Energetik und Stoffwechsel des letzten gemeinsamen Urvorfahren

Inaugural–Dissertation

zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Jessica Lidwina Elisabeth Wimmer
geboren in Dorsten

Düsseldorf, März 2022

Aus dem Institut für Molekulare Evolution
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

Berichterstatter:
1. Prof. Dr. William F. Martin
2. apl. Prof. Dr. Ing. Gerhard Steger

Tag der mündlichen Prüfung: 15.06.2022

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass diese Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die Arbeit wurde bisher keiner Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Ich habe bisher keinen erfolglosen Promotionsversuch unternommen.

_____

Jessica Lidwina Elisabeth Wimmer, Düsseldorf, 2022

Für Angelika.

Für Joseph Hans.

Im Laufe dieser Arbeit wurden mit Zustimmung des Betreuers folgende Beiträge veröffentlicht:

Publikationen in Fachzeitschriften thematisiert in dieser Thesis

I **Jessica L. E. Wimmer**, Andrey d. N. Vieira, Joana C. Xavier, Karl Kleinermanns, William F. Martin, Martina Preiner (2021). The autotrophic core: An ancient network of 404 reactions converts H_2, CO_2, and NH_3 into amino acids, bases, and cofactors. *Microorganisms* 9, 458.

II **Jessica L. E. Wimmer**, Joana C. Xavier, Andrey d. N. Vieira, Delfina P. H. Pereira, Jacqueline Leidner, Filipa L. Sousa, Karl Kleinermanns, Martina Preiner, William F. Martin (2021). Energy at origins: Favorable thermodynamics of biosynthetic reactions in the Last Universal Common Ancestor (LUCA). *Frontiers in Microbiology* 12, 793664.

III **Jessica L. E. Wimmer**, Karl Kleinermanns, William F. Martin (2021). Pyrophosphate and irreversibility in evolution, or why PP_i is not an energy currency and why nature chose triphosphates. *Frontiers in Microbiology* 12, 759359.

IV Joana C. Xavier, Rebecca E. Gerhards, **Jessica L. E. Wimmer**, Julia Brueckner, Fernando D. K. Tria, William F. Martin (2021). The metabolic network of the Last Bacterial Common Ancestor. *Communications Biology* 4, 413.

Weitere Publikationen in Fachzeitschriften

V Fernando D. K. Tria, Julia Brueckner, Josip Skejo, Joana C. Xavier, Nils Kaput, Michael Knopp, **Jessica L. E. Wimmer**, Falk S. P. Nagies, Verena Zimorski, Sven B. Gould, Sriram G. Garg, William F. Martin (2021). Gene duplications trace mitochondria to the onset of eukaryote complexity. *Genome Biology and Evolution* 13, evab055.

Inhaltsverzeichnis

1 Zusammenfassung

Das Leben besteht aus einer Abfolge chemischer Reaktionen, welche im Zusammenspiel ein metabolisches Netzwerk bilden. Die grundlegenden Reaktionen am Ursprung des Stoffwechsels, noch bevor Leben existierte, müssen energetisch günstig gewesen sein, um spontan entstanden sein zu können und damit den Grundstein für einen hocheffizienten Metabolismus zu setzen. Am Beginn dieses Prozesses vor etwa 4 Milliarden Jahren stehen möglicherweise geochemische Vorgänge in der Erdkruste in serpentinisierenden Systemen am Meeresgrund, den Hydrothermalquellen.

Im Rahmen dieser Arbeit wird durch Konzipierung eines biosynthetischen Reaktionsnetzwerkes der autotrophe Grundstoffwechsel des letzten gemeinsamen Vorfahren aller Lebewesen (LUCA) charakterisiert. Dieser umfasst mehr als 400 metabolische Reaktionen, welche zur Synthese von Aminosäuren, Basen und Cofaktoren benötigt werden, ausgehend von den Metaboliten Wasserstoff (H_2), Kohlenstoffdioxid (CO_2) und Ammoniak (NH_3), sowie Schwefelwasserstoff (H_2S) und Phosphat. Diese Komponenten waren vermutlich bereits im frühen Erdzeitalter in Hydrothermalquellen verfügbar. Eine thermodynamische Untersuchung dieser Reaktionen unter Hydrothermalbedingungen zeigt, dass 95–97 % exergon sind ($\Delta G \leq 0$ $kJ \cdot mol^{-1}$), d. h. Energie freisetzen und somit spontan ablaufen. Über 75 % der zentralen Reaktionen beziehen ihre Energie nicht aus der Hydrolyse von Triphosphaten, stattdessen werden diese thermodynamisch durch Energie aus involvierten Kohlenstoffverbindungen angetrieben.

Für die vieldiskutierte Möglichkeit einer Energiebereitstellung durch die Hydrolyse von anorganischem Pyrophosphat (PP_i) finden sich keinerlei Hinweise im biosynthetischen Grundstoffwechsel. Im Gegenteil wird PP_i in zentralen Biosynthesewegen ausschließlich als Produkt generiert und kann somit nicht als Energielieferant gedient haben. Die Hydrolyse von Triphosphaten mit Produktion von PP_i ist überwiegend kinetisch gesteuert und fungiert als Mechanismus, um die Synthese der Grundbausteine des Lebens irreversibel in Richtung Biosynthese auszurichten, wohingegen die Möglichkeit der Hydrolyse mit Bildung von P_i einen thermodynamischen Effekt darstellt. Diese doppelte Funktion ist spezifisch für Triphosphate.

Schlussendlich beschäftigt sich diese Arbeit mit einem weiteren Schritt in der Evolution, indem der Stoffwechsel des letzten gemeinsamen Vorfahren aller Bakterien (LBCA) charakterisiert wird. Dazu wurden aus 1.089 vollständigen bakteriellen Genomen anaerober

Prokaryoten mittels phylogenetischer Analysen 146 Proteinfamilien identifiziert welche bereits in LBCA vorhanden gewesen sein könnten und einen konservierten Zentralmetabolismus bilden. Dieser benötigt nur neun weitere Gene um alle universellen Metaboliten (Aminosäuren, Basen, tRNA, Cofaktoren und Lipide) generieren zu können.

2 Abstract

Life is organized as a series of chemical reactions forming a metabolic network. The underlying reactions at the origin of metabolism before the emergence of life itself had to be energetically far from equilibrium to occur spontaneously, setting up the cornerstone for a highly efficient metabolism. Potentially, this process began 4 billion years ago with geochemical processes in the Earth's crust in serpentinizing systems, namely hydrothermal vents.

In the course of this thesis the autotrophic metabolism of the last universal common ancestor (LUCA) is characterized by conceiving a biosynthetic core network comprising more than 400 metabolic reactions necessary to synthesize amino acids, bases and cofactors from hydrogen (H_2), carbon dioxide (CO_2) and ammonia (NH_3), as well as hydrogen sulfide (H_2S) and phosphate. These components were potentially already available in hydrothermal vents on the early Earth. Thermodynamic investigations of these reactions under hydrothermal conditions reveal that 95–97 % are exergonic ($\Delta G \leq 0$ kJ·mol^{-1}), therefore energy releasing and running spontaneously. Over three-third of the core reactions obtain their energy not from hydrolysis of triphosphates but instead are thermodynamically driven by energy from involved carbon bonds.

The biosynthetic core network provides no indication for the highly debated possibility of energy supply via hydrolysis of inorganic pyrophosphate (PP_i). On the contrary, in the core network PP_i is generated as a product exclusively and therefore cannot serve as an energy supply. The hydrolysis of triphosphates producing PP_i is kinetically driven and provides a mechanism for rendering the synthesis of basic building blocks irreversible in the direction of biosynthesis whereas the hydrolysis producing P_i is induced by a thermodynamic effect. This dual function is specific to triphosphates.

In addition, this thesis illuminates a subsequent step in evolution by characterizing the metabolism of the last bacterial common ancestor (LBCA). Therefore, 1,089 complete bacterial genomes of anaerobic prokaryotes were analyzed phylogenetically, uncovering 146 potential protein families leading to LBCA. This conserved core metabolism needs only nine additional genes to be equipped to synthesize all universal metabolites (amino acids, bases, tRNA, cofactors and lipids).

3 Einleitung

> *„Wie entstanden jene ersten und einfachsten Lebewesen, aus denen sich alle übrigen, vollkommeneren Organismen allmählich entwickelten?“*
>
> Ernst Haeckel (1866a), Seite 168

Die Frage nach dem Ursprung des Lebens ist eine der spannendsten obgleich es schwierig zu sagen ist, ob es darauf jemals eine valide Antwort geben wird. Nichtsdestotrotz war es nicht nur Ernst Haeckel, der sich im 19. Jahrhundert eben diese Frage stellte (Haeckel, 1866a). Bereits Charles Darwin wagte in seinem berühmten Standardwerk *Über die Entstehung der Arten* (engl. Originalfassung von 1859: *On the origin of species by means of natural selection*) den Ursprung alles Lebendigen zu hinterfragen (Darwin, 1860). Er formulierte seine fundamentale Evolutionstheorie und das Konzept der natürlichen Selektion, welche die nachfolgenden Generationen nachhaltig prägten und den Grundstein für die Evolutionsforschung legten. Während Darwin zwar Antworten auf die Artbildung und Evolution von bereits existierendem Leben hatte, ließ er die Frage nach dem Ursprung des Lebens selbst offen. Er sah es nicht als seine Aufgabe, diese zu beantworten, und nahm an, dass „das Leben zuerst vom Schöpfer eingehaucht worden ist“ (Darwin, 1860, S. 488; Haeckel, 1866a, S. 169). Wenige Jahre später sah sich Ernst Haeckel in der Pflicht, sich der Frage nach dem eigentlichen Ursprung des Lebens zu widmen. Er formulierte eine Hypothese über die Selbstzeugung, welche er *Autogonie* nannte und im Rahmen derer sich strukturlose Urorganismen, die *Moneren*, aus anorganischem Material bildeten (Haeckel, 1866a, S. 179f.). Diese ersten Organismen bedeuteten dabei die letzten gemeinsamen Vorfahren aller darauffolgenden Lebensformen, welche sich im Späteren daraus differenzierten (vgl. Abbildung 1):

> *„Wir nehmen endlich an, dass alle jetzt lebenden Organismen-Formen und alle, welche jemals die Erde bewohnt haben, die Nachkommen einer geringen Anzahl verschiedener Moneren sind, und dass jede der Hauptgruppen der Organismen-Welt, welche wir unter dem Namen Stamm oder Phylon als eine zusammengehörige genealogische Einheit aufstellen, einer besonderen Moneren-Art ihre Entstehung verdankt.“*
>
> Ernst Haeckel (1866a), Seite 185

Der von Haeckel erwähnte Übergang zwischen der abiotischen und der biotischen Phase, heutzutage Abiogenese genannt, bei welcher Energie aus anorganischen chemischen Elementen zur Synthese von Biomasse genutzt wird, bildet zugleich den Ursprung des Lebens.

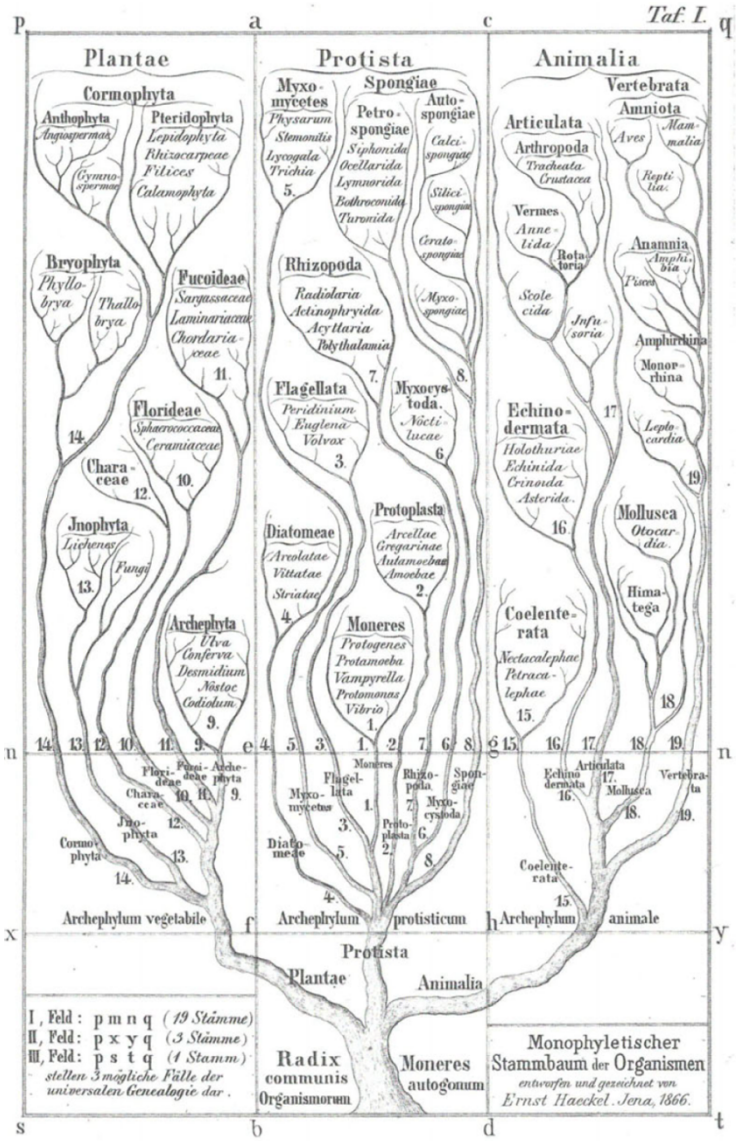

Abbildung 1 | Monophyletischer Stammbaum der Organismen nach Ernst Haeckel. Dargestellt sind die Verwandtschaftsbeziehungen zwischen den abgebildeten Taxa. Basal ist der Lebensbaum in die drei Reiche *Plantae*, *Protista* und *Animalia* eingeteilt. Das Phylum der *Moneres* befindet sich an der *Radix communis*—der gemeinsamen Wurzel—und bildet somit den Urvorfahren aller Nachkommen (Haeckel, 1866b).

3.1 Die Frage nach dem Ursprung des Lebens

3.1.1 Das Salz in der Ursuppe

Ein halbes Jahrhundert nach der Theorie von Ernst Haeckel waren es der Biochemiker Aleksandr I. Oparin und der Biologe John B. S. Haldane, welche unabhängig voneinander die Theorie einer Ursuppe beschrieben. Haldane nahm an, dass in der Atmosphäre der frühen Erde wenig beziehungsweise kein Sauerstoff vorhanden war, die UV-Strahlung der Sonne somit nicht von Ozon absorbiert werden konnte und deshalb ungehindert auf das Land und den Ozean traf (Haldane, 1929). Dieses UV-Licht diente als extrinsische Energiequelle in Reaktionen des Ozeanwassers mit in der Atmosphäre gelöstem Kohlenstoff und Ammoniak, bei welchen organische Komponenten entstanden. Diese akkumulierten, bis die Ozeane schließlich zu einer Art Ursuppe wurden. Die Komplexität dieser Moleküle stieg an, bis erste lebendige Partikel entstanden, welche sich von dieser Suppe ernährten (Haldane, 1929). Oparin war ebenfalls der Meinung, dass die ersten einfachen Organismen schrittweise über lange Zeit aus organischer Materie evolvierten (Oparin, 1938, S. 60). Für ihn schien es „absolut undenkbar, dass so komplexe Strukturen wie Organismen jemals spontan generiert werden konnten, direkt aus Kohlendioxid (CO_2), Wasser (H_2O), Sauerstoff (O_2), Stickstoff (N_2) und Mineralsalzen“ (Oparin, 1938, S. 62). Es müsse vorab eine organische Substanz entstanden sein, in welcher sich schlussendlich strukturierte Makromoleküle abgrenzten, die er als Koazervate bezeichnete und welche die Vorläufer der ersten Organismen darstellten (Oparin, 1938, S. 163). Die Oparin-Haldane-Theorie beschreibt ein heterotrophes Ursprungsszenario, da sich die beschriebenen Protoorganismen von den organischen Verbindungen in der Ursuppe ernährten.

Im Jahr 1952 veröffentlichte der Chemiker Harold C. Urey einen Artikel über die chemischen Bedingungen auf der frühen Erde, in welchem er sich analog zu Oparin gegen CO_2 in der Atmosphäre aussprach. Seiner Meinung nach bestand diese stattdessen aus Methan, Ammoniak, Wasser und Wasserstoff (Urey, 1952). Im darauffolgenden Jahr simulierte Ureys Doktorand Stanley L. Miller das Ursuppenszenario unter der vermuteten präbiotischen reduzierenden Atmosphäre mithilfe einer Apparatur, durch welche oben genannte Gase bei 70–80 °C zirkulierten und elektrische Entladungen in Form von Lichtbögen als Energiequelle fungierten (Miller, 1953). Im Rahmen des Versuches, bekannt als Miller-Urey-Experiment, entstanden geringe Mengen der Aminosäuren Glycin, α- und β-Alanin, Asparaginsäure und 2-Methylalanin neben anderen zu dem Zeitpunkt nicht identifizierten Verbindungen (Miller, 1953). In weiterführenden Experimenten konnten zusätzlich die gasförmigen Reaktionsprodukte Kohlenstoffmonoxid (CO), CO_2 und N_2, flüchtige Komponenten wie Ameisensäure,

Essigsäure, Propionsäure, Glycolsäure und Milchsäure, sowie weitere Aminosäuren wie Sarkosin und α-Aminobuttersäure detektiert werden (Miller, 1955). Somit war es erstmalig gelungen, eine Hypothese zum Ursprung des Lebens im Labor zu testen (McCollom, 2013) und die Ursuppentheorie wurde in vielen Fachkreisen bis dato zur plausibelsten Erklärung.

3.1.2 Die Entdeckung der Schwarzen Raucher

Zu Beginn der 1980er Jahre, kurz nach der Entdeckung geothermisch aktiver Tiefseehydrothermalquellen auf dem Meeresgrund im Galápagos Rift (Corliss *et al.*, 1979), gerieten heiße Quellen auf dem Meeresgrund erstmals in den Fokus der Suche nach dem Ursprung des Lebens (Corliss *et al.*, 1981). Die submarinen Quellen finden sich gehäuft an den Spreizungszonen tektonischer Platten wo heißes Magma durch Risse in der Erdkruste aufsteigen kann, auf durch Ozeanwasser gekühltes Gestein trifft und sich dadurch ein thermischer Gradient bildet, welcher Interaktionen des Gesteins mit Wasser bei über 300 °C energetisch vorantreibt. Das in den Ozean strömende Fluid ist reich an CO_2, Ammoniak (NH_3) und Wasserstoff (H_2). Das vorgefundene Milieu in den Schloten der sogenannten Schwarzen Raucher (engl. *Black Smokers*) wird als stark reduzierend beschrieben (Corliss *et al.*, 1981). Obwohl die extrem heißen Bedingungen lebensunfreundlich scheinen, besiedeln Populationen unterschiedlichster Tierarten wie Muscheln, Wasserschnecken und Röhrenwürmer diese Quellen (Corliss *et al.*, 1979). Die benötigte Nahrung wird durch schwefeloxidierende Bakterien bereitgestellt, welche in großer Zahl die Gesteinsrisse bevölkern und chemolithotroph vorhandenen Schwefelwasserstoff verstoffwechseln (Corliss *et al.*, 1979).

Aufgrund der in den Schwarzen Rauchern vorgefundenen Umgebungsbedingungen wurden heiße Hydrothermalquellen in der Tiefsee erstmals mit dem Ursprung und der Evolution des Lebens assoziiert (Corliss *et al.*, 1981). Laut der Theorie von Corliss *et al.* bildeten sich aus CO_2 mittels Hitze zuerst Komponenten wie Methan (CH_4), Ammoniak, Schwefelwasserstoff (H_2S), CO und Reduktionsmittel wie Cyanwasserstoff und Formaldehyd, später monomere Aminosäuren, darauffolgend Zucker, Purine und Pyrimidine und schließlich komplexe Polymere aus denen sodann die ersten Protozellen entstanden (Corliss *et al.*, 1981). Fossile Funde erster Lebewesen, dessen Alter auf ~3,8 Milliarden Jahre datiert wurde, decken sich mit Mikroorganismen in Hydrothermalfeldern moderner Ozeane (Baross und Hoffman, 1985; Arndt und Nisbet, 2012).

Die Entdeckung von Hydrothermalquellen und die idealen Voraussetzungen für die dortige Manifestierung von Leben sind handfeste Argumente gegen die bis dato vorherrschende und für valide geltende Ursuppentheorie von Oparin und Haldane. Die extreme Hitze in den Quellen und die damit einhergehende Instabilität von Aminosäuren bedeutet allerdings ein lebensfeindliches Attribut (Bernhardt *et al.*, 1984), welches Anhänger der Oparin-Haldane-Theorie als ein Argument gegen den Ursprung des Lebens in hydrothermalen Schloten werteten (Miller und Bada, 1988; Lazcano und Miller, 1996).

3.1.3 Die verlorene Stadt im Atlantis-Massiv

Der Geologe Michael J. Russell stellte sich gegen die Argumente bezüglich der Unvereinbarkeit der Abiogenese mit heißen Hydrothermalquellen (Russell *et al.*, 1988) nachdem ein Forschungsteam von dem Fund etwa 3,6 Milliarden Jahre alter fossiler Strukturen hydrothermaler Quellen in Irland berichtet hatte (Larter *et al.*, 1981). Die Bildung der eisenschwefelhaltigen Pyritröhren wurde nicht von Magma angetrieben, sondern von im Gestein gespeicherter Wärme, sie bildeten sich somit bei unter 150 °C (Boyce *et al.*, 1983). Moderne Hydrothermalquellen ähnlicher Konstitution im Ostpazifischen Rücken waren zwar bereits bekannt, diese waren allerdings viel heißer mit etwa 380 °C (Larter *et al.*, 1981). Russell *et al.* sahen in den moderat heißen Hydrothermalquellen eine Möglichkeit, dass durch dortige Bedingungen das Leben entstanden sein könnte und ließen die Diskussion über den Ursprung des Lebens in hydrothermalen Tiefseequellen wieder aufleben (Russell *et al.*, 1989). In einer Publikation von 1997 diskutierten sie die Bildung von Eisensulfidblasen, sogenannten „Probotryoiden“, welche sich durch hydrostatische Ausdehnung ausbildeten (Russell und Hall, 1997). Laut der Theorie traf heißes hinaufströmendes sulfidhaltiges, alkalisches Wasser auf den sauren, kühleren, eisenhaltigen Ozean. Im Rahmen einer Fällungsreaktion bildeten sich eisenschwefelhaltige (FeS) Membranen aus, welche als semipermeable Barriere zwischen beiden Fluiden fungierten, katalytische Fähigkeiten aufwiesen und an deren Oberfläche sich organische Komponenten aus im Ozean gelöstem Kohlenstoff bildeten. Der steigende osmotische Druck fungierte dabei als Triebkraft für die Ausdehnung der Probotryoide. Gemäß der Hypothese übernahmen mit der Zeit organische Moleküle die Rolle der FeS-Membranen und ein erster Protometabolismus formte sich (Russell und Hall, 1997).

Zu Beginn des neuen Jahrtausends fand ein Team um Deborah Kelley ein Hydrothermalfeld namens Lost City (engl. für *Verlorene Stadt*) auf dem Meeresgrund im

Atlantis-Massiv (Kelley *et al.*, 2001). Die aus den bis zu 60 Meter hohen Schloten austretende Flüssigkeit wurde als alkalisch (pH 9–11) und stark reduzierend beschrieben. Die Temperaturen dort sind mit 40–90 °C viel kühler verglichen mit denen in Schwarzen Rauchern. Des Weiteren weisen die Tiefseequellen Mikrokompartimente auf (Martin und Russell, 2003), welche an Probotryoide erinnern. Somit stimmen die Quellen von Lost City stark mit dem Szenario überein, welches Russell bereits in seiner Theorie beschrieben hatte (Russell und Hall, 1997). Das Gestein unterhalb von Lost City besteht großteils aus Serpentinit (Kelley *et al.*, 2005), welches bei der Hydratation der Gesteine Olivin und Pyroxen entsteht (Boyd *et al.*, 2020). Ein Nebenprodukt dieses geochemischen Vorgangs in der Erdkruste, auch Serpentinisierung genannt, ist neben Magnetit und Brucit auch molekularer Wasserstoff (H_2; Bach *et al.*, 2006; McCollom und Seewald, 2013). Durch den Serpentinisierungsprozess wird das Milieu alkalischer und es entstehen in den Hydrothermalquellen Ionengradienten (Schrenk *et al.*, 2013; Wimmer *et al.*, 2021b). Das durch die Gesteinsrisse aufsteigende Fluid vermischt sich mit dem Ozeanwasser sodass sich Temperatur-, Redox- sowie pH-Gefälle bilden, welche das elektrochemische Potential in der Umgebung steigern (Sleep *et al.*, 2011; Sousa *et al.*, 2013). Die H_2-produzierende Eigenschaft rückt hydrothermale Tiefseequellen wie Lost City ins Zentrum der Forschung zum Ursprung des Lebens, da Wasserstoff als ursprüngliches Reduktionsmittel für die Fixierung von CO_2 vermutet wird, welches potentiell den Weg zur Synthese von komplexeren organischen Molekülen ebnete (Martin und Russell 2003; Preiner *et al.*, 2020; Wimmer *et al.*, 2021c) und weil moderne chemolithoautotrophe Prokaryoten wie Acetogene und Methanogene H_2 als Elektronendonor für die Reduktion von CO_2 im Zuge der Synthese von Adenosintriphosphat (ATP) nutzen (Thauer *et al.*, 1977; Fuchs, 2011; Martin, 2020). Der Einfluss serpentinisierender Systeme auf den Ursprung metabolischer Komplexität wird in Kapitel 3.2.2 konkretisiert.

3.2 Die Entstehung des Stoffwechsels

Das Leben ist ein Netzwerk aus hintereinander geschalteten biochemischen Reaktionen. Dabei fungieren Reaktionsprodukte wiederum für andere Reaktionen als Ausgangsprodukte, aus welchen abermals neue Produkte gebildet werden. Dabei wird zwischen Auf- und Abbaureaktionen unterschieden. Mittels definierter Stoffwechselwege können Zellen effizient hintereinander geschaltete Reaktionen für den Aufbau von biochemischen Komponenten nutzen, als auch diese abbauen. Dieses Reaktionsnetzwerk ist gemeinhin als Stoffwechsel oder Metabolismus bekannt. Essentiell für den Zellaufbau ist die Synthese von monomeren Grundbausteinen, welche lebensnotwendig für den Organismus sind. Dazu zählen die 20 essentiellen Aminosäuren als Bestandteile der Proteine, die vier Purin- und Pyrimidinbasen, welche nötige Bausteine für Nukleinsäuremoleküle darstellen und Cofaktoren. Letztere können organische Moleküle oder Metallionen sein und als Reduktions- oder Oxidationsmittel fungieren oder katalytisch in einer Reaktion mitwirken (Berg *et al.*, 2013). Zur Zellsynthese brauchen Organismen primär das Element Kohlenstoff (C), welches etwa die Hälfte der Trockenmasse einer Zelle darstellt (Madigan *et al.*, 2020, S. 95). Des Weiteren bedarf es Wasserstoff (H), welcher zusammen mit Sauerstoff circa 25 % der Zellmasse darstellt (Madigan *et al.*, 2020, S. 95) und Stickstoff (N), dessen Quelle zumeist molekularer Stickstoff (N_2) oder Ammoniak (NH_3) ist und welcher 13 % der Zelltrockenmasse ausmacht (Madigan *et al.*, 2020, S. 95).

3.2.1 Ein System im energetischen Ungleichgewicht

> *„Dieser Grundsatz, auf welchem die ganze folgende Entwicklung beruht, lautet: es kann nie Wärme aus einem kälteren in einen wärmeren Körper übergehen, wenn nicht gleichzeitig eine andere damit zusammenhängende Änderung eintritt.“*
>
> Rudolf Clausius (1854), Seite 488

Der hier zitierte zweite Hauptsatz der Thermodynamik in der Fassung von Rudolf Clausius (Clausius, 1854) beschreibt, dass Wärme nicht aus einem Bereich mit niedriger Temperatur in einen Bereich höherer Temperatur übertragen werden kann, ohne dass Arbeit verrichtet wird. Übertragen auf biochemische Reaktionen können diese zwar spontan in Richtung eines

Energiegefälles ablaufen, entgegengesetzt dieses Gradienten wird hingegen zusätzliche Energie benötigt. Dieses Verhalten kann mittels der Änderung der freien Enthalpie, auch Gibbs-Energie ΔG genannt, beschrieben werden (Berg *et al.*, 2013). Bei spontanen Reaktionen entlang eines Energiegefälles ist diese negativ ($\Delta G < 0$ kJ·mol^{-1}), Energie wird freigesetzt und die Reaktion ist exergon. Im Umkehrschluss bedeutet dies, dass Stoffwechselreaktionen welche Energie benötigen niemals spontan ohne Energiezufuhr ablaufen können, da das Energiegefälle sonst nicht überwunden werden kann. Diese Reaktionen sind endergon ($\Delta G > 0$ kJ·mol^{-1}; Madigan *et al.*, 2020, S. 101). Ein thermodynamischer Gleichgewichtszustand ist erreicht, wenn kein thermodynamisches Potential mehr im System vorhanden ist ($\Delta G = 0$ kJ·mol^{-1}). Die Gibbs-Energie ist abhängig von verschiedenen physikalischen Größen wie der Enthalpie und Entropie, aber auch von der Temperatur, dem pH-Wert, den Stoffkonzentrationen und der Ionenstärke (Berg *et al.*, 2013).

Das Leben bedingt energetische Vorgänge jenseits des thermodynamischen Gleichgewichtszustandes. Ohne konstanten Energiefluss herrscht metabolischer Stillstand, welcher den Tod für lebende Systeme darstellt. Offene Systeme wie Organismen befinden sich daher im Fließgleichgewicht (engl. *steady state*; Richter, 1998, S. 10f.; Wimmer *et al.*, 2021c), es kommt soviel in das System wie es das System auch verlässt, was einen stetigen Reaktionsfluss gewährleistet (Richter, 1998, S. 10f.).

3.2.2 Aus Geochemie wird Biochemie

In chemischen Reaktionen, egal welcher Art, ist somit immer Energie involviert. Doch welche Energie fungierte als Zündkerze am Ursprung des metabolischen Zusammenspiels von Reaktionen? Es macht Sinn anzunehmen, dass der Stoffwechsel selbst Hinweise auf die Antwort zu dieser Frage versteckt hält (Wimmer *et al.*, 2021b). Verortet man die Wiege des Lebens in hydrothermalen Tiefseequellen wie Lost City, so muss ein weiterer Teil der Antwort auf die Frage nach dem Beginn des Lebens dort zu finden sein. Die Umweltbedingungen in H_2-produzierenden Hydrothermalquellen am heutigen Meeresboden decken sich mit dem, was man von den Bedingungen in urzeitlichen Tiefseequellen zu der Zeit des späten Hadaikums, dem Ende des ersten Äons der Erdgeschichte vor etwa 4,2–3,8 Milliarden Jahren, annimmt (Amend und McCollom, 2009; Arndt und Nisbet, 2012). Der Kohlenstoff aus der urzeitlichen Atmosphäre war im Ozean gelöst, das Milieu in serpentinisierenden Hydrothermalquellen reduzierend aufgrund der stetigen Synthese von Wasserstoff, das ausströmende Thermalwasser

alkalisch und warm (Amend und McCollom, 2009; Schrenk *et al.*, 2013). Die negative Änderung der Gibbs-Energie in der Nettosynthese von Aminosäuren unter Hydrothermalquellbedingungen legt nahe, dass sowohl der alkalische pH-Wert (circa 9), die heißen Temperaturen (circa 100 °C), als auch das reduzierende Milieu innerhalb der Quellen die Bildung von Biomasse energetisch begünstigen (Amend und Shock, 1998; Amend und McCollom, 2009; LaRowe und Amend, 2016).

Wie bereits in Kapitel 3.1.3 erwähnt, brauchte es Reduktionsmittel, um das im primordialen Ozean des Hadaikums gelöste CO_2 biologisch verfügbar zu machen (Martin und Russell, 2007). Der durch die serpentinisierenden geochemischen Prozesse in den alkalischen Hydrothermalquellen entstandene Wasserstoff diente hierfür wohlmöglich als Elektronendonor (Martin und Russell, 2003; Kelley *et al.*, 2005). Im Hydrothermalquellszenario liegt das Gleichgewicht der Reduktion auf der Seite der reduzierten Kohlenstoffkomponenten und die initiale Reaktion der CO_2-Fixierung ist thermodynamisch günstig (Shock, 1990). Energetisch angetrieben von dieser Reaktion bildeten sich unter anderem intermediäres Formiat, Acetat und daraus schließlich das Schlüsselintermediat Pyruvat (Preiner *et al.*, 2020). Im aufsteigenden Fluid in Lost City konnten bereits erhöhte Konzentrationen an Formiat abiogenen Ursprungs nachgewiesen werden (Lang *et al.*, 2010). In Abbildung 2 ist das Prinzip schematisch dargestellt und entspricht dem abgebildeten Prozess auf der linken Seite. Aufgrund einer kinetischen Hürde innerhalb der Reaktion wird allerdings ein Katalysator benötigt, um diese zu überwinden (Shock, 1990). Am Ursprung des Lebens fungierten anorganische Komponenten wie Metallionen (Muchowska *et al.*, 2020) und möglicherweise Minerale wie Magnetit (Fe_3O_4), Greigit (Fe_3S_4) und vor allem Awaruit (Ni_3Fe) als Antriebskraft (Preiner *et al.*, 2020). Im späteren Verlauf der Evolution wurde diese Aufgabe von Enzymen übernommen (Sousa und Martin, 2014).

Ein gegenwärtiges Analogon zu der anorganisch katalysierten Reaktion $CO_2 + H_2$ findet sich in modernen anaerob autotroph lebenden Acetogenen und Methanogenen (Ragsdale und Pierce, 2008; Sousa und Martin, 2014). Diese Organismen nutzen den reduktiven Acetyl-CoA-Weg zur Energiegewinnung und verstoffwechseln H_2 und CO_2 mittels enzymatischer Reaktionen (Fuchs, 2011; Martin, 2020). Die Produktion einfacher Intermediate des reduktiven Acetyl-CoA-Weges wie Acetat, Formiat, Pyruvat und Methan konnte bereits experimentell unter hydrothermalen Bedingungen einzig aus CO_2 und H_2 in Anwesenheit von H_2O mithilfe von als Katalysatoren fungierenden Übergangsmetallen synthetisiert werden (Barge *et al.*, 2019; Preiner *et al.*, 2020). Die in Kapitel 3.1.1 zitierte Aussage von Aleksandr Oparin, dass eine spontane Bildung biologischer Komponenten aus unter anderem CO_2 und H_2O nicht ohne

ein intermediäres Medium möglich sei (Oparin, 1938, S. 62) scheint dadurch tendenziell widerlegt.

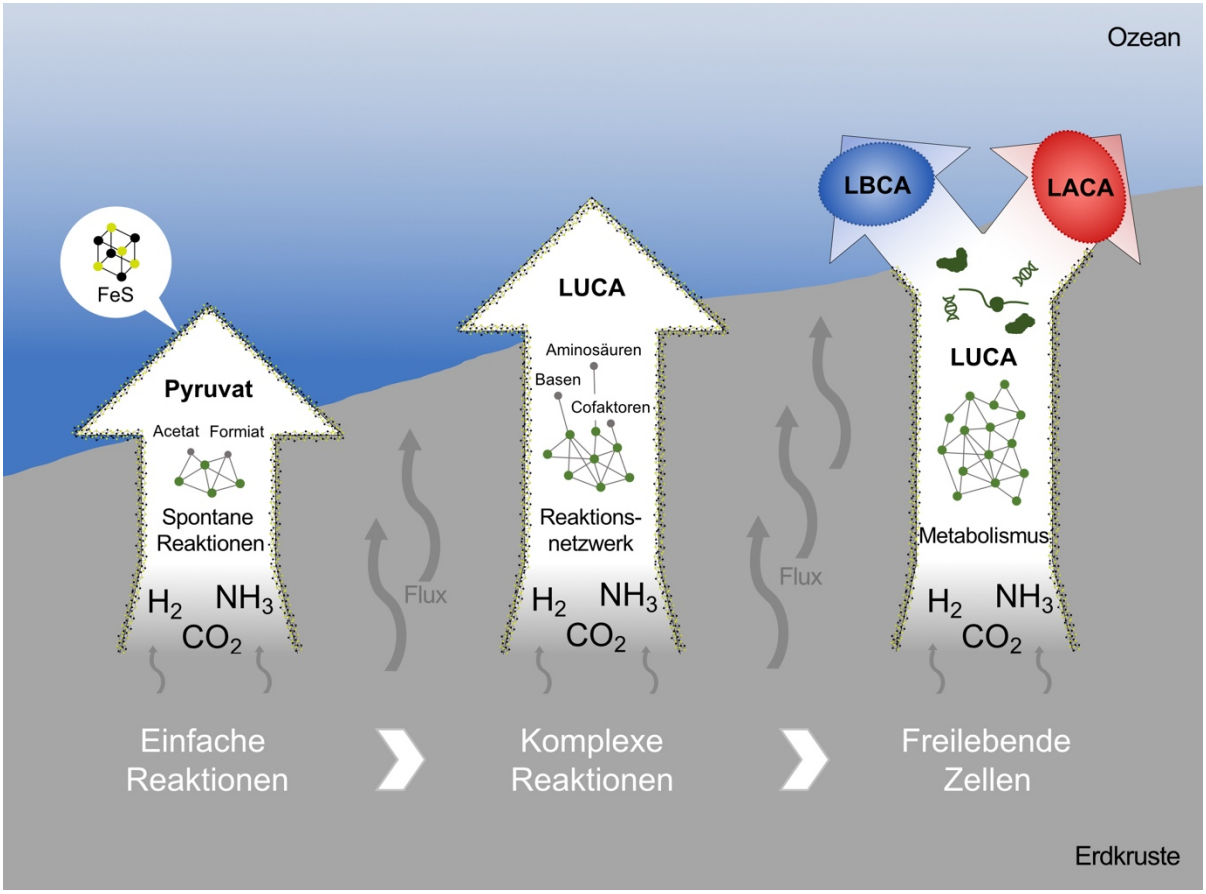

Abbildung 2 | Schematische Darstellung der Hypothese über die schrittweise Entwicklung freilebender Zellen in H_2-produzierenden Hydrothermalquellen. Die Ausgangskomponenten für den in Hydrothermalquellen ablaufenden Prozess sind umweltverfügbarer Wasserstoff (H_2), Kohlenstoffdioxid (CO_2) und Ammoniak (NH_3) im aufsteigenden Hydrothermalfluid. Innerhalb der Hydrothermalquellkompartimente, welche durch Übergangsmetall-Mineralien begrenzt sind, bilden sich anfangs einfache spontane Reaktionen aus welchen unter anderem Acetat, Formiat und Pyruvat hervorgehen. Im nächsten Schritt werden die Reaktionen komplexer, ein Netzwerk aus überwiegend energetisch günstigen Reaktionen entsteht und die Grundbausteine zur Zellsynthese—Aminosäuren, Cofaktoren und Basen—bilden sich aus. Aus diesem Grundgerüst bildet sich nachfolgend der Stoffwechsel des letzten gemeinsamen Urvorfahren LUCA, welcher allerdings noch räumlich an seine Umwelt gebunden ist. Schließlich teilt sich die gemeinsame Stammeslinie in Bakterien (LBCA) und Archaeen (LACA) auf.

Aus den ersten auf abiotischem Wege synthetisierten organischen Molekülen entstanden, der Theorie nach, im weiteren Verlauf durch ebenso energetisch getriebene spontane Reaktionen sukzessive Aminosäuren, Basen und Cofaktoren (Kaufmann, 2009). Diese monomeren Moleküle stellten die Grundbausteine des Lebens dar und werden zum Aufbau von Zellen benötigt (siehe Abbildung 2 Mitte). Selbsterhaltende autokatalytische Netzwerke (engl. *reflexively autocatalytic food-generated set*; RAF) bilden sich, in denen Moleküle zur Unterstützung ihrer eigenen Synthese benötigt werden (Xavier *et al.*, 2020). Fortlaufend beginnen Monomere zu stabileren Makromolekülen zu polymerisieren, Proteine und der genetische Code entstehen. Vermutlich musste all dies in irgendeiner Form vorhanden sein, damit die erste Zelle—der Urvorfahr aller Lebewesen—die Bühne der Evolution betreten konnte.

3.2.3 Der Urahn allen Lebens

Das bereits von Ernst Haeckel gesuchte Bindeglied zwischen der abiotischen und der biotischen Phase unseres Planeten ist äquivalent zum letzten gemeinsamen Urvorfahren aller Lebewesen (engl. *last universal common ancestor*, kurz LUCA; Weiss *et al.*, 2016). Analysen der Zusammensetzung von δ^{13}C-Isotopen in Sedimentgesteinsproben aus dem Isua-Gürtel nahe Grönland datieren das Alter erster Spuren mikrobiellen Lebens auf der Erde auf etwa 3,8 Milliarden Jahre (Rosing, 1999; Arndt und Nisbet, 2012), was dem Beginn des zweiten Äons der Erdgeschichte, dem Archaikum, entspricht (Amend und McCollom, 2009). LUCA ist ein hypothetisches Konstrukt und gilt als halb-lebendig da es noch keine eigenständige Zelle war. Phylogenetische Untersuchungen legen eine Verbindung LUCAs zu Tiefseehydrothermalquellen nahe (Weiss *et al.*, 2016). Sein Stoffwechsel lief vermutlich innerhalb von Hydrothermalquellkompartimenten ab, er war an seine Umwelt gebunden (Martin und Russell, 2003). Durch die Analyse von rekonstruierten Stammbäumen aus 5.655 vollständig sequenzierten prokaryotischen Referenzgenomen konnten 355 LUCA-Gene identifiziert werden (Weiss *et al.*, 2016). Diese legen nahe, dass LUCA anaerob autotroph lebte. Es fanden sich Gene des reduktiven Acetyl-CoA-Weges, welche auf Energiegewinnung beginnend mit Wasserstoff und Kohlenstoffdioxid hinweisen wie es in Acetogenen und Methanogenen der Fall ist (siehe Kapitel 3.2.2; Weiss *et al.*, 2016; Martin, 2020). In den Topologien der rekonstruierten Stammbäume standen acetogene Clostridia und methanogene Archaeen an der Wurzel. Des Weiteren befinden sich ribosomale Gene und Aminoacyl-tRNA-Synthetasen unter den

identifizierten Genen, beides wichtig für die Informationsverarbeitung und Umsetzung des genetischen Codes. Dies weist darauf hin, dass das Auftreten dessen eng mit LUCA verknüpft ist (Weiss *et al.*, 2016).

Einer der ersten evolutionären Schritte nach LUCA war die Entwicklung zweier distinkter Domänen, wie auf der rechten Seite in Abbildung 2 dargestellt. Es kam zu einer Aufspaltung der bis dato gemeinsamen Stammeslinie in Bakterien und Archaeen. Der hypothetische bakterielle Urvorfahr (kurz LBCA, engl. *last bacterial common ancestor*) ist hier, analog zu LUCA (Weiss *et al.*, 2016), der letzte gemeinsame Vorfahr aller Bakterien. Groben Schätzungen zufolge soll er vor mindestens 3,4 Milliarden Jahren (Battistuzzi *et al.*, 2004; Busch *et al.*, 2016) existiert haben. Da sich die Sauerstoffkonzentration in der Atmosphäre erst durch die Große Sauerstoffkatastrophe (GOE, engl. *great oxygenation event*) vor ~2,5–2,3 Milliarden Jahren drastisch erhöht hatte (Holland, 2002; Barth *et al.*, 2018), wird LBCA wie LUCA anaerob gelebt haben (Martin und Sousa, 2016).

4 Zielsetzung

Zwischen dem erstmaligen Auftreten von flüssigem Wasser auf der Erde vor ungefähr 4,3 Milliarden Jahren (Mojzsis *et al.*, 2001) und den ersten Lebewesen vor 3,8–3,6 Milliarden Jahren (Arndt und Nisbet, 2012; Weiss *et al.*, 2016; Tashiro *et al.*, 2017) existiert eine Zeitspanne, in welcher der Übergang von der abiotischen zur biotischen Phase, die Abiogenese, stattgefunden haben muss. Für die Erforschung der frühen Evolution ist dabei von Interesse, wie die ersten Schritte in der Entwicklung von Leben hin zum letzten gemeinsamen Urvorfahren aller Lebewesen (kurz LUCA; Weiss *et al.*, 2016) ausgesehen haben. Zahlreiche Studien beschreiben geochemische Vorgänge in serpentinisierenden Hydrothermalquellen am Meeresgrund als möglichen Ausgangspunkt für die Ausbildung eines Protometabolismus (Baross und Hoffman, 1985; Martin und Russell, 2007; Preiner *et al.*, 2020). Wenn der Urstoffwechsel so entstanden ist, sollten seine Spuren in der Biochemie moderner Zellen zu finden sein.

Fragen, die sich im Hinblick auf den Übergang von geochemischen Reaktionen auf frühe metabolische Netzwerke stellen, sind vor allem: i) Was waren die ersten organischen Moleküle, ii) wie formten sich diese aus den vorhandenen Komponenten und iii) woher kam die dafür benötigte Energie? Diese Fragen betreffen den Ursprung des Stoffwechsels als eine Voraussetzung für die Entstehung der ersten Lebewesen (Lipmann, 1965; Wächtershäuser, 1988).

Diese Arbeit verfolgt das Ziel, die Lücke im Verständnis der Bildung von frühen metabolischen Strukturen aus einfachen chemischen Reaktionen am Ursprung des Lebens zu schließen und damit eine mögliche Entwicklung des Metabolismus selbst zu beschreiben. Mittels Analyse universeller metabolischer Reaktionen, welche im heutigen Stoffwechsel die Synthese von Aminosäuren, Basen und Cofaktoren gewährleisten, wird ein biosynthetischer Grundstoffwechsel entworfen, der aufgrund seiner Konservierung auf LUCA zurückzuführen ist. Durch Kombination dieses metabolischen Netzwerkes mit thermodynamischen Parametern unter verschiedensten Umweltbedingungen wird die energetische Realisierbarkeit und die Quelle der benötigten Energie für Biosynthese und Wachstum untersucht. Zudem wird ein weiterer wichtiger Aspekt der frühen Evolution nach dem Auftreten von LUCA, der Ursprung des letzten gemeinsamen Vorfahren aller Bakterien (LBCA), im Hinblick auf dessen metabolisches Repertoire und seine Physiologie untersucht.

5 Publikationen

I The autotrophic core: An ancient network of 404 reactions converts H_2, CO_2, and NH_3 into amino acids, bases, and cofactors

Jessica L. E. Wimmer, Andrey d. N. Vieira, Joana C. Xavier, Karl Kleinermanns, William F. Martin, Martina Preiner.

Institut für Molekulare Evolution, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Deutschland.

Dieser Artikel wurde am 23.02.2021 in *Microorganisms* Ausgabe 9 veröffentlicht.

Beitrag von Jessica L. E. Wimmer (Erstautor und Co-Korrespondenz):
Ich war an der Konzeptualisierung und Methodologie beteiligt. Der präsentierte Datensatz, bestehend aus 404 Stoffwechselreaktionen im KEGG-Format, wurde zu einem Großteil von mir zusammengestellt. Die dazu erforderliche Recherche wurde von mir durchgeführt. Die Analysen wurden von mir umgesetzt und das metabolische Netzwerk von mir visualisiert. Des Weiteren habe ich die Überarbeitung des Manuskriptes durchgeführt.

microorganisms

MDPI

Article

The Autotrophic Core: An Ancient Network of 404 Reactions Converts H_2, CO_2, and NH_3 into Amino Acids, Bases, and Cofactors

Jessica L. E. Wimmer [1,*], Andrey do Nascimento Vieira [1], Joana C. Xavier [1], Karl Kleinermanns [2], William F. Martin [1] and Martina Preiner [1]

[1] Department of Biology, Institute for Molecular Evolution, Heinrich-Heine-University of Düsseldorf, 40225 Düsseldorf, Germany; nascima@uni-duesseldorf.de (A.d.N.V.); xavier@hhu.de (J.C.X.); bill@hhu.de (W.F.M.); preiner@hhu.de (M.P.)

[2] Department of Chemistry, Institute for Physical Chemistry, Heinrich-Heine-University of Düsseldorf, 40225 Düsseldorf, Germany; kkleinermanns@yahoo.de

* Correspondence: jessica.wimmer@hhu.de

Abstract: The metabolism of cells contains evidence reflecting the process by which they arose. Here, we have identified the ancient core of autotrophic metabolism encompassing 404 reactions that comprise the reaction network from H_2, CO_2, and ammonia (NH_3) to amino acids, nucleic acid monomers, and the 19 cofactors required for their synthesis. Water is the most common reactant in the autotrophic core, indicating that the core arose in an aqueous environment. Seventy-seven core reactions involve the hydrolysis of high-energy phosphate bonds, furthermore suggesting the presence of a non-enzymatic and highly exergonic chemical reaction capable of continuously synthesizing activated phosphate bonds. CO_2 is the most common carbon-containing compound in the core. An abundance of NADH and NADPH-dependent redox reactions in the autotrophic core, the central role of CO_2, and the circumstance that the core's main products are far more reduced than CO_2 indicate that the core arose in a highly reducing environment. The chemical reactions of the autotrophic core suggest that it arose from H_2, inorganic carbon, and NH_3 in an aqueous environment marked by highly reducing and continuously far from equilibrium conditions. Such conditions are very similar to those found in serpentinizing hydrothermal systems.

Keywords: chemolithoautotrophy; early metabolism; serpentinizing systems; hydrothermal vents; origins of life

Citation: Wimmer, J.L.E.; Vieira, A.d.N.; Xavier, J.C.; Kleinermanns, K.; Martin, W.F.; Preiner, M. The Autotrophic Core: An Ancient Network of 404 Reactions Converts H_2, CO_2, and NH_3 into Amino Acids, Bases, and Cofactors. *Microorganisms* **2021**, *9*, 458. https://doi.org/10.3390/microorganisms9020458

Academic Editor: Françoise Bringel

Received: 28 January 2021
Accepted: 19 February 2021
Published: 23 February 2021

1. Introduction

Biologists have traditionally linked the topic of C1 metabolism to thoughts about life's origins. Haeckel (1902) posited that the first cells probably lived from CO_2 [1], perhaps in a manner similar to organisms discovered by Winogradsky (1888), growing from CO_2 with the help of electrons from inorganic donors [2]. The chemolithoautotrophic lifestyle—converting inorganic carbon into cell mass with inorganic electron donors using chemical energy instead of light—is common among modern microbes that inhabit environments similar to those on the early Earth [3]. Although microbiologists have traditionally favored the view that the first cells were anaerobic autotrophs [4–7], the electric spark experiments of Miller shifted the focus in origins literature from microbiology to nucleic acid chemistry [8]. The facile synthesis of nucleobases from cyanide condensations [9], Spiegelman's in vitro RNA replication experiments using Qβ replicase [10], and the demonstration that RNA has catalytic activity [11] led to the concept of an RNA world [12] in which RNA molecules became synthesized by abiotic chemical means and then competed with one another for resources (activated ribonucleoside triphosphate monomers) via replication [13]. This line of thinking diverted attention away from the more challenging problem concerning the origin of living cells and toward the more tractable problem concerning the origin of nucleic

acid monomers [14,15]. However, for critics, the allure of the RNA world concept has caveats, as RNA synthesis, regardless how good, does not alone solve the problem of how living cells arose [16]; *Escherichia coli* is clearly alive; RNA is clearly not.

1.1. Metabolism vs. Genetics?

As an alternative to the concept of an RNA world, Wächtershäuser's theory of surface metabolism rekindled the idea of a chemolithoautotrophic start of life and brought energy into the origins debate, positing that the exergonic conversion of iron-sulfur (FeS) minerals to pyrite provided the thermodynamic drive to fuel the origin of biochemical pathways and the first autotrophic cells [17]. Wächtershäuser's theory ignited a "genetics first vs. metabolism first" debate [18–21] that Lipmann had presciently perceived decades in advance [22]. While the theory of surface metabolism interfaced well with catalytic mechanisms in autotrophic cells, it did not interface well with energy conservation in the currency of high-energy phosphate bonds [23] nor did it offer inroads to accounting for the fundamental property of life that the living cell maintains itself in a thermodynamic state that is far from equilibrium. The discovery of deep-sea hydrothermal vents [24] and alkaline hydrothermal vents of serpentinizing systems [25] impacted the origins issue in that they harbor geochemically continuous far from equilibrium conditions that help to define the living state of cells [5].

The idea that life started from CO_2 is appealing, but it only solves half the problem because both for life and for organic synthesis, CO_2 requires a reductant. This is why serpentinizing hydrothermal systems are so interesting in the origins context. Serpentinization synthesizes H_2, the main energy and electron source of chemolithoautotrophs, from protons and electrons within the Earth's crust through the reduction of H_2O by Fe(II) minerals. The amount of H_2 generated by serpentinization is substantial, on the order of 16 mmol/kg in some modern systems [26], which is orders of magnitude more H_2 than modern chemolithoautotrophs require for growth [27]. The synthesis of H_2 during serpentinization is a continuous process that has been going on since there was water on Earth [28]. The further researchers explored the properties of serpentinizing systems, the more similarities they revealed to the life process [29], with compartmentation, energy harnessing, catalysis, and the chemical reactions of C1 compounds at vents converging on processes that comprise the core of carbon and energy metabolism in primitive autotrophic cells [30].

From a biological perspective, the genetics first vs. metabolism first debate misses the point because neither by itself is sufficient for life. Countering the genetics first view, cells are made of far more than RNA. Cells consist by dry weight of about 50% protein and 20% RNA, with DNA, lipids, cell wall, reserves, and metabolites making up the rest [31]. Most of the RNA resides in the ribosome, which synthesizes proteins in a process that consumes about 75% of the cell's ATP investment in biosynthetic processes [32], whereby a large portion of the ATP that a cell synthesizes is not used for the synthesis of cell mass—it goes to what is called ATP spilling and maintenance energy [33,34].

Countering the metabolism first view, a handful of small molecules reacting with each other do not qualify as metabolism. A cell is a very complicated chemical system composed of more than 1000 individual partial reactions that harness energy and synthesize building blocks as well as polymers. A decisive property of metabolism is redox balance: the number of electrons that enter the cell in substrates has to be equal to the number of electrons that leave the cell in waste products plus those that remain sequestered in compounds of cell mass; otherwise, metabolism and life come to a halt [35]. Although most reactions in cells are catalyzed by enzymes, enzymes do not perform feats of magic; they just accelerate reactions that tend to occur anyway. Many core metabolic reactions of cells readily take place without enzymes [36–38]. The sum of the chemical reactions in the cell (metabolism) runs both the synthesis and the operation of the cellular machinery that produces progeny, harboring a new copy of instructions in DNA, hence heredity, which over generations forms the process called genetics.

Yet the main thing that cells do is neither genetics nor metabolism but energy harnessing because without energy, neither metabolism nor genetics can take place. Metabolism and genetics are merely manifestations of the actions of sustained sets of exergonic chemical reactions over generations. There is a third option in the genetics first vs. metabolism first debate, namely "energy first", because it is hands-down obvious that energy has to come first [39], for without favorable energetics and energy release, neither genetics, metabolism, nor anything at all will take place, so says the 2nd law of thermodynamics. Cells themselves underscore that view, because the amount of ATP that a cell synthesizes always exceeds the amount of energy required for the synthesis of new cells during growth, often by about a factor of 3 [40] (the converse would violate the 2nd law), underscoring the point that life is an energy-releasing process. For a cell, staying alive means staying far from equilibrium, which it achieves by merely running the exergonic reactions that synthesize ATP: maintenance energy or ATP spilling [33,34]. In low-energy environments, where survival becomes more important than growth [41], maintenance energy becomes the main process of life.

1.2. Autotrophic Origins and Energy First Link C1 Metabolism to Vents

Is there an origins option that starts with energy first? Yes, and it is seated firmly in C1 metabolism and autotrophic origins. In 2021, serpentinizing systems have gone a long way to closing the gap between CO_2 and cells. Convergent lines of evidence indicate that reactions of C1 compounds were not only the source of carbon for the first cells but also the source of energy at the origin of the first metabolic reactions. This is because in the reaction of H_2 with CO_2, the equilibrium lies on the side of the simple reduced carbon compounds that comprise the backbone of carbon metabolism in organisms that use the acetyl Coenzyme A (CoA) pathway of CO_2 fixation—formate, acetate, and pyruvate. The synthesis of these acids from H_2 and CO_2 is exergonic under standard conditions [39], in cells that use the acetyl CoA pathway [6] and under conditions of simulated hydrothermal vents [30]. Hydrothermal vents are generally of interest in modern theories for origins [3] because they present continuously far from equilibrium conditions, with geochemically catalyzed redox reactions and gradients that could be tapped by the first cells for energy harnessing [29].

In hydrothermal systems, both modern and on the early Earth, the key to redox reactions, catalyst synthesis, and the formation of ion gradients, is molecular hydrogen, H_2, which is generated by the spontaneous geochemical process of serpentinization [28,42,43]. During serpentinization, mineral catalysts awaruite (Ni_3Fe) and magnetite (Fe_3O_4) are formed in situ in serpentinizing hydrothermal vents [44]. These minerals catalyze the synthesis of formate, acetate, and pyruvate as well as methane [30] in the laboratory from H_2 and CO_2 in the presence of only water and the mineral catalyst overnight at 100 °C and only 24 bar. It is likely, but not directly demonstrated, that hydrothermally formed awaruite and magnetite catalyze the synthesis of formate and methane found in the effluent of modern serpentinizing systems [45–48]. Serpentinization also renders the effluent of hydrothermal systems alkaline [48], generating the ion gradients that form at hydrothermal vents.

The synthesis of simple organics from C1 precursors in hydrothermal systems is the only known geochemical process that follows the same chemical route as a modern core pathway of carbon and energy metabolism [30,49]. In addition, modern organisms that use the acetyl CoA pathway for carbon and energy metabolism, acetogens and methanogens, exhibit a physiology that, among known life forms, is most similar to that inferred from genomic reconstructions for the last universal common ancestor of all cells, LUCA [50]. This implicates acetogens and methanogens that lack cytochromes and quinones as very primitive microbial lineages, in line with early predictions from physiology [4] and with predictions based on similarities between geochemical and biochemical reactions [36]. It is also consistent with the identification of overlapping autocatalytic networks in the

metabolism of acetogens and methanogens that implicate a role for small molecule reaction systems prior to the advent of both protein and RNA [51].

2. Methods

2.1. Reaction Data Collection

Metabolic reactions were gathered and curated from the Kyoto Encyclopedia of Genes and Genomes (KEGG) reaction database [52] (version December 2020) manually. Synthesis pathways for 46 target compounds (Table S1) were obtained and curated by hand with the help of KEGG pathways [53] and KEGG modules. The 46 target compounds comprise 20 amino acids, four ribonucleoside triphosphates, four deoxyribonucleoside triphosphates, and 18 cofactors shown in Figure 1. KEGG lacked biosynthetic pathway information on iron–sulfur clusters, so these were not included. Although depicted in Figure 1, polymers and the genetic code are also not part of the target set. The reductive acetyl CoA pathway as well as the reverse tricarboxylic acid cycle (rTCA) cycle were added to the reaction set, covering the basal CO_2 fixation along with the gluconeogenesis and pentose phosphate cycle, allowing for the synthesis of key intermediates needed to produce amino acids, nucleic acids, and cofactors from α-ketoacids, sugars, and aldehydes. Nitrogen fixation pathways were not included, since NH_3 gets incorporated via amino acid synthesis. If a pathway was unavailable in KEGG, it was manually reconstructed based on KEGG pathway maps. The collection of reactions unfolds in a short example: For methanofuran biosynthesis, KEGG module M00935 was used to add reactions R10935, R11038, R11039, R00736, R10902, and R11040. The very last step of producing methanofuran is missing in the module. This reaction from APMF-Glu is depicted in pathway map00680; thus, it was added manually.

In all pathways collected, oxygen-dependent reactions were either replaced with an anaerobic alternative if possible or omitted if not. This was the case for the synthesis of dimethylbenzimidazole, which is a precursor for cobamide. Although an anaerobic synthesis pathway for this precursor is known, starting from 5-aminoimidazole ribotide (short AIR) [54], several other intermediaries are not implemented in KEGG yet. Neither was there an anaerobic alternative for the production of 2-phospholactate as a precursor in the F_{420} synthesis pathway available. For both precursors, dimethylbenzimidazole and 2-phospholactate, as well as reduced ferredoxin (involved in the reductive acetyl CoA pathway) and reduced flavodoxin (in the rTCA cycle), we assume them to be producible in an unknown way in early metabolism. Assuming the reactions in question arose before the genetic code, the according proteins were presumably replaced by an alternative at that early period. Three reactions were constructed manually, because they appear as a dashed line in KEGG pathways with no corresponding reaction identification number. The reactions named *RMAN1-3* are presumed to be incomplete, since only the key compounds were listed. Two reactions are affected within tetrahydromethanopterin synthesis and the very last step was within methanofuran synthesis. Involved chemical elements such as molybdenum, sulfur donors, cobalt, and nickel were assumed to be present in the environment. During curation of the final reaction set, redundant reactions occurring in multiple syntheses were reduced to a single occurrence, such as the reaction chorismate <=> prephenate that occurs in both tyrosine and phenylalanine syntheses.

For the detection of autocatalytic cycles within cofactor biosynthetic pathways, catalysis rules (indicating which cofactors are used as catalysts in each reaction) were gathered from [51] (Supplemental Dataset S1A) [51]. Autocatalysis is assumed if a target is needed as a catalyst within its own biosynthetic pathway.

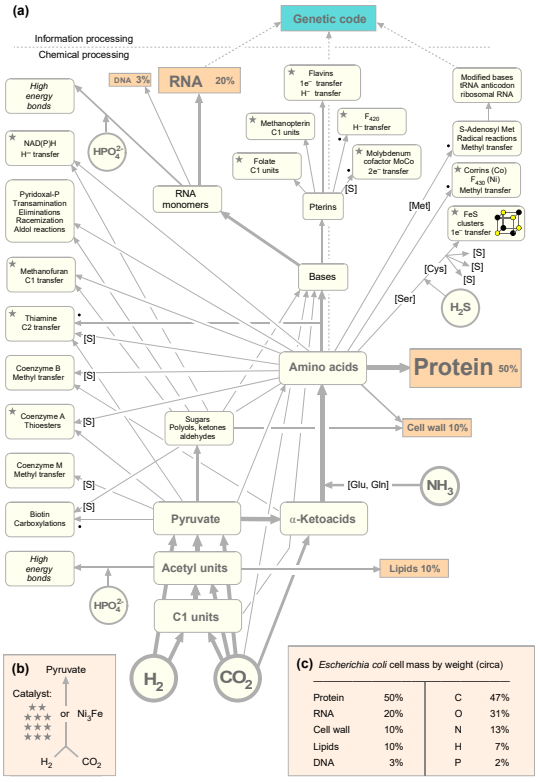

Figure 1. (**a**) A general map of core metabolism. The arrows in the map do not cover every atom in every cofactor, amino acid, or base, showing main mass contributions instead. A dot indicates that radical S-adenosyl methionine (SAM) enzymes are involved in the biosynthetic pathway leading to the product. [S] indicates that sulfur is incorporated in the biosynthetic pathway. (**b**) Cofactors indicated by a star are required in the pathway from H_2 and CO_2 to pyruvate in either acetogens or methanogens or both. (**c**) The composition of cells in terms of its main components and elemental contributions to dry weight (from [31]).

2.2. Visualization of the Autotrophic Core Network

An undirected metabolic network showing the autotrophic core was generated in simple interaction format (sif) using a custom Python script. The resulting network consists of the given 404 metabolic reactions and 380 involved compounds. The bipartite network was visualized using CytoScape [55] v. 3.8.0. One partition class corresponds to reaction nodes (diamonds), the other one corresponds to compound nodes (circle-shaped). The latter were sized according to their node degree. Target compounds were colored in blue, whereas reaction nodes are depicted smaller and in gray.

2.3. Different Core Reaction Sets Based on Distinct Identification Approaches

Two additional reaction datasets were used to determine their intersection with the 404 reactions of the autotrophic core. The LUCA set, containing 355 genes, was identified via the phylogenetic approach [50], translating to 163 metabolic reactions and the ancient 'reflexively autocatalytic food-generated' (RAF) set with 172 reactions (from [51] Figure 4). The intersection between the autotrophic core, LUCA, and the ancient RAF was determined by examining which reactions overlap in the respective analyzed datasets. In addition, the overlap of reactions between all three datasets was determined. The intersection for each comparison (Figure S1) is available in Supplemental Table S3A and the initial reaction lists are in Table S3B.

2.4. Statistical Analysis

A contingency table for each highly connected compound (Table 1) was built, comparing the compound frequency in the autotrophic core with the frequency in the global prokaryote anaerobic network consisting of 5994 reactions (from [51] S1A). A significant enrichment of compound frequency in the autotrophic core compared to the global prokaryotic set was observed for *p*-values smaller than 0.05. One-tailed Fisher tests were performed using the package scipy.stats in Python 3.6 (Table S4).

Table 1. Highly connected nodes.

Compound	Frequency
H_2O	125
ATP	77
H^+	76
P_i	66
ADP	55
CO_2	49
Glutamate	44
PP_i	37
NAD^+	37
$NADP^+$	35
NADPH	34
NADH	33
2-Oxoglutarate	24
Pyruvate	22
NH_3	21

3. Results

3.1. The Autotrophic Core of Biosynthesis Requires 19 Cofactors

For the purpose of this paper, let us assume for the sake of argument that life really did start from exergonic reactions of H_2 and CO_2 along the acetyl CoA pathway. Why do we assume the acetyl CoA pathway as the starting point of CO_2 fixation? Among the six known pathways of CO_2 fixation [6,56,57], it is the only one that occurs in both bacteria and archaea, the only one that traces in part to LUCA [50], and it is the only one that has been shown in the laboratory to produce acetate and pyruvate from H_2 and CO_2 without enzymes, using only hydrothermal minerals as catalysts [30]. In that sense, it is the obvious choice as the starting point for metabolic evolution investigations based upon current laboratory evidence. The horseshoe (incomplete) rTCA cycle follows in Figure 1 because it is the pathway that autotrophs using the acetyl CoA pathway employ to generate C4 and C5 precursors for amino acid and other syntheses [6,36,58]. Although the incomplete horseshoe the rTCA cycle occurs in bacteria and archaea, it is fed by the acetyl CoA pathway, which is the only pathway of CO_2 fixation that is known to occur in bacteria and archaea. The other five are known to operate in only one domain [6,56]. The rTCA cycle is also an ancient pathway [59,60], and most of its reactions also operate in the laboratory in the absence of enzymes provided that pyruvate and glyoxylate are supplied as starting material, but the non-enzymatic reaction sequence operates in the oxidative direction, that is, in the absence of H_2 and CO_2 [37]. The acetyl CoA pathway, the rTCA cycle, and the dicarboxylate/4-hydroxybutyrate cycle, which is a derivative of the rTCA cycle and occurs only in archaea, employ O_2 sensitive enzymes, an ancient trait [6]. The other three CO_2 fixation pathways have no O_2-sensitive enzymes and are typically found in aerobes, occur in only one domain each, and they arose more recently in evolution, using enzymes co-opted from preexisting pathways [6].

We also assume that the first living things on the path to cells we recognize today required the universal amino acids and bases of life, the modern synthesis of which requires in turn cofactors as catalysts. We asked: How big, exactly, is the set of reactions required for the synthesis of the building blocks of cells and the cofactors needed to make them? This

gives us an impression of how challenging it would be to generate the main compounds of life at origins, with or without enzymes. We started by sketching out Figure 1a, in which the main pathways of biosynthesis in anaerobes and the amounts of main biosynthetic end products are summarized. Cofactors are usually not present in amounts that would contribute appreciably to cell mass, but they are required as catalysts. Along the acetyl CoA pathway, there are differences in the methanogenic and acetogenic versions [61].

If we look at the cofactors required to get from H_2 and CO_2 to pyruvate in acetogens and methanogens [6,62], we find that methanofuran, NAD(P)H, corrins, coenzyme A, thiamine, flavins, F_{420}, three pterins—folate, methanopterin, and the molybdenum cofactor MoCo—as well as FeS clusters, a prosthetic group of proteins that we count as a cofactor here, are required. That is a substantial cofactor requirement, not to mention the enzymes that hold those cofactors in place for function in the pathway. The mere requirement for those 11 complicated organic cofactors would appear to make the proposition that C1 metabolism from CO_2 to pyruvate could be the first pathway [36] seem downright absurd, were it not for the recent observation that a bit of metal, awaruite (Ni_3Fe), or a piece of iron oxide magnetite (Fe_3O_4), can also catalyze the synthesis of pyruvate from H_2 and CO_2 [30] overnight at 100 °C in water. As a set of chemical reactions, the acetyl CoA pathway is older than the genes that encode its enzymes [58], and it is also older than the cofactors required by those enzymes.

By the foregoing count, it takes 11 cofactors to synthesize pyruvate in the modern pathway, whereby we have not counted the two steps requiring pyruvoyl enzymes at decarboxylation steps in the CoA (pantothenate) synthesis pathway; the pyruvoyl cofactor is synthesized from serine residues in the polypeptide chain of the enzyme [63]. Very surprisingly, only three additional cofactors (biotin, pyridoxal phosphate, and SAM) are required for the synthesis of the 11 other cofactors plus the main nucleosides of nucleic acids and the 20 amino acids, whereby only two more (coenzyme M and coenzyme B) are required specifically in the methanogenic pathway of energy conservation. That makes a total of 19 cofactors (counting NAD and NADP separately as well as the flavins flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) along with two corrins F_{420} and cobamide) to support their own synthesis plus the synthesis of four ribonucleoside triphosphates, four deoxyribonucleoside triphosphates, and 20 amino acids. The genetic code and polymers are not included in the autotrophic core. In total, that makes a list of 47 target compounds (19 cofactors, 8 nucleotides, and 20 amino acids; Table S1) that would be required to synthesize the substance of cells, as summarized in Figure 1.

The starting point of Figure 1a is H_2 and CO_2. Critics of autotrophic origins will be quick to point out that cyanide chemistry can readily give rise to amino acids and bases under laboratory conditions [14,64], such that we need not worry about Figure 1. However, in reply, we would be equally quick to point out that there are 415 distinct reactions in microbial metabolism involving CO_2 as a substrate in either the forward or reverse direction [65], but there are no reactions known to us in which cyanide serves as a main source of carbon in core anabolic metabolism. Some bacteria can convert CN^- to CO_2 and NH_3 or formate and NH_3 for growth [66,67], because CO_2, formate, and NH_3 readily enter metabolism, whereas cyanide does not. In other words, CO_2 directly enters and exits the organic chemistry of the cell substance at 415 reactions, where cyanide puts up a zero. We interpret the fact that cyanide has nothing to do with modern metabolism as a clear metabolic fossil: no role for cyanide in modern metabolism indicates that cyanide had nothing to do with primordial metabolism either, or was at best <1/415th as important as CO_2. In that sense, the main message of Figure 1 is the overall scheme, the metabolism of cells, not that it contains amino acids and bases as products.

That brings us to nitrogen. If carbon did not enter metabolism via cyanide, then the same must be true for nitrogen. If not via cyanide, how did N enter metabolism? All the amino acids, bases, and cofactors contain nitrogen (except coenzyme M). Nitrogen enters metabolism as NH_3 (NH_4^+ is very unreactive) with N atoms replacing O atoms in amino acids, either via an acyl phosphate intermediate in the glutamine synthase reaction or

via reductive aminations of 2-oxoacids [36,68]. An exception is the carbamoyl phosphate synthase reaction, in which NH_3 reacts with carboxyphosphate to form carbamate in pyrimidine and arginine biosynthesis. Of course, NH_3 is synthesized from N_2 by nitrogenase to make it available for incorporation into organic compounds. However, N does not enter metabolism as N_2; it enters metabolism as NH_3, which is why we selected NH_3 as the source of nitrogen in Figure 1. Similarly, sulfur enters metabolism as H_2S in cysteine synthesis from serine, either via serine activation as *O*-acetylserine or *O*-phosphoserine [69]. N and S enter metabolism as dissolved gasses (NH_3 and H_2S) via amino acid synthesis [36]. In cells that live from H_2 and CO_2, C, N, S, and electrons (H_2) enter metabolism as gasses.

3.2. Enzymatic Reactions in the Autotrophic Core

Figure 1a depicts the relationships among reactions that underpin the core synthesis of cells from H_2, CO_2, and NH_3, but it does not depict the reactions themselves. To find out which, what kind of, and how many reactions are required to synthesize 18 cofactors, 8 nucleotides, and 20 amino acids from H_2, CO_2, NH_3, and H_2S, we turned to KEGG pathways using Figure 1 as a framework to identify the reactions and enzymes that catalyze them. The metabolic network for the 404 reactions (Table S2) that comprise the autotrophic core is shown in Figure 2.

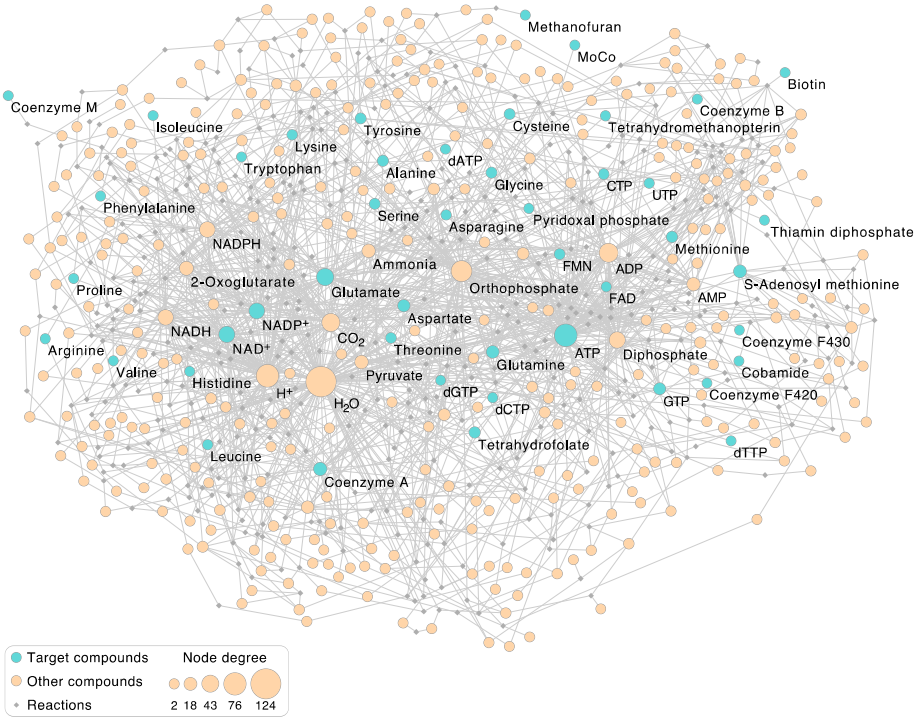

Figure 2. The autotrophic core network of 404 reactions underlying Figure 1. The undirected bipartite graph comprises 404 reaction nodes (displayed as gray diamonds) and 380 compound nodes (circles). The 46 target compounds are colored blue; other compounds involved in the reactions appear orange. Target compounds correspond to the core compounds in Figure 1. Each compound participating in a reaction is connected to the respective reaction node with an edge. Compounds are sized according to node degree (number of reactions the compound takes place in). For example, H_2O appears either as reactant or product in 125 reactions and is the most frequent compound in the 404 reactions (see also Table 1). In primordial metabolic processes, before the existence of enzymes, a more limited spectrum of compounds than those in Figure 1 was provided by the environment. Compound nodes are labeled if they are targets or if the node degree is $\geq$20. Note that FeS clusters are not included in this figure since their synthesis cannot be reconstructed using KEGG. The network contains only L-amino acids.

Other than supplying a greater level of detail than Figure 1, and showing the relative size of nodes, the network itself in Figure 2 is not hugely informative, but some of its properties are. Keeping in mind that Figure 2 comprises the marrow of modern metabolism, hence reactions that were present in life's common ancestor, we first asked what the most highly connected metabolites are. The fifteen most common metabolites are given in Table 1. The most common compound in the autotrophic core is by far H_2O. As stated above, water is the solvent of life's chemistry and its most common reaction partner. Proponents of the RNA world generally view water as a poison for origins, because it promotes the hydrolysis of RNA [70]. However, the host rocks of serpentinizing hydrothermal systems are replete with environments of low water activity, mainly because water is consumed by rock in the serpentinization process [71,72]. Furthermore, fluctuating water activities that occur during serpentinization can be conducive to polymerization reactions [71]. Life counteracts the hydrolysis problem by coupling nucleic acid and protein polymerization reactions to exergonic reactions via ATP synthesis and hydrolysis such that polymer synthesis vastly outpaces hydrolysis [58]. Accordingly, ATP is the second most common reactant in the autotrophic core (Table 1), right before protons. Protons are of course normally bound to water as H_3O^+, although they are not counted as water here. Protons arise in hydride transfer reactions involving NADH and NADPH which yield NAD^+ and $NADP^+$, respectively. The frequency of protons in the network mainly reflects the frequency of NAD(P)H-dependent redox reactions in the autotrophic core (Table 1).

Among reactions that involve the formation or alteration of bonds with carbon atoms, the most common carbon-containing compound in the autotrophic core is, fittingly, CO_2, which underscores the CO_2-dependent nature of core metabolism. Among the 404 reactions in the core, 49, or every eighth reaction, involves CO_2. This can be seen as physiological evidence in favor of autotrophic origins. The next most common carbon backbone in the core is glutamate, which is the main workhorse of nitrogen metabolism. Glutamate arises as a product in amidotransferase reactions involving glutamine as an amino donor and in transamination reactions that produce 2-oxoglutarate, which is also among the top 15 reactants in the core. ATP hydrolysis products P_i and PP_i round out the list as well as pyruvate, which links the acetyl CoA pathway to sugar synthesis and the reverse citric acid cycle [73] and is a common starting point for cofactor synthesis in the autotrophic core (Figure 2). Last among the top fifteen is NH_3, which is often donated to biosynthetic reactions from glutamine via an amidotransferase [74] during the enzymatic reaction, without being released as free NH_3 in the cytosol.

We identified five autocatalytic cycles in the network, that is, cofactors that are required for their own biosynthesis: pyridoxal phosphate and thiamine, whose biosyntheses were previously identified as autocatalytic cycles [36], plus ATP, NAD, and NADP. Though not contained within our set, Davidson recently reported that coenzyme A is required for activation of the complex that synthesizes the active moiety of decarboxylating pyruvoyl enzymes, which are involved in CoA biosynthesis [63]. That would make a sixth autocatalytic cycle.

3.3. Comparison of the Autotrophic Core with LUCA's Genes and Ancient Autocatalytic Sets

Other recent papers have addressed the nature of ancient metabolism by looking at phosphate-independent reactions among all KEGG reactions [75], the properties of thioester-dependent reactions [76] or chemical investigations of metabolic reactions without enzymes [30,37,49,77]. A different approach has been to focus on evidence for the nature of ancient microbial metabolism that is recorded in the genomes and metabolism of bacteria and archaea. A phylogenetic approach to ancient microbial metabolism uncovered 355 genes present in bacteria and archaea trace to LUCA on the basis of vertical intradomain inheritance as opposed to archaeal–bacteria transfer [50]. Autocatalytic networks called RAFs, for reflexively autocatalytic food-generated networks, have been identified in the metabolism of anaerobic autotrophs, with an ancient RAF of 172 genes that overlaps in the

metabolism of H_2-dependent acetogens and methanogens [51]. Do these sets overlap with the autotrophic core, and if so, how?

A comparison of these three sets (Table S3; Figure S1) reveals that among the 404 reactions of the autotrophic core, only 24 are represented among the 355 genes (6%) that trace to LUCA. That low degree of overlap is not surprising for two reasons. First, only a fraction of genes that trace to LUCA by phylogenetic criteria were involved in amino acid or cofactor biosynthesis, most being involved in ribosomal biogenesis or other categories. Second, only 3% of all genes shared by bacteria and archaea were not subjected to bacterial–archaeal transfers by the measure of phylogenetic trees [50], which is a criterion that played no role in the construction of Figure 1. However, it is very noteworthy that all of the cofactors shown in Figure 1, with the exception of the archaeal-specific cofactors CoM and CoB, do trace to LUCA via phylogeny, because enzymes that trace to LUCA possessed those cofactor requirements for activity [50]. In that sense, there is excellent agreement between the physiology of LUCA as inferred from phylogeny and the present autotrophic core, their commonality being cofactors, organic catalysts that are smaller and involved in a greater number of reactions than any individual enzyme.

Among the 172 reactions present in the ancient autocatalytic network shared by acetogen and methanogen RAFs [51], 81 (47%) are present in the autotrophic core. This substantial overlap also makes sense, because all cells use the same amino acids and because both this study and Xavier et al. [51] focused on bacteria and archaea that use the acetyl CoA pathway, which by itself involves almost all of the cofactors shown in Figure 1 as it operates in bacteria and archaea. That is again noteworthy, because even though pyruvate, the central C3 product of the acetyl CoA pathway [6], can be obtained from H_2 and CO_2 using only simple minerals as catalysts [30], the biological pathway requires about a dozen enzymes and cofactors. These cofactors trace to LUCA [50], are well represented in RAFs [51], and comprise the basal foundation of the ancient autotrophic core (Figure 1b). Clearly, in early metabolism, cofactors and the catalytic minerals that were their inorganic precursors were very important [78]. Although self-evident, this indicates that there existed a vectorial progression in metabolic evolution that centered around the nature of catalysts: from transition metal minerals to organic cofactors to enzymes, each adding specificity and rate enhancement to exergonic reactions that tend to occur anyway. The retention of transition metal centers in some enzymes, such as carbon monoxide dehydrogenase, acetyl CoA synthetase, hydrogenases, or nitrogenase, suggests that microbes have been unable to invent catalysts that can perform the same reactions without the help of electrons in the *d*-orbitals of transition metals.

The comparison with 5994 anaerobic prokaryotic reactions (see S1A in [51]) tells us which compounds are enriched in the autotrophic core. Table S4 shows that this is true for ATP (and ADP plus P_i), CO_2, glutamate, pyruvate, and 2-oxoglutarate. This suggests a more crucial role of these compounds in the origin of the core subsequent to later evolution in anaerobes, reflecting a process of carbon backbone elongation from CO_2 at the heart of the core as a supply of precursors for cofactor and amino acid biosynthesis, the latter being the starting point for nucleotide biosynthesis [78].

4. Conclusions

It is human nature to wonder about the origin of life, which is an issue that is among the most debated of all scientific questions. However, in comparison to questions concerning the existence of dark matter or how consciousness works, the origins process lies in the ancient past, and its events are only accessible through inference. Debates within the origins field can be fierce and have a long history. They hinge upon definitions about what qualifies as being alive, what one assumes to be the habitat that brought forth the first biochemical reactions, what came first, small molecule metabolism and proteins or nucleic acids and genetics, what the nature of first energy source(s) was that the early life forms harnessed in order to grow, and what kinds of chemical compounds existed before the first energy-releasing reactions germane to modern metabolism started taking

place [21]. The literature harboring those debates is generally exhausting, because the same arguments resurface over and over again. The more broadly one reads the literature on early evolution, the more one gets the impression that scientists not only do not agree about origins and the nature of the first forms of life, but worse, that scientists know little about early evolution, leaving the topic open to unconstrained speculation and argument. That puts the origins field at risk of defining scientific progress in the units of debate preparation and presentation skills rather than units of empirical findings that are linked to the explanandum (real life); it also risks vulnerability to criticisms about the role of dogma in science.

Biologists tend to hold that there are traces of early evolution preserved in metabolism itself [4,6]. While there is no obvious proof for that conjecture, the nature of basic building blocks of life is dramatically well conserved across all cells [79]. All life forms we know use proteins made of amino acids, nucleic acids made of purines, pyrimidines, sugars, and phosphate. That means that the first forms of life from which all modern forms descend had that core chemistry in place, in addition to the universal genetic code to transfer information from nucleic acids to protein at the ribosome. This adds direly needed constraints to the origins problem. By looking at metabolism from a comparative standpoint, one can distill insights into the nature of early cells.

Here, we have identified 404 reactions that comprise the autotrophic core. It contains five small autocatalytic cycles in which cofactors participate in their own synthesis. The core represents a collection of reactions that underpin the synthesis of RNA and proteins. It was present in the first cells, but it can hardly have arisen all at once. The aqueous synthesis of pyruvate from H_2 and CO_2 using only solid-state metal or metal oxides as a catalyst [30] indicates that the core itself likely started from H_2 and CO_2 and grew outwards from pyruvate while incorporating nitrogen from NH_3. How complex the core could have become prior to the origin of enzymes is a question for future study. However, let us keep in mind that enzymes just accelerate reactions that tend to occur anyway. It is well known that many enzymatic reactions take place without enzymes [36], although sometimes, the non-enzymatic reaction rates can be so slow as to be irrelevant [80]. However, it was also demonstrated that citric acid cycle reactions [49,81] and a number of reactions involving sugars in central metabolism [77,82] can be catalyzed non-enzymatically. This suggests that a fairly complex system of reactions, yet with far less specificity than that in the core, could have arisen before the advent of genes and proteins.

H_2O is the most common reactant in the autotrophic core, indicating an aqueous environment during its formation. That environment was not only aqueous but also reducing, as revealed by the abundance of redox reactions in the autotrophic core, the central role of CO_2, and the circumstance that the core's main products (amino acids and nucleic acids) are far more reduced than CO_2. Furthermore, the number of central reactions depending upon the hydrolysis of high-energy phosphate bonds indicates that the core arose in the presence of a continuous and highly exergonic chemical reaction capable of continuously synthesizing high-energy phosphate bonds, both before and after the origin of enzymes; here, an H_2-dependent CO_2 reduction to acetate [30] forming acyl phosphate bonds [58] is the proposition.

Thus, the chemical reactions of the autotrophic core suggest that it formed in an aqueous environment that supplied H_2, CO_2, and NH_3, was highly reducing, and harboring continuously far from equilibrium conditions. Those conditions are very similar to those found in serpentinizing hydrothermal systems [44,77], and furthermore, they are very similar to those inferred from the functions of enzymes that vertically trace to the last universal common ancestor [50,83].

Notwithstanding pyrrolysine [84], selenocysteine [85], and a number of modified bases [86], the lack of fundamental deviation among modern life forms from the core building blocks of life, core information processing, and the core repertoire of cofactors [87] indicates that whatever chemical processes occurred at origin did not give rise to alternative cores with enough staying power to persist to the present. "Still, other

cores could have existed" the critic might interject, which is true. "But even if they existed, they are irrelevant", we would counter, because they are disjunct from the biologist's explanandum: the autotrophic core that we can observe in modern life forms.

Supplementary Materials: The following are available online at https://www.mdpi.com/2076-2607/9/2/458/s1, Table S1: List of 47 target compounds of the autotrophic core metabolism, Table S2: Autotrophic core consisting of 404 metabolic reactions needed to synthesize amino acids, nucleic acids, cofactors and intermediate precursors, Table S3: (A) Reaction lists for 404 autotrophic core reactions, 163 LUCA reactions, 172 Core RAF reactions, (B) their respective intersections and (C) intersection of all three reaction sets with functional KEGG annotation, Table S4: Frequency of highly connected nodes among the autotrophic core and the global prokaryote anaerobic network and results of statistical tests for significant compound enrichment using Fisher's exact test, Figure S1: Venn diagram showing the proportion of intersecting reactions of three different core datasets.

Author Contributions: Conceptualization, W.F.M., J.L.E.W., J.C.X., M.P.; methodology, J.L.E.W., J.C.X.; formal analysis, J.L.E.W.; data curation, J.L.E.W., J.C.X.; writing—original draft preparation, W.F.M., M.P.; writing—review and editing, M.P., J.L.E.W., W.F.M., A.d.N.V., J.C.X., K.K.; visualization, J.L.E.W. and W.F.M.; supervision, W.F.M.; funding acquisition, W.F.M. and M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Research Foundation (Ma 1426/21-1) as well as the Volkswagen Foundation (VW 96742).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Metabolic data that supports the findings of this study is available in KEGG [52]. The data presented in this study are available in supplementary material.

Acknowledgments: We thank Harun Tüysüz, Joseph Moran and Fernando Tria for discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Haeckel, E. *Natürliche Schöpfungs-Geschichte. Gemeinverständliche Wissenschaftliche Vorträge über die Entwickelungslehre Zehnte verbesserte Auflage. Zweiter Theil: Allgemeine Stammesgeschichte*; Georg Reimer Verlag: Berlin, Germany, 1902.
2. Winogradsky, S. *Beiträge zur Morphologie und Physiologie der Bakterien. H. 1. Zur Morphologie und Physiologie der Schwefelbakterien*; A. Felix: Leipzig, Germany, 1888.
3. Madigan, M.T.; Martinko, J.M.; Bender, K.S.; Buckley, D.H.; Stahl, D.A. *Brock Biology of Microorganisms*, 15th ed.; Pearson Education: Harlow, UK, 2019.
4. Decker, K.; Jungermann, K.; Thauer, R.K. Energy production in anaerobic organisms. *Angew. Chem. Int. Ed.* **1970**, *9*, 138–158. [CrossRef]
5. Baross, J.A.; Hoffmann, S.E. Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. *Orig. Life Evol. Biosph.* **1985**, *15*, 327–345. [CrossRef]
6. Fuchs, G. Alternative pathways of carbon dioxide fixation: Insights into the early evolution of life? *Annu. Rev. Microbiol.* **2011**, *65*, 631–658. [CrossRef]
7. Kowallik, K.V.; Martin, W.F. The origin of symbiogenesis: An annotated English translation of Mereschkowsky's 1910 paper on the theory of two plasma lineages. *Biosystems* **2021**, *199*, 104281. [CrossRef]
8. Miller, S.L. A production of amino acids under possible primitive earth conditions. *Science* **1953**, *117*, 528–529. [CrossRef] [PubMed]
9. Oró, J.; Kimball, A.P. Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide. *Arch. Biochem. Biophys.* **1961**, *94*, 217–227. [CrossRef]
10. Haruna, I.; Spiegelman, S. Specific template requirements of RNA replicases. *Proc. Natl. Acad. Sci. USA* **1965**, *54*, 579–587. [CrossRef]
11. Cech, T.R. The chemistry of self-splicing enzymes. *Science* **1986**, *236*, 1532–1539. [CrossRef]
12. Gilbert, W. The RNA world. *Nature* **1986**, *319*, 618. [CrossRef]
13. Lincoln, T.A.; Joyce, G.F. Self-sustained replication of an RNA enzyme. *Science* **2009**, *323*, 1229–1232. [CrossRef]
14. Powner, M.W.; Gerland, B.; Sutherland, J.D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **2009**, *459*, 239–242. [CrossRef]
15. Sutherland, J.D. The origin of life—Out of the blue. *Angew. Chem. Int. Ed.* **2016**, *55*, 104–121. [CrossRef]

16. Shapiro, R. Small molecule interactions were central to the origin of life. *Q. Rev. Biol.* **2006**, *81*, 105–126. [CrossRef] [PubMed]
17. Wächtershäuser, G. Before enzymes and templates: Theory of surface metabolism. *Microbiol. Rev.* **1988**, *52*, 452–484. [CrossRef]
18. Orgel, L.E. The implausibility of metabolic cycles on the prebiotic earth. *PLoS Biol.* **2008**, *6*, e18–e19. [CrossRef] [PubMed]
19. Kamminga, H. Historical perspective: The problem of the origin of life in the context of developments in biology. *Orig. Life Evol. Biosph.* **1988**, *18*, 1–11. [CrossRef]
20. Pascal, R.; Pross, A. Stability and its manifestation in the chemical and biological worlds. *Chem. Commun.* **2015**, *51*, 16160–16165. [CrossRef]
21. Preiner, M.; Asche, S.; Becker, S.; Betts, H.C.; Boniface, A.; Camprubi, E.; Chandru, K.; Erastova, V.; Garg, S.G.; Khawaja, N.; et al. The future of origin of life research: Bridging decades-old divisions. *Life* **2020**, 10. [CrossRef] [PubMed]
22. Lipmann, F. Projecting backward from the present stage of evolution of biosynthesis. In *The Origin of Prebiological Systems and of Their Molecular Matrices*; Fox, S.W., Ed.; Academic Press: New York, NY, USA, 1965; pp. 259–280.
23. Ferry, J.G.; House, C.H. The stepwise evolution of early life driven by energy conservation. *Mol. Biol. Evol.* **2006**, *23*, 1286–1292. [CrossRef] [PubMed]
24. Corliss, J.B.; Dymond, J.; Gor, L.I.; Edmond, J.M.; Von Herzen, R.P.; Bal, R.D.; Green, K.; Williams, D.; Bainbri, A.; Crane, K.; et al. Submarine thermal springs on the Galapágos rift. *Science* **1979**, *203*, 1073–1083. [CrossRef] [PubMed]
25. Kelley, D.S.; Karson, J.A.; Blackman, D.K.; Früh-Green, G.L.; Butterfield, D.A.; Lilley, M.D.; Olson, E.J.; Schrenk, M.O.; Roe, K.K.; Lebon, G.T.; et al. An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30° N. *Nature* **2001**, *412*, 145–149. [CrossRef]
26. Charlou, J.L.; Donval, J.P.; Fouquet, Y.; Jean-Baptiste, P.; Holm, N. Geochemistry of high H_2 and CH_4 vent fluids issuing from ultramafic rocks at the Rainbow hydrothermal field (36°14′ N, MAR). *Chem. Geol.* **2002**, *191*, 345–359. [CrossRef]
27. Thauer, R.K.; Kaster, A.-K.; Seedorf, H.; Buckel, W.; Hedderich, R. Methanogenic archaea: Ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* **2008**, *6*, 579–591. [CrossRef]
28. Sleep, N.H.; Bird, D.K.; Pope, E.C. Serpentinite and the dawn of life. *Philos. Trans. R. Soc. B* **2011**, *366*, 2857–2869. [CrossRef] [PubMed]
29. Martin, W.; Baross, J.; Kelley, D.; Russell, M.J. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **2008**, *6*, 6–805. [CrossRef] [PubMed]
30. Preiner, M.; Igarashi, K.; Muchowska, K.B.; Yu, M.; Varma, S.J.; Kleinermanns, K.; Nobu, M.K.; Kamagata, Y.; Tüysüz, H.; Moran, J.; et al. A hydrogen dependent geochemical analogue of primordial carbon and energy metabolism. *Nat. Ecol. Evol.* **2020**, *4*, 534–542. [CrossRef]
31. Schönheit, P.; Buckel, W.; Martin, W.F. On the origin of heterotrophy. *Trends Microbiol.* **2016**, *24*, 12–25. [CrossRef] [PubMed]
32. Harold, F.M. *The Vital Force: A Study of Bioenergetics*; WH Freeman: New York, NY, USA, 1986.
33. Russell, J.B.; Cook, G.M. Energetics of bacterial growth: Balance of anabolic and catabolic reactions. *Microbiol. Rev.* **1995**, *59*, 48–62. [CrossRef]
34. Russell, J.B. The energy spilling reactions of bacteria and other organisms. *J. Mol. Microbiol. Biotechnol.* **2007**, *13*, 1–11. [CrossRef]
35. Van Dijken, J.P.; Scheffers, W.A. Redox balances in the metabolism of sugars by yeasts. *FEMS Microbiol. Rev.* **1986**, *32*, 199–224. [CrossRef]
36. Martin, W.F.; Russell, M.J. On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. B* **2007**, *362*, 1887–1925. [CrossRef] [PubMed]
37. Muchowska, K.B.; Varma, S.J.; Moran, J. Synthesis and breakdown of universal metabolic precursors promoted by iron. *Nature* **2019**, *569*, 104–107. [CrossRef] [PubMed]
38. Ralser, M. An appeal to magic? The discovery of a non-enzymatic metabolism and its role in the origins of life. *Biochem. J.* **2018**, *475*, 2577–2592. [CrossRef]
39. Thauer, R.K.; Jungermann, K.; Decker, K. Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol. Rev.* **1977**, *41*, 100–180. [CrossRef] [PubMed]
40. Stouthamer, A.H. Energy-yielding pathways. In *The Bacteria*; Gunsalus, I.C., Ed.; Academic Press: New York, NY, USA, 1978; Volume 6, pp. 389–462.
41. Hoehler, T.M.; Jørgensen, B.B. Microbial life under extreme energy limitation. *Nat. Rev. Microbiol.* **2013**, *11*, 83–94. [CrossRef]
42. Arndt, N.T.; Nisbet, E.G. Processes on the young earth and the habitats of early life. *Annu. Rev. Earth Planet. Sci.* **2012**, *40*, 521–549. [CrossRef]
43. Russell, M.J.; Hall, A.J.; Martin, W. Serpentinization as a source of energy at the origin of life. *Geobiology* **2010**, *8*, 355–371. [CrossRef]
44. Preiner, M.; Xavier, J.C.; Sousa, F.; Zimorski, V.; Neubeck, A.; Lang, S.Q.; Greenwell, H.C.; Kleinermanns, K.; Tüysüz, H.; McCollom, T.M.; et al. Serpentinization: Connecting geochemistry, ancient metabolism and industrial hydrogenation. *Life* **2018**, *8*, 41. [CrossRef]
45. Proskurowski, G.; Lilley, M.D.; Seewald, J.S.; Früh-Green, G.L.; Olson, E.J.; Lupton, J.E.; Sylva, S.P.; Kelley, D.S. Abiogenic hydrocarbon production at lost city hydrothermal field. *Science* **2008**, *319*, 604–607. [CrossRef]
46. Lang, S.Q.; Butterfield, D.A.; Schulte, M.; Kelley, D.S.; Lilley, M.D. Elevated concentrations of formate, acetate and dissolved organic carbon found at the Lost City hydrothermal field. *Geochim. Cosmochim. Acta* **2010**, *74*, 941–952. [CrossRef]
47. Konn, C.; Charlou, J.L.; Holm, N.G.; Mousis, O. The production of methane, hydrogen and organic compounds in ultramafic-hosted hydrothermal vents of the Mid-Atlantic Ridge. *Astrobiology* **2015**, *15*, 381–399. [CrossRef] [PubMed]

48. Schrenk, M.O.; Brazelton, W.J.; Lang, S.Q. Serpentinization, carbon, and deep life. *Rev. Mineral. Geochem.* **2013**, *75*, 575–606. [CrossRef]
49. Varma, S.J.; Mochowska, K.B.; Chatelain, P.; Moran, J. Native iron reduces CO_2 to intermediates and endproducts of the acetyl-CoA pathway. *Nat. Ecol. Evol.* **2018**, *2*, 1019–1024. [CrossRef]
50. Weiss, M.C.; Sousa, F.L.; Mrnjavac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W.F. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **2016**, *1*, 16116. [CrossRef] [PubMed]
51. Xavier, J.C.; Hordijk, W.; Kauffman, S.; Steel, M.; Martin, W.F. Autocatalytic chemical networks preceded proteins and RNA in evolution. *Proc. R. Soc. Lond. B* **2020**, 287. [CrossRef]
52. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef] [PubMed]
53. Qiu, Y.Q. KEGG pathway database. In *Encyclopedia of Systems Biology*; Dubitzky, W., Wolkenhauer, O., Cho, K.H., Yokota, H., Eds.; Springer: New York, NY, USA, 2013; pp. 1068–1069.
54. Hazra, A.B.; Han, A.W.; Mehta, A.P.; Mok, K.C.; Osadchiy, V.; Begley, T.P.; Taga, M.E. Anaerobic biosynthesis of the lower ligand of vitamin B12. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10792–10797. [CrossRef]
55. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]
56. Berg, I.A.; Kockelkorn, D.; Ramos-Vera, W.H.; Say, R.F.; Zarzycki, J.; Hügler, M.; Alber, B.E.; Fuchs, G. Autotrophic carbon fixation in archaea. *Nat. Rev. Microbiol.* **2010**, *8*, 447–460. [CrossRef] [PubMed]
57. Hügler, M.; Sievert, S.M. Beyond the Calvin cycle: Autotrophic carbon fixation in the ocean. *Ann. Rev. Mar. Sci.* **2011**, *3*, 261–289. [CrossRef]
58. Martin, W.F. Older than genes: The acetyl CoA pathway and origins. *Front. Microbiol.* **2020**, *11*, 11–817. [CrossRef]
59. Mall, A.; Sobotta, J.; Huber, C.; Tschirner, C.; Kowarschik, S.; Bačnik, K.; Mergelsberg, M.; Boll, M.; Hügler, M.; Eisenreich, W.; et al. Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. *Science* **2018**, *359*, 563–567. [CrossRef] [PubMed]
60. Nunoura, T.; Chikaraishi, Y.; Izaki, R.; Suwa, T.; Sato, T.; Harada, T.; Mori, K.; Kato, Y.; Miyazaki, M.; Shimamura, S.; et al. A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. *Science* **2018**, *359*, 559–563. [CrossRef]
61. Maden, B.E. Tetrahydrofolate and tetrahydromethanopterin compared: Functionally distinct carriers in C1 metabolism. *Biochem. J.* **2000**, *350*, 609–629. [CrossRef] [PubMed]
62. Sousa, F.L.; Martin, W.F. Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochim. Biophys. Acta* **2014**, *1837*, 964–981. [CrossRef]
63. Davidson, V.L. Protein-derived cofactors revisited: Empowering amino acid residues with new functions. *Biochemistry* **2018**, *57*, 3115–3125. [CrossRef]
64. Canavelli, P.; Islam, S.; Powner, M.W. Peptide ligation by chemoselective aminonitrile coupling in water. *Nature* **2019**, *571*, 546–549. [CrossRef] [PubMed]
65. Xavier, J.C.; Preiner, M.; Martin, W.F. Something special about CO-dependent CO_2 fixation. *FEBS J.* **2018**, *285*, 4181–4195. [CrossRef]
66. Knowles, C.J. Cyanide utilization and degradation by microorganisms. *Ciba. Found. Symp.* **1988**, *140*, 3–15. [CrossRef] [PubMed]
67. Fernandez, R.F.; Dolghih, E.; Kunz, D.A. Enzymatic assimilation of cyanide via pterin-dependent oxygenolytic cleavage to ammonia and formate in Pseudomonas fluorescens NCIMB 11764. *Appl. Environ. Microbiol.* **2004**, *70*, 121–128. [CrossRef]
68. Huber, C.; Wächtershäuser, G. α-hydroxy and α-amino acids under possible hadean, volcanic origin-of-life conditions. *Science* **2006**, *314*, 630–632. [CrossRef]
69. Liu, Y.; Beer, L.L.; Whitman, W.B. Sulfur metabolism in archaea reveals novel processes. *Environ. Microbiol.* **2012**, *14*, 2632–2644. [CrossRef]
70. Marshall, M. How the first life on Earth survived its biggest threat—Water. *Nature* **2020**, *588*, 210–213. [CrossRef]
71. Do Nascimento Vieira, A.; Kleinermanns, K.; Martin, W.F.; Preiner, M. The ambivalent role of water at the origins of life. *FEBS Lett.* **2020**, *594*, 2717–2733. [CrossRef] [PubMed]
72. Lamadrid, H.M.; Rimstidt, J.D.; Schwarzenbach, E.M.; Klein, F.; Ulrich, S.; Dolocan, A.; Bodnar, R.J. Effect of water activity on rates of serpentinization of olivine. *Nat. Commun.* **2017**, *8*, 16107. [CrossRef] [PubMed]
73. Berg, I.A.; Kockelkorn, D.; Buckel, W.; Fuchs, G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in archaea. *Science* **2007**, *318*, 1782–1786. [CrossRef] [PubMed]
74. Raushel, F.M.; Thoden, J.B.; Holden, H.M. The amidotransferase family of enzymes: Molecular machines for the production and delivery of ammonia. *Biochemistry* **1999**, *38*, 7891–7899. [CrossRef]
75. Goldford, J.E.; Hartman, H.; Smith, T.F.; Segrè, D. Remnants of an ancient metabolism without phosphate. *Cell* **2017**, *168*, 1126–1134. [CrossRef] [PubMed]
76. Semenov, S.N.; Kraft, L.J.; Ainla, A.; Zhao, M.; Baghbanzadeh, M.; Campbell, V.E.; Kang, K.; Fox, J.M.; Whitesides, G.M. Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. *Nature* **2016**, *537*, 656–660. [CrossRef] [PubMed]
77. Muchowska, K.B.; Varma, S.J.; Moran, J. Nonenzymatic metabolic reactions and life's origins. *Chem. Rev.* **2020**, *120*, 7708–7744. [CrossRef]

78. Preiner, M.; Xavier, J.C.; Vieira, A.D.N.; Kleinermanns, K.; Allen, J.F.; Martin, W.F. Catalysts, autocatalysis and the origin of metabolism. *Interface Focus* **2019**, *9*, 9–20190072. [CrossRef]
79. Morowitz, H.J.; Kostelnik, J.D.; Yang, J.; Cody, G.D. The origin of intermediary metabolism. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7704–7708. [CrossRef] [PubMed]
80. Wolfenden, R. Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Annu. Rev. Biochem.* **2011**, *80*, 645–667. [CrossRef]
81. Muchowska, K.B.; Varma, S.J.; Chevallot-Beroux, E.; Lethuillier-Karl, L.; Li, G.; Moran, J. Metals promote sequences of the reverse Krebs cycle. *Nat. Ecol. Evol.* **2017**, *1*, 1716–1721. [CrossRef]
82. Messner, C.B.; Driscoll, P.C.; Piedrafita, G.; De Volder, M.F.L.; Ralser, M. Nonenzymatic gluconeogenesis-like formation of fructose 1,6-bisphosphate in ice. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 7403–7407. [CrossRef]
83. Sousa, F.L.; Nelson-Sathi, S.; Martin, W.F. One step beyond a ribosome: The ancient anaerobic core. *Biochim. Biophys. Acta.* **2016**, *1857*, 1027–1038. [CrossRef]
84. Krzycki, J.A. The direct genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* **2005**, *8*, 706–712. [CrossRef] [PubMed]
85. Böck, A.; Forchhammer, K.; Heider, J.; Leinfelder, W.; Sawers, G.; Veprek, B.; Zinoni, F. Selenocysteine: The 21st amino acid. *Mol. Microbiol.* **1991**, *5*, 515–520. [CrossRef] [PubMed]
86. Grosjean, H.; de Crécy-Lagard, V.; Marck, C. Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes. *FEBS Lett.* **2010**, *584*, 252–264. [CrossRef] [PubMed]
87. Xavier, J.C.; Patil, K.R.; Rocha, I. Integration of biomass formulations of genome-scale metabolic models with experimental data reveals universally essential cofactors in prokaryotes. *Metab. Eng.* **2017**, *39*, 200–208. [CrossRef] [PubMed]

II Energy at origins: Favorable thermodynamics of biosynthetic reactions in the Last Universal Common Ancestor (LUCA)

Jessica L. E. Wimmer, Joana C. Xavier, Andrey d. N. Vieira, Delfina P. H. Pereira, Jacqueline Leidner, Filipa L. Sousa, Karl Kleinermanns, Martina Preiner, William F. Martin

Institut für Molekulare Evolution, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Deutschland.

Dieser Artikel wurde am 13.12.2021 in *Frontiers in Microbiology* Ausgabe 12 veröffentlicht.

Beitrag von Jessica L. E. Wimmer (Erstautor und Co-Korrespondenz):
Ich war in die Entwicklung des Studienkonzepts involviert, habe die Methodologie mitausgearbeitet und den Datensatz zusammengestellt sowie überarbeitet. Des Weiteren wurde die überwiegende Mehrheit der Analysen von mir durchgeführt und Abbildungen durch mich erstellt. Das initiale Manuskript wurde von mir überarbeitet.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 December 2021
doi: 10.3389/fmicb.2021.793664

Energy at Origins: Favorable Thermodynamics of Biosynthetic Reactions in the Last Universal Common Ancestor (LUCA)

Jessica L. E. Wimmer[1], Joana C. Xavier[1], Andrey d. N. Vieira[1], Delfina P. H. Pereira[1], Jacqueline Leidner[1], Filipa L. Sousa[2], Karl Kleinermanns[3], Martina Preiner[1] and William F. Martin[1]**

[1] Department of Biology, Institute of Molecular Evolution, Heinrich Heine University Düsseldorf, Düsseldorf, Germany, [2] Department of Functional and Evolutionary Ecology, University of Vienna, Vienna, Austria, [3] Department of Chemistry, Institute of Physical Chemistry, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

OPEN ACCESS

Edited by:
Martin G. Klotz, Washington State University, United States

Reviewed by:
Ivan A. Berg, University of Münster, Germany
Bernhard Schink, University of Konstanz, Germany

***Correspondence:**
Jessica L. E. Wimmer
jessica.wimmer@hhu.de
William F. Martin
bill@hhu.de

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *20 October 2021*
Accepted: *24 November 2021*
Published: *13 December 2021*

Citation:
Wimmer JLE, Xavier JC, Vieira AdN, Pereira DPH, Leidner J, Sousa FL, Kleinermanns K, Preiner M and Martin WF (2021) Energy at Origins: Favorable Thermodynamics of Biosynthetic Reactions in the Last Universal Common Ancestor (LUCA). Front. Microbiol. 12:793664. doi: 10.3389/fmicb.2021.793664

Though all theories for the origin of life require a source of energy to promote primordial chemical reactions, the nature of energy that drove the emergence of metabolism at origins is still debated. We reasoned that evidence for the nature of energy at origins should be preserved in the biochemical reactions of life itself, whereby changes in free energy, ΔG, which determine whether a reaction can go forward or not, should help specify the source. By calculating values of ΔG across the conserved and universal core of 402 individual reactions that synthesize amino acids, nucleotides and cofactors from H_2, CO_2, NH_3, H_2S and phosphate in modern cells, we find that 95–97% of these reactions are exergonic ($\Delta G \leq 0$ kJ·mol^{-1}) at pH 7-10 and 80-100°C under nonequilibrium conditions with H_2 replacing biochemical reductants. While 23% of the core's reactions involve ATP hydrolysis, 77% are ATP-independent, thermodynamically driven by ΔG of reactions involving carbon bonds. We identified 174 reactions that are exergonic by –20 to –300 kJ·mol^{-1} at pH 9 and 80°C and that fall into ten reaction types: six pterin dependent alkyl or acyl transfers, ten S-adenosylmethionine dependent alkyl transfers, four acyl phosphate hydrolyses, 14 thioester hydrolyses, 30 decarboxylations, 35 ring closure reactions, 31 aromatic ring formations, and 44 carbon reductions by reduced nicotinamide, flavins, ferredoxin, or formate. The 402 reactions of the biosynthetic core trace to the last universal common ancestor (LUCA), and reveal that synthesis of LUCA's chemical constituents required no external energy inputs such as electric discharge, UV-light or phosphide minerals. The biosynthetic reactions of LUCA uncover a natural thermodynamic tendency of metabolism to unfold from energy released by reactions of H_2, CO_2, NH_3, H_2S, and phosphate.

Keywords: origin of life, energetics, bioenergetics, metabolism, early evolution, biosynthesis, thermodynamics, last universal common ancestor

INTRODUCTION

Between the first appearance of liquid water on the Earth roughly 4.3 billion years ago (Mojzsis et al., 2001) and the appearance of the first signs of life roughly 3.8 billion years ago (Rosing, 1999), simple spontaneous geochemical reactions gave rise to the enzymatically catalyzed reaction network of microbial metabolism: a highly organized set of specific organic reactions that provides the amino acids, nucleotides and cofactors to sustain ribosomal protein synthesis and growth. How metabolism arose is a keystone issue for understanding how the first microbes arose from the elements. It is a complex problem with many facets, several approaches to investigate the issue are current.

From the standpoint of theory, autocatalytic networks provide a useful framework for the study of metabolic origin (Kauffman, 1986; Hordijk and Steel, 2004). In autocatalytic sets, elements of the set can catalyze the synthesis of other elements of the set, potentially giving rise to molecular self-organization provided that a food source is supplied to drive the network forward (Hordijk et al., 2010). Autocatalytic sets are not purely theoretical objects because they can be identified in the metabolism of both modern cells and their inferred ancestors (Sousa et al., 2015; Xavier et al., 2020).

From the standpoint of individual reactions, inorganic catalysts have long been known to catalyze many metabolic reactions under laboratory conditions (Wächtershäuser, 1992; Huber and Wächtershäuser, 1997; Martin and Russell, 2007; Sousa et al., 2018). More recently, complex reaction sets approximating biochemical pathways (Muchowska et al., 2019, 2020) and in some cases even exactly retracing biochemical pathways (Preiner et al., 2020) have been reported, uncovering a natural tendency of numerous metabolic reactions to unfold in the presence of transition metal catalysis. From the computational standpoint, simulations have been widely employed to study metabolic origin, particularly network expansion algorithms. These have been shown to generate small molecule networks consisting of up to hundreds of compounds with properties that resemble metabolism, with the caveat that networks so generated are not manifest as natural pathways in modern cells (Goldford et al., 2017, 2019; Tian et al., 2019).

Independent of the methodological approach, current investigations of metabolic origin tend to start from the acetyl-CoA pathway of CO_2 fixation (Fuchs and Stupperich, 1985; Fuchs, 2011) for a number of reasons. It is the only pathway of CO_2 fixation (i) that is both linear and exergonic (Berg et al., 2010), (ii) that occurs in both bacteria and archaea (Berg et al., 2010; Fuchs, 2011), and (iii) that traces to the last universal common ancestor (LUCA) (Weiss et al., 2016). Its exergonic nature allows coupling of H_2-dependent CO_2 reduction to ion pumping and ATP synthesis, as in acetogens (Schuchmann and Müller, 2014) and methanogens (Thauer et al., 2008), strict anaerobes that obtain both their carbon and energy from the reduction of CO_2 with H_2. Organisms that use the acetyl-CoA pathway still inhabit H_2-producing geochemical systems (Magnabosco et al., 2018; Smith et al., 2019), habitats that existed on the early Earth (Sleep et al., 2011). The first intermediate of the acetyl-CoA pathway, formate, is synthesized geochemically via abiotic reactions in modern hydrothermal systems (Lang et al., 2010; Schrenk et al., 2013), as are the endproducts of energy metabolism via the pathway in acetogens (acetate; Sherwood Lollar et al., 2021) and in methanogens (methane; Proskurowski et al., 2008). In carbon metabolism, the acetyl-CoA pathway generates pyruvate as the main product (Fuchs, 2011) via reactions that require 10 enzymes and cofactors, yet those enzymes can be replaced by simple hydrothermal minerals such as awaruite (Ni_3Fe), which convert H_2 and CO_2 into formate, acetate and pyruvate overnight at 100°C in water (Preiner et al., 2020). Such findings connect the carbon and energy metabolism of acetogens and methanogens to spontaneous geochemical processes in H_2-producing hydrothermal vents via the chemical reactions of the acetyl-CoA pathway (Martin, 2020).

Thermodynamic studies in geochemical systems also point to an origin of metabolism from H_2 and CO_2 in a hydrothermal setting, as the synthesis of amino acids (Amend and Shock, 1998) and even prokaryotic cell mass (Amend and McCollom, 2009) from H_2, CO_2 and NH_3 is exergonic under the chemical conditions germane to H_2-producing hydrothermal vents. However, calculating ΔG for a one-step geochemical reaction that converts H_2, CO_2 and NH_3 into amino acids (Amend and Shock, 1998; Amend et al., 2013) does not begin to capture the thermodynamic landscape of metabolism, either modern or ancient, because the biosynthesis of amino acids and all other cell constituents involve the entry of H_2, CO_2, and NH_3 at a very small number of very specific enzymatic reactions, followed by their distribution in activated form as hydride, organic carbon or amino moieties in highly connected networks of intermediate conversions. For example, over 20 distinct reactions are involved in the synthesis of either tryptophan or purines from H_2, CO_2, and NH_3 (Kanehisa and Goto, 2000). Studies of thermodynamics at metabolic origin ideally need to address the thermodynamics of individual metabolic reactions as they are organized in modern cells or in the inferred ancestors thereof.

Our present investigation into metabolic origin is based on comparative physiology. Wimmer et al. (2021a) identified roughly 400 reactions that are used by bacteria and archaea to synthesize the amino acids, nucleotides and cofactors required for growth. Because these reactions are universal, they represent core biosynthetic metabolism in the last universal common ancestor (LUCA). As such, they can be seen as the endpoint of metabolic origin on the one hand and the starting point of physiological diversification on the other. Here we have updated this set of reactions, which we designate as the metabolic core, to include the two-enzyme reaction sequence of substrate level phosphorylation used by acetogens and some methanogens (Rother and Metcalf, 2004) as an acetyl-CoA dependent source of cytoplasmic (membrane independent) ATP synthesis. Although the acetyl-CoA pathway is not universal, having been replaced by many other autotrophic (Berg et al., 2010; Fuchs, 2011; Hügler and Sievert, 2011; Steffens et al., 2021) and heterotrophic (Schönheit et al., 2016) carbon assimilation pathways during

evolution, it traces to LUCA (Weiss et al., 2016) and, like many of LUCA's biochemical reactions (Sousa et al., 2018), is older than the enzymes that catalyze its reactions (Martin, 2020). Though the remaining chemical reactions of the core do not occur in all genomes, as auxotrophies arise recurrently in evolution, they are universal at the level of primary production, the process that has fueled all ecosystems from origins to today (Hamilton et al., 2016; Martin et al., 2018). However, the enzymes that catalyze the reactions of the core are not universal, such that the core cannot be identified through purely genomic comparisons because (i) reactions that arose post LUCA, in particular O_2-dependent reactions (Dailey et al., 2017; Jabłońska and Tawfik, 2021), need to be filtered out (Wimmer et al., 2021a), (ii) because lateral gene transfers of recently arisen pathways have to be filtered out (Weiss et al., 2016), and (iii) because the enzymes that catalyze these reactions are often unrelated across the archaeal-bacterial divide (Sousa et al., 2013), suggesting independent origins of enzymatic pathways from LUCA en route to the last common ancestors of archaea (Williams et al., 2017) and bacteria (Xavier et al., 2021), respectively.

Despite many unknowns concerning the process of metabolic origin, one factor provides stringent constraint: The chemical reactions that comprised LUCA's metabolism, and those from which it arose, were perforce exergonic, for without energy release, no reactions will take place. It has long been recognized that energy was required to promote reactions at metabolic origin, but the nature of that energy has been debated. Many possible environmental sources of energy at origins have been suggested, including pyrophosphate (PP_i; Schramm et al., 1962), cyclic polyphosphates (Ozawa et al., 2004), reduced phosphorous minerals (Pasek, 2020), ultraviolet light (Patel et al., 2015), radioactive decay (Ebisuzaki and Maruyama, 2017), lightning (Ducluzeau et al., 2009), geochemical pyrite synthesis (Wächtershäuser, 1992), geochemical ion gradients (Russell and Cook, 1995), geoelectrical potential (Kitadai et al., 2021), bolide impacts (Ferus et al., 2015), and heat (Muller, 1995). Modern cells in nature, however, harness none of those environmental energy sources, they harness redox reactions instead (Mitchell, 1961; Thauer et al., 1977; Müller et al., 2018), and conserve energy for metabolic use in the chemically accessible currency of ATP (Decker et al., 1970) or reduced ferredoxin (Herrmann et al., 2008; Buckel and Thauer, 2013; Müller et al., 2018). The fact that only a fraction of core biosynthetic reactions entail ATP hydrolysis (Wimmer et al., 2021a) leads to a seldom formulated question: What drove the majority of LUCA's metabolic reactions forward? We reasoned that ATP-independent biosynthetic reactions might themselves be a possible primordial energy source, one that would be particularly conducive to the formation of autocatalytic networks (Xavier et al., 2020). To investigate further, we polarized the core biosynthetic network of LUCA in the direction of cell synthesis and estimated the changes of Gibbs energy for each individual reaction using the component contribution method (Flamholz et al., 2012; Noor et al., 2013; Beber et al., 2021) to identify the nature of ATP-independent exergonic reactions endogenous to LUCA's biosynthetic metabolism.

MATERIALS AND METHODS

Biosynthetic Network

The 402 metabolic reactions comprising the core were manually polarized in the direction of cell synthesis (Wimmer et al., 2021a; **Supplementary Table 1**). Reactions of the acetyl-CoA pathway in the CO_2 fixing reductive direction (Fuchs, 2011) [the archaeal pathway is missing in The Kyoto Encyclopedia of Genes and Genomes (KEGG)], gluconeogenesis (Say and Fuchs, 2010), the reverse citric acid cycle (Steffens et al., 2021) and the pentose phosphate pathway generate most key intermediates. No anaerobic synthesis was available in KEGG (the standard database for microbial metabolic pathways; Kanehisa and Goto, 2000) for dimethylbenzimidazole, 2-phospholactate and flavins. Three cofactors (CoB, CoM, and F_{430}) that are not required in biosynthesis but are essential for ATP synthesis in methanogenic archaea (Thauer et al., 2008) are included in the core. The rare amino acids selenocysteine and pyrrolysine were not included, nor were modified amino acids in proteins as cofactors, including pyruvoyl enzymes. Reactions were obtained from KEGG (Kanehisa and Goto, 2000), version December 2020, excluding degradation reactions and oxygen-dependent reactions (Wimmer et al., 2021a), including H_2-dependent substrate level phosphorylation (Martin and Thauer, 2017), ferredoxin:NAD(P)H interconversion, and H_2-dependent CO_2 reductase (Schuchmann and Müller, 2014). Of the 18 cofactors in **Figure 1**, 10 are required by the acetyl-CoA pathway in archaea and bacteria from H_2 and CO_2 to pyruvate (Fuchs and Stupperich, 1985; Martin, 2020). The biosynthetic pathway to iron-guanylylpridinol, required for H_2-dependent methenyl H_4MPT reduction in methanogenesis under nickel limitation (Huang et al., 2020), is not represented in KEGG and missing in the network, leaving only two entry points of H_2 into metabolism via ferredoxin-reducing hydrogenases (Huang et al., 2020) and H_2-dependent CO_2 reductase (Schuchmann and Müller, 2014). Except biotin, the compounds clockwise from Trp to methanofuran in **Figure 1** contain at least one aromatic ring. Aromatic ring forming reactions in the core entail five rings in amino acids, six in nucleoside bases, seven in pterins (two shared and five specific), eight in tetrapyrroles (four in pyrrole formation and four leading to F_{430} and cobalamin), two for methanofuran, one each for thiamine, pyridoxal, and pyridine dinucleotides. Each aromatic compound requires a ring formation reaction plus two non-aromatic rings in biotin and one each in ribose and proline. Modern chemolithoautotrophs live from the components shown on the left in **Figure 1** plus trace elements (Magnabosco et al., 2018; Smith et al., 2019), growing on biotic H_2 from fermentations (Wolfenden, 2011) or abiotic H_2 from hydrothermal systems (Schrenk et al., 2013; Dick, 2019; Lang and Brazelton, 2020).

Estimation of Gibbs Energy for Individual Reactions

A few words are needed concerning the component contribution (or group contribution) method. Traditionally, biochemists determine the change of Gibbs energy, ΔG, in a physiological

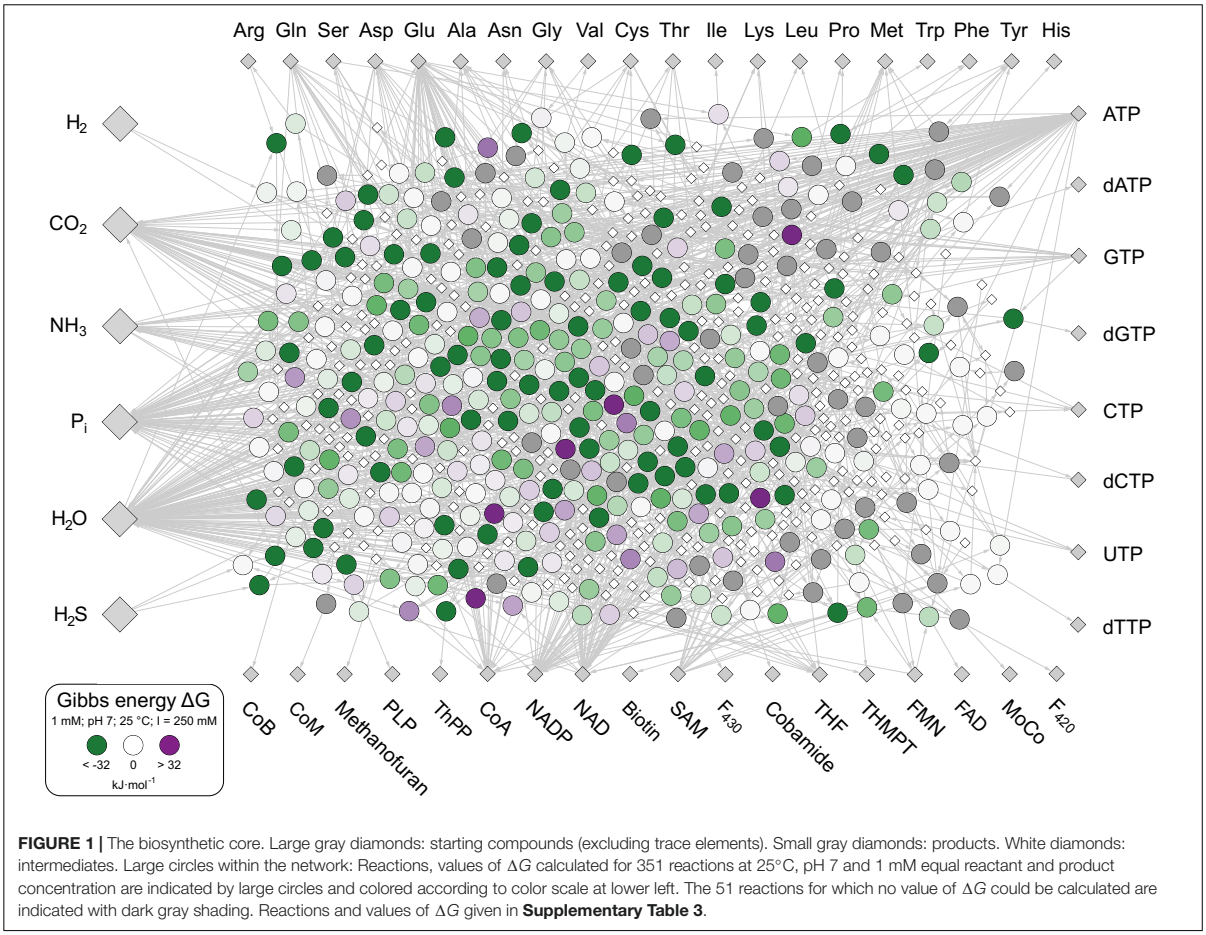

FIGURE 1 | The biosynthetic core. Large gray diamonds: starting compounds (excluding trace elements). Small gray diamonds: products. White diamonds: intermediates. Large circles within the network: Reactions, values of ΔG calculated for 351 reactions at 25°C, pH 7 and 1 mM equal reactant and product concentration are indicated by large circles and colored according to color scale at lower left. The 51 reactions for which no value of ΔG could be calculated are indicated with dark gray shading. Reactions and values of ΔG given in **Supplementary Table 3**.

reaction by measuring the concentrations of reactants and products in the presence of the enzyme. The change in Gibbs energy ΔG for the reaction $A + B \rightleftharpoons C + D$ is obtained from the equation:

$$\Delta G = \Delta G^{\circ\prime} + \mathrm{RT} \ln \frac{[C]\,[D]}{[A]\,[B]} \quad (1)$$

Where R is the gas constant, T is the temperature in Kelvin and [A], [B], [C], and [D] are the molar concentrations (more precisely activities) of reactants and products forming the reaction quotient. ΔG' is the change of free standard enthalpies during reaction in water at physiological pH 7, 25°C, 1 M molar concentrations and 1 atm gas pressure. If H^+ is involved in the reaction, its activity is 1 in eq. (1) at pH 7. If water is involved in the reaction, its activity in eq. (1) is also 1 because ΔG' is obtained from measurements in water and the water concentration in water as the solvent does not change appreciably by reaction water.

At equilibrium, $\Delta G = 0$ (no net driving force and therefore no change of reactant and product concentrations anymore), resulting in:

$$0 = \Delta G^{\circ\prime} + \mathrm{RT} \ln \frac{[C]_{eq}\,[D]_{eq}}{[A]_{eq}\,[B]_{eq}} \quad (2)$$

$$\Delta G^{\circ\prime} = -\,\mathrm{RT} \ln \frac{[C]_{eq}\,[D]_{eq}}{[A]_{eq}\,[B]_{eq}} = -\,\mathrm{RT} \ln K' \quad (3)$$

Therefore, ΔG°' can be obtained from the reactant and product concentrations measured at reaction equilibrium in water at pH 7. K' is the equilibrium constant at pH 7. The increments used in the component contribution method to obtain ΔG°' derive their values from measurements of K' in water, hence the activity of water is already taken into account in ΔG°' and can be set to 1 in the reaction quotient. At physiological conditions, concentrations are generally different from 1 M and eq. (1) with the reaction quotient term is used to calculate the Gibbs energy ΔG'. For clarity, we manually polarized the reactions toward synthesis by writing the KEGG reactions from left to right such that the flux of carbon and nitrogen starts from CO_2 and NH_3 and proceeds within the KEGG pathways in the direction of amino acid, nucleotide and cofactor synthesis. To estimate ΔG

under nonequilibrium conditions, unequal reactant to product concentration ratios were inserted into in the reaction quotient for the polarized reaction, see below.

For many reactions catalogued in large biochemical databases such as KEGG (Kanehisa and Goto, 2000) the equilibrium concentrations are not known or not readily obtained, but the value of $\Delta G'$ can still be estimated using the component contribution method, which is based on the group contribution method originally developed by Benson (1968) to study the equilibria of chemical reactions in the gas phase and later adapted by Alberty (1998) and others to the study of aqueous reactions. It is an indirect method for estimating the position of the equilibrium in a reaction based on the thermodynamic contributions of the moieties in the compounds in question (Jankowski et al., 2008). In this paper we will use $\Delta G'$ to indicate 1 M reactant and product concentrations and 1 bar pressure for gasses at 25°C, in the strict sense. When we refer to conditions that deviate from $\Delta G'$, for example different temperatures or different reactant and/or product concentrations, we use the generic term ΔG, whereby its parameters are then unambiguous by context.

Gibbs energies were calculated using eQuilibrator API (Flamholz et al., 2012; Beber et al., 2021) version 0.4.1 under Python v. 3.6.7 which bases its estimates on the component

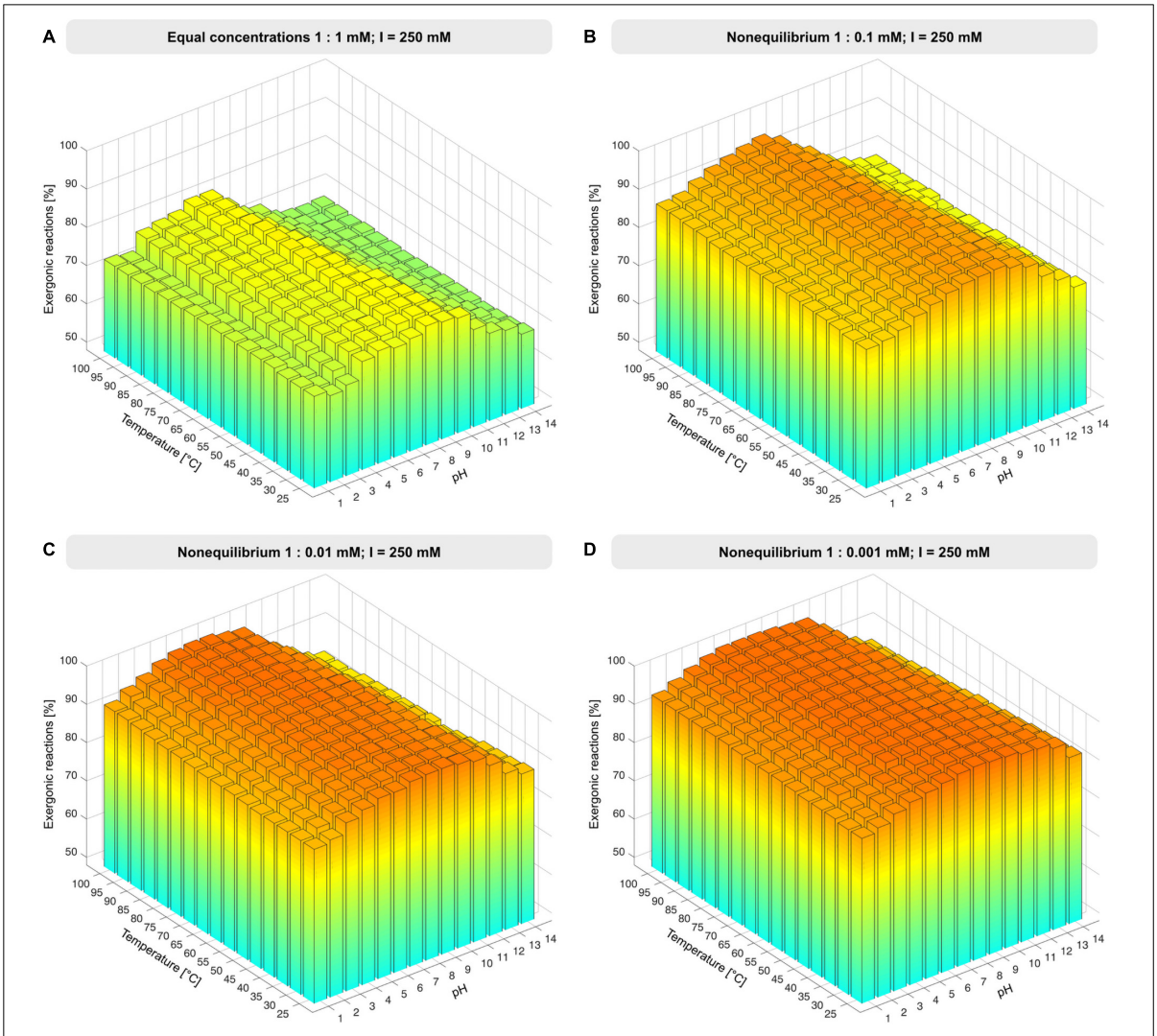

FIGURE 2 | Proportion of exergonic reactions under nonequilibrium conditions. **(A)** At constant ionic strength, I, and reactant concentration of 1 mM, the concentration of products was decreased from 1 mM to 1 μM **(B–D)**, ΔG was calculated for the 351 reactions that deliver a value of ΔG across the pH and temperature range and the proportion of exergonic reactions was plotted for each condition set (see also **Supplementary Figure 1** and **Supplementary Table 3**).

contribution method (Noor et al., 2013). eQuilibrator is widely used in biochemical and genome-based investigations, inter alia because it is capable of operating with reactions and compounds in the KEGG database. To cross check the current set, we compared values obtained using eQuilibrator to those determined by the traditional biochemical method for core carbon metabolism (**Supplementary Table 2**; Fuchs, 2011). As in earlier studies (Alberty, 1998; Jankowski et al., 2008; Flamholz et al., 2012), the agreement was good, usually within a few $kJ{\cdot}mol^{-1}$, indicating that the method delivers useful estimates.

Unless otherwise specified, environmental conditions were simulated by varying the pH from 1 to 14 in increments of 1 and temperature from 25 to 100°C in increments of 5°C at constant ionic strength of 250 mM, Mg^{2+} concentration fixed to 3 mM, and reactant concentrations set to 1 mM. Nonequilibrium conditions were simulated by altering the reactant to product ratio from 1:1 to 1:0.1 mM, 1:0.01 mM, 1:0.001 mM, 1:0.0001 mM and 1:10 mM (**Figure 2**, **Supplementary Figure 1**, and **Supplementary Table 3**). Atomic balancing was checked prior to calculation, such that ΔG was only calculated for balanced reactions, excluding partial reactions. For 351 reactions Gibbs energy was calculable, for the remaining 51 reactions ΔG calculation failed due to involvement of KEGG compounds undefined in the eQuilibrator database, compounds having ambiguous structures, or unbalanced reactions.

Even though the reactions of biosynthetic metabolism are interconnected, we can consider each reaction individually with regard to its change in free energy in the biosynthetic direction, because the value for change of free energy for a given enzymatic reaction results from the physicochemical properties of its reactants and products under the specified conditions as in eq. (2). A directed metabolic network representing the 402 reactions was created in simple interaction format (sif). The bipartite graph was drawn with CytoScape (Shannon et al., 2003) v. 3.8.0. Reaction nodes and compound nodes were labeled as indicated in **Figure 1**.

Substitution of Biochemical Reductants With Hydrogen

To investigate the influence of environmental H_2 in the 73 reactions involving biochemical reductants, NAD(P)H, reduced ferredoxin and reduced flavodoxin were replaced with H_2, generating a reduced product and protons in the balanced equation (reaction equations are given in **Supplementary Table 4**), simulating H_2 as a reductant present in an environmental setting. Ferredoxin:NADH oxidoreductase and ferredoxin reducing hydrogenase reactions were excluded from H_2 substitution because H_2 would have appeared on both sides of the reaction. Gibbs energies were calculated as for the altered set. In the substituted set, two additional reactions (353 total) yielded a value for ΔG, 49 did not. The compound concentration ratio was set to nonequilibrium 1:0.01 mM with fixed H_2 reactant and product concentrations 1 μM, 10 μM, 100 μM, 1 mM, 10 mM, and 100 mM (**Figure 3**, **Supplementary Figure 2**, and **Supplementary Table 4**). The influence of ionic strength, I, was probed by altering I from standard 250 mM to 2.5 mM, 25 mM, 2.5 M, and 0 M under a nonequilibrium concentration ratio of 1:0.01 mM and with H_2 fixed to 1 μM (see **Supplementary Figure 3** and **Supplementary Table 5**). For all calculations, reactions are classified as exergonic if $\Delta G \leq 0$.

Metal Catalyzed NAD^+ Reduction With H_2

NAD^+ solution (3 mM) was prepared in a phosphate buffer at pH 8.5. Both glass reaction vials containing 4 ml NAD^+ solution (no catalyst) and vials containing 4 ml NAD^+ solution and nickel (Alfa Aesar) and iron powder (Alfa Aesar) as solid phase catalysts, added as 26 mg Fe plus 28 mg Ni powder per ml solution, were placed in a stainless-steel reactor (Berghof). The vials were closed with PTFE septum lids which were penetrated with syringe needles (Sterican) to ensure the reaction gas could enter the vials. The closed reactor was pressurized with 5 bar of hydrogen gas and heated up to 40°C for a total of 4 h. After depressurizing the reactor, samples were transferred to 2 ml Eppendorf tubes, centrifuged for 15 min at 13,000 rpm (Biofuge fresco, Heraeus) and the supernatant was collected to spectrophotometrically observe NADH synthesis (characteristic maximum absorbance at 339 nm; Cary 3500 UV-Vis, Agilent) (see **Supplementary Figure 4**). For convenience, conversion tables relating H_2 partial pressures and H_2 concentrations in water at different temperatures are given in **Supplementary Table 6**.

Energetics of Amino Acid Synthesis

Energetics of synthesis pathways for the 20 canonical amino acids consisting of KEGG reactions starting from key intermediates pyruvate, oxalacetate, 2-oxoglutarate, phosphoenolpyruvate, 3-phosphoglycerate, and C5 sugars (Martin, 2020; **Supplementary Table 7** and **Figure 4**) were analyzed. The pathways, when expressed as linear sets of reactions, are detached from the biosynthetic core network by the removal of edges. Alternative pathway branches and reactions are indicated by numbers, for example 2.1 corresponds to the first reaction in the second pathway alternative. Gibbs energies for 1 mM reactant and product concentrations, pH 7 and 25°C are given in **Supplementary Table 3** and for vent-like conditions (nonequilibrium 1:0.01 mM, pH 9 and 80°C) in **Supplementary Table 4**.

RESULTS

Thermodynamics in the Metabolism of the Last Universal Common Ancestor

Theories of autotrophic origin posit that the first free living cells grew from CO_2 and inorganic compounds without the help of light (Mereschkowsky, 1910; Fuchs and Stupperich, 1985; Wächtershäuser, 1992; Fuchs, 2011). For such chemolithoautotrophic cells to arise at a specific environmental site, the reactions underpinning their origin, that is, the overall

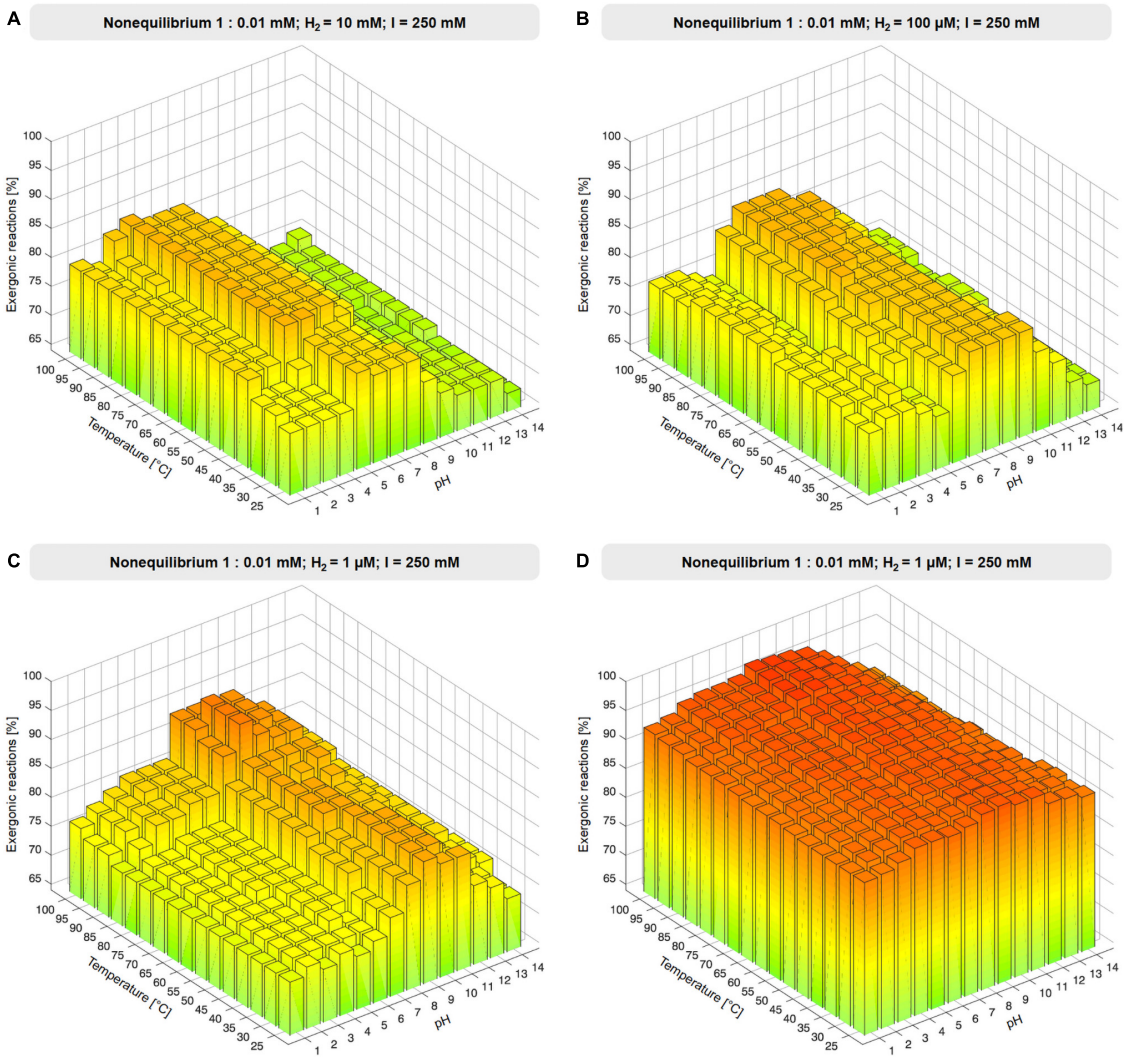

FIGURE 3 | Effect of environmental H_2. **(A–C)** Proportion of exergonic reactions at 1 µM, 100 µM, 10 mM H_2 and 1:0.01 nonequilibrium at constant ionic strength I expressed as % of the 67 redox reactions of the core that yield values of ΔG. **(D)** Proportion of exergonic reactions at 1 µM H_2 and 1:0.01 nonequilibrium expressed as % of 353 reactions for which values of ΔG were obtained (see also **Supplementary Figure 3** and **Supplementary Table 4**).

set of reactions that synthesize the cell needs to be exergonic and no individual reaction should be so endergonic as to block the reaction network under physiological conditions. The source of energy that allows those reactions to go forward is of interest here. The synthesis of the amino acids, nucleotides and cofactors germane to life from H_2, CO_2, NH_3, H_2S, and P_i requires only 402 reactions (Wimmer et al., 2021a; **Supplementary Table 1**) which are listed in KEGG (Kanehisa and Goto, 2000). We polarized those reactions so that carbon flux through each reaction proceeds from H_2 and CO_2 in the direction of monomer synthesis. We then employed the component contribution method (Noor et al., 2013) to estimate the change in Gibbs energy, ΔG, for the 402 reactions in the biosynthetic direction (see section "Materials and Methods").

The set of 402 polarized reactions in KEGG format contained 51 entries that yielded no value of ΔG because one or more reactants are poorly defined or have ambiguous structures, that is, they were not among the underlying data with which eQuilibrator works (see section "Materials and Methods"). The remaining 351 reactions yield thermodynamic estimates, providing a very broad sample for changes in ΔG, covering 87% of reactions in the core (see **Supplementary Table 3**). We started with the simple case of all reactants (compounds on the left side of reactions) and products (right side) at 1 mM concentration, a value well

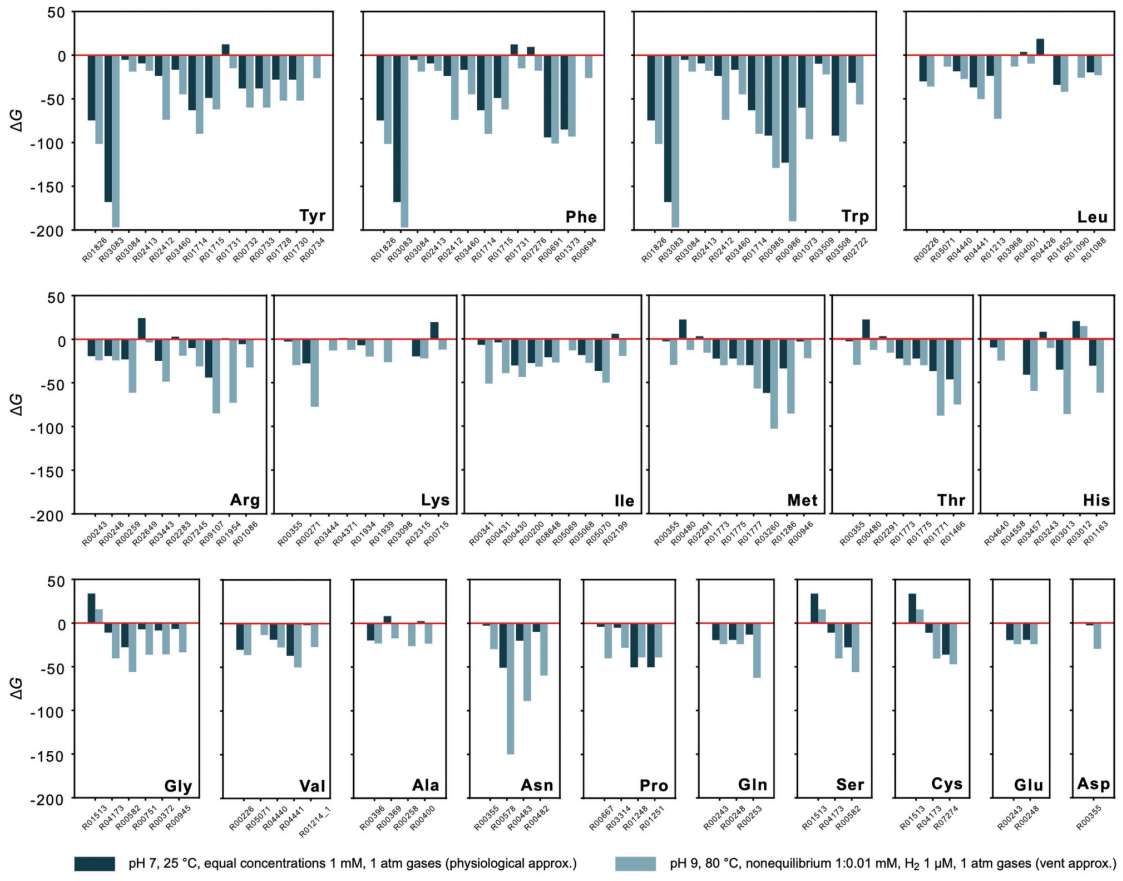

FIGURE 4 | Energetics of amino acid synthesis. Gibbs energies ΔG were calculated for reactions involved in the synthesis of the 20 canonical amino acids. Only reactions yielding a value for ΔG are plotted. Values of ΔG are indicated for physiological (dark blue) and vent conditions (light blue) as defined in the figure. Underlying synthesis pathways are listed in **Supplementary Table 7**. Only pathways starting from key intermediates pyruvate, oxalacetate, 2-oxoglutarate, phosphoenolpyruvate, 3-phosphoglycerate, and C5 sugars (as proposed in Martin, 2020) are examined. Additional synthesis alternatives from KEGG are not taken into consideration.

within the 1 μM to 10 mM range of metabolite concentrations in *Escherichia coli* during exponential growth (Bennett et al., 2009) to examine the effect of pH and temperature regarding metabolic origins under hot (Stetter, 2006) vs. cold (Miyakawa et al., 2002) or acidic (Wächtershäuser, 1988) vs. alkaline (Martin and Russell, 2007) conditions. Roughly 77% of core reactions are exergonic at pH 6-7, with temperature exerting little effect (**Figure 2A**). Note that the component contribution method does not obtain values for ΔG as a function of temperature, and that temperature effects are considered by the reaction quotient (see eq. (1) in section "Materials and Methods").

Nonequilibrium Conditions

Metabolism in cells is a connected series of far from equilibrium reactions in which reactants continuously react to products at every step (Decker et al., 1970; Battley, 1987; Dai and Locasale, 2018), whereby the products of one reaction become the reactants of the next in succession. As it concerns calculations of thermodynamic values, this presents a stark difference to geochemical thermodynamics, where one step reactions are the rule, for example balanced single step reactions for the synthesis of amino acids from H_2, CO_2, and NH_3 (Amend and Shock, 1998, 2001). In the context of metabolic origin, the process to model concerns a situation in which compounds supplied by the environment (H_2, CO_2, and NH_3 for example) react to generate products that do not initially exist (Martin and Russell, 2007), such as formate and pyruvate (Preiner et al., 2020) or amino acids. In a hydrothermal vent context, such compounds can either react further, or be eluted from their site of synthesis via hydrothermal effluent by convection and/or thermal or concentration diffusion. In cells, the products can either react further, or be excreted as an end product, generating steady state equilibrium (German: *Fließgleichgewicht*), or they can be converted to biological polymers—proteins, sugars, nucleic

acids—exiting the metabolic network as cell mass. In acetogens, for example, roughly 24 molecules of CO_2 are converted to acetate as an end product for every atom of carbon that is incorporated into cell mass (Daniel et al., 1990). We designate the situation of higher reactant concentrations relative to product concentrations as nonequilibrium conditions.

When examined using the component contribution method, the effect of nonequilibrium conditions is large. Increasing the product concentration 10-fold relative to reactant concentrations renders most reactions of the core endergonic (**Supplementary Figure 1B**). This is because many reactions in metabolism are close to equilibrium in terms of ΔG, with every 10-fold reduction in product concentration relative to reactant concentration corresponding to a change in ΔG of –5.7 kJ·mol^{-1} at 25°C (Walsh et al., 2018) for reactions having equal stoichiometric coefficients of reactants and products. Increasing product concentrations shows that the reactions of the core have little tendency to run backward (**Supplementary Figure 1B**), which is in line with the concept of autotrophic origins (Fuchs, 2011).

Lowering the concentration of products relative to reactants approximates the situation in an environmental setting in which H_2, CO_2, H_2S, NH_3, and phosphate (**Figure 1**) are continuously supplied in roughly constant amounts, while the products of reactions are allowed to react further or removed by flow processes. To model nonequilibrium conditions, we reduced the product concentrations in steps of 10-fold change relative to reactants (**Figures 2B-D** and **Supplementary Figure 1**). At 100-fold less product than reactant, 98% of core reactions become exergonic (**Figure 2C**), with marginal increase at higher ratios and no marked effect of temperature except at very high pH.

Regardless of the specific environment within which LUCA arose, the reactions fueling the synthesis of its building blocks underwent a transition during the origin of metabolism: Reactions that were initially either uncatalyzed or catalyzed by substances in the environment eventually came to be catalyzed by cofactors and enzymes encoded by genes. During that transition, it is possible, and cannot be excluded, that some or many of the chemical reactions themselves might have changed. But it is also possible, and cannot be excluded, that the reaction set remained essentially the same, as in the example of the acetyl-CoA pathway (Preiner et al., 2020) and reverse citric acid cycle (Muchowska et al., 2020). In that case, only the nature of the catalysts changed from inorganic to organic, adding specificity and rate to preexisting reactions that tend to occur anyway.

Because the core constitutes a minimal set of enzymatic reactions required for the synthesis of amino acids, nucleotides and cofactors, it contains neither a rotor stator ATPase, nor cytochromes, quinones, or even membrane-associated reactions. Although the rotor stator ATPase is as universal in cells as the ribosome itself, and was present in LUCA (Weiss et al., 2016), is not essential for the biosynthetic core to operate. Net ATP synthesis can be derived within the core from substrate level phosphorylation via acetate synthesis from H_2 and CO_2 in soluble reactions, similar to the situation of *Methanosarcina mazei* growing on CO (Rother and Metcalf, 2004). Also note that we are considering each reaction individually, not as a system of interconnected reactions set in series, in which case reactant concentrations would approach zero under nonequilibrium conditions. We are not querying the extent to which the overall balanced one-step reactions from H_2, CO_2, and NH_3 to the individual amino acids, bases and cofactors are energy releasing, which for amino acids and nucleotides is known to be the case under the conditions of H_2-producing hydrothermal vents interfacing with ocean water (Amend and McCollom, 2009). Instead, we are investigating the exergonic nature of the individual reactions in LUCA's biosynthetic pathways, as they are manifest in modern enzymatic reactions, which are intensely interconnected in a metabolic network (**Figure 1**), applying the same concentration gradient to each, so that the individual chemical reactions underlying energy release within the network, as opposed to energy release for the network as a whole as in the energetics of growth (Battley, 1987; Hansen et al., 2009), can be identified.

The finding that 98% of the reactions in the core that deliver a value of ΔG using the component contribution method are exergonic under nonequilibrium conditions starting from H_2 and CO_2, with 100-fold less product than substrate, is noteworthy. It also reminds us that the reactions of metabolism as they operate in modern cells are generally exergonic, otherwise metabolism would not run. Yet even with equal substrate and product concentrations, on average 78% of the reactions in the core are exergonic under the conditions sampled here (**Figure 5**). As an caveat, many enzymatic reactions in the core might not go forward under prebiotic conditions for lack of suitable catalysts, for reasons of inhibitory inorganic compounds, due to substrate sequestration on surfaces, or for other reasons. Favorable thermodynamics are thus a necessary but not sufficient condition for the emergence of metabolism. We also note that our study addresses only monomer synthesis, not polymerization reactions. Notwithstanding, the present findings indicate that there is a natural thermodynamic tendency for the reactions of LUCA's biosynthetic network to unfold from H_2, CO_2, NH_3, H_2O, and P_i. This is not self-evident, because it introduces the possibility that the energy needed at the origin of metabolism simply stemmed from within metabolism itself, as opposed to some external source.

The Effect of Environmental H_2

At the very onset of the process that gave rise to LUCA's metabolism, it is reasonable to assume there were no preformed organic redox cofactors in supply in the environment, as these are products of organic synthesis. Microbiologists have, however, long held that reduced low potential FeS centers such as those in ferredoxin were the source of reducing power in the early stages of biochemical evolution (Eck and Dayhoff, 1966; Hall et al., 1971). In line with that view, all hydrogenases in modern chemolithoautotrophs that use H_2 as a reductant reduce FeS clusters, with only one known exception, the Fe hydrogenase of methanogens that transfers electrons from the active site of the iron-guanylylpyridinol (FeGP) cofactor directly to F_{420}, generating $F_{420}H_2$ without the involvement of FeS or other intermediate electron carriers (Huang et al., 2020).

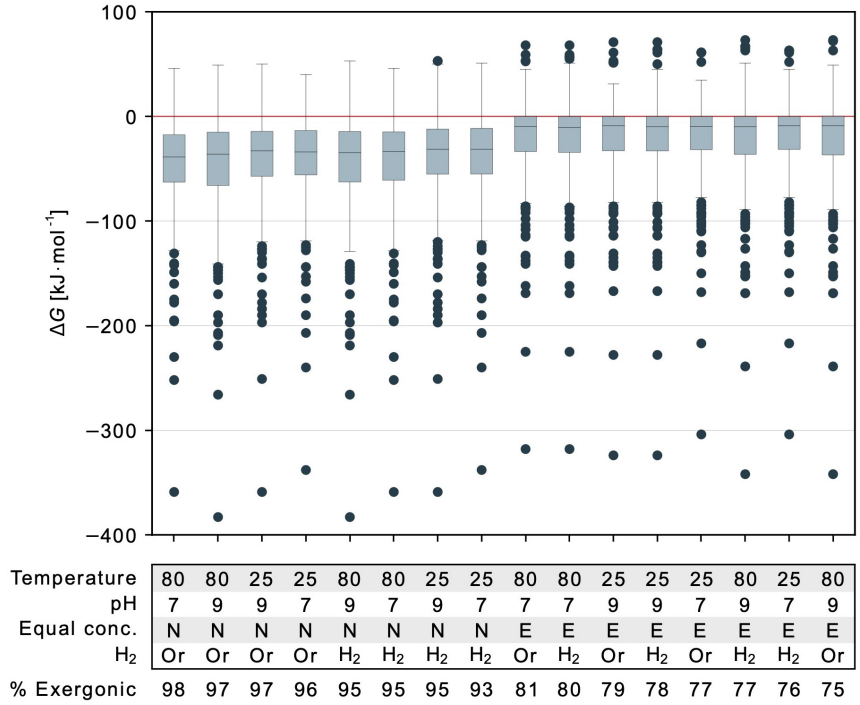

FIGURE 5 | Energetic effect of temperature, pH, reductants and reactant to product concentration ratios. Gibbs energy is indicated for 351 calculable reactions that deliver a value of ΔG for different parameter combinations. Values for each parameter of each calculation are specified in the table below the figure and varied with respect to physiological and vent conditions. Temperature is described in degree Celsius. Equal conc.: *E* indicates all concentrations set to 1 mM, *N* indicates nonequilibrium 1:0.01 mM reactant to product ratio. Retention of organic reductants (NADH, NADPH, flavodoxin$_{red}$, ferredoxin$_{red}$) is indicated as *Or* while the replacement of these organic reductants with hydrogen is marked by H_2. Note that in this case, two additional reactions yield a value for ΔG (353 reactions). Proportions of exergonic reactions ($\Delta G \leq 0$ kJ·mol^{-1}) across the biosynthetic core are listed below the table. In each boxplot, the horizontal line indicates the median ΔG among calculable reactions. The colored boxes represent the interquartile range (IQR) with ΔG within quartile 1 (Q1, median of lower half of the data) and quartile 3 (Q3, median of upper half of the data). The range bars mark the minimum (Q1 – 1.5·IQR) and maximum (Q3 + 1.5·IQR) value of the data excluding any outliers. Outliers are indicated by individual dots and do not fall into the defined range between minimum and maximum.

Hydrogen gas is also the source of electrons for chemolithoautotrophic archaea and bacteria that synthesize ATP by reducing CO_2 (Thauer et al., 1977; Fuchs, 2011; Schuchmann and Müller, 2014; Preiner et al., 2020). In modern geological environments that generate abiotic hydrogen (Schrenk et al., 2013), H_2 is synthesized in amounts that generate midpoint potentials on the order of –700 to –900 mV (Boyd et al., 2020), more than sufficient to substitute for known biochemical reductants such as NAD(P)H or reduced ferredoxins (**Supplementary Table 6**). The very low midpoint potentials come from an interplay of two factors: serpentinization generates H_2 in a geochemical process that also generates metal hydroxides such as $Mg(OH)_2$, which in turn generate alkalinity. Alkaline solutions foster the release of protons from H_2 via heterolytic cleavage, leading to the release of electrons onto suitable acceptors. Some modern microbes that inhabit such H_2-rich alkaline environments even appear to lack known hydrogenase enzymes (Suzuki et al., 2018), suggesting that there might be alternative or bypass entry points for H_2 into their metabolism. To investigate the effect of environmental redox potential on the thermodynamics of the biosynthetic core, we replaced biological reductants by the environmental source of electrons in CO_2-reducing autotrophs, H_2, in all reactions of the core. This captures the thermodynamic effect of an environmental redox buffer, but entails the premise that mineral catalysts naturally occurring in hydrothermal vents (Fontecilla-Camps, 2019) can readily replace hydrogenases and ferredoxin to reduce the main biochemical hydride carrier, $NAD(P)^+$, with H_2. To that end, we tested H_2-dependent $NAD(P)^+$ reduction in the laboratory using simple transition mineral catalysts. The reaction is facile under hydrothermal conditions (**Supplementary Figure 4**).

As an environmental parameter, H_2 reactant and product concentrations must be equal. This impacts redox reactions of the core under nonequilibrium conditions. The core encompasses 73 redox reactions involving NAD(P)H, flavins, or ferredoxin. Reduced cofactors occur on the left in 48 reactions and on the right in 27 (**Table 1**). We replaced biochemical reductants on both sides of the reactions with H_2 at concentrations corresponding to an E_0 of –600 to –800 mV at pH 10 around

TABLE 1 | Most frequent reactants and products in the core.

Compound	Reactant						Product			
	Left	Right	Physiological[a]		Vent[b]		Physiological[a]		Vent[b]	
			$\Delta G \leq 0$	$\Delta G > 0$	$\Delta G \leq 0$	$\Delta G > 0$	$\Delta G \leq 0$	$\Delta G > 0$	$\Delta G \leq 0$	$\Delta G > 0$
H_2O	71	49	51	15	64	2	35	10	43	2
H^+	48	31	33	7	37	3	18	10	26	2
ATP	77	2	55	14	69	0	2	0	2	0
P_i	2	64	0	2	2	0	54	5	59	0
ADP	3	54	3	0	3	0	36	12	48	0
CO_2	12	37	3	8	10	1	34	0	34	0
Glutamate	26	18	14	10	23	1	16	1	17	0
NAD^+	21	16	12	9	20	1	16	0	15	1
$NADP^+$	6	30	5	1	5	1	20	3	22	1
PP_i	0	36	0	0	0	0	30	3	33	0
NADPH	29	6	19	3	21	1	5	1	5	1
NADH	14	19	14	0	13	1	10	9	18	1
2-Oxoglutarate	6	19	4	2	6	0	7	11	17	1
Pyruvate	12	10	9	3	12	0	8	1	9	0
AMP	2	19	1	0	1	0	13	2	15	0
NH_3	13	7	12	1	13	0	6	0	6	0
CoA	4	17	4	0	4	0	8	7	15	0
SAM	16	1	9	0	9	0	1	0	1	0
Acetyl-CoA	13	3	8	4	12	0	3	0	3	0
Glutamine	14	1	13	0	13	0	1	0	1	0

Reactant and product frequency for each compound of the 402 core reactions is given and further classified into participation in exergonic/endergonic reactions for physiological and vent conditions. Number of occurrences on the left or right side of reactions can exceed numbers of reactions for which values of ΔG are obtained.
[a]Physiological condition is defined as pH 7 at 25°C and concentrations at 1 mM.
[b]Vent condition is defined as pH 9 at 80°C and nonequilibrium 1:0.01 mM reactant to product concentrations.

100°C (**Supplementary Table 6**). For the reactant:product ratio of 1:0.01 mM under nonequilibrium conditions, we adjusted the H_2 concentration on both sides of the reaction from 1 μM to 100 mM (**Figure 3**, **Supplementary Figure 2**, and **Supplementary Table 4**). For reactions in which H_2 is a reactant, large H_2 concentrations are favorable, whereas for reactions where H_2 is product, low H_2 concentrations are favorable. *Vice versa*, for reactions where H_2 is reactant and H^+ is product, high pH is favorable.

The effect of H_2 across the core is substantial, with 337–342 out of 353 (95–97%) of core reactions that deliver a value for ΔG being exergonic ($\Delta G \leq 0$ kJ·mol^{-1}) under nonequilibrium conditions with H_2 at 1 μM, 80–100°C, and pH 7–10. Under these conditions, only 12 out of 353 core reactions are endergonic by ≥5 kJ·mol^{-1} (**Supplementary Table 8**). It is noteworthy that alkalinity impacts the thermodynamics of metabolic origin because it strongly affects the electron donating potential of H_2 (**Supplementary Table 6**). Modern geochemical systems synthesize formate (Lang et al., 2018) and acetate (Sherwood Lollar et al., 2021) in abiotic reactions that blueprint the CO_2-fixing reactions of microbes.

Are the conditions that we are investigating realistic in a primordial geochemical context? We have investigated the temperature range 25 to 100°C, the pH range 1–14, and H_2 concentrations from 1 μM to 100 mM. Those ranges span conditions existing today at the serpentinizing Lost City hydrothermal field, where the temperature range is 40–90°C, the pH is 9–11, and H_2 concentrations range from 1 to 15 mM (Kelley et al., 2005). Are such conditions primordial? Serpentinizing systems have existed since there was liquid water on earth (Sleep et al., 2011). We observed a tendency for the largest proportion of reactions to be exergonic around pH 9, 80°C and at low H_2 concentrations, very much in line with, but not constrained by, modern conditions at Lost City, which provide a window into conditions on the early Earth (Sleep et al., 2011; Schrenk et al., 2013). We allowed the concentration of CO_2 to vary freely across analyses, having a substrate concentration of 1 mM under nonequilibrium conditions. In natural environments, CO_2 and bicarbonate concentrations vary across extremes. While Lost City itself has very low inorganic carbon and CO_2, Kelley et al. (2002) report CO_2 concentrations in vent fluids from 3 to 215 mM, while modern sea water contains roughly 11 μM CO_2 and 2 mM HCO_3^-, some modern hydrothermal systems emit pure CO_2 gas (Steffens et al., 2021) and other submarine hydrothermal vents emit pure, supercritical CO_2 as bubbles at 1.4 km depth and high pressure (Zhang et al., 2020). On the early earth, global CO_2 levels were generally very high (Zahnle et al., 2007; Sossi et al., 2020), but local CO_2 concentrations might have varied as much as they do in modern environments. In general, submarine hydrothermal systems exist under very high pressure and therefore allow gasses to dissolve up to very high concentrations, today and on the early earth.

In laboratory scale model vents (Preiner et al., 2020), a 10 bar partial pressure of H_2 at pH 9 and 100°C corresponding to 5 mM H_2 in solution (**Supplementary Table 6**) and within the range of 1–15 mM H_2 concentrations observed at Lost City (Kelley et al., 2005), will reduce CO_2 to formate, acetate and pyruvate, although much lower H_2 partial pressures will also suffice for the same reaction (Preiner et al., 2020). That is, geochemical H_2 and CO_2 spontaneously generate central compounds of autotrophic metabolism in the acetyl-CoA pathway in the presence of metal catalysts (Preiner et al., 2020). This is noteworthy not only because of the congruence between the products of the abiotic and enzymatic products from H_2 and CO_2 but also because earlier studies of H_2-dependent CO_2 reduction under higher pressures and temperatures, but performed in inert titanium reactors in the absence of catalysts, did not detect the synthesis of either acetate or pyruvate among the products (McCollom and Seewald, 2003), whereas inclusion of iron or nickel, either as native metal or as oxide or sulfide minerals effectively replace the pathway to pyruvate, yielding physiologically relevant concentrations of pyruvate (∼40 μM) overnight. From the outset of the first abiotic reactions to the origin of an enzymatically catalyzed metabolism in LUCA, redox reactions were integral to metabolic origin, whereby H_2 provided an ample and biochemically accessible supply of electrons throughout that process, particularly under the alkaline conditions of serpentinization (Preiner et al., 2019).

Water

Views concerning the role of water at origins differ widely. One view has it that water is inhibitory at the origin of life because reactions that generate water, in particular polymerization reactions, proceed against the pushback of a 55 M product (Marshall, 2020). The other view is that water is essential to origins because it is both the solvent of all molecules of life and the most common reactant in microbial metabolic networks (do Nascimento Vieira et al., 2020). An underappreciated aspect of hydrothermal systems is that they harbor abundant local microenvironments of low water activity (Lamadrid et al., 2017). The serpentinization process that underpins the formation of H_2 for CO_2 reduction at metabolic origin entails rock-water interactions that consume about 20 molecules of H_2O per molecule of H_2 formed and about 100 molecules of H_2O per molecule of abiotic methane formed from CO_2 (Preiner et al., 2018). In the present calculations, water concentration is fixed at 55 M and cannot be changed in these calculations (Alberty, 1998; Flamholz et al., 2012). Water is furthermore the most common compound in the reactions of the core, appearing in 120 reactions, 97% of which are exergonic regardless of whether water is consumed or produced against the 55 M gradient (**Table 1**). From the thermodynamic perspective H_2O exerts no inhibitory effect upon the reactions of core biosynthesis. That, and the frequency of water as a reactant (**Table 1**) suggest that the reactions that gave rise to LUCA's metabolism arose in an aqueous environment, a premise preferable to the proposition that the chemistry of life began in non-aqueous environments, and only later transformed *en masse* into the aqueous reactions of the cytosol.

Salt

Salt concentrations differ in marine vs. freshwater origin environments, and some origin of life theories posit that life arose in freshwater environments based on arguments relating to concentrations of K^+ (Korolev, 2021) as opposed to arguments based upon reactions of carbon (Preiner et al., 2020). Seawater has an ionic strength of ca. 700 mM, while cytosol has a variable ionic strength on the order of 20–900 mM in *E. coli* (Richey et al., 1987) but exceeding 2,000 mM in some archaea (Ginzburg et al., 1970). Hydrothermal effluent has an ionic strength on the order of 20-800 mM (Kelley et al., 2002). Across the range from 0 to 2.5 M, ionic strength has very little impact on ΔG of core reactions as estimated by implementation of the component contribution method employed here (see **Supplementary Figure 3** and **Supplementary Table 5**).

Nonequilibrium Conditions Have a Pronounced but Not a Dominant Effect

Using eQuilibrator (Noor et al., 2013), water activity cannot be perturbed but is already taken into account in ΔG. The effect of ionic strength was small (**Supplementary Figure 3** and **Supplementary Table 5**). To compare the effects for parameters investigated here that did show effects, we plotted the mean and range of values of ΔG for comparison of temperature (25°C vs. 80°C), pH (7 vs. 9), nonequilibrium vs. equal reactant and product concentrations, and organic reductants vs. H_2. Nonequilibrium conditions have the most pronounced effect across reactions of the core (**Figure 5**). But even for conditions of 1 mM reactant and product concentrations, the mean of the 351 reactions that deliver an estimate of ΔG is still negative. For reactions that are only slightly endergonic, the effect of nonequilibrium conditions can render the value of ΔG negative (**Figure 5**).

Though nonequilibrium conditions have a pronounced effect, they do not fundamentally distort the picture for individual reactions. This is shown in **Figure 4**, where the estimate of ΔG for amino acid synthesis is compared for physiological conditions (with 1 mM reactant and product concentrations, pH 7, 25°C, gasses at 1 atm) vs. conditions more similar to those in serpentinizing hydrothermal systems (nonequilibrium with 1:0.01 mM concentrations, 1 μM of H_2 instead of organic reductants, other gasses at 1 atm) for 111 reactions of amino acid metabolism starting from the key intermediates for the biosynthesis of the families of amino acids: pyruvate, oxalacetate, 2-oxoglutarate, phosphoenolpyruvate, 3 phosphoglycerate and C5 sugars (**Supplementary Table 7**). The main effect is observed for reactions that are close to equilibrium ($\Delta G \approx 0$) to begin with. This indicates that there is a natural tendency for the individual reactions of amino acid metabolism from H_2, CO_2 and NH_3 in the core to go forward both in physiological and vent conditions, a finding that does not follow from calculations of one-step amino acid syntheses from the same reactants (Amend and Shock, 1998; Amend et al., 2013). It is also important because amino acids are essential sources of C and N for the biosynthesis of bases and cofactors (Wimmer et al., 2021a). Note that the reactions in **Figure 4** correspond to KEGG reactions

and are detached from the overall metabolic network, such that the products of an upstream reaction do not necessarily generate all of the reactants required for the subsequent reaction. Despite that caveat, the general exergonic nature of the individual reactions is evident.

Phosphate

Phosphate is a component of many of the metabolic energy currencies. It forms high energy bonds which are cleaved in exergonic reactions that, when enzymatically coupled to endergonic reactions of metabolism, allow the latter to go forward. The entry of phosphate into metabolism is a heavily debated topic. One view has it that high energy phosphorous minerals reacted with inert carbon compounds (Pasek, 2020), another view has it that inert phosphate reacted with highly reactive carbonyl groups (Martin, 2020), yet another view, based on computer simulations, is that simple protometabolic networks might have been possible without phosphate (Goldford et al., 2017, 2019), though subsequent work identified contrary effects (Tian et al., 2019). In the conserved core of microbial metabolism, LUCA's metabolism, phosphate is indispensable. Of 402 core reactions, 260 (65%) involve phosphate or phosphorylated compounds. For comparison, 83% of the core reactions contain nitrogen. Moreover, 80 reactions (20%) involve ATP hydrolysis or phosphoanhydride hydrolysis of other nucleoside triphosphates in the biosynthetic direction. Among those NTP hydrolyzing reactions, 26 generate ADP and P_i, 10 generate AMP and PP_i, while 33 generate phosphorylated products (**Supplementary Table 9**). Furthermore, all of the cofactors that generate amino acids, bases and cofactors themselves, except biotin, contain phosphate. There can be no question that the biosynthetic core as it existed in LUCA had phosphate inextricably hard wired into its fabric.

That phosphate was part of the core and LUCA's metabolism seems difficult to debate, but how did it enter the core? Net ATP synthesis in the core is afforded by substrate level phosphorylation involving acetyl phosphate via acetyl-CoA (Ferry and House, 2006; Martin and Russell, 2007). ATP is synthesized there by energy conserving reactions that, like thioester synthesis (Huber and Wächtershäuser, 1997; Kitadai et al., 2021), can proceed without enzymes (Sousa et al., 2018; Whicher et al., 2018). Under nonequilibrium vent conditions, the reaction of acetyl-CoA with P_i—the reaction of phosphate with carbonyl groups—to yield acetyl phosphate is exergonic by –18.6 kJ·mol^{-1}, the subsequent reaction of acetyl phosphate with ADP to yield ATP and acetate is exergonic by –40.1 kJ·mol^{-1} (**Supplementary Table 8**). The energetics of acetyl-CoA synthesis from H_2, CO_2, and coenzyme A are, however, strongly dependent upon the H_2 partial pressure (Fuchs, 2011). Under nonequilibrium conditions, the reaction is endergonic by +37 kJ·mol^{-1} at 1 μM H_2 and pH 9 but at 1 mM H_2 it becomes exergonic by –44 kJ·mol^{-1}. This crucial CO_2 activating reaction requires H_2 partial pressures corresponding to potentials on the order of –660 mV at metabolic origin, which abound in natural H_2-producing vents (Boyd et al., 2020). At pH 9 and 100°C, –660 mV corresponds to ca. 1 atm H_2 or 10^5 Pa H_2 or 560 μM H_2 (**Supplementary Table 6**), less H_2 than is found in serpentinizing systems, which contain typically 1 mM H_2 or more, with 1–15 mmol H_2 per kg aqueous effluent observed in the case of Lost City (Kelley et al., 2005).

Of the 351 core reactions that deliver a value of ΔG, 80 involve hydrolysis of anhydride bonds in ATP or other triphosphates as an energy currency (**Supplementary Table 9**). None of the reactions in the core utilize pyrophosphate (PP_i) as an energy source, but 36 reactions generate PP_i from nucleoside triphosphates (**Table 1**). In contrast to many traditional views, PP_i was not a source of energy in early metabolism (Wimmer et al., 2021b). If we subtract the contribution of phosphoanhydride hydrolysis from those 80 reactions, 63 become endergonic by more than 20 kJ·mol^{-1} (**Supplementary Table 9**), a very steep energetic barrier, even under nonequilibrium vent conditions. High energy phosphate bonds are thus essential integral components of the core, apparently as old as metabolism itself and likely the result of inert phosphate reacting with carbonyl groups generated as intermediates of CO_2 reduction. The pressing question remains, however: What is the driving force behind ∼75% of the core reactions that are exergonic independent of ATP?

The Dark Energy at Origins Resides in Carbon

Because our starting compounds are H_2, CO_2, NH_3, H_2S, H_2O, and P_i (**Figure 1** and **Supplementary Table 1**), because no other sources of energy are introduced into the system, and because no N–N or O–O bonds are formed in the core, reactions of carbon are the only candidate for the source of free energy change in core reactions without ATP. We identified 10 organic reaction types that together account for half of ATP-independent exergonic reactions (**Table 2**). Among the 351 reactions that deliver values of ΔG, 10 involve S-adenosylmethionine dependent alkyl transfers ($\Delta G'$ = −24 kJ·mol^{-1}; Lewis and Wolfenden, 2018). Six reactions involve folate dependent alkyl transfers ($\Delta G'$ = −30 kJ·mol^{-1}; Thauer et al., 1977) or acyl transfers ($\Delta G'$ = −26 kJ·mol^{-1}; Decker et al., 1970). Acyl thiol ester (thioester) hydrolysis ($\Delta G'$ = −32 kJ·mol^{-1}; Buckel and Eggerer, 1965) drives 14 reactions and acyl phosphate hydrolysis ($\Delta G'$ = −45 kJ·mol^{-1}; Decker et al., 1970) drives four reactions (**Supplementary Table 10**).

The only input compound that is reduced in reactions of the core is CO_2 (**Figure 1**). For the 44 reactions involving reductions of carbon with reduced nicotinamide, flavin, ferredoxin or formate, reactions that are exergonic under physiological conditions (Decker et al., 1970; Thauer et al., 1977), the average ΔG in the core is –28 kJ·mol^{-1} under nonequilibrium 1:0.01 mM conditions at 80°C and pH 9 (**Table 2**). Decarboxylations, with a $\Delta G°'$ on the order of –20 kJ·mol^{-1} (Dimroth and Schink, 1998) occur in 30 reactions, 10 of which are oxidative decarboxylations (**Supplementary Table 10**). In addition, many reactions of the core generate aromatics from non-aromatic substrates. Aromaticity entails very large changes in ΔG, on the order of –60 to –150 kJ·mol^{-1} or more (Morrison and Boyd, 1977). The amino acids, bases and cofactors produced by the core involve the synthesis of 31 aromatic rings and 35 ring closure reactions (Goldberg and Tewari, 1989) that are involved in their formation.

TABLE 2 | Energy release in the core.

Gibbs energy changes for exergonic carbon based reactions in the core			
Reaction	***N***	**ΔG^a**	**References**
Pyruvate formation from H_2+CO_2	1	−57	Preiner et al. (2020)
Ring formations	35	−10 to −25	Goldberg and Tewari (1989)
Decarboxylations	30	−20	Dimroth and Schink (1998)
SAM dependent alkyl transfers	10	−24	Lewis and Wolfenden (2018)
Folate dependent acyl transfers	2	−26	Decker et al. (1970)[b]
Reductions	44	−28	
Folate dependent alkyl transfers	4	−30	Thauer et al. (2008)
Acyl thiol ester hydrolyses	14	−32	Buckel and Eggerer (1965)
Acyl phosphate hydrolyses	4	−45	Decker et al. (1970)
Aromatic formation	31	−60 to −150	Morrison and Boyd (1977)

Estimated ΔG for different reaction types. N is the number of reactions among 351 reactions in the core for which values of ΔG are obtained.
[a] ΔG [kJ·mol^{-1}] as given in references.
[b] Average ΔG for 44 NAD(P) H-, ferredoxin-, and formate-dependent reductions in the core under conditions specified in ***Supplementary Table 10****.*

Including the exergonic synthesis of pyruvate from H_2 and carbon dioxide (Preiner et al., 2020), these sources of carbon-based energy (**Table 2**) contribute to favorable thermodynamics in 50% of core reactions (175/351), more than twice the number of reactions (80/351) driven by ATP hydrolysis, though sometimes with a smaller contribution to ΔG per reaction. The core's remaining 84 exergonic conversions (24%) are driven by other energy releasing reactions of carbon that do not fall into the 10 categories listed in **Table 2**. At the energetic extremes, only 12 reactions in the core (3%) are endergonic by >5 kJ·mol^{-1} under nonequilibrium conditions at 80°C and pH 9 (**Supplementary Table 8**). The most highly exergonic reaction in the core is catalyzed by pyridoxal phosphate synthase, the mechanism of which (Laber et al., 1999) requires no ATP and eliminates 3 H_2O against a 55 M gradient but with a ΔG of −383 kJ·mol^{-1} (**Supplementary Table 8**) because of the reaction product's aromaticity relative to its reactants. In the simplest interpretation, the carbon-based sources of energy shown in **Table 2** are identical to the sources of energy that gave rise to metabolism, which in turn gave rise to LUCA. The overall flow of energy through the core from high energy substrate H_2 plus low energy CO_2 to reactive carbon compounds and its thermodynamically more stable products is schematically summarized in **Supplementary Figure 5**.

CONCLUSION

The individual biochemical reactions underpinning the synthesis of amino acids, nucleotides and cofactors in modern cells trace to LUCA because of their universality. These reactions are exergonic under the conditions of H_2-producing geochemical systems, where formate (Lang et al., 2010), acetate (Sherwood Lollar et al., 2021) and methane (Proskurowski et al., 2008) are synthesized in abiotic reactions today. In the present work, we have not investigated the role of high hydrostatic pressure exerted by the water column in deep water. This is because the tool we employed to estimate values of ΔG through the component contribution method is designed for studies of microbial metabolism at ambient pressures. At higher hydrostatic pressures, as are found in hydrothermal vents (Kelley et al., 2002), a shift in equilibria toward the formation of more products for reactions of the type A + B → C might be expected according to Le Chatelier's principle. However, it is noteworthy that autotrophic microbes isolated from hydrothermal vents at depths of 2.4 km (ca. 240 bar hydrostatic pressure) grow well under ambient pressure (Beatty et al., 2005), such that in the presence of excellent catalysts (enzymes), high pressure might not be a decisive factor whereby in the presence of only mineral catalysts, hydrostatic pressure might play an important role. Indeed, gasses are compressed considerably at 240 bar and dissolve better in water so that reactant concentrations of dissolved gasses are higher than at ambient pressure. On the contrary, liquid water is compressed only very little (<1% at 240 bar) so that microbes without gas inclusions stay essentially untouched. Methanogens that lack cytochromes require only 10^{-4} to 10^{-5} atm of H_2 for growth (Thauer et al., 2008). Like acetogens, their main energy harnessing reaction results in the conversion of about 20 molecules of CO_2 into waste product (methane for methanogens and acetate for acetogens) for every molecule of CO_2 that is incorporated into cell mass (Martin, 2020). That is, cell mass, the product of metabolism, is just a byproduct of the main energy releasing reaction of the cell. The environment where metabolism arose must therefore have harbored a constantly out of equilibrium supply of carbon, electrons, and transition metal catalysts to promote energy releasing reactions. Reactions of H_2 and CO_2 in serpentinizing hydrothermal systems fulfill those criteria (Schrenk et al., 2013) in a manner that directly connects to the metabolism of modern cells (Preiner et al., 2018; Xavier et al., 2020).

The present data uncover a hitherto unique thermodynamic link between core biochemistry as a whole and the conditions of a geochemical environment known to have existed on the early Earth. The reactions of the core require neither membrane proteins, cytochromes, quinones, nor light. Their thermodynamics indicate that the core biosynthetic reactions of microbial metabolism could have arisen from soluble (Martin and Russell, 2007; Muchowska et al., 2019) and surface-catalyzed (Wächtershäuser, 1988; Preiner et al., 2020) reactions in the dark, under hot, aqueous, H_2-bearing geochemical environments, independent of exposed land masses (light) or the existence of water with a low ionic strength. Though ATP provides energy for roughly one fourth of the core's reactions, a three fourth's majority of reactions derive their energy release from reactions of carbon compounds germane to metabolism itself, sources of chemical energy that, with the exception of thioesters (Semenov et al., 2016) and acyl phosphates (Martin and Russell, 2007; Whicher et al., 2018), have escaped the focus of previous investigations into early metabolic evolution. While estimates of ΔG are, of course, silent on reaction rates, activation

energy, and catalysts (Wolfenden, 2011), the crucial energetic role of hydrogen (Thauer et al., 2008; Fuchs, 2011; Amend et al., 2013; Boyd et al., 2020; Preiner et al., 2020) and the exergonic biochemical reactions of carbon reported here uncover a natural thermodynamic tendency for the individual reactions of metabolism to arise from H_2, CO_2, NH_3, and H_2S in the presence of phosphate.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

WM and JW: conceptualization and visualization. JW, WM, FS, KK, and JX: methodology. JW, JX, and WM: data curation. JW, WM, AV, DP, JL, MP, and KK: formal analysis. WM: writing—original draft, supervision, and funding acquisition. WM, JW, JX, AV, DP, JL, FS, KK, and MP: writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (grant agreement nos 666053 and 101018894). The work was also supported by the Deutsche Forschungsgemeinschaft (Ma 1426/21-1 to WM) and Volkswagen Foundation (96 742 to WM).

ACKNOWLEDGMENTS

We thank Peter Schönheit (Kiel) and Wolfgang Buckel (Marburg) for discussions, Bernhard Schink and Ivan Berg for critical and constructive comments on the manuscript, Masaru K. Nobu (Tsukuba) for critical comments on an earlier version of the manuscript, and computer facilities of University of Düsseldorf (ZIM) for technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.793664/full#supplementary-material

REFERENCES

Alberty, R. A. (1998). Calculation of standard transformed Gibbs energies and standard transformed enthalpies of biochemical reactants. *Arch. Biochem. Biophys.* 353, 116–130. doi: 10.1006/abbi.1998.0638

Amend, J. P., LaRowe, D. E., McCollom, T. M., and Shock, E. L. (2013). The energetics of organic synthesis inside and outside the cell. *Phil. Trans. R. Soc. B* 368:20120255. doi: 10.1098/rstb.2012.0255

Amend, J. P., and McCollom, T. M. (2009). Energetics of biomolecule synthesis on early Earth. In "Chemical Evolution II: from the Origins of Life to Modern Society". *Am. Chem. Soc. Chapter* 4, 63–94.

Amend, J. P., and Shock, E. L. (1998). Energetics of amino acids synthesis in hydrothermal ecosystems. *Science* 281, 1659–1662. doi: 10.1126/science.281.5383.1659

Amend, J. P., and Shock, E. L. (2001). Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and bacteria. *FEMS Microbiol. Rev.* 25, 175–243. doi: 10.1111/j.1574-6976.2001.tb00576.x

Battley, E. H. (1987). *Energetics of Microbial Growth*. New York, NY: Wiley.

Beatty, J. T., Overmann, J., Lince, M. T., Manske, A. K., Lang, A. S., Blankenship, R. E., et al. (2005). An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9306–9310. doi: 10.1073/pnas.0503674102

Beber, M. E., Gollub, M. G., Mozaffari, D., Shebek, K. M., and Noor, E. (2021). Equilibrator 3.0 – a platform for the estimation of thermodynamic constants. *arXiv [Preprint]* Available online at: https://arxiv.org/pdf/2103.00621.pdf (accessed October 4, 2021),

Bennett, B., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* 5, 593–599. doi: 10.1038/nchembio.186

Benson, S. W. (1968). *Thermochemical Kinetics; Methods for the Estimation of Thermochemical Data and Rate Parameters*. New York, NY: Wiley.

Berg, I. A., Kockelhorn, D., Ramos-Vera, W. H., Say, R. F., Zarzyncki, J., Hügler, M., et al. (2010). Autotrophic carbon fixation in archaea. *Nat. Rev. Microbiol.* 8, 447–460. doi: 10.1038/nrmicro2365

Boyd, E. S., Amenabar, M. J., Poudel, S., and Templeton, A. S. (2020). Bioenergetic constraints on the origin of autotrophic metabolism. *Phil. Trans. Roy. Soc. A* 378:20190151. doi: 10.1098/rsta.2019.0151

Buckel, W., and Eggerer, H. (1965). Zur optischen bestimmung von citrat-synthase und von acetyl-coenzym A. [On the optical determination of citrate synthase and acetyl-coenzyme A.]. *Biochem. Z.* 343, 29–43.

Buckel, W., and Thauer, R. K. (2013). Energy conservation via electron bifurcating ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation. *Biochim. Biophy. Acta.* 1827, 94–113. doi: 10.1016/j.bbabio.2012.07.002

Dai, Z., and Locasale, J. W. (2018). Thermodynamic constraints on the regulation of metabolic fluxes. *J. Biol. Chem.* 293, 19725–19739. doi: 10.1074/jbc.RA118.004372

Dailey, H. A., Dailey, T. A., Gerdes, S., Jahn, D., Jahn, M., O'Brian, M. R., et al. (2017). Prokaryotic heme biosynthesis: multiple pathways to a common essential product. *Microbiol. Mol. Biol. Rev.* 81:e00048-16. doi: 10.1128/MMBR.00048-16

Daniel, S. L., Hsu, T., Dean, S. I., and Drake, H. L. (1990). Characterization of the H2- and CO-dependent chemolithotrophic potentials of the acetogens *Clostridium thermoaceticum* and *Acetogenium kivui*. *J. Bacteriol.* 172, 4464–4471. doi: 10.1128/jb.172.8.4464-4471.1990

Decker, K., Jungerman, K., and Thauer, R. K. (1970). Energy production in anaerobic organisms. *Angew. Chem. Int. Ed. Engl.* 9, 138–158. doi: 10.1002/anie.197001381

Dick, G. J. (2019). The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nat. Rev. Microbiol.* 17, 271–283. doi: 10.1038/s41579-019-0160-2

Dimroth, P., and Schink, B. (1998). Energy conservation in the decarboxylation of dicarboxylic acids by fermenting bacteria. *Arch. Microbiol.* 170, 69–77. doi: 10.1007/s002030050616

do Nascimento Vieira, A., Kleinermanns, K., Martin, W. F., and Preiner, M. (2020). The ambivalent role of water at the origins of life. *FEBS Lett.* 594, 2717–2733. doi: 10.1002/1873-3468.13815

Ducluzeau, A. L., van Lis, R., Duval, S., Schoepp-Cothenet, B., Russell, M. J., and Nitschke, W. (2009). Was nitric oxide the first deep electron sink? *Trend Biochem. Sci.* 34, 9–15. doi: 10.1016/j.tibs.2008.10.005

Ebisuzaki, T., and Maruyama, S. (2017). Nuclear geyser model of the origin of life: driving force to promote the synthesis of building blocks of life. *Geosci. Front.* 8, 275–298. doi: 10.1016/j.gsf.2016.09.005

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

Ferry, J. G., and House, C. H. (2006). The step-wise evolution of early Life driven by energy conservation. *Mol. Biol. Evol.* 23, 1286–1292. doi: 10.1093/molbev/msk014

Ferus, M., Nesvorný, D., Šponer, J., Kubelík, P., Michalčíková, R., Shestivská, V., et al. (2015). High-energy chemistry of formamide: a unified mechanism of nucleobase formation. *Proc. Natl. Acad. Sci. U. S. A.* 112, 657–662. doi: 10.1073/pnas.1412072111

Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012). EQuilibrator – The biochemical thermodynamics calculator. *Nucleic Acids Res.* 40, 770–775. doi: 10.1093/nar/gkr874

Fontecilla-Camps, J. C. (2019). Geochemical continuity and catalyst/cofactor replacement in the emergence and evolution of Life. *Angew. Chem.* 58, 42–48. doi: 10.1002/anie.201808438

Fuchs, G. (2011). Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu. Rev. Microbiol.* 65, 631–658. doi: 10.1146/annurev-micro-090110-102801

Fuchs, G., and Stupperich, E. (1985). "Evolution of autotrophic CO2 fixation. In Evolution of Prokaryotes," in *FEMS Symposium*, Vol. 29, eds K. Schleifer and E. Stackebrandt (London: Academic Press), 235–251.

Ginzburg, M., Sachs, L., and Ginzburg, B. Z. (1970). Ion metabolism in a Halobacterium. I. Influence of age of culture on intracellular concentrations. *J. Gen. Physiol.* 55, 187–207. doi: 10.1085/jgp.55.2.187

Goldberg, N., and Tewari, Y. B. (1989). Thermodynamic and transport properties of carbohydrates and their monophosphates: the pentoses and hexoses. *J. Phys. Chem. Ref. Data* 18, 809–880. doi: 10.1063/1.555831

Goldford, J. E., Hartman, H., Marsland, R. III, and Segrè, D. (2019). Environmental boundary conditions for the origin of life converge to an organo-sulfur metabolism. *Nat. Ecol. Evol.* 3, 1715–1724. doi: 10.1038/s41559-019-1018-8

Goldford, J. E., Hartman, H., Smith, T. F., and Segrè, D. (2017). Remnants of an ancient metabolism without phosphate. *Cell* 168, 1126–1134. doi: 10.1016/j.cell.2017.02.001

Hall, D. O., Cammack, R., and Rao, K. K. (1971). Role for ferredoxins in the origin of life and biological evolution. *Nature* 233, 136–138. doi: 10.1038/233136a0

Hamilton, T. L., Bryant, D. A., and Macalady, J. L. (2016). The role of biology in planetary evolution: cyanobacterial primary production in low-oxygen Proterozoic oceans. *Environ. Microbiol.* 18, 325–340. doi: 10.1111/1462-2920.13118

Hansen, L. D., Criddle, R. S., and Battley, E. H. (2009). Biological calorimetry and the thermodynamics of the origination and evolution of Life. *Pure Appl. Chem.* 81, 1843–1855. doi: 10.1351/PAC-CON-08-09-09

Herrmann, G., Jayamani, E., Mai, G., and Buckel, W. (2008). Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. *J. Bacteriol.* 190, 784–791. doi: 10.1128/JB.01422-07

Hordijk, W., Hein, J., and Steel, M. (2010). Autocatalytic sets and the origin of life. *Entropy* 12, 1733–1742. doi: 10.3390/e12071733

Hordijk, W., and Steel, M. (2004). Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J. Theor. Biol.* 227, 451–461. doi: 10.1016/j.jtbi.2003.11.020

Huang, G., Wagner, T., Ermler, U., and Shima, S. (2020). Methanogenesis involves direct hydride transfer from H2 to an organic substrate. *Nat. Rev. Chem.* 4, 213–221. doi: 10.1038/s41570-020-0167-2

Huber, C., and Wächtershäuser, G. (1997). Activated acetic acid by carbon fixation on (Fe,Ni)S under primordial conditions. *Science* 276, 245–248. doi: 10.1126/science.276.5310.245

Hügler, M., and Sievert, S. M. (2011). Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Ann. Rev. Mar. Sci.* 3, 261–289. doi: 10.1146/annurev-marine-120709-142712

Jabłońska, J., and Tawfik, D. S. (2021). The evolution of oxygen-utilizing enzymes suggests early biosphere oxygenation. *Nat. Ecol. Evol.* 5, 442–448. doi: 10.1038/s41559-020-01386-9

Jankowski, M. D., Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2008). Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.* 95, 1487–1499. doi: 10.1529/biophysj.107.124784

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kauffman, S. A. (1986). Autocatalytic sets of proteins. *J. Theor. Biol.* 119, 1–24. doi: 10.1016/S0022-5193(86)80047-9

Kelley, D. S., Baross, J. A., and Delaney, J. R. (2002). Volcanoes, fluids, and life at mid-ocean ridge spreading centers. *Annu. Rev. Earth Planet. Sci.* 30, 385–491. doi: 10.1146/annurev.earth.30.091201.141331

Kelley, D. S., Karson, J. A., Früh-Green, G. L., Yoerger, D. R., Shank, T. M., Butterfield, D. A., et al. (2005). A serpentinite-hosted ecosystem: the Lost City hydrothermal field. *Science* 307, 1428–1434. doi: 10.1126/science.1102556

Kitadai, N., Nakamura, R., Yamamoto, M., Okada, S., Takahagi, W., Nakano, Y., et al. (2021). Thioester synthesis through geoelectrochemical CO2 fixation on Ni sulfides. *Commun. Chem.* 4:37. doi: 10.1038/s42004-021-00475-5

Korolev, N. (2021). How potassium came to be the dominant biological cation: of metabolism, chemiosmosis, and cation selectivity since the beginnings of Life. *BioEssays* 3:2000108. doi: 10.1002/bies.202000108

Laber, B., Maurer, W., Scharf, S., Stepusin, K., and Schmidt, F. S. (1999). Vitamin B6 biosynthesis: formation of pyridoxine 5'-phosphate from 4-(phosphohydroxy)-L-threonine and 1-deoxy-D-xylulose-5-phosphate by PdxA and PdxJ protein. *FEBS Lett.* 449, 45–48. doi: 10.1016/s0014-5793(99)00393-2

Lamadrid, H. M., Rimstidt, J. D., Schwarzenbach, E. M., Klein, F., Ulrich, S., Dolocan, A., et al. (2017). Effect of water activity on rates of serpentinization of olivine. *Nat. Commun.* 8:16107. doi: 10.1038/ncomms16107

Lang, S. Q., and Brazelton, W. J. (2020). Habitability of the marine serpentinite subsurface: a case study of of the Lost City hydrothermal field. *Philos. Trans. A Math. Phys. Eng. Sci.* 378:20180429. doi: 10.1098/rsta.2018.0429

Lang, S. Q., Butterfield, D. A., Schulte, M., Kelley, D. S., and Lilley, M. D. (2010). Elevated concentrations of formate, acetate and dissolved organic carbon found at the Lost City hydrothermal field. *Geochim. Cosmochim. Acta* 74, 941–952. doi: 10.1016/j.gca.2009.10.045

Lang, S. Q., Früh-Green, G. L., Bernasconi, S. M., Brazelton, W. J., Schrenk, M. O., and McGonigle, J. M. (2018). Deeply-sourced formate fuels sulfate reducers but not Methanogens at Lost City hydrothermal field. *Sci. Rep.* 8:755. doi: 10.1038/s41598-017-19002-5

Lewis, C. A. Jr., and Wolfenden, R. (2018). Sulfonium ion condensation: the burden borne by SAM synthetase. *Biochemistry* 57, 3549–3551. doi: 10.1021/acs.biochem.8b00477

Magnabosco, C., Lin, L.-H., Dong, H., Bomberg, M., Ghiorse, W., Stan-Lotter, H., et al. (2018). The biomass and biodiversity of the continental subsurface. *Nat. Geosci.* 11, 707–717. doi: 10.1038/s41561-018-0221-6

Marshall, M. (2020). How the first Life on Earth survived its biggest threat—water. *Nature* 588, 210–213.

Martin, W., and Russell, M. J. (2007). On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 1887–1925. doi: 10.1098/rstb.2006.1881

Martin, W. F. (2020). Older than genes: the acetyl CoA pathway and origins. *Front. Microbiol.* 11:817. doi: 10.3389/fmicb.2020.00817

Martin, W. F., Bryant, D. A., and Beatty, J. T. (2018). A physiological perspective on the origin and evolution of photosynthesis. *FEMS Microbiol. Rev.* 42, 201–231. doi: 10.1093/femsre/fux056

Martin, W. F., and Thauer, R. K. (2017). Energy in ancient metabolism. *Cell* 168, 953–955. doi: 10.1016/j.cell.2017.02.032

McCollom, T., and Seewald, J. S. (2003). Experimental constraints on the hydrothermal reactivity of organic acids and acid anions: I. Formic acid and formate. *Geochim. Cosmochim. Acta* 67, 3625–3644. doi: 10.1016/S0016-7037(03)00136-4

Mereschkowsky, C. (1910). Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen. Biol. Centralbl. 30, 278-288; 289-303; 321-347; 353-367. [English translation in: Kowallik, K. V., and Martin, W. F. (2021). The origin of symbiogenesis: an annotated English translation of Mereschkowky's 1910 paper on the theory of two plasma lineages. *Biosystems* 199:104281. doi: 10.1016/j.biosystems.2020.104281

Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* 191, 144–148. doi: 10.1038/191144a0

Miyakawa, S., James Cleaves, H., and Miller, S. L. (2002). The cold origin of Life: a. Implications based on the hydrolytic stabilities of hydrogen cyanide and formamide. *Orig. Life Evol. Biosph.* 32, 195–208. doi: 10.1023/A:1016514305984

Mojzsis, S. J., Harrison, T. M., and Pidgeon, R. T. (2001). Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* 409, 178–181. doi: 10.1038/35051557

Morrison, R. T., and Boyd, R. N. (1977). *Organic Chemistry*. Boston, MA: Allyn and Bacon.

Muchowska, K. B., Varma, S. J., and Moran, J. (2019). Synthesis and breakdown of universal metabolic precursors promoted by iron. *Nature* 569, 104–107. doi: 10.1038/s41586-019-1151-1

Muchowska, K. B., Varma, S. J., and Moran, J. (2020). Nonenzymatic metabolic reactions and Life's origins. *Chem. Rev.* 120, 7708–7744. doi: 10.1021/acs.chemrev.0c00191

Muller, A. W. (1995). Were the first organisms heat engines? A New model for biogenesis and the early evolution of biological energy conversion. *Prog. Biophys. Mol. Biol.* 63, 193–231. doi: 10.1016/0079-6107(95)00004-7

Müller, V., Chowdhury, N. P., and Basen, M. (2018). Electron bifurcation: a long-hidden energy-coupling mechanism. *Annu. Rev. Microbiol.* 72, 331–353. doi: 10.1146/annurev-micro-090816-093440

Noor, E., Haraldsdóttir, H. S., Milo, R., and Fleming, R. M. T. (2013). Consistent estimation of Gibbs energy using component contributions. *PLoS Comput. Biol.* 9:e1003098. doi: 10.1371/journal.pcbi.1003098

Ozawa, K., Nemoto, A., Imai, E. I., Honda, H., Hatori, K., and Matsuno, K. (2004). Phosphorylation of nucleotide molecules in hydrothermal environments. *Orig. Life Evol. Biosph.* 34, 465–471. doi: 10.1023/B:ORIG.0000043121.65714.05

Pasek, M. A. (2020). Thermodynamics of prebiotic phosphorylation. *Chem. Rev.* 120, 4690–4706. doi: 10.1021/acs.chemrev.9b00492

Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D., and Sutherland, J. D. (2015). Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* 7, 301–307. doi: 10.1038/nchem.2202

Preiner, M., Igarashi, K., Muchowska, K. B., Yu, M., Varma, S. J., Kleinermanns, K., et al. (2020). A hydrogen dependent geochemical analogue of primordial carbon and energy metabolism. *Nature Ecol. Evol.* 4, 534–542. doi: 10.1038/s41559-020-1125-6

Preiner, M., Xavier, J. C., Sousa, F. L., Zimorski, V., Neubeck, A., Lang, S. Q., et al. (2018). Serpentinization: connecting geochemistry, ancient metabolism and industrial hydrogenation. *Life* 8:41. doi: 10.3390/life8040041

Preiner, M., Xavier, J. C., Vieira, A. D. N., Kleinermanns, K., Allen, J. F., and Martin, W. F. (2019). Catalysts, autocatalysis and the origin of metabolism. *Interface Focus* 9:20190072. doi: 10.1098/rsfs.2019.0072

Proskurowski, G., Lilley, M. D., Seewald, J. S., Früh-Green, G. L., Olson, E. J., Lupton, J. E., et al. (2008). Abiogenic hydrocarbon production at Lost City hydrothermal field. *Science* 319, 604–607. doi: 10.1126/science.1151194

Richey, B., Cayley, D. S., Mossing, M. C., Kolka, C., Anderson, C. F., Farrar, T. C., et al. (1987). Variability of the intracellular ionic environment of *Escherichia coli*. Differences between in vitro and in vivo effects of ion concentrations on protein DNA interactions and gene expression. *J. Biol. Chem.* 262, 7157–7164. doi: 10.1016/S0021-9258(18)48218-X

Rosing, M. T. (1999). 13C-depleted carbon microparticles in >3700-Ma sea-floor sedimentary rocks from West Greenland. *Science* 283, 674–676. doi: 10.1126/science.283.5402.674

Rother, M., and Metcalf, W. W. (2004). Anaerobic growth of *Methanosarcina acetivorans* C2A on carbon monoxide: an unusual way of life for a methanogenic archaeon. *Proc. Natl. Acad. Sci. U.S.A.* 101, 16929–16934. doi: 10.1073/pnas.0407486101

Russell, J. B., and Cook, G. M. (1995). Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol. Rev.* 59, 48–62. doi: 10.1128/mr.59.1.48-62.1995

Say, R. F., and Fuchs, G. (2010). Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* 464, 1077–1081. doi: 10.1038/nature08884

Schönheit, P., Buckel, W., and Martin, W. (2016). On the origin of heterotrophy. *Trends Microbiol.* 24, 12–25. doi: 10.1016/j.tim.2015.10.003

Schramm, G., Grotsch, H., and Pollmann, W. (1962). Nicht-enzymatische Synthese von Polysacchariden, Nucleosiden und Nucleinsäuren und die Entstehung selbst vermehrungsfähiger Systeme. *Angew. Chem.* 74, 53–92. doi: 10.1002/ange.19620740202

Schrenk, M. O., Brazelton, W. J., and Lang, S. Q. (2013). Serpentinization, carbon and deep Life. *Rev. Mineral. Geochem.* 75, 575–606. doi: 10.2138/rmg.2013.75.18

Schuchmann, K., and Müller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* 12, 809–821. doi: 10.1038/nrmicro3365

Semenov, S. N., Kraft, L. J., Ainla, A., Zhao, M., Baghbanzadeh, M., Campbell, V. E., et al. (2016). Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. *Nature* 537, 656–660. doi: 10.1038/nature19776

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sherwood Lollar, B., Heuer, V. B., McDermott, J., Tille, S., Warr, O., Moran, J. J., et al. (2021). A window into the abiotic carbon cycle – Acetate and formate in fracture waters in 2.7 billion year-old host rocks of the Canadian Shield. *Geochim. Cosmochim. Acta.* 294, 295–314. doi: 10.1016/j.gca.2020.11.026

Sleep, N. H., Bird, D. K., and Pope, E. C. (2011). Serpentinite and the Dawn of Life. *Philos. Trans. R. Soc. B* 366, 2857–2869. doi: 10.1098/rstb.2011.0129

Smith, A. R., Kieft, B., Mueller, R., Fisk, M. R., Mason, O. U., Popa, R., et al. (2019). Carbon fixation and energy metabolisms of a subseafloor olivine biofilm. *ISME J.* 13, 1737–1749. doi: 10.1038/s41396-019-0385-0

Sossi, P. A., Burnham, A. D., Badro, J., Lanzirotti, A., Newville, M., and O'Neill, H. S. C. (2020). Redox state of Earth's magma ocean and its Venus-like early atmosphere. *Sci. Adv.* 6:eabd1387. doi: 10.1126/sciadv.abd1387

Sousa, F. L., Hordijk, W., Steel, M., and Martin, W. F. (2015). Autocatalytic sets in E. coli metabolism. *J. Syst. Chem.* 6, 15–21. doi: 10.1186/s13322-015-0009-7

Sousa, F. L., Preiner, M., and Martin, W. F. (2018). Native metals, electron bifurcation, and CO2 reduction in early biochemical evolution. *Curr. Opin. Microbiol.* 43, 77–83. doi: 10.1016/j.mib.2017.12.010

Sousa, F. L., Thiergart, T., Landan, G., Nelson-Sathi, S., Pereira, I. A. C., Allen, J. F., et al. (2013). Early bioenergetic evolution. *Phil. Trans. R Soc. Lond. B* 368, 20130088. doi: 10.1098/rstb.2013.0088

Steffens, L., Pettinato, E., Steiner, T. M., Mall, A., König, S., Eisenreich, W., et al. (2021). High CO2 levels drive the TCA cycle backwards towards autotrophy. *Nature* 592, 784–788. doi: 10.1038/s41586-021-03456-9

Stetter, K. O. (2006). Hyperthermophiles in the history of Life. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1837–1842. doi: 10.1098/rstb.2006.1907

Suzuki, S., Nealson, K. H., and Ishii, S. (2018). Genomic and in-situ transcriptomic characterization of the candidate phylum NPL-UPL2 from highly alkaline highly reducing serpentinized groundwater. *Front. Microbiol.* 9:3141. doi: 10.3389/fmicb.2018.03141

Thauer, R. K., Jungermann, K., and Decker, K. (1977). Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol. Rev.* 41, 100–180.

Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W., and Hedderich, R. (2008). *Methanogenic* Archaea: ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* 6, 579–591. doi: 10.1038/nrmicro1931

Tian, T., Chu, X.-Y., Yang, Y., Zhang, X., Liu, Y.-M., Gao, J., et al. (2019). Phosphates as energy sources to expand metabolic networks. *Life* 9:43. doi: 10.3390/life9020043

Wächtershäuser, G. (1988). Pyrite formation, the first energy source for life: a hypothesis. *Syst. Appl. Microbiol.* 10, 207–210. doi: 10.1016/S0723-2020(88)80001-8

Wächtershäuser, G. (1992). Groundworks for an evolutionary biochemistry – the iron sulfur world. *Prog. Biophys. Mol. Biol.* 58, 85–201. doi: 10.1016/0079-6107(92)90022-x

Walsh, C. T., Tu, B. P., and Tang, Y. (2018). Eight kinetically stable but thermodynamically activated molecules that power cell metabolism. *Chem. Rev.* 118, 1460–1494. doi: 10.1021/acs.chemrev.7b00510

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., et al. (2016). The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* 1:16116. doi: 10.1038/nmicrobiol.2016.116

Whicher, A., Camprubi, E., Pinna, S., Herschy, B., and Lane, N. (2018). Acetyl phosphate as a primordial energy currency at the origin of Life. *Orig. Life Evol. Biosph.* 48, 159–179. doi: 10.1007/s11084-018-9555-8

Williams, T. A., Szöllősi, G. J., Spang, A., Foster, P. G., Heaps, S. E., Boussau, B., et al. (2017). Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc. Natl Acad. Sci. U.S.A* 114, E4602–E4611. doi: 10.1073/pnas.1618463114

Wimmer, J. L. E., Vieira, A. D. N., Xavier, J. C., Kleinermanns, K., Martin, W. F., and Preiner, M. (2021a). The autotrophic core: an ancient network of 404 reactions converts H2, CO2, and NH3 into amino acids, bases, and cofactors. *Microorganisms* 9:458. doi: 10.3390/microorganisms9020458

Wimmer, J. L. E., Kleinermanns, K., and Martin, W. F. (2021b). Pyrophosphate and irreversibility in evolution, or why PPi is not an energy currency and why nature chose triphosphates. *Front. Microbiol.* 12:759359. doi: 10.3389/fmicb.2021.759359

Wolfenden, R. (2011). Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Annu. Rev. Biochem.* 80, 645–667. doi: 10.1146/annurev-biochem-060409-093051

Xavier, J. C., Gerhards, R. E., Wimmer, J. L. E., Brueckner, J., Tria, F. D. K., and Martin, W. F. (2021). The metabolic network of the last bacterial common ancestor. *Commun. Biol.* 4:413. doi: 10.1038/s42003-021-01918-4

Xavier, J. C., Hordijk, W., Kauffman, S., Steel, M., and Martin, W. F. (2020). Autocatalytic chemical networks at the origin of metabolism. *Proc. Biol. Sci.* 287:20192377. doi: 10.1098/rspb.2019.2377

Zahnle, K., Arndt, N., Cockell, C., Halliday, A., Nisbet, E., Selsis, F., et al. (2007). Emergence of a habitable planet. *Space Sci. Rev.* 129, 35–78.

Zhang, X., Li, L.-F., Du, Z.-F., Hao, Z.-L., Cao, L., Luan, Z.-D., et al. (2020). Discovery of supercritical carbon dioxide in a hydrothermal system. *Sci. Bull.* 65, 958–964. doi: 10.1016/j.scib.2020.03.023

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

III Pyrophosphate and irreversibility in evolution, or why PP_i is not an energy currency and why nature chose triphosphates

Jessica L. E. Wimmer, Karl Kleinermanns, William F. Martin

Institut für Molekulare Evolution, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Deutschland.

Dieser Artikel wurde am 06.10.2021 in *Frontiers in Microbiology* Ausgabe 12 veröffentlicht.

Beitrag von Jessica L. E. Wimmer (Erstautor und Co-Korrespondenz):
Der Datensatz wurde von mir zusammengestellt und modifiziert. Die Analysen wurden von mir durchgeführt mit Ausnahme der kinetischen Simulation. Ich war in die Entwicklung des Studiendesigns involviert und habe das initiale Manuskript überarbeitet.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 06 October 2021
doi: 10.3389/fmicb.2021.759359

Pyrophosphate and Irreversibility in Evolution, or why PP_i Is Not an Energy Currency and why Nature Chose Triphosphates

Jessica L. E. Wimmer[1], Karl Kleinermanns[2] and William F. Martin[1]**

[1]Institute for Molecular Evolution, Department of Biology, Heinrich Heine University Duesseldorf, Duesseldorf, Germany,
[2]Institute for Physical Chemistry, Department of Chemistry, Heinrich Heine University Duesseldorf, Duesseldorf, Germany

The possible evolutionary significance of pyrophosphate (PP_i) has been discussed since the early 1960s. Lipmann suggested that PP_i could have been an ancient currency or a possible environmental source of metabolic energy at origins, while Kornberg proposed that PP_i vectorializes metabolism because ubiquitous pyrophosphatases render PP_i forming reactions kinetically irreversible. To test those ideas, we investigated the reactions that consume phosphoanhydride bonds among the 402 reactions of the universal biosynthetic core that generates amino acids, nucleotides, and cofactors from H_2, CO_2, and NH_3. We find that 36% of the core's phosphoanhydride hydrolyzing reactions generate PP_i, while no reactions use PP_i as an energy currency. The polymerization reactions that generate ~80% of cell mass – protein, RNA, and DNA synthesis – all generate PP_i, while none use PP_i as an energy source. In typical prokaryotic cells, aminoacyl tRNA synthetases (AARS) underlie ~80% of PP_i production. We show that the irreversibility of the AARS reaction is a kinetic, not a thermodynamic effect. The data indicate that PP_i is not an ancient energy currency and probably never was. Instead, PP_i hydrolysis is an ancient mechanism that imparts irreversibility, as Kornberg suggested, functioning like a ratchet's pawl to vectorialize the life process toward growth. The two anhydride bonds in nucleoside triphosphates offer ATP-cleaving enzymes an option to impart either thermodynamic control (P_i formation) or kinetic control (PP_i formation) upon reactions. This dual capacity explains why nature chose the triphosphate moiety of ATP as biochemistry's universal energy currency.

Keywords: energetics, bioenergetics, chemical evolution, origin of life, early evolution, metabolism, kinetics, thermodynamics

OPEN ACCESS

Edited by:
Christiane Dahl,
University of Bonn, Germany

Reviewed by:
Jose Roman Perez-Castineira,
Sevilla University, Spain
Adrian Goldman,
University of Helsinki, Finland

****Correspondence:***
Jessica L. E. Wimmer
jessica.wimmer@hhu.de
William F. Martin
bill@hhu.de

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *16 August 2021*
Accepted: *15 September 2021*
Published: *06 October 2021*

Citation:
Wimmer JLE, Kleinermanns K and Martin WF (2021) Pyrophosphate and Irreversibility in Evolution, or why PP_i Is Not an Energy Currency and why Nature Chose Triphosphates. Front. Microbiol. 12:759359. doi: 10.3389/fmicb.2021.759359

INTRODUCTION

Starting in the 1960s, thoughts on the possible evolutionary significance of inorganic pyrophosphate (PP_i) have centered around two main concepts: irreversibility and energy. Kornberg, who worked on nucleic acid polymerization, recognized that PP_i producing biochemical steps confer the property of irreversibility upon reactions under physiological conditions because ubiquitous pyrophosphatases constantly degrade PP_i in the cytosol of cells (Kornberg, 1962). His reasoning

was straightforward: By degrading PP_i, a substrate required for the enzymatic back reaction of the PP_i producing step, the rate of the back reaction effectively approaches zero. In this way, pyrophosphatases would render PP_i producing reactions irreversible by means of kinetics, rather than thermodynamics. Though Kornberg's mechanism of irreversibility was later called into question because PP_i concentrations in exponentially growing cells were reported to be too high for this principle to work (Kukko and Heinonen, 1982), as soon as cells leave the exponential growth phase, Kornberg's principle immediately applies, as we will see during the course of this paper, because PP_i production is strictly linked to growth, while PP_i hydrolysis is not. Kornberg's list of such irreversible PP_i producing reactions included nucleic acid polymerization, translation, and cofactor biosynthetic routes (Kornberg, 1962) and this function, irreversibility, was seen as harboring the significance of PP_i.

Lipmann, who worked on high energy bonds, suggested that PP_i could have served as a possible energy currency in primordial metabolism, and that modern PP_i-dependent enzymes represent fossils from a time in which prebiotic metabolism extracted energy from environmentally available phosphate minerals (Lipmann, 1965). In that view, the evolutionary significance of PP_i is sought in its possible role as a source of biochemical energy in prebiotic chemical reactions resembling those of physiology. Aspects of both Kornberg's and Lipmann's views are germane to Schramm's proposal that environmental polyphosphates could have powered early nucleic acid synthesis (Schramm et al., 1962).

In 1966, Baltscheffsky reported a membrane-associated pyrophosphatase (mPPase) that reversibly couples proton translocation to PP_i hydrolysis (Baltscheffsky et al., 1966), thereby linking PP_i to Mitchell's then new chemiosmotic theory of ATP synthesis involving ion gradients and electron transfer chains (Mitchell, 1961). That finding, together with Reeves' report of a PP_i-dependent glycolytic enzyme (Reeves, 1968), now called pyruvate orthophosphate dikinase, seemed to support an ancient bioenergetic role behind the possible evolutionary significance of PP_i. Based on such findings, the view that PP_i's evolutionary significance resides in primordial energetics established a long tradition that is still widely embraced (de Duve, 1991; Russell and Hall, 1997; Russell et al., 2013, 2014; Wang et al., 2019; Piast et al., 2020) though seldom critically inspected (Martin, 2020).

Comparatively few enzymatic reactions involve PP_i. Kornberg (1962) listed 35 enzymatic reactions that release PP_i in the physiological reaction. Heinonen (2001) listed 173 PP_i producing reactions. By contrast, Kyoto Encyclopedia of Genes and Genomes (KEGG) list 194 reactions among prokaryotes that involve ATP. Since the book of Heinonen (2001), some new PP_i producing reactions have been reported (Nagata et al., 2018), yet the precise roles of PP_i in physiology and evolution are still discussed (Heinonen, 2001; Pérez-Castiñeira et al., 2021). Soluble pyrophosphatases (sPPases) are ubiquitous in distribution (Lahti, 1983). Ion-pumping mPPases are found in various microbes and plants (Serrano et al., 2007), and PP_i-dependent glycolytic enzymes occur as alternatives of ATP-dependent forms, at the phosphofructokinase (PFK) and pyruvate kinase (PYK) steps (Heinonen, 2001; Siebers and Schönheit, 2005; Bräsen et al., 2014; Holwerda et al., 2020). Though PP_i-dependent glycolysis is often interpreted as an adaptation that reduces ATP expense (Heinonen, 2001) or that salvages energy from PP_i produced from translation (Reeves, 1968), PP_i-utilizing glycolytic enzymes have a conspicuous tendency to occur among microbes that have specialized to sugar-rich environments. Such specialists include human parasites, such as *Entamoeba* (Reeves, 1984), *Giardia* (Müller et al., 2012), and trypanosomes (Michels et al., 2006), as well as non-parasitic cellulose-, saccharose-, and sugar-degrading bacteria (Bielen et al., 2010; Holwerda et al., 2020) and archaea (Bräsen et al., 2014). In addition, PP_i-dependent enzymes are particularly common in the strictly sugar-based carbon metabolism of plants (Serrano et al., 2007). This pattern of occurrence might be suggestive of an ecological rather than energetic basis behind the distribution of PP_i-dependent glycolytic pathways.

In line with that view, the use of PP_i-dependent glycolytic enzymes generally coincides with loss of allosteric regulation through the pathway (Siebers and Schönheit, 2005; Bräsen et al., 2014). In the well-studied example of trypanosomes, loss of regulation allows flux through the pathway to be governed by sugar concentrations in the medium (blood sugar), an ecological adaptation of growth rates to substrate availability, not energetic efficiency, especially as trypanosomes excrete the energy rich compound pyruvate as a metabolic end product (Michels et al., 2006). Even in the well-studied glucose fermenting bacterium *Clostridium thermocellum*, which also excretes pyruvate, a clear energetic advantage of its PP_i-dependent glycolysis is not evident (Holwerda et al., 2020). Moreover, deletion of *C. thermocellum*'s mPPase has no impact on growth (Holwerda et al., 2020), a finding that is hard to reconcile with a central role for energy conservation *via* PP_i in energy metabolism of the bacterium, although sPPase activity was not reported in the mPPase mutant. By contrast, deletion mutants of sPPases are lethal in *Escherichia coli* (Chen et al., 1990) and in yeast (Pérez-Castiñeira et al., 2002). This finding is very notable because from an energetic standpoint, because sPPases effectively "waste" phosphoanhydride bonds *via* rapid PP_i hydrolysis, raising the question: why should elimination of the "energy wasting" reaction catalyzed by sPPase be lethal? The growth inhibiting phenotype of sPPase deletion mutants is, however, readily reconciled with Kornberg's kinetic view of PP_i function, because sPPase knockouts in *E. coli* and yeast yield cells that cannot grow mainly because protein synthesis comes to a halt through product inhibition *via* PP_i accumulation at the amino acyl tRNA synthesis step.

Our present interest in PP_i stems from comparative physiological investigations into the energetics of primordial metabolism (Martin and Russell, 2007; Sousa et al., 2013; Sousa and Martin, 2014; Preiner et al., 2020). We reasoned that if PP_i had played any role in primordial energetics, as is widely assumed (de Duve, 1991; Russell et al., 2013, 2014; Wang et al., 2019; Piast et al., 2020), evidence for that role should be preserved in the conserved core of metabolism within modern cells. This is the same conventional logic that is used

to interpret other aspects of physiology as relicts of ancient metabolism: metal sulfide clusters in proteins (Eck and Dayhoff, 1966; Wächtershäuser, 1992; Heinen and Lauwers, 1997), the use of organic cofactors as catalysts (White, 1976), carbon metal bonds in enzyme active sites (Martin, 2019), thioesters as energy currencies (Semenov et al., 2016; Kitadai et al., 2021), or anaerobic chemolithoautotrophy (Mereschkowsky, 1910; Decker et al., 1970). Though this line of reasoning (comparative physiology) can be questioned, it is the same reasoning that underlies the view that PP_i is an ancient energy currency. The conserved core of metabolism is a set of roughly 400 reactions that generates the 20 canonical amino acids, the four bases of RNA and DNA, and the cofactors required for their synthesis from H_2, CO_2, NH_3, H_2S, P_i, and inorganic salts (Wimmer et al., 2021). Because of its universally conserved nature, this biosynthetic core of chemical reactions (though not necessarily all of it enzymes) was present in the last universal common ancestor, LUCA, and has persisted in all lineages throughout evolution over the last 4 billion years since their divergence from LUCA (Weiss et al., 2016). The universal core thus harbors insights not only into LUCA's physiology, but also into the primordial set of reactions that gave rise to the building blocks from which LUCA was assembled. Because our present investigation probes metabolism itself, our insights into the role of PP_i in evolution differ from those based in the study of phosphorous minerals (Pasek et al., 2017). And because our investigation is based in the comparative physiology of living cells, our insights into the role of thermodynamics and kinetics in evolution differ from those based in studies of chemical nucleic acid synthesis (Pascal et al., 2013). A fresh look at the role of PP_i in ancient metabolism suggests that Lipmann was probably wrong, that Kornberg was probably right, and furthermore reveals why nature chose triphosphates as the universal energy currency.

MATERIALS AND METHODS

Reactions Involving Inorganic Pyrophosphate

The 36 metabolic reactions involving inorganic pyrophosphate (PP_i; **Supplementary Table S1**) in the biosynthetic core were taken from supplemental data of Wimmer et al. (2021). Reaction R00720 involving IMP synthesis was removed from the core because it is not essential. The reactions were initially collected from the KEGG (Kanehisa and Goto, 2000), version December 2020 and polarized in the direction of cell synthesis.

ATP Hydrolysis Among Prokaryotes

Reactions involving ATP hydrolysis of the reaction scheme $X + ATP \leftrightarrow Y + ADP + P_i$ were obtained from KEGG (Kanehisa and Goto, 2000). *X* and *Y* are placeholders for variable compounds. Additional compounds on both sides can be present. A total of 15,339 KEGG reactions were searched for the reaction scheme in the forward and back direction since KEGG reactions are not polarized in general. A total of 131 reactions involving ATP hydrolysis were obtained in the data, whereas 61 reactions are specific to prokaryotes (**Supplementary Table S2**). The domain check was performed by parsing the Enzyme Commission (EC) numbers of each reaction, gathering a list of genes and their respective organisms leading to the domain.

Collection of Michaelis–Menten Constants for Pyrophosphatases

Michaelis–Menten constants (K_m) for inorganic pyrophosphatase activity in *E. coli* wildtypes were obtained from BRENDA (Schomburg et al., 2002) *via* EC number 3.6.1.1. *Escherichia coli* mutants were removed from the data (**Supplementary Table S3**).

Kinetic Effect of PP_i in Translation

To investigate the effect of pyrophosphate hydrolysis on the product yield (adenylated amino acid) in aminoacyl-tRNA synthetase (AARS) reactions more quantitatively, kinetic simulations of substrate binding and activation of isoleucine by adenylation in isoleucyl-tRNA synthetase were performed using Mathcad 2001 (Mathsoft Engineering & Education, Inc.). The underlying kinetic scheme is taken from Pope et al. (1998) with hydrolysis of PP_i added (see **Supplementary Figure S1A**). The rate equations were used to obtain the concentration vs. time profiles by numerical integration (see **Supplementary Figure S1B**). Experimental rate constants were obtained from Pope et al. (1998) and Stockbridge and Wolfenden (2011). Initial concentrations of 1 mM amino acid, enzyme, and ATP were used, and integration was carried out up to 20 s. These calculations provide an empirical basis for the intuitive effect of product removal during the PP_i forming step of translation (see **Supplementary Figure S2**).

RESULTS

Pyrophosphate Polarized LUCA's Core Biosynthetic Metabolism

To see whether PP_i might have had a role in primordial energetics, the reactions of the core (Wimmer et al., 2021) that involve PP_i or ATP were identified and polarized in the biosynthetic direction, that is, from H_2 and CO_2 toward cell mass synthesis. In Lipman's view, PP_i was an environmental energy source, a substrate that assumes a thermodynamic role as an educt residing on the left side of an enzymatic reaction, while in Kornberg's view PP_i is synthesized in metabolism *via* ATP hydrolysis and assumes a kinetic role as a product that is removed from the right side of the reaction. Writing the reactions from left to right in the direction of CO_2 to products as the pathways are mapped in KEGG (Kanehisa and Goto, 2000) brings the role of PP_i in the core into focus. Among the 36 reactions of the core in which PP_i occurs, it is always a reaction product occurring on the right side of the reaction, serving as an energy source in zero reactions (**Supplementary Table S1**). The reactions of the biosynthetic core of metabolism thus speak 36:0 in favor of PP_i conferring irreversibility, as Kornberg (1962)

suggested, and harbor no traces of Lipmann's proposal for an ancient energetic or thermodynamic role for PP_i.

In the metabolism of modern cells, PP_i is always produced from ATP by reaction sequences that sum to $ATP + H_2O \rightarrow AMP + PP_i$ ($\Delta G_o' = -46\,kJ{\cdot}mol^{-1}$), slightly more exergonic than $ATP + H_2O \rightarrow ADP + P_i$ ($\Delta G_o' = -32\,kJ{\cdot}mol^{-1}$). This opens the possibility that PP_i formation might have played an energetic role in the core, but not as a source of high energy bonds. Were the role of PP_i in the core thermodynamic, it could have readily been replaced in evolution by compounds with a similar or higher free energy of hydrolysis, such as acetyl phosphate ($\Delta G_o' = -43\,kJ{\cdot}mol^{-1}$), 1,3-bisphosphoglycerate ($\Delta G_o' = -52\,kJ{\cdot}mol^{-1}$), or phosphoenolpyruvate ($\Delta G_o' = -62\,kJ{\cdot}mol^{-1}$), the high energy bonds in all three of which are synthesized in metabolism using one ATP each (the same cost as PP_i hydrolysis). Because ATP hydrolysis to adenosine monophosphate (AMP) and PP_i is not replaced by alternative energy currencies with a higher free energy of hydrolysis, and because PP_i is always a product in the core, not an educt, the function of PP_i in the core can hardly be thermodynamic.

Keeping in mind that Kornberg's suggestion for the role of PP_i was based on nucleic acid polymerization and translation, the occurrence of PP_i in the core solely as a product suggests that its role is kinetic, lowering the rate of back reactions, rather than thermodynamic. Is this true more generally in metabolism, that is, outside the core? We consulted KEGG. If PP_i had any role during early evolution as an energy currency, then some reactions should persist in which PP_i hydrolysis is coupled to an otherwise endergonic reaction. Though a handful of PP_i consuming reactions phosphorylate substrates in the physiological reaction (Nagata et al., 2018), among 15,339 reactions in KEGG, we found no PP_i hydrolyzing, non-phosphorylating reactions at all that provide energetic coupling to an otherwise thermodynamically unfavorable reaction. That is, there were no reactions of the type $X + PP_i \leftrightarrow Y + 2\,P_i$, whereby we note that the pyrophosphatase reaction, KEGG reaction number R00004, employs H_2O as X but has no Y component. This is a noteworthy observation. It indicates that PP_i serves at best as a phosphorylating agent in metabolism, but never as a source of pure thermodynamic impetus to help push unfavorable reactions forward *via* coupling to PP_i hydrolysis. By contrast, a number of metabolic reactions (61 prokaryote specific reactions among 15,339 total reactions in KEGG; **Supplementary Table S2**) go forward because they are coupled to non-phosphorylating ATP hydrolysis in reactions of the type $X + ATP \leftrightarrow Y + ADP + P_i$. The lack of such reactions for PP_i in KEGG clearly indicates that PP_i is not a dedicated energy currency in biosynthesis, notwithstanding the existence of PP_i-dependent glycolytic pathways, as outlined in the introduction. Note that KEGG does not include the myriad reactions in which ATP (or GTP) phosphorylates proteins, and we know of no examples in which PP_i is used to phosphorylate proteins as true energy currencies do. These findings indicate that PP_i is not a dedicated energy currency and that by inference, in the simplest interpretation, it never has been.

PP_i producing reactions are generally seen as being irreversible under physiological conditions because of the ubiquitous presence of high activities of sPPases in cells (Lahti, 1983; Danchin et al., 1984; Heinonen, 2001), which catalyze the reaction $PP_i + H_2O \rightarrow 2P_i$ ($\Delta G_o' = -21\,kJ{\cdot}mol^{-1}$), thereby continuously removing a substrate for PP_i producing reactions in the reverse direction. Notably, aqueous Mg^{2+} ions alone accelerate the rate of spontaneous PP_i hydrolysis by three orders of magnitude in water and PP_i hydrolysis in dimethyl sulfoxide/water by six orders of magnitude (Stockbridge and Wolfenden, 2011), such that inorganically catalyzed PP_i hydrolysis might have been a mechanism of irreversibility even before the advent of enzymes. Irreversibility at 36 PP_i-dependent enzymatic reactions in the core – nine in amino acid pathways, three in nucleotide synthesis, and 19 in cofactor synthesis (**Supplementary Table S1**) – functions in modern metabolism as a system of check valves (valves that close to prevent backward flow) that, individually and in concert, act as a ratchet's pawl, inching the reactions of the core unidirectionally forward toward product synthesis. We suggest that this has been the case since the availability of ATP as the universal energy currency.

Pyrophosphate Polarized Metabolism *in toto* Throughout All of Evolution

In the metabolism of LUCA, PP_i forced the reactions of the core forward in the direction of monomer synthesis for cell mass synthesis. The effect of PP_i, however, extended well beyond LUCA's core biosynthesis because PP_i renders nucleic acid and protein synthesis irreversible (Kornberg, 1962), and because LUCA possessed the genetic code and was able to synthesize RNA, DNA, and proteins (Weiss et al., 2016). To get a better picture of the polarizing role of PP_i in the central dogma of molecular biology, we generated estimates for its quantitative contribution to the overall ATP budget based on the classical estimates of Stouthamer (1978), which are still in wide use today. Protein synthesis requires activated amino acids, rRNA, tRNA, and mRNA. In a modern cell growing from H_2, CO_2, and NH_3, the synthesis of protein comprises the combined energetic cost of making RNA and protein, consuming roughly 76% of the biosynthetic ATP budget (**Table 1**). The quantitative contribution of PP_i forming reactions to the cellular energy budget is surprisingly large. In amino acid biosynthesis, 47% of the ATP consuming reactions (8/17) generate PP_i, whereas in nucleotide synthesis 13.6% (3/22) generate PP_i. In polymerization reactions, the contributions are greater.

Guanosine triphosphate (GTP) hydrolyzing reactions are not uncommon in LUCA's biosynthetic core (**Figure 1**), in line with its ancient role in metabolism (Martin and Russell, 2007) and the observation that in some organisms where it has been investigated, GTP is readily used as a substrate in reactions that are typically regarded as ATP dependent (Holwerda et al., 2020). About 26% of a cell's energy budget is consumed in the GTP-dependent steps of translation. The main biosynthetic ATP expense in protein synthesis is translation, which consumes four ATP per peptide bond (Stouthamer, 1978). Peptide chain elongation at the ribosome has two P_i forming GTP hydrolysis steps catalyzed by EF-Tu and EF-G (Satpati et al., 2014), while

TABLE 1 | ATP expense per gram of cells during biosynthetic processes[a] [mol·10^4].

Component	Monomer synthesis[b]		Polymerization[c]	
	Total	PP_i- forming	Total	PP_i- forming
Protein	14	6.6	191	91.3
RNA	34	4.6	23	20.3
DNA	9	1.2	2	1.2
Lipid[d]	1	–	–	–
Polysaccharide[d]	21	–	–	–
Import[e]	52	_	_	_
Sum	131	12.4	216	113

Energetic cost of protein synthesis incl. ribosome biosynthesis: 262/347 = 76%
Energetic contribution of PP_i forming steps in cell biosynthesis: 125/347 = 36%
Energetic contribution of GTP-dependent biosynthetic steps[f]: 98.5/347 = 28%

[a]Values are for Escherichia coli from Stouthamer (1978) as tabulated by Harold (1986). Lever et al. (2015) calculate $\Delta G_o'$ for the synthesis of monomers from H_2, CO_2, and NH_3 based on the values of Neidhardt et al. (1990) but not the ATP expense per monomer. Neidhardt et al. (1990) estimate ATP expense for monomer synthesis as 42 ATP per 20 amino acids and 40 ATP per four nucleotides in E. coli. A dash (–) indicates that the value is zero or negligible. Note that these calculations entail only biosynthetic costs and do not consider energy spilling (Tempest and Neijssel, 1984) or maintenance energy, which can exceed biosynthesis by a factor of 3 in exponential growth (Russell and Cook, 1995; Russell, 2007), and by more under energy limitation (Lever et al., 2015).

[b]The proportion of PP_i forming steps in monomer biosynthesis was calculated as the total cost for the monomer multiplied by the fraction of PP_i forming steps among ATP hydrolyzing steps en route to nucleoside monophosphates (Wimmer et al., 2021).

[c]The proportion of PP_i forming steps in polymerization takes the costs of proofreading, assembly and modification from Lever et al. (2015) into account. These are not PP_i-forming reactions.

[d]LPS, lipopolysaccharide. These values are for E. coli. There is a PP_i forming component of lipid synthesis in archaea that is neglected here. Archaea also lack LPS and possess no murein, though sometimes pseudomurein (Albers and Meyer, 2011), and have a larger protein component in the cell wall (S-Layer).

[e]Stouthamer (1978) calculates the cost of import for precursors, mainly ammonium, Lever et al. (2015) neglect import. If one considers a functional core before the origin of free-living cells, no costs for import are incurred.

[f]In the core, 13% of the triphosphate expense for amino acid and NMP/dNMP monomer synthesis (55 10^{-4} mol ATP per gram of cells) is GTP-dependent (7.2 10^{-4} mol ATP per gram of cells), plus two GTP-dependent steps in translation (91.3 10^{-4} mol ATP per gram of cells) yield ca. 98.5 10^{-4} mol GTP per gram of cells.

amino acyl tRNA synthesis requires the expense of two ATP through amino acid activation by amino acyl tRNA synthetase (AARS) enzymes. Activation generates aminoacyl adenylate and PP_i, followed by aminoacylation of tRNA and AMP release (Gomez and Ibba, 2020). Because half of the energy cost for translation resides in the PP_i producing nature of the AARS reactions, roughly 26% of the cell's total biosynthetic ATP expense (91/347, **Table 1**) is incurred to pay the price of irreversibility at the formation of aminoacyl tRNAs for translation.

PP_i has an ancient and conserved function in metabolism as a mediator of irreversibility (Kornberg, 1962) that clearly traces to LUCA (**Figure 1**). By contrast, not a single reaction in the core uncovers a role for PP_i as an energy currency in primordial metabolism. Based upon the conserved nature of the core across all modern lineages, we can infer that PP_i generating reactions have vectorialized monomer synthesis of ABC compounds throughout evolution, acting as a ratchet's pawl (**Figure 1B**), rendering monomer synthesis unidirectional, even in low energy environments.

PP_i producing steps in RNA monomer and polymer synthesis account for about 7% of the overall biosynthetic energy budget (**Figure 1**). As with translation, thermodynamics do not strictly demand PP_i generation and hydrolysis, as transcription can operate with NDPs *in vitro* (Gottesman and Mustaev, 2019). PP_i production during DNA synthesis only accounts for 1% of the cell's energy budget (**Figure 1**), whereby some DNA polymerases can also operate with NDP substrates (Burke and Lupták, 2018). In addition, some polymerases use the irreversible effect of PP_i hydrolysis by possession of a pyrophosphatase domain that cleaves PP_i as the enzyme moves forward (Kottur and Nair, 2018).

PP_i generation and hydrolysis render both the reactions of the core and polymerization reactions during replication, transcription and translation irreversible under the physiological conditions of the cell. In the core, and in the cytosol, this ratchet has locked biochemistry in the direction of cell synthesis during the 4 billion years since metabolic origin. The insight of Kornberg (1962, p. 261) "*Hydrolysis of the latter* [pyrophosphate] *by inorganic pyrophosphatases promotes the irreversibility of the synthetic route to coenzymes, nucleic acids, proteins, and structural carbohydrates and lipids*" still stands.

The Effect of PP_i in Translation Is Demonstrably Kinetic

The vectorializing effect of PP_i in translation is not thermodynamic, it is the same as in the biosynthetic core: a kinetic ratchet afforded by ubiquitous pyrophosphatases that render protein synthesis unidirectional toward growth. We were able to demonstrate this effect by calculating the kinetics of aminoacyl tRNA synthesis using published rate constants obtained from Pope et al. (1998) and from Stockbridge and Wolfenden (2011). The equations are given in **Supplementary Figure S1**, into which we introduced values from the literature to obtain the kinetics (the time dependence of reactant and product concentrations) for the isoleucyl-tRNA synthetase reaction in *E. coli* in the presence of inorganic pyrophosphatase. The result is shown in **Supplementary Figure S2**. The *E. coli* enzyme belongs to the family I or type I PPase, which typically have rate constants on the order of 200–400 s^{-1} (Kajander et al., 2013).

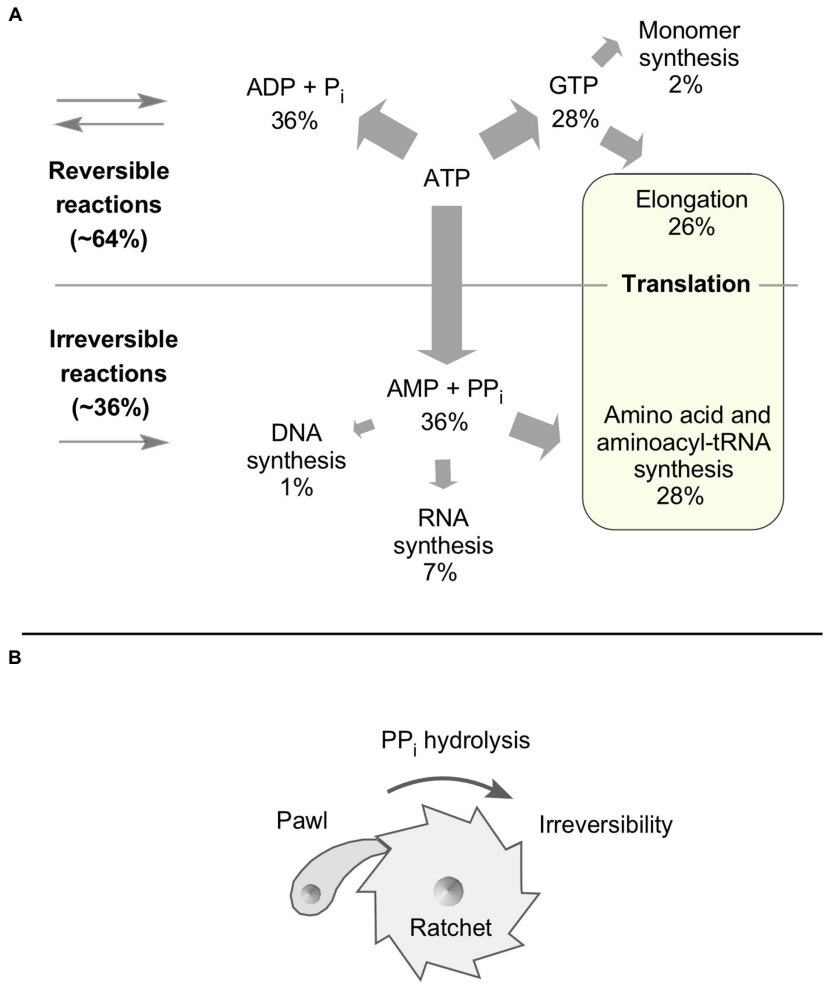

FIGURE 1 | A divide in biosynthetic energy expense. **(A)** Based on **Table 1**, a summary of triphosphate expenses across biosynthetic processes. **(B)** A ratchet and a ratchet's pawl as a mechanism of irreversibility. In metabolism, PP_i hydrolysis functions as the pawl.

Using the *E. coli* sPPase I rate constant of 570 s^{-1} provided by Stockbridge and Wolfenden (2011) in our calculations, PP_i hydrolysis is essentially complete after 0.4 s (**Supplementary Figure S2**). This is a clear result: Pyrophosphatase activity drives the overall reaction of aminoacyl tRNA synthesis forward by removing PP_i at a high rate relative to other steps of the reaction such that the adenylated amino acid is formed irreversibly. The reaction kinetics provides a clear empirical basis for the intuitive effect of product removal during the PP_i forming step of translation.

Different PPases have, however, different rate constants for PP_i hydrolysis. In particular, the membrane bound mPPases are extremely slow; hence, it was of interest to see if they could still provide a similar kinetic effect in the AARS reaction. The mPPases occur in roughly 25% of prokaryotes; the enzyme is also common among protists and is ubiquitous among land plants, where it couples PP_i hydrolysis to the pumping of ions (Na^+ or H^+) out of the cell or, in the case of vacuolar PPases (vPPases), from the cytosol into vacuoles or acidocalcisomes, organelles rich in calcium and polyphosphate (Kajander et al., 2013). The mPPases and vPPases are one to two orders of magnitude slower than type I sPPases, with rates on the order of 3.5–20 s^{-1} (Kajander et al., 2013). Thus, we lowered the rate constant of PP_i hydrolysis by a factor of 100 in our calculations ($k_7 = 5.7\ s^{-1}$; **Supplementary Figure S2**). The kinetic effect was basically the same: Nearly complete conversion (96%) of PP_i to P_i is reached at 20 s ($\tau_{1/2} = 900$ ms). Hence, even very slow pyrophosphatases such as mPPases or vPPases can still drive amino acid activation by AARS enzymes

to near completion. The reaction takes slightly longer than in the case of the type I sPPase, but is still complete in well under a minute. The reasons why the rate constants of the membrane bound PPases are so low is not known (Kajander et al., 2013), and it might be because their ion pumps work reversibly, synthesizing PP_i from $2P_i$ when the cations flow back through the membrane.

There are also type II sPPase (type II sPPase) that are much faster, and they hydrolyze PP_i with rate constants on the order of 1,700–3,000 s^{-1} (Kajander et al., 2013). While all type I sPPases use Mg^{2+} ions to bind PP_i and water to negatively charged amino acids like aspartic acid and polarize water for PP_i hydrolysis by nucleophilic OH^- attack, type II sPPases additionally use Mn^{2+} (or Co^{2+}) for binding and polarization, which probably relates to their higher rate constants. Type II sPPases have only been found in prokaryotes so far; they are common among clostridia and bacilli. Using the rate of 3,000 s^{-1} for a type II sPPase (Kajander et al., 2013), we find that PP_i hydrolysis is 96% complete at 0.35 s and 98% complete at 0.5 s ($\tau_{1/2} = 36$ ms). The reaction rate of type II sPPase drives PP_i hydrolysis to completion in less than second, removing all PP_i substrate for the AARS back reaction. PPases I and II release the PP_i hydrolysis energy of −20 to −25 $kJ{\cdot}mol^{-1}$ as heat in the cytosol. Hence, it is not pyrophosphatase thermodynamics, but its kinetics which drive the amino acid adenylation to high yield.

We point out that that in *E. coli*, PP_i can accumulate to transient concentrations on the order of 1 mM (Kukko and Heinonen, 1982) in exponentially growing cells, which would seem to create a conflict with the idea of a kinetic effect. Yet in natural environments, exponential growth is rarely if ever attained, as discussed in more detail in the next section. In the context of weighing kinetic vs. thermodynamic effects of PP_i production, we also recall the "uncomfortable" observation that deletion of *C. thermocellum* mPPase has no impact upon exponential growth (Holwerda et al., 2020), whereby chemostat cultures of *C. thermocellum*, which have high PP_i concentrations in the cytosol and a PP_i-dependent glycolytic pathway, show clear signs of increased reversibility at the AARS reactions in the form of high concentrations of excreted amino acids (Holwerda et al., 2020). Whether such altruistic amino acid excretion into the environment *via* the AARS reaction would be manifested or sustainable in natural cellulose degrading environments over evolutionary timescales as opposed to chemostat growth conditions designed for biofuel yield is currently not known. It is also noteworthy that cells expend four ATP to generate a peptide bond even though one ATP would suffice, as peptide synthesis from aminoacyl phosphates (Kachalsky and Paecht, 1954) or non-ribosomal peptide synthesis (Martínez-Núñez and López y López, 2016) shows. The energetic difference between the one ATP required to form a peptide bond in solution vs. the four ATP that cells expend to make peptide bonds during translation can be seen as the energetic cost of structural information that is specified within a protein sequence (Haber and Anfinsen, 1962) plus the cost of its irreversible synthesis (Kornberg, 1962).

DISCUSSION

Irreversibility in the Long-Term Evolution of Cells in Nature

What are the consequences of a PP_i irreversibility ratchet over geological timescales? From a physiological and energetic standpoint, the function of PP_i is always subordinate to ATP, because the source of the anhydride bond in PP_i in modern metabolism is always ATP, generated either *via* ion gradients or *via* substrate level phosphorylation. A critic might interject that thylakoid pyrophosphatases might be able to conserve energy as PP_i (Jiang et al., 1997), but if they do, it would be at the expense of one ATP per PP_i formed.

A critic might interject that Heinonen (2001) has summarized evidence to suggest that the measured cellular PP_i levels on the order of 1 mM in logarithmically growing *E. coli* cells are too high to exert a kinetic effect of the kind that Kornberg had in mind. This issue can be illustrated with a passage from Kukko and Heinonen (1982), who measured the intracellular PP_i concentration of *E. coli* grown in batch culture with a doubling time of roughly 1 h. They found that the PP_i concentration was constant at about 0.5 mM during exponential growth. From this, they concluded that "*[…] the metabolic role of PP_i has been clouded by the widespread belief that PP_i formed in the metabolism is rapidly hydrolyzed in cells to inorganic phosphate and the concentration of PPi thus approaches zero in the cytoplasm. This view must be in error.*" We do not doubt their observations; their interpretation is the issue. Kukko and Heinonen (1982) cite several other papers where PP_i concentrations on the order of 0.1–2 mM are reported, always from exponentially growing cells. Why are PP_i concentrations in exponentially growing cells misleading in the context of irreversibility?

When growth stops, so does PP_i production in the cytosol. But even after growth-dependent production of PP_i has ceased, PP_i continues to be hydrolyzed by pyrophosphatase activity. In a rare report, Danchin et al. (1984) measured PP_i after blocking ATP (hence PP_i) synthesis in *E. coli*, and they found that PP_i levels dropped exponentially to 100 µM within a minute and to 10 µM within 10 min, in line with our kinetic calculations (**Supplementary Figure S2**). Bielen et al. (2010) also noted that PP_i levels dropped when cells ceased exponential growth. Why do PP_i levels drop when growth is arrested? It is because PP_i is produced by growth processes (Klemme, 1976), but is hydrolyzed to phosphate by pyrophosphatases continuously, also in resting cells, independent of growth. This leads to a rapid drop in PP_i concentrations, which do in fact approach zero in the cytoplasm, once PP_i production is halted. This is why Danchin et al. (1984) observed a precipitous drop in PP_i concentrations once PP_i production was arrested.

The sources of PP_i in metabolism in typical cells (here, typical means cells that lack PP_i-dependent glycolysis) have been known for decades. Klemme (1976) summarized the main sources of PP_i production in growing *E. coli* and the rates are which it is produced. In the units of µmol per 100 mg biomass, the contributions to PP_i production in exponentially growing *E. coli* were: synthesis of protein (545), nucleic acids (67),

polysaccharides (60), and lipids (60) for a total of 740. Using the fixed relationship between PP_i synthesis and growth in either rich or minimal medium, he was able to obtain good estimates for the rate of PP_i synthesis for several bacteria, which he set in relationship to the measured PPase activity for the same bacteria, allowing him to calculate the ratio of rates ($\mu moles \cdot h^{-1} \cdot mg\ protein^{-1}$) for PP_i production and PP_i hydrolysis. For eight exponentially growing bacterial species (six Gram negative, two Gram positive), he found that the ratio of PP_i hydrolysis to PP_i synthesis was 79, 73, 57, 33, 14, 10, 8, and 1 (Table II of Klemme, 1976). The average value was 35; the value for *E. coli* was 14. That is, the rate of *E. coli* PP_i hydrolysis is 14 times higher than the rate of PP_i production from growth processes. Even if the ratio is only 1, when exponential growth is arrested, pyrophosphatase activity remains, which will relentlessly gnaw away at PP_i concentrations until they essentially reach zero or until rapid growth is resumed, such that rates of cytosolic PP_i production can exceed PP_i hydrolysis. A number of microbes possess Mn^{2+}- or Co^{2+}-dependent sPPases, called family II sPPases, that have a catalytic rate higher than that of typical sPPases (Kajander et al., 2013).

The foregoing raises two important points. The first is that PP_i levels in exponentially growing cells are not a good proxy for the function of PP_i over evolutionary time. This is because sustained exponential growth is never attained for microbes in the environment or during evolution, as they are mainly starved for nutrients in the wild. In the largest microbial community known, marine sediment, cells do not actually grow, they just slowly die as organic nutrients become limiting (Orsi et al., 2020). In such environments, the standard concept of doubling times does not apply to growth or survival, as cell mass never doubles, it just turns over from one living cell to another, with turnover times on the order of tens to thousands of years (Hoehler and Jørgensen, 2013). In starved cells as they exist in sediment, ATP synthesis is orders of magnitude slower than in exponentially growing cells and PP_i production is governed by the rate of protein synthesis, meaning that on time scales of days, months, and years, trace pyrophosphatase activity will hold the cytosolic PP_i concentration close to zero, even if the enzyme's affinity for PP_i is comparatively low. However, measured values of K_m (the substrate concentration at half maximal enzymatic reaction rate) for PP_i for sPPases are not high, and they tend to be in the range of 1 µM to 1 mM in *E. coli* (**Supplementary Table S3**) and values of catalytic rate tend to be on the order of 200–400 s^{-1} (Avaeva, 2000; Kajander et al., 2013), meaning that over long time scales, PPases keep PP_i levels in cells too low to permit the enzymatic reactions of translation or nucleic acid polymerization from running backwards, especially at the extremely slow PP_i production rates of starved microbial communities.

We say "extremely slow PP_i production rates." How slow is slow? We can provide an estimate. Starved cells are small. An exponentially growing *E. coli* cell has about 2 million proteins with on average 300 amino acids each (Milo et al., 2010). If a starved cell is half that size, roughly 300 million peptide bonds are required for its formation. There are 32 million seconds in a year, such that if the turnover time of starved cell is on the order of 10 years, one peptide bond per second is formed, on average, during the formation of the cell. That might not sound too slow, because a ribosome can form about 10 peptide bonds per second. But a small *E. coli* cell has on the order of 10,000 ribosomes, such that in a starved cell with a 10-year turnover time, an individual ribosome might perform an elongation step on the order of roughly once every 3 h or 10^{-4} peptide bonds per second. In terms of molecular processes that are extremely slow, the rate of PP_i production is 100,000 times slower than from translation during exponential growth. This example underscores the value to the cell of translation (aminoacyl tRNA synthesis) being an irreversible process over evolutionary timescales, and the essential function that irreversibility plays by acting as a ratchet's pawl to prohibit the back reaction of aminoacyl tRNA formation, quantitatively the most important energetic expense a cell encounters (**Figure 1**).

A second important point concerns (rare) examples in which sPPases are lacking in the genome such that PP_i reaches high cytosolic concentrations rendering translation theoretically reversible. This can occur in cells that have PP_i-based glycolysis, such as *C. thermocellum*, where ion-pumping membrane-bound PPases (mPPases) are present (Holwerda et al., 2020). In such cells, which are sugar specialists, PP_i levels can exceed 20 mM during chemostat growth, whereby the metabolic source of such high PP_i levels is still unclear and deletion of the *C. thermocellum* mPPase has no impact upon growth (Holwerda et al., 2020). The existence of cells with PP_i-dependent carbon metabolism lacking high activities of sPPases in chemostat growth on very rich medium do not invalidate Kornberg's principle of irreversibility over evolutionary timescales. Rather they constitute an evolutionarily derived special case of adaptation to growth on sugar. As outlined in the introduction, PP_i utilizing glycolytic enzymes tend to occur among microbes that have specialized to sugar-rich environments, including human parasites such as *Entamoeba*, *Giardia*, and trypanosomes, or cellulose- and saccharose-degrading bacteria and archaea, but also in plants with their specialized sugar synthesizing compartment, the plastid. Proton-pumping mPPases are often associated with acidocalcisomes, membrane-bounded compartments that occur in some prokaryotes and in some eukaryotes including parasites and plants (Docampo et al., 2005). The functions discussed for acidocalcisomes include among other things storage of cations, phosphate and polyphosphate, calcium signaling, and osmoregulation but no evidence for an involvement in energy metabolism of acidocalcisomes or their mPPase has emerged so far (Docampo and Huang, 2016).

In the bigger picture of microbial evolution, sugar-dependent lifestyles cannot be ancestral, and they have to be derived. The early earth was barren and offered CO_2, not glucose, as the main environmental carbon source (Schönheit et al., 2016). In the modern crust (Smith et al., 2019) and marine sediments (Orsi et al., 2020), where most cells on Earth have always resided (McMahon and Parnell, 2018), net growth is almost non-existent due to nutrient limitations (Orsi et al., 2020). Particularly in low-energy environments, PP_i irreversibility at the quantitatively dominant (in terms of

PP_i synthesis) AARS reaction acts like a ratchet's pawl that keeps aminoacyl tRNAs moving in solely in the direction of translation, even if translation is slow for reasons of substrate limitation or prolonged starvation.

CONCLUSION

Why Nature Chose Triphosphates

The role of PP_i in evolution raises a question similar to Westheimer's "why phosphate," namely "why triphosphates?" Westheimer (1987) proposed that phosphates became energy carriers because of the metastability of the various bonds that phosphate forms with organic compounds under physiological conditions. By examining the role of PP_i in the core and in the central dogma, we found that the central function of PP_i producing reactions is not that of an energy currency in any case. In metabolism, PP_i is always generated from nucleoside triphosphates. This is also true for mPPases, where the ion gradients that are required for PP_i synthesis are generated at ATP expense. Formulated directly, we find no evidence that PP_i served as a primordial energy currency or that it serves as an energy currency today. Rather, it appears that the role of PP_i is to impart direction upon the most essential operations of life: biosynthesis of cofactors, the biosynthesis of the monomeric building blocks of proteins and nucleic acids, and the polymerization of those building blocks into the catalysts and information carriers of cells (**Figure 1**).

Why did nature specifically choose nucleoside triphosphates as the universal energy currency? That is a fundamentally different question from why nature chose phosphate (Westheimer, 1987; Liu et al., 2019) because many biological compounds harbor phosphate bonds with a large free energy of hydrolysis (Decker et al., 1970), but only triphosphates are the universal energy currency in all lineages today. Irreversibility provides the answer. Triphosphates can generate either P_i or PP_i. This subtle property reveals why triphosphates became the universal energy currency in cells and why they have not been replaced in 4 billion years of evolution (**Figure 2**). How so?

Nucleoside triphosphates such as ATP have two phosphoanhydride bonds. The β-phosphate in ATP can be cleaved on either side (**Figure 2**). ATP-dependent enzymatic reactions that release P_i utilize ATP as a currency of thermodynamic control, making ΔG of the reaction sufficiently negative (or the net activation energy sufficiently low) to allow the reaction to go forward. ATP-dependent reactions that release PP_i also have a thermodynamic component, but the irreversibility of the reaction conferred by PP_i hydrolysis under physiological conditions places the reaction under kinetic rather than thermodynamic control.

No biochemical energy currency other than (nucleoside) triphosphates offers, within the same compound, the alternative of exerting either thermodynamic or kinetic control over a reaction. This property is specific to triphosphates. It allowed primordial enzymes to exert either kinetic control or thermodynamic control over catalyzed reactions, depending upon which anhydride bond of the β-phosphate was cleaved.

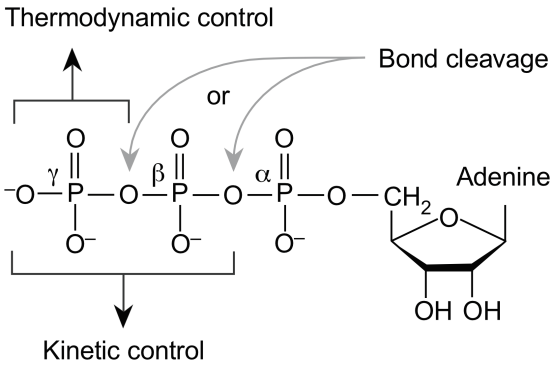

FIGURE 2 | Role of nucleoside triphosphates in metabolic evolution. Coupling of ATP hydrolysis to P_i and ADP or to phosphorylation releases $-32\,kJ{\cdot}mol^{-1}$ under standard physiological conditions, shifting the equilibrium of otherwise mildly endergonic reactions in the direction of product formation. Such reactions are under thermodynamic control, the products are more stable than the educts, but such reactions are usually reversible under physiological conditions of the cell, unless ΔG is very large. Coupling of ATP hydrolysis to PP_i and AMP or to adenylation releases $-45\,kJ{\cdot}mol^{-1}$ under standard physiological conditions, similar to the free energy change for acyl phosphate hydrolysis, but the ubiquitous presence of pyrophosphatase activity in cells leads to immediate PP_i hydrolysis, such that a product for the reverse reaction is removed. Even if the reverse reaction was thermodynamically favorable, it cannot take place because an educt (PP_i) is lacking, making the reverse reaction orders of magnitude slower than the forward reaction, placing it under kinetic control.

This in turn imparted the option of evolutionary refinement of an initial catalytic activity among ATP utilizing enzymes according to the prevailing selective forces in a given cellular environment. An early onset of PP_i-dependent irreversibility in metabolism would not require the presence of a pre-existing inorganic pyrophosphatase enzyme activity at the site of origins, because the reaction can be catalyzed by inorganic ions alone, such as Mg^{2+} (Stockbridge and Wolfenden, 2011), which catalyzes hydrolysis in the active site of many modern pyrophosphatase enzymes (Varfolomeev and Gurevich, 2001).

This, in turn, is the reason why nucleoside triphosphates became fixed in both monomer and polymer biosynthesis in the metabolism of LUCA and have not been displaced since. From an ancestral state in which acyl phosphates provided thermodynamic impetus and a means of energetic coupling in enzymatic reactions, the advent of nucleoside triphosphates changed the nature of early biochemical evolution by introducing the option of kinetic irreversibility. Triphosphates offered primordial enzymes a means to exert either kinetic control or thermodynamic control over catalyzed reactions with one and the same energy currency. The only evident alternative solutions would have been (i) to maintain two distinct energetic currencies in the cell, one for energetic and one for kinetic purposes (an event for which there is no evidence) or (ii) to abandon one of the functions (which is not a viable option over evolutionary time). The ability of triphosphates to function in roughly 2/3 of phosphoanhydride bond expenditure as a currency of energy (thermodynamic drive) and in roughly 1/3 of phosphoanhydride

bond expenditure as a currency of irreversibility (kinetic drive; **Figure 1**) is the reason they became – and remained – life's universal energy currency.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

JW collected and analyzed data, participated in project design, and revised the manuscript. KK performed the kinetic calculations and contributed in data interpretation. WM wrote the first manuscript draft, performed literature research, visualization, and conceived and supervised the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the European Research Council (Advanced Grants eMicrobevol and EcolMetabOrigin to WM), the Deutsche Forschungsgemeinschaft (Ma 1426/21-1 to WM), and the Volkswagen Foundation (VW 96742 to WM).

ACKNOWLEDGMENTS

We thank Harun Tüysüz, William Orsi, Martina Preiner, Joana Xavier, Delfina Pereira, and Andrey do Nascimento Vieira for helpful discussions and the team of the central computing facility at the University of Düsseldorf (ZIM) for their support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.759359/full#supplementary-material

REFERENCES

Albers, S. V., and Meyer, B. H. (2011). The archaea cell envelope. *Nat. Rev. Microbiol.* 9, 414–426. doi: 10.1038/nrmicro2576

Avaeva, S. M. (2000). Active site interactions in oligomeric structures of inorganic pyrophosphatases. *Biochemistry* 65, 361–372.

Baltscheffsky, H., von Stedingk, L. V., Heldt, H. W., and Klingenberg, M. (1966). Inorganic pyrophosphate: formation in bacterial photophosphorylation. *Science* 153, 1120–1122. doi: 10.1126/science.153.3740.1120

Bielen, A. A. M., Willquist, K., Engman, J., van der Oost, J., van Niel, E. W. J., and Kengen, S. W. M. (2010). Pyrophosphate as a central energy carrier in the hydrogenproducing extremely thermophilic Caldicellulosiruptor saccharolyticus. *FEMS Microbiol. Lett.* 307, 48–54. doi: 10.1111/j.1574-6968.2010.01957.x

Bräsen, C., Esser, D., Rauch, B., and Siebers, B. (2014). Carbohydrate metabolism in archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiol. Mol. Biol. Rev.* 78, 89–175. doi: 10.1128/MMBR.00041-13

Burke, C. R., and Lupták, A. (2018). DNA synthesis from diphosphate substrates by DNA polymerases. *Proc. Natl. Acad. Sci. U. S. A.* 115, 980–985. doi: 10.1073/pnas.1712193115

Chen, J., Brevet, A., Fromant, M., Lévêque, F., Schmitter, J.-M., Blanquet, S., et al. (1990). Pyrophosphatase is essential for growth of *Escherichia coli*. *J. Bacteriol.* 172, 5686–5689. doi: 10.1128/jb.172.10.5686-5689.1990

Danchin, A., Dondon, L., and Daniel, J. (1984). Metabolic alterations mediated by 2-ketobutyrate in *Escherichia coli* K12. *Mol. Gen. Genet.* 193, 473–478. doi: 10.1007/BF00382086

de Duve, C. (1991). *Blueprint for a Cell: The Nature and Origin of Life*. Burlington, USA: Neil Patterson Publishers

Decker, K., Jungerman, K., and Thauer, R. K. (1970). Energy production in anaerobic organisms. *Angew. Chem. Int. Ed.* 9, 138–158. doi: 10.1002/anie.197001381

Docampo, R., de Souza, W., Miranda, K., Rohloff, P., and Moreno, S. N. (2005). Acidocalcisomes—conserved from bacteria to man. *Nat. Rev. Microbiol.* 3, 251–261. doi: 10.1038/nrmicro1097

Docampo, R., and Huang, H. (2016). Acidocalcisomes of eukaryotes. *Curr. Opin. Cell Biol.* 41, 66–72. doi: 10.1016/j.ceb.2016.04.007

Eck, R. V., and Dayhoff, M. O. (1966). Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science* 152, 363–366. doi: 10.1126/science.152.3720.363

Gomez, M. A. R., and Ibba, M. (2020). Aminoacyl-tRNA synthetases. *RNA* 26, 910–936. doi: 10.1261/rna.071720.119

Gottesman, M. E., and Mustaev, A. (2019). Ribonucleoside-5′-diphosphates (NDPs) support RNA polymerase transcription, suggesting NDPs may have been substrates for primordial nucleic acid biosynthesis. *J. Biol. Chem.* 294, 11785–11792. doi: 10.1074/jbc.RA119.009074

Haber, B., and Anfinsen, C. B. (1962). Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.* 237, 1839–1844. doi: 10.1016/S0021-9258(19)73945-3

Harold, F. M. (1986). *The Vital Force: A Study of Bioenergetics*. New York, USA: WH Freeman

Heinen, W., and Lauwers, A. M. (1997). The iron-sulfur world and the origins of life: abiotic synthesis from metallic iron, H_2S and CO_2: A comparison of the thiol generating $FeS/HCl(H_2S)/CO_2$-system and its $Fe^0/H_2S/CO_2$-counterpart. *Proc. K. Ned. Akad. van Wet.* 100, 11–25.

Heinonen, J. K. (2001). *Biological Role of Inorganic Pyrophosphate*. New York, USA: Springer US

Hoehler, T. M., and Jørgensen, B. B. (2013). Microbial life under extreme energy limitation. *Nat. Rev. Microbiol.* 11, 83–94. doi: 10.1038/nrmicro2939

Holwerda, E. K., Zhou, J., Hon, S., Stevenson, D. M., Amador-Noguez, D., Lynd, L. R., et al. (2020). Metabolic fluxes of nitrogen and pyrophosphate in chemostat cultures of *clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*. *Appl. Environ. Microbiol.* 86, e01795–e01720. doi: 10.1128/AEM.01795-20

Jiang, S. S., Fan, L. L., JingYang, S., Kuo, S. Y., and Pan, R. L. (1997). Purification and characterization of thylakoid membrane-bound inorganic pyrophosphatase from *Spinacia oleraciaL*. *Arch. Biochem. Biophys.* 346, 105–112. doi: 10.1006/abbi.1997.0279

Kachalsky, A., and Paecht, M. (1954). Phosphate anhydrides of amino acids. *J. Biol. Chem. Soc.* 76, 6042–6044. doi: 10.1021/ja01652a053

Kajander, T., Kellosalo, J., and Goldman, A. (2013). Inorganic pyrophosphatases: one substrate, three mechanisms. *FEBS Lett.* 587, 1863–1869. doi: 10.1016/j.febslet.2013.05.003

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kitadai, N., Nakamura, R., Yamamoto, M., Okada, S., Takahagi, W., Nakano, Y., et al. (2021). Thioester synthesis through geoelectrochemical CO_2 fixation on Ni sulfides. *Commun. Chem.* 4:37. doi: 10.1038/s42004-021-00475-5

Klemme, J.-H. (1976). Regulation of intracellular pyrophosphatase-activity and conservation of the phosphoanhydride-energy of inorganic pyrophosphate in microbial metabolism. *Z. Naturforsch.* 31, 544–550. doi: 10.1515/znc-1976-9-1011

Kornberg, A. (1962). "On the metabolic significance of phosphorolytic and pyrophosphorolytic reactions," in *Horizons in Biochemistry*. eds. H. Kasha and B. Pullman (New York, USA: Academic Press), 251–264.

Kottur, J., and Nair, D. T. (2018). Pyrophosphate hydrolysis is an intrinsic and critical step of the DNA synthesis reaction. *Nucleic Acids Res.* 46, 5875–5885. doi: 10.1093/nar/gky402

Kukko, E., and Heinonen, J. (1982). The intracellular concentration of pyrophosphate in the batch culture of *Escherichia coli*. *Eur. J. Biochem.* 127, 347–349. doi: 10.1111/j.1432-1033.1982.tb06878.x

Lahti, R. (1983). Microbial inorganic pyrophosphatases. *Microbiol. Rev.* 47, 169–178. doi: 10.1128/mr.47.2.169-178.1983

Lever, M. A., Rogers, K. L., Lloyd, K. G., Overmann, J., Schink, B., Thauer, R. K., et al. (2015). Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. *FEMS Microbiol. Rev.* 39, 688–728. doi: 10.1093/femsre/fuv020

Lipmann, F. (1965). "Projecting backward from the present stage of evolution of biosynthesis," in *The Origin of Prebiological Systems and of their Molecular Matrices*. ed. S. W. Fox (New York, USA: Academic Press), 259–280.

Liu, Z., Rossi, J. C., and Pascal, R. (2019). How prebiotic chemistry and early life chose phosphate. *Lifestyles* 9:26. doi: 10.3390/life9010026

Martin, W. F. (2019). Carbon-metal bonds: rare and primordial in metabolism. *Trends Biochem. Sci.* 44, 807–818. doi: 10.1016/j.tibs.2019.04.010

Martin, W. F. (2020). Older than genes: The acetyl CoA pathway and origins. *Front. Microbiol.* 11:817. doi: 10.3389/fmicb.2020.00817

Martin, W., and Russell, M. J. (2007). On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. B* 362, 1887–1925. doi: 10.1098/rstb.2006.1881

Martínez-Núñez, M. A., and López y López, V. E. (2016). Nonribosomal peptide synthetases and their applications in industry. *Sustain. Chem. Process.* 4:13. doi: 10.1186/s40508-016-0057-6

McMahon, S., and Parnell, J. (2018). The deep history of earth's biomass. *J. Geol. Soc.* 175, 716–720. doi: 10.1144/jgs2018-061

Mereschkowsky, C. (1910). Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen. *Biol. Centralbl.* 30, 353–367.

Michels, P. A. M., Bringaud, F., Herman, M., and Hannaert, V. (2006). Metabolic functions of glycosomes in trypanosomatids. *Biochim. Biophys. Acta* 1763, 1463–1477. doi: 10.1016/j.bbamcr.2006.08.019

Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* 38, D750–D753. doi: 10.1093/nar/gkp889

Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* 191, 144–148. doi: 10.1038/191144a0

Müller, M., Mentel, M., van Hellemond, J. J., Henze, K., Woehle, C., Gould, S. B., et al. (2012). Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* 76, 444–495. doi: 10.1128/MMBR.05024-11

Nagata, R., Fujihashi, M., Sato, T., Atomi, H., and Miki, K. (2018). Identification of a pyrophosphate-dependent kinase and its donor selectivity determinants. *Nat. Commun.* 9:1765. doi: 10.1038/s41467-018-04201-z

Neidhardt, F. C., Ingraham, J. L., and Schaechter, M. (1990). *Physiology of the Bacterial Cell*. Sunderland, MA, USA: Sinauer Associates

Orsi, W. D., Schink, B., Buckel, W., and Martin, W. F. (2020). Physiological limits to cell mass conversion in anoxic subseafloor sediment. *FEMS Microbiol. Rev.* 44, 219–231. doi: 10.1093/femsre/fuaa004

Pascal, R., Pross, A., and Sutherland, J. D. (2013). Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open Biol.* 3:130156. doi: 10.1098/rsob.130156

Pasek, M. A., Gull, M., and Herschy, B. (2017). Phosphorylation on the early earth. *Chem. Geol.* 475, 149–170. doi: 10.1016/j.chemgeo.2017.11.008

Pérez-Castiñeira, J. R., Docampo, R., Ezawa, T., and Serrano, A. (2021). Pyrophosphates and polyphosphates in plants and microorganisms. *Front. Plant Sci.* 26:653416. doi: 10.3389/fpls.2021.653416

Pérez-Castiñeira, J. R., Lopez-Marques, R. L., Villalba, J. M., Losada, M., and Serrano, A. (2002). Functional complementation of yeast cytosolic pyrophosphatase by bacterial and plant H^+-translocating pyrophosphatases. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15914–15919. doi: 10.1073/pnas.242625399

Piast, R. W., Garstka, M., Misicka, A., and Wieczorek, R. M. (2020). Small cyclic peptide for pyrophosphate dependent ligation in prebiotic environments. *Lifestyles* 10:103. doi: 10.3390/life10070103

Pope, A. J., Moore, K. J., McVey, M., Mensah, L., Benson, N., Osbourne, N., et al. (1998). Characterization of isoleucyl-tRNA synthetase from *Staphylococcus aureus* II. Mechanism of inhibition by reaction intermediate and pseudomonic acid analogues studied using transient and steady-state kinetics. *J. Biol. Chem.* 273, 31680–31690. doi: 10.1074/jbc.273.48.31691

Preiner, M., Igarashi, K., Muchowska, K. B., Yu, M., Varma, S. J., Kleinermanns, K., et al. (2020). A hydrogen dependent geochemical analogue of primordial carbon and energy metabolism. *Nat. Ecol. Evol.* 4, 534–542. doi: 10.1038/s41559-020-1125-6

Reeves, R. E. (1968). A new enzyme with the glycolytic function of pyruvate kinase. *J. Biol. Chem.* 243, 3202–3204. doi: 10.1016/S0021-9258(18)93395-8

Reeves, R. E. (1984). Metabolism of *Entamoeba histolytica* Schaudinn, 1903. *Adv. Parasitol.* 23, 105–142. doi: 10.1016/s0065-308x(08)60286-9

Russell, J. B. (2007). The energy spilling reactions of bacteria and other organisms. *J. Mol. Microbiol. Biotechnol.* 13, 1–11. doi: 10.1159/000103591

Russell, M. J., Barge, L. M., Bhartia, R., Bocanegra, D., Bracher, P. J., Branscomb, E., et al. (2014). The drive to life on wet and icy worlds. *Astrobiology* 14, 308–343. doi: 10.1089/ast.2013.1110

Russell, J. B., and Cook, G. M. (1995). Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol. Rev.* 59, 48–62. doi: 10.1128/mr.59.1.48-62.1995

Russell, M. J., and Hall, A. J. (1997). The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc. Lond.* 154, 377–402. doi: 10.1144/gsjgs.154.3.0377

Russell, M. J., Nitschke, W., and Branscomb, E. (2013). The inevitable journey to being. *Philos. Trans. R. Soc. B Biol. Sci.* 368:20120254. doi: 10.1098/rstb.2012.0254

Satpati, P., Sund, J., and Åqvist, J. (2014). Structure-based energetics of mRNA decoding on the ribosome. *Biochemistry* 53, 1714–1722. doi: 10.1021/bi5000355

Schomburg, I., Chang, A., and Schomburg, D. (2002). BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* 30, 47–49. doi: 10.1093/nar/30.1.47

Schönheit, P., Buckel, W., and Martin, W. (2016). On the origin of heterotrophy. *Trends Microbiol.* 24, 12–25. doi: 10.1016/j.tim.2015.10.003

Schramm, G., Grotsch, H., and Pollmann, W. (1962). Nicht-enzymatische Synthese von Polysacchariden, Nucleosiden und Nucleinsäuren und die Entstehung selbst-vermehrungsfähiger Systeme. *Angew. Chem.* 74, 53–92. doi: 10.1002/ange.19620740202

Semenov, S. N., Kraft, L. J., Ainla, A., Zhao, M., Baghbanzadeh, M., Campbell, V. E., et al. (2016). Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. *Nature* 537, 656–660. doi: 10.1038/nature19776

Serrano, A., Pérez-Castiñeira, J. R., Baltscheffsky, M., and Baltscheffsky, H. (2007). H^+-PPases: yesterday, today and tomorrow. *IUBMB Life* 59, 76–83. doi: 10.1080/15216540701258132

Siebers, B., and Schönheit, P. (2005). Unusual pathways and enzymes of central carbohydrate metabolism in archaea. *Curr. Opin. Microbiol.* 8, 695–705. doi: 10.1016/j.mib.2005.10.014

Smith, A. R., Kieft, B., Mueller, R., Fisk, M. R., Mason, O. U., Popa, R., et al. (2019). Carbon fixation and energy metabolisms of a subseafloor olivine biofilm. *ISME J.* 13, 1737–1749. doi: 10.1038/s41396-019-0385-0

Sousa, F., and Martin, W. F. (2014). Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochim. Biophys. Acta* 1837, 964–981. doi: 10.1016/j.bbabio.2014.02.001

Sousa, F. L., Thiergart, T., Landan, G., Nelson-Sathi, S., Pereira, I. A. C., Allen, J. F., et al. (2013). Early bioenergetic evolution. *Philos. Trans. R. Soc. B* 368:20130088. doi: 10.1098/rstb.2013.0088

Stockbridge, R. B., and Wolfenden, R. (2011). Enhancement in the rate of pyrophosphate hydrolysis by nonenzymatic catalysts and by inorganic pyrophosphatase. *J. Biol. Chem.* 286, 18538–18546. doi: 10.1074/jbc.M110.214510

Stouthamer, A. H. (1978). "Energy-yielding pathways," in *The Bacteria*. ed. I. C. Gunsalus (New York, USA: Academic Press), 389–462.

Tempest, D. W., and Neijssel, O. M. (1984). The status of Y_{ATP} and maintenance energy as biologically interpretable phenomena. *Annu. Rev. Microbiol.* 38, 459–486. doi: 10.1146/annurev.mi.38.100184.002331

Varfolomeev, S. D., and Gurevich, K. G. (2001). Enzyme active sites: bioinformatics, architecture, and mechanisms of action. *Russ. Chem. Bull.* 50, 1709–1717. doi: 10.1023/A:1014353724442

Wächtershäuser, G. (1992). Groundworks for an evolutionary biochemistry—the iron sulfur world. *Prog. Biophys. Mol. Biol.* 58, 85–201. doi: 10.1016/0079-6107(92)90022-X

Wang, Q., Barge, L. M., and Steinbock, O. (2019). Microfluidic production of pyrophosphate catalyzed by mineral membranes with steep pH gradients. *Chem. Eur. J.* 25, 4732–4739. doi: 10.1002/chem.201805950

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., et al. (2016). The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* 1:16116. doi: 10.1038/nmicrobiol.2016.116

Westheimer, F. H. (1987). Why nature chose phosphates. *Science* 235, 1173–1178. doi: 10.1126/science.2434996

White, H. B. 3rd (1976). Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* 7, 101–104. doi: 10.1007/BF01732468

Wimmer, J. L. E., Vieira, A. D. N., Xavier, J. C., Kleinermanns, K., Martin, W. F., and Preiner, M. (2021). The autotrophic core: An ancient network of 404 reactions converts H_2, CO_2, and NH_3 into amino acids, bases, and cofactors. *Microorganisms* 9:458. doi: 10.3390/microorganisms9020458

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

IV The metabolic network of the Last Bacterial Common Ancestor

Joana C. Xavier, Rebecca E. Gerhards, **Jessica L. E. Wimmer**, Julia Brueckner, Fernando D. K. Tria, William F. Martin

Institut für Molekulare Evolution, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Deutschland.

Dieser Artikel wurde am 26.03.2021 in *Communications Biology* Ausgabe 4 veröffentlicht.

Beitrag von Jessica L. E. Wimmer (Zweitautor):
Die Rekonstruktion des metabolischen Netzwerkes wurde von mir durchgeführt. Dazu zählen die Konstruktion einer Reaktionsdatenbank, die Anwendung des Expansions-Algorithmus und die Visualisierung der Ergebnisse. Des Weiteren habe ich die Verwandtschaftsrekonstruktion verschiedener Attribute (engl. *ancestral state reconstruction*) und deren Visualisierung durchgeführt. Ebenfalls habe ich die Vertikalität der potentiellen LBCA-Gene analysiert, diese mit dem metabolischen Netzwerk kombiniert und visualisiert. Außerdem habe ich zum Design weiterer Abbildungen und Tabellen beigetragen und Teile des Manuskriptes geschrieben.

communications biology

ARTICLE

https://doi.org/10.1038/s42003-021-01918-4 OPEN

The metabolic network of the last bacterial common ancestor

Joana C. Xavier[1,2✉], Rebecca E. Gerhards[1,2], Jessica L. E. Wimmer[1], Julia Brueckner[1], Fernando D. K. Tria[1] & William F. Martin[1]

Bacteria are the most abundant cells on Earth. They are generally regarded as ancient, but due to striking diversity in their metabolic capacities and widespread lateral gene transfer, the physiology of the first bacteria is unknown. From 1089 reference genomes of bacterial anaerobes, we identified 146 protein families that trace to the last bacterial common ancestor, LBCA, and form the conserved predicted core of its metabolic network, which requires only nine genes to encompass all universal metabolites. Our results indicate that LBCA performed gluconeogenesis towards cell wall synthesis, and had numerous RNA modifications and multifunctional enzymes that permitted life with low gene content. In accordance with recent findings for LUCA and LACA, analyses of thousands of individual gene trees indicate that LBCA was rod-shaped and the first lineage to diverge from the ancestral bacterial stem was most similar to modern Clostridia, followed by other autotrophs that harbor the acetyl-CoA pathway.

[1] Institute for Molecular Evolution, Heinrich-Heine-University, 40225 Düsseldorf, Germany. [2] These authors contributed equally: Joana C. Xavier, Rebecca E. Gerhards. ✉email: xavier@hhu.de

Among all cells on Earth[1], bacteria are not only the most abundant, they comprise the most diverse domain in terms of physiology and metabolism[2] and are generally regarded as ancient[3–5]. Isotopic signatures trace autotrophy 3.9 billion years back in time[6]. Based on the universality of the genetic code, amino acid chirality, and universal metabolic currencies, there is an agreement that a last universal common ancestor (LUCA) predated the divergence of bacteria and archaea. Because the bacterial and archaeal domains are monophyletic, there is evidence for one clear ancestor for each domain—the last bacterial common ancestor (LBCA) and the last archaeal common ancestor (LACA). Phylogenomic reconstructions indicate that LUCA was a thermophilic anaerobe that lived from gasses in a hydrothermal setting[7], notwithstanding contrasting views[8,9]. Both phylogenomics and geological evidence indicate that LACA was a methanogen[10–12], or a similar anaerobic autotroph that fixed carbon via the Wood–Ljungdahl (also known as acetyl-CoA) pathway[12]. Reconstructing the habitat and lifestyle of LBCA is, however, impaired by lateral gene transfer (LGT)[13], which decouples physiological evolution from ribosomal phylogeny. Like LUCA and LACA, LBCA must have been an anaerobe, because the accrual of atmospheric oxygen occurred much later in Earth's history, as a product of cyanobacterial metabolism[14–16]. Although some details of Earth's oxygenation continue to be debated, it is generally accepted that the Great Oxidation Event occurred ~2.4 billion years ago[4,16,17]. The most important difference between anaerobes and aerobes is related to energy; anaerobic pathways such as fermentation, sulfate reduction, acetogenesis, and methanogenesis yield only a fraction of the energy when compared to aerobic pathways[18], but this is compensated by the circumstance that the synthesis of biomass costs 13 times more energy per cell in the presence of O_2 than under anoxic conditions. This is because, in the reaction of cellular biomass with O_2, the thermodynamic equilibrium lies very far on the side of CO_2. That is, the absence of O_2 offers energetic benefits of the same magnitude as the presence of oxygen does[19–21]. Although the advent of O_2 expanded routes for secondary metabolism, allowed novel O_2-dependent steps in existing biosynthetic pathways, and allowed the evolution of new heterotrophic lifestyles by enabling the oxidation of unfermentable substrates, the advent of O_2 did not alter the nature of life's basic building blocks nor did it redesign their biosynthetic pathways[22,23]. It did, however, promote LGT for genes involved in O_2 utilization[24]. In other words, the fundamentals of biochemistry, metabolism, and physiology were invented in a time when the Earth was anoxic.

Both from the geochemical and the biological standpoint, looking back into the earliest phases of evolution ca. 4 billion years ago is challenging. The geological challenge is that rocks of that age are generally rare, and those that bear traces of life are extremely scarce. The biological challenge is that LGT has reassorted genes across genomes for 4 billion years. As an alternative to reconstructing gene history, metabolic networks themselves harbor independent inroads to the study of early evolution[25]. Metabolic networks represent the set of chemical transformations that occur within a cell, leading to both energy and biomass production[26]. Genome-scale metabolic networks are inferred from a full genome and the corresponding full set of functional (metabolic) annotations[27], allowing for predictive models of growth and insights into physiology[28]. Furthermore, metabolism itself is connected to the informational processing machine in the cell, because enzymes are coded in DNA, transcribed, and translated, while they also produce the building blocks of DNA and RNA and polymerize them. However, metabolism is much more versatile than information processing. Metabolic networks include multiple redundant paths, and in different species, different routes can lead to the same functional outcome. Because metabolism is far more variable across lineages than the information processing machinery, the genes coding for enzymes are not universal across genomes and are much more prone to undergo LGT than information processing genes are[29]. This circumstance has impaired the use of metabolic enzymes for the study of early prokaryotic evolution.

Metabolic networks and metabolic enzymes unquestionably bear witness to the evolutionary process, but methods to harness their evolutionary information are so far lacking. Here we take a simple but effective approach at inferring the metabolism of LBCA, by focusing on anaerobic genomes and genes that are widely distributed among them. We reconstruct the core metabolic network of LBCA independent of any single backbone phylogenetic tree[30] for the lineages in question. In doing so, we harness the information in thousands of individual trees for gene families of anaerobic prokaryotes, analyze converging signals, and point to the modern groups most similar, in terms of metabolism, to the groups that diverged earliest from LBCA.

Results

Conservation in anaerobic groups unveils LBCA's physiology. To identify genes tracing to the LBCA, we started from 5443 reference genomes from bacteria and selected those 1089 classified as anaerobic by virtue of lacking oxygen reductases[31] and having >1000 protein sequences (to exclude energy parasites; Supplementary Data 1 and Supplementary Table 1). The resulting genomes contained 2,465,582 protein sequences that were then clustered into 114,326 families. Of these, 146 families have at least one sequence present in all the 25 major taxonomic groups analyzed. These groups correspond roughly to phyla in GenBank taxonomy, with the exception of Proteobacteria and Firmicutes, which we split into Classes due to their high representation in the dataset. It is worth mentioning that the abundance of Firmicutes and Proteobacteria is not only a result of taxonomic oversampling but is also a reflection of their orders-of-magnitude larger abundance in natural habitats[32]. Upon closer inspection, the families were present in most of the genomes in the analysis, with 122 of the 146 present on average in at least 90% of all genomes in a group (Supplementary Data 2 and Supplementary Fig. 1). These genes are nearly universal and are among the most vertically inherited genes in prokaryotes (Table 1). These 146 families were rechecked manually with regards to functional annotation (Supplementary Data 3) to provide a list of gene functions that trace to LBCA. Around half of those families are involved in information processing, protein synthesis, or other structural functions (Table 1), and the other half can be mapped to at least one metabolic reaction in KEGG, the Kyoto Encyclopedia of Genes and Genomes (even if often also involved in information processing, e.g., the transfer RNA (tRNA) charging category), thus providing insights into LBCA's physiology and lifestyle.

Various lines of evidence suggest that the first cells were autotrophs that generated acetyl-CoA and pyruvate via the acetyl-CoA pathway[33–35] and sugars via gluconeogenesis[36–38]. LBCA possessed a nearly complete trunk gluconeogenetic pathway with pyruvate kinase (PK), enolase, phosphoglycerate kinase (PGK), glyceraldehyde 3-phosphate dehydrogenase, and triosephosphate isomerase. Phosphoglycerate mutases, which can be either 2,3-bisphosphoglycerate-dependent or cofactor-independent, escape the criteria of universality, but are highly distributed, the former in 21, the latter in 18 of the 25 bacterial groups sampled. Because the PK reaction is reversible in eukaryotes in vivo[39] and in

Table 1 Functional categories for the 146 LBCA protein families.

Functional category	Number of protein families	Average family size	Average verticality
Ribosomal proteins	27	1082	12.260
Translation	17	1083	11.803
tRNA charging	16	1058	12.618
DNA recombination and repair	10	1055	13.165
DNA replication	9	1025	12.669
tRNA modification	9	1075	11.036
Transcription	3	1091	16.123
rRNA modification	5	1056	9.513
Carbohydrate and energy metabolism	10	1062	9.422
Protein modification, folding, sorting, and degradation	9	1113	9.727
Lipid and cell wall metabolism	8	1020	9.473
Nucleotide metabolism	7	1073	10.712
Metabolism of cofactors and vitamins	6	901	7.797
Amino acid metabolism	5	917	9.765
Membrane protein targeting	3	984	13.823
Cell division	2	1060	14.946

For each category, the number of protein families annotated, the average family size, and the average verticality (higher meaning less subject to LGT; see "Methods") are shown.

bacteria[40], bacterial PK likely functioned in the gluconeogenetic direction to provide LBCA with phosphoenolpyruvate for amino acid and peptidoglycan synthesis[41] and carbon backbones with more than three carbon atoms in an early Earth environment rich in CO_2[42]. Four other kinases in addition to PK and PGK trace to LBCA, two involved in cofactor metabolism and two in phosphorylating ribonucleotides to nucleoside diphosphates, whose further activation to LBCA's NTPs could have been carried out via substrate promiscuity of PK, as it occurs in anaerobically grown *Escherichia coli*[43]. Also tracing to LBCA are two enzymes involved in cell division, FtsH and FtsY, which however also fulfill a number of other functions in the cell including protein degradation and assembly[44] and correct targeting of proteins and ribosomes to the membrane[45]. Three other membrane-targeting proteins can be traced to LBCA: Ffh, YidD, and SecA of the sec pathway. One validation of our analysis is the absence of important genes in LBCA's families that were lost in the ancestor of particular groups, for example, FtsZ, present in only 24 out of 25 of the taxonomic groups in our dataset, consistently with previous reports of its loss in Chlamydiae[46].

Only nine compounds were required to complete intermediary metabolism in LBCA. The list of LBCA genes is conservative because our criteria, although not imposing bacterial universality, do require the presence in 25 higher taxonomic groups. However, even though the list is short, the 146 protein families of LBCA generate a tightly connected metabolic network (Supplementary Fig. 2) of 243 compounds with only one reaction (diaminopimelate epimerase) out of 130 disconnected from the rest (Supplementary Data 4A). The network is close to complete in that it generates 48 of the 57 universally essential prokaryotic metabolites[47]: the 20 amino acids, four DNA bases, four RNA bases, eight universal cofactors, glycerol 3-phosphate as a lipid precursor, and 20 charged tRNAs (Supplementary Data 4B). The compounds missing are the charged tRNAs for Lys, Met, Ile, Pro, Asn, Gly, and Gln and two cofactors (thiamine diphosphate and pyridoxal 5-phosphate). Using a network expansion algorithm[48], adding all reactions encoded by non-LBCA genes to the network, and then sequentially and gradually removing them until the production of all universal metabolites was possible with the minimal set of reactions (see "Methods"), we found that the addition of only nine genes—seven aminoacyl tRNA synthetases (aaRS), ADP: thiamine diphosphate phosphotransferase and D-ribulose 5-phosphate, D-glyceraldehyde 3-phosphate pyridoxal 5′-phosphate-lyase—completes the network to generate all 57 universal compounds (Fig. 1 and Supplementary Data 4). It is likely that ancestors of the two classes of aaRS enzymes acted promiscuously in charging tRNA in LBCA[49]. The network is not self-generated from an initial set of nutrients[50]. It would have required additional genes derived from LUCA[7] and lost in some lineages of anaerobic bacteria (including transporters, completely absent in the set of 146 genes) and compounds from geochemical synthesis[34,35] to be a completely functional genome-scale metabolic network. However, the majority of the core of cellular metabolism is represented in the network.

LBCA's network is highly structured around three major metabolic hubs: (i) ATP/diphosphate, (ii) NADP(H)/H+, and (iii) CO_2/ACP/malonyl-ACP. These represent the cores of (i) energy, (ii) hydride transfer, and (iii) carbon metabolism of LBCA (Fig. 1). Malonyl-ACP is central in the initiation and regulation of fatty acid biosynthesis[51]. When we remove PK from the set of enzymes, the phosphorylation of dADP to dATP is no longer possible, suggesting that PK may have acted promiscuously in early nucleotide phosphorylation[43,52]. The connectivity of ATP mainly involves tRNA charging and protein synthesis (Fig. 1), which might seem unexpected at first, because ATP is the universal currency in all of the metabolism. In modern anaerobes, although, roughly 90% of the cell's energy budget is devoted to protein synthesis[21], and similar appears to have applied to LBCA as well.

The first lineages to diverge were most similar to modern Clostridia. The deepest split in the bacterial trees can identify lineages and traits that reflect LBCA's lifestyle. Lineages such as Aquificae and Thermotogae were long considered early branching based on trees of ribosomal proteins and ribosomal RNA (rRNA)[53], but the ribosome cannot speak to the physiology of LBCA because LGT decouples ribosomal evolution from physiology. LGT is extremely frequent within and between most bacterial groups[13], it hinders the inference of the bacterial root via traditional phylogenetic analysis by introducing conflicting signals that reduce verticality. To mitigate the effect of LGT, we examined the relative order of emergence for the 25 bacterial groups using 63,324 trees rooted with minimal ancestor deviation (MAD)[54]. In current practice, the majority of root inferences for the domain Bacteria have been done with outgroup rooting[55,56]. Our choice of an outgroup-independent rooting

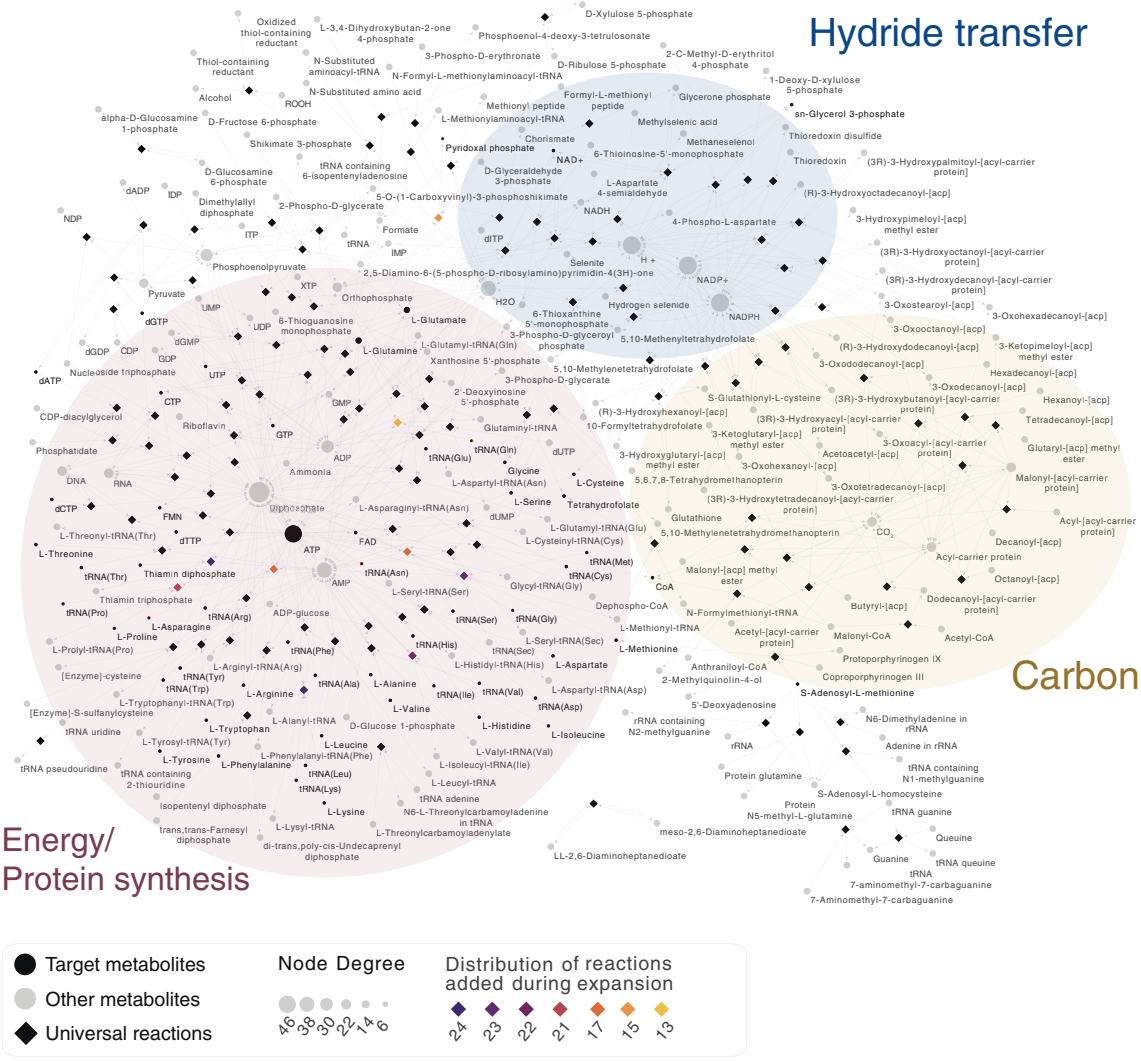

Fig. 1 Metabolic network of LBCA expanded with 9 genes to include 57 universal biomolecules. Metabolic interconversions encoded by 146 LBCA genes plus 9 genes present in fewer groups are shown in a bipartite graph, with 243 metabolites (circular nodes) and 130 reactions (diamond nodes). Black circles represent the 57 universal target metabolites and gray circles represent the remaining metabolites. Note, however, that some of these are also universal (e.g., NADH), but directly connected to the chosen targets (e.g., in that case NAD^+). Node sizes increase according to node degree. Diamonds (reactions) are colored according to the presence of genes encoding for those reactions in different taxonomic groups: in black, reactions present in all taxa; in a gradient from purple to orange reactions added during network expansion and distributed in fewer taxa (target compounds are highlighted with the same outline color if they were introduced with network expansion). Transparent colored ellipses highlight the core of energy (red) hydride transfer (blue) and carbon (yellow) metabolism.

method applied to multiple gene trees is threefold: (i) LGT between Archaea and Bacteria confounds results[13,57,58]; outgroup sequences are notoriously prone to long-branch phylogenetic artifacts[59]; and lack of criteria to assess the quality of different roots, which is possible with MAD. Independent studies have recently shown that the MAD method is more efficient than other rooting methods and robust to a wide spectrum of phylogenetic parameters, both with simulated and empirical prokaryotic gene trees[60].

We started by focusing on the trees for the 146 LBCA protein families, and we analyzed the divergence accumulated from the bacterial root to each modern genome, measured as root-to-tip distance in terms of (i) sequence divergence (branch length) and (2) node depth (Fig. 2) (15 trees with ambiguous root inferences were discarded; root ambiguity indexes given in Supplementary Data 3; see "Methods"). The results identify clostridial genomes as the least diverged both in terms of sequence divergence (Wilcoxon's signed-rank test with Bonferroni correction, largest p value $< 1e-5$, average normalized distance 0.299) and node depth (Wilcoxon's signed-rank test with Bonferroni correction, largest p value < 0.05, average normalized distance 0.116; Supplementary Fig. 3), followed by Deltaproteobacteria (average normalized divergence 0.354, and average normalized depth 0.156). Anaerobic members of Aquificae also show significant proximity to the root as judged by branch length (average normalized distance 0.382, Supplementary Fig. 3). There are only three genomes of (anaerobic) Aquificae in our

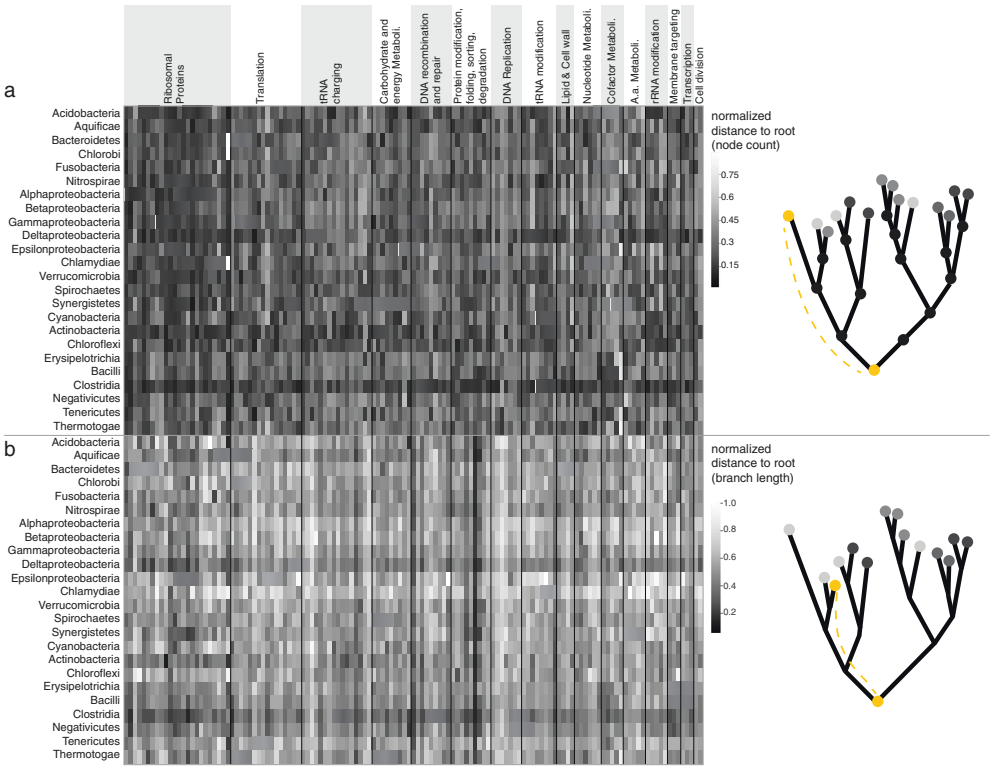

Fig. 2 Divergence analyses for 1089 anaerobic genomes using 131 universal trees reveal clostridial species are closer to the root. Analysis of 131 rooted trees of genes universally present in bacterial anaerobic taxa spanning major functional categories (sorted horizontally according to curated classifications shown on top; order as in Supplementary Data 3). Illustrative trees on the side portray the metric used in each analysis and identify the group at the root in each with yellow nodes. **a** Root-to-tip distance measured as node depth (normalized by the largest distance in each tree). **b** Root-to-tip distance measured as branch length (normalized by the largest distance in each tree).

dataset, and all three belong to chemolithoautotrophs isolated from hydrothermal vents that can grow on H_2 and CO_2[61]. The divergence values for all genomes in all trees ranked from least to most distant show that the top-ranking 12 genomes are all thermophilic species belonging to the class Clostridia, several possessing the acetyl-CoA pathway (Supplementary Table 2). The results shown in Fig. 2 are not dependent on genome abundance in the dataset (the most abundant group is Bacilli, with 38% of all genomes; Supplementary Table 1).

Prokaryotic gene trees differ from the species tree due both to random phylogenetic errors and to the cumulative impact of LGT[62]. In the absence of LGT, gene lineages branch together (monophyletic) and the phylogenetic diversity of sister clades reflects the time since their origin, with older lineages having higher sister diversity. In the context of gene evolution with LGT, gene lineages branch into multiple clades, with the number of clades increasing with gene transfer prevalence. Because LGT is a continuous phenomenon in prokaryotic evolution, the taxonomic labels of sister lineages change dynamically, but their phylogenetic diversity gives us the means to infer the relative timing for the origin of lineages. To integrate the information of sister relation from all gene trees spanning the 25 bacterial groups, we scored the phylogenetic diversity for sister clades of each group in the individual trees permitting as many inter-group LGT as necessary in the trees (5402 trees with at least six groups, Fig. 3 and Supplementary Data 5). The analyses show Clostridia as the group with the highest sister clade diversity, measured as the maximum number of phyla in a sister clade (on average five), followed by a tie between Deltaproteobacteria, Bacilli, Actinobacteria, and Spirochaetes all with three distinct groups on average present in sister clades. The result stands when looking at the 131 universal trees only, where Clostridia has on average nine distinct sister groups, followed by Actinobacteria with seven and Deltaproteobacteria with five (Supplementary Data 6). Maximum-likelihood ancestral state reconstructions using 131 universal trees indicate that LBCA was a rod-shaped cell (Supplementary Fig. 4) and reconstructs Clostridia as the most ancestral lineage (Supplementary Fig. 5) in agreement with the previous analyses.

The analyses so far suggest that the 146 protein families conserved in all groups of anaerobic bacteria were present in LBCA, not only due to their ubiquitous and nearly universal nature (Supplementary Fig. 1) but also because they form a functional unit: a highly connected, nearly complete core metabolic network (Fig. 1). But is the ubiquitous nature of these genes caused by their antiquity, or is it the result of LGT? To address this question, we obtained all values of verticality for prokaryotic gene families[29] as a proxy to measure the gene's tendency to undergo or resist LGT. LBCA's protein families are distinctively and significantly (Kolgomorov–Smirnov statistic = 0.99, p value = 2.4e − 318) more vertical than the average prokaryotic protein family (Fig. 4a, Supplementary Data 7, and Table 1). The metabolic network annotated with verticality values shows that genes involved both in metabolism and information processing (as aaRSs) are highly vertical (Fig. 4b and Supplementary Data 7). Although the most vertically evolving genes in prokaryotic genomes, those for ribosomal proteins, are not involved in specific biosynthesis and

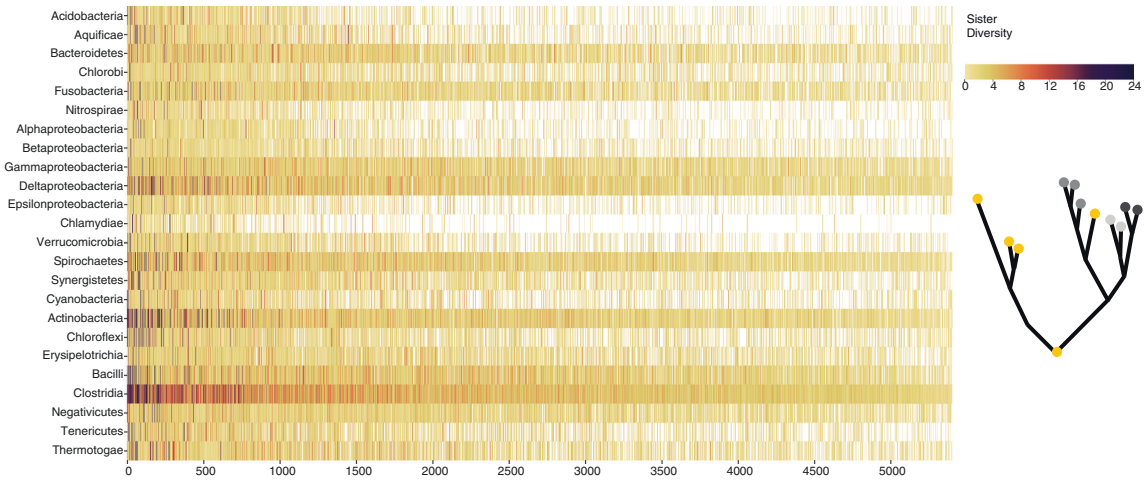

Fig. 3 Sister diversity analysis of 5402 phylogenetic trees reveals Clostridia is the most ancestral group. Sister diversity (maximum number of different groups in the sister clade) for each group (rows) for 5402 trees with at least six groups (columns). An illustrative tree portrays the question asked in the analyses, where the yellow group is the one with the highest sister diversity score and therefore inferred as most ancestral.

hence not represented in metabolic maps, the metabolic functions most closely associated with protein synthesis, those of aaRSs, build the core of a metabolic network that is vertical in nature and thus ubiquitous due to antiquity, not transfer (Fig. 4) and hence ancestral to the domain Bacteria.

Discussion

By investigating the genomes of anaerobic bacteria, we were able to obtain inferences about the metabolism and physiology of LBCA. Our results indicate that LBCA was autotrophic, gluconeogenetic, and rod-shaped. Our analyses of trees for all genes, not just those universally present in all genomes, point to Clostridia (a class within the phylum Firmicutes) as the modern bacterial group most similar to the first lineages, which diverged from LBCA. This result contrasts with previous analyses placing other groups at the root based on concatenated protein phylogeny[53,56,63,64], but it is consistent with early proposals based on the evolution of tetrapyrrole synthesis[65], with studies that place the broader taxon of Firmicutes deep-branching in bacterial trees[37,66] and with the proposal of a rod-shaped Gram-positive ancestor for bacteria[67], and, more recently, for Firmicutes[68]. Why do our inferences on the root of the bacterial tree contrast with different roots[63,64] proposed in other recent analyses? First, our results are based on genome data for cultured organisms with high-quality and complete genomes, and are therefore independent of binning procedures inherent to metagenomic data[69]. In addition, our data are based on genomes for anaerobic bacteria available to date, and is thus less prone to LGT effects associated with the rise of oxygen[24]. The assumption that LBCA was anaerobic is supported by geochemical[14,17] and phylogenomic[4,16,24] evidence, and it undoubtedly reduces phylogenetic noise that would be introduced with late-coming aerobic sequences. Furthermore, our results do not rest upon one or two branches in a single concatenated or consensus tree based on ribosomal sequences, an approach that notwithstanding long tradition has strong potential problems[30], not the least of which is that with concatenated alignments, different methods give fully resolved but conflicting trees, making the results dependent on ad hoc site filtering procedures and specific maximum-likelihood parameters[70].

Our results are internally consistent, based on the convergence of signals from multiple individual trees for individual protein families (with statistical support, Supplementary Fig. 3). In addition, the core set of 146 families trace to LBCA through multiple lines of evidence: (i) the families are universally present in all taxonomic groups analyzed, and (ii) nearly universally present in all genomes analyzed (Supplementary Fig. 1); (iii) they enable a highly connected and nearly complete core metabolic network (Fig. 1); (iv) they are enriched in information processing genes, known to be ancient (Table 1); (v) their functional repertoire (including RNA modifications, multifunctionality, and gluconeogenesis-early) is in accordance with independent studies for LUCA[7] and LACA[12,37]; and (vi) they are among the most vertical genes known (Table 1, Supplementary Data 7, and Fig. 4). The metabolic network enabled by the 146 LBCA genes can be completed for universal essential metabolites with only nine genes, all nine of which are present both in Clostridia and Deltaproteobacteria (Supplementary Data 2).

It has been proposed that Gram-negative bacteria originated from Gram-positive bacteria by an early sporulation event[71], a hypothesis that is compatible with our results. Endospore formation is specific to Firmicutes, implying that if sporulation was an ancient trait, it was subsequently lost before the divergence of most other anaerobic lineages. Spores could have survived in the geologically challenging environments of early Earth[3], and the loss of sporulation in more moderate environments is facile[72].

Other groups showing proximity to the root in the phylogenomic tests we performed are Deltaproteobacteria (all tests), three anaerobic species of Aquificae that are significantly closer to the root by branch length (Figs. 2 and 3 and Supplementary Fig. 3) than other lineages, and Actinobacteria, which rank higher than both Deltaproteobacteria and Aquificae in the sister diversity analysis (Fig. 3). What do these groups have in common? Members of all have the acetyl-CoA pathway for carbon fixation and/or energy metabolism[73]; the only carbon fixation pathway present in both archaea and bacteria that traces to LUCA[7] and that is also present in methanogens, the root of the archaeal tree[10–12]. This physiological trait links LBCA both to LUCA and LACA, and also to anaerobic H_2-dependent growth in hydrothermal environments[7]. Whereas most Deltaproteobacteria use the acetyl-CoA pathway solely for carbon fixation while reducing sulfate for energy metabolism, recent reports show that some members can use the acetyl-CoA pathway for ATP supply as well[74,75]. The divergence patterns herein inferred are fully

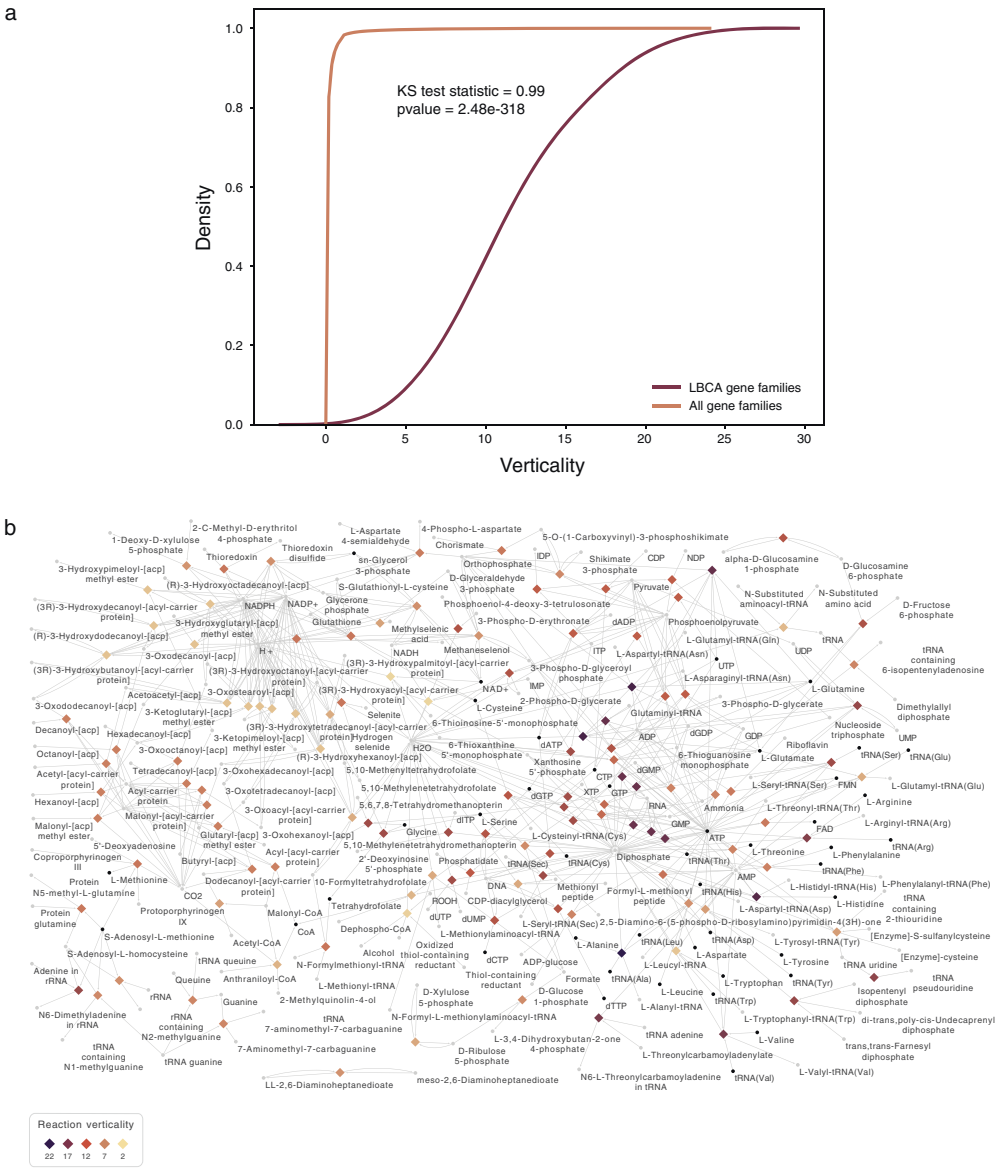

Fig. 4 Analysis of verticality for LBCA gene families. a Verticality for all prokaryotic gene families (light brown) and for LBCA gene families (dark brown) and Kolmogorov–Smirnov statistics between the two distributions. **b** LBCA metabolic network annotated with verticality value for each reaction node.

consistent with the observation that both Clostridia and Deltaproteobacteria are known to be remarkably polyphyletic. Recently, a proposal to divide Deltaproteobacteria into new phyla has been published, confirming that sulfate/sulfite reduction within the class is ancient[76]. Deep-branching Actinobacteria with the Wood–Ljungdahl pathway have recently been uncovered in serpentinizing systems[77]. In terms of physiology, the acetyl-CoA pathway is undoubtedly an ancient biochemical route[78]. By the measure of analyses presented here, several lineages that use it for survival appear to be ancient as well. The reconstruction of LBCA's metabolism reveals the presence of several multifunctional enzymes, reducing the number of genes required for its viability, an important evolutionary consequence of ancestral enzyme promiscuity[79] and possibly a general strategy among the earliest prokaryotes. The physiology of LBCA reconstructed from anaerobes reveals traits well suited to the inhospitable environment of the early Earth[42].

Methods

Data collection and clustering. Bacterial genomes were collected from NCBI, version September 2016[80]. Genomes were classified as anaerobic or aerobic as done elsewhere[31], rendering 1089 bacterial genomes from anaerobes. Briefly, a dataset of 1784 sequences labeled as heme-copper oxygen reductases (HCOs) and nitric oxide reductases (NORs) was blasted against our dataset of prokaryotic genomes. If one homolog (>25% identity, e value $<10^{-10}$, coverage of at least 300 amino acids) for HCOs and NORs was found, the genome was classified as aerobic.

Genomes were assigned their corresponding phyla in NCBI taxonomy, except for (i) Firmicutes and Proteobacteria (the size of which exceeded other phyla by an order of magnitude) where species were assigned to classes for resolution, and (ii) phyla with fewer than 5 species, assigned to "Other Bacteria." Pairwise local alignments for all protein sequences were calculated with a reciprocal blastp (BLAST+ version 2.5.0)[81], followed by the calculation of global identities with an

adaptation of EMBOSS needle[82]. Pairs of sequences with a minimum global identity of 25% and an *e* value ≤1E − 10 were then used to create protein families with the MCL algorithm[83,84]. For the creation of protein families with the MCL algorithm, the parameters --abc -P 180000 -S 19800 -R 25200 were used, resulting in 114,326 families. Of these, 64,149 were present in at least three species and at least four genomes, and were retained for further analyses.

Functional annotation. All protein sequences were aligned against the KEGG Orthology (KO) database[26] (accessed August 2017) using BLAST searches. The best query-subject hits as judged by *E* value, query coverage, and length ratio (cut-off: query coverage ≥80%, *E* value ≤1E − 10, and length ratio between 0.7 and 1.3) were used to annotate the protein sequences individually. We assigned the functional category to each gene family according to the most frequent annotation for the protein sequences in the family. If two or more functional categories occurred with the same frequency, the gene family was annotated within all equally supported categories. For the 146 universal protein families, the annotation of each family in its corresponding functional categories was rechecked manually (Supplementary Data 3).

Sequence alignment, tree reconstruction, and root inferences. For each gene family, the protein sequences were aligned using MAFFT (Multiple Alignment with Faster Fourier Transform) version 7.130[85] (parameters: --maxiterate 1000 --localpair; alignments not predictable this way were constructed using the parameter --retree 2). The resulting alignments were used to reconstruct maximum-likelihood trees with RAxML version 8.2.8[86] (parameters: -m PROTCATWAG -p 12345). Trees were rooted with MAD[54]. Trees with more than one possible MAD root were ignored, leaving 63,324 trees for the subsequent analyses (available in Supplementary Data 5).

Tree analysis

Divergence analysis. To quantify divergence since the LBCA split for each bacterial genome, we calculated root-to-tip distances for all tips in all gene trees measured as (i) the sum of branch lengths (phenetic distance) along the path connecting each operational taxonomic unit to the root and (ii) the sum of branch splits (node depth). To allow for comparisons among trees we normalized the root-to-tip distances for each tree according to the largest distance attained in the tree, so that distance values are bound to the unity interval, with large values indicating more divergence. We scored divergence values to each taxonomic group across all the trees according to the affiliated genome with the smallest root-to-tip distance, independently for each metric (phenetic and node depth). All analyses were performed with custom Python scripts using the Environment for Tree Exploration[87] (ETE3, version 3.1.1).

Sister diversity. We analyzed the distribution of sister relationships for each taxonomic group across the rooted trees as follows: for a given tree with the leaves labeled according to the taxonomic group, we retrieved the set of pure clades for each taxonomic group represented by at least one species in the tree. Note that even though some taxa may not branch as a single clade in the tree, the minimal set of pure (monophyletic) clades can be identified. For each pure clade, the number of taxonomic groups present in the sister clade was recorded (a value in the range of [1–24]) and the sister clade with maximal diversity (in terms of the number of taxonomic groups) was used as sister diversity score. All analyses were performed with custom Python scripts using ETE3[87] (version 3.1.1).

Verticality. All 261,058 values of verticality for all prokaryotic gene families were obtained from Nagies et al.[29], where the highest possible value is 24 and the lowest is zero. All LBCA protein families were ranked from most to least vertical (Supplementary Data 7). For reactions encoded by multiple protein families, the average value of verticality was taken.

LBCA metabolic network

Network construction. For all 6164 anaerobic bacteria KOs the respective reactions were downloaded from the KEGG reaction database[26] (version 16-08-2019), 2414 KOs had at least one reaction associated, resulting in 3550 reactions. Reaction reversibility was determined by parsing KGML (KEGG Markup Language) files from 165 KEGG pathway maps. Reactions that did not occur in the KGML files were assigned as irreversible. Seventy-three reactions containing ambiguous stoichiometries (characters n and m) or unknown compounds were discarded. The final set consisted of 3477 reactions.

Metabolic network expansion. Twenty proteinogenic amino acids, four DNA bases, four RNA bases, eight universal cofactors, one lipid, and 20 uncharged tRNAs were investigated as targets in the network. The algorithm[48] started with a complete reaction network containing all 3477 LBCA candidate reactions regardless of their taxonomic distribution. A score was assigned to each reaction, reflecting the likelihood of their presence in LBCAs metabolic network. Reactions with low distribution among taxonomic groups were scored lower, whereas the score increased with the higher taxonomic distribution. The reactions were sorted increasingly by their score. Starting with low scores, reactions were removed temporarily from the full network sequentially. If neither the presence of the target compounds nor the core network was violated, the respective reaction was removed permanently. The reduction algorithm stopped when no further reaction could be removed. The network was visualized with Cytoscape[88] (version 3.7.2).

Ancestral state reconstruction. Ancestral state reconstruction for cell shape and taxonomic groups was performed with PastML[89] version 1.9.20 using the 131 trees with all taxonomic groups as independent estimates of the prokaryotic phylogeny. The underlying metadata for the tip states was downloaded from JGI GOLD[90] v.6. The maximum-likelihood-based prediction method MPPA (marginal posterior probabilities approximation) with model F81 was used to reconstruct the states at the root of each tree. The reconstructed states at the root of the trees occurring in the highest frequencies were considered the most likely state for LBCA.

Statistics and reproducibility. Statistical tests were performed to assess differences of root-to-tip distances between all 276 possible taxon pairs. For a given taxon pair *a* and *b*, all 131 trees with all taxonomic groups were used and the representative species with smallest root-to-tip distance were recorded for each tree resulting in two distance vectors D_a and D_b. Statistical tests were performed with one-sided Wilcoxon's signed-rank test for paired samples, such that:

H_0: $D_a = D_b$
H_1: $D_a < D_b$

Across all taxon pairs, the tests generated a *p* value matrix (24-by-24), and *p* values were considered significant <0.05 after Bonferroni correction (Supplementary Fig. 3). The tests were conducted using the scipy.stats[91] implementation of the Wilcoxon's signed-rank test in Python. The Kolmogorov–Smirnov test used to measure significance in the comparison of verticalities was also conducted with the default parameters in the scipy.stats implementation in Python. No random sampling was made in the analyses conducted in this paper.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Sequence data that supports the findings of this study are available in NCBI RefSeq[80] (GCF identifiers used are provided in Supplementary Data 1). Metabolic data is available in KEGG[26]. Metadata is available from JGI GOLD[90]. Phylogenetic trees and all other relevant data are provided as Supplementary Datasets.

Code availability

All data sources, software packages, and their usage are described in the "Methods" with the corresponding versions and references, including NCBI, KEGG, JGI GOLD v. 6, BLAST v. 2.5.0, EMBOSS needle, MAFFT v. 7.130, RAxML v. 8.2.8, MCL, MAD, ETE3 v. 3.1.1, PastML v. 1.9.20, and Cytoscape v. 3.7.2. New codes used here consisted of batch subroutines to run the aforementioned algorithms multiple times, calculations, and statistical analyses thoroughly described in the "Methods". The data and results presented in this paper do not result from new software development.

Received: 24 August 2020; Accepted: 26 February 2021;
Published online: 26 March 2021

References

1. Flemming, H. C. & Wuertz, S. Bacteria and archaea on Earth and their abundance in biofilms. *Nat. Rev. Microbiol.* **17**, 247–260 (2019).
2. Madigan, M. T., Bender, K. S., Buckley, D. H., Sattley, W. M. & Stahl, D. A. *Brock Biology of Microorganisms* (Pearson, 2017).
3. Sleep, N. H. Geological and geochemical constraints on the origin and evolution of life. *Astrobiology* **18**, 1199–1219 (2018).
4. Betts, H. C. et al. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* **2**, 1556–1562 (2018).
5. Javaux, E. J. Challenges in evidencing the earliest traces of life. *Nature* **572**, 451–460 (2019).
6. Tashiro, T. et al. Early trace of life from 3.95 Ga sedimentary rocks in Labrador, Canada. *Nature* **549**, 516–518 (2017).
7. Weiss, M. C. et al. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1**, 16116 (2016).
8. Berkemer, S. J. & McGlynn, S. E. A new analysis of Archaea-Bacteria domain separation: variable phylogenetic distance and the tempo of early evolution. *Mol. Biol. Evol.* **37**, 2332–2340 (2020).
9. Catchpole, R. J. & Forterre, P. The evolution of reverse gyrase suggests a nonhyperthermophilic last universal common ancestor. *Mol. Biol. Evol.* **36**, 2737–2747 (2019).

10. Sousa, F. L. & Martin, W. F. Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochim. Biophys. Acta* **1837**, 964–981 (2014).
11. Raymann, K., Brochier-Armanet, C. & Gribaldo, S. The two-domain tree of life is linked to a new root for the Archaea. *Proc. Natl Acad. Sci. USA* **112**, 6670–6675 (2015).
12. Williams, T. A. et al. Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc. Natl Acad. Sci. USA* **114**, E4602–E4611 (2017).
13. Popa, O. & Dagan, T. Trends and barriers to lateral gene transfer in prokaryotes. *Curr. Opin. Microbiol.* **14**, 615–623 (2011).
14. Kump, L. R. The rise of atmospheric oxygen. *Nature* **451**, 277–278 (2008).
15. Martin, W. F. & Sousa, F. L. Early microbial evolution: the age of anaerobes. *Cold Spring Harb. Perspect. Biol.* **8**, a018127 (2016).
16. Fischer, W. W., Hemp, J. & Johnson, J. E. Evolution of oxygenic photosynthesis. *Annu. Rev. Earth Planet. Sci.* **44**, 647–683 (2016).
17. Lyons, T. W., Reinhard, C. T. & Planavsky, N. J. The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307–315 (2014).
18. Müller, V. Energy conservation in acetogenic bacteria. *Appl. Environ. Microbiol.* **69**, 6345–6353 (2003).
19. Zimorski, V., Mentel, M., Tielens, A. G. M. & Martin, W. F. Energy metabolism in anaerobic eukaryotes and Earth's late oxygenation. *Free Radic. Biol. Med.* **140**, 279–294 (2019).
20. McCollom, T. M. & Amend, J. P. A thermodynamic assessment of energy requirements for biomass synthesis by chemolithoautotrophic micro-organisms in oxic and anoxic environments. *Geobiology* **3**, 135–144 (2005).
21. Lever, M. A. et al. Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. *FEMS Microbiol. Rev.* **39**, 688–728 (2015).
22. Raymond, J. & Segrè, D. The effect of oxygen on biochemical networks and the evolution of complex life. *Science* **311**, 1764–1767 (2006).
23. Sousa, F. L., Nelson-Sathi, S. & Martin, W. F. One step beyond a ribosome: the ancient anaerobic core. *Biochim. Biophys. Acta* **1857**, 1027–1038 (2016).
24. Soo, R. M., Hemp, J., Parks, D. H., Fischer, W. W. & Hugenholtz, P. On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science* **355**, 1436–1440 (2017).
25. Xavier, J. C., Patil, K. R. & Rocha, I. Metabolic models and gene essentiality data reveal essential and conserved metabolism in prokaryotes. *PLoS Comput. Biol.* **14**, e1006556 (2018).
26. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
27. Durot, M., Bourguignon, P. Y. & Schachter, V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol. Rev.* **33**, 164–190 (2009).
28. Liu, L., Agren, R., Bordel, S. & Nielsen, J. Use of genome-scale metabolic models for understanding microbial physiology. *FEBS Lett.* **584**, 2556–2564 (2010).
29. Nagies, F. S. P., Brueckner, J., Tria, F. D. K. & Martin, W. F. A spectrum of verticality across genes. *PLoS Genet.* **16**, e1009200 (2020).
30. Thiergart, T., Landan, G. & Martin, W. F. Concatenated alignments and the case of the disappearing tree. *BMC Evol. Biol.* **14**, 266 (2014).
31. Sousa, F. L., Alves, R. J., Pereira-Leal, J. B., Teixeira, M. & Pereira, M. M. A bioinformatics classifier and database for heme-copper oxygen reductases. *PLoS ONE* **6**, e19117 (2011).
32. Magnabosco, C. et al. The biomass and biodiversity of the continental subsurface. *Nat. Geosci.* https://doi.org/10.1038/s41561-018-0221-6 (2018).
33. Fuchs, G. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu. Rev. Microbiol.* **65**, 631–658 (2011).
34. Varma, S. J., Muchowska, K. B., Chatelain, P. & Moran, J. Native iron reduces CO2 to intermediates and end-products of the acetyl-CoA pathway. *Nat. Ecol. Evol.* **2**, 1019–1024 (2018).
35. Preiner, M. et al. A hydrogen-dependent geochemical analogue of primordial carbon and energy metabolism. *Nat. Ecol. Evol.* **4**, 534–542 (2020).
36. Ronimus, R. S. & Morgan, H. W. Distribution and phylogenies of enzymes of the Embden-Meyerhof-Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. *Archaea* **1**, 199–221 (2003).
37. Say, R. F. & Fuchs, G. Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* **464**, 1077–1081 (2010).
38. Schönheit, P., Buckel, W. & Martin, W. F. On the origin of heterotrophy. *Trends Microbiol.* **24**, 12–25 (2016).
39. Dobson, G. P., Hitchins, S. & Teague, W. E. Thermodynamics of the pyruvate kinase reaction and the reversal of glycolysis in heart and skeletal muscle. *J. Biol. Chem.* **277**, 27176–27182 (2002).
40. Ueda, S. & Sakasegawa, S. Pyruvate kinase from *Geobacillus stearothermophilus* displays an unusual preference for Mn^{2+} in a cycling reaction. *Anal. Biochem.* **570**, 27–31 (2019).
41. Sperber, A. M. & Herman, J. K. Metabolism shapes the cell. *J. Bacteriol.* **199**, e00039–17 (2017).
42. Nisbet, E. & Sleep, N. H. The habitat and nature of early life. *Nature* **409**, 1083–1091 (2001).
43. Saeki, T., Hori, M. & Umezawa, H. Pyruvate kinase of *Escherichia coli*. *J. Biochem.* **76**, 631–637 (1974).
44. Schumann, W. FtsH - a single-chain charonin? *FEMS Microbiol. Rev.* **23**, 1–11 (1999).
45. Bahari, L. et al. Membrane targeting of ribosomes and their release require distinct and separable functions of FtsY. *J. Biol. Chem.* **282**, 32168–32175 (2007).
46. Pilhofer, M. et al. Characterization and evolution of cell division and cell wall synthesis genes in the bacterial phyla Verrucomicrobia, Lentisphaerae, Chlamydiae, and Planctomycetes and phylogenetic comparison with rRNA genes. *J. Bacteriol.* **190**, 3192–3202 (2008).
47. Xavier, J. C., Patil, K. R. & Rocha, I. Integration of biomass formulations of genome-scale metabolic models with experimental data reveals universally essential cofactors in prokaryotes. *Metab. Eng.* **39**, 200–208 (2017).
48. Ebenhöh, O., Handorf, T. & Heinrich, R. Structural analysis of expanding metabolic networks. *Genome Inf.* **15**, 35–45 (2004).
49. Carter, C. W. Coding of class I and II aminoacyl-tRNA synthetases. *Adv. Exp. Med. Biol.* **966**, 103–148 (2017).
50. Xavier, J. C., Hordijk, W., Kauffman, S., Steel, M. & Martin, W. F. Autocatalytic chemical networks at the origin of metabolism. *Proc. R. Soc. Ser. B* **287**, 20192377 (2020).
51. Martinez, M. A. et al. A novel role of malonyl-ACP in lipid homeostasis. *Biochemistry* **49**, 3161–3167 (2010).
52. Gao, S. et al. Substrate promiscuity of pyruvate kinase on various deoxynucleoside diphosphates for synthesis of deoxynucleoside triphosphates. *Enzyme Microb. Technol.* **43**, 455–459 (2008).
53. Bocchetta, M., Gribaldo, S., Sanangelantoni, A. & Cammarano, P. Phylogenetic depth of the bacterial genera *Aquifex* and *Thermotoga* inferred from analysis of ribosomal protein, elongation factor, and RNA polymerase subunit sequences. *J. Mol. Evol.* **50**, 366–380 (2000).
54. Tria, F. D. K., Landan, G. & Dagan, T. Phylogenetic rooting using minimal ancestor deviation. *Nat. Ecol. Evol.* **1**, 0193 (2017).
55. Achenbach-Richter, L., Gupta, R., Stetter, K. O. & Woese, C. R. Were the original eubacteria thermophiles? *Syst. Appl. Microbiol.* **9**, 34–39 (1987).
56. Brochier, C. & Philippe, H. A non-hyperthermophilic ancestor for Bacteria. *Nature* **417**, 244–244 (2002).
57. Nelson-Sathi, S. et al. Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature* **517**, 77–80 (2015).
58. Boucher, Y. et al. Lateral gene transfer and the origins of prokaryotic groups. *Annu. Rev. Genet.* **37**, 283–328 (2003).
59. Brinkmann, H., van der Giezen, M., Zhou, Y., de Raucourt, G. P. & Philippe, H. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst. Biol.* **54**, 743–757 (2005).
60. Wade, T., Rangel, L. T., Kundu, S., Fournier, G. P. & Bansal, M. S. Assessing the accuracy of phylogenetic rooting methods on prokaryotic gene families. *PLoS ONE* **15**, e0232950 (2020).
61. Vetriani, C., Speck, M. D., Ellor, S. V., Lutz, R. A. & Starovoytor, V. Thermovibrio ammonificans sp. nov., a thermophilic, chemolithotrophic, nitrate-ammonifying bacterium from deep-sea hydrothermal vents. *Int. J. Syst. Evol. Microbiol.* **54**, 175–181 (2004).
62. Dagan, T., Artzy-Randrup, Y. & Martin, W. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc. Natl Acad. Sci. USA* **105**, 10039–10044 (2008).
63. Taib, N. et al. Genome-wide analysis of the Firmicutes illuminates the diderm/monoderm transition. *Nat. Ecol. Evol.* **4**, 1661–1672 (2020).
64. Coleman, G. et al. A rooted phylogeny resolves early bacterial evolution. *bioRxiv* https://doi.org/10.1101/2020.07.15.205187 (2020).
65. Decker, K., Jungermann, K. & Thauer, R. K. Energy production in anaerobic organisms. *Angew. Chem. Int. Ed. Engl.* **9**, 138–158 (1970).
66. Ciccarelli, F. D. Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283–1287 (2006).
67. Koch, A. L. Were Gram-positive rods the first bacteria? *Trends Microbiol.* **11**, 166–170 (2003).
68. El Baidouri, F., Venditti, C. & Humphries, S. Independent evolution of shape and motility allows evolutionary flexibility in Firmicutes bacteria. *Nat. Ecol. Evol.* **1**, 0009 (2017).
69. Garg, S. G. et al. Anomalous phylogenetic behavior of ribosomal proteins in metagenome assembled asgard archaea. *Genome Biol. Evol.* https://doi.org/10.1093/gbe/evaa238 (2020).
70. Fan, L. et al. Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. *Nat. Ecol. Evol.* **4**, 1213–1219 (2020).
71. Tocheva, E. I., Ortega, D. R. & Jensen, G. J. Sporulation, bacterial cell envelopes and the origin of life. *Nat. Rev. Microbiol.* **14**, 535–542 (2016).

72. Maughan, H., Masel, J., Birky, C. W. & Nicholson, W. L. The roles of mutation accumulation and selection in loss of sporulation in experimental populations of *Bacillus subtilis*. *Genetics* **177**, 937–948 (2007).
73. Xavier, J. C., Preiner, M. & Martin, W. F. Something special about CO-dependent CO_2 fixation. *FEBS J.* **285**, 4181–4195 (2018).
74. Schink, B., Thiemann, V., Laue, H. & Friedrich, M. W. *Desulfotignum phosphitoxidans* sp. nov., a new marine sulfate reducer that oxidizes phosphite to phosphate. *Arch. Microbiol.* **177**, 381–391 (2002).
75. Ikeda-Ohtsubo, W. et al. '*Candidatus* Adiutrix intracellularis', an endosymbiont of termite gut flagellates, is the first representative of a deep-branching clade of *Deltaproteobacteria* and a putative homoacetogen. *Environ. Microbiol.* **18**, 2548–2564 (2016).
76. Waite, D. W. et al. Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. *Int. J. Syst. Evol. Microbiol.* https://doi.org/10.1099/ijsem.0.004213 (2020).
77. Merino, N. et al. Single-cell genomics of novel actinobacteria with the Wood–Ljungdahl pathway discovered in a serpentinizing system. *Front. Microbiol.* **11**, 1031 (2020).
78. Martin, W. F. Older than genes: the acetyl CoA pathway and origins. *Front. Microbiol.* **11**, 817 (2020).
79. Khersonsky, O., Roodveldt, C. & Tawfik, D. S. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508 (2006).
80. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
81. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
82. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
83. van Dongen, S. *A Cluster Algorithm for Graphs*. Technical Report INS-R0010 (National Research Institute for Mathematics and Computer Science in the Netherlands, 2000).
84. Enright, A. J. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
85. Katoh, K. & Standley, D. M. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
86. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
87. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
88. Shannon, P. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
89. Ishikawa, S. A., Zhukova, A., Iwasaki, W. & Gascuel, O. A fast likelihood method to reconstruct and visualize ancestral scenarios. *Mol. Biol. Evol.* **36**, 2069–2085 (2019).
90. Mukherjee, S. et al. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res.* **45**, D446–D456 (2017).
91. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

Acknowledgements

This work was supported by grants from the Deutsche Forschungsgemeinschaft (MA-1426/21-1); the European Research Council (666053); and the Volkswagen Foundation (93046). We thank Madeline Weiss for comments on the clustering and tree analysis, Oliver Ebenhoh and Nima Saadat for comments on the network expansion algorithm and Nathalie Brenner for help with the classification of anaerobes.

Author contributions

J.C.X. analyzed data, curated annotations, performed the statistical analysis, performed the sister diversity calculations, visualizations, and wrote the first manuscript draft. R.E.G. performed data filters, clustering of proteins in families, multiple alignments, tree inferences, initial annotations, and distance calculations. J.L.E.W. reconstructed LBCA's network, performed the network expansion and ancestral reconstructions, and contributed to visualizations and verticality analysis. J.B. performed the initial BLASTs for the clustering in protein families. F.D.K.T. participated in project design and supervision and tree, verticality, and statistical analysis. J.C.X. and W.F.M. designed and supervised the project. All authors contributed to the writing of the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-01918-4.

Correspondence and requests for materials should be addressed to J.C.X.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

6 Zusammenfassung der Ergebnisse

Der Ursprung des Lebens fand vor mehr als 4 Milliarden Jahren statt. Dieses einmalige und längst vergangene Ereignis wird wohl ein immerwährendes Rätsel bleiben, was sich vermutlich niemals in Gänze lösen lässt. Universell konservierte Eigenschaften und Bausteine des Lebens gewähren jedoch einen empirischen Zugang zum Thema. Da LUCA, der letzte gemeinsame Urvorfahr aller Lebewesen, als Brücke zwischen der abiotischen und der biotischen Phase fungiert, ist dieser mit seinen Charakteristika von besonderem Interesse für die Evolutionsbiologie, welche sich mit der Genese von Leben und dessen Entwicklung beschäftigt. Durch die Untersuchung und Definition von LUCAs Physiologie können Rückschlüsse auf das Ursprungshabitat, mögliche Energiequellen und die Lebensweise gezogen werden (Weiss *et al.*, 2016). In den letzten Jahrzehnten rückten dadurch besonders H_2-produzierende Tiefseehydrothermalquellen in den Vordergrund (Martin und Russell, 2007; Amend und McCollom, 2009; Weiss *et al.*, 2016).

Infolgedessen wird in **Publikation I** ein Netzwerk aus 404 biochemischen Stoffwechselreaktionen definiert, welches einen autotrophen Grundstoffwechsel am Ursprung des Metabolismus beschreibt (Wimmer *et al.*, 2021b). Beginnend mit den Komponenten Wasserstoff (H_2), Kohlenstoffdioxid (CO_2) und Ammoniak (NH_3), welche noch heute in Hydrothermalfeldern wie Lost City (Kelley *et al.*, 2005; Martin *et al.*, 2008) zur Verfügung stehen, ist die Synthese der Grundbausteine des Lebens—der 20 Aminosäuren, 4 Nukleoside, 4 Desoxyribonukleoside und 18 Cofaktoren—mittels dieser Reaktionen möglich. Ergänzt wird das Reaktionsnetzwerk durch Reaktionen des reduktiven Acetyl-CoA-Weges, der mutmaßlich ursprünglichsten Alternative zur Fixierung von CO_2 (Pereto *et al.*, 1999; Fuchs, 2011; Martin, 2020), welche als einzige in Bakterien und Archaeen vorkommt und teilweise auf LUCA zurückzuführen ist (Weiss *et al.*, 2016; Preiner *et al.*, 2020; Wimmer *et al.*, 2021b). Des Weiteren beinhaltet das metabolische Zentralnetzwerk die Reaktionen des reversen Citratzyklus (Steffens *et al.*, 2021), der Gluconeogenese (Say und Fuchs, 2010) und des Pentosephosphatweges (Keller *et al.*, 2016), um Schlüsselintermediate wie Pyruvat, Phosphoenolpyruvat und 3-Phosphoglycerat bilden zu können. Das Gesamtnetzwerk beinhaltet 380 Komponenten von denen Wasser die häufigste darstellt, gefolgt von ATP, H^+ und Phosphat. Kohlenstoffdioxid ist die häufigste Kohlenstoffkomponente im Reaktionsdatensatz, was als Hinweis für einen autotrophen Lebensursprung gewertet werden kann. Da erste

Reaktionsnetzwerke noch keinen genetischen Code und somit keine Enzyme aufwiesen, benötigten diese anorganische Katalysatoren wie Metallionen und Cofaktoren für die Reaktionen des Zellaufbaus (Muchowska *et al.*, 2020). Beispielsweise werden elf Cofaktoren benötigt, um aus H_2 und CO_2 das Intermediat Pyruvat zu bilden (Sousa und Martin, 2014). Das autotrophe Reaktionsnetzwerk enthält fünf autokatalytische Zyklen, welche die Synthese von Pyridoxalphosphat, Thiamin, ATP, NAD^+ und $NADP^+$ umfassen. Ein Vergleich des Grundstoffwechsels mit den 355 identifizierten Genen von LUCA (Weiss *et al.*, 2016) und den 172 Reaktionen im Ur-RAF (Xavier *et al.*, 2020), welche die Schnittmenge aus H_2-abhängigen Acetogenen und Methanogenen bilden, zeigt nur wenige Übereinstimmungen. Dies liegt daran, dass phylogenetische Parameter zur Identifikation der LUCA-Gene herangezogen wurden, diese Gene allerdings weniger in die Synthese der Grundbausteine involviert sind und stattdessen mehr in die ribosomale Biogenese (Weiss *et al.*, 2016). Der Grundstoffwechsel überlappt hingegen zur Hälfte mit dem rekonstruierten RAF-Netzwerk (Xavier *et al.*, 2020), die Schnittmenge kommt vor allem durch den Fokus auf Bakterien und Archaeen, welche den reduktiven Acetyl-CoA-Weg nutzen (Fuchs, 2011), zustande. Die Schnittmengen der drei Datensätze zeigen, dass Cofaktoren und deren Vorläufer (katalytisch aktive Mineralien) wohl essentiell für den frühen Stoffwechsel waren.

Da die über 400 Reaktionen des Grundstoffwechsels kaum gleichzeitig entstanden sein können wird vermutet, dass sich das Reaktionsnetzwerk stattdessen schrittweise ausbreitete (Ferry, 2006; Ragsdale und Pierce, 2008). Beginnend mit geochemischen Vorgängen in der Erdkruste in hydrothermalen Tiefseequellen, sogenannten Serpentinisierungsprozessen, wird Wasserstoff freigesetzt (Bach *et al.*, 2006; McCollom und Seewald, 2013; Preiner *et al.*, 2018). Mit dessen Hilfe kann im Ozeanwasser gelöster Kohlenstoff fixiert und reduziert werden (Martin und Preiner, 2017; Preiner *et al.*, 2020). Da keine grundlegende Abweichung hinsichtlich der chemischen Intermediate des Grundstoffwechsels in modernen Mikroben bekannt ist (Kanehisa und Goto, 2000), lassen die konservierten Reaktionen Rückschlüsse auf das chemische Geschehen in LUCAs Grundstoffwechsel zu. Die zuvor in **Publikation I** vorgestellten ungerichteten Reaktionen des autotrophen Grundstoffwechsels werden in **Publikation II** in Richtung der Biosynthese ausgerichtet (Wimmer *et al.*, 2021c). Durch Berechnung der Änderung der Gibbs-Energie (ΔG) für jede Reaktion unter verschiedenen Umweltbedingungen wird deren energetischer Beitrag charakterisiert. Unter Bedingungen, wie sie in hydrothermalen Tiefseequellen wie beispielsweise in Lost City zu finden sind (neutraler bis alkalischer pH-Wert 7–10, Temperatur etwa 80 °C, ein Überschuss an Reaktanten zu Produkten—unterstützen die These, dass sich das Reaktionsnetzwerk graduell entwickelte und

nicht alle Reaktionen gleichzeitig entstanden—und Substitution von organischen Reduktionsmitteln mit Wasserstoff) benötigt die große Mehrheit aller Reaktionen (95–97 %) keinerlei äußere Energie um abzulaufen, die Reaktionen setzen vielmehr Energie frei (Wimmer *et al.*, 2021c). Anstatt der Notwendigkeit vieldiskutierter externer Energiequellen wie Ultraviolettstrahlung (Patel *et al.*, 2015), Meteoriteneinschläge (Ferus *et al.*, 2015) oder Radioaktivität (Ebisuzaki und Maruyama, 2017) zeigt der Grundstoffwechsel vielmehr eine natürliche Tendenz sich von selbst zu entfalten, angetrieben von dem in Hydrothermalquellen produzierten Wasserstoff (Wimmer *et al.*, 2021c).

Auch anorganisches Pyrophosphat (PP_i) wurde lange als verfügbare Energiequelle am Ursprung des Stoffwechsels diskutiert (Schramm *et al.*, 1962; Lipmann, 1965). **Publikation III** zeigt jedoch, dass PP_i im Rahmen des modernen Zellaufbaus keinerlei Energie bereitstellt, da es in den 36 betroffenen Reaktionen stets als Produkt und in keinem Fall als Reaktant partizipiert (Wimmer *et al.*, 2021a, 2021c). Stattdessen wird gezeigt, dass Pyrophosphat-involvierende Reaktionen durch ubiquitäre Pyrophosphataseaktivität im Cytosol irreversibel in Richtung der Biosynthese gezwungen werden, da die Phosphoanhydridbindung des Pyrophosphats hydrolysiert wird und somit die Rückreaktion in Ermangelung an Substrat blockiert wird. Dieser kinetische Effekt wird am Beispiel der Isoleucyl-tRNA Synthese für unterschiedliche PPase-Raten simuliert und bestätigt. Anstelle von PP_i hat sich Adenosintriphosphat (ATP) als universeller Energielieferant in biologischen Prozessen etabliert. Aufgrund der zwei vorhandenen Phosphoanhydridbindungen kann aus ATP, je nachdem welche Bindung aufgespalten wird, entweder Pyrophosphat (PP_i) oder Phosphat (P_i) gebildet werden. Der kinetische Effekt der PP_i-Bildung steht einem thermodynamischen Effekt im Rahmen der P_i-Bildung gegenüber, bei welchem die Änderung der Gibbs-Energie negativ genug ist, um Energie freizusetzen und so diese Reaktion ablaufen zu lassen. Dieser zweifache Effekt ist spezifisch für Triphosphate und erklärt, warum sich diese in der Natur als Energielieferanten durchgesetzt haben (Wimmer *et al.*, 2021a).

Schließlich geht **Publikation IV** in der Evolution der prokaryotischen Lebewesen einen Schritt weiter und untersucht die Physiologie und den Metabolismus des letzten gemeinsamen Vorfahren aller Bakterien (LBCA; Xavier *et al.*, 2021). Aus 1.089 anaeroben bakteriellen Genomen wurden mittels phylogenetischer Kriterien 146 Gene herausgefiltert, welche Rückschlüsse auf LBCA zulassen. Jedes dieser Gene lässt sich allen bakteriellen taxonomischen Gruppen im Datensatz zuordnen. Das Reaktionsnetzwerk braucht nur neun weitere Gene um alle universellen Metaboliten wie Aminosäuren, Nukleoside, Desoxyribonukleoside, Cofaktoren, Lipide und transfer-RNAs zu synthetisieren. Eine Rekonstruktion ursprünglicher

Merkmalszustände (Ishikawa *et al.*, 2019) schlägt ein stabförmiges Erscheinungsbild für LBCA vor. Die Rekonstruktion phylogenetischer Stammbäume von 131 universellen Genfamilien weist darauf hin, dass heutige Clostridien dem bakteriellen Urvorfahren genetisch am nächsten zu sein scheinen. Das Vorhandensein von Reaktionen welche mit dem reduktiven Acetyl-CoA-Weg assoziiert sind, indiziert eine autotrophe Lebensweise von LBCA (Xavier *et al.*, 2021).

Zusammenfassend lässt sich aus den Ergebnissen der in dieser Arbeit vorgestellten **Publikationen I–IV** schlussfolgern, dass H_2-produzierende Tiefseehydrothermalquellen ein plausibles Szenario für den Ursprung des Lebens darstellen. Die in den Quellen gegebenen Umweltbedingungen, welche sich seit Milliarden von Jahren kaum verändert haben, unterstützten die ersten metabolischen Reaktionen und ließen diese energetisch günstig werden, sodass die Grundbausteine des Lebens spontan mithilfe von rein chemischer Energie synthetisiert werden konnten. Der biosynthetische Grundstoffwechsel des Urvorfahren konnte sich ohne extrinsische Energie ausbilden und ein komplexer werdendes Reaktionsnetzwerk bilden.

Die Ergebnisse der angewandten bioinformatischen Untersuchungen decken sich mit denen von experimentell im Labor durchgeführten Versuchen (Preiner *et al.*, 2020). Obgleich sich die Forschung zum Ursprung des Lebens hauptsächlich mit weit in der Vergangenheit zurückliegenden Ereignissen beschäftigt, lassen sich die hier gewonnenen Erkenntnisse zu natürlich auftretenden energiefreisetzenden Reaktionen zwischen H_2 und CO_2 in einer Umformulierung des in der Einleitung angeführten Haeckel-Zitates auf die Ableitung einiger biochemischer Merkmale übertragen:

> Wir haben jetzt Grund anzunehmen, dass die zentralen Stoffwechselreaktionen aller derzeit lebenden Organismen-Formen die Nachkommen einer geringen Anzahl von Stoffwechselreaktionen im letzten gemeinsamen Vorfahren aller Zellen sind und dass diese, energetisch angetrieben durch Wasserstoff und Kohlenstoffdioxid, spontan ablaufenden geochemischen Reaktionen ihre Entstehung verdanken.

7 Literaturverzeichnis

Amend, J. P. und McCollom, T. M. (2009). Energetics of biomolecule synthesis on early Earth. In: *ACS Symposium Series*. S. 63–94.

Amend, J. P. und Shock, E. L. (1998). Energetics of amino acid synthesis in hydrothermal ecosystems. *Science* 281:1659–62. doi: 10.1126/science.281.5383.1659.

Arndt, N. T. und Nisbet, E. G. (2012). Processes on the young Earth and the habitats of early life. *Annual Review of Earth and Planetary Sciences* 40:521–49. doi: 10.1146/annurev-earth-042711-105316.

Bach, W., Paulick, H., Garrido, C. J., Ildefonse, B., Meurer, W. P. und Humphris, S. E. (2006). Unraveling the sequence of serpentinization reactions: Petrography, mineral chemistry, and petrophysics of serpentinites from MAR 15°N (ODP Leg 209, Site 1274). *Geophysical Research Letters* 33:L13306. doi: 10.1029/2006GL025681.

Barge, L. M., Flores, E., Baum, M. M., VanderVelde, D. G. und Russel, M. J. (2019). Redox and pH gradients drive amino acid synthesis in iron oxyhydroxide mineral systems. *Proceedings of the National Academy of Sciences of the United States of America* 116:4828–33. doi: https://doi.org/10.1073/pnas.1812098116.

Baross, J. und Hoffman, S. (1985). Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. *Origins of life and evolution of the biosphere* 15:327–45. doi: 10.1007/BF01808177.

Barth, C., Weiss, M. C., Roettger, M., Martin, W. F., und Unden, G. (2018). Origin and phylogenetic relationships of [4Fe–4S]-containing O_2 sensors of bacteria. *Environmental Microbiology* 20:4567–86. doi: 10.1111/1462-2920.14411.

Battistuzzi, F. U., Feijao, A. und Hedges, S. B. (2004). A genomic timescale of prokaryote evolution: Insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evolutionary Biology* 4:44. doi: 10.1186/1471-2148-4-44.

Berg, J. M., Tymoczko, J. L. und Stryer, L. (2013). Stryer Biochemie. Berlin, Heidelberg: Springer Berlin Heidelberg.

Bernhardt, G., Lüdemann, H. D., Jaenicke, R., König, H. und Stetter, K. O. (1984). Biomolecules are unstable under “black smoker” conditions. *Naturwissenschaften* 71:583–6.

Boyce, A. J., Coleman, M. L. und Russell, M. J. (1983). Formation of fossil hydrothermal chimneys and mounds from Silvermines, Ireland. *Nature* 306:545–50. doi: 10.1038/306545a0.

Boyd, E. S., Amenabar, M. J., Poudel, S. und Templeton, A. S. (2020). Bioenergetic constraints on the origin of autotrophic metabolism. *Philosophical Transactions of the Royal Society A* 378:20190151. doi: 10.1098/rsta.2019.0151.

Busch, F., Rajendran, C., Heyn, K., Schlee, S., Merkl, R. und Sterner, R. (2016). Ancestral tryptophan synthase reveals functional sophistication of primordial enzyme complexes. *Cell Chemical Biology* 23:709–15. doi: 10.1016/j.chembiol.2016.05.009.

Clausius, R. (1854). Ueber eine veränderte Form des zweiten Hauptsatzes der mechanischen Wärmetheorie. *Annalen der Physik* 169:481–506. doi: 10.1002/andp.18541691202.

Corliss, J. B., Baross, J. A. und Hoffman, S. (1981). An hypothesis concerning the relationship between submarine hot springs and the origin of life on Earth. *Oceanologica Acta* 1980:59–69.

Corliss, J. B., Dymond, J., Gordon, L. I., Edmond, J. M., von Herzen, R. P., Ballard, R. D., Green, K., Williams, D., Bainbridge, A., Crane, K. und van Andel, T. H. (1979). Submarine thermal springs on the Galápagos Rift. *Science* 203:1073–83. doi: 10.1126/science.203.4385.1073.

Darwin, C. (1860). Über die Entstehung der Arten im Thier- und Pflanzen-Reich durch natürliche Züchtung. Stuttgart: Schweizerbart.

Ebisuzaki, T. und Maruyama, S. (2017). Nuclear geyser model of the origin of life: Driving force to promote the synthesis of building blocks of life. *Geoscience Frontiers* 8:275-98. doi: 10.1016/j.gsf.2016.09.005.

Ferry, J. G. und House, C. H. (2006). The stepwise evolution of early life driven by energy conservation. *Molecular Biology and Evolution* 23:1286–92. doi: 10.1093/molbev/msk014.

Ferus, M., Nesvorn,Ý. D., Šponer, J., Kubelík, P., Michalčíková, R., Shestivská, V., Šponer, J. und Civiš, S. (2015). High-energy chemistry of formamide: A unified mechanism of nucleobase formation. *Proceedings of the National Academy of Sciences of the United States of America* 112:E339. doi: 10.1073/pnas.1412072111.

Fuchs, G. (2011). Alternative pathways of carbon dioxide fixation: Insights into the early evolution of life? *Annual Review of Microbiology* 65:631–58. doi: 10.1146/annurev-micro-090110-102801.

Haeckel, E. (1866a). Generelle Morphologie der Organismen Erster Band: Allgemeine Anatomie der Organismen. Berlin: Verlag von Georg Reimer.

Haeckel, E. (1866b). Generelle Morphologie der Organismen Zweiter Band: Allgemeine Entwicklungsgeschichte der Organismen. Berlin: Verlag von Georg Reimer.

Haldane, J. B. S. (1929). The origin of life. *The Rationalist Annual* 148:3–10.

Holland, H. D. (2002). Volcanic gases, black smokers, and the great oxidation event. *Geochimica et Cosmochimica Acta* 66:3811–26. doi: 10.1016/S0016-7037(02)00950-X.

Ishikawa, S. A., Zhukova, A., Iwasaki, W. und Gascuel, O. (2019). A fast likelihood method to reconstruct and visualize ancestral scenarios. *Molecular Biology and Evolution* 36:2069–85. doi: 10.1093/molbev/msz131.

Kanehisa, M. und Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28:27–30. doi: 10.1093/nar/28.1.27.

Kaufmann, M. (2009). On the free energy that drove primordial anabolism. *International Journal of Molecular Sciences* 10:1853–71. doi: 10.3390/ijms10041853.

Keller, M. A., Zylstra, A., Castro, C., Turchyn, A. V., Griffin, J. L. und Ralser, M. (2016). Conditional iron and pH-dependent activity of a non-enzymatic glycolysis and pentose phosphate pathway. *Science Advances* 2:–. doi: 10.1126/sciadv.1501235.

Kelley, D. S., Karson, J. A., Blackman, D. K., Früh-Green, G. L., Butterfield, D. A., Lilley, M. D., Olson, E. J., Schrenk, M. O., Roe, K. K., Lebon, G. T., Rivizzigno, P. und AT3-60 Shipboard Party (2001). An off-axis hydrothermal vent field near the mid-atlantic ridge at 30° n. *Nature* 412:145–9. doi: 10.1038/35084000.

Kelley, D. S., Karson, J. A., Früh-Green, G. L., Yoerger, D. R., Shank, T. M., Butterfield, D. A., Hayes, J .M., Schrenk, M. O., Olson, E. J., Proskurowski, G., Jakuba, M., Bradley, A., Larson, B., Ludwig, K., Glickson, D., Buckman, K., Bradley, A. S., Brazelton, W., J., Roe, K., Elend, M. J., Delacour, A., Bernasconi, S. M., Lilley, M. D., Baross, J. D., Summons, R. E. und Sylva, S. P. (2005). A serpentinite-hosted ecosystem: The Lost City hydrothermal field. *Science* 307:1428–34. doi: 10.1126/science.1102556.

Lang, S. Q., Butterfield, D. A., Schulte, M., Kelley, D. S. und Lilley, M. D. (2010). Elevated concentrations of formate, acetate and dissolved organic carbon found at the Lost City hydrothermal field. *Geochimica et Cosmochimica Acta* 74:941–52. doi: 10.1016/j.gca.2009.10.045.

LaRowe, D. E. und Amend, J. P. (2016). The energetics of anabolism in natural settings. *The ISME Journal* 10:1285–95. doi: 10.1038/ismej.2015.227.

Larter, R. C. L., Boyce, A. J. und Russel, M. J. (1981). Hydrothermal pyrite chimneys from the Ballynoe baryte deposit, Silvermines, County Tipperary, Ireland. *Ineralium Deposita* 16:309–17. doi: 10.1007/BF00202742.

Lazcano, A. und Miller, S. L. (1996). The origin and early evolution of life: Prebiotic chemistry, the pre-RNA world, and time. *Cell* 85:793–8. doi: 10.1016/s0092-8674(00)81263-5.

Lipmann, F. (1965). Projecting backward from the present stage of evolution of biosynthesis. In: Fox SW, editor. *The origin of prebiological systems and of their molecular matrices*. New York: Academic Press, S. 259–80.

Madigan, M. T., Bender, K. S., Buckley, D. H., Sattley, W. M. und Stahl, D. A. (2020). Brock Mikrobiologie. München: Pearson Deutschland.

Martin, W., Baross, J., Kelley, D. und Russell, M. J. (2008). Hydrothermal vents and the origin of life. *Nature Reviews Microbiology* 6:805–14. doi: 10.1038/nrmicro1991.

Martin, W. und 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* 358:59–85. doi: 10.1098/rstb.2002.1183.

Martin, W. und Russell, M. J. (2007). On the origin of biochemistry at an alkaline hydrothermal vent. *Philosophical Transactions of the Royal Society B* 362:1887–925. doi: 10.1098/rstb.2006.1881.

Martin, W. F. (2020). Older than genes: The acetyl CoA pathway and origins. *Frontiers in Microbiology* 11:1–21. doi: 10.3389/fmicb.2020.00817.

Martin, W. F. und Preiner, M. (2017). Origin of life, theories of. In: *Reference Module in Life Sciences.* Oxford: Elsevier Inc. doi: 10.1016/B978-0-12-809633-8.02403-1.

Martin, W. F. und Sousa, F. L. (2016). Early microbial evolution: The age of anaerobes. *Cold Spring Harbor Perspectives in Biology* 8:a018127. doi: 10.1101/cshperspect.a018127.

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* 41:207–29. doi: 10.1146/annurev-earth-040610-133457.

McCollom, T. M. und Seewald, J. S. (2013). Serpentinites, hydrogen, and life. *Elements* 9:129–34. doi: 10.2113/gselements.9.2.129.

Miller, S. L. (1953). A production of amino acids under possible primitive Earth conditions. *Science* 117:528–9. doi: 10.1126/science.117.3046.528.

Miller, S. L. (1955). Production of some organic compounds under possible primitive Earth conditions. *Journal of the American Chemical Society* 77:2351–61. doi: 10.1021/ja01614a001.

Miller, S. L. und Bada, J. L. (1988). Submarine hot springs and the origin of life. *Nature* 334:609–11. doi: 10.1038/334609a0.

Mojzsis, S. J., Harrison, T. M. und Pidgeon, R. T. (2001). Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* 409:178–81. doi: 10.1038/35051557.

Muchowska, K. B., Varma, S. J. und Moran, J. (2020). Nonenzymatic metabolic reactions and life's origins. *Chemical Reviews* 120:7708–44. doi: 10.1021/acs.chemrev.0c00191.

Oparin, A. I. (1938). The origin of life. Translation with annotations by Sergius Morgulis. New York: The Macmillan Company.

Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. und Sutherland, J. D. (2015). Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nature Chemistry* 7:301–7. doi: 10.1038/nchem.2202.

Pereto, J., Velasco, A., Becerra, A. und Lazcano, A. (1999). Comparative biochemistry of CO_2 fixation and the evolution of autotrophy. *International Microbiology* 2:3–10.

Preiner, M., Igarashi, K., Muchowska, K. B., Yu, M., Varma, S. J., Kleinermanns, K., Nobu, M. K., Kamagata, Y., Tüysüz, H., Moran, J. und Martin, W. F. (2020). A hydrogen-dependent geochemical analogue of primordial carbon and energy metabolism. *Nature Ecology and Evolution* 4:534–42. doi: 10.1038/s41559-020-1125-6.

Preiner, M., Xavier, J. C., Sousa, F. L., Zimorski, V., Neubeck, A., Lang, S. Q., Greenwell, H. C., Kleinermanns, K., Tüysüz, H., McCollom, T. M., Holm, N. G. und Martin, W. F. (2018). Serpentinization: Connecting geochemistry, ancient metabolism and industrial hydrogenation. *Life* 8:41. doi: 10.3390/life8040041.

Ragsdale, S. W. und Pierce, E. (2008). Acetogenesis and the Wood–Ljungdahl pathway of CO_2 fixation. *Biochimia et Biophysica Acta (BBA) – Proteins and Proteomics* 1784:1873–98. doi: 10.1016/j.bbapap.2008.08.012.

Richter, G. (1998). Stoffwechselphysiologie der Pflanzen: Physiologie und Biochemie des Primär- und Sekundärstoffwechsels. 6. Auflage. Stuttgart, New York: Thieme.

Rosing, M. T. (1999). ^{13}C-depleted carbon microparticles in ≥3700-Ma sea-floor sedimentary rocks from west Greenland. *Science* 283:674–6. doi: 10.1126/science.283.5402.674.

Russell, M. J. und 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* 154:377–402. doi: 10.1144/gsjgs.154.3.0377.

Russell, M. J., Hall, A. J., Cairns-Smith, A. G., und Braterman, P. S. (1988). Submarine hot springs and the origin of life. *Nature* 334:609–11. doi: 10.1038/334609a0.

Russell, M. J., Hall, A. J. und Turner, D. (1989). In vitro growth of iron sulphide chimneys: Possible culture chambers for origin-of-life experiments. *Terra Nova* 1:238–41. doi: 10.1111/j.1365-3121.1989.tb00364.x.

Say, R. F. und Fuchs, G. (2010). Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* 464:1077–81. doi: 10.1038/nature08884.

Schramm, G., Grötsch, H. und Pollmann, W. (1962). Nicht-enzymatische Synthese von Polysacchariden, Nucleosiden und Nucleinsäuren und die Entstehung selbst-vermehrungsfähiger Systeme. *Angewandte Chemie* 74:53–9. doi: 10.1002/ange.19620740202.

Schrenk, M. O., Brazelton, W. J. und Lang, S. Q. (2013). Serpentinization, carbon, and deep life. *Reviews in Mineralogy and Geochemistry* 75:575–606. doi: 10.2138/rmg.2013.75.18.

Shock, E. L. (1990). Geochemical constraints on the origin of organic compounds in hydrothermal systems. *Origins of Life and Evolution of the Biosphere* 20:331–67. doi: 10.1007/BF01808115.

Sleep, N. H., Bird, D. K. und Pope, E. C. (2011). Serpentinite and the dawn of life. *Philosophical Transactions of the Royal Society B* 366:2857–69. doi: 10.1098/rstb.2011.0129.

Sousa, F. L. und Martin, W. F. (2014). Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochimica et Biophysica Acta (BBA) – Bioenergetics* 1837:964–81. doi: 10.1016/j.bbabio.2014.02.001.

Sousa, F. L., Thiergart, T., Landan, G., Nelson-Sathi, S., Pereira, I. A. C., Allen, J. F., Lane, N. und Martin, W. F. (2013). Early bioenergetic evolution. *Philosophical Transactions of the Royal Society B* 368:20130088. doi: 10.1098/rstb.2013.0088.

Steffens, L., Pettinato, E., Steiner, T. M., Mall, A., König, S., Eisenreich und Berg, I. A. (2021). High CO_2 levels drive the TCA cycle backwards towards autotrophy. *Nature* 592:784–8. doi: 10.1038/s41586-021-03456-9.

Tashiro, T., Ishida, A., Hori, M., Igisu, M., Koike, M., Méjean, P., Takahata, N., Sano, Y. und Komiya, T. (2017). Early trace of life from 3.95 Ga sedimentary rocks in Labrador, Canada. *Nature* 549:516–8. doi: 10.1038/nature24019.

Thauer, R. K., Jungermann, K. und Decker, K. (1977). Energy conservation in chemotrophic anaerobic bacteria. *Bacteriological Reviews* 41:100–80. doi: 10.1128/br.41.1.100-180.1977.

Urey, H. C. (1952). On the early chemical history of the Earth and the origin of life. *Proceedings of the National Academy of Sciences of the United States of America*. 38:351–63. doi: 10.1073/pnas.38.4.351.

Wächtershäuser, G. (1988). Before enzymes and templates: Theory of surface metabolism. *Microbiological Reviews* 52:452–84.

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S. und Martin, W. F. (2016). The physiology and habitat of the Last Universal Common Ancestor. *Nature Microbiology* 1:1–8. doi: 10.1038/nmicrobiol.2016.116.

Wimmer, J. L. E., Kleinermanns, K. und Martin, W. F. (2021a). Pyrophosphate and irreversibility in evolution, or why PP_i is not an energy currency and why nature chose triphosphates. *Frontiers in Microbiology* 12:759359. doi: 3389/fmicb.2021.759359.

Wimmer, J. L. E., do Nascimento Vieira, A., Xavier, J. C., Kleinermanns, K., Martin, W. F. und Preiner, M. (2021b). The autotrophic core: An ancient network of 404 reactions converts H_2, CO_2, and NH_3 into amino acids, bases, and cofactors. *Microorganisms* 9:458. doi: 10.3390/microorganisms9020458.

Wimmer, J. L. E., Xavier, J. C., do Nascimento Vieira, A., Pereira, D. P. H., Leidner, J., Sousa, F. L., Kleinermanns, K., Preiner, M. und Martin, W. F. (2021c). Energy at origins: Favorable thermodynamics of biosynthetic reactions in the Last Universal Common Ancestor (LUCA). *Frontiers in Microbiology* 12:793664. doi: 10.3389/fmicb.2021.793664.

Xavier, J. C., Gerhards, R. E., Wimmer, J. L. E., Brueckner, J., Tria, F. D. K. und Martin, W. F. (2021). The metabolic network of the Last Bacterial Common Ancestor. *Communications Biology* 4:1–10. doi: 10.1038/s42003-021-01918-4.

Xavier, J. C., Hordijk, W., Kauffman, S., Steel, M. und Martin, W. F. (2020). Autocatalytic chemical networks at the origin of metabolism. *Proceedings of the Royal Society of London B* 287:20192377. doi: 10.1098/rspb.2019.2377.

Danke

Diese Arbeit wäre nicht möglich gewesen ohne eine Reihe von Menschen, die mich in den letzten Jahren begleitet, unterstützt und gestützt haben. Jede/r Einzelne von ihnen hat ein großes Dankeschön verdient:

Allen voran bedanke ich mich bei Prof. Dr. William F. Martin, der mich bereits zum Bachelor in sein Institut aufnahm, mich stets förderte, forderte und mir großes Vertrauen entgegenbrachte. Ein herzliches Dankeschön gilt zudem Zweitgutachter apl. Prof. Dr. Gerhard Steger.

I would also like to thank my former postdoc supervisor Dr. Joana C. Xavier who taught me a ton, first of all self-confidence. Your advice was helpful and you always cared for my work and my welfare. Besides, special thanks to every single coauthor I was lucky to work with. Your input was extremely valuable and supportive. I learned a lot from the interdisciplinary expertise and discussions.

Ebenfalls danke ich Prof. Dr. Karl Kleinermanns für seine Geduld, seinen Enthusiasmus und seine exzellenten Ratschläge und Erklärungen bezüglich physikalischer Chemie.

Bei meiner Korrekturleserin Andrea Alexa bedanke ich mich für deine Zeit und Hilfe.

Darüber hinaus danke ich meinen Bürokolleginnen aus dem Ober-Office, dem gesamten MolEvol-Team und ganz besonders den Kollegen, aus denen gute Freunde wurden. Ihr alle seid mehr als ein guter Grund, dass ich jeden Morgen gerne ins Institut komme.

Ein besonderer Dank gilt außerdem meinem Partner Julian. Ohne deinen Rückhalt wäre ich heute keine Biologin (buchstäblich nicht!) und erst recht nicht der Mensch, der ich jetzt bin. Danke für deine immerwährende Unterstützung.

Zu guter Letzt danke ich meinen Eltern, die mich stets bedingungslos unterstützt haben und mir versucht haben alles zu ermöglichen. Ihr habt mir immer das Gefühl gegeben, alles schaffen zu können und habt selbst niemals im Geringsten daran gezweifelt.