °C. A fresh aliquot was used at each analytical calibration. The MALDI mass spectrum of the oligonucleotide standard is shown in Figure 2.3.

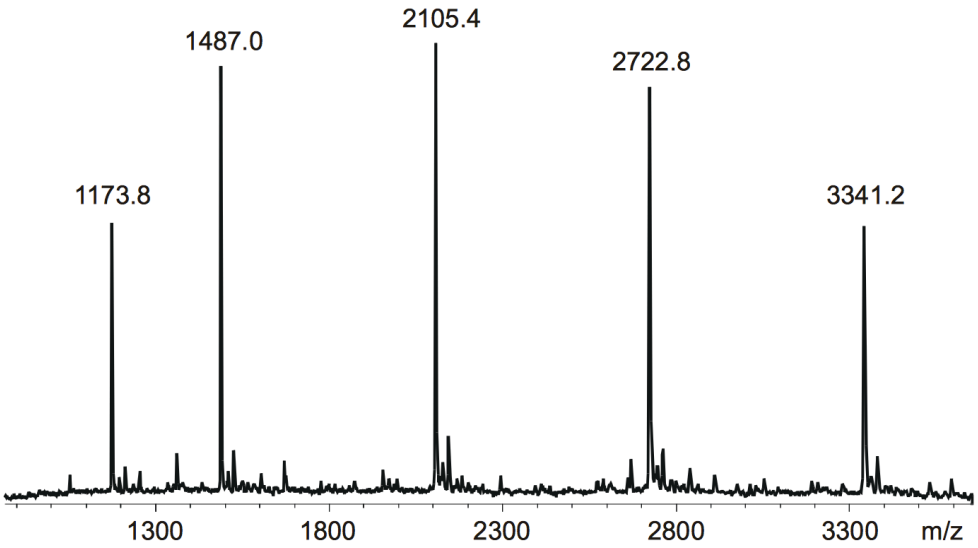


Figure 2.3. MALDI mass spectrum of the oligonucleotide calibration standard. The peak at 1487.0 m/z (pentamer) was used for the calibration of the sample spectra due to its higher intensity compared the other four under the operational conditions shown in Table

2.3 Glycine polymerisation

2.3.1 Glycine polymerisation experiments

Solutions of glycine (0.5 M) were prepared, pH adjusted to 7, 8, 9, 10, 11 or 12 using NaOH (1 M), and AcP (0.5 M; prepared as in Crans & Whitesides (1983), or using a commercial preparation) added to each pH condition. Samples were incubated at 20 °C for 7 days. Glycine (≥99%), N-acetylglycine (99%), diketopiperazine (≥99%) and diglycine (≥99%) were purchased from Sigma-Aldrich.

2.3.2 Nuclear magnetic resonance

Analysis was carried out by ^{1}H-NMR (Bruker Avance 300 MHz; 10% D_2O, 16 scans, water suppression). All peaks in NMR were identified by analysis of both pure standards and by spiked experimental samples. Percentage conversion of glycine to N-acetylglycine was calculated using the integrated ratio of protons by comparable integration of peak areas.

2.4 Formose reaction

2.4.1 Formose reaction experiments

Conditions for the formose reactions were adapted from Kopetzki & Antonietti (2011); no high pressure system was used. Calcium salts were used as a source of Ca^{2+} and to naturally adjust the initial reaction pH to 12.6-12.9; calcium ions have been shown to be crucial for a successful formose reaction at moderate temperature (<100 °C) (Kopetzki & Antonietti, 2011). Formaldehyde (Fischer, 16%), $Ca(OH)_2$ (Sigma-Aldrich, ≥95%) and $CaCO_3$ (Sigma-Aldrich, ≥99%) were diluted in HPLC water to achieve a final reaction volume of 2 mL. The initial reaction concentrations were of 0.5 M for formaldehyde, and 0.167 M for both calcium-bearing salts. Reactions were carried out at 60 °C in glass vials

in a metallic block heater. Samples were taken at 30 minutes, and 1, 2, 3 and 5 hours and immediately frozen. When applicable, commercial AcP was added to the formose reactions (400 mM) in order to study its effect.

For later GC-MS analysis (i), the sampling volumes were 99 µL plus immediate addition of 1 µL of myo-Inositol (Sigma-Aldrich, ≥99%) 0.025 M aqueous solution as internal standard (IS) to achieve a final concentration of IS of 250 µM in each sample. Samples were then immediately frozen at –80 °C. For later LC-MS or HPLC-UV analysis (ii) the sampling volumes were of 80 µL and were immediately frozen at –80 °C. All samples were thawed right before derivatisation. Sugar/sugar phosphates/alditol standards (glycerol, 99.5%; dihydroxyacetone, USP reference standard; D-ribose, 99%; D-arabinose, ≥98%, D-lyxose, 99%; D-xylose, ≥99%; D-fructose, ≥99%, 2-deoxy-D-ribose, 97%; D-erythrose 4-phosphate, ≥98%; D-glyceraldehyde 3-phosphate, ≥97%; D-glucose 6-phosphate, ≥98%, and D-ribulose, ≥97%) were purchased from Sigma-Aldrich. Glyoxal 40% solution was purchased from Sigma-Aldrich. Ethylene glycol was purchased from Fisher Chemical and was of general-purpose grade. Meso-erythritol was purchased from Alfa Aesar, ≥99%.

2.4.2 Sugar derivatisation for GC-MS

Carbohydrate analysis was achieved by transforming the sugars to their alditol acetates (Figure 2.4) (Blakeney et al., 1983). Sugars were first reduced using $NaBH_4$ (Fisher Chemical, general purpose grade) in dimethyl sulfoxide (DMSO) (Sigma-Aldrich, ≥99.6%) at 0.02 g/mL. 1 mL of this solution was added to 0.1 mL of samples/standards and incubated in 10 mL glass vials at 40 °C for 90 minutes. After the incubation, 0.1 mL of acetic acid (Sigma-Aldrich, ≥99%) was added to destroy excess $NaBH_4$. Then 0.2 mL N-methyl imidazole (Fisher Chemical, general purpose grade), which was used as a reaction catalyst was added followed by 2 mL acetic anhydride, which served as a source of acetyl groups. After 10 minutes, 5 mL of HPLC water was added to destroy excess acetic anhydride. Once cool, 1 mL of dichloromethane (DCM) (Sigma-Aldrich, ≥99.9%) was added; after vortexing the vial to separate the phases, the lower

phase was taken using a Pasteur pipette. Later, to improve sensitivity, this liquid-liquid extraction was repeated three times for each sample and the resulting ~3 mL evaporated down to a final volume of 0.5 mL. Afterwards, approximately 0.3 mL was transferred to a capped GC-MS vial.

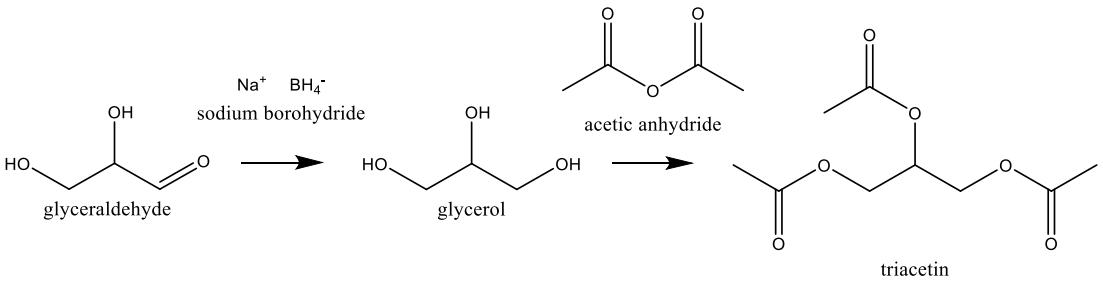


Figure 2.4. Derivatisation reaction of an aldehyde (glyceraldehyde) for GC-MS using reduction to its alditol (glycerol) and full acetylation to its alditol-acetate (triacetin).

2.4.3 Gas chromatography – mass spectrometry (GC-MS)

Once the samples had been derivatised according to the previous method, they were analysed using GC-MS, which is a good method for analysing volatile analytes such as the alditol-acetate derivatives of sugars. GC-MS analysis was carried out at the mass spectrometry laboratory at UCL's Christopher Ingold building. A Trace 1310 GC with ISQ quadrupole MS was used for the analysis and a TriPlus RSH autosampler was used for liquid injection of samples. Between samples, dichloromethane blanks were injected to prevent cross-contamination. Scanning began after 3 minutes, as before this time peaks correspond to derivatisation reagents, and reconstructed ion chromatograms were produced for identification of relevant sugar peaks. The general conditions used are summarised in Table 2.6.

Table 2.6. Operational conditions of the Trace GC-MS.

GC-MS	
Column	Rxi-5Sil MS (Thames Restek) Fused silica capillary column 15 m x 0.25 mm x 0.25 µm
Oven temperature	50 °C (for 1 min), then increase 30 °C /min to 280 °C (and hold 10 min)
Carrier gas	Helium
Injection temperature	220 °C
Injection volume	2 µL
Ion source temperature	280 °C
Transfer line temperature	280 °C
Injection mode	Splitless
Monitorised ions	Glyceraldehyde: 103, 145, 116 m/z Erythrose: 103, 115, 128 m/z Ribose: 103, 115, 145 m/z Myo-Inositol (internal standard): 115, 168, 199 m/z
Ion source	EI

2.4.4 High performance liquid chromatography (HPLC) – UV spectroscopy

HPLC-UV was used to quantify the sugars synthesised by the formose reaction, as well as to provide a peak assignation by comparison with commercial standards, which was later reinforced using LC-MS-MS (Section 2.4.5). The formose reaction samples were derivatised for liquid chromatography as explained in Section 2.1.3, and SPE-purified as explained in Section 2.1.4. Only one extra step was added to this sample preparation: once thawed, the samples were completely dried using N_2 gas (for 22 minutes) in order to eliminate all unreacted formaldehyde (which started at an abundance of 0.5 M); this was done in order to prevent the large quantity of formaldehyde from competing with the smaller amount of sugars for the derivatisation reagents, which would preclude an accurate quantification.

Once dried, the samples were rehydrated with HPLC water in order to achieve the same initial volume (80 µL). Once prepared, the samples were analysed using an Agilent 1260 Infinity II LC system with a UV detector. The analytical

method was adapted from Han et al. (2013); a series of columns and mobile phases were tested in order to maximise peak separation, since this is the primary kind of information that an HPLC-UV system provides. The chosen mobile phases were A: 0.1 M ammonium acetate in water (pH 4.5 titrated using HCl), and B: pure acetonitrile. The operational conditions are summarised in Table 2.7.

Table 2.7. Operational conditions of the Agilent 1260 Infinity II HPLC-UV.

HPLC-UV	
Column	Agilent InfinityLab Poroshell 120 EC-C18 column 4.6 x 150 mm, 4 µm particle size
Column temperature	40 °C
Flow rate	0.5 mL/min
Mobile phase	Isocratic 75% A during 50 minutes
Injection volume	10 µL
UV detector wavelength	254 nm

2.4.5 Liquid chromatography – mass spectrometry (LC-MS)

MS-MS assays were also performed on the formose reaction samples in order to provide an unequivocal peak assignation for each sugar. These analyses were performed using a Thermo HPLC Accela 600 pump and autosampler LC system connected to a Thermo Finnigan LTQ mass spectrometer. The used mobile phases were A: 0.05 M ammonium acetate in water (pH 4.5 titrated using formic acid), and B: pure acetonitrile. The analytical conditions used are summarised in Table 2.8.

Table 2.8. Operational conditions of the Thermo LC-MS (LTQ) for the MS-MS assays.

LC-MS-MS	
Column	Thermo Hypersil Gold C18 1.9 µm, 150 mm x 2.1 mm
Column temperature	40 °C
Flow rate	0.18 mL/min
Mobile phase	Isocratic 75% A during 50 minutes
Injection volume	10 µL
Ion source	ESI (+)
Normalised collision energy	35
MS-MS parent masses (m/z)	375.4 (glucose), 455.4 (glucose 6-phosphate), 345.2 (ribose), 425.2 (ribose 5-phosphate), 315.3 (erythrose), 395.1 (erythrose 4-phosphate), 285.1 (glyceraldehyde), 365 (glyceraldehyde 3-phosphate) 359.2 (fucose; used as external standard)

The MS-MS analyses were qualitative and were performed once per formose reaction experiment since they were time consuming and, due to the large number of scan instances (a total of 9), the resolution of the retention time values was poor (Figure 2.5c), so this could not be used for quantification purposes. Despite that, these tests reinforced the peak assignation for each sugar that was previously achieved using the HPLC-UV instrument. An example of the full scanned and MS-MS spectra of commercial standards versus experimentally synthesised sugars is shown in Figure 2.5a-b.

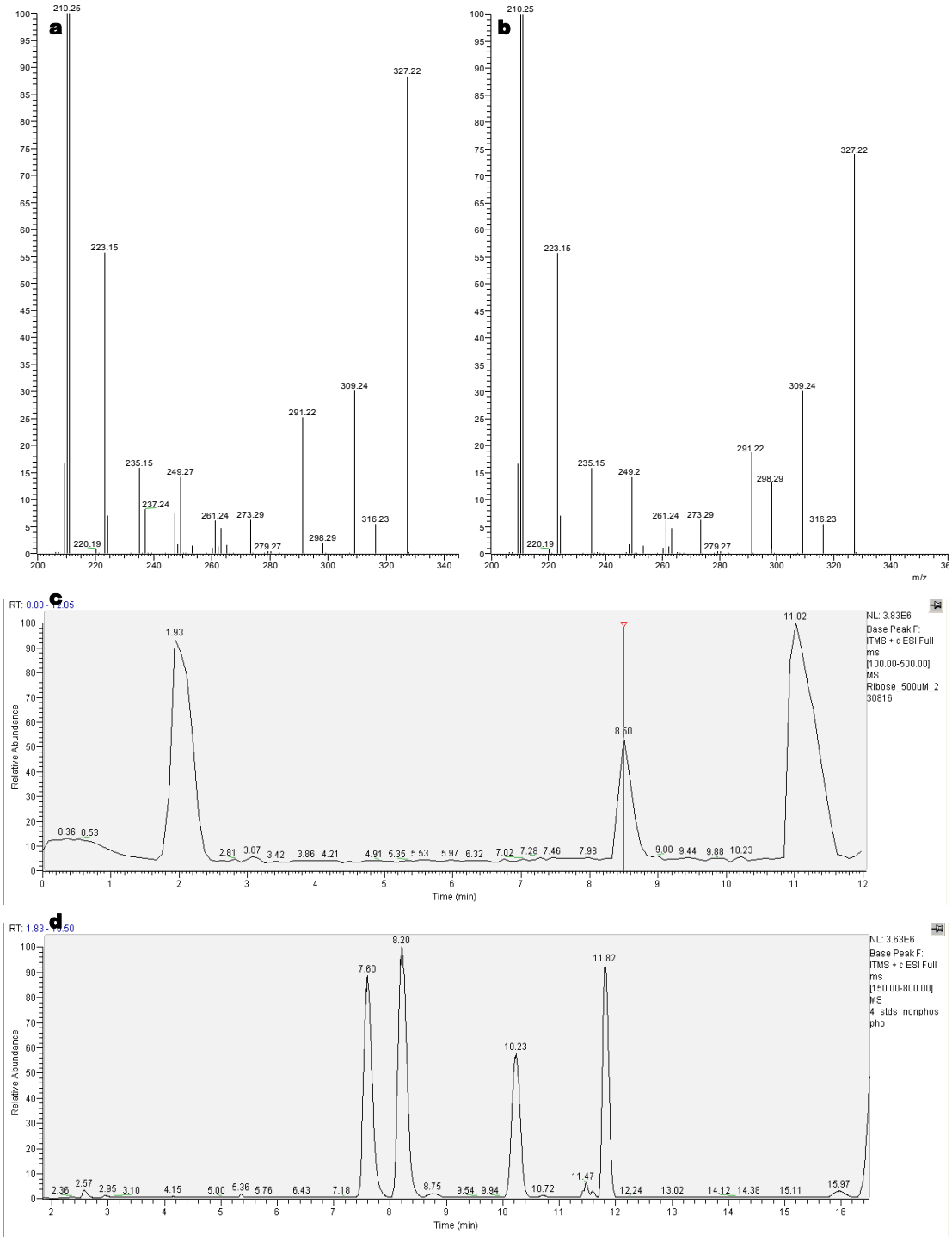
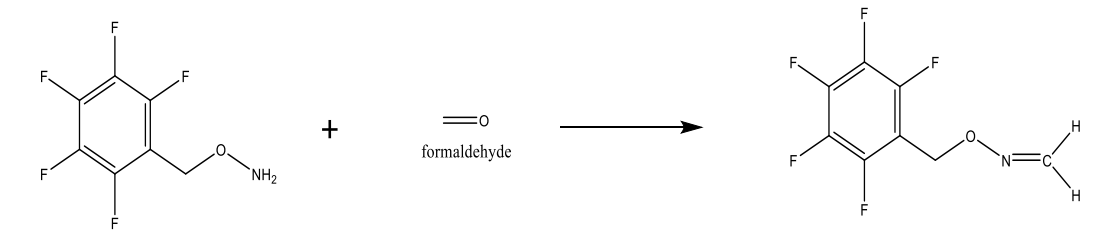


Figure 2.5. An example of the MS-MS data acquisition. (**a**-**b**) Comparison of the MS-MS spectra of a D-ribose commercial standard (**a**), and the peak corresponding to ribose synthesised by the formose reaction (**b**). LC-MS chromatograms corresponding to commercial sugar standards (**c**-**d**). Note the poor time-resolution of the chromatogram in (**c**) when scanning for 9 precursor ions. The resolution of (**d**) is much better as it was a general full scan (1 scanning instance). The method shown in (**c**) was used for peak identification in the formose reaction samples.

2.4.6 Formaldehyde quantification

A method for formaldehyde quantification was necessary in order to make sure that the formose reaction samples (for LC-MS and HPLC-UV analysis) were formaldehyde-free, as it would be derivatised (since it has a carbonyl group) when using AEC, and therefore its high concentrations would interfere with the quantification of sugars and sugar phosphates. Formaldehyde was derivatised with O-(2,3,4,5,6-pentafluorobenzyl) hydroxylamine hydrochloride (PFBOA) (Figure 2.6) (Sugaya et al., 2001). Formaldehyde is a small organic difficult to analyse as it normally escapes the chromatographic column as fast as the solvent and becomes masked by it; derivatisation with PFBOA adds molecular weight and solves this problem. Also, it improves the volatility of formaldehyde, which allows for the usage of headspace injection (injection of the gas phase of the sample vial). Headspace injection of the sample is normally preferable to liquid injection as the former only introduces volatile compounds to the machine, greatly improving sensitivity and minimizing irrelevant signals as well as allowing for a faster sample preparation making any purification steps during derivatisation unnecessary. A 0.1 mM standard stock solution of 4-bromofluorobenzene (Sigma Aldrich, ≥99%) in methanol was prepared for use as an internal standard for these analyses. Derivatised formaldehyde signal was always corrected to the internal standard signal to minimize random fluctuations affecting quantification. PFBOA (Sigma Aldrich, ≥98%) standard stock solution at 4 mM was prepared with HPLC water.



O-(2,3,4,5,6-Pentafluorobenzyl)hydroxylamine

formaldehyde *O*-((perfluorophenyl)methyl) oxime

Figure 2.6. Derivatisation reaction of formaldehyde using PFBOA, resulting in an oxime (right), which is volatile.

A volume of 1 µL of formaldehyde standard stock solution (when preparing calibration curves) or 1 µL of experimental sample was added to a PTFE-lined

septum and aluminium cap vial, followed by 1 mL of IS stock solution and 8,999 mL of HPLC grade water. A volume of 0.6 mL of PFBOA standard stock solution was added immediately followed by 3 g of sodium chloride (VWR, laboratory reagent grade). The vial was sealed and the samples were left for 4 h at 20 °C to allow for the derivatisation reaction. After the incubation time, the samples were analysed using GC-MS in headspace injection mode. The analytical conditions used are summarised in Table 2.9.

Table 2.9. Operational conditions of GC-MS and headspace autosampler.

Headspace autosampler	
Needle temperature	120 °C
Agitator temperature	60 °C
Agitator time	10 min
Injection depth	40 mm
GC-MS	
Column	Trace TR-5MS 260F047P (Thermo Scientific) Fused silica capillary column 30 m x 0.25 mm x 0.25 µm
Oven temperature	40 °C for 1 min, followed by an increase of 10 °C /min until reaching 220 °C; then hold at 220 °C for 20 min
Carrier gas	Helium
Injection temperature	200 °C
Sample draw	1 mL
Ion source	EI
Ion source temperature	240 °C
Transfer line temperature	220 °C
Injection mode	Split 10:1, 15 mL/min
Scan range	30-300 m/z

Chapter 3

Proto-metabolism driven by pH gradients and FeS minerals

The work described in this Chapter was published in Camprubi et al. (2017) and corresponds to theoretical predictions on the abiogenic role of Fe(Ni)S minerals at Hadean alkaline hydrothermal vents.

3.1 Introduction

Hadean alkaline hydrothermal vents present a number of interesting biogenic properties, which make them particularly suitable as hatcheries for life on Earth. Even though many of these characteristics are important (e.g. thermophoresis), both the presence of natural pH gradients and Fe(Ni)S minerals stand out, as these bear a great degree of congruence with life as we know it.

Hadean alkaline hydrothermal vents have long been postulated as the hatcheries of life and initially this was mostly due to the pH gradients naturally appearing there. Alkaline hydrothermal effluents are (and were) alkaline (pH 10-12), whilst Hadean oceans were mildly acidic (pH 5-6). As explained earlier, this proton gradient is congruent with life, which uses electrochemical gradients across membranes in order to conserve energy in the form of ATP. Proton gradients also drive CO_2 reduction via membrane Fe(Ni)S proteins such as the energy converting hydrogenase (see below). Interestingly, the approximate value of these natural gradients is remarkably reminiscent to that of the one found in modern mitochondria, which inspired modelling studies determining whether this could have been a successful source of energy at the origin of life (Sojo et al., 2016). Modern life uses intricate molecular machines (e.g. ATP synthase) in order to tap into the electrochemical gradient, but how could a

similar natural gradient be useful if ATP synthase, together with most other complex biological inventions such as a lipid membrane, was not present yet?

Iron (and nickel) sulphide minerals are also very closely related to life. It is calculated that as much as 50% of all known enzymes are metalloproteins (Thomson & Gray, 1998): they use inorganic metals as necessary cofactors. A prevalent group of these inorganic cofactors are Fe(Ni)S clusters. Iron-sulphur proteins are ancient and drive fundamental processes in cells, most notably electron transfer and CO_2 fixation (Fuchs, 2011; Ferry, 2010). It is reasonable then to postulate that similar iron-sulphur minerals could have played an important role in the origin of life because electron transfer (redox) and CO_2 fixation reactions were definitely some of the most important abiotic chemical reactions, which potentially led to the synthesis of organic molecules. But, how could simple minerals catalyse what nowadays is mediated using a series of complex and well-regulated enzymes? As discussed previously, Wächtershäuser (1988b; 1988a) proposed that abiotic proto-metabolism could have been powered through pyrites-pulling without the need of complex organic catalysts. Despite Wächtershäuser's novelty at aiming to explain the biological importance of Fe(Ni)S clusters, little evidence of this type of prebiotic reactivity has been experimentally provided.

Phylogenetic studies (Morowitz et al., 2000; Smith & Morowitz, 2004) suggest that the earliest cells drove carbon and energy metabolism via the acetyl CoA pathway, which is replete in Fe(Ni)S proteins and is the only known carbon fixation pathway found in both archaea and bacteria (Fuchs, 2011). This pathway works mechanistically different in bacteria and archaea, which makes it difficult to assess its ancestral (pre-enzymes) state (Russell & Martin, 2004). As will be discussed later, these differences could arguably be explained if proto-metabolism depended on the natural electrochemical gradients found in alkaline hydrothermal vents, and domain-specific adaptations were invented alongside the later differentiation of domain-specific membrane types. If this is true, it would be reasonable to envision that the acetyl CoA pathway diverged with the origins of active ion pumping (Sojo et al., 2016) (during the split between bacteria and archaea), and ancestral CO_2 fixation could have been very similar

to methanogens, which depend on a membrane-bound NiFe hydrogenase, the energy converting hydrogenase (Ech), which uses the proton-motive force to reduce ferredoxin, which in turn reduces CO_2. This mechanism suggests that pH modulates the reduction potential at the active site of the enzyme, making the unfavourable reduction of CO_2 by H_2 (when both reacting couples are at the same pH) feasible. It is then reasonable to postulate that under abiotic conditions, steep pH differences across semi-conducting Fe(Ni)S barriers could drive CO_2 reduction in a similar way.

It is also possible that not just CO_2 reduction could be achieved, but also a range of similar carbonylation and hydrogenation reactions could take place forming longer chain carboxylic acids such as pyruvate and oxaloacetate, mimicking the incomplete reverse Krebs cycle as found in methanogens (see Section 3.2.7). It is possible that the later closure of a complete reverse Krebs cycle, by regenerating acetyl CoA (or its prebiotic analogue) directly, could have displaced the acetyl CoA pathway in many modern groups (where this pathway is missing). A later reliance on acetyl CoA and ATP could have eliminated the need for the proton-motive force to drive most steps of the reverse Krebs cycle, yielding the modern version we observe today.

The work presented in this Chapter is theoretical and was carried out during several months when most analytical instruments were either out of service or could not be used due to, for example, the impracticability of switching daily between an LC-MS analytical setup for large protein analysis to one for small molecules. This work has been based on the analysis of reaction mechanisms as they occur in modern (or ancestral) enzymes, and their correlation with mineral-based catalysis of organic compounds. In line with de Duve's congruence principle, there has been a focus on proposing mechanisms that could have later developed into the modern forms by using the least amount of changes as possible.

3.2 Results and discussion

3.2.1 A primordial iron-sulphur world?

Two papers published in 1966 produced a remarkable shift of the paradigm on the research on the origin of life. Eck and Dayhoff (1966) reconstructed the evolutionary history of ferredoxin on the basis of its amino acid sequence. These authors proposed that ferredoxin, which contains very simple FeS cofactors, was formed through the doubling of a shorter protein, containing only the simplest eight amino acids, which in turn developed from a repeating sequence of only four amino acids (glycine, proline, aspartic acid and serine). The author's conclusion was that ferredoxin had been incorporated into metabolism even before complex proteins and the complete modern genetic code existed. Another article by Evans et al. (1966) showed that the Krebs cycle could operate in reverse, driving a ferredoxin-dependent carbon-reduction cycle in a photosynthetic bacterium. This pathway remained controversial for a long time, but was eventually accepted as an alternative to the Calvin-Benson cycle as a CO_2-fixing pathway. The reverse Krebs cycle is autocatalytic (from oxaloacetate, a single spin of the cycle generates two oxaloacetates). This discovery reinforced ideas on autocatalytic cycles and hypercycles as a means of driving stable growth at the origin of life, dating back to Eigen (1971), and foreshadowing the current widely accepted view on the biogenic importance of autocatalytic sets.

Almost 20 years later, Wächtershäuser proposed his often regarded as radical conception of a pyrites-pulled reverse Krebs cycle (Wächtershäuser, 1990). Wächtershäuser dismissed a heterotrophic origin of life in the form of a primordial soup and attempted to reconstruct the first metabolic cycles by retrodiction based on the tenets that (i) the reverse Krebs cycle was the ancestral state of the oxidative cycle, that (ii) it was common in bacteria and some archaea, and that (iii) it could be strictly chemoautotrophic (i.e. phototrophy was not required). Wächtershäuser was aware that, unlike the easier chemistry proceeding from electrical discharges or UV radiation acting on reduced gases such as methane and ammonia, the reduction of CO_2 by H_2 was not trivial (Wächtershäuser, 1988b). The reaction between CO_2 and H_2 is

overall exergonic under anoxic conditions, but the first steps to generate CO, $HCOO^-$ or CH_2O are markedly endergonic and inhibit the synthesis of organic molecules using this reaction scheme (Amend & McCollom, 2009). Autotrophic cells that grow from this reaction alone (methanogens and acetogens) must then use some form of energetic coupling to lower the kinetic barrier that would otherwise inhibit this form of metabolism. Wächtershäuser proposed that, before the advent of modern enzymes, the coupling mechanism was pyrites pulling, in which oxidation of FeS to pyrites (FeS_2) is coupled to the reduction of CO_2 by H_2:

$$FeS + H_2S \longrightarrow FeS_2 + H_2 \qquad \Delta G^{\circ\prime} = -38.4\ kJmol^{-1}$$
$$H_2 + CO_2 \longrightarrow HCOOH \qquad \Delta G^{\circ\prime} = +30.2\ kJmol^{-1}$$
$$FeS + H_2S + CO_2 \longrightarrow FeS_2 + HCOOH \qquad \Delta G^{\circ\prime} = -8.2\ kJmol^{-1}$$

Some studies have supported this mechanism, at least to form simple (but reactive) organics such as CH_3SH (methanethiol) (Heinen & Lauwers, 1996), but most successful studies testing this hypothesis, including those from Wächtershäuser, start with the most reactive substrate CO rather than CO_2 (Huber & Wächtershäuser, 1997; 1998), and therefore ironically avoid pyrites pulling. Parallel work on primordial soup was beginning to take account of geochemical evidence which suggested that the Earth's atmosphere was never rich in reduced gases such as CH_4 and NH_3, but was instead relatively oxidising, composed mainly of CO_2 and N_2 (Arndt & Nisbet, 2012; Trail et al., 2011; Zahnle et al., 2007). Unlike previous results, electrical discharges through such relatively oxidising atmospheres did not readily form amino acids; but inclusion of high concentrations of CO did indeed facilitate organic chemistry (Miyakawa et al., 2002). Accordingly, both Wächtershäuser and his opponents, who favoured a heterotrophic origin of life, called on the reactivity of CO. The synthesis of 'activated acetate' (methyl thioacetate) by Huber and Wächtershäuser in 1996 was achieved under an atmosphere of 1 bar of CO, which is geologically highly unlikely, and occurred readily in the absence of FeS; $NiSO_4$ was also effective as a catalyst for this reaction, despite being incapable of pyrites pulling.

3.2.2 Problems with primordial metabolic cycles

Most discussion of prebiotic chemistry has historically been framed in terms of specificity and yield. Most synthetic chemists working on the origins of life tend to see the problem in terms of achieving high yields of specific products, and distrust chemistry that produces low yields of mixed products (Patel et al., 2015). Biologists tend to be more inclined to the opposite view, arguing instead that selection will favour higher yields of specific products over time; hence the earliest prebiotic chemistry should have produced low yields of mixed products, otherwise natural selection would have no space to operate. Particular catalysts such as metal ions or minerals favour certain reactivity landscapes over others, and increase yields somewhat when compared to a non-catalysed reaction, hinting at the beginnings of geochemical proto-metabolic pathways (Copley et al., 2007). Chelation by simple organics such as amino acids, and later short non-coded oligopeptides, and ultimately genetically encoded proteins, would each favour increasing yields of more specific products (Copley et al., 2007; de Duve, 2005). As a rule of thumb, cells (using enzymes) lower kinetic barriers to thermodynamically favoured reactions. This applies equally to the tardy reaction between H_2 and CO_2; methanogens and acetogens that depend on the acetyl CoA pathway gain all the carbon and energy they need to grow from this reaction alone (Ferry & House, 2006; Maden, 2000; Shock et al., 1998). Because this reaction barely occurs as abiotic chemistry, the reactants remain naturally far from equilibrium, an opportunity waiting to be exploited. Most prebiotic reactions performed throughout this thesis display low yields and specificity. In essence, we propose that this is not a problem prebiotic chemists need to work around, but rather it is what should be sought after, since otherwise natural selection would have had no ‘room’ for selecting for catalytic improvements over time.

From this point of view, the fact that Wächtershäuser’s hypothetical pyrites-pulled Krebs cycle has never been shown to work is not unexpected as, amongst other considerations, a theoretical energetic coupling is technically not catalysis, unless a catalyst (in almost a mechanical way) couples both reactions. But there is a deeper problem with the conception of primordial

metabolic cycles, as pointed out by Orgel (2008), which applies specifically to the idea of cycles rather than linear pathways. The problem is that if abiotic reactions occur at low yield, and the product of one reaction acts as the substrate of the next reaction, then the more reaction steps there are, the more catastrophic the decline in yield. That much is true for any metabolic pathway, and must presumably favour the shortest pathways of carbon fixation. But it becomes far more of a serious problem for metabolic cycles, in part because they usually comprise many steps (Wächtershäuser's pyrite-pulled reverse Krebs cycle has 16), but especially because the substrate concentration for the first step depends (basically in its totality) on the yield of the final step. This is surely unworkable as abiotic chemistry; Orgel eloquently named it as 'if pigs could fly' hypothetical chemistry (Orgel, 2008).

This last issue, incidentally, is not solved by a recent study suggesting that Krebs cycle intermediates are favoured by sulphate radical chemistry (Keller et al., 2017). This study demonstrated a decomposition of long chain Krebs cycle intermediates yielding simpler molecules such as pyruvate, but not a proto-metabolic step-wise turning of the cycle. The article is nevertheless interesting in that it shows that Krebs cycle intermediates are thermodynamically metastable (not unlike many other small organic molecules such as amino acids), but does not solve the problem of declining yield at each steps of a full prebiotic cycle. An interesting point of this study is that it shows that radical chemistry (that does not involve UV radiation) could be a source of prebiotic chemical reactivity. The outcome of this study broadens the possibility of exploring prebiotic radical chemistry in a way more akin to life; for instance using FeS clusters as those involved in the reactivity of pyruvate synthase (also known as pyruvate:ferredoxin oxidoreductase (PFOR)), which is a key enzyme linking the acetyl CoA pathway to the reductive Krebs cycle catalysing the formation of pyruvate from acetyl CoA in anaerobic archaea and bacteria (Charon et al.,1999).

3.2.3 The acetyl CoA pathway is ancient

The complete reverse Krebs cycle is found only in bacteria so, despite the appeal of FeS catalysis in CO_2 fixation, its deep antiquity is not sustained by phylogenetics (unless it was lost in all archaea) (Campbell & Cary, 2004; Fuchs, 2011; Martin & Russell, 2003; Weiss et al., 2016). There are six pathways of CO_2 fixation known to date (Fuchs, 2011). Of these, only the acetyl CoA pathway (or Wood/Ljungdahl pathway) is found in both bacteria and archaea (Campbell & Cary, 2004; Fuchs, 2011; Martin & Russell, 2003; Weiss et al., 2016); so it likely dates back to their common ancestor, the last universal common ancestor of all life (LUCA), which arguably might have lived in submarine hydrothermal vents (Martin et al., 2014).

There are other factors that are consistent with the early evolution of the acetyl CoA pathway. It is short and linear, which avoids the problems discussed above (Russell & Martin, 2004). Unlike other pathways of CO_2 fixation, it is exergonic overall, so there is no net input of ATP needed in order to drive the pathway; on the contrary, the reaction of H_2 with CO_2 yields of the energy required for cell growth (Russell & Martin, 2004). As mentioned earlier, this pathway is also replete in Fe(Ni)S cofactors, to the extent that the pathway has been said to have ‘rocky roots’ (Russell & Martin, 2004). The groups of microorganisms that rely on the acetyl CoA pathway for carbon fixation (such as methanogens, acetogens and sulphate-reducers) have notably more FeS-bearing proteins in their genomes than other prokaryotes (Sousa et al., 2013). The key enzyme of this pathway, the carbon monoxide dehydrogenase/acetyl CoA synthase (CODH/ACS), is an unusual hydrogenase with the Fe_4S_4 clusters, two of which bridge via sulphide to nickel (Ni-S-Fe_4S_4) (Ragsdale & Kumar, 1996). This enzyme catalyses the synthesis of acetyl CoA, the hub of metabolism, directly from CO bound to an iron atom, and a methyl ($–CH_3$) group bound to a nickel one (Ragsdale & Kumar, 1996). Similar abiotic reactions have been proposed to take place on Fe(Ni)S surfaces, via Fischer-Tropsch-type hydrogenations and Koch-type carbonylations (Cody et al., 2004; Cody, 2004), as shown in Figure 3.1. This reaction mechanism also depicts the first difficult step, the

reduction of CO_2 to CO on an iron-sulphide surface (Dzade et al., 2015), which is discussed later in the relation to pH modulation of reduction potential.

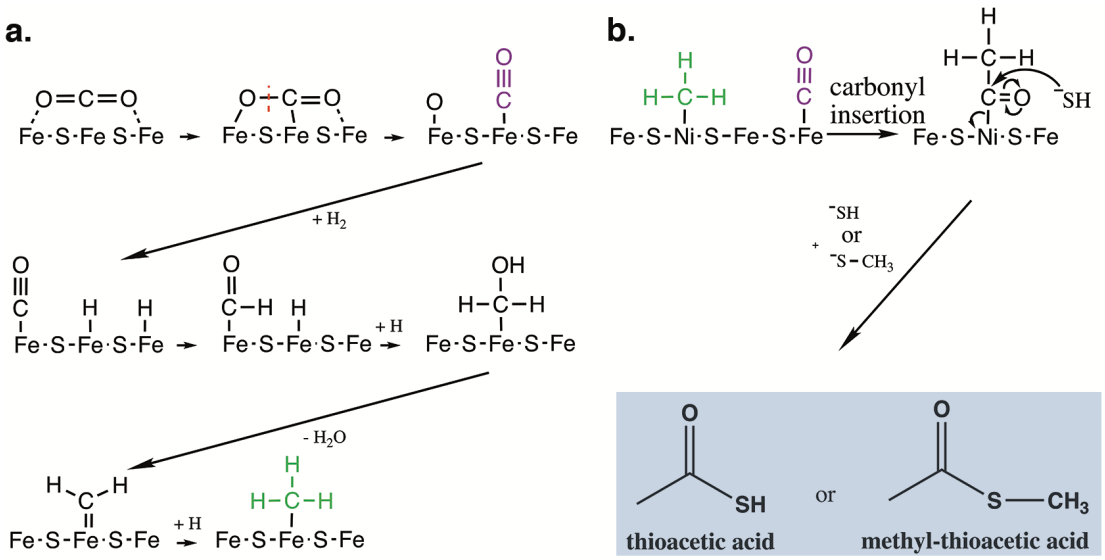


Figure 3.1. Primordial CO_2 reduction on the surface of a Fe(Ni)S mineral (Camprubi et al., 2017). (**a**) Fe-catalysed reduction of CO_2 into surface-bound CO (purple) via the cleavage of a C-O bond, followed by subsequent hydrogenations and a dehydration yielding surface-bound methyl group (green). (**b**) Carbonylation of a Ni-bound methyl group by Fe-bound CO, followed by elution through nucleophilic attack by a sulphydryl or methyl-sulphydryl ion yielding thioacetic acid or methyl-thioacetic acid, respectively. This abiotic mechanism is analogous to the carbon monoxide dehydrogenase/acetyl CoA synthase (CODH/ACS) enzymatic mechanism, where coenzyme A acts as the eluting nucleophile.

Others (Varma et al., 2017) have achieved similar reactions yielding acetate (and pyruvate) using metals (Fe^0, Ni^0 and Co^0) as catalysts. But, even though this is impressive chemistry and shows the feasibility of the non-enzymatic reduction of CO_2, the choice of metals are of limited geological plausibility, mainly due to their highly reduced state.

3.2.4 Deep differences between bacteria and archaea

While the previous factors are consistent with the primordial origins of the acetyl CoA pathway, there are deep differences in the methyl synthesis between bacteria and archaea, which make it difficult to discern the evolution of the pathway (Nitschke & Russell, 2013; Sojo et al., 2016; Sousa & Martin, 2014).

One possibility is that reactive methyl groups (perhaps in the form of CH_3SH, methanethiol) were so plentiful in Hadean hydrothermal systems that there was no requirement for genetically encoded methyl synthesis in the earliest stages of evolution (Sousa & Martin, 2014). If that were the case, then the true main substrate for the origin of life was not CO_2, as in most autotrophs, but CH_3SH. To date, little abiotic CH_3SH has been detected in modern hydrothermal systems (Reeves et al., 2014), but it could be consumed by cells lower in the crust, or rapidly oxidised under modern aerobic conditions. If so, the concentration of CH_3SH could have been far greater in anoxic Hadean vents.

An alternative possibility is that H_2 reduced CO_2 to both CO and to $–CH_3$ at the vent-ocean interface, through a process similar to that which occurs in methanogens (Sojo et al., 2016). The logic here relates to the process of flavin-based electron bifurcation, which is similar in concept, but differs in many details, between methanogens and acetogens (archaea and bacteria, respectively) (Buckel & Thauer, 2013; Herrmann et al., 2008; Kaster et al., 2011). Electron bifurcation accomplishes the endergonic reduction of a low-potential ferredoxin by coupling it to the exergonic reduction of a heterodisulphide (in methanogens) (Kaster et al., 2011) or NAD^+ (in acetogens) (Herrmann et al., 2008). The specific steps of electron bifurcation are discussed in detail elsewhere (Buckel & Thauer, 2013; Herrmann et al., 2008; Kaster et al., 2011; Nitschke & Russell, 2013; Sojo et al., 2016) and it is not the focus of this Chapter, so it will not be discussed here. The main point is that, in both groups, electron bifurcation achieves two outcomes, albeit via distinct pathways: (i) the difficult reduction of CO_2 by H_2 to form reactive methyl groups and (ii) the generation of an electrochemical ion gradient via a membrane-bound pump, Mtr in the case of methanogens, and Rnf or energy converting hydrogenase (Ech) in the case of acetogens (Buckel & Thauer, 2013) (Figure 3.2).

Critically, for electron bifurcation to continue operating, the exergonic reduction products (thiols or NADH) must be reoxidised, and this is achieved by the excretion of either methane (in methanogens) or acetate (in acetogens) (Buckel & Thauer, 2013). So electron bifurcation generates an electrochemical

membrane potential (which is then used to drive both ATP synthesis and CO_2 fixation, as discussed below), but it does not directly generate biomass (Figure 3.2a-b, left panels). This is a key point, which is worth reiterating. While electron bifurcation achieves the difficult reduction of ferredoxin, all that is actually conserved is an electrochemical ion gradient across a membrane. In other words, electron bifurcation drives active ion pumping (Sojo et al., 2016).

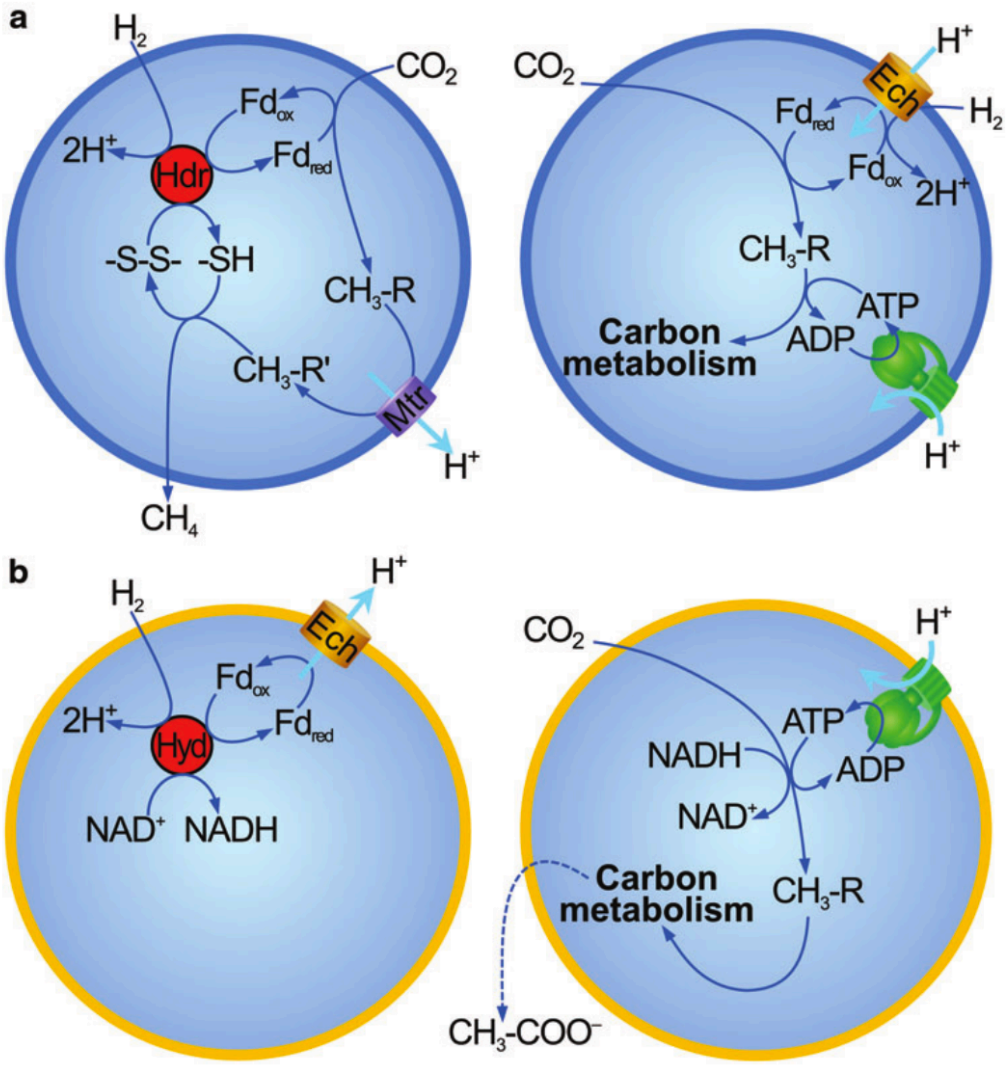


Figure 3.2. Schematic depiction of methanogenesis and acetogenesis (Sojo et al., 2016). (**a**) Methanogenesis. Left panel: Flavin-based electron bifurcation in methanogenesis generates reduced ferredoxin and reduced thiols. Ferredoxin is used to reduce CO_2 to a methyl group; its transfer from one cofactor to another releases enough energy to generate an ion gradient (depicted here as a H^+ gradient for simplicity) via the membrane protein Mtr. Methane is excreted as waste. Right panel: The ion gradient is used to drive ferredoxin reduction directly to form a methyl group, which is reacted with CO to form acetyl CoA (not shown for simplicity); ATP is synthesized via a standard rotor-stator ATP synthase. Acetyl CoA and ATP together drive intermediary metabolism. (**b**) Acetogenesis. Left panel: Electron bifurcation in acetogens generates reduced ferredoxin and NADH. Oxidation of ferredoxin drives the extrusion of ions (typically Na^+ but shown here as H^+ for consistency) via the membrane protein Ech or Rnf. Right panel: CO_2 is reduced using NADH and ATP to form a methyl group. This is reacted with CO (not shown) to form acetyl CoA, which together with ATP is used to drive carbon metabolism. Acetate is excreted to maintain redox balance.

3.2.5 Alkaline hydrothermal vents

The fact that electron bifurcation generates electrochemical ion gradients allows us to structure the problem of its origin in a specific environment: alkaline hydrothermal vents. As explained in the Introduction to the thesis, these vents have many properties that make them conducive to the origins of life, which have been developed in detail by Russell and colleagues over two decades (Martin & Russell, 2007; Martin et al., 2008; Russell & Hall, 1997; Russell et al., 2014; Russell et al., 1994); due to this, the details will not be discussed here. Suffice to say that, in anoxic Hadean oceans, alkaline vents should have formed labyrinths of micropores with thin, catalytic walls containing Fe(Ni)S minerals, through which hydrothermal fluids and ocean waters percolated. Like modern vents, the hydrothermal fluids should have been warm (70-90 °C), rich in H_2 (15–200 mM) and strongly alkaline (about pH 11), whereas in the Hadean the ocean waters were probably mildly acidic (pH 6) and saturated in CO_2. Within vents, laminar flow of hydrothermal fluids and ocean waters through interconnected micropores should have generated proton gradients across the thin semi-conducting Fe(Ni)S walls, with CO_2 in acidic ocean waters in pores off from the main hydrothermal flow, and H_2 in alkaline hydrothermal fluids in more actively venting regions. In summary: these vents should have provided everything needed to drive the acetyl CoA pathway: H_2, CO_2 and natural proton gradients (up to 5 pH units) across thin, semi-conducting barriers containing Fe(Ni)S minerals, with structures similar to the FeS clusters in ferredoxin of Ech. There would have been no need to actively pump ions via electron bifurcation if hydrothermal flow provided equivalent proton gradients across natural Fe(Ni)S barriers (Lane & Martin, 2012; Sojo et al., 2016; Sojo et al., 2014); no need to actively generate proton gradients if the vents provided them for free.

Both methanogens and acetogens use electrochemical ion gradients to grow. Acetogens use them to generate ATP via the ATP synthase, and then use ATP and NADH to drive CO_2 fixation (Buckel & Thauer, 2013; Herrmann et al., 2008) (Figure 3.2b, right panel). Since the ATP synthase is a sophisticated protein, a rotating nanomotor, it is unlikely to have been primordial (Martin & Thauer,

2017). In contrast, methanogens stand out: they use the proton-motive membrane-bound NiFe hydrogenase Ech in order to drive the reduction of ferredoxin directly (Kaster et al., 2011) (Figure 3.2a, right panel). This is then used to reduce CO_2, to form a methyl group bound to a cofactor, as well as CO, which then react together to form acetyl CoA on the Ni-S-Fe_4S_4 clusters of CODH/ACS (Ragsdale & Kumar, 1996). The entire path of CO_2 fixation in methanogens is driven by the proton gradient (likely converted from a Na^+ gradient, generated by Mtr, via an obligate Na^+/H^+ antiporter (Šurín et al., 2007)) and does not require ATP (Buckel & Thauer, 2013; Kaster et al., 2011). While the proteins involved today are moderately complex, all the critical electron transfers are achieved by the Fe(Ni)S cofactors, with structures similar to Fe(Ni)S minerals found in hydrothermal systems (Russell & Martin, 2004). In principle, the fact that alkaline vents contain H_2, CO_2 and natural proton gradients across thin semi-conducting Fe(Ni)S barriers means that they could in principle drive prebiotic CO_2 fixation via a mechanism analogous to methanogens (Sojo et al., 2016). If so, that was the ancestral pathway of carbon fixation, and the detailed differences in electron bifurcation and methyl synthesis between methanogens and acetogens could be ascribed to the independent evolution of active pumping in the two groups, as proposed in detail elsewhere (Lane & Martin, 2012; Sojo et al., 2016; Sojo et al., 2014).

3.2.6 How proton gradients could drive CO_2 fixation

Why does the reduction of ferredoxin via Ech depend on the proton-motive force? The answer is as yet unknown, but cannot relate to reverse electron flow (as originally proposed (Hedderich, 2004)) as these methanogens do not possess an electron-transport chain that could operate in the opposite direction (Buckel & Thauer, 2013; Kaster et al., 2011). A more likely possibility is that pH modulates reduction potential at the active site of the enzyme. The flux of protons through Ech from the relatively acidic exterior could lower the pH at the active site of the enzyme, which should facilitate reductions that depend on protons, including CO_2 as well as some ferredoxins (Chen et al., 2002). As Wächtershäuser pointed out in his proposal for pyrites pulling, H_2 is not

sufficiently electropositive to reduce CO_2 to CO, formate or formaldehyde (Wächtershäuser, 1988b, 1990), even though the later steps to methanol and methane are strongly exergonic (Amend & McCollom, 2009). This is the reason that electron bifurcation works: the endergonic initial steps are coupled to the exergonic final steps, so that the final steps pull through the first steps (Buckel & Thauer, 2013). But that is not the mechanism of Ech (Buckel & Thauer, 2013); something else must be going on.

As introduced in Chapter 1, we propose that pH modulates the reduction potential of H_2 and CO_2 (Herschy et al., 2014; Sojo et al., 2016). When protons are involved in a reduction (e.g. balancing charges), the Nernst equation shows that the reduction potential falls by approximately –59 mV per pH unit (Nicholls & Ferguson, 2013). At pH 0, the reduction potential of H_2 is 0 (as defined by the hydrogen electrode). At pH 7, it is –414 mV; and at pH 11, it falls to –650 mV, strongly reducing. The reason is simple: if H_2 gives up its electrons under alkaline conditions, the remaining H^+ ions will react swiftly with OH^- to form water, thermodynamically a highly favoured reaction under these circumstances, as there is an overwhelming excess of hydroxide ions ready to react. Conversely, CO_2 is more easily reduced under acidic conditions. At pH 7, the mid-point reduction potential of the CO_2/CH_2O couple is about –580 mV (Herschy et al., 2014; Sojo et al., 2016), but at pH 6 it rises to –520 mV. Again, the reason is simple: CO_2 will pick up the four electrons (in order to form CH_2O) more easily if protons are readily available (at low pH) to balance the negative charges resulting from the direct reduction. Much the same applies to some FeS minerals such as mackinawite, which can be protonated under mild acidic conditions (below pH 7.5) but deprotonate under alkaline conditions (Wolthers et al., 2005); and at least some FeS clusters, in certain ferredoxins, seem to follow a similar behaviour (Chen et al., 2002). If so, ferredoxin is easier to reduce under mildly acidic conditions, and in the context of the active site of Ech it should be readily reduced in the vicinity of proton channels that could lower the pH locally. Conversely, in the more alkaline interior of the cell, ferredoxin should deprotonate and become more reducing (more able to reduce CO_2 to CO and beyond). It is worth noting that some literature points to ferredoxin binding to the other side of Ech from the H^+ pore, so it seems

implausible that it could be protonated; therefore, it is probable that another of the FeS clusters in Ech is (Hedderich & Forzi, 2006; Sapra et al., 2003).

These principles could have significant connotations for electrochemistry at the origin of life. While the discussion above relates to the active site of Ech, the same principles should hold in an abiotic context, given a structure that is capable of separating two phases of different pH. This is precisely what alkaline vents provide. While H_2 cannot reduce CO_2 to CH_2O at pH 7 (because the reduction potentials of the H_2/H^+ and CO_2/CH_2O couples are –414 mV and –580 mV respectively), in alkaline vents H_2 is dissolved in alkaline solution at pH 11-11.5, giving it a reduction potential as low as –680 mV, whereas CO_2 is dissolved in ocean waters at a pH 5.5-6, conferring it a reduction potential of about –500 mV. As long as the two pH phases are separate (which is aided by laminar flow in vents (Möller et al., 2017)), and as long as they are separated by a semi-conducting barrier capable of transferring electrons from one pH phase to the other, then H_2 in the alkaline phase should reduce CO_2 in the acidic phase via transfer of electrons across the semi-conducting barrier. Not only are Fe(Ni)S minerals semi-conducting (Nakamura et al., 2010) but they can also be protonated under mildly acidic conditions (Wolthers et al., 2005), making the acid-facing surfaces more electronegative, hence drawing electrons across the barrier. While such a process would necessarily generate an electrical charge on the barrier (Barge et al., 2015), this should be dissipated by the physical mixing of the hydrothermal fluids and ocean waters elsewhere in the vent. This phenomenon could also be somewhat counterproductive since the synthesis of H_2 on the acidic side would be promoted (from H^+ from the acidic side, and e^- from the alkaline), and this could then compete with the synthesis of organics. In sum: natural pH gradients across thin, semi-conducting catalytic Fe(Ni)S barriers should facilitate the reduction of CO_2 by H_2 to form CO and more highly reduced organics such as $–CH_3$ groups, as depicted in Figure 3.1, through a mechanism analogous to Ech in methanogens.

3.2.7 Analogous reductions in the incomplete reverse Krebs cycle

The reduction of CO_2 to reactive thioesters such as acetyl CoA, or abiotic equivalents such as thioacetate or methyl thioacetate is an important first step, but is only the first step of proto-metabolism. Amino acids and nucleotides are synthesised from carboxylic acids with 3-6 carbons (which are Krebs cycle intermediates) and producing these requires a series of further carbonylation and hydrogenation reactions. Considering this in terms of prebiotic chemistry, Martin and Russell have previously suggested that the acetyl CoA pathway could feed into the reverse incomplete Krebs cycle, as in methanogens (Martin & Russell, 2007).

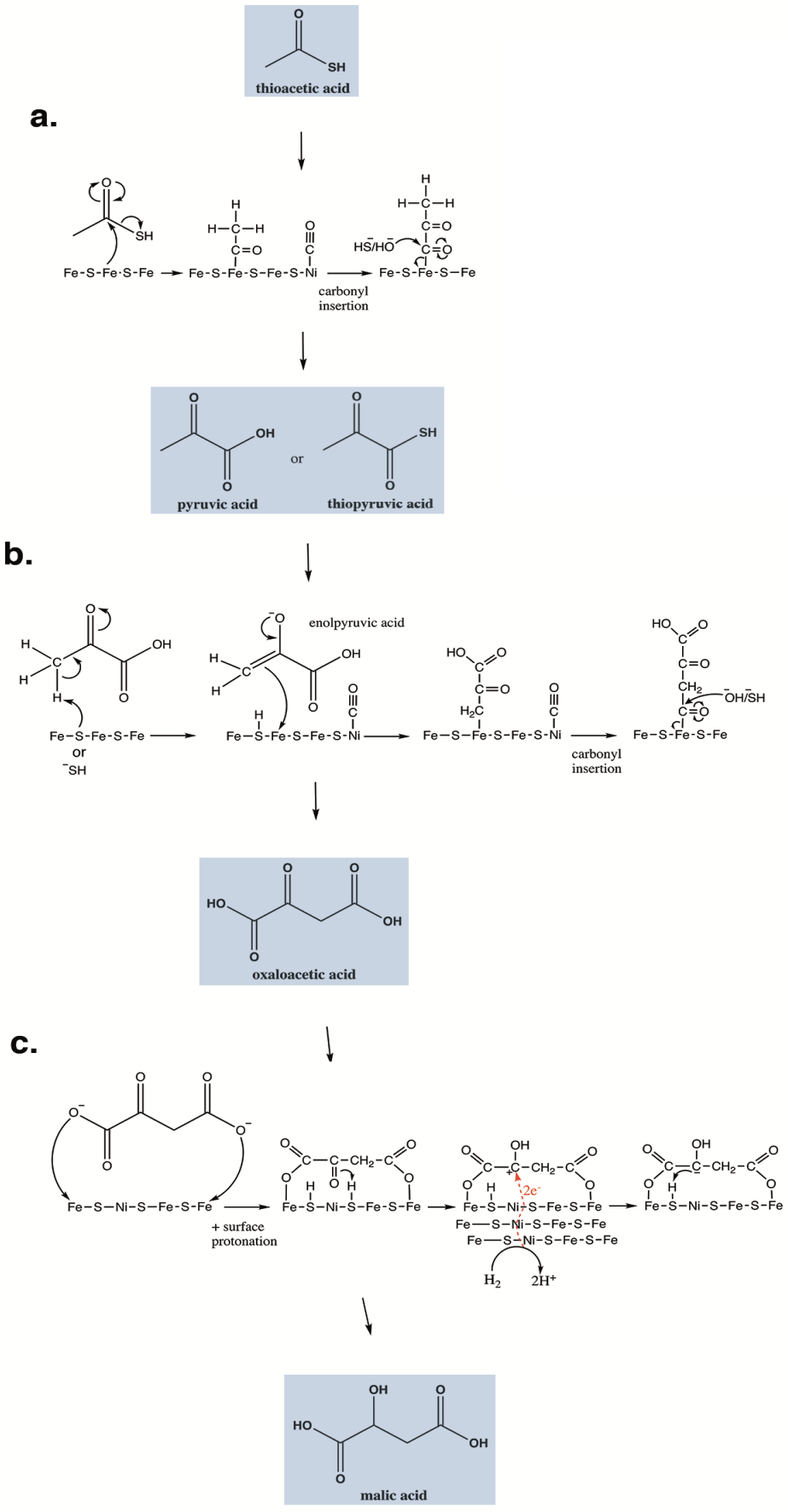

thioacetic acid
a.
Fe-S-Fe-S-Fe
Fe-S-Fe-S-Fe-S-Ni
carbonyl insertion
Fe-S-Fe-S-Fe
pyruvic acid
or
thiopyruvic acid
b.
enolpyruvic acid
Fe-S-Fe-S-Fe
or
SH
Fe-S-Fe-S-Fe-S-Ni
Fe-S-Fe-S-Fe-S-Ni
carbonyl insertion
Fe-S-Fe-S-Fe
oxaloacetic acid
c.
Fe-S-Ni-S-Fe-S-Fe
+ surface protonation
Fe-S-Ni-S-Fe-S-Fe
2e-
Fe-S-Ni-S-Fe-S-Fe
Fe-S-Ni-S-Fe-S-Fe
Fe-S-Ni-S-Fe-S-Fe
H2
2H+
Fe-S-Ni-S-Fe-S-Fe
malic acid

Figure 3.3. Primordial incomplete reverse Krebs cycle facilitated by pH gradients across Fe(Ni)S semipermeable barriers (Camprubi et al., 2017). (**a**) Carbonylation of a surface-bound acetyl group (from thioacetic acid), and elution from the mineral surface by a hydroxide or sulphydryl ion yielding pyruvic acid or thiopyruvic acid, respectively. To simplify, the non-thiolated version of the products is shown for the remaining reactions. (**b**) Adsorption of pyruvic acid through its enolate form onto the mineral surface followed by carbonylation and elution to form oxaloacetic acid. (**c**) Reversible adsorption of oxaloacetate followed by hydrogenation of its keto group by two electrons (from H_2 oxidation by Ni^+ in the alkaline phase) and H^+ ions from surface mineral protonation yielding malic acid. Nickel atoms channel electrons from H_2 catalysing a two-electron

The incomplete reverse Krebs cycle has two advantages over the complete version. First, given high H_2 concentrations and natural proton gradients transecting semi-conducing Fe(Ni)S barriers, the short linear acetyl CoA pathway should provide a continuous supply of reactive thioesters (far more than would be attained through a complete turn of the cycle, with its calamitous decline in yield at each step). Second, we might expect that the high concentration of thioesters could drive at least the next couple of steps, to pyruvate or oxaloacetate (Figure 3.3), both of which are important precursors for amino acid and nucleotide biosynthesis. These in turn are likely to chelate Fe(Ni)S minerals, enhancing their catalytic properties, by mimicking the active site of the enzyme and increasing the catalytic surface area (Dzade et al., 2016; Wang et al., 2015; Yamaguchi et al., 2014), driving faster CO_2 reduction; a valuable positive feedback (West et al., 2017). As catalytic properties improve, the yield of individual steps should increase, ultimately extending the pathway (still linear) through to C6 tricarboxylic acids, such as isocitrate.

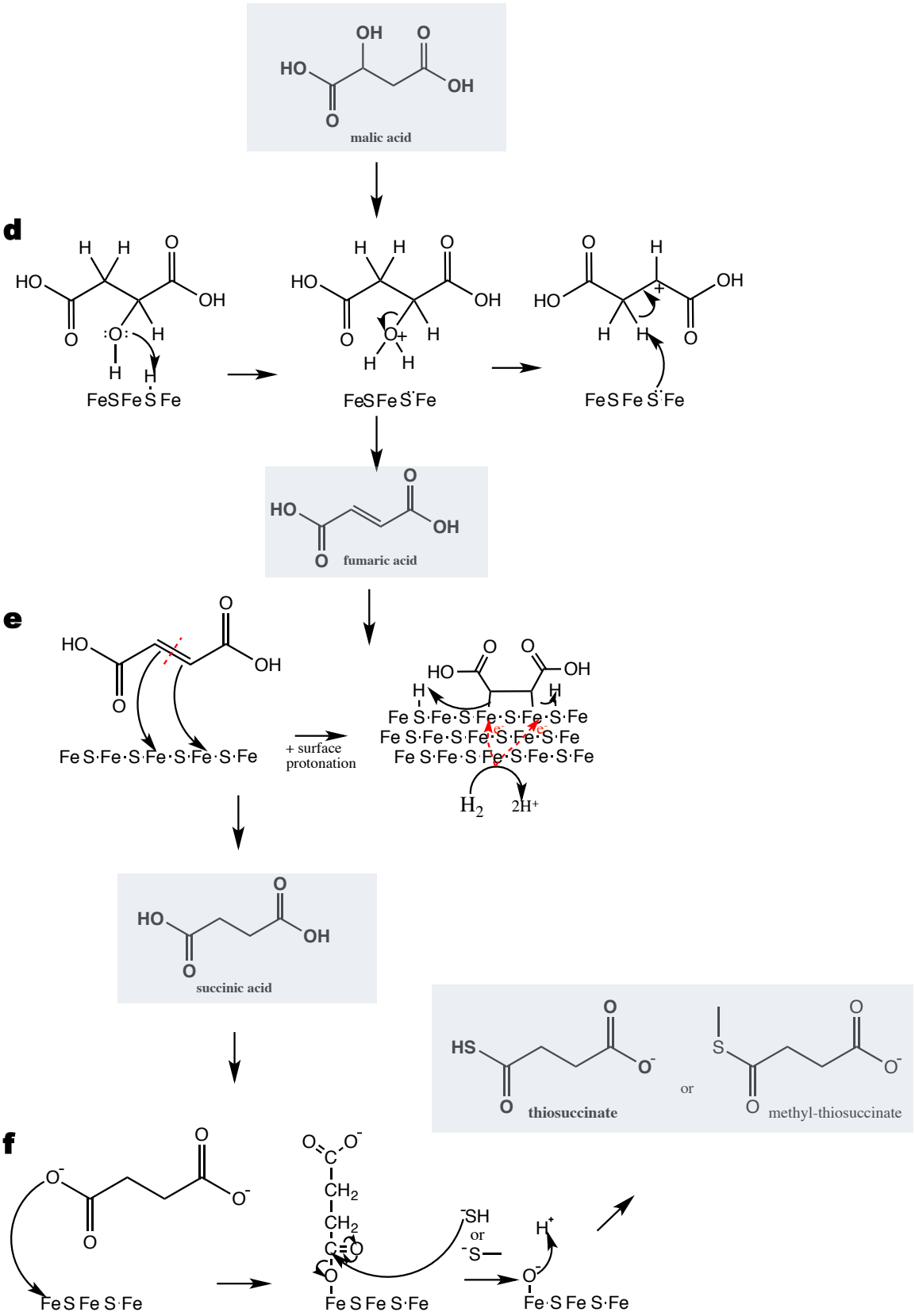

malic acid
d
FeS Fe S Fe
FeS Fe S Fe
FeS Fe S Fe
fumaric acid
e
Fe S Fe S Fe S Fe S Fe
+ surface protonation
Fe S Fe S Fe S Fe S Fe
Fe S Fe S Fe S Fe S Fe
Fe Fe
H2
2H+
succinic acid
thiosuccinate
or
methyl-thiosuccinate
f
Fe S Fe S Fe
Fe S Fe S Fe
-SH
or
-S—
Fe S Fe S Fe

Figure 3.4. Primordial incomplete reverse Krebs cycle facilitated by pH gradients across Fe(Ni)S semipermeable barriers (Camprubi et al., 2017). (**d**) Acid–base catalysed dehydration of malic acid by the mineral surface yielding fumaric acid. (**e**) Adsorption of fumaric acid though π bond cleavage followed by hydrogenation of its secondary carbons (with electrons from H_2 oxidation in the alkaline phase and H^+ ions from surface mineral protonation) yielding free succinic acid. Iron atoms channel electrons from H_2 catalysing two one-electron reductions (**f**) Thiolation or methyl-thiolation of succinic acid yielding thiosuccinate or methyl-thiosuccinate. The collateral hydroxylation of the catalyst surface can be readily removed by dehydration at low pH.

While this mechanism is pleasing in itself, we note that the carbonylations and hydrogenations of the reverse incomplete Krebs cycle are closely analogous to those in the acetyl CoA pathway itself. The general idea that there are two branches of the acetyl CoA pathway, one which generates CO and the other a methyl group, is chemically misleading. As shown in Figure 3.1, the formation of thioesters is best seen as the carbonylation of a methyl group, which is formed through the partial reduction of CO_2 on an Fe(Ni)S surface by Fischer-Tropsch and Koch-type reactions. The reverse incomplete Krebs cycle as a series of analogous carbonylations and hydrogenations on an Fe(Ni)S surface is shown in Figures 3.3, 3.4, and 3.5 (originally part of one large figure, but divided for better detail clarity). For simplicity, CO is shown bound to the surface, but this is not calling for a 1 bar atmosphere of CO, or even high concentrations in hydrothermal vents; the reduction of CO_2 to CO (and $–CH_3$) on the Fe(Ni)S surface is inferred, as shown in Figure 3.1. From thioacetate in Figures 3.3 to 3.5, the carboxylic acids of the reverse Krebs cycle are depicted as being successively carbonylated by nickel-bound CO, and then hydrogenated by electrons from H_2, which are transferred across a thin, semi-conducting Fe(Ni)S barrier as described above.

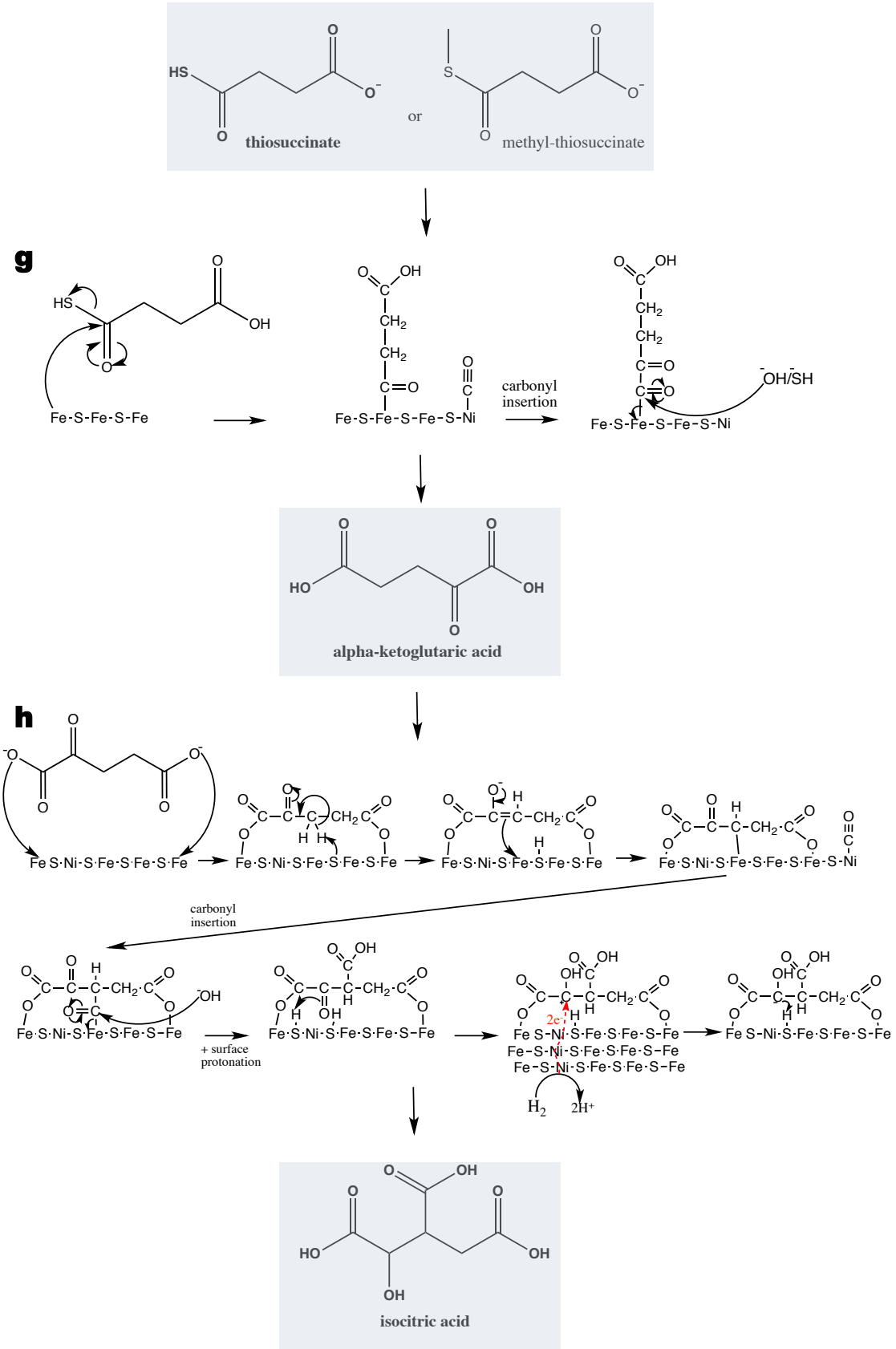


Figure 3.5. Primordial incomplete reverse Krebs cycle facilitated by pH gradients across Fe(Ni)S semipermeable barriers (Camprubi et al., 2017). (**g**) Thiosuccinate carbonylation yielding α-ketoglutaric acid. (**h**) Isocitric acid synthesis by carbonylation of its 3' carbon and Ni-mediated keto group hydrogenation by a mechanism equivalent to reaction in Figure 3.3c.

It is worth stressing that each carbonylation reaction should be facilitated by the pH gradient across a barrier, as discussed above, to form CO from CO_2; and each hydrogenation reaction should likewise be aided by the pH difference, with protons deriving either from the acidic ocean waters or from adjoining –S– groups in the Fe(Ni)S barrier, which are protonated below pH 7.5. This is admittedly hypothetical chemistry, and some steps might not be facile, but it is not of the 'if pigs could fly' type (Orgel, 2008), as there is no requirement for a complete cycle with declining yield at each step. On the contrary, each step is driven independently by the high concentration of H_2 and proton gradients across semi-conducting barriers, both of which are sustained by hydrothermal flow. It is also probable that the proposed abiotic reactions would yield other products apart from the specific ones required for the incomplete reverse Krebs cycle; we see this as a valid criticism. As discussed earlier, we accept that early proto-metabolic pathways probably presented low specificity and propose that available catalysts (both inorganic minerals and organically-chelated ones) would have been crucial at directing (and precluding) certain reactivity pathways, in time generating the ancient core of metabolism. This being said, one could predict that the study of how simple catalysts allowed for certain reaction pathways (and not others), will slowly reveal how the core of ancient metabolism chemically developed before (or alongside) the advent of genetic coding. These predictions are experimentally testable and are currently being so in a microfluidic reactor. Roldan et al., (2015) have achieved equivalent reductions of CO_2 to pyruvate, using direct electrochemical potential rather than H_2 and pH modulation of reduction potential, as discussed here. Others (Muchowska et al., 2017) have synthesised large portions of the reverse Krebs cycle using metals (e.g. Fe^0) as reductants but, even though this shows that the reduction of CO_2 is possible, the choice of the electron donor is of limited (compared to H_2) geological and biological plausibility.

3.2.8 Collapse to the complete reverse Krebs cycle

The steps discussed above concern abiotic chemistry, driven by the far-from-equilibrium conditions and topological structure of pores in alkaline

hydrothermal vents, based on a deep congruence with the carbon and energy metabolism of methanogens. We propose that the ancestral metabolism of the first cells was similar to methanogens (Sojo et al., 2016), and the early divergence of bacteria and archaea arose from the independent origins of electron bifurcation and active pumping in the methanogens and acetogens (Lane & Martin, 2012; Sojo et al., 2016; Sojo et al., 2014). This Chapter will conclude with some thoughts on the origins of the complete reverse Krebs cycle in bacteria, which obviously do not suffer from the abiotic problem of declining yields.

So what is the advantage of a complete reductive Krebs cycle, given that the modern cycle requires an investment of ATP at several steps? One factor might be thermodynamic. The rate of flux through a linear metabolic pathway depends (amongst other factors) on the concentration of end products relative to substrates; an accumulation of end products inhibits flux, as does a low concentration of substrates. In the case of the incomplete reverse Krebs cycle, an accumulation of citrate would slow flux through the pathway. But if the citrate were split into oxaloacetate and acetyl CoA, the overall effect would be to remove the end product and replenish the substrates, providing a thermodynamic drive and balancing availability of precursors for the synthesis of amino acids, fatty acids, sugars and nucleotides needed for growth (Eigen, 1971). As noted earlier, the reverse Krebs cycle is autocatalytic, providing two oxaloacetates for each turn of the cycle, which promotes stable growth. The input of acetyl CoA depends stoichiometrically on the spinning of the cycle itself, meaning there is no longer a need for input from the acetyl CoA pathway.

Other factors might have contributed to displacing the acetyl CoA pathway from most bacteria. C1 analytical chemistry is notoriously difficult, and we wonder whether cells also struggle to react two C1 molecules to form the acetyl moiety of acetyl CoA. It could be simpler for enzymes to recognize substrates if the C1 carbons are attached to longer chain carboxylic acids, rather than other C1 molecules. The acetyl CoA pathway is unique among the six known pathways of carbon fixation in dealing solely with C1 chemistry (Fuchs, 2011). Another advantage of the reverse Krebs cycle is that the acetyl CoA pathway combines

carbon and energy metabolism in the same pathway, whereas the reverse Krebs cycle has to derive the necessary reducing power and ATP from alternative sources (Martin & Russell, 2007), notably phototrophy as originally reported by Evans et al. (1966). H_2 and CO_2 can then be used solely for carbon fixation, rather than for ATP synthesis too, enabling growth at lower partial pressures of H_2 or in its absence (Campbell & Cary, 2004; Thauer et al., 2008). Given either phototrohpy or relatively high concentrations of anaerobic electron acceptors for a respiratory chain, ATP synthesis ceases to be rate limiting. And with abundant ATP, there would no longer be any need for a proton-motive force to power the chemistry proposed in Figures 3.3, 3.4 and 3.5, freeing membrane surface for other purposes, including faster ATP synthesis. That could explain why most steps of the reverse Krebs cycle no longer depend on the proton-motive force.

3.3 Conclusions

Ever since its discovery in 1966, the reverse Krebs cycle has been proposed as an ancient pathway of CO_2 fixation (Evans et al., 1966), potentially even prebiotic (Smith & Morowitz, 2004; Wächtershäuser, 1990). However, the conception of primordial metabolic cycles was challenged by Orgel on the basis that a steep decline in yield at each step would preclude non-enzymatically catalysed flux through the full cycle (Orgel, 2008); we see this is a valid criticism. The complete reverse Krebs cycle is also phylogenetically restricted almost entirely to bacteria, and so is unlikely to date back to the common ancestor of bacteria and archaea (LUCA) (Campbell & Cary, 2004; Fuchs, 2011; Martin & Russell, 2003; Weiss et al., 2016). The short, linear acetyl CoA pathway, which uses Fe(Ni)S proteins and electrochemical ion gradients across membranes to drive both carbon and energy metabolism from H_2 and CO_2, solves many of these problems. This pathway is found in both methanogenic archaea and acetogenic bacteria, but deep differences in the mechanism of electron bifurcation and methyl synthesis obscure the ancestral pathway of CO_2 fixation (Nitschke & Russell, 2013; Sojo et al., 2016; Sousa & Martin, 2014). Our group has proposed elsewhere that these differences arose with the origin of

active pumping (Lane & Martin, 2012; Sojo et al., 2014), and that ancestral carbon fixation in the first cells resembled that of methanogens (Sojo et al., 2016), which use electrochemical ion gradients to drive ferredoxin reduction via the membrane-bound NiFe hydrogenase Ech.

These factors point towards the topology of pores in submarine alkaline hydrothermal vents, in which natural proton gradients across Fe(Ni)S barriers could drive the reduction of CO_2 by H_2. I suggest that the initial reduction of CO_2 to a methyl group proceeded via a Fischer Tropsch-type mechanism on the acid side of an Fe(Ni)S surface, followed by carbonylation via a Koch-type mechanism. These endergonic reactions ought to be facilitated by pH differences between phases, which modulate the reduction potential of H_2, CO_2 and Fe(Ni)S minerals such as mackinawite (Herschy et al., 2014; Sojo et al., 2016; Wolthers et al., 2005). Similar hydrogenations and carbonylations, facilitated by natural proton gradients across Fe(Ni)S barriers, could generate longer chain carboxylic acids (up to 6 carbons) via a prebiotic equivalent to the reverse incomplete Krebs cycle in methanogens. Finally, after the evolution of cells with genes and proteins, closure of the complete reverse Krebs cycle should displace the acetyl CoA pathway from bacteria and archaea, except under anoxic conditions and at high partial pressures of H_2. This framework helps explain the origins of autotrophic CO_2 fixation in alkaline vents, the early divergence of bacteria and archaea, and the later evolution of the reductive Krebs cycle in bacteria.

Chapter 4

Prebiotic phosphorylations mediated by acetyl phosphate

The work described in this Chapter has been published in Whicher et al. (2018) as part of a first co-authored publication with Dr Alexandra Whicher. The work on the synthesis of AcP presented in Section 4.2.1 was independently performed by Dr Alexandra Whicher, whilst the work on the stability of AcP was performed jointly. This work on the synthesis and characteristics of AcP is presented in this Chapter as it is crucial for the overall understanding of the AcP-mediated phosphorylation experiments detailed in further Sections.

4.1 Introduction

In the previous Chapter I discussed in detail how hydrothermal vents could drive the synthesis of organic molecules, such as methyl thioacetate, pyruvate and Krebs cycle intermediates, from CO_2 and H_2 by Fischer Tropsch-type hydrogenations and Koch-type carbonylations. I also mentioned that the reaction of CO_2 and H_2 facilitated by the natural pH gradients present at alkaline hydrothermal vents could also yield a more immediate reactive organic molecule, formaldehyde. Lane has argued that proton gradients across thin semi-conducting barriers could drive the reduction of CO_2 to formaldehyde directly in a manner analogous to modern autotrophic cells such as methanogens (Lane, 2014; Sojo et al., 2016). In short, an abiotic mechanism analogous to that of Ech (from methanogens) could theoretically catalyse the reduction of CO_2 (by H_2) based on differences in reduction potential, which are pH-dependant. Thus while it is essentially 'impossible' for H_2 to reduce CO_2 at pH 7 (because the reduction potential is not low enough) in vents it could be possible. The FeS catalytic barrier would act as a semiconductor by 'feeling' the

reduction potentials in each side and should transfer electrons from H_2 to CO_2 (Figure 1.3), since the acidic side would be more electronegative.

Under the conditions found in alkaline hydrothermal vents, formaldehyde undergoes the formose reaction. The formose reaction was introduced earlier; in short, in 1861 Alexander Butlerow discovered the formose reaction where formaldehyde undergoes a series of aldol condensation and isomerisation reactions to form sugars, including ribose, in alkaline medium (Figure 1.4). Ribose is the structural backbone of RNA and many other key organic compounds (e.g. several vitamins and cofactors); for this reason, the formose reaction has been of interest as an abiotic source of sugar backbones. As explained in the Introduction, one of the prebiotically interesting next steps after obtaining ribose (in order to generate nucleotides) would be to phosphorylate it. Furthermore, it is possible that phosphorylation could favour the accumulation of pentose sugars in the complex formose reaction (see Chapter 6). A historical criticism of the relevance of the formose reaction, even under ideal hydrothermal conditions, is that it is frequently claimed to yield little more than tars and complex mixtures of biologically irrelevant sugars (Gollihar et al., 2014). Mellersh & Smith (2010) suggested that phosphorylation could drive the synthesis of ribose phosphate (from ribose from the formose reaction), but until this project there was no experimental verification of this. The formose reaction has been extensively studied during this project and will be discussed in detail during Chapter 6; suffices to say here that there is a plausible abiotic source of sugars at Hadean alkaline hydrothermal vents, and that sugar phosphorylation during the formose reaction could funnel the formose reaction towards pentoses such as ribose.

This Chapter focuses on the study of the phosphorylating capabilities of AcP. As mentioned earlier, both acetyl CoA and ATP are universally conserved across life, hence are most probably ancient (Fuchs, 2011; Martin et al., 2014; Morowitz et al., 2000; Smith & Morowitz, 2004). Nonetheless, both are complex molecules produced by genetically encoded enzymes, and so are unlikely to have driven the emergence of biochemistry at the origin of life. Plausible prebiotic precursors to acetyl CoA and ATP have been proposed to operate in a

'thioester world' (de Duve, 1988, 1991, 1998; Goldford et al., 2017; Sousa et al., 2013). Prebiotic thioesters such as methyl thioacetate could arguably phosphorolyse to generate a phosphoester bond equivalent to that in ATP, as in the simple 2-carbon molecule acetyl phosphate, AcP (Ferry & House, 2006; Martin et al., 2014; Sojo et al., 2016). This is analogous to modern cells, where acetyl CoA is readily phosphorolysed to generate AcP in both bacteria and archaea (Decker et al., 1970; Martin & Russell, 2007; Thauer et al., 1977). AcP then phosphorylates ADP to ATP (Ferry & House, 2006; Schönheit, 2016) making it the fulcrum between thioester and phosphate metabolism (de Duve, 1991). In terms of prebiotic chemistry, AcP could have been formed from simple thioesters such as methyl thioester, and it could have then driven phosphorylation and condensation reactions in a similar fashion to ATP (Figure 4.1) (de Duve, 1991; Ferry & House, 2006; Martin & Russell, 2007), narrowing the gap between geochemistry and the origins of intermediary metabolism.

However, simple thioesters such as methyl thioacetate and putative ATP analogues like AcP have generally been considered to be too unstable to persist for long enough to drive prebiotic chemistry, especially at the high or low pH and hot temperatures found in hydrothermal systems (Barge et al., 2014; Chandru et al., 2016). AcP has long been overlooked experimentally as a prebiotic precursor to ATP, except in the synthesis of more stable (but less reactive) products such as pyrophosphate (Baltscheffsky & Baltscheffsky, 1995; Barge et al., 2014; de Zwart et al., 2004; Kornberg et al., 1999). This perception stems largely from the 'heterotrophic' approach to prebiotic chemistry: stable organics are assumed to accumulate in a primordial soup or warm geothermal pond, perhaps with different precursors formed under distinct conditions before being brought together by some circumstance (Patel et al., 2015), all of which demands long-term stability. In contrast, autotrophic origins imply a continuous flux of inorganic substrates (e.g. H_2 and CO_2) in a specific environment, in which organic synthesis is driven by the continuous flux, and the rates of synthesis and breakdown or reaction at least balance over time (Barge et al., 2017; Branscomb & Russell, 2013; Lane, 2015). It is of little relevance if prebiotic thioesters or AcP react within minutes or hours as long as they are formed at a similar rate; indeed it is more parsimonious if they do, as they are

then approaching the lability of biochemical intermediates. In this chapter then, the synthesis and reactivity of AcP under mild alkaline hydrothermal conditions will be examined with this lability in mind.

For these experiments simulating Hadean alkaline hydrothermal vents, temperatures from 20 to 60 °C and pH values from 6 to 12 have been considered, conditions that are equivalent to the ranges found in these (Lane & Martin, 2012; Martin et al., 2008; Nitschke & Russell, 2009; Russell et al., 2014, 1994). Such vents are composed of labyrinths of interconnected micropores with thin inorganic walls, which separate hydrothermal fluids (pH 9 – 12) and ocean waters (Kelley et al., 2005; Russell & Hall, 1997; Russell et al., 1994). The oceans were probably mildly acidic (pH 5–7) in the Hadean, as CO_2 levels were higher (Arndt & Nisbet, 2012; Pinti, 2005). Mixing within the vent is driven by convection and thermal diffusion (Baaske et al., 2007; Braun & Libchaber, 2002; Herschy, et al., 2014; Kreysing et al., 2015; Mast, 2013; Mast & Braun, 2010), which is important for three reasons. First, conditions for organic synthesis are not necessarily warm and strongly alkaline, but could equally be cool and neutral pH, or even mildly acidic. Second, convective cycling means that products formed in cool, neutral conditions can cycle through warm, alkaline conditions, potentially driving other reactions, including simple hydrolysis. Third, thermal cycling can concentrate small organics such as nucleotides by at least 5000-fold via thermophoresis (Baaske et al., 2007; Braun & Libchaber, 2002; Herschy, et al., 2014; Kreysing et al., 2015; Mast, 2013; Mast & Braun, 2010); such extreme concentration by thermophoresis partially justifies the relatively high concentrations of reagents used in the work presented in this Chapter.

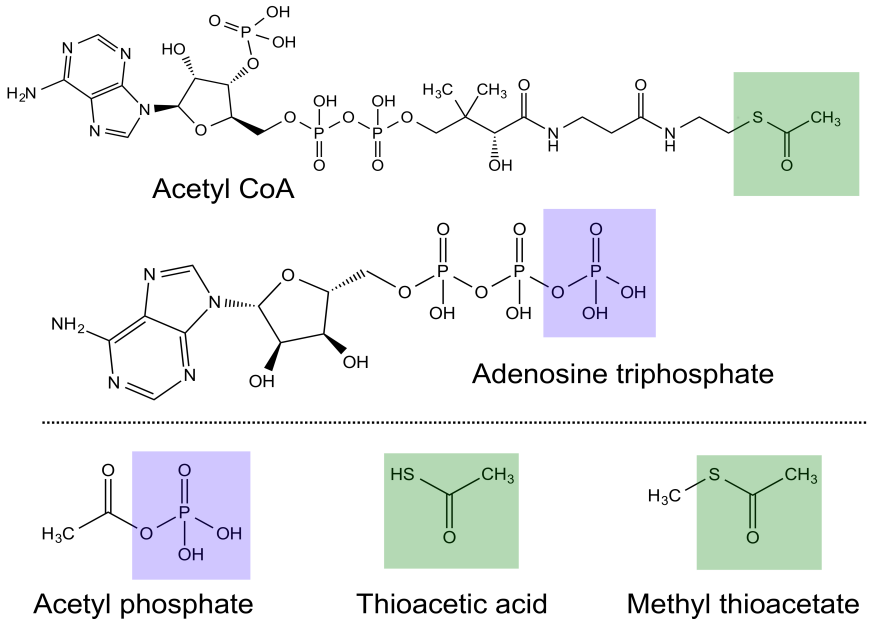


Figure 4.1. Molecular structures of key biotic and prebiotic molecules (Whicher et al., 2018). Acetyl CoA is the hub of metabolism in all six known pathways of carbon fixation. The 'business end' of the molecule is the reactive thioester group (green), which has prebiotic analogues in methyl thioacetate/thioacetic acid. ATP is the universally conserved energy currency in modern metabolism; AcP could have been a simple precursor, driving similar reactions. The lilac shading highlights equivalent phosphate groups.

4.2 Results

4.2.1 Synthesis of AcP under alkaline hydrothermal vent conditions

We expected to synthesise AcP by phosphorolysis of methyl thioacetate (CH_3COSCH_3). This simple prebiotic thioester has been argued to have properties equivalent to acetyl CoA (de Duve, 1991, 1988; Lane & Martin, 2012; Martin & Russell, 2007), and has indeed been formed under hydrothermal conditions from CO and CH_3SH alone (Huber & Wächtershäuser, 1997). In fact, no AcP from CH_3COSCH_3 in the presence of 20 mM Na_2HPO_4 was detected in our experiments, but we rather generated yields of up to 2% within 1–2 h from the even simpler precursor thioacetate (CH_3COSH, 40 mM) (Figures 4.1 and 4.2). This probably reflects the different hydrogen bonding capabilities of both substrates: the electronegativity of the S atom of thioacetate is increased by

hydrogen bonding of its proton with the (water) solvent, since this causes the H atom to pull the electron cloud of the S-H bond towards H. In turn this effect favours the nucleophilic attack (from Pi) to the carbonyl group of the thioacetate, favouring the synthesis of AcP. In contrast, the methyl moiety of methyl thioacetate cannot establish hydrogen bonding with the solvent, making its productive phosphorolysis (yielding AcP) less likely. In order to confirm this, a hydrogen-deuterium exchange experiment using deuterated water as the solvent could be performed. The rate of exchange of the S-H proton of thioacetate with the solvent depends on the hydrogen bonding capabilities of such proton; this can be monitored using H-NMR or mass spectroscopy. If this proton is heavily involved in hydrogen bonding with the solvent, it would easily exchange with deuterium atoms, therefore its rate of deuterium incorporation would be fast. In contrast, a proton bonded to a more electropositive atom (e.g. C instead of O) would have much lower hydrogen bonding capabilities, and thus would exchange with deuterium considerably slower. In conclusion, if the rate of deuteration of the S-H proton of thioacetate is higher than that of the methyl protons of a similar control molecule unable of hydrogen-bonding (e.g. acetone), this would indicate that the proton was indeed heavily involved in hydrogen-bonding, which in turn favoured the phosphorolysis of thioacetate, yielding AcP.

The yield of AcP synthesis depended on pH, temperature, and ions present. Under mildly acidic conditions and cooler temperatures (pH 6, 20 °C), equimolar mixtures of Ca^{2+} and Mg^{2+} ions promoted synthesis of AcP, compared with no ions (Figure 4.2a, red line). Under more alkaline conditions, Ca^{2+} and Mg^{2+} ions lowered AcP synthesis, because Ca^{2+} precipitated out some phosphate as apatite, so less was available for AcP formation (Figure 4.2b–c; data for Ca^{2+} alone are not shown, as nanoparticles interfered with NMR measurements). AcP was not formed at all at pH 11, even in the absence of ions (not shown). In contrast, Mg^{2+} alone precipitated relatively little phosphate below pH 11, and doubled yields of AcP under neutral or mildly alkaline conditions, presumably promoting synthesis relative to breakdown of AcP (Figure 4.2d–f). These data could be interpreted as foreshadowing the close association of Mg^{2+} ions with AcP, ATP and nucleotides, and the exclusion of Ca^{2+} ions from modern cells.

Synthesis under anaerobic conditions (mimicking Hadean conditions) did not affect the yield of AcP, indicating that thioacetate is equally reactive in the absence of oxygen (Figure 4.3b). The addition of Fe^{2+} was also tested in order to see if such a prevalent metal at Hadean vents would have any effect on the synthesis of AcP, but it did not (Figure 4.3a).

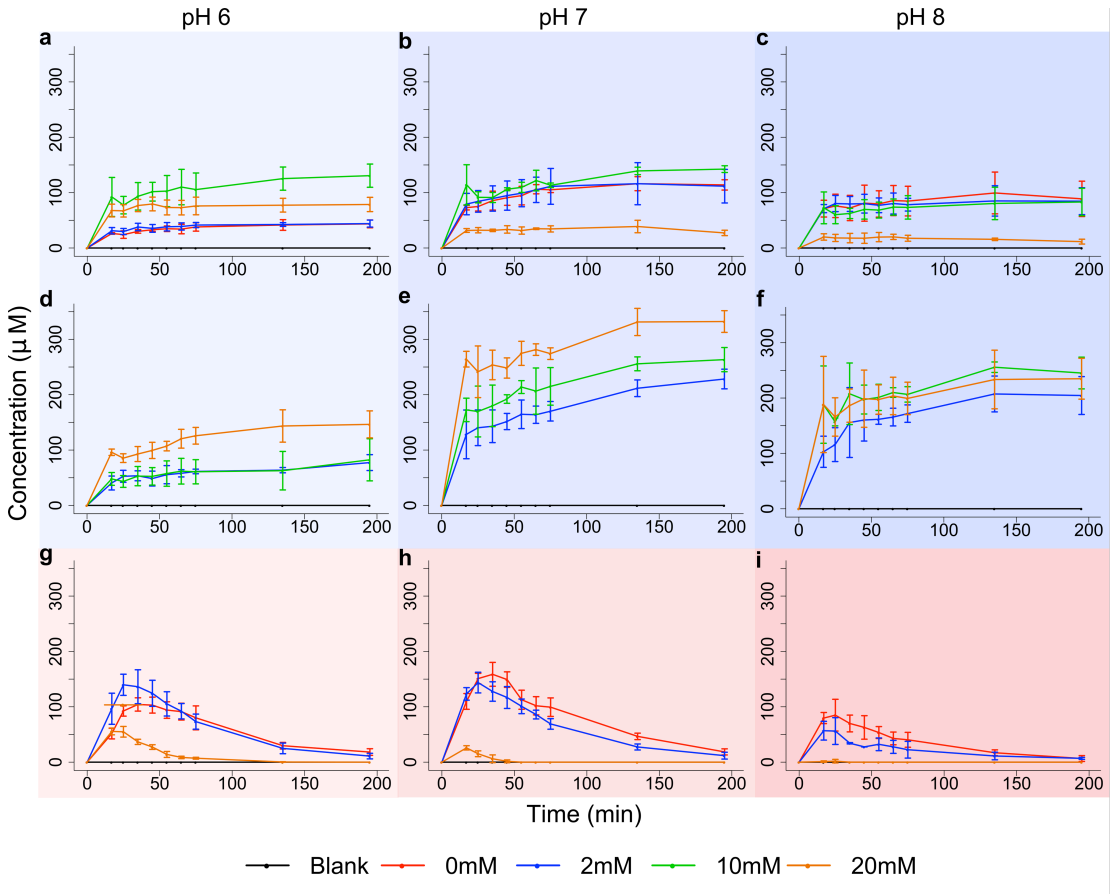


Figure 4.2. Synthesis of AcP from orthophosphate and thioacetate (Whicher et al., 2018). (**a**–**c**) AcP synthesis at pH 6, 7 and 8, respectively, with varying equimolar concentrations of Ca^{2+} and Mg^{2+} ions at 20 °C (blue). (**d**–**f**) Varying concentrations of Mg^{2+} ions alone. (**g**–**i**) Varying concentrations of both Ca^{2+} and Mg^{2+} ions at 50 °C (red). The upper two rows represent experiments performed at 20 °C, whilst the lower row those performed at 50 °C. Darker shading of the vertical columns to the right represent growing reaction pH (6 to 8). The maximum yield of AcP obtained was of over 300 μM (corresponding to a ~2% conversion from starting 40 mM of thioacetate). N=3± SD. Data provided by Dr Alexandra Whicher.

However, when performing the experiment at 50 °C with equimolar mixtures of Ca^{2+} and Mg^{2+}, the balance between AcP formation and hydrolysis shifted markedly towards hydrolysis (Figure 4.2g–i). Although the initial rate of synthesis was slightly faster, AcP was almost completely hydrolysed within 3 h,

especially under mildly alkaline conditions (pH 8; Figure 4.2i). In the vent setting, therefore, AcP synthesis should occur mostly in cooler, more neutral regions, precisely the regions where organics tend to accumulate by thermophoresis (Baaske et al., 2007; Herschy et al., 2014), and should be stable under aqueous neutral to alkaline conditions over at least several hours.

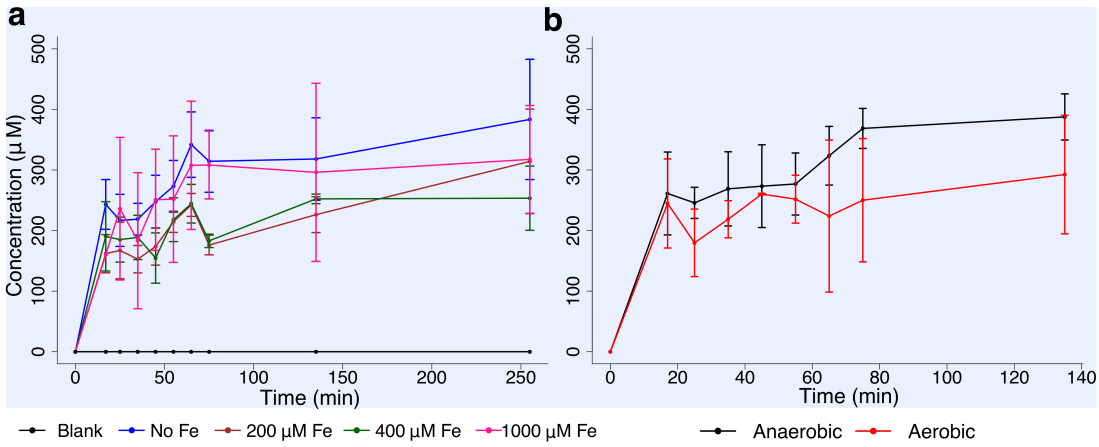


Figure 4.3. Anaerobic synthesis of AcP from orthophosphate and thioacetate at pH 7 and 20°C (Whicher et al., 2018). (**a**) Anaerobic experiments carried out in the presence of different concentrations of Fe^{2+} ions (0, 200, 400 and 1000 µM). In the absence of thioacetic acid no AcP was produced (black); and (**b**) under anaerobic and aerobic conditions with no ions present. N=3 ±SD. Data provided by Dr Alexandra Whicher.

A study testing the stability of AcP in water was performed in order to help us understand the dynamics of acetyl phosphate synthesis (Figure 4.2), and that of ribose phosphorylation (see next Section (4.2.2)). At ambient temperatures and non-extreme pH, AcP is reasonably stable, as previously reported by others (Etaix & Buvet, 1974; Koshland, 1952) with ~20% hydrolysed over 5 h at 20 °C under the range of pH conditions tested (Figure 4.4a, circles). The rate of hydrolysis depends strongly on temperature, with AcP completely hydrolysed within 3–5 h at 50 °C, and within 90 min at 60 °C (Figure 4.4a, squares and triangles respectively). In contrast, pH had little effect at any given temperature (Figure 4.4a). The presence of Mg^{2+} (Figure 4.4b) and Ca^{2+} (Figure 4.4c) ions also had little effect, slightly speeding the initial rate of hydrolysis at both 20 °C and 50 °C, as reported previously for concentrated salt solutions by Di Sabato and Jencks (1961b), but not changing the overall proportion hydrolysed over 5 h (Figure 4.4b-c).

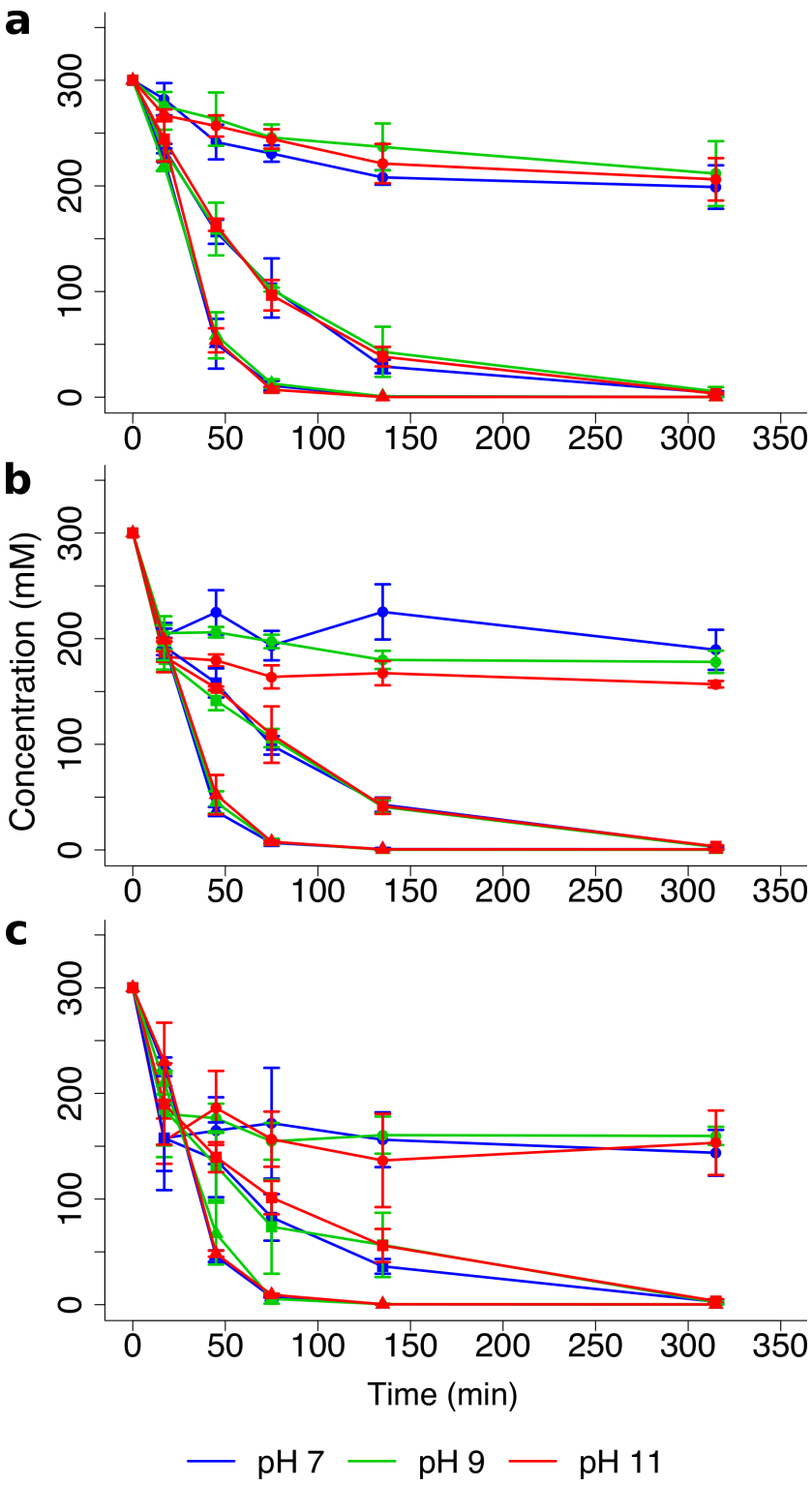


Figure 4.4. Stability of AcP depending on pH, temperature and ions (Whicher et al., 2018). Degradation profile for AcP over 5 hours with and without ions at pH 7, 9 and 11, stored at 20 °C (circles), 50 °C (squares) and 60 °C (triangles). (**a**) no ions added, (**b**), with 20 mM Mg^{2+} ions added, (**c**), with 20 mM Ca^{2+} ions added. Initial concentration for every condition was 300 mM verified with the protocol stipulated in Section 2.1.2. N=3 ±SD. With thanks to Dr Alexandra Whicher for providing the data shown in (b) and (c).

4.2.2 Phosphorylation of ribose by acetyl phosphate

The experiments on the AcP-mediated phosphorylation of D-ribose occurred as two sets, since several issues with the analytical instruments took place during the months these tests were carried out (see below). Despite this, both sets of experiments were successful under a wide range of pH, temperature and ions presence conditions; the main differences between both sets were (i) the maximum yields of ribose phosphate obtained, as well as (ii) the overall reaction profile over time.

For all ribose phosphate synthesis experiments, the quantification was achieved by interpolating the experimental integrated areas with a calibration curves prepared in triplicate using commercial ribose 5-phosphate. A vial with 250 μM of commercial D-fucose was thawed, derivatised and analysed as an external standard alongside the analysis of both calibration curve samples, as well as experimental ones. This was done in order to minimise variation or noise in experimental data generated due to random changes in the derivatisation efficiency and instrumental precision.

4.2.2.1 First experimental set

An example of the calibration curve used is shown in Figure 4.5; and the calibration parameters are displayed in Table 4.1

Table 4.1. Ribose 5-phosphate calibration curve (1-500 μM).

Compound	Linearity	Lower limit of quantification	r^2
Ribose 5-phosphate	As/Ai = 0.007c - 0.0568[a]	1 μM	0.99234

[a]As/Ai: peak area ratio of the analyte to the external standard; c: concentration in μM

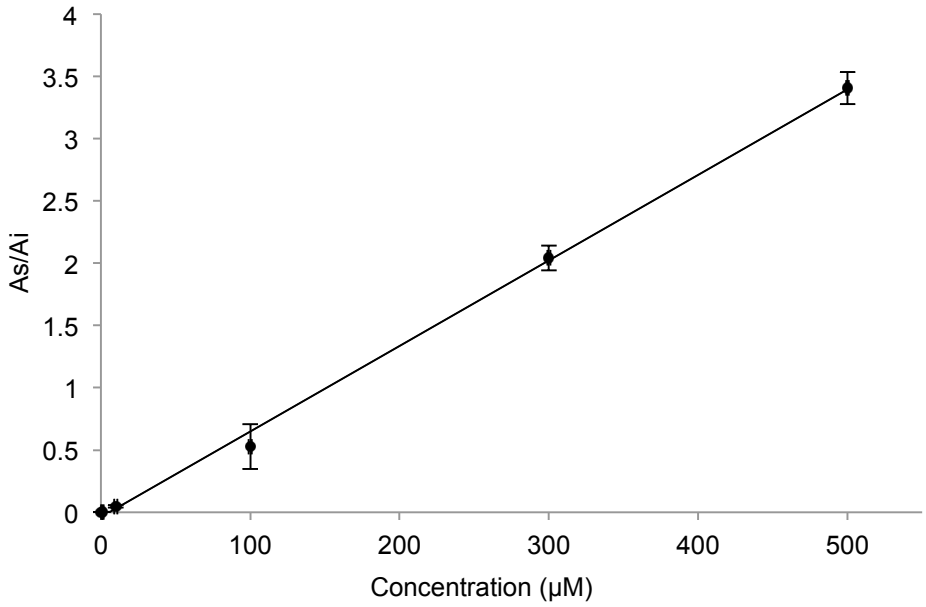


Figure 4.5. Calibration curve (1-500 μM) for commercial ribose 5-phosphate, corrected using fucose as external standard (y axis). N=3 ±SD.

Ribose + *Acetyl phosphate* ⇄ *Ribose phosphate* + *Acetate*

Reaction rate (pH 11, 20 °C): $\frac{\Delta\,[Ribose\ phosphate]}{\Delta\,[t]} = \frac{209.43\mu M}{120\ s} = 0.00174\ mM/s$

The maximum yield of ribose phosphate obtained in these experiments with respect to the concentration of ribose was low (~0.1%), but the rate of synthesis was rapid. Concentrations of ribose phosphate over 200 μM were present after only 2 minutes at 20 °C when no ions were present (Figure 4.6a) (samples were impossible obtain before 2 minutes due the time it took for the samples to freeze at –80 °C). The yield was higher at 50 °C, with concentrations reaching 300 μM (Figure 4.6b). The reaction pH seemed to have little effect on the efficiency of this phosphorylation.

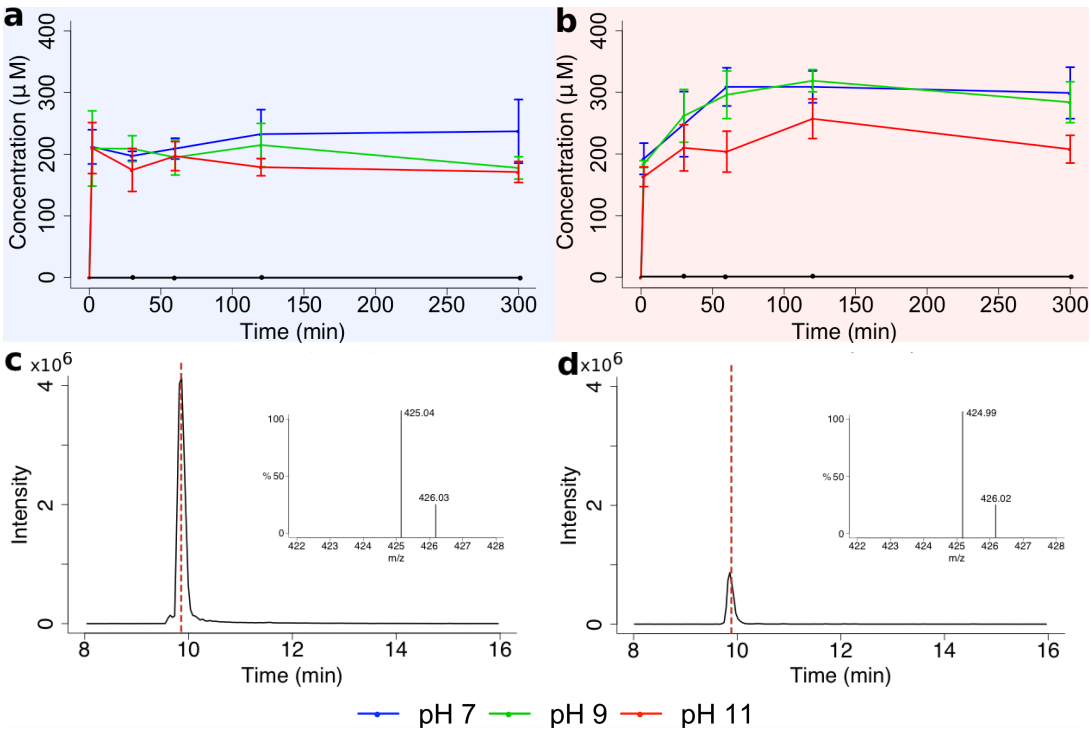


Figure 4.6. Phosphorylation of ribose by Na^+AcP. (**a**–**b**) Synthesis of ribose phosphate from D-ribose and AcP (300 mM each) at pH 7, 9 and 11 at: (**a**) 20 °C and (**b**) 50 °C. Black line indicates a negative control containing ribose only (no AcP). (**c**–**d**) LC-MS chromatogram for derivatised ribose 5-phosphate (monoisotopic mass: 425.14 m/z): (**c**) standard and (**d**) experimental sample. Graph inserts show mass spectra for each peak. N=3 ±SD.

Despite these encouraging results, it seemed strange that the concentration of ribose phosphate reached its maximum concentration so quickly, and that it hardly changed over the course of 5 hours. In order to ensure that these results were not reflecting an analytical artefact, I planned to repeat the experiments using another analytical set up which had a more restricted access (Agilent Q-TOF), as the first LC-MS setup (Acquity UPLC-MS) was on an open access regime and was often out of order. Initially, a repetition of the reaction at pH 11 (at 20 and 50 °C) using the same concentration of reagents used in Figure 4.6 (300 mM each) was performed. The acetyl phosphate stock solution used this time was of commercial origin (Sigma-Aldrich, ≥85%) and not synthesised in the lab; this was done in order to exclude the possibility of the high concentration of acetate and orthophosphate present in the lab-prepared AcP to affect the synthesis of ribose phosphate. The commercial version of AcP was a Li^+/K^+ salt whereas the synthesised in the lab was a Na^+ salt. The results of this experiment are shown in Figure 4.7.

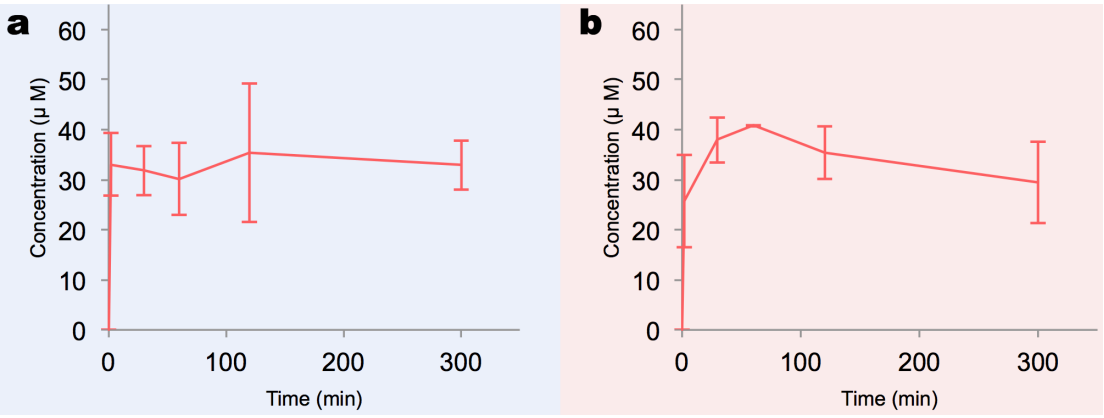


Figure 4.7. Synthesis of ribose phosphate from D-ribose and commercial Li^+AcP at pH 11 at: (**a**) 20 °C and (**b**) 50 °C. N=3 ±SD. Precursor ion monitored for derivatised ribose-5-phosphate (425.1472 m/z).

The overall shape of the concentration/time curve was very similar to the data shown in Figure 4.6. Despite this, the yield was about 7 times lower. The lower yield was not due to a problem with the new calibration curve (1-750 μM concentration range, r^2 = 0.99551) as a new one was generated again with an almost identical result; something else was responsible for the concentration decrease. During the months this experiment was being performed, the Agilent Q-TOF LC-MS instrument underwent several weeks of inoperability and days of a marked reduced sensibility. More frequent disruptions meant that the planned experiments using different pH values had to be postponed as the data acquired lacked the proper sturdiness. The cause of the instrumental problems could not be identified without doubt as there were several research groups using the instrument, but the large amount of phosphate and other ions present in our samples could, in theory, contribute to block the internal tubing of the instrument causing multiple issues such as lower signal intensity over short periods of time, as well as the complete halt of a run (due to high internal pressure) causing the next run to have material (and thus chromatographic peaks) from the previous sample, which made posterior analysis impossible. Due to these problems, and in order to prevent more issues in the future, a solid phase extraction protocol (SPE) was designed with the aid of undergraduate student Jeanne Bonnel under my supervision. This purification protocol used a C18 matrix and a range of solvents in order to separate the non-polar derivatised sugars from the polar ions (phosphates, Na^+, Li^+, K^+, etc.) present in the samples. Unfortunately this protocol inevitably diluted the samples (1/15)

but, in turn, ensured that our samples could not damage any instrumental part in the future. This protocol is detailed in the Section 2.1.4 of Chapter 2. From this point onwards, any sample bound for HPLC-UV or LC-MS analysis was purified using this SPE protocol.

Apart from the analytical problems mentioned above, the lack of reproducibility of the data from Figure 4.6 when the LC-MS set up was changed (Figure 4.7) could be explained by another three factors: (i) the type stock AcP solution used, (ii) the potential effect of ion suppression during the analysis, and (iii) problems with the derivatisation protocol. These three potential factors were tested and the results are discussed below.

(i) The AcP stock solution used in the previous experiments of Figure 4.6 was prepared in the lab and always kept close to 3-4 °C during its preparation and frozen at –80 °C directly. The commercial AcP used for the experiments in Figure 4.7 was dissolved in water at ambient temperature just before freezing it at –80 °C. The commercial AcP solution was very far from the equilibrium of its hydrolysis reaction, which yields acetate and phosphate, and thus would have been comparatively more labile than the AcP prepared in the lab, which contained about 0.1 M acetate and unknown but probably high levels of phosphate as it was one of the reagents used to synthesise it. This intrinsically higher instability of the commercial AcP coupled to its dissolution at ambient temperature could have contributed to the lower reaction yields observed in Figure 4.7 (compared to Figure 4.6).

(ii) When two chromatographic peaks overlap even a bit, ion suppression can take place; this means that the whole population of analyte molecules are not being ionised completely because there is competition from another (often more abundant) molecule being ionised at the same time. This effect causes an under-quantification of the analyte of interest. Because ribose was present in concentrations close to 300 mM, much higher than the potential amounts of ribose phosphate synthesised; the chromatographic peaks of ribose and ribose phosphate did overlap a bit; and it was impossible to separate the peaks further, it is possible that ion suppression within the instrument's ionisation

chamber occurred during my analyses and thus the concentration of ribose phosphate was routinely underestimated. An experiment using solutions of commercial ribose 5-phosphate (100 μM) incubated with either 300 mM of a solution of commercial ribose (the same concentration as in the polymerisation experiments) or the same volume of water was performed in triplicate (all samples were corrected using fucose as the external standard). The result was an 84% reduction in the amount of ribose 5-phosphate being quantified when 100 uM was incubated with 300 mM of ribose. Of course, the extent of this suppression depends on the relative amount of both analytes and the instrumental analytical conditions. The relative concentration of both analytes was never exactly the same in each experimental condition or reaction time point; therefore it would be wrong to correct all ribose phosphate concentration values presented in Figures 4.6 and 4.7 by this amount, but it would be justified to generally accept that the real value of ribose phosphate in each experiment was higher than the values given.

This reduction on the intensity of the ribose 5-phosphate signal could also have been caused by (iii) a derivatisation protocol not optimised to deal with high a concentration of sugars (300 mM); this could result in a systematic underestimation of ribose phosphate concentrations causing disruption on both the overall yields, and on the shape of the concentration/time curves in Figures 4.6 and 4.7. An experiment was designed in order to study whether the derivatisation protocol used so far was optimised for the mid mM range of sugars present in the samples:

First a dilution series of commercial ribose was prepared from 100 μM to 500 mM, and derivatised using the first derivatisation protocol (25 mM of 3-amino 9-ethylcarbazole (AEC) in methanol). The integrated peak areas were compared to the theoretical concentrations and the results indicated that above 1 mM the derivatisation efficiency collapsed dramatically (Table 4.2, middle column). This indicated that the derivatisation issue was probably the main reason why the concentration of ribose phosphate in Figures 4.6 and 4.7 was unchanging over time: only a small proportion of ribose phosphate experimentally synthesised was being derivatised. The outcome was that the instrument could detect a

signal (the analyte concentration was above the detection limit), but the signal was basically non-responsive to changes in the real concentration of analyte due to its incomplete derivatisation. In order to rectify this issue, the stock solution of 3-amino 9-ethylcarbazole in methanol was brought to its saturation concentration (238 mM), and the concentration of the other reagent ($NaCHBH_3$) was brought up to 476 mM (keeping the ratio of both reagents unchanged). The same experiment was performed again, but this time using the new derivatisation method: solutions of commercial ribose from 10 mM to 300 mM were prepared and analysed. The results (Table 4.2, right column) indicated that, this time, the derivatisation efficiency was maintained throughout the concentration scale; up to 300 mM, which was the maximum concentration of sugars present in the ribose phosphorylation experiments. Therefore, using this new derivatisation method should allow a much more accurate ribose phosphate quantification, as well as a better overview of the reaction dynamics over time. Both initial and final derivatisation protocols are detailed in Section 2.1.3 of Chapter 2.

Table 4.2. Comparison of the derivatisation efficiencies of both methods (data acquired by Jeanne Bonnel under my supervision).

Ribose concentration (mM)	Efficiency (first method)	Efficiency (second method)
0.1	**1.111**	---
0.5	**1.021**	---
0.75	**0.918**	---
1	**1.040**	---
5	**0.948**	---
10	---	**0.866**
20	---	**0.948**
50	**0.216**	**0.970**
100	**0.169**	**0.871**
300	**0.157**	**0.870**

4.2.2.2 Second experimental set

With this in mind, the experiments resulting in Figure 4.6 were repeated, this time using (i) commercial AcP stock solutions, and (ii) the final derivatisation protocol that could derivatise up to 300 mM of sugars (which was the initial ribose concentration, therefore this was also the maximum concentration of all sugars present in the samples). The analytical setup was also changed (Thermo LTQ) in order to improve the quantification reproducibility, as this instrument was more commonly available than the Agilent Q-TOF used previously. The results of these experiments are shown in Figure 4.8; the analytical parameters are shown in Table 4.3.

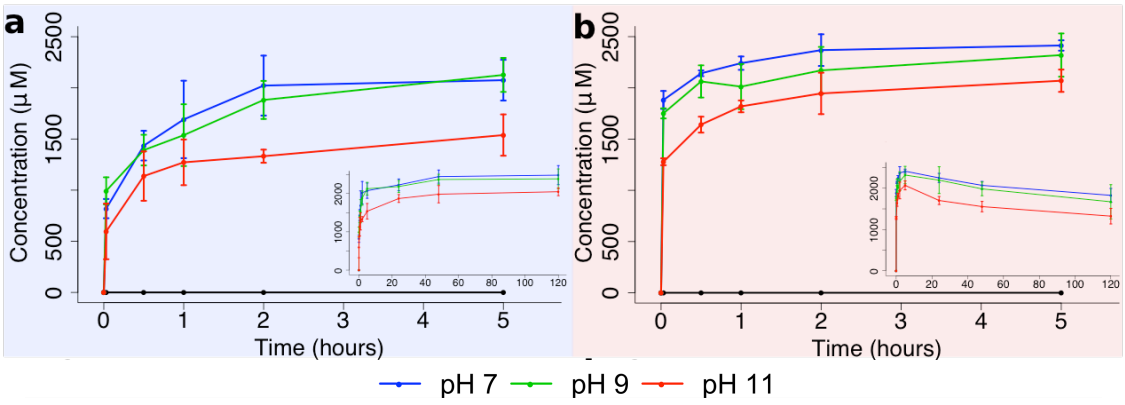


Figure 4.8. Phosphorylation of ribose by commercial AcP (Whicher et al., 2018). Synthesis of ribose phosphate from D-ribose and AcP (300 mM each) at pH 7, 9 and 11 at: (**a**) 20 °C and (**b**) 50 °C. Black line indicates a negative control containing ribose only (no AcP). Graph inserts show full reaction profile over 120 hours (5 days). N=3 ±SD.

Table 4.3. Ribose 5-phosphate calibration curve (0.8-1000 µM).

Compound	Linearity	Lower limit of quantification	r^2
Ribose 5-phosphate	As/Ai = 0.0008c + 0.0043[a]	0.8 µM	0.99386

[a]As/Ai: peak area ratio of the analyte to the external standard; c: concentration in µM

Reaction rate (pH 11, 20 °C): $\frac{\Delta\,[Ribose\ phosphate]}{\Delta\,[t]} = \frac{596.42\mu M}{120\ s} = 0.00497\ mM/s$

These results confirm that in previous experiments the yield of ribose phosphate was being underestimated systematically, and indeed reinforces the putative role of acetyl phosphate as an abiotic phosphorylating agent. The overall yield (relative to ribose) was still modest (~2%), but significantly higher than in the first set of experiments, and the rate of synthesis was rapid, with >500 µM of ribose phosphate formed within 2 min at 20 °C in the absence of

Ca^{2+} and Mg^{2+} ions, and concentrations peaking at ~2500 μM after 120 h (Figure 4.8a). The fast initial reaction rate could reflect a sugar-catalysed hydrolysis of AcP (see below), which together with the slower hydroxyl-mediated phosphorylation could explain the power curve shown in Figure 4.8. The proposed mechanism for this sugar-catalysed hydrolysis of AcP follows a nucleophilic attack by the phosphate onto ribose's carbonyl followed by the hydrolysis of AcP, yielding ribose phosphate and acetate. This would require ribose to exist in its open (aldehyde) form, which is abundant only initially: in water most of sugar molecules rapidly form an intramolecular hemiacetal. To test this, the same reaction should be performed using ribulose (the ketose form of ribose), which due to its lower electrophilic properties should impede this effect, thus slowing down the initial reaction rate. The expected reaction profile using ribulose should be similar to the phosphorylation of adenosine monophosphate by AcP (see section 4.2.3.1) since both lack a free aldehyde.

The synthesised ribose phosphate was surprisingly stable at 20 °C, with no loss of yield over 5 days (Figure 4.8a inset). The rate of synthesis of ribose phosphate was even faster at 50 °C, with concentrations reaching ~1500 μM within 2 min and ~2300 μM after 2 h, giving a similar total yield (Figure 4.8b). However, ~20% of this yield was lost by hydrolysis over the following 5 days at 50 °C (Figure 4.8b inset). The apparent stability of ribose phosphate to hydrolysis over 5 days, even at 50 °C, was unexpected, so I measured the stability of commercial ribose 5-phosphate (R5P) under equivalent pH and temperature conditions; I confirmed that R5P is indeed relatively stable to hydrolysis under mild hydrothermal conditions (Figure 4.9).

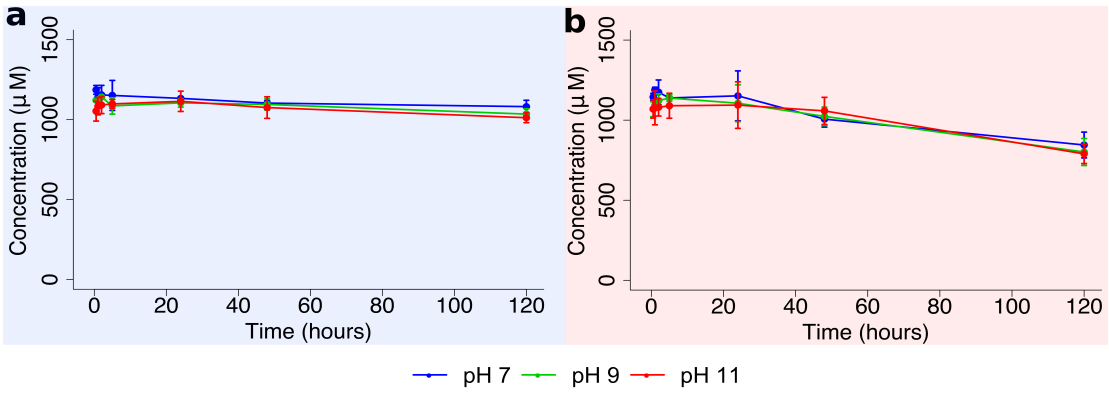


Figure 4.9. Degradation profile for commercial ribose 5-phosphate over 5 days at pH 7, 9 and 11, stored at (**a**) 20 °C and (**b**) 50 °C (Whicher et al., 2018). N = 3 ±SD.

While pH had little effect, the synthesis of ribose phosphate was slightly lower at pH 11 (Figure 4.8a-b, in red), reflecting the formation of tetra-acetylated ribose at pH 11 alone (Figure 4.10). This corresponds to the known pKa of the hydroxyl groups of ribose, which is ~11.8 (Sen et al., 2014) hence these are acetylated only at quite strongly alkaline pH. This type of pH-dependent acetylation reaction has been reported in the literature (Di Sabato & Jencks, 1961a) on amino groups; in our case the acetylation occurred on hydroxyl groups, most probably following an analogous reaction mechanism (Figure 4.10e). Small chromatographic peaks corresponding to traces of tri-acetylated ribose (Figure 4.10d) could also be found at pH 11.

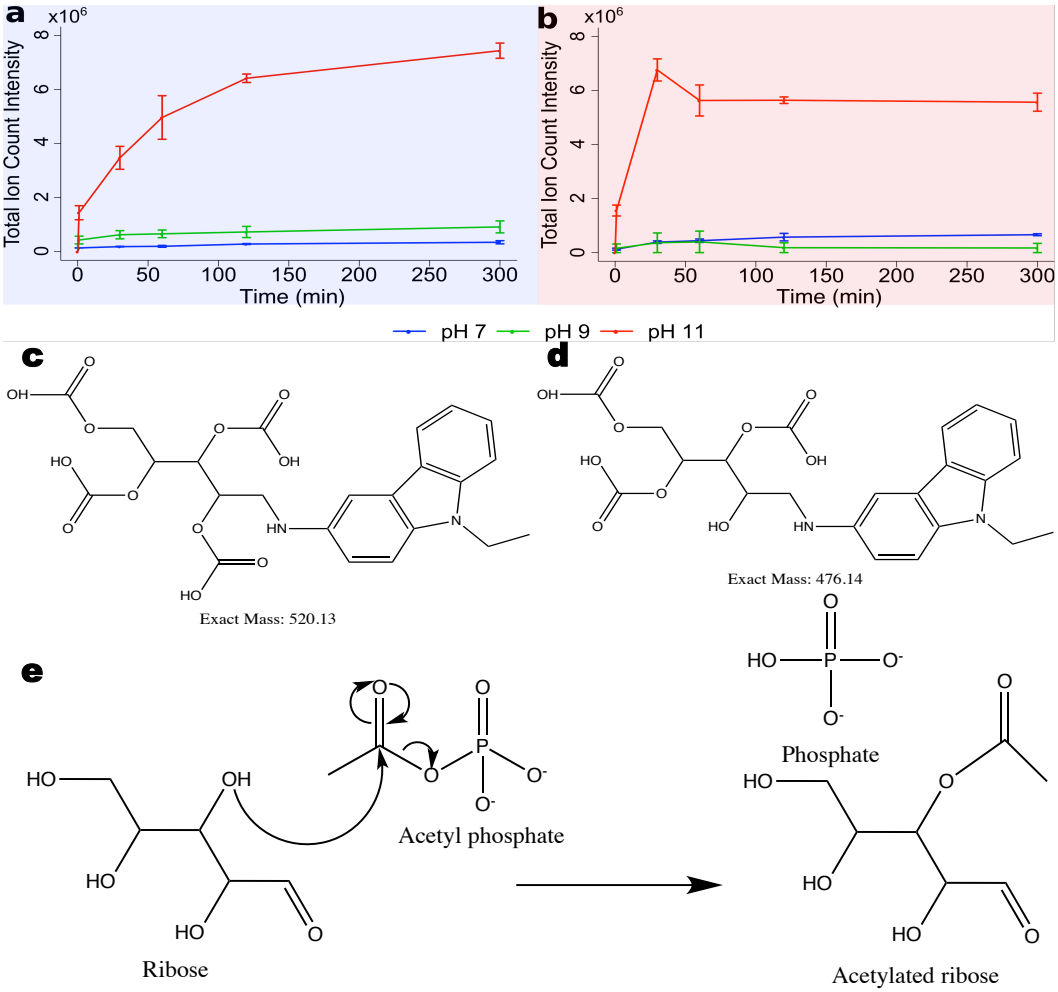


Figure 4.10. Formation of tetra-acetylated ribose from the reaction of ribose and AcP: (**a**) 20 °C, (**b**) 50 °C (amended from Whicher et al., 2018). This ion was predominantly detected at pH 11. Precursor ion monitored for tetra-acetylated ribose was 521.13 m/z (520.13 (**c**) + $1H^+$). Y-axis units are given as total ion count giving a relative abundance of the monitored ion at each experimental condition. N = 3 ±SD. Ribose molecules with differing number of acetylations (**d**) were detected also only at pH 11. The proposed reaction mechanism is depicted in (**e**).

The presence of other ions lowered the rate of ribose phosphate synthesis and tended to increase its rate of hydrolysis at 50 °C (Figure 4.11). Ca^{2+} ions (0.15 M) halved the initial rate of synthesis but subsequently had a limited effect on hydrolysis. In contrast, Mg^{2+} ions barely effected the initial rate of synthesis but tended to promote hydrolysis, with ~40% loss over 48 h (Figure 4.11). Again this points to the differing roles of Mg^{2+} and Ca^{2+} in the cell, whereby Mg^{2+} promotes both the formation (acetyl phosphate synthesis) and reactivity (phosphorylation of ribose) of phosphorylated intermediates. Borate ions have previously been claimed to stabilise ribose (Ricardo et al., 2004) potentially favouring phosphorylation to ribose 5-phosphate (Mellersh & Smith, 2010) but had little effect in our hands (Figure 4.11).

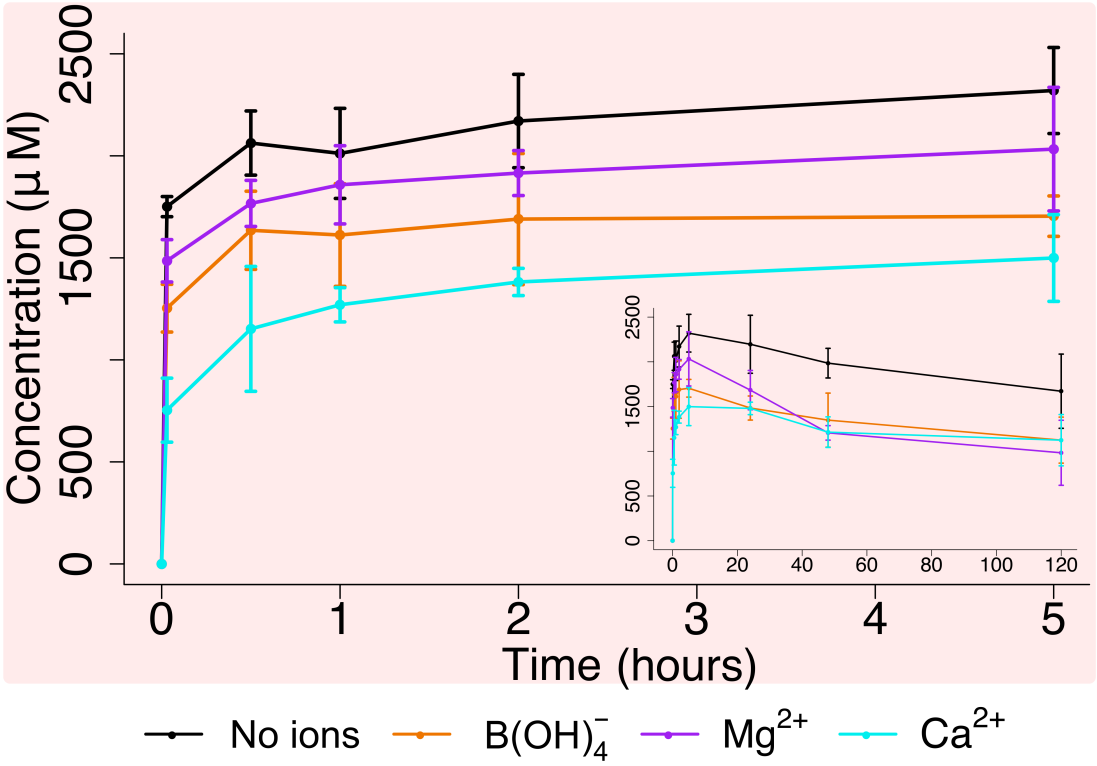


Figure 4.11. Synthesis of ribose phosphate from D-ribose and commercial AcP at pH 9 and 50 °C in the presence of different ions (0.15 M) (Whicher et al., 2018). An equivalent run without ions (black) is shown for comparison purposes. Graph insert shows the full reaction profile over 120 hours. N = 3 ±SD.

4.2.2.3 Unambiguous identification of ribose phosphate

The identification of ribose phosphate as a product of the reaction of ribose with AcP was initially achieved by comparison of the liquid chromatography retention time and mass spectrum of the sample to a ribose 5-phosphate commercial standard. The mass of derivatised ribose 5-phosphate was of 425.14 m/z (425.2 m/z when using the first analytical setup). The correspondence in virtually exact m/z values is shown Figure 4.6 between the derivatised commercial standard Figure 4.6c and the derivatised sample Figure 4.6d; secondary MS peaks at +1 m/z correspond to normal isotopic fractionation, with the relative height of the monoisotopic (main) mass peak and the +1 m/z peak depending on the number of carbon atoms of the molecule. When only D-ribose or only AcP was introduced under the same reaction conditions, no signal at 425.2 m/z developed but a signal at ~421.5 m/z was visible. This ion corresponded to a dimer of the 3-amino 9-ethylcarbazole derivatisation agent (210.27 g/mol x 2) + $1H^+$. Regardless, this ion would not interfere with the quantification of ribose phosphate as the scanning window was set to +/-1 m/z, therefore only counting ions from 424.2 to 426.2 m/z.

If, by chance, a molecule formed during the sugar derivatisation (a derivatisation artefact) presented the same molecular mass as the derivatised ribose phosphate, then this would interfere in the quantification of the actual sugar phosphate. One of the main reasons for the choice of a more accurate LC-MS instrument (Agilent Q-TOF) for the first experimental repetition (Figure 4.7) was to obtain a better separation of all reaction components. Figure 4.12 depicts LC-MS traces of a mixture of 1 mM of commercial standards of ribose 5-phosphate (Figure 4.12a) and fucose (which was used as the external standard in all samples and calibration curves) (Figure 4.12b). Their separation was not the best possible but sufficient for our tests; also no ion at 425.1472 m/z or closely similar ion was present in the fucose chromatogram, and vice versa. Most importantly, Figure 4.12c shows that the bulk of the derivatisation matter (whose most common ion fragment is 211.1230 m/z) was well separated chromatographically from the sugars, and thus did not interfere at all with their quantification.

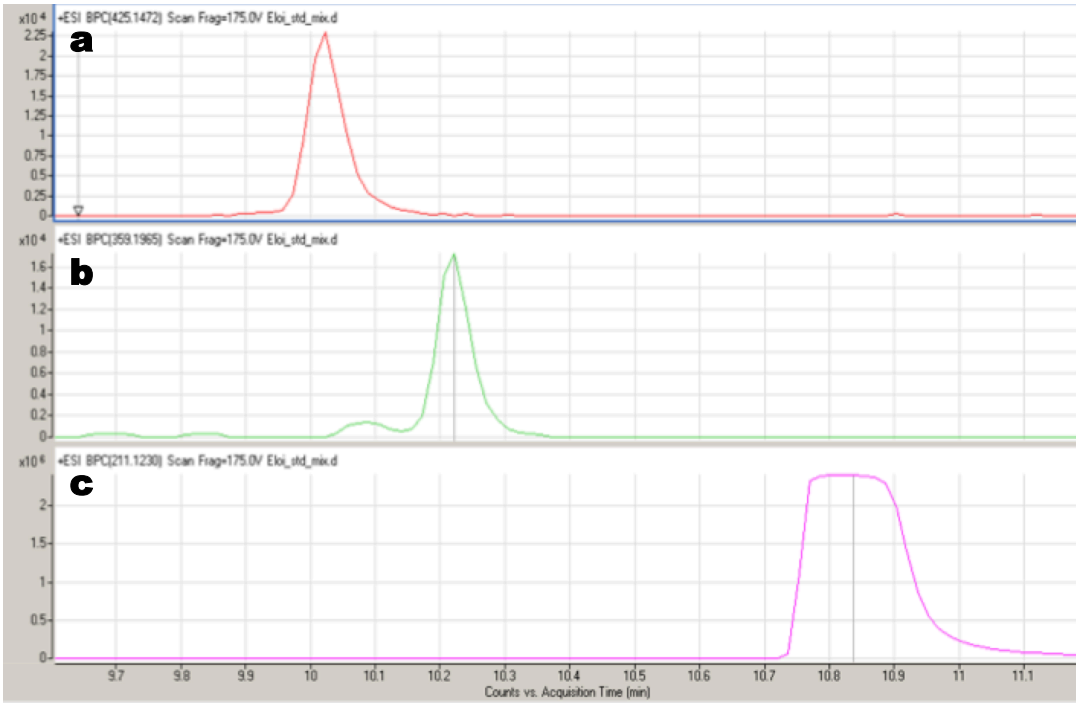
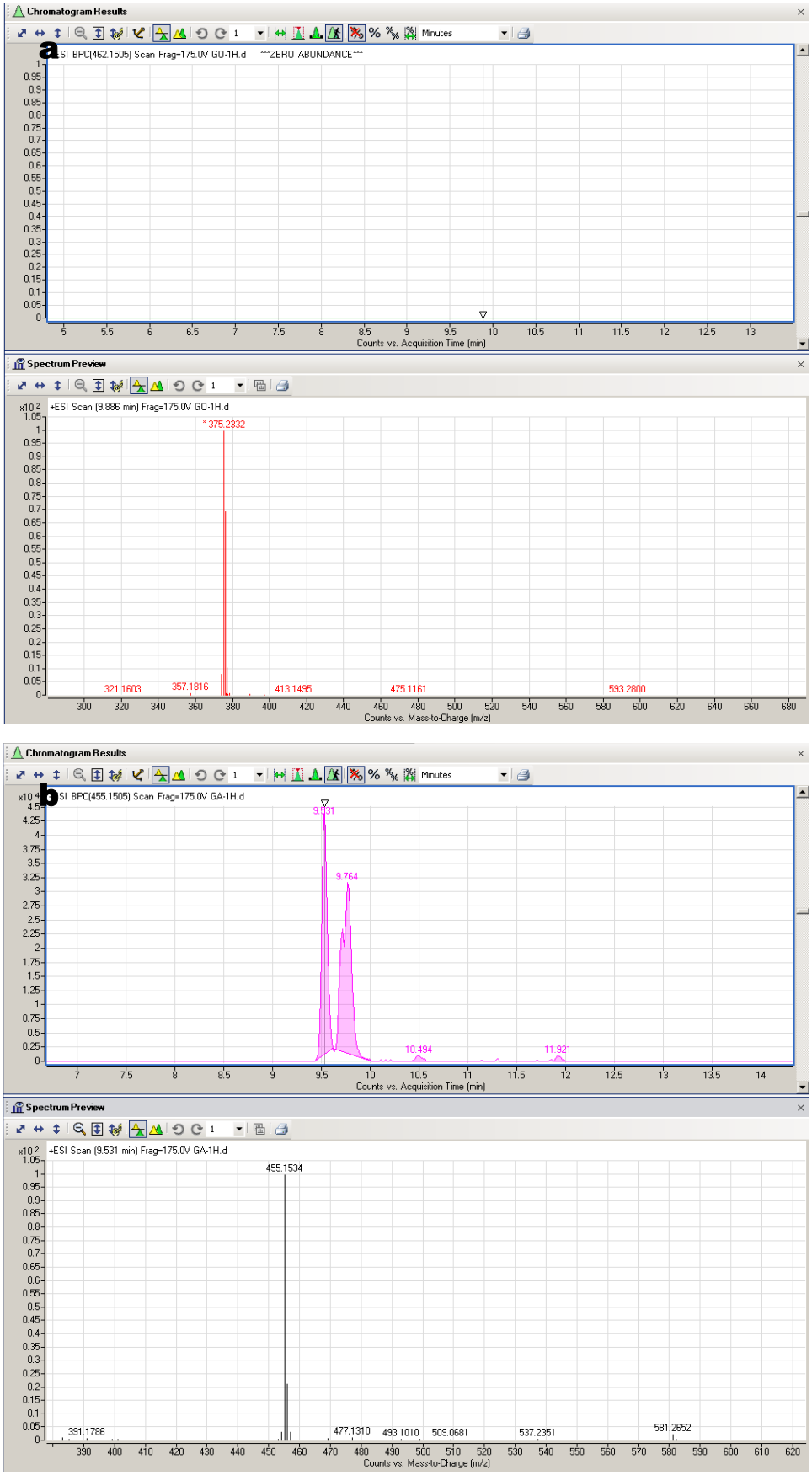
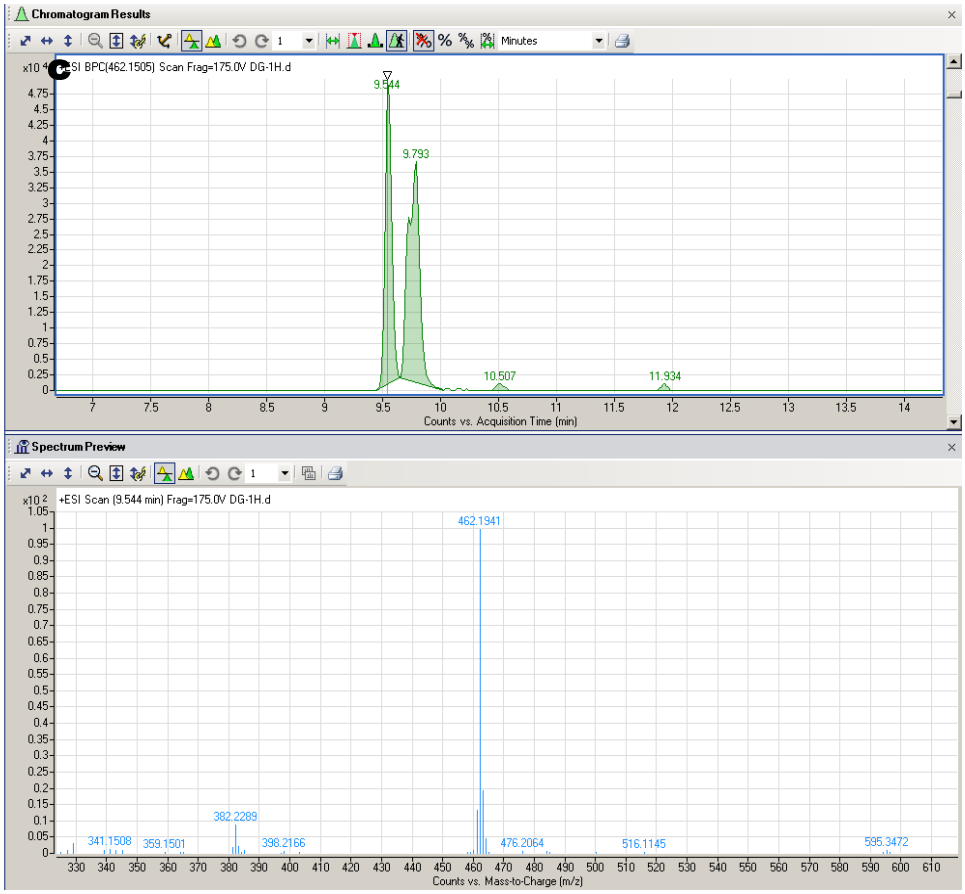


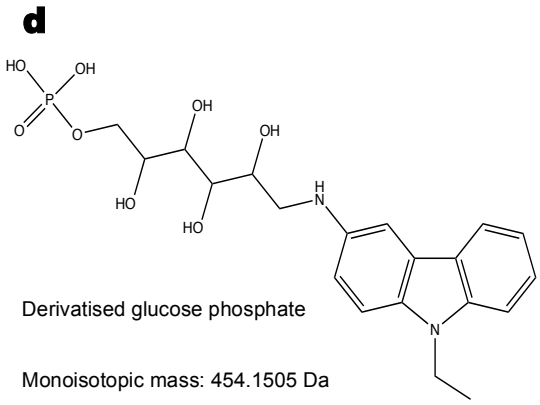
Figure 4.12. LC-MS traces of a derivatised 1 mM mixture of commercial standards of ribose-5-phosphate and fucose (internal standard) scanned for ions: (**a**) 425.1472 m/z (ribose-5-phosphate), (**b**) 359.1965 m/z (fucose) and (**c**) 211.1230 m/z (derivatisation reagent (AEC) precursor ion).

Further to a good chromatographic separation, a phosphorylation experiment using the same solution of commercial AcP (300 mM) and the same experimental conditions as in Figure 4.7 was performed. This time, instead of commercial ribose, AcP was incubated together with commercial glucose and, in another experiment, with 7-deuterated glucose (+7 Da). This way, if the 425.1472 m/z ion observed when AcP was incubated with ribose, which I attributed to ribose phosphate, was not ribose phosphate and instead was the product of another unknown reaction taking place, it would still appear when AcP was incubated with another sugar. If the 425.1472 m/z ion did not appear when AcP was incubated with glucose or deuterated glucose and instead I saw the expected mass of phosphorylated glucose or phosphorylated deuterated glucose, then I could say that indeed ribose was being phosphorylated in the first place, just as glucose and deuterated glucose were now. As expected, when glucose was incubated alone, no signal from its phosphorylation product (455.1505 m/z, which is 454.1505 m/z plus 1 H^+) appeared (Figure 4.13a). Also as expected, when glucose was incubated with AcP (pH 11, 50 °C, 1h), an ion

with the mass of phosphorylated glucose (455.1505 m/z) appeared (Figure 4.13b); as expected, no signal from 425.1472 m/z (phosphorylated ribose) was present. Furthermore, when 7-deuterated glucose was incubated with AcP, the corresponding phosphorylated ion +7 m/z appeared: 462.1505 m/z (Figure 4.13c); also, no signal from phosphorylated ribose was present either.


a
Chromatogram Results
+ESI BPC(462.1505) Scan Frag=175.0V G0-1H.d ***ZERO ABUNDANCE***
Minutes
Counts vs. Acquisition Time (min)
Spectrum Preview
+ESI Scan (9.886 min) Frag=175.0V G0-1H.d
* 375.2332
321.1603
357.1816
413.1495
475.1161
593.2800
Counts vs. Mass-to-Charge (m/z)
b
Chromatogram Results
+ESI BPC(455.1505) Scan Frag=175.0V GA-1H.d
9.531
9.764
10.494
11.921
Counts vs. Acquisition Time (min)
Spectrum Preview
+ESI Scan (9.531 min) Frag=175.0V GA-1H.d
455.1534
391.1786
477.1310
493.1010
509.0681
537.2351
581.2652
Counts vs. Mass-to-Charge (m/z)





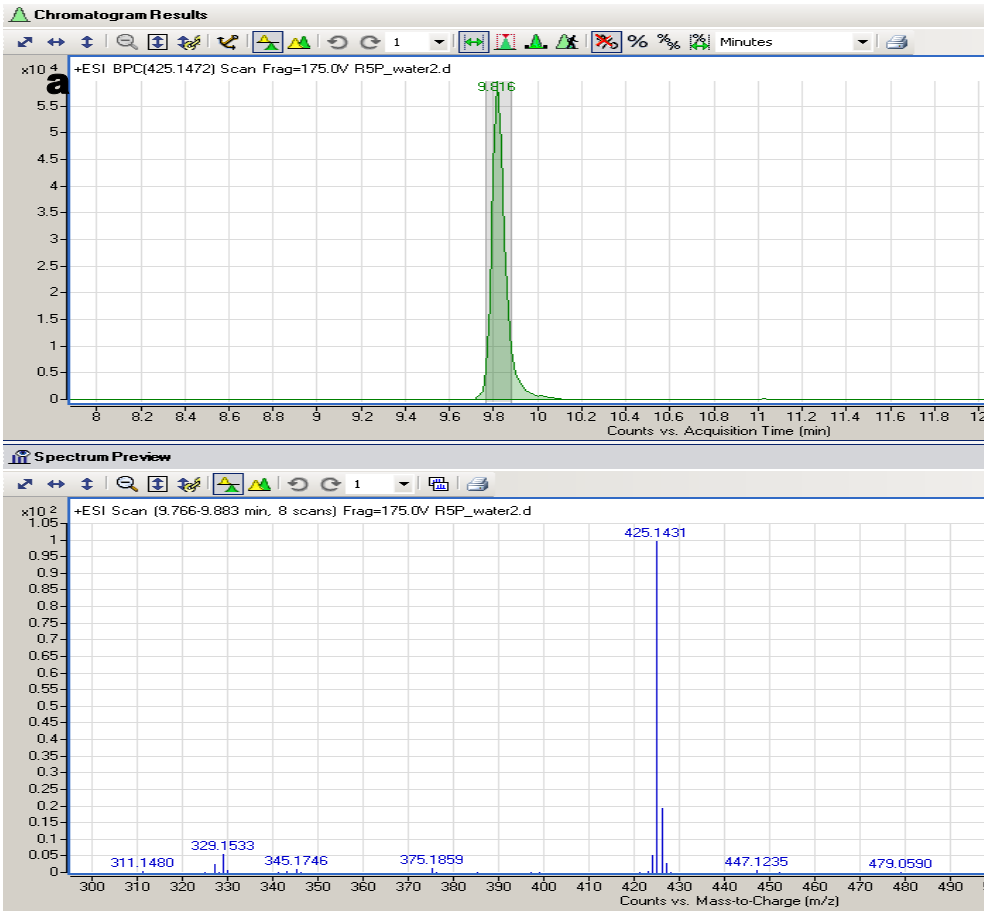
Derivatised glucose phosphate

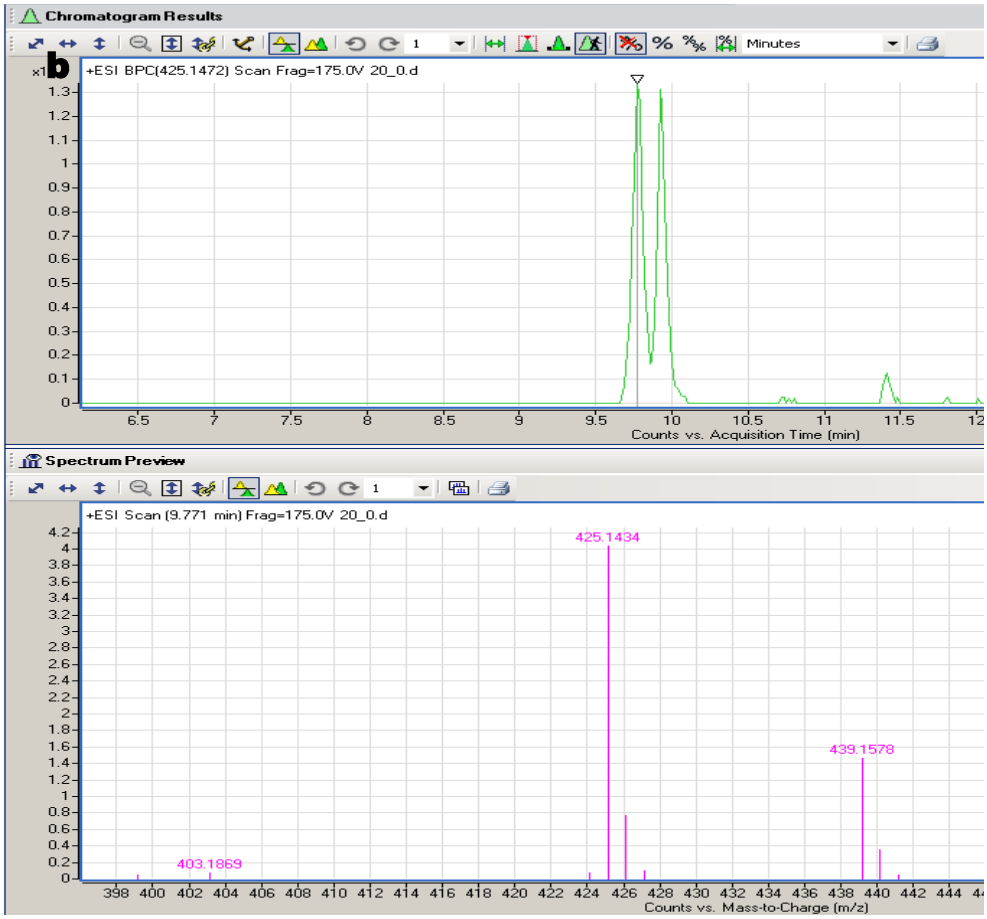
Monoisotopic mass: 454.1505 Da

Chemical Formula: $C_{20}H_{27}N_2O_8P$
Molecular Weight: 454.42

Figure 4.13. LC traces (upper cells) and mass spectra (lower cells) of reaction mixtures (pH 11, 50 °C, 1h) of: (**a**) glucose alone (300 mM), (**b**) glucose and AcP (both 300 mM) and (**c**) 7-deuterated glucose and AcP (both 300 mM). Chemical structure of derivatised (non-deuterated) glucose phosphate (**d**). All mass spectra shown (lower cells) are reconstructed ion chromatograms for each expected phosphorylated sugar monoisotopic ion mass: 455.1505 m/z (454.1505 (**d**) + $1H^+$) for (**a**) and (**b**); and 462.1505 m/z (455.1505 + 7 ^2H) for (**c**). Note the absence of phosphorylated sugar when there was only glucose in the sample (**a**).

All the previous experiments, coupled to the fact that AcP also mediates the abiotic phosphorylation of adenosine (to AMP) and adenosine diphosphate (to ATP) (Whicher et al., 2018, see Section 4.2.3), strongly suggest that ribose was also non-enzymatically phosphorylated by AcP. Interestingly, when the phosphorylation experiments were analysed with the second analytical setup (Agilent Q-TOF) the ribose (and glucose/deuterated glucose) phosphate peak appeared as a double peak (Figure 4.13 and 4.14), each sub-peak with the same mass as the ribose 5-phosphate commercial standard. This was surely happening when using the other analytical setups (as the experimental chemical reactions were identical) but, because the sugar chromatographic separation was worse previously (due to logistical constraints), all forms of phosphorylated ribose appeared as a single peak. This double sub-peak feature strongly suggests that more than one constitutional isomer of the phosphorylated sugar was being synthesised when incubated with AcP, which is not surprising given that there was more than one hydroxyl group from the sugar which could nucleophilically attack the phosphorus atom of AcP. Figure 4.15 depicts the proposed reaction mechanism. When quantifying ribose phosphate using the second analytical setup (data presented in Figure 4.7) the areas of both peaks were added as, undoubtedly, both represent ribose phosphate despite most probably representing different bond constitutions (constitutional isomers). For this purpose, the assumption that the calibration curves for each ribose phosphate isomer would be the same or very similar to the ribose 5-phosphate one, had to be made. Given the chemical similarities of the four potential constitutional isomers synthesised and the fact we are not dealing with enzymatic active sites, which in many cases can differentiate between isomers, it is reasonable to make this assumption. The lack of resolution between ribose phosphate isomers during quantification is the reason I specifically refer to ‘ribose phosphate’ (instead of ‘ribose 5-phosphate’) when dealing with the experimentally synthesised version throughout this document; despite the fact that ribose 5-phosphate is almost undoubtedly the most abundantly produced isoform of them all (as discussed in the next sub-Section).


Chromatogram Results
a
+ESI BPC(425.1472) Scan Frag=175.0V R5P_water2.d
9.816
Counts vs. Acquisition Time (min)
Spectrum Preview
+ESI Scan (9.766-9.883 min, 8 scans) Frag=175.0V R5P_water2.d
425.1431
329.1533
311.1480
345.1746
375.1859
447.1235
479.0590
Counts vs. Mass-to-Charge (m/z)


Chromatogram Results
b
+ESI BPC(425.1472) Scan Frag=175.0V 20_0.d
Counts vs. Acquisition Time (min)
Spectrum Preview
+ESI Scan (9.771 min) Frag=175.0V 20_0.d
425.1434
439.1578
403.1869
Counts vs. Mass-to-Charge (m/z)

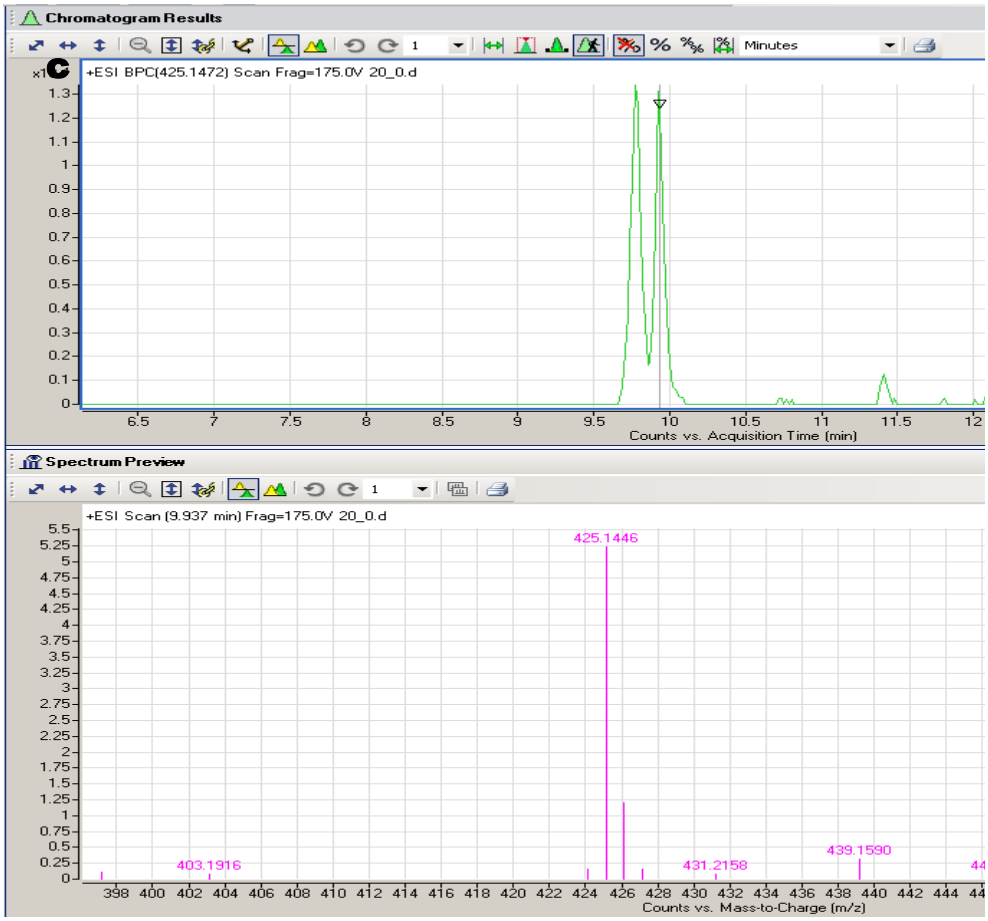


Figure 4.14. LC traces and mass spectra of (**a**) ribose 5-phosphate commercial standard, (**b**) experimental ribose phosphate (first peak), and (**c**) experimental ribose phosphate (second peak). Experimental conditions of (**b**) and (**c**) are: pH 11, 50 °C and 1 h of incubation time. Other small MS peaks appearing in the experimental spectra are due to the imperfect chromatographic separation and background contamination. Note that only the commercial phosphorylated sugar standard (**a**) lacks the double sub-peak feature, since it is a pure monoisomeric preparation.

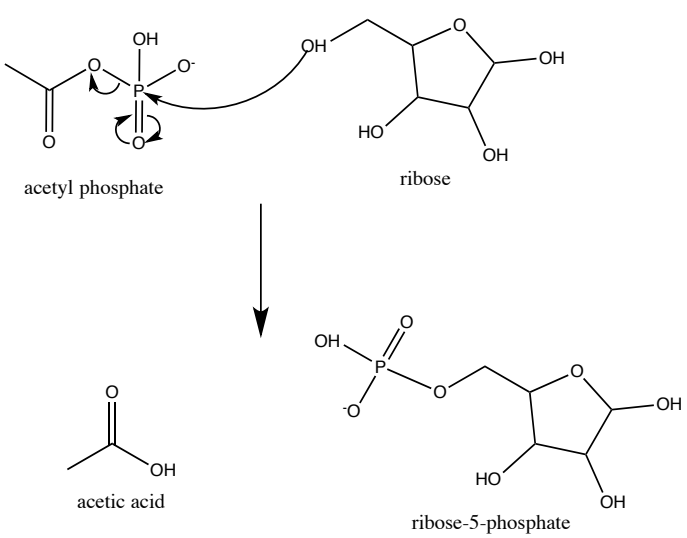


Figure 4.15. Proposed reaction mechanism of the non-enzymatic phosphorylation of ribose by AcP. Any other hydroxyl group from ribose can be the nuclophilic attacker onto the phosphorus atom of AcP, resulting in the 4 possible different ribose phosphate constitutional isomers (ribose 1/2/3/5-phosphate).

4.2.2.4 Competing sites for ribose phosphorylation

As mentioned earlier, the experimentally synthesised ribose phosphate could have been phosphorylated in several positions generating: ribose 5-phosphate, ribose 1-phosphate, ribose 2-phosphate and ribose 3-phosphate. The 4' carbon of ribose could not be phosphorylated due to its lack of a hydroxyl group after the cyclation of the sugar. If synthesised by our studied reaction, any ribose 1-phosphate would not be detectable using this derivatisation method, as a bond from the anomeric carbon from the sugar (in its aldehyde form) must be created with the derivatisation agent, and it would not be free if it was already phosphorylated. In the unlikely event of the derivatising bond being able to form when the initial substrate was ribose 1-phosphate, the final derivatisation product would not have weighed 425.2 m/z anyway, as it would have exchanged with the phosphate group (it would have weighed 345.2 m/z just like naked ribose). This way, it was not possible for any putative ribose 1-phosphate experimentally synthesised to be quantified as ribose 5-phosphate (the main sugar of interest).

So, the three possible molecules giving rise to the 425.2 m/z signal could have been ribose 5-phopshate, ribose 2-phosphate and ribose 3-phosphate. Another possibility is that the secondary peak corresponded to a cyclic ribose 2,3-phosphate; the putative formation of this cyclic isomer would be interesting as it is deemed as an activated monomer, ready for further otherwise unfavoured reactions, such as polymerisations (Powner et al., 2009). HMBC (heteronuclear multiple bond correlation) 2-dimensional ^{31}P /^{1}H NMR was used to try determine the phosphorylation sites of the synthesised ribose phosphate. Unfortunately, as shown in Figure 4.16, the signal/noise ratio of the results was too small to be informative. Despite the fact that generally speaking phosphorus NMR is more sensitive than carbon NMR, concentrations of analyte below 1 mM are challenging to robustly quantify. This coupled to the measurable amount of sensitivity being lost when switching from 1-D to 2-D NMR, accounts for the low signal/noise ratio obtained. Furthermore, if more than one of the ribose phosphate isoforms was present in the sample, this would have divided the signal into two or three different couplings, thus lowering the overall signal

intensity even more. The validity of this last factor in contributing to the poor results of this 2-D test was supported by the visualisation of a double peak when using the Agilent Q-TOF analytical setup (Figure 4.14) suggesting that several of the ribose hydroxyl groups are good candidates for the AcP-facilitated phosphorylation. A still better chromatographic separation of all sugars (and derivatisation material) could theoretically generate non-overlapping peaks, which could be exploited by future analyses using MS-MS in order to further elucidate the phosphorylation positions. Unfortunately, due to time constraints and numerous analytical instrument setbacks, these further analyses were not carried out; this issue is discussed further in the Conclusions and future work Section (7.1) of Chapter 7.

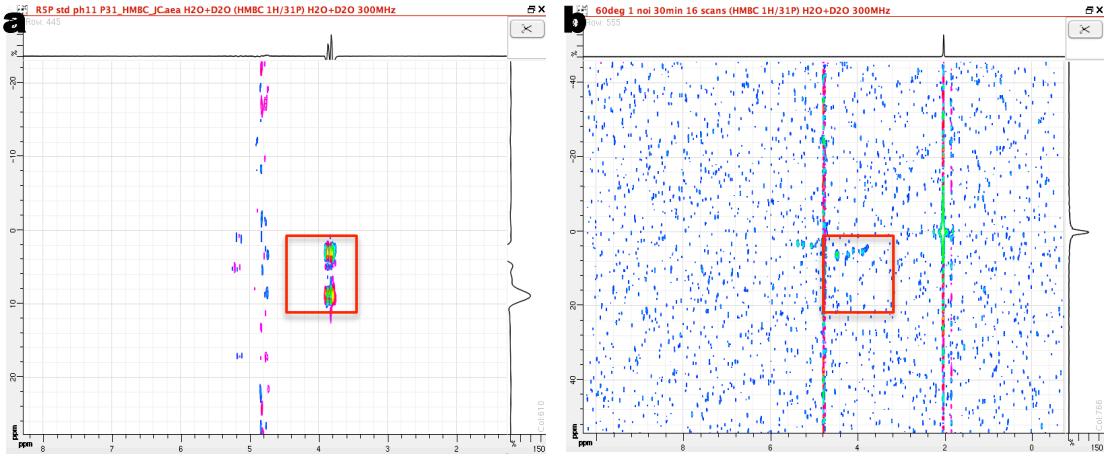


Figure 4.16. HMBC 2-D $^{31}P/^{1}H$ NMR of (**a**) D-ribose 5-phosphate commercial standard, and (**b**) the reaction product of D-ribose and AcP at 60 °C, pH 11 and after 30 min of reaction. The red square marks the area where ribose phosphates yield signal. The signal arising from (**b**) was too weak to calculate the corresponding J couplings. The strong signal arising at 2 ppm (x axis) in (**b**) corresponds to AcP.

4.2.3 Phosphorylation of other nucleotide precursors by acetyl phosphate

Apart from the phosphorylation of ribose and due to their relevance for nucleotide synthesis, we also considered whether AcP could phosphorylate other nucleotide precursors under a range of alkaline hydrothermal conditions.

In this Section I will consider the synthesis of adenosine monophosphate (AMP) from adenosine, and ATP from ADP. The experimentation on AcP-mediated phosphorylation of AMP from adenosine was performed independently by Dr Alexandra Whicher, whilst the AcP-mediated phosphorylation of ATP from ADP was performed under my supervision by Silvana Pinna.

4.2.3.1 Phosphorylation of adenosine to AMP by AcP

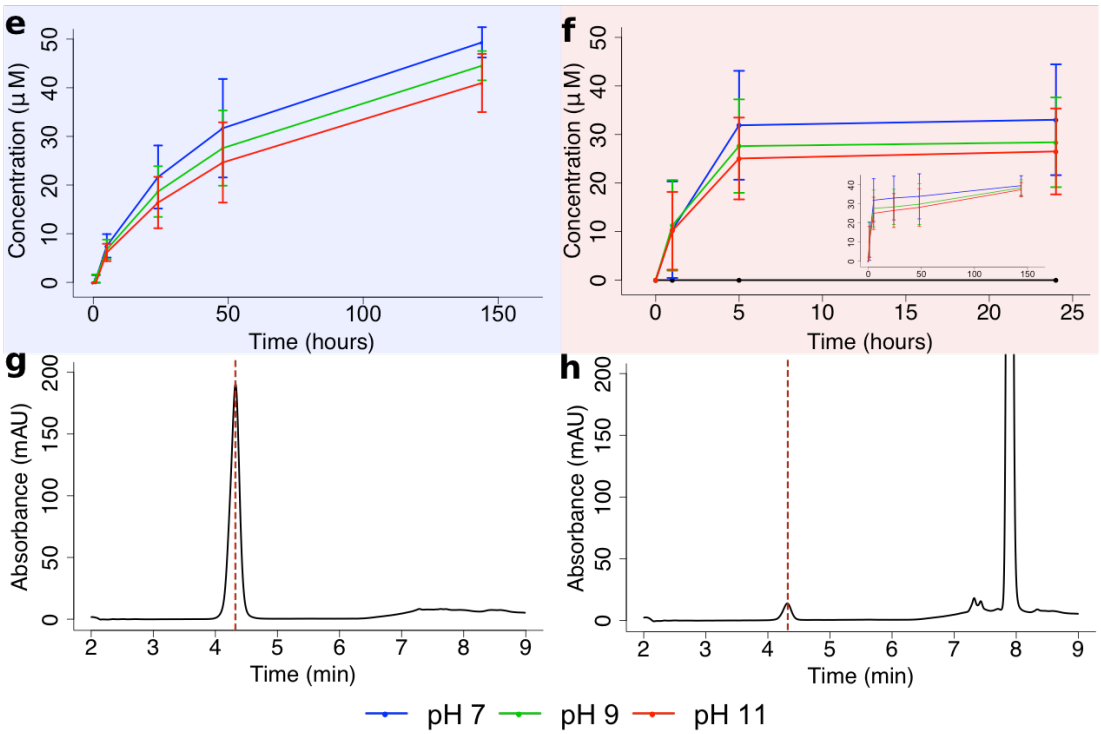


Figure 4.17. (**e**–**f**) Phosphorylation of adenosine to AMP by commercial AcP (Whicher et al., 2018) at pH 7, 9 and 11; at (**e**) 20 °C, and (**f**) 50 °C. (**g**–**h**) HPLC-UV chromatograms for detection of AMP at 4.3 min at 254 nm of UV wavelength: (**g**) commercial standard and (**h**) experimental sample. Graph insert shows full reaction profile over 144 h. The large peak at 7.9 min is adenosine. N=3 ±SD. Data provided by Dr Alexandra Whicher.

The phosphorylation of adenosine to AMP by AcP was markedly slower than that of ribose, taking several hours even at 50 °C (Figure 4.17e–h), but ultimately achieving a similar yield (relative to adenosine) of ~2%. These differences may reflect steric or chemical hindrances by the nucleobase, or the reaction conditions, which differed due to the limited solubility of adenosine

(maximum solubility in water at 20 °C ~10 mM). Again, the concentration of AMP synthesised at pH 11 was slightly lower; but in general the pH had little effect on phosphorylation in the case of either ribose phosphate or AMP. Nor did the presence or absence of Mg^{2+} or Fe^{2+} ions under anaerobic or aerobic conditions (Figure 4.18). Although the rate of reaction was slower in this case compared with ribose phosphate, all these phosphorylations took place within minutes to hours, rather than days to weeks as commonly reported (Chung et al., 1971; Lohrmann & Orgel, 1968). These relatively rapid reaction rates are more commensurate with cellular biochemistry, as catalysed by enzymes, than with much slower geological processes, and so offer a first step towards the far-from-equilibrium dynamics of living cells.

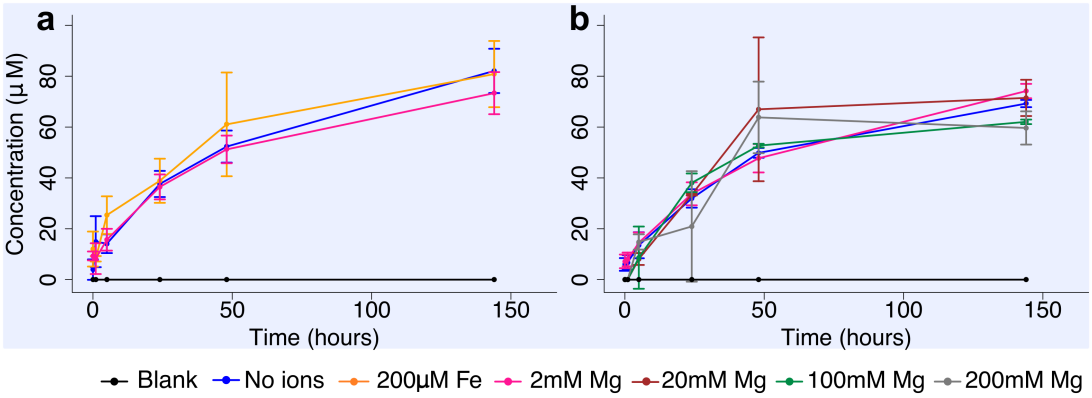


Figure 4.18. Phosphorylation of adenosine to AMP by AcP under anaerobic conditions with the addition of metal ions (Fe^{2+} and Mg^{2+}). Experiments at pH 7 and 20 °C (Whicher et al., 2018). (**a**) Anaerobic conditions with Fe^{2+} at 200 µM and Mg^{2+} ions at 2 mM. (**b**) Aerobic conditions with Mg^{2+} ions at 2, 20, 100 and 200 mM. In both experiments phosphorylations did not occur in the absence of AcP (black lines). N = 3 ±SD. Data provided by Dr Alexandra Whicher.

4.2.3.2 Phosphorylation of ADP to ATP by AcP

AcP could also phosphorylate ADP to ATP in water, but we were unable to quantify the ATP yields using ^{31}P–NMR, as the concentrations were close to the limits of detection of our instrument (Figure 4.19).

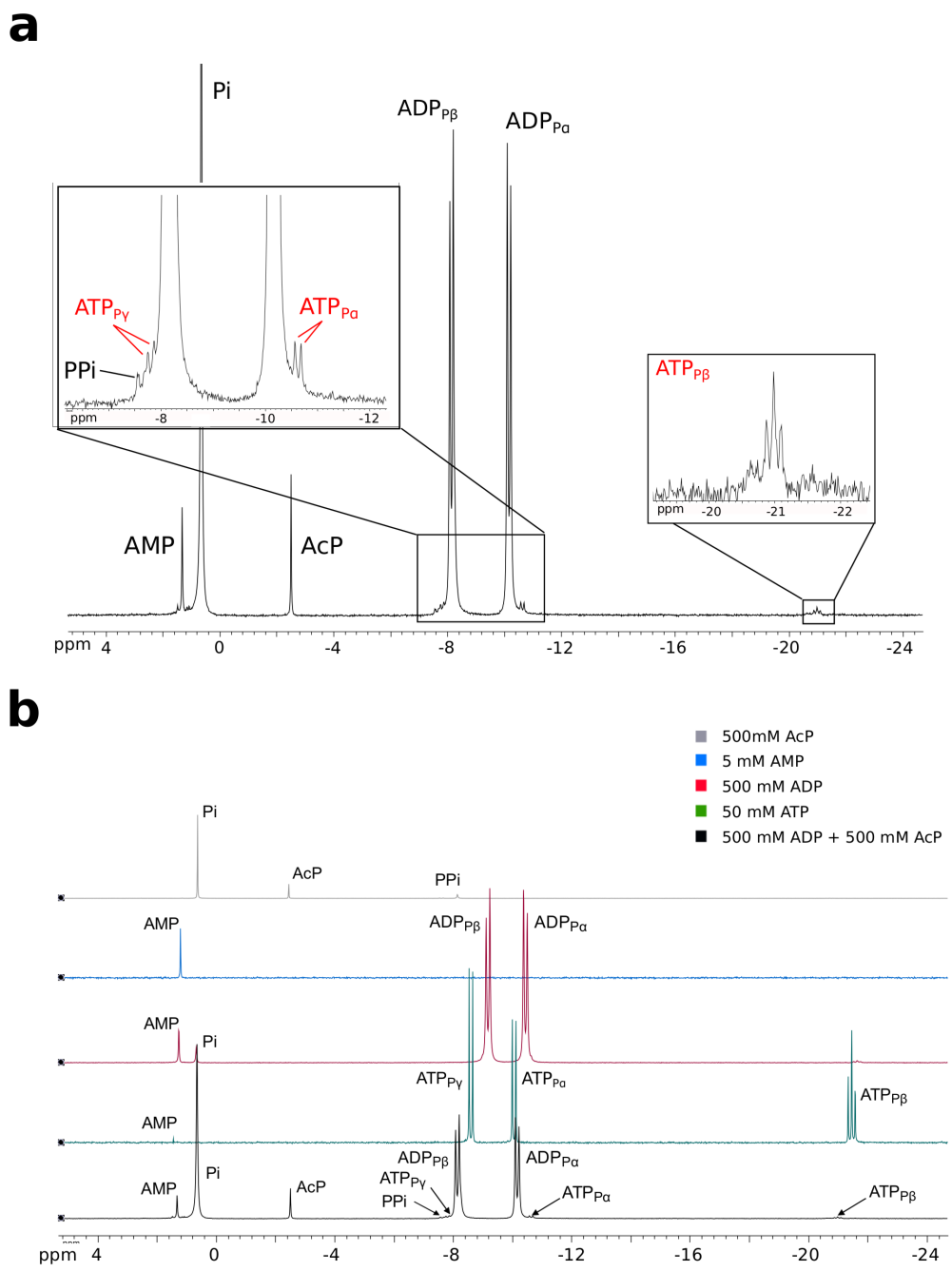


Figure 4.19. Phosphorylation of ADP by AcP (Whicher et al., 2018). (**a**) ^{31}P–NMR spectrum of the phosphorylation of ADP by AcP at 50 °C and pH 5.5 after 2 h, yielding ATP. All peaks labelled for clarity. Graph inserts show zoomed in areas of ATP signal. (**b**) ^{31}P–NMR spectra of the phosphorylation of ADP by AcP, compared with commercial ATP, commercial ADP, commercial AMP, and commercial AcP. Small pH differences account for the imperfect alignment of the peaks between −8 and −10 ppm. Data provided by Silvana Pinna under my supervision.

ATP could not be detected over 24 h at 20 °C, but ATP synthesis was detected after 1–5 h by both ^{31}P–NMR (Figure 4.19) and HPLC (Figure 4.20) at 50 °C, again suggesting reactivity over time periods of minutes to hours. Our group is currently developing a more sensitive HPLC analysis for ATP synthesis in the near future (this issue is discussed further in the Conclusions and future work Section (7.1) of Chapter 7), so simply report its successful synthesis here. Our findings corroborate those of Kitani et al. (1995) who reported the synthesis of ATP from ADP using AcP, especially when catalysed by Fe^{3+}. Unlike Kitani et al. we found higher yields at higher temperatures.

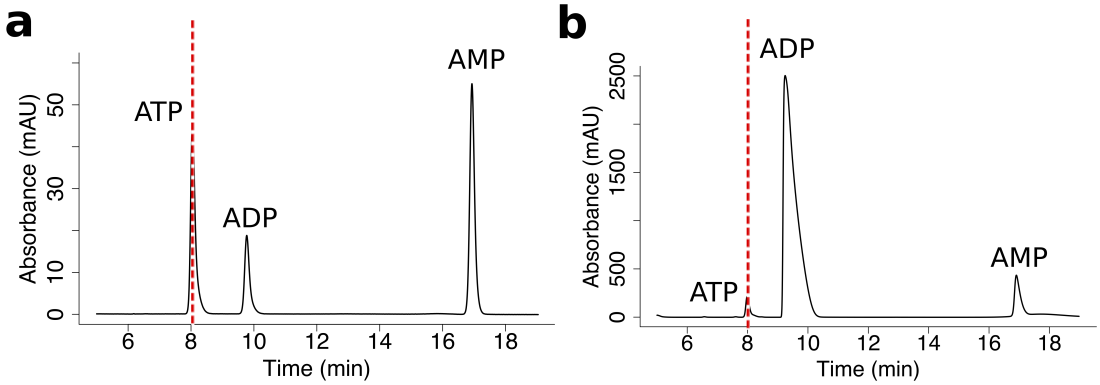


Figure 4.20. HPLC-UV chromatograms of (**a**) commercial ATP, ADP and AMP; and (**b**) experimental phosphorylation of ADP by AcP (Whicher et al, 2018). Data provided by Silvana Pinna under my supervision.

4.3. Discussion

While the yields of phosphorylated products were generally low (~2%), as mentioned in Section 3.2.2 of Chapter 3, I do not consider that to be a problem; rather the reverse. As argued by de Duve (2005), the 'only scientifically plausible' explanation for the emergence of biological catalysts, whether ribozymes or enzymes, is selection. The first biological catalysts must have been selected because they enhanced flux through proto-metabolic pathways. These proto-metabolic pathways presumably had much the same substrates and products, and probably many equivalent intermediates, as later genetically encoded pathways (Copley et al., 2007; Keller et al., 2017a; Martin et al., 2014;

Sojo et al., 2016). Tardy reactions should have been promoted by mechanisms equivalent to those still observed in modern cells, such as phosphorylation. That is what we did find. AcP is formed under mild prebiotic conditions; it is capable of phosphorylating biologically meaningful molecules, including ADP to ATP; and it is still used by cells today, critically as the fulcrum between thioester and phosphate metabolism (de Duve, 1991; Ferry & House, 2006; Schönheit et al., 2016). If metabolism did indeed emerge through selection, then low yields of intermediates and products would be expected to form in water in the absence of catalysts. Work on prebiotic forms of glycolysis and the pentose phosphate pathway (Keller et al., 2014), Krebs cycle (Keller et al., 2017b) and gluconeogenesis (Messner et al., 2017) suggests that the intermediates do indeed form spontaneously under prebiotic conditions. These intermediates are necessarily relatively reactive, or they would accumulate and could not be part of metabolic networks (Keller, et al., 2017; Smith & Morowitz, 2004).

The first steps of catalysis should be promoted by ions or mineral clusters equivalent to the cofactors of modern proteins. Simple ions in solution should improve yields slightly. Accordingly, we found that Mg^{2+} promoted the synthesis of AcP from thioacetate relative to other ions. That is important because Mg^{2+} is unusual in being able to complex with two phosphate groups (Holm, 2012); it is still required in cells today to maintain the structure of RNA and enzymes, as well as the activity of phosphorylated molecules, including ATP. Chelation of ions or small mineral crystals (such as FeS clusters) by amino acids or short polypeptides should enhance catalysis further, as these complexes ought to mimic the active site of enzymes for reasons of physical chemistry, increasing yields further (Copley et al., 2007; Milner-White & Russell, 2005; Nitschke & Russell, 2013; West et al., 2017). Ultimately, genetically encoded ribozymes or enzymes would increase yields considerably more. In other words, the explicit prediction of a selectionist approach to the origin of life is that the yields of prebiotic reactions should be low in the absence of organic catalysts in water. From this perspective, the synthetic chemists' preoccupation with high prebiotic yields of specific products is misguided.

The phosphorylation reactions reported here required high concentrations of reagents, and could be criticised for being prebiotically implausible (McCollom, 2013). While the concentrations we used are similar to those reported for most other studies (Bujdák & Rode, 1996; Burcar et al., 2015; Chung et al., 1971; Costanzo et al., 2007; DeGuzman et al., 2014; Huber et al., 2003; Leman et al., 2004; Liu & Orgel, 1997; Lohrmann & Orgel, 1968; Morasch et al., 2014; Rabinowitz, 1970; Yamagata et al., 1991), three factors make these high concentrations more realistic in alkaline hydrothermal systems. First, deep-sea hydrothermal systems are at high pressure (100– 300 bars), which increases the solubility, and so concentration, of gases such as H_2. Pressure also increases the likelihood of molecular collisions, which is equivalent to increasing the concentration of reactants or raising the temperature. Continuous hydrothermal flux sustains the critical disequilibria in H_2, CO_2 and proton concentrations, and should theoretically drive organic synthesis, including formation of thioacetate and acetyl phosphate, under these conditions (Amend & Shock, 1998; Amend et al., 2013; Shock & Canovas, 2010; Sojo et al., 2016). Second, as mentioned previously, thermal gradients and convection currents in alkaline hydrothermal vents can concentrate organics by thermophoresis. Steep heat gradients produced by lasers in closed glass capillaries concentrate nucleotides and small RNAs by up to 10^3–10^{12}-fold (Baaske et al., 2007; Braun & Libchaber, 2002; Mast et al., 2013; Mast & Braun, 2010). In simulated hydrothermal systems, we have shown that thermophoresis across an open, inert microporous matrix can concentrate small organics equivalent to nucleotides by at least 5000-fold (Herschy et al., 2014). Third, the charge on phosphorylated biomolecules such as acetyl phosphate makes them more likely to interact with mineral surfaces within hydrothermal vents, promoting their retention in vent pores rather than their discharge from the vent along with uncharged ‘waste’ molecules such as methane (Pratt, 2006; Mellersh & Smith, 2010; Westheimer, 1987). Similarly, if monomers were formed inside fatty acid vesicles, with ‘autotrophic’ growth driven by geologically sustained gradients across the membranes (West et al., 2017), molecular crowding and membrane surface associations inside growing protocells would also increase local concentrations. These three factors mean that the concentrations used in the

work presented here (as well as in the other Chapters), while admittedly high, are not necessarily unreasonable.

4.4 Conclusions

The work in this Chapter shows that AcP is formed at modest (~2%) yields within minutes under ambient conditions (neutral pH and 20 °C) and mild hydrothermal (pH 8, 50 °C) conditions, from the simple 2-carbon precursor thioacetate. The formation of AcP from the simple thioester methyl thioacetate was not detected under any of the conditions tested. While most theoretical work (de Duve, 1991; Martin & Russell, 2007) has considered methyl thioacetate as a prebiotic equivalent to acetyl CoA, thioacetate has also been synthesised from CO and CH_3SH under mild hydrothermal conditions (Huber & Wächtershäuser, 1997). Being even simpler, thioacetate is arguably a more plausible prebiotic precursor, if not a thioester. Once formed, AcP exhibits an ideal poise between stability and reactivity: it is stable over >5 h under ambient conditions, even at pH 11 and in the presence of Mg^{2+} and Ca^{2+} ions. Earlier work showed that AcP is hydrolysed much more rapidly above pH 11 (Etaix & Buvet, 1974; Koshland, 1952), but those conditions would be rare even in alkaline hydrothermal systems (Kelley et al., 2001, 2005), so these were not studied here. AcP is less stable at warmer temperatures: it is completely hydrolysed over 3–5 h at 50 °C, and within 90 min at 60 °C.

It has also been shown that AcP does indeed act as an ATP mimetic, phosphorylating several intermediates in nucleotide synthesis, in water, over minutes to hours: ribose to ribose phosphate, adenosine to AMP, and ADP to ATP. AcP also is shown to act as an acetylating agent on the hydroxyl groups of ribose at alkaline pH. The acetylating properties of AcP are discussed in detail in the next Chapter, so this will not be discussed further here.

In conclusion, AcP is readily synthesised under ambient and mild hydrothermal conditions, and promotes phosphorylation reactions in water, without catalysts, under a wide range of conditions. Most of the conclusions of this Chapter apply

equally to other prebiotic environments, and do not necessarily need to be interpreted in a hydrothermal context. Unlike the majority of chemical pathways to biomolecules proposed so far in abiogenesis research, the fast reaction rates shown throughout this Chapter are more commensurate with cellular biochemistry, catalysed by enzymes, than with slower geological processes. Therefore our proposed pathways offer a possible first step towards the far-from-equilibrium dynamics of living cells. Furthermore AcP is still used by bacteria (Thauer et al., 1977). These findings coupled with what is shown in Chapter 5 (exploring the condensing properties of AcP) suggest that AcP constitutes a credible primordial energy currency, coupling carbon and energy flux at the origin of life.

Chapter 5

Acetyl phosphate as a primordial condensing agent

5.1 Introduction

The non-enzymatic phosphorylation capabilities of acetyl phosphate have been discussed in the previous Chapter. This Chapter focuses on whether these AcP-mediated phosphorylations could promote the respective non-enzymatic condensation reactions of nucleotides and amino acids in water. In modern cells, the condensation of amino acids and nucleotides is achieved by their phosphorylation; either directly, as in the case of nucleotide monophosphates, or, indirectly by adenylation (through a phosphate group) in the case of amino acids to be integrated onto the aminoacyl-tRNA complex. Abiotically, some limited phosphorylation can be achieved by phosphate minerals and other compounds, including ATP, in the absence of enzymes (Burcar et al., 2016; Gibard et al., 2017; Gull et al., 2015; Costanzo et al., 2007; Saygin, 1983; Schieven & Martin, 1988; Vasant et al., 2005). As shown in the previous Chapter, AcP is a strong abiotic phosphorylating reagent, is much simpler than ATP, and is synthesized from the reaction of thiolated compounds (such as acetyl CoA biotically or simpler thioesters abiotically) with inorganic phosphate (de Duve, 1991). As predicted theoretically, in our lab we have succeeded in synthesising AcP from thioacetate and inorganic phosphate under simulated alkaline hydrothermal vent conditions (Figure 4.2) (Whicher et al., 2018). Thiolated compounds, such as thioacetate, could have been synthesized in alkaline hydrothermal vents from CO and CH_3SH using FeS minerals as catalysts (Huber & Wächtershäuser, 1997), or from CO_2, H_2 and SH^-/CH_3S^- using Fe(Ni)S minerals as catalysts (Figure 3.1).

Focusing specifically on the abiotic polymerisation of nucleotides, there have been many attempts to answer this question from a heterotrophic origins point of view. Most authors propose that either life started somewhere with little water

(e.g. Mars) or through dehydrating mechanisms (e.g. tidal beaches or low water activity environments (Burcar et al., 2015)) whilst others opt for primordial solvents other than water (e.g. formamide) (Saladino et al., 2009) which eliminate the product inhibition problem of having a dehydration in water altogether. Whilst this approach has generated some promising results, according to de Duve's congruence principle it also widens the gap between geochemistry and biochemistry. As in the rest of this thesis, following de Duve's congruence principle I assume water to be the initial solvent of life; alkaline hydrothermal vents mostly occur on the ocean floor after all. The proposed solutions to this problem from the perspective of autotrophic origins generally consists of using previously activated monomers (e.g. imidazole-activated nucleotides (Burcar et al., 2015) or cyclic nucleotides (Costanzo et al., 2009)) which are in effect pre-condensed versions of the monomer, and thus facilitate further condensations to form the polymer. The main problem with this approach is that the way of obtaining the condensed monomers remains unclear.

The literature on prebiotic condensation of amino acids has also yielded some interesting positive results (Chen & Yang, 2007; Huber & Wächtershäuser, 1998; Le Son et al., 1998; Martra et al., 2014), but suffers from a similar problem whereby most proposed mechanisms (e.g. sulphur trioxide/salt-induced/oxide surface-mediated condensations) do not remotely resemble ancient life (based on any broad reconstruction of LUCA). This disconnect between prebiotic geochemistry and ancient biochemistry is, in my view, one of the most crippling issues that the research on the origins of life currently faces. Our approach is to propose a prebiotic reactivity that resembles an analogous (but simpler) core metabolism of all life. This approach somewhat limits the reactivity landscape (e.g. using water as the only solvent available), but offers testable predictions and, if successful, could being to reconnect geo- and biochemistry.

Therefore, I anticipated that AcP would promote the condensation of nucleotides and simple amino acids under aqueous conditions, albeit producing small oligomers in low yields. Higher yields would be expected in future

experimentation when adding chelated ions (e.g. Mg^{2+} chelated by aspartate oligomers) or mineral surfaces (e.g. the $Mg(OH)_2$ mineral brucite), as these conditions would roughly mimic the predicted natural evolution steps occurring in vents: from unspecific and low-yielding reactions in an aqueous medium, to highly productive and stereospecific reactions occurring at the active centres of enzymes or ribozymes (in this case the Mg^{2+}-dependent RNA polymerase).

5.2 Results

5.2.1 AcP-mediated polymerisation of adenosine monophosphate

5.2.1.1 Perspective

Previous experiments within our lab suggested that AcP favoured the condensation of glycine into the cyclic peptide diketopiperazine at high yield under simulated alkaline hydrothermal vent conditions. This reaction generated a condensed cyclic form of diglycine (diketopiperazine or DKP) with yields up to 80%. Peer reviewers raised critical questions on this work and after careful assessment I concluded that what we thought was DKP, was in reality N-acetylated glycine rather than condensed glycine; more on this is detailed later in this Chapter (Section 5.2.2). I decided to test the condensing activity of AcP on nucleotides, encouraged by the apparent success on AcP-mediated amino acid condensation. Successive adenosine monophosphate (AMP) (monohydrate, ≥97%, Sigma) polymerisation attempts were attempted. AMP was chosen as our monomer as it has been the most successful of all nucleotides in polymerisation tests (Burcar et al., 2015), presumably due to the purine stacking capacity which may also favour its polymerisation (Costanzo et al., 2009). Acetyl phosphate was tested as a condensing agent as it is known to phosphorylate several organic molecules on hydroxyl groups, including ribose as described earlier; this phenomenon could potentially favour the subsequent polymerisation of the phosphorylated monomers in water. This activating effect mainly occurs (i) as the phosphate group is generally a good leaving group in S_N2 reaction mechanisms (since charges in the generated inorganic (poly)phosphates are stabilised due to their multiple resonant structures); and

(ii) to the fact that phosphorylated molecules are already a condensed form of the monomer so that the subsequent condensation with another monomer would, in theory, represent a lower energetic jump (Chung et al., 1971; Hulshof & Ponnamperuma, 1976).

Three general reaction mechanisms for the AcP-mediated AMP polymerisation, depicted in Figure 5.1, were conceivable at the start of this project. Figure 5.1a shows the first possible reaction mechanism where a molecule of AMP would generate a nucleophilic attack from its 3'-OH onto the α-phosphate of a molecule of ADP, which in turn would have been synthesised in situ due to an AcP-mediated phosphorylation on the 5'-phosphate of an AMP molecule. This potential AcP-mediated phosphorylation onto the 5'-phosphate of an AMP molecule is depicted in the first reaction step (top horizontal arrow) of Figure 5.1b. This putative first reaction is in line with the results showing AcP-mediated phosphorylation of ADP to ATP (Section 4.2.3 of Chapter 4). Figure 5.1b depicts a second potential reaction pathway onto a dinucleotide: once the potential ADP molecule was synthesised, it could undergo an intramolecular cyclation generating cyclic AMP (cAMP). This cyclic derivative of AMP, very common in biology as a secondary messenger, represents a more condensed version of the molecule, which can be hydrolytically opened again generating a metastable ionised intermediate; its charged hydroxyl group could act as a good nucleophile and thus attack the phosphate of another molecule of AMP, resulting in a dehydrating condensation. Finally, we envisioned the possibility of AcP facilitating the condensation of AMP molecules by its reaction with the leaving group (a hydroxide ion) of the condensation (Figure 5.1c). The presence of a high-energy phosphoryl donor in the reaction mixture could add a dynamism whose implications are difficult to predict theoretically. For example, it is conceivable that the 3' OH of a molecule of AMP could perform a nucleophilic attack onto the α-phosphate of another molecule of AMP: normally the leaving group would be a OH^- (a bad leaving group at high pH), but in the presence of AcP the free OH^- could instead attack a molecule of AcP, yielding acetate and orthophosphate (good final leaving groups under these experimental conditions).

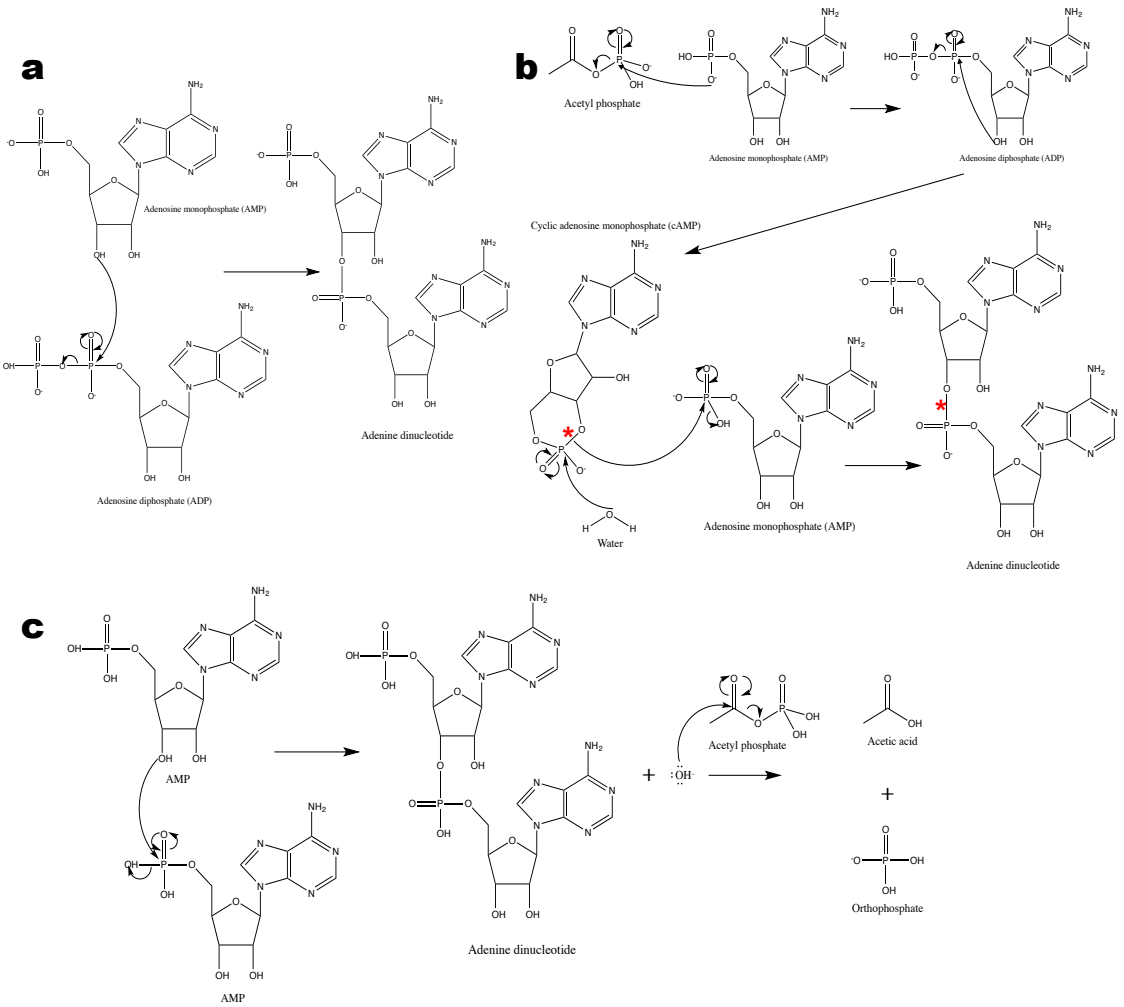


Figure 5.1. Three proposed reaction mechanisms leading to AMP polymerisation. The first (**a**) assumes an ADP intermediate produced by an AcP-mediated phosphorylation of AMP; the second (**b**) depicts the intramolecular cyclisation of ADP also generated through AcP-mediated phosphorylation of AMP; and finally (**c**) a reaction between two molecules of AMP facilitated by AcP, which allows for the final production of acetic acid and orthophosphate molecules instead of water/hydroxide ion. The red asterisk indicates the bond of cAMP (intramolecular P-O bond) that would hydrolytically open and trigger a nucleophilic attack onto another AMP molecule by the now negatively charged (reactive) AMP molecule, generating a dinucleotide (intermolecular P-O bond).

5.2.1.2 Analytical improvements for MALDI-TOF analysis

In order to reduce salt-induced adduct formation, which can easily confound the MALDI-TOF peak assignation and therefore the analysis of the results, a

desalting protocol was applied to the samples before the mixing with the MALDI matrix. In order to first test whether the desalting protocol (Section 2.2.1) was useful to desalt our samples and also to make sure it retained our analytes (small nucleic acid oligomers) it was tested on four 6 µL of standard oligomer aliquots (prepared as described in Section 2.2.2). Two of them were not desalted, whist the other two were. In each case, 6 µL of a 0.02 g/mL solution of NaCl was added to one of the aliquots in order to simulate a sample with a heavy ion load. Figure 5.2a shows the spectrum of the non-filtrated and without NaCl standard with the five expected peaks (black arrows). Figure 5.2b shows the spectrum of the non-desalted aliquot with added NaCl, the resulting Gaussian bell shape for each peak was typical for a sample with an ion excess. The Na^+ cations create numerous adducts with the sample molecule generating many secondary peaks at +/- 22 m/z (corresponding to the analyte +/- 23 m/z (Na^+) – 1 m/z (H^+)) and successive higher and lower further peaks depending on the number of cations the analyte molecules can accept. Figure 5.2c shows the aliquot with added NaCl, which was then desalted. This test shows the desalting protocol worked as planned because the five expected peaks appeared again as desired. I noted also that when the desalting protocol was applied, some of the analyte was lost in the procedure; in the case of the DNA standard the loss was of about a 10–20% of the starting material. It is reasonable to assume this would also apply to any RNA oligomers generated experimentally.

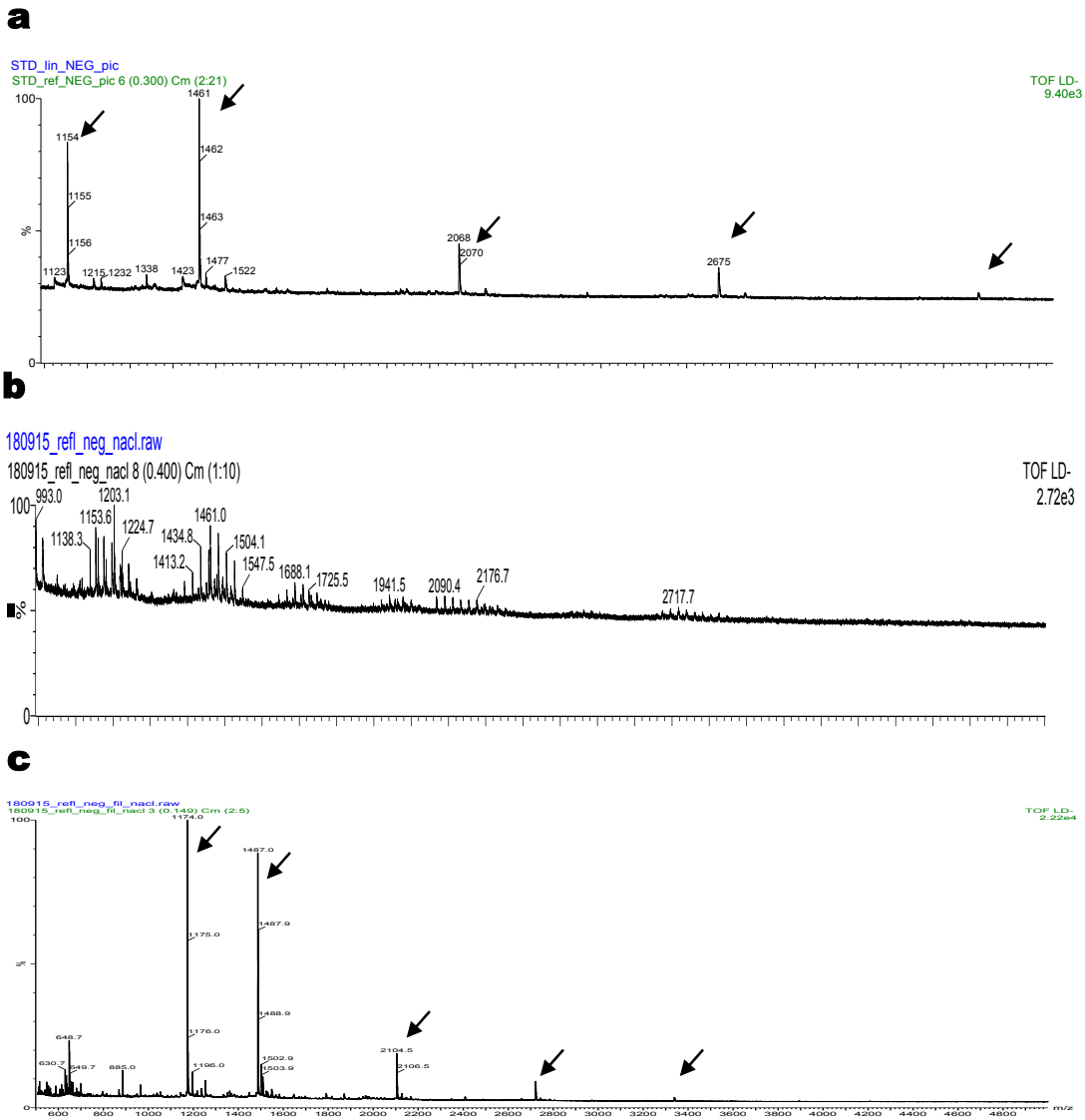


Figure 5.2. MALDI-TOF mass spectra of the DNA standard oligomers: (**a**) non-desalted and without NaCl added, (**b**) non-desalted with NaCl added, and (**c**) desalted and with NaCl added. Black arrows indicate the expected five DNA oligomer peaks. Abscises correspond to m/z values and ordinates to relative ionic abundance in all presented MALDI spectra.

5.2.1.3 Experimentation on AMP polymerisation

The final concentration of AMP used in all polymerisation experiments was of 3 mM (lower than the previous phosphorylation tests using ribose at 300 mM) due to the low solubility of AMP in water (stock solution concentration of 5 mM). The acetyl phosphate concentration used was of 450 mM (slightly higher than in

previous experiments (300 mM)) in order to try to maximise the reaction yield. As in the rest of the early experimentation presented throughout this thesis, the stock solution of AcP synthesised in the laboratory (Section 2.1.1 of Chapter 2) was used for these experiments. The temperature conditions chosen were those previously tested, 20 °C and 50 °C, as well as 75 °C as it has been claimed that some condensation reactions would only take place from 75 °C and above (Šponer et al., 2015). All three temperatures used can easily be found in a vent scenario. The pH values tested were of 11 and 5, as both would have been present in Hadean alkaline vents (pH 5–6 would be found in ocean waters percolating into the vent). No additional ions were added during the initial tests. Reactions were carried out in 1.5 mL Eppendorf tubes with a final volume of 1 mL. The tubes were heated when applicable in a block heater. The pH of each reaction was adjusted using solutions of 5 M HCl and 10 M NaOH. The incubation time for each experiment was a maximum of 5 hours (as little effect of AcP would be expected any longer, especially at high temperature) although some samples were taken after 24 and 48 hours. To account for the time required between sampling and data acquisition, 30 minutes have been added to each sample time. Samples were not frozen (in order to store them) as the freeze-thaw process (akin to a wet-dry cycle) might account for some polymerisation, which would become an artefact in this set of experiments.

The ion fragments which would be indicative of polymerisation are the following: dimer (675.44 m/z), trimer (1004.66 m/z), tetramer (1333.88 m/z) and successive additions of 329.22 m/z which correspond to an AMP residue (AMP (347.22 m/z) – H_2O (18.01 m/z)) incorporated into the growing RNA polymer. The spectra in Figure 5.3 are representative of the data that the first set of experiments yielded.

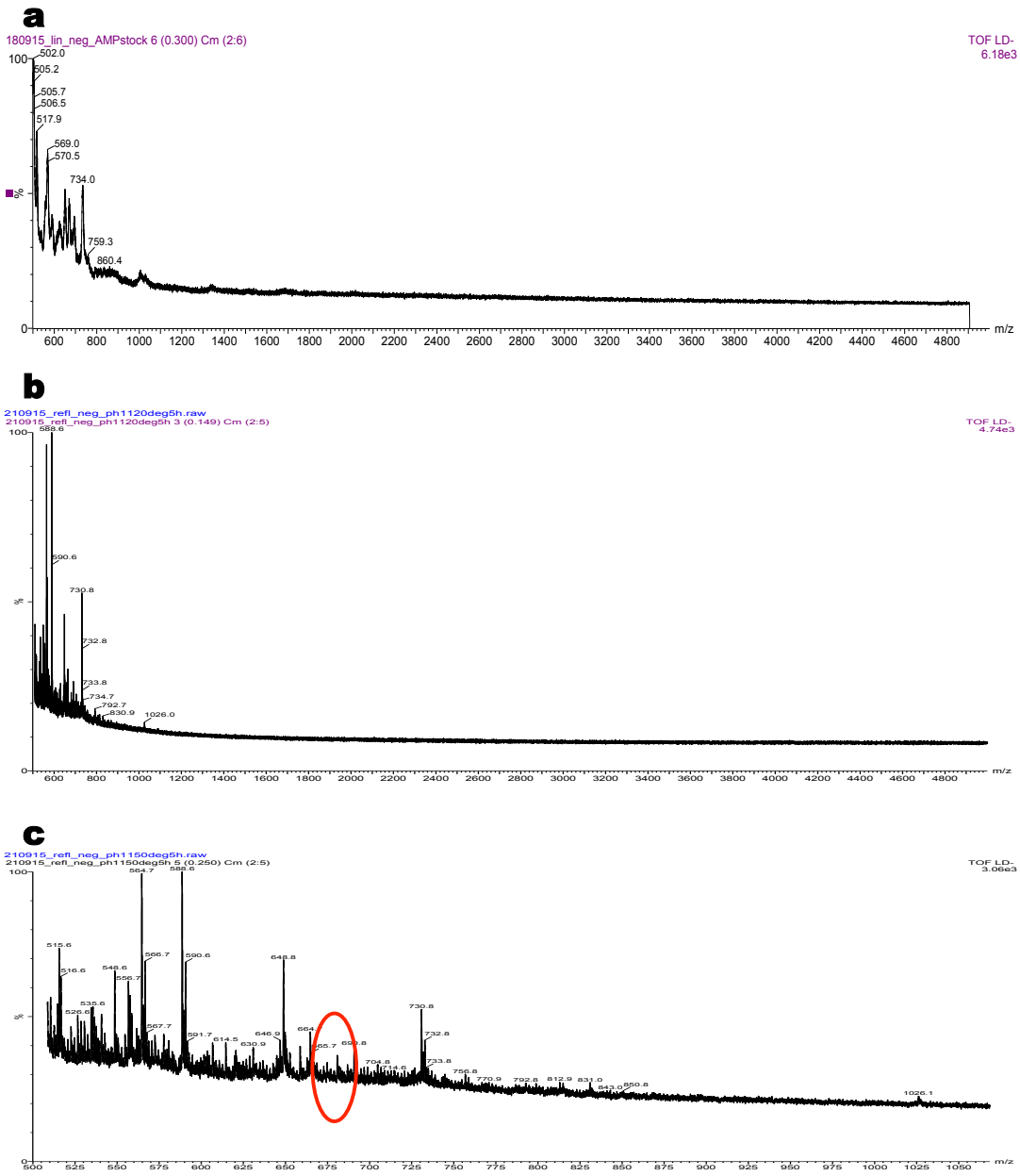


Figure 5.3. MALDI-TOF mass spectra of the products of the reaction of AMP and AcP, at pH 11 and after 5 hours of incubation: (**a**) negative control containing only AMP and HPLC water at 20 °C (range: 500–5000 m/z); (**b**) polymerisation reaction at 20 °C (range: 500–5000 m/z); and (**c**) polymerisation reaction at 50 °C (range: 500–1075 m/z). The red oval shows the absence of the dimer mass (675.44 m/z), which would indicate polymerisation. All samples were desalted before their analysis.

No polymerisation products could be observed at any pH, temperature or incubation-time combination. The red oval in Figure 5.3c shows the absence of the smallest oligomer I was trying to synthesise. Some minor peaks appeared in several cases but their masses were attributed to adducts formed by the matrix

(THAP) and cations such as K^+ and Na^+. These minor peaks could also be observed in the negative controls done for every experimental combination of reaction conditions, reinforcing the interpretation that they were analytical artefacts. Two controls with 3 or 300 mM AcP and 200 µL of standard oligomer solution were also performed to ensure that a high concentration of AcP did not mask the signal arising from potential oligomers, as this is a common occurrence in mass spectrometry due to ionisation suppression; the results (Figure 5.4) indicated that a high concentration of AcP did not interfere with the signal arising from nucleic acid oligomers.

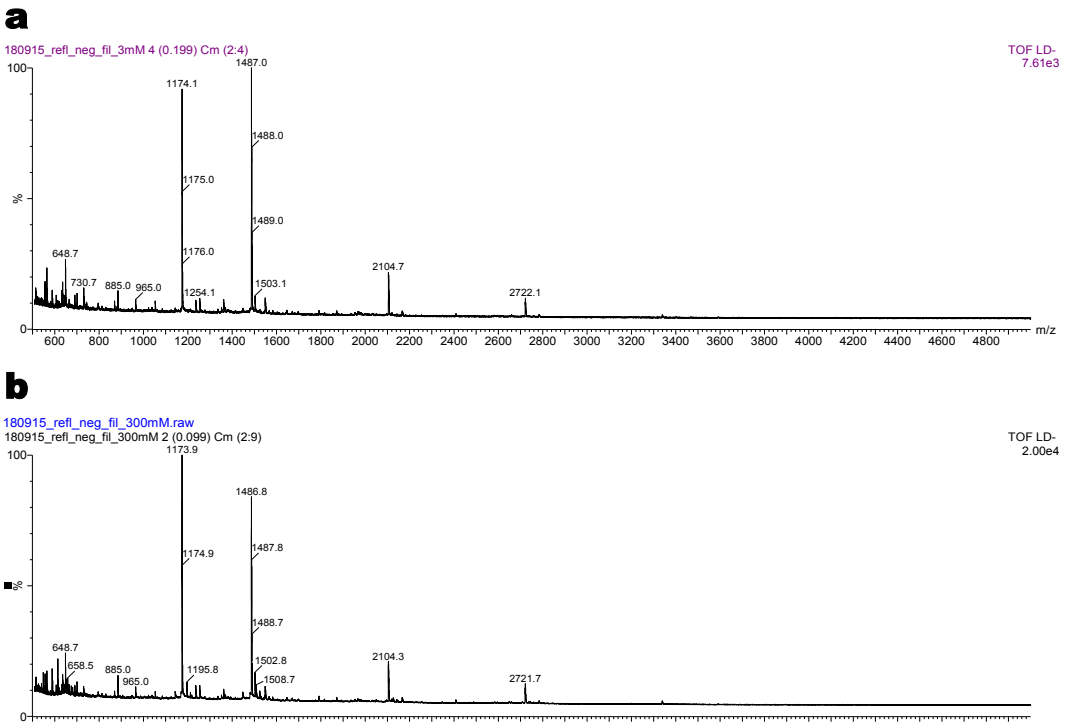


Figure 5.4. MALDI-TOF mass spectra of the DNA standard oligomers incubated during 1 h at 20 °C with (**a**) 3 mM AcP, and (**b**) 300 mM AcP. Peak intensity did not vary depending on the AcP concentration. All samples were desalted before their analysis.

One of the samples from the previous experiments was accidentally not desalted and a pattern that could be attributed to a successful AMP polymerisation emerged. This striking result rendered the question of whether I had been successfully synthesising small oligomers at modest yields before, but could not detect them due to the known loss of 10–20% of analyte during the desalting protocol. Alternatively it was also possible that the oligomer loss

during desalting was not a relative percentage but a given amount (e.g. 200 nmol); if the reaction between AcP and AMP yielded oligomer quantities of up to 200 nmol, none would be detected. Because of this, all the previous experiments (which yielded negative results) were performed again and analysed without the posterior desalting procedure. The data shown in Figure 5.5 are representative of the data, at pH 5, acquired for these non-desalted set of samples under all the other experimental conditions tested (temperature and incubation time). Under all the combinations of experimental conditions, marked peaks at 716.3, 1085.1, 1453.4, 1821.74, 2190.8, 2558,37 m/z appeared. These peaks never appeared in any of the negative controls performed for each of the combinations of experimental conditions; instead, small peaks appeared around the area corresponding to the 1085.1 and 1453.4 peaks from the experimental runs, but not matching exactly (red oval in Figure 5.5b). These observations strongly suggested that the peaks found in the experimental runs arose from AMP polymers. Both in the experimental and control conditions, secondary peaks at +/- 22 m/z (Na^+ - H^+) appeared as expected when not desalting the samples.

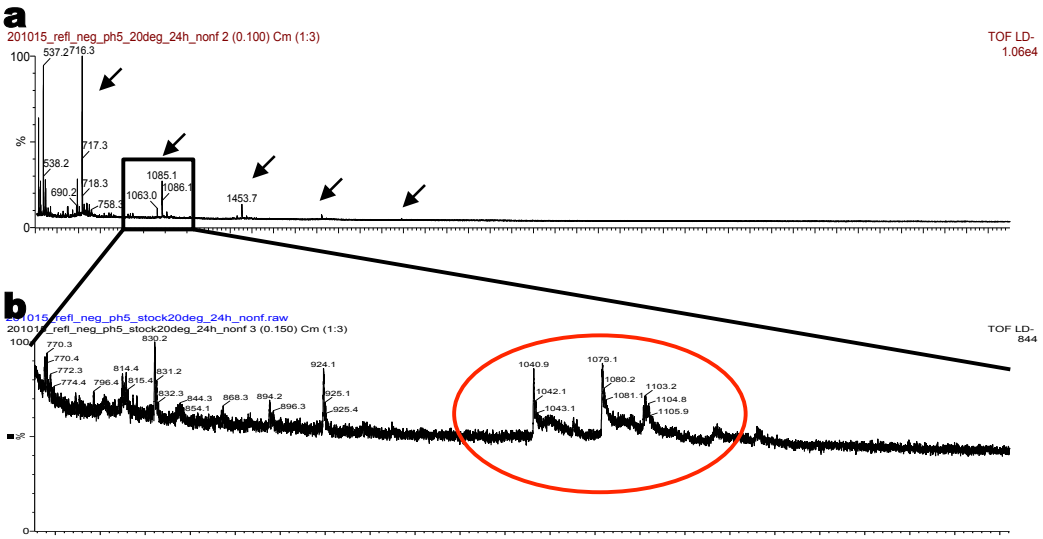


Figure 5.5. MALDI-TOF mass spectra of the products of the reaction of AMP (3 mM) and AcP (450 mM), at pH 5 and after 24 hours of incubation: (**a**) polymerisation reaction at 20 °C, and (**b**) zoomed in view of a negative control containing only AMP at 20 °C. None of the samples were desalted before analysis. The red oval highlights the zoomed in area in the negative control experiment corresponding to the 1085.1 m/z peak of the putative trimer from the experiment in (a). None of the small peaks in (b) corresponded to the putative trimer: 1085.1 m/z.

Nucleotide polymerisation experiments are known to generate a wide range of artefacts that make interpretation of data difficult (Burcar et al., 2013). As mentioned earlier in this Chapter, work done in the past in which the authors claimed polymerisation had been achieved (Costanzo et al., 2012, 2009), has been questioned by Morasch et al., (2014). Purine nucleotides tend to aggregate due to interactions between the rings of the nucleobases (Figure 5.6). Some authors claim this phenomenon boosts the chances of a successful polymerisation (Costanzo et al., 2012, 2009; Pino et al., 2015). Other authors consider it merely an artefact that masks potential true polymers (Burcar et al., 2013) due to the polymers and stacks displaying very similar m/z values. Polymers differ from stacks in that they present a loss of a water molecule for every monomer added to the polymer due to the condensation necessary to generate the new covalent bond.

The peaks obtained in the experimental runs could be assigned to AMP polymers according the following scheme:

716.3 = 347.22 m/z (AMP) x 2 – 18.01 m/z (H_2O) = 713.54 m/z, which is close to the experimentally found 716.3 m/z

1085.1 = 1004.66 (AMP trimer) + 80.971 m/z (phosphate radical) = 1085.63 m/z

↓ + 329.22 m/z (AMP monomer) + 39.098 m/z (K^+) – 1 m/z (H^+)

1453.4

↓ + 329.22 m/z (AMP monomer) + 39.098 m/z (K^+) – 1 m/z (H^+)

1821.74, and so on

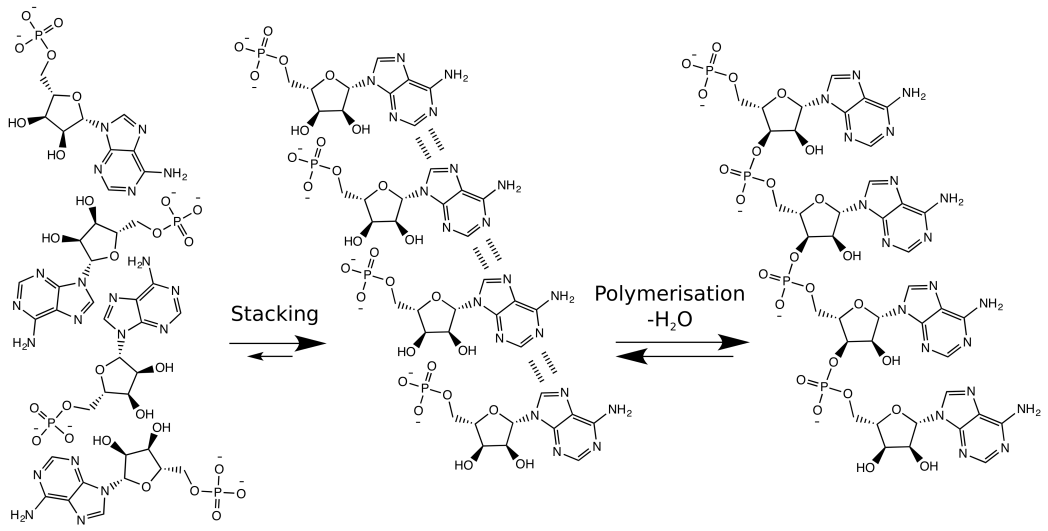


Figure 5.6. Stacking (middle) and polymerisation (right) of AMP monomers (amended from Whicher et al., 2018). Stacking is energetically favoured in water due to nucleobase interactions, and even more so when dissolved cations balance the negative charges of AMP.

The peaks appearing in the same zones of the negative control experiments (Figure 5.5b) could be explained as aggregates. For example the 1079.1 m/z peak = 3 x 347.22 m/z + 39.09 m/z (K^+) – 1 m/z (H^+). As pointed out before, aggregates of purine nucleotides are known to be common due their bases being more energetically favoured piled up in water rather than each one in solution (Šponer et al., 2015). This is due to an entropic effect: fewer water molecules need to be structured around the surface of the relatively hydrophobic surfaces of the nucleobases of a dinucleotide stack than around two separate nucleobases in solution; fewer structured water molecules implies a higher entropy (and thus a lower energy or more stable) state. Because both sets of peaks did not weigh the same, and the aggregates did not appear in the experimental runs (or appeared as very small peaks), I initially considered them to be analytical artefacts that did not interfere with the peak assignation done for our experimental runs.

But this presumptive peak assignation had several peculiarities. First, from the AMP trimer onwards, the polymers would contain an extra phosphate, presumably at their 3' hydroxyl end (Figure 5.7), or on one of the several 2'

hydroxyls. The possibility of this extra phosphate being present in the polymers was plausible if we consider the phosphorylating capability of AcP on hydroxyl groups of ribose (and other sugar derivatives) demonstrated in Chapter 4 when mixing ribose and AcP. Despite this, the reason why only the trimer (and onwards) presented this 3' phosphate, but the dimer did not, was unclear. Considering the proposed reaction mechanisms shown in Figure 5.1, there is no obvious mechanistic reason explaining the presumptive presence of this extra 3' phosphate group only from the trimer onwards.

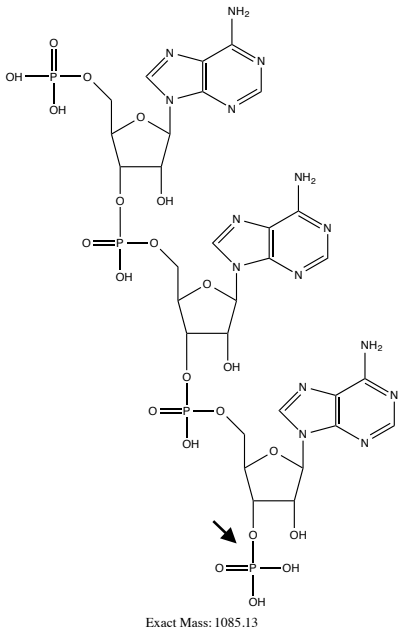


Figure 5.7. Chemical structure of the presumptive AMP trimer, which contains an extra phosphate group at its 3' hydroxyl (black arrow).

The same experiment performed at pH 11 yielded negative controls (experiments with only AMP) containing (albeit lower in intensity) the 1085.1 m/z and 1453.4 m/z peaks. As shown earlier, these peaks were previously only present in the experimental runs at pH 5, and were undetectable in the negative controls containing only AMP. After generating new stock solutions of all reagents used (including the pH-adjusting solutions) to exclude cross-contamination, the same surprising results at pH 11 appeared. This suggested that the Na^+ ions from the NaOH (not added at pH 5) used to adjust the pH could potentially generate the peaks, which I had previously assigned to AMP polymers. NaOH was the only source of Na^+ ions of these experiments containing solely commercial AMP (which is a Li^+/K^+ salt).

Other observations point to this possibility: (i) the peaks which were assigned to AMP polymers were accompanied by secondary peaks which would correspond to +/- Na^+ adducts; if that had been the case, the main peak should have contained Na^+ as well, but that was not the case in our peak assignation. (ii) The putative dimer m/z value did not fully match the m/z values repetitively observed in the experimental runs; the difference was of almost 3 m/z units, which was greater than could be expected for a low resolution MALDI-TOF instrument (1-1.5 m/z units of variation could have been explained by small differences in the spectrum calibration). (iii) For any of the presumed polymerisation peaks obtained, there should have been accompanying peaks at – 39.09 m/z (K^+) and/or at – 80.97 m/z (phosphate radical) which would have been a result of the parent molecule losing one or both of these moieties (– K^+ should be more common due to the ionic bond fragility compared with the covalent bond); these secondary peaks were never observed. (iv) Despite the possibility that a significant portion of the (small) quantity of polymerised oligomers generated were lost during the desalting process, it was strange that the presumed polymers were only visible when the samples were not desalted, as the desalting method (Section 2.2.1) was tailored specifically to small nucleic acid oligomers and with the advice of an expert who participated in a relevant publication. (v) At all temperatures, pH and incubation times tested, it was strange that the presumed positive polymerisation results were almost identical to each other; what seems more reasonable is that each different combination of all three factors would direct the polymerisation in a slightly different way, some generating longer polymers, some generating shorter ones, and some precluding their formation or disaggregating them over longer incubation times. In conclusion, all these points, coupled to the negative controls containing solely AMP presenting the putative polymerisation peaks, strongly suggest that what I had previously assigned to polymerisation products were, in reality, aggregates product of AMP stacking.

In order to test the hypothesis that all the experimental peaks obtained were AMP aggregates (not polymerised), the experiments were run again and their pH adjustment was performed using KOH (10 M) and HCl (5 M), this way any differences caused by the main cation present (K^+ or Na^+) would be evident.

Figure 5.8 shows that the peak at 1085.1 m/z appeared only when the AcP solution (which contained a lot of Na^+ due to its preparation protocol) was added (Figure 5.8a-b) to the AMP; and also that the peaks already appearing in the negative control experiment were still visible when the AcP solution was added (Figure 5.8c). These results support the aggregate hypothesis as the negative control with only AMP and titrated with KOH showed different peaks from the same control titrated with NaOH. In conclusion, it seems that all the peaks obtained changed according to the most common ion present (generating different aggregates) and not as a result of polymerisation.

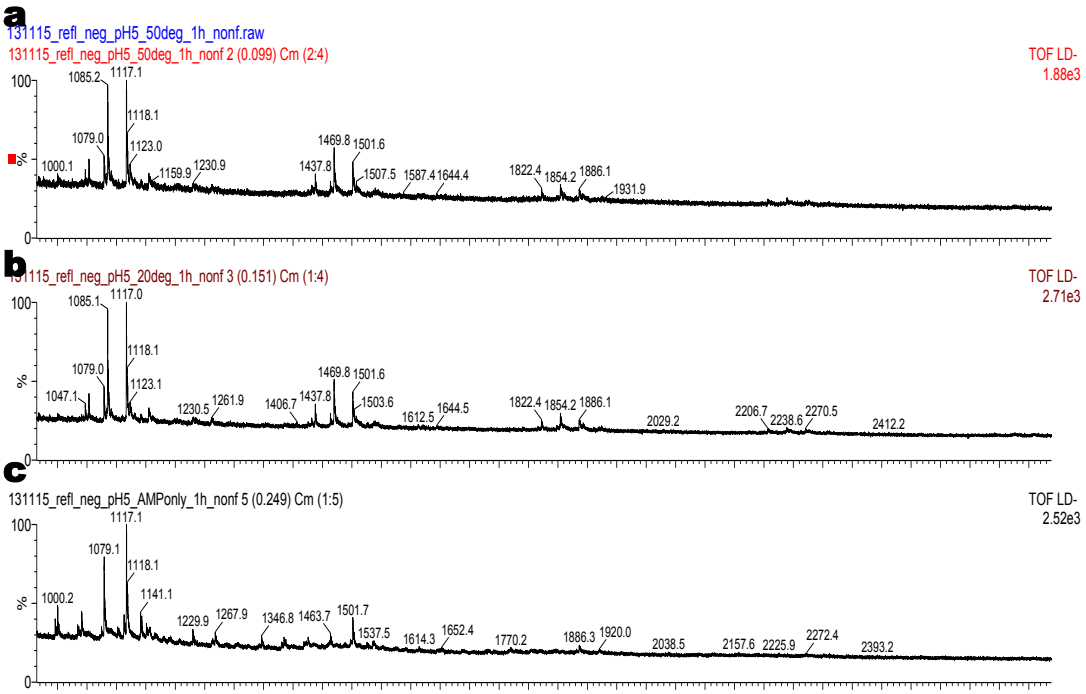


Figure 5.8. MALDI-TOF mass spectra of polymerisation experiments using 3 mM AMP and 450 mM AcP, at pH 5 and after 1 hour of incubation at (**a**) 50 °C, (**b**) 20 °C and (**c**) a negative control with only AMP at 20 °C. The pH titration was performed with 10 M KOH and 5 M HCl. None of the samples were desalted prior to analysis. Peak at 1085.1 m/z, previously assigned to an AMP trimer, only appeared when the AcP solution (which is rich in Na^+) was added to the AMP.

Furthermore, our synthesised AcP solution was known to contain around 0.1 M acetate as well as undetermined but relevant levels of Na^+, K^+ and orthophosphate due to the solution preparation protocol (Section 2.1.1 of Chapter 2). In order to assess the potential impact of the usage of a 'dirty' AcP stock solution for the polymerisation experiments, a fresh 1.1 M solution of

commercial AcP (lithium-potassium salt, ≥85%, Sigma) in HPLC water was prepared, aliquots were prepared and stored at –80 °C. Due to commercial constraints, Li^+ was introduced into the equation (cation-free AcP was not available) so, in order to also study its effect on the potential aggregation of AMP and how that compares with the effect of Na^+ and K^+ ions, a set of negative controls made of 3 mM AMP and 3 mM LiCl (≥99.99%, Aldrich) or 3 mM NaOH or 3 mM KOH were prepared and analysed.

The results of these experiments confirmed that the mass of the aggregates when there was only AMP (not AcP) in solution depended on which cation was dominating in the solution. For instance, when the AMP solution was not titrated, the resulting main peak in the dimer area was 694.4 m/z, which corresponded to two molecules of AMP (347.22 m/z) stacked together (not condensed) (Figure 5.9d). The commercial AMP I used contained ions giving rise to a secondary peak at 756.4 m/z, which corresponded to the previous stack of two AMP plus Na^+ and K^+. The spectra from Figure 5.9a-c show the main aggregate peak at 694.4 m/z plus secondary peaks depending on the ion added Na^+, K^+ or Li^+ at 716.4 m/z, 732.4 m/z or 700.5 m/z respectively. The difference from all the secondary peaks to the main peak was attributable precisely to the mass of each ion minus 1 m/z corresponding to the displaced proton.

This effect was applicable to the results obtained for the area around 1085.1 m/z and 1453.4 m/z in each experimental run; the chart depicted in Figure 5.10 shows the final peak assignation for the peaks from the aggregate consisting of 3 AMP molecules onwards. Only aggregates with K^+ and Na^+ have been assigned, as Li^+ aggregates in experimental runs were practically non-existent. K^+ seemed to outcompete Li^+ when generating AMP stacks when both are present in roughly equimolar quantities (commercial AMP contained only trace amounts of K^+). The reason for this remains highly speculative but it may have to do with Li^+ presenting a higher charge density compared to K^+, which seems to be inconvenient for its chelation onto phosphate groups when larger cations, such as K^+, are present.

The AMP polymerisation experiments were repeated, this time using commercial AcP. Comparing the usage of Li^+K^+AcP with the experiments previously performed (using Na^+AcP), I observed the same phenomenon: pH (the same ion solution at different pH yielded the same result); temperature and incubation time did not have an impact on which peaks were generated. Instead, the ion that was dominant at each time point determined the resulting peaks. With all of the previous observations I can conclude that acetyl phosphate did not drive the condensation of AMP under the experimental conditions tested.

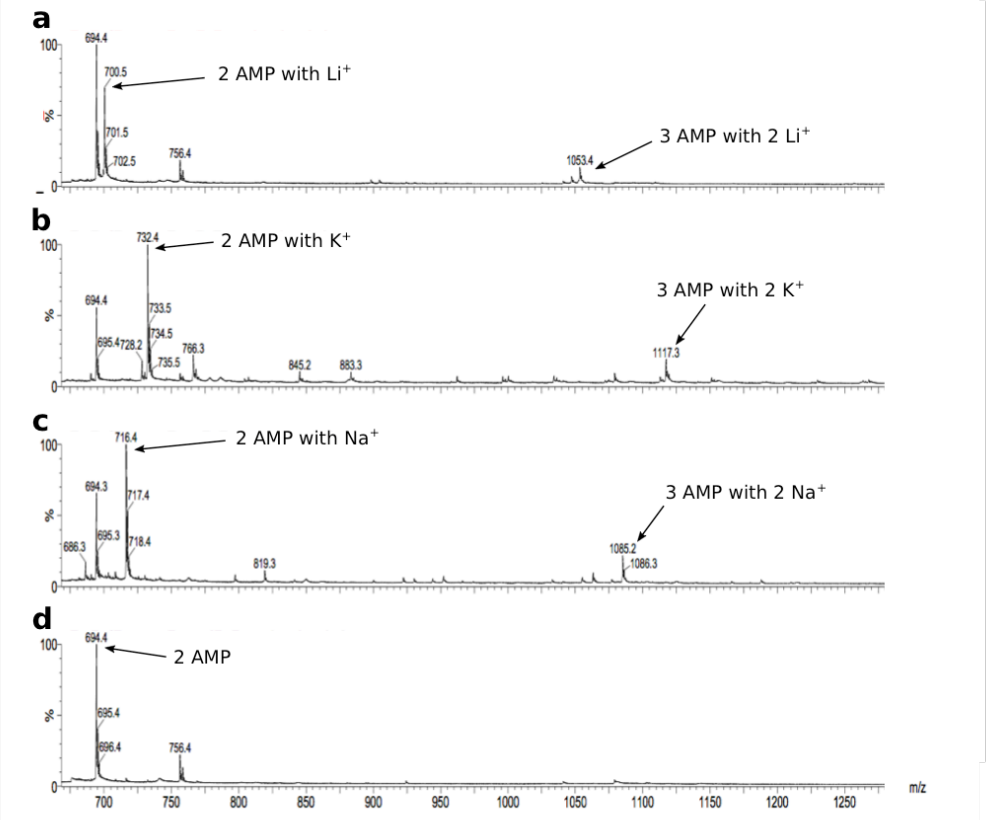


Figure 5.9. MALDI-TOF mass spectra of solutions of 3 mM AMP with added (**a**) 3 mM LiCl, (**b**) 3 mM KOH, (**c**) 3 mM NaOH and (**d**) water (Whicher et al., 2018). Main peaks corresponding to stacks of AMP molecules with each cation are shown for clarity. The size of an AMP monomer is of 347.22 m/z.

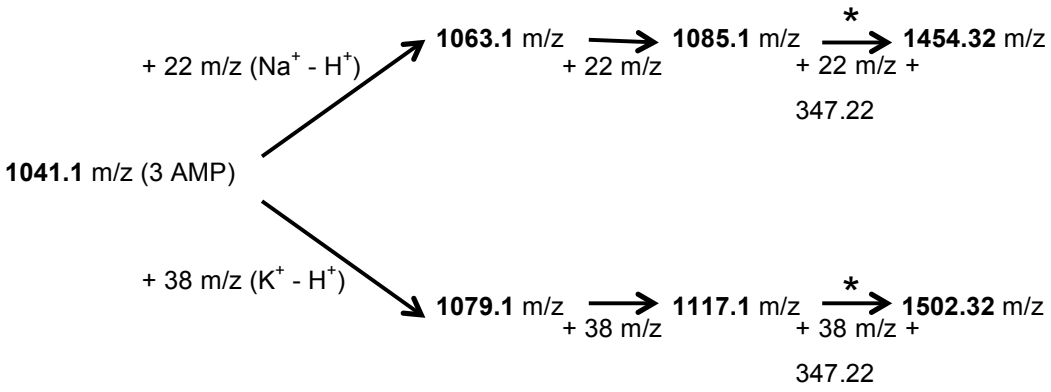


Figure 5.10. Assignation of the main peaks observed during experimental runs. All the successive main aggregate peaks observed are a repetition of the last additive step (marked with a *).

Another experiment was performed to determine whether the orthophosphate present in the initial AcP solution used (Section 2.1.1 of Chapter 2) was responsible for any of peaks observed during the experimental runs. The results in Figure 5.11 show this was not the case; no new peaks could be observed.

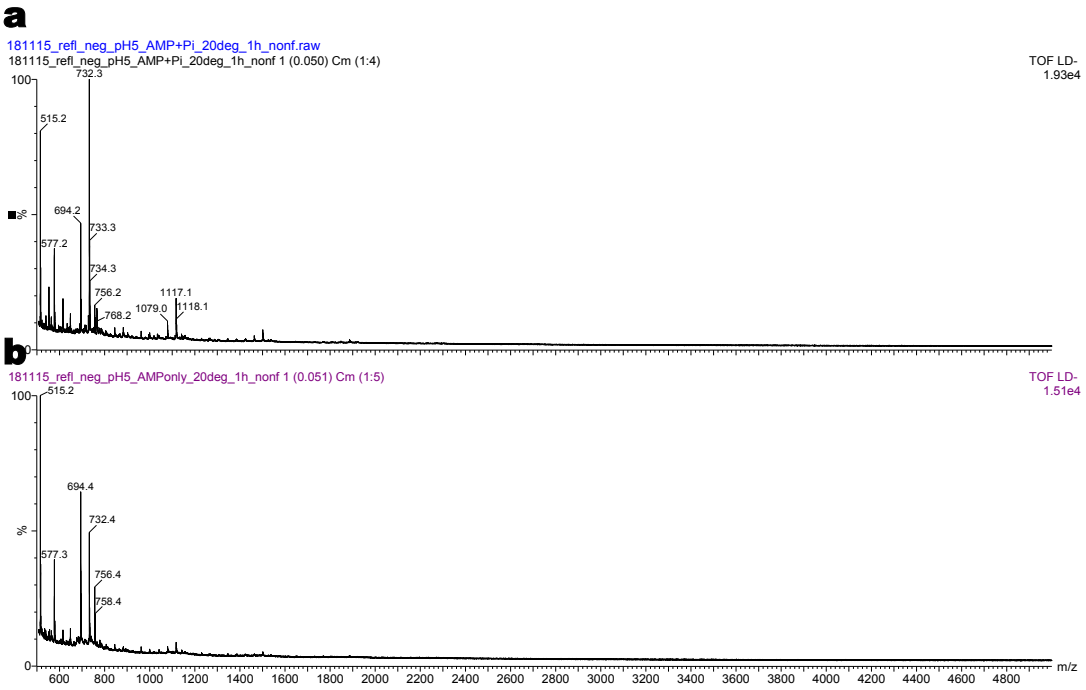


Figure 5.11. MALDI-TOF mass spectra of a solution with (**a**) 3 mM AMP and 45 mM phosphoric acid or (**b**) negative control with only 3 mM AMP. The solutions were not titrated or desalted. The phosphoric acid concentration used was chosen on the basis that after 1 hour at 20 °C, 10% of the initial 450 mM of AcP (in an experimental run) would have been degraded into orthophosphate (Figure 4.4). No AcP was added to either of these runs.

Another set of experiments was designed to assess the effect of Mg^{2+} and Ca^{2+} on the aggregation or potential polymerisation of AMP. These ions were tested as they are known to be necessary for several cellular and biochemical processes, and had also been tested for other experiments in this project. The concentration of both ions was 10 mM (as in other publications for polymerisation purposes (Burcar et al., 2015)). The AMP concentration was 2.5 mM and AcP (commercial version) 366 mM. The results (not shown) showed no difference to the aggregates seen in previous experiments (either at pH 5 or 11) and no new peaks, corresponding to aggregates of AMP with Mg^{2+} or Ca^{2+}, appeared.

Finally, an experiment was performed using adenosine triphosphate (ATP) instead of AMP. ATP was selected for these experiments as it is one of the monomers used to polymerise nucleotides in cells (not AMP), and pyrophosphate (from ATP) is a better leaving group than orthophosphate (or water, obviously) under the tested experimental conditions. An experiment with 3 mM ATP (dihydrate, 99%, Aldrich) and 450 mM AcP was prepared. The mixture was incubated for 1 hour at pH 11, at 50 °C (Figure 5.12a) and 20 °C (Figure 5.12b). Another control using only 3 mM ATP (Figure 5.12c) under the same experimental conditions was also performed. The results in Figure 5.12c show a marked aggregation (but not polymerisation) pattern of the ATP monomers due to nucleobase stacking; aggregates up to a 9-mer could be observed. Stacks of ATP were expected to almost disappear at high pH, as the phosphates would be fully ionised, thus making the phosphoanhydride bonds less stable). In fact this did not happen, probably reflecting that the cation excess present in all AcP solutions used compensated for any extra charges present at high pH. When ATP was incubated with AcP (Figure 5.12a-b) the aggregates only reached the size of a tetramer and present much larger secondary peaks corresponding to K^+ and Na^+ adducts, probably reflecting the capacity of the triphosphate to chelate cations better than the single phosphate of AMP. No polymerisation products (aggregate mass – water mass) could be observed at any combination of pH, temperature or incubation time.

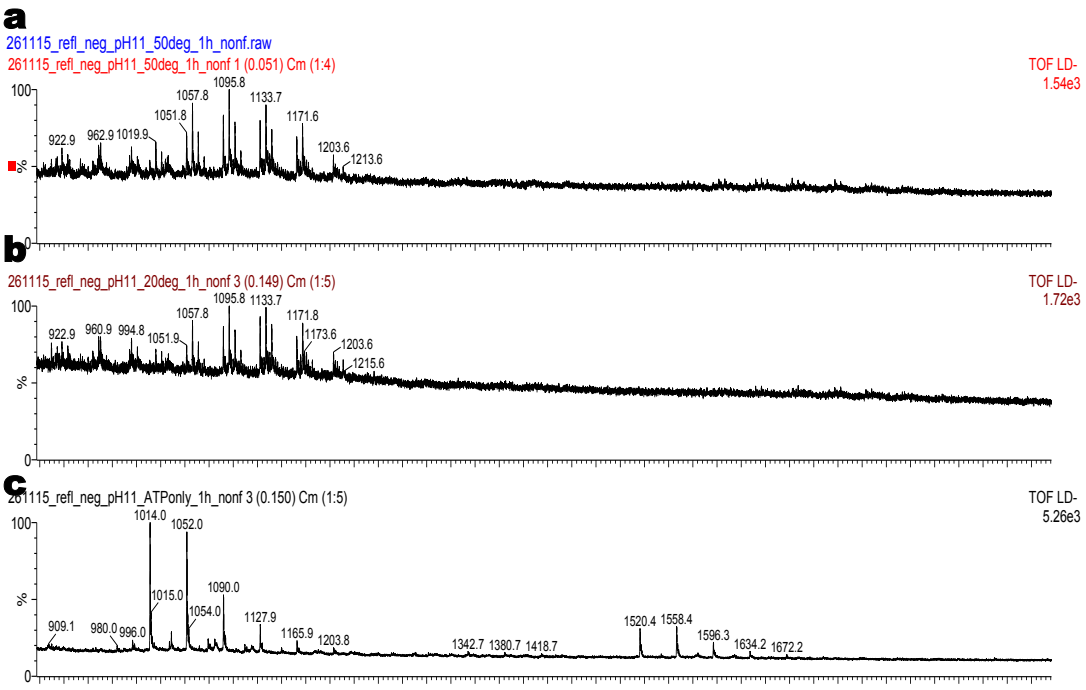


Figure 5.12. MALDI-TOF spectra of 3 mM ATP and 450 mM AcP at pH 11, incubated for 1 hour at (**a**) 50 °C, and (**b**) 20 °C. Negative control containing only 3 mM ATP (**c**). None of the samples were desalted.

5.2.2 AcP-mediated polymerisation of glycine

The polymerization of glycine (0.5 M) by AcP (0.5 M) in water was attempted but I was unable to detect any linear diglycine, cyclic diketopiperazine, or any other short polymer; instead I corroborated the findings of others (Di Sabato & Jencks, 1961a), that AcP efficiently acetylates the amino group of glycine to form N-acetyl glycine (NAG), especially under alkaline conditions.

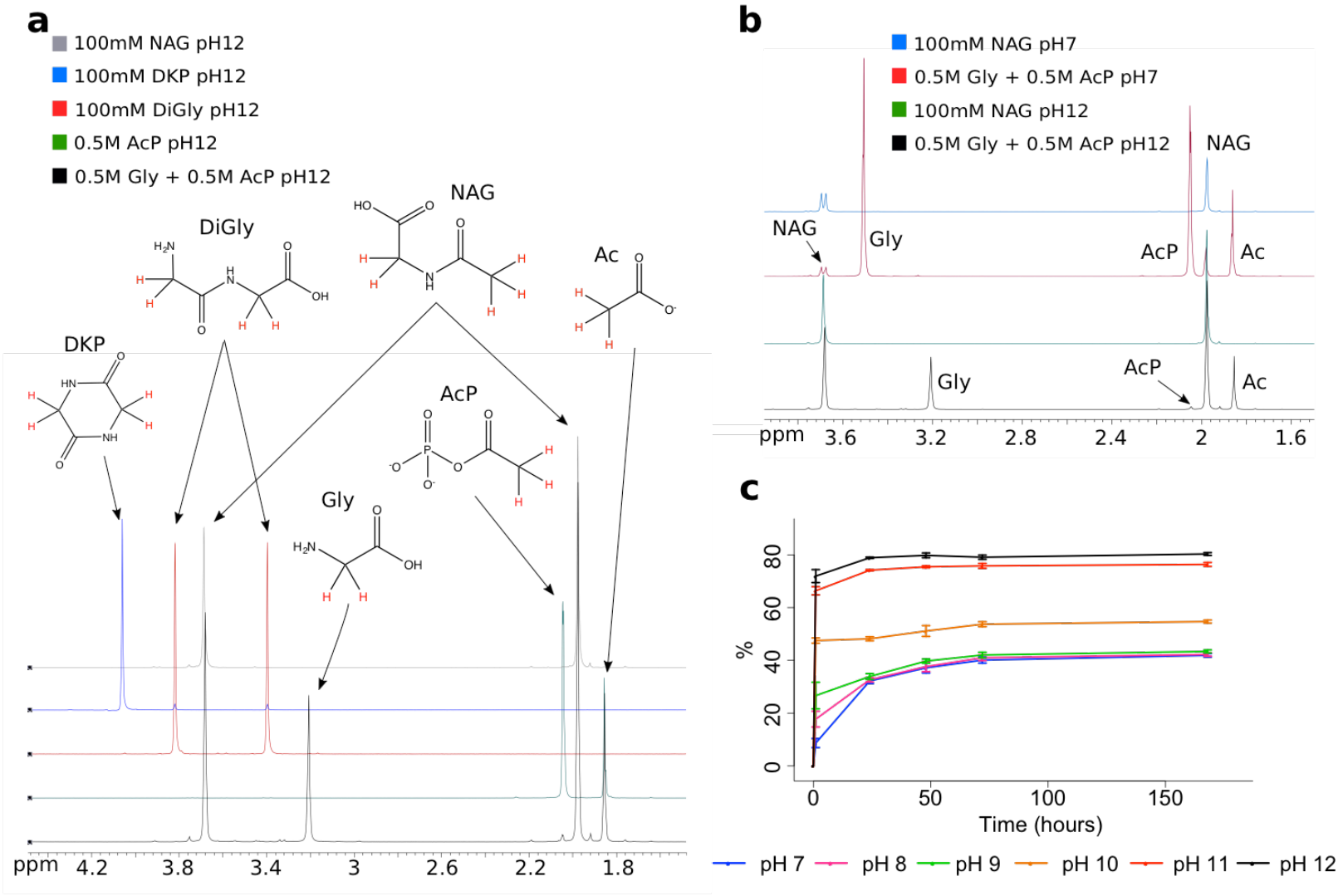

a
100mM NAG pH12
100mM DKP pH12
100mM DiGly pH12
0.5M AcP pH12
0.5M Gly + 0.5M AcP pH12
DiGly
NAG
Ac
DKP
AcP
Gly
ppm 4.2 3.8 3.4 3 2.6 2.2 1.8
b
100mM NAG pH7
0.5M Gly + 0.5M AcP pH7
100mM NAG pH12
0.5M Gly + 0.5M AcP pH12
NAG
NAG
Gly
AcP
Ac
Gly
AcP
Ac
ppm 3.6 3.2 2.8 2.4 2 1.6
c
%
0 20 40 60 80
0 50 100 150
Time (hours)
pH 7 pH 8 pH 9 pH 10 pH 11 pH 12

Figure 5.13. Acetylation of glycine in water by AcP (Whicher et al., 2018). (**a**) ^{1}H-NMR spectra of acetylation of glycine to N-acetylglycine (NAG) by AcP at 20 °C and pH 12 after 2 hours, compared with AcP, diglycine, diketopiperazine (DKP), and NAG; all commercial standards were titrated to pH 12. Molecular structures highlighting the protons giving rise to each spectroscopic peak are added for clarity. (**b**) ^{1}H-NMR spectra of the acetylation of glycine to NAG by AcP at pH 12 and at pH 7 compared with the spectra of commercial NAG at both pH 12 and pH 7. (**c**) Acetylation of glycine by AcP at 20 °C at different initial pH values over 168 hours. N=3 ±SD. With thanks to Dr Barry Herschy for providing the data shown in (c). All peaks were identified by comparison with the chemical shift of commercial standards (at the same pH), as well as by spiking experimental samples with standards and observing the growth of the assigned peak.

Diglycine forms two peaks on ^{1}H–NMR (Figure 5.13) whereas the condensed form of diglycine, diketopiperazine (DKP) forms a single peak. Under alkaline conditions (pH 9–12) two peaks were detected at around 3.7 ppm and 2 ppm (with the exact chemical shift varying slightly with pH) within minutes of adding AcP to glycine in solution. These peaks corresponded to the formation of NAG (Figure 5.13a). This interpretation was confirmed by the behaviour of NAG by ^{1}H–NMR under more acidic conditions (Figure 5.13b and Figure 5.14). At pH 7 and below, NAG still forms two peaks, but the peak at 3.7 ppm now splits into a doublet. The splitting of these methylene protons at low pH relates to their interaction with the proton on the neighbouring N. This H-N proton exchanges with protons in the solvent at a rate that is dependent on pH. At low pH the rate of exchange of the H-N proton is lower than the speed at which that proton magnetically affects the methylene protons (by shielding and un-shielding), so the methylene signal splits. This pH-dependent peak splitting at 3.7 ppm is, in this context, diagnostic of NAG and confirms that this was the major product formed, rather than diglycine or DKP. We found an 80% yield (relative to glycine) after 24 h at an initial pH 12 (with ~70% yield after 1 h) falling to <40% yield at an initial pH 7–9 after 24h (Figure 5.13c).

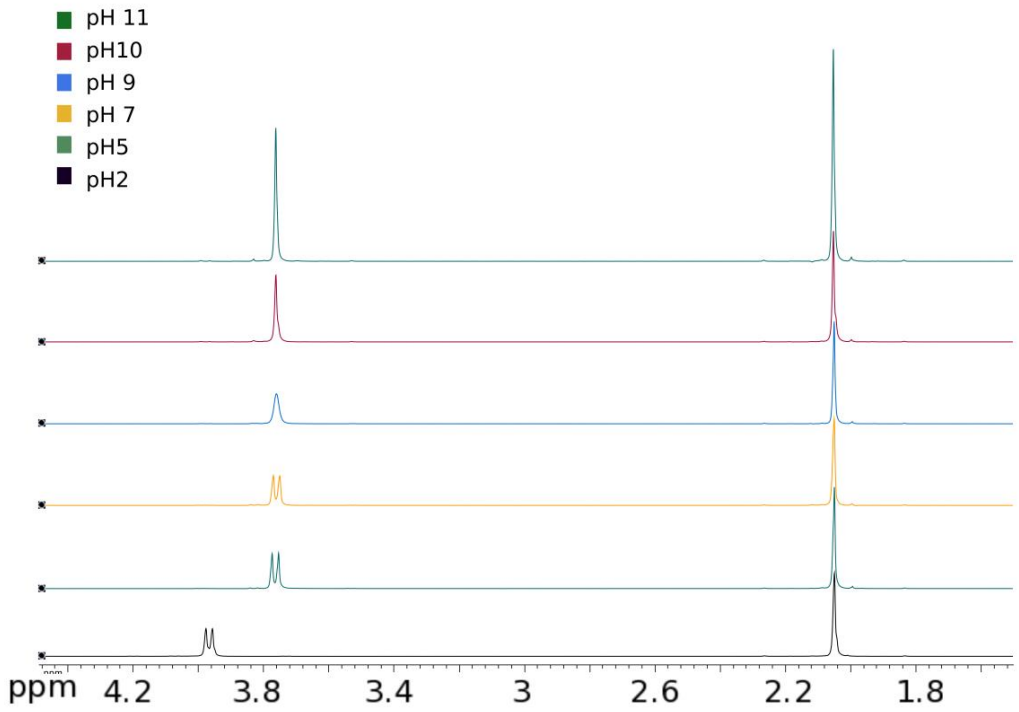


Figure 5.14. ^{1}H-NMR spectra of commercial N-acetyl glycine (100 mM) at pH 2, 5, 7, 9, 10, and 11 (Whicher et al., 2018). In alkaline pH (9-11), the peak at ~3.8 ppm is a singlet, whereas in acidic pH (2-7) it splits into a doublet (see text).

Apart from the aforementioned N-acetylation of glycine, I was also able to observe trace signal of its N-phosphorylation, but only at pH 11 (Figure 5.15a and b(ii)). The O-phosphorylation of glycine was not detected (Figure 5.15b(iii) at any pH. Had this reaction taken place, it is easy to envision a subsequent nucleophilic attack from the amino group of a free glycine, yielding glycylglycine (or diglycine), the successful polymerisation product. While these findings were disappointing, it remains possible that AcP could promote nucleotide polymerisation on mineral surfaces or at lower water activity.

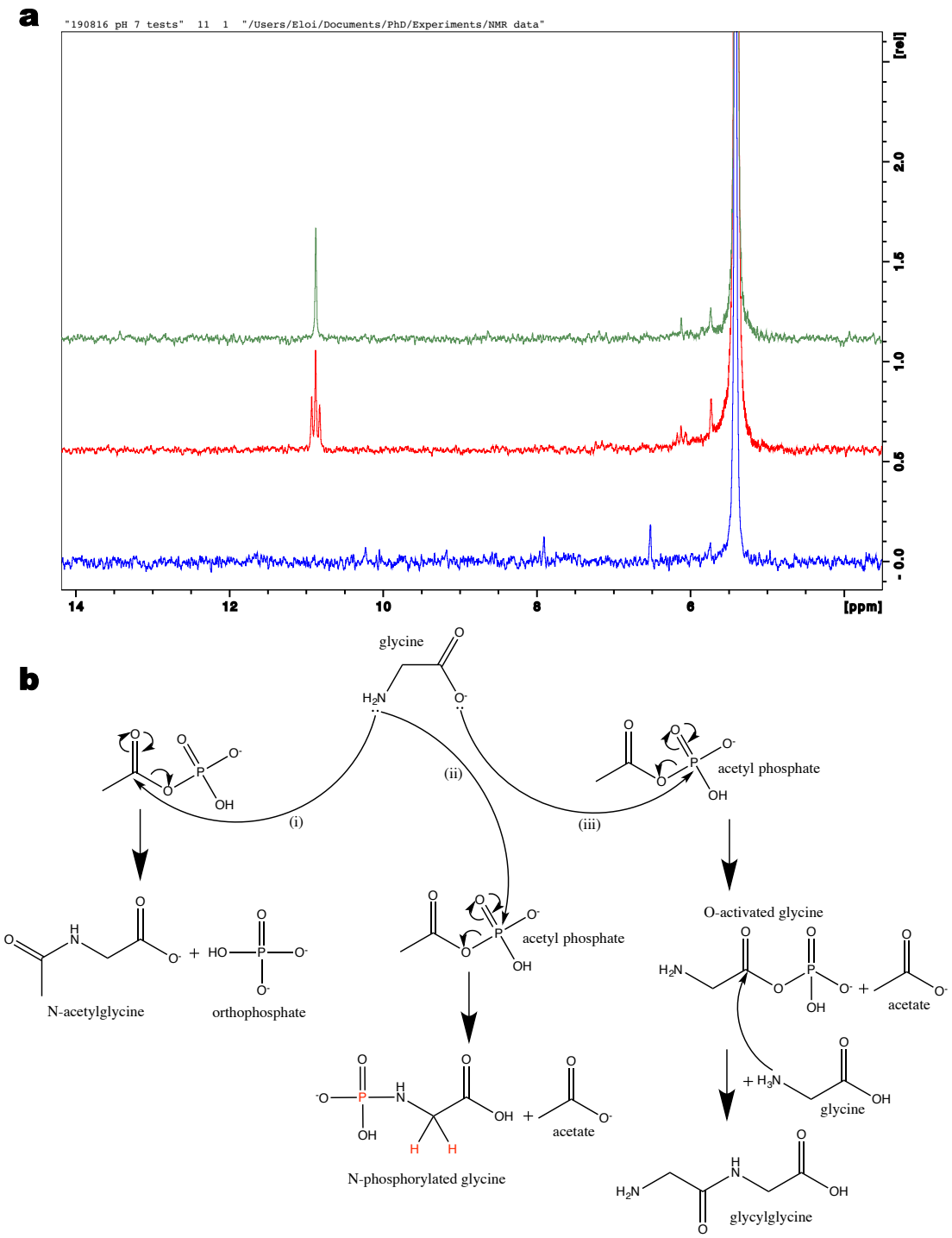


Figure 5.15. (**a**) ^{31}P-NMR spectra of a mixture of glycine (0.5 M) and AcP (0.5 M) at 20 °C and after 2 hours of reaction. The reaction pH was titrated to 11 in the upper two spectra (green and red), and to pH 7 in the lower spectrum (blue). Green and blue spectra show ^{1}H-decoupled ^{31}P-NMR, whilst the red one shows ^{1}H-coupled ^{31}P-NMR (note the splitting of the signal at 11 ppm). (**b**) Proposed reaction mechanisms: (i) N-acetylation, (ii) N-phosphorylation, and (iii) O-phosphorylation (or activation) of glycine. The highlighted atoms (in red) for N-phosphorylated glycine are responsible for the spitting of the ^{31}P-NMR signal in (a, in red).

5.3 Discussion

In the previous Chapter I showed that AcP readily promotes the abiotic phosphorylation of ribose, adenosine and ADP, making it a strong candidate for a prebiotic analogous to ATP. In this Chapter I have shown that AcP does not promote the polymerization of either nucleotides (Section 5.2.1) or amino acids (Section 5.2.2.) in aqueous solution.

Under all the reaction conditions tested, AcP did not favour the polymerisation of AMP. There is likely to be a range of reasons why this did not happen. For instance, so far I have not tested the potential capacity of surface-based catalysts aiding the polymerisation of AMP (or indeed glycine). The polymerisation of nucleotides in extant life occurs at the active centre of enzymes such as DNA and RNA polymerases. Focusing now on RNA polymerases, it is well known that their two-metal-ion reaction mechanism (Sosunov et al., 2003; Steitz & Steitz, 1993) involves three conserved aspartate residues which coordinate two Mg^{2+} that are crucial for catalysis. It is thought that the first Mg^{2+} is directly coordinated by the aforementioned aspartates whilst the second is kept soluble and joins the catalytic intermediary structure when catalysis is taking place (Carvalho et al., 2011). Figure 5.16 shows the proposed enzymatic mechanism involving the 3' oxygen of the previous nucleotide attacking the α-phosphate of the incoming nucleotide and thus creating a new phosphodiester bond whilst liberating pyrophosphate. The catalytic centre residues (the conserved aspartates and others) play an important role by potentially affecting the pK_a of involved chemical groups and stabilizing ionic charges of reaction intermediates and phosphate groups.

It is very possible that the proposed reaction mechanism proposed in Figure 5.1a, despite being chemically possible, did not occur, at least partly, due to the lack of surface catalysts present in our reactions. Catalysts are important for life and for all sorts of chemical reactions (such as industrial ones) as they lower (or eliminate) the activation energy of the reaction, speeding them up considerably. For example, for the RNA polymerisation reaction shown in Figure 5.1 it seems clear that the pK_a of the 3' hydroxyl group is modified at the active site, which

results in its ionisation favouring the reaction, as now it can behave as a stronger nucleophile. The same phenomenon happens with the phosphorus atom of the α-phosphate of the incoming nucleotide: it becomes a better electrophile as the electrons from the surrounding oxygens are being extruded by the positively charged magnesium ions generating a stronger polarity of the P-O bonds. Even though in our experiments I introduced metal ions such as magnesium occasionally, these were in solution and therefore, the optimal or sub-optimal distances between the reacting atoms were virtually impossible to achieve.

As mentioned, the outcome that I had hoped to obtain was a poor, albeit existing, polymerisation yielding AMP dimers or trimers when using AcP in water. Then with the addition of simple mineral catalysts (e.g. brucite ($Mg(OH)_2$) the expected yield and/or length of the oligomers should increase; later successive catalytic improvements such as the addition of Mg^{2+} chelated by aspartate (and later by polypeptide chains) would also potentially improve the yield/length of products. By performing this succession of catalytic improvements I had hoped to roughly reproduce what could have ocurred naturally at the origin of life over time (Copley et al., 2007; Milner-White & Russell, 2005; Nitschke et al., 2013; West et al., 2017). The active site of enzymes has been evolutionarily tuned for thousands of millions of years so that all catalytic atoms are disposed in the optimal positions for catalysis. This surely would not be the case at the origin of life or very early afterwards but, probably, catalysts that were in a solid phase (such as inorganic minerals like brucite and small nucleic acid/peptide sequences) could have provided the initial scaffolding on which life later built surrounding nucleic acid and protein scaffolds, which, over time, improved the yield and selectivity of the reactions as these offered a selective advantage. It is debatable whether the crucial initial reactions at the origin of life (e.g. condensed phosphate synthesis, polymerisation of nucleotides and amino acids, etc.) started with the aid of mineral surface catalysis or if they occurred incipiently in solution and were later improved by inorganic and organic catalysis; this is currently one of the hot topics in the research field of the origin of life. In my view both are plausible and what most likely happened was a mix of both: some reactions probably required

a minimal surface catalysis input (e.g. early phosphate condensing mechanisms could have involved a sort of molecular physical press made of clays, reminiscent of modern ATP synthase), whilst some other reactions could have occurred spontaneously in solution and were only later evolutionarily transferred into ribo/enzymes when/if the corresponding selective advantage paid off.

It is possible that the proposed reaction mechanisms shown in Figures 5.1a and 5.1c did not progress because the condensation of AMP (and potentially of all other nucleotides) at the origin of life was one of the reactions that initially required a reactivity based (at least minimally) on surface catalysis. This conclusion does not stem just from my experimental data, but certainly would fit with my observations. Furthermore this is the direction the research field is slowly tending towards, as some other nucleotide polymerisation studies based on soluble chemistry are now seen with rather unconcealed scepticism. This is, by no means, definitive, as some striking work pointing to soluble polymerisation chemistry could be made public any time. For this reason it would be beneficial in the future to explore the effect of solid state catalysts such as magnesium-bearing minerals such as brucite ($Mg(OH)_2$), which has a similar geometry between Mg^{2+} and O^- as in the active centre of the RNA polymerase; and also small peptide sequences such as small polyaspartate peptides, which could potentially chelate magnesium ions from solution. The catalytic mineral arrays we will mainly focus on will consist of small-scale mineral arrays chelated by amino acids (e.g. Mg^{2+} chelated by serine or Fe_3S_4 clusters chelated by cysteine) instead of large-scale mineral surfaces (e.g. $Mg(OH)_2$ or mackinawite (FeS) sheets), as we consider the former scenario more congruent with life. Metabolism involving large-scale mineral catalysis is not present in extant life or in the core metabolism of LUCA (Weiss et al., 2016) so, probably, it was not either at the origin of life.

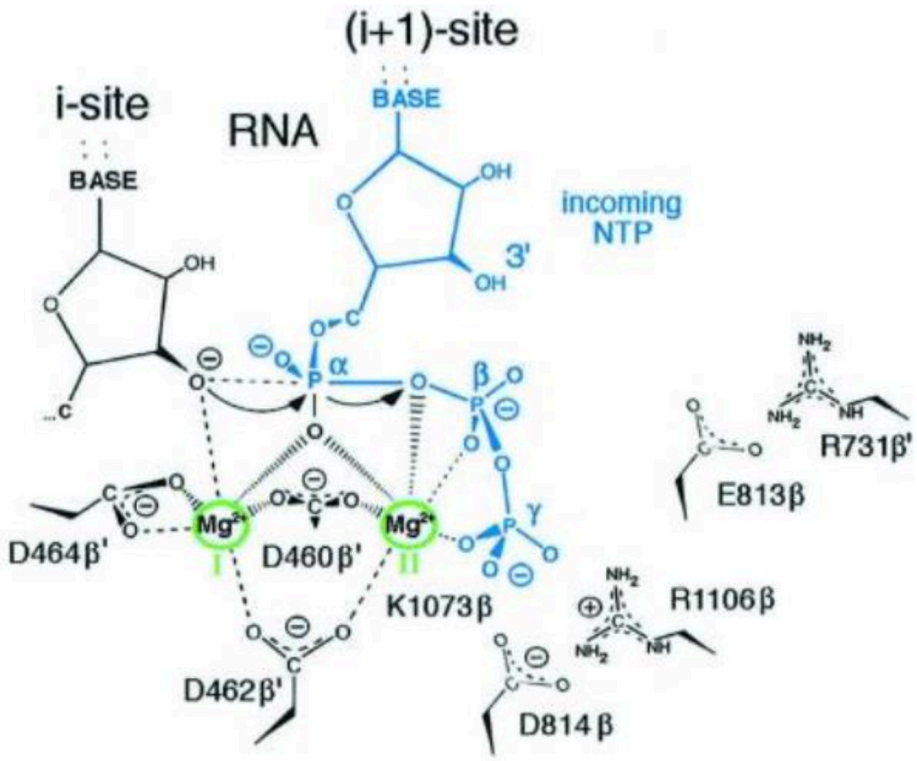


Figure 5.16. Proposed reaction mechanism in detail showing the coordination system for Mg^{2+} I and Mg^{2+} II in the active site of RNA polymerases (amended from Sosunov et al., 2003). Incoming nucleotide is shown in blue. Other residues shown to affect catalysis are also illustrated.

Another factor, which probably contributed to the unsuccessful nucleotide polymerisation, was the fact that some of the reaction mechanisms that theoretically could have worked, would have needed to occur via dehydration (in an aqueous environment). For instance, the mechanism shown in Figure 5.1b (as well as other potential dehydrating mechanisms that are not shown) implies a dehydration reaction. This dehydration, energetically unfavoured due to the pervasive presence of water molecules, could potentially occur due to (i) the well-known phenomenon of purine nucleobase stacking, which would lower the energy requirements for the trans-phosphorylation reaction generating the condensed dimer (Costanzo et al., 2012, 2009); and (ii) due to the fact that a molecule of water would be incorporated during the hydrolytic step of the reaction, this way there would not be a net production of water on the overall reaction. The polymerisation of cAMP in water reported by Costanzo et al. (2012, 2009), which as the authors argue followed the reaction mechism

depicted in Figure 5.1b, has since been under scrutiny as no other research group has been able to replicate their results. Later research (Morasch et al., 2014) compellingly suggested that short adenine polymers are generated during the freeze drying process of the industrial synthesis of commercial cAMP, therefore concluding that the polymerising capabilities of cAMP reported by Costanzo et al. were an experimental artefact. I agree with the view from Morasch et al. (2014), but Costanzo et al. have been robustly defending their work, and in any case the reaction pathway depicted in Figure 5.1b was, at least theoretically, a possibility.

As mentioned earlier, dehydration reactions are a subset of condensation reactions in which a molecule of water is being produced as a reaction product. Due to the equilibrium that exists between a dehydration reaction and the hydrolysis of the dehydration product, these reactions tend not to occur when the reaction solvent is water as the equilibrium is strongly shifted towards the reaction substrates (in our case, AMP monomers), and the concentration of dehydration product (small nucleotides) remains undetectable. This factor has strongly contributed to a number of researchers interested in the origin of life on Earth to postulate that life's first solvent was not water. Other researchers think water must have been the solvent where life first evolved and propose a series of mechanisms by which polymerisation reactions could happen in water. One of the possibilities are wet-dry cycles (happening on beaches for instance) where water is evaporated intermittently favouring polymerisation in the process; another is the usage of low water activity settings such as concentrated silicates submarine settings. Another possibility is that polymerisations were able to occur in water due to the chemical activation of the monomers by phosphates, thioesters and a series of other chemical groups. This monomer activation offers mainly a good leaving group, which crucially is not a water molecule.

The prebiotic reactions that have been successful in our research group and, generally speaking, on the research on the origin of life in hydrothermal vents occur as condensation reactions not occurring via dehydrations. In a condensation reaction, two molecules join together whilst a moiety other than

one of the previous molecules acts as a leaving group (water, orthophosphate, pyrophosphate, a thiol, etc.); if the leaving group is water, the reaction is commonly named 'dehydration'; if it is not, the reaction remains with the more generic name of 'condensation'. The synthesis of AcP from thioacetate and orthophosphate (Whicher et al., 2018) under simulated aqueous hydrothermal vent conditions was successful partly because the reactions occurred via the loss of H_2S, not water. The phosphorylation of adenosine and ribose by AcP (Whicher et al. 2018) also occurs via a non-dehydrating condensation with acetate as a leaving group.

The pool of biotic reactions in extant life in which two molecules join to form a larger product (condensations) occur via non-dehydrating mechanisms with, so far as one can tell from a consideration of intermediary metabolism, a single intriguing exception: ATP synthesis by the ATP synthase. For instance, in the amino acid polymerisation reactions occurring in the ribosome, the leaving group is the tRNA, which was covalently bonded to the incoming amino acid; in this case the leaving group is definitely not a small molecule, but it is still smaller than the growing polypeptide-tRNA nascent chain. Another example would be the synthesis of the amino acyl-tRNA complex itself, where first the amino acid is activated by an adenylation, with pyrophosphate acting as a leaving group; then the amino acid is bound to the tRNA and AMP (or adenylyl group) acts as a leaving group. There are plenty of other examples: in glycogen and starch polymers synthesis the leaving groups are UDP and ADP respectively. In the formation of the N-glycosidic bond of nucleosides the leaving group is pyrophosphate again. Nucleotide polymerisation is another example; the pyrophosphate of the incoming nucleotide is the leaving group. The list is endless but what all these condensation reactions have in common is that they do not happen via dehydrations. I am not proposing that all these biotic reactions are the way they are just because of the dehydration vs condensation dilemma, but it is certainly a factor; for instance it is well known that life keeps the concentration of pyrophosphate and ADP (and the other diphospho-nucleosides) to much lower concentrations (50-100 μM for ADP (Williams et al., 1993)) than ATP (1-10 mM); this strong disequilibrium is responsible for ADP and pyrophosphate acting as excellent leaving groups.

There are some biotic reactions where water is being produced: for instance the citrate – *cis*aconitate – D-isocitrate isomerisations, the enolase dehydration in glycolysis, and the fumarase reaction in the Krebs cycle. Despite these examples, dehydrating reactions are scarce in life and most probably only occur thanks to the catalytic aid of enzymes; furthermore none of these reactions are condensing (or polymerising) reactions.

The scarcity of dehydration reactions in life suggests that life was indeed cradled in a water-based solution and that the condensation reactions it needed had to have their genesis as non-dehydrating condensation reactions (otherwise they would not have occurred at all) and they were maintained as such as evolution went on. If life started in formamide (Pino et al., 2015) or urea (Powner et al., 2009) or within (or with the aid of) other organic solvents these reactions would have probably started as dehydrations as there would be no thermodynamic impediment that way, and water would act as the leaving group in many of the aforementioned reactions. If that was the case, then why did life later completely switch from one paradigm to the other? Some authors propose that life had to adapt to a water-based biochemistry as cataclysmic events (such as the late heavy bombardment) forced it to the ocean depths in order to survive. The very existence of the late heavy bombardment has been questioned (Boehnke & Harrison, 2016), and in the case that it actually occurred, this would imply in effect a second origin of life, as virtually all biochemistry would need to be invented all over again. It is obvious that, despite being possible, it is by far not the most parsimonious explanation.

As mentioned earlier, the ATP synthase represents a genuine dehydrating machinery, one that strikingly acts almost as a physical press in order to extrude water molecules and uses a flow of protons in order to turn, in much the same way that a turbine uses the physical force of water (falling due to the gravitational potential energy it possesses) to turn in a hydroelectric plant. Other reactions yielding ATP (substrate level phosphorylations) are not dehydrations: for instance in the reaction of 1,3-bisphosphoglycerate with ADP yielding ATP and 3-phosphoglycerate, the leaving group is always the molecular moiety that

was initially phosphorylated (1,3-bisphosphoglycerate in this case, leaving as 3-phosphoglycerate). The structure and reaction mechanism of the ATP synthase look to be a standout process: it seems as if the energy accumulated in these initial dehydrations is transferred to the rest of the cell as a series of non-dehydrating condensations (de Duve, 1991). The implications of this for the origin or life are beyond the scope of this thesis but would conceivably point to a mechanical origin of phosphate-mediated energy coupling at the origin of life (Branscomb et al., 2017; Branscomb & Russell, 2013), which has been maintained throughout evolution in the form of the ATP synthase, a molecular machine conserved across essentially all extant life. It is admittedly difficult to envision a direct series of homologous steps from the mechanical action of a mineral (such as green rust ($Fe(OH)_2$) and a molecular machine such as the ATP synthase; this apparent discontinuity hinders the acceptance of this idea. Maybe acetyl phosphate could be envisioned as a solution as it can be synthesised in solution and it can phosphorylate monomers in solution; unfortunately this did not lead to polymerisation under the experimental conditions tested. Despite these limitations, the biogenic importance of AcP is supported by triple-phosphorylated molecules (such as ATP) being successfully synthesised using AcP (Figures 4.19 and 4.20 in Section 4.2.3 of Chapter 4), as this narrows the discontinuity problem (from mono to tri-phosphorylated molecules); work on phosphorylating AMP to ADP is on-going within the research group but it is proving to be quite challenging. The formation of thioesters (such as methyl thioacetate) or thioacids (such as thioacetate), which are necessary for the hydrothermal synthesis of AcP (Figure 4.2) (Whicher et al., 2018) is arguably a more difficult critical point, since these molecules represent a pre-condensed state. We propose that they could be synthesised on Fe(Ni)S barriers (at low water activity setting) by tapping the natural pH gradient present at alkaline hydrothermal vents (Figure 3.1) (Camprubi et al., 2017).

I expected AcP to promote the polymerisation of the simple amino acid glycine. As in the case of AMP polymerisation, what I expected to obtain was a poor, albeit existing, polymerisation yielding glycylglycine or slightly longer oligomers when using solely AcP. Then, with the successive addition of simple mineral

catalysts or organically-chelated ones, the expected yield and/or length of the oligomers should increase, mimicking the putative natural evolution of catalysts. Instead, AcP tended to acetylate the amino group of amino acids (Figure 5.13), especially at more alkaline pH, as reported by others (Di Sabato & Jencks, 1961a). Its tendency to acetylate amino groups could partially explain why AcP is used less commonly than ATP to catalyse phosphorylation reactions in modern cells, and indeed in archaea AcP remains bound to the active site of acetyl CoA synthetase (Bräsen et al., 2008; Schonheit et al., 2016). If instead of N-phosphorylation of glycine, the reaction that had taken place were its O-phosphorylation, the AcP-mediated polymerisation of glycine would have happened more easily. The O-phosphorylation (technically an adenylation) by ATP of the carboxyl group of amino acids is the biological activating mechanism of amino acids, which will be subsequently mounted onto tRNA, yielding the aminoacyl-tRNA complex. The O-adenylation of amino acids (and not N-adenylation) is achieved by the catalytic properties of the active site of the enzyme (aminoacyl-tRNA synthetase), which modifies the pKa of the amino acid groups, promoting their ionisation (Carter, 1993). This ionisation allows for the amino group (which is now charged positively) to cease competing with the carboxyl (now negatively charged) for the subsequent adenylation/activation. An uncharged amino group is a strong nucleophile; even when the carboxyl is ionised, an uncharged amino group would still act as a competing nucleophile. Therefore, a putative prebiotic mechanism that promotes the double ionisation of amino acids (e.g. a neutral or slightly acidic pH) should positively impact the chances of a phosphorylation-mediated polymerisation; in our hands, this did proved not to be enough (Figure 5.13), presumably due to the strong N-acetylating properties of AcP.

Finally, it could be that condensation reactions really did begin through radically different mechanisms, which were later overwritten by enzymatic catalysis. If metabolism were ‘invented’ in an RNA world that would arguably need to be the case. On the other hand, if intermediary metabolism arose initially from geochemical flux, via conserved intermediates, then it might be that polymerizations in water are genuinely difficult to achieve, and arose later in a world of organic monomer catalysts. If that were the case, the fact that AcP

promoted the synthesis of ATP from ADP (as it still does in modern cells (Ferry & House, 2006; Schönheit et al., 2016)) might be important, as both amino acid and nucleotide polymerization depends on nucleotide triphosphates today. The concept of a 'monomer world' is not new (Copley et al., 2007), and could be highly structured, as the formation of fatty acids is favoured under a wide range of conditions (Amend et al., 2013), and spontaneously form bilayer vesicles when above a threshold concentration (Morowitz et al., 1988; Segré et al., 2001). The possibility of a rudimentary form of heredity, based on several linked positive feedbacks in growing protocells, is consistent with the idea that polymerization to form macromolecules occurred relatively late (West et al., 2017). If so, then AcP could potentially drive prebiotic chemistry towards a form of proto-metabolism that prefigures the metabolic pathways of modern cells, in which relatively unreactive precursors are activated through phosphorylation. AcP therefore begins to bridge the gap between prebiotic chemistry and monomer biochemistry in living cells, but does not directly promote the formation of macromolecules.

5.4 Conclusions

The fact that AcP did not drive the condensation of amino acids or nucleotides, even to a limited degree, was disappointing, but possibly significant. I had hoped that short polymers would form spontaneously in water in the presence of AcP. In the case of glycine, high yields (up to 80%) of N-acetylglycine (NAG) were generated, as reported by others (Di Sabato & Jencks, 1961a). NAG did not react any further, although acetylation of amino groups can promote the formation of short peptides in the presence of amino acid esters, AMP, a carbodiimide and 1-ethylimidazole, as acetylation of the amino group facilitates reaction of the carboxylic acid group instead (Griesser et al., 2017; Jauker et al., 2015). But modern cells do not promote polymerization using any of these condensing agents, nor through acetylation or phosphorylation of amino groups, as reported for diamidophosphate (Gibard et al., 2017), or acetylation of hydroxides (Bowler et al., 2013); nor do they do so through wet-dry cycles (Da Silva et al., 2015; Deguzman et al., 2014; Forsythe et al., 2015; Rajamani et al.,

2008). Polymerization of cyclic nucleotides in water at temperatures above 75 °C has been reported (Costanzo et al., 2012, 2009) but these findings have proved 'erratic' to repeat (Morasch et al., 2014; Šponer et al., 2015). Polymerization apparently occurs only in the absence of Na^+ ions, and ideally in a gradually desiccating environment (Šponer et al., 2015) that has little in common with polymerization in modern cells.

I believe that these findings (together with those reported in Chapter 4) make AcP a credible primordial energy currency, coupling carbon and energy flux at the origin of life. The fact that AcP promotes phosphorylation but not condensation reactions in water points to a period of monomer biochemistry before the emergence of polymeric enzymes or ribozymes at the origin of life (Copley et al., 2007). This conception is consistent with autotrophic origins and rudimentary heredity in growing protocells before the emergence of RNA, polypeptides, and true genetic heredity (West et al., 2017). The biochemical utility of AcP may be limited by its tendency to acetylate amino groups. But in driving the synthesis of ATP from ADP, AcP might have helped bridge the gap between monomer biochemistry and the origin of genetic replicators.

Chapter 6

The formose reaction as a primordial source of sugars

The work described in this Chapter will be published in Camprubi et al. (MS in preparation; 2018).

6.1 Introduction

Sugars have plenty of functions in living organisms but, despite this, are rarely deemed to be central on the origin of life. In cells, they are paramount to the conservation of energy both by photosynthesis and glycolysis; as polymers they serve as an energy store and have structural roles, such as cellulose and chitin; sugar derivatives (e.g. nucleotides) constitute the basis for the coding molecules of life (RNA and DNA) and also form part of many cofactors, essential for enzymatic catalysis. One of the most obvious route of prebiotic synthesis of sugars is through the formose reaction, first performed by Butlerow in (1861). In the formose reaction, formaldehyde undergoes a series of aldol condensation and isomerisation reactions to form sugars, including ribose, in alkaline medium (Figure 1.4). There is contention as to whether it is only formaldehyde that slowly reacts (since it is a poor nucleophile) to form glycolaldehyde and then these two react much faster, yielding the rest of the sugars, or whether the formose reaction cannot proceed as mentioned and indeed trace amounts of C2 or C3 sugars need to be added in order to get the reaction started in the first place (Kieboom & VanBekkum, 1984; Socha, 1980); this contention is not the focus of this Chapter so this will not be discussed further here. For my experiments I used pure formaldehyde (Fischer, 16%; commercial formulations containing higher % contain methanol as a stabilizing agent) but I did not assess whether trace amounts of other aldehydes were present as contaminants. In any case, in a heavily reducing mild hydrothermal environment, it is reasonable to assume that trace quantities of small (C2-3)

molecules could be abiotically synthesised by mechanisms analogous to those depicted in Figure 1.3 and 3.1.

As mentioned previously, a historical criticism of the relevance of the formose reaction, even under ideal hydrothermal conditions, is that it is frequently claimed to yield little more than tars and complex mixtures of biologically irrelevant sugars (Gollihar et al., 2014). Orgel (2004) acknowledged that if by a plausible mechanism the formose reaction was to be directed towards the synthesis of ribose, this would represent an ideal prebiotic route to the synthesis of the precursors of nucleotides. Later on, Mellersh & Smith (2010) suggested that this putative mechanism could be phosphorylation and that this would funnel the formose reaction towards the synthesis of ribose phosphate. Mellersh & Smith propose that the polarity of the phosphate moiety of glyceraldehyde 3-phosphate (phosphorylated from glyceraldehyde) would promote its retention on a mineral surface and its keto-enol isomerisation to dihydroxyacetone phosphate. At alkaline pH DHAP would easily enolise, which would promote the formation of a new C-C bond with a G3P molecule through a canonical aldol condensation; a series of stepwise analogous reactions would logically result in the synthesis of pentose (or hexose) phosphates, such as ribose 5-phosphate or glucose 6-phophate (Figure 6.1).

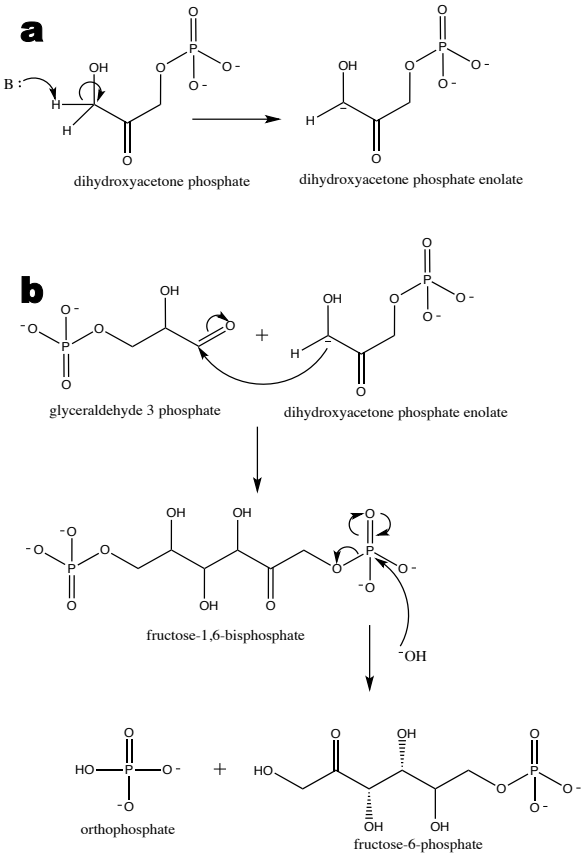


Figure 6.1. (**a**) Enolisation of dihydroxyacetone phosphate (DHAP) creating a reactive negatively charged carbon species. B represents a strong base, such as a hydroxide ion. (**b**) Enol condensation of glyceraldehyde 3-phosphate (G3P) with DHAP yielding fructose 1.6-bisphosphate, which can be subsequently hydrolysed to fructose 6-phosphate. Analogous reactions would yield a wide spectrum of sugars and sugar phosphates.

In essence, this series of phosphorylation-promoted isomerisations and aldol condensations could help transition from the *a priori* hardly biological formose reaction, to a 'sugar cascade' more reminiscent of anabolism (Figure 6.2). Mellersh and Smith proposed, rather cryptically, that a mechanism that selectively uses ribose 5-phosphate (and therefore removes it from the equilibrium) would funnel the formose reaction towards this biologically important phosphorylated sugar (Mellersh & Smith, 2010). Despite the alluring aspect of this hypothesis, it is admittedly difficult to envision a prebiotic mechanism that would selectively remove ribose 5-phosphate amongst all the possible products (and isomers) of this complex set of reactions, even more so in the absence of modern enzymes. In this Chapter I will explore whether the formose reaction takes place under simulated alkaline hydrothermal vent conditions, and whether AcP-mediated phosphorylations specifically promotes the synthesis of biologically important sugars, such as ribose or ribose 5-phosphate.

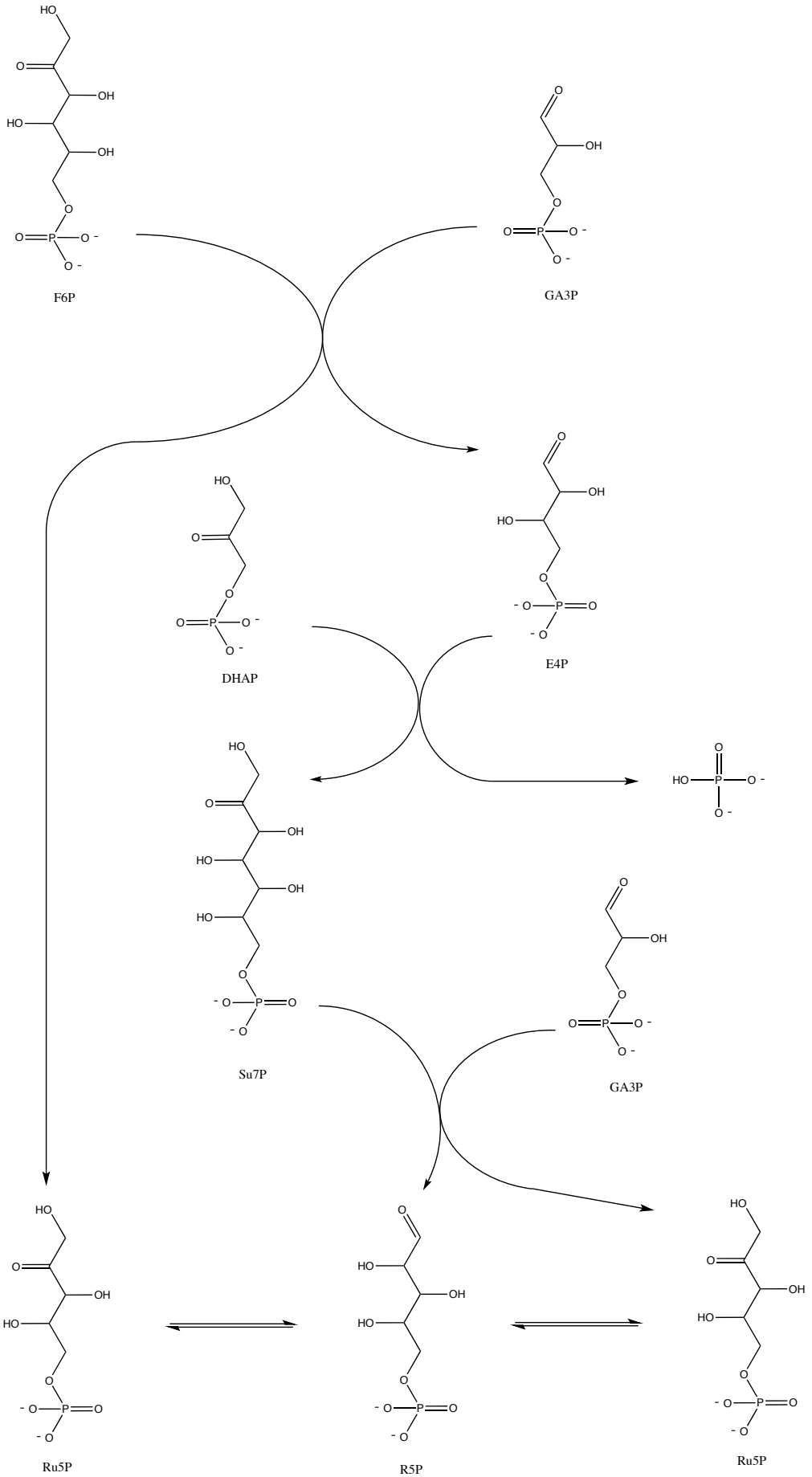

F6P
GA3P
DHAP
E4P
Su7P
GA3P
Ru5P
R5P
Ru5P

Figure 6.2. Sugar cascade promoted by phosphorylation during the formose reaction (amended from Mellersh & Smith (2010). Fructose 6-phosphate (F6P), synthesised as in Figure 6.1, reacts with G3P yielding ribulose 5-phosphate (Ru5P) and erythrose 4-phosphate (E4P). E4P reacts with DHAP, yielding sedoheptulose 7-phosphate (Su7P), which can react with G3P giving rise to Ru5P and ribose 5-phosphate (R5P). Alkaline-promoted tautomerisation causes ketoses (such as Ru5P) and aldoses (such as R5P) to readily interchange.

6.2 Results

My work on the formose reaction occurred in two sets of experiments. The first set was performed in order to assess whether the formose reaction could occur under alkaline hydrothermal vent conditions and whether it yields individual sugars, rather than the often-claimed complex (and biologically useless) tar; this data was analysed using GC-MS. The second set was analysed using HPLC-UV and LC-MS due to recurrent problems with the GC-MS setup used previously (see below); this set studied whether the addition of AcP on the formose reaction had any effect on the synthesised sugar distribution, and was also used to test for the presence of sugar phosphates.

6.2.1 Formose reaction

A GC-MS analytical method from Kopetzki & Antonietti (2011) was adapted for the first set of experiments, and a good chromatographic separation was achieved using this method (Figure 6.3).

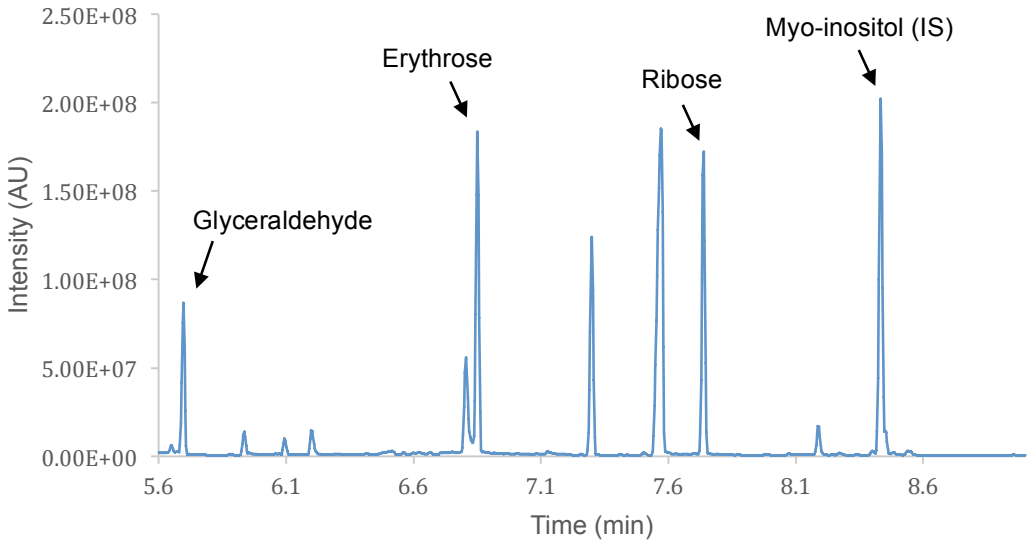


Figure 6.3. GC-MS chromatogram (acquired on full scan mode) of a mixture of all derivatised sugar standards and internal standard (IS). Solutions of 1 mM of each standard and IS were derivatised and 60 µL of the resulting DCM-based derivaitsation solutions were mixed together in a liquid injection vial. Other peaks correspond to derivatisation artefacts (mostly carbonated cyclic structures) but, because of their good separation and later usage of single ion monitoring (SIM) mode for acquisition, they did not affect quantification in experimental samples.

Using the NIST 14 Mass Spectral library, which is a fully evaluated collection of electron ionization (EI) mass spectra, it was possible to assign many chromatographic peaks found experimentally. For instance, one of the first peaks corresponded to the derivatisation reagent N-methyl imidazole (the derivatisation reaction catalyst) (Figure 6.4). Peaks from 4.5–7.5 minutes corresponded to the various sugars, which were further confirmed by comparison of the retention times of commercial standards (as their fragmentation patterns were almost identical). Some peaks were also attributed to derivatisation artefacts (mostly cyclic carbonated structures (Nygren et al., 1996)).

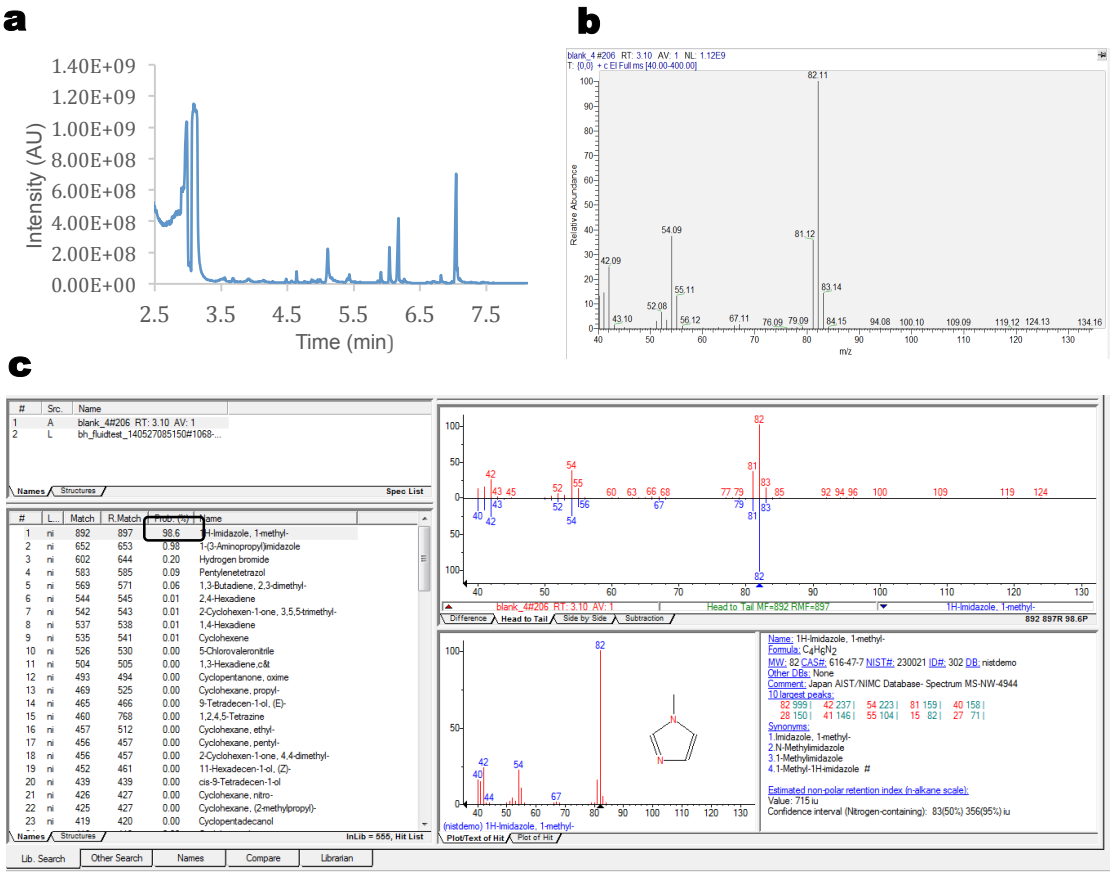


Figure 6.4. A GC-MS peak assignation example. (**a**) GC-MS trace of a derivatised HPLC water sample (negative control) shows the early peaks corresponding to the derivatisation reagents (2.8–3.1 min), as well as scattered peaks that appeared later corresponding to derivatisation artefacts. (**b**) MS spectrum of the peak at retention time 3.1 min. (**c**) Example of NIST-14 Mass Spectrometry library usage where the spectrum from (b) is attributed to N-methyl imidazole with a probability of 98.6% (black square). The red spectrum corresponds to the NIST-recorded spectrum of commercial N-methyl imidazole; the blue one corresponds to the peak experimentally found at 3.1 min.

An example of a peak assignation is depicted in Figure 6.5, showing the comparison between commercial D-ribose and experimentally synthesised ribose (this analytical method could not separate D and L isomers of a given sugar, but it is reasonable to assume that the D/L ratios of non-enzymatically synthesised sugars would be of 50% each).

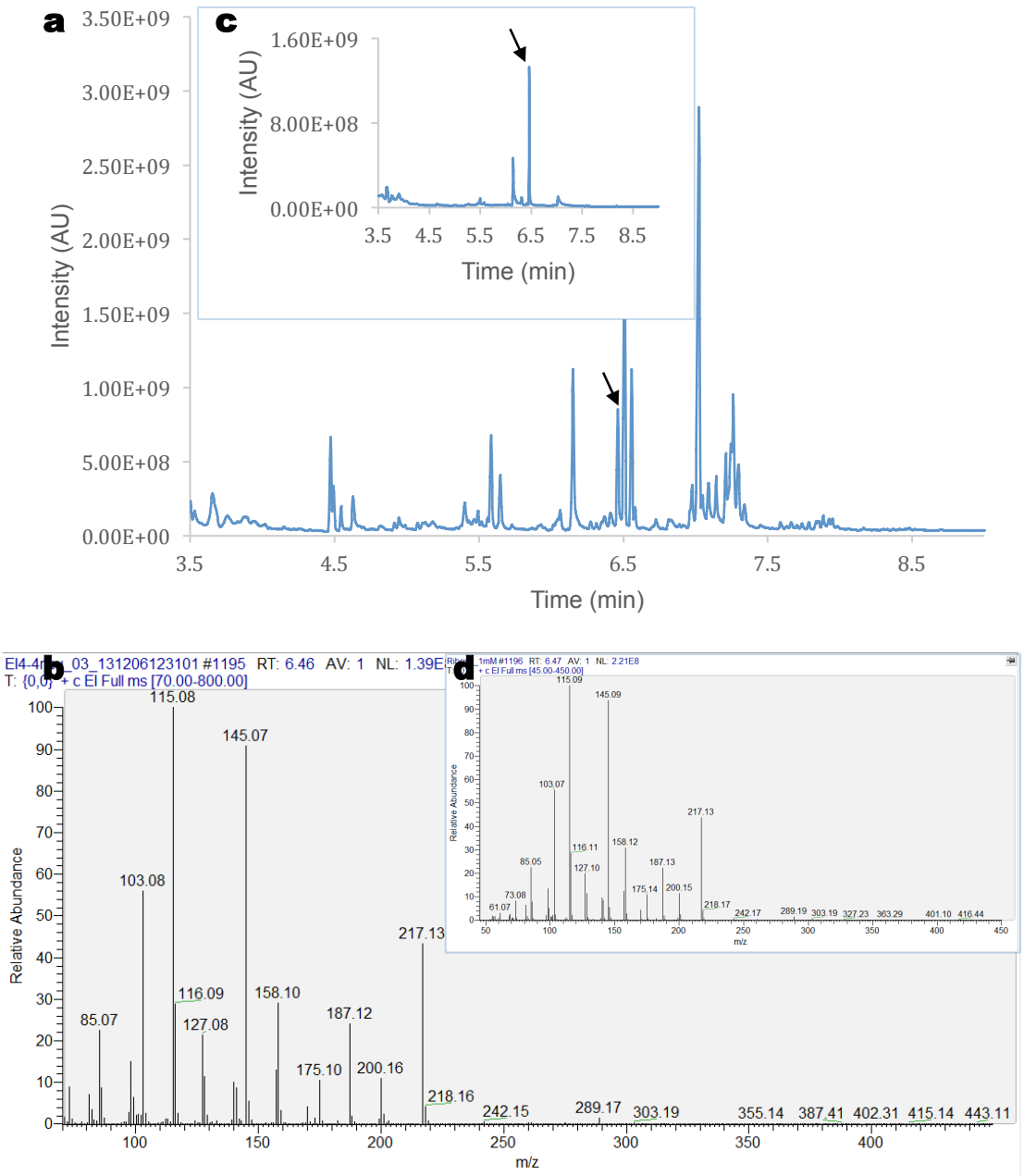


Figure 6.5. GC chromatograms and mass spectra comparison between formose reaction product eluting at 6.46 min (**a**-**b**) and D-ribose standard (500 μM) (**c**-**d**). Peak eluting at 6.46 min (±0.07 min) was assigned to ribose (black arrows) due to the coincident retention times and fragmentation patterns of the reaction product and standard.

The retention times of the commercial standards of all sugars assessed is shown in Table 6.1. This analytical method was not suitable for C2 sugars, as the derivatised product of glycolaldehyde (C2) was not detected, even as a commercial standard.

Table 6.1. GC retention times of commercial standards of sugars.

Compound (C atoms)	Retention time (min)
Glyceraldehyde (3C)	4.47
Erythrose (4C)	5.65
Deoxyribose (5C)	6.02
Ribose (5C)	6.47
Arabinose (5C)	6.51
Lyxose (5C)	6.51
Xylose (5C)	6.56
Myo-inositol (IS) (6C)	6.82
Fructose (6C)	7.25

The ratio of the peak area of the analyte to the internal standard (myo-inositol) was routinely calculated for quantification purposes in order to improve data reproducibility. Due to time constraints, not all sugars were systematically quantified using the GC-MS setup; the quantified sugars were glyceraldehyde, erythrose and ribose. Only trace amounts of deoxyribose were found experimentally and, because the integrated experimental areas were smaller than the analytical quantification limit, it was not quantified (Table 6.2). Amongst the aldopentoses (containing an aldehyde group, not a ketone) analysed, ribose was successfully separated from the other diastereomers (arabinose, lyxose and xylose). In all these experiments, ribose constituted roughly ¼ of all aldopentoses, as expected.

Ketoses (sugars containing a ketone group) could not be quantified using this method since some of their derivatisation products were shared with those from some others. For instance, fructose, which is a ketopentose, yielded two overlapping peaks corresponding to mannitol and glucitol when derivatised. This is due to the fact that when a ketone group is reduced, two possible isomers are produced depending on which side of the carbonyl plane the nucleophile (acetic anhydride) attacks from. This is not an issue for quantification, as the integrated areas could be calculated and treated as one; the problem is the interference with glucose, which is also a product of the formose reaction. Glucose is an aldohexose that, when reduced, also yields

glucitol, which would be quantified together with the fructose reduction products, thus hindering quantification of the latter. Following this logic, a ketopentose, such as ribulose (5C), would yield the same peak as ribose (an aldopentose). The fact that ketoses could not be quantified independently during this first set of GC-MS experiments helps explain the results of the experiments that were analysed using the LC-MS (Section 6.2.2); this will be discussed in detail later (Section 6.3).

Table 6.2. Calibration curves for sugar standards (50 nM - 1 mM).

Compound	Linearity	Lower limit of quantitation	r^2
Glyceraldehyde	As/Ai = 4223.9c + 0.1072[a]	1 µM	0.999
Erythrose	As/Ai = 6241.5c + 0.0349	0.5 µM	0.999
Deoxyribose	As/Ai = 7793.1c - 0.0188	0.5 µM	0.996
Ribose	As/Ai = 7318.5c + $9x10^{-5}$	0.1 µM	1

[a]As/Ai: peak area ratio of the analyte to the internal standard; c: concentration in µM

Figure 6.6 shows the first quantification data acquired for a formose reaction. The experimental conditions were the following: 0.5 M formaldehyde, 0.16 M $Ca(OH)_2$, 0.16 M $CaCO_3$, 60 mL reaction volume, 60 °C, and initial pH = 12.3. Due to recurrent problems with the GC-MS instrument, these data only represents one experimental run, so exact quantification values should not be taken into account; instead, these results suggest that individual (monomeric) sugars can indeed be synthesised by the formose reaction, and that they are maintained as such for a long time (5 days) under cooler conditions (which, as discussed, hydrothermal vents provide). The maximum yield of ribose was of only 0.06%. Note that this, which seems minor, is only one of the multiple sugars synthesised by a large set of reactions, which are commonly simplified (misleadingly) as a single reaction pathway.

5 HCHO (formaldehyde, 30.03 g/mol) $\longrightarrow$ $C_5H_{10}O_5$ (ribose, 150.13 g/mol)

Initial formaldehyde: $0.5\,\frac{moles}{L}\,x\,30.03\,\frac{g}{mol}\,x\,0.06\,L = 0.9009\,grams$

Final ribose: $60.35\,\mu\frac{moles}{L}\,x\,1\,\frac{mol}{1E6\,\mu moles}\,x\,150.13\,\frac{g}{mol}\,x\,0.06\,L = 0.00054\,grams$

$(\frac{0.00054\ grams\ ribose}{0.9009\ grams formaldehyde})\ x\ 100$ = 0.06 % of initial formaldehyde transformed into ribose.

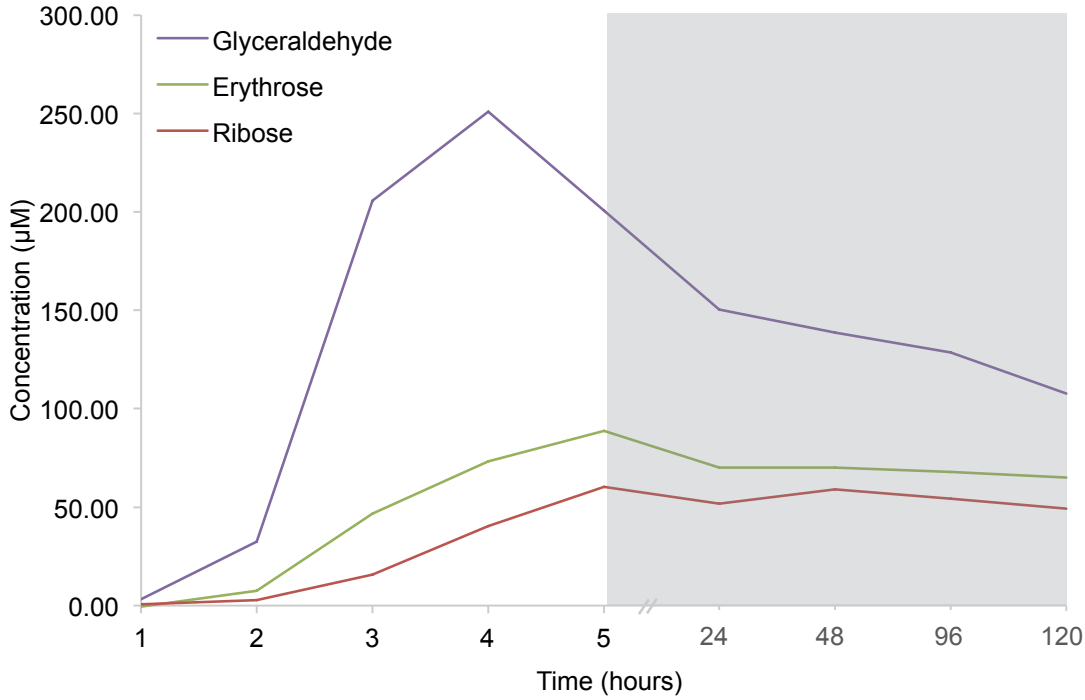


Figure 6.6. Change in concentration of formose reaction products over 120 h. The shaded area (5–120 h) represents the reaction mixture left at ambient temperature (20 °C), whilst the unshaded area represents the reaction mixture at 60 °C. N=1.

A high initial pH of around 12.3 was necessary for this reaction since reducing the initial pH to 10, either by using lower concentrations of $CaCO_3$/$Ca(OH)_2$ or titrating the pH using HCl, lowered the yield of sugars by 10-fold (not shown). Furthermore, the addition of boron (Furukawa et al., 2013; Holm et al., 2006; Ricardo et al., 2004) has been associated in the literature with the funnelling of the formose reaction towards pentoses but, in our hands, it had no such effect (not shown).

6.2.2 AcP addition to the formose reaction

6.2.2.1 Initial tests using GC-MS analysis

AcP was added to the formose reaction in order to test the predictions made by Mellersh & Smith (2010). Undergraduate student Hebe Wildi performed these initial exploratory tests under my supervision. We decided to use commercial AcP in order to eliminate the potential effects of the high phosphate and acetate concentrations that the prepared AcP could have. Due to the high cost of commercial AcP, the reaction volume was reduced to 2 mL for these experiments. Volumes of 800 µL of 1.5 M AcP were added to the formose reactions when applicable, giving a final AcP concentration of 400 mM. AcP solutions act as a strong buffer (around pH 7) so, once AcP was added the pH of the reaction had to immediately be titrated up using 5 M NaOH (approximately 200 µL). So, adding AcP to a formose reaction changed its final volume (to 3 mL), and its pH (albeit just for a few seconds). In order to ensure that these two factors did not affect the outcome of these experiments (and that any changes measured were due only to the chemical properties of AcP), every time that AcP was added to a formose reaction, a set of control reactions were also performed: (i) an unaltered formose reaction whose reaction pH was monitored in order to adjust the pH of all the other experimental reactions accordingly; and (ii) a formose reaction where AcP was not added (nor the 200 µL of 5 M NaOH), but the same volume of HPLC water (to a total of 3 mL) was added in order to mimic the dilution effect that the addition of AcP (plus subsequent titration) had on the experimental reactions. The addition of HPLC water to the second type of control experiments did not alter the reaction pH so, their pH never had to be adjusted. Therefore, when assessing the effect of the addition of AcP to a formose reaction, these results will always be compared to the corresponding control reaction.

In the first experiments, AcP was added immediately (time 0) but this seemed to prevent the formose reaction from beginning at all, and no sugars could be detected at any point during the next 5 hours (not shown), even after titrating the pH immediately back to 12.5. This presumably reflects the fragility of the formose reaction during its first minutes, when formaldehyde needs to act as a

nucleophile (whilst not being a good one); it is conceibable that the addition of a large amount of AcP provides better nucleophiles, so that the reaction cascade between formaldehyde molecules never starts.

Figure 6.7 shows the results from the first set of experiments in which AcP was added after 30 minutes of reaction (black arrow). In these control formose reactions (Figure 6.7, dashed lines) the yields for each sugar quantified are about 6 times greater than those shown in Figure 6.6; not just that, but the overall reaction dynamics are slower in Figure 6.6, reaching the maximum yield of ribose at 5 h (rather than at 1 h as in Figure 6.7). These discrepancies most probably reflect the change in the reaction volume (from 60 mL to 3 mL); the high volume of 60 mL would be completely heated up to 60 °C considerably slower than only 3 mL, therefore slowing down the whole reaction cascade did. Remarkably, the maximum yield of each sugar quantified was considerably higher when AcP was added to the reaction (Figure 6.7, solid lines). The increase in maximum yield was of 2.7-fold in the case of glyceraldehyde, of 3.3-fold in the case of erythrose, and of 5.24-fold in the case of ribose. So, not only all the maximum yields increase, but also there seems to have been an AcP-mediated selective effect, with ribose concentration being preferentially raised.

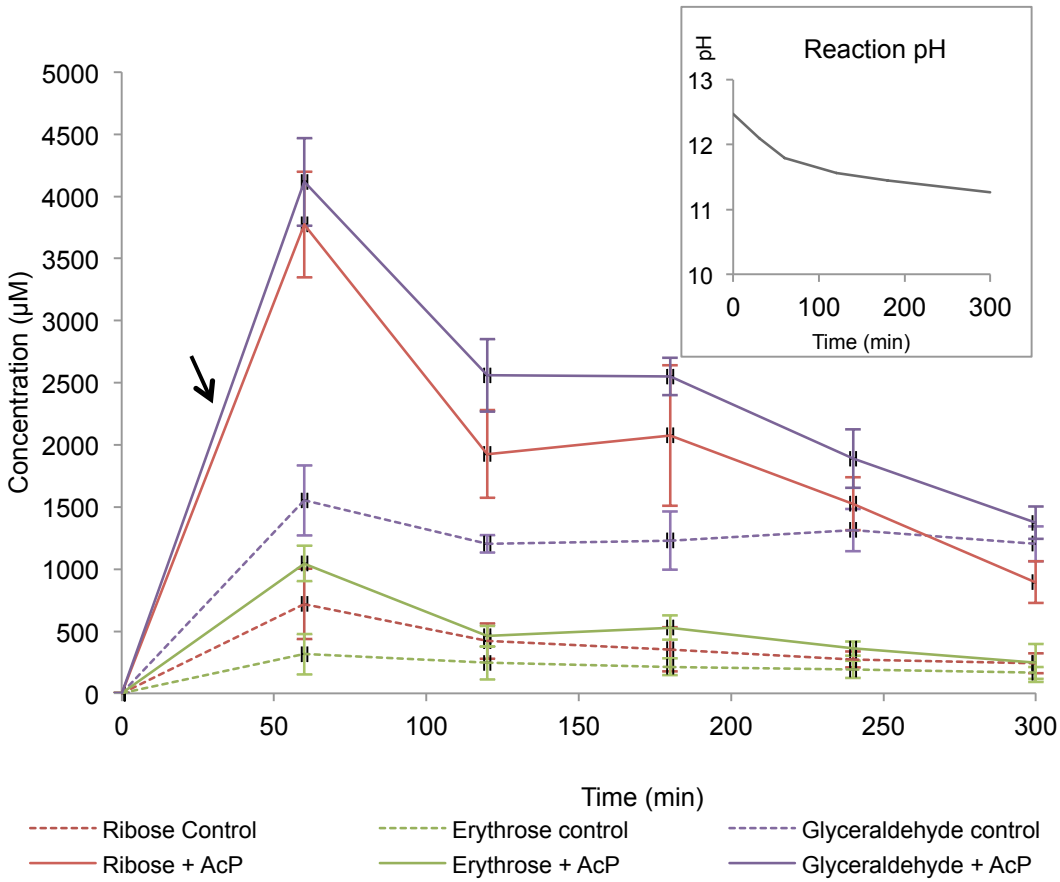


Figure 6.7. Quantification of formose reaction products ribose (5 carbon), erythrose (4 carbon) and glyceraldehyde (3 carbon) over 5 hours of reaction (0.5 M formaldehyde, and 0.167 M $Ca(OH)_2$ and $CaCO_3$, at 60 °C). Dashed lines show control formose reactions; solid lines show experimental runs in which AcP was added to the reaction. The black arrow indicates the point of AcP addition (400 mM). The graph insert shows the pH of the reactions. The pH of the experimental reactions was maintained accordingly to the pH of the control reactions. N=2 ±SD. Data provided by Hebe Wildi under my supervision.

These results seemed to be in line with what Mellersh & Smith (2010) had predicted: that the addition of a phosphorylating agent would funnel the formose reaction towards pentose sugars. Unfortunately, due to continuous instrumental problems with the GC-MS apparatus, the analytical method had to be changed. This setback offered the opportunity to test for the presence of sugar phosphates (putatively phosphorylated by AcP) since the GC-MS analytical method was not suitable for the analysis of sugar phosphates.

6.2.2.2 Later tests using HPLC-UV analysis

Some time later, professor Nick Lane's research group purchased a new HPLC instrument (Agilent 1260 Infinity II) and, together with sporadic Thermo LTQ LC-MS tests, this allowed for the reprisal of the formose reaction experiments. The change in analytical method was a setback time-wise, but also allowed for a deeper study of the AcP-mediated changes on the formose reaction. For instance, sugar phosphates were hard to quantify using the GC-MS setup since both their volatility and ionisation efficiencies were negligible, and thus the detection of putative AcP-synthesised sugar phosphates during the formose reaction was never successful on the GC-MS. As discussed in Chapter 5, AcP is a strong acetylating agent and, even though acetylated molecules are very GC-MS-friendly, because the derivatisation for GC-MS analysis consisted in acetylating the hydroxyl groups of sugars, it would have been virtually impossible to differentiate between experimental acetylations produced by AcP and acetic anhydride-mediated acetylations produced by the derivatisation procedure. The derivatisation method for HPLC-UV and LTQ LC-MS is described in Section 2.1.5 of Chapter 2; suffice to say here that it involved the derivatisation of sugars using 3-amino 9-ethylcarbazole (AEC) as in the AcP-mediated ribose phosphorylation experiments (Section 4.2.2, Chapter 4), with the addition of a N_2-based drying of the samples prior to derivatisation, which allowed for the complete removal of formaldehyde (which is volatile and was present in high concentrations within the formose reaction samples and would have interfered with the derivatisation using AEC).

In order to ensure that N_2-drying would remove the formaldehyde from the formose reaction samples two experiments were performed. The (i) first consisted of assessing the percentage of formaldehyde removed after the drying (and later rehydration) process. A total of 10 samples consisting of 80 µL of 1 mM formaldehyde were prepared; 5 of them were completely N_2-dried (for 22 min) and rehydrated with 80 µL of water, whilst the remaining 5 were kept capped in the fridge as controls. All 10 samples were then derivatised and analysed on the GC-MS instrument as detailed in Section 2.4.6 of Chapter 2. The integrated areas of each group were averaged and their comparison

showed that only 8.16% of the initial formaldehyde was present in the samples that were N_2-dried, so that a 91.83% was removed. The (ii) second test consisted of assessing whether some sugars could be lost during the N_2-drying process (this should not happen as sugars are barely volatile, but it was worth checking in case they experienced unexpected reactions during the dehydration-rehydration process). A total of ten 80 µL aliquots of 1 mM of both commercial ribose and commercial ribose 5-phosphate were prepared. Five samples of each sugar were N_2-dried (for 22 min) and then rehydrated, whilst the remaining five were kept capped in the fridge as controls. Then, all 20 samples were derivatised and analysed on the HPLC-UV instrument as explained in Sections 2.1.3 and 2.4.4 of Chapter 2 respectively, and the integrated areas were compared with the respective controls. The results were that 97.23 % of initial ribose, and 95.66 % of initial ribose 5-phosphate were recovered after the drying process. In conclusion, the results of both tests were satisfying: most formaldehyde was being removed during the drying, whilst the sugars and sugar phosphates remained intact, so the experiments on the effect of the addition of AcP on the formose reaction moved on.

As usual, calibration curves were prepared in triplicate for each sugar (Table 6.3); an example of a calibration curve is shown in Figure 6.8. Commercial fucose was again used as an external standard. Because no sugar phosphates could be detected in any formose reaction samples, neither by HPLC-UV or LTQ LC-MS, calibration curves for them were not prepared (see below).

Table 6.3. Calibration curves for HPLC-UV quantification (1 µM - 2 mM).

Compound	Linearity	Lower limit of quantitation	r^2
Glyceraldehyde	As/Ai = 0.0015c – 0.024[a]	1 µM	0.9980
Erythrose	As/Ai = 0.0014c – 0.001	1 µM	0.9998
Ribose	As/Ai = 0.0021c + 0.0014	1 µM	0.9999
Glucose	As/Ai = 0.002c – 0.0062	10 µM	0.9997

[a]As/Ai: peak area ratio of the analyte to the external standard; c: concentration in µM

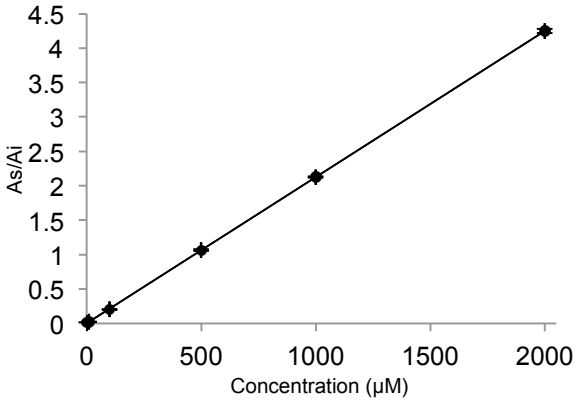


Figure 6.8. Calibration curve (1-2000 µM) for commercial ribose, corrected using fucose as external standard (y axis). N=3 ±SD.

An example of the HPLC chromatograms for the 8 sugar and sugar phosphate standards, and a real formose reaction sample is shown in Figure 6.9.

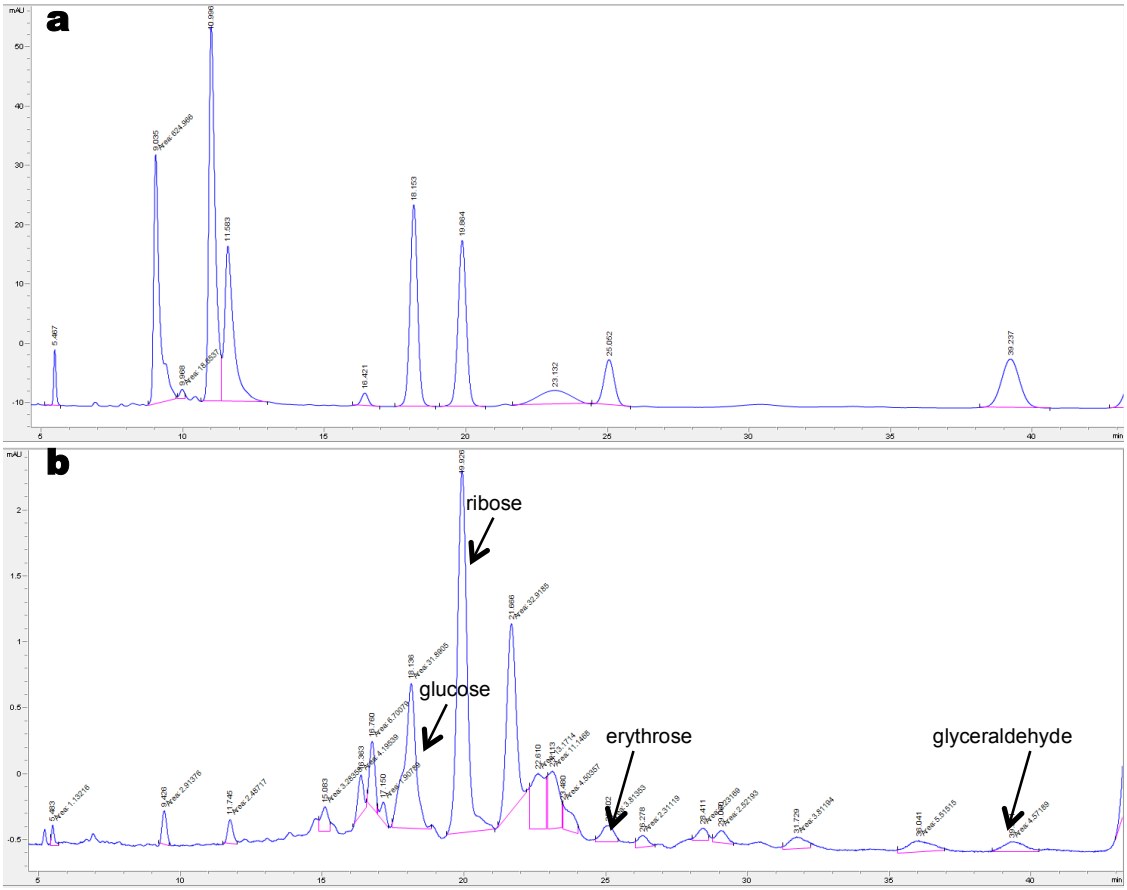


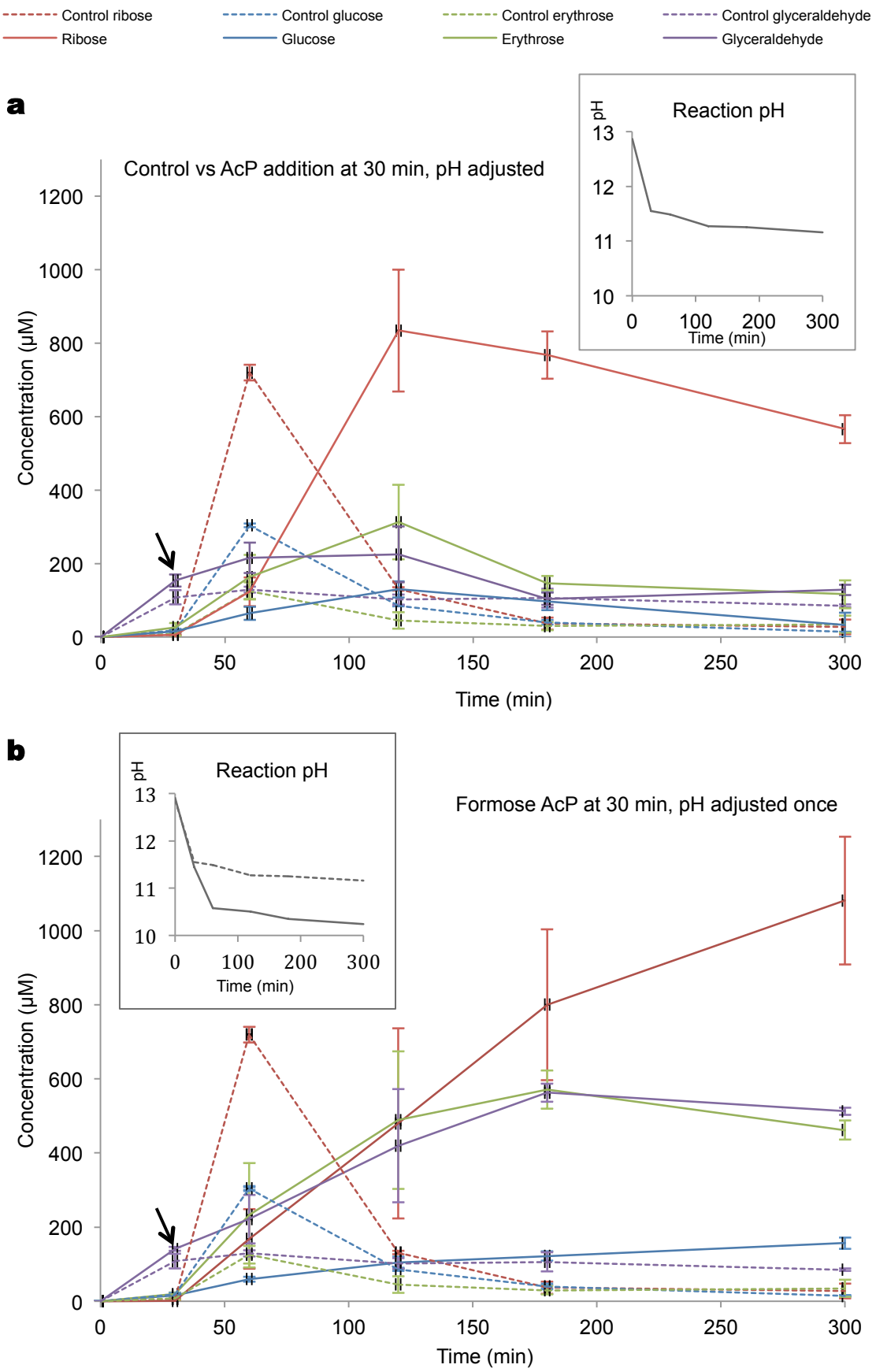
Figure 6.9. HPLC-UV chromatograms of: (**a**) a mixture (250 µM) of 8 sugar standards (glucose 6-phosphate, 9.035 min; ribose 5-phosphate, 9.968 min; erythrose 4-phosphate, 10.996; glyceraldehyde 3-phosphate, 11.583; glucose, 18.153; ribose, 19.864; erythrose, 25.052; glyceraldehyde, 39.237). (**b**) A formose reaction sample (0.5 M formaldehyde, and 0.167 M $Ca(OH)_2$ and $CaCO_3$, at 60 °C) after 1 hour of reaction.

A series of formose reaction experiments were performed in order to (i) repeat the results from Figure 6.7, and to assess the individual effects of (ii) AcP-mediated chemistry on the reaction, and of (iii) the effect of pH changes during the addition of AcP; these results are shown on the four graphs of Figure 6.10. When adding AcP at 30 minutes of reaction and adjusting the pH of the vials where AcP was added throughout, the result was a markedly different reaction profile overall (Figure 6.10a, solid lines). The whole reaction seemed to proceed slower, with sugars peaking at later reaction times compared with the controls (Figure 6.10a, dashed lines). The maximum yields of glyceraldehyde, erythrose and ribose marginally increased, whereas that of glucose decreased. A striking effect of the addition of AcP was that high concentrations of ribose were maintained during 5 hours whereas the concentration of the other sugars quickly collapsed. In order to assess the effect of the reaction pH, the same experiments were repeated but this time, the pH of the reactions where AcP was added were only adjusted (according to the pH of the control reactions) right after the addition of AcP (at 30 minutes) but left un-titrated thereafter (Figure 6.10b). The result was that the reaction profile slowed down even more, with maximum yields of glyceraldehyde and erythrose at 3 hours, and ribose peaking at 5 hours (and potentially kept growing thereafter); the effect on glucose was also positive, but much more modest. Looking at Figure 6.10b, it seems clear that the reaction pH played a role in slowing down the formose reaction, and allowing for the accumulation of intermediate sized sugars (C3-4), such as those tested.

The measured slowdown of the synthesis of ribose shown in Figure 6.10a-b could also be explained by the differing ionic strengths present between the experimental and control formose reactions. The addition of AcP (charged at high pH, and added as a Li^+/K^+ salt) increases the ionic strength of the reaction, an effect that could not be controlled by the addition of the same volume of water in the control reaction. Higher ionic strength could have slowed down the rate of ribose phosphorylation because the additional charged species interact with charged AcP, thus reducing the chances of AcP productively colliding with ribose. In order to elucidate whether ionic strength played a role in the observed effect, a series of control formose experiments with the addition of salty water

should be performed. To this end, a range of aqueous solutions with 1-800 mM NaCl in order to mimic the addition of AcP ionic strength-wise should be used (400 mM of AcP implied 400 mM of both K^+ and Li^+ ions). If the addition of high concentration of salt were to decrease the rate of ribose phosphorylation this would strongly suggest that, at least partially, the effect of the addition of AcP to the formose reaction could be explained by the modification of the ionic strength of the solution.

When 400 mM of AcP were added to a formose reaction vial, the reaction pH dropped down to 8.5 for around 30 seconds, which is the time it took to titrate it back up according to a control reaction performed alongside (and where the same volume of HPLC water had been added). In order to make sure that this brief drop in pH was not responsible for the effects observed in Figure 6.10a-b, the same experiment was repeated but, this time, no AcP was added (the same volume of HPLC water was added instead), and the pH of the reaction was artificially titrated down for 30 seconds using HCl (5 M), mimicking the effect of AcP addition, followed by a titration up to 11.5 (using 5 M NaOH), as it was done when adding AcP (Figure 6.10c). The difference in pH changed the reaction profile only marginally: the concentration of each sugar was maintained slightly higher for longer periods, but the yield peaks for each sugar remained the same as the control reaction. These results support the interpretation that the effects seen in Figure 6.10a-b reflect a chemical AcP-mediated effect on the formose reaction, rather than just a pH-related effect linked to its addition. In order to make sure that the results shown in Figure 6.10c were not due to the specific usage of HCl, this experiment was repeated using phosphoric acid (5 M) instead (Figure 6.11a). When using phosphoric acid instead of HCl the results were basically identical, reinforcing the interpretation that the phenomena observed in Figure 6.10a-b were due to the chemical reactivity of AcP and not its buffering effect. Finally, another experiment, leaving the reaction totally un-titrated after the addition of AcP (instead of titrating it right up to 11.5) was preformed (Figure 6.11b). As expected, the results indicated that the formose reaction slows down and then stops when the pH drops below 8.5.


Control ribose
Ribose
Control glucose
Glucose
Control erythrose
Erythrose
Control glyceraldehyde
Glyceraldehyde
a
Control vs AcP addition at 30 min, pH adjusted
Reaction pH
pH
Time (min)
Concentration (μM)
Time (min)
b
Formose AcP at 30 min, pH adjusted once
Reaction pH
pH
Time (min)
Concentration (μM)
Time (min)

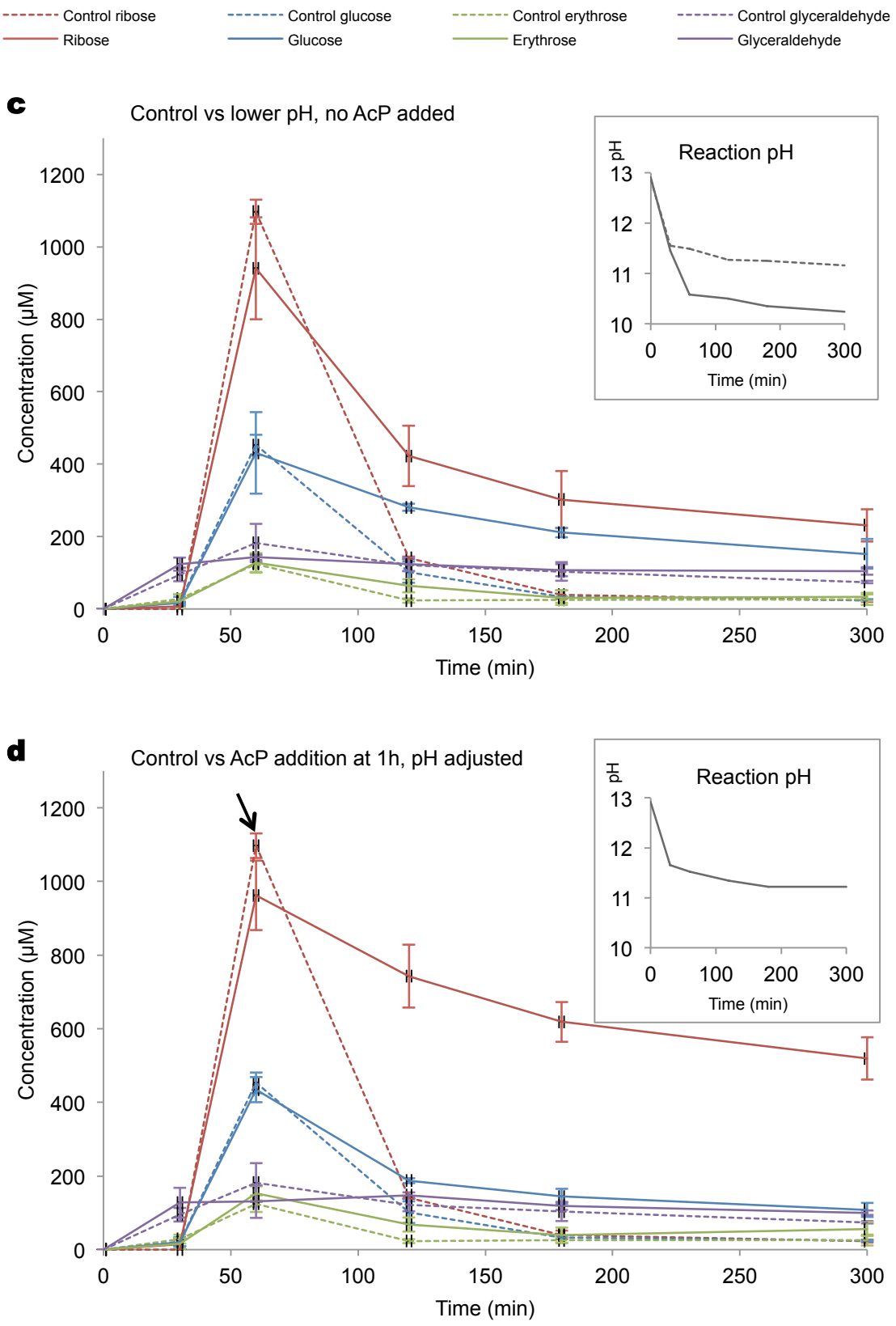

Control ribose
Ribose
Control glucose
Glucose
Control erythrose
Erythrose
Control glyceraldehyde
Glyceraldehyde
c
Control vs lower pH, no AcP added
Concentration (µM)
Time (min)
Reaction pH
pH
Time (min)
d
Control vs AcP addition at 1h, pH adjusted
Concentration (µM)
Time (min)
Reaction pH
pH
Time (min)

Figure 6.10. Quantification of formose reaction products glucose (6 carbon), ribose (5 carbon), erythrose (4 carbon) and glyceraldehyde (3 carbon) over 5 hours of reaction (0.5 M formaldehyde and 0.167 M $Ca(OH)_2$ and $CaCO_3$, at 60 °C). Dashed lines show control formose reactions; solid lines show experimental runs. Black arrows indicate the point of AcP addition (400 mM). Graph inserts show the pH of each reaction; dashed lines indicate pH of controls if different from experimental reactions. (**a**) Formose reaction with the addition of AcP after 30 minutes of reaction; the reaction pH was maintained as in the control reaction. (**b**) Formose reaction with the addition of AcP after 30 minutes of reaction; the pH was titrated immediately after the AcP addition only. (**c**) Formose reaction where the pH was titrated down (with 5 M HCl) mimicking the effect of the addition of AcP; no AcP was added. (**d**) Formose reaction with the addition of AcP after 1 hour of reaction; the reaction pH was maintained as in the control reaction. All 'reaction pH' graph inserts should technically reflect a 30 seconds pH dip for each experimental run as shown in Figure 6.11a; this is not shown here for clarity. N=3 ±SD.

Going back to Figure 6.10, the effect of the addition of AcP after 1 hour of reaction (instead of at 30 minutes) was also tested (Figure 6.10d). As expected, before the AcP addition there were no differences between experimental and control runs. Afterwards, the concentration of each sugar was similar to each respective control, again with the striking exception of ribose, whose concentration declined only slowly over 5 hours.

In conclusion, the reaction pH did slow down the formose reaction in an inversely proportional fashion, and affected each sugar equally. In contrast, the addition of AcP seemed to preferentially boost the ribose concentrations. This effect could be achieved through promoting its synthesis, inhibiting its further polymerisation into bigger sugars, promoting the retro-aldol breakdown of larger sugars yielding ribose, or a mixture of these effects. Interestingly, apart from the experiment shown in Figure 6.10b (where reaction pH was not corrected for), no experimental treatment increased the yield of ribose compared with the control reaction – a striking difference to the earlier results using GC-MS performed by Hebe Wildi under my supervision (previous sub-Section). This discrepancy is discussed in detail in Section 6.3 of this Chapter.

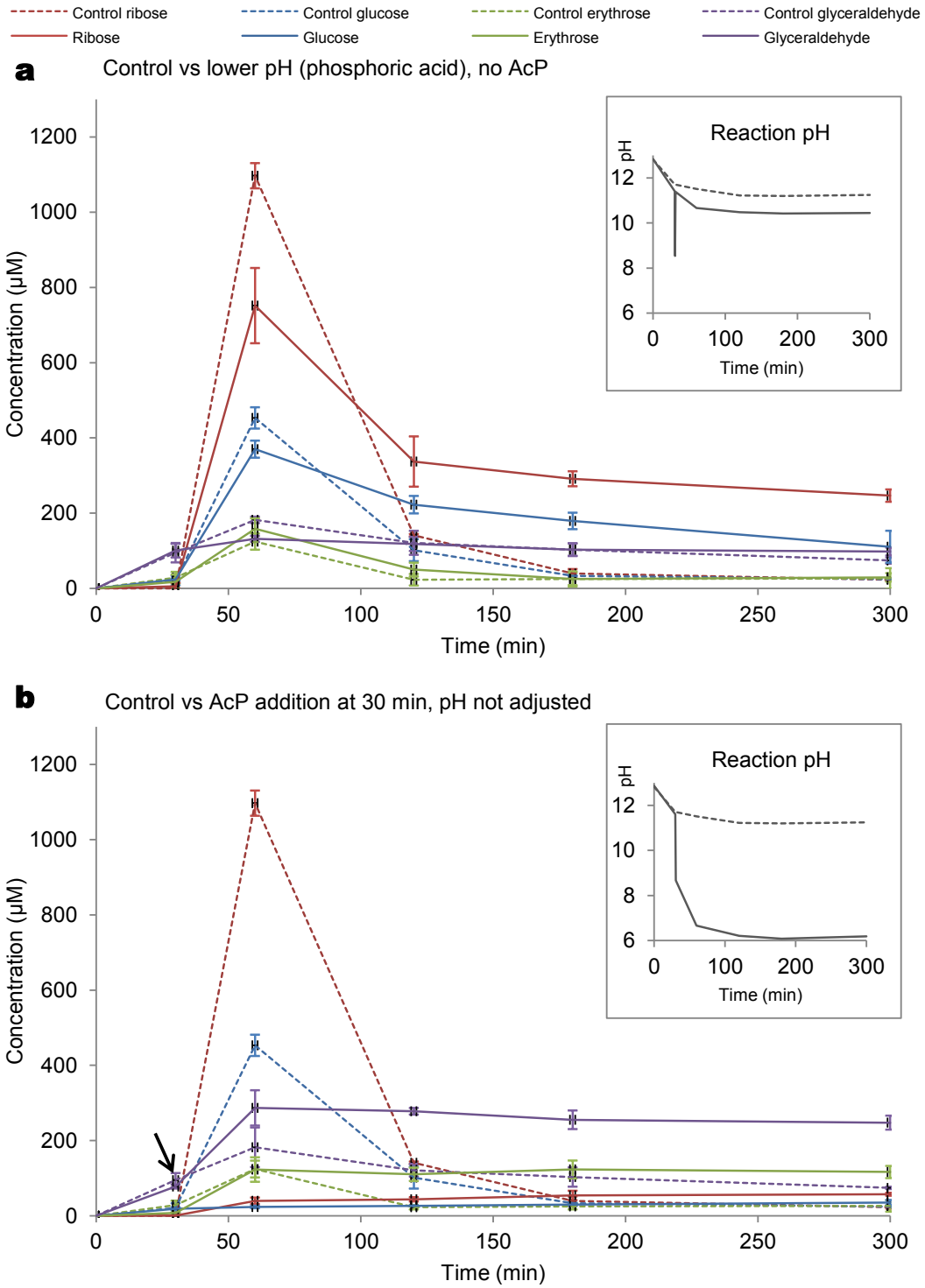


Figure 6.11. Quantification of formose reaction products over 5 hours of reaction (0.5 M formaldehyde and 0.167 M $Ca(OH)_2$ and $CaCO_3$, at 60 °C). Dashed lines show control formose reactions; solid lines show experimental runs (this also applies for graph inserts showing reaction pH). The black arrow indicates the point of AcP addition (400 mM). (**a**) Formose reaction where the pH was titrated down (with 5 M H_3PO_4) mimicking the effect of the addition of AcP; no AcP was added. (**d**) Formose reaction with the addition of AcP after 30 min of reaction; the reaction pH was not titrated at any point. N=3 ±SD.

6.2.2.3 Final tests using LC-MS analysis

The experiments on the addition of AcP to the formose reaction analysed using the HPLC-UV setup yielded interesting data, but the depth of information that a UV detector can provide is rather limited. Due to this, further exploratory tests were planned on the LTQ LC-MS setup. The samples were N_2-dried in order to eliminate formaldehyde, derivatised and purified as explained in the previous Section for HPLC-UV analysis; the LC-MS analytical conditions are detailed in Section 2.4.5 of Chapter 2. Due to time constraints, no quantification was attempted for these experiments. Figure 6.12 shows a chromatogram for each non-phosphorylated (Figure 6.12a) and phosphorylated (Figure 6.12b) sugars analysed using this method.

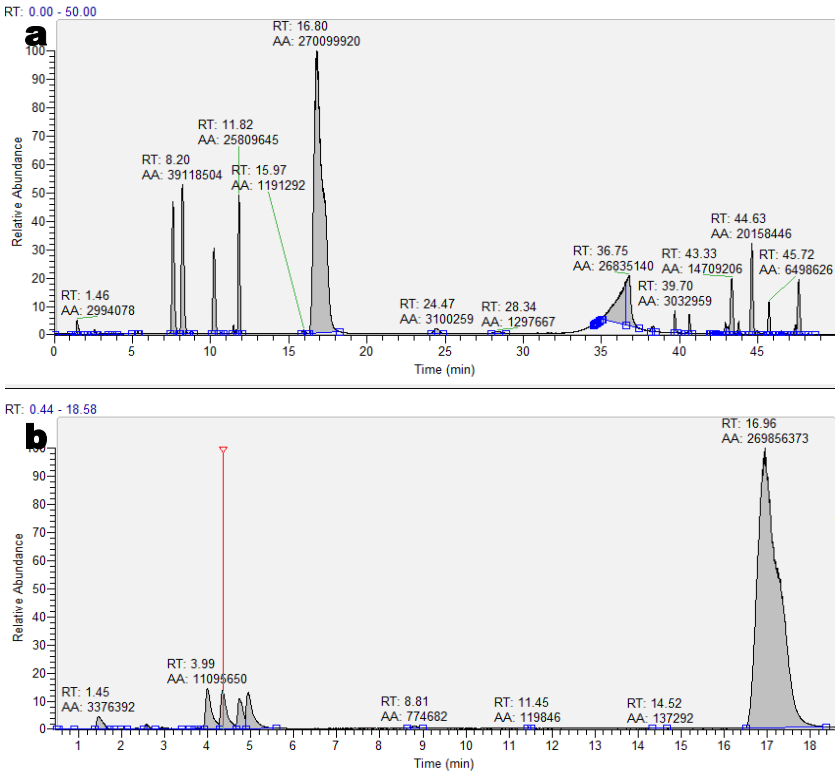
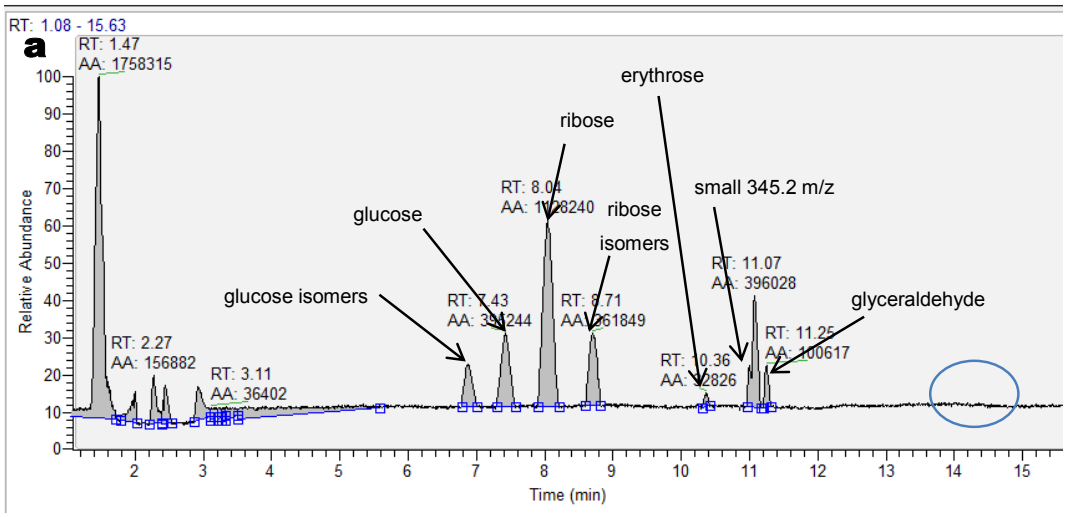


Figure 6.12. LC-MS chromatograms of (**a**) 4 commercial sugar standards (1 mM): glucose (7.6 min), ribose (8.2 min), erythrose (10.2 min), glyceraldehyde (11.82 min); and (**b**) 4 commercial sugar phosphates standards (500 µM): glucose 6-phosphate (3.99 min), ribose 5-phosphate (4.34 min), erythrose 4-phosphate (4.73 min), glyceraldehyde 3-phosphate (4.94 min). Peak at 16.8-16.96 corresponds to the derivatisation reagent (AEC).

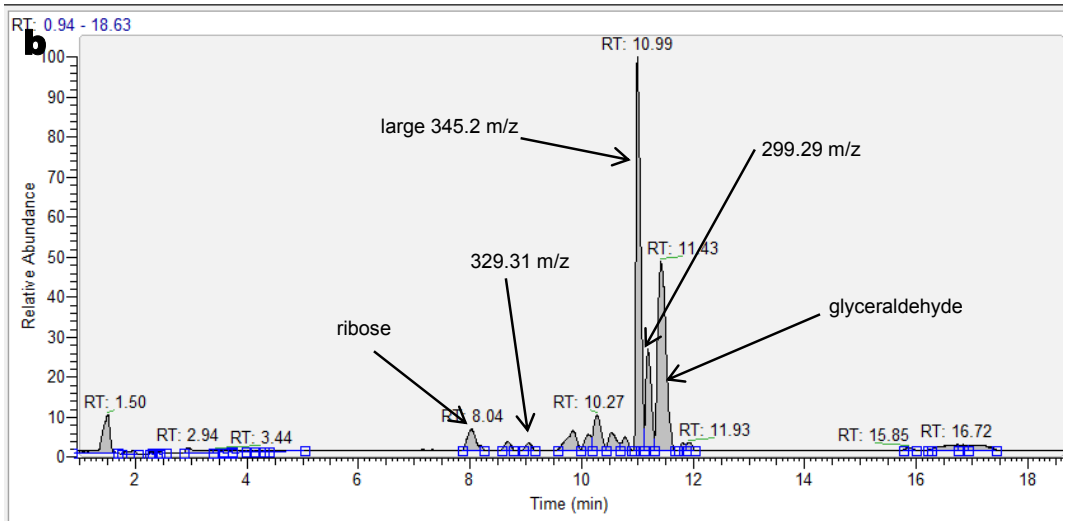
Note that in Figure 6.12 the retention time of the derivatisation reagent (AEC) was not the same in both chromatograms. These discrepancies were mostly due to different research groups using the instrument: over time, the usage of different solvents and not thoroughly purified samples generated a random drift in the retention times of analytes. This was the main reason for performing the quantification of the formose products using our HPLC-UV instrument (Section 6.2.2.2) as, even though the information it yielded was limited, the instrument was under our daily supervision and was kept in perfect state. In order to quantify this drift, a solution containing 250 μM of commercial ribose was derivatised and analysed 15 times; the resulting peaks had a mean retention time of 8.115 minutes ±0.101 SD. There were relatively large peaks appearing from minute 34 onwards (Figure 6.12a), corresponding to unidentified contaminants adsorbed into the chromatography column; despite its thorough cleaning, these always appeared in every chromatogram, probably contributing to the moderate retention time drift.

Examples of chromatograms for a control formose reaction, and formose reactions where AcP had been added at minute 30, and 60 are shown in Figure 6.12. Despite using different analytical instruments and a different chromatographic column (albeit both being C18-based), both control formose reactions from Figure 6.9b and 6.13a are remarkably similar. The identification of the peaks was performed by comparison with the retention times of commercial standards and with their MS-MS fragmentation patterns. When AcP was added after 30 minutes of reaction (Figure 6.13b), the peaks corresponding to the four non-phosphorylated sugars were smaller than in the control chromatogram, with the exception of glyceraldehyde (Figure 6.13a); this fits with the quantification data shown in Figure 6.10a-b. Interestingly, two new peaks appeared with the addition of AcP at 30 minutes: one with a mass of 299.29 m/z, and another with a size of 329.31 m/z (see below). In contrast, when AcP was added after 1 hour or reaction (Figure 6.13c), the resulting peak distribution was more reminiscent of the control reaction (Figure 6.13a) with a few additions: two peaks with a mass of 359.2 m/z, and again peaks with a mass of 329.31 and 299.29 m/z.

Control_formose 26/01/2018 22:06:06



Formose_plusAcP_at30min_90min 26/01/2018 23:47:32



Formose_plusAcP_at1h_90min 27/01/2018 01:28:54

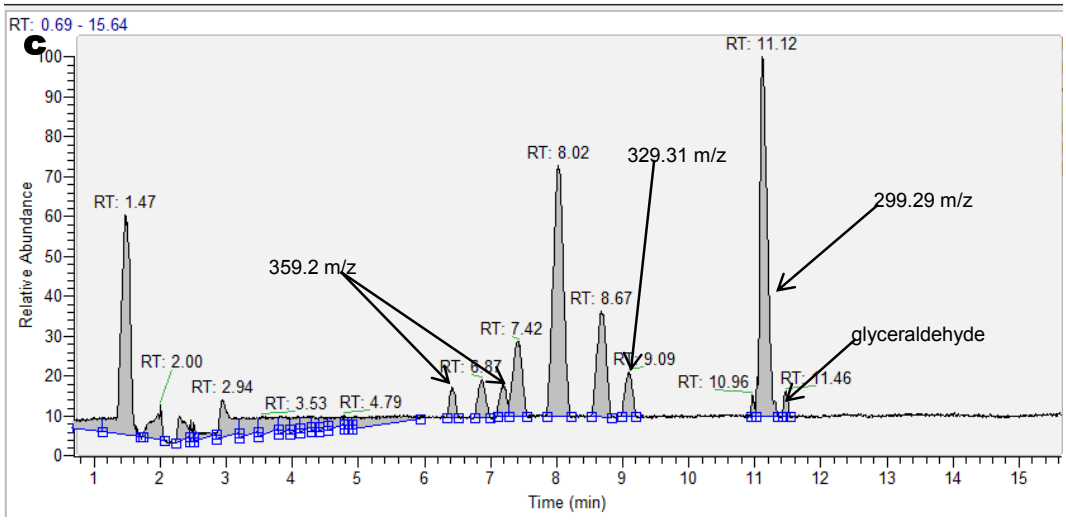


Figure 6.13. Full scanned LC-MS chromatograms of formose reaction (0.5 M formaldehyde and 0.167 M $Ca(OH)_2$ and $CaCO_3$, at 60 °C) samples after 90 minutes of reaction. (**a**) Control formose reaction without the addition of AcP. The blue oval highlits the lack of a dihydroxyacetone signal. (**b**) Formose reaction with the addition of AcP after 30 minutes of reaction. (**c**) Formose reaction with the addition of AcP after 60 minutes of reaction. The pH of the experimental reactions (b-c) was titrated according to the pH of the control one. Peaks labeled as 'isomers' present the same mass as the labeled isomer (e.g. ribose), but not its retention time. When no commercial standard was available for a given putative molecule (see text), the corresponding peak is labeled with just its mass.

None of the expected phosphorylated sugars (Figure 6.12b) appeared in any of the samples where AcP was added (Figure 6.13). The presence of acetylated sugars was also tested for even though commercial standards are not available for most (e.g. 436.14 m/z for mono-acetylated erythrose), but none of the predicted masses were observed either. Tetra-acetylated ribose was commercially available (513.2 m/z), but its derivatisation product could not be chromatographically observed using this analytical method (not shown).

The peaks corresponding to 299.29 and 329.31 m/z did not match up with the mass of any derivatised sugars (the putative masses of the un-derivatised molecules should be between 100 and 200 m/z). Interestingly, these peaks only appeared when AcP was added to the formose reaction, which suggested that the un-derivatised parent molecules were a result of the reaction between AcP and a formose reaction product. The only two possible masses for the un-derivatised molecules are of 104.05 m/z and 134.06 m/z each; these masses can only conceivably correspond to mono-acetylated ethylene glycol and glycerol respectively (Figure 6.14). The masses of other conceivable un-derivatised parent molecules, such as phosphorylated or (mono/poly)acetylated small sugars (e.g. glycolaldehyde or glyceraldehyde) did not match the experimentally determined masses. Ethylene glycol and glycerol are the products of the reduction of glycolaldehyde and glyceraldehyde respectively; this reduction is known to occur in the Cannizzaro reaction, which is a base-induced disproportionation of aldehydes (Figure 6.14). The Cannizzaro reaction is well known to take place during the formose reaction, and is mainly responsible for the pH drop that the reaction undergoes naturally, as it

generates formic acid from formaldehyde with up to 40% efficiency (Kopetzki & Antonietti, 2011).

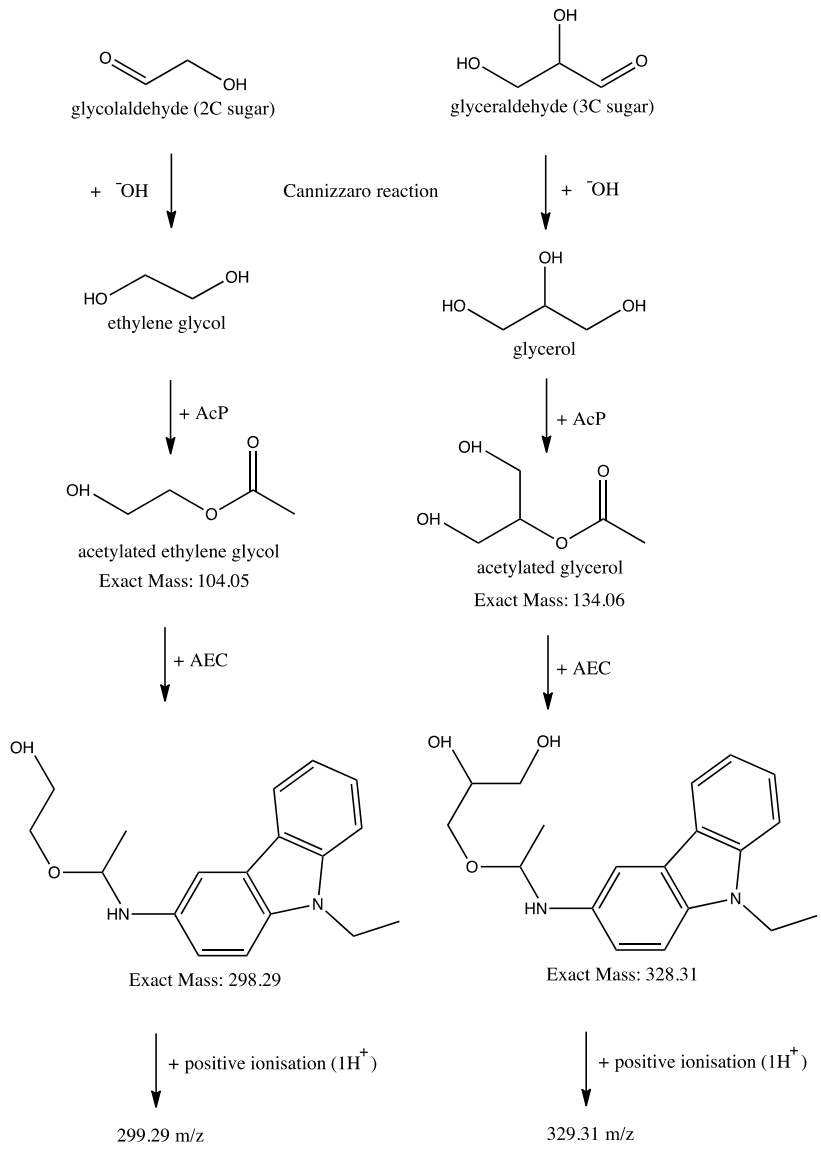


Figure 6.14. Proposed reaction pathways from the formose reaction products glycolaldehyde and glyceraldehyde leading to the experimentally measured masses of 299.29 and 329.31 m/z respectively. Other chemically possible precursor molecules, such as acetylated glycolaldehyde (102.03 Da) would give rise to masses (297.15 m/z) not corresponding to those observed experimentally.

Probably, the most striking feature of Figure 6.13 is the large peak with a mass of 345.2 m/z (Figure 6.13b), which is exactly the mass of ribose (and its

isomers). The corresponding precursor molecule could not be ribose since, as shown earlier, it presented a very different retention time (more than 2 minutes apart). In order to assess what molecule the peak at 10.99 min was, 500 μM solutions of the commercial diastereomers of ribose (xylose, arabinose and lyxose), as well as the ketopentose ribulose, were prepared and analysed (Figure 6.15).

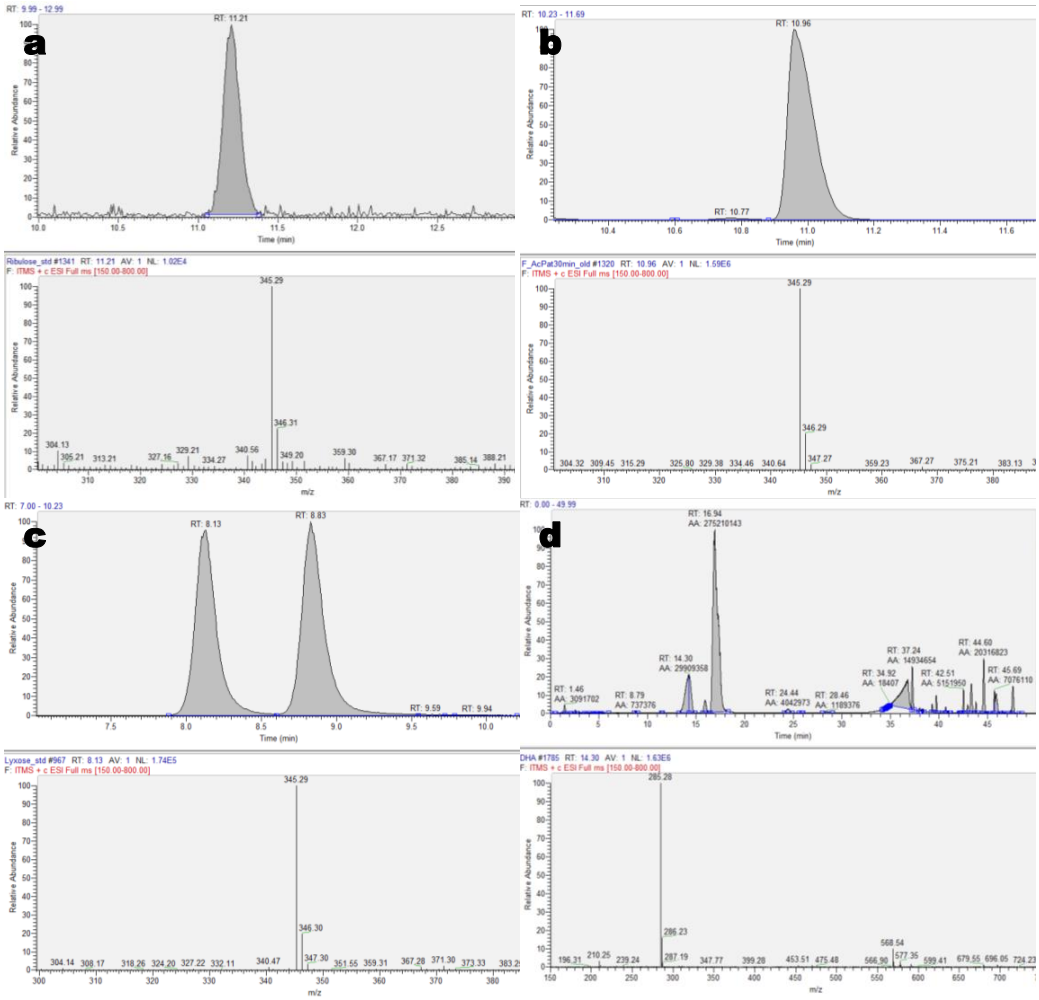


Figure 6.15. LC chromatograms and MS spectra of commercial (**a**) ribulose (345.29 m/z), (**b**) experimental signal attributed to ribulose (345.29 m/z), (**c**) a mixture of commercial xylose, arabinose and lyxose (both peaks with a mass of 345.29 m/z), and (**d**) commercial dihydroxyacetone (285.28 m/z).

The data shown in Figure 6.15a-b strongly suggests that the 345.2 m/z signal detected when AcP was added to the formose reaction was added after 30 minutes (Figure 6.13b) corresponded to ribulose (a ketopentose). It is

impossible that this signal at about 11 minutes corresponds to any of the isomeric aldopentoses of ribose (xylose, arabinose and lyxose), since the retention time of their combined two peaks was of 8.13-8.83 minutes (Figure 6.15c). The signal from commercial dihydroxyacetone (DHA) showed a retention time of 14.3 minutes, which was not detected in any of the chromatograms shown in Figure 6.13 (blue circle). It is noticeable that the retention times of both commercial (Figure 6.15a) and experimental (Figure 6.15b) ribulose (11.21 vs 10.96 min), and of commercial (Figure 6.15c) and experimental (Figure 6.13a,c) aldopentose isomers of ribose (8.13-8.83 vs 8.04-8.71 min) were not completely coincident. As explained earlier, these small discrepancies are caused by random retention time drifts of the instrument, and are within the expected range. The identity of these peaks was further reinforced by the perfect coincidence in the masses of their molecular ions (Figure 6.15a-c, lower cells).

6.3 Discussion

Looking at all the data presented in the previous Section, it is clear that the formose reaction constitutes a plausible prebiotic source of sugars under simulated alkaline hydrothermal vent conditions. None of the formose reactions performed developed a dark colour or a tarry texture, which is associated with the synthesis of large (and often ramified) sugars. Kim et al. (2011) showed that the development of a tar occurs when all initial formaldehyde has been used; taking this into account, even though the amount of formaldehyde in each sample was not quantified, it is safe to assume that the formose reactions performed never exhausted the formaldehyde 'fuel'. From this perspective, it is tempting to think that this represents a laboratory artefact, and that in a natural environment the initial concentration of formaldehyde would have been lower, which would get depleted rather quickly, resulting in the reaction yielding mostly the classical biologically useless tars. This may not have been the case, however, if the thin semiconducting Fe(Ni)S mineral walls of Hadean alkaline hydrothermal vents catalysed the reduction of CO_2 into formaldehyde (Lane, 2014; Martin & Russell, 2007; Martin et al., 2008; Russell et al., 2014; Sojo et

al., 2016). This would continuously do so (as long as the vent is active), constantly supplying the nascent proto-metabolic network with fresh reactants.

As expected, the addition of AcP to the formose reaction altered its products qualitatively and quantitatively. The early experimentation using GC-MS analysis showed that indeed the addition of AcP to the formose reaction (whilst controlling for the changes in volume and pH that it entailed) altered the production of sugars. When AcP was added initially (0 minutes), the reaction did not produce any sugars at all. With the addition of AcP after 30 minutes of reaction, the yield of all sugars increased, with ribose experiencing the largest amplification (Figure 6.7). As explained earlier, the experiments using GC-MS had to be abandoned due to problems with the chromatographic column, so the next set of experiments were analysed using HPLC-UV.

Using the HPLC-UV setup, it was observed that the control formose reaction runs (Figure 6.10a-d, dashed lines), showed distinct sugar patterns to those from previous experiments using the GC-MS setup (Figures 6.7): ribose was now the sugar with the higher yield (without the addition of AcP), but only after 1 hour of reaction; at 30 minutes glyceraldehyde was the most prominent sugar but with modest yields. Erythrose followed the same pattern as in the experiments analysed with GC-MS: modest yields and peaking at 1 hour of reaction. Glucose, a C6 sugar, was quantified for the first time, and its pattern was similar to that of ribose, but with decreased yields throughout. These differences probably reflect the analytical method used: when analysed using the GC-MS setup, the samples were derivatised by acetylation, which was preceded by a reduction of the aldehyde group using $NaBH_4$. Commercial ribose was used as such (aldehyde form) when generating calibration curves, but erythrose and glyceraldehyde were purchased as erythritol and glycerol respectively, as done in relevant literature (Kopetzki & Antonietti, 2011). It is possible that the reduction step of the derivatisation protocol did not present 100% efficiency and thus, the sugars which were purchased as such (and not as already-reduced polyalcohols), such as ribose, were infraquantified systematically, yielding the reduced ribose yields seen in Figure 6.7 (red dashed line). Finally, the data shown in Figure 6.7 represents only a total of 2

repetitions, whereas the data shown in Figure 6.10 represents 9 independent experimental runs (where AcP was not added), conferring on the latter significantly higher credibility. The sugar yields in the control formose reactions shown in Figure 6.10c-d were about 20% higher than those in Figure 6.10a-b; this probably just reflected a random fluctuation in the reaction parameters since both sets were performed identically. These differences are not very important since both reaction profiles are almost identical (apart from the absolute yields), and each experimental run should be compared only to the control reactions performed alongside on the same day (which is the way it is displayed for each graph in Figure 6.10). Continuing with the HPLC-UV-analysed samples, the addition of AcP at 30 minutes seemed to slow down the formose reaction overall, delaying the peak of maximum concentration for each sugar. Not only that, but ribose seemed to (again) be most affected by the addition of AcP compared to the other sugars (Figure 6.10a,b and d). It is clear that this was not just a pH-related effect, since all appropriate controls were performed (Figures 6.10c and 6.11a), strongly indicating that the observed AcP-mediated effect was genuinely due to its chemical activity. Finally, a striking difference when comparing the data acquired using the HPLC-UV setup with the one analysed by GC-MS, is that no experimental treatment consisting only of AcP addition increased the maximum yields of ribose compared with the control reaction (Figure 6.10a and d), but instead seemed to boost its stability (or hinder its reactivity) over time; Figure 6.10b constitutes an exception since the addition of AcP was accompanied with lower reaction pH from 30 minutes onwards. All the aforementioned discrepancies were further clarified by studying the samples on an LC-MS setup.

Using the analytical capabilities of the LTQ LC-MS setup, the results suggested that the main effects of AcP addition to the formose reaction occur via the (i) acetylation of reduced glycolaldehyde (ethylene glycol) and glyceraldehyde (glycerol), and (ii) via the promotion of the synthesis of ribulose. The chemically possible synthesis of phosphorylated and acetylated sugars was examined, but no signal was detected. Evidence of the synthesis of DHA (an important formose reaction intermediate) was also sought, but was unsuccessful.

The combined results discussed in the previous Section suggest that AcP promoted the channelling of the formose reaction towards certain sets of products via acetylation and in a time-dependant fashion. The conveniently named formose reaction (Butlerow, 1861) is in reality a complexly reaction cascade with multiple autocatalytic cycles and competing branched side reactions (Breslow, 1959; Holm et al., 2006; Orgel, 2004). The first reaction of the formose reaction (the condensation of formaldehyde with another molecule of formaldehyde) is controversial, with some authors claiming that, because formaldehyde is such a poor nucleophile, the reaction rate is so slow that it basically does not happen, and that successful formose reactions in reality start with the condensation of formaldehyde with trace (down to 3 ppm) amounts of glycolaldehyde (2C) or glyceraldehyde (3C) (Kieboom & VanBekkum, 1984; Socha, 1980). This controversy has not been studied in this thesis, so the first reaction will be treated as a dimerization of two molecules of formaldehyde. Regardless, the controversy exemplifies how the early stages of the formose reaction (which are arguably the most important) are not fully understood.

The results shown throughout Section 6.2 of this Chapter can be interpreted following the diagram shown in Figure 6.16. As just mentioned, the first reaction yielding glycolaldehyde is a slow one and helps explaining the slow pace of the reaction during its first 30 minutes. Once glycolaldehyde has been synthesised, it can follow three distinct pathways; obviously it could undergo other reactions, but for clarity, and because these putative reactions do not help explaining our results, these have not been taken into account. First, glycolaldehyde can undergo the well-known Cannizzaro reaction, yielding ethylene glycol and glycolic acid (which lowers the reaction pH). It is clear that AcP acetylated ethylene glycol (299.29 m/z peaks at 11.12 min in Figure 6.13b,c), which constitutes a poorly reactive molecule that tends to accumulate. With the addition of AcP, ethylene glycol becomes readily acetylated (in a similar fashion to glycine, as discussed in Chapter 5) and accumulates; this removes ethylene glycol from the equilibrium, which in turn would funnel more glycolaldehyde into the Cannizzaro reaction replenishing the lost ethylene glycol (Figure 6.16, green reactions). This interpretation explains why with the addition of AcP at 0 minutes of the formose reaction, no sugars were formed at all; this probably

reflects that formaldehyde and/or glycolaldehyde were mostly funnelled towards their Cannizzaro products, bringing the reaction to a halt.

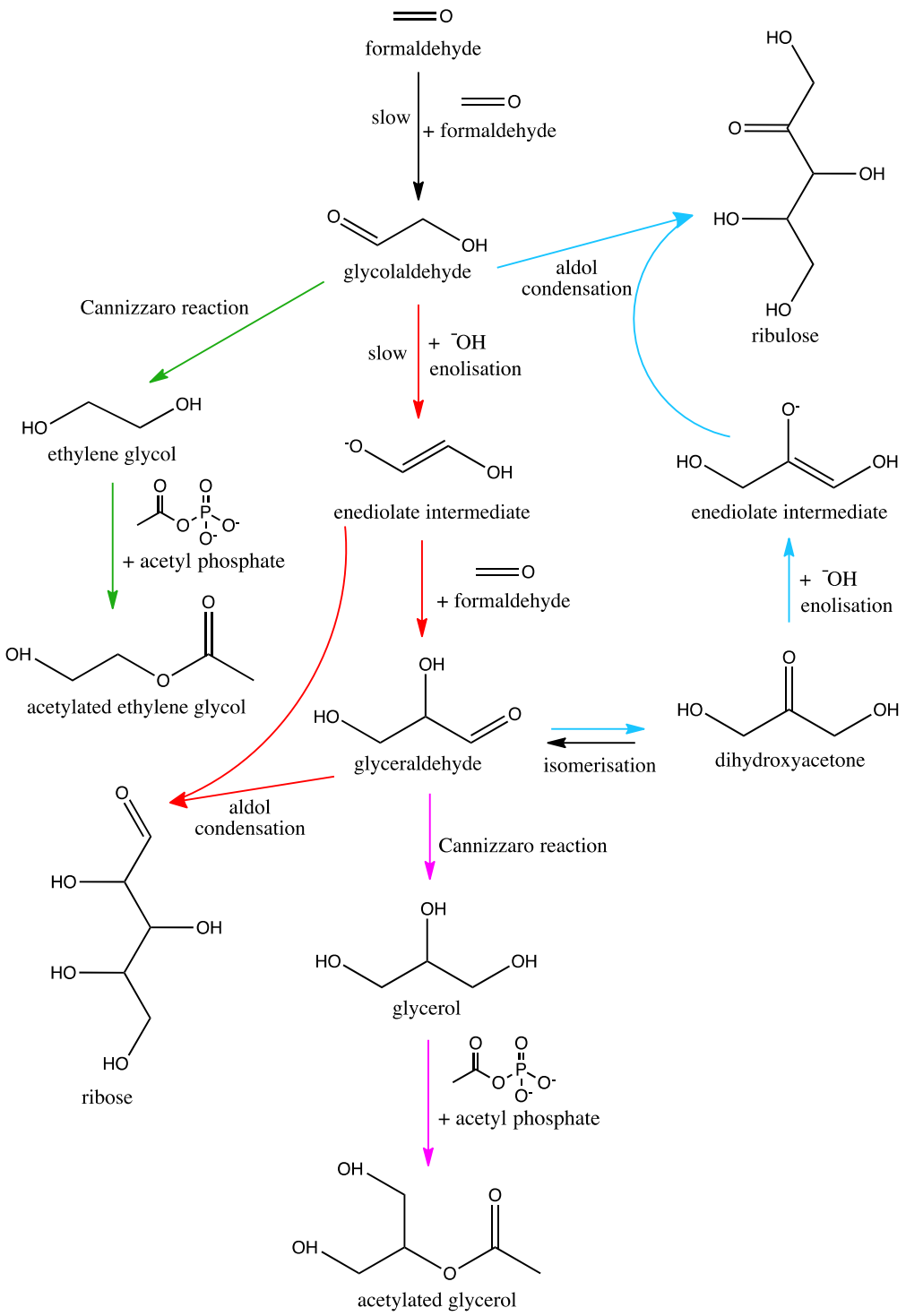


Figure 6.16. Initial steps of the formose reaction cascade, highlighting the unproductive Cannizzaro reactions followed by AcP-mediated acetylations (green and violet), and the aldol condensations yielding ribose (red) and ribulose (blue).

The finding that AcP-mediated acetylation indirectly favoured the synthesis of glycerol (it was not detected, as the analytical method used was not suitable) in the formose reaction is as interesting one, which has implications for the research on the prebiotic synthesis of ancient biological membranes, as these often involve the synthesis of glycerolipids (esters of glycerol and fatty acids).

In control formose reactions (without the addition of AcP), I observed ribose being synthesised plentifully (Figures 6.10 and 6.11, dashed lines). It is worth noting here that I could not differentiate whether only ribose was being synthesised (which is unlikely) when using the HPLC-UV setup, since ribose shared its retention time with its three aldo-pentose isomers (xylose, lyxose and arabinose). Therefore, I should technically be using the term aldopentose when referring to that data (which includes all four isomers), but for clarity I will use the term 'ribose' in this Chapter. As shown in Table 6.1, on the GC-MS setup the retention times of all 4 aldopentoses were different (with the exception of arabinose and lyxose), and in all the quantification attempts, the integrated area of ribose was roughly ¼ of the sum of all three peaks. This is an expected outcome since our experiments did not contained any known chiralgenic factor; it is then reasonable to assume that the expected ¼ ratio ribose/aldopentose found via GC-MS analysis applies to any formose reaction analysed. Returning to the control formose reactions: without the addition of AcP, glycolaldehyde would not be solely funnelled towards its Cannizzaro products, so the formose reaction advances normally, generating aldopentoses such as ribose, as only 3 reactions are needed (Figure 6.16, red reactions); note that only small amounts of ribulose were synthesised under control conditions (small 345.2 m/z peat at 10.99 min in Figure 6.13a), which involves a total of 5 reactions (Figure 6.16, blue reactions).

With the addition of AcP at 30 minutes of reaction, a different reactivity landscape takes place. In these experiments the formose reaction was left unaltered during the first 30 minutes, which is enough time for glyceraldehyde to be synthesised (Figure 6.10a,b). Glyceraldehyde isomerises to its ketose form (dihydroxyacetone (DHA)) easily under alkaline conditions via a Ca^{2+}-mediated 1,2-hydride shift mechanism (Appayee & Breslow, 2014). DHA in turn

is a good nucleophile once it enolises (which is favoured at high pH) and thus it readily reacts; this would help explain why dihydroxyacetone was not detected (Figure 6.13, blue oval). DHA has been notoriously difficult to detect in similar studies (Kim et al., 2011), even though it is most definitely a required intermediate of the formose reaction (Appayee & Breslow, 2014; Deng et al., 2013; Meinert et al., 2016). The enolised dihydroxyacetone can then nucleophilically attack glycolaldehyde, yielding ribulose. Conversely, the enolisation of glycolaldehyde (yielding an enediolate intermediate) and its subsequent nucleophilic attack onto formaldehyde is a very slow process (Mellersh & Smith, 2010). To recap, the addition of AcP after 30 minutes of reaction encounters a complex system where glyceraldehyde and dihydroxyacetone have already been synthesised in sizeable portions; then glycolaldehyde starts being funnelled towards its Cannizzaro products due to AcP-mediated acetylation. This AcP-mediated effect could conceivably force the remaining glycolaldehyde to be directed mostly towards the synthesis of ribulose instead of ribose (which would require glycolaldehyde to act as a nucleophile via its slow enolisation, an unflavoured reaction when under competition). This interpretation is further supported by the large (unreacted) glyceraldehyde peak observed when AcP was added after 30 minutes of reaction (Figure 6.13b), since it is required for the (now slower) synthesis of ribose.

This (putatively indirect) AcP-mediated promotion of the synthesis of ribulose found experimentally would help explain two of the other phenomena, which were left unexplained earlier. First, it would help explain why when using the GC-MS analytical setup, the addition of AcP caused a substantial increase in the yield of ribose compared to the control reactions (Figure 6.7), but when using the HPLC-UV setup and controlling the reaction pH throughout (Figure 6.10a) the maximum yield or ribose did not increase. The sample derivatisation protocol for subsequent GC-MS analysis involved a $NaBH_4$-mediated reduction of any aldehydes and ketones present in the sample; thus, any ribose (aldose) and ribulose (ketose) synthesised would yield the same acetylated polyalcohol as a final derivative (pentol). In contrast, the derivatisation protocol for HPLC-UV (and LC-MS) analysis derivatised ribose (and its aldose isomers) and

ribulose as separate molecules, and their retention times on the C18 column were different (8 and 11 minutes respectively). This means that the GC-MS-acquired data shown in Figure 6.7 was representative of the added concentrations of ribose and ribulose, so the synthesis of ribulose promoted by AcP was seen as an increased ribose concentration; whilst HPLC-UV-acquired data represented ribose (or as explained earlier, it represents aldopentoses, with ¼ being ribose), and is independent of ribulose quantification. Second, the synthesis of large quantities of ribulose (and less ribose) when AcP was added after 30 minutes could help explain why the ribose peak in Figure 6.10 was delayed compared to the control reaction: the synthesis of ribose was indeed bypassed by the preeminent synthesis of ribulose delaying the peak but, over time, ribulose can isomerise into ribose easily at high pH (Appayee & Breslow, 2014; Hamada et al., 2003; Yoshida et al., 2012), either through an enediol intermediate or hydride-shift mechanisms. Therefore, ribulose would effectively act as a mid-term reservoir of ribose, resulting in the measured stability of the ribose concentrations after its peak (Figure 6.10a), since the large ribulose reservoir would slowly but steadily replenish it.

With the addition of AcP at 60 minutes of reaction, ribose also seemed to have its stability boosted (Figure 6.10d), this effect was not observed for the other quantified sugars. The experimental LC-MS chromatogram (Figure 6.13c) is remarkably similar to the control formose reaction one, but with acetylated Cannizzaro reaction products (329.31 and 299.29 m/z), and two new peaks appearing around the glucose ones with a mass of 359.2 m/z each. Coincidentally, a precursor molecule that matches this mass is the C6 deoxysugar fucose and its isomers (precursor mass of 164.16 Da), which I had used in other experiments as external standard (but not in these since quantification was not attempted). With the information at hand, it is hard to envision how AcP could promote the synthesis of deoxysugars, such as fucose, but it is probable that with the addition of AcP at 1 hour, the formose reaction bottleneck around glycolaldehyde (Figure 6.16) had been already bypassed; this way the effects of AcP would be mostly on the chemistry of larger sugars (which were not quantified in this study) and effectively either promoted the

break down of these heavier sugars into ribose (by retro-aldol reactions), or hindered reactions that use ribose as a substrate to build larger sugars.

During this Section, all the mechanisms proposed so far involve AcP-mediated acetylations since these have been detected experimentally (Figure 6.13). Despite this, I cannot dismiss the possibility that some of the AcP-mediated effects on the formose reaction occurred through phosphorylation, as proposed by Mellersh & Smith (2010); even though their proposed phosphorylations would have affected the reaction mostly by retention in the positively charged mineral surface of a vent. An intriguing possibility is that these phosphorylated intermediates would become so reactive that were readily depleted (as noted previously must have been the case for dihydroxyacetone); this way they would not be detected. It is worth noting though that this does not square well with the apparent long-term stability of ribose phosphate (Figure 4.9), a feature which presumably applies to the other sugar phosphates. It is possible that the reaction between a phosphorylated tetrose (e.g. erythrose 4-phosphate) with the enediolate of glycolaldehyde could yield a fucose isomer, which would yield a mass of 359.2 m/z (Figure 6.17); again, this is highly speculative since the putative phosphorylated intermediate was not detected. How this putative reaction would be related to the measured increase in ribose stability (Figure 6.10d) remains unsolved.

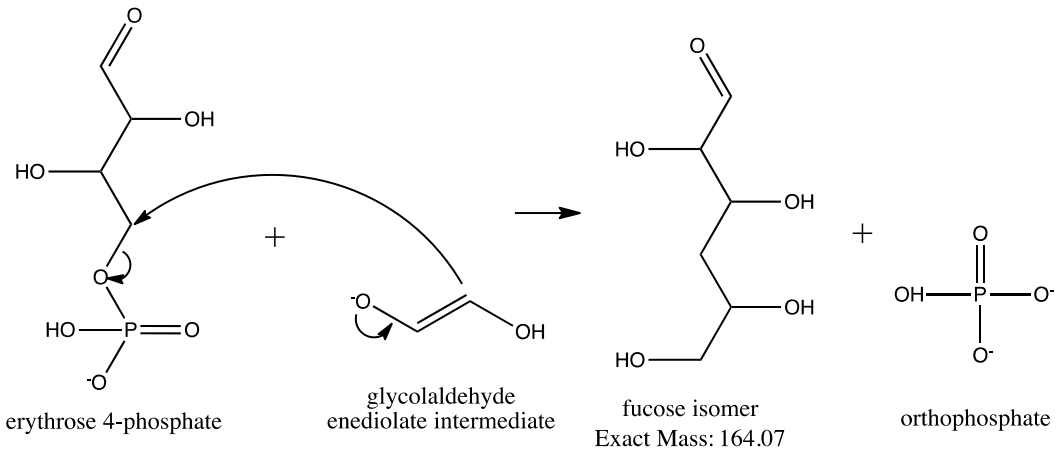


Figure 6.17. Reaction between the enediolate of glycolaldehyde and putative (AcP-synthesised) erythrose 4-phosphate, yielding a fucose isomer and orthophosphate.

In conclusion, acetyl phosphate did indeed have an effect on the product palette resulting from the formose reaction, albeit possibly not through the phosphorylating mechanism proposed by Mellersh & Smith (2010). Admittedly, I did not test the effects of adding a positively charged metal surface to the reaction, which according to the authors, together with a moderate hydrothermal flow would wash out undesired products, those which are non-phosphorylated, and retain phosphorylated ones, this way promoting a biologically-reminiscent phosphorylated sugar cascade. A key question though, would be whether my findings are relevant for the research field, since the addition of AcP had interesting effects, such as promoting pentose abundance throughout the formose reaction, but these only presented biologically-interesting outcomes in a time-dependant manner (e.g. no sugars were formed when AcP was added at the very beginning). It is indeed hard to envision a way by which AcP is naturally added to the formose reaction some time after it has started and not before. As mentioned earlier, the natural layout of alkaline hydrothermal vents is intrinsically heterogeneous, with areas whose physico-chemical parameters are substantially different to some other. Taking this into account, it is possible to envision the synthesis of AcP taking place mostly in cooler and more neutral areas of the vent (its outer layers), whilst the formose reaction would predominantly occur within its inner layers (hotter and strongly alkaline); this way, AcP could be 'delivered' via thermal convection to the formose reaction in volumes and in a temporality that favours a biologically-synergetic outcome of the formose reaction. The rate at which the freshly synthesised AcP would be delivered to the formose reaction would basically depend on the physico-chemical structure (which would have been randomly assembled) of that part of the vent. The biologically interesting synergism found experimentally in this study would then only occur in very specific areas of a vent, and would probably be completely impossible in other alkaline vents. The implied difficulty of having the right set of reactions happening in the right order, magnitude and time lapses may be hinting to why life seems to have only started once on Earth (that we know of); perhaps it was indeed the result of a rare combination of chemical reactions that needed a rather specific set of randomly-assembled conditions, which were rarely met. Of course, such a conclusion does not stem directly from these results, but would fit with my observations. Further experimentation on a

continuous-flow reactor simulating more realistically the interactions between AcP and the formose reaction is detailed in the Conclusions and future work Section of Chapter 7.

6.4 Conclusions

The formose reaction constitutes a plausible source of sugars under alkaline hydrothermal vents conditions. The synthesised sugars remained in their monomeric form for up to five hours and, under the experimental conditions tested, did not lead to biologically inert tars. The addition of AcP to the formose reaction resulted in an altered product range. Its addition at the very beginning of the reaction arrested it so that it never yielded sugars, whilst its addition later on (30 or 60 minutes) directed it towards the synthesis of pentoses. These effects seem to be mostly mediated by the acetylating properties of AcP, which are well studied in the literature (Di Sabato & Jencks, 1961a), and were experimentally corroborated during this project on other substrates such as the amino acid glycine. Acetylation is hardly a biologically conserved mechanism, unlike phosphorylation, so these results are somewhat disappointing. It is worth nothing that the fact that AcP remains bound to the active site of CODH/ACS in archaea (Bräsen et al., 2008; Schonheit et al., 2016) could be suggestive of acetylation causing problems in modern cells; this way, it is conceivable that acetylation played a larger role at the origin of life, but was reigned in after the switch to ATP. These acetylation-mediated effects could be inferred due to the detection of the acetylated early Cannizzaro products (ethylene glycol and glycerol) of the formose reaction; surprisingly, no acetylated sugars could be detected. AcP boosted the synthesis of ribulose when it was added at 30 minutes, which seemed to act indirectly as a ribose reservoir since its isomerisation into ribose is base-catalysed. With the addition of AcP at 60 minutes of reaction, the stability (or inertness) of ribose was promoted by a yet undefined mechanism; fucose (and its isomers), which is a C6 deoxysugar, was detected for the first time as a reaction product.

The initial aim of this study was to assess the validity of theoretical predictions on the putative effect of phosphorylation on the formose reaction (Mellersh & Smith, 2010). Albeit the possibility that the synthesis of deoxysugars (such as fucose) reflected AcP-mediated phosphorylations, this could not be corroborated since sugar phosphates could not be detected during this study. Admittedly, I used AcP because, as explained previously, is a very simple molecule that can be synthesised under hydrothermal vent conditions and it is still used by life today. The choice of AcP, which as extensively discussed does not only possess phosphorylating capabilities, means that I technically did not falsify Mellersh & Smith predictions, since they specifically proposed that it is the combined effects of phosphorylation plus the presence mineral surfaces that could direct the formose reaction into biologically interesting sugars. It would be indeed interesting to explore the effect of other purely phosphorylating prebiotic reagents, such as pyrophosphate or trimetaphosphate, on the formose reaction, with and without the presence of mineral surfaces.

Chapter 7

Conclusions and future work

7.1 Conclusions and future work

The aims of this project were to:

1. Assess whether Fe(Ni)S minerals could have catalysed chemical reactions analogous to those of the acetyl CoA pathway and the incomplete reverse Krebs cycle, by acting as semiconducting barriers promoting redox reactions between reactants, whose reduction potentials could have been attuned by the steep natural pH gradients present at Hadean alkaline hydrothermal vents. The synthesised intermediate molecules would include thioacetate, which is conducive to the synthesis of AcP.

2. Study whether AcP could have acted as a primordial non-enzymatic ATP analogue by promoting substrate-level phosphorylations on simple organics, such as ribose, adenosine and ADP, under simulated alkaline hydrothermal vent conditions.

3. Evaluate whether similar non-enzymatic phosphorylations on small organics (AMP and glycine) would have been conducive to their non-enzymatic condensation yielding oligomers under simulated alkaline hydrothermal vent conditions.

4. Assess whether the formose reaction could have acted as a credible source of biologically important sugars, such as ribose, at Hadean alkaline hydrothermal vents; and appraise whether the addition of a phosphorylating agent (such as AcP) could have channelled it towards the synthesis of pentoses, as suggested in previous literature.

In this project I found that Hadean alkaline hydrothermal vents could indeed have promoted non-enzymatic reactions analogous to those (later catalysed by enzymes) in the acetyl CoA pathway and incomplete reverse Krebs cycle. I suggest that the initial reduction of CO_2 to a methyl group proceeded via a Fischer Tropsch-type mechanism on the acid side of an Fe(Ni)S surface, followed by carbonyl migration and insertion (carbonylation) via a Koch-type mechanism. These endergonic reactions ought to be facilitated by pH differences between phases, which modulate the reduction potential of H_2, CO_2 and Fe(Ni)S minerals such as mackinawite (Herschy et al., 2014; Sojo et al., 2016; Wolthers et al., 2005). Similar hydrogenations and carbonylations, facilitated by natural proton gradients across Fe(Ni)S barriers, could in principle generate longer chain carboxylic acids (up to 6 carbons) via a prebiotic equivalent to the reverse incomplete Krebs cycle in methanogens.

Whilst promising, these are of course theoretical predictions and experimental work to ascertain these claims is needed in the future. Current work on developing an open-flow Y-shaped microfluidics chip is underway in the Lane Lab. In the chip, each of the two branches of the Y-shaped channel supplies one of the two fluid phases (either alkaline outflow or acidic ocean), which meet in the middle channel and generate a thin Fe(Ni)S precipitate in-between both laminar flows. This device allows for a more precise control of the reaction parameters (pH, temperature, Fe(Ni)S mineral thickness, reduction potentials, etc.), which in turn helps defining the required parameters for these tardy reactions to take place. Plans for a high-pressure (up to 500 bar) microfluidics chip are in place, since this should maximise the percentage of H_2 gas that gets dissolved in the alkaline fluid. It is suspected that some of the CO_2 reduction experiments in the past were unsuccessful because only a marginal amount of H_2 (the ultimate reducing agent in these experiments) would be dissolved at atmospheric pressure, hindering the yields of formaldehyde produced. The solubility of H_2 in water is proportional to the pressure in the system (Wiebe et al., 1932) (Figure 7.1); therefore the usage of a high-pressure reactor should clear this limitation. This setup should also help with the experimentation on the reduction and carbonylation reactions of larger organic molecules, such as those forming part of the incomplete reverse Krebs cycle.

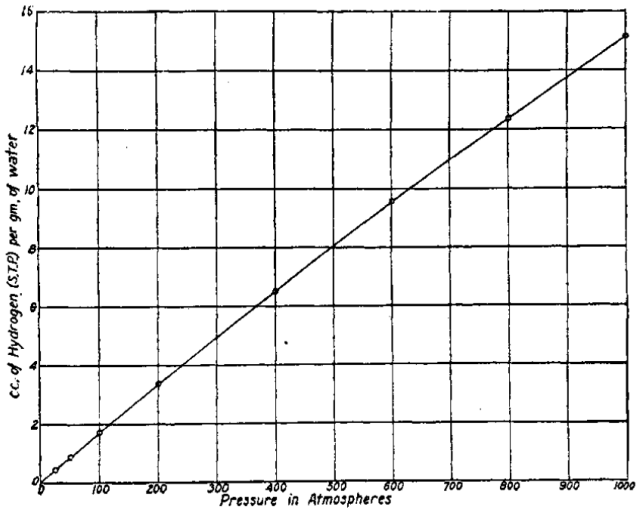


Figure 7.1. Solubility of H_2 gas in water at 25 °C at various pressures (Wiebe et al., 1932).

I have also shown that acetyl phosphate is a good non-enzymatic phosphorylating agent, capable of phosphorylating the hydroxyl groups of several sugars (ribose and glucose) and sugar derivatives (adenosine); therefore confirming that AcP could have acted as a prebiotic precursor to ATP in this regard. Although the yields found were modest I explain why, from an evolutionary point of view, low yields should be the expected outcome for non-enzymatic prebiotic reactions. Unfortunately, the exact phosphorylation positions of ribose could not be determined due to (i) the low experimental yields, which did not allow for a good 2D ^{31}P-NMR signal, and (ii) the achieved LC chromatographic separation not being optimal. A better chromatographic separation of all sugars (and derivatisation material) in the future would theoretically generate non-overlapping peaks; this feature could be exploited in future analyses using MS-MS in order to further elucidate the exact ribose phosphorylation positions, since each constitutional isomer would generate secondary fragments with slightly different abundances.

We have also shown that AcP can phosphorylate already phosphorylated nucleotides, such as ADP. We confirmed that ATP (and not other ATP isomers, such as for example a putative 3-phospho-ADP) had been synthesised using ^{31}P-NMR and HPLC-UV, since our results can only be compatible with three phosphate groups disposed in tandem. This is a relevant finding, since this way

AcP could then act as an intermediate between an acetyl CoA precursor (thioacetate) and ATP, which (not)coincidentally is its role in modern cells. Unfortunately, at the time of writing of this thesis, the quantification of the synthesis of ATP (from ADP) could not be performed due to the Dionex HPLC-UV used being often out of order, mostly due to column blockages caused by the usage of phosphate-based buffers. A new method using Lane Lab's Agilent 1260 Infinity II LC system with a UV detector is being developed; it will entail the usage of a C18 column this time (Agilent InfinityLab Poroshell 120 EC-C18, 4.6 x 150 mm, 4 µm particle size), and mobile phases consisting of 0.1 M ammonium acetate in water (A) (pH 6, adjusted with the addition of 10 mM of acetic acid), and methanol (B). I have also shown that AcP acetylated hydroxyl groups easily at high pH. Quantification was not attempted for the acetylated products, but it is safe to say that the competition between acetylation and phosphorylation reactions pH probably accounted for the lower yields of ribose phosphate observed.

AcP did not promote the polymerisation of either glycine or AMP under the experimental conditions tested. In the case of glycine this had to do with the strong N-acetylating capabilities of AcP. Acetylation is generally deemed as prebiotically non-productive, since it is a rather stable chemical group, making the acetylated molecule less prone to react further. Despite this, it could be theoretically possible to promote the condensation of amino acids by first acetylating the amino group of glycine in order to promote further reactions of other non-acetylated groups (e.g. carboxyl) of the same molecule; this way, the synthesis of a phosphor-carboxylate intermediate would be promoted, which could yield di-glycine with the addition of free glycine molecules. I have observed that differences in pH affect the phosphorylating/acetylating capabilities of AcP. In order to better understand the conditions that promote each capacity, more experimentation should take place assessing the effects of: pH and temperature, the presence of ions (e.g. Mg^{2+} ions could promote phosphorylation reactions since these increase the electrophilicity of phosphate groups), and the addition of mineral surfaces which, due to their interaction with charged groups, could promote or hinder certain reactions.

In the case of AMP I could not detect any acetylated products, so it is possible that other reaction parameters (e.g. lack of surface-based catalysis) account for the negative results. Future work should focus mainly on the potential effect of prebiotically plausible surface catalysts on this reaction. As theorised earlier, extant nucleotide polymerising enzymes have a conserved two-metal (Mg^{2+}) ion reaction mechanism, which modifies the pK_a of the chemical groups of the substrates allowing for a great yield and selectivity of the reaction. I realistically do not aim for this as I predict that prebiotic reactions had to generate low yields, so that successive evolutionary steps could be selected because they offered a higher selectivity for substrate/product and overall yield. What I aim to achieve is to get the reaction going with prebiotic catalysts that plausibly existed in alkaline hydrothermal vents. For example, 0.01-2 g/mL of ground (to increase the catalytic surface area) brucite ($Mg(OH)_2$ (Figure 7.2)) could be added to solutions of 3-10 mM AMP, ADP and ATP, hoping to mimic, in a very crude way, the conditions of the active site of the RNA polymerase. For the same purpose, small 2-10 aspartate oligomers (since these coordinate Mg^{2+} and ligands in the active site) could be introduced into the equation at different stoichiometric ratios with free Mg^{2+} ions (e.g. from $MgCl_2$) in order to assess their catalytic properties on nucleotide polymerisation.

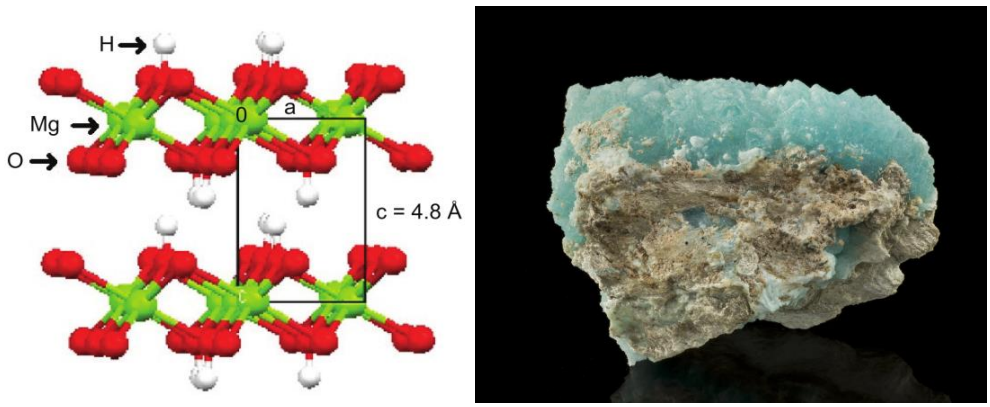


Figure 7.2. Lattice structure of brucite crystals (left) (Saifullah & Hussein, 2015), and a macroscopic brucite deposit (right).

I showed that the formose reaction indeed acted as a primordial source of monomeric sugars under alkaline hydrothermal vent conditions, with ribose (or more accurately, aldopentoses) being one of the main products during the first hours of reaction. More strikingly, I also found that the addition of AcP had a strong effect on the product palette. The maximum yields of ribose were not higher when compared to the control conditions (as initially suggested by the results using GC-MS) and the reaction was slowed overall, with the maximum yield peaks shifted to later reaction times. Despite this, the concentration of ribose remained high for a long time (both when AcP had been added after 30 or 60 minutes of reaction), whereas the other sugars measured were affected much more moderately. Due to the detection of acetylated Cannizzaro products, I attributed these effects to AcP-mediated acetylations, which most probably altered the fine equilibrium around glycolaldehyde in the early stages of the formose reaction; even so, the possible synthesis of reactive phosphorylated intermediates could not be discarded since their presence could explain the synthesis of experimentally found deoxysugars. I propose that the observed increased stability of the concentrations of ribose over time reflects an AcP-mediated promotion of the synthesis of ribulose, which slowly isomerises into ribose, replenishing its abundance. The reason why ribose seems more stable (compared with the respective control) when AcP was added after 60 minutes of reaction remains unknown, but probably reflects AcP having an effect onto the rates of hydrolysis (via retro-aldol condensations) of larger sugars, which were not measured during this study.

Admittedly, it is possible that limitations in the time to carry out these experiments are concealing other AcP-mediated effects; therefore further testing would be required. (i) A test assessing the rate of ribulose to ribose (and vice versa) isomerisation under the experimental conditions used would help explain whether a high amount of ribulose acting as a reservoir could explain the sustained high concentrations of ribose observed. This could be performed by simply monitoring the concentrations of ribulose and ribose over time, starting with a known amount of ribulose. (ii) Assessing the rate at which equimolar amounts of dihydroxyacetone and glycolaldehyde react to form ribulose under the experimental conditions used. This would shed light onto

whether this reaction is faster than the one between glycolaldehyde and formaldehyde, which is known to be slow (Mellersh & Smith, 2010) – one of the requirements if the proposed mechanism (Figure 6.16) is true. (iii) Future testing on the addition of AcP to the formose reaction should broaden the studied sugars in order to comprise both glycolaldehyde and sugars larger than 6C, since this should help having a more accurate understanding of its effect. (iv) Further experimentation using the aforementioned Y-shaped continuous-flow reactor could help better simulate the delivery of AcP to the formose reaction in time- and magnitude-wise fashions commensurate with the thermal cycling of hydrothermal fluids within a vent. (v) Finally, in order to better test the predictions made by Mellersh & Smith (2010) it would be of great interest to explore the effect of purely phosphorylating prebiotic reagents (unlike AcP), such as pyrophosphate or trimetaphosphate, on the formose reaction, with and without the presence of mineral surfaces.

In conclusion, alkaline hydrothermal vents constitute great candidates as the hatcheries of life on Hadean Earth. Their naturally occurring pH gradients would have constituted a source of energy congruent with how energy is conserved across all life today (chemiosmosis). Their mineral composition, rich in Fe(Ni)S, is also congruent with the prevalence of this type of cofactors in metalloproteins, particularly in CO_2-fixing metabolic pathways. In essence, these vents could have acted as giant aqueous flow reactors, continuously supplying simple organic molecules (at low yields), which could accumulate due to thermophoresis, increasing the chances of further proto-anabolic reactions; at the same time, waste materials would be washed out into the ocean due to their un-reactivity (e.g. acetate). In essence, the sustained far-from-equilibrium geochemical state of alkaline hydrothermal vents could have conceivably lead to the step-wise emergence of the far-from-equilibrium biochemical phenomenon we know as ‘life’. Of course, the work presented in this thesis is only a snapshot of a much larger set of events, but hopefully it adds another grain of sand necessary to construct a larger mountain.

7.2 Impact statement

The work presented in this thesis has a beneficial impact on an academic level, but as mentioned previously, the end goals of this field of research goes well beyond that.

On an academic level, this work represents pieces of evidence congruent with an autotrophic origin of life and, even though marine alkaline hydrothermal vents were the proposed setting here, many of the findings here could be extrapolated elsewhere (e.g. alkaline hydrothermal vents in fresh water volcanic lakes). This work aims to put in perspective the often-exaggerated importance of achieving high prebiotic yields, and advocates the contrary: congruence with ancient life and low non-enzymatic yields should be sought after. The methodological protocols detailed here, albeit somewhat specific in their scope, represent highly refined analytical methods developed across years, and could easily be transposed to other research fields.

Outside of academia, this work helps with the understanding of 'how and where' life started on Earth, which hopefully one day will help grasp 'why' life appeared on Earth, and potentially elsewhere in the Universe; these are fundamental questions that far exceed science. Furthermore, this research will contribute influencing political decision-making regarding future space exploration programmes, both in national and international institutions, since it is in line with evidence suggesting that Enceladus and Europa (satellites orbiting Saturn and Jupiter) could contain internal oceans, which in contact with an olivine-rich sea floor should promote the emergence of alkaline hydrothermal vents, and therefore potentially life.

The impact of this research has been further brought about thanks to several disseminating outputs in the form of: (i) four academic papers, two of which were first-authored (Camprubi et al., 2017; Whicher et al., 2018); (ii) numerous national and international scientific conference presentations and posters; (iii) and several public-engagement events, such as a stand in the Royal Society's Summer Science Exhibition, as well as talks to primary and secondary schools.

Bibliography

Amend, J. P., Larowe, D. E., Mccollom, T. M., & Shock, E. L. (2013). The energetics of organic synthesis inside and outside the cell The energetics of organic synthesis inside and outside the cell. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*(June).

Amend, J. P., & McCollom, T. M. (2009). Energetics of biomolecule synthesis on early Earth. In *Chemical Evolution II: From the Origins of Life to Modern Society* (Vol. 1025, pp. 63–94). ACS Symposium Series.

Amend, J. P., & Shock, E. L. (1998). Energetics of amino acid synthesis in hydrothermal ecosystems. *Science*, *281*(5383), 1659–1662.

Appayee, C., & Breslow, R. (2014). Deuterium studies reveal a new mechanism for the formose reaction involving hydride shifts. *Journal of the American Chemical Society*, *136*(10), 3720–3723.

Arndt, N. T., & Nisbet, E. G. (2012). Processes on the young Earth and the habitats of early life. *Annual Review of Earth and Planetary Sciences*, *40*(1), 521–549.

Baaske, P., Weinert, F. M., Duhr, S., Lemke, K. H., Russell, M. J., & Braun, D. (2007). Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proceedings of the National Academy of Sciences*, *104*(22), 9346–9351.

Baltscheffsky, M., & Baltscheffsky, H. (1995). Alternative photophosphorylation, inorganic pyrophosphate synthase and inorganic pyrophosphate. *Photosynthesis research* (Vol. 46).

Barge, L. M., Abedian, Y., Russell, M. J., Doloboff, I. J., Cartwright, J. H. E., Kidd, R. D., & Kanik, I. (2015). From chemical gardens to fuel cells:

generation of electrical potential and current across self-assembling iron mineral membranes. *Angewandte Chemie - International Edition*, *54*(28), 8184–8187.

Barge, L. M., Branscomb, E., Brucato, J. R., Cardoso, S. S. S., Cartwright, J. H. E., Danielache, S. O., … Sobron, P. (2017). Thermodynamics, disequilibrium, evolution: far-from-equilibrium geological and chemical considerations for origin-of-life research. *Origins of Life and Evolution of Biospheres*, *47*(1), 39–56.

Barge, L. M., Doloboff, I. J., Russell, M. J., VanderVelde, D., White, L. M., Stucky, G. D., … Kanik, I. (2014). Pyrophosphate synthesis in iron mineral films and membranes simulating prebiotic submarine hydrothermal precipitates. *Geochimica et Cosmochimica Acta*, *128*, 1–12.

Benner, S. A. (2013). Planets, minerals and life's origin. *Mineralogical Magazine*, *77*(5), 686.

Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., … Fuchs, G. (2010). Autotrophic carbon fixation in archaea. *Nature Reviews Microbiology*, *8*(6), 447–460.

Blakeney, A. B., Harris, P. J., Henry, R. J., & Stone, B. A. (1983). A simple and rapid preparation of alditol acetates for monosaccharide analysis. *Carbohydrate Research*, *113*, 291–299.

Boehnke, P., Harrison, T. (2016). Illusory late heavy bombardments. Proceedings of the National Academy of Sciences. 113 (39), 10802-10806.

Bowler, F. R., Chan, C. K. W., Duffy, C. D., Gerland, B., Islam, S., Powner, M. W., … Xu, J. (2013). Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. *Nature Chemistry*, *5*(5), 383–389.

Branscomb, E., Biancalani, T., Goldenfeld, N., Russell, M. (2017). Escapement

mechanisms and the conversion of disequilibria; the engines of creation. *Physics Reports*. 677 (October), 1-60.

Branscomb, E., & Russell, M. J. (2013). Turnstiles and bifurcators: The disequilibrium converting engines that put metabolism on the road. *Biochimica et Biophysica Acta - Bioenergetics*, *1827*(2), 62–78.

Bräsen, C., Schmidt, M., Grötzinger, J., & Schönheit, P. (2008). Reaction mechanism and structural model of ADP-forming acetyl-CoA synthetase from the hyperthermophilic archaeon Pyrococcus furiosus: Evidence for a second active site histidine residue. *Journal of Biological Chemistry*, *283*(22), 15409–15418.

Braun, D., & Libchaber, A. (2002). Trapping of DNA by thermophoretic depletion and convection. *Physical Review Letters*, *89*(18), 188103.

Breslow, R. (1959). On the mechanism of the formose reaction. *Tetrahedron Letters*, *1*(21), 22–26.

Buckel, W., & Thauer, R. K. (2013). Energy conservation via electron bifurcating ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation. *Biochimica et Biophysica Acta*, *1827*(2), 94–113.

Bujdák, J., & Rode, B. M. (1996). The effect of smectite composition on the catalysis of peptide bond formation. *Journal of Molecular Evolution*, *43*, 326–333.

Burcar, B., Pasek, M., Gull, M., Cafferty, B. J., Velasco, F., Hud, N. V, & Menor-salvan, C. (2016). Darwin's warm little pond: a one-pot reaction for prebiotic phosphorylation and the mobilization of phosphate from minerals in a urea-based solvent. *Angewandte Chemie - International Edition*, *55*, 1–6.

Burcar, B. T., Barge, L. M., Trail, D., Watson, E. B., Russell, M. J., & McGown, L. B. (2015). RNA Oligomerization in laboratory analogues of alkaline

hydrothermal vent systems. *Astrobiology*, (July), 1–14.

Burcar, B. T., Cassidy, L. M., Moriarty, E. M., Joshi, P. C., Coari, K. M., & McGown, L. B. (2013). Potential pitfalls in MALDI-TOF MS analysis of abiotically synthesized RNA oligonucleotides. *Origins of Life and Evolution of Biospheres*, *43*(3), 247–261.

Butlerow, A. (1861). Bildung einer zuckerartigen Substanz durch Synthese. *Justus Liebigs Annalen Der Chemie*, *120*(3), 295–298.

Campbell, B. J., & Craig Cary, S. (2004). Abundance of reverse tricarboxylic acid cycle genes in free-living microorganisms at deep-sea hydrothermal vents. *Appl Environ Microbiol*, *70*(10), 6282–6289.

Camprubi, E., Harrison, S., Pinna, S., Jordan, S. F., Lane, N. (MS in preparation; 2018). Acetyl phosphate directs the formose reaction towards ribose.

Camprubi, E., Jordan, S. F., Vasiliadou, R., & Lane, N. (2017). Iron catalysis at the origin of life. *IUBMB Life*, *69(6)*(March), 373–381.

Carter, C. W. (1993). Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. *Annual Review of Biochemistry*, *62*(1), 715–748.

Carvalho, A. T. P., Fernandes, P. A., & Ramos, M. J. (2011). The catalytic mechanism of RNA polymerase II. *Journal of Chemical Theory and Computation*, *7*(4), 1177–1188.

Chandru, K., Gilbert, A., Butch, C., Aono, M., & II, H. J. C. (2016). The abiotic chemistry of thiolated acetate derivatives and the origin of life. *Scientific*, (June), 1–11.

Charon, M. H., Volbeda, A, Chabriere, E., Pieulle, L., & Fontecilla-Camps, J. C.

(1999). Structure and electron transfer mechanism of pyruvate:ferredoxin oxidoreductase. *Current Opinion in Structural Biology*, *9*(6), 663–669.

Chen, F., & Yang, D. (2007). Condensation of amino acids to form peptides in aqueous solution induced by the oxidation of sulfur(iv): An oxidative model for prebiotic peptide formation. *Origins of Life and Evolution of Biospheres*, *37*(1), 47–54.

Chen, K., Bonagura, C. A., Tilley, G. J., McEvoy, J. P., Jung, Y.-S., Armstrong, F. A., … Burgess, B. K. (2002). Crystal structures of ferredoxin variants exhibiting large changes in [Fe-S] reduction potential. *Nature Structural Biology*, *9*(3), 188–192.

Chung, N. M., Lohrmann, R., Orgel, L. E., & Babinowitz, J. (1971). The mechanism of the trimetaphosphate-induced peptide synthesis. *Tetrahedron*, *27*, 1205–1210.

Cleland, C. E., & Chyba, C. F. (2002). Defining "Life." *Origins of Life and Evolution of the Biosphere*, *32*(4), 387–393.

Cody, G. D. (2004). Transition metal sulfides and the origins of metabolism. *Annual Review of Earth and Planetary Sciences*, *32*(1), 569–599.

Cody, G. D., Octor, N. Z. B., Randes, J. A. B., Illey, T. R. F., Azen, R. M. H., & Oder, H. S. Y. (2004). Assaying the catalytic potential of transition metal sulfides for abiotic carbon fixation. *Geochimica et Cosmochimica Acta*, *68*(10), 2185–2196.

Copley, S. D., Smith, E., & Morowitz, H. J. (2007). The origin of the RNA world: co-evolution of genes and metabolism. *Bioorganic Chemistry*, *35*(6), 430–43.

Costanzo, G., Pino, S., Ciciriello, F., & Di Mauro, E. (2009). Generation of long RNA chains in water. *Journal of Biological Chemistry*, *284*(48), 33206–

33216.

Costanzo, G., Saladino, R., Botta, G., Giorgi, A., Scipioni, A., Pino, S., & Di Mauro, E. (2012). Generation of RNA molecules by a base-catalysed Click-like reaction. *ChemBioChem*, *13*(7), 999–1008.

Costanzo, G., Saladino, R., Crestini, C., Ciciriello, F., & Di Mauro, E. (2007). Nucleoside phosphorylation by phosphate minerals. *Journal of Biological Chemistry*, *282*(23), 16729–16735.

Crans, D. C., & Whitesides, G. M. (1983). A convenient synthesis of disodium acetyl phosphate for use in in situ ATP cofactor regeneration. *J. Org. Chem*, *48*, 3130–3134.

Da Silva, L., Maurel, M.-C., & Deamer, D. (2015). Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. *Journal of Molecular Evolution*, *80*(2), 86–97.

de Duve, C. (1991). *Blueprint for a cell: The nature and origin of life*. New York: Neil Patterson Publishers.

de Duve, C. (1995). *Vital Dust: Life as a Cosmic Imperative*. Basic Books.

de Duve, C. (2005). *Singularities: Landmarks on the Pathways of Life* (First). New York: Cambridge University Press.

de Duve, C. (1998). Clues from present-day biology: the thioester world. In A. Brack (Ed.), *The Molecular Origins of Life: Assembling Pieces of the Puzzle* (pp. 219–236). Cambridge: Cambridge University Press.

de Duve, C. (1988). Did God make RNA? *Nature*, *336*(6196), 209–210.

de Zwart, I. I., Meade, S. J., & Pratt, A. J. (2004). Biomimetic phosphoryl transfer catalysed by iron(II)-mineral precipitates. *Geochimica et*

Cosmochimica Acta, *68*(20), 4093–4098.

Decker, K., Jungermann, K., & Thauer, R. K. (1970). Energy production in anaerobic organisms. *Angewandte Chemie (International Ed. in English)*, *9*(2), 138–158.

Deng, L., Pan, T., Xu, Q., Chen, M., Zhang, Y., Guo, Q., Fu, Y. (2013). Linked strategy for the prouction of fuels via the formose reaction. *Scientific reports*. 3, 1-7.

Deguzman, V., Vercoutere, W., Shenasa, H., & Deamer, D. (2014). Generation of oligonucleotides under hydrothermal conditions by non-enzymatic polymerization. *Journal of Molecular Evolution*, *78*(5), 251–262.

Di Sabato, G., & Jencks, W. P. (1961a). Mechanism and catalysis of reactions of acyl phosphates. I. Nucleophilic reactions. *J. Am. Chem. Soc.*, *83*, 4400–4405.

Di Sabato, G., & Jencks, W. P. (1961b). Mechanism and catalysis of reactions of acyl phosphates. II. Hydrolysis. *Journal of the American Chemical Society*, *83*(21), 4400–4405.

Dzade, N. Y., Roldan, A., & de Leeuw, N. H. (2016). Surface and shape modification of mackinawite (FeS) nanocrystals by cysteine adsorption: a first-principles DFT-D2 study. *Phys. Chem. Chem. Phys.*, *18*(47), 32007–32020.

Dzade, N. Y., Roldan, A., & De Leeuw, N. H. (2015). Activation and dissociation of CO on the (001), (011), and (111) surfaces of mackinawite (FeS): A dispersion-corrected DFT study. *Journal of Chemical Physics*, *143*(9).

Eck, R. V., & Dayhoff, M. O. (1966). Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science*, (152), 363–366.

Eigen, M. (1971). Die naturwissenschaften 58. *Naturwissenschaften*, *58*, 465–562.

Etaix, E., & Buvet, R. (1974). Conditions of occurrence for primeval processes of transphosphorylations - Cosmochemical Evolution and the Origins of Life. J. Oró, S. L. Miller, C. Ponnamperuma, & R. S. Young (Eds.), (pp. 175–183). Dordrecht: Springer Netherlands.

Evans, M. C. W., Buchanan, B. B., & Arnon, D. I. (1966). A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proceedings of the National Academy of Sciences of the United States of America*, *55*, 928–934.

Ferry, J. G. (2010). How to make a living by exhaling methane. *Annu. Rev. Microbiol. Lett.* 64, 453-473.

Ferry, J. G., & House, C. H. (2006). The stepwise evolution of early life driven by energy conservation. *Mol. Biol. Evol.*, *23*, 1286–1292.

Forsythe, J. G., Yu, S. S., Mamajanov, I., Grover, M. A., Krishnamurthy, R., Fernández, F. M., & Hud, N. V. (2015). Ester-mediated amide bond formation driven by wet-dry cycles: a possible path to polypeptides on the prebiotic Earth. *Angewandte Chemie - International Edition*, *54*(34), 9871–9875.

Fowler, M. L., Ingram-Smith, C. J., & Smith, K. S. (2011). Direct detection of the acetate-forming activity of the enzyme acetate kinase. *Journal of Visualized Experiments : JoVE*, (58), 1–7.

Fuchs, G. (2011). Alternative Pathways of Carbon Dioxide Fixation: Insights into the Early Evolution of Life? *Annual Review of Microbiology* (Vol. 65).

Furukawa, Y., Horiuchi, M., & Kakegawa, T. (2013). Selective stabilization of ribose by borate. *Origins of Life and Evolution of the Biospheres*, *43*(4–5),

353–361.

Gibard, C., Bhowmik, S., Karki, M., Kim, E., & Krishnamurthy, R. (2017). Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nature Chemistry*, (November), 1–6.

Goldford, J. E., Hartman, H., Smith, T. F., & Segrè, D. (2017). Remnants of an ancient metabolism without phosphate. *Cell*, *168*(6), 1126–1134.e9.

Goldschmidt, V. M. (1952). Geochemical aspects of the origin of complex organic molecules on the Earth, as precursors to organic life. *New Biol.*, (12), 97–105.

Gollihar, J., Levy, M., & Ellington, A. D. (2014). Many paths to the origin of life. *Science*, *343*(6168), 259–60.

Griesser, H., Bechthold, M., Tremmel, P., Kervio, E., & Richert, C. (2017). Amino acid-specific, ribonucleotide-promoted peptide formation in the absence of enzymes. *Angewandte Chemie - International Edition*, *56*(5), 1224–1228.

Gull, M., Mojica, M. A., Fernández, F. M., Gaul, D. A., Orlando, T. M., Liotta, C. L., & Pasek, M. A. (2015). Nucleoside phosphorylation by the mineral schreibersite. *Scientific Reports*, *5*, 17198.

Haldane, J. (1929). The Origin of Life. *Rationalist Annual*, 3–10.

Hamada, K., Ago, H., Sugahara, M., Nodake, Y., Kuramitsu, S., & Miyano, M. (2003). Oxyanion hole-stabilized stereospecific isomerization in ribose-5-phosphate isomerase (Rpi). *The Journal of Biological Chemistry*, *278*(49), 49183–49190.

Han, J., Tschernutter, V., Yang, J., Eckle, T., & Borchers, C. H. (2013). Analysis of selected sugars and sugar phosphates in mouse heart tissue by

reductive amination and liquid chromatography-electrospray ionization mass spectrometry. *Analytical Chemistry*, *85*(12), 5965–73.

Hedderich, R. (2004). Energy-converting [NiFe] hydrogenases from archaea and extremophiles: Ancestors of complex I. *Journal of bioenergetics and biomembranes* (Vol. 36).

Hedderich, R., Forzi, L. (2006). Energy-converting [NiFe] hydrogenases: More than just H_2 activation. *Journal of Molecular Microbiology and Biotechnology*. 10 (2-4), 92-104.

Heinen, W., & Lauwers, A. M. (1996). Organic Sulfur compounds resulting from the interaction of iron sulfide, hydrogen sulfide and carbon dioxide in an anaerobic aqueous environment. *Origins of Life and Evolution of the Biosphere*, *26*, 131–150.

Herrmann, G., Jayamani, E., Mai, G., & Buckel, W. (2008). Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. *Journal of Bacteriology*, *190*(3), 784–791.

Herschy, B., Whicher, A., Camprubi, E., Watson, C., Dartnell, L., Ward, J., … Lane, N. (2014). An origin-of-Life reactor to simulate alkaline hydrothermal vents. *Journal of Molecular Evolution*, *79*(5–6), 213–227.

Holm, N. G. (2012). The significance of Mg in prebiotic geochemistry. *Geobiology*, *10*(4), 269–279.

Holm, N. G., Dumont, M., Ivarsson, M., & Konn, C. (2006). Alkaline fluid circulation in ultramafic rocks and formation of nucleotide constituents: A hypothesis. *Geochemical Transactions*, *7*, 14–16.

Huber, C., Eisenreich, W., Hecht, S., & Wächtershäuser, G. (2003). A possible primordial peptide cycle. *Science*, *301*(August), 938–940.

Huber, C., & Wächtershäuser, G. (1997). Activated acetic acid by carbon fixation on (Fe,Ni)S under primordial conditions. *Science*, *276*(5310), 245–247.

Huber, C., & Wächtershäuser, G. (1998). Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces : implications for the origin of life. *Science*, *281*(July), 670–673.

Hulshof, J., & Ponnamperuma, C. (1976). Prebiotic condensation reactons in an aqueous meduim: a review of condensing agents. *Origins Of Life*, *6*(1970), 197–224.

Pratt, A. J. (2006). *The curious case of phosphate solubility*. *Chem. N.Z.*

Jauker, M., Griesser, H., & Richert, C. (2015). Spontaneous formation of RNA strands, peptidyl RNA, and cofactors. *Angewandte Chemie - International Edition*, *54*(48), 14564–14569.

Kaster, A.-K., Moll, J., Parey, K., & Thauer, R. K. (2011). Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *Proceedings of the National Academy of Sciences*, *108*(7), 2981–2986.

Keller, M. A., Driscoll, P. C., Messner, C., & Ralser, M. (2017). Primordial Krebs-cycle-like non-enzymatic reactions detected by mass spectrometry and nuclear magnetic resonance. *Wellcome Open Research*.

Keller, M. A., Kampjut, D., Harrison, S. A., & Ralser, M. (2017). Sulfate radicals enable a non-enzymatic Krebs cycle precursor. *Nature Ecology and Evolution*, *1*.

Keller, M. A, Turchyn, A. V, & Ralser, M. (2014). Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archean ocean. *Molecular Systems Biology*, *10*, 1–12.

Kelley, D. S., Karson, J. A, Blackman, D. K., Früh-Green, G. L., Butterfield, D. A, Lilley, M. D., … Rivizzigno, P. (2001). An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30 degrees N. *Nature*, *412*(6843), 145–9.

Kelley, D. S., Karson, J. A, Fru, G. L., Yoerger, D. R., Shank, T. M., Butterfield, D. A, … Sylva, S. P. (2005). A Serpentinite-Hosted Ecosystem: The Lost City Hydrothermal Field. *Science*, *307*(March), 1428–1434.

Kieboom, A. P., & VanBekkum, H. (1984). Aspects of the chemical conversion of glucose. *Recueil Des Travaux Chimiques Des Pays-Bas*, *103*(1), 1–12.

Kim, H. J., Ricardo, A., Illangkoon, H. I., Kim, M. J., Carrigan, M. A., Frye, F., & Benner, S. A. (2011). Synthesis of carbohydrates in mineral-guided prebiotic cycles. *Journal of the American Chemical Society*, *133*(24), 9457–9468.

Kitani, A., Tsunetsugu, S. ichi, Suzuki, A., Ito, S., & Sasaki, K. (1995). Fe(III)-ion-catalysed non-enzymatic transformation of adenosine diphosphate into adenosine triphosphate part II. Evidence of catalytic nature of Fe ions. *Bioelectrochemistry and Bioenergetics*, *36*(1), 47–51.

Kopetzki, D., & Antonietti, M. (2011). Hydrothermal formose reaction. *New Journal of Chemistry*, *35*(9), 1787.

Kornberg, A., Rao, N. N., & Ault-Riché, D. (1999). Inorganic polyphosphate: a molecule of many functions. *Annual Review of Biochemistry*, *68*(1), 89–125.

Koshland, D. E. (1952). Effect of catalysts on the hydrolysis of acetyl phosphate. Nucleophilic displacement mechanisms in enzymatic reactions. *Journal of the American Chemical Society*, *74*(9), 2286–2292.

Kreysing, M., Keil, L., Lanzmich, S., & Braun, D. (2015). Heat flux across an open pore enables the continuous replication and selection of

oligonucleotides towards increasing length. *Nature Chemistry*, *7*(3), 203–208.

Lane, N. (2014). Bioenergetic constraints on the evolution of complex life. *Cold Spring Harbor Perspectives in Biology*, *6*(5), a015982.

Lane, N. (2015). The vital question: energy, evolution and the origins of complex life. New York: WWNorton.

Lane, N., & Martin, W. F. (2012). The origin of membrane bioenergetics. *Cell*, *151*(7), 1406–16.

Le Son, H., Suwannachot, Y., Bujdak, J., & Rode, B. M. (1998). Salt-induced peptide formation from amino acids in the presence of clays and related catalysts. *Inorganica Chimica Acta*, *272*, 89–94.

Leduc, S. (1911). *The Mechanism of Life*. London: Rebman Ltd.

Leman, L., Orgel, L., & Ghadiri, M. R. (2004). Carbonyl sulfide-mediated prebiotic formation of peptides. *Science (New York, N.Y.)*, *306*(5694), 283–286.

Ligon, B. L. (2002). Louis Pasteur : a controversial figure in a debate on scientific ethics. *Seminars in Pediatric Infectious Diseases*, *13*(2), 134–141.

Liu, R., & Orgel, L. E. (1997). Oxidative acylation using thioacids. *Nature*, *389*(6646), 52–54. http://doi.org/10.1038/37944

Lohrmann, R., & Orgel, L. E. (1968). Prebiotic synthesis: phosphorylation in aqueous solution. *Science*, *161*, 64–66. Retrieved from http://dx.doi.org/10.1126/science.161.3836.64

Maden, B. E. H. (2000). Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C1 metabolism. *Biochem. J*, *350*,

609–629.

Martin, W., Baross, J., Kelley, D., & Russell, M. J. (2008). Hydrothermal vents and the origin of life. *Nature Reviews Microbiology*, *6*(11), 805–814. http://doi.org/10.1038/nrmicro1991

Martin, W. F. (2012). Hydrogen, metals, bifurcating electrons, and proton gradients: the early evolution of biological energy conservation. *FEBS Letters*, *586*(5), 485–93.

Martin, W. F., & Thauer, R. K. (2017). Energy in ancient metabolism. *Cell*, *168*(6), 953–955.

Martin, W., & Russell, M. J. (2003). On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1429), 59–85.

Martin, W., & Russell, M. J. (2007). On the origin of biochemistry at an alkaline hydrothermal vent. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1486), 1887–1926.

Martin, W., Sousa, F., & Lane, N. (2014). Energy at life's origin. *Science*, *344*(6188), 1092–1093.

Martra, G., Deiana, C., Sakhno, Y., Barberis, I., Fabbiani, M., Pazzi, M., & Vincenti, M. (2014). The formation and self-assembly of long prebiotic oligomers produced by the condensation of unactivated amino acids on oxide surfaces. *Angewandte Chemie - International Edition*, *53*(18), 4671–4674.

Mast, C. B., & Braun, D. (2010). Thermal trap for DNA replication. *Physical Review Letters*, *104*(18), 188102.

Mast, C. B., Schink, S., Gerland, U., & Braun, D. (2013). Escalation of polymerization in a thermal gradient. *Proceedings of the National Academy of Sciences*, *110*(20), 8030–8035.

McCollom, T. M. (2013). Miller-Urey and beyond: what have we learned about prebiotic organic synthesis reactions in the past 60 years? *Annual Review of Earth and Planetary Sciences*, *41*(1), 207–229.

Meinert, C., Myrgorodska, I., de Marcellus, P., Buhse, T., Nahon, L., Hoffmann, S. V, … Meierhenrich, U. J. (2016). Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science*, *publishing*(0), 1–5.

Mellersh, A. R., & Smith, P. M. (2010). The alkaline world and the origin of life. *Journal of Cosmology*, *10*, 3230–3242.

Messner, C., Driscoll, P. C., Piedrafita, G., De Volder, M. F. L., & Ralser, M. (2017). Nonenzymatic gluconeogenesis-like formation of fructose 1,6-bisphosphate in ice. *Proc. Natl Acad. Sci. USA*, (June), 1–5.

Miller, S. L. (1953). A production of amino acids under possible primitive Earth conditions. *Science*, *117*(3046), 528–529.

Milner-White, E. J., & Russell, M. J. (2005). Sites for phosphates and iron-sulfur thiolates in the first membranes: 3 to 6 residue anion-binding motifs (Nests). *Origins of Life and Evolution of Biospheres*, *35*(1), 19–27.

Mitchell, P. (1959). The origin of life and the formation and organizing functions of natural membranes. In C. F. & R. L. M. Synge (Eds.), *Proceedings of the First International Symposium on the Origin of Life on the Earth* (pp. 437–443). New York: Pergamon Press.

Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*, *191*, 144–148.

Miyakawa, S., Yamanashi, H., Kobayashi, K., Cleaves, H. J., & Miller, S. L. (2002). Prebiotic synthesis from CO atmospheres : Implications for the origins of life. *Proceedings of the National Academy of Sciences*, *99*(23), 14628–14631.

Möller, F. M., Kriegel, F., Kieß, M., Sojo, V., & Braun, D. (2017). Steep pH gradients and directed colloid transport in a microfluidic alkaline hydrothermal pore. *Angewandte Chemie - International Edition*, *56*(9), 2340–2344.

Morasch, M., Mast, C. B., Langer, J. K., Schilcher, P., & Braun, D. (2014). Dry polymerization of 3',5'-cyclic GMP to long strands of RNA. *ChemBioChem*, 879–883.

Morowitz, H. J., Heinz, B., & Deamer, D. W. (1988). The chemical logic of a minimum protocell. *Origins of Life and Evolution of the Biosphere*, *18*(3), 281–287.

Morowitz, H. J., Kostelnik, J. D., Yang, J., & Cody, G. D. (2000). The origin of intermediary metabolism. *Proceedings of the National Academy of Sciences*, *97*(14), 7704–7708.

Muchowska, K. B., Varma, S. J., Chevallot-beroux, E., Lethuillier-Karl, L., Li, G., & Moran, J. (2017). Metals promote sequences of the reverse Krebs cycle. *Nature Ecology & Evolution*, 1–6.

Nakamura, R., Takashima, T., Kato, S., Takai, K., Yamamoto, M., & Hashimoto, K. (2010). Electrical current generation across a black smoker chimney. *Angewandte Chemie - International Edition*, *49*(42), 7692–7694.

Nicholls, D. G., & Ferguson, S. J. (2013). *Bioenergetics 4*. London: Academic Press.

Nitschke, W., McGlynn, S. E., Milner-White, E. J., & Russell, M. J. (2013). On

the antiquity of metalloenzymes and their substrates in bioenergetics. *Biochimica et Biophysica Acta - Bioenergetics*, *1827*(8–9), 871–881.

Nitschke, W., & Russell, M. J. (2009). Hydrothermal focusing of chemical and chemiosmotic energy, supported by delivery of catalytic Fe, Ni, Mo/W, Co, S and Se, forced life to emerge. *Journal of Molecular Evolution*, *69*(5), 481–496.

Nitschke, W., & Russell, M. J. (2013). Beating the acetyl coenzyme A-pathway to the origin of life. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1622), 20120258–20120258.

Nygren, Y., Fredriksson, S.-A., & Nilsson, B. (1996). Identification of sialic acid and related acids as acetylated lactones by gas chromatography/mass spectrometry. *Journal of Mass Spectrometry*, *31*(September 1995), 267–274.

Oparin, A. I. (1924). The origin of life. *Proiskhozhdenie Zhizny*.

Orgel, L. E. (2004). Prebiotic chemistry and the origin of the RNA world. *Critical Reviews in Biochemistry and Molecular Biology*, *39*, 99–123.

Orgel, L. E. (2008). The implausibility of metabolic cycles on the prebiotic Earth. *PLoS Biology*, *6*(1), e18.

Oro, J. (1960). Synthesis of adenine from ammonium cyanide. *Biochem. Biophys. Res. Commun.*, *2*(6), 407–412.

Pasteur, L. (1862). *Mémoire sur les corpuscules organisés qui existent dans l'atmosphère*. Mallet-Bachelier.

Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D., & Sutherland, J. D. (2015). Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nature Chemistry*, *7*(4), 301–307.

Pepi, F., Garzoli, S., Tata, A., & Giacomello, P. (2010). Low-energy collisionally activated dissociation of pentose–borate complexes. *International Journal of Mass Spectrometry*, *289*(2–3), 76–83.

Pino, S., Sponer, J. E., Costanzo, G., Saladino, R., & Di Mauro, E. (2015). From formamide to RNA, the path is tenuous but continuous. *Life*, *5*(1), 372–84.

Pinti, D. L. (2005). The origin and evolution of the oceans. In *Lectures in Astrobiology,Vo* (Vol. I, pp. 83–112).

Powner, M. W., Gerland, B., & Sutherland, J. D. (2009). Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, *459*(7244), 239–242.

Pratt, A. J. (2011). Prebiological evolution and the metabolic origins of life. *Artificial Life*, *17*(3), 203–17.

Proskurowski, G., Lilley, M. D., Seewald, J. S., Früh-green, G. L., Olson, E. J., Lupton, J. E., … Kelley, D. S. (2008). Abiogenic hydrocarbon production at Lost City hydrothermal field. *Science*, *319*(1), 604–607.

Rabinowitz, J. (1970). Peptide and amide bond formation in aqueous solutions of cyclic or linear polyphosphates as a possible prebiotic process. *Helvetica Chimica Acta*, *53*(6), 1350–1355.

Ragsdale, S. W., & Kumar, M. (1996). Nickel-containing carbon monoxide dehydrogenase/acetyl-CoA synthase. *Chemical Reviews*, *96*(7), 2515–2539.

Rajamani, S., Vlassov, A., Benner, S., Coombs, A., Olasagasti, F., & Deamer, D. (2008). Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Origins of Life and Evolution of Biospheres*, *38*(1), 57–74.

Reeves, E. P., McDermott, J. M., & Seewald, J. S. (2014). The origin of methanethiol in midocean ridge hydrothermal fluids. *Proceedings of the National Academy of Sciences*, *111*(15), 5474–5479.

Ricardo, A., Carrigan, M. A., Olcott, A. N., & Benner, S. A. (2004). Borate minerals stabilize ribose. *Science*, *303*(5655), 196.

Roldan, A., Hollingsworth, N., Roffey, A., Islam, H.-U., Goodall, J. B. M., Catlow, C. R. A., … de Leeuw, N. H. (2015). Bio-inspired CO2 conversion by iron sulfide catalysts under sustainable conditions. *Chem. Commun.*, *51*(35), 7501–7504.

Ruiz-Mirazo, K., Peretó, J., & Moreno, A. (2004). A universal definition of life: autonomy and open-ended evolution. *Origins of Life and Evolution of the Biosphere*, *34*(3), 323–346.

Russell, M. J., Barge, L. M., Bhartia, R., Bocanegra, D., Bracher, P. J., Branscomb, E., … Kanik, I. (2014). The drive to life on wet and icy worlds. *Astrobiology*, *14*(4), 308–343.

Russell, M. J., Daniel, R. M., Hall, A. J., & Sherringham, J. A. (1994). A hydrothermally precipitated catalytic iron sulphide membrane as a first step toward life. *Journal of Molecular Evolution*, *39*(3), 231–243.

Russell, M. J., & Hall, A. J. (1997). The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *Journal of the Geological Society*, *154*(3), 377–402.

Russell, M. J., Hall, A. J., Cairns-Smith, A. G., & Braterman, P. S. (1988). Submarine hot springs and the origin of life. *Nature*, *336*(10), 117.

Russell, M. J., Hall, A. J., & Turner, D. (1989). In-vitro growth of iron sulphide chimneys: Possible culture chambers for origin-of-life experiments. *Terra Nova*, *1*, 238–241.

Russell, M. J., & Martin, W. (2004). The rocky roots of the acetyl-CoA pathway. *Trends in Biochemical Sciences*, *29*(7), 358–363.

Saifullah, B., & Hussein, M. Z. B. (2015). Inorganic nanolayers: Structure, preparation, and biomedical applications. *International Journal of Nanomedicine*, *10*, 5609–5633.

Saladino, R., Crestini, C., Ciciriello, F., Pino, S., Costanzo, G., & Di Mauro, E. (2009). From formamide to RNA: the roles of formamide and water in the evolution of chemical information. *Research in Microbiology*, *160*(7), 441–448.

Sapra, R., Bagramyan, K., Adams, M. (2003). *Proceedings of the National Academy of Sciences of the United States of America*. 100 (13), 7545-7550.

Saygin, Ö. (1983). Nonenzymatic phosphorylation of acetate by carbamyl phosphate. *Origins of Life*, *13*(1), 43–48.

Schieven, G., & Martin, G. S. (1988). Nonenzymatic phosphorylation of tyrosine and serine by ATP is catalyzed by manganese but not magnesium. *The Journal of Biological Chemistry*, *263*(30), 15590–3.

Schonheit, P., Buckel, W., & Martin, W. F. (2016). On the origin of heterotrophy. *Trends in Microbiology*, *24*(1), 12–25.

Segré, D., Ben-Eli, D., Deamer, D. W., & Lancet, D. (2001). The lipid world. *Origins of Life and Evolution of the Biosphere*, *31*(1), 119–145.

Sen, S., Pal, U., & Maiti, N. C. (2014). pKa determination of D-ribose by raman spectroscopy. *The Journal of Physical Chemistry B*, 909–914.

Shock, E., & Canovas, P. (2010). The potential for abiotic organic synthesis and biosynthesis at seafloor hydrothermal systems. *Frontiers in Geofluids*, 161–192.

Shock, E. L., McCollom, T. M., & Schulte, M. D. (1998). The emergence of metabolism from within hydrothermal systems. In Thermophiles: The Keys to Molecular Evolution and the Origin of Life? In M. W. W. Wiegel, J. and Adams (Ed.). London: Taylor and Francis.

Smith, E., & Morowitz, H. J. (2004). Universality in intermediary metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), 13168–13173.

Socha, R. F. (1980). Autocatalysis in the formosereaction. *React. Kinet. Catal. Lett.*

Sojo, V., Herschy, B., Whicher, A., Camprubi, E., & Lane, N. (2016). The origin of life in alkaline hydrothermal vents. *Astrobiology*, *16*(2), 181–197.

Sojo, V., Pomiankowski, A., & Lane, N. (2014). A bioenergetic basis for membrane divergence in archaea and bacteria. *PLoS Biology*, *12*(8).

Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., & Goldfarb, A. (2003). Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *The EMBO Journal*, *22*(9), 2234–44.

Sousa, F. L., & Martin, W. F. (2014). Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochimica et Biophysica Acta - Bioenergetics*, *1837*(7), 964–981.

Sousa, F. L., Thiergart, T., Landan, G., Nelson-Sathi, S., Pereira, I. A. C., Allen, J. F., … Martin, W. F. (2013). Early bioenergetic evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1622),

Šponer, J. E., Šponer, J., Giorgi, A., Di Mauro, E., Pino, S., & Costanzo, G. (2015). Untemplated nonenzymatic polymerization of 3′,5′cGMP: A plausible route to 3′,5′-linked oligonucleotides in primordia. *The Journal of*

Physical Chemistry B, *119*(7), 2979–2989.

Steitz, T. A, & Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(14), 6498–502.

Sugaya, N., Nakagawa, T., Sakurai, K., Morita, M., & Onodera, S. (2001). Analysis of aldehydes in water by head space-GC/MS. *Journal of Health Science*, *47*(1), 21–27.

Šurín, S., Čuboňová, L., Majerník, A. I., McDermott, P., Chong, J. P. J., & Šmigáň, P. (2007). Isolation and characterization of an amiloride-resistant mutant of Methanothermobacter thermautotrophicus possessing a defective Na+/H+antiport. *FEMS Microbiology Letters*, *269*(2), 301–308.

Thauer, R. K., Jungermann, K., & Decker, K. (1977). Energy conservation in chemotrophic anaerobic bacteria. *Bacteriological Reviews*, *41*(1), 100–180.

Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W., & Hedderich, R. (2008). Methanogenic archaea: Ecologically relevant differences in energy conservation. *Nature Reviews Microbiology*, *6*(8), 579–591.

Thomson, A. J., Gray, H. B. (1998). Bioinorganic chemistry. *Current opinion in Chemical Biology*. 2: 155-158.

Trail, D., Watson, E. B., & Tailby, N. D. (2011). The oxidation state of Hadean magmas and implications for early Earth's atmosphere. *Nature*, *480*, 79–82.

Varma, S. J., Muchowska, K. B., Chatelain, P., & Moran, J. (2017). Metals enable a non-enzymatic acetyl CoA pathway. *bioRxiv*, 235523.

Vasant, C., Sankaramanivel, S., Jana, M., Rajaram, R., & Ramasami, T. (2005). Non-enzymatic phosphorylation of bovine serum albumin by Cr(V)

complexes: role in Cr(VI)-induced phosphorylation and toxicity. *Molecular and Cellular Biochemistry*, *275*(1–2), 153–64.

Wächtershäuser, G. (1988a). Before enzymes and templates: theory of surface. *Microbiol. Rev.*, *52*(4), 452.

Wächtershäuser, G. (1988b). Pyrite Formation, the First Energy Source for Life: a Hypothesis. *Systematic and Applied Microbiology*, *10*(3), 207–210.

Wächtershäuser, G. (1990). Evolution of the first metabolic cycles. *Proceedings of the National Academy of Sciences*, *87*(1), 200–204.

Wang, W., Song, Y., Wang, X., Yang, Y., & Liu, X. (2015). Alpha-oxo acids assisted transformation of FeS to Fe3S4 at low temperature: Implications for abiotic, biotic, and prebiotic mineralization. *Astrobiology* (Vol. 15).

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-sathi, S., & Martin, W. F. (2016). The physiology and habitat of the last universal common ancestor. *Nature Microbiology*, *1*(July), 1–8.

West, T., Sojo, V., Pomiankowski, A., & Lane, N. (2017). The origin of heredity in protocells. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*.

Westheimer, F. (1987). Why nature chose phosphates. *Science*, *235*(4793), 1173–1178.

Whicher, A., Camprubi, E., Pinna, S., Herschy, B., & Lane, N. (2018). Acetyl phosphate as the primordial energy currency at the origin of life. *Origins of Life and Evolution of Biospheres*.

Wiebe, R., Gaddy, V. L., & Heins Jr, C. (1932). Solubility of hydrogen in water at 25c from 25 to 1000 atmospheres. *Industrial & Engineering Chemistry*, *24*(7), 823–825.

Williams, S. P., Fulton, A M., & Brindle, K. M. (1993). Estimation of the intracellular free ADP concentration by 19F NMR studies of fluorine-labeled yeast phosphoglycerate kinase in vivo. *Biochemistry*, *32*(18), 4895–4902.

Wolthers, M., Charlet, L., van Der Linde, P. R., Rickard, D., & van Der Weijden, C. H. (2005). Surface chemistry of disordered mackinawite (FeS). *Geochimica et Cosmochimica Acta*, *69*(14), 3469–3481.

Wong, J. T.-F., & Lazcano, A. (2009). Biomolecules. In *Prebiotic Evolution and Astrobiology* (pp. 65–75). Austin, Texas: Landes Bioscience.

Yamagata, Y., Watanabe, H., Saitoh, M., & Namba, T. (1991). Volcanic production of polyphosphates and its relevance to prebiotic evolution. *Nature*.

Yamaguchi, A., Yamamoto, M., Takai, K., Ishii, T., Hashimoto, K., & Nakamura, R. (2014). Electrochemical CO2 reduction by Ni-containing iron sulfides: How is CO2 electrochemically reduced at bisulfide-bearing deep-sea hydrothermal precipitates? *Electrochimica Acta*, *141*, 311–318.

Yoshida, H., Yoshihara, A., Teraoka, M., Yamashita, S., Izumori, K., & Kamitori, S. (2012). Structure of l-rhamnose isomerase in complex with l-rhamnopyranose demonstrates the sugar-ring opening mechanism and the role of a substrate sub-binding site. *FEBS Open Bio*, *3*, 35–40.

Zahnle, K., Arndt, N., Cockell, C., Halliday, A., Nisbet, E., Selsis, F., & Sleep, N. H. (2007). Emergence of a habitable planet. In K. E. Fishbaugh, P. Lognonné, F. Raulin, D. J. Des Marais, & O. Korablev (Eds.), *Geology and Habitability of Terrestrial Planets* (pp. 35–78). New York, NY: Springer New York.