



29
20

UNIVERSIDAD NACIONAL AUTÓNOMA
DE MEXICO

FACULTAD DE CIENCIAS

LA SECUENCIA Gly-Gly-Arg-Glu-Gly-Leu-Val-Asp: UN
NUEVO SITIO FUNCIONAL CONSERVADO EN TODAS
LAS RNA POLIMERASAS DNA-DEPENDIENTES
CELULARES EN LA SUBUNIDAD β' DE *E. coli*
Y SUS HOMOLOGAS,

T E S I S

QUE PARA OBTENER EL TITULO DE:

B I O L O G O

P R E S E N T A:

RENATO EUGENIO CAPPELLO GONZALEZ

MEXICO, D. F.

NOVIEMBRE DE 1992

TESIS CON
FALLA DE ORIGEN



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

ÍNDICE

RESUMEN

1. INTRODUCCIÓN	1
1.1 EVOLUCIÓN BIOLÓGICA Y ENTIDADES HISTÓRICAS	1
1.2 LA VIDA COMO ENTIDAD HISTÓRICA	2
1.3 LAS MOLÉCULAS COMO REGISTRO HISTÓRICO	4
1.4 ¿LAS PROTEÍNAS HACEN HONOR A SU ASCENDENCIA?	13
1.5 PROPIEDADES EMERGENTES DE LAS PROTEÍNAS	15
1.6 LA ESTRUCTURA Y ORGANIZACIÓN DE LAS PROTEÍNAS	16
1.7 DOMINIOS Y SITIOS CONSERVADOS: "LAS PROTEÍNAS SE FORMAN EN MOSAICO"	23
1.8 SITIOS QUE SE UNEN A NUCLEÓTIDOS Y/O POLINUCLEÓTIDOS	28
 2.0 EVOLUCIÓN DE LAS RNA-POLIMERASAS	 38
2.1 EL MUNDO DEL RNA	39
2.2 LOS PRIMEROS PÉPTIDOS Y EL GENOMA	40
2.3 LA EVOLUCIÓN DE LAS POLIMERASAS Y EL RNA	43
2.4 EVOLUCIÓN DE β' Y SUS DOMINIOS VESTIGIALES	44
 3.0 LA PROTEÍNA RIBOSOMAL S1	 51
 4.0 LA PROTEÍNA QUE POLIMERIZA AL AZAR: PNPasa	 59
 5.0 MATERIAL Y MÉTODO	 67
 6.0 RESULTADOS	 74
 7.0 DISCUSIÓN Y CONCLUSIONES	 83
 8.0 REFERENCIAS	 95

Resumen

En este trabajo se describe la presencia de un motivo muy conservado, formado por la secuencia GGREGLVD, en la subunidad β' de la RNA polimerasa DNA-dependiente de *E. coli* y sus homólogas en otros organismos. Dicho motivo forma parte de una secuencia previamente descrita en la secuencia de la PNPasa de *E. coli* y en cuatro dominios de la proteína ribosomal S1 de eubacterias (Régner et al., 1987). Debido a que esta secuencia forma parte del sitio funcional que une el RNA a la PNPasa (Régner et al., 1987), es posible que su presencia en las RNA polimerasas celulares se deba a restricciones funcionales equivalentes. La comparación de las secuencias de RNA polimerasas, PNPasa y la Proteína S1 ribosomal, con la estructura primaria de la subunidad β de las ATPasas celulares, reveló la existencia de una secuencia homóloga. Ello apoya la hipótesis de que el motivo GGREGLVD es parte de un sitio de unión a ribonucleótidos, ya sea aislados o formando parte de una cadena.

1.0 INTRODUCCIÓN

1.1 EVOLUCIÓN BIOLÓGICA Y ENTIDADES HISTÓRICAS

El hombre siempre ha querido explicarse cuál es su origen (de dónde viene), cómo es que se encuentra en este planeta junto con todos los organismos que lo acompañan. Estas son preguntas que en los últimos 200 años se han transformado en ¿cómo apareció la vida? y ¿por qué permanecemos?. Gracias a nuevas concepciones evolutivas, nos hemos dado cuenta de que las entidades biológicas se caracterizan por el cambio y la continuidad, lo que resulta en una increíble diversidad biológica.

Estas tres nociones sencillas, cambio, continuidad y diversidad, rompen con el pensamiento que a menudo permea otras disciplinas científicas, tales como la física y la química donde tradicionalmente se proponen tanto constancia como discontinuidad. El estudio comparativo de las entidades biológicas ha permitido el reconocimiento no sólo de la tremenda diversidad, sino también de la similitud que guardan muchos organismos entre sí. El reconocimiento de esta similitud permitió establecer el concepto de homología, y reconocer las entidades biológicas como entidades históricas, apartir de la teoría de la evolución.

1.2 LA VIDA COMO ENTIDAD HISTÓRICA

En la actualidad los organismos se conciben como entidades que cambian y que son sujeto de la evolución biológica. ¿Cómo podemos reconocer que un organismo tiene características históricas?. ¿De qué manera se puede observar cómo ha sido el cambio de las entidades biológicas a través del tiempo?. El registro fósil es una de las fuentes que ha permitido deducir que, efectivamente, los organismos han sufrido cambios a lo largo del tiempo. De esta manera se ha establecido el parentesco que guardan muchas de las especies actuales, pudiéndose determinar la genealogía de éstas, cuándo existieron y cómo fueron evolucionando.

Otra de las disciplinas esenciales para determinar las relaciones evolutivas entre los organismos, es la anatomía comparada, con la cual, a partir de individuos actuales, se pueden reconocer similitudes que sugieren una ascendencia común entre los organismos en cuestión. Lamarck pudo hacer este tipo de observaciones en la colección de invertebrados que estaba a su cargo (Moulton y Schifferes, 1947), lo que le permitió darse cuenta de la continuidad y ascendencia común que existe entre todos los organismos.

En la actualidad se pretende establecer cuáles de las estructuras o funciones son homólogas, es decir, tienen un origen común y cuáles, por lo contrario son análogas. Reconocer la homología en

los diferentes niveles biológicos provee los datos primarios para hacer de la biología una ciencia histórica (Gould, 1986). Un ejemplo de estructuras homólogas son las alas de los murciélagos y las manos del hombre; en contraparte, las alas de los murciélagos y las de los insectos son análogas. En la biología existe una infinidad de ejemplos similares. La importancia del concepto de homología es que permite establecer descendencia u origen común, con lo cual se ha podido concluir que todos los organismos actuales descienden de un ancestro común (Darwin, 1859; Oparin, 1938; Woese et al., 1990); es decir, la vida es un continuo.

Una de las mayores aportaciones de Darwin a la biología fue el haber planteado una vía para entender cómo ocurren los cambios organizados en las especies, al proponer a la selección natural como un mecanismo para la transmutación de las especies (Huxley, 1983). En términos de la síntesis moderna, ello se traduce como "la fuerza constructora en la evolución es la selección natural, que actúa sobre la variación que es heredada por las leyes de Mendel y mantenida en la escala del tiempo por mutaciones aleatorias" (Huxley, 1983). No obstante, en la actualidad surgen escuelas estructuralistas no-darwinistas que pretenden explicar la evolución sin el componente histórico, estas postula que los fenómenos biológicos evolucionan de una estructura determinada de dicho fenómeno, para dilucidar el origen de las especies. (Lambert y Hughes et al., 1988).

Otra aportación tal vez mucho más profunda de Darwin para el futuro de la biología, fue el haber fundado un sistema metodológico comparativo que permitiese estudiar científicamente la historia de la vida, así como la evolución biológica y por ende sus entidades (Gould, 1986). Es decir, permitió establecer un programa de investigación, el cual podemos utilizar para hacer estudios evolutivos de las diferentes jerarquías de las entidades históricas de la vida. Así, extrapolando sus efectos a pequeña escala, los fenómenos pueden ser observados e investigados para intentar inferir su historia. Una de las jerarquías en donde se puede aplicar este programa es a nivel molecular (Zuckerkandl y Pauling, 1962, 1965). Con la Biología Molecular se ha podido establecer una fuente de datos apropiada para hacer comparaciones e intentar establecer homologías, lo cual permite deducir o reconstruir la historia evolutiva de las entidades biológicas (Gould 1986).

1.3 LAS MOLÉCULAS COMO REGISTRO HISTÓRICO

En el estudio de la evolución biológica y molecular se sigue una metodología similar a la que aplican los historiadores. En ambos casos, a partir de una serie de fenómenos y hechos se realiza una reconstrucción histórica. Cuando un cronista investiga un hecho histórico, lo primero que hace es plantearse preguntas al respecto; después, para responder a estas interrogantes, indaga sobre todo lo que rodea al hecho, consultando las fuentes del mismo, es decir, los diferentes registros o legados, generados

por el acontecimiento, a fin de documentar las circunstancias del hecho. Lo importante es que en estos documentos ha quedado plasmada la historia de la humanidad a través del tiempo. Los documentos emitidos permiten narrar e interpretar la sucesión de los acontecimientos.

En sentido estricto, la evolución siempre es divergente en el tiempo; en un sentido más laxo, la evolución ha tenido convergencias, tanto en sus ramas como en algunas estructuras de los organismos. El primer caso se ejemplifica en las ocasiones en que ocurre simbiosis; el segundo con estructuras análogas como las alas de insectos, aves y mamíferos. La divergencia crea a las novedades evolutivas ya sea a través del cambio de una especie a otra (transmutación) o por medio de la simbiosis, pues existen casos donde distintas ramas del árbol de la vida se juntan en una sola, como ocurrió con el surgimiento de los eucariontes. En biología existen diversas jerarquías o niveles de organización, que van desde los ácidos nucleicos hasta un ecosistema (Gould y Eldredge, 1977; Mayr, 1982). Cada nivel tiene diferentes características, y cada una de ellas tiene un valor diferente al usarse para el estudio de la historia evolutiva de las entidades biológicas. Estos principios son aplicables al material de la herencia como son los ácidos nucleicos (DNA, RNA), y a los productos codificados por ellos como las proteínas.

El inicio de la biología molecular se inauguró con tres acontecimientos importantes en la década de los 50's. El primero

de ellos fue la deducción por parte de L. Pauling de que las proteínas contienen al menos dos tipos de estructuras, las hélices α y las hojas plegadas β (Von Heijne, 1987). Poco después Sanger secuenció la insulina (Sanger y Tuppy 1951) y por último, la elucidación de la estructura del DNA por Watson y Crick (1953).

El conocimiento de la organización de las macromoléculas mostró claramente que ni las proteínas ni los ácidos nucleicos presentan una secuencia generada al azar, sino que tienen una estructura primaria determinada. Esto permitió concebir a las macromoléculas como estructuras organizadas que han sido modificadas durante la evolución biológica. Las distintas biomoléculas fueron reconocidas como estructuras capaces de almacenar la historia evolutiva de las entidades biológicas (Zuckerklund y Pauling, 1962; 1965). Por la misma época, Crick (1958) propuso que algún día se podría llegar a hacer taxonomía de las biomoléculas basándose en patrones estructurales organizados.

Para los biólogos las macromoléculas biológicas (ácidos nucleicos y proteínas) son fuentes documentales, análogas a los que usan los historiadores. En ellas queda marcada al menos parte de la historia del organismo. Cada documento molecular guarda o perpetúa partes diferentes de la historia evolutiva de las entidades biológicas. Es decir, extrae de distintas moléculas mayor o menor información histórica sobre un evento evolutivo en particular. Sin embargo, hay preguntas que no pueden ser

contestadas por ciertas moléculas, pues no todas poseen las características que les permitan dar luz sobre las mismas. ¿Cómo pueden las macromoléculas ser utilizadas como documentos históricos? Zuckerkandl y Pauling (1962, 1965) propusieron extraer de las biomoléculas la historia evolutiva. Para ello clasificaron cada grupo de moléculas a partir de su valor como marcador evolutivo, que quedaron divididos en moléculas asemánticas, episemánticas y semánticas.

Las moléculas asemánticas son aquéllas que no son producidas por el organismo, por lo que no expresan ni directa ni indirectamente la información que los seres vivos contienen. Son moléculas como las vitaminas, los fosfatos, el oxígeno, CO_2 , Fe^{++} , los cationes y los aniones elementales, etc., que utiliza el organismo para llevar a cabo sus funciones metabólicas.

Las moléculas episemánticas se sintetizan bajo el control de semánticas terciarias. Todas las moléculas construídas por enzimas en la ausencia de un molde, se encuentran en esta clase. Se les denomina episemánticas porque aún cuando no poseen la información contenida en las semánticas, sí son productos de éstas últimas. Este grupo incluye, por ejemplo, a los carotenoides, a la celulosa, al ATP, carbohidratos, etc.

Dentro de las moléculas semántidas quedan comprendidos el DNA (semántidas primarias), RNA (semántidas secundarias) y las proteínas (semántidas terciarias).

La clasificación anterior muestra el valor evolutivo que ofrece cada categoría. Las semántidas guardan y mantienen el significado informativo del mensaje genético; por ende, son las que mejor representan las relaciones evolutivas entre las biomoléculas. Las semántidas primarias y secundarias corresponden al material genético *per se*, mientras que las terciarias comprenden al efector que se codifica a partir de las dos anteriores.

En las semánticas terciarias existen diferentes niveles a los que se les puede extraer información evolutiva. Las proteínas tienen cuatro niveles estructurales: a) la estructura primaria (que es la secuencia lineal de aminoácidos), b) la secundaria (que son las unidades formadas a partir de las interacciones de aminoácidos cercanos como las alfa hélices y las hojas β plegadas); c) la terciaria (es decir, las que resultan de las interacciones lejanas de los aminoácidos, lo que forma la estructura tridimensional de las proteínas) y d) la cuaternaria (que se refiere a la interacción de varias moléculas organizadas en una sola estructura).

Por otro lado, en las semántidas secundarias (RNA) también existen cuatro niveles de estructuras: la primaria, que es la secuencia de nucleótidos; la secundaria, que resulta del plegamiento de la molécula por la complementariedad entre sus distintas bases; la terciaria, que es la estructura tridimensional que adquiere la molécula de RNA al plegarse en el espacio y, finalmente, la cuaternaria, formada por complejos de

diferentes moléculas de RNA o de RNA con proteínas, como sería los casos del ribosoma, de la RNasa P y los U1 SnRNPs (Busch et al., 1982).

Las características que permiten utilizar a ciertas biomoléculas como documentos históricos incluyen tanto su papel de transmisoras de la información a través del tiempo, así como su distribución biológica. Las semántidas son las que mejor cumplen estos requisitos, lo que les permite servir como base para reconstruir la historia evolutiva de un grupo de organismos o taxa. Las semántidas terciarias presentan diversos niveles de complejidad, cada uno de los cuales proporcionan mayor o menor información evolutiva. Por ejemplo, las proteínas estructurales presentan repeticiones internas de un mismo módulo, a diferencia de las proteínas globulares que presentan una mayor diversidad de módulos (Zuckerkandl y Pauling 1962, 1965). En este caso el número de repeticiones o de no repeticiones afecta directamente la cantidad de información disponible, que será obtenida a partir de la secuencia de los motivos y de las veces que la repetición haya ocurrido.

La selección y la deriva génica hacen que los polímeros sufran constantes sustituciones de sus monómeros, por lo que la secuencia original se modifica. Surgen así variaciones entre las que se pueden comparar sus similitudes y/o diferencias.

Las episemántidas tienen un valor diferente en la reconstrucción de la historia evolutiva. La información que se les puede extraer depende de su contexto biológico, por lo que las episemántidas

permiten completar partes de un árbol genealógico, pero no pueden ser la base a partir de la cual se deduzca una filogenia. Su utilidad queda demostrada en el caso de las bacterias fotosintéticas, donde usando los carotenoides se ha podido refinar una rama del árbol (*In*: Zuckerkandl y Pauling, 1965). Este tipo de estudios requieren de una confirmación independiente como podría ser el uso de una semántida tal como el rRNA o los citocromos (Zuckerkandl y Pauling, 1965).

Algunas de las moléculas episemántidas pueden ser universales, como el ATP, aunque también ocurre lo contrario como sustancias que se encuentran sólo en un organismo. Tienen distintos grados de complejidad, y en algunos organismos dicha complejidad puede ser una fuente de información, pues para producir una molécula episemántida intervienen varias enzimas.

Por último, las asemántidas son moléculas que no nos dicen mucho acerca de la historia evolutiva, pues prácticamente no guardan información biológica, no presentan variación y tienen poca complejidad.

Del análisis anterior se desprende que las moléculas pierden relevancia evolutiva, es decir, que ésta decrece, conforme se al siguiente esquema:

semántidas -----> episemántidas -----> asemántidas.

Las proteínas y los genes que las codifican, sirven como herramientas para la reconstrucción filogenética y, en algunos casos al completar la información con el registro fósil, para fechar los eventos de divergencia de los grupos biológicos (Zuckerkandl y Pauling 1962, 1965). A continuación se enumeran algunas características que pueden deducirse a partir de la comparación de las secuencias de monómeros. Al realizar las comparaciones se buscan sitios conservados que muestren similitud significativa entre dos o más polipéptidos que permitan establecer:

- 1) La descendencia común entre dos o varias moléculas; 2) el cálculo del momento aproximado en el que ocurrió la divergencia a partir del último ancestro molecular común; 3) la probable secuencia del ancestro; 4) las líneas de descendencia que surgieron por el cambio de monómeros; 5) la identificación de los sitios funcionales y/o estructurales importantes; 6) los módulos y motivos comunes a proteínas con función y/o origen distintos.

Se debe tomar en cuenta que dentro de los diferentes tipos de proteínas que hay en los organismos, cada proteína tiene un valor diferente como marcador evolutivo, por dos razones. En la primera existen diferencias que se presentan entre las proteínas debido a su historia evolutiva, en cuanto a su función y al momento en que surgen; es decir, hay proteínas más antiguas que otras y ello puede limitar o no su uso en la reconstrucción filogenética. La segunda razón está implícita en la hipótesis

del reloj molecular, planteada por Zuckerkandl y Pauling (1962), y que lleva al establecimiento de una relación entre el tiempo y el número de sustituciones. Esto supone que la tasa de sustituciones es constante y que la mayoría de estos cambios son neutros. Sin embargo, al ponerse a prueba esta hipótesis, se ha observado que en cierto casos el reloj se acelera y en otros se desacelera. Así, la tasa de cambio no puede ser constante si se tienen diferentes ritmos de cambio, por lo que en la actualidad, rara vez se utiliza como marcador para fechar eventos evolutivos (Melnick, 1990; Behe, 1990).

La hipótesis del reloj molecular que observa las diferentes tasas de sustitución que presentan las proteínas en sus monómeros, ayuda a valorar qué tan importante es una proteína para una pregunta evolutiva específica, es decir, existen proteínas que son más adecuadas que otras para establecer una divergencia, pues hay moléculas que tienden a cambiar rápidamente impidiendo relacionar a dos proteínas, como ocurre con los fibrinopéptidos. En el caso contrario, cuando una proteína es muy lenta en su tasa de cambio, no podría registrar las diferencias entre dos proteínas de organismos muy cercanos filogenéticamente, como ocurre con las histonas.

Finalmente, diremos que las técnicas de evolución molecular presentan dos ventajas: 1) permite comparar filogenéticamente a organismos pertenecientes a taxa muy lejanos, como aves, hongos y eubacterias (lo cual sería imposible con la morfología); y 2)

permite hacer un análisis más objetivo, que la anatomía comparada o el registro fósil, debido a que los estudios son de carácter cuantitativo, siempre y cuando se tenga un polipéptido lo suficientemente largo para elaborar la estadística correspondiente, pues de lo contrario, el análisis para polipéptidos más cortos requiere elementos diferentes al estadístico por lo cual se aplican criterios subjetivos.

1.4 ¿LAS PROTEÍNAS HACEN HONOR A SU ASCENDENCIA?

En los organismos actuales cada gen de cada proteína tiene un retrospectiva particular, es decir, cada biomolécula tiene una diferente. El legado de estas moléculas se delimita, por el momento de su surgimiento evolutivo, como por las formas de cambio en su secuencia de nucleótidos y en la de aminoácidos en el tiempo. Estos cambios son dependientes de las restricciones funcionales que existen sobre el gen *per se*, en su producto (el papel funcional que juega en la célula) y de las variaciones aleatorias en la secuencia del gen. Dichos cambios quedarían contenidos dentro de la biomolécula, de esta manera se lograría deducir la historia de la molécula, que a su vez reflejaría incierta mediada la de los organismos.

La riqueza que tiene cada biomolécula como marcador de la historia evolutiva de los sistemas biológicos, dependerá de su retrospectiva y de su capacidad de mantener o perder sus rasgos y características ancestrales (hasta el límite de transmutarse)

esto último debido a la tasa de cambio, lenta o acelerada, que tiene una secuencia.

Al seleccionar una molécula en particular se elige, de alguna manera, el reloj de una molécula cuya tasa de cambio se ajuste a cierto evento evolutivo, está es una retrospectiva adecuada. En consecuencia, la semántica conservaría los dominios funcionales que se ajustarán a los eventos de la historia evolutiva de los organismos que se pretenda estudiar. Se construye así una crónica aproximada de la evolución molecular y biológica, pero también se puede correr el riesgo de terminar relatando más bien una aventura de ciencia ficción.

El alcance que tiene una molécula para mostrar los eventos del pasado tiene límites. Cada molécula tiene diferentes posibilidades para mostrar el pasado cercano, lejano, y muy lejano. Si se desea establecer una serie de eventos evolutivos lejanos se debe optar por una molécula conservada evolutivamente que posea una distribución generalizada dentro de los taxa en cuestión. Existen moléculas que pueden perder rápidamente las huellas del origen común en la secuencia primaria, pero no necesariamente en otros niveles estructurales. Por otra parte, existen sistemas biológicos que se prestan con mayor facilidad para tener un reloj molecular acelerado, como los virus de RNA, en los cuales la RNA polimerasa RNA dependiente no tiene capacidad correctora, por lo que acumula más mutaciones.

Por ejemplo, al comparar las diferentes secuencias de citocromo b, se puede relacionar a organismos filogenéticamente muy lejanos que utilizan la cadena respiratoria para producir ATP, debido a que se trata de una proteína de grandes dimensiones con una tasa de cambio relativamente lenta. En cambio, si tomamos a una gen que tiene una tasa de sustitución más rápida, como los fibrinopéptidos, será difícil encontrar la similitud y por ende la ascendencia común en organismos cercanos (Doolittle, 1988).

1.5 PROPIEDADES EMERGENTES DE LAS PROTEÍNAS

De manera similar a lo que ocurre en todos los niveles de organización biológicos, donde las propiedades de un nivel no se definen por la suma de sus componentes sino por la emergencia de éstos, las proteínas poseen características propias que no se determinan únicamente por la suma de sus monómeros, siendo sus propiedades químicas mucho más amplias.

Si se pudieran definir las propiedades de las proteínas tomando en cuenta únicamente sus aminoácidos, entonces, tanto la estructura terciaria como las funciones de una proteína se podrían deducir fácilmente a partir de la estructura primaria lo cual sigue siendo imposible. Las propiedades de un aminoácido cualquiera son de un tipo. No obstante, al hallarse en una secuencia, dichas propiedades pueden ser modificadas por el ambiente (Williams, 1985). La importancia del encadenamiento de los aminoácidos en una secuencia radica en la combinación de los

misimos, lo cual genera alternativas en las propiedades químicas. Como consecuencia se conforma una química diferente de la que se comienzan a dilucidar las reglas que determinan la función biológica de estas moléculas. Además, existen otros elementos que contribuyen a la química de las biomoléculas como las porfirinas, los metales, las coenzimas, etc. En la biosíntesis y plegamiento intervienen otras moléculas, como las chaperoninas, que son determinantes para que las proteínas adquieran su conformación funcional.

Una de las principales propiedades de las proteínas es la capacidad que tienen sus secuencias de aminoácidos para plegarse, adoptando diferentes formas que se traducen en estructuras complejas, capaces de llevar a cabo diversas funciones. Además, salvo en casos excepcionales, existe un hueco en el conocimiento entre la relación que guardan la estructura y la función de una proteína (Williams, 1985).

La evolución de las distintas funciones de una proteína y su especialización se reflejan en la estructura, en la cual quedan marcados los eventos del cambio, aunque éstos no indican cómo es que se lleva a cabo la evolución de la función.

1.6 LA ESTRUCTURA Y ORGANIZACION DE LAS PROTEÍNAS

Si se conocen algunos de los niveles estructurales de las proteínas, se puede intentar establecer la relación que éstos guardan entre sí, para describir cómo se encuentra organizada una proteína. Si hiciéramos una separación arbitraria entre cada nivel estructural de las proteínas, donde los niveles fueran independientes unos de otros, tendríamos que aceptar que cada nivel representa una jerarquía con sus características propias. ¿Cómo serían afectadas a través del tiempo sus propiedades?. La respuesta a esta interrogante se ha estudiado de manera empírica desde los inicios de la biología molecular, y sólo hasta hace poco se ha intentado hacerlo con modelos predictivos (Creighton, 1984). Se ha visto que cada nivel estructural tiene su propio comportamiento en el tiempo (Zuckerandl y Pauling, 1962 y 1965; Rossmann y Argos, 1981). Las modificaciones que se suceden tienen dos modalidades de cambio, una a lo largo del tiempo, en toda su historia evolutiva y otra en su propio tiempo, que corresponde a el período en que se encuentra en un determinado linaje.

Al estudiar las secuencias de aminoácidos pueden observar algunas relaciones realizando un análisis cuantitativo que determine el grado de similitud. Si a través de este análisis se demuestra la existencia de una genuina relación, se dice que las secuencias son homólogas, es decir, que comparten una ascendencia común. Las posibilidades son las siguientes:

I) Al comparar diferentes secuencias entre sí, tanto de tamaño igual como diferente, se observa que contienen distintos grados de similitud. Puede ocurrir que dos secuencias se identifiquen como homólogas o bien que, a primera vista, no sea posible concluir que existe un parentesco, por lo que habría casos en que estudiando solamente la estructura primaria, sería imposible determinar si la similitud se conservó en otro nivel.

II) Si se tiene la secuencia de una proteína de función conocida y se compara contra el banco de secuencias de proteínas, a fin de buscar otras proteínas relacionadas que lleven a cabo la misma función, es posible hallar una proteína cuya similitud sea significativa y que tenga una función biológica diferente, o bien, que no exista ninguna relación aparente con proteínas de función similar. En este último caso es posible que la proteína haya cambiado sólo en su estructura primaria, conservado su estructura terciaria (Ver tablas 1.6.1 y 1.6.2). Asimismo se

Tabla 1.6.1 Similitudes obtenidas al comparar la estructura primaria de los residuos de aminoácidos de dos proteínas A y B.

I a	A es similar a B y ambas poseen la misma función
I b	A es similar a B y tienen distinta función
II a	A no es similar a B y tienen la misma función
II b	A no es similar a B y tienen diferente función
III a	Un segmento de A es similar a un segmento de B y tienen función equivalente
III b	Un segmento de A es similar a un segmento de B y tienen función diferente
III c	El segmento A está repetido dentro de la proteína y tiene función similar
III d	Un segmento de A está repetido y tiene distinta función

plantea la posibilidad de que exista una convergencia evolutiva en la estructura terciaria por lo cual no correspondería a una estructura terciaria.

III) Consideremos ahora el caso en donde la comparación de la totalidad de las secuencias de aminoácidos de las proteínas A y B no demuestra que existe una relación entre ambas. Sin embargo, si comparamos segmentos de la secuencia A contra la secuencia completa de B, puede ocurrir que éstas guarden similitud con una parte de B. A partir de este tipo de hallazgos se ha descubierto que las proteínas se construyen a base de módulos, denominados dominios, elementos que se combinan para dar origen a proteínas de diferente función. Ésto explica que distintas proteínas compartan dominios posiblemente con funciones similares.

Un caso análogo es aquel en el cual las proteínas pueden aumentar de tamaño debido a la duplicación de un dominio dentro del mismo gen. En la Tabla 1.6.1 se resumen los tres casos anteriores y sus diferentes variantes al comparar algunas estructuras primarias.

En el caso I las secuencias no han cambiado lo suficiente, por lo que es posible reconocer su ascendencia común y por lo tanto, concluirse que son secuencias homólogas. Ejemplo de la son las diferentes cadenas de globinas (mioglobinas y hemoglobinas) en las que encontramos ligeras variaciones, aún cuando se conserva la función (Zuckerkandl y Pauling, 1962; Stryer, 1988)). En cambio, la variante Ib está representada por el caso de la

Tabla 1.6.2 Ejemplos de los diferentes casos de la Tabla 1.6.1, donde se ha encontrado similitud entre diferentes proteínas.

CASO	EJEMPLO	REFERENCIA
Ia	Mioglobina con Hemoglobina	Rossmann y Argos (1981); Lehninger (1975)
FUNCIÓN	Une O_2 a un grupo prostético (Grupo Hemo)	
Ib	Lisozima con α -lactoalbúmin. Rompe el enlace glucosídico β de los azúcares NAM y NAG. Cambia la especificidad de la galactosil transferasa para formar lactosa en vez de unir galactosa con la NAG. Tripsina con Quimotripsina. Hidroliza el enlace peptídico en Arg o Lys. Insulina con Factor de crecimiento nervioso. Se une a un receptor de la membrana que promueve la internalización de la glucosa. Haptoglobina con Serina proteasa. Se une a un receptor de membrana que influye en el crecimiento de la neuritas, promotoras de las conexiones neurales.	Lehninger (1975) Stryer (1988) Lehninger (1975) Lehninger (1975)
IIa	Lisozima de Pollo con Lisozima de fago T4	Rossmann y Argos (1982); Stryer (1988)
FUNCIÓN	Rompe el enlace glucosídico β de los azúcares NAM y NAG.	
IIb	Ribonucleasa de bovino con Lisozima de pollo	Rossmann y Argos (1981); Lehninger (1975)
FUNCIÓN	Hidroliza RNA en el extremo 5' siempre que haya una pirimidina en el extremo 3'.	
IIIa	Proteína C con Factor XII con tPa Comparten los módulos tipo EGF (G), "Kring" (K), fibronectina (F1) y módulo tipo proteasa no catalítico (P). Intervienen en la coagulación y la fibrólisis; G actúa como factor de crecimiento y liga Ca^{++} . F1 actúa en la interacción proteína-proteína. P determina la especificidad de las proteínas K módulo que se ligan a los módulos de tipo proteasa.	Baron et al., (1991); Patthy (1985).
FUNCIÓN		
IIIb	Superóxidasa dismutasa (bovino) comparte un dominio con una inmunoglobulina (IgG)	Rossmann y Argos (1982).
FUNCIÓN		
IIIc	Ovomucoide, miosina, ferredoxina, tiene módulos repetidos internamente que inhibe la tripsina, tiene actividad de ATPasa y se liga a la actina en el músculo. Participa en el transporte de electrones en la fotosíntesis.	L1 y Graur (1991) Lehninger (1975)

N-acetil N-acetilmurámico (NAM), glucosamina (NAG).

lisozima de la clara de huevo y la α -lactoalbúmina, las cuales, si bien son homólogas, no conservaron la misma función (Rossmann y Argos, 1981 ; Phillips et al., 1983).

En los casos Ila y IIb la secuencia primaria no evidencia un origen común, ya que pudo haberse perdido o no haber existido nunca, a pesar de lo cual la función es igual. Un ejemplo de esto ocurre entre la lisozima de pollo y la del fago T4, donde las secuencias de aminoácidos son diferentes, pero la función es la misma (Phillips et al., 1983).

Un ejemplo del tipo IIb serían las secuencias de ribonucleasa de bovino y la lisozima de pollo, que no presenta una similitud estadísticamente significativa (Doolittle, 1988). El caso III es el más importante para los fines de este trabajo, pues en los últimos años se ha descubierto que algunas proteínas poseen segmentos homólogos entre ellas, lo cual ha llevado a proponer que las proteínas están construídas a base de módulos funcionales y estructurales que se pueden intercambiar para dar origen a proteínas con distintas funciones (Gilbert, 1987; Phillips et al., 1983).

Las proteínas se organizan y se pliegan con base en sus estructuras secundarias y supersecundarias. Al parecer estos dos niveles estructurales pueden formar otros niveles de organización que pueden mantenerse a lo largo de la evolución de una proteína. Se ha observado que algunas de estas estructuras básicas se

presentan en diferentes proteínas. Por ejemplo, en las hidrogenasas el sitio de unión al NAD de las piruvato cinasa y probablemente en la aldolasa comparten el dominio $\beta\alpha\beta$, (estructura supersecundaria) llamada el pliegue de Rossmann. Se ha discutido la posibilidad de que las estructuras comunes a varias proteínas provengan de una estructura ancestral o bien, que sean resultado de una convergencia evolutiva (Rossman y Argos, 1981).

Al estudiar la estructura de las proteínas puede pensarse que la misma determina su evolución, olvidándose de que uno de los factores centrales de la evolución de las proteínas es la función.

Aparentemente existen sólo unas cuantas estructuras básicas con las que se han podido armar las proteínas para funciones específicas, que son heredadas y se reutilizan en diferentes proteínas. Por ello es posible que las proteínas posean unos cuantos bloques básicos (Smódulos o dominios estructurales o funcionales) que son los componentes que posibilitan la construcción de diferentes enzimas con actividades y propiedades distintas (Rossman y Argos 1981; Phillips et. al., 1983). Lo anterior dificulta diferenciar si estos niveles estructurales son convergentes o divergentes, sobre todo en aquella en que las estructuras primarias no muestran una similitud discernible, y la función es diferente (ver Phillips et al., 1983; Thornton y Gardner, 1989). Ejemplos de estructuras compartidas son las superestructuras secundarias de barril $\alpha\beta$ y $\beta\alpha\beta$, que se

encuentran en distintas proteínas (Rossmann y Argos, 1981; Phillips et al., 1983).

Al comparar la estructura tridimensional de proteínas que llevan a cabo la misma función, se ha visto que este nivel estructural es el que tiene mayor conservación, es decir, que sufre cambios con mayor lentitud, mientras que la estructura primaria tiene cambios más rápidos en el tiempo. Ello significa que las estructuras de las proteínas perdieron evidencia de su homología a nivel de la secuencia de aminoácidos pero que ello, se ha conservado en el nivel de la estructura terciaria. Un ejemplo de este caso es la lisozima de pollo y la del fago T4 (Ver Tabla 1.6.2), que sólo comparten las partes de los dos sitios catalíticos, y el sitio de unión al sustrato. Otro caso es el de las enzimas malato deshidrogenasa (MDH) y lactato deshidrogenasa (LDH), que tienen una estructura terciaria similar, catalizan reacciones parecidas (reducen al oxaloacetato y al piruvato, respectivamente) y funcionan con sustrato diferentes. Por lo tanto, desde un punto de vista biológico, comparten algunas similitudes estructurales pero no funcionales, (Rossmann y Argos, 1981). Un experimento que logró revelar la cercanía de la especificidad funcional y estructural que guardan estas dos enzimas, que al determinarse mediante la mutagenesis dirigida, que con sólo hacer una sustitución de un residuo en la LDH, de Gln 102 por Arg, se logró cambiar la especificidad por el sustrato (Wilks et al., 1988). Esta mutación transformó la actividad de LDH a MDH, sintetizando malato en vez de lactato,

teniendo una eficiencia catalítica de 4.2×10^6 Kcat/Km (rango catalítico/constante de Michaelis) hacia su sustrato natural (piruvato) y de 4×10^1 para el oxaloacetato. Con el cambio de residuo, la enzima se vuelve 8400 veces más activa hacia el sustrato no natural (oxaloacetato) que para el natural (Wilks et al., 1988).

Si se tiene una similitud significativa en la estructura primaria es fácil predecir que ésta se observará también en la estructura terciaria; el ejemplo clásico es el de las globinas, donde se comparte la forma estructural general con pequeñas variaciones entre las hemoglobinas α y β (Rossmann y Argos, 1981; Phillips et al., 1983).

1.7 DOMINIOS Y SITIOS CONSERVADOS: LAS PROTEÍNAS SE FORMAN EN MOSAICO

Cuando Sanger y Tuppy (1951) secuenciaron la insulina se demostró que las proteínas contienen una secuencia única de aminoácidos, lo que abrió la posibilidad de suponer que las proteínas están organizadas en diferentes niveles estructurales. A partir de este descubrimiento la sucesión en que se dió el hallazgo de la conformación de los otros niveles estructurales fue rápido.

El primer paso fue dado por Pauling y Corey (1951), quienes lograron resolver las estructuras α hélice y β lámina, usando la

difracción de rayos X (Pauling et al., 1951 y Lehninger, 1975). Sin embargo, desde 1948 Pauling ya había dilucidado con papel y lápiz la estructura de las α hélices, tomando los datos de rayos X existentes de la α -queratina (Lehninger, 1975 y von Heijne, 1987).

Más tarde, se obtuvo, con mucha mayor nitidez, la estructura tridimensional de la mioglobina (Kendrew et al., 1960), la cual resultó estar constituida por varias α hélices. Este descubrimiento seguramente hizo muy feliz a Pauling (Juddson, 1980).

Sin embargo, una vez que se conoció la estructura tridimensional de dicha proteína, se descubrió que no tenía la estructura regular que algunos esperaban, y que no había simetría. Fue hasta 1959 cuando se propusieron los términos de estructura "primaria", "secundaria" y "terciaria" (Lindertrom-Lang y Shellman, 1959). Por otro parte, el pensamiento evolutivo ha permitido establecer que parte de las causas en el ordenamiento de una secuencia esta correlacionada con la historia evolutiva de la proteína; es decir, que el ordenamiento se debe a causas de origen biológico (Crick, 1958).

Al aplicarse la perspectiva evolutiva a la comparación de secuencias de aminoácidos, se pudo establecer, cuando sólo se contaba con secuencias de hemoglobina de diferentes especies, que algunas proteínas tienen una tasa constante de sustitución de

aminoácidos en el tiempo (Pauling y Zuckerkandl, 1962). Por otra parte, utilizando la similitud en la secuencia de proteínas homólogas fue factible realizar una reconstrucción del camino seguido por un polipéptido a lo largo de la evolución.

La pregunta sobre la secuencia de aminoácidos planteada anteriormente puede ser extendida a los otros niveles estructurales: ¿Qué reglas sigue el ordenamiento de los diferentes niveles estructurales?. Una de las estrategias que ha permitido ir encontrando respuesta ha sido el estudio de la diversidad de secuencias y estructuras que presentan las proteínas. Esto se ha podido lograr gracias al desarrollo simultáneo de la informática, y a la rapidez con que ahora se dispone de secuencias de genes. Se han podido dar así pasos más sólidos para discriminar las diferentes estructuras que presentan las proteínas (que se pueden definir en módulos o dominios estructurales y funcionales dentro de los cuales se localizan los sitios funcionales). Toda esta redundancia de información ha enriquecido y permitido que se reconstruyan al menos en parte los eventos del pasado evolutivo de las proteínas (Doolittle, 1988 y 1989).

El desarrollo de la computación ha permitido almacenar gran cantidad de las secuencias que se producen en la actualidad. Dicha información está concentrada y ordenada en bancos de datos de diferentes tipos. Por ejemplo, existen bancos de genes, de proteínas y de la estructura terciaria de las proteínas. Por

otro lado, las computadoras actuales permiten realizar cálculos rápidamente, por lo que se pueden ejecutar programas muy complejos con acceso a una mayor memoria. Ello permite, en principio, analizar las secuencias con mucha facilidad.

El discernimiento de las estructuras se ha realizado utilizando una de las herramientas de mayor tradición en la Biología: la comparación, gracias a la cual se puede establecer, en las proteínas, si dos secuencias de aminoácidos tienen un porcentaje de similitud significativo para ser consideradas homólogas o no, precisándose si las proteínas están en alguna de las distintas familias estructurales. Dichas familias están formadas por un número limitado de dominios que se organizan en distintos juegos (Rossmann y Argos, 1981).

Tabla 1.7.1 Sitios funcionales (Tomado de Doolittle, 1988).

	MOTIVO	PROTEÍNA
Sitios Activos	GDSSG TMSA GXCH/Y UDTG	Sitio activo de la proteasa serina Subtilisina-tipo de proteasa de serina Proteasa sulfhidrica Proteasa Ácida
	DFG(X)nAPE TCP(X)nNLGT CTTNC UAYEP	Hidroxiquinasa Guanidinin-quinasa Fosfogliceraldehído deshidrogenasa Triosa fosfato isomerasa
Sitios de ligado	TASHD CCXXHD SXXK	Fosfoglucomutasa Fosfolipasa A Lactamasa
	NXS/T C/H(X)n C/H(X)n C/H(X)n C/H RGD CXXCH GXGXXG	Sitios que se unen Asn con carbohidratos Dedo de zinc Adhesión celular Citocromo C tioéster Motivo de Rossmann, sitios que unen nucleótidos
"U" = residuo no polar : "X" = cualquier residuo		

El poder reconocer un bloque funcional de una proteína que se ha caracterizado posibilita, en otros casos, un estudio más rápido de todas las secuencias que se van conociendo; de no ser así, el estudio de las secuencias sería extremadamente lento debido a que habría que estudiar la estructura de cada proteína por separado. En cambio, al caracterizar un módulo sólo hay que descubrir quiénes lo comparten, estableciendo así la función posible de la proteína, bajo la hipótesis de que los dominios específicos están regularmente asociados a una determinada actividad (Thornton y Gardner, 1989).

Gracias a las comparaciones realizadas se ha encontrado que las proteínas están constituidas en mosaicos de dominios y sitios funcionales conservados, lo cual permite suponer que muchas proteínas tienen una ascendencia de origen mezclado o compartido (véase la Tabla 1.6.2). De hecho, se ha sugerido que el número de dominios funcionales con los que se construyen las proteínas es limitado (Gilbert 1987). Con base en lo anterior, pueden hacerse comparaciones sobre la composición o proporción de los dominios y sitios funcionales dentro de las proteínas, abriéndose así la posibilidad de que su clasificación refleje una ascendencia común.

El centro activo de una proteína se puede encontrar en la interfase de los dominios (Rossmann y Argos, 1981). Empero, hay dominios funcionales que sólo se encuentran en una sola parte del polipéptido (Li y Graur, 1991). Dentro del dominio la función se

encuentra asociada a pequeñas secuencias específicas denominadas motivo o sitio funcional.

En los dominios se pueden encontrar secuencias más pequeñas que forman uno o varios bloques muy conservados. Estas secuencias suelen tener 30 aminoácidos o menos. Debido a que son segmentos muy pequeños (con respecto al dominio), podrían mostrar un sitio funcional, el cual se puede compartir entre diferentes proteínas que muestran actividades similares. Un buen ejemplo lo constituyen los sitios funcionales descritos por Doolittle (1988) (Tabla 1.7.1)

1.8 SITIOS QUE SE UNEN A NUCLEÓTIDOS Y/O POLINUCLEÓTIDOS

Varias de las proteínas que intervienen en los procesos metabólicos usan nucleótidos como sustrato o cofactores para llevar acabo su función. Por lo tanto, presentan dominios o sitios a donde se ligan sustratos (nucleótidos), coenzimas (aceptores o donadores de H^+), como reguladores (AMPc) o moldes (polinucleótidos).

Las proteínas que requieren unión a nucleótidos están presentes en los sistemas que transfieren energía y que regulan la replicación de los ácidos nucleicos. Tal es el caso de las polimerasas, las tRNA-sintetasas, la proteínas ribosomales, de la cápside viral y de la ribonucleoproteínas (RNP-CS).

Las proteínas que actúan con polirribonucleótidos pueden tener un origen diferente al de las que lo hacen con ribonucleótidos como serían algunos dominios que están presentes en las proteínas ribosomales como S1 y PNPasa (Ver los capítulos correspondientes).

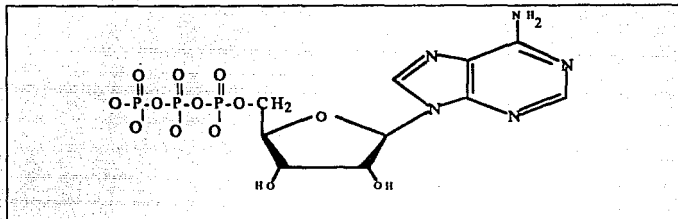


Figura 1.8.1. El ATP y algunos aminoácidos se unen mediante enlaces negativos con el oxígeno del fosfato: por medio de los oxígenos de los extremos hidroxilo 2' y 3' de la furosa, y con cargas positivas a través de los hidrógenos del grupo amino del anillo de purina.

Las fuerzas que intervienen en la unión polipéptido-nucleótido son atracciones electrostáticas (puentes de metales), fuerzas dipolares (puentes de hidrógeno) y fuerzas entrópicas (efecto hidrofóbico).

Algunas de las proteínas que se unen a ácidos nucleicos son aquellas que requieren de cofactores como las ATPasas, proteínas G, factores de elongación, quinasas, deshidrogenasas, etc. Existen otras que se unen a polirribonucleótidos como las polimerasas, ribonucleoproteínas, las endonucleasas y las exonucleasas, los factores de transcripción, los factores de

represión, las histonas, las topoisomerasas, las proteínas de la envoltura viral y aminoaciltransferasas.

Algunas de las proteínas anteriormente mencionadas incluyen en su estructura segmentos polipeptídicos que contienen residuos de aminoácidos que en su cadena lateral presentan cargas capaces de establecer, a través de las fuerzas electrostáticas, enlaces con las regiones polares de los ácidos nucleicos. La cadena de fosfatos puede formar complejos con aminoácidos básicos mediante puentes salinos, con el grupo ϵ -amino de la Lys, el grupo guanidín de la Arg y el anillo del imidazol protonado de la His. Los puentes de hidrógeno se forman con el grupo amino de la cadena lateral de un aminoácido del polipéptido. Algunos de los aminoácidos que pueden establecer dicho enlace la son la Asn y el ácido Gln (polares neutros), y la Lys y Arg. El ácido Asp y el Gln pueden enlazarse a los hidroxilos de la ribosa a través del grupo carboxilo de la cadena lateral. Se ha encontrado que la Gly está presente en las secuencias de sitios de unión a nucleótidos debido a que permite que el polipéptido se plegue (Blackburn y Gait, 1990).

La coenzima NAD⁺ es una molécula formada por dos nucleótidos que corresponden a un anillo de nicotinamida y una adenina. Este dinucleótido es utilizado por todas las deshidrogenasas para llevar acabo la catálisis del sustrato. Rossmann *et al.* (1975) descubrieron que todas estas proteínas comparten un módulo

estructural que se une a la parte nucleotídica de la coenzima. A este módulo se le denominó pliegue de Rossmann.

El pliegue de Rossmann corresponde a una estructura supersecundaria $\beta\alpha\beta$ (Rossmann y Argos, 1981), que forma una oquedad hidrofóbica en la molécula que permite que la inserción de la adenina de la NAD^+ . La secuencia de aminoácidos conservados de la malato deshidrogenasa-soluble (s-MDH), la lactato deshidrogenasa (LDH), la alcohol deshidrogenasa ADH y la gliceraldehído 3-fosfato deshidrogenasa (GAPDH) que permiten la unión del nucleótido al pliegue son: Gly 28, Gly 33 y Asp 53 de la secuencia de la LDH y las correspondientes en el resto de las hidrogenasas. Como la glicina presenta un hidrógeno lateral que no representa un impedimento estérico, la adenina se acomoda en la cavidad. El ácido aspártico se une a través un puente de hidrógeno que se establece entre el aldehído de su cadena lateral y el grupo 2'-hidroxilo de la ribosa de adenina (Rossmann et al., 1975).

Los dos nucleótidos del NAD^+ están unidos mediante dos pirofosfatos. Es a través de uno de éstos que se establece un sitio más de unión entre el NAD^+ y la deshidrogenasa. Una diferencia que existe entre las estructuras de las deshidrogenasas es que la LDH contiene una asa funcional, que no comparte con la ADH y con la GAPDH, que forma el pliegue de Rossmann.

Tabla 1.8.1 Algunos motivos funcionales que se comparten en distintas proteínas, que ligan Nucleótidos.

	PROTEÍNA QUE COMPARTEN SITIOS FUNCIONALES	MOTIVOS CONSENSO QUE SE LIGAN CON NUCLEÓTIDOS				REFERENCIAS
		GXXXXGKS/T	NKXD	DXXGH	ETSAK	
1	GTPasa GRANDE subunidad α	SI	GTP	GDP	GTP	Kaziro et al., 1991
2	GTPasa PEQUERA (RAS) P21 oncogen	GTP/GDP	GTP	SI	GTP	Argos y Leberman, 1985 Saraste et al., 1990
3	EF-Tu, EF-1-a, EF-2 y EF-G	GTPGDP	GTP	SI	GTP	Saraste et al., 1990
4	ATPasa subunidad β	ADP/ATP/GTP	NO	NO	NO	Hsu et al., 1987 Saraste et al., 1990
5	ATPasa subunidad α	ADP/ATP/GTP	NO	NO	NO	Saraste et al., 1990
6	REC A	ATP	NO	NO	NO	Saraste et al., 1990
7	Miosina cadena pesada	ATP	NO	NO	NO	Saraste et al., 1990
8	Alcohol deshidrogenasa	NAD ⁺	NO	NO	NO	Argos y Leberman, 1985
9	Gliceraldehído-3-fosfato	NAD ⁺	NO	NO	NO	Argos y Leberman, 1985
10	Lactato deshidrogenasa	NAD ⁺	NO	NO	NO	Argos y Leberman, 1985
11	Glutación reductasa	FAD ⁺	NO	NO	NO	Argos y Leberman, 1985
12	p-hidroxibenzoato hidroxilasa	FAD ⁺	NO	NO	NO	Argos y Leberman, 1985
13	Fosfogliceraldehído cinasa	ATP	NO	NO	NO	Saraste et al., 1990
14	Adenilato cinasa	AMP/ATP	NO	NO	NO	Saraste et al., 1990
15	Guanilato cinasa	ATP	NO	NO	NO	Saraste et al., 1990
16	V-SRC oncogen	ATP	NO	NO	NO	Argos y Leberman, 1985
17	RNA polimerasa celular β	?	NO	NO	NO	Cornelissen et al., 1985
18	RNA polimerasa celular β'	?	NO	NO	NO	Cornelissen et al., 1985
19	Oxidasa D-aminoácido	Posibles	NO	NO	NO	Argos y Leberman, 1985
20	Proteína ribosomal S9	Posibles	NO	NO	NO	Argos y Leberman, 1985
21	Proteína ribosomal S5	Posibles	NO	NO	NO	Argos y Leberman, 1985
22	Histona H1 y H5	Posibles	NO	NO	NO	Argos y Leberman, 1985
23	Cinasa AMPc dependiente	SI	NO	NO	NO	Saraste et al., 1990
24	Tubulina	SI	?	?	?	Saraste et al., 1990
MOTIVOS		LF1GL(RNP2)	[Xn]	RG6FUUY(RNP1)		
25	AMP-CS y tipos relacionados	Polímero monocatenario de RNA o DNA				Bandzulis et al., 1989
26	Subunidad Sigma (RNA polimerasa)	NO	NO	SI		Helmann y Chamberlin, 1988

La región de la coenzima donde se encuentra presente la adenina, interactúa de manera similar en todas las deshidrogenasas mencionadas anteriormente.

En las proteínas GOGAT, *p*-hidroxibenzoato hidroxilasa, girasa, rubredoxin reductasa, flavoproteína oxidoreductasa, alcali hidroxilasa y glutatión reductasa, el pliegue de Rossmann se presenta dos veces, una en el extremo amino y otra en parte media de la proteína. En la primera se une a FAD⁺ y en el segundo a NAD⁺ o NADP⁺.

Se ha encontrado también que el dominio que une a la coenzima FAD⁺ presenta la secuencia típica del motivo de Rossmann, aunque de las tres Gly conservadas una puede ser sustituida por una Ala. Los sitios de interacción en FAD⁺ son un poco diferentes a los del NAD⁺, en cuanto a la longitud de aminoácidos de la estructura $\beta\alpha\beta$. Al igual que el sitio de unión de NAD⁺, la Gly permite que se aproxime el pirofosfato de la coenzima al extremo amino terminal de la hélice. La parte ácida, en este caso el ácido Asp, forma un puente de hidrógeno con el extremo 2'-OH de la ribosa del ADP, aunque en el caso de NADPH es con His o Arg, como ocurre con la Girasa (Wierenga *et al.*, 1983). La forma en que se inserta el FAD⁺ es muy similar al del NAD⁺ en las deshidrogenasas.

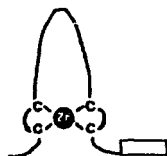
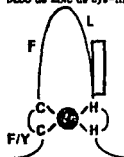
En el dominio de Rossmann se forma un pliegue (asa) que se encuentra entre la estructura α y la β . Este pliegue tiene un tamaño variable, dependiendo si se une NAD⁺ o FAD⁺. Por otro

lado, se ha sugerido que también las Gly se encuentran muy cercanas al pirofosfato del nucleótido (Wierenga et al., 1983). De hecho, Eggink et al. (1990) sugirieron que existen otros segmentos de la secuencia que pueden establecer interacciones con la coenzima, ya sea formando un ambiente hidrofóbico o bien llegando a establecer enlaces con diferentes partes de la coenzima como puede ser una Thr y un Asp en la secuencia TXXXIYIIGD.

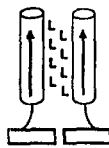
Hélice-vuelta-Hélice



DEDO DE ZINC de Cys-His



DEDO DE ZINC de Cys-Cys



CIERRE DE LEUCINA

Figura 1.8.2 Algunas representaciones estructurales típicas de unión a DNA presentes en las polimerasas (Tomado de Struhl, 1989).

El motivo que forma parte del pliegue de Rossmann está presente en todas aquellas proteínas que utilizan los cofactores mencionados anteriormente, además de GTP y ATP en GTPasas y cinasas (Wierenga et al., 1983; Eggink et al., 1990). De lo anterior se puede deducir que el sitio consenso es el motivo de Rossmann. A través de la comparación de las quinasas y GTPasas con este motivo, se encontró que las ATPasas también lo presentan (Duncan

et al., 1986; Hsu et al., 1987; Lochrie y Simon, 1988; Rozen et al., 1989; Kemp y Pearson, 1990; Saraste et al., 1990; Kaziro et al., 1991).

En las enzimas como la RNA y la DNA polimerasas, existen varios sitios de unión a nucleótidos que se unen al molde recién sintetizado, y uno de nucleótidos para incorporarlos a la cadena durante la reacción de elongación (Tabla 3.4.2). Ambas polimerasas presentan estructuras similares de unión a DNA. Para el caso particular de la RNA polimerasa DNA dependiente se ha sugerido la presencia del motivo de Rossmann (Tabla 3.4.2), que pudiera ser un sitio de unión al sustrato (Allison et al., 1985; Cornelissen et al., 1988; Mémet et al., 1988 a, b; Pühler et al., 1989 a, b; Igloi et al., 1990; Nawrath et al., 1990).

De la comparación del motivo de Rossmann de la adenilato quinasa, con las secuencias de las ATPasas, se encontró que el motivo GXXXXGKT/S (la segunda Gly no siempre se conserva), también está presente en estas últimas, que podría corresponder sitio de unión al ATP (Duncan et al., 1986; Hsu et al., 1987; Rozen et al., 1989; Kemp y Pearson, 1990). En las ATPasas se ha logrado identificar a este motivo que posiblemente se une cerca del fosfato α del nucleótido como sucede en las quinasas e hidrogenasas. Además, se sabe que para que el ATP se una no es estrictamente necesaria la presencia de magnesio en la F1 (en el caso del motivo de Rossmann). Se ha propuesto pues que deben

existir otros motivos que interactúen directamente con el nucleótido (Garboczi *et al.*, 1988 b; Ysern *et al.*, 1988).

Debido a que las ATPasas tienen cierto parecido funcional con las GTPasas, en estas últimas se han logrado determinar otros motivos que interactúan con el GTP, se piensa que deben existir secuencias equivalentes en ATPasas, que hasta el momento no se han identificado. Se ha argumentado que como esta región está muy conservada en las ATPasas, el sitio de unión debe ser más grande que los 8 aminoácidos (Lochrie y Simon, 1988; Saraste *et al.*, 1990; Kaziro *et al.*, 1991).

Una caracterización más amplia del dominio en las GTPasas incluye, además de la secuencia consenso GXG/XXXGK más de 11 a 17 residuos de cualquier tipo, continuándose con dos aminoácidos hidrofóbicos seguido por cualquier otro aminoácido, rematando con una Lys que puede ser sustituida por Asp o Glu (Argos y Leberman, 1985) (Tabla 1.8.2). Se han sugerido la existencia de al menos tres sitios más que se ligan al GTP/GDP, con las secuencias DXXGH, NKXD y ETSK (Lochrie y Simon, 1988; Kaziro *et al.*, 1991)..

Se ha propuesto también que los sitios DXXGH, NKXD y ETSK son parte de dominios funcionales donde ocurre hidrólisis del nucleótido, e intervienen en el cambio conformacional, cuando se une la guanina (Kaziro *et al.*, 1991), aunque cada uno de ellos puede interactuar con alguna parte del nucleótido el conjunto es

el que permite establecer la función (Kaziro et al., 1991). Con el uso de la cristalografía de rayos X se observó que para las proteínas *ras* y los factores de elongación -TU (EF-FU) que el motivo de Rossmann se encuentra cerca de los fosfatos β y α del GDP ligado. Además, en el caso del GTP, se observó que el grupo amino de la Lys de este motivo forma puentes de hidrógeno con el fosfato β y/o γ (neutraliza la carga negativa del fosfato) y con los oxígenos de la cadena principal de los residuos 10 y 11 de la vuelta inversa sobre sí misma (asa P) (Saraste et al., 1990). En la proteína *ras* la Ser interactúa con Mg^{++} y con el hidrógeno de su oxígeno que se liga con el fosfato del ATP (Kaziro et al., 1991). Para el motivo DXXGH se ha sugerido que el Asp puede formar enlaces coordinados a través del Mg^{++} con el fosfato β del GDP (Lochrie y Simon, 1988); el hidrógeno de la amida de la Gly forma puentes de hidrógeno con el fosfato γ del GTP. Esta Gly al parecer juega un papel relevante en el cambio de la conformación de la proteína al momento de intercambiar el GDP por GTP. Por lo tanto, a este dominio se le asigna la función, no solamente de ligar al nucleótido, sino también la de sufrir un cambio conformacional debido al cambio del sustrato. El motivo NKXD interactúa directamente con el anillo de la guanina del GTP. Lo mismo sucede con el motivo ETSK en las proteínas *ras*, se propone que la secuencia correspondiente en las GTPasas de subunidad mayor es H(F/M)TCA(I/V)DT (Lochrie y Simon 1988, Kaziro et al., 1991).

2.0 EVOLUCIÓN DE LAS RNA POLIMERASAS

Tradicionalmente el estudio de la célula se ha centrado con más fuerza en la perspectiva de una biología funcional. Esto deja muy poco espacio a la intervención de otro paradigma, que es fundamental para generar el conocimiento de la célula: el estudiarla desde o conjuntamente con la perspectiva evolutiva. El descubrimiento de las arqueobacterias y la utilización de la evolución molecular han permitido en la actualidad hacer una biología celular comparada entre los diferentes linajes celulares. Lo anterior permite reconocer cuáles rasgos de los tres linajes (eubacterias, arqueobacterias y eucariontes), son los más antiguos, y deducir así algunas de las características del ancestro común (Lazcano *et al.*, 1992a y b).

Por otro lado, al ir penetrando en el estudio de la evolución de las células ancestrales, se abre la posibilidad de determinar qué complejos y actividades macromoleculares de las células actuales pudieron preceder al ancestro común. El estudio de lo anterior, llevaría a plantearnos preguntas de mayor precisión sobre qué es y cómo surgió la vida. Así, aunque durante mucho tiempo se sostuvo con ahínco que las primeras células tenían que estar basadas exclusivamente en el DNA y proteínas. Desde el punto de vista evolutivo, surge la hipótesis de células basadas en RNA y aún en polinucleótidos con azúcares más sencillos como el glicerol.

3.1 EL MUNDO DEL RNA

En la década de los 80's se desarrolló la hipótesis de que tanto las proteínas como el DNA fueron precedidos por el RNA en las funciones, catalíticas e informativas. Lo que dió el giro a favor del RNA fue el descubrimiento de la actividad catalítica que este posee (Kruger *et al.*, 1982; Guerrier-Takada *et al.*, 1983). Posteriormente se han ido caracterizando las diferentes actividades que presentan las ribozimas el RNA (Cech y Bass 1986; Joyce, 1989). Además, desde tiempo atrás, se había observado que el RNA tiene la capacidad de transmitir información genética, como sucede en algunos retroides, virus y retrovirus (Reaney, 1982; Diener, 1984).

Lo anterior llevó a Alberts (1986), a Gilbert (1986) y a Lazcano (1986) ha plantear la posibilidad de que durante el Arqueano temprano, existieran poblaciones celulares con un metabolismo basado en RNA. Ello niega el surgimiento tanto de las proteínas como del DNA a partir de la evolución prebiótica. Según este punto de vista, dichos elementos aparecieron en las células como producto de la evolución biológica y no química. A esta etapa de la evolución celular se le denominó El Mundo del RNA (Gilbert, 1986).

3.2 LOS PRIMEROS PÉPTIDOS Y EL GENOMA

Dos de los problemas básicos que se debieron presentar en la etapa de las células basadas en RNA, fue el éxito para la reproducción (que implica la replicación propia de su genoma de RNA) y para llevar a cabo la catálisis de sus sustratos (metabolismo).

Una alternativa a la hipótesis anterior es la que propone que los primeros polipéptidos que surgieron codificados por el código genético no presentaban actividad catalítica, sino que tenían la función de estabilizar al RNA para que éste aumentara su actividad catalítica (Lazcano, 1986; Lazcano et al., 1992 a y b). Dicha característica adaptativa podría explicar el origen y evolución de las primeras proteínas a través de la selección natural. A partir de esta idea se infiere que la evolución de las primeras proteínas se halla estrechamente asociada con la de la unión a los ácidos nucleicos. Es decir, muchos de los posibles dominios funcionales se originaron para unirse estructuralmente al RNA y a nucleótidos para llevar a cabo sus diferentes funciones.

Dado que un complejo molecular puede intervenir en diferentes procesos biológicos, es posible suponer que tanto el aparato de traducción y replicación como el de la catálisis de sustratos, evolucionaron con mecanismos de origen común. Tal es el caso de

la proteína ribosomal S1, la cual participa en la síntesis de proteínas y en la replicación de algunos virus de RNA.

La principal evidencia que lleva a postular que la respuesta adaptativa de los primeros polipéptidos corresponde a la capacidad de los mismo para unirse a polirribonucleótidos y ribonucleótidos es la función que realiza el segmento proteico la proteína RNAasa P de *E. coli* (Lazcano, 1986; Lazcano et al., 1992 a y b). La RNAasa P es un complejo que formado por una cadena de aminoácidos, y por una cadena de RNA catalítico (Hansen et al., 1985). Si se extrae el polipéptido y se deja sólo al RNA, se observa que la actividad catalíticase encuentra en el RNA y que ésta es baja (Altman, 1984; Guerrier.Takada et al., 1983). Para realizar normalmente su actividad, el complejo molecular requiere del polipéptido para potenciar la catálisis. Esto se consigue a través de la estabilización del RNA en una estructura terciaria, permitiéndole así aumentar su actividad.

Otra evidencia correspondé al hecho de que muchas proteínas actuales requieren de coenzimas para llevar a cabo su función catalítica. La parte ribonucleica de las coenzimas se considera como una reminiscencia del RNA catalítico ancestral (White, 1976). Las coenzimas presentan las siguientes características: a) la mayoría proviene de precursores ribonucleicos, b) más del 50% de éstas son indispensables para el funcionamiento de la proteína y c) algunas presentan actividad catalítica, ya sea como parte

del sitio activo, como aceptor de electrones. Para que surgiera y se estableciera la relación funcional proteína-coenzima, ambas partes tuvieron que coevolucionar. Ello apoya la hipótesis de que los primeros dominios polipeptídicos evolucionaron a partir de la interacción con los polirribonucleótidos.

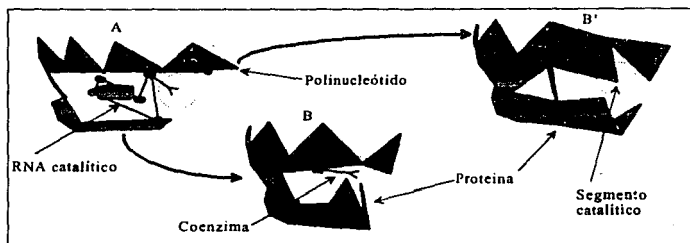
Una evidencia adicional más la constituyen los aminoácidos histidina y cisteína, que posiblemente se originaron a partir de ribonucleótidos catalíticos coenzimas que se transformaron en dichos aminoácidos (White, 1976). Se podría pensar así que las proteínas actuales se convirtieron en un sustituto estructural del RNA catalítico ancestral, conservándose en el polipéptido la posibilidad de asociarse a aquel o aquellos nucleótidos que tenían la actividad catalítica. Por lo tanto, se ha conservado la posición del sitio activo en la proteína con la coenzima, como se encontraba en el complejo polipéptido-RNA catalítico (Figura 3.2.2). Otro ejemplo, pueden ser los ribosomas los cuales contienen una parte de proteína y otra de RNA.

La catálisis que lleva a cabo el RNA no es muy eficiente para acelerar los procesos en los que interviene. Por ello se propone que el RNA requirió de polipéptidos que fueran capaces de estabilizarlo, logrando así hacer más eficientes el metabolismo de replicación y el de transformación de un sustrato.

la aceleración de la actividad catalítica trajo como consecuencia poblaciones celulares más competitivas al menos en cuanto a la eficiencia en la reproducción.

Es posible que los primeros polipéptidos que evolucionaron dando origen a las proteínas, son aquellos dominios funcionales capaces de unirse a polirribonucleótidos y nucleótidos. A la luz de lo anterior, puede ser probable que las proteínas que en la actualidad presentan sitios de unión a nucleótidos y polinucleótidos, compartan un origen común, siendo posibles descendientes de dominios que ancestralmente se encontraban unidos a un RNA catalítico.

Figura 3.2.2 Evolución de los dominios y el RNA catalítico. A) Polipéptido no catalítico unido a un RNA catalítico. B) Polipéptido unido a una coenzima que contribuye a la catálisis. B') El último paso evolutivo de la proteína es la pérdida de la coenzima, para que esta lleve a cabo la catálisis.



3.3 LA EVOLUCIÓN DE LAS POLIMERASAS Y EL RNA

La capacidad de las poblaciones celulares de perpetuar su información genética permite mantener su identidad y por ende su sobrevivencia. Ello requiere de una replicación eficiente. De hecho, se puede decir que una parte de la evolución celular ha sido moldeada o seleccionada entorno a la mayor eficiencia y fidelidad de replicación posible, sin haber llegado a la perfección.

Considerando lo anterior, al estudiar la evolución de la maquinaria que fue seleccionada para replicar a los primeros genomas celulares, como la RNA polimerasa, podemos reconstruir el camino seguido por las células en sus diferentes etapas evolutivas.

Si la molécula que antecedió al DNA como reservorio de la información genética fue el RNA y ésta misma precedió a las proteínas en la catálisis, entonces debieron aparecer primero polipéptidos que contribuyeron a catalizar la autorreplicación del RNA, haciéndola más eficiente. Dichos polipéptidos pudieron derivar en una replicasa o en una RNA polimerasa RNA-dependiente que hubiera sido seleccionada a favor rápidamente. Por lo anterior, se considera que las RNA replicasas debieron ser una de las primeras proteínas que evolucionaron en la célula, y de la

cual surgen luego todas polimerasas actuales (Lazcano 1986; Lazcano et al., 1988b) (Figura 3.3.1).

De acuerdo con Lazcano (1986) y Lazcano et al. (1988b) los vestigios de polimerasa ancestral se encuentran en la subunidad β' de las eubacterias de las RNA polimerasas DNA-dependientes, y en sus homólogos en los distintos linajes celulares.

3.4 EVOLUCIÓN DE β' Y SUS DOMINIOS VESTIGIALES

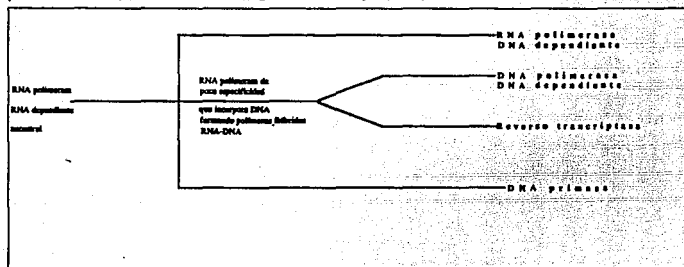
Todas las distintas RNA polimerasas DNA-dependientes celulares tienen subunidades homólogas a β y β' de *E. coli* (Tabla 3.4.1). Las correspondencias con estas subunidades son las siguientes: en eubacterias se les denominan β' y β ; en arqueobacterias A, C y B (aunque esta última subunidad puede encontrarse dividida en B' y B''); en eucariontes se presentan tres RNA polimerasas la I (A), II (B) y la III (C), cada una de las cuales contiene la subunidad mayor y la segunda más grande, que son homólogas a β' y β , respectivamente (Pühler et al., 1989) (Tabla 3.4.1).

En el caso de las cianobacterias, las subunidades equivalentes a β' pueden encontrarse en dos subunidades separadas, que son denominadas β' y γ (Xie et al., 1989; Mèmet et al., 1988a y b).

En las arqueobacterias la subunidad homóloga a la subunidad β' de *E. coli* se encuentra dividida en las subunidades A y C. En

eubacterias y arqueobacterias los genes homólogos a β y β' se encuentran codificados, uno al lado del otro, en un sólo operón (Zillig *et al.*, 1989).

Figura 3.3.1 Diagrama del posible origen común y evolución de las polimerasas. (Modificado de Lazcano *et al.*, 1988b).



El surgimiento del resto de las subunidades particulares de cada linaje, serían adaptaciones posteriores para regular la transcripción (Armaleo, 1987; Lazcano *et al.*, 1988b). De acuerdo con la hipótesis anterior, la polimerasa ancestral debió ser una RNA polimerasa RNA-dependiente monomérica, lo que ha llevado a sugerir que las subunidades β' y β son consecuencia de una amplificación del gen β' (Lazcano *et al.*, 1992a).

Los ancestros de esta polimerasa pudieron ser dominios funcionales con actividad para reconocer a polirribonucleótidos y ribonucleótidos con sus derivados. Por ello debería ser posible

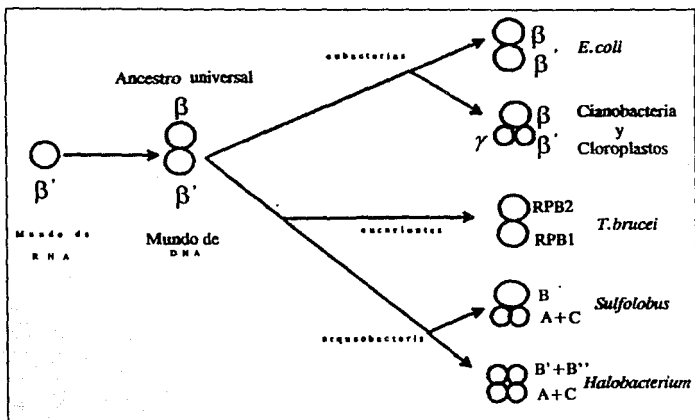
identificar en β' dominios capaces de unirse a polirribonucleótidos que tomen a la hebra recién elongada.

Tabla 3.4.1 Correspondencia de las subunidades, β y β' en los distintos linajes celulares.

Especie	<i>E. coli</i>	<i>Nostoc</i> <i>comune</i>	<i>Sulfolobus</i> <i>acidocaldarius</i>	<i>Halobacterium</i> <i>halobium</i>	<i>Trypanosoma</i> <i>brucei</i>
subunidad equivalente	β	β	B	B' B''	2a. sub unidad
	β'	$\beta'+\gamma$	A+C	A+C	subuni- dad mayor
Eubacteria Cianobacterias Arqueobacteria Arqueobacteria Eucariote					

La RNA polimerasa DNA-dependiente es una proteína de dimensiones muy grandes, razón por la cual se ha dificultado su estudio. Hasta el momento sólo se ha logrado hacer un modelo basado en la cristalografía de rayos X de la estructura terciaria de muy baja resolución (16Å). Debido a ésto se imposibilita asignar con claridad qué dominios y sitios corresponden a una función. Sin embargo, la subunidad β' de *E.coli* y sus homólogas en los tres linajes celulares (eubacteria, arqueobacteria y eucariote) comparten por lo menos siete regiones conservadas, existiendo dentro de éstas once motivos a los cuales se les han atribuido funciones específicas (Allison *et al.*, 1985; Cornelissen *et al.*, 1988; Mémet *et al.*, 1988a y b; Pühler *et al.*, 1989a y b; Igloi *et al.*, 1990; Nawrath *et al.*, 1990) (Tabla 3.4.2).

Figura 3.4.1. Evolución de las subunidades mayores de la RNA polimerasas. Las subunidades $\beta + \beta'$ son homólogas de β y $\alpha + \alpha'$ son homólogas de β' . RPB2 es homólogo de β . RPB1 es homólogo a β' . γ es homóloga de una parte de β' . (Modificado de LLaca, 1988).



En las células actuales, las RNA polimerasas DNA-dependientes juegan el papel de sintetizar a los intermediarios que comunican la información genética del DNA a la síntesis de proteínas, es decir, llevan a cabo el proceso de transcripción. En ellos existen varios dominios que se ligan a nucleótidos, que podrían ser aquellos que interaccionan con DNA de una y dos hélices, permitiendo la lectura para formar un RNA específico, a la vez que interactúa con él. Por lo tanto, es una proteína que está muy ligada a la interacción con nucleótidos, sean estos monómeros o parte de una cadena polinucleotídica.

Tabla 3.4.2 Dominios conservados en la RNA polimerasa DNA-dependiente de la subunidad β' y sus homólogas celulares. Dominio (Dom)

Región Dom	Estructura	Función	Dominio homólogo a
A ó I	D1 Dedo de Zinc	Interacción con DNA	Factor de transcripción IIA y SV40
B ó II	D2 Dominio D-D	Se cree que es un: - sitio catalítico y/o - unión a nucleótidos	Se encuentra en todas las RNA polimerasas DNA y RNA-dependiente, y en la PNPasa.
C ó III	D3 Motivo de Rossmann	Sitio de unión a nucleótido	ATPasas, cinasas, lactato deshidrogenasas, etc.
	D3' Motivo hélice J vuelta hélice K	Se une a DNA	DNA polimerasa I <i>E. coli</i> DNA polimerasa fago T7
D ó IV	D4 Motivo de Mèmet	función y estructura desconocidas	Secuencia más conservada en las diversas RNA polimerasas de origen
	D4' hélice P	Se une a DNA	DNA polimerasa I
E ó V	D5	Estructura y función desconocidas	Es equivalente a la región III descrita por Mèmet et al. (1988b)
	D6	Estructura desconocida. Está involucrada con la respuesta de α -amanitina, una toxina que inhibe la elongación (Cochet-Meilhac y Chambon, 1974), por lo que esta región posiblemente se halla relacionada con alguno(s) aspecto(s) de la elongación (Jokerst et al., 1989).	Determinado a partir de experimentos con la toxina α -amanitina.
F ó VI	D7	Estructura desconocida. Este dominio es homólogo a cuatro de S1 y a uno de la PNPasa. Apparently interactúa con polirribonucleótidos. Este segmento es el más conservado en las RNA polimerasas II de eucariotes y el segundo más conservado en las homólogas.	Las proteínas ribosomales S1 y PNPasa, son equivalentes a la región IV descrita por Mèmet et al. (1988b)
G ó VII	D8	Estructura y función desconocida	
H ó VIII	D9	Estructura y función desconocida	Es equivalente a la región V descrita por Mèmet et al. (1988b)

En la subunidad β' de *E. coli* y sus homólogas RNA polimerasas DNA-dependientes celulares, existen varios dominios y sitios funcionales en su estructura primaria que se han conservado a lo largo de la evolución, y que se les ha tratado de asignar la posible función que desempeñan. Sus posibles funciones se han deducido a partir de la similitud que guardan con los dominios de otras proteínas cuya función de sus dominios sí se ha caracterizado. Al analizar dichos dominios, es muy difícil distinguir los que son reminiscentes, que existieron en la RNA-replicasa ancestral, de aquéllos que se modificaron o aparecieron posteriormente para permitir la transcripción (de DNA a RNA) y una regulación más eficiente de la misma. Un criterio que puede ayudar a discernir que dominios son más antiguos, será aquél que se encuentre más extendido en las proteínas.

3.0 LA PROTEÍNA RIBOSOMAL S1

De acuerdo con Kolb et al. (1977), la proteína ribosomal S1 es una proteína ácida. Presenta 557 aminoácidos en *E.coli* y 568 en *Rhizobium meliloti* (Schnier et al., 1988). Su peso molecular en *E.coli* es de 76 Kd (Suryanarayana y Subramanian, 1979), y es la más grande de las 21 proteínas que constituyen la subunidad 30S de los ribosomas de las eubacterias Gram negativo (Higo et al., 1982). A parte de mediciones hidrodinámicas por dispersión de rayos X se ha estimado que posee un tamaño aproximado de 200 a 280Å, siendo un poco más grande que el resto del ribosoma, que mide 200Å (Giorginis y Subramanian, 1980; Suryanarayana y Subramanian, 1979).

S1 fue una de las primeras proteínas ribosomales de la subunidad 30S a la que se le pudo asignar una función específica. Resulta indispensable para formar el complejo 30S ribosomal y está involucrada en el reconocimiento y pegado del mRNA al ribosoma. Al parecer S1 estimula la traducción al unirse al ribosoma, por lo que se deduce que tiene un importante papel en la iniciación de la síntesis de proteínas (Draper y von Hippel, 1978a).

La proteína S1 se ha descrito únicamente en eubacterias Gram negativas, como *E. coli*, donde se ha estudiado extensamente, y en otras eubacterias Gram negativas, como las eubacterias púrpura *Proteus*, *Providencia*, *Serratia*, *Pseudomonas*, *Agrobacterium* y

Rhizobium que también son eubacterias Gram negativas (Schnier et al., 1988). Se sabe que la ausencia de S1 es letal en *E. coli* (Kitakawa y Isono 1982). En eubacterias Gram positivas, como *Bacillus subtilis*, *B. stearothermophilus* y *Micrococcus phlei*, se han buscado proteínas homólogas con resultados negativos (Isono e Isono 1976; Higo et al., 1982; Schnier y Faist 1985). Dicha búsqueda se realizó mediante inmunorréplicas tipo Western y a través de la caracterización bioquímica en los ribosomas. Hasta la fecha no se ha podido detectar una proteína con funciones equivalentes en eubacterias Gram positivas. En el caso de las arqueobacterias, como *Halobacterium halobium* y otras halofilas, tampoco se obtuvieron resultados positivos.

Aunque la proteína S1 no ha sido encontrada en eubacterias Gram positivo ni en arqueobacterias, no debe descartarse la posibilidad de existan otras proteínas no homólogas que juegan un papel similar al de S1, como han argumentado Schnier y Feist (1985).

Se ha observado que las eubacterias Gram positivas requieren de una interacción más extendida entre mRNA y rRNA, para dar inicio a la síntesis de proteínas (Murray y Rabinowitz, 1982). *In vitro* los mRNA de eubacterias Gram negativas no son traducidos eficientemente por los ribosomas de las eubacterias Gram positivo (Isono e Isono, 1976) y en el caso inverso, las eubacterias Gram positivas por lo general no pueden traducir ningún mRNA de *E. coli* (Murray y Rabinowitz 1982). Asimismo, se ha hallado que el genoma

del $\phi 29$ (fago de Gram positivas) es traducido ineficientemente por el ribosoma de *E. coli* (Murray y Rabinowitz, 1982).

La proteína ribosomal S1 no interviene exclusivamente en la síntesis de proteínas. Se ha demostrado que también participa en sistemas virales como Q β , MS2, R17 y f2, donde se integra y forma parte de la replicasa, el complejo oligomérico que replica el genoma viral.

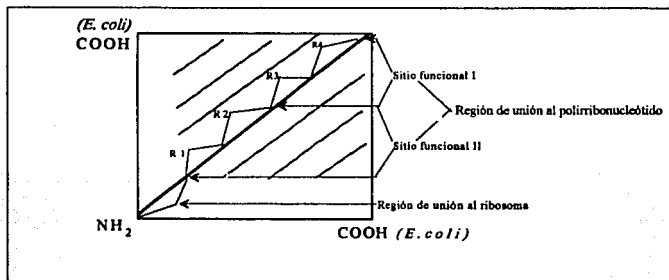
La Q β replicasa está formada por una subunidad codificada por el virus, y tres más que son tomadas de la eubacteria infectada (*E. coli*). La subunidad I es precisamente la proteína S1 ribosomal, la subunidad II es codificada por el virus, y se agregan a ellas los factores de elongación EF-Tu III y EF-Ts IV (Wahba *et al.*, 1974; Draper y von Hippel, 1978a).

Es esencial que S1 se una a la cadena positiva de Q β para iniciar la replicación de la cadena negativa ya que además, confiere actividad represora en la traducción de la replicasa de Q β (Wahba *et al.*, 1974). Al suprimirse S1 en el complejo de la replicasa, ésta no reconoce al RNA positivo, pudiendo entoces utilizar poly(rC) o la cadena de RNA negativo de Q β . Se puede concluir, por lo tanto, que S1 interviene de algún modo en el reconocimiento específico del RNA positivo de Q β (Draper y von Hippel, 1978a,b).

A diferencia de lo que ocurre con otras proteínas ribosomales, el gen de S1 no se codifica junto con alguna subunidad de

polimerasa, ni el gen *rpsA* de S1 en *E. coli* forma parte de un operón de transcripción donde estén presentes otras proteínas ribosomales. Hasta el momento se disponen de cuatro secuencias de S1, dos para *E. coli* (Schnier e Isono 1982; Schnier et al., 1982), otra para *Providencia sp.* (Schnier et al., 1985) y otra más para *Rhizobium meliloti* (Schnier et al., 1988).

Figura 4.1 Representación gráfica de una matriz de puntos. Regiones funcionales de la proteína ribosomal S1. En el extremo amino se encuentra la región que interactúa con la partícula 30S ribosomal. En el extremo medio y COOH terminal se encuentran las cuatro repeticiones de los dominios que se unen al polirribonucleótido; R1, R2, R3 y R4. Los sitios funcionales I y II se caracterizaron bioquímicamente (Draper y von Hippel, 1978 a, b).



Una vez secuenciados estos genes se pudo ver que la proteína S1 posee dos regiones funcionales (Fig 4.1). Una de ellas se localiza en el extremo amino terminal y tiene la función de asociarse con el ribosoma extendiéndose del aminoácido 1 al 191. En el extremo carboxilo terminal se encuentra el segundo segmento, que se une al RNA. Este último sitio se caracteriza

por tener cuatro repeticiones internas homólogas entre sí (Giorginis y Subramanian 1980; Schiner y Isono, 1982 Schiner et al., 1985). Cada una de estas repeticiones son a su vez homólogas de un dominio de 69 aminoácidos de la PNPasa (Régner et al., 1987). Como veremos en este trabajo, al parecer este dominio también está presente en la subunidad β' de la RNA-polimerasa DNA dependiente de *E.coli* y sus homólogos celulares.

La región que se une a los ácidos nucleicos (Tabla 4.1), que corresponde a la región del extremo carboxilo. Al estudiar la experimentalmente esta región, se caracterizaron a su vez en dos sitios funcionales, denominados I y II. El sitio I se subdivide en dos segmentos que poseen las mismas propiedades, y que funcionan pegándose a DNA o RNA monocatenario y reconociendo los enlaces fosfato del esqueleto del polinucleótido, respectivamente, teniendo propiedades de cooperatividad (el primer sustrato que se liga a la proteína facilita que otros se unan más fácilmente). El sitio II es más específico, pues sólo

Tabla 4.1 Características y propiedades de los dominios del extremo COOH⁻ terminal de S1. La región del extremo NH₂⁺ se une al ribosoma.

Región del extremo COOH ⁻ tiene la propiedad de unirse a polinucleótidos monocatenario				
Regiones	SITIO II		SITIO I	
Dominios	R1	R2	R3	R4
Características de los dominios				
Tipo de polinucleótido. al que se une	RNA A través de sales con la base y el azúcar		DNA y RNA	DNA y RNA
Tipo de enlace			Espina de fosfato	Espina de fosfato
Presenta cooperatividad			SI	SI

reconoce RNA monocatenario. El enlace se realiza a través de intermediarios como Mg^{++} y Na^+ , que se enlazó con la base y el azúcar, y no a través del fosfato (Draper y von Hippel, 1978 a y b). Se cree que el sitio II está asociado con la unión al ribosoma (Draper y von Hippel, 1978 a).

Al parecer el sitio II corresponde a la parte intermedia de la proteína, que abarca las dos primeras secuencias repetidas que, además, son las más conservadas (Schnier et al., 1988, ver Tabla 4.1). El sitio I es el que está más cerca del extremo carboxilo terminal, y corresponde a las dos últimas repeticiones, a diferencia del sitio II, en donde la secuencia de aminoácidos de ésta no se encuentra tan conservada. El mismo fenómeno ha sido observado en la secuencia de S1 de *Rhizobium meliloti* (Schnier et al., 1982a, b; Schnier et al., 1988).

Se ha postulado que la proteína S1 es capaz de acoplar el mRNA con la subunidad rRNA 16S por su extremo 3' (von Dieijen et al., 1976). Sin embargo, Laughrea & Moore (1978) han demostrado que esto no ocurre directamente. Ello llevó a preguntarse si los cuatro dominios funcionales, además de unirse al mRNA, también tienen la posibilidad de reconocer al 16S rRNA.

Las evidencias experimentales sugieren que al momento de ligarse al RNA la proteína ribosomal S1, impide que éste adopte una estructura secundaria, estabilizándolo linealmente, de esta manera las bases quedan expuestas para facilitar tanto la

replicación (en el caso de la Q β replicasa), al impedir que se formen asas que pudiesen entorpecer la lectura del molde, como la traducción a proteína (von Dieijen et al., 1976; Kolb et al., 1977). Esta posibilidad está confirmada con los resultados experimentales usando formaldehído, que también impide que el mRNA adopte una estructura secundaria, y que facilita que se inicie la síntesis de proteína en ausencia de S1 (Kolb et al., 1977; von Dieijen et al. 1976; Draper y von Hippel, 1978 a y b).

Como en la secuencia S1 de *E.coli* se detectó la presencia de cisteína, se sugirió que existía la posibilidad que este aminoácido se relacionara con la actividad de unión a polinucleótidos, debido a los grupos -SH que presenta. Se realizaron experimentos donde se bloqueó el grupo -SH del aminoácido, pero no se encontró ninguna inhibición de la actividad de unión a polinucleótidos. Además, no se hallaron cisteínas en la secuencia de S1 de *Rhizobium*, por lo que se descartó algún papel importante de este aminoácido en el pegado de los polinucleótidos (Schnier et al., 1988).

De lo anterior se puede deducir que S1 más que desempeñar un papel funcional en el mismo ribosoma, está relacionada con el reconocimiento de RNA. A primera vista uno podría pensar que S1 tiene dos funciones que serían disímiles y sin relación entre sí. Sin embargo, se pueden relacionar sus actividades tanto en la traducción como en la replicación de los genomas virales. Éstos dos se pueden conjuntar en un sólo mecanismo, que consiste en una

sola propiedad: impedir que se pliegue el RNA para formar estructuras secundarias y terciarias. Esto permite linearizar al RNA, permitiendo así la iniciación tanto de la replicación como de la traducción. Aparentemente SI lo logra a través de la estabilización de la estructura del RNA en forma lineal, y es posible que al exponer las bases le permita interactuar con los otros rRNA o en la lectura de sus bases. No se le ha podido asociar con algún motivo funcional que se una directamente con el polinucleótido dentro de los sitios I y II y sus subsecuentes dominios. Lo excepcional de esta proteína es que interviene tanto en la síntesis de proteínas como en la replicación del RNA genómico de $\phi\beta$. Estas dos actividades de la proteína están relacionadas con el RNA, cumpliendo así con un sólo papel.

4.0 LA PROTEÍNA QUE POLIMERIZA AL AZAR: PNPasa

La polinucleótido fosforilasa (PNPasa-polirribonucleótido orotofosfato nucleotidiltransferasa, EC. 2.7.7.8) fue descubierta, en 1955, en el laboratorio de Severo Ochoa por Marianne Grunberg-Manago (Grunberg-Manago, 1989). El hallazgo de dicha proteína permitió desarrollar una serie de experimentos que respondieran a una serie de preguntas específicas relacionadas con los ácidos nucleicos, debido a que la enzima puede sintetizar oligonucleótidos polimerizándolos al azar, sin la necesidad de copiar un molde y sin requerir de cebador. En *E.coli* la PNPasa consiste de un trímero de subunidades idénticas denominadas α .

Se considera que la PNPasa interviene en la degradación de los diferentes tipos de RNA (Godefroy-Colburn y Grunberg-Manago, 1972, función sobre la cual aún existen dudas.

La PNPasa se purificó inicialmente en la eubacteria (*Azotobacter vinelandi*), y desde entonces se ha reportado en los restantes linajes celulares, incluyendo eubacterias Gram positivas (*Mycobacterium*) y Gram negativas (*Neisseria meningitidis*, *Salmonella plantarum*); arqueobacterias, como es el caso de *Halobacterium* (Louis et al., 1971; Francino, 1992) y en eucariontes como *Chlorella*, plantas y animales (vertebrados e invertebrados) y así como en cloroplastos de frijol y lechuga (In: Godefroy-Colburn y Grunberg-Manago, 1972).

Existen observaciones de microscopía electrónica que sugieren que la PNPasa tiene un tamaño de unas 85Å (Godefroy-Colburn y Grunberg-Manago, 1972).

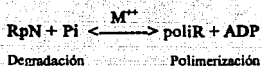
El carácter ácido de la PNPasa se ha determinado gracias a una serie de experimentos donde el pK y el pI lo muestran (Godefroy-Colburn y Grunberg-Manago, 1972). Esta caracterización se ve reforzada por la secuencia primaria en *E.coli*, que en su mayoría son de tipo ácido. Los pesos moleculares reportados se muestran en la Tabla 4.1.

Tabla 4.1 Pesos moleculares de la PNPasa de diferentes organismos, calculadas por métodos bioquímicos y por la secuencia del gen que codifica para la enzima.

Organismos	Estimación bioquímica del tamaño de la subunidad α (KDa)	Estimación por secuencia de la subunidad α (Da)	Referencias
<i>E.coli</i>	----- 86	77,122	Règnier <i>et al.</i> , 1987
<i>E.coli</i>	84 95	-----	Portier <i>et al.</i> , 1975; Soreq y Littauer, 1977; Lehrach <i>et al.</i> , 1971
<i>Micrococcus luteus</i>	70	-----	Klee, 1969
<i>Halobacterium</i> sp.	12 75-90	-----	Louis <i>et al.</i> , 1971; Francino, 1992
<i>Clostridium perfringens</i>	65	-----	Dietz y Grunberg-Manago, 1967

El descubrimiento y estudio de la PNPasa se ha caracterizado por investigaciones que han establecido que las diferentes PNPasas, presentes en los tres linajes celulares, comparten la misma actividad y posiblemente la misma función (Francino, 1992). En

los estudios bioquímicos de la enzima se ha determinado que la PNPasa lleva a cabo, degradación, fosforólisis y polimerización. Es decir, las PNPasas no sólo tienen reacciones de intercambio, sino también intervienen en reacciones donde se obtienen productos. La reacción es reversible y se representa de la siguiente manera:



RpN: polirribonucleótido con un grupo hidroxilo en el sitio 3'
 Pi: fosfato inorgánico
 R: polinucleótido (elongación de novo)
 M: metal divalente (Mg⁺⁺ o Mn⁺⁺)
 (Godefroy-Colburn y Grunberg-Manago, 1972)

La reacción de polimerización se lleva a cabo en el sentido 5'---> 3', a partir de la polimerización de ribonucleosidodifosfatos, en donde el fosfato β se libera como P_i, y el fosfato α se enlaza con el sitio 3' de la ribosa, pudiéndose formar polímeros de diferentes tamaños. La reacción inversa sucede cuando un ortofosfato se adiciona al P_i α del nucleósido, lo cual permite que éste se separe del polímero. La degradación del polímero se lleva a cabo con la polaridad 3'--->5' (Godefroy-Colburn y Grunberg-Manago, 1972) (Figura 4.1).

La PNPasa forma polímeros usando de preferencia como sustrato a ribodinucleósidos, pero como no es tan específica puede usar desoxirribonucleósidos para formar DNA monocatenario de tamaño corto. La enzima no tiene especificidad por las bases (Hsieh, 1971).

El gen que codifica para la PNPasa se encuentran a 69 minutos en el genóforo de *E.coli*. Forma parte del operón de la proteína ribosomal S15, por lo que se regulan juntos (Fig. 4.2). Como ya se mencionó, la subunidad α de la PNPasa está codificada por un gen que tiene un marco de lectura de 2133 nucleótidos y codifica para una proteína de 711 aminoácidos. La holoenzima es un trímero de subunidades α y con un solo sitio catalítico por subunidad (Règnier *et al.*, 1987).

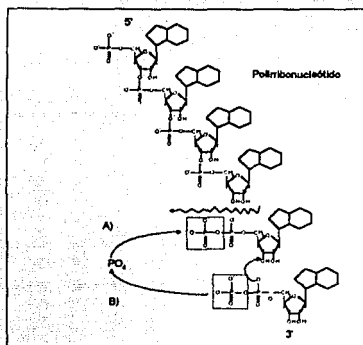


Figura 4.1 Representación de cabo la polimerización y la degradación catalizada por la PNPasa.

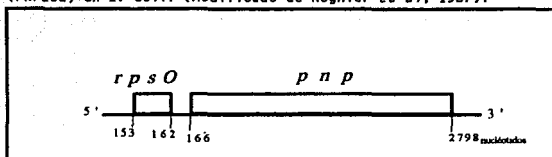
A) En la fosforólisis entra el P_i para convertirse en el β fosfato del ADP, cortando al polímero en el extremo 3'.

B) Sale el $Pi\beta$ del ADP y éste se polimeriza formándose un polirribonucleótido.

La función que tienen las PNPasas dentro de los sistemas vivos es poco clara. Por ejemplo, una doble mutante de *pnp-* y *rnb-* (RNasa II) muere rápidamente (Grunberg-Manago, 1989). Sin embargo, si se repara la célula con un plásmido que contenga el gen de *pnp+* se restaura el crecimiento. Cuando se repara el gen *rnb+* con una mutante *pnp-*, también se restaura el crecimiento. Esto sugiere que la PNPasa y la RNasa II tienen un papel

fisiológico equivalente en la célula, lo que apoya la idea de que la PNPasa degrada mRNA y permite el reciclamiento de los ribonucleótidos (Grunberg-Manago, 1989).

Fig 5.2 Orden de la secuencia del Operón *rpsO* (S15) y *pnp* (PNPasa) en *E. coli*. (Modificado de R  gnier et al, 1987).



El uso de codones sugiere que es una enzima de alta expresi  n. Se calcula que las c  lulas de *E.coli* presentan las mismas cantidades de mol  culas de PNPasa que de RNA-polimerasa DNA dependiente, cuyo n  mero se calcula en alrededor de 1700 mol  culas por c  lula (Godefroy-Colburn y Grunber-Manago 1972).

Como ya vimos por estudios bioqu  micos se ha determinado que la PNPasa contiene un sitio catal  tico por subunidad y uno de uni  n a RNA, en posiciones separadas dentro de la secuencia de cada subunidad α . El sitio de uni  n al RNA se encuentra en el extremo carboxilo terminal, hacia el exterior de la subunidad. Ello se determin   en experimentos donde se observ   que cada una de las subunidades de la PNPasa se proteolizaba, dividi  ndose las subunidades α de 85 KDa en dos trozos de 62 KDa y 20 KDa, respectivamente (Portier, 1975). La primera pierde la capacidad

para unirse a RNA de más de 11 nucleótidos, y la segunda se relaciona con la unión al RNA (Thang et al., 1971; Régnier et al., 1987). Esto sugiere que el sitio catalítico se encuentra en el extremo NH₂ terminal, y el sitio de unión en el extremo carboxilo de la PNPasa.

Después de determinar la secuencia de la PNPasa, Régnier et al., (1987), descubren la existencia de un segmento de 69 aminoácidos que se localiza en el extremo carboxilo terminal, y que es homólogo a cada una de las cuatro repeticiones internas que presenta la proteína ribosomal S1, cuyas longitudes van de 73 aminoácidos a 75 aminoácidos. Debido a que comparten dominios homólogos, ambas proteínas (S1 y PNPasa) pueden interactuar con RNA monocatenario, de manera similar y sin ninguna especificidad por la secuencia.

De hecho, si comparamos el funcionamiento entre la proteína S1 y la PNPasa, es posible equiparar la función que lleva a cabo cada dominio de S1 para ligar al RNA, con la que realizaría cada subunidad de la PNPasa. Por ejemplo, se ha observado que S1 puede seguir realizando la función con sólo tres de los cuatro dominios homólogos que interactúan con el RNA monocatenario. Por su parte, la forma activa de la PNPasa es de tres subunidades α , lo cual podría indicar que ambos grupos de dominios pueden tener interacciones funcionales dinámicas que les permiten moverse a lo largo de una sola molécula de RNA (Régnier et al., 1987).

Tabla 5.2 Porcentajes de similitud entre el dominio de 69 aminoácidos de la PNPasa y cada uno de los cuatro segmentos de los dominios de la proteína ribosomal S1 (Régnier, *et al.*, 1987).

	Dominio de 69 a.a. de PNPasa	% de a.a. conservados	Substituciones de a.a. equivalentes
1º SEGMENTO DE S1	PNPasa	42%	22%
2º SEGMENTO DE S1	PNPasa	39%	17%
3º SEGMENTO DE S1	PNPasa	39%	10%
4º SEGMENTO DE S1	PNPasa	40%	8%

Dentro de los 69 aminoácidos que constituyen el dominio de la PNPasa, existe una subregión de 27 aminoácidos más conservada, que va del aminoácido 626 al 652. La conservación de aminoácidos idénticos o equivalentes en las mismas posiciones de los 5 dominios, uno de PNPasa y cuatro de S1 (Tablas 5.2 y 6.1), y en particular la conservación de 6 Gly y 9 residuos hidrofóbicos (señalados en la misma Tabla 6.1 con un asterisco), hicieron pensar a Régnier *et al.* (1987) que el dominio de la PNPasa se pliega de la misma manera que los de S1. Esta conclusión se ve apoyada por la observación de que en las proteínas que se unen a DNA, los sitios conservados, constituidos por aminoácidos como la glicina y los residuos hidrofóbicos, se pliegan de manera similar a la que poseen estructuras terciarias de dominios que interactúan con DNA de otras proteínas con funciones equivalentes (Régnier *et al.*, 1987).

La recurrencia de dominios funcionales homólogos en diferentes proteínas, como ocurre con la enzima PNPasa y la proteína S1, es

un fenómeno que se puede interpretar a la luz de la aparición de dominios surgidos a partir de la amplificación de un minigen ancestral que codificaría para secuencias que se unieran a RNA, y podrían tener un origen común muy antiguo (Régner et al., 1987). Es posible, por lo tanto, que esta región corresponda a uno de los primeros dominios que evolucionaron que se hallaban en células del Arqueano y que se ligaban a RNA (Lazcano, 1986 y Lazcano et al., 1992a).

Aquí cabría preguntarse si se puede responder ¿en qué momento y bajo qué condiciones de la evolución biológica surgió dicho dominio? ¿ya se encontraba el dominio en el último ancestro común a todos los linajes celulares, o su aparición data de más atrás, los organismos del mundo del RNA? ¿es posible que el dominio fuese una adquisición más reciente? Las respuestas a estas y otras preguntas las trataremos de contestar en los capítulos finales de esta tesis.

En este trabajo se caracteriza un dominio que se comparte con otras polimerasas y se hace un método para dilucidar su función. Para ello se compara la PNPasa, que posee una región de 69 aminoácidos de la PNPasa homóloga a las cuatro repeticiones del extremo amino de la proteína ribosomal S1 (Régner et al., 1987). A estas regiones se les ha caracterizado la capacidad de unirse a RNA. g Debido a que la PNPasa polimeriza y degrada RNA, esta proteína debe presentar un sitio de unión al RNA. Por ello,

Régnier et al. (1987) dedujeron que este dominio repetido tiene la función de unirse al RNA que va a ser degradado o sintetizado por la PNPasa, debido al tipo de secuencia que presenta el dominio, donde se encuentran aminoácidos como la Gly, Arg, Lys, His y Asp, que son aminoácidos típicos de sitios de unión a nucleótidos.

En este trabajo se caracteriza un dominio que se comparte con otras polimerasas y se hace un método para dilucidar su función. Para ello se compara la PNPasa, que posee una región de 69 aminoácidos de la PNPasa homóloga a las cuatro repeticiones del extremo amino de la proteína ribosomal S1 (Régnier et al., 1987). A estas regiones se les ha caracterizado la capacidad de unirse a RNA. Debido a que la PNPasa polimeriza y degrada RNA, esta proteína debe presentar un sitio de unión al RNA. Por ello, Régnier et al. (1987) dedujeron que este dominio repetido tiene la función de unirse al RNA que va a ser degradado o sintetizado por la PNPasa, debido al tipo de secuencia que presenta el dominio, donde se encuentran aminoácidos como la Gly, Arg, Lys, His y Asp, que son aminoácidos típicos de sitios de unión a nucleótidos.

Es posible que los sitios similares en ambas proteínas, la subunidad β' de la RNA polimerasa y sus homólogas y en la PNPasa, tengan la misma función de unión a ribonucleótidos.

5.0 MATERIAL Y MÉTODO

Se utilizaron y se analizaron las secuencias de la subunidad β' de la RNA polimerasa DNA-dependiente de *E.coli* (2.7.7.6) y sus homólogas (Tabla 5.2), las secuencias de las subunidades de la ATPasa F1 (α y β) de *E.coli* (3.6.1.34) y sus homólogas en los diferentes linajes celulares (Tablas 5.3 y 5.4, respectivamente), la secuencia de la proteína ribosomal S1 de *E.coli*, presentes sólo en eubacterias Gram negativas (Tabla 5.1), así como la secuencia de la polinucleótido fosforilasa de *E.coli* (2.7.7.8) (Tabla 5.1).

La comparación entre las diferentes secuencias se realizó utilizando los programas FASTP, TFAST y FASTA (Lipman y Pearson, 1985), basados en el algoritmo de Wilbur y Lipman (1983). Estos programas permiten hacer búsquedas tanto para secuencias de aminoácidos como de nucleótidos. Las secuencias de aminoácidos de las diferentes proteínas se alinearon con los programas "GAP", "BESTFIT", "LINEUP", "PROFILE" y "PROFILEEGAP" del paquete UWGGC de la Universidad de Wisconsin (Devereux et al., 1984), y se corrigieron manualmente.

Se utilizó un programa elaborado por Luis Medrano en el Laboratorio de Microbiología del Departamento de Biología de la Facultad de Ciencias, para observar el perfil de hidropatía y carga eléctrica que presentan las distintas secuencias de aminoácidos.

Todos los programas se corrieron en una PC-80386DX y una VAX (pbr232) del Instituto de Biotecnología de la UNAM. En la primera se formó el banco de polimerasas y se corrieron los programas FASTP y el de perfiles de hidropatía. En la segunda se corrieron los programas FASTT, FASTA y los del paquete de Wisconsin (UWGCG). Los bancos que se utilizaron para llevar a cabo la comparación y búsqueda de secuencias fueron el "GENBANK" (versión 70), formado por nucleótidos; el banco de residuos de aminoácidos "PIR" versión 7 (Protein Identification Resource) compilado por la "National Biomedical Research Foundation" y uno de polimerasas formado por Victor LLaça y Renato Cappello, en el Laboratorio de Microbiología de la Facultad de Ciencias, UNAM. Se utilizó el código de aminoácidos de una sola letra (Tabla 5.6).

En lo general, se trabajó, con el método de Doolittle (1988) para alinear las secuencias de aminoácidos. Sin embargo, como Doolittle (1988) ha señalado, el valor mínimo de similitud de una pareja de secuencias de 100 aminoácidos es del 15%. Debido a que en este trabajo se estudiaron un dominio y un sitio funcional cuya longitud es menor que 100 aminoácidos, se tomaron las recomendaciones adicionales propuestas por Argos y Leberman (1985) y Doolittle (1988). Los criterios permiten verificar si la correlación es legítima o no, pues se podría deber a un artificio. Una de las recomendaciones es hacer la comparación del patrón a través de los bancos de secuencias disponibles. Para interpretar estos resultados hay que pensar si tienen sentido biológico o no, es decir, si hay una relación natural y con sentido biológico entre el dominio en cuestión y los resultados obtenidos en la búsqueda. La presencia de la secuencia en cuestión en sitios conservados de todas las proteínas homólogas constituye un criterio adicional de validez.

Se calculan perfiles de hidropatía de acuerdo a Hopps y Woods (1981) por considerarlos elementos adicionales para verificar la autenticidad de la correlación entre las secuencias en cuestión. Para lo anterior se usaron tanto el dominio de 69 aminoácidos, reportado por Régner et al. (1987), como el octapéptido de aminoácidos que se obsevo como parte de una subsecuencia de búsqueda contra los diferentes bancos.

Sólo se tomaron en cuenta aquellas proteínas que tuvieran características biológicas adecuadas, como la presencia de dominios que se unan a ribonucleótidos y polirribonucleótidos y que permitieran establecer una relación razonable. Otros criterios adicionales para determinar la validez de un resultado en la búsqueda contra los bancos de secuencias es el no haber encontrado falsos positivos, es decir, proteínas que tuvieran el octapéptido y presentaran una función que no estuviese relacionada con la unión a nucleótidos. Además si no se encuentre ninguna otra proteína que se uniera a nucleótidos y que posea el octapéptido, se puede tomar con mayor confianza los resultados.

Tabla 5.1 Fuente de la secuencia de nucleótidos y aminoácidos de la polinucleótidosforilasa de *E. coli*. Fuente de la secuencia de nucleótidos y aminoácidos de la subunidad ribosomal S1 de *E. coli* y sus homólogas.

Organismo	Abreviación	Referencia
Eubacteria (Gram negativa)		
<i>Escherichia coli</i>	PNPasa	Régner et al. (1987)
Eubacterias (Gram negativas)		
<i>Escherichia coli</i>	EcRPS1	Schnier et al. (1982); Schnier y Isono (1982)
<i>Providencia sp.</i>	P-RPS1	Schnier et al. (1988)
<i>Rhizobium melinotti</i>	RmRPS1	Schnier et al. (1985)

TABLA 5.2 Listado de organismos de donde se tomaron las secuencias de aminoácidos y nucleótidos de las subunidades β' de la RNA polimerasa DNA-dependiente de *E.coli* y sus homólogos.

Organismo	Abreviación	Referencia
Eubacteria		
<i>Escherichia coli</i>	Ec	Ovchinnikov et al., (1982)
<i>Pseudomonas putida</i>	Pp	Danilovich et al., (1988)
<i>Nostoc commune</i>	Nc	Xie et al., (1989)
<i>Marchantia polymorpha</i> (Cloroplasto)	MpCl	Ohyama et al., (1986)
<i>Spinacia oleracea</i> (Cloroplasto)	SoCl	Hudson et al., (1988); Igloi et al., (1990)
<i>Nicotiana tabacum</i> (Cloroplasto)	NtCl	Shinozaki et al., (1986)
<i>Pisum sativum</i> (Cloroplasto)	PsCl	Cozens y Walker., (1986)
<i>Zea mays</i> (Cloroplasto)	ZmCl	Igloi et al., (1990)
Arqueobacterias		
<i>Sulfolobus acidocaldarius</i>	Sac	Pühler et al., (1989b)
<i>Methanobacterium thermoautotrophicum</i>	Mth	Berghöfer et al., (1988); Schallenberg et al., (1988)
<i>Halobacterium halobium</i>	Hh	Leffers et al., (1989); Pühler et al., (1989a)
<i>Halococcus morrhue</i>	Hm	Leffers et al., (1989)
Eucariontes		
<i>Trypanosoma brucei</i> I (A)	TbI	Jess et al., (1989); Smith et al., (1989)
<i>Trypanosoma brucei</i> II (B)	TbII	Evers et al., (1989a); Smith et al., (1989)
<i>Trypanosoma brucei</i> III (C)	TbIII	Köck et al., (1988); Smith et al., (1989)
<i>Plasmodium falciparum</i> II (B)	PfII	Li et al., (1989)
<i>Plasmodium falciparum</i> III (C)	PfIII	Li et al., (1991b)
<i>Crithidia fasciculata</i> II (B)	CfII	Evers et al., (1989)
<i>Saccharomyces cerevisiae</i> I (A)	ScI	Mémet et al., (1988 a y b); Zillig et al., (1989)
<i>Saccharomyces cerevisiae</i> II (B)	ScII	Sweetser et al., (1987); Mémet et al., (1988b)
<i>Saccharomyces cerevisiae</i> III (C)	ScIII	Allison et al., (1985); Mémet et al., (1988)
<i>Schizosaccharomyces pombe</i> I (A)	SpI	Yamagishi y Nomura, (1988); Azuma et al., (1991)
<i>Schizosaccharomyces pombe</i> II (B)	SpII	
<i>Caenorhabditis elegans</i> II (B)	Cell	Bird y Riddle, (1989)
<i>Drosophila melanogaster</i> II (B)	DmII	Jokerst et al., (1989)
<i>Mus musculus</i> II (B)	MmII	A'hearn et al., (1987)
<i>Arabidopsis thaliana</i> II (B)	AtII	Nawrath et al., (1990)
Virus de Eucariontes		
Virus de Vaccinia	Vacc	Broyles y Moss, (1986); Goebel et al., (1990)

Tabla 5.3 Listado de organismos de donde se tomaron las secuencias de aminoácidos y nucleótidos de la subunidad β de las ATPasas Fien *E.coli* y sus homólogas (Mt: mitocondria; Cl: cloroplasto).

Organismo	Abreviación	Referencia
Rubacterias		
<i>Escherichia coli</i>	Ec	Kanazawa et al., (1982); Walker et al., (1984)
<i>Cytophaga lytica</i>	Cl	Amann et al., (1988)
<i>Vibrio alginolyticus</i>	Va	Krumholz et al., (1989)
<i>Bacillus firmus</i>	Bfir	Ivey y Krulwich et al., (1991)
<i>Bacillus megaterium</i>	Bm	Hawthorne y Brusilow, (1988)
<i>Bacteroides fragilis</i>	Bfra	Amann et al., (1988)
<i>Bacteria Termofilica</i> PS3	Tb	Kagawa et al., (1986)
<i>Rhodospirillum rubrum</i>	Rr	Falk y Walker, (1985)
<i>Rhodopseudomonas blastica</i>	Rb	Tybulewicz et al., (1984)
<i>Cyanobacterium synechococcus</i> 6301	Cs	Cozens et al., (1987)
<i>Anabaena</i> sp.	Asp	Curtis, (1987)
<i>Synechococcus</i> sp.	Ssp	Cozens y Walker, (1987)
<i>Saccharomyces cerevisiae</i> (Mitocondria)	ScMt	Takeda et al., (1986)
<i>Schizosaccharomyces pombe</i>	SpMt	Falson et al., (1991)
<i>Neurospora crassa</i> (Mt)	NcMt	Rassow et al., (1990)
<i>Homo sapiens</i> (Mt de corazón)	HsMt	Ohta et al., (1988)
<i>Rattus norvegicus</i> (Mt)	RatMt	Boulet et al., (1989)
<i>Bos taurus</i> (Mt de corazón)	BtMt	Walker et al., (1985)
<i>Zea mays</i> (Mt)	ZmMt	Ysern et al., (1988)
<i>Nicotiana plumbaginifolia</i> (Mt)	NpMt	Winning et al., (1990)
<i>Marchantia polymorpha</i> (Cl)	MpCl	Boutry et al., (1985)
<i>Cebada</i> (Cl)	CebCl	Ohyama et al., 1986
<i>Zea mays</i> (Cl)	ZmCl	Ysern et al., 1988)
<i>Nicotiana tabacum</i> (Cl)	NsCl	Krebbers et al., (1982)
<i>Spinacia oleracea</i> (Cl)	SoCl	Shinotaki et al., (1983)
<i>Triticum aestivum</i> (Cl)	TeCl	Zurawski et al., (1982)
<i>Chlamydomonas reinhardtii</i> (Cl)	CrCl	Howe et al., (1985)
<i>Hordeum vulgare</i> (Cl)		Woessner et al., (1986)
<i>Ipomoea batatas</i> (Cl)	HvCl	Zurawski y Clegg et al., (1984)
<i>Oryza sativa</i> (Cl)	IbCl	Kobayashi et al., (1987)
<i>Pisum Sativum</i> (Cl)	OsCl	Hiratsuka et al., (1989)
		Zurawski et al., (1986)
Eucarlontes		
<i>Daucus carota</i> (vacuolar)	DcVr	Zimniak et al., (1988)
<i>Neurospora crassa</i> (vacuolar)	NcVr	Bowman et al., (1988a)
<i>Saccharomyces cerevisiae</i>	ScVr	Hirata et al., (1990)
Arqueobacterias		
<i>Sulfolobus acidocaldarius</i>	Sac	Denda et al., (1988b)
<i>Halobacterium salinarum</i>	Hs	Ihara y Mukohata, (1991)
<i>Methanosarcina barkeri</i>	Mb	Inatomi et al., (1989)

Tabla 5.4 Listado de organismos de donde se tomaron las secuencias de aminoácidos y nucleótidos de la subunidad α de las ATPasas F₁ en *E. coli* y sus homólogas (Mt: mitocondria; Cl: cloroplasto).

Organismo	Abreviación	Referencia
Eubacterias		
<i>Escherichia coli</i>	Ec	Walker et al., (1984)
<i>Bacillus firmus</i>	Bfir	Ivey y Krulwich et al., (1991)
<i>Bacillus megaterium</i>	Bm	Brusilow et al., (1989)
<i>Bacteria Termofilica</i> PS3?	Tb	Ohta et al., (1988)
<i>Rhodospirillum rubrum</i>	Rr	Falk et al., (19885)
<i>Rhodopseudomonas blastica</i>	Rb	Tybuliewicz et al., (1984)
<i>Anabaena</i> sp.	Asp.	Mccarn et al., (1988)
<i>Synechococcus</i> 6301	Cs	Cozens y Walker., (1987)
<i>Vibrio alginolyticus</i>	Va	Krumholz et al., (1989): Dmitirev et al., (1989)
<i>Saccharomyces cerevisiae</i> (Mt)	ScMt	Takeda et al., (1986)
<i>Schizosaccharomyces pombe</i> (Mt)	SpMt	Falson et al., (1991)
<i>Homo sapiens</i> (Mt de corazón)	HsMt	Kataoka y Biswas., (1991)
<i>Rattus norvegicus</i> (Mt)	RnMt	Lee et al., (1990)
<i>Xenopus laevis</i> (Mt)	XlMt	Weeks y Melton., (1987)
<i>Bos taurus</i> (Mt de corazón)	BtMt	Walker et al., (1989)
<i>Bos taurus</i> (Mt de hígado)	BtMth	Breen., (1988)
<i>Brassica napus</i> (Mt)	BnMt	Handa y Nakajima., (1991)
<i>Helianthus annuus</i> (Mt)	HaMt	Koehler et al., (1990)
<i>Zea mays</i> (Mt)	ZmMt	Braun y Levings., (1985): Isaac et al., (1985)
<i>Nicotiana plumbaginifolia</i> (Mt)	NpMt	Chaumont et al., (1988)
<i>Oenothera biennis</i> (Mt)	ObMt	Schuster y Brennicke., (1986)
<i>Oryza sativa</i> (Mt)	OsMt	Kadowaki et al., (1990)
<i>Pisum sativum</i> (Mt)	PsMt	Morikami y Nakamura., (1987)
<i>Phaseolus vulgaris</i> (Mt)	PvMt	Swiss protein libreria
<i>Raphanus sativus</i> (Mt)	RsMt	Makaroff et al., (1990)
<i>Triticum aestivum</i> (Mt)	TaMt	Schulte et al., (1989)
<i>Marchantia polymorpha</i> (Cl)	MpCl	Ohyama et al., (1986)
<i>Zea mays</i> (Cl)	ZmCl	Roddermel y Bogorao., (1987)
<i>Nicotiana tabacum</i> (Cl)	NsCl	Deno et al., (1984)
<i>Spinacia oleracea</i> (Cl)	SoCl	Hudson et al., (1987)
<i>Triticum aestivum</i> (Cl)	TaCl	Howe et al., (1985)
<i>Oryza sativa</i> (Cl)	OsCl	Hiratsuka et al., (1989)
<i>Pisum sativum</i> (Cl)	PsCl	Hudson et al., (1987)
Eucariontes		
<i>Homo sapiens</i> (vacuola del riñón)	HsVrr	Suedhof et al., (1989)
<i>Homo sapiens</i> (vacuola del cerebro)	HsVrc	Bernasconi et al., (1990)
<i>Arabidopsis thaliana</i> (vacuolar)	AtVr	Manolson et al., (1988)
<i>Candida tropicalis</i> (vacuolar)	CtVr	Gu et al., (1990)
<i>Neurospora crassa</i> (vacuolar)	NcVr	Bowman et al., (1988b)
Arqueobacterias		
<i>Sulfolobus acidocaldarius</i>	Sac	Denda et al., (1988a)
<i>Halobacterium salinarum</i>	Hs	Ihara y Mukohata., (1991)
<i>Methanosarcina barkeri</i>	Hb	Inatomi et al., (1989)

Tabla 5.5 Código de tres y una sola letra para los aminoácidos

Alanina	Ala	A	Glicina	Gly	G
Arginina	Arg	R	Histidina	His	H
Asparagina	Asn	N	Isoleucina	Ile	I
Ácido aspártico	Asp	D	Leucina	Leu	L
Cisteína	Cys	C	Lisina	Lys	K
Ácido glutámico	Glu	E	Metionina	Met	M
Glutamina	Gln	Q	Fenilalanina	Phe	F
Prolina	Pro	P	Serina	Ser	S
Treonina	Thr	T	Triptofano	Trp	W
Tirosina	Tyr	Y	Valina	Val	V

6.0 RESULTADOS

En la Figura 6.1 se muestran las comparaciones entre el dominio de Régnier *et al.* (1987) y los dominios de la proteína ribosomal S1 de *E. coli* y sus homólogas en las bacterias Gram negativo; la subunidad b' de *E. coli* y sus homólogas eubacterianas, que incluyen cloroplastos; la subunidad mayor de las RNA polimerasa I, II y III nucleocitoplásmicas de eucariontes y, por último, con las unidades C de arqueobacterias. Señaladas con un asterisco se indica la posición de las Gly conservadas, que comparten la proteína S1 y la PNPasa y, a su vez, las que se preservan en las RNA polimerasas. Los recuadros indican los sitios idénticos con respecto a la PNPasa y se destacan (en negrillas) las sustituciones conservadas. De la posición 27 a la 34 se localiza un octapéptido muy conservado con secuencia GGREGLVD. Con esta Tabla se muestra el grado de conservación del dominio que comparten las proteínas mencionadas anteriormente. En las RNA polimerasas este dominio corresponde a la región VI en los tres linajes celulares.

La Figura 6.2 corresponde a los perfiles de hidropatía (A y B) y de la carga eléctrica (C y D). En las gráficas A y C se compara la PNPasa (línea negra gruesa) con los cuatro dominios de la proteína ribosomal S1 de *E. coli*. Con líneas se indica la región que corresponde al octapéptido conservado. En las gráficas B y D se presenta la comparación de la PNPasa contra las RNA polimerasa

de la subunidad β' de *E. coli* y la subunidad mayor de las RNA polimerasa I, II y III de levadura (líneas delgadas) y del virus de Vaccinia (línea punteada). Se observa que el comportamiento de la polaridad en el octapéptido más conservado es igual en todos los casos, excepto en Vaccinia donde la polaridad es menor.

En las Tablas 6.3 y 6.4 se compara el octapéptido de la PNPasa y la proteína ribosomal S1 de *E. coli*, con la subunidad α y β (respectivamente) de la ATPasa F1 y sus homólogas en mitocondrias, cloroplastos, eubacterias, arqueobacterias y vacuolas. Los recuadros indican los sitios idénticos con respecto a la PNPasa y se destacan (en negrillas) las sustituciones conservadas.

RmRPS1	N	A	V	V	T	S	A	G	V	I	G	V	N	D	G	G	I	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S	D	S
P-RPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S	D	
EcRPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
RmRPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
EcRPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
RmRPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
P-RPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
EcRPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
RmRPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
P-RPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
EcRPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
RmRPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
P-RPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
EcRPS1	G	A	I	V	T	S	E	K	V	T	T	A	V	D	A	K	G	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
PNPase	H	R	V	V	T	S	A	G	V	I	G	V	N	D	G	G	I	A	T	F	V	R	L	L	V	D	H	E	D	V	G	E	A	F	I	R	A	S		
M.th.	G	A	K	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
H.h.	S	A	E	A	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
H.mo.	S	A	E	A	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
S.ac.	A	P	E	L	E	G	G	F	V	E	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
Vacc.	S	P	E	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
S.c.II	S	P	E	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
S.p.II	S	P	E	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
A.th.II	S	P	E	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
C.e.II	S	P	E	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
Hm.II	S	P	E	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
D.m.II	S	P	E	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
P.f.II	S	P	E	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
T.b.II	S	P	E	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
C.f.II	S	P	E	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
T.b.III	A	P	A	R	G	F	V	E	S	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
P.f.III	D	S	E	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
S.c.III	T	P	Q	S	R	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
S.c.I	D	A	M	A	G	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
S.p.I	S	A	R	K	S	R	G	F	V	E	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
T.b.I	R	A	R	S	F	G	F	V	E	S	S	S	Y	K	E	D	S	I	L	L	E	F	F	F	H	A	H	G	G	R	E	G	L	V	D	-	T	A	V	R
E.c.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
P.s.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
N.c.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
M.p.Ch	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
S.o.Ch	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
N.t.Ch	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
Z.m.Ch	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
O.s.Ch	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
E.g.Ch	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
P.s.Ch	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			

Figura 6.1 La secuencia GGREGLYVD es un octapéptido muy conservado, encontrado en el dominio que se une al RNA de la PNPasa de *E. coli*, la proteína ribosomal S1 y en la subunidad β' de la RNA polimerasa en *E. coli* y sus homólogas en arqueobacterias y en eucariontes. Solamente los aminoácidos idénticos en la PNPasa, S1 y en la RNA polimerasa se han enmarcado. Las sustituciones de aminoácidos conservados se muestran en negritas.

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

RmRPS1	RD - RDEQ PERFSVGHTVDR	T	NFSKKDRK	OL	-	K
P-RPS1						
EcRPS1	RV - EDA	TLVLSVGD	VEVA	KFTGV	DKR	NATISLS
RmRPS1	RPGE	WYIEEFN	RGD	VRA	VVLDV	VVKER
P-RPS1	VAGE	EAVREYK	GGDEI	RAV	LQVDA	ERE
EcRPS1	VAGE	EAVREYK	GGDEI	RAV	LQVDA	ERE
RmRPS1	KKNVHP	GLK	LST	SQEV	VVVL	LQVDP
P-RPS1	NKNIHPS	KVVNV	VG	DVEV	MLD	TD
EcRPS1	NKNIHPS	KVVNV	VG	DVEV	MLD	TD
RmRPS1	RVL - HPSE	TONIG	GOQV	LV	QIT	R
P-RPS1	RVK - HPSE	IVNV	GD	ET	VKVL	KFD
EcRPS1	RVK - HPSE	IVNV	GD	ET	VKVL	KFD
PNPase	RV - E	KVTD	YLG	MGO	EV	PVKVL
H.th.	T	A	QS	GYM	ORRL	VNA
H.h.	TS	VS	GYL	ORRL	INA	LS
H.mo.	TS	KS	GYL	ORRL	INA	LS
S.ac.	TS	QS	GYM	ORRL	INA	LS
Vacc	TS	RT	GL	ARK	IKK	MED
S.c.II	T	A	ET	GYI	ORRL	VKA
S.p.II	T	A	ET	GYI	ORRL	VKA
A.th.II	TS	ET	GYI	ORRL	VKA	MED
C.e.II	T	A	ET	GYI	ORRL	VKA
Mm.II	T	A	ET	GYI	ORRL	VKA
D.m.II	TS	ET	GYI	ORRL	VKA	MED
P.f.II	TS	ET	GYI	ORRL	VKA	MED
T.b.II	TS	DT	GYL	ORRL	VKA	MED
C.f.II	T	A	ET	GYI	ORRL	VKA
T.b.III	T	A	ET	GYI	ORRL	VKA
P.f.III	T	A	ET	GYI	ORRL	VKA
S.c.III	T	A	ET	GYI	ORRL	VKA
S.c.I	TS	RS	GYL	ORCL	TK	OLE
S.p.I	TS	RS	GYL	ORCL	TK	OLE
T.b.I	TS	RS	GYL	ORCL	TK	OLE
E.c.	T	A	NS	GYL	TRRL	VVD
P.s.	T	A	NS	GYL	TRRL	VVD
N.c.	T	A	NS	GYL	TRRL	VVD
H.p.Ch	TS	DA	GYL	TRRL	VVD	VS
S.o.Ch	TS	DA	GYL	TRRL	VVD	VS
N.t.Ch	TS	DA	GYL	TRRL	VVD	VS
Z.m.Ch	T	A	DA	GYL	TRRL	VVD
O.s.Ch	TS	DA	GYL	TRRL	VVD	VS
E.g.Ch	T	A	DA	GYL	TRRL	VVD
P.s.Ch						

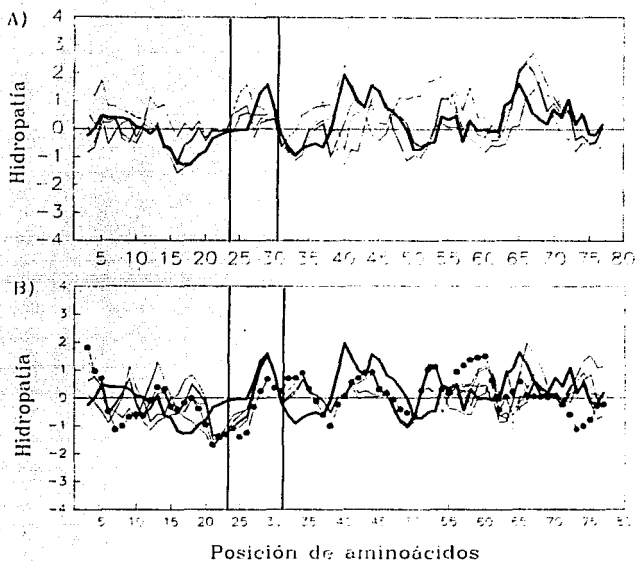


Figura 5.2

- A) y C) Comparación de los perfiles de hidropatía, con la escala de Hopp y Woods (1981) de los dominios que unen RNA monocatenario de las secuencias de PNPase (línea gruesa) y las cuatro repeticiones de la proteína ribosomal S1 (línea delgada). Las líneas verticales muestran la región de la secuencia GGREGLVH.
- B) y D) Comparación de los perfiles de la carga eléctrica con la escala de Hopp y Woods (1981) de los dominios que unen RNA monocatenario de las secuencias de la PNPase (línea gruesa, *E. coli*), de las RNA polimerasas DNA dependientes de origen celular (línea delgada, Sa; Sc I, II y III y Ec) y RNA polimerasa DNA-dependiente del Virus de vaccinia (línea con círculos negros). Las líneas verticales muestran la región de la secuencia GGREGLVH.

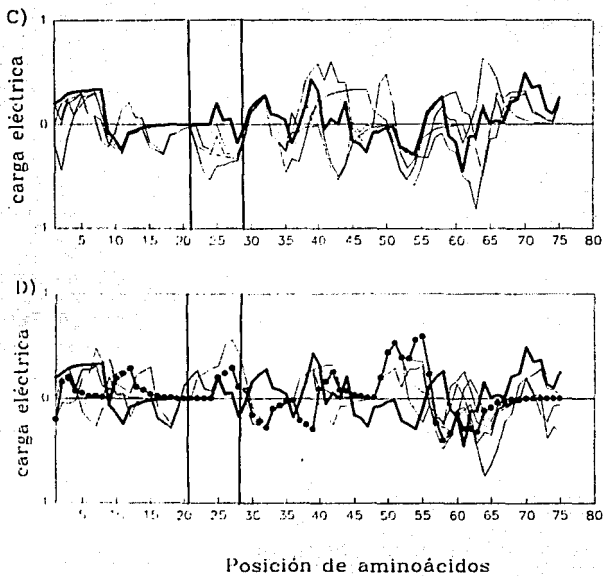


Tabla 6.3 Alineamiento del octapéptido de la PNPasa con una parte de las secuencias de las ATPasas F1 de la subunidad α que corresponde a la región homóloga al sitio de Rossmann. Sólo se alinea el octapéptido donde los residuos enmarcados son sitios conservados y los residuos en negritas indican sustituciones conservadas.

RPS1	ATVEL-----A	D	G	V	E	G	Y	L	RASEASRDV-
RPS1	IFIGL-----D	G	G	I	D	G	L	V	HLSDISWHVAG
RPS1	CFVEI-----E	F	G	V	E	G	L	V	HYSEMDWTNKN
RPS1	AFVDL-----	G	G	V	D	G	L	L	HJTDMAWKRVK
PNPasa	AFVAI-----G	G	G	K	E	G	L	V	HISQIADKRV-
Ec	KAVDSMIPGR	G	G	R	E	L	I	I	GDROTGKTALA
Ss	TAIDAMIPGR	G	G	R	E	L	I	I	GDROTGKTALA
T.A	KAIDALVPIGR	G	G	R	E	L	I	I	GDROTGKTSA
Bm b	KAIDALVPIGR	G	G	R	E	L	I	I	GDROTGKTSA
Va	KSVDSMIPGR	G	G	R	E	L	I	I	GDROIGKTALA
Rb	KAVDAMIPGR	G	G	R	E	L	I	I	GDROTGKTAVA
Rr	KAIIDSLIPGR	G	G	R	E	L	I	I	GDROTGKTAVI
BF	KSIDTHIPGR	G	G	R	E	L	I	I	GDROTGKTALA
BnM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTIIA
Rm	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTIIA
HaM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTALA
ZmM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTALA
NpM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTALA
HsM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTSA
BtM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTSA
RnM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTSA
SpM	KAIIDSHVPIGR	G	G	R	E	L	I	I	GDROTGKTALA
ScM	KAVDALVPIGR	G	G	R	E	L	I	I	GDROTGKTAVA
Odm	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTALA
OsM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTALA
PsM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTALA
PvM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTALA
TaM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTALA
XtM	KAVDSLVPGR	G	G	R	E	L	I	I	GDROTGKTSA
A. spC	TAIDSMIPGR	G	G	R	E	L	I	I	GDROTGKTALA
ZmC	IAIDSMIPGR	G	G	R	E	L	I	I	GDROTGKTAVA
MpC	IAIDSMIPGR	G	G	R	E	L	I	I	GDROTGKTAVA
OSC	IAIDSMIPGR	G	G	R	E	L	I	I	GDROTGKTAVA
PsC	IAIDSMIPGR	G	G	R	E	L	I	I	GDROTGKTAVA
SoC	IAIDAMIPVGR	G	G	R	E	L	I	I	GDROTGKTAVA
NtC	IAIDSMIPGR	G	G	R	E	L	I	I	GDROTGKTAVA
TaC	IAIDSMIPGR	G	G	R	E	L	I	I	GDROTGKTAVA
AtV57	STIDVHNSIAR	G	G	K	I	P	I	F	SAAGLPHNEIA
CtV57	SAIDTHNSIAR	G	G	K	I	P	I	F	SASGLPHNEIA
HsV58	SPIDVHNSIAR	G	G	K	I	P	I	F	SAAGLPHNEIA
NcV57	SAIDTHNSIAR	G	G	K	I	P	I	F	SAAGLPHNEIA
ScV57	SAIDTHNSIAR	G	G	K	I	P	I	F	SASGLPHNEIA
SaB	SAIDGLNSLLR	G	G	S	K	I	T	D	LSGSLPANTLA
Mba	RILDGLFPVAK	G	G	T	A	A	I	P	GPFGSGKTVTO
Hsa	RILDGLFPIAK	G	G	T	A	A	I	P	GPFGSGKTVTO

ATPasa
Eubacteriana

ATPasa
Mitocondrial

ATPasa de
Cianobacteria y
Cloroplasto

ATPasa
Vacuolar

ATPasa
Arqueobacteriana

Tabla 6.4

Alineamiento del octapéptido de la PNPasa con una parte de las secuencias de la subunidad β catalítica de la ATPasas F1 que corresponde a la región homóloga al sitio de Rossmann. Solo se alinea el octapéptido, donde los residuos enmarcados son sitios conservados y los residuos en negritas indican sustituciones conservadas.

RPS1	ATVEL-----A	D	G	I	E	G	Y	L	RASEASDRV-
RPS1	IFIGL-----D	G	I	D	G	L	V	HLSDISWVAG	
RPS1	CFVEI-----E	E	G	V	E	G	L	VHVSEMDITNKN	
RPS1	AFVDL-----G	G	V	D	G	L	L	HITDMANKRYK	
PNPasa	AFVAI-----G	G	G	R	E	G	L	VHISOTADKRV-	
Ec	KVIDLMCPFAK	G	G	K	V	G	L	GGAGVGKTVNM	
S.sp	KVIDLLAPYRQ	G	G	K	I	G	L	GGAGVGKTVLI	
T.B	KVDVLLAPYIK	G	G	K	I	G	L	GGAGVGKTVLI	
Bm	KVDVLLAPYIK	G	G	K	I	G	L	GGAGVGKTVLI	
Va	KVIDLICPFAK	G	G	K	I	G	L	GGAGVGKTVNM	
Rb	KVIDLLAPYSK	G	G	K	I	G	L	GGAGVGKTVLI	
Rr	KVIDLIAPYTK	G	G	K	V	G	L	GGAGVGKTVLI	
Bf1	KVDVLLAPYIK	G	G	K	I	G	L	GGAGVGKTVLI	
Bf	KVIDLLEPYSK	G	G	K	I	G	L	GGAGVGKTVLI	
Hv	KVDVLLAPYRR	G	G	K	I	G	L	GGAGVGKTVLI	
C1	KVIDLIEPYAK	G	G	K	I	G	L	GGAGVGKTVLI	
ZmM	KVDVLLAPYOR	G	G	K	I	G	L	GGAGVGKTVLI	
NpM	KVDVLLAPYOR	G	G	K	I	G	L	GGAGVGKTVLI	
HsM	KVDVLLAPYAK	G	G	K	I	G	L	GGAGVGKTVLI	
BtM	KVDVLLAPYAK	G	G	K	I	G	L	GGAGVGKTVLI	
RnM	KVDVLLAPYAK	G	G	K	I	G	L	GGAGVGKTVLI	
SpM	KVDVLLAPYAR	G	G	K	I	G	L	GGAGVGKTVFI	
ScM	KVDVLLAPYAR	G	G	K	I	G	L	GGAGVGKTVFI	
NcM	KVDVLLAPYAR	G	G	K	I	G	L	GGAGVGKTVFI	
A.SP	KVDVLLTPYRR	G	G	K	I	G	L	GGAGVGKTVIM	
ZmC	KVDVLLAPYRR	G	G	K	I	G	L	GGAGVGKTVLI	
MpC	KVDVLLAPYRR	G	G	K	I	G	L	GGAGVGKTVLI	
OsC	KVDVLLAPYRR	G	G	K	I	G	L	GGAGVGKTVLI	
PsC	KVDVLLAPYRR	G	G	K	I	G	L	GGAGVGKTVLI	
SoC	KVVNLLAPYRR	G	G	K	I	G	L	GGAGVGKTVLI	
NtC	EVDVLLAPYRR	G	G	K	I	G	L	GGAGVGKTVLI	
TaC	KVDVLLAPYRR	G	G	K	I	G	L	GGAGVGKTVLI	
CrC	KVDVLLAPYRP	G	G	K	I	G	L	GGAGVGKTVLI	
IbC	KVDVLLAPYRR	G	G	K	I	G	L	GGAGVGKTVLI	
DCV69	RVLDAIFPSVL	G	G	T	C	A	I	PAFGCGKTVIS	
NCV67	RVLDAIFPSVO	G	G	T	V	A	I	PAFGCGKTVIS	
SCV67	RVLDAIFPCVO	G	G	T	T	C	I	PAFGCGKTVIS	
Saa	RVLDTVFPIAK	G	G	T	A	A	I	PAPFGSGKTVTL	
MtB	STIDGNTLVR	G	Q	K	L	P	I	FSGSGLPHNTLA	
MbB	STIDGNTLVR	G	Q	K	L	P	I	FSGSGLPHNEIA	
HsB	SGIDGMNTLVR	G	Q	K	L	P	I	FSGSGLPHSDLA	
ATPasa									
Bacteriana									
ATPasa									
Mitochondrial									
ATPasa de									
Cianobacteria y									
Cloroplasto									
ATPasa									
Vacuolar									
ATPasa									
Arqueobacteriana									

ATPasa
Bacteriana

ATPasa
Mitocondrial

ATPasa de
Cianobacteria y
Cloroplasto

ATPasa
Vacuolar

ATPasa
Arqueobacteriana

7.0 DISCUSIÓN Y CONCLUSIONES

En este trabajo se encontró que la RNA polimerasa y la PNPasa comparten un dominio homólogo de 69 aminoácidos, que a su vez es homólogo a las cuatro repeticiones de la proteína ribosomal S1 (ver Figura 6.1). Dicho dominio fue reportado originalmente por Régnier et al. (1987), y está presente en la secuencia de la polinucleótido fosforilasa (PNPasa).

El dominio de Régnier et al. (1987) se comparó con las secuencias de la subunidad β' de la RNA polimerasa DNA dependiente y sus homólogas, se encontró que también está presente. Además, se determinó que dentro del dominio existe un octapéptido un posible (motivo funcional), cuya secuencia es GGREGLVD. Este motivo está muy conservado en los tres linajes celulares dentro de la subunidad β' de la RNA-polimerasa DNA dependiente de *E.coli* y sus homólogas. Finalmente, al comparar el sitio funcional contra los bancos de secuencias se determinó que la secuencia GGREGLVD es similar a la secuencia que se encuentra presente cerca del sitio de unión del nucleótido en la subunidad β de la ATPasa F1 (ver Figura 6.3). Los perfiles de hidropatía realizados muestran que el dominio tiene características polares similares (Ver Figura 6.2).

Cuando el grupo de Régnier *et. al.* (1987) encontró que un segmento de la PNPasa es homólogo a cuatro dominios repetidos de la proteína ribosomal S1, se discutió la posibilidad de que esta secuencia de 69 aminoácidos tuviera las mismas propiedades funcionales que sus homólogas en S1, incluyendo el de unirse a polirribonucleótidos. En este sentido Régnier *et. al.* (1987) argumentaron que el dominio encontrado puede presentar propiedades similares a través del complejo macromolecular que forma el trimero de subunidades α , que es la forma activa; así, es posible que cada dominio que está presente en la subunidad haga las veces de cada una de las repeticiones que se presentan en la proteína ribosomal S1 pudiendo así unirse al RNA, con la diferencia que ésta última presenta un dominio de más, que se sabe es prescindible para su funcionamiento (Régnier *et al.*, 1987). Entre los cuatro segmentos de S1 y la PNPasa existen conservados 6 Glicinas y 9 aminoácidos hidrofóbicos, lo que hace pensar que se puede plegar de la misma manera que los dominios de S1. Al compararse con la subunidad β' de la RNA polimerasa de *E. coli* y sus homólogas, se ve que se conservan 6 Glicinas, 5 de ellas en la misma posición, y uno en diferente posición, compartido sólo con la PNPasa y no con S1. Además, se comparten 16 residuos hidrofóbicos o sinónimos de éstos. Por lo tanto, se puede pensar que este dominio también adoptaría una estructura secundaria igual a la S1 y la PNPasa.

Debido a que ambas proteínas interaccionan con moléculas de RNA, como también sucede con la subunidad β' (y sus homólogas) de las RNA polimerasas celulares, se ha sugerido que este dominio funciona para unirse al RNA que está recién sintetizado por la PNPasa o que será degradado por ella (Régner et al., 1987). Una explicación propuesta por Régner et al. (1987) supone que estos dominios evolucionaron a partir de la duplicación de un gen que se codificaba para unirse con un polirribonucleótido. Esta última hipótesis es coherente con el hallazgo del dominio homólogo en la subunidad β' de *E.coli* de la RNA polimerasa DNA-dependiente reportado en este trabajo.

La secuencia GGREGLVD forma parte de uno de los dominios más conservados de la subunidad β' de la RNA polimerasa DNA dependiente de *E.coli* (eubacterias) y sus homólogas celulares, en arqueobacterias y en eucariontes (Sawadogo y Sentenac 1990). En especial de la subunidad mayor de las RNA polimerasa II de eucariontes (Mémet et al., 1988a, b; Cornelissen et al., 1988).

De hecho, el motivo GREGL(I/V)DTAVKT, se ha encontrado que es el segundo bloque más conservado en todas las RNA polimerasas de origen. Este motivo es parte de la región VI de esta misma proteína, la cual ya había sido señalado por Mémet et al. (1988a, b) como un sitio muy conservado, y a la cual no se ha atribuido ninguna función (ver Tabla 3.4.2). Se sabe también que ésta región es compartido por un dominio donde la toxina α -amanitina

inhibe la elongación, por lo que se presume que la región VI interviene en la elongación (Jokerst et al., 1989) (ver Tabla 3.4.2). Este dominio se encuentra presente en las RNA polimerasas de los tres linajes celulares (eucariontes, eubacterias y arqueobacterias).

Al analizar la secuencia del motivo GGREGLVLD se puede observar lo siguiente que: de los tres linajes celulares, son las eubacterias quienes conservan menos este motivo, presentando, por ejemplo, la secuencia GARKGLAD. En eucariontes se presentan dos casos: el primero se observa en la RNA polimerasa II de los kinetoplástidos (T.b.II), y en el apicomplejo (P.f.II), en la levadura (S.p.II), en las RNA polimerasas III de T.b.III, en *Plasmodium* (P.f.III) y en la levadura (S.c.III) y en la RNA polimerasa I de T.b.I y las levaduras S.c.I y S.p.I. La secuencia de la variación es (A/S)GREGL(V/I)D. En el segundo caso se presenta la secuencia GGREGLVLD, y se observa en el resto de las RNA polimerasas II de los organismos que aparecen en la Tabla 6.1. Las arqueobacterias comparten este último motivo. La única secuencia de este motivo homólogo que difiere radicalmente en los aminoácidos que la componen es la del virus de Vaccinia, el que presenta la secuencia VARSQSTD, aunque como se sabe, la subunidad mayor de la RNA polimerasa de este virus es homóloga a la subunidad más grande de la RNA polimerasas II de eucariontes y a β' de *E. coli*. (Broyles y Moss, 1986).

Tabla 7.1 Comparación de las diferencias de la composición de aa del octapéptido GGKGLVH. * como a: S.p., T.b. y C.f.

a) PNPasa	G G K E G L V H
b) Archeobacterias y eucariontes RNA pol. II	G G R E G L V D
c) Eubacterias	G A R K G L A D
d) Eucariontes RNA pol. I y III, y algunas II*	A G R E G L I D
e) Vaccinia	V A R S O S T D

Como se puede observar en la Tabla 7.1; y 6.1, los aminoácidos más conservados del motivo son aquellos que presentan la mayor polaridad que son sustituidas por un equivalente polar, como es el caso de la H (por D), o la K (por R). El aminoácido más conservado es la glicina en la posición 5, seguido de la arginina en la posición 3. En tercer lugar están las glicinas de las posiciones 1 y 2. La secuencia de vaccinia no cumple con el patrón de las otras RNA polimerasas. Tal vez la RNA polimerasa del virus puede adoptar una secuencia alternativa que le dé el mismo efecto funcional. Para comprobar lo anterior se calculo un índice de variación que permite cuantificar que tanto un aminoácido ha sido cambiado, no considera las sutituciones conservativas. En la Tabla 7.2 se obseva que aquellas posiciones que tiene un valor bajo son las más conservadas como son los casos de las posiciones: 3 con la arginina (R), 5 con la glicina

Tabla 7.2 Comparación del índice de variabilidad según Alizon et al., (1986), del octapeptido GGKGLVH.

Posición	1	2	3	4	5	6	7	8
Índice $V_{(x)} = \left(\frac{D-1}{A/N} \right)$	3.33	1.66	1.25	3.33	1.25	1.25	7.5	1.25

(G), 6 con la leucina (L) y con el ácido aspártico (D). Como se ve los aminoácidos más conservados son dos polares como es R y D, y no polares como es la G y L. Quedando en consecuencia con la conclusión anterior.

Es probable que aquellos aminoácidos de cada grupo que no se conservan (como es el caso de las Gly en las posiciones I y II, que se sustituyen por A ó S) no representan necesariamente un cambio en la actividad funcional que realiza el motivo. Por ejemplo, se ha visto que el motivo de Rossmann puede tener algunos aminoácidos cambiados, sin que ello implique una alteración en la unión del nucleótido (no siempre aparecen las tres Gly) (Hsu *et al.*, 1987; Saraste *et al.*, 1990; Kaziro *et al.*, 1991). No obstante, valdría la pena realizar estudios experimentales que puedan derimir este punto. Como la sustitución de una Gly por Ala o por Ser en aquellas secuencias de proteínas que no tiene sustituidas sus Gly.

Hasta este momento no se había atribuido ninguna función a este motivo en la literatura científica. Sin embargo, en este trabajo se encontró que tanto la región IV como el dominio al que corresponde son homólogos al dominio de 69 aa de la PNPasa (Régner *et al.*, 1987). A su vez, este último dominio se repite cuatro veces en la proteína ribosomal S1 de las bacterias Gram negativo. La función de ésta última proteína sí está caracterizada (Régner *et al.*, 1987). Además, existen muchas

evidencias que sugieren que la subunidad β' de *E.coli* tiene la función de elongar el RNA (Lazcano, 1986; Lazcano et al., 1988a y b; 1992a; Jøkerst et al., 1989). Debido a lo anterior, se propone que el dominio de 69 aa de la RNA polimerasa β' celular y las proteínas homólogas se une al polirribonucleótido que está siendo elongado. En la alineación de las tres proteínas mencionadas se observa que el octapéptido GGREGLVD es el motivo más conservado (ver Tabla 6.1). Se puede suponer, por lo tanto, que el octapéptido GGREGLVD forma parte del sitio de unión al polímero de RNA recién sintetizado en la transcripción.

En efecto, el motivo GGREGLVD tiene las propiedades químicas que le permiten establecer una unión con los nucleótidos del polímero. Este octapéptido tiene un ambiente polar, por lo que en principio resulta idóneo para interactuar con ribonucleótidos y polirribonucleótidos. Por ejemplo, presenta Gly y Arg/Lys, que son características de motivos que se unen a nucleótidos (Doolittle, 1988). La Leu y la Val, por su parte, pueden contribuir a estas interacciones debido a que son aminoácidos neutros, no interponiendo cargas electrostáticas. Cabe mencionar que tanto la pareja Arg/Lys, como el Glu y el Asp. Los primeros son aminoácidos polares que pueden interaccionar con las cargas negativas del fosfato de los nucleótidos a través y los segundos lo pueden hacer a través de un metal como Mg^{++} o Mn^{++} . La Gly es el aminoácido que está presente con más frecuencia en todas las

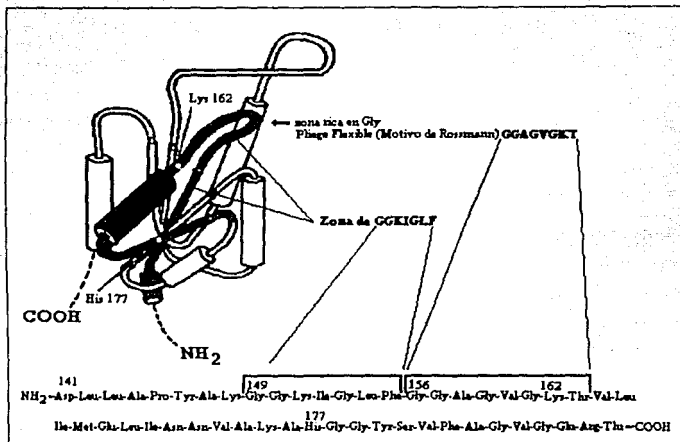


Figura 7.1 Gráfica que muestra donde se encuentra el octapéptido conservado en un modelo de ATPasas F₁. (tomada de Garboezi et al., 1986).

proteínas, pero también siempre se le encuentra asociado a sitios de unión a nucleótidos, como ocurre con el motivo de Rossmann. Esto se puede deber a que éste residuo de aminoácidos permite que el polipéptido tenga plegamientos, cooperando así a definir la estructura terciaria

Por otra parte, se comparó la secuencia GGREGLVLD contra los bancos de secuencias "PIR" (de proteínas) y "Genbank" (de nucleótidos). Se encontró que dicho octapéptido está presente en la secuencia de la subunidad β de la ATPasa F₁ celular de origen eubacteriano. A ésta última se le atribuye la función catalítica (ver Tabla 6.3 en los Resultados). El sitio de unión a los

nucleótidos de la subunidad β de la ATPasa F₁ de *E. coli* se ha determinado a través del estudio de algunas mutantes. Para ello, se han comparado la subunidad β de la ATPasas. Según Fry *et al* (1985, 1986) se ha determinado que el sitio de unión al ATP se encuentra en la secuencia consenso GXGXXG. Esta, a su vez, se ha identificado en la adenilato cinasa (Argos y Leberman 1985), cuya estructura terciaria sí se conoce. De la comparación del motivo aquí estudiado se desprende que este se encuentra muy cercano al sitio de unión al ATP (GXGXXG). Finalmente, Duncan, *et al* (1986) y Hsu, *et al* (1987) han propuesto que la Lys 162 de la Figura 7.1 se puede encontrar muy cercana al sitio de unión para ATP.

En este trabajo se ha observado que las subunidades β en las ATPasas F₁ de origen eubacteriano presentan el motivo GGREGLV. Sin embargo, cabría preguntarse, porqué no sucede lo mismo para las ATPasas F₁ de origen eucariote y sus homólogas arqueobacterianas. En primer lugar, la ausencia en ambos linajes es consistente con las que indican la cercanía filogenética de arqueobacterias con eucariontes (Gogarten *et al.*, 1989; Iwabe *et al.*, 1989). En segundo término, su ausencia se podría explicar por que se tratara de una regulación diferencial y que se haya propagado hasta la actualidad.

De acuerdo al modelo estructural propuesto por Fry *et al.* (1985 y Garboezi *et al.*, 1986), para la subunidad β de la ATPasa tipo F₁, el motivo que han descrito se encontraría como parte de un doblez

en la estructura terciaria de la proteína (ver Figura tomada de Garboezi *et al.*, 1986) donde se observa que se encuentra cerca del sitio donde se une al fosfato del β ATP).

La secuencia GGREGLVD que se estudió en este trabajo es adyacente al motivo que se había descrito previamente como parte del sitio de unión del ATP en las ATPasas GGPGSGK(T/S) en la subunidad β (Duncan *et al.*, 1986). Es posible, por lo tanto, que esta secuencia sea también un componente hasta ahora ignorado del sitio con el que estas enzimas se unen al ATP. Esta conclusión es congruente con la hipótesis de que en la subunidad β de la ATPasa el sitio de unión al ATP es mas grande, basado en el grado de conservación en el extremo C-terminal de la subunidad catalítica de la ATPasa F1- β (donde se encuentra el motivo de Rossmann, GXGXXG) (Ysern *et al.*, 1988, Senior, 1990). Por ello, cobra mayor relevancia el hecho de que los dos motivos, el descrito en este trabajo y el GXGXXG son adyacentes.

La comparación evolutiva de sitios y dominios funcionales siempre conlleva un riesgo, particularmente cuando se trata de sitios de unión a ribonucleótidos y polirribonucleótidos. En este caso, el riesgo radica en el tamaño reducido de las secuencias comparadas. Este problema ha sido discutido en detalle por Argos y Leberman (1985), quienes han comprobado que se puede validar la similitud de una secuencia tan corta a través de la comparación contra todos los bancos de secuencias existentes tanto para genes como

para proteínas. Ejemplos de ello son los casos comprobados para una secuencia de 12 aminoácidos a la que Doolittle (1988) diagnosticó como transferrina, y para el motivo de Rossmann en otras proteínas. Como ya se mencionó, este último ha sido reconocido como un sitio de unión a nucleótidos. Es importante señalar lo anterior, ya que fue gracias a la comparación con dos bancos de datos que en este trabajo se pudo llegar a la conclusión de que el motivo se encuentra presente en el sitio de unión de la subunidad β de la ATPasa F₁. Al buscar en el "GenBank" no se detectó la similitud en la subunidad, pero al comparar con el "PIR", sí se encontró que había identidades con el motivo arriba mencionado. Cabe recalcar, entonces, que es recomendable buscar en los dos tipos de bancos debido a que la sensibilidad tanto del algoritmo como de la constitución en sí del tipo de banco puede dar como resultado hallazgos que sean complementarios, o quizás dé hallazgos contradictorios.

POSIBLES IMPLICACIONES FUNCIONALES DE LA EVOLUCION DE LOS DOMINIOS QUE SE UNEN A RNA

Según la hipótesis de Lazcano (1986, 1988a y 1992), las primeras proteínas que surgieron en células basadas en RNA fueron polipéptidos que tenían la función de ligarse a nucleótidos. En esta forma estabilizaban al RNA catalítico incrementando, así, la eficiencia de la catálisis.

Aplicando esta idea, a continuación se propone un modelo de replicación del material genético para las células basadas en RNA en el arqueano:

Se sabe que el virus Q β requiere de la proteína ribosomal S1 para replicarse eficientemente. Esta proteína la obtiene al secuestrarla de su hospedero. Se sospecha que esta proteína tiene la función de estabilizar la estructura primaria del genoma de RNA. De esta manera queda expuesto el molde pudiéndose así copiar. Si tomamos los resultados de este trabajo y se les aplica la hipótesis de Lazcano (1986), se podría postular un modelo tentativo que explicase un posible mecanismo primitivo de replicación, a través de la existencia de dominios que se uniesen a polirribonucleótidos catalíticos en el mundo del RNA. Este podría ser el caso del dominio encontrado en la subunidad β' de la RNA polimerasa DNA dependiente de *E.coli* y sus contrapartes en PNPasa, así como las repeticiones que se presentan a lo largo

de la proteína ribosomal S1. El dominio anterior podría contribuir a aumentar la capacidad de replicación del genoma basado en RNA, funcionando de la siguiente manera: el polipéptido o un conjunto de éstos tendría la capacidad de desestabilizar la estructura secundaria y terciaria del RNA, como es el caso de S1. En consecuencia, el RNA quedaría estabilizado en su estructura primaria. Así podría mantener expuesto el molde del RNA el tiempo suficiente para que se lleve a cabo la replicación del genoma con mayor facilidad. Este dominio facilitaría, además, que el proceso se llevara a cabo de manera más eficiente. Se puede pensar que este mecanismo, al facilitar la replicación, la haría más rápida y que en consecuencia, el dominio se seleccionaría sobre otros. Las células que hubieran presentado este mecanismo también se seleccionarían debido al aumento de su capacidad reproductiva. En esta forma la población con esta característica predominaría sobre las demás poblaciones que no lo posean.

8.0 REFERNCIÀS

- 1) Alberts B. M., Bray D., Lewis J., Raff M., Roberts K., Watson J. D., (1983) *Molecular Biology of The Cell*. 1ª Edition, Garland Publishing. N.Y. 1146pp.
- 2) Alberts B. M. (1986) The function of the hereditary materials: biological catalyses reflect the cell's evolutionary history. *Amer. Zool.* 26: 781-796.
- 3) A'hearn J. M. Jr., Bartolomei M. S., West M. L., Cisek L. J. y Corden J. L. (1987) Cloning and sequence analysis of the mouse genomic locus encoding the largest subunit of RNA polymerase II. *J. Biol. Chem.* 262(22): 10695-10705.
- 4) Alizon M., Wain-Hobson S., Montagnier L. y Sonigo P. (1986) Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* 46(1):63-74.
- 5) Allison L.A., Moyle M., Shales M., e Ingles C. J. (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* 42(2):599-610.
- 6) Allison L. A., Wong J. K., Fitzpatrick V. D., Moyle M. e Ingles C. J. (1988) The C-Terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. *Mol. Cell. Biol.* 8(1):321-329
- 7) Altman S. (1984) Aspects of biochemical catalysis. *Cell* 36(2):237-239.
- 8) Amann R., Ludwig W., Schleifer K.H. (1988) β -subunit of ATP-synthase: a useful marker for studying the phylogenetic relationship of eubacteria. *J. Gen. Microbiol.* 134:2815-2821.
- 10) Argos P. y Leberman R. (1985) Homologies and anomalies in primary structural patterns of nucleotide binding proteins. *Eur. J. Biochem.* 152(3): 651-656.
- 11) Argos P. (1988) A sequence motif in many polymerases. *Nucleic Acids Res.* 16(21):9909-9916.
- 12) Armaleo D. (1987) Structure and evolution of prokaryotic and eukaryotic RNA polymerases: a model. *J. Theoret. Biol.* 127(3):301-314.

- 13) Armaleo D. (1988) A note on RNA polymerase phylogeny. *J. Theoret. Biol.* **131**(1):135-136.
- 14) Azuma Y., Yamagishi M., Ueshima R., y Ishihama A. (1991) Cloning and sequence determination of the *Schizosaccharomyces pombe* *rpbl* gene encoding the largest subunit of RNA polymerase II. *Nucleic Acids Res.* **19**(3):461-468.
- 15) Bandziulis R. J., Swanson M. S. y Dreyfuss G. (1989) RNA-binding proteins as developmental regulators. *Genes & Develop.* **3**(4):431-437.
- 16) Baron M., Norman D. G. y Campbell I. D. (1991) Protein modules. *Trends Biochem. Sci.* **16**(1):13-17.
- 17) Beal R. B., Pillai R. P., Chuknyisky P. P., Levy A., Terien E. y Eichhorn G. (1990) Structural studies on the active site of *Escherichia coli* RNA polymerase 2. Geometrical relationship of the interacting substrates. *Biochemistry*. **29**:5994-6002.
- 18) Behe M. J. (1990) Histone deletion mutants challenge the molecular clock hypothesis. *Trends Biochem. Sci.* **15**(10):374-376.
- 20) Berg J. M. (1988) Proposed structure for the zinc-binding domain from transcription factor IIIA and related proteins. *Proc. Natl. Acad. Sci. USA* **85**(1):99-102.
- 21) Berghöfer B., Kröckel L., Körtner C., Truss M., Schallenberg J. y Klein A. (1988) Relatedness of archaeobacterial RNA polymerase core subunits to their eubacterial and eukaryotic equivalents. *Nucleic Acids Res.* **16**(16):8113-8128.
- 22) Bergsland K. J. y Haselkorn R. (1991) Evolutionary relationships among eubacteria, cyanobacteria and chloroplasts: evidence from the *rpoCI* gene of *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **173**(11):3446-3455.
- 23) Bernasconi P., Rausch T., Gogarten J.P., Taiz L. (1989) The H⁺ ATPase regulatory subunit of *Methanoccus thermolithotrophicus*: amplification of an 800bp fragment by polymerase chain reaction. *FEBS Lett.* **251**(1-2):132-136.
- 24) Bernasconi P., Rausch T., Struve I., Morgan L., Taiz L. (1990) An mRNA from human brain encodes an isoform of the B subunit of the vacuolar H⁺-ATPase. *J. Biol. Chem.* **265**(29):17428-17431.
- 25) Bhargava P. y Chatterji D. (1988) Spectroscopic studies on the mode of binding of ATP, UTP and α -amanitin with yeast RNA polymerase II. *FEBS Lett.* **241**(1-2):33-37.

- 26) Biggs J., Searles L. L. y Greenleaf A. L. (1985) Structure of the eukaryotic transcription apparatus: features of the gene for the largest subunit of *Drosophila* RNA polymerase II. *Cell* 42(2):611-621.
- 27) Bird D. M. y Riddle D. L. (1989) Molecular cloning and sequencing of *ama-1*, the gene encoding the largest subunit of *Caenorhabditis elegans* RNA Polymerase II. *Mol. Cell. Biol.* 9(10):4119-4130.
- 28) Blackburn G. M. y Gait M. J. (eds.) (1990) *Nucleic acids in chemistry and biology*. IRL Press. (Oxford University Press, Oxford). 745pp.
- 29) Boni I. V., Isaeva D. M., Masychenko M. L. y Tzareva N. V. (1991) Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucleic Acids Res.* 19(1):155-162.
- 30) Boulet D., Poirier J., y Cote C. (1989) Studies on the biogenesis of the mammalian ATP synthase complex: isolation and characterization of a full-length cDNA encoding the rat F₁- β -subunit. *Biochem. Biophys. Res. Commun.* 159(3):1184-1190.
- 31) Boutry M., y Chua N.H. (1985) A nuclear gene encoding the beta subunit of the mitochondrial ATP synthase in *Nicotiana plumbaginifolia*. *EMBO J.* 4(9):2159-2165.
- 32) Bowman B. J., Allen R., Wechsler M. A. y Bowman E. J. (1988b) Isolation of genes encoding the *Neurospora* vacuolar ATPase. Analysis of *vma-2* encoding the 57-kDa polypeptide and comparison to *vma-1*. *J. Biol. Chem.* 263(28):14002-14007.
- 33) Bowman E. J., Tenney K., y Bowman B.J. (1988a) Isolation of the genes encoding the *Neurospora* vacuolar ATPase. Analysis of *vma-1* encoding the 67-kDa subunit reveals homology to other ATPases. *J. Biol. Chem.* 263(28):13994-14001.
- 34) Braun C. J. y Levings C. S. III (1985) Nucleotide sequence of the F₁-ATPase α subunit gene from maize mitochondria. *Plant Physiol.* 79(1-2):571-577.
- 35) Breen G. A. (1988) Bovine liver cDNA clones encoding a precursor of the α -subunit of the mitochondrial ATP synthase complex. *Biochem. Biophys. Res. Commun.* 152(1):264-269.
- 36) Breen G. A., Holmans P. L. y Garnett K. E. (1988) Isolation and characterization of a complementary DNA for the nuclear-coded precursor of the β -subunit of bovine mitochondrial F₁-ATPase. *Biochemistry.* 27(11):3955-3961.

- 37) Broyles S. S. y Moss B. (1986) Homology between RNA polymerases of poxviruses, prokaryotes and eukaryotes: nucleotide sequence and transcriptional analysis of vaccinia virus genes encoding 147-kDa and 22 kDa subunits. *Proc. Natl. Acad. Sci. USA* 83(10):3141-3145
- 38) Brusilow W. S., Scarpetta M. A., Hawthorne C. A., y Clark W. P. (1989) Organization and sequence of the genes coding for the proton-translocating ATPase of *Bacillus megaterium*. *J. Biol. Chem* 264(3): 1528-1533
- 39) Busch H., Reddy R., Rothblum L. y Choi Y. C. (1982) SnRNA, SnRNPs and RNA processing. *Annu. Rev. Biochem.* 51:617-654.
- 40) Cech T. R., y Bass B. L., (1986) Biological catalysis by RNA. *Annu. Rev. Biochem.* 55:599-629.
- 41) Cohet-Meilhac M. y Chambon P. (1974) Animal DNA-dependent RNA polymerases II. Mechanism of the inhibition of RNA polymerase B by amatoxins. *Biochim. Biophys. Acta* 353(2):160-184.
- 42) Corden J. L. (1990) Tails of RNA Polymerase II. *Trends Biochem. Sci.* 15(10):383-387.
- 43) Cornelissen A. W., Evers K. y Köck J. (1988) Structure and sequence of genes encoding subunits of eukaryotic RNA polymerases. In: Maclean N. (ed) *Oxford Surveys on Eukaryotic Genes*. (Oxford University Press, Oxford) 5:91-131.
- 44) Cozens A. L. y Walker J. E. (1986) Pea chloroplast DNA encodes homologues of *Escherichia coli* ribosomal subunit S2 and β' -subunit of RNA polymerase. *Biochem J.* 236(2):453-460.
- 45) Cozens A. L. y Walker J. E. (1987) The organization and sequence of genes for ATP synthase subunits in the cyanobacterium *Synechococcus* 6301. Support for an endosymbiotic origin of chloroplasts *J. Mol. Biol.* 194 (3):359-383.
- 46) Creighton T. E. (1984) Proteins: structures and molecular properties. Freeman and Company, San Francisco. 514pp.
- 47) Crick F. H. C. (1958) On protein synthesis. Symposia of the Society for Experimental Biology 12:138-163.
- 48) Cross R. L. (1988) The number of functional catalytic sites on F1-ATPases and the effects of quaternary structural asymmetry on their properties. *J. Bioenerg. Biomembr.* 20(4):395-405.

- 49) Curtis S.E. (1987) Genes encoding the β and ϵ subunits of the proton-translocating ATPase from *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **169**(1):80-86.
- 50) Chaumont F., Boutry M., Briquet M., y Vassarotti A. (1988) Sequence of the gene encoding the mitochondrial F₁-ATPase alpha subunit from *Nicotiana plumbaginifolia*. *Nucleic Acids Res.* **16**(13):6247-6247.
- 51) Chothia C. y Lesk A. M. (1987) The evolution of protein structures. *Cold Spring Harbor Symposia on Quantitative Biology*. **52**:399-405
- 52) Dani M., Benatti L., Lorenzetti R., Martini D., Minganti C., Sassano M., Sidoli A. y Soria M. (1988) A DNA sequence from *Saponaria officinalis* is similar to various RNA polymerase genes. *Nucleic Acids Res.* **16**(7): 3103.
- 53) Danilkovich A. V., Borodin A. M., Allkmets R. L., Rostapshov V. M., y Chernov I. P. (1988) Nucleotide sequence of the *rpoC* gene coding for RNA polymerase β' subunit of *Pseudomonas putida*. *Dokl. Akad. Nauk USSR* **303**(1):241-245.
- 54) Darwin C. (1859) *Origin of Species*. Avenel Books New York.
- 55) Delarue M., Poch O., Tordo N., Moras D. y Argos P. (1990) An attempt to unify the structure of polymerases. *Protein Eng.* **3**(6):461-467.
- 56) Denda K., Kinishi J., Oshima T., Date T. y Yoshida M. (1988a) The membrane-associated ATPase from *Sulfolobus acidocaldarius* is distantly related to F₁ as assessed from the primary structure of its α -subunit *J. Biol. Chem.* **263**(13):6012-6015.
- 57) Denda K., Kinishi J., Oshima T., Date T. y Yoshida M. (1988b) Molecular cloning of the β -subunit of a possible non-FoF₁ type ATP synthase from the acidothermophilic archaeobacterium, *Sulfolobus acidocaldarius*. *J. Biol. Chem.* **263**(33):17251-17254.
- 58) Deno H., Shinozaki K., y Sugiura M. (1983) Nucleotide sequence of tobacco chloroplast gene for the α subunit of proton-translocating ATPase. *Nucleic Acids Res.* **11**(7):2185-2191.
- 59) Deno H. y Sugiura M. (1984) Chloroplast tRNA^{Glu} gene contains a long intron in the D stem: nucleotide sequence of tobacco chloroplast genes for tRNA^{Glu} (UCC) and tRNA^{Glu} (UCU). *Proc. Natl. Acad. Sci. USA* **81**:405-408.
- 60) Devereux J., Haeberli P. y Smithies O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**(1):387-395.

- 61) Diener T. O. (1984) Portraits of viruses: the viroid. *Intervirology* 22(1):1-16.
- 62) Dietz G. W. Jr., y Grunberg-Manago M. G. (1967) *Clostridium perfringens* polynucleotide phosphorylase: two molecular species. *Biochem. Biophys. Res. Commun.* 28(2):146-151.
- 63) Dmitriev O. Yu., Grinkevich V. A., y Skulachev V. P. (1989) The F₁-ATPase of *Vibrio alginolyticus*. Purification and N-terminal sequence of major subunits. *FEBS Lett.* 258:219-222
- 64) Doolittle R. F. (1985) The genealogy of some recently evolved vertebrate proteins. *Trends Biochem. Sci.* 10(6):233-237.
- 65) Doolittle R. F. (1988) *OF URFS AND ORFS*. A primer on how to analyze derived amino acid sequences. (University Science Books, La Jolla California) 1-103pp.
- 66) Doolittle R. F. (1989) Similar amino acid sequences revisited *Trends Biochem. Sci.* 14(7):244-245.
- 67) Doolittle W. F. (1987) What introns have to tell us: hierarchy in genome evolution. *Cold Spring Harbor Symposia on Quantitative Biology*, 52: 907-913.
- 68) Draper D. E. y von Hippel P. H. (1978a) Nucleic acid binding properties of *Escherichia coli* ribosomal protein S1. Structure and interactions of binding site I. *J. Mol. Biol.* 122(3):321-338.
- 69) Draper D. E. y von Hippel P. H. (1978b) Nucleic acid binding properties of *Escherichia coli* ribosomal protein S1. Co-operativity and specificity of binding site II. *J. Mol. Biol.* 122(3):339-359.
- 70) Duncan T. M., Parsonage D. y Senior A. E. (1986) Structure of the nucleotide-binding domain in the β -subunit *Escherichia coli* F₁-ATPase. *FEBS. Lett.* 208(1):1-6.
- 71) Eggink G., Engel H., Vriend G., Terpstra P. y Witholt B. (1990) Rubredoxin reductase of *Pseudomonas oleovorans* structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. *J. Mol. Biol.* 212(1):135-142.
- 72) Evans R. M. y Hollenberg S. M. (1988) Zinc fingers: gilt by association. *Cell* 52(1):1-3.

- 73) Evers R., Hammer A., Köck J., Jess W., Borst P., Mémet S. and Cornelissen A. W. (1989a) *Trypanosoma brucei* contains two RNA polymerase II largest subunit genes with an altered C-terminal domain. *Cell* 56(4): 585-597.
- 74) Evers R., Hammer A. y Cornelissen A. W. (1989b) Unusual C-terminal domain of the largest subunit of RNA polymerase II of *Crithidia fasciculata*. *Nucleic Acids Res.* 17(9):585-597.
- 75) Falk G., Hampe A., y Walker J. E. (1985) Nucleotide sequence of the *Rhododpirillum rubrum* *atp* operon. *Biochem. J.* 228(2):391-407.
- 76) Falson P., Maffey L., Conrath K., y Boutry M., (1991a) α subunit of mitochondrial F₁-ATPase from the fission yeast. Deduced sequence of the wild type and identification of a mutation that alters apparent negative cooperativity. *J. Biol. Chem.* 266(1):287-293.
- 77) Falson P., Maffey L., Conrath K., y Boutry M., (1991b) β subunit of mitochondrial F₁-ATPase from the fission yeast. Deduced sequence of the wild type protein and identification of a mutation that increases nucleotide binding. *Eur. J. Biochem.* 200(1):61-67.
- 78) Feng D. F. y Doolittle R. F. (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* 25(4): 351-360.
- 79) Francino P. (1992) Origen común de las polinucleótido fosforilasas de eubacterias y arqueobacterias. Tesis de Licenciatura (Biología). Facultad de Ciencias. UNAM. (México)
- 80) Fry D. C., Kuby S. A. y Mildvan A. S. (1985) NMR studies of the MgATP binding site of adenylate kinase and of a 45-residue peptide fragment of the enzyme. *Biochemistry* 24(17):4680-4694.
- 81) Fry D. C., Kuby S. A., y Mildvan A. S. (1986) ATP-binding site of adenylate kinase: mechanistic implications of its homology with *ras*-encoded p21, F₁-ATPase, and nucleotide-binding proteins. *Proc. Natl. Acad. Sci. USA* 83(4):907-911.
- 82) Futai H., Noumi I. y Maeda M. (1989) ATP synthase (H⁺-ATPase): results by combined biochemical and molecular biological approaches. *Annu. Rev. Biochem.* 58:111-136.
- 83) Gaber R. F., Styles C. A. y Fink G. R. (1988) *TRK1* encodes a plasma membrane protein required for high-affinity potassium transport in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 8(7):2848-2859.

- 84) Garboczi D. N., Fox A. H., Gerring S. L. y Pedersen P. L. (1988a) β -subunit of rat liver mitochondrial ATP synthase: cDNA cloning, amino acid sequence, expression in *Escherichia coli*, and structural relationship to adenylate kinase. *Biochemistry* 27(2):553-560.
- 85) Garboczi D. N., Shenbagamurthi P., Kirk W., Hulihan J. y Pedersen P. L. (1988b) Mitochondrial ATP synthase interaction of a synthetic 50-amino acids, β -subunit peptide with ATP. *J. Biol. Chem.* 263(2):812-816.
- 86) Gardner M. J., Williamson D. H. y Wilson R. J. (1991) A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. *Mol. Biochem. Parasitol.* 44(1):115-123.
- 87) Gautheron D. C. y Godinot C. (1986) Evidence from immunological studies of structure-mechanism relationship of F₁ and F₁F₀. *J. Bioenerg. Biomembr.* 20(4):451-468.
- 88) Geiduschek E. P. y Tocchini-Valentini G. P. (1988) Transcription by RNA polymerase III. *Annu. Rev. Biochem.* 57:873-914.
- 89) Ghislain M., Schlessner A. y Goffeau A. (1987) Mutation of a conserved glycine residue modifies the vanadate sensitivity of the plasma membrane H⁺-ATPase from *Schizosaccharomyces pombe*. *J. Biol. Chem.* 262(36):17549-17555.
- 90) Gilbert W. (1986) The RNA World. *Nature* 319:618.
- 91) Gilbert W. (1987) The exon theory of genes. *Cold Spring Harbor Symposia on Quantitative Biology.* 52:901-905.
- 92) Giorginis S. y Subramanian A. R. (1980) The major ribosome binding site of *Escherichia coli* ribosomal protein S1 is located in its N-terminal segment. *J. Mol. Biol.* 141(4):393-408.
- 93) Godefroy-Colburn T. y Grunberg-Manago M. (1972) Polynucleotide phosphorylase. In: Boyer, P. D. (Ed) *The Enzymes*. (Academic Press, New York) 7:533-574.
- 94) Goebel S. J., Johnson G.P., Perkus M. E., Davis S. W., y Wilson J. P. Paoletti E. (1990) The complete DNA sequence of vaccinia virus. *J. Virology* 179(1):247-266 y 517-563.
- 95) Gogarten J. P., Kibak H., Dittrich P., Taiz L., Bowman E. J., Bowman B. J., Manolson M. F., Poole R. J., Date T., Oshima T., Konishi J., Denda K. y Yoshida M. (1989) Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. USA* 86(17):6661-6665.

- 96) Gould S. J. y Eldredge N. (1977) Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3:115-151.
- 97) Gould S. J. (1986) Evolution and the triumph of homology. or why history matters. *Amer. Sci.* 74:60-69.
- 98) Gu H.H., Gallagher M.J., Rupkey S., y Dean G.E. (1990) Gene and derived peptide sequences for *C. tropicalis* vacuolar ATPase subunit B. *Nucleic Acids Res.* 18(24):7446.
- 99) Guerrier-Takada C., Gardiner K., Marsh T., Pace N., y Altman S. (1983) The RNA miety of ribonuclease P is the catalytic subunit of the enzyme. *Ce11*(3)35:849-857.
- 100) Gujssani A. y Portier C. (1976) Study on the structure-function relationship of polynucleotide phosphorylase: model of a proteolytic degraded polynucleotide phosphorylase. *Nucleic Acids Res.* 3(11):3015-3024.
- 101) Grachev M. A., Mustáev A. A., Zaychikov E. F., Lindner A. J. y Hartmann G. R. (1989) Localization of the binding site for the initiating substrate on the RNA polymerase from *Sulfolobus acidocaldarius*. *FEBS Lett.* 250(2):317-322.
- 102) Grunberg-Manago M. (1989) Recollections on studies of polynucleotide phosphorylase: a commentary on 'Enzymic Synthesis of Polynucleotids. I. Polynucleotide Phosphorylase of *Azobacter vinelandii*. *Biochim. Biophys. Acta* 1000:59-64.
- 103) Handa H., y Nakajima K. (1991) Nucleotide sequence and transcription analyses of the rapeseed (*Brassica napus* L.) mitochondrial F1-ATPase α -subunit gene. *Plant. Mol. Biol.* 16(2):361-364.
- 104) Hansen F. G., Hansen E. B., Atlung T. (1985) Physical mapping and nucleotide sequence of the *rnpA* gene that encodes the protein component of ribonuclease P in *Escherichia coli*. *Gene* 38(1-3):85-93.
- 105) Harrison S. C. (1991) A structural taxonomy of DNA-binding domains. *Nature* 353:715-719.
- 106) Hawthorne C. A. y Brusilow W.S. (1988) Sequence of the genes for the β and ϵ subunits of the ATP Synthase of *Bacillus megaterium* QM B1551. *Biochem. Biophys. Res. Commun.* 151(2):926-931.
- 107) Helmann J. D. y Chamberlin M. J. (1988) Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* 57:839-872.

- 108) Higo K., Otaka E., y Osawa S. (1982) Purification and characterization of 30S ribosomal proteins from *Bacillus subtilis*: correlation to *Escherichia coli* 30S proteins. *Mol. Gen. Genet.* 185(2):239-244.
- 109) Hirata R., Ohsumi Y., Nakano A., Kawasaki H., Susuki K., y Anraku Y. (1990) Molecular structure of a gene, *VMA1*, Encoding the catalytic subunit of H⁺-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 265(12): 6726-6733.
- 110) Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M., Mori M., Kondo C., Honji Y., Sun C. R., Meng B. Y., Li Y. Q., Kanno A., Nishizawa Y., Hirai A., Shinozaki K., y Sugiura M. (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* 217(2-3):185-194.
- 111) Hopp T. P. y Woods K. R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* 78(6):3824-3828.
- 112) Howe C.J., Fearnley J. M., Walker J. E. Dyer T. A., y Gray J. C. (1985) Nucleotide sequences of the genes for the α , β , and ϵ subunits of wheat chloroplast ATP synthase. *Plant Mol. Biol.* 4(6):333-345.
- 113) Hsieh W. T. (1971) Polymerization of deoxyribonucleoside diphosphates with an enzyme from an *Escherichia coli* mutant lacking deoxyribonucleic acid polymerase activity. *J. Biol. Chem.* 246(6):1780-1784.
- 114) Hsu S-Y., Noumi T., Takeyama M., Maeda M., Ishibashi S. y Futai M. (1987) β -subunit of *Escherichia coli* F₁-ATPase. An amino acid replacement within a conserved sequence (G-X-X-X-X-G-K-T/S) of nucleotide-binding proteins. *FEBS Lett.* 218(2):222-226.
- 115) Hudson G. S., Mason J. G., Holton T. A., Koller B., Cox G. B., Whitfield P. R., y Bottomley W. (1987) A gene cluster in the spinach and Pea chloroplast genomes encoding one cfi and three cfo subunits of the H⁺-ATP synthase complex and the ribosomal protein S2. *J. Mol. Biol.* 196(2): 283-298.
- 116) Hudson G. S., Holton T. A., Whitfield P. R. y Bottomley W., (1988) Spinach chloroplast *rpoBC* genes encode three subunits of the chloroplast RNA polymerase. *J. Mol. Biol.* 200(4):639-654.
- 117) Huxley A.F. (1983) How far will Darwin take us? IN: D.S. Bendall (ed) *Evolution from Molecules to Man*. Cambridge, University Press, p.3-19.

- 118) Igloi G. L., Meinke A., Döry I. y Kössel H. (1990) Nucleotide sequence of the maize chloroplast *rpo* B/C1/C2 operon: Comparison between the derived protein primary structures from various organisms with respect to functional domains. *Mol. Gen. Genet.* 221(3):379-394.
- 119) Ihara K. y Mukohata Y. (1991) The ATP synthase of *Halobacterium salinarum* (halobium) is an archaeobacterial type as revealed from the amino acid sequences of its two major subunits. *Arch. Biochem. Biophys.* 286(1):111-116.
- 120) Inatomi K. I., Eya S., Maeda M., y Futai M. (1989) Amino acid sequence of the α and β subunits of *Methanosarcina barkeri* ATPase deduced from cloned genes. Similarity to subunits of eukaryotic vacuolar and FoF₁-ATPases. *J. Biol. Chem.* 264(19):10954-10959.
- 121) Isaac P. G., Brennicke A., Dunbar S. M., y Leaver C. J., (1985) The mitochondrial genome of fertile maize (*Zea mays* L.) contains two copies of the gene encoding the alpha-subunit of the F₁-ATPase. *Curr. Genet.* 10(4):321-328.
- 122) Isono K. e Isono S (1976) Lack of ribosomal protein S1 in *Bacillus stearothermophilus*. *Proc. Natl. Acad. Sci. USA* 73(3):767-770.
- 123) Ivey D. M. y Krulwich T. A. (1991) Organization and nucleotide sequence of the *atp* genes encoding the ATP synthase from alkaliphilic *Bacillus firmus* 0F4. *Mol. Gen. Genet* 229(2):292-300.
- 124) Iwabe N., Kuma K-I., Hasegawa M., Osawa S. y Miyata T. (1989) Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. USA* 86(23):9355-9359.
- 125) Jess W., Hammer A. y Cornelissen A. W. (1989) Complete sequence of the gene encoding the largest subunit of RNA polymerase I of *Trypanosoma brucei*. *FEBS Lett.* 249(1):123-128.
- 126) Jess W., Paml P., Evers R., Köck J. y Cornelissen A. W. (1990) Phylogenetic analysis of the RNA polymerases of *Trypanosoma brucei*, with special reference to class-specific transcription. *Curr. Genet.* 18(6):547-551.
- 127) Jockerst R. S., Weeks J. R. Zehring W. A. y Greenleaf A. L. (1989) Analysis of the gene encoding the largest subunit of RNA polymerase II in *Drosophila*. *Mol. Gen. Genet.* 215(2):266-275.
- 128) Joyce G. F. (1989) RNA evolution and the origins of life. *Nature* 338 (6212):217-224.

- 129) Juddson H. F., (1980) *The eighth day of creation*. Simon & Schuster, Inc. N.Y. EUA. 746 pp.
- 130) Kadowaki K., Kazama S. Suzuki T. (1990) Nucleotide sequence of the F₁-ATPase α subunit gene from rice mitochondria. *Nucleic Acids Res.* 18(5):1302.
- 131) Kagawa Y., Ishizuka M., Saishu T., Nakao S. (1986) Stable structure of thermophilic proton ATPase beta subunit. *J. Biochem. Tokyo* 100(4): 923-934.
- 132) Kanazawa H., Kayano T., Kiyasu T., y Futai M. (1982) Nucleotide sequence of the genes for β and ϵ subunits of proton-translocating ATPase from *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 105(4):1257-1264.
- 133) Kataoka H. y Biswas C. (1991) Nucleotide sequence of a cDNA for the α subunit of human mitochondrial ATP synthase. *Biochem. Biophys. Acta* 1089(3):393-395.
- 134) Kaziyo Y., Itoh H., Kozasa T., Nakafuku M. y Satoh A. T. (1991) Structure and function of signal-transducing GTP-binding proteins. *Annu. Rev. Biochem.* 60:349-400.
- 135) Kemp B. E. y Pearson R. B. (1990) Protein kinase recognition sequence motifs. *Trends Biochem. Sci.* 15(9):342-346.
- 136) Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart R. G., Davies D. R., Phillips D. C., y Shore V. C. (1960) Structure of Myoglobin. A three-dimensional fourier synthesis at 2Å resolution. *185(4711):422-427.*
- 137) Kitakawa M. e Isono K. (1982) An amber mutation in the gene *rpsA* for Three-dimensional fourier synthesis at 2Å. resolution. *Nature* ribosomal protein S1 in *E. coli*. *Mol. Gen. Genet.* 185(3):445-447.
- 138) Klee C. B. (1969) The proteolytic conversion of polynucleotide phosphorylase to a primer-dependent form. *J. Biol. Chem.* 244(10):2558-2566.
- 139) Klug A. y Rhodes D. (1987) Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harbor Symposia on Quantitative Biology.* 52:473-482.
- 140) Kobayashi K., Nakamura K., Asahi T. (1987) cF₁ ATPase β - and ϵ -subunit genes are separated in the sweet potato chloroplast genome. *Nucleic Acids Res.* 15(17):7177.

- 141) Köhler R. H., Lössel A., y Zetscher K. (1990) Nucleotide sequence of the F₁-ATPase α subunit gene of sunflower mitochondria. *Nucleic Acids Res.* 18(15):4588.
- 142) Köck J., Evers R. y Cornelissen A. W. (1988) Structure and sequence of the gene for the largest subunit of trypanosomal RNA polymerase III. *Nucleic Acids Res.* 16(18):8753-8772.
- 143) Kolb A., Hermoso J. M., Thomas J. O. y Szer W. (1977) Nucleic acid helix-unwinding properties of ribosomal protein S1 and the role of S1 in mRNA binding to ribosomes. *Proc. Natl. Acad. Sci. USA* 74(6):2379-2383.
- 144) Krebbers E. T., Larrinua I. M., McIntosh L., y Bogorad L. (1982) The maize chloroplast genes for the β and ϵ subunit of the photosynthetic coupling factor cF₁ are fused. *Nucleic Acids Res.* 10(16):4985-5002.
- 145) Kruger K., Grabowski P. J., Zaug A. J., Sands J., Gottschling D. E. y Cech T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31(1):147-157.
- 146) Krumholz L. R., Esser U., y Simoni R. D. (1989) Nucleotide sequence of the *unc* operon of *Vibrio alginolyticus*. *Nucleic Acids Res.* 17(19):7993-7994.
- 147) Krumholz L. R., Esser U., Simoni R.D. (1990) Characterization of the H⁺-pumping F₁F₀ ATPase of *Vibrio alginolyticus*. *J. Bacteriology* 172(12): 6809-6817.
- 148) Lambert D. M. y Hughes A. J. (1988) Keywords and concepts in structuralist and functionalist biology. *J. Theor. Biol.* 133:133-145.
- 149) Laughrea M. y Moore P. B. (1978) Ribosomal components required for binding protein S1 to the 30S subunit of *Escherichia coli*. *J. Mol. Biol.* 122(1):109-112.
- 150) Lazcano A. (1986) Prebiotic evolution and the origin of cells. *Trends Soc. Cat. Biol.* 39:73-103
- 151) Lazcano A., Guerrero R., Margulis L., y Oró J. (1988a) The evolution transition from RNA to DNA in early cells. *J. Mol. Evol.* 27(4):283-290.
- 152) Lazcano A., Fastag J., Gariglio P., Ramirez C. y Oró J. (1988b) On the early evolution of RNA polymerase. *J. Mol. Evol.* 27(4):365-376.

- 153) Lazcano A., Llaca V., Fox G.E. y Oró J. (1989) A classification of RNA polymerases based on their evolutionary relatedness. 6th ISSOL Meeting (Prague, Czechoslovakia).
- 154) Lazcano A., Llaca V., Cappello R., Valverde V. y Oró J. (1992a) The origin and early evolution of nucleic acid polymerases. *Adv. Space Res.* **12(4)**:(4)207-(4)216.
- 155) Lazcano A., Fox G. E. y Oró J. (1992b) Life before DNA: the origin and evolution of early archaean cells. In Mortlock R. P. (ed), *The evolution of metabolic function* (CRC Press, Boca Raton.), 237-295 pp.
- 156) Lee J. H., Gaarboeci D. N., Thomas P. J., y Pedersen P. L. (1990) Mitochondrial ATP synthase. cDNA cloning, amino acid sequence, overexpression, and properties of the rat liver α subunit. *J. Biol. Chem.* **265**(8):4664-4669.
- 157) Leffers H., Gropp F., Lottspeich F., Zillig W. y Garrett R. A. (1989) Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles *Halobacterium halobium* and *Halococcus morrhuae*. *J. Mol. Biol.* **206**(1):1-17.
- 158) Lehninger A. L. (1975) *Biochemistry*. 2^o ed. Worth Publishers. N.Y. 1104 pp.
- 159) Lehrach H., Schäfer K., y Scheit K. H. (1971) Concerning the subunit structure of polynucleotide phosphorylase from *E. coli*. *FEBS Lett.* **14**(5):343-345.
- 160) Li W-B, Bzik D. J., Gu H., Tanaka M., Fox B. A. y Inselburg J. (1989) An enlarged subunit of *Plasmodium falciparum* RNA polymerase II define conserved and variable RNA polymerase domains. *Nucleic Acids Res.* **17** (23):9621-9636.
- 161) Li W-B, Bzik D. J., Tanaka M., Gu H. M., Fox B. A. y Inselburg J. (1991) Characterization of the gene encoding the largest subunit of *Plasmodium falciparum* RNA polymerase III. *Mol. Biochem. Parasitol.* **46**(2):229-239.
- 162) Li W-H. y Graur D. (1991) *Fundamentals of Molecular Evolution*. Sinauer Assoc. Sunderland, Massachusetts. 284 pp.
- 163) Linderström-Lang K. U. y Schellman J. A. (1959) Protein structure and enzyme activity. In: Boyer, P. D. (Ed) *The Enzymes* Academic Press. New York. 1:444-510.

- 164) Lipman D. J. y Pearson W. R. (1985) Rapid and sensitive protein similarity searches. *Science* 227(4693):1435-1441.
- 165) Lisitsyn N. A., Monastyrskaya G. S. y Sverdlov E. D. (1988) Genes coding for RNA polymerase β subunit in bacteria. Structure/function analysis. *Eur. J. Biochem.* 177(2):363-369.
- 166) Little M. C. y Hallick R. B. (1988) Chloroplast *rpoA*, *rpoB*, and *rpoC* genes specify at least three components of a chloroplast DNA-dependent RNA polymerase active in tRNA and mRNA transcription. *J. Biol. Chem.* 263(28):14302-14307.
- 167) Llaca V. (1988) Caracterización de la actividad de replicasa y determinación de la subunidad catalítica de la RNA polimerasa DNA-dependiente de *Escherichia coli*. Tesis de licenciatura (Biología) Facultad de Ciencias UNAM. (México) 95pp.
- 168) Lochrie M. A. y Simon M. I. (1988) G protein multiplicity in eukaryotic signal transduction systems. *Biochemistry* 27(14):4957-4965.
- 169) Logan K. A., Dahmus M. E. y Bradbury E. M. (1988) Histones H10 and H5 share common epitopes with RNA Polymerase II. *J. Biol. Chem.* 263(20):9658-9662.
- 170) Louis B. G., Peterkin P. I. y Fitt P. S. (1971) Nucleic acid enzymology of extremely halophilic bacteria. Gel-filtration and density-gradient-centrifugation studies of the molecular weights of *Halobacterium cutirubrum* polynucleotide phosphorylase and deoxyribonucleic acid and ribonucleic acid-dependent ribonucleic acid polymerase. *Biochem. J.* 121(4):635-641.
- 171) Makaroff C. A., Apel I. J., y Palmer J. D. (1990) Characterization of radish mitochondrial *atpA*: influence of nuclear background on transcription of *atpA*-associated sequence and relationship with male sterility. *Plant Mol. Biol.* 15(5):735-746.
- 172) Manolson M.F., Ouellette B. F., Filion M., y Poole R. J. (1988) cDNA sequence and homologies of the "57-kDa" nucleotide-binding subunit of the vacuolar ATPase from *Arabidopsis*. *J. Biol. Chem.* 263(34):17987-17994.
- 173) Martin C., Okamura S. y Young R. (1990) Genetic exploration of interactive domains in RNA Polymerase II subunits. *Mol. Cell. Biol.* 10(5):1908-1914.
- 174) Mayr E. (1982) The growth of biological thought: Diversity, Evolution, and Inheritance. Harvard University Press, Cambridge, Mass. 977 pp.

- 175) McCarn D. F., Whitaker R. A., Alam J., Vrba-J. M., y Curtis S. E. (1988) Genes encoding the alpha, gamma, delta and four Fo subunits of ATP synthase constitute an operon in the Cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **170**(8):3448-3458.
- 176) McEnery M. W., Hüllihem J. y Pedersen P. L. (1989) Fo "proton channel" of rat liver mitochondria. Rapid purification of a functional complex and a study of its interaction with the unique probe diethylstilbestrol. *J. Biol. Chem.* **264**(20):12029-12036.
- 177) Melnick D.J. (1990) Molecules, evolution and time. *Trends Ecol. Evol.* **5**(6):171-173.
- 178) Hémet S., Saurin W. y Sentenac A. (1988a) RNA polymerases B and C are more closely related to each other than to RNA polymerase A. *J. Biol. Chem.* **263**(21):10048-10051.
- 179) Hémet S., Gouy M., Mark C., Sentenac A. y Buhler J-M. (1988b) *RPA190*, the gene coding for the largest subunit of yeast RNA polymerase A. *J. Biol. Chem.* **263**(6):2830-2839.
- 180) Mildvan A. S. (1989) NMR studies of the interactions of substrates with enzymes and their peptide fragments. *FASEB J.* **3**(6):1705-1714.
- 181) Moradi-Ameli M., Clerc F. F., Cieur F., Seiberras G., Godinot C. (1991) Localization on the mitochondrial F₁ ATPase α subunit of an epitope masked in the membrane-bound enzyme using a monoclonal antibody and synthetic peptides. *Eur. J. Biochem.* **199**(3):671-676.
- 182) Morikami A., y Nakamura K. (1987) Structure and expression of pea mitochondrial F₁ ATPase α -subunit gene and its pseudogene involved in homologous recombination. *J. Biochem. Tokyo* **101**(4):967-976.
- 183) Moulton F. R. y Schifferes J. J. (1947) Autobiografía de la Ciencia. Ed. Fondo de Cultura Económica. México. 642 pp.
- 184) Murray C. L. y Rabinowitz J. C. (1982) Nucleotide sequences of transcription and translation initiation regions in *Bacillus* phage ϕ 29 early genes. *J. Biol. Chem.* **257**(2):1053-1062.
- 185) Nawrath C., Schell J. y Koncz C. (1990) Homologous domains of the largest subunit of eucaryotic RNA polymerase II are conserved in plants. *Mol. Gen. Genet.* **223**(1):65-75.
- 186) Nelson H., Mandiyan S. y Nelson H. (1989) A conserved gene encoding the 57-kDa subunit of the yeast vacuolar H⁺-ATPase. *J. Biol. Chem.* **264**(3):1775-1778.

- 187) Nelson N. y Taiz L. (1989) The evolution of H⁺-ATPases. *Trends Biochem. Sci.* **14**(3):113-116.
- 188) Noller H. F. (1984) Structure of ribosomal RNA. *Annu. Rev. Biochem.* **53**: 119-162.
- 189) Oeser B. (1988) S2 plasmid from *Zea Mays* encodes a specific RNA polymerase: an alternative alignment. *Nucleic Acids Res.* **16**(17):8729
- 190) Ohta S., Yohda M., Ishizuka M., Hirata H., Hamamoto T., Otawara-Hamamoto Y., Matsuda K., y Kagawa Y. (1988a) Sequence and over-expression of subunits of adenosine triphosphate synthase in thermophilic bacterium PS3. *Biochim. Biophys. Acta* **933**(1):141-155.
- 191) Ohta S., Tomura H., Matsuda K., y Kagawa Y. (1988b) Gene structure of the human mitochondrial adenosine triphosphate synthase β subunit. *J. Biol. Chem.* **263**(23):11257-11262.
- 192) Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S., Umesono K., Shiki Y., Takeuchi M., Chang Z., Aotas S-I, Inokuchi H. y Ozeki H. (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* **322**:572-574.
- 193) Oparin A. I. (1938) *Origin of life*. Macmillan, New York. 270 pp.
- 194) Ovchinnikov Yu. A., Monastyrskaya G. S., Gubanov, V. V., Guryev S. O., Salomatina I. S., Shuvaeva T. M. Lipkin V. M. y Sverdlov E. D. (1982) The primary structure of *E. coli* RNA polymerase, nucleotide sequence of the *rpoC* gene and amino acid sequence of the β' -subunit. *Nucleic Acids Res.* **10**(13):4035-4044.
- 195) Patthy L. (1985) Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. *Cell* **41**(3):657-663.
- 196) Pauling L., Corey R. B. y Branson H. R. (1951) The Structure of proteins: two hydrogen bonded helical configurations of the polypeptide. *Proc. Natl. Acad. Sci. USA* **37**:205-211.
- 197) Phillips D.C., Sternberg M.J.E. y Sutton B.J. (1983) Intimations of evolution from the three-dimensional structures of proteins. In: Bendall, D. S. (ed). *Evolution from molecules to men*. Cambridge University Press, Great Britain. p. 145-174.
- 198) Poch O., Sauvaget I., Delarue M. y Tordo N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8**(12):3867-3874.

- 199) Portier C., (1975) Quaternary structure of polynucleotide phosphorylase from *Escherichia coli*: evidence of a complex between two types of polypeptide chains. *Eur. J. Biochem* 55(1-2):573-582.
- 200) Pühler G., Leffers H., Gropp F., Palm P., Klenk H-P., Lottspeich F., Garrett R. A. y Zillig W. (1989a) Archaeobacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proc. Natl. Acad. Sci. USA* 86(12):4569-4573.
- 201) Pühler G., Lottspeich F. y Zillig W. (1989b) Organization and nucleotide sequence of the genes encoding the large subunits A, B and C of the DNA-dependent RNA polymerase of the archaeobacterium *Sulfolobus acidocaldarius*. *Nucleic Acids Res.* 17(12):4517-4534.
- 202) Quivey R.G. Jr., Faustoferri R.C., Belli W.A., y Flores J.S. (1991) Polymerase chain reaction amplification, cloning, sequence determination and homologies of streptococcal ATPase-encodings DNAs. *Gene* 97(1):63-68.
- 203) Rasso J., Harmey M. A., Muller H. A., Neupert W., y Tropschug M. (1990) Nucleotide sequence of a full-length cDNA coding for the mitochondrial precursor protein of the β -subunit of F₁-ATPase from *Neurospora crassa*. *Nucleic Acids Res.* 18(16):4922.
- 204) Reanney D. C. (1982) The evolution of RNA viruses. *Annu. Rev. Microbiol.* 36:47-73.
- 205) Régnier P., Grunberg-Manago M. y Portier C. (1987) Nucleotide sequence of the *pnp* gene of *Escherichia coli* encoding polynucleotide phosphorylase: homology of the primary structure of the protein with the RNA-binding domain of ribosomal protein S1. *J. Biol. Chem.* 262(1):63-68.
- 206) Rodermeil S.R. y Bogorad L. (1987) Molecular evolution and nucleotide sequences of the maize plastid genes for the α subunit of cF₁(*atpA*) and the proteolipid subunit of cF_o (*atpH*). *Genetics* 116(1):127-139.
- 207) Rogalski T. M., Bullerjahn A. M. y Riddle D. L. (1988) Lethal and amanitin-resistance mutations in the *Caenorhabditis elegans ama-1* and *ama-2* genes. *Genetics* 120(2):409-422.
- 208) Rossmann M.G., Liljas A., Brändén C-I., y Banaszak L. J. (1975) Evolutionary and structural relationships among dehydrogenase. In: P.D. Boyer (ed). *The Enzymes* Academic Press, New York. 11:61-102.
- 209) Rossmann M. G. y Argos P. (1981) Protein folding. *Annu. Rev. Biochem.* 50:497-532.

- 210) Rozen F., Pelletier J., Tracnsel H. y Sonenberg N. (1989) A lysine substitution in the ATP-binding side of eucaryotic initiation factor 4A abrogates nucleotide-binding activity. *Mol. Cell. Biol.* 9(9):4061-4063.
- 211) Sanger F. y Tuppy H. (1951) The amino acid sequence in the phenylalanyl chain of insulin. *Biochem J.* 49:463-490.
- 212) Saraste M., Sibbald P.R. y Wittinghofer A. (1990) The P-loop-a common motif in ATP-and GTP-binding proteins *Trends Biochem. Sci.* 15(11):430-434.
- 213) Sawadogo M. y Sentenac A. (1990) RNA polymerase B (II) and general transcription factors. *Annu. Rev. Biochem.* 59:711-754.
- 214) Schleif R. (1988) DNA binding by proteins. *Science* 241(4870):1182-1187.
- 215) Schallenberg J., Moes M., Truss M., Reiser W., Thomm M., Stetter K. O. y Klein A. (1988) Cloning and physical mapping of RNA polymerase genes from *Methanobacterium thermoautotrophicum* and comparison of homologies and gene orders with those of RNA polymerase genes from other methanogenic archaeobacteria. *J. Bacteriol.* 170(5):2247-2253.
- 216) Schneider G. J. y Hasekorn R. (1986) RNA polymerase subunit homology among cyanobacteria, other eubacteria and archaeobacteria. *J. Bacteriol.* 170(9):4136-4140.
- 217) Schnier J. e Isono K. (1982) *rpsA* sequence of the gene *rpsA* of *Escherichia coli* coding for ribosomal protein S1. *Nucleic Acids. Res.* 10(6):1857-1865.
- 218) Schnier J., Kimura M., Foulaki K., Subramanian A.-R., Isono K., y Wittmann-Liebold B. (1982) Primary structure of *Escherichia coli* ribosomal protein S1 and of its gene *rpsA*. *Biochemistry* 79:1008-1011.
- 219) Schnier J. y Faist G. (1985) Comparative studies on the structural gene for the ribosomal protein S1 in ten bacterial species. *Mol. Gen. Genet.* 200(3):476-481.
- 220) Schnier J., Thamm S., Lurz R., Hussain A., Faist G. y Dobrinski B. (1988) Cloning and characterization of a gene from *Rhizobium meliloti* 2011 coding for ribosomal protein S1. *Nucleic Acids. Res.* 16(7):3075-3089.
- 221) Schulte E., Staebach S., Laser B., y Kuck U. (1989) Wheat mitochondrial DNA: organization and sequences of the *atpA* and *atpB* genes. *Nucleic Acids Res.* 17(18):7531.

- 222) Schuster W. y Brennicke A. (1986) Pseudocopies of the ATPase α -subunit gene in *Oenothera* mitochondria are present on different circular molecules. *Mol. Gen. Genet.* 204(1):29-35.
- 223) Senior A.E. (1990) The proton-translocating ATPase of *Escherichia coli*. *Annu. Rev. Biophys. Chem.* 19:7-41.
- 224) Shinozaki K., Deno H., Kato A. y Sugiura M. (1983) Overlap and cotranscription of the genes for the β and ϵ subunits of tobacco chloroplast ATPase. *Gene* 24(2-3):147-155.
- 225) Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N., Matsubayashi T., Zalta H., Chunwongse J., Obokata J., Yamaguchi-Shinozaki K., Ohto C., Torazawa K., Meng B. Y., Sugita M., Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A., Tohdoh N., Shimada H. y Sugiura M. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* 5(9):2043-2049.
- 226) Shimada H., Fukuta M., Ishikawa M. y Sugiura M. (1990) Rice chloroplast RNA polymerase genes: The absence of an intron in *rpoC1* and the presence of an extra sequence in *rpoC2*. *Mol. Gen. Genet.* 221(3):395-402.
- 227) Smith J. L., Levin J. R. y Agabian N. (1989) Molecular characterization of the *Trypanosoma brucei* RNA polymerase I and III largest subunit genes *J. Biol. Chem.* 264(30):18091-18099.
- 228) Soreq H. y Littauer U. Z. (1977) Purification and characterization of polynucleotide phosphorylase from *Escherichia coli*. Probe for the analysis of 3' sequence of RNA. *J. Biol. Chem* 252(19):6885-6888.
- 229) Stahl D., Rodermerl S. R., Subramanian A. R. y Bogorad L. (1990) Nucleotide sequence of a 3.46 Kb region of maize chloroplast DNA containig the gene cluster *rpoC2-rps2-atpI-atpH*. *Nucleic Acids Res.* 18(10): 3073-3074.
- 230) Steitz J. A. y Jakes K. (1975) How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 72(12):4734-4738.
- 231) Struhl K. (1989) Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *Trends Biochem. Sci.* 14(4):137-140.
- 232) Stryer L. (1988) *Biochemistry*. 3rd ed. Freeman & Company. San Francisco. 706 pp.

- 233) Södhof T. C., Fried V. A., Stone D. K., Johnston P. A. y Xie X-S. (1989) Human endomembrane H⁺ pump strongly resembles the ATP-synthetase of achaeobacteria. *Proc. Natl. Acad. Sci. USA* 86(16):6067-6071.
- 234) Sugiura M. (1989) The chloroplast chromosomes in land plants. *Annu. Rev. Cell. Biol.* 5:51-70.
- 235) Suryanarayana T. y Subramanian A.R. (1979) Functional domains of *E.coli* ribosomal protein S1. Formation and characterization of a fragment with ribosome-binding properties. *J. Mol. Biol.* 127(1):41-54.
- 236) Sweetser D., Nonet M. y Young R.A. (1987) Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc. Natl. Acad. Sci. USA* 84(5):1192-1196.
- 237) Takeda M., Chen W. J., Saltzgaber J., y Douglas M. G. (1986) Nuclear genes encoding the yeast mitochondrial ATPase complex. Analysis on ATP1 coding the F₁-ATPase α -subunit and its assembly *J. Biol. Chem.* 261(32): 15126-15133.
- 238) Taylor W. R. y Oregon C. A. (1989) Protein structure alignment. *J. Mol. Biol.* 208(1):1-22.
- 239) Thang M. N., Dondon L. y Godefroy-Colburn T. (1971) Degradation of *Escherichia coli* polynucleotide phosphorylase by *E. coli* endogenous proteases and by trypsin. *Biochimie* 53(3):291-302.
- 240) Thornton J. M. y Gardner S. P. (1989) Protein motifs and data-base searching. *Trends Biochem. Sci.* 14(7):300-304.
- 241) Tittawella I. P. (1985) *Salmonella typhimurium* *rpoB* and *rpoC* genes cloned on λ phages. *FEBS Lett.* 185(1):33-36.
- 242) Tybulewicz V.L.J., Falk G. y Walker J.E. (1984) *Rhodopseudomonas blautica* *atp* operon nucleotide sequence and transcription. *J. Mol. Biol.* 179(2):185-214.
- 243) van Dieijen G., van Knippenberg P. H. y van Duin J. (1976) The specific role of ribosomal protein S1 in the recognition of native phage RNA. *Eur. J. Biochem.* 64(2):511-518.
- 244) von Heijne G. (1987) Sequence Analysis in Molecular Biology. Treasure trove or trivial pursuit. Academic Press, Inc. San Diego, California.

- 245) Wahba A. J., Miller M. J., Niveleau A., Landers T. A., Carmichael G. G., Weber K., Hawley D. A., Slobin L. I. (1974) Subunit I of Q β replicase and 30S ribosomal protein S1 of *Escherichia coli*. Evidence for the identity of the two proteins *J. Biol. Chem.* **249**(10):3314-3316.
- 246) Walker J.E., Gay N.J., Saraste M., y Eberle A.N. (1984) DNA sequence around the *Escherichia coli unc* operon. *Biochem. J.* **224**(3):799-815.
- 247) Walker J. E., Fearnley I. M., Gay N. J., Gibson B. W., Northrop F. D., Powell S. J., Runswick M. J., Saraste M., y Tybulewicz V. L. (1985) Primary structure and subunit stoichiometry of F₁-ATPase from bovin mitochondria. *J. Mol. Biol.* **184**(4):677-701.
- 248) Walker J.E., Powell S. J., Vinas O. y Runswick M. J. (1989) ATP synthase from bovine mitochondria: complementary DNA sequence of the import precursor of a heart isoform of the α subunit. *Biochemistry* **28**(11):4702-4708.
- 249) Wang T. S., Wong. S. W. y Korn D. (1989) Human DNA polymerase α : predicted funtional domains and relationships with viral DNA polymerases. *FASEB J.* **3**(1):14-21.
- 250) Watson J. D. y Crick F. H. C. (1953) Molecular structure of Nucleic Acids. A structure for deoxyribose nucleic acid. *Nature* **171**:737-738.
- 251) Watson J.D., Hopkins N.H. Roberts J.W., Steitz J. A. y Weiner A.M. (1987) Molecular Biology of the Gene Vol.1 Ed. The Benjamin Cummings Publishing Company, Inc. Menlo Park, California. 789 pp.
- 252) Weeks D.L. y Melton D.A. (1987) A maternal mRNA localized to the animal pole of *Xenopus* eggs encodes a subunit of mitochondrial ATPase. *Proc. Natl. Acad. Sci. USA* **84**(9):2798-2802.
- 253) White H. B. (1976) Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**:101-104.
- 254) Wierenga R.K., Drenth J. y Schulz G. E. (1983) Comparison of the three-dimensional protein and nucleotide structure of the FAD-binding domain of p-hydroxybenzoate hydroxylase with the FAD as well as NADPH-binding domains of glutathione reductase. *J. Mol. Biol.* **167**(3):725-739.
- 255) Wilbur W.J. y Lipman D. J. (1983) Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA* **80**(3):726-730.
- 256) Williams R. J. P. (1985) The symbiosis of metal and protein functions. *Eur. J. Biochem.* **150**:231-248.

- 257) Wilks, H. M., Hart, K. W., Feeney, R., Dunn, C. R., Muirhead, H., Chia, W. N., Barstow D. A., Atkinson, T., Clarke, A. R. y Holbrook, J. J. (1988) A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. *Science* 242(4885):1541-1544.
- 258) Wierenga R.K., Tarpstra P. y Hol W. G. (1986) Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$ -fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* 187(1):101-107.
- 259) Winning B.M., Bathgate B., Purdue P. E., y Leaver C. J. (1990) Nucleotide sequence of a cDNA encoding the beta subunit of the mitochondrial ATP synthase from *Zea mays*. *Nucleic Acids Res.* 18(19):5885.
- 260) Woese C. R. (1987) Bacterial evolution. *Microbiol. Rev.* 51(2):221-271.
- 261) Woese C. R. y Gutell R. R. (1989) Evidence for several higher order structural elements in ribosomal RNA. *Proc. Natl. Acad. Sci. USA* 86(9):3119-3122.
- 262) Woese C. R., Kandler O. y Wheelis M. L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* 87(12):4576-4579.
- 263) Woessner J. P., Gillham N. W., y Boynton J. E. (1986) The sequence of the chloroplast *atpB* gene and its flanking regions in *Chlamydomonas reinhardtii*. *Gene* 44(1):17-28.
- 264) Woychik N. A., Liao S.-M., Kolodziej P. A. y Young R. A. (1990) Subunits shared by eukaryotic nuclear RNA polymerases. *Genes & Development* 4(3):313-323.
- 265) Woychik N. A. y Young R. A. (1990) RNA polymerase II: subunit structure and function. *Trends Biochem. Sci.* 15(9):347-351.
- 266) Xie W.-Q., Jäger K. y Fotts M. (1989) Cyanobacterial RNA polymerase genes *rpoC1* and *rpoC2* correspond to *rpoC* of *Escherichia coli*. *J. Bacteriol.* 171(4):1967-1973.
- 267) Yamagishi M. y Nomura M. (1988) Cloning and sequence determination of the gene encoding the largest subunit of the fission yeast *Schizosaccharomyces pombe* RNA polymerase I. *Gene* 74(2):503-515.
- 268) Yepiz-Plascencia G. M., Radebaugh C. A. y Hallick R. B. (1990) The *Euglena gracilis* chloroplast *rpoB* gen. Novel gene organization and transcription of the RNA polymerase subunit operon. *Nucleic Acids. Res.* 18(7):1869-1878.

- 269) Young R.A. (1991) RNA Polymerase II. *Annu. Rev. Biochem.* 60:689-715.
- 270) Ysern X., Amzel L. M., and Pedersen P. L., (1988) ATP Synthases-structure of the F₁-moiety and its relationship to function and mechanism. *J. Bioenerg. Biomembr.* 20(4):423-450.
- 271) Zillig W., Klenk H-P., Palm P., Pühler G. y Gropp F. Garrett R. A., Leffers H. (1989) The phylogenetic relations of DNA-dependent RNA polymerases of archaeobacteria, eukaryotes and eubacteria. *Can. J. Microbiol.* 35(1):73-80.
- 272) Zimniak L., Dittrich P., Gogarten J. P., Kibak H. y Taiz L. (1988) The cDNA sequence of the 69-kDa subunit of the carrot vacuolar H⁺-ATPase. Homology to the beta-chain of FoF₁-ATPase. *J. Biol. Chem.* 263(19): 9102-9112.
- 273) Zuckerkandl E. y Pauling L. (1962) Molecular disease, evolution, and genic heterogeneity. In: M. Kasha y B. Pullman (eds) *Horizons in Biochemistry* Academic Press, New York. p.189-225.
- 274) Zuckerkandl E. y Pauling L. (1965) Molecules as documents of evolutionary history. *J. Theoret. Biol.* 8:357-366.
- 275) Zuckerkandl E. (1987) On the molecular evolutionary clock. *J. Mol. Evol.* 26:34-46.
- 276) Zurawski G., Bottomley W., y Whitfield P.R. (1982) Structure of the genes for the β and ϵ subunits of spinach chloroplast ATPase indicate a dicistronic mRNA and an overlapping translation stop/start signal. *Proc. Natl. Aca. Sci. USA* 79:6260-6264.
- 277) Zurawski G. y Clegg M. T. (1984) The berley chloroplast DNA *atpBE*, *trnM2*, and *trnV1* loci. *Nucleic Acids Res.* 12(5):2549-2559.
- 278) Zurawski G., Bottomley W., y Whitfield P. R. (1986) Sequence of the genes for the β and ϵ subunits of ATP synthase from pea chloroplasts. *Nucleic Acids Res.* 14(9):3974.

THE ORIGIN AND EARLY EVOLUTION OF NUCLEIC ACID POLYMERASES

A. Lazcano,** V. Llaca*,** R. Cappello,** V. Valverde** and J. Oro***

**Departamento de Biología, Facultad de Ciencias-UNAM Apdo, Postal 70-407, Cd. Universitaria, Mexico 04511, D.F., Mexico

***Department of Biochemical and Biophysical Sciences, The University of Houston, Houston, TX 77004, U.S.A.

*Present address: Department of Agronomy and Range Sciences, University of California-Davis, CA 95616, U.S.A.

ABSTRACT

The hypothesis that vestiges of the ancestral RNA-dependent RNA polymerase involved in the replication of RNA genomes of Archean cells are present in the eubacterial RNA polymerase β' subunit and its homologues is discussed. We show that in the DNA-dependent RNA polymerases from the three cellular lineages a very conserved sequence of eight amino acids also found in a small RNA-binding site previously described for the *E. coli* polynucleotide phosphorylase and the S1 ribosomal protein is present. The optimal conditions for the replicase activity of the avian myeloblastosis virus reverse transcriptase are presented. The evolutionary significance of the *in vitro* modifications of substrate and template specificities of RNA polymerases and reverse transcriptases is also discussed.

INTRODUCTION

There is considerable experimental evidence suggesting that under the anoxic conditions of the primitive Earth environment /1,2/, cyanamide /3/, metallic cations /4,5/, histidyl-histidine /6/, and catalytic RNA molecules /7/, may have catalyzed non-template and template-directed polymerization reactions involving nucleotides and their analogues.

The understanding of the evolutionary transitions of cellular genetic material from nucleic acid-like compounds /8,9/ to RNA genomes, and from those to double-stranded DNA molecules found in all extant cells /10/ allows us to infer some of the characteristics of early nucleic acid polymerases. Thus, the compelling evidence suggesting that RNA preceded DNA as cellular genetic material /10/ implies that a RNA-dependent RNA polymerase, i.e., a replicase, must have been one of the first enzymes to appear in early Archean cells /11,12/.

In this paper we discuss some possible characteristics of the ancestral RNA replicase and its role in the evolutionary transition from RNA to DNA genomes. First we review the hypothesis that important vestiges of the ancestral RNA-dependent RNA polymerase are still found in the contemporary eubacterial DNA-dependent RNA polymerase β' subunit and its homologues /11, 12/.

We also show that an eight amino acid sequence that forms part of a RNA-binding site formed by a 69 amino acid stretch previously described in the *Escherichia coli* polynucleotide phosphorylase and four homologous stretches of the ribosomal protein S1 of the same prokaryote /13/, is also found in eubacterial, eukaryotic and archaeobacterial DNA-dependent RNA polymerases. Finally, the substrate and template ambiguities of extant nucleic acid polymerases both *in vivo* and *in vitro* are discussed in terms of their evolutionary significance concerning our understanding of the evolution of enzymes /14/. Our conclusions are presented in the final section.

THE ANCESTRAL RNA POLYMERASE

Although there are many experiments in which the formation of peptide bonds has been demonstrated under possible prebiotic conditions /2/, Alberts /15/, Gilbert /16/ and Lazcano /11/ suggested that protein biosynthesis did not appear before the appearance of cells, but rather in cells in which both metabolism and reproduction depended on the catalytic and replicative properties of RNA molecules. This hypothesis implies that the basic selective pressure for the origin and stabilization of the primitive translation apparatus was the enhancement and diversification of the catalytic activities of RNA-based cells, thus increasing both their dynamic stability and survival. It is possible, of course, that oligopeptides of biological origin were initially selected for because their interaction with RNA molecules may have enhanced the catalytic properties of the latter, simply by stabilizing their structure /11/. Thus, the molecular ancestors of the different functional sites of extant nucleic acid polymerases were not necessarily part of Archean polymerases.

We know nothing of the transition from genetic material based on nucleic acid-like molecules to cellular RNA genomes /9, 17/. However, there is compelling evidence suggesting that RNA preceded DNA as genetic material in early Archean cells. This evidence has been reviewed recently /10, 14/, and includes the replicative and catalytic properties of RNA molecules, the central role of RNA in protein synthesis, the universal requirement of ribonucleotides as precursors for deoxyribonucleotides, and the chemical stability of double stranded DNA.

The Q β replicase has sometimes been used as a working model of an ancestral replicase /18, 19/. However, one of us /11/ has argued that the eubacterial β' subunit and its homologues correspond to the oldest component of extant RNA polymerases. Conclusive evidence on the evolutionary conservation of homologues of the eubacterial RNA polymerase β' subunit among other cellular lineages has been provided by primary sequence determinations that have demonstrated their presence in eukaryotes, archaeobacteria, chloroplasts /12, 20/ and viruses /21, 22/.

The hypothesis that the evolutionary conservation of the β' subunit and its homologues is due to their catalytic role in RNA elongation is contrary to the common assumption, which states that the catalytic site of the *E. coli* RNA polymerase is located in its β subunit. According to this scheme, β' is assumed to be involved solely in binding to the DNA template /23/. This traditional view is supported by a number of experiments involving metal substitution /24, 25/, photolabeling /26/, affinity-labeling /27/, and antibody-induced inhibition studies /23, 28/.

However, there is compelling evidence suggesting that the *E. coli* RNA polymerase β' subunit also plays a direct role in RNA elongation /11, 12/. This evidence includes a) the well established fact that in eukaryotic RNA polymerases the catalytic site involved in the elongation reaction is located in the largest subunit /29-33/. Therefore, the striking similarity between the largest subunits of eukaryotic RNA polymerases and the *E. coli* β' subunit discussed above suggest that these polypeptides may have comparable functions. It follows logically that the RNA elongation site must be located in the A+C component of archaeobacterial RNA polymerases; b) streptolydigin is known to inhibit RNA elongation in *E. coli* by interacting with the β subunit. However, in *Bacillus subtilis* streptolydigin sensitivity is located in the β' subunit; c) the Zn²⁺ atoms required for RNA polymerase activity are bound only to the *B. subtilis* β' subunit; and d) monoclonal antibodies directed against the *E. coli* β' subunit inhibit RNA elongation /34/.

These two apparently contradictory hypotheses on the roles of the *E. coli* RNA polymerase β and β' subunits can be reconciled by assuming that the genes coding for these two polypeptides are the result of an early tandem duplication event. This hypothesis explains their similar involvement in RNA synthesis summarized above, and is supported by a) the genetic linkage of the DNA sequences

coding for the two largest subunits of prokaryotic RNA polymerases. In the *E. coli* *rpoBC* operon the gene for the β subunit is followed by that coding for β' /35/. In archaeobacteria the A plus C genes are also colinear, and homologous to the β' subunit, and the organization of their operon is similar to that of *E. coli* /36/; b) both the *E. coli* β and β' subunits have a DD domain, a hydrophobic stretch of 14 amino acids centered around an Asp-Asp pair, which is highly conserved among viral and cellular RNA polymerases /37, 38/, which apparently plays an essential catalytic and/or structural role in nucleic acid polymerases /14/; and c) the existence of a segment formed by 113 amino acids, located towards the C-terminal end of the *E. coli* β' subunit, homologous to a region found towards the central part of β (Figure 1). Alignment of these two segments shows that 20 amino acids in the corresponding positions are identical, and 48 amino acids are conservative replacements.

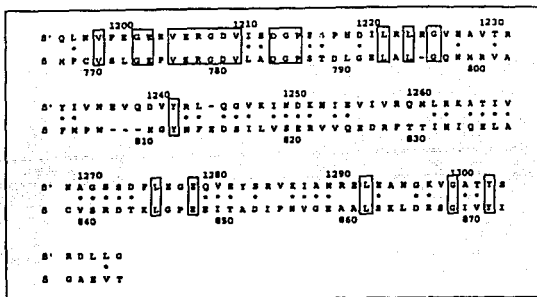


Fig. 1. Homologous regions of *E. coli* DNA-dependent RNA polymerase β and β' subunits. Identical amino acids are boxed. Conservative substitutions are marked with an asterisk (*).

AN ANCIENT RNA-BINDING SITE?

Polynucleotide phosphorylase (PNPase) is an enzyme first discovered in *E. coli* by Grunberg-Manago and Ochoa /39/, and later found in many bacteria and plant cells /40, 41/. PNPase is an oligomeric enzyme formed by three identical alpha subunits /42/, which are required to catalyze *in vitro* a) the polymerization of ribonucleoside diphosphates; b) the phosphorolysis of polyribonucleotides in the presence of phosphate; and c) an exchange reaction between free orthophosphate and the terminal phosphate of ribonucleosides /40/.

Kinetic analysis has shown that PNPase has a catalytic center and a RNA binding site /43-45/. The *E. coli* *pnp* gene was cloned by Portier and his collaborators /46/ and sequenced by Regnier et al /13/, who reported the existence of a 69 amino acid-stretch of the C-end of PNPase, homologous with a RNA-binding site represented by four stretches located in the middle and C-terminal portions of the *E. coli* ribosomal protein S1 /47/.

We have compared this RNA-binding region with the available amino acid sequences of eubacterial, archaeobacterial and eukaryotic RNA polymerases. Sequence alignment was performed using standard programs and is described in detail elsewhere /48/. As shown in Figure 2, at least part of the

putative RNA-binding site previously described for the *E. coli* PNPase and the S1 ribosomal protein /13/ can also be observed in one of the domains of the eubacterial RNA polymerase β' subunit and its homologues among the archaeobacterial and eukaryotic enzymes. This is particularly true for the sequence GGREGLVLD, in which charged amino acid residues (R, E, and D) are associated with glycine.

RPS1	G	A	I	V	E	G	K	V	T	A	V	D	A	K	G	A	T	V	A	L	-	-	-	-	A	D	G	V	E	G	V	L	R
RPS1	G	T	V	E	G	K	I	A	K	I	A	K	I	D	F	G	I	F	I	G	L	-	-	-	-	D	G	V	E	G	V	L	R
RPS1	G	T	V	E	G	K	I	A	K	I	A	K	I	D	F	G	I	F	I	G	L	-	-	-	-	D	G	V	E	G	V	L	R
RPS1	G	T	V	E	G	K	I	A	K	I	A	K	I	D	F	G	I	F	I	G	L	-	-	-	-	D	G	V	E	G	V	L	R
PNPase	G	T	V	E	G	K	I	A	K	I	A	K	I	D	F	G	I	F	I	G	L	-	-	-	-	D	G	V	E	G	V	L	R
M.th.	G	A	I	V	E	G	K	V	T	A	V	D	A	K	G	A	T	V	A	L	-	-	-	-	A	D	G	V	E	G	V	L	R
H.h.	S	A	E	A	H	G	F	V	E	N	A	Y	T	R	G	L	I	P	K	E	F	F	H	A	M	Q	G	R	E	G	V	L	D
S.ac.	S	A	E	A	H	G	F	V	E	N	A	Y	T	R	G	L	I	P	K	E	F	F	H	A	M	Q	G	R	E	G	V	L	D
S.c.II	S	A	E	A	H	G	F	V	E	N	A	Y	T	R	G	L	I	P	K	E	F	F	H	A	M	Q	G	R	E	G	V	L	D
S.c.III	T	P	Q	S	K	G	F	V	R	M	F	F	S	G	L	P	P	E	F	L	F	H	A	L	A	G	R	E	G	V	L	D	
D.m.II	G	T	V	E	G	K	I	A	K	I	A	K	I	D	F	G	I	F	I	G	L	-	-	-	-	D	G	V	E	G	V	L	D
E.c.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
RPS1	A	S	E	A	H	G	F	V	E	N	A	Y	T	R	G	L	I	P	K	E	F	F	H	A	M	Q	G	R	E	G	V	L	D
RPS1	A	S	E	A	H	G	F	V	E	N	A	Y	T	R	G	L	I	P	K	E	F	F	H	A	M	Q	G	R	E	G	V	L	D
RPS1	A	S	E	A	H	G	F	V	E	N	A	Y	T	R	G	L	I	P	K	E	F	F	H	A	M	Q	G	R	E	G	V	L	D
RPS1	A	S	E	A	H	G	F	V	E	N	A	Y	T	R	G	L	I	P	K	E	F	F	H	A	M	Q	G	R	E	G	V	L	D
RPS1	A	S	E	A	H	G	F	V	E	N	A	Y	T	R	G	L	I	P	K	E	F	F	H	A	M	Q	G	R	E	G	V	L	D
M.th.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
H.h.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S.ac.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S.c.II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S.c.III	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
D.m.II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
E.c.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
RPS1	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
RPS1	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
RPS1	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
RPS1	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PNPase	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
M.th.	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
H.h.	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S.ac.	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S.c.II	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S.c.III	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
D.m.II	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
E.c.	M	A	I	S	L	S	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

Fig. 2. A highly conserved eight amino acid stretch formed by GGREGLVLD found in the RNA-binding domain present in the *E. coli* polynucleotide phosphorylase (PNPase), the ribosomal protein S1 (RPS1), and in the *E. coli* RNA polymerase β' subunit (E.c.) and its homologues from *Methanobacterium thermoautotrophicum* (M.th.), *Halobacterium halobium* (H.h.), *Sulfolobus acidocaldarius* (S.ac.), *Saccharomyces cerevisiae* RNA polymerase II (S.c.II), *S. cerevisiae* RNA polymerase III (S.c.III), and the *Drosophila melanogaster* RNA polymerase II (D.m.II). Only identical amino acids in both the RNA polymerases and PNPase and the ribosomal protein S1 (RPS1) are boxed. Conservative amino acid replacements are shown in upper-case letters.

The domains in which the GGREGLVD sequences are found are among the most conserved in the RNA polymerase, and no particular function has been assigned to them. The homologies presented in Figure 2 suggest that PNPase, the ribosomal protein S1, and RNA polymerases may interact in the same way with RNA molecules. Since only one copy of the RNA-binding domain is present in RNA polymerases, it is possible that this is enough to bind to RNA molecules, as occurs with the yeast poly(A)-binding protein /49/, while PNPase and S1 protein may require multiple copies of this domain /13/. Moreover, the evolutionary conservation of the GGREGLVD in the RNA polymerases of the three cellular lineages suggests that it was present in their common ancestor, i.e., the progenote.

THE SUBSTRATE AND TEMPLATE AMBIGUITIES OF POLYMERASES

Both DNA- and RNA-polymerases are metalloenzymes that in addition to Zn^{++} require additional metallic cations such as Mg^{++} , without which the biological synthesis of polynucleotides cannot take place /50/. It is a well-established fact that the substitution of Mg^{++} by Mn^{++} increases the misincorporation of deoxyribonucleotides by DNA-dependent DNA polymerases and alters their substrate specificities /50/. The same effect has been reported for DNA-dependent RNA polymerases /51/. Additional evidence of the *in vitro* and *in vivo* alteration of substrate and template specificities of nucleic acid polymerases is shown in Table 1.

TABLE 1. Alternate functions of nucleic acid polymerases

Enzyme	Normal Activity	Alternate Role	Reference
<i>E. coli</i> DNA polymerase I	DNA-dependent DNA polymerase	incorporates NTPs in nascent DNA chains (with Mg^{++})	/61/
		reverse transcriptase (with Mn^{++})	/62-64/
<i>E. coli</i> RNA polymerase	DNA-dependent RNA polymerase	RNA-dependent RNA polymerase (replicase) (with Mn^{++})	/51/
Coliphage T7	DNA-dependent RNA polymerase	RNA-dependent RNA polymerase (replicase)	/65/
<i>E. coli</i> DNA primase	DNA-dependent RNA polymerase	incorporates NTPs producing mixed primer	/66-68/
HIV-1 RT AMV-RT	RNA-dependent DNA polymerase	RNA-dependent RNA polymerase (replicase) (with Mn^{++})	/14/

The hypothesis that the catalytic component of extant cellular RNA polymerases was initially RNA dependent, i.e., was a replicase, is strongly supported by the observation that DNA-dependent RNA polymerases can utilize different classes of RNA and synthetic polyribonucleotides as templates /51/. Thus, the evidence that suggest that the catalytic site involved in RNA elongation is located

in the *E. coli* β' subunit and its homologues, implies that this polypeptide is directly involved in RNA synthesis reading RNA molecules or polyribonucleotide templates /12/.

Although Mg^{++} appears to be universally required for all known RNA- and DNA polymerases, this is not the case for reverse transcriptases (RTs), i.e., RNA-dependent DNA polymerases /52/ involved in the replication of viral and non-viral retrotransposons /53/. Some RTs, like the avian myeloblastosis virus (AMV) RT require Mn^{++} . It is likely that the cationic requirements are not randomly distributed but reflect phylogenetic relationships among different RTs /54, 55/.

We have examined the role of Mn^{++} in the modification of the substrate specificity of RTs, using a human immunodeficiency virus (HIV1) partially purified RT, and AMV RT, which is commercially available from Sigma Chemical Co. (USA). Since it has been shown that reverse transcription can be accurately studied *in vitro* using poly (rC)-oligo (dG) as a template-primer system /56, 57/, this system was employed in the presence of Mn^{++} and labeled ribonucleoside guanosine triphosphate as a substrate. Our experimental procedures have been described in further detail elsewhere /14/. Our results clearly indicate that the substitution of Mn^{++} for Mg^{++} modifies the substrate specificity of these two RTs, transforming them into replicases, i.e., into RNA-dependent RNA polymerases. As shown in Table 2, the major factor influencing the AMV RT substrate specificity is the metallic cation added to the incubation medium.

TABLE 2. Optimal conditions for the enzymatic activity of the AMV-RT*

	Reverse Transcriptase	RNA Replicase
pH	8.3	8.3
MgCl ₂	2 to 5 mM	---
MnCl ₂	---	1.0 mM
KCl	40 mM	40 mM

* With poly(c)-oligo(dG) as template-primer as described in /14/.

The results summarized in Table 1 clearly indicate that nucleic acid polymerases are very sensitive to cation changes, i.e., limited modifications of the catalytic and binding sites induced by Mn^{++} can alter the substrate and template specificities of polymerases. As noted recently by Visscher and Schwartz /5/, who have studied the catalytic role of Mn^{++} in template-directed oligomerization of nucleotide analogues, this metallic cation reduces base stacking, and may interact with the substrate. Furthermore, in the experiments we have described Mn^{++} must be forming non-covalent bonds with the metal-binding sites that exist in RTs molecules. Thus, the role of Mn^{++} in the system we have described may be multifunctional, involving simultaneous interactions with the template, the substrate and the RTs /14/.

These observations on the ambiguous specificities of extant nucleic acid polymerases also have evolutionary implications. As noted elsewhere /14/, the transition from RNA cellular genomes to double-stranded DNA molecules that is usually assumed to have taken place during the early Archean, could have involved the existence of a) different nucleic acid polymerases (at least, a RT, one DNA-dependent DNA polymerase and a DNA primase); or b) nucleic acid polymerases of broad, ambiguous template and substrate specificities. The results summarized here support the second alternative, i.e., they suggest that novel catalytic properties of nucleic acid polymerases required only minor tertiary structural modifications, and that the specificity observed in RNA polymerases, DNA-

dependent DNA polymerases and RTs probably resulted after gene duplication events. Therefore, although there is convincing evidence of the monophyletic origin of extant viral and non-viral RTs /58/, the hypothesis that these polymerases are directly descended from an ancestral Archean RT cannot be automatically supported, since our results suggest that primordial reverse transcription activity could have been mediated by early proteinic polymerases of broad substrate and template specificity /14/.

CONCLUSIONS

There is a considerable gap in our understanding of the evolutionary transition from small catalytic peptides like His-His, and the enzymes ancestral to contemporary nucleic acid polymerases. This gap may be filled in when detailed models of the origin and early evolution of ribosome-mediated protein biosynthesis become available. Such a problem is beyond the scope of this article (and, perhaps, beyond our theoretical and experimental abilities). However, we have argued here that vestiges of one of the oldest enzymes, i.e., the ancestral replicase, are found in the eubacterial RNA polymerase β' subunit and its homologues among the archaeobacterial and eukaryotic DNA-dependent RNA polymerases.

The hypothesis that the β' subunit and its homologues are directly involved in the RNA elongation reaction is also supported by the presence in these polypeptides of a highly conserved eight amino acid stretch found in a small RNA-binding sequence previously described for PNPase and the S1 protein by Regnier and his collaborators /13/.

AMBIGUOUS SUBSTRATE RECOGNITION



SPECIFIC SUBSTRATE RECOGNITION



UNREGULATED CATALYSIS



REGULATED CATALYSIS

Fig. 3. Evolution of enzymatic activity

The results summarized here on the substrate specificity of nucleic polymerases support Waley's /59/ and Jensen's /60/ models of enzyme evolution, which suggest that primitive enzymes were unregulated, substrate-ambiguous catalysts. Thus, in the course of biological evolution primitive biological catalysts have evolved into highly specific, regulated, and in many cases, oligomeric enzymes (Figure 3).

ACKNOWLEDGEMENTS

We are indebted to Drs. Jaime Sepulveda Amor and Federico Chavez Peon, of the Mexican Ministry of Health, for their constant support and encouragement. One of us (A.L.) would like to acknowledge the stimulating environment of the 1990 Congreso Universitario (UNAM, Mexico), during which parts of this work were written. We are grateful to Drs. Alan W. Schwartz and Andre Brack for their careful reading of the manuscript. We would also like to thank Tamara Smith and Thomas Mills for their work in the preparation of this paper in its final form. This work has been supported in part by CONASIDA-Mexico (A.L.) and NASA Grant 44-005-002 (J.O.).

REFERENCES

1. A. Lazcano, J. Oro and S.L. Miller, *Precambrian Res.* 20, 259 (1983)
2. J. Oro, S.L. Miller and A. Lazcano, *Annu. Rev. Earth and Planet. Sci.* 18, 317 (1990)
3. E. Stephen-Sherwood, D. Odom and J. Oro, *J. Mol. Evol.* 3, 323 (1974)
4. P.K. Bridson and L.E. Orgel, *J. Mol. Biol.* 144, 567 (1980)
5. J. Visscher and A.W. Schwartz, *J. Mol. Evol.* 29, 284 (1989)
6. S. Shen, A. Lazcano and J. Oro, *J. Mol. Evol.* (in press) (1990)
7. T.R. Cech, *Science* 236, 1532 (1987)
8. A.W. Schwartz and L.E. Orgel, *Science* 228, 585 (1985)
9. G.F. Joyce, A.W. Schwartz, S.L. Miller and L.E. Orgel, *Proc. Natl. Acad. Sci. USA* 84, 4398 (1987)
10. A. Lazcano, R. Guerrero, L. Margulis and J. Oro, *J. Mol. Evol.* 27, 283 (1988)
11. A. Lazcano, *Treballs Soc. Cat. Biol.* 39, 73 (1986)
12. A. Lazcano, J. Fastag, P. Gariglio, C. Ramirez and J. Oro, *J. Mol. Evol.* 27, 365 (1988)
13. P. Regnier, M. Grunberg-Manago and C. Portier, *J. Biol. Chem.* 262, 63 (1987)
14. A. Lazcano, V. Valverde, Greco-Hernandez, P. Gariglio, G.E. Fox and J. Oro, *J. Mol. Evol.*, in press (1990)
15. B. Alberts, *Am. Zool.* 26, 781 (1986)
16. W. Gilbert, *Nature* 319, 618 (1986)
17. G.F. Joyce, *Nature* 338, 217 (1989)
18. M. Eigen and P. Schuster, *Naturwissenschaften* 64, 541 (1987)
19. P. Schuster, *Adv. Space Res.* 4, 143 (1984)
20. G. Puhler, H. Leffers, F. Gropp, P. Palm, H.P. Klenk, F. Lottspeich, R.A. Garrett and W. Zillig, *Proc. Natl. Acad. Sci. USA* 86, 4569 (1989)
21. S.S. Broyles and B. Moss, *Proc. Natl. Acad. Sci. USA* 83, 3141 (1986)
22. P. Roy, A. Fukusho, D.G. Ritter and D. Lyon, *Nucleic Acid Res.* 16, 11759 (1988)
23. T. Yura and A. Ishihama, *Annu. Rev. Genet.* 13, 59 (1979)

24. D. Chatterji and F.Y.H. , *Biochemistry* 21, 4651 (1982)
25. D. Chatterji and F.Y.H. Wu, *Biochemistry* 21, 4657 (1982)
26. N. Reusch and D. Dennis, *J. Biol. Chem.* 260, 16310 (1985)
27. V.W. Armstrong, H. Sternbach and F. Eckstein, *Biochemistry* 15, 2086 (1976)
28. A. Heil and W. Zillig, *FEBS Lett.* 11, 165 (1976)
29. J. Matial, J. Zaldivar, P. Bull, A. Venegas and P. Valenzuela, *Biochemistry* 14, 4907 (1975)
30. A. Ruet, A. Sentenac, P. Fromageot, B. Windson and F. Lacroute, *J. Biol. Chem.* 255, 6450 (1980)
31. A.L. Greenleaf, *J. Biol. Chem.* 258, 13403 (1983)
32. K.M. Rose, K.A. Maguire, J.N.D. Worpel, D.A. Stetler and E.D. Marquez, *J. Biol. Chem.* 258, 12976 (1983)
33. D.E. Coulter and A.L. Greenleaf, *J. Biol. Chem.* 260, 13190 (1985)
34. P. Rockwell, E. Beasley and J.S. Krakow, *Biochemistry* 24, 3240 (1985)
35. J. Scaife, in *RNA Polymerase*, eds. R. Losick and M. Chamberlin, Cold Spring Harbor Laboratory, New York, p. 207 (1976)
36. W. Zillig, P. Palm, W.D. Reiter, F. Gropp, G. Puhler and H.P. Klenk, *Eur. J. Biochem.* 173, 473 (1988)
37. G. Kamer and P. Argos, *Nucleic Acid Res.* 12, 7269 (1984)
38. P. Argos, *Nucleic Acid Res.* 16, 9909 (1988)
39. M. Grunberg-Manago and S. Ochoa, *J. Am. Chem. Soc.* 77, 3165 (1955)
40. T. Godefroy-Colburn and M. Grunberg-Manago, in *The Enzymes*, ed P.D. Boyer, Academic Press, New York, vol. 7, 533 (1972)
41. Z. Kahn and H. Fraenkel-Conrat, *Proc. Natl. Acad. Sci. USA* 82, 1311 (1985)
42. C. Portier, *Eur. J. Biochem.* 55, 573 (1975)
43. M.N. Thang, R.A. Harvey and M. Grunberg-Manago, *J. Mol. Biol.* 53, 261 (1971)
44. T. Godefroy, *Eur. J. Biochem.* 79, 233 (1970)
45. A. Guissani, *Eur. J. Biochem.* 86, 473 (1977)
46. C. Portier, C. Migot and M. Grunberg-Manago, *Mol. Gen. Genet.* 183, 298 (1981)
47. B. Wittman-Liebold, L. Ashman and M. Dzionara, *FEBS Lett.* 1154, 31 (1983)
48. A. Lazzano, V. Liaca, R. Cappello, T. Mills, G.E. Gox and J. Oro, *J. Mol. Evol.* (in preparation)
49. A.B. Sachs, R.W. Davis and R. Kornberg, *Mol. Cell. Biol.* 7, 3268 (1987)

50. A.S. Mildvan and L.A. Loeb, *CRC Crit. Rev. Biochem.* 6, 219 (1979)
51. V. Llaca, E. Silva, A. Lazcano, L.M. Rangel, P. Gariglio and J. Oro, in *Prebiological Organization*, eds. F. Eirich and C. Ponnampuruma, A. Deepak Publ. Co., Hampton, Virginia, 247 (1990)
52. H. Varmus and P. Brown, in *Mobile DNA Elements*, eds. M. Howe and M. Berg, American Soc. of Microbiology, Washington, p. 53 (1989)
53. A.M. Weiner, P.L. Dieninger and A. Efstratiadis, *Annu. Rev. Biochem.* 55, 631 (1986)
54. N. Sagata, T. Yasunaga, J. Tsuzuku-Kawamura, K. Oshini, Y. Ogawa and Y. Ikawa, *Proc. Natl. Acad. Sci. USA* 82, 677 (1985)
55. P. Sonigo, M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, E. Retzel, P. Tiollais, A. Haase and S. Wain-Hobson, *Cell* 42, 369 (1985)
56. G.F. Gerard, F. Rottman and M. Green, *Biochemistry* 13, 1632 (1976)
57. A.A. Kiessling and M. Goulian, *Biochem. Biophys. Res. Comm.* 71, 1069 (1976)
58. R.F. Doolittle, D.F. Feng, M.S. Johnson and M.A. McClure, *Quart. Rev. Biol.* 64, 1 (1989)
59. S.G. Waley, *Comp. Biochem. Physiol.* 30, 1 (1969)
60. R.A. Jensen, *Annu. Rev. Microbiol.* 30, 409 (1976)
61. P. Berg, H. Fancher and M. Chamberlin, in *Informational Macromolecules*, eds. HJ. Vogel, V. Bryson and J.O. Lampen, Academic Press, New York, p. 467 (1963)
62. J.F. Karkas, G.S. Jannis and E. Chargaff, *Proc. Natl. Acad. Sci. USA* 69, 398 (1972)
63. J.D. Karkas, *Proc. Natl. Acad. Sci. USA* 70, 3834 (1973)
64. L.A. Loeb, K.D. Tartof and E.C. Travaglini, *Nature* 242, 66 (1973)
65. M.M. Konarska and P.A. Sharp, *Cell* 57, 423 (1987)
66. A. Kornberg, *Biochem. Soc. Trans.* 5, 359 (1977)
67. L. Rowen and A. Kornberg, *J. Biol. Chem.* 253, 770 (1978)
68. R. McMacken and A. Kornberg, *J. Biol. Chem.* 253, 3313 (1978)