UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
DOCTORADO EN CIENCIAS BIOMÉDICAS
FACULTAD DE MEDICINA

EVOLUCIÓN MOLECULAR DEL ANABOLISMO DE HISTIDINA, CISTEÍNA Y ÁCIDO ASPÁRTICO. SU PAPEL EN SITIOS CATALÍTICOS

TESIS
QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS

PRESENTA:
ALBERTO VÁZQUEZ SALAZAR

DIRECTOR DE TESIS
DR. ANTONIO EUSEBIO LAZCANO ARAUJO REYES
FACULTAD DE MEDICINA

COMITÉ TUTOR
DRA. LORENZO PATRICK SEGOVIA FORCELLA
INSTITUTO DE BIOTECNOLOGÍA
DR. JUAN PABLO PARDO VÁZQUEZ
FACULTAD DE MEDICINA

CIUDAD DE MÉXICO, OCTUBRE, 2019

PDCB/FM/90/2019.

C.P. AGUSTIN MERCADO
Director de Certificación y Control Documental, UNAM
P r e s e n t e.

Nos permitimos informarle que con base al Artículo 31 del RGEP el Comité Académico de **DOCTORADO EN CIENCIAS BIOMÉDICAS**, en su reunión 455 celebrada el 21 de agosto del 2019, designó el siguiente jurado para examen de grado de **DOCTOR EN CIENCIAS** de **ALBERTO VÁZQUEZ SALAZAR** candidato al grado de **DOCTOR EN CIENCIAS** con número de cuenta **306665686** con la tesis titulada **"EVOLUCIÓN MOLECULAR DEL ANABOLISMO DE HISTIDINA, CISTEÍNA Y ÁCIDO ASPÁRTICO. SU PAPEL EN SITIOS CATALÍTICOS "**, dirigida por el Dr. Antonio Eusebio Lazcano Araujo.

PRESIDENTE:	DRA. ADELA RODRÍGUEZ ROMERO
SECRETARIO:	DR. ANTONIO EUSEBIO LAZCANO ARAUJO
VOCAL:	DR. ALFREDOTORRES LARIOS
VOCAL:	DR. ENRIQUE MERINO PÉREZ
VOCAL:	DRA. ALICIA NEGRÓN MENDOZA

A t e n t a m e n t e
"Por mi raza hablará el espíritu"
Ciudad Universitaria, Cd. de Mx., a 21 de agosto de 2019.

DRA. AUREA OROZCO RIVAS
COORDINADORA DEL PROGRAMA

C.c.p. Mtro. Javier de la Torre Segura, Subdirector de Asuntos Escolares del Posgrado. UNAM.
Tutor (a) Dr. (a) Antonio Eusebio Lazcano Araujo.
Coordinación de PDCB.

AOR/RGG/evb

AGRADECIMIENTOS Y RECONOCIMIENTOS INSTITUCIONALES

El presente trabajo se realizó bajo la dirección del Dr. Antonio Lazcano Araujo, en el Laboratorio de Origen de la Vida del Departamento de Biología Evolutiva de la Facultad de Ciencias, Universidad Nacional Autónoma de México.

Se agradecen las valiosas contribuciones del comité tutor, formado por los doctores Juan Pablo Pardo Vázquez, Lorenzo Segovia Forcella y Antonio Lazcano Araujo, a lo largo de todo el proyecto doctoral.

Se agradecen la revisión crítica y las valiosas aportaciones al presente trabajo por parte del jurado, conformado por las doctoras Adela Rodríguez Romero y Alicia Negrón Mendoza, así como por los doctores Enrique Merino Pérez y Alfredo Torres Larios.

Alberto Vázquez Salazar fue alumno del Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), y fue beneficiario del Programa de Becas Nacionales CONACYT (No. de Apoyo: 404507), así como del Programa de Becas de Movilidad en el Extranjero.

Se agradece el apoyo técnico recibido por parte de los doctores Ricardo Hernández Morales y José Alberto Campillo Balderas.

La presente tesis forma parte del proyecto de investigación financiado por el Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) a través del Proyecto IN223916.

Se agradece el apoyo económico recibido del Programa de Apoyo a Estudios de Posgrado (PAEP), gracias al cual se cubrieron, parcialmente, gastos para asistir a:

1) Gordon Research Conference. Origins of Life: “Bridging Disciplinary Perspectives to See Further Into Life's Origins”. Hotel Galvez, Texas, EUA, del 7 al 22 de enero de 2016
2) XVIIIth International Conference on the Origin of Life. Universidad de California en San Diego, California, EUA, del 16 al 22 de Julio de 2017
3) Estancia de investigación en el Departamento de Química y Bioquímica, Universidad de California en Santa Bárbara, California, EUA, del 31 de agosto al 20 de noviembre de 2018
4) FMASOEL: French-Mexican Advanced School on the Origins and Evolution of Life. Muséum National d'Histoire Naturelle, París, Francia, del 1 al 5 de julio de 2019

ÍNDICE

RESUMEN

Una descripción adecuada del origen y la evolución temprana de la vida requiere de la comprensión de la evolución de la catálisis biológica. Durante las primeras etapas de la evolución, los sistemas biológicos primitivos pudieron haber dependido no solo de ribozimas sino también de varios catalizadores simples, tales como péptidos, iones metálicos, arcillas, minerales y moléculas orgánicas; una de estas moléculas orgánicas es el grupo imidazol. Este grupo químico se encuentra presente en varios tipos de biomoléculas: forma parte de la estructura química de las purinas, y juega un papel central en la catálisis biológica contemporánea como parte de la cadena lateral de histidina, el aminoácido que se encuentra con mayor frecuencia en el sitio catalítico de las enzimas. Sorprendentemente, los motivos imidazólicos de la histidina y purinas son biosintetizados por enzimas no homólogas. Como se discute aquí, este fenómeno puede entenderse como un caso de convergencia evolutiva de tipo funcional. En este trabajo, analizamos este proceso polifilético empleando análisis filogenéticos, estructurales y de bioquímica evolutiva, y argumentamos que el origen independiente de las enzimas puede ser explicado por las diferencias en la función bioquímica y catalítica de la histidina en comparación con la de los ribonucleótidos de purina.

CAPITULO I. ORIGEN DE LA CATÁLISIS BIOLÓGICA: UNA PREGUNTA FUNDAMENTAL PARA ENTENDER EL ORIGEN DE LA VIDA

Introducción: sobre el estudio del origen de la vida

La idea de que los organismos vivos son el resultado histórico de una transformación gradual a partir de materia inanimada se comenzó a popularizar después de la publicación de la primera edición de 'El Origen de las Especies' de Charles Darwin (1859). En este marco teórico se defendía que los primeros organismos tenían la capacidad de fijar CO_2 atmosférico, el cual usaban, junto con el agua, para sintetizar compuestos orgánicos; es decir, se proponía un origen autotrófico de la vida (Tirard, 2010).

La publicación de los trabajos de Alexander Oparin sobre el origen de la vida (Oparin, 1924; 1938) disputan la propuesta de un origen autotrófico y, mediante un análisis impregnado por un pensamiento darwinista que involucraba una evolución lenta y gradual de lo simple a lo complejo, proponen un origen heterotrófico, resultado de un largo proceso de evolución química y pre-biológica, donde las primeras formas de vida debieron haber sido microorganismos dependientes de las moléculas y substancias orgánicas presentes en su medio externo (Oparin, 1938; Tirard, 2010).

Oparin basó la idea de un origen heterotrófico, en parte, en la universalidad de las reacciones fermentativas, las cuales, según su razonamiento, debieron haber aparecido primero en la evolución debido a su sencillez, y es que para Oparin era imposible conciliar la propuesta de organismos dotados de pigmentos fotosintéticos primigenios con las ideas de evolución darwinista donde se parte de lo simple hacia lo complejo (Lazcano, 2010).

A partir del análisis detallado de los datos geoquímicos y astronómicos conocidos a esa fecha, Oparin propuso también una atmósfera primitiva, altamente reductora, desprovista de O_2 y compuesta de CH_4, NH_3 y H_2O; bajo esas condiciones se señalaba que el origen de la vida había sido precedido por un periodo de síntesis abiótica y posterior acumulación de diversos compuestos orgánicos en los mares de la Tierra primitiva (Oparin, 1938). Dicha acumulación dio lugar a la formación de un caldo primigenio (o sopa prebiótica) que contenía una

amplia variedad de moléculas. Ahí, según el razonamiento de Oparin, debido a la conglomeración de moléculas orgánicas y otros polímeros con cargas opuestas se formó un tipo particular de coloide: los coacervados. Oparin sugirió que la evolución de los coacervados dio lugar a la aparición de los primeros organismos (Oparin, 1938).

Los coacervados probaron ser un modelo inexacto para la evolución de las primeras formas de vida en la Tierra, por lo que fueron abandonados. De la misma manera, experimentos modernos han demostrado que la atmósfera primitiva probablemente no era altamente reductora sino neutra (Cleaves, et al., 2008), también entendemos que los procesos graduales no son necesariamente lentos, e incluso sabemos, gracias al registro fósil, que el origen y la evolución temprana de la vida ocurrieron en lapsos de tiempo geológico cortos (Lazcano & Miller, 1994; 1996).

Sin embargo, el planteamiento general de la teoría de Oparin tuvo grandes implicaciones para la biología, ya que su trabajo logró la transformación del estudio del origen de la vida, que pasó de ser un campo puramente especulativo a un programa de investigación estructurado y amplio (Lazcano, 2010). Dentro de este nuevo programa de investigación hay, incluso, reformulaciones de un origen autotrófico de la vida que disputan las propuestas de Oparin (cf. Weiss, et al., 2018).

La teoría del origen y evolución temprana de la vida de Oparin ha experimentado también una reestructuración desde la segunda mitad del siglo XX, en ésta encuentran cabida los hallazgos experimentales de la biología molecular, así como las aportaciones teóricas de la biología evolutiva.

Un punto de convergencia entre estas dos ramas de la biología y que ha sido perfectamente incorporado a la teoría del origen heterotrófico es el planteamiento del papel que pudo haber desempeñado la molécula de ácido ribonucleico en las etapas más tempranas de la vida, una hipótesis que se conoce como “el mundo de RNA” (Gilbert, 1986).

Mundos de RNA

En términos generales se entiende por mundo de RNA al conjunto de modelos que buscan explicar un periodo hipotético en la evolución temprana de la vida en la Tierra y donde las moléculas de ácido ribonucleico tenían un papel central. La concepción más conservadora de un mundo de RNA sustenta que dichas moléculas actuaban como catalizadores y como polímeros genéticos en formas primitivas de vida, las cuales carecían de DNA y enzimas proteínicas (Joyce, 2002). Sin embargo, tal y como lo discuten Robertson y Joyce (2012), el mundo de RNA implica diversas premisas para diferentes investigadores, donde cada uno tiene su propio juicio de lo que representa, creando así no uno, sino varios *mundos de RNA*. Por ejemplo, para Gilbert (1986) el mundo de RNA representa un sinónimo del origen de la vida mientras que para Williams y colaboradores representa un estadio implausible en la evolución de la misma (Bowman et al., 2015). Para el presente trabajo se establece que el mundo de RNA es una etapa en la evolución temprana de la vida, donde la continuidad de la información genética estaba dada por la auto-replicación del RNA, y donde los sistemas biológicos primitivos dependían de dicha molécula y de sus interacciones con iones metálicos, minerales, así como con un amplio catálogo de moléculas orgánicas para llevar a cabo las reacciones químicas necesarias para mantener su metabolismo. El catálogo de dichas moléculas pudo haber incluido compuestos capaces de formar membranas, aminoácidos e incluso péptidos pequeños, todos estos compuestos pudieron haber contribuido a moldear el ambiente químico en el cual evolucionó la vida (Jadhav & Yarus, 2002; Hsiao, et al., 2013; Wieczorek, et al., 2013; Vázquez-Salazar & Lazcano, 2018).

Las primeras propuestas de una etapa temprana de evolución biológica donde el RNA tenía un papel trascendental vienen de los trabajos independientes de Alexander Rich (1962), Carl Woese (1967), Francis Crick (1968) y Leslie Orgel (1968). Estos autores analizaron detalladamente, bajo una óptica evolutiva, las evidencias bioquímicas y de la biología molecular que demostraban el papel del RNA en diversos procesos celulares, principalmente en la traducción, y concluyeron que dicha molécula debió haber precedido la aparición del DNA en la evolución,

sugiriendo que los primeros organismos en habitar la Tierra empleaban al RNA como polímero genético y, posiblemente, como catalizador.

Gracias al trabajo independiente de los grupos liderados por Sidney Altman y Thomas Cech en la década de 1980 se descubrió que el RNA también puede llevar a cabo reacciones químicas (Kruger, et al., 1982; Guerrier-Takada, et al., 1983). El hallazgo de moléculas de RNA catalíticas (ribozimas) dio, prácticamente sin buscarlo, la evidencia experimental que necesitaban las propuestas teóricas de Woese, Rich, Crick y Orgel para profundizar y enfatizar el papel primigenio del RNA.

Además de las ribozimas celulares, producto de la evolución biológica, diversos laboratorios en el mundo han logrado seleccionar artificialmente un amplio catálogo de ribozimas. Estas ribozimas artificiales, producto de experimentos de evolución *in vitro*, presentan una variedad de actividades catalíticas no encontradas en el repertorio de los RNA biológicos (Wilson & Szostak, 1999; Martin, et al., 2015). Con esta aproximación se ha logrado concluir que, en conjunto, las ribozimas, tanto naturales como artificiales, pueden llevar a cabo las seis clases de actividades catalíticas presentes en las enzimas proteínicas (Hernández-Morales, et al. 2019). Un listado variado de ribozimas robustece y hace más atractiva la hipótesis del mundo del RNA, pues permite construir un escenario donde se explique cómo una proto-célula desprovista de enzimas proteínicas podría mantener un proto-metabolismo (suponiendo también una manera de replicar la información genética contenida en el mismo RNA).

Una de las evidencias más usadas para dar peso y plausibilidad al mundo de RNA proviene del ribosoma (cf. Noller, 2012). Este complejo ribonucleoproteínico (RNP) representa una de las maquinarias más importantes para la célula, pues es ahí donde la traducción de las proteínas se lleva a cabo.

Fue precisamente el interés en la síntesis de proteínas lo que llevó al temprano reconocimiento de que la molécula RNA jugaba un papel importante en este proceso, pues las mediciones experimentales demostraban una correlación entre la cantidad de RNA en la célula con la capacidad de síntesis de proteínas de dichas células (Caspersson, 1941; Brachet; 1942); sin embargo, no fue sino hasta la descripción bioquímica de la fracción microsomal del citosol (Claude, 1946)

vinculada a la síntesis de proteínas que se pudo establecer una asociación entre RNA y proteínas en el proceso de síntesis proteínica (Keller, et al., 1954).

Para la década de 1950, George Palade, usando microscopía electrónica, logró caracterizar morfológicamente gránulos compuestos por RNA y proteína (Palade, 1955), que fueron más tarde bautizados como ribomas (Roberts, 1958)

Desde entonces, la evolución del ribosoma y de todo el aparato de traducción ha acaparado la atención de diversos científicos (Woese, 1965; Crick, 1968; Orgel, 1968). Aproximaciones bioquímicas que buscaban entender el proceso de la formación del enlace peptídico apuntaban hacía una posible implicación del RNA ribosomal (rRNA) en este proceso (Noller & Chaires, 1972); sin embargo, fue gracias a los trabajos pioneros de Ada Yonath (Yonath, et al., 1980) que la disponibilidad de cristales de calidad se estandarizó y se hizo posible el escrutinio a nivel molecular de las interacciones RNA-proteína que componen al ribosoma (Ban, et al., 1999; 2000; Cate, et al., 1999; Clemons, et al., 1999; Nissen, et al., 2000; Schluenzen, et al., 2000; Wimberly, et al., 2000;). La conclusión más importante a la que se llegaría eventualmente es el reconocimiento de que el centro catalítico del ribosoma (donde se forma el enlace peptídico y que se encuentra altamente conservado en los tres dominios de la vida) está constituido enteramente por RNA (cf. Moore & Steitz, 2011); es decir, el ribosoma es una ribozima (Cech, 2000).

Evolutivamente, esta evidencia sugiere que la síntesis de proteínas se originó en un mundo de RNA como resultado de un proceso muy temprano (Fox, 2010).

Evolución de la catálisis biológica: de ribozimas a enzimas

Una descripción apropiada del origen y la evolución temprana de la vida tiene que englobar, necesariamente, una explicación del origen de la catálisis biológica. En el presente trabajo, y en el contexto del metabolismo contemporáneo, se entiende por catálisis biológica al conjunto de reacciones químicas que ocurren en la célula y que son necesarias para el mantenimiento de ésta, las cuales son posibles gracias a: 1) proteínas (enzimas); 2) ácidos ribonucleicos (ribozimas), y 3) reacciones que ocurren de manera espontánea bajo condiciones bioquímicas específicas; por ejemplo, la ciclación espontánea de Δ^1-pirrolina-2-carboxilato en la biosíntesis de

prolina (Fichman, et al., 2015). El papel desempeñado por cofactores orgánicos, i.e. coenzimas, así como inorgánicos, i.e. iones metálicos, en cada uno de los tres puntos anteriores también es fundamental para entender la catálisis biológica de la célula contemporánea.

El hecho de que un número considerable de enzimas contemporáneas coordinen complejos metálicos similares a los constituyentes de los minerales (por ejemplo, la similitud estructural entre la greigita, Fe_3S_4, y los grupos Fe_3S_4 y Fe_4S_4 presentes en sitios activos) ha sido de suma importancia para formular hipótesis que centran a los minerales como posibles catalizadores prebióticos; más aún, se ha formulado que existe una relación entre la geoquímica pre-biológica y la bioquímica contemporánea (cf. Sahai, et al., 2016).

El estudio de minerales que catalizan reacciones relevantes para el origen y la evolución temprana de la vida también ha arrojado resultados interesantes. Se ha demostrado que la polimerización de nucleótidos activados puede llevarse a cabo en presencia de minerales como la hidroxiapatita o la montmorillonita y en ausencia de otros catalizadores (cf. Orgel & Lohrmann, 1974); la polimerización de alanina también ha sido obtenida empleando calcita y dolomita como caltalizadores (Kawamura et al. 2011). En soluciones de iones de hierro, sulfuro inorgánico y ligandos de azufre adicionales se forman fácilmente grupos de Fe-S que presentan actividad catalítica (Ogino, et al., 1998); estos grupos son análogos a los encontrados en la ferredoxina De la misma forma, se ha observado que distintos minerales (montmorillonita, silicato de aluminio, hidrotalquita, caolinita, etc.) pueden catalizar la formación de vesículas (Hanczyc, et al., 2007), lo cual es particularmente importante pues la separación del medio externo debió haber sido un paso fundamental para la evolución de la vida celular.

También es evidente que, en un contexto prebiótico, los iones metálicos pudieron haber actuado como catalizadores sin la necesidad de estar en complejos de minerales, justamente como ocurre en la célula contemporánea, donde la catálisis depende de un amplio listado de iones metálicos para mantener su metabolismo. De hecho, tanto proteínas como RNA, requieren cationes metálicos para plegarse y funcionar.

En este marco teórico, se ha investigado el desempeño catalítico de los iones de Fe^{2+}, los cuales pueden descomponer peróxido de hidrógeno (Cohen, et al., 1981). Otros cationes divalentes como el Ca^{2+}, Cd^{2+}, Mn^{2+} pueden catalizar reacciones de transfosforilación (Lowenstein, 1958), dando pie a formulaciones de cómo los iones metálicos pudieron haber sido importantes para la catálisis primitiva.

Sin embargo, la forma en que los iones metálicos podrían haberse utilizado para catalizar reacciones metabólicas antes del origen de un aparato de traducción que produjera proteínas complejas sigue sin ser clara (Belmonte & Mansy, 2016). De la misma manera, a pesar de las hipótesis en las que se plantea una continuidad geoquímica entre la Tierra primitiva y la biología contemporánea (Sahai, et al., 2016), resulta complicado trazar una línea evolutiva que dé continuidad histórica a la química de los minerales y los iones metálicos con los complejos biológicos que emplean dichos compuestos hoy día. A pesar de dicha dificultad es muy posible que los iones metálicos hayan sido sumamente importantes tanto para la actividad catalítica primitiva como en el contexto de un mundo de RNA.

Como se ha descrito, es posible que el RNA haya desempeñado un papel fundamental como molécula catalizadora en las etapas más tempranas de la vida. De esta manera, una de las preguntas más interesantes que surge es ¿cómo se dio la transición entre la actividad catalítica de las moléculas de RNA a las enzimas proteínicas que engloban casi el 100% de los catalizadores biológicos en la célula actual?

Harold White III estudió otro de los componentes fundamentales de la catálisis biológica mediada por enzimas: los cofactores orgánicos. En 1976 publicó una hipótesis sobre la transición de lo que llamó “un estado metabólico temprano” (i.e. mundo de RNA) hacia un metabolismo dotado de enzimas. Al igual que Handler (1963) y Eakin (1963), White III (1976, 1982) se interesó en la estructura química de las coenzimas, sin las cuales se pierde por completo la función catalítica de las enzimas que las emplean. En su análisis, White III observó que un número elevado de coenzimas contiene un motivo ribonucleotídico en sus estructuras químicas, lo cual, según discute él mismo, podría deberse a un proceso temprano de

diversificación en las capacidades catalíticas del RNA. De esta manera, tomando en cuenta: i) la amplia distribución de las coenzimas en los seres vivos; ii) la dependencia de las enzimas hacia dichas moléculas para llevar a cabo catálisis, y iii) el hecho de que una gran cantidad de coenzimas son ribonucleótidos o derivados de ribonucleótidos, White III propuso que estas moléculas son los fósiles moleculares de un mundo de RNA (White III, 1976, 1982).

De esta forma, en el esquema de White III (1976, 1982) (Figura 1), la catálisis biológica contemporánea dependiente de coenzimas es el resultado de un proceso evolutivo donde las enzimas proteínicas fueron precedidas por ribozimas. Particularmente, los ribonucleótidos que componían los sitios catalíticos de dichas ribozimas fueron las únicas porciones que se conservaron durante el proceso evolutivo. Esta conservación dio origen a las coenzimas de naturaleza ribonucleotídica que se encuentran involucradas en la catálisis biológica contemporánea (Figura 1).

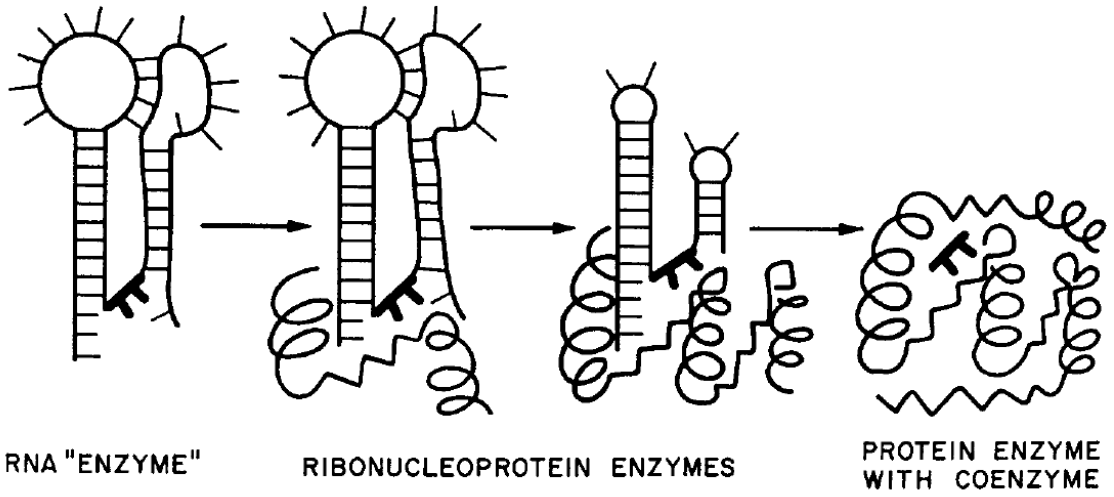

Figura 1. Modelo evolutivo de sustitución propuesto por White III (1976). En este modelo, porciones de las enzimas de ácidos nucleicos (ribozimas) fueron paulatinamente sustituidas por péptidos, siendo la parte catalítica la única que se conservó durante la evolución, encontrándola hoy día como coenzimas. Tomado de White III, 1982

Thomas Cech propuso un esquema evolutivo semejante al de White III (Cech 2009a; 2009b). En éste se incluye a los complejos ribonucleoproteínicos (tales como el ribosoma o la RNAsa P), los cuales, propone, evolucionaron a partir de moléculas de RNA que interaccionaban con péptidos. Dichos péptidos conferían estabilidad a las ribozimas, lo cual fue importante pues permitió la evolución a complejos RNA-proteína de mayor tamaño y, probablemente, con más actividades catalíticas (Figura 2) (Cech 2009a; 2009b).

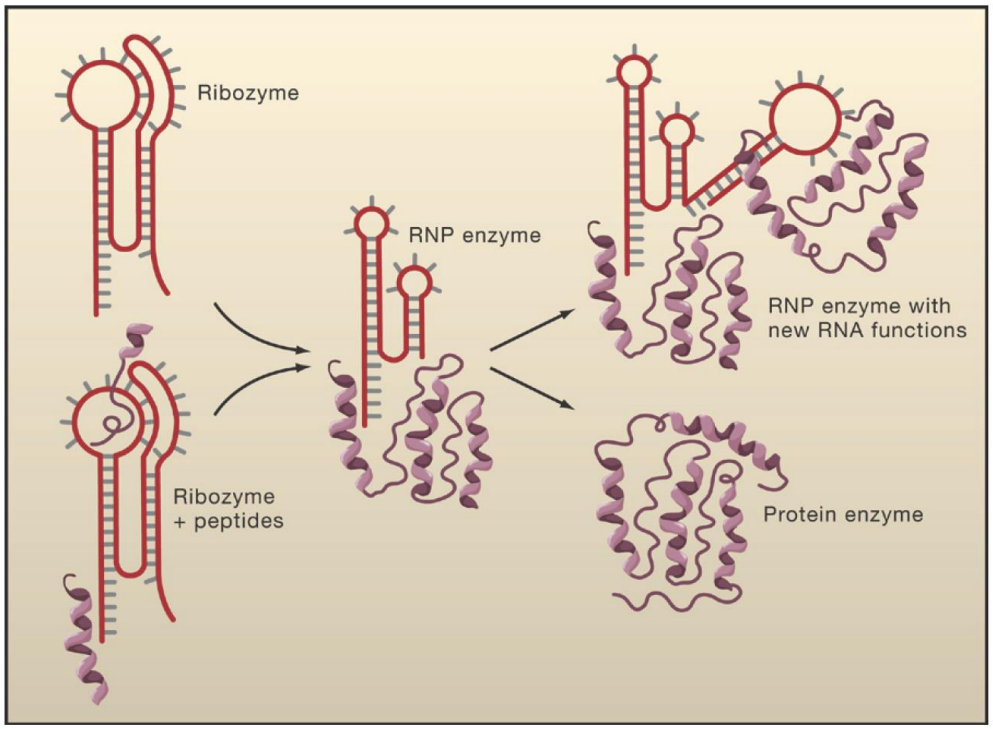

Figura 2. Esquema evolutivo de Cech (2009a; 2009b). La concepción de mundo de RNA de Cech engloba péptidos que interaccionaban con el RNA, estabilizándolo. Estos complejos evolucionaron a complejos ribonucleoproteínicos (RNPs) que poseían actividad catalítica y que, a su vez, fueron seleccionados positivamente, evolucionando en RNPs más grandes (e. g. ribosoma). En este escenario evolutivo, las proteínas pudieron haber substituido también completamente la porción catalítica de RNA, dando lugar a enzimas compuestas enteramente por proteína. Tomada de Cech, 2009a

Una revisión detallada de la propuesta de Cech (Figura 2) revela que, al igual que en el esquema propuesto por White III, (Figura 1) hay un requerimiento indispensable para que la transición de un mundo de RNA a uno con RNA/proteínas (mundo RNP) se pueda llevar a cabo: el establecimiento de una maquinaria de síntesis de proteínas, esto es, un proto-ribosoma (Davidovich, et al., 2010).

El origen de un ribosoma primitivo garantizaría una reserva suficientemente grande y constante de péptidos y proteínas con cargas positivas y demás características químicas necesarias para interaccionar con el RNA. De esta manera, los complejos formados por RNA y péptidos (y/o proteínas) podrían perdurar a pesar de que las secuencias no estuvieran codificadas como lo están en los sistemas biológicos modernos; es decir, diferentes secuencias de cadenas polipeptídicas podrían unirse al RNA y fungir como estabilizadores o andamios del mismo, optimizando su actividad catalítica (Vázquez-Salazar et al., *in prep*). La evolución del proceso de traducción, así como el origen del código genético marcarían, entonces, el final del mundo de RNA para dar lugar a un metabolismo donde las proteínas y las enzimas comienzan a tener un papel cada vez más predominante en la catálisis biológica, esto es, un mundo de RNA/proteínas (Figura 3).

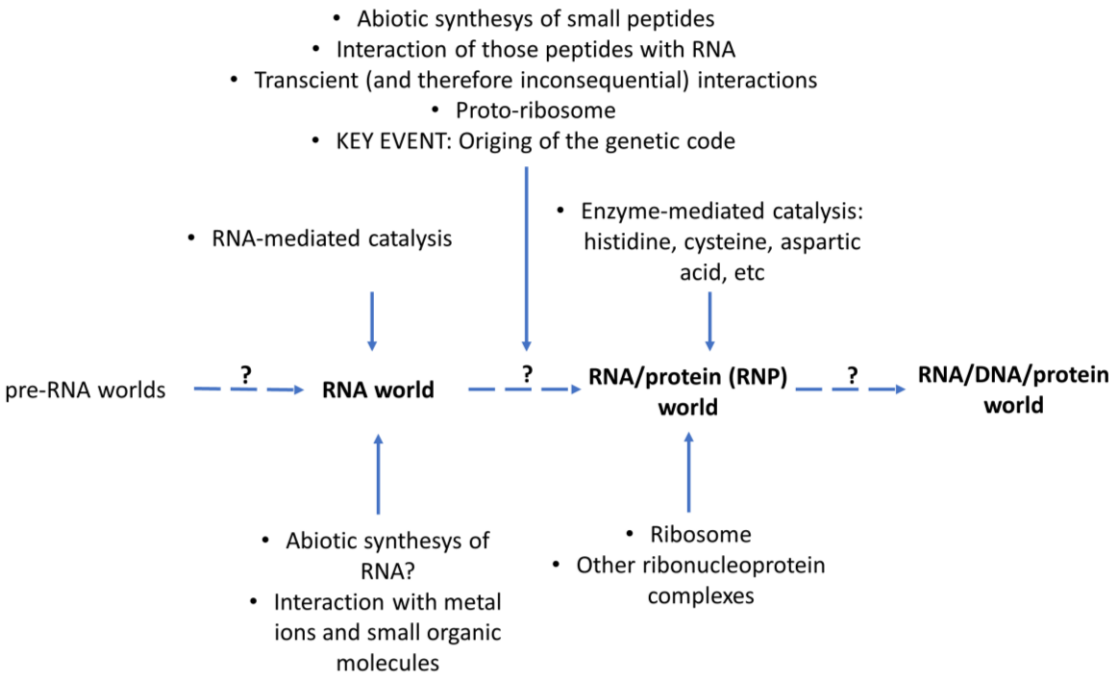

Figura 3. Transición de un mundo de RNA a un mundo de RNA/proteínas (RNP).

Figura 3 (Continuación). Para describir de manera detallada esta transición se debe de describir la evolución de la colaboración entre las moléculas de RNA y las proteínas, donde el evento más importante debió haber sido el origen de la maquinaria de traducción primitiva (i.e. el proto-ribosoma) y el posterior origen del código genético

En su artículo de 1976, White III también escribió “es concebible que los grupos catalíticos que anteriormente formaban parte de las enzimas de ácido nucleico fueran tan importantes para la catálisis general que el grupo catalítico se incorporó a aminoácidos específicos en lugar de ser retenido como una coenzima. En particular, el grupo imidazol de la histidina puede ser un ejemplo” (White III, 1976). Resulta sumamente importante resaltar, tal y como lo hizo White III, que la histidina es el único aminoácido cuya biosíntesis comienza a partir de un ribonucleótido, pues el primer paso de su biosíntesis corresponde a una reacción de condensación entre el fosforribosil pirofosfato (PRPP) y una molécula de ATP, catalizado por la enzima HisG, formando fosforribosil-ATP (Figura 4). No obstante, el anillo imdazólico presente en la estructura química de la histidina no es el mismo que el de la molécula de ATP involucrada en el primer paso; en realidad, el grupo imidazol del ATP se recicla en forma de AICAR hacia la síntesis *de novo* de purinas (Vázquez-Salazar, et al., 2018). Si bien la estructura química de la histidina contiene átomos que provienen del ATP (C y N en color verde en la Figura 4) y puede considerarse como el derivado de un ribonucleótido, el anillo imidazólico tiene también átomos provenientes del PRPP y un átomo de nitrógeno procedente de la glutamina (Figura 4).

Así, aunque White III está en lo cierto al mencionar la relación entre la histidina y los ribnocleótidos, los anillos imidazólicos de cada molécula son distintos (como se discutirá a detalle en las secciones: Capítulo II, ARTÍCULO I y ARTÍCULO II)

De esta manera, la presencia del grupo imidazol en histidina puede ser entendida si se toma en cuenta un origen temprano de la biosíntesis de dicho aminoácido (Alifano, et al., 1996; Vázquez-Salazar, et al., 2018). Esto es, la ruta biosintética de histidina debe de haber sido una de las primeras en establecerse.

Esto implica necesariamente un proceso de evolución biológica pues, como se discute a continuación, la histidina pudo haber estado ausente en el caldo prebiótico de la Tierra primitiva.

A pesar de que Shen y colaboradores reportaron la síntesis de la histidina en posibles condiciones prebióticas (Shen, et al., 1987; 1990), un examen detallado de esta reacción sugiere que las condiciones de síntesis no son realistas prebióticamente. La reacción propuesta comienza a partir de D-eritrosa y formamidina, la condensación de estos dos compuestos químicos da lugar a imidazol-4-acetaldehído, que luego forma histidina a través de una síntesis de Strecker con un rendimiento relativamente alto (3.5% basado en la proporción de His/eritrosa). Aunque la selección *a priori* del enantiómero D de la eritrosa es irrelevante (los estereocentros no participan en la síntesis, y la reacción funcionaría si se usaran otras aldosas de 4 carbonos como precursores), la D-eritrosa y la D-treosa son productos menores e inestables de la reacción de formosa (Decker, et al., 1982), y por lo tanto no pudieron estar disponibles en la Tierra primitiva en la concentración requerida para la reacción propuesta por Shen et al. (1987; 1990). Además, la formamidina es un compuesto lábil que se hidroliza rápidamente en ácido fórmico y amoníaco (Bada, et al., 2016), lo que también hace que sea muy difícil lograr la concentración requerida para la reacción propuesta (0.3 M). Todos estos datos dificultan la plausibilidad prebiótica de esta síntesis (Vázquez-Salazar, et al., 2017).

A pesar de que existe un amplio catálogo de aminoácidos sintetizados bajo condiciones prebióticas (Tabla 1), tanto en experimentos que simulan una atmósfera reductora (CH_4, NH_3, H_2O y H_2) (Miller, 1953), como en experimentos que simulan una atmósfera neutra (H_2O, N_2, CO_2) (Cleaves, et al., 2008), o una atmósfera rica en H_2S (Parker, et al., 2011a; 2011b), a la fecha, no se ha detectado ningún aminoácido que contenga el grupo imidazol; más aún, la búsqueda de histidina y sus productos de degradación en condritas carbonáceas también ha arrojado resultados negativos (Burton, et al., 2012). Estos resultados sugieren que la histidina pudo haber estado totalmente ausente del ambiente prebiótico y que, como se argumente aquí, puede ser un producto muy temprano de la evolución biológica.

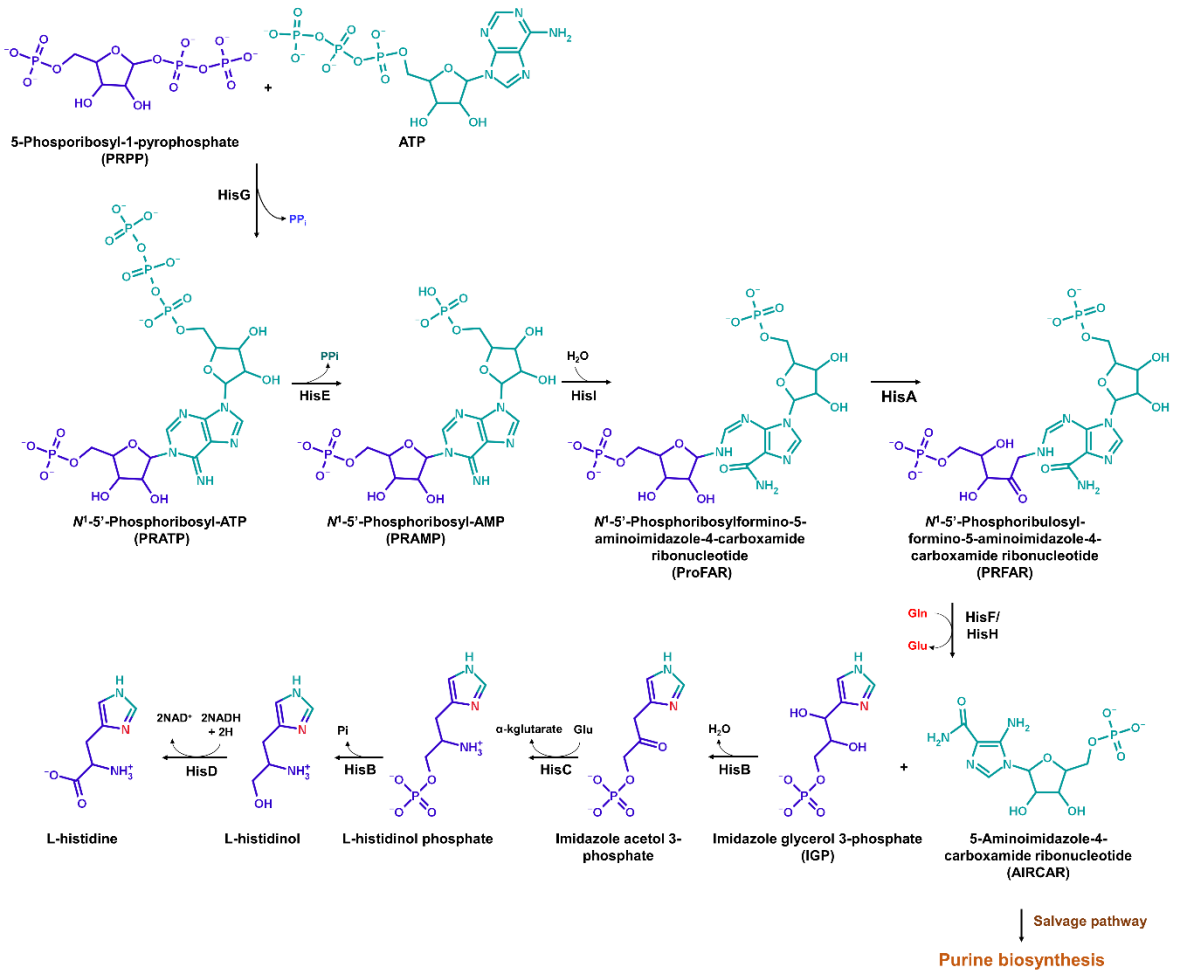

Figura 4. Biosíntesis de la histidina. Se muestran los nueve pasos enzimáticos necesarios para transformar al PRPP (azul) y al ATP (verde) en una molécula de histidina. En la molécula de histidina se pueden apreciar en colores los átomos que provienen de cada una de las moléculas precursoras, así como el átomo de nitrógeno (en rojo) que proviene de la glutamina (Gln) y que se incorpora en la reacción de ciclación del grupo imidazol.

Glu, glutamato; Pi, fosfato inorgánico; PPi, pirofosfato; NAD^+, dinucleótido de nicotinamida y adenina oxidado; NADH, dinucleótido de nicotinamida y adenina reducido; HisG, ATP fosforibosiltransferasa; HisE, PRATP pirofosfohidrolasa; HisI, PRAMP ciclohidrolasa; HisA, isomerasa ProFAR; HisF/HisH, imidazol glicerol fosfato sintasa; HisF, IGPS: subunidad ciclasa; HisH, IGPS: subunidad glutamina amidotransferasa (GATasa tipo I); HisB, imidazol-glicerol-fosfato deshidratasa; HisC, istidinol-fosfato aminotransferasa; HisD, histidinol deshidrogenasa

Tabla 1. Algunos aminoácidos sintetizados bajo condiciones que asemejan la Tierra primitiva (síntesis prebióticas)

Aminoácido	**Referencias**	**Presente en meteoritos** (Burton et al., 2012)
Glicina	Miller, 1953; Cleaves, et al., 2008; Parker, et al., 2011a	Sí
Alanina	Miller, 1953; Cleaves, et al., 2008; Parker, et al., 2011a	Sí
Serina	Cleaves, et al., 2008; Parker, et al., 2011a	Sí
Treonina	Parker, et al., 2011a	Sí
Ác. aspártico	Miller, 1953; Cleaves, et al., 2008; Parker, et al., 2011a	Sí
Ác. glutámico	Miller, 1953; Cleaves, et al., 2008; Parker, et al., 2011a	Sí
Valina	Parker, et al., 2011a	Sí
Isoleucina	Parker, et al., 2011a	Sí
Leucina	Parker, et al., 2011a	Sí
Metionina	Parker, et al., 2011a; Parker, et al., 2011b	No
Cisteamina*****	Parker, et al., 2011a; Parker, et al., 2011b	No
Ác. homocisteico	Parker, et al., 2011b	No
Metionina sulfóxido	Parker, et al., 2011b	No
S-metil cisteína	Parker, et al., 2011b	No
Etionina	Parker, et al., 2011b	No

La coincidencia del listado de aminoácidos sintetizados en experimentos que simulan la Tierra primitiva con el catálogo de aminoácidos encontrados en condritas puede ser un indicador que robustece la hipótesis de que dichos aminoácidos estaban presentes en la Tierra primitiva. Nótese la ausencia de histidina en ambos listados

Catálisis biológica en la célula contemporánea: histidina, cisteína y ácido aspártico

La idea elemental detrás de los esquemas evolutivos planteados por White III (1976; 1982) y Cech (2009a; 2009b) es que el sitio catalítico de ribozimas y de enzimas se encuentra bajo presiones de selección muy estrictas. Esto, durante el proceso evolutivo, genera un nivel de conservación muy elevado en estas regiones comparado con el resto de la secuencia o estructura (por ejemplo, regiones estructurales que no están involucradas en la catálisis o en la unión de substratos, reguladores alostéricos, etc.). Es decir, el sitio catalítico puede catalogarse como el componente más antiguo de ribozimas, de complejos ribonucleproteínicos y de enzimas.

Por esta razón, el análisis de los componentes de los sitios catalíticos en enzimas (iones metálicos, coenzimas y residuos de aminoácidos) arroja datos relevantes no sólo sobre la química de la catálisis biológica, sino también sobre su evolución. Por ejemplo, diversos estudios de caracterización bioquímica (Liao, et al., 2013), así como análisis bioinformáticos recientes (Bartlett, et al., 2002; Holliday, et al., 2007; 2011; Ribeiro, et al., 2018), han demostrado que la histidina es el aminoácido que se encuentra con más frecuencia como residuo catalítico en sitios activos (Figura 5) (Vázquez-Salazar, et al. 2017; 2018).

En un estudio publicado en 2002, Barlett y colaboradores argumentan que existe una inconsistencia para la definición de un residuo catalítico; por esta razón, los autores proponen que, para que un residuo pueda ser considerado como catalítico, éste debe cumplir las siguientes condiciones: i) debe tener una participación directa en el mecanismo catalítico; por ejemplo, como un nucleófilo; ii) debe ejercer un efecto sobre otro residuo (o molécula de agua) que esté directamente involucrado en el mecanismo catalítico; por ejemplo, ejercer una acción electrostática o de tipo ácido-base); iii) estabilizar el estado de transición, y iv) ejercer un efecto sobre un sustrato o cofactor que contribuya a la catálisis; por ejemplo, polarizando un enlace que se rompa en el mecanismo catalítico; esto incluye efectos estéricos y electrostáticos (Barlett, et al., 2002). Con esta definición. Barlett y colaboradores (2002) analizaron una base de datos con información

cristalográfica y bioquímica para 178 sitios activos; en ellos, la distribución de frecuencia observada para cada residuo indica que la histidina constituye el porcentaje más alto (18%) de todos los residuos catalíticos en las proteínas, aunque tiene un porcentaje de abundancia global bajo en la base de datos (2.7%). El segundo residuo más frecuente es el aspartato (15%, con una abundancia global de 5.7%), mientras que el tercero es la arginina (11%, con abundancia global de 4.9%). Es interesante notar que el residuo de cisteína constituye el 5.6% de los residuos catalíticos de la base de datos, con abundancia de 1.2% (Barlett et al., 2002).

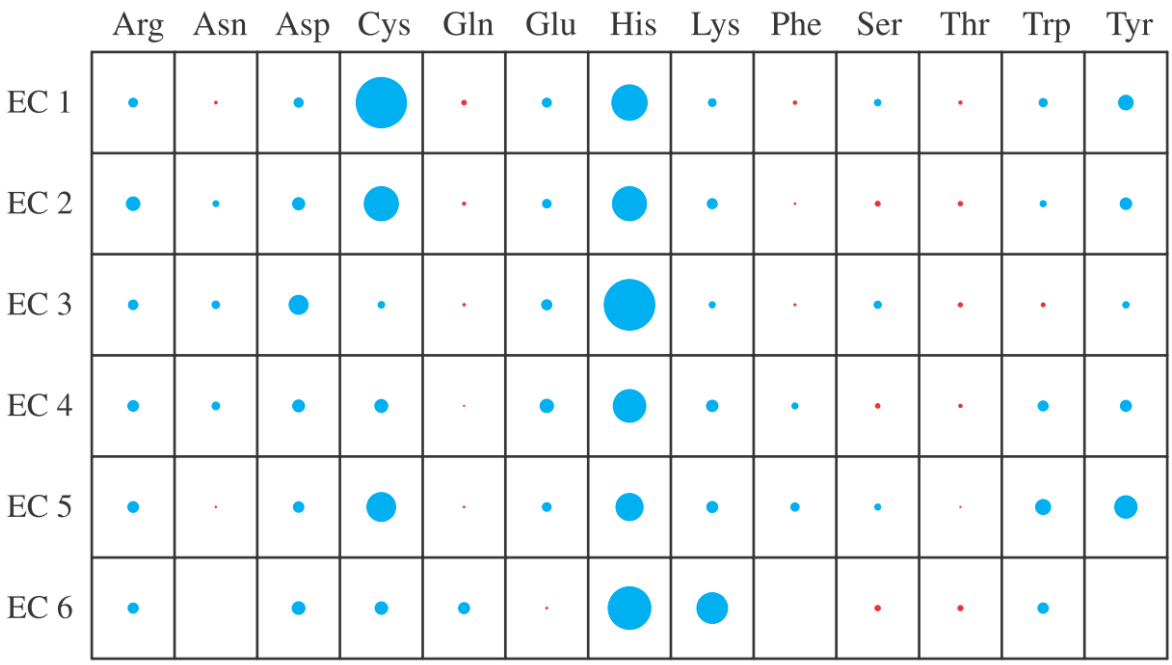

Figura 5. Representación gráfica en donde se muestra la propensión de 13 aminoácidos para ser catalíticos en los sitios activos de las distintas clases de enzimas. Se observa como la histidina aparece significativamente representada en los sitios activos de todas las clases, siendo el más propenso en tres (EC3, EC4 y EC 6); mientras que la cisteína es el segundo aminoácido más representado. La cuantificación de los datos con los que fue preparada esta figura muestra que el tercer aminoácido más representado es el ácido aspártico. EC 1, oxidoreductasas;

EC 2, transferasas; EC 3, hidrolasas; EC 4, liases; EC 5, isomerasas; EC 6, ligasas. Tomada de Holliday, et al., 2011.

Los análisis con una muestra más amplia a las 178 proteínas de la base de datos de Barlett et al. (2002) demuestran la misma tendencia para la histidina (Holliday, et al., 2007; 2011; Ribeiro, et al., 2018).

En 2007, Holliday y colaboradores analizaron una base de datos compuesta de 202 diferentes mecanismos catalíticos; al incrementar más la muestra y analizar 280 mecanismos de reacción únicos (cubriendo 268 números de la *Enzyme Commission* - EC), los autores demuestran que la histidina es el residuo más propenso a ser catalítico, donde la propensión catalítica es una medida de la frecuencia con la que un residuo "*x*" es catalítico en comparación con sus niveles de fondo en una proteína, y se calcula dividiendo el porcentaje de cada residuo que es catalítico entre el porcentaje total de ese residuo en todo el conjunto de datos de proteínas (Holliday, et al., 2007; 2011). En los resultados publicados por Holliday y colaboradores, el residuo de cisteína fue el segundo más propenso a ser catalítico, mientras que aspartato fue el tercero (Holliday, et al., 2007; 2011). Esto representa un cambio en el orden del listado con respecto a los análisis publicados por Barlett y colaboradores en 2002, y significa que el número de sitios y mecanismos analizados sí tiene un efecto en las estadísticas, excepto para la histidina.

La base de datos *Mechanism and Catalytic Site Atlas* (Ribeiro, et al., 2018) contiene (hasta el 9 de septiembre de 2019) 964 entradas curadas a mano, 684 con el mecanismo catalítico detallado. El análisis de esta base de datos continúa demostrando que la histidina es el más frecuentemente encontrado en sitios activos como residuo catalítico, como se resume en la Tabla 2.

En la Tabla 2 también se aprecia el caso de la cisteína que, junto con el triptófano, es el residuo con menor frecuencia en las proteínas de la base de datos (1.4%); esto genera que, a pesar de tener una frecuencia no tan elevada en sitios activos (6.2%), su propensión catalítica sea la segunda más grande (~4).

La disponibilidad de todos estos datos indica, entonces, que el análisis evolutivo de los sitios activos debe, al menos en parte, enfocarse en describir la historia evolutiva de los residuos que los componen.

Tabla 2. Propensión catalítica de diferentes residuos en la base de datos *Mechanism and Catalytic Site Atlas* (Ribeiro, et al., 2018)

Residuo	Frecuencia en sitios catalíticos (%)	Frecuencia en proteínas (%)	Propensión catalítica
Histidina	19.2	2.4	8.08
Aspartato	17.4	5.8	3.01
Glutamato	13.4	6.4	2.09
Arginina	9.5	5.1	1.87
Lisina	8.9	5.5	1.63
Tirosina	6.2	3.4	1.82
Cisteína	6.2	1.4	4.54
Triptófano	1.5	1.4	1.05

Con base en este razonamiento, en el presente trabajo se analizó a detalle la biosíntesis del aminoácido más propenso a ser catalítico en enzimas: histidina. Los resultados publicados en revistas indexadas y que componen la presente tesis doctoral se enfocan en el análisis de la química prebiótica, bioquímica e implicaciones evolutivas del grupo imidazol (la cadena lateral de la histidina), y su importancia en la evolución de la catálisis biológica.

CAPITULO II. EL GRUPO IMIDAZOL DE LA HISTIDINA: IMPLICACIONES CATALÍTICAS Y EVOLUTIVAS

El anillo imidazólico (IUPAC: 1H-Imidazol) es un sistema heterocíclico de cinco miembros, que contiene un nitrógeno terciario y un grupo imino. Este grupo químico se encuentra presente en la estructura de varias moléculas importantes para el metabolismo celular, entre ellas, las purinas y la histidina. Son precisamente las características fisicoquímicas del grupo imidazol las que permiten comprender las propiedades catalíticas de la histidina. Por esta razón, el análisis evolutivo de la biosíntesis de dicho grupo químico resulta esencial para comprender no sólo las implicaciones de este aminoácido en la catálisis biológica, sino también, para indagar el papel que pudo haber tenido en etapas evolutivas anteriores, esto es, la química prebiótica, el mundo de RNA y el mundo de RNA/proteínas (Vázquez-Salazar, et al., 2017).

La biosíntesis de histidina ha sido estudiada ampliamente y ha servido como un modelo de procesos tales como el control por retroalimentación, interacciones alostéricas y la canalización de metabolitos (Alifano, et al., 1996; Winkler & Ramos-Montañez, 2008). Además, existen características en la expresión del operón de histidina en bacterias (operón *his*), así como reacciones bioquímicas e intermediarios dentro de su biosíntesis que han sido catalogadas como “inusuales”, con lo cual su estudio resulta de sumo interés para entender rasgos excepcionales en la biología. Adicionalmente, como muestra de la importancia de los iones metálicos en la catálisis biológica, la propia biosíntesis de histidina depende de Mg^{2+}, Zn^{2+} y Mn^{2+} en varios pasos (Winkler & Ramos-Montañez, 2008).

El análisis de la biosíntesis de la histidina también demuestra que dicha ruta metabólica se encuentra altamente relacionada con la síntesis *de novo* de purinas (Alifano, et al., 1996; Vázquez-Salazar, et al., 2018). Ambas rutas comienzan a partir del PRPP y, pasos más adelante, están conectadas por el intermediario 5-aminoimidazol-4-carboxamida ribonucleotido (AICAR). Como se muestra en la Figura 6, el AICAR se sintetiza en ambas rutas, pero se usa solamente en la síntesis *de novo* de purinas. Sin embargo, la enzima que cataliza la síntesis del motivo imidazólico de la histidina no guarda una relación de ancestría-descendencia con la

enzima que sintetiza el grupo imdazol de purinas. Este fenómeno puede ser comprendido como un caso de convergencia bioquímica funcional (Vázquez-Salazar et al. 2018)

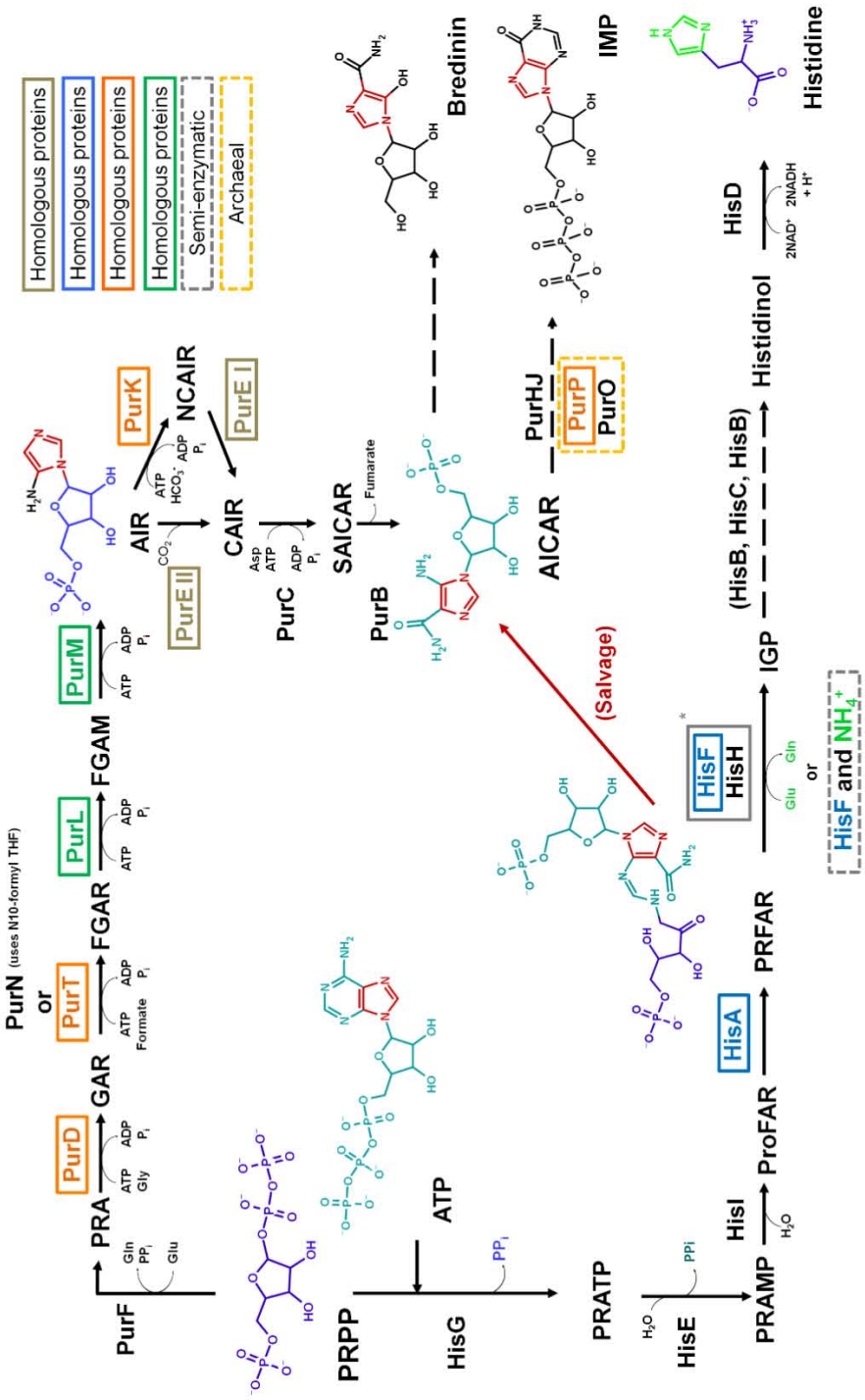

Figura 6. Rutas de biosíntesis de purinas (parte superior) y de histidina (parte inferior).

Figura 6 (Continuación). Se puede observar que ambas rutas se interconectan a través del PRPP y de AICAR, un importante intermediario imidazólico. Se ilustra cómo el grupo imidazol de la molécula de ATP se incorpora a la estructura química de AICAR, el cual pasa a la síntesis *de novo* de purinas por medio de una ruta de salvamento; mientras que el imidazol en IGP se forma independientemente. Se resaltan con recuadros de colores a las proteínas que son homólogas dentro de las rutas, así como las reacciones semi-enzimáticas.

Gln, glutamina; Glu, glutamato; Pi, fosfato inorgánico; PPi, pirofosfato; N^{10}-formil THF; HCO^{3-}, bicarbonato; Asp, ácido aspártico;NAD^{+}, dinucleótido de nicotinamida y adenina oxidado; NADH, dinucleótido de nicotinamida y adenina reducido; $NH4^{+}$, amonio; PRA, 5-fosforibosilamina; GAR, 5'-Fosforibosilglicinamida; FGAR, N-Formyl-GAR; FGAM, 1- (5'-fosforibosil) -N-formilglicinamidina; AIR, 1- (5'-Fosforibosil) -5-aminoimidazol; CAIR, 1- (5'-Fosforibosil) -5-amino-4-imidazolcarboxilato; NCAIR, 5-carboxiamino-1- (5-fosfo-D-ribosil) imidazol; SAICAR, succinil-AICAR; AICAR, ribonucleótido de 5-aminoimidazol-4-carboxamida; PurF, Amidophosphoribosyltransferase; PurD, GAR sintetasa; PurN, GAR transformilasa; PurT, sintetasa FGAR; PurL, fosforibosilformilglicinamidina sintasa; PurM, fosforibosilformilglicinamidina ciclo-ligasa; PurK, NCAIR sintetasa; PurE I, NCAIR mutase; PurE II, AIR carboxilasa; PurC, SAICAR sintasa; PurB, adenilosuccinato liasa; PurHJ, AICAR transfromilasa / IMP ciclohidrolasa; PurP, sintetasa FAICAR; PurO, IMP ciclohidrolasa II

Tomada de Vázquez-Salazar et al. 2018.

Formación del grupo imidazol en la biosíntesis de histidina y purinas

Posterior a la reacción de condensación entre PRPP y ATP, catalizada por la enzima HisG, los siguientes pasos en la vía implican la hidrólisis irreversible de PRATP a PRAMP y pirofosfato, que cataliza el dominio carboxilo-terminal de la enzima HisI (denominado en este trabajo como HisE); a continuación, el dominio amino terminal de HisI cataliza la apertura del anillo de pirimidina del ATP, formando ProFAR. La ribosa proveniente del PRPP y presente en ProFAR sufre un reordenamiento de Amadori catalizado por HisA (Figura 3). Este reordenamiento de Amadori genera

PRFAR, el cual es el substrato para el complejo HisH/HisF, dicho complejo es de particular importancia para el presente trabajo, como se describe a contuniación.

El quinto paso de la biosíntesis de histidina corresponde a la reacción de ciclación del grupo imidazol (Figura 3; 6). Dicha reacción es catalizada por la enzima bifuncional imidazol glicerol fosfato (IGP) sintasa. En bacterias, esta enzima es un heterodímero formado por los productos de los genes *hisH* y *hisF*, donde *hisH* codifica para una amidotransferasa de glutamina (GATasa tipo I), y *hisF* codifica para una ciclasa que adopta el plegamiento de barril $(\beta/\alpha)_8$, también referido como barril TIM (Chaudhuri, et al., 2001; 2003). En eucariontes estos dos genes sufrieron un proceso de fusión que resultó en la formación del gen *His7* (Brilli & Fani, 2004). La IGP sintasa cataliza la conversión de N '- ((5'-fosforibulosil) formimino) -5-aminoimidazol-4-carboxamida-ribonucleótido (PRFAR) y glutamina a IGP, AICAR, y glutamato.

La unión de PRFAR al sitio activo de HisF estimula la actividad de glutaminasa en la enzima HisH, revelando una regulación alostérica (Rivalta, et al, 2012). Sin embargo, cabe señalar que la síntesis de IGP puede tener lugar en ausencia de la proteína HisH *in vivo* en la gammaproteobacteria *Klebsiella pneumoniae*, donde el gen que codifica HisH fue mutado experimentalmente y la reacción restaurada mediante la adición de una alta concentración de NH_4^+ (Reider, et al., 1994). Esta evidencia fue analizada por Lazcano y Miller (1999), quienes propusieron la idea de un metabolismo semi-enzimático, en el cual las vías metabólicas evolucionaron explotando el repertorio preexistente de compuestos sintetizados de forma abiótica, así como los péptidos catalíticos existentes (Lazcano & Miller, 1999).

Los pasos finales en la biosíntesis de histidina incluyen una reacción de deshidratación catalizada por el dominio carboxilo terminal de HisB, una reacción de transaminación catalizada por HisC, así como la desfosforilación de L-histidinol catalizada por el dominio amino terminal de HisB. El L-histidinol es transformado en histidina por la deshidrogenasa HisD (Winkler & Ramos Montañez, 2008) (Figura 3).

Por su parte, los ribonucleótidos de purina se biosintetizan *de novo* en una ruta en la que el PRPP proporciona las unidades de fosfo-ribosa para los ribonucleótidos; de esta manera, las purinas se construyen a partir de una serie de reacciones de adición al azúcar que aporta el PRPP. En la ruta de las purinas el anillo imidazólico es sintetizado en el sexto paso (Figura 6) y la reacción es catalizada por la enzima fosforibosilformilglicinamidina ciclo ligasa, codificada por el gen *purM*. Dicha proteína cataliza la formación de 1-(5'-fosforibosil)-5-aminoimidazol (AIR) a partir de 1-(5'-fosforibosil)-N-formilglicinamidina (FGAM), en una reacción dependiente de ATP y que es común para todas las variantes de la ruta de síntesis de purinas conocidas en la naturaleza.

El producto del gen *purM* consiste en una proteína que se divide estructuralmente en dos dominios: el dominio A o N-terminal y el B o C-terminal (Li, et al., 1999). En bacterias, la enzima activa es un homodimero conformado por la interacción de dos láminas-β (una de cada subunidad) formadas por cuatro hojas-β, constituyendo así una estructura semejante a un barril-β de ocho hojas. En algunos eucariontes, incluyendo a los mamíferos, PurM es parte de un complejo que presenta tres distintas funciones enzimáticas, involucrando también la de glicinamida ribonucleotido sintetasa (PurD) y glicinamida ribonucleotido transformilasa (PurN) (Daubner, et al., 1985).

PurM sintetiza el grupo imidazol de AIR, el cual sufre tres transformaciones enzimáticas hasta dar lugar a AICAR. Se debe subrayar que AICAR no sólo es el precursor de la inosina monofosfato (IMP), sino que también es el precursor de la alarmona 5-amino carboxamida ribósido 4-imidazol 5'-trifosfato (ZTP) la cual puede promover la detención del ciclo celular cuando existe una limitación de folato (Bochner & Ames, 1982; Hernández-Morales et al. 2019).

En el presente trabajo fueron analizadas, usando herramientas bioinformáticas (ver Anexo 2) y un enfoque de bioquímica evolutiva, las enzimas HisF y PurM, con la finalidad de describir la relación evolutiva que existe entre ellas.

Resumen de resultados

La distribución filogenética de las enzimas que sintetizan a la histidina y las purinas, y que incluyen a HisF y PurM, es amplia, lo cual indica que el ancestro común de todos los seres vivos (LUCA) ya poseía las rutas para biosintetizar ambas moléculas. Este resultado es consistente con la idea de un establecimiento temprano en la evolución de la biosíntesis de ambas moléculas (Lazcano & Miller, 1999).

Dada la estrecha conexión entre la biosíntesis de histidina y la síntesis *de novo* de purinas, sobre todo en el intercambio del compuesto imidazólico AICAR, así como el hecho de que PurM y HisF sintetizan la ciclación del mismo grupo químico en dichas rutas, es factible hipotetizar que las enzimas puedan ser homólogas. Sin embargo, HisF y PurM poseen plegamientos tridimensionales diferentes y no relacionados evolutivamente. Esto se hace evidente mediante la superposición de las estructuras cristalográficas, donde ninguno de los dominios estructurales se superpone de manera adecuada. El análisis de esta evidencia apoya la idea de que la evolución de HisF y PurM corresponde a un caso de convergencia de tipo funcional.

El ejemplo clásico de convergencia evolutiva es el de las proteasas de serina. Estas proteínas adoptan plegamientos estructurales distintos; sin embargo, poseen los mismos aminoácidos catalíticos e incluso el mismo mecanismo de reacción para llevar a cabo la hidrólisis del enlace peptídico (Wright, et al., 1969; Reardon & Farber, 1995).

Este tipo de convergencia funcional entre plegamientos proteínicos no relacionados evolutivamente ha sido descrito también en las cinasas que catalizan la fosforilación de azúcares, en particular en las familias hexocinasa, ribocinasa y galactocinasas (Bork, et al., 1993). En este caso, se pueden identificar dos tipos de convergencia: por un lado encontramos a las enzimas que evolucionaron de funciones distintas hacia la misma función dentro de un mismo plegamiento proteínico, como es el caso de la especificidad a glucosa, que ha evolucionado varias veces dentro de la familia de hexocinasa; y, por otro lado están las enzimas que evolucionaron a partir de estructuras proteínicas diferentes hacia la misma

función catalítica, como lo ocurrido con la especificidad a fructosa, que ha evolucionado independientemente en la familia de hexocinasa y en la de ribocinasa; este último caso de convergencia es el que presentan también HisF y PurM.

Teóricamente, la función catalítica es la que se encuentra sujeta a la selección natural, no la molécula específica (e. g. proteína o enzima) que presenta dicha función; por esta razón, es posible que diferentes tipos de estructuras proteínicas puedan ser seleccionadas para llevar a cabo la misma función catalítica (McGhee, 2011).

La hipótesis que se propone para explicar este caso de convergencia funcional radica en la diferencia en funciones bioquímicas de las dos moléculas que poseen el grupo imidazol (Vázquez-Salazar, et al., 2018).

En el RNA, el grupo imidazol de las purinas no juega un papel catalítico, pues la actividad catalítica de las ribozimas radica en la estructura terciaria que éstas adopten en el espacio, en su interacción con iones metálicos y en el grupo 2'-hidroxilo de la ribosa. A diferencia de la histidina, el grupo imidazol presente en purinas se encuentra unido covalentemente a un azúcar por medio de un enlace β-glicosídico. Este enlace se encuentra bloqueando al grupo amino ionizable y que es altamente reactivo en el anillo imidazólico de la histidina. Como resultado de ello, los ribonucleótidos de purina son incapaces de participar en reacciones químicas por medio de su imidazol, pues a pesar de que existen otros grupos ionizables dentro de la estructura de las purinas, éstos no son buenos catalizadores debido a que el p*K*a de cada uno de los grupos químicos depende de la forma tautomérica en la que se encuentre la molécula de purina, que por lo general es inferior o superior a un valor catalíticamente óptimo en el contexto intracelular (Steenken, 1989).

En cambio, el anillo imidazólico de la histidina es, desde que es biosintetizado, capaz de catalizar reacciones ácido/base gracias a que se encuentra libre para reaccionar mediante un grupo amino, el cual posee un p*K*a cercano al pH fisiológico (Steenken, 1989). Esta característica, que tiene que ver con la actividad catalítica de cada una de las moléculas, es quizá una de las presiones más

importantes para la evolución de dos sistemas enzimáticos no homólogos para la síntesis del grupo imidazol.

El grupo imidazol en la química prebiótica

Como se mencionó en el Capítulo I, a la fecha no se ha logrado sintetizar ningún aminoácido que contenga el grupo imidazol en experimentos de tipo Miller-Urey o en otro tipo de simulaciones de condiciones de la Tierra primitiva; sin embargo, la historia es totalmente distinta para el grupo imidazol no substituido, pues su síntesis prebiótica se ha logrado en un número extenso de experimentos, como se detalla a continuación.

La adenina fue la primera molécula imidazólica sintetizada en condiciones abióticas a partir de una solución acuosa de cianuro de amonio (NH_4CN) (Oró, 1960). A partir de ese experimento, muchas otras purinas también se han sintetizado en simulaciones prebióticas, incluyendo guanina y xantina (Oró, 1963), hipoxantina (Lowe, et al., 1963), 8-hidroximetiladenina (Schwartz & Bakker, 1989) y 2,6-diaminopurina (Borquez, et al., 2005). Es importante destacar que el grupo imidazol, así como su derivado 2-metilado, fueron obtenidos por Oró y colaboradores en 1984 (Oró, et al., 1984) empleando glioxal, aldehídos y amoníaco como precursores. Estos experimentos son extremadamente relevantes, pues el imidazol no sustituido también pudo haber jugado un papel importante durante el mundo de RNA, ya que se ha publicado que el grupo imidazol puede actuar como cofactor de la ribozima mutada del virus de la hepatitis delta (HDV) (Perrota, et al., 1999). En estos experimentos, la incubación con una solución amortiguadora de imidazol no sustituido (200 mM, pH 7,4) restaura la actividad catalítica de la ribozima de HDV, que contiene una mutación en la citosina catalítica. Esto demuestra que el imidazol por sí solo puede actuar como un co-catalizador en reacciones catalizadas por RNA, lo que podría ser de suma importancia en un mundo de RNA, y de considerable importancia en nuestra comprensión de la evolución temprana de la catálisis biológica, pues demuestra que el mismo principio que gobierna la catálisis ácido-base que desempeña el imidazol de la histidina en enzimas es aplicable a un anillo imidazólico no substituido que interacciona con RNA.

Toda esta evidencia nos permite proponer al grupo imidazol como un puente evolutivo entre la química prebiótica, el mundo de RNA y la catálisis biológica contemporánea, pues en cada una de estas etapas evolutivas dicho grupo químico tiene una importancia significativa para la evolución de la vida y de la catálisis (Figura 7).

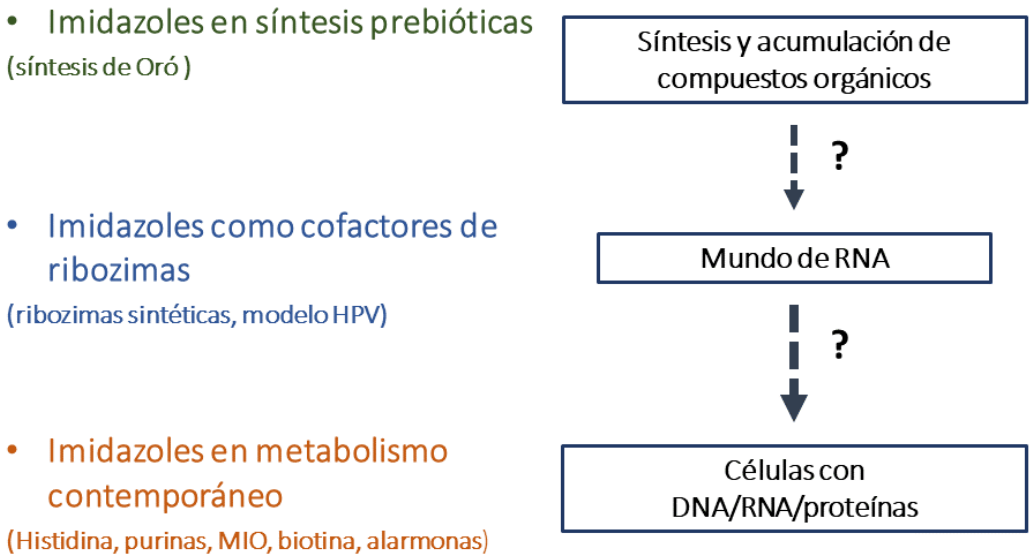

Figura 7. Importancia evolutiva del grupo imidazol. Se esquematizan las etapas principales de la evolución de la vida según la propuesta del origen heterotrófico. A pesar de que no se sabe cómo ocurrieron las transiciones entre cada etapa propuesta, es posible plantear que el grupo imidazol tuvo un papel importante en cada una de ellas, siendo una molécula decisiva en la evolución de la catálisis biológica

Conclusiones

I. El análisis de las enzimas biosintéticas de la histidina muestra que están altamente conservadas en los tres dominios celulares; los eventos de duplicación y fusión de genes que marcaron la evolución de esta ruta demuestran que en etapas anteriores pudo haber tenido menos enzimas, así como reacciones semi-enzimáticas acopladas. Estos datos, aunados a la estrecha relación que existe con la biosíntesis de purinas y la importancia de este aminoácido en la catálisis

biológica podrían implicar que la biosíntesis de histidina fue una de las primeras en establecerse en los sistemas biológicos.

II. PurM y HisF son proteínas con diferentes plegamientos tridimensionales y diferentes secuencias primarias; sin embargo, catalizan reacciones químicamente equivalentes sobre sustratos similares, lo cual habla de una evolución convergente en la función catalítica de dichas enzimas.

III. El origen polifilético de PurM y HisF se entiende mejor cuando se analiza la actividad catalítica del grupo imidazol de la histidina en comparación con el de los ribonucleótidos de purinas, en donde el enlace β-glicosídico entre la ribosa y el nucleótido desempeña un papel clave para atenuar la actividad catalítica del grupo imidazol presente en las purinas.

IV. Las distintas maneras de sintetizar al grupo imidazol, prebióticas, enzimáticas y no enzimáticas, así como la importancia de este grupo químico en diversos procesos celulares, lo colocan como un vínculo evolutivo entre la química prebiótica, el mundo de RNA y la aparición del mundo de RNA/proteínas, resaltando el papel fundamental que jugó para la evolución de la catálisis, así como en diversos procesos celulares.

V. Hay diferentes mecanismos biológicos que producen el grupo imidazol. Esto es una demostración de la redundancia bioquímica que habla, primeramente, de lo químicamente favorecido que se encuentra la ciclación del anillo imidazólico y, por otra parte, de la importancia de los derivados de imidazol en los procesos biológicos, reflejando el papel evolutivo trascendental de dichas moléculas.

Artículo requisito: **Evolutionary convergence in the biosyntheses of the imidazole moieties of histidine and purines**

PLOS ONE

RESEARCH ARTICLE

Evolutionary convergence in the biosyntheses of the imidazole moieties of histidine and purines

Alberto Vázquez-Salazar[1], Arturo Becerra[1], Antonio Lazcano[1,2]*

1 Facultad de Ciencias, Universidad Nacional Autónoma de México, Cd. Universitaria, Cd. de México, México, **2** Miembro de El Colegio Nacional, Centro, Cd. de México, México

* alar@ciencias.unam.mx

OPEN ACCESS

Citation: Vázquez-Salazar A, Becerra A, Lazcano A (2018) Evolutionary convergence in the biosyntheses of the imidazole moieties of histidine and purines. PLoS ONE 13(4): e0196349. https://doi.org/10.1371/journal.pone.0196349

Editor: Israel Silman, Weizmann Institute of Science, ISRAEL

Received: February 8, 2018

Accepted: April 11, 2018

Published: April 26, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Alberto Vázquez-Salazar is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship 404507 from CONACYT. Financial support to Arturo Becerra from PAPIIT-UNAM (IN223916) is gratefully acknowledged.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Background

The imidazole group is an ubiquitous chemical motif present in several key types of biomolecules. It is a structural moiety of purines, and plays a central role in biological catalysis as part of the side-chain of histidine, the amino acid most frequently found in the catalytic site of enzymes. Histidine biosynthesis starts with both ATP and the pentose phosphoribosyl pyrophosphate (PRPP), which is also the precursor for the *de novo* synthesis of purines. These two anabolic pathways are also connected by the imidazole intermediate 5-aminoimidazole-4-carboxamide ribotide (AICAR), which is synthesized in both routes but used only in purine biosynthesis. Rather surprisingly, the imidazole moieties of histidine and purines are synthesized by different, non-homologous enzymes. As discussed here, this phenomenon can be understood as a case of functional molecular convergence.

Results

In this work, we analyze these polyphyletic processes and argue that the independent origin of the corresponding enzymes is best explained by the differences in the function of each of the molecules to which the imidazole moiety is attached. Since the imidazole present in histidine is a catalytic moiety, its chemical arrangement allows it to act as an acid or a base. On the contrary, the *de novo* biosynthesis of purines starts with an activated ribose and all the successive intermediates are ribotides, with the key β-glycosidic bondage joining the ribose and the imidazole moiety. This prevents purine ribonucleotides to exhibit any imidazole-dependent catalytic activity, and may have been the critical trait for the evolution of two separate imidazole-synthesizing-enzymes. We also suggest that, in evolutionary terms, the biosynthesis of purines predated that of histidine.

Conclusions

As reviewed here, other biosynthetic routes for imidazole molecules are also found in extant metabolism, including the autocatalytic cyclization that occurs during the formation of creatinine from creatine phosphate, as well as the internal cyclization of the Ala-Ser-Gly motif of

some members of the ammonia-lyase and aminomutase families, that lead to the MIO cofactor. The diversity of imidazole-synthesizing pathways highlights the biological significance of this key chemical group, whose biosyntheses evolved independently several times.

Introduction

The imidazole group is a five-membered heterocyclic chemical compound containing a tertiary nitrogen and an imino group. It is present in the structure of several key molecules of major biological significance, most notably purines and histidine. The structure of the unsubstituted imidazole ring has two nitrogen atoms in positions 1 and 3 (also named pyrrole- and pyridine-nitrogen atoms, respectively) (Fig 1). The pyrrole nitrogen is a very weak acid with a $pK_a = 14.4$. In contrast, the pK_a of the pyridine nitrogen is 6.9, which allows it to accept a proton under neutral pH conditions [1]. This phenomenon does not occur in the imidazole moiety of purines, but it takes place in the imidazole side-chain of histidine, whose pK_a value is in the range of 6 to 7 as a result of the electron-withdrawing inductive effect of the protonated amino group [2,3]. Bioinformatic analyses have shown that histidine is the most frequently found catalytic residue in the active sites of enzymes [4]. This phenomenon may be explained by the level of ionization of the imidazole moiety of histidine, which is estimated to be between 9–50%, resulting in its well-known ability to act as a general acid or base to donate or accept

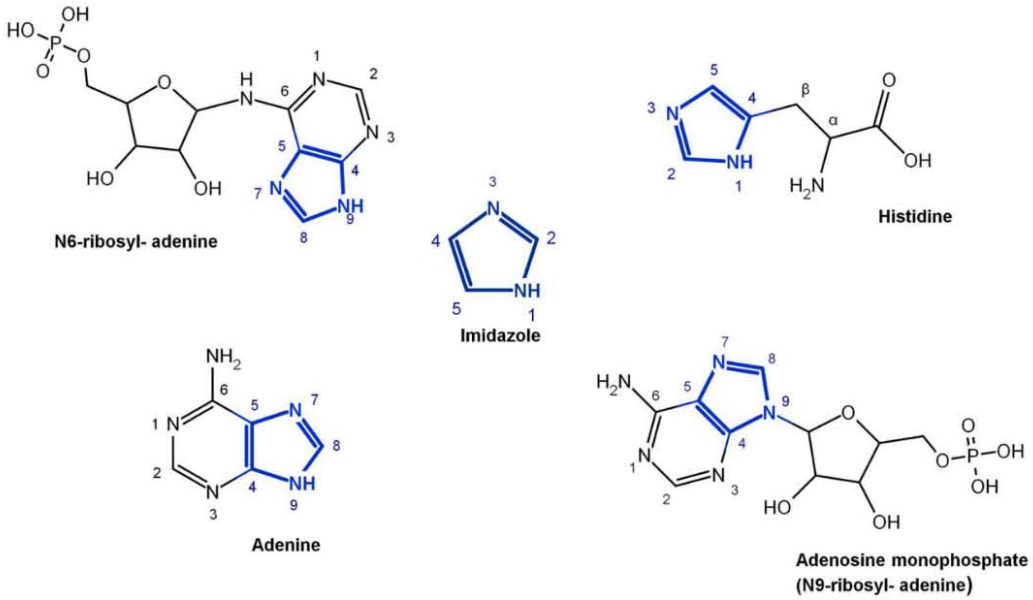

Fig 1. Structure of imidazole and imidazole-derivatives with atom numbering. The unsubstituted imidazole group is shown in the center. Other important imidazole molecules are shown: histidine (2-amino-3-(1H-imidazol-4-yl), adenine (9H–purin-6-amine), adenosine monophosphate (N9-ribosyl-adenine), and N6-ribosyladenine (which synthesis has been reported in prebiotic chemistry experiments).

https://doi.org/10.1371/journal.pone.0196349.g001

protons during a chemical reaction. This property, together with the stable complexes that it can form with different metallic cations such as Cu^{2+}, Co^{2+}, Zn^{2+}, and Mn^{2+} [1,5,6], plays a key role in our understanding of the catalytic propensity of histidine.

Abiotic synthesis of adenine and histidine

Adenine was the first imidazole-bearing molecule synthetized under abiotic conditions from an aqueous solution of ammonium cyanide (NH_4CN) [7]. Since then, many other purines have also been synthesized in prebiotic simulations, including guanine and xanthine [8], hypoxanthine [9], 8-hydroxymethyladenine [10], and 2,6-diaminopurine [11]. Outstandingly, the imidazole group, as well as its 2-methyl derivative were obtained by Oró et al. [12] using glyoxal, aldehydes and ammonia as precursors.

The first attempt to synthesize histidine under possible prebiotic conditions was reported by Shen et al. [13,14]. In the proposed reaction, D-erythrose and formamidine are condensed, leading to imidazole-4-acetaldehyde, which then forms histidine via a Strecker synthesis with a relatively high yield of 3.5% based on the ratio of His/erythrose (Fig 2). As argued below, a detailed examination of this reaction suggests that the conditions of synthesis are prebiotically unrealistic, and that it is unlikely it took place in the primitive Earth.

Even though the *a priori* selection of the D-enantiomer of erythrose is irrelevant since the stereocenters do not partake in the synthesis, and the reaction would work if other 4-carbon aldoses were used as precursors, D-erythrose and D-threose are minor, unstable products of the formose reaction [15], and therefore could not have been available in the concentration required for the reaction reported by Shen et al. [13,14] to take place. In addition, formamidine is a labile compound that is rapidly hydrolyzed into formic acid and ammonia [16], which also makes it very difficult to achieve the concentration required for the proposed reaction (0.3 M). As concluded elsewhere, all this data strongly hinders the prebiotic significance of this synthesis [17].

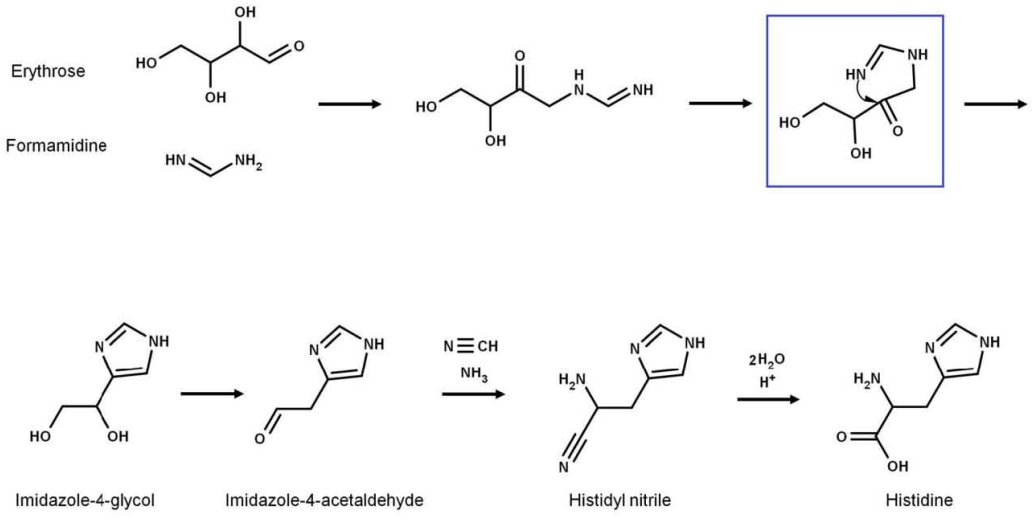

Fig 2. Abiotic synthesis of histidine. The synthesis was proposed by [13,14] and occurs via Strecker synthesis. As shown in the upper part of the scheme, an Amadori rearrangement gives place to the histidine imidazole moiety [cf. ref 17].

https://doi.org/10.1371/journal.pone.0196349.g002

Nevertheless, it is quite interesting that in the reaction mechanism proposed by Shen et al. [13,14] the formation of the imidazole group is the product of an Amadori rearrangement (Fig 2), which is also the mechanism involved in the synthesis of the imidazole moiety of imidazole glycerol phosphate (IGP) by the IGP synthase (HisHF) enzyme during histidine biosynthesis. However, as argued elsewhere [17], all the evidence suggests that there is no direct evolutionary connection between the abiotic synthesis reported by Shen et al. [13,14] and the extant histidine anabolism.

As of today, no imidazole-bearing amino acid has been produced in Miller-Urey-type experiments or in other type of laboratory simulations, and the search of histidine and its degradation products in carbonaceous chondrites has also yielded negative results. These results suggest that histidine as such may have not been present in the prebiotic environment [17–19] but, as argued here, may be in fact a product of very early biological evolution.

Role of imidazolides in the evolution of catalysis

Although it has been shown that clays, metallic cations, and relatively simple compounds like proline and other amino acids can affect the rates of chemical reactions under possible primordial conditions, the emergence of biological catalysis remains one of the key questions in modern biology.

The discovery of the catalytic properties of RNA [20,21] provided support for the RNA world proposal, where it is assumed that RNA molecules played a key role in both heredity and catalysis during the very early stages of biological evolution.

Although how the RNA world evolved from the prebiotic environment remains an open issue, there have been advances in our understanding of the role that imidazole and imidazole-derivatives could have played during this evolutionary stage. For instance, it has been demonstrated that 2-aminoimidazole can act as an activating group for non-enzymatic RNA copying by a mechanism involving an imidazolium-bridged dinucleotide intermediate at pH = 7 [22]. Unsubstituted imidazole may have also played a role during the RNA world, as suggested by the demonstration that RNA cleavage can be catalyzed by imidazole in buffer [23]. The unsubstituted imidazole group can also act as a cofactor for the hepatitis delta virus (HDV) self-cleaving ribozyme [24]. The reaction mechanism of this ribozyme involves a nucleophilic attack of the 2′-OH on the phosphorus in the phosphodiester backbone, where the catalytic cytosine accepts the proton from the attacking 2′-hydroxyl group. As shown by Perrotta et al. [24], the incubation with unsubstituted imidazole buffer (200 mM, pH 7.4) restores the catalytic activity of a mutated HDV ribozyme. This shows that imidazole by itself can act as a co-catalyst in reactions involving RNA molecules, which could be of key significance in an RNA world scenario, and of considerable significance in our understanding of the early evolution of biological catalysis.

The catalytic activity of His-containing di- and tri-peptides has also been investigated. Histidyl-histidine (His-His) catalyzes the dephosphorylation of deoxyribonucleoside monophosphate, the hydrolysis of oligo $(A)_{12}$, and the oligomerization of 2',3'-cAMP under cyclic wet-dry reaction conditions [25]. More recent studies have shown that seryl-histidine (Ser-His) catalyzes the oligomerization of trimers of imidazole-activated nucleotides [26,27], and catalyze peptide bond formation between two amino acids, an activity also found in seryl-histidyl-glycine (Ser-His-Gly) [28,29].

Evolution of enzyme catalytic activity from an RNA world

Although there have been alternative proposals to explain the origin of ribosome-mediated protein synthesis based on the simultaneous appearance of both RNA and proteins [30,31], it

can also be argued that the ribosome first appeared as an evolutionary outcome of the interaction of prebiotically synthesized amino acids with catalytic RNA molecules. How enzymes first evolved remains an open question, but it is reasonable to assume that their active sites must be one of their most conserved portions. However, if histidine was absent in the primitive Earth, how can its ample distribution in the active sites of enzymes be explained?

In what may be the first attempt to explain the origin of histidine biosynthesis, over 40 years ago Harold White III correlated the start of the histidine biosynthetic pathway, which involves a condensation reaction between ATP and the sugar phosphoribosyl pyrophosphate (PRPP), with the possibility of an ancient metabolism mediated by nucleic acid enzymes [32]. In his evolutionary biochemistry approach, White proposed that histidine was, in fact, the molecular vestige of an ancient catalytic nucleotide that was part of the RNA world [32,33]. According to White, this possibility is reinforced by a) the well-known fact that the histidine anabolic pathway is connected with the *de novo* synthesis of purines by the usage of PRPP and the 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) intermediate; b) His is the only known amino acid with a ribonucleotide-starting biosynthesis; and c) histidine is the only imidazole-bearing amino acid.

While it is true that histidine is the only biological amino acid with an imidazole side chain, the other two issues merit a reexamination. As underlined by White [32], during the first step of the histidine biosynthesis, PRPP, whose synthesis in prebiotic conditions has been recently reported by Akouche et al. [34], is condensed with a molecule of ATP to form N'5- phosphoribosyl -ATP (PRATP) (Fig 3). Quite surprisingly, however, the imidazole component of the ATP does not partake in the biosynthesis of the imidazole group of histidine. Instead, the histidine's imidazole moiety is biosynthesized *de novo* by a mechanism completely different from that of purines.

As shown on Fig 3, histidine- and purine biosyntheses are indeed connected by the imidazole intermediate AICAR, which is formed independently in the two pathways. In the purine route it is formed from succinyl-AICAR (SAICAR), and from N'-((5'-phosphoribulosyl) formimino)-5-aminoimidazole-4-carboxamide-ribonucleotide (PRFAR) in the histidine pathway, but it does not play any further role in the subsequent steps of the histidine anabolism. Instead, it is recycled directly into the purine pathway, where it is used as a substrate by the bifunctional enzyme AICAR transfromylase/inosine monophosphate (IMP) cyclo-hydrolase (PurHJ) [EC: 2.1.2.3/EC: 3.5.4.10], that catalyzes the final two steps in the route and synthesizes IMP (Fig 3). We suggest that the incorporation of AICAR from the histidine biosynthetic route into the *de novo* synthesis of purines can be best understood as an intracellular salvage pathway (Fig 3), in which the imidazole compound is used to supply the requirements of purine synthesis.

Biosynthesis of the imidazole group in histidine and purines

In the histidine biosynthetic pathway, the enzyme imidazole glycerol phosphate synthase (IGPS) [EC: 2.4.2.-/EC:4.1.3.-] catalyzes the conversion of PRFAR and glutamine to imidazole glycerol phosphate (IGP), AICAR, and glutamate [35]. As mentioned above, AICAR is discarded. In other words, in what appears to be a rather convoluted, non-parsimonious route, the histidine's imidazole moiety is synthesized *de novo* as part of IGP, while AICAR is recycled into purine synthesis.

The bacterial IGPS enzyme is a heterodimer formed by the *hisH* and *hisF* gene products (Fig 4A), where *hisH* encodes a glutamine amidotransferase (GATase type I) subunit [EC: 2.4.2.-], and *hisF* encodes a cyclase subunit [EC:4.1.3.-] that adopts the $(\beta/\alpha)_8$-barrel fold [36,37] (Fig 4B). Binding of PRFAR to the HisF active site stimulates the glutaminase activity

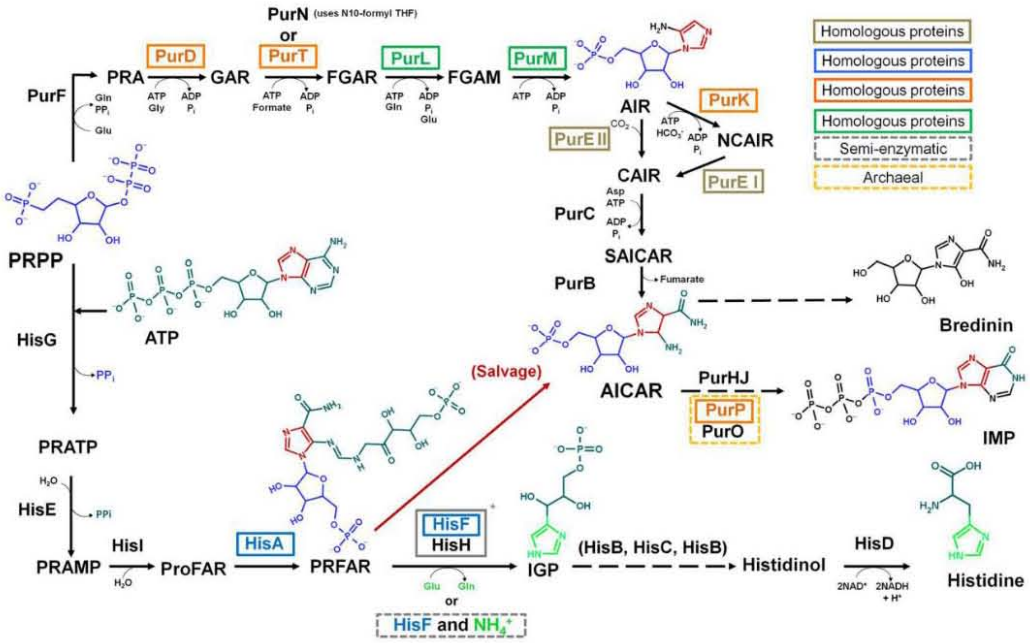

Fig 3. Pathways for the biosynthesis of histidine (bottom) and purines (top). Both routes start with PRPP and are connected by AICAR. The enzymes for each step are depicted. The enzyme name's color and the boxes indicate homologous proteins within pathways as well as the semi-enzymatic reaction. It is evident how the imidazole group of the ATP molecule along with the pentose of PRPP become part of AICAR, and goes into the purine synthesis, while the imidazole in IGP forms independently. The red arrow indicates the salvage pathway in which AICAR is recycled into the de novo synthesis of purines. The dashed arrows indicate two or more steps in the route, and for simplicity the cofactors and coenzymes used in these steps have been omitted. *The complex formed by the HisF and HisH proteins is named imidazole glycerol phosphate synthase (IGPS). The full names of the most relevant metabolites and enzymes are given in the text.

https://doi.org/10.1371/journal.pone.0196349.g003

in the HisH enzyme, revealing an allosteric regulation [38]. The synthesis of IGP can also take place *in vitro* in the absence of the HisH protein under a high concentration of NH_4^+ [39]. This phenomenon was also shown to occur *in vivo* in the gammaproteobacteria *Klebsiella pneumoniae*, where the *hisH* gene was mutated but the reaction was restored in high concentrations of NH_4^+ [40] (Fig 5). As argued by Lazcano and Miller [41], the fact that the pathway can take place in the absence of HisH suggests that a simpler pathway with lesser number of enzymes may have existed in the distant past. This provides an example of the semi-enzymatic origin of metabolic pathways, in which a primitive biosynthetic route could have exploited a favorable chemical environment as well as available enzymes (Fig 5) [41].

As summarized in Fig 3, the ATP-dependent enzyme phosphoribosylformylglycinamidine cyclo-ligase [EC:6.3.3.1] (PurM) synthesizes 1-(5'-Phosphoribosyl)-5-aminoimidazole (AIR) from 1-(5'-phosphoribosyl)-N-formylglycinamidine (FGAM). This enzyme is present in all purine biosynthetic pathway variations found in the three domains of life. Structurally, each PurM monomer is divided into two domains, an A- or N-terminal domain, and a B- or C-terminal domain (Fig 4c) [42]. The active enzyme is a homodimer formed by the interaction of two four-stranded β-sheets (one from each subunit) forming an eight-stranded β-barrel-like structure (Fig 4C) [42]. AIR is then enzymatically processed into SAICAR, which is the

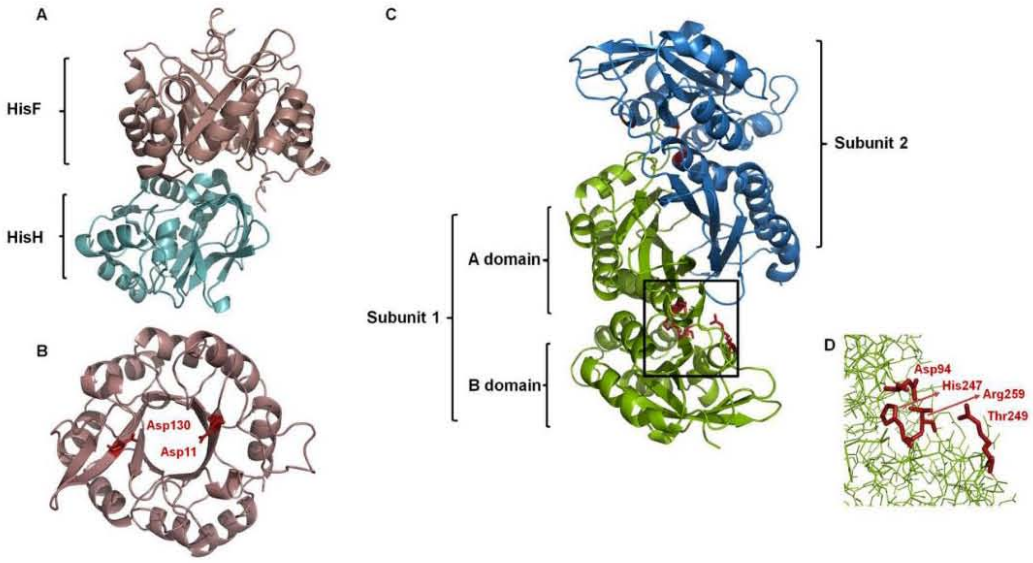

Fig 4. Crystallographic structures of HisF and PurM. (A) Structure of the imidazole glycerol phosphate synthase (IGPS) of *Thermotoga maritima* (PDB:1GPW). (B) Structure of the cyclase (HisF) subunit of *Thermotoga maritima* (PDB: 1THF). Catalytic residues (Asp11 and Asp130) are indicated. (C) Structure of the phosphoribosylformylglycinamidine cyclo-ligase (PurM) of *Escherichia coli* (PDB: 1CLI). (D) Proposed active site for PurM, which comprises the residues Asp94, His247, Thr249, and Arg259.

https://doi.org/10.1371/journal.pone.0196349.g004

substrate of the enzyme adenylosuccinate lyase [EC:4.3.2.2] (PurB), that catalyzes the synthesis of AICAR [43].

In summary, HisF and PurM are not homologues but they catalyze the biosyntheses of the imidazole motif in the histidine and purine pathways (Fig 3) and do so by completely different

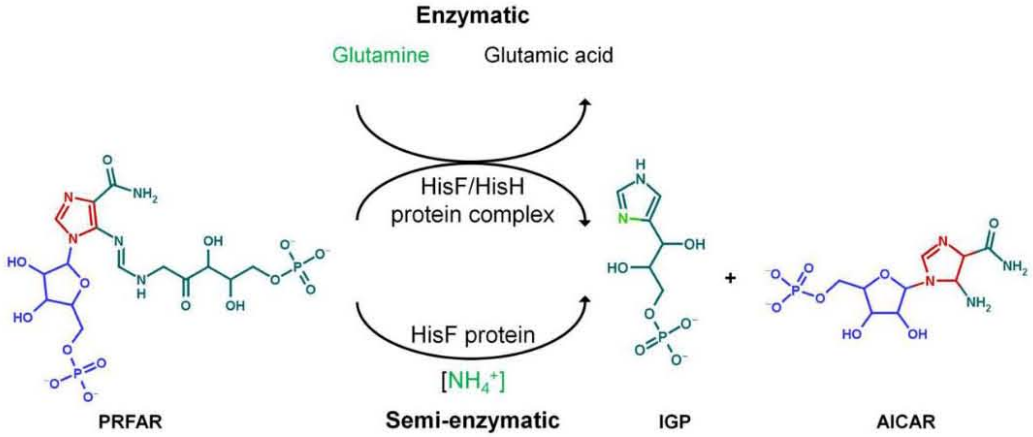

Fig 5. Enzymatic and semi-enzymatic synthesis of imidazole glycerol phosphate (IGP). The enzymatic branch needs both HisF and HisH proteins. HisH transfers nitrogen from glutamine, forming glutamate. In contrast, the semi-enzymatic branch only requires the HisF protein, since the high concentration of NH_4^+ in the medium can replace the transferase activity of HisH. In both cases, IGP and AICAR are formed (modified from ref. [41]).

https://doi.org/10.1371/journal.pone.0196349.g005

mechanisms. The presence of two different independent syntheses of imidazole intermediates in highly conserved, widely distributed routes is rather intriguing and becomes an interesting question from an evolutionary perspective. How can the existence of such biochemically redundant enzymatic complexes in highly connected, early pathways be understood? Here we analyze the active sites of both enzymes and the reactions they catalyze, and propose an explanation for their polyphyletic origin.

Materials and methods

Sequence analysis

The sequences of bacterial HisF (*Escherichia coli* K-12 MG1655: b2025), and PurM (*Escherichia coli* K-12 MG1655: b2499) proteins were retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [44]. Pairwise sequence alignment was performed using the Needleman-Wunsch algorithm as implemented in the EMBOSS Needle program of the European Molecular Biology Open Software Suite [45].

Structural analysis

Visualization and analysis of protein structures were performed using the PyMol package [46]. Structure comparisons were performed using the FATCAT algorithm [47] included in the protein comparison tool of the Protein Data Bank (PDB) web server. DaliLite v3 was used to search for structural related proteins [48].

Structure-based dendogram construction

The available crystal structures of the HisF, PurM, and their paralogues were retrieved from the PDB. A total of 38 structures available as of January 2018 (S1 Table) were analyzed following a previously described procedure [49], only structures with no bounded ligands and no reported mutations were considered in this work. To avoid redundancy, for each protein only the structures with the best resolution were selected if two or more of them were available for the same species. Pairwise structural comparisons were performed using the Secondary Structure Matching (SSM) program of the European Protein Data Bank (PDBe) webserver [50]. The structural alignment score (SAS) for each protein-protein comparison was calculated following the formula: (RMSD x 1000)/number of aligned residues [51]. Geometric distance values were calculated using the SAS and then transformed into evolutionary distances using the FITCH program of the PHYLIP package [52]. Finally, dendograms were constructed using the DRAWTREE program included in PHYLIP. Dendograms were visualized and edited using FigTree version 3.2 [53].

Results

Imidazole-biosynthetic enzymes HisF and PurM: A case of evolutionary convergence

The *E. coli* PurM (354 aa) and HisF (258 aa) proteins share only 12.2% of overall sequence identity (21.4% similarity) (S1 Fig). When the crystallographic structures of the two proteins are superimposed, the root-mean-square deviation (RMSD) of the α-carbons is 8.27 Å (Fig 6A). This very high RMSD value reflects of the very different domain architectures of the two enzymes. As noted above, HisF has the $(\beta/\alpha)_8$-barrel fold and belongs to the ribulose-phoshate binding barrel superfamily, while PurM adopts a completely different fold and is part of the PurM superfamily. Considering these structural differences, it can be readily concluded that

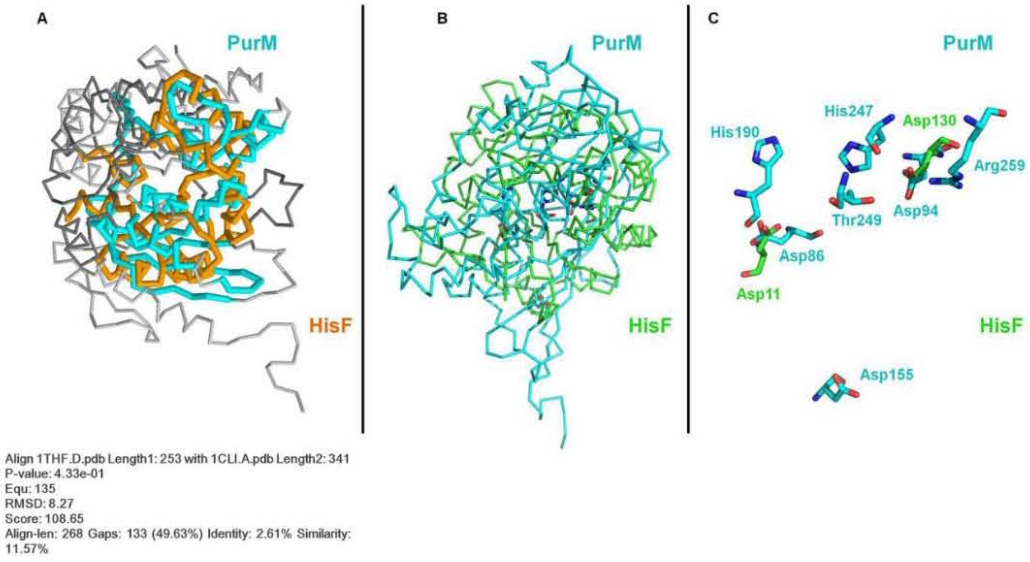

Fig 6. Structural analysis of the HisF and PurM proteins. (A) Structural alignment of HisF from *Thermotoga maritima* (PDB: 1THF) and PurM from *Escherichia coli* (PDB: 1CLI) using the FATCAT algorithm (B) Structural alignment of the active sites of HisF and PurM using the *align* command of PyMol. (C) Catalytic residues that comprise the active sites of HisF and PurM isolated from the alignment in B.

https://doi.org/10.1371/journal.pone.0196349.g006

these are different enzymes with different evolutionary histories that share no common ancestor. The evolutionary relationship between HisF and its structural homologues is depicted in the structural phylogeny shown in Fig 7. The branches more closely related to HisF are those corresponding to the enzyme N'-[(5′-phosphoribosyl) formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase [EC:5.3.1.16] (HisA). It has been shown that HisF and HisA share a similar internal organization that can be explained as two paralogous modules half the size of the extant sequences [54]. An ancestral module duplicated and gave rise to the ancestral *hisA* gene which, in turn, underwent another duplication and fused, giving rise to the *hisF* gene [35,54,55]. This shared evolutionary history explains the closeness between the branches.

Interestingly, HisA catalyzes an Amadori rearrangement that isomerizes the aminoaldose motif in ProFAR into the aminoketose PRFAR. The same chemical rearrangement is responsible for the imidazole ring closure by HisF, and also occurs in the reaction catalyzed by the tryptophan biosynthetic enzyme phosphoribosylanthranilate isomerase [EC:5.3.1.24] (TrpF), which transforms N-(5-phospho-beta-D-ribosyl) anthranilate (PRAn) into 1-(2-carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate (CdRP) [56,57]

Once synthesized, CdRP is then processed by the enzyme indole-3-glycerol phosphate synthase [EC:4.1.1.48] (TrpC), which catalyzes the formation of 1-C-(indol-3-yl) glycerol 3-phosphate (rCdRP), CO_2, and H_2O. Both TrpF and TrpC are homologues to the HisF/HisA pair, and their evolutionary relationship is shown in Fig 7. Although a prebiotic synthesis of tryptophan has been reported [58], it is likely that this amino acid is a latecomer in evolution [59]. This implies that, in fact, the HisF/HisA pair is more ancient than the TrpF and TrpC proteins. This assumed antiquity is concordant with the key catalytic role played by histidine in many extant enzymes (see above).

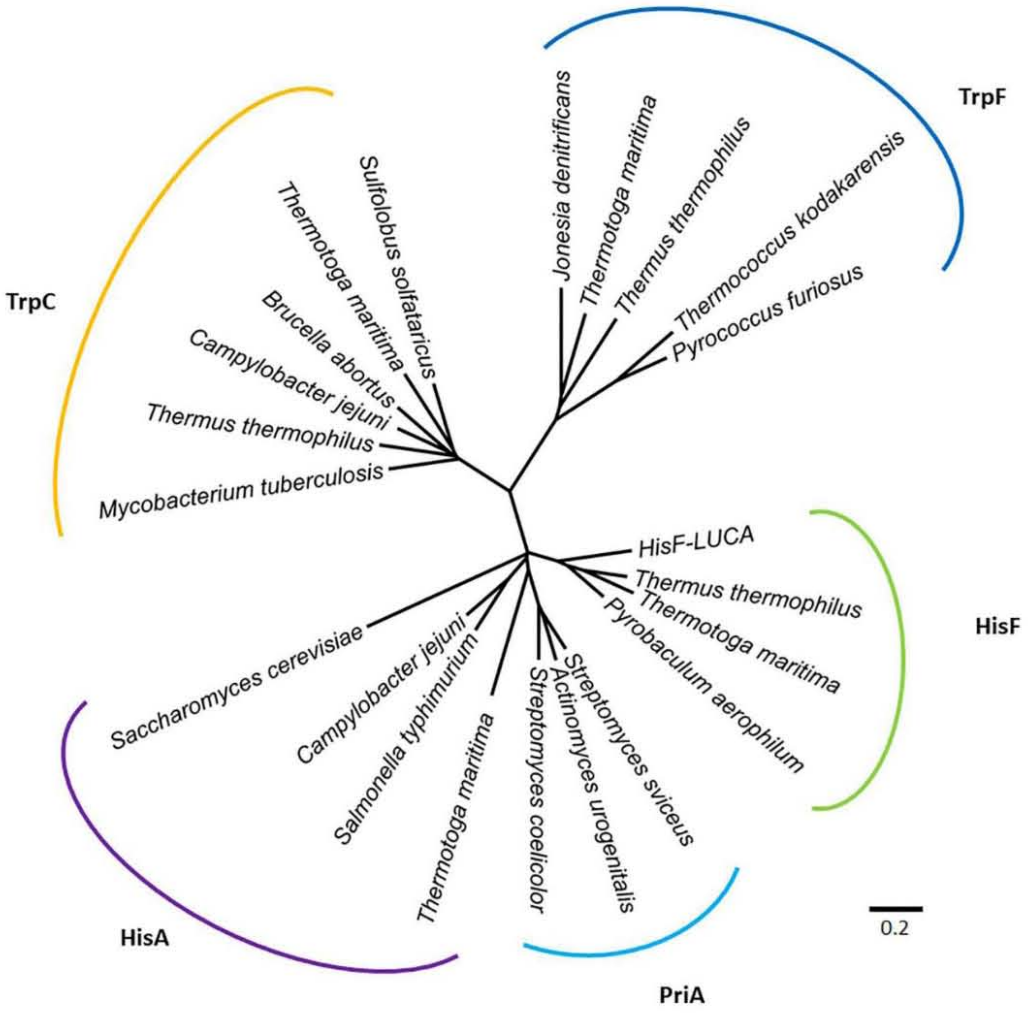

Fig 7. Structural phylogeny of HisF and its homologues. The unrooted tree shows the evolutionary relationships between the different HisF protein paralogs. Of particular interest is the position of the PriA branch, which suggests that this protein arose after diversification of HisA and TrpF. The tree was constructed from comparison of tertiary structures using the Fitch-Margoliash algorithm included in the PHYLIP package [52].

https://doi.org/10.1371/journal.pone.0196349.g007

PriA is a bisubstrate enzyme capable of catalyzing the isomeration, also via an Amadori rearrangement, of both ProFAR and PRAn into the corresponding aminoketoses. It was first described in *Streptomyces coelicolor* and *Mycobacterium tuberculosis* [60] and, as can be seen in the tree, its branch is projected from the HisA branch, suggesting that PriA evolved from HisA (Fig 7). These results confirm the antiquity of the HisA/HisF pair respect to PriA, and support previous conclusions of a close evolutionary relationship between PriA and HisA [61].

On the other hand, PurM is part of the PurM superfamily, and it is homologous to phosphoribosylformylglycinamidine synthase [EC: 6.3.5.3] (PurL), the enzyme which catalyzes the previous step in the *de novo* synthesis of purines. Both enzymes share structural domains and catalyze biochemical equivalent processes using a very similar ATP-dependent reaction mechanism [43]. The available evidence suggest that PurL is the outcome of a gene duplication followed by a fusion event [62], confirming the role of gene duplication events in the early evolution of metabolic pathways [63].

The search for other structural homologues of PurM identified several proteins of the PurM superfamily, such as the Ni-Fe hydrogenase maturation protein (HypE), selenophosphate synthetase [EC: 2.7.9.3] (SelD), and thiamin phosphate kinase [EC: 2.7.4.16] (ThiL). The evolutionary relationship between all these enzymes is depicted in Fig 8.

Thus, the HisA/HisF pair are homologous proteins that catalyze similar reactions using the same protein scaffold. The same occurs in the PurL/PurM pair. However, the Hisa/HisF and the PurL/PurM pairs are not homologous. In fact, it is well known that analogous catalytic functions can be present in two or more unrelated protein folds. In other words, similar or even identical catalytic functions can appear independently more than once during the evolutionary process, i.e. they are the outcome of convergent evolution [64,65]. As argued here, this is the case of the HisA/HisF and PurL/PurM pairs.

Active sites and reaction mechanisms of HisF and PurM

There are several types of convergent evolution, including enzymes with similar protein folds that catalyze the identical or similar reactions, as well as convergent evolution of different protein folds catalyzing chemically equivalent reactions on similar or identical substrates, i.e. functional convergent evolution [66].

One of the best-known examples of evolutionary functional convergence at a biochemical level is the case of the subtilisin-like and chymotrypsin-like families of serine proteases. These enzymes adopt different three-dimensional structures, but their active sites evolved independently and resulted in the same catalytic amino acids (the Ser-His-Asp triad), as well as a very similar catalytic mechanism for the hydrolysis of the peptide bond [67,68]. As described below, we have searched for this kind of convergence in the HisF and PurM enzymes.

Biochemical studies of HisF of *Thermotoga maritima* have shown that the essential residues for catalytic activity are Asp11 and Asp130 (Fig 4B). It has been proposed that the carboxylate groups of both aspartate residues are involved in general acid/base catalysis [69]. On the other hand, based on sequence conservation between several PurM enzymes, residues Asp94, His247, Thr249, Arg259 (Fig 4D) have been identified as key components of the active site for the PurM of *Escherichia coli*, where these residues may be involved in general acid and base catalysts, as well as in the coordination of metal ions [42]. As shown in Fig 6C, the use of the *align* command of PyMol to superimpose the active sites of HisF and PurM demonstrates that no spatial correspondence exists between them (Fig 6C). Our comparison reveals that the similarities between the two enzymes are limited to the structural correspondence between the Asp130 of HisF and the Asp94 of PurM and, to a certain extent, to the proximity between the Asp11 of HisF and Asp86 of PurM. However, the rest of the active site residues of PurM have no structural counterpart in the HisF structure. This implies that the architecture of the active sites of HisF and PurM is completely different, and thus it can be safely concluded that their biochemical activities are not due to the convergence of their catalytic sites.

Moreover, analysis of the reaction mechanisms proposed for each of the imidazole-synthesizing enzymes [42,69] demonstrates their profound differences, and allows to describe the specific details of the cyclization of the imidazole motif in each case. As noted above, the HisF

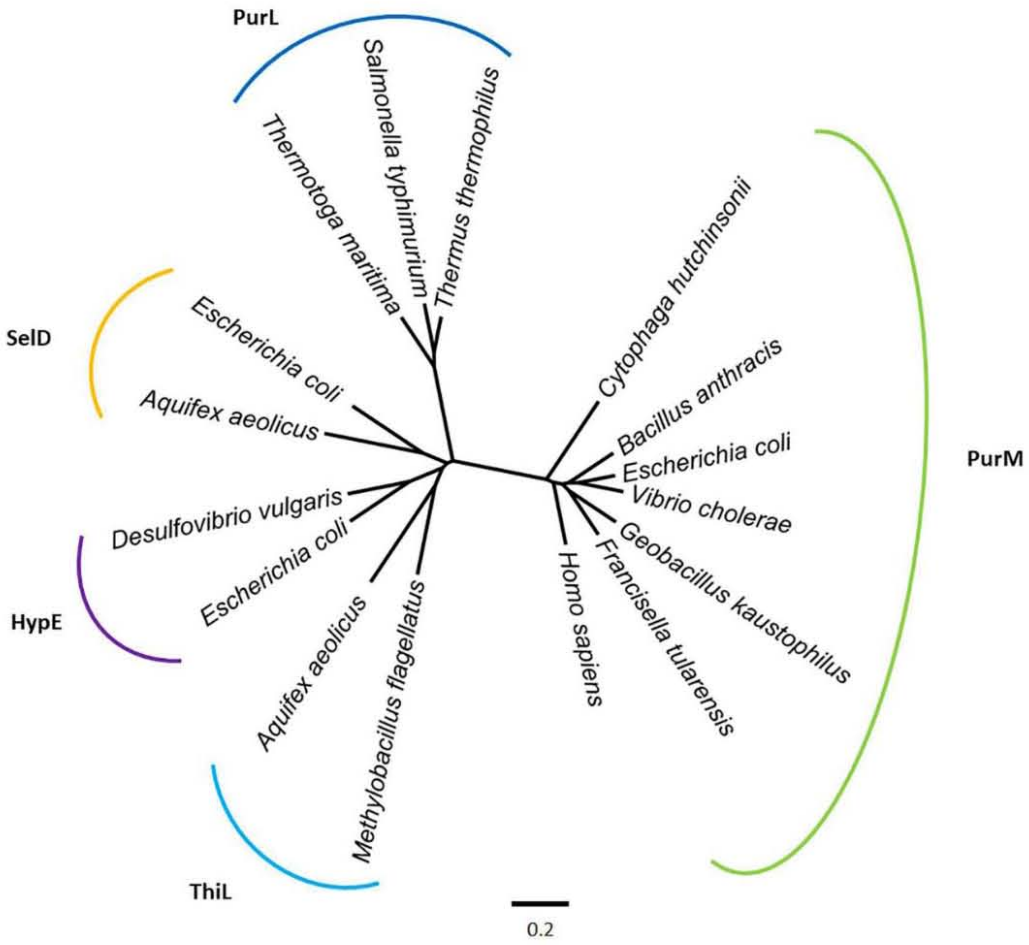

Fig 8. Structural phylogeny of PurM and its homologues. The unrooted tree shows the relationships among the different paralogs of the protein PurM. The tree was constructed from comparison of tertiary structures using the Fitch-Margoliash algorithm included in the PHYLIP package [52].

https://doi.org/10.1371/journal.pone.0196349.g008

enzyme catalyzes an Amadori rearrangement that is responsible for the formatiom of the imidazole ring, where the cyclization occurs after the hydrolysis of AICAR from the substrate PRFAR. This exposes a carbonyl group that can then react with the amino group of the same substrate (Fig 9). On the other hand, PurM uses one ATP molecule to activate a carbonyl group present in the FGAM structure, generating an iminophosphate intermediate. Immediately after this activation, an amino group makes a nucleophilic attack towards the previously-activated carbonyl group, leading to the cyclization of the imidazole ring. As depicted in Fig 9, the imidazole cyclization in both enzymes occurs by means of a nucleophilic attack of a nitrogen towards a carbonyl group. However, the overall mechanisms are completely different, with HisF catalyzing an Amadori rearrangement, and PurM catalyzing a strictly ATP-

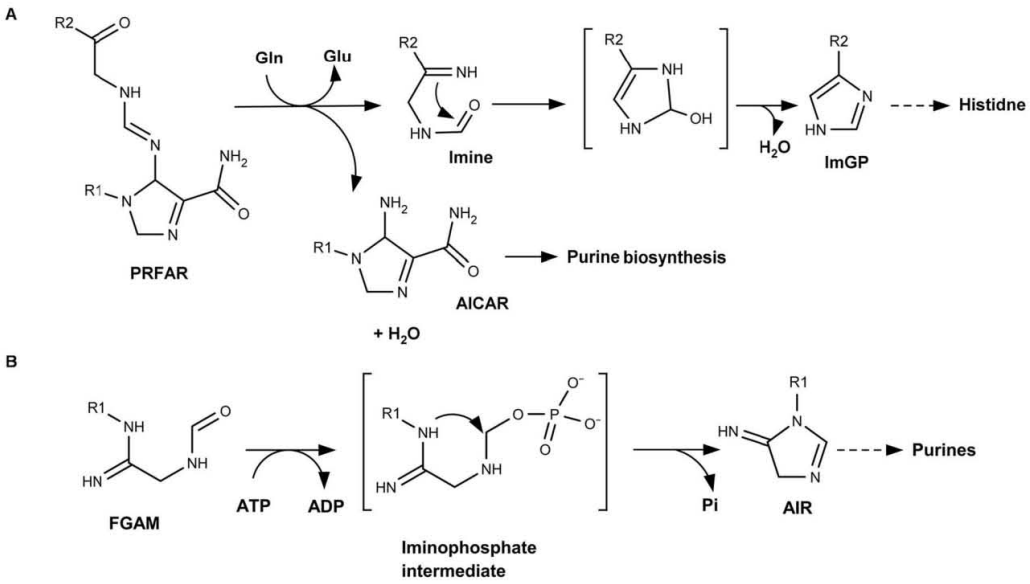

Fig 9. Proposed reaction schemes for HisF and PurM. The reaction schemes are based on [42, 69]. The nucleophilic attack of amino to the carbonyl group is depicted. R1: ribose phosphate; R2: triose phosphate.

https://doi.org/10.1371/journal.pone.0196349.g009

dependent reaction. In other words, the two enzymes are not homologous and the biosyntheses of the two imidazole groups proceed via different mechanisms.

Discussion

Evolutionary convergence in biochemistry

An example of biochemical redundancy is found in the biosyntheses of lysine, where three different anabolic pathways have been reported, i.e., the diaminopimelic acid pathway (DAP), the α-aminoadipic acid route (AAA) [70,71], and the recently-discovered TK0280-TK0283 variant found in *Thermococcus kodakarensis* [72].

Evolutionary functional convergence has also been described in the families of kinases that catalyze the phosphorylation of sugars, specifically in the hexokinase, ribokinase and galactokinase families, where two types of convergence can be recognized [66]. In the first case, a common trait originated independently several times in the same protein fold, such as the specificity to phosphorylate glucose, which clearly appeared more than once within the hexokinase family. In the second case, different protein structures convergently evolve to the same catalytic function, as occurred with the use of fructose as a substrate, which appeared independently in both the ribokinase and the hexokinase enzyme families [66]. The case of the imidazole-synthesizing HisF and PurM enzymes discussed here not only represents another example of this second kind of convergent evolution, but also underlines the role of redundancy in the biochemical syntheses of the imidazole group. As summarized below, this is best understood in terms of the extraordinary biological significance of this chemical moiety.

Evolution of two different imidazole-synthesizing enzymatic systems

As described above, histidine biosynthesis starts with ATP. This ribonucleotide is condensed with PRPP to form PRATP (Fig 3), which is then enzymatically processed to form AICAR. As underlined above, rather surprisingly the imidazole moiety in AICAR (which is derived from ATP) is not used to form histidine. Instead, a new imidazole moiety is biosynthesized and becomes part of histidine. The obvious question is why the available imidazole moiety already present in AICAR is not used for such task. As explained below, we suggest that the answer is found in the differences between the enhanced catalytic properties of the imidazole moiety of histidine as compared to that of purines.

Adenine has a well-known ability to coordinate a series of metal ions, including Cu^{2+} and Ag^{+} [73]. It has been shown that artificially constructed adenine-containing matrices can coordinate such metals and exhibit catalytic activities, including hydrolysis of ester substrates such as *p*-nitrophenyl phosphate (*p*NPP), bis-(*p*-nitrophenyl) phosphate (bNPP) and 2-hydroxypropyl-*p*-nitrophenyl phosphate (hNPP), as well as the hydrolysis of 2'-3'-cAMP [73]. Metal binding is equally significant for the catalytic capabilities of RNA, together with the highly reactive ribose 2'-hydroxyl group and the tertiary structure that these molecules adopt. Although it has been demonstrated that the catalytic activity of the ribosome is dependent on an adenine base [74,75], the reaction mechanism involves a proton abstraction by the N3 of the adenine with no direct participation of the imidazole ring, which, as of today, has not been shown to participate in any of the known catalytic activities of RNA.

These results are explained by the fact that in both N6 ribosyl adenine and histidine the imidazole rings remain unbound, which leaves them free to react [76,77]. Nevertheless, this could change if the covalent bond that joins the ribose with the base is changed from position N9 to the position N6. In fact, as demonstrated by Maurel and Ninio [76], N6 ribosyl adenine can accelerate the hydrolysis of *p*-nitrophenyl acetate (PNPA) in solution at pH 7.7, a reaction that neither adenosine nor AMP can catalyze. The rate at which N6 ribosyl adenine accelerates the reaction is comparable to that observed when histidine is used as catalyst.

In sharp contrast with histidine, the β-glycosidic bond between the N9 of the imidazole ring in purine nucleotides and the C1' of the ribose blocks the otherwise acidic N9 atom, completely inhibiting the catalytic capabilities of its imidazole moiety. In other words, if a molecule such as N6-ribosyl adenine would be incorporated into RNA, we would expect that its imidazole ring would exhibit catalytic activities. This possibility is amenable to experimental analysis.

It is worth noting that in sharp contrast with pyrimidines, in the *de novo* synthesis of purines this β-N9-glycosidic bond is formed in the very first step of the anabolic pathway, and along the entire anabolic process the purine biosynthetic intermediates are ribotide derivatives, constraining the catalytic capabilities of the imidazole moiety from the start of the route.

We surmise that these structural changes in the imidazole moieties account for the major differences in the catalytic properties between histidine and adenine, and were probably the critical issue in the emergence of histidine biosynthesis, as well as an important selection pressure for the evolution of two non-homologous routes for the synthesis of imidazole group.

Evolution of histidine and purine biosynthesis

The phylogenetic distribution of the enzymes involved in histidine and purine *de novo* biosyntheses shows that both pathways are highly conserved through the three domains of life, suggesting that they originated and were well established before the diversification of the Last Universal Common Ancestor (LUCA) into major kingdoms [78]. Taking into consideration the functional convergence of the enzymes that catalyze imidazole syntheses, the two pathways

must have evolved independently during very early stages of protein evolution, using the same pentose derivative as precursor, and later on became connected through AICAR, which as underlined above can be considered as a mere intracellular salvage step. Given the key role of purines in RNA properties, including catalysis, it is reasonable to conclude that purine biosynthesis appeared before histidine anabolism.

Sequence analysis of the enzymes that catalyze the different steps in the pathways provide information of their evolution. Considering the number of paralogous genes that both HisF and PurM have within their respective routes (Fig 3), it is reasonable to hypothesize that the present enzymatic activities of these two proteins evolved from ancestral enzymes with less specificity for substrates. For instance, in addition to the PurL/PurM pair, the archeal *purD*, *purT*, *purK* and *purP* genes are also homologous [43]. This suggests the existence of at least two ancestral enzymes with relaxed substrate specificity, that could catalyze all the corresponding activities of the modern paralogous enzymes. This possibility is certainly consistent with the idea of a patchwork assembly of the routes [79]. In histidine biosynthesis, the experimental evidence indicating that the synthesis of IGP can take place, both *in vivo* and *in vitro*, in the absence of the HisH protein [40], as well as the likely existence of an ancestral enzyme with both HisA and HisF activities, suggest that primitive histidine biosynthesis was simpler in the early stages of its evolution, during which spontaneous transformations coexisted with enzymes with low substrate specificity catalyzing similar chemical reactions [41]. As summarized in Fig 3, a simpler histidine anabolism can thus be envisioned.

AICAR has an evolutionary adaptive value by its own

The manifold roles of AICAR in biochemical process go beyond the proposed salvage pathway connecting the histidine and the *de novo* synthesis of purines. For example, it has been shown that AICAR accumulates in *purH* mutants and inhibits the conversion of AIR to the 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) moiety of thiamine [80], and that can also decrease total coenzyme A pools inhibiting the activity of the enzyme pantoate-alanine ligase in *Salmonella enterica* [81]. Perhaps more significantly, both AICAR and its triphosphorylated form (5-amino 4-imidazole carboxamide riboside 5'-triphosphate, also known as ZTP) can act by themselves as alarmones in C-1-folate deficiency [82]. AICAR accumulation also leads to histidine auxotrophy and even growth arrest in yeast [83]. AICAR is also the precursor for the antibiotic bredinin (5-hydroxyl-1-beta-D-ribofuranosyl-1H-imidazole-4-carboxamide), which is part of the naturally occurring family of nucleotide antibiotics first isolated from *Eupenicillium brefeldianum* M2166 [84]. This underlines the significance of AICAR, which during evolution has acquired other functions beyond that of an intermediate in purine biosynthesis (Fig 10). This may be seen as a case of exaptation, in which evolution selected AICAR into other roles, with that of alarmone being among the earliest. The fact that both AICAR and ZTP can bind to functional RNA molecules [85], and that 4-amino-5-imidazole carboxamide (AICA) can be synthesized under prebiotic conditions [86,87] may also speak of the antiquity of this molecule.

Other imidazole syntheses in the cell

In addition to the histidine and purines pathways there are other imidazole-synthesizing routes present in the extant cells, which also underline the key role of imidazole and imidazole-derivatives. All the biological and abiotic forming mechanisms of imidazolides represent a case of redundancy for the synthesis of this chemical group and highlight its catalytic significance.

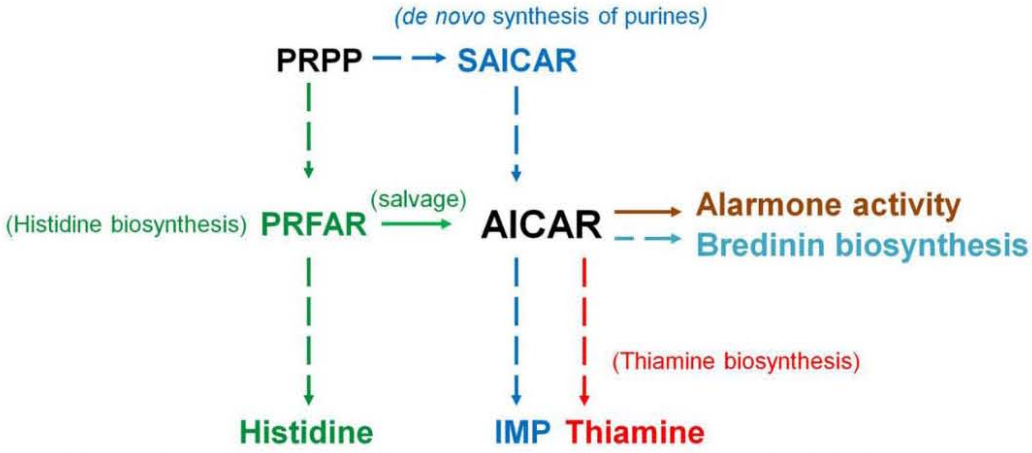

Fig 10. Biological significance of a metabolic intermediary. Scheme showing the roles of the imidazole intermediary AICAR in extant metabolism. From the salvage biosynthesis pathway from histidine to the *de novo* synthesis of purines.

https://doi.org/10.1371/journal.pone.0196349.g010

Some examples include the anabolic routes of (i) the protein cofactor biotin; (ii) the molecule creatinine; and (iii) the autocatalytically-formed prosthetic group 4-methylideneimidazole-5-one (MIO). Biotin is a vitaminic cofactor that transfers CO_2 as well as C-2 units, and its imidazole moiety is formed enzymatically by action of the dethiobiotin synthetase [EC:6.3.3.3] (BioD) enzyme. Quite significantly, both creatinine and MIO are imidazole molecules formed non-enzymatically. Creatinine is formed via the spontaneous cyclization of creatine-phosphate [88]. On the other hand, MIO participates as a prosthetic group in the histidine-, tyrosine- and phenylalanine ammonia-lyases (HAL, TAL and PAL, respectively) enzymes, that catalyze the α, β-elimination of ammonia from the corresponding amino acid using MIO as an electrophile, and are structurally homologous to the tyrosine and phenylalanine amininomutases (TAM and PAM), which catalyze the interconversion of a β-amino acid from an α-amino acid using the same MIO cofactor. This imidazole cofactor is formed by condensation of a highly conserved alanine-serine-glycine motif located within the active site of the enzymes and is used as an electrophile [89]. The reaction starts with the attack of the Gly amine to the carbonyl group of Ala, generating a five-membered intermediate, which is then transformed into the MIO cofactor (Fig 11) [90]. Given the well-established fact that Ala, Ser and Gly are three of

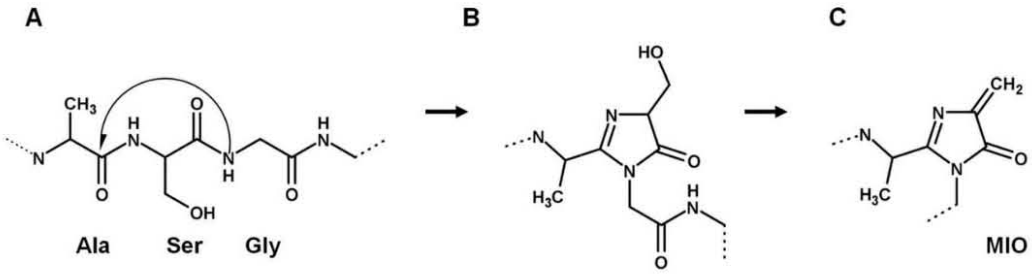

Fig 11. Mechanism of MIO formation. MIO formation via cyclization of an Ala-Ser-Gly- motif. (Modified from ref. [89]). Chemical structures were drawn using MarvinSketch v16.1.4 ChemAxon [94].

https://doi.org/10.1371/journal.pone.0196349.g011

the most abundant amino acids found in prebiotic experiments [91,92] as well as in carbonaceous chondrites [93] a possible model prebiotic experiment using the Ala-Ser-Gly in order to obtain an imidazolide has recently been proposed [17].

Conclusions

In contrast to other cases of convergence in biosynthetic pathways such as those reported for polyphyletic caffeine biosyntheses [65], in the different routes for imidazole formation discussed here the final products are not the outcome of enzyme recruitment via a patchwork mechanism, but rather examples of deterministic chemical processes that lead via different processes to the same functional product.

Like all other biological traits, the properties of biochemical pathways have been shaped by their evolutionary history. History, however, is not a mere sequence of chance events. Natural selection can overcome contingent effects to an extent, and chemical and biochemical constraints may have had a greater significance than is normally recognized in shaping metabolic traits. This appears to be the case of the manifold ways in which imidazole moieties are formed in extant cells.

As reviewed here, analysis of metabolic processes and biochemical properties demonstrate that several independent mechanisms of imidazole biosyntheses have evolved. The significance of this chemical group for living organisms is evident as their metabolism became redundant in respect to their syntheses, although the specific reaction mechanisms of each pathway are clearly different. In the purine and histidine biosynthesis, HisF and PurM have different sequences and unrelated three-dimensional structures; however, the two enzymes catalyze analogous reactions that lead to the formation of the imidazole group, representing a case of convergence at a biochemical level. This speaks for the catalytical and structural significance of imidazole and its derivatives.

The two-different enzyme-dependent syntheses of histidine and purines, catalyzed by PurM and HisF, together with the autocatalytic formation of creatinine (from creatine-phosphate) and MIO (from the Ala-Ser-Gly tripeptide) are unique examples of biochemical redundancy that underline for the significance of imidazole derivatives in biological processes. As argued here, the polyphyletic origin of PurM and the HisF is best understood by the enhanced catalytic activity of the histidine imidazole group compared to that of purines, where the β-glycosidic bond between the ribose and the imidazole ring of purines is established from the very start of IMP biosynthesis and results in the inhibition of purine catalytic activities.

The formation of the imidazole motifs in purines and histidine proceeds by means of a nucleophilic attact of a nitrogen atom in an amino group to a carbon atom in the substrate. However, as detailed here, the corresponding imidazole-synthesizing enzymes have different amino acid sequences and different three-dimensional folds, but catalyze chemically equivalent reactions. Thus, in spite of chemical restrictions, considerable degrees of freedom in the sequence space were available during the early stages of metabolic evolution.

Supporting information

S1 Fig. Pairwise alignment of bacterial (*Escherichia coli*) HisF and PurM sequences. The corresponding HisF and PurM sequences of *E. coli* were aligned using the Needleman-Wunsch algorithm as implemented in the EMBOSS Needle program of the European Molecular Biology Open Software Suite [45]. The aligment shows 58.5% of gaps with only 12.2% of sequence identity and 21.4% of sequence similarity.
(TIF)

S1 Table. Crystallographic structures used for the analysis. Complete list of the crystallographic structures and their PDB codes used to construct the dendograms.
(DOCX)

Acknowledgments

Alberto Vázquez-Salazar is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship 404507 from CONACYT. Financial support from PAPIIT-UNAM (IN223916) is gratefully acknowledged. We are grateful to Ada Yonath, Marie-Christine Maurel, and Renato Fani for many conversations over the years on the origin of histidine biosynthesis, and thank Ricardo Hernandez-Morales for help with the manuscript.

Author Contributions

Conceptualization: Alberto Vázquez-Salazar, Arturo Becerra, Antonio Lazcano.

Investigation: Alberto Vázquez-Salazar, Arturo Becerra, Antonio Lazcano.

Writing – review & editing: Alberto Vázquez-Salazar, Arturo Becerra, Antonio Lazcano.

References

1. Sundberg RJ, Martin RB. Interactions of histidine and other imidazole derivatives with transition metal ions in chemical and biological system. Chem Rev. 1974; 74, 4:471–517.
2. Edgcomb SP, Murphy KP. Variability in the pKa of histidine side-chains correlates with burial within proteins. Proteins. 2002; 1; 49(1):1–6. https://doi.org/10.1002/prot.10177 PMID: 12211010
3. Grimsley GR, Scholtz JM, Pace CN. A summary of the measured pK values of the ionizable groups in folded proteins. Protein Sci. 2009; 18(1):247–251. https://doi.org/10.1002/pro.19 PMID: 19177368
4. Holliday GL, Fischer JD, Mitchell JB, Thornton JM. Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis. FEBS J. 2011; 278 (20):3835–45. https://doi.org/10.1111/j.1742-4658.2011.08190.x PMID: 21605342
5. Maley LE, Mellor DP. Stability of some metal complexes of histidine. Nature. 1950; 165: 4194: 453.
6. Andreini C, Bertini I, Cavallaro G, Holliday GL, Thornton JM. Metal ions in biological catalysis: from enzyme databases to general principles. J Biol Inorg Chem. 2008; 13(8):1205–18. https://doi.org/10.1007/s00775-008-0404-5 PMID: 18604568
7. Oró J. Synthesis of adenine from ammonium cyanide. Biochem Biophys Res Commun. 1960; 2:407–412.
8. Oró J. Synthesis of organic compounds by electric discharges. Nature. 1963; 197:862–67.
9. Lowe CU, Ress A, Markham FRS. Synthesis of complex organic compounds from simple precursors: Formation of amino-acids, amino-acid polymer, fatty acids and purines from ammonium cyanide. Nature. 1963; 19:219–222.
10. Schwartz AW, Bakker CG. Was adenine the first purine? Science. 1989; 245:1102–1104. PMID: 11538344
11. Borquez E, Cleaves HJ, Lazcano A, Miller SL. An investigation of prebiotic purine synthesis from the hydrolysis of HCN polymers. Orig Life Evol Biosph. 2005; 35(2):79–90. PMID: 16010991
12. Oró J, Basile B, Cortes S, Shen C, Yamrom T. The prebiotic synthesis and catalytic role of imidazoles and other condensing agents. Orig Life. 1984; 14(1–4):237–42. PMID: 6462668
13. Shen C, Yang L, Miller SL, Oró J. Prebiotic syntheses of imidazole-4-acetaldehyde and histidine. Orig Life. 1987 17: 295–305.
14. Shen C, Yang L, Miller SL, Oró J. Prebiotic synthesis of histidine. J Mol Evol. 1990; 31:167–174. PMID: 11536478
15. Decker P, Schweer H, Pohlmann R. Bioids. X. Identification of formose sugars, presumable prebiotic metabolites using capillary gas chromatography/gas chromatography-mass spectrometry of *n*-butoxime trifluoroacetates on OV-225. Journal of Chromatography. 1982; 244:281–291.

16. Bada JL, Chalmers JH, Cleaves HJ. Is formamide a geochemically plausible prebiotic solvent? Phys Chem Chem Phys. 2016; 18:20085–90. https://doi.org/10.1039/c6cp03290g PMID: 27253848

17. Vázquez-Salazar A, Tan G, Stockton A, Fani R, Becerra A, Lazcano A. Can an imidazole be formed from an alanyl-seryl-glycine tripeptide under possible prebiotic conditions? Orig Life Evol Biosph. 2017; 47:345–354. https://doi.org/10.1007/s11084-016-9525-y PMID: 27771860

18. Álvarez-Carreño C, Becerra A, Lazcano A. Norvaline and norleucine may have been more abundant protein components during early stages of cell evolution. Orig Life Evol Biosph. 2013; 43:363–75. https://doi.org/10.1007/s11084-013-9344-3 PMID: 24013929

19. Raggi L, Bada JL, Lazcano A. On the lack of evolutionary continuity between prebiotic peptides and extant enzymes. Phys Chem Chem Phys. 2016; 18:20028–20032. https://doi.org/10.1039/c6cp00793g PMID: 27121024

20. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell. 1982; 31:147–157. PMID: 6297745

21. Guerrier-Takada C, Gardiner K, Marsh T, Pace T, Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell. 1983; 35:849–857. PMID: 6197186

22. Walton T, Szostak JW. A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. J Am Chem Soc. 2016; 138(36):11996–200222. https://doi.org/10.1021/jacs.6b07977 PMID: 27552367

23. Breslow R. Bifunctional acid-base catalysis by imidazole groups in enzyme mimics. J Mol Cat. 1994; 91 (2): 161–174.

24. Perrotta AT, Shih I, Been MD. Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. Science. 1999; 286 (5437):123–6. PMID: 10506560

25. Shen C, Lazcano A, Oró J. The enhancement activities of histidyl-histidine in some prebiotic reactions. J Mol Evol. 1990; 31:445–52. PMID: 11540924

26. Wieczorek R, Dörr M, Chotera A, Luisi PL, Monnard PA. Formation of RNA phosphodiester bond by histidine-containing dipeptides. Chembiochem. 2013; 14 (2):217–23. https://doi.org/10.1002/cbic.201200643 PMID: 23255284

27. Wieczorek R, Adamala K, Gasperi T, Polticelli F, Stano P. Small and random peptides: an unexplored reservoir of potentially functional primitive organocatalysts. The case of seryl-histidine. Life (Basel). 2017; 7(2). pii: E19.

28. Gorlero M, Wieczorek R, Adamala K, Giorgi A, Schininà ME, Stano P, et al. Ser-His catalyses the formation of peptides and PNAs. FEBS Lett. 2009; 583:153–6. https://doi.org/10.1016/j.febslet.2008.11.052 PMID: 19071124

29. Adamala K, Anella F, Wieczorek R, Stano P, Chiarabelli C, Luisi PL. Open questions in origin of life: experimental studies on the origin of nucleic acids and proteins with specific and functional sequences by a chemical synthetic biology approach. Comput Struct Biotechnol J. 2014; 9:e201402004. https://doi.org/10.5936/csbj.201402004 PMID: 24757502

30. Petrov AS, Gulen B, Norris AM, Kovacs NA, Bernier CR, Lanier KA, et al. History of the ribosome and the origin of translation. Proc Natl Acad Sci U S A. 2015; 112(50):15396–401. https://doi.org/10.1073/pnas.1509761112 PMID: 26621738

31. Kovacs NA, Petrov AS, Lanier KA, Williams LD. Frozen in time: the history of proteins. Mol Biol Evol. 2017; 34(5):1252–1260. https://doi.org/10.1093/molbev/msx086 PMID: 28201543

32. White HB. Coenzymes as fossils of an earlier metabolic state. J Mol Evol. 1976; 7:101–104. PMID: 1263263

33. White HB. Evolution of coenzymes and the origin of pyridine nuclectides. In: Everse J, Anderson B, You KS, editors. The pyridine nucleotide coenzymes. New York: Academic Press; 1982. pp. 1–17.

34. Akouche M, Jaber M, Maurel MC, Lambert JF, Georgelin T. Phosphoribosyl pyrophosphate: a molecular vestige of the origin of life on minerals. Angew Chem Int Ed Engl. 2017; 56(27):7920–7923. https://doi.org/10.1002/anie.201702633 PMID: 28558156

35. Alifano P, Fani R, Liu P, Lazcano A, Bazzicalupo M, Carlomagno MS, et al. Histidine biosynthetic pathway and genes: structure, regulation, and evolution. Microbiol Rev. 1996; 60:44–69. PMID: 8852895

36. Chaudhuri BN, Lange SC, Myers RS, Chittur SV, Davisson VJ, Smith JL. Crystal structure of imidazole glycerol phosphate synthase: a tunnel through a (beta/alpha)$_8$ barrel joins two active sites. Structure. 2001; (10):987–97. PMID: 11591353

37. Chaudhuri BN, Lange SC, Myers RS, Davisson VJ, Smith JL. Toward understanding the mechanism of the complex cyclization reaction catalyzed by imidazole glycerolphosphate synthase: crystal structures of a ternary complex and the free enzyme. Biochemistry. 2003; 42(23):7003–12. https://doi.org/10.1021/bi034320h PMID: 12795595

38. Rivalta I, Sultan MM, Lee NS, Manley GA, Loria JP, Batista VS. Allosteric pathways in imidazole glycerol phosphate synthase. Proc Natl Acad Sci U S A., 2012; 109 (22): E1428–E1436. https://doi.org/10.1073/pnas.1120536109 PMID: 22586084

39. Smith DW, Ames BN. Intermediates in the early steps of histidine biosynthesis. J Biol Chem. 1964; 239:1848–1855. PMID: 14213364

40. Rieder G, Merrick MJ, Castorph H, Kleiner D. Function of *hisF* and *hisH* gene products in histidine biosynthesis. J Biol Chem. 1994; 269 (20):14386–14390. PMID: 8182043

41. Lazcano A, Miller SL. On the origin of metabolic pathways. J Mol Evol. 1999; 49(4): 424–31 PMID: 10486000

42. Li C, Kappock TJ, Stubbe J, Weaver TM, Ealick SE. X-ray crystal structure of aminoimidazole ribonucelotide synthetase (PurM), from the *Escherichia coli* purine biosynthetic pathway at 2.5 Å resolution. Structure. 1999; 7:1155–1166. PMID: 10508786

43. Zhang Y, Morar M, Ealick SE. Structural biology of the purine biosynthetic pathway. Cell Mol Life Sci. 2008; 65 (23):3699–724. https://doi.org/10.1007/s00018-008-8295-8 PMID: 18712276

44. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. 2011; Nucleic Acids Res. 40(D1): D109–D114.

45. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. 2015; 43(W1):W580–W584.

46. The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC.

47. Prlic A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A, et al. Precalculated Protein Structure Alignments at the RCSB PDB website Bioinformatics. 2010; 26(23): 2983–2985. https://doi.org/10.1093/bioinformatics/btq572 PMID: 20937596

48. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010; 38(Web Server issue): W545–W549. https://doi.org/10.1093/nar/gkq366 PMID: 20457744

49. Jácome R, Becerra A, Ponce de León S, Lazcano A. Structural analysis of monomeric RNA-dependent polymerases: evolutionary and therapeutic implications. PLoS One. 2015; 10(9):e0139001. https://doi.org/10.1371/journal.pone.0139001 PMID: 26397100

50. Krissinel E, Henrick K. Secondary-structure matching (PDBeFold), a new tool for fast protein structure alignment in three dimensions. Acta Cryst. 2004; D60, 2256–2268.

51. Subbiah S, Laurents DV, Levitt M. Structural similarity of DNA-binding domains of bacteriophage repressors and the globin core. Curr Biol. 1993; 3:141–148. PMID: 15335781

52. Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle

53. FigTree v1.4.2. http://tree.bio.ed.ac.uk/software/figtree/

54. Fani R, Liò P, Chiarelli I, Bazzicalupo M. The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for the *hisA* and *hisF* genes. J Mol Evol. 1994; 5:489–95.

55. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. Science. 2000; 289(5484):1546–1550. PMID: 10968789

56. Henn-Sax M, Höcker B, Wilmanns M, Sterner R. Divergent evolution of (beta/alpha)8-barrel enzymes. Biol Chem. 2001; 382(9):1315–1320. https://doi.org/10.1515/BC.2001.163 PMID: 11688714

57. Henn-Sax M, Thoma R, Schmidt S, Hennig M, Kirschner K, Sterner R. Two (beta/alpha)(8)-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. Biochemistry. 2002; 41(40):12032–12042. PMID: 12356303

58. Friedmann N, Haverland WJ, Miller SL. Prebiotic synthesis of the aromatic and other amino acids. In: Buvet R, Ponnamperuma C, editors. Chemical evolution and the origin of life: Proceedings of the international conference on the origin of life. Amsterdam: North-Holland Publishing Company; 1971. pp 123–135.

59. Jukes TH. Molecules and evolution. New York: Columbia University Press; 1966.

60. Barona-Gómez F, Hodgson DA. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. EMBO Rep. 2003; 4(3):296–300. https://doi.org/10.1038/sj.embor.embor771 PMID: 12634849

61. Due AV, Kuper J, Geerlof A, von Kries JP, Wilmanns M. Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. Proc Natl Acad Sci U S A. 2011); 108(9):3554–3559. https://doi.org/10.1073/pnas.1015996108 PMID: 21321225

62. Suzuki S, Yanai H, Kanagawa M, Tamura S, Watanabe Y, Fuse K, et al. Structure of N-formylglycinamide ribonucleotide amidotransferase II (PurL) from *Thermus thermophilus* HB8. Acta Crystallogr Sect

F Struct Biol Cryst Commun. 2012; 68(Pt 1):14–9. https://doi.org/10.1107/S1744309111048184 PMID: 22232163

63. Becerra A, Lazcano A. The role of gene duplication in the evolution of purine nucleotide salvage pathways. Orig. Life Evol. Biosph. 1998; 28:539–553. PMID: 9742728
64. Todd AE, Orengo CA, Thornton JM. Evolution of protein function, from a structural perspective. Curr Opin Chem Biol. 1999 3:548–556. PMID: 10508675
65. Huang R, O'Donnell AJ, Barboline JJ, Barkman TJ. Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. Proc Natl Acad Sci U S A. 2016; 113(38):10613–10618. https://doi.org/10.1073/pnas.1602575113 PMID: 27638206
66. Bork P, Sander C, Valencia A. Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. Protein Sci. 1993; 2: 31–40. https://doi.org/10.1002/pro.5560020104 PMID: 8382990
67. Reardon D, Farber GK. The structure and evolution of alpha/beta barrel proteins. FASEB J. 1995; 7: 497–503.
68. Frey PA, Hegeman AD. Enzymatic reaction mechanisms. New York: Oxford University Press; 2007.
69. Beismann-Driemeyer S, Sterner R. Imidazole glycerol phosphate synthase from *Thermotoga maritima*. Quaternary structure, steady-state kinetics, and reaction mechanism of the bienzyme complex. J Biol Chem. 2001; 276(23): 20387–20396. https://doi.org/10.1074/jbc.M102012200 PMID: 11264293
70. Kobashi N, Nishiyama M, Tanokura M. Aspartate kinase-independent lysine synthesis in an extremely thermophilic bacterium, *Thermus thermophilus*: lysine is synthesized via alpha-aminoadipic acid not via diaminopimelic acid. J Bacteriol. 1999; 181(6):1713–1718. PMID: 10074061
71. Velasco AM, Leguina JI, Lazcano A. Molecular evolution of the lysine biosynthetic pathways. J Mol Evol. 2002; 55(4):445–459. https://doi.org/10.1007/s00239-002-2340-2 PMID: 12355264
72. Yoshida A, Tomita T, Atomi H, Kuzuyama T, Nishiyama M. Lysine biosynthesis of *Thermococcus kodakarensis* with the capacity to function as an ornithine biosynthetic system. J Biol Chem. 2016; 291(41): 21630–21643. https://doi.org/10.1074/jbc.M116.743021 PMID: 27566549
73. Verma S, Mishra AK, Kumar J. The many facets of adenine: coordination, crystal patterns, and catalysis. 2010; Acc Chem Res. 43(1): 79–91. https://doi.org/10.1021/ar9001334 PMID: 19719100
74. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. Science. 2000; 289(5481): 920–930. PMID: 10937990
75. Yonath A. Ribosomal crystallography: peptide bond formation, chaperone assistance and antibiotics activity. Mol Cells. 2005; 20(1): 1–16. PMID: 16258236
76. Maurel MC, Ninio J. Catalysis by a prebiotic nucleotide analog of histidine. Biochimie. 1987; 69:551–553. PMID: 2445388
77. Décout JL, Maurel MC. N6-substituted adenine derivatives and RNA primitive catalysts. Orig Life Evol Biosph. 1993; 23: 299–306. PMID: 7509474
78. Delaye L, Becerra A, Lazcano A. The last common ancestor: what's in a name? Orig Life Evol Biosph. 2005; 35(6): 537–554. https://doi.org/10.1007/s11084-005-5760-3 PMID: 16254691
79. Jensen RA. Enzyme recruitment in evolution of new function. Ann. Rev. Microbiol. 1976; 30: 409–425.
80. Allen S, Zilles JL, Downs DM. Metabolic flux in both the purine mononucleotide and histidine biosynthetic pathways can influence synthesis of the hydroxymethyl pyrimidine moiety of thiamine in *Salmonella enterica*. J Bacteriol. 2002; 184(22): 6130–6137. https://doi.org/10.1128/JB.184.22.6130-6137.2002 PMID: 12399482
81. Bazurto JV, Downs DM. Amino-4-imidazolecarboxamide ribotide directly inhibits coenzyme A biosynthesis in *Salmonella enterica*. J Bacteriol. 2014; 196(4): 772–779. https://doi.org/10.1128/JB.01087-13 PMID: 24296672
82. Bochner BR, Ames BN. ZTP (5-amino 4-imidazole carboxamide riboside 5'-triphosphate): a proposed alarmone for 10-formyl-tetrahydrofolate deficiency. Cell. 1982; 29(3): 929–937. PMID: 6185232
83. Daignan-Fornier B, Pinson B. 5-Aminoimidazole-4-carboxamide-1-beta-D-ribofuranosyl 5'-monophosphate (AICAR), a highly conserved purine intermediate with multiple effects. Metabolites. 2012; 2(2): 292–302. https://doi.org/10.3390/metabo2020292 PMID: 24957512
84. Suhadolnik RJ. Naturally occurring nucleoside and nucleotide antibiotics. Prog Nucleic Acid Res Mol Biol. 1979; 22: 193–291. PMID: 230535
85. Jones CP, Ferré-D'Amaré AR. Recognition of the bacterial alarmone ZMP through long-distance association of two RNA subdomains. Nat Struct Mol Biol. 2015; (9): 679–685. https://doi.org/10.1038/nsmb.3073 PMID: 26280533
86. Oró J, Kimball AP. Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide. Arch Biochem Biophys. 1961; 94: 217–227. PMID: 13731263

87. Joyce GF. RNA evolution and the origins of life. Nature. 1989; 338: 217–224. https://doi.org/10.1038/338217a0 PMID: 2466202

88. Borsook H, Dubnoff JW. The hydrolysis of phosphocreatine and the origin of urinary creatinine. J Biol Chem. 1947; 168(2): 493–510. PMID: 20238607

89. Cooke HA, Chistianson CV, Bruner SD. Structure and chemistry of 4-methylideneimidazole-5-one containing enzymes. Curr Opin Chem Biol. 2009; 4: 460–468.

90. Rétey J. Discovery and role of methylidene imidazolone, a highly electrophilic prosthetic group. Biochim Biophys Acta. 2003; 1647: 179–184. PMID: 12686130

91. Miller SL. A production of amino acids under possible primitive earth conditions. Science. 1953; 117: 528–529. PMID: 13056598

92. Cleaves HJ, Chalmers JH, Lazcano A, Miller SL, Bada JL. A reassessment of prebiotic organic synthesis in neutral planetary atmospheres. Orig Life Evol Biosph. 2008; 38(2): 105–115. https://doi.org/10.1007/s11084-007-9120-3 PMID: 18204914

93. Burton AS, Stern JC, Elsila JE, Glavin DP, Dworkin JP. Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. Chem Soc Rev. 2012; 41: 5459–5472. https://doi.org/10.1039/c2cs35109a PMID: 22706603

94. MarvinSketch v16.1.4 ChemAxon. https://chemaxon.com/products/marvin-js

Presentación artículo: **Can an imidazole be formed from an alanyl-seryl-glycine tripeptide under possible prebiotic conditions?**

Las enzimas histidina-, tirosina- y fenilalanina-amonioliasas (HAL, TAL y PAL), así como las fenilalanina- y tirosina-aminomutasas (PAM y TAM) son enzimas homólogas con un alto grado de conservación estructural. Todas ellas comparten la característica de usar el grupo prostético 4-metilideno-imidazol-ona (MIO) como un electrófilo para llevar a cabo sus actividades catalíticas. Este cofactor se forma de manera autocatalítica a partir de un motivo de Ala-Ser-Gly que se encuentra conservado en la secuencia del sitio activo de dichas enzimas. Tal característica hace de este cofactor orgánico un objeto de estudio interesante desde el punto de vista bioquímico y, como se discutirá aquí, evolutivo.

La caracterización bioquímica de la histidina-amonioliasa (HAL) (Givot, et al., 1969) y de la fenilalanina-amonioliasa (PAL) (Hodgins, 1971) llevó al reconocimiento de que un componente electrofílico era esencial para su actividad catalítica. Durante varios años se pensó que dicho componente catalítico era una deshidroalanina. Sin embargo, fue hasta la década de 1990 cuando la primera estructura de una HAL (*Pseudomonas putida*) fue resuelta por cristalografía de rayos-X, y que el grupo prostético MIO fue descrito por primera vez (Schwede, et al., 1999). A partir de su descripción en HAL y PAL, MIO también ha sido encontrado en la estructura de la tirosina amonioliasa (TAL) (Louie, et al., 2006) y de las tirosina- y fenilalanina aminomutasas (TAM y PAM, respectivamente) (Walker, et al., 2004; Feng, et al., 2011; Christenson, et al., 2003).

Las enzimas que emplean al cofactor MIO forman un subgrupo dentro de las familias de las amonioliasas y aminomutasas, pues dicha molécula imdazólica ha sido descrita como inusual en la biología (Cooke, et al., 2009). En ambos casos MIO es empleado como un grupo electrófilo.

Desde que la estructura química de MIO fue descrita por primera vez (Schwede, et al., 1999) se propuso también su mecanismo de biosíntesis. Éste consiste en un proceso autocatalítico que involucra un motivo altamente conservado

en la secuencia de alanina (Ala), serina (Ser) y glicina (Gly). La reacción comienza con el ataque del amino de Gly al carbonilo de Ala, generando un intermedio de cinco miembros, que se transforma en el cofactor MIO (Rétey, 2003). Por ello, MIO ha sido catalogado como un cofactor derivado de un polipéptido, y corresponde a un caso especial de grupos protésicos (Fisher, et al., 2010). Como se detalla a continuación, aquí subrayamos la importancia experimental de este mecanismo de reacción para un posible experimento modelo en condiciones prebióticas.

Se han llevado a cabo numerosos experimentos que simulan la síntesis prebiótica de compuestos orgánicos bajo posibles condiciones de Tierra primitiva. En ellos se han empleado combinaciones de gases que van desde el altamente reductor (CH_4, NH_3, H_2O y H_2) (Miller, 1953) hasta neutro (H_2O, N_2, CO_2) (Cleaves, et al., 2008). Tres de los productos que siempre se identifican en este tipo de experimentos son alanina, serina y glicina, lo que sugiere fuertemente su presencia en la Tierra primitiva. La caracterización química de estos aminoácidos en condritas carbonosas también es consistente con su presencia en el inventario prebiótico de moléculas orgánicas (Burton, et al., 2012).

Los aminoácidos que dan origen al cofactor MIO son precisamente Ala, Ser y Gly. De este modo, la posible condensación espontánea de un péptido pequeño con un motivo de Ala-Ser-Gly en su secuencia pudo haber representado una fuente extra de moléculas imidazólicas para la vida temprana.

La principal crítica para esta propuesta es la ausencia de un mecanismo de codificación primitivo aunado a una baja probabilidad de formación de cantidades significativas de un tripéptido específico. Sin embargo, se sabe que la ciclación intramolecular de péptidos es favorecida por altas diluciones y que las ciclaciones tienden a ser reacciones inherentemente lentas (Davies, 2003). De esta forma, si se supone un proceso abiótico que conduzca a la síntesis de aminoácidos y péptidos pequeños que incluyan secuencias tales como el motivo Ala-Ser-Gly, un mecanismo de formación, semejante a la biosíntesis de MIO, de una molécula imidazólica puede ser plausible dadas las condiciones adecuadas. Esto puede incluir reacciones de condensación entre aminoácidos en lagunas y otros cuerpos

de agua sometidos a ciclos húmedos y secos, así como la interacción con arcillas, cationes metálicos e incluso otras moléculas orgánicas.

Orig Life Evol Biosph (2017) 47:345–354
DOI 10.1007/s11084-016-9525-y

HYPOTHESIS

Can an Imidazole Be Formed from an Alanyl-Seryl-Glycine Tripeptide under Possible Prebiotic Conditions?

Alberto Vázquez-Salazar[1] · George Tan[2] · Amanda Stockton[2] · Renato Fani[3] · Arturo Becerra[1] · Antonio Lazcano[1,4]

Received: 13 September 2016 / Accepted: 12 October 2016 /
Published online: 22 October 2016

Abstract The five-membered heterocyclic imidazole group, which is an essential component of purines, histidine and many cofactors, has been abiotically synthesized in different model experiments that attempt to simulate the prebiotic environment. The evolutionary significance of imidazoles is highlighted not only by its presence in nucleic acid components and in histidine, but also by experimental reports of its ability to restore the catalytic activity of ribozymes. However, as of today there are no reports of histidine in carbonaceous chondrites, and although the abiotic synthesis of His reported by Shen et al. (1987, 1990a) proceeds via an Amadori rearrangement, like in the biosynthesis of histidine, neither the reactants nor the conditions are truly prebiotic. Based on the autocatalytic biosynthesis of 4-methylidene-imidazole-one (MIO), a cofactor of some members of the amino acid aromatic ammonia-lyases and aminomutases, which occur via the self-condensation of a simple Ala-Ser-Gly motif within the sequence of the enzymes, we propose a possible prebiotic synthesis of an imidazolide.

Keywords Imidazole group · Prebiotic evolution · Prebiotic synthesis of imidazolides · Evolution of catalysis

Paper dedicated to the memory of Jim Ferris.

✉ Antonio Lazcano
alar@ciencias.unam.mx

[1] Facultad de Ciencias, Universidad Nacional Autónoma de México, Apdo. Postal 70-407, Cd. Universitaria, 04510 Mexico City, Mexico

[2] School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30309, USA

[3] Department of Biology, University of Florence, Via Madonna del Piano 6, I-50019 Sesto F. no, Florence, Italy

[4] Miembro de El Colegio Nacional, Mexico City, Mexico

investigada experimentalmente aún.

Introduction

A proper description of the origin and early evolution of life requires the understanding of the evolution of biological catalysis. During the very early stages of evolution, i.e. the RNA world, primitive biological systems could have depended not only on ribozymes but also on a number of simple catalysts including simple peptides, metallic ions, clays, and organic molecules that may have helped to catalyze the chemical reactions required to maintain their metabolism and replication. The roles played by the imidazole group and its derivatives in these processes, to which our friend and colleague James P. Ferris and his associates devoted considerable attention, are addressed here.

The imidazole ring is a five-membered heterocyclic molecule containing a tertiary nitrogen and an imino group. It is present in the structure of many different molecules and intermediates of biological significance, including purines, many cofactors, and histidine (Fig. 1). Imidazole derivatives are of major evolutionary interest, particularly during the early stages of biological evolution. The synthesis of a wide array of imidazole compounds, including some that are present in extant metabolism, has been reported in several prebiotic chemistry experiments (Table 1). The fact that imidazole derivatives (a) are formed in the prebiotic syntheses of purines and other molecules; (b) that these type of compounds interact with RNA molecules; and (c) the manifold roles that imidazole molecules play in extant metabolism, strongly suggest a case of evolutionary continuity between prebiotic chemistry and the emergence of the biosphere, and raises the question on the role of this chemical group during the origin and early evolution of life.

A classic prebiotic synthesis producing imidazole derivatives is that of adenine, the first abiotically obtained purine (Oró 1960). Oró used an aqueous solution of ammonium cyanide (NH_4CN), in which the condensation of hydrogen cyanide (HCN) molecules yielded 4-aminoimidazole-5-carboxamide, and 4-aminoimidazole-5-carboxamidine, which in turn reacted with formamidine to produce adenine (Oró 1960; Oró and Kimball 1961, 1962). In a surprising experiment, Ferris and Orgel (1966) demonstrated the formation of 4-aminoimidazole-5-

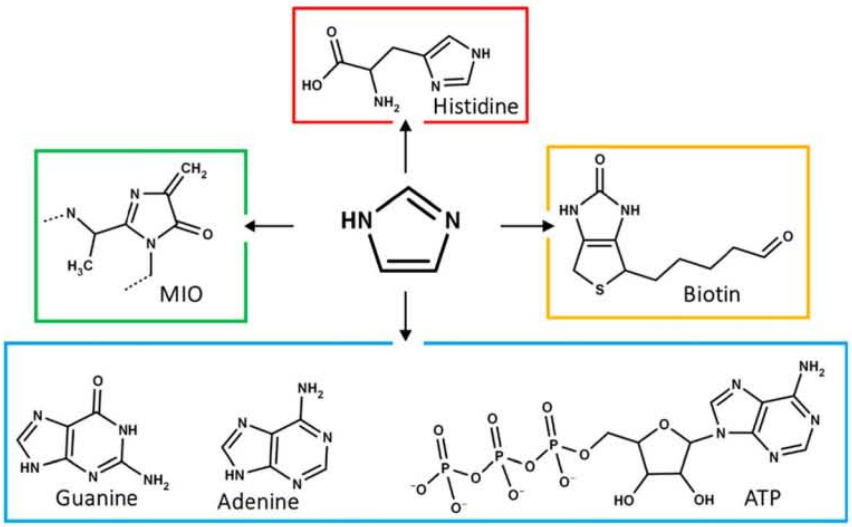

Fig. 1 Imidazole compounds in contemporary metabolism. The chemical structures of some relevant imidazole derivatives are shown. Histidine (2-amino-3-(1H-imidazol-4-yl) propanoic acid), biotin (5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanoic acid), ATP (adenosine triphosphate), adenine (9H–purin-6-amine), guanine (2-amino-9H-purin-6(1H)-one), and the MIO cofactor (4-methylidene-imidazole-one). Chemical structures were drawn using MarvinSketch v16.1.4 ChemAxon (http://www.chemaxon.com)

Table 1 Abiotically-synthesized imidazole compounds

Precursos	Compound	Reference
Hydrogen cyanide (HCN)	Adenine	Oró 1960
	4-aminoimidazole-5-carboxamide	Oró and Kimball 1962
	4-aminimidazole-5-carboxamidine	Oró and Kimball 1962
	Guanine	Oró and Kimball 1962
	4-aminoimidazole-5-carbonitrile	Ferris and Orgel 1966
	Xanthine	Oró 1963
	Hipoxanthine	Lowe et al. 1963
	2,6-diaminopurine	Borquez et al. 2005
	8-hydroxymethyladenine	Schwartz and Bakker 1989
Glyoxal, formaldehyde, ammonia	Imidazole	Oró et al. 1984
Glyoxal, acetaldehyde, ammonia	2-methylimidazole	
D-erythrose, formamidine	Histidine	Shen et al. 1987, 1990a

carbonitrile by the photochemical rearrangement of β-amino acrylonitrites in the synthesis of adenine. It has also been demonstrated that the hydrolysis of HCN oligomers yields adenine and 4-aminoimidazole-5-carboxamide (Ferris et al. 1978). Since the first reports on the abiotic formation of adenine, the synthesis of other purines have also been achieved, including guanine and xanthine (Oró 1963), hypoxanthine (Lowe et al. 1963), 2,6-diaminopurine (Borquez et al. 2005), and 8-hydroxymethyladenine (Schwartz and Bakker 1989). More recently a novel formation of purine nucleosides has been described, in which the condensation of formamidopyrimidines (FaPy) with ribose produces adenosine with up to 20 % yield (Becker et al. 2016). This experimental scheme represents an alternative to the classical prebiotic synthesis (Fuller et al. 1972). The imidazole group, as well as its 2-methyl derivative were obtained by Oró et al. (1984) using glyoxal, aldehydes and ammonia as precursors. It is worth noting that hydrogen cyanide is the precursor for a large number of molecules, including purines, pyrimidines and amino acids (Ferris and Hagan 1984), but not histidine. The broad catalogue of molecules synthesized from HCN makes this compound a very important reactant in the origin and early evolution of life (Ruiz-Bermejo et al. 2013; Sutherland 2016).

In addition to their prebiotic synthesis, a wide array of nucleobases has been identified in meteorites, supporting the possibility that they were present in the early Earth. This is the case for purines, adenine, hypoxantine, xanthine, guanine, 2,6-diaminopurine, and 6,8-diaminopurine, among others (Callahan et al. 2011; Rios and Tor 2013). However, as of today no imidazole-bearing amino acid has been detected in meteorites (Burton et al. 2012) nor in Miller-Urey-type experiments, which may indicate that His was not present in the prebiotic set of molecules (Álvarez-Carreño et al. 2013; Raggi et al. 2016).

Catalytic Activities of Imidazolides

Different imidazole molecules have been tested for catalytic activities. Most notably, the role of activation group carried by the 2-methyl-imidazole in the formation of oligo (G)s in a non-enzymatic template-directed synthesis (Inoue and Orgel 1981). Carbonyl diimidazole was employed as condensing agent in peptide synthesis in aqueous solution, which led to 55mer-long polypeptides in clay surfaces (Ferris et al. 1996).

It has also been shown that histidine-containing di- and tri-peptides possess catalytic properties. For instance, histidyl-histidine (His-His) catalyzes the dephosphorylation of deoxyribonucleoside monophosphate, the hydrolysis of oligo $(A)_{12}$, and the oligomerization of 2′,3′-cAMP under cyclic wet-dry reaction conditions (Shen et al. 1990b). Moreover, seryl-histidine (Ser-His) catalyzes the oligomerization of trimers of imidazole-activated nucleotides (Wieczorek et al. 2013), and can also cleave several molecules, including bovine serum albumin protein, bacteriophage λ-DNA and the ester *p*-nitrophenyl acetate (Li et al. 2000). Ser-His also catalyzes peptide bond formation between two amino acids, an activity also exhibited by seryl-histidine-glycine (Ser-His-Gly) (Gorlero et al. 2009; Adamala et al. 2014). The binding of oligoarginine peptides to complementary strands of RNA helps in the template-directed non-enzymatic RNA polymerization (Jia et al. 2016), which demonstrate the catalytic and structural potential of small peptides which may have been present in the prebiotic environment.

It has also been shown that the imidazole group can act as a cofactor for ribozymes. As demonstrated by Perrotta et al. (1999), the incubation of an inactivated hepatitis delta virus (HDV) ribozyme with unsubstituted imidazole (200 mM, pH 7.4) restores the catalytic activity of the ribozyme, which is essential in the process of the rolling-circle replication of the viral RNA. This replacement shows that the imidazole can act as a co-catalyst in RNA molecules, which could be of key significance in an RNA world scenario, and of considerable significance in the evolution of biological catalysis.

In extant cells perhaps the most significant example of a catalytic imidazole moiety is histidine. The key role of histidine resides in protein mediated catalysis, a well-established biochemical fact. How proteins actually originated remains unclear. However, it is reasonable to assume that the active site of enzymes must be one of their oldest components, and biochemical and bioinformatic analyses have shown that histidine is the catalytic residue most frequently found in the active sites of enzymes (Holliday et al. 2011). This can be explained by its imidazole side-chain. The unsubstituted imidazole has a $pK_a = 7$, but on the cellular environment, the pK_a values for histidine's imidazole are estimated to be in the range of 6 to 7, reflecting the electron-withdrawing inductive effect of the protonated amino group (Edgcomb and Murphy 2002; Grimsley et al. 2009). Depending on the enzyme, the level of ionization in the imidazole of histidine is estimated to range between 9 to 50 %. This particular feature allows it to serve as a general acid or a general base, donating or accepting protons during the course of a chemical reaction in acid-base catalysis (McMurry and Tadhg 2005).

Given its significance in enzyme catalysis, the understanding of how histidine was incorporated into active sites is critical. Study of the molecular evolution of the histidine biosynthetic pathway has shown that a complex series of gene duplications and fusion events shaped the route (Alifano et al. 1996; Fani et al. 1994, 2006), but its ultimate origin remains unknown. In an effort to address this issue, a non-enzymatic synthesis of histidine was reported (Shen et al. 1987, 1990a). The proposed reaction starts with D-erythrose and formamidine. The condensation of these two compounds gives imidazole-4-acetaldehyde, which then forms histidine via a Strecker synthesis (Fig. 2). As John Oró described in considerable detail to one of us (A.L.), the cyclization reaction leading to the imidazole derivative used in their work was based on the chemical synthesis of imidazole glycerol phosphate (IGP) from ribose-5-phosphate and formamidine described by Ames (1957) as part of the efforts that led to the elucidation of histidine biosynthesis. A detailed examination of the synthesis proposed by Shen et al. (1987, 1990a), demonstrates that the formation of the imidazole group is the outcome of an Amadori rearrangement (Fig. 2, upper panel), a mechanism also involved in the biosynthesis of imidazole glycerol phosphate by the IGP synthase (HisHF) enzyme during

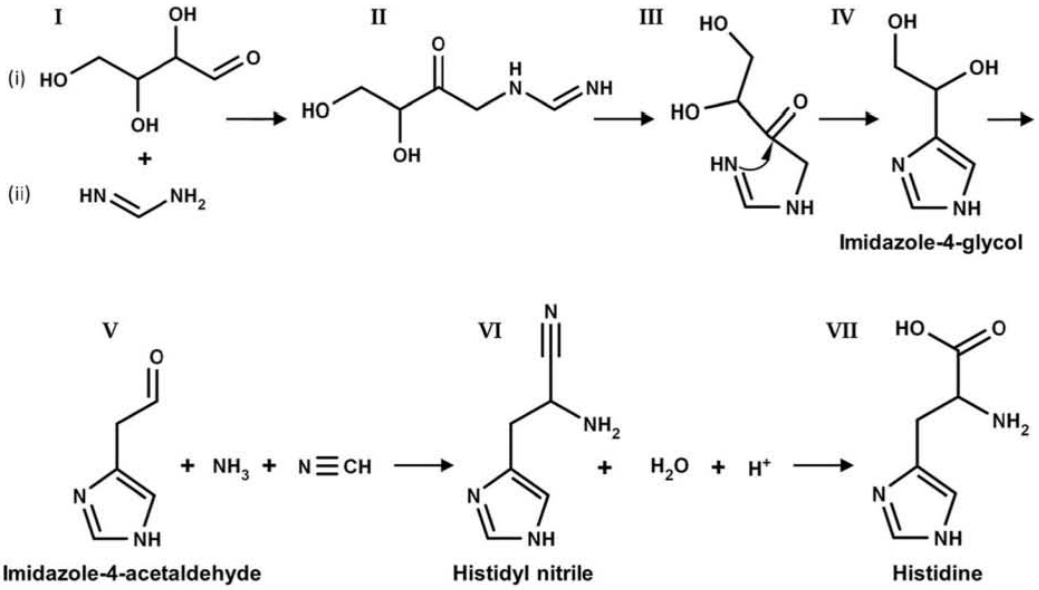

Fig. 2 Abiotic synthesis of histidine proposed by Shen et al. (1987, 1990a). The first step in the reaction is the condensation of erythrose (i) with formamidine (ii). Subsequently to the formation of imidazole-4-acetaldehyde, histidine is formed via Strecker synthesis. As shown in the upper part of the scheme, an Amadori rearrangement gives place to the imidazole ring. Chemical structures were drawn using MarvinSketch v16.1.4 ChemAxon (http://www.chemaxon.com)

histidine biosynthesis. However, both routes are completely different and there is clearly no direct evolutionary connection between the abiotic synthesis of His reported by Shen et al. (1987, 1990a) and the extant histidine anabolic route. As underlined by a generous reviewer, since the stereocenters of D-erythrose do not partake in the reaction described by Shen et al. (1987, 1990a), the a priori selection of this enantiomer does not diminish the prebiotic significance of the proposed histidine synthesis. In fact, the same would be true if other 4-carbon aldoses were used as precursors. However, both erythrose and threose are minor products in the formose reaction (Decker et al. 1982). On the other hand, the reaction reported by Shen et al. (1987, 1990a) is inefficient (3.5 % yield based on the ratio of His/erythrose), and the conditions of formation are not truly prebiotic because of the requirement of high concentration of formamidine (0.3 M), which is a labile compound that is rapidly hydrolyzed into formic acid and ammonia (Bada et al. 2016). This strongly hinders the prebiotic significance of this synthesis, and leaves open the issue of the absence of His from the prebiotic catalog of molecules.

On the Autocatalytic 4-Methylidene-Imidazole-one (MIO) Biosynthesis

In addition to its presence in purines and histidine, the imidazole moiety is also present in the cell as the 4-methylidene-imidazole-one cofactor (MIO) (Fig. 1), which plays a key role, together with other amino acids, in the aromatic amino acid lyase family of enzymes, i.e. histidine-, tyrosine- and phenylalanine-ammonia-lyases (HAL, TAL and PAL, respectively) (Fig. 3). These homologous enzymes catalyze the α, β-elimination of ammonia from the corresponding amino acid using MIO as an electrophile. HAL (EC 4.3.1.3) is the first enzyme in the degradation pathway of L-histidine, and catalyzes its deamination with the formation of *trans*-urocanic acid and ammonia (Baedeker and Schulz 2002). TAL (EC 4.3.1.23) deaminates

L-tyrosine into *p*-coumaric acid, which is a precursor of the chromophore of photoactive yellow protein in purple phototropic bacteria, and in *Saccharothrix espanaensis* coumarate is used in the biosynthesis of saccharomicin, which is an oligosaccharide antibiotic (Berner et al. 2006). PAL (EC 4.3.1.24) is part of the degradative metabolism of phenylalanine, and deaminates the amino acid into *trans*-cinnamic acid and ammonia in both plants and fungi (Kong 2015).

The phylogenetic distribution of the MIO-dependent enzymes suggest that they are quite recent. The MIO-dependent ammonia-lyases are structurally homologous to the tyrosine and phenylalanine aminomutases (TAM and PAM, respectively) (Fig. 3), which catalyze the interconversion of a β-amino acid from an α-amino acid using the same MIO cofactor. In this manner, TAM (EC 5.4.3.6) catalyzes the formation of 3-amino-3-(4-hydroxyphenyl) propanoate, while PAM (EC 5.4.3.11) transforms phenylalanine into β-phenylalanine.

As noted in Fig. 3, the active site of these enzymes is highly conserved, and the mechanism of MIO formation is the same in all of them. MIO formation is an autocatalytic process that involves a highly conserved motif consisting of alanine (Ala), serine (Ser), and glycine (Gly) found in the sequence. Because of this, MIO has been catalogued as a polypeptide-derived cofactor, which corresponds to a special case of prosthetic groups (Fischer et al. 2010). As summarized in Fig. 4, the reaction starts with the attack of the amine of Gly to the carbonyl of Ala, generating a five membered intermediate, which is then transformed into the MIO cofactor (Fig. 4) (Rétey 2003). As detailed below, here we underline the experimental significance of this reaction mechanism for a possible model prebiotic experiment.

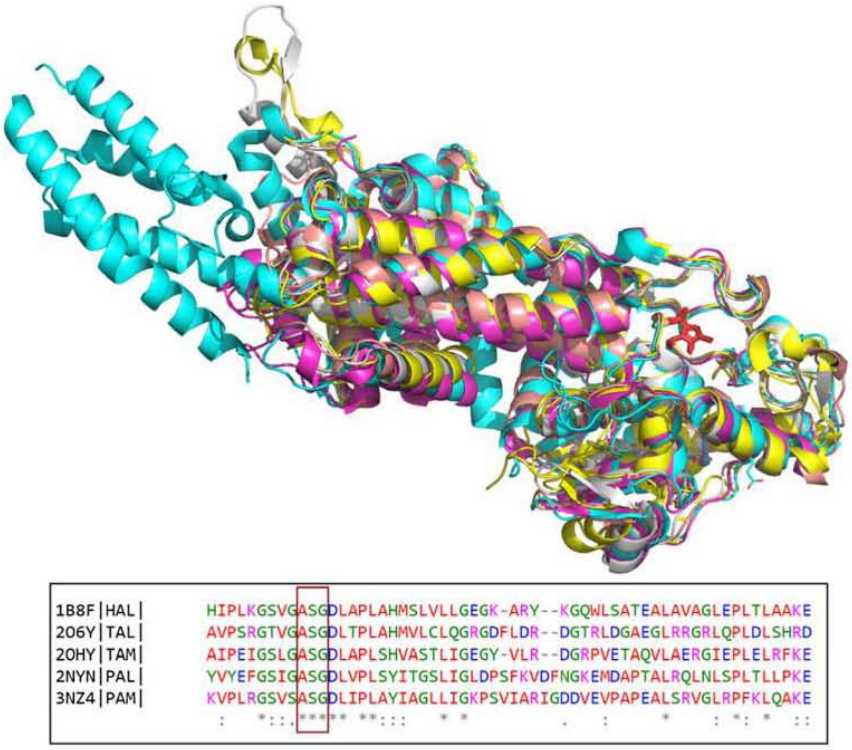

Fig. 3 Structural alignment of the HAL (PDB: 1B8F), PAL (PDB: 2NYN), TAL (PDB: 2O6Y), PAM (PDB: 2NZ4) and TAM (PDB: 2OHY) enzymes. In sticks view and red is shown the MIO cofactor. Below it is shown a fragment of the sequence alignment of the same enzymes, the box indicates the Ala-Ser-Gly motif, which is required for the MIO biosynthesis. The alignment of the crystallographic structures was performed using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.)

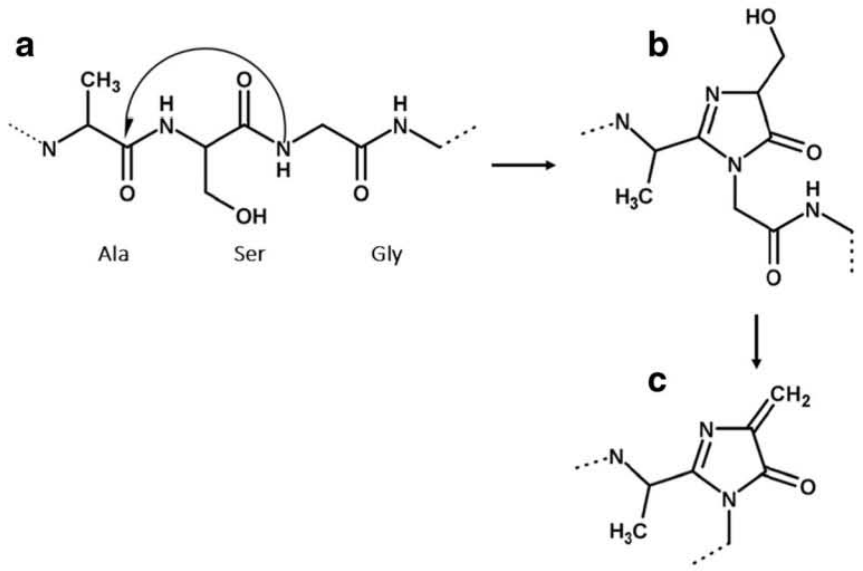

Fig. 4 MIO formation via cyclization of an Ala-Ser-Gly- motif. (Modified from Rétey 2003). Chemical structures were drawn using MarvinSketch v16.1.4 ChemAxon (http://www.chemaxon.com)

Could the Abiotic Condensation of Ala-Ser-Gly Give Rise to Imidazolides?

Numerous experiments simulating the prebiotic synthesis of organic compounds under possible primitive Earth conditions have been carried out using combinations of gases ranging from the highly reducing (CH_4, NH_3, H_2O and H_2) (Miller 1953), to neutral (H_2O, N_2, CO_2) mixtures (Cleaves et al. 2008). In all cases, alanine, serine and glycine are always found, strongly suggesting their occurrence in the primitive Earth. The chemical characterization of these amino acids in carbonaceous chondrites is also consistent with their presence in the prebiotic inventory of organic molecules (Burton et al. 2012). Additionally, peptides including these amino acids have been found in irradiated interstellar ice analogues (Kaiser et al. 2013).

The amino acids that give rise to the MIO cofactor are precisely three of the most abundant ones found in prebiotic experiments. However, the laboratory syntheses of functional products of cyclization reactions of small peptides is challenging and rather inefficient (Lambert et al. 2001; Skropeta et al. 2004). While the overall peptidic component might have been relatively large (Brack 2007) on the early Earth, we are aware that in the absence of a primitive coding mechanism, the likelihood of formation of significant amounts of a specific tripeptide is certainly small. However, it is known that intramolecular cyclization of peptides is favored by high dilutions and that cyclizations tend to be inherently slow reactions (Davies 2003). If an abiotic process leading to the synthesis of amino acids and small peptides that included sequences such as the Ala-Ser-Gly motif is assumed, the possibility of the formation of an imidazole molecule in a MIO-like fashion may be plausible, given the right conditions. This may include condensation reactions between prebiotic amino acids in lagoons and other water bodies undergoing wet-and-dry cycles, as well as the interaction with clays, metal cations and even other organic molecules. The micro-environmental conditions are major factors which may be of pivotal importance for the proposed synthesis, since the biogenesis of the MIO cofactor is not only dependent on the existence of the Ala-Ser-Gly sequence, but also on the chemical microenvironment experienced by the peptide due to conformation of the larger peptide within which the cyclization happens. The interaction with other peptides may promote the cyclization process.

Here we propose an empirical exploration of a cyclization of the Ala-Gly-Ser tripeptide after exposure to dry (by heating up the sample at 85C) and wet phases (in which sample is re-hydrated at 65C). The dry phase and wet phase combined could complete one cycle. The wet-dry cycle believed to be associated with day/night and/or seasonal cycles on early Earth was proven experimentally to promote polymerization (Mamajanov et al. 2014; Forsythe et al. 2015). At the end of each cycle, a sample will be analyzed using mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR). As noted above, the Ala-Ser-Gly motif does not occur isolated in the enzymes, therefore the use of capped hydrophobic termini on the experimental tripeptides may also be a good approach to the formation of the imidazole molecule. The effect of pH is important and hence the synthesis in different pH ranges (1–2, 6–7, 9–10, 12–13) should be explored, as should the inclusion of other prebiotically-produced organics, including tars, to help mimic the microenvironment of the biogenesis reaction. Proof-of-principle experiment is currently underway.

Conclusions

How the transition from prebiotic chemistry to the first forms of life took place is unknown. However, the prevalence of imidazole molecules in prebiotic chemistry experiments and contemporary metabolism suggest that this chemical group played a unique role due to its catalytic properties.

The processes discussed here show the existence of multiple abiotic and prebiotic mechanisms for the synthesis of imidazole derivatives, some of which present similarity, at some extent, to the biosynthesis of imidazole compounds found in extant metabolism. Particularly in the case of the abiotic synthesis of histidine reported by Shen et al. (1987, 1990a) and the biosynthesis of the same amino acid, both routes coincide in the Amadori rearrangement that produces the imidazole ring. Nevertheless, this does not indicate by itself an evolutionary continuity between them. The major differences between the reaction mechanisms involved in the biosynthetic processes leading to the formation of purines, histidine and the MIO cofactor are evidence for the polyphyletic origin of these imidazole-bearing compounds in biological systems.

Although in general there are few similarities between abiotic processes and extant biosynthetic routes (Lazcano and Miller 1999), biological processes can be a key inspiration for the empirical exploration of possible prebiotic synthesis. This is particularly true for non-enzymatic processes. Perhaps the best example is the successful selective abiotic synthesis of 2,4 (6) sugar phosphates using glycoaldehyde phosphate, based on the recognition that in extant carbohydrate metabolic pathways sugars are generally phosphorylated (Mueller et al. 1990). Using the same reasoning, and taking into account the biosynthesis of the MIO cofactor, we propose the synthesis of an imidazole compound under putative primitive conditions starting from small peptides containing the Ala-Ser-Gly motif. Empirical demonstration for its synthesis and accumulation in the primitive environment is required.

The proposed synthesis of an imidazolide by condensation of an Ala-Ser-Gly-containing peptide does not represent the emergence of an intermediate state between a prebiotic imidazole ring and histidine, and hence does not solve in any way the problem of the possible absence of this amino acid in the prebiotic environment. The origin and posterior incorporation of histidine into the active sites of enzymes remains an open issue.

Acknowledgments Alberto Vázquez-Salazar is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship 404507 from CONACYT. Financial support from the PAPIIT-UNAM (IN223916) is gratefully acknowledged. George Tan

and Amanda Stockton were supported by a start-up fund from the Georgia Institute of Technology. We thank Sara Islas and Ricardo Hernandez-Morales for help with the manuscript. We also thank Dr. Ramanarayanan Krishnamurthy for helpful discussions and his critical reading of the manuscript, as well to an anonymous reviewer for many kind suggestions.

References

Adamala K, Anella F, Wieczorek R, Stano P, Chiarabelli C, Luisi PL (2014) Open questions in origin of life: experimental studies on the origin of nucleic acids and proteins with specific and functional sequences by a chemical synthetic biology approach. Comput Struct Biotechnol J 9:e201402004

Alifano P, Fani R, Liu P, Lazcano A, Bazzicalupo M, Carlomagno MS, Bruni CB (1996) Histidine biosynthetic pathway and genes: structure, regulation, and evolution. Microbiol Rev 60:44–69

Álvarez-Carreño C, Becerra A, Lazcano A (2013) Norvaline and norleucine may have been more abundant protein components during early stages of cell evolution. Orig Life Evol Biosph 43(4–5):363–375

Ames BN (1957) The biosynthesis of histidine: D-erythro-imidazole glycerol phosphate dehydrase. J Biol Chem 228(1):131–143

Bada JL, Chalmers JH, Cleaves HJ (2016) Is formamide a geochemically plausible prebiotic solvent? Phys Chem Chem Phys 18(30):20085–20090

Baedeker M, Schulz GE (2002) Autocatalytic peptide cyclization during chain folding of histidine ammonialyase. Structure 10:61–67

Becker S, Thoma I, Deutsch A, Gehrke T, Mayer P, Zipse H, Carell T (2016) A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. Science 352(6287):833–836

Berner M, Krug D, Bihlmaier C, Vente A, Müller R, Bechthold A (2006) Genes and enzymes involved in caffeic acid biosynthesis in the actinomycete *Saccharothrix espanaensis*. J Bacteriol 188(7):2666–2673

Borquez E, Cleaves HJ, Lazcano A, Miller SL (2005) An investigation of prebiotic purine synthesis from the hydrolysis of HCN polymers. Orig Life Evol Biosph 35(2):79–90

Brack A (2007) From interstellar amino acids to prebiotic catalytic peptides: a review. Chemistry and Biochemistry 4(4):665–679

Burton AS, Stern JC, Elsila JE, Glavin DP, Dworkin JP (2012) Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. Chem Soc Rev 41(16):5459–5472

Callahan MP, Smith KE, Cleaves HJ, Ruzicka J, Stern JC, Glavin DP, House CH, Dworkin JP (2011) Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. Proc Natl Acad Sci U S A 108:13995

Cleaves HJ, Chalmers JH, Lazcano A, Miller SL, Bada JL (2008) A reassessment of prebiotic organic synthesis in neutral planetary atmospheres. Orig Life Evol Biosph 38(2):105–115

Davies JS (2003) The cyclization of peptides and depsipeptides. J Pept Sci 9:471–501

Decker P, Schweer H, Pohlmann R (1982) Bioids. X. Identification of formose sugars, presumable prebiotic metabolites using capillary gas chromatography/gas chromatography-mass spectrometry of *n*-butoxime trifluoroacetates on OV-225. J Chromatogr 244:281–291

Edgcomb SP, Murphy KP (2002) Variability in the pKa of histidine side-chains correlates with burial within proteins. Proteins 1 49(1):1–6

Fani R, Lió P, Chiarelli I, Bazzicalupo M (1994) The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for the *hisA* and *hisF* genes. J Mol Evol 38:489–495

Fani R, Brilli M, Fondi M, Lió P (2006) The role of gene fusions in the evolution of metabolic pathways: the histidine biosynthesis case. BMC Evol Biol 7(Suppl 2):S4

Ferris JP, Hagan WJ Jr (1984) HCN and chemiclal evolution. Tetrehedron 40:1093–1120

Ferris JP, Orgel LE (1966) An unusual photochemical rearrangement in the synthesis of adenine from hydrogen cyanide. J Am Chem Soc 88(5):1074–1074

Ferris JP, Joshi PC, Edelson E, Lawless JG (1978) HCN: a plausible source of purines, pyrimidines and amino acids on the primitive earth. J Mol Evol 11:293–311

Ferris JP, Hill AR Jr, Liu R, Orgel LE (1996) Synthesis of long prebiotic oligomers on mineral surfaces. Nature 381(6577):59–61

Fischer JD, Holliday GL, Rahman SA, Thornton JM (2010) The structures and physicochemical properties of organic cofactors in biocatalysis. J Mol Biol 403(5):803–824

Forsythe JG, Yu SS, Mamajanov I, Grover MA, Krishnamurthy R, Fernández FM, Hud NV (2015) Ester-mediated amide bond formation driven by wet–dry cycles: a possible path to polypeptides on the prebiotic earth. Angew Chem Int Ed 54:9871–9875

Fuller WD, Sanchez RA, Orgel LE (1972) Studies in prebiotic synthesis. VI Synthesis of purine nucleosides J Mol Biol 67(1):25–33
Gorlero M, Wieczorek R, Adamala K, Giorgi A, Schininà ME, Stano P, Luisi PL (2009) Ser-his catalyses the formation of peptides and PNAs. FEBS Lett 583(1):153–156
Grimsley GR, Scholtz JM, Pace CN (2009) A summary of the measured pK values of the ionizable groups in folded proteins. Protein Sci 18(1):247–251
Holliday GL, Fischer JD, Mitchell JB, Thornton JM (2011) Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis. FEBS J 278(20):3835–3845
Inoue T, Orgel LE (1981) Substituent control of the poly(C)-directed oligomerization of guanosine 5′-phosphoroimidazolide. J Am Chem Soc 103:7666–7667
Jia TZ, Fahrenbach AC, Kamat NP, Adamala KP, Szostak JW (2016) Oligoarginine peptides slow strand annealing and assist non-enzymatic RNA replication. Nat Chem. doi:10.1038/nchem.2551
Kaiser RI, Stockton AM, Kim YS, Jensen EC, Mathies RA (2013) On the formation of dipeptides in interstellar model ices. Astrophysical Journal 765:111–120
Kong J-Q (2015) Phenylalanine ammonia-lyase, a key component used for phenylpropanoids production by metabolic engineering. RSC Adv 5:62587–62603
Lambert JN, Mitchell JP, Roberts KD (2001) The synthesis of cyclic peptides. J. Chem. Soc. Perkin Trans 1:471–484
Lazcano A, Miller SL (1999) On the origin of metabolic pathways. J Mol Evol 49(4):424–431
Li Y, Zhao Y, Hatfield S, Wan R, Zhu Q, Li X, McMills M, Ma Y, Li J, Brown KL, He C, Liu F, Chen X (2000) Dipeptide seryl-histidine and related oligopeptides cleave DNA, protein, and a carboxyl ester. Bioorg Med Chem 12:2675–2680
Lowe CU, Ress A, Markham FRS (1963) Synthesis of complex organic compounds from simple precursors: formation of amino-acids, amino-acid polymer, fatty acids and purines from ammonium cyanide. Nature 19:219–222
Mamajanov I, MacDonald PJ, Ying J, Duncanson DM, Dowdy GR, Walker CA, Engelhart AE, Fernández FM, Grover MA, Hud NV, Schork FJ (2014) Ester formation and hydrolysis during wet–dry cycles: generation of far-from-equilibrium polymers in a model prebiotic reaction. Macromolecules 47(4):1334–1343
McMurry J, Tadhg B (2005) The organic chemistry of biological pathways. Roberts and Company Publishers
Miller SL (1953) A production of amino acids under possible primitive earth conditions. Science 117:528–529
Mueller D, Pitsch S, Kittaka A, Wagner E, Wintner CE, Eschenmoser A (1990) Chemistry of alpha-aminonitriles. Aldomerisation of glycolaldehyde phosphate to *rac*-hexose 2,4,6-triphosphates and (in presence of formaldehyde) *rac*-pentose 2,4-diphosphates: *rac*-allose 2,4,6-triphosphate and *rac*-ribose 2,4-diphosphate are the main reaction products. Helv Chim Acta 73:1410–1468
Oró J (1960) Synthesis of adenine from ammonium cyanide. Biochem Biophys Res Commun 2:407–412
Oró J (1963) Synthesis of organic compounds by electric discharges. Nature 197:862–867
Oró J, Kimball AP (1961) Synthesis of purines under possible primitive earth conditions. I Adenine from hydrogen cyanide Arch Biochem Biophys 94:217–227
Oró J, Kimball AP (1962) Synthesis of purines under possible primitive earth conditions. II. Purine intermediates from hydrogen cyanide. Arch Biochem Biophys 96:293–313
Oró J, Basile B, Cortes S, Shen C, Yamrom T (1984) The prebiotic synthesis and catalytic role of imidazoles and other condensing agents. Orig Life 14(1–4):237–242
Perrotta AT, Shih I, Been MD (1999) Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. Science 286(5437):123–126
Raggi L, Bada JL, Lazcano A (2016) On the lack of evolutionary continuity between prebiotic peptides and extant enzymes. Phys Chem Chem Phys. doi:10.1039/c6cp00793g
Rétey J (2003) Discovery and role of methylidene imidazolone, a highly electrophilic prosthetic group. Biochim Biophys Acta 1647(1–2):179–184
Rios AC, Tor Y (2013) On the origin of the canonical nucleobases: an assessment of selection pressures across chemical and early biological evolution. Isr J Chem 53(6–7):469–483
Ruiz-Bermejo M, Zorzano MP, Osuna-Esteban S (2013) Simple organics and biomonomers identified in HCN polymers: an overview. Life (Basel) 3(3):421–448
Schwartz AW, Bakker CG (1989) Was adenine the first purine? Science 245:1102–1104
Shen C, Yang L, Miller SL, Oró J (1987) Prebiotic syntheses of imidazole-4-acetaldehyde and histidine. Origins of Life 17:295–305
Shen C, Yang L, Miller SL, Oró J (1990a) Prebiotic synthesis of histidine. J Mol Evol 31:167–174
Shen C, Lazcano A, Oro J (1990b) The enhancement activities of histidyl-histidine in some prebiotic reactions. J Mol Evol 31:445–452
Skropeta D, Jolliffe KA, Turner P (2004) Pseudoprolines as removable turn inducers: tools for the cyclization of small peptides. J Org Chem 69:8804–8809
Sutherland JD (2016) The origin of life-out of the blue. Angew Chem Int Ed Engl 55(1):104–121
Wieczorek R, Dörr M, Chotera A, Luisi PL, Monnard PA (2013) Formation of RNA phosphodiester bond by histidine-containing dipeptides. Chembiochem 14(2):217–223

ANEXO 1

8. Gott, J.A., Liley, D.T.J., and Hobson, J.A. (2017). Towards a functional understanding of PGO waves. Front. Hum. Neurosci. *11*, 89.

9. Steriade, M. (2009). Sleep Oscillations and PGO Waves. In Encyclopedia of Neuroscience, pp. 1035–1042.

10. Fernández-Mendoza, J., Lozano, B., Seijo, F., Santamarta-Liébana, E., Ramos-Platón, M.J., Vela-Bueno, A., and Fernández-González, F. (2009). Evidence of subthalamic PGO-like waves during REM sleep in humans: a deep brain polysomnographic study. Sleep *32*, 1117–1126.

11. Parrino, L., Ferri, R., Bruni, O., and Terzano, M.G. (2012). Cyclic alternating pattern (CAP): The marker of sleep instability. Sleep Med. Rev. *16*, 27–45.

12. Lecci, S., Fernandez, L.M.J., Weber, F.D., Cardis, R., Chatton, J.-Y., Born, J., and Lüthi, A. (2017). Coordinated infraslow neural and cardiac oscillations mark fragility and offline periods in mammalian sleep. Sci. Adv. *3*, e1602026.

13. Menicucci, D., Piarulli, A., Allegrini, P., Bedini, R., Bergamasco, M., Laurino, M., Sebastiani, L., and Gemignani, A. (2015). Looking for a precursor of spontaneous sleep slow oscillations in human sleep: The role of the sigma activity. Int. J. Psychophysiol. *97*, 99–107.

14. Rolls, A., Colas, D., Adamantidis, A., Carter, M., Lanre-Amos, T., Heller, H.C., and de Lecea, L. (2011). Optogenetic disruption of sleep continuity impairs memory consolidation. Proc. Natl. Acad. Sci. USA *108*, 13305–13310.

15. Maingret, N., Girardeau, G., Todorova, R., Goutierre, M., and Zugaro, M. (2016). Hippocampo-cortical coupling mediates memory consolidation during sleep. Nat. Neurosci. *19*, 959–964.

16. Khodagholy, D., Gelinas, J.N., and Buzsáki, G. (2017). Learning-enhanced coupling between ripple oscillations in association cortices and hippocampus. Science *358*, 369–372.

17. Penttonen, M., Nurminen, N., Miettinen, R., Sirviö, J., Henze, D.A., Csicsvári, J., and Buzsáki, G. (1999). Ultra-slow oscillation (0.025 Hz) triggers hippocampal afterdischarges in Wistar rats. Neuroscience *94*, 735–743.

18. Seibt, J., Richard, C.J., Sigl-Glöckner, J., Takahashi, N., Kaplan, D.I., Doron, G., de Limoges, D., Bocklisch, C., and Larkum, M.E. (2017). Cortical dendritic activity correlates with spindle-rich oscillations during sleep in rodents. Nat. Commun. *8*, 684.

19. Reimer, J., McGinley, M.J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D.A., and Tolias, A.S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. Nat. Commun. *7*, 13289.

20. Naji, M., Krishnan, G., McDevitt, E.A., Bazhenov, M., and Mednick, S.C. (2017). Coupling of autonomic and central events during sleep boosts declarative memory consolidation. bioRxiv *195586*, https://doi.org/10.1101/195586.

Early Life: Embracing the RNA World

Alberto Vázquez-Salazar[1] and Antonio Lazcano[1,2,*]
[1]Facultad de Ciencias, Universidad Nacional Autónoma de México, Apdo. Postal 70-407, Cd. Universitaria, 04510, CDMX, Mexico
[2]Miembro de El Colegio Nacional, Mexico
*Correspondence: alar@ciencias.unam.mx
https://doi.org/10.1016/j.cub.2018.01.055

Protein complexes with artificial aptamers exhibit a bias towards arginine, tryptophan and tyrosine, providing insights into physicochemical traits of the interactions between amino acids and RNA that may have led to the origin of the genetic code and the ribosome.

By the late 1960s the ample biological distribution of RNA and its biochemical properties had started to catch the attention of evolutionists, and in an extraordinary combination of scholarship and conjecture Francis Crick, Leslie Orgel and a few others independently suggested that the first living beings were devoid of both DNA and proteins, and dependent on RNA both as a genetic polymer and as a catalyst. As Crick wrote, the primitive ribosome "had not protein at all and consisted entirely of RNA" [1], while Orgel (1968) stated that "it seems quite possible that polynucleotide chains could make a primitive selection among organic molecules such as amino acids, by forming stereospecific complexes stabilized by hydrogen-bonding and hydrophobic interactions. This will be important in any discussion of the evolution of the genetic code" [2].

The groundbreaking discovery of ribozymes gave considerable credibility to the idea that RNA had played a key role in the appearance of life. The discovery of ribozymes took everyone by surprise, because no one was looking for them, but the stunning widening of their functional repertoire under *in vitro* conditions has demonstrated that they can catalyze the same classes of chemical reactions as enzymes. The structural and regulatory properties of RNA molecules and ribonucleotides, combined with their catalytic activities and ubiquity in cellular processes, suggest that during an early, perhaps primordial, stage RNA molecules played a much more conspicuous role in heredity and metabolism (Figure 1).

We still do not know how the RNA World first appeared. However, it should not be understood as a mere collection of catalytic and replicative molecules stripped of all other recognized attributes of extant life, but rather as an evolutionary stage during which polyribonucleotides interacted with a wide array of compounds of biochemical significance that may have been present on the primitive Earth, including amino acids, membrane-forming compounds and metallic cations. As suggested by a number of experiments, the inventory probably included small functional peptides that may have helped to shape the chemical environment in which life appeared [3]. However, such peptides

would have come and gone without becoming the direct ancestors of extant proteins. Even if prebiotic processes replenishing them are postulated, in the absence of hereditary mechanisms ensuring their production and diversification, sooner or later they would be exhausted.

The evidence that the catalytic center of the ribosome is a ribozyme strongly supports the hypothesis that nucleic acid-directed protein synthesis appeared in the RNA World. In a paper published in a recent issue of *Current Biology*, Blanco *et al.* [4] follow the steps of Orgel, and argue that the physico-chemical properties of amino acids must have played a key role at the start of the stepwise evolution of protein biosynthesis. This includes their ability to interact with RNA molecules, as shown by reports on the efficient *in vitro* selection of different aptamers against arginine, valine and isoleucine, and tyrosine and tryptophan, which are cationic, hydrophobic and aromatic amino acids, respectively. Since it is unlikely that the first oligopeptides were already endowed with catalytic properties, it is possible that they may have been selected at first because of their ability to bind to RNA and stabilize functional conformations of ribozymes (Figure 2).

Extant RNA-binding protein domains may provide insights into these early stages, but as noted by Blanco *et al.*, they are the outcome of a lengthy evolutionary history that reflects distinct selective pressures. To avoid such bias, they analyzed interactions between nucleic acids and proteins, but with an ingenious twist based on the characterization of a population of complexes of proteins with artificial nucleic acid aptamers derived from *in vitro* evolution. A meticulous study using bioinformatics tools and molecular dynamics analysis of the amino acid composition and interfacial residues of the complexes allowed Blanco *et al.* to show that positively charged arginine and lysine residues are overrepresented in the data set and, quite specifically, in the interfaces where the amino acid residues are directly involved in the nucleic acid–protein interaction.

By leaving out large proteins from their experimental design, Blanco *et al.* excluded sequences that do not partake in nucleic acid binding. They selected instead two other sets of smaller proteins, one with less than 500 amino acids, and the other with less than 200 amino acids. In both cases, the aliphatic amino acids valine and leucine, as well as histidine, are strongly underrepresented in the binding sites. In these two sets, a compositional bias was observed towards positively charged residues, specifically arginine and lysine, as well as towards the aromatic residues tryptophan and tyrosine, which are among the top-most catalytic residues of extant enzymes [5].

The same trends are observed in the amino acids present in the nucleic acid–protein interaction interface and the solvating areas of the complexes. A fascinating observation is that the properties of biological proteins are not quite the same as those of random sequences of amino acids, suggesting that in spite of the selective pressures, considerable degrees of freedom may exist in the sequence space of the components of peptide–RNA complexes. The results reported by Blanco *et al.* are consistent with the views of Fox and Naik [6] and of Kovacs *et al.* [7], who have argued that the direct association of oligopeptides and proteins with primitive ribosomal RNA resulted in their mutual chaperoning and coevolution. The association of RNA and an arginine-rich proto-peptidic scaffold would provide protection to both components against environmental insults and chemical degradation. Because the binding between RNA and oligopeptides involves weak electrostatic interactions, the process described by Chen and her coworkers would have been most efficient in low temperature primitive environments.

A major caveat of the model is the prebiotic availability of arginine and lysine. While valine, leucine, and tyrosine are present in organic-rich chondritic meteorites and are readily formed in experimental simulations of the primitive soup, the same is not true of arginine, lysine or histidine. There are reports of a

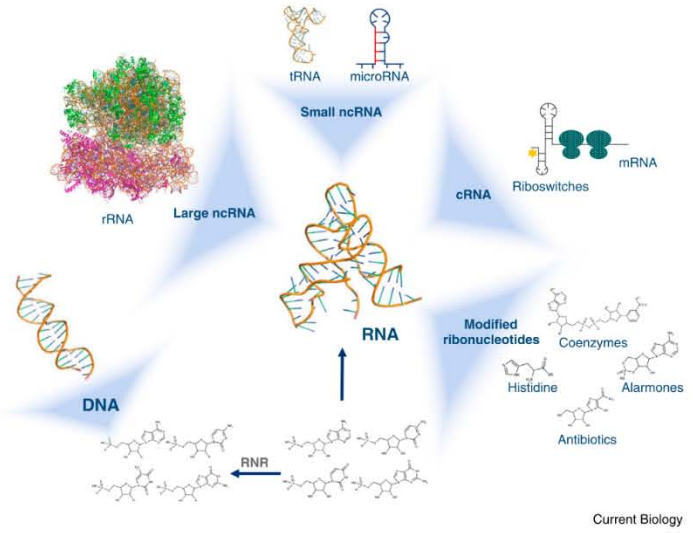

Figure 1. The key roles of RNA and ribonucleotides in cellular processes.
RNA and ribonucleotides are ubiquitous and play key catalytic, structural and regulatory roles in biological processes. The highly conserved ribonucleotide biosynthetic pathway very likely appeared prior to the divergence of the three major lineages. As shown in the lower part of the figure, ribonucleotides are reduced by ribonucleotide reductase (RNR) to deoxyribonucleotides, and some of their derivatives like coenzymes participate in biological catalysis and, as alarmones, in the activation or inhibition of cellular processes under stress conditions. RNA molecules can be divided into coding (cRNA) and non-coding (ncRNA). Like riboswitches, long ncRNA and small ncRNA play a key role in regulating gene expression (based on [12]).

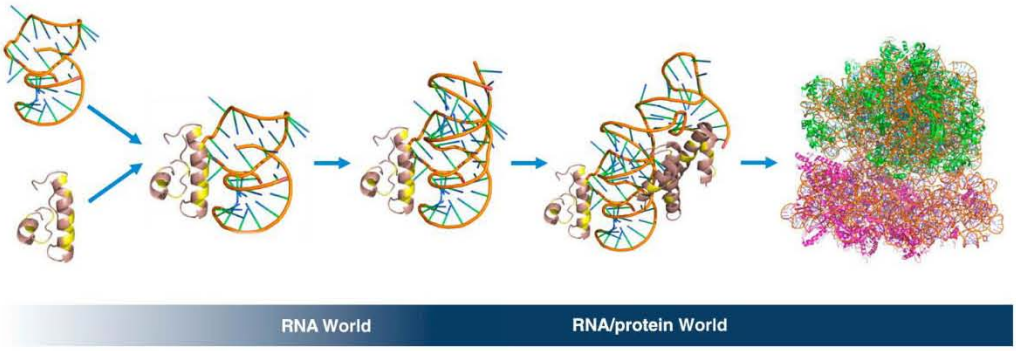

Figure 2. The association between RNA and cationic amino acid-rich oligopeptides.
During primordial times the binding of prebiotic oligopeptides with basic amino acids (in yellow) would stabilize the catalytic and functional conformation of RNAs, leading to the formation of complexes that would coevolve into more complex, efficient ribozymes from which primitive ribosomes are hypothesized to have evolved.

high energy abiotic synthesis of arginine by a Fischer–Tropsch-type process [8] and, more recently, from alpha-aminopropionitrile in a multi-pot low-energy cyanosulfidic chemical system [9], but the absence of arginine and other basic amino acids in meteorites raises the issue of which compounds were available in the prebiotic Earth that could interact with RNA.

It is of course possible that some amino acids like arginine and histidine are the outcome of early biological evolution and not of chemical evolution. This does not invalidate the conclusions of Blanco *et al.*, since their work is based on the overall physicochemical properties of amino acids, including those that are not found in extant living beings. The available evidence suggests that the prebiotic broth was a chemical wonderland, and may have included oligomers containing both amide and ester linkages [10], as well as chemical proxies of arginine and lysine, such as 2,4-diaminobutanoic acid, a diamino acid found in the Murchison meteorite [11]. It is also possible to picture metal ion-mediated interactions between the phosphate backbone of RNA with negatively charged amino acids, like aspartic and glutamic acids, which are among the most abundant compounds synthesized under laboratory conditions and are present in carbon-rich meteorites.

Looking back across time since the origin of life occurred is a scientific exercise fraught with guesswork, since the evidence required to understand the prebiotic environment and the nature of the events that led to the first living systems is scant and sometimes open to opposite interpretations. However, the system-level approach developed by Blanco *et al.* is a good example of how to design experiments that can help to overcome at least in part the reductionism that haunts our understanding of the emergence of life.

REFERENCES

1. Crick, F.H.C. (1968). The origin of the genetic code. J. Mol. Biol. *39*, 367–379.
2. Orgel. (1968). Evolution of the genetic apparatus. J. Mol. Biol. *39*, 381–393.
3. Wieczorek, R., Dörr, M., Chotera, A., Luisi, P.L., and Monnard, P.A. (2013). Formation of RNA phosphodiester bond by histidine-containing dipeptides. Chembiochem. *14*, 217–223.
4. Blanco, C., Bayas, M., Yan, F., and Chen, I.A. (2018). Analysis of evolutionary independent protein-RNA complexes yield a criterion to evaluate the relevance of prebiotic scenarios. Curr. Biol. *28*, 526–537.e5.
5. Holliday, G.L., Fischer, J.D., Mitchell, J.B., and Thornton, J.M. (2011). Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis. FEBS J. *278*, 3835–3845.
6. Fox, G.E., and Naik, A.K. (2004). The evolutionary history of the translation machinery. In The Genetic Code and the Origin of Life, L.R. de Pouplana, ed. (New York: Kluwer Academic/Plenum Publishers), pp. 92–105.
7. Kovacs, N.A., Petrov, A.S., Lanier, K.A., and Williams, L.D. (2017). Frozen in time: the history of proteins. Mol. Biol. Evol. *34*, 1252–1260.
8. Yoshino, D., Hayatsu, R., and Anders, E. (1971). The origin of organic matter in early solar system. III. Amino acids: Catalytic synthesis. Geochim. Cosmochim. Acta *3l5*, 927–993.
9. Sutherland, J.D. (2016). The origin of life – out of the blue. Angew. Chem. Int. Ed. *55*, 104–121.
10. Forsythe, J.G., Yu, S.S., Mamajanov, I., Grover, M.A., Krishnamurthy, R., Fernández, F.M., and Hud, N.V. (2015). Ester-mediated amide bond formation driven by wet-dry cycles: a possible path to polypeptides on the prebiotic Earth. Angew. Chem. Int. Ed. *54*, 9871–9875.
11. Meierhenrich, U.J., Muñoz Caro, G.M., Bredehöft, J.H., Jessberger, E.K., and Thiemann, W.H. (2004). Identification of diamino acids in the Murchison meteorite. Proc. Natl. Acad. Sci. USA *101*, 9182–9186.
12. Lazcano, A. (2014). The RNA World: stepping out of the shadows. In Why Does Evolution Matter? The Importance of Understanding Evolution, G. Trueba, ed. (Newcastle upon Tyne: Cambridge Scholars Publishing), pp. 101–119.

ANEXO 2. Materiales y método extendidos de Vázquez-Salazar, et al. 2018

Análisis filogenético basado en estructura primaria

Con la finalidad de estudiar la distribución de las dos enzimas que sintetizan el grupo imidazol, tanto en la ruta biosintética de histidina (HisF) como en la síntesis *de novo* de purinas (PurM), se crearon matrices de presencia/ausencia de dichas enzimas usando las secuencias encontradas en genomas completamente secuenciados.

Identificación y colección de secuencias

Para comenzar el análisis se identificaron las secuencias proteínicas de cada una de las enzimas a analizar. Los mapas metabólicos de la base de datos *Kyoto Encyclopedia of Genes and Genomes* (KEGG) (Kanehisa, et al., 2017) fueron empleados para identificar y descargar las secuencias. Los números de identificación de las secuencias empleadas son los siguientes: bacterias (*Escherichia coli*) PurM: b2499, HisF: b2025; arqueas (*Methanococcus jannaschii*) PurM: MJ_0203, HisF: MJ_0411; eucariontes (*Saccharomyces cerevisiae*) PurM: YGL234W, HisF: YBR248C.

Búsqueda de homólogos en genomas completamente secuenciados

Las secuencias mencionadas fueron empleadas para realizar la búsqueda de homólogos en la base de datos local de genomas completamente secuenciados usando el algoritmo *Basic Local Aligment Search Tool* (BLAST) versión 2.2.28 (Altschul, *et al.*, 1990). Los parámetros de BLAST que se usaron fueron los predeterminados. El valor de corte empleado para considerar homología fue un e-value ≤ 1e-5 con porcentaje de identidad ≥ 30%.

La información sobre homólogos en genomas completamente secuenciados fue adicionalmente arreglada en matrices de presencia/ausencia.

Reconstrucción filogenética basada en estructura terciaria

Durante el proceso evolutivo, las estructuras terciarias de las proteínas se encuentran más conservadas que las secuencias de aminoácidos; por esta razón, la comparación de estructuras proteínicas puede ser empleada para describir de manera más apropiada las relaciones evolutivas entre proteínas homólogas cuyas secuencias hayan divergido substancialmente durante la evolución (Sillitoe & Orengo, 2007). De esta manera, la comparación estructural de estructuras cristalográficas de proteínas se convierte en una alternativa novedosa para el estudio de la historia evolutiva de proteínas.

Identificación y colección de estructuras cristalográficas de proteínas

Las estructuras cristalográficas disponibles de las proteínas HisF de *Thermotoga maritima* (código: 1THF, 1.45 Å de resolución) así como para PurM de *Escherichia coli* (código: 1CLI, 2.50 Å de resolución) fueron descargados del Protein Data Bank (PDB) (http: // www.rcsb.org/pdb). Todas las estructuras fueron visualizadas y editadas mediante el programa PyMol versión 1.6 (The PyMOL Molecular Graphics System, Version 1.6 Schrödinger, LLC)

Búsqueda de homólogos estructurales

Las estructuras cristalográficas de HisF y PurM descargados previamente fueron empleados para realizar una búsqueda de homólogos estructurales en la base de datos PDB mediante el algoritmo DaliLite versión 3 (Hasegawa & Holm, 2009). La lista de posibles homólogos fue curada a mano, y los mejores resultados fueron descargados para su posterior análisis, considerando únicamente las estructuras de las proteínas de interés que no tuvieran ligandos unidos.

Algoritmos de comparación de estructuras terciarias

Empleando el algoritmo *Secondary structure matching*-SSM (Krissinel & Henrick, 2004) que forma parte del servidor web del Protein Data Bank-Europe (PDBe) (http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver), se llevó a cabo la comparación de las estructuras terciarias de las distintas proteínas identificadas como homólogos estructurales de HisF y PurM. Dicho algoritmo realiza un análisis pareado de los elementos de la estructura secundaria de cada proteína, así como una alineación tridimensional iterativa del esqueleto de carbonos alfa en las cadenas polipeptídicas; de esta manera es posible calcular la distancia entre dichos átomos en cada par de proteínas analizadas. El resultado obtenido es un valor numérico, denominado *root mean square deviation* (RMSD), mientras menor sea este valor (cero indicaría que las proteínas comparadas son idénticas), mayor es el parecido global de las estructuras proteínicas comparadas, lo cual, evolutivamente, implicaría que están más cercanamente relacinadas.

Construcción de árboles filogenéticos

Un total de 38 estructuras cristalográficas (Tabla A2.1) fueron analizadas de manera pareada mediante el algoritmo *Secondary structure matching*-SSM (Krissinel & Henrick, 2004). El valor del alineamiento estructural (SAS) para cada comparación proteína-proteína se calculó según la fórmula: *(RMSD * 1000) / # de residuos alineados* (Subbiah, *et al.*, 1993; Jácome, et al. 2015). Posteriormente fueron construidas matrices de distancia empleando los valores calculados de SAS, dichas matrices fueron analizadas utilizando el programa FITCH, basado en el algoritmo de Fitch-Margoliash, que se incluye en la paquetería PHYLIP (Felsenstein, 2005); finalmente, se construyeron dendogramas utilizando el programa DRAWTREE del mismo paquete de análisis. Los árboles fueron visualizados y editados usando FigTree versión 3.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Tabla A2.1. Estructuras cristalográficas empleadas en la construcción de dendogramas estructurales

Código PDB	Enzima	Especie	Resolusión [Å]
3P4E	PurM	*Vibrio cholerae*	1.77
2Z01	PurM	*Geobacillus kaustophilus*	2.2
3M84	PurM	*Francisella tularensis*	1.699
2V9Y	PurM	*Homo sapiens*	2.1
1CLI	PurM	*Escherichia coli (strain K12)*	2.5
2BTU	PurM	*Bacillus anthracis*	2.31
3KIZ	PurM	*Cytophaga hutchinsonii (strain ATCC 33406 / NCIMB 9469)*	1.5
3VIU	PurL	*Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579)*	2.35
1VK3	PurL	*Thermotoga maritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099)*	2.15
1T3T	PurL	*Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)*	1.9
1VQV	ThiL	*Aquifex aeolicus (strain VF5)*	2.65

3MCQ	ThiL	*Methylobacillus flagellatus (strain KT / ATCC 51484 / DSM 6875)*	1.91
2Z1T	HypE	*Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303)*	2.6
2I6R	HypE	*Escherichia coli (strain K12)*	2.51
3U0O	SelD	*Escherichia coli (strain K12)*	2.25
2ZOD	SelD	*Aquifex aeolicus (strain VF5)*	1.98
4EVZ	HisF	Ancestral reconstruction	1.46
1H5Y	HisF	*Pyrobaculum aerophilum (strain ATCC 51768 / IM2 / DSM 7523 / JCM 9630 / NBRC 100827)*	2
1THF	HisF	*Thermotoga maritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099)*	1.45
1KA9	HisF	*Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579)*	2.3

5AHE	HisA	*Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)*	1.7
1QO2	HisA	*Thermotoga maritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099)*	1.85
4GJ1	HisA	*Campylobacter jejuni subsp. jejuni serotype O:2 (strain ATCC 700819 / NCTC 11168)*	2.152
2AGK	HisA	*Saccharomyces cerevisiae (strain ATCC 204508 / S288c)*	1.3
1VC4	TrpC	*Thermus thermophilus*	1.8
3QJA	TrpC	*Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv)*	1.29
1IGS	TrpC	*Sulfolobus solfataricus (strain ATCC 35092 / DSM 1617 / JCM 11322 / P2)*	2
3TSM	TrpC	*Brucella abortus (strain 2308)*	2.15
6BMA	TrpC	*Campylobacter jejuni subsp. jejuni serotype O:2 (strain ATCC 700819 / NCTC 11168)*	1.98

1I4N	TrpC	*Thermotoga maritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099)*	2.5
4WUI	TrpF	*Jonesia denitrificans (strain ATCC 14870 / DSM 20603 / CIP 55134)*	1.09
1NSJ	TrpF	*Thermotoga maritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099)*	2
5LHE	TrpF	*Thermococcus kodakarensis (strain ATCC BAA-918 / JCM 12380 / KOD1)*	1.85
1V5X	TrpF	*Thermus thermophilus*	2
4AAJ	TrpF	*Pyrococcus furiosus (strain ATCC 43587 / DSM 3638 / JCM 8422 / Vc1)*	1.75
1VZW	PriA	*Streptomyces coelicolor (strain ATCC BAA-471 / A3(2) / M145)*	1.8
4X2R	PriA	*Actinomyces urogenitalis DSM 15434*	1.05
4U28	PriA	*Streptomyces sviceus ATCC 29083*	1.33

ANEXO 3. Anabolismo de histidina, cisteína y ácido aspártico. Resultados preliminares

El primer paso para estudiar la evolución de la biosíntesis de los tres aminoácidos catalíticos (histidina, cisteína y ácido aspártico) fue hacer un análisis de distribución filogenética siguiendo el método que a continuación se describe.

Identificación de las enzimas que participan en la biosíntesis de His, Cys y Asp

Las enzimas involucradas en la biosíntesis de histidina, cisteína y ácido aspártico se identificaron usando los mapas metabólicos de la base de datos Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa, et al., 2017) y la base MetaCyc (Caspi, et al., 2018). Las secuencias proteínicas de cada una de las enzimas fueron descargadas de KEGG. Para todos los casos se emplearon organismos modelo, donde se han realizado caracterizaciones bioquímicas de las enzimas mencionadas. Los organismos de los cuales fueron extraídas las secuencias son los siguientes: **eco**: *Escherichia coli K-12 MG1655;* **mja**: *Methanocaldococcus jannaschii;* **ape**: *Aeropyrum pernix;* sce: *Saccharomyces cerevisiae.*

Números de identificación (KEGG) de las secuencias empleadas en la búsqueda de homólogos:

I. Histidina

EC, Identificador
1. 2.4.2.17, eco b2019
2. 3.6.1.31, eco b2026
3. 3.5.4.19, eco b2026
4. 5.3.1.16, eco b2024
5. 4.1.3.-, eco b2025
6. 2.4.2.-, eco b2023
7. 4.2.1.19, eco b2022
8. 2.6.1.9, eco b2021
9. 3.1.3.15, eco b2022
10. 1.1.1.23, eco b2020

II. Cisteína

EC, Identificador
1. 2.3.1.30, eco b3607
2. 2.5.1.47, eco b2414
3. 2.5.1.47, eco b2421
4. 6.1.1.27, mjaMJ_1660
5. 2.5.1.73, mja MJ_1678
6. 2.5.1.65, ape APE_1586
7. 2.3.1.31, sce YNL277W
8. 2.5.1.49, sce YLR303W
9. 4.2.1.22, sce YGR155W
10. 4.4.1.1, sce YAL012W

III. Ácido aspártico

EC, identificador

1. 2.6.1.1, ecob0928

Búsqueda de homólogos en genomas completamente secuenciados

Las secuencias proteínicas recuperadas fueron empleadas para realizar una búsqueda de homólogos en la base de datos local de genomas completamente secuenciados (basada en KEGG), usando el algoritmo *Basic Local Aligment Search Tool* (BLAST) versión 2.2.26+ (Altschul, et al., 1990). Los parámetros de BLAST que se usaron fueron los predeterminados. El valor de corte empleado para considerar homología fue un e-value ≤ 1e5 con porcentaje de identidad ≥ 25 %.

Curación de matrices de presencia/ausencia

La información sobre homólogos en genomas completamente secuenciados fue adicionalmente arreglada en matrices de presencia/ausencia, las cuales fueron analizadas con el fin de encontrar patrones de distribución. Las matrices se encuentran disponibles en la siguiente liga: https://drive.google.com/drive/folders/1hu1bm5z-UE-6AfOwdVF-Y69K8NMuU1Gw

Resultados Preliminares. Biosíntesis de histidina

Como se describió con anterioridad (ver Capítulo II), las enzimas involucradas en la biosíntesis de histidina (Tabla A3.1; Figura A3.1) están ampliamente distribuidas en los tres dominios celulares. Esto sugiere que su biosíntesis estaba presente en el último ancestro común de todos los seres vivos (LUCA). A diferencia de lo que ocurre con cisteína (Tabla A3.2), la biosíntesis de histidina se encuentra ampliamente conservada en los tres dominios; es decir, bacterias, arqueas y eucariotas emplean las mismas enzimas para realizar las mismas reacciones bioquímicas. Las únicas diferencias inter-dominio que se encuentran son casos de fusión de genes en el dominio Eukarya, generando enzimas bifuncionales; por ejemplo, el gen *His7* producto de la fusión de los genes *hisF* y *hisH* bacterianos.

Tabla A3.1. Enzimas que participan en la biosíntesis de Histidina

Nombre	**EC**
ATP fosforibosiltransferasa, HisG	2.4.2.17
Fosforibosil-ATP pirofosfatasa, HisI-C	3.6.1.3.1
Fosforibosil-AMP ciclohidrolasa, HisI-N	3.5.4.19
Fosforibosilformimino-5-aminoimidazol carboxamida ribotido isomerasa, HisA	5.3.1.16
Imidazol glicerol-fosfato sintasa, subunidad ciclasa, HisF	4.1.3.-
Imidazol glicerol-fosfato sintasa, subunidad glutamina amidotransferasa, HisH	2.4.2.-
Imidazol glicerol-fosfato dehidratasa, HisB-C	4.2.1.19
Histidinol-fosfato aminotransferasa, HisC	2.6.1.9
Histidinol-fosfatasa, HisB-N	3.1.3.15
Histidinal/histidinol dehidrogenasa, HisD-N	1.1.1.23
Histidinal/histidinol dehidrogenasa, HisD-C	1.1.1.23

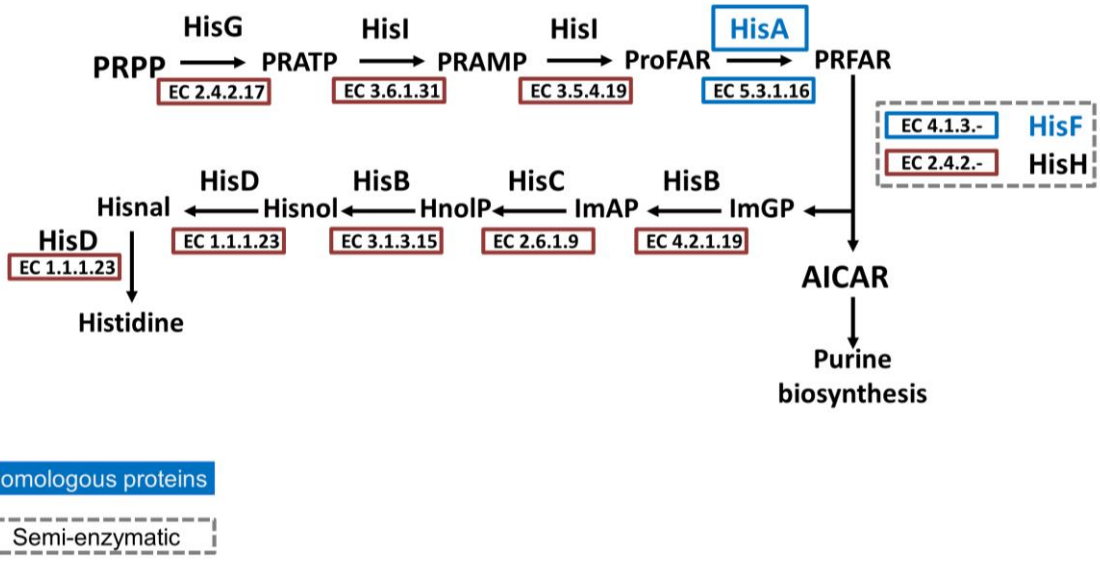

Figura A3.1. Ruta de biosíntesis de la histidina. Se trata de 10 reacciones

enzimáticas distintas catalizadas por 8 enzimas conservadas en los tres dominios celulares.

Resultados preliminares.Biosíntesis de cisteína

La cisteína cumple funciones importantes como antioxidante y como fuente de azufre para biomoléculas como la tiamina, los grupos de Fe-S, la biotina, el CoA, la metionina y varios tioles antioxidantes (glutatión, micotiol, trypanothione) (Giles, et al., 2003). De la misma manera, como se discutió en secciones anteriores (ver Capítulo I), la cisteína es un residuo catalítico muy importante en las enzimas.

En el presente trabajo, cinco distintas rutas de biosíntesis de cisteína fueron identificadas, según la base de datos MetaCyc (Caspi, et al., 2018), en el metabolismo contemporáneo (Tabla A3). Del análisis de estas enzimas se puede establecer que i) las cinco rutas de biosíntesis de cisteína están estrechamente relacionadas con la serina; ii) no hay homología entre las enzimas de las distintas rutas (todas tienen plegamientos distintos), y iii) las enzimas de las rutas de serina (EC 2.3.1.30 y 2.5.1.47), y de la ruta de homoserina (EC 2.3.1.31 y 2.5.1.49) realizan reacciones equivalentes en substratos semejantes; sin embargo, no son homólogas.

Tabla A3.2 Enzimas que participan en la biosíntesis de cisteína

I Canónica (bacteriana)	
Serina *O*-acetiltransferasa	2.3.1.30
Cisteína sintasa	2.5.1.47
II Dependiente de RNA	
O-fosfoseril-tRNA ligasa	6.1.1.27
O-fosfo-L-seryl-tRNA:Cys-tRNA sintasa	2.5.1.73
III De la *O*-fosfoserina sulfhidrilasa (OPSS)	
O-fosfoserina sulfhidrilasa	2.5.1.65
IV Presente en hongos	
Homoserina *O*-acetiltransferasa	2.3.1.31
O-acetilhomoserina aminocarboxipropiltransferasa	2.5.1.49
V A partir de homocisteína	

Cistathionina beta-sintasa 4.2.1.22

Cistathionina gamma-liasa 4.4.1.1

El análisis de la distribución de las enzimas en cada uno de los diferentes dominios de la vida indica lo siguiente.

Eucariontes

- Los animales, incluyendo al humano, tienen los genes para sintetizar cisteína a partir de la homocisteína (EC 4.2.1.22 y 4.4.1.1)
- Las plantas tienen las proteínas para sintetizar cisteína de tres maneras, a partir de de la serina (EC 2.3.1.30 y 2.5.1.47), de la fosfoserina (EC 2.5.1.65) y de la homocisteína (EC 4.2.1.22 y 4.4.1.1)
- La mayoría de los hongos tiene las enzimas para sintetizar cisteína a partir de la homoserina (EC 2.3.1.31, 2.5.1.49, 4.2.1.22 y 4.4.1.1). La excepción son los Microsporidia
- El caso interesante son los protistas, donde hay un número considerable de organismos que parecen no tener ninguna de las enzimas de las rutas de biosíntesis. La búsqueda no arrojó ningún resultado para Apicomplexa ni para el diplomonado *Giardia lamblia*
- *Entamoeba histolytica* tiene los genes para la síntesis a partir de serina (EC 2.3.1.30 y 2.5.1.47). Nuestro análisis también encuentra hits para *Phaeodactylum tricornutum, Thalassiosira pseudonana* (Diatomeas), así como para *Trypanosoma cruzi, Leishmania, Naegleria gruberi* y *Trichomonas vaginalis.*
- Los Choanoflagellados, así como los géneros de Amoebozoa *Dictyostelium y Acanthamoeba* posiblemente poseen la ruta de la homocisteína (EC 4.2.1.22 y 4.4.1.1)

Arqueas

- Las Euryarchaeota metanogénicas tienen la ruta de biosíntesis dependiente de RNA (EC 6.1.1.27 y 2.5.1.73), excepto *Methanosphaera stadtmanae*,

Methanobrevibacter smithii y *Methanobrevibacter ruminantium,* que no dan hit para dicha ruta, pero sí para las de serina (EC 2.3.1.30 y 2.5.1.47), la de fosfoserina (EC 2.5.1.65) y de homocisteína (EC 4.2.1.22 y 4.4.1.1).

- Todas las Crenarchaeota, excepto *Staphylothermus hellenicus*, *Ignicoccus hospitalis* y *Fervidicoccus fontis* dan hit para la ruta de la fosfoserina (EC 2.5.1.65).
- *Cenarchaeum symbiosum* (Thaumarchaeota), *Nanoarchaeum equitans* (Nanoarchaeota) y Candidatus *Korarchaeum cryptofilum* (Korarchaeota) no dan hit para ninguna enzima.

Bacterias

- En general, al menos una ruta completa para la biosíntesis de cisteína está presente en los grupos de Gammaproteobacteria - Enterobacteria, Gammaproteobacteria - Others, Betaproteobacteria, Epsilonproteobacteria, Deltaproteobacteria
- Ninguno de los miembros analizados de Rickettsia, Orientia, Wolbachia, Anaplasma, Ehrlichia, Neorickettsia, Midichloria, así como Bartonella (todas Alphaproteobacteria) dio hit para alguna enzima
- Tenericutes y Chlamydiae son grupos bacterianos que no dan hit para ninguna de las enzimas.
- Las Spirochaetes Borreliella, Borrelia y Treponema tampoco dan hit para ninguna enzima

Sin lugar a duda, la evolución de la biosíntesis de cisteína es compleja, como lo revelan las múltiples rutas que se encuentran distribuidas en los diferentes dominios de la vida; de hecho, la base de datos MetaCyc en sus últimas actualizaciones (Caspi, et al. 2018) ha incrementado el número de rutas biosintéticas para cisteína a nueve, debido a variantes que se encuentran en organismos como *Trichomonas vaginalis* o *Thermococcus kodakarensis*. Esto sirve

también como un recordatorio de que la biología de muchos organismos (algunos con estilos de vida extremófilo) no está completamente descrita.

Resultados preliminares. Biosíntesis de ácido aspártico

En completo contraste con lo que sucede para la biosíntesis de histidina o cisteína, para esta ruta biosintética únicamente una enzima es necesaria: la aspartato transaminasa (EC 2.6.1.1) (Figura A3.2). Esta enzima está muy poco representada en el dominio Archaea. En Bacteria se encuentra bien distribuida en γ-proteobacteria y β-protobacteria pero pobremente distribuida en los demás grupos. Interesantemente, se encuentra presente en todos los géneros de Chlamydiae analizados. Estos resultados hacen dudar sobre la biosíntesis de Asp en los organismos donde no fue detectada.

Uno de los puntos que guiará la discusión en torno al ácido aspártico como residuo catalítico es el porqué de la preferencia de este aminoácido sobre el ácido glutámico, que es bioquímicamente semejante.

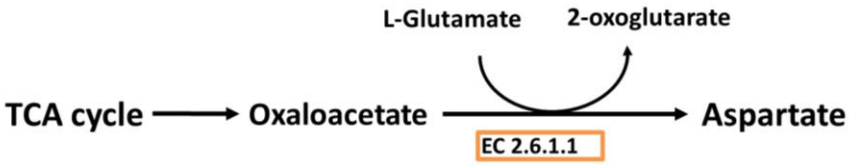

Figura A3.2. Ruta de biosíntesis del ácido aspártico. La única enzima asociada a la biosíntesis del ácido aspártico, la aspartato transaminasa (EC 2.6.1.1) emplea oxaloacetato y glutamato para formar aspartato y 2-oxoglutarato.

REFERENCIAS

Alifano, P.; Fani, R.; Liò, P.; Lazcano, A.; Bazzicalupo, M.; Carlomagno, M. S. & Bruni, C. B. (1996) **Histidine biosynthetic pathway and genes: structure, regulation, and evolution**. *Microbiol Rev*. 60(1):44-69

Altschul, S.F.; Gish, W.; Miller, W.; Myers, E. W. & Lipman, D. J. (1990) **Basic local alignment search tool.** *J Mol Biol*. 215: 403–410

Bada, J. L.; Chalmers, J. H. & Cleaves, H. J. (2016) **Is formamide a geochemically plausible prebiotic solvent?** *Phys Chem Chem Phys*. 18: 20085-20090

Baedeker, M. & Schulz, G. E. (2002) **Autocatalytic peptide cyclization during chain folding of histidine ammonialyase.** *Structure.* 10: 61-67

Ban, N.; Nissen, P.; Hansen, J.; Capel, M.; Moore, P. B. & Steitz, T. A. (1999) **Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit.** *Nature*. 400(6747): 841-847

Ban, N.; Nissen, P.; Hansen, J.; Capel, M.; Moore, P. B. & Steitz, T.A. (1999) **Placement of protein and RNA structures into a 5 Å -resolution map of the 50S ribosomal subunit.** *Nature*. 400: 841-847

Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B. & Steitz, T. A. (2000) **The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution.** *Science* 289: 905-920

Bartlett, G. J.; Porter, C. T.; Borkakoti, N. & Thornton, J. M. (2002) **Analysis of catalytic residues in enzyme active sites.** *J Mol Biol.* 324(1):105-121

Belmonte, L. & Mansy, S. S. (2016) **Metal catalysis and the origin of life.** *Elements*. 12: 413-418

Bochner, B. R. & Ames, B. N. (1982) **ZTP (5-amino 4-imidazole carboxamide riboside 5'-triphosphate): a proposed alarmone for 10-formyl-tetrahydrofolate deficiency.** *Cell*. 29(3): 929-37

Bork, P.; Sander, C. & Valencia, A. (1993) **Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases.** *Protein Sci.* (1): 31-40

Borquez, E; Cleaves, H. J.; Lazcano, A. & Miller, S. L. (2005) **An investigation of prebiotic purine synthesis from the hydrolysis of HCN polymers**. *Orig Life Evol Biosph.* 35(2):79-90

Bowman, J. C.; Hud, N. V. & Williams, L. D. (2015) **The ribosome challenge to the RNA world.** *J Mol Evol.* 80(3-4): 143-161

Brachet, J. (1942) **La localisation des acides pentosenucleiques dans les tissues animaux et les oeufs d'amphibiens en voie de developement.** *Arch. Biol.* 53: 207-257

Brilli, M. & Fani, R. (2004) **The origin and evolution of eucaryal HIS7 genes: from metabolon to bifunctional proteins?** *Gene.* 15(339): 149-60

Burton, A. S.; Stern, J. C.; Elsila, J. E.; Glavin, D. P. & Dworkin, J. P. (2012) **Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites.** *ChemSoc Rev.* 41 (16): 5459–5472

Casperson, T. (1941) **Studien uber den eiweissumsatz der zelle.** *Naturwissenschaften*. 29: 33-43

Caspi, R.; Billington, R.; Fulcher, C. A.; Keseler, I. M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Midford, P. E.; Ong, Q.; Ong, W. K.; Paley, S.; Subhraveti, P. & Karp, P. D. (2018) **The MetaCyc database of metabolic pathways and enzymes.** *Nucleic Acids Res*. 46(D1): D633-D639

Cate, J. H.; Yusupov, M. M.; Yusupova, G. Z.; Earnest, T. N.; & Noller, H. F. (1999). **X-ray crystal structures of 70S ribosome functional complexes.** *Science* 285(5436): 2095-104

Cech T. R. (2000) **Structural biology. The ribosome is a ribozyme.** *Science*. 289(5481):878-879

Cech T. R. (2009a) **Crawling out of the RNA world**. *Cell*. 136(4):599-602

Cech T. R. (2009b**) Evolution of biological catalysis: ribozyme to RNP enzyme**. *Cold Spring Harb Symp Quant Biol*. 74:11-16

Chaudhuri, B. N.; Lange, S.C.; Myers, R. S.; Chittur, S. V.; Davisson, V. J. & Smith, J. L. (2001) **Crystal structure of imidazole glycerol phosphate synthase: a**

tunnel through a (beta/alpha)$_8$ barrel joins two active sites. *Structure*. (10): 987-97

Chaudhuri, B. N.; Lange, S.C.; Myers, R. S.; Davisson, V. J. & Smith J. L. (2003) **Toward understanding the mechanism of the complex cyclization reaction catalyzed by imidazole glycerolphosphate synthase: crystal structures of a ternary complex and the free enzyme.** *Biochemistry*. 42(23): 7003-12

Christenson, S.D.; Liu, W.; Toney, M. D. & Shen, B. (2003) **A novel 4-methylideneimidazole-5-one-containing tyrosine aminomutase in enediyne antitumor antibiotic C-1027 biosynthesis**. *J Am Chem Soc*. 21; 125(20): 6062-6063

Claude, A. (1946) **Fractionation of mammalian liver cells by differential centrifugation; problems, methods, and preparation of extract.** *J. Exp. Meg.* 84:51-59

Cleaves, H. J.; Chalmers, J. H.; Lazcano, A.; Miller, S. L. & Bada, J. L. (2008) **A reassessment of prebiotic organic synthesis in neutral planetary atmospheres**. *Orig Life Evol Biosph*. 38(2): 105-115

Clemons, W. M. Jr.; May, J. L.; Wimberly, B. T.; McCutcheon, J. P.; Capel, M. S. & Ramakrishnan, V. (1999) **Structure of a bacterial 30S ribosomal subunit at 5.5 A resolution.** *Nature*. 400(6747): 833-840

Cohen, G.; Lewis, D. & Sinet P. M (1981**) Oxygen consumption during the fenton-type reaction between hydrogen peroxide and a ferrous-chelate (Fe^{2+}-DTPA)**. *J Inorg Biochem*. 15(2): 143-151

Cooke, H. A.; Christianson, C. V. & Bruner, S. D. (2009) **Structure and chemistry of 4-methylideneimidazole-5-one containing enzymes**. *Curr Opin Chem Biol.* (4): 460-468

Crick, F. H. C. (1968) **The origin of the genetic code.** *J. Mol. Biol.* 38: 367-379

Darwin, C. (1859) **On the origin of species by means of natural selection, or The preservation of favoured races in the struggle for life.** Primera Edición. London: John Murray

Daubner, S. C.; Schrimsher, J. L.; Schendel, F. J.; Young, M.; Henikoff, S.; Patterson, D.; Stubbe, J. & Benkovic, S. J. (1985) **A multifunctional protein possessing glycinamide ribonucleotide synthetase, glycinamide ribonucleotide transformylase, and aminoimidazole ribonucleotide synthetase activities in de novo purine biosynthesis.** *Biochemistry*. (24): 7059–7062

Davidovich, C.; Belousoff, M.; Wekselman, I.; Shapira, T.; Krupkin, M.; Zimmerman, E.; Bashan, A. & Yonath, A. (2010) **The Proto-Ribosome: an ancient nano-machine for peptide bond formation.** *Isr J Chem.* 50(1):29-35

Davies, J. S. (2003) **The cyclization of peptides and depsipeptides.** *J Pept Sci.* 9: 471-501

Decker, P.; Schweer, H. & Pohlmann, R. **Bioids. X. Identification of formose sugars, presumable prebiotic metabolites using capillary gas chromatography/gas chromatography-mass spectrometry of *n*-butoxime trifluoroacetates on OV-225**. *J Chromatogr*. 244: 281-291

Eakin, R. E. (1963) **An approach to the evolution of metabolism.** *PNAS*. 49: 360-366

Felsenstein, J. (2005) **PHYLIP (Phylogeny Inference Package) version 3.6.** Distributed by the author. Department of Genetics, University of Washington, Seattle

Feng, L.; Wanninayake, U.; Strom, S.; Geiger, J. & Walker, K. D. (2011) ***Mechanistic, mutational, and structural evaluation of a Taxus phenylalanine aminomutase.*** Biochemistry. 50(14): 2919-2930

Fichman, Y.; Gerdes, S. Y.; Kovács, H.; Szabados, L.; Zilberstein, A. & Csonka, L. N. (2015) **Evolution of proline biosynthesis: enzymology, bioinformatics, genetics, and transcriptional regulation.** *Biol Rev Camb Philos Soc.* 90(4): 1065-1099

Fischer, J. D.; Holliday, G. L.; Rahman, S. A. & Thornton, J. M. (2010**) The structures and physicochemical properties of organic cofactors in biocatalysis.** *J Mol Biol.* 403(5): 803-824

Fox G. E (2010) **Origin and evolution of the ribosome**. *Cold Spring Harb Perspect Biol.* 2(9): a003483

Gilbert, W. (1986) **Origin of life-The RNA world**. *Nature.* 319: 618

Giles, N. M.; Watts, A. B.; Giles, G. I.; Fry, F. H.; Littlechild, J. A. & Jacob, C. (2003) **Metal and redox modulation of cysteine protein function.** *Chem Biol.* (8): 677-693

Givot, I. L.; Smith, T. A. & Abeles, R. H. (1969) **Studies on the mechanism of action and the structure of the electrophilic center of histidine ammonia lyase.** *J Biol Chem*. 244(23): 6341-6353

Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, T. & Altman, S. (1983) **The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme**. *Cell.* 35: 849–857

Hanczyc, M. M.; Mansy, S. S. & Szostak, J. W. (2007) **Mineral surface directed membrane assembly.** *Orig Life Evol Biosph.* 37(1): 67-82

Handler, P. (1963) **Evolution of the coenzymes**. En: Oparin, A. I. (Ed) Proceedings of the Fifth International Congress of Biochemistry, Vol. III. Biochemistry. Pergamon Press/Macmillan Company, 149-157

Hasegawa, H. & Holm, L. (2009) **Advances and pitfalls of protein structural alignment.** *Curr. Opin. Struct. Biol.* 19: 341-348

Hernández-Molares, R.; Becerra, A. & Lazcano, A. (2019) **Alarmones as vestiges of a bygone RNA World.** *J Mol Evol.*87(1):37-51

Hodgins, D. S. (1971) **Yeast phenylalanine ammonia-lyase. Purification, properties, and the identification of catalytically essential dehydroalanine.** *J Biol Chem.* 246(9): 2977-2985

Holliday, G. L.; Almonacid, D. E.; Mitchell, J. B. & Thornton, J. M. (2007) **The chemistry of protein catalysis.** *J Mol Biol.* 372(5): 1261-1277

Holliday, G. L.; Fischer, J. D.; Mitchell, J. B. & Thornton, J. M. (2011) **Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis.** *FEBS J*. 278(20): 3835-45

Hsiao, C.; Chou, I.-C.; Okafor, C. D.; Bowman, J. C.; O'Neill, E. B.; Athavale, S. S.; Petrov, A. S.; Hud, N. V.; Wartell, R. M.; Harvey, S. C. & Williams, L. D. (2013)

RNA with iron(II) as a cofactor catalyses electron transfer. *Nat. Chem.* 5: 525-528

Jácome, R.; Becerra, A.; Ponce de León, S. & Lazcano, A. (2015) **Structural analysis of monomeric RNA-dependent polymerases: evolutionary and therapeutic implications.** *PLoS One*. 10(9): e0139001

Jadhav, V. R. & Yarus, M. (2002) **Coenzymes as coribozymes.** *Biochimie.* 84(9): 877-88

Joyce, G. F. (2002) **The antiquity of RNA-based evolution.** *Nature.* 418: 214–221

Kack, H.; Sandmark, J.; Gibson, K. J.; Schneider, G. & Lindqvist, Y. (1998) **Crystal structure of two quaternary complexes of dethiobiotin synthetase, enzyme-MgADP-AlF3-diaminopelargonic acid and enzyme-MgADP-dethiobiotin-phosphate.** *Protein Sci.* 7: 2560-2566

Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y. & Morishima, K. (2017) **KEGG: new perspectives on genomes, pathways, diseases and drugs.** *Nucleic Acids Res*. 45(D1): D353-D361

Kawamura, K.; Takeya, H.; Kushibe, T. & Koizumi, Y. (2011) **Mineral-enhanced hydrothermal oligopeptide formation at the second time scale.** *Astrobiology.* 11: 461-469

Keller, E. B; Zamecnik, P. C & Loftfield, R. B (1954) **The role of microsomes in the incorporation of amino acids into proteins.** *J Histochem Cytochem.* 1954 Sep;2(5): 378-86

Krissinel, E. & Henrick, K. (2004). **Secondary-structure matching (PDBeFold), a new tool for fast protein structure alignment in three dimensions.** *Acta Cryst.* D60: 2256-2268

Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E. & Cech, T. R. (1982) **Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.** *Cell*. 31: 147–157

Lazcano A. & Miller S.L. (1994) **How long did it take for life to begin and evolve to cyanobacteria?** *J Mol Evol*. 39(6):546-554

Lazcano A. & Miller S.L. (1996) **The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time.** *Cell*. 85(6):793-798

Lazcano, A. & Miller, S. L. (1999) **On the origin of metabolic pathways.** *J Mol Evol.* 49(4): 424-31

Lazcano, A. (2010) **Historical development of origins research.** *Cold Spring Harb Perspect Biol.* 2(11): a002089

Li, C.; Kappock, T. J.; Stubbe, J.; Weaver, T. M. & Ealick, S. E. (1999) **X-ray crystal structure of aminoimidazole ribonucelotide synthetase (PurM), from the Escherichia coli purine biosynthetic pathway at 2.5 Å resolution.** *Structure.* 7: 1155-1166

Liao, S. M.; Du, Q. S.; Meng, J. Z.; Pang, Z. W. & Huang, R. B. (2013) **The multiple roles of histidine in protein interactions.** *Chem Cent J*. 7(1):44

Louie, G. V.; Bowman, M. E.; Moffitt, M. C.; Baiga, T. J.; Moore, B. S. & Noel, J. P. (2006) **Structural determinants and modulation of substrate specificity in phenylalanine-tyrosine ammonia-lyases.** *Chem Biol.* (12): 1327-1338

Lowe, C.U.; Ress, A & Markham, F. R. S. (1963) **Synthesis of complex organic compounds from simple precursors: formation of amino-acids, amino-acid polymer, fatty acids and purines from ammonium cyanide**. *Nature*. 19:219–222

Lowenstein, J. M. (1958) **Transphosphorylations catalysed by bivalent metal ions.** *Biochem J.* 70: 222-230

Martin, L.L.; Unrau, P. J. & Müller, U. F. (2015) **RNA synthesis by in vitro selected ribozymes for recreating an RNA world.** *Life* (Basel). 5(1):247-268

McGhee, G. (2011) **Convergent evolution: limited forms most beautiful.** MIT Press

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

Mizuhara, S. & Handler, P. (1954) **Mechanism of thiamine-catalyzed reactions**. *J. Am. Chem. Soc.* 76(2): 571–573

Moore, P. B. & Steitz, T. A. (2011) **The roles of RNA in the synthesis of protein**. *Cold Spring Harb Perspect Biol.* 3(11):a003780

Nissen, P.; Hansen, J.; Ban, N.; Moore, P.B. & Steitz, T.A. (2000) **The structural basis of ribosome activity in peptide bond synthesis.** *Science*. 289: 920-30

Noller, H. F. (2012) **Evolution of protein synthesis from an RNA world.** *Cold Spring Harb Perspect Biol.* 4(4): a003681

Noller, H.F. & Chaires, J.B. (1972) **Functional modification of 16S ribosomal RNA by kethoxal.** *Proc. Nat. Acad. Sci., USA.* 69: 3115-3118

Ogino, H.; Inomata, S. & Tobita, H. (1998) **Abiological iron sulfur clusters.** *Chem Rev.* 98(6): 2093-2122

Oparin, A. I. (1924) **The origin of life.** En: Bernal, J. D. (Ed.) (1967). Weidenfeld and Nicolson

Oparin, A. I. (1938) **The origin of life**. MacMillan, New York

Orgel, L. E. & Lohrmann, R. (1974) **Prebiotic chemistry and nucleic acid replication.** *Acc Chem Res.* 7: 368-377

Orgel, L. E. (1968) **Evolution of the genetic apparatus.** *J. Mol. Biol*. 38: 381-393

Oró, J. (1960) **Synthesis of adenine from ammonium cyanide**. *Biochem Biophys Res Commun.* 2:407-412

Oró, J. (1963) **Synthesis of organic compounds by electric discharges.** *Nature*. 197:862-67

Oró, J.; Basile, B.; Cortes, S.; Shen, C. & Yamrom, T. (1984) **The prebiotic synthesis and catalytic role of imidazoles and other condensing agents**. *Orig Life.* 14(1-4):237-42

Palade G. E. (1955**). A small particulate component of the cytoplasm.** *J Biophys Biochem Cyto*l.;1(1):59-68

Parker, E T.; Cleaves, H. J.; Callahan, M. P.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A. & Bada, J. L. (2011a) **Prebiotic synthesis of methionine and other sulfur-containing organic compounds on the primitive Earth**. *Orig Life Evol Biosph.* 41(3): 201–21

Parker, E. T.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Callahan, M.; Aubrey, A.; Lazcano, A. & Bada, J. L. (2011b) **Primordial synthesis of amines and**

amino acids in a 1958 Miller H2S-rich spark discharge experiment. *PNAS.* 108(14): 5526-31

Peretó, J.; Bada, J.L. & Lazcano, A. (2009) **Charles Darwin and the origin of life.** *Orig Life Evol Biosph*. 39(5):395-406

Perrotta, A. T.; Shih, I. & Been, M. D. (1999) **Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme.** *Science*. 286 (5437):123-126

Reardon, D. & Farber, G. K. (1995) **The structure and evolution of alpha/beta barrel proteins**. *FASEB J.* (7): 497-503

Reider, G.; Merrick, M. J.; Castorph, H. & Kleiner, D. (1994) **Function of *hisF* and *hisH* products in histidine biosynthesis**. *J Biol Chem.* 269: 14386–14392

Rétey, J. (2003**) Discovery and role of methylideneimidazolone, a highly electrophilic prosthetic group.** *Biochim Biophys Acta* 1647(1-2): 179-184

Ribeiro, A. J. M.; Holliday, G. L.; Furnham, N.; Tyzack, J. D.; Ferris, K. & Thornton, J. M. (2018) **Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites.** *Nucleic Acids Res.* 46(D1): D618-D623

Rich, A. (1962) **On the problems of evolution and biochemical information transfer**. En: Kasha, M. & Pullman, B. (Eds) Horizons in biochemistry. New York: Academic Press. 103-126

Rivalta, I.; Sultan, M. M.; Lee, N. S.; Manley, G. A.; Loria, J. P. & Batista, V. S. (2012) **Allosteric pathways in imidazole glycerol phosphate synthase**. *PNAS.* 109(22): E1428-36

Roberts, R. B., ed. (1958) **Microsomal particles and protein synthesis**. First Symposium of the Biophysical Society. Pergamon Press, New York

Robertson, M. P. & Joyce, G. F. (2012) **The origins of the RNA world.** *Cold Spring Harb Perspect Biol.* 4(5). pii: a003608

Sahai, N.; Kaddour, H. & Dalai, P. (2016) **Transition from geochemistry to bio(geo)chemistry: outlining the problem of the origins of life.** *Elements.* 12: 389-394

Schlesinger, G. & Miller, S. L. (1983) **Prebiotic synthesis in atmospheres containing CH4, CO, and CO2**. *J Mol Evol*. 19(5): 376-382

Schluenzen, F.; Tocilj, A.; Zarivach, R.; Harms, J.; Gluehmann, M.; Janell, D.; Bashan, A.; Bartels, H.; Agmon, I.; Franceschi, F. & Yonath, A. (2000). **Structure of functionally activated small ribosomal subunit at 3.3 Å resolution.** *Cell*. 102: 615-623

Schwartz, A. W. & Bakker, C. G. (1989) **Was adenine the first purine?** *Science*. 245:1102–1104

Schwede, T. F.; Rétey, J. & Schulz, G. E. (1999) **Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile.** *Biochemistry*. 38(17): 5355-5361

Shen, C.; Yang, L.; Miller, S. L. & Oró, J. (1987**) Prebiotic syntheses of imidazole-4-acetaldehyde and histidine.** *Orig Life*. 17: 295-305

Shen, C.; Yang, L.; Miller, S. L. & Oró, J. (1990) **Prebiotic synthesis of histidine.** *J Mol Evol.* 31: 167–174

Sillitoe, I. & Orengo, C. (2007) **Protein structure comparison**. En: Orengo C.; Jones, D. & Thornton, J. (Eds) Bioinformatics: genes, proteins, and computers. Oxford

Steenken, S. (1989) **Purine bases, nucleosides, and nucleotides: aqueous solution redox chemistry and transformation reactions of their radical cations and e- and OH adducts**. *Chem. Rev*. 89(3): 503–520

Subbiah, S.; Laurents, D. V. & Levitt, M. (1993) **Structural similarity of DNA-binding domains of bacteriophage repressors and the globin core.** *Curr Biol.* (3): 141-8

Tirard, S. (2010) **Origin of life and definition of life, from Buffon to Oparin.** *Orig Life Evol Biosph*. 40: 215–220

Todd; A. E.; Orengo, C. A. & Thornton, J. M. (1999) **Evolution of protein function, from a structural perspective.** *Curr Opin Chem Biol.* (5): 548-56

Vázquez-Salazar, A. & Lazcano, A. (2018) **Early life: embracing the RNA World**. *Curr Biol* 5:220–222

Vázquez-Salazar, A.; Becerra, A. & Lazcano, A. (2018) **Evolutionary convergence in the biosyntheses of the imidazole moieties of histidine and purines**. *PLoS ONE* 13: e0196349

Vázquez-Salazar, A.; Becerra, A. & Lazcano, A. (*In prep*) **Scaffolding the RNA world.**

Vázquez-Salazar, A; Tan, G; Stockton, A; Fani, R; Becerra, A & Lazcano A (2017) **Can an imidazole be formed from an alanyl-seryl-glycine tripeptide under possible prebiotic conditions?** *Orig Life Evol Biosph* 47:345–354

Visser, C. M. & Kellogg, R. M. (1978) **Biotin. Its place in evolution.** *J Mol Evol.* 20; 11(2): 171-87

Walker, K. D.; Klettke, K.; Akiyama, T. & Croteau, R. (2004) **Cloning, heterologous expression, and characterization of a phenylalanine aminomutase involved in taxol biosynthesis**. *J Biol Chem.* 279(52): 53947-53954

Weiss, M. C.; Preiner, M.; Xavier, J. C.; Zimorski, V. & Martin, W . F. (2018) **The last common ancestor between ancient Earth chemistry and the onset of genetics.** *PLoS Genet.* 14:e1007518

White III, H. B (1976) **Coenzymes as fossils of an earlier metabolic state.** *J Mol Evol.* 7: 101-104

White III, H. B. (1982) **Evolution of coenzymes and the origin of pyrimidine nucleotides**, En: Everse, J.; Anderson, M. B. & You, K. (Eds) The pyrimidine coenzymes. New York: Academic Press

Wieczorek, R.; Dörr, M.; Chotera, A.; Luisi, P. L. & Monnard, P. A. (2013) **Formation of RNA phosphodiester bond by histidine-containing dipeptides.** *Chembiochem*. 14(2): 217-23.

Wilson, D. S. & Szostak, J. W. (1999) **In vitro selection of functional nucleic acids.** *Annu Rev Biochem.* 68: 611-47

Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T. & Ramakrishnan, V. (2000) **Structure of the 30S ribosomal subunit.** *Nature*. 407: 327-39.

Winkler, M. E. & Ramos-Montañez, S. (2008) **Biosynthesis of histidine.** *EcoSal Plus*. 3(2): 1-34

Woese, C. R. (1965) **On the evolution of the genetic code.** *Proc Natl Acad Sci U S A.* 54(6): 1546–1552

Woese, C. R. (1967) **The genetic code: The molecular basis for genetic expression.** Harper & Row, New York

Wright, C. S.; Alden, R. A. & Kraut, J. (1969) **Structure of subtilisin BPN' at 2.5 angström resolution**. *Nature*. 221: 235-242

Yonath, A. (2004) **Ribosomal crystallography: dynamics, flexibility and peptide bond formation**, En: Cheng, R. H. & Hammar, L. (Eds) Conformational proteomics of macromolecular architecture. World Scientific Publishing

Yonath, A.; Muessig, J.; Tesche, B.; Lorenz, S.; Erdmann, V. A. & Wittmann, H.G. (1980). **Crystallization of the large ribosomal subunit from B. stearothermophilus**. *Biochem Int*. 1, 428-35