



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS

ANÁLISIS COEVOLUTIVO DE LAS PROTEÍNAS SBDS y EFL1

INFORME DE TRABAJO DE INVESTIGACIÓN
PARA OPTAR POR EL GRADO DE

MAESTRA EN CIENCIAS

PRESENTA

Q. CLAUDIA DANIELA TORRES ZULUETA

Dra. NURIA VICTORIA SÁNCHEZ PUIG
INSTITUTO DE QUÍMICA, UNAM

CIUDAD UNIVERSITARIA, CD. MX., Abril 2022



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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 esta 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.

Jurado Asignado

Presidente	Dr. Roberto Arreguín Espinosa de los Monteros	Instituto de Química, UNAM
Vocal	Dr. Roberto Sánchez Olea	Universidad Autónoma de San Luis Potosí
Vocal	Dr. Carlos Becerra Bracho	Facultad de Ciencias, UNAM
Vocal	Dra. Karina Martínez Mayorga	Instituto de Química, UNAM
Secretario	Dra. Corina Diana Ceapa	Instituto de Química, UNAM

El presente trabajo se desarrolló bajo la tutela de la Dr. Nuria Victoria Sánchez Puig en el Laboratorio 7 del Departamento de Química de Biomacromoléculas del Instituto de Química. UNAM.

AGRADECIMIENTOS

- Agradezco a la coordinación del programa de posgrado de Maestría y Doctorado en Ciencias Químicas por darme la oportunidad de formar parte del programa de maestría y por ayudarme con todos los trámites para venir a estudiar a México.
- Agradezco a CONACYT por la beca (CVU 1009093) brindada para cubrir la totalidad de mis gastos durante mis estudios.
- Al financiamiento otorgado en el proyecto CONACYT 283909 “Escrutinio de moléculas orgánicas pequeñas contra la GTPasa EFL1 como blanco en el Síndrome Shwachman-Diamond” y DGAPA-PAPIIT IN215122 “Entendiendo los cambios conformacionales en EFL1 y su impacto en el Síndrome de Shwachman-Diamond” para la realización de este trabajo.
- Agradezco grandemente a mi tutora, la Dr. Nuria Victoria Sánchez Puig, por darme la oportunidad de trabajar en su grupo de investigación y por todo el apoyo que tuve de su parte durante mi maestría, y durante los tiempos difíciles de la pandemia.
- Especial agradecimiento a mis padres y mis amigos. Que a pesar de la distancia física estuvieron presentes emocionalmente, y me acompañaron durante todo mi proceso, especialmente en esos días en los que sentía que ya todo estaba perdido durante el confinamiento.
- Me siento muy feliz y agradecida con el universo por todas las nuevas experiencias que tuve durante estos dos años.

Índice

1	Antecedentes	1
1.1	Coevolución: de la evolución de las especies a la evolución molecular.....	1
1.1.1	Coevolución a nivel de los residuos de una proteína.....	2
1.1.2	Coevolución a nivel de proteínas.....	2
1.1.3	Predicción de contactos entre residuos mediante coevolución	3
1.2	El ribosoma	4
1.2.1	Biogénesis ribosomal.....	5
1.2.2	Las proteínas SBDS y EFL1	7
2	Planteamiento del problema.....	8
2.1	Hipótesis	8
2.2	Objetivos	8
3	Materiales y métodos	9
3.1	Cálculos para inferir coevolución a nivel intra e inter-proteína	9
4	Resultados	10
4.1	Coevolución a nivel intra-proteína	10
4.2	Coevolución a nivel inter-proteína	14
5	Discusión.....	18
6	Conclusiones.....	21
7	Perspectivas.....	22
8	Bibliografía	22
9	Anexos	26

RESUMEN

Los ribosomas son sintetizados mediante un proceso altamente coordinado en tiempo y espacio que comienza en el nucléolo y termina en citoplasma celular. Los ribosomas están compuestos por proteínas y RNA ribosomal que se arreglan en dos subunidades; la menor que decodifica la información genética y la mayor encargada de la formación del enlace peptídico. La producción de ribosomas requiere de la participación de múltiples biomoléculas accesorias que participan en su síntesis, ensamblaje, modificación y transporte. Las proteínas SBDS y EFL1 participan en la última etapa de maduración ribosomal de la pre-subunidad 60S liberando al factor de anti-asociación eIF6 para permitir la formación de la partícula ribosomal madura 80S. SBDS está altamente conservada en Eukarya y Archaea, mientras que EFL1 está presente solo en Eukarya.

Estudios anteriores desarrollados en nuestro grupo de investigación evaluaron la complementación funcional de ortólogos de las familias de proteínas SBDS y EFL1 en células de levadura carentes de los genes correspondientes. Estos demostraron que ninguno de los pares de ortólogos evaluados pudo complementar la función de los genes equivalentes de levadura, lo que sugirió una posible coevolución entre estas dos familias de proteínas. Por lo anterior en el presente trabajo se llevó a cabo un estudio bioinformático para predecir pares de aminoácidos que han coevolucionado en cada una de las familias de proteínas SBDS y EFL1, así como predecir la coevolución inter-proteína entre ambas.

El aporte de este trabajo fue la predicción, a través de la interpretación de los resultados obtenidos con el servidor EVcouplings, de los residuos que han coevolucionado en ambas familias de proteínas, así como entre ellas. Además, se predijeron residuos en contactos entre las dos proteínas que hasta este momento no se han dilucidado, debido a la falta de estudios estructurales de suficiente resolución. Por lo que se hace necesario comprobar experimentalmente la calidad de dichas predicciones. Los resultados obtenidos predijeron coevolución a nivel intra-proteína en los tres dominios estructurales de SBDS y en los cinco dominios de EFL1. También entre los tres dominios de SBDS y los dominios II, III y IV de EFL1. Los aminoácidos R126, N121, T129, K118, R218 de SBDS y C883 y R970 de EFL1 son posiciones que han coevolucionado a nivel inter-proteína, y en las que además se han encontrado mutaciones en pacientes con el Síndrome Shwachman Diamond.

1 Antecedentes

1.1 Coevolución: de la evolución de las especies a la evolución molecular

El origen de la coevolución se remonta al estudio de las interacciones entre especies en sistemas ecológicos. Los primeros estudios de coevolución reportados en la literatura moderna se deben a Darwin, quien observó la relación existente entre las longitudes de los nectarinos de las orquídeas y las probóscides de sus polinizadores (1). Sin embargo, este término fue realmente definido en 1964 por Ehrlich y Raven (2), quienes lo definieron como las interacciones evolutivas halladas entre diferentes especies u organismos cuyo intercambio de información genética se asume mínimo o nulo. Otra definición fue planteada por John N. Thompson (3), el cual decía que es el proceso de evolución recíproca entre especies que interactúan dirigido por la selección natural. Otro concepto relevante en el contexto de la coevolución es la hipótesis de la Reina Roja formulada por van Valen en 1973 (4), el cual formuló dos ideas aparentemente contradictorias: la conservación y el cambio. van Valen propuso que las especies están cambiando constantemente para poder mantenerse o conservarse en un medio en el que todo cambia, incluyendo el resto de las especies con las que están relacionadas. La convergencia entre los conceptos de conservación y cambio se debe a que actúan a distintos niveles. Desde un punto de vista, se conserva la interacción entre las especies y, por otro lado, es necesario adaptarse a los cambios que otras especies sufren, y al resto del entorno (5).

- *Jerarquía de niveles evolutivos y coevolución molecular*

Durante el siglo XX se desarrollaron las primeras investigaciones moleculares sobre coevolución gracias al almacenamiento de gran cantidad de datos de secuencias de proteínas y genomas. Algunas de estos estudios encontraron que la interacción entre especies está relacionada a nivel coevolutivo entre ciertos genes de cada especie (6), los cuales tienen entre otras funciones, mantener la interacción entre las dos especies. Otros estudios mostraron que también podía existir entre residuos de la misma proteína. Los cambios recíprocos entre los residuos de la misma proteína o entre un par de proteínas permiten conservar tanto el correcto funcionamiento de la proteína o del complejo proteico, como su plegamiento (7).

Una jerarquía simplificada de niveles evolutivos puede presentarse como la coevolución entre especies, entre proteínas y entre residuos. De igual forma, este fenómeno ocurre gen a gen entre especies. Y se cree que una de sus principales funciones es mantener la interacción entre ellas (7,8). También se podría considerar en otros niveles macroscópicos como ecosistemas y poblaciones.

En el presente trabajo solo se utilizará la terminología coevolución para referirse a interacciones entre residuos en una misma proteína y entre residuos de dos proteínas. Un concepto simplificado de coevolución puede entenderse como “*los cambios coordinados que ocurren en pares de organismos o biomoléculas para mantener interacciones funcionales entre esos pares*” (7)

1.1.1 Coevolución a nivel de los residuos de una proteína

Las contribuciones pioneras en el campo de la coevolución realizadas por Altschuh *et al.* en 1987 (9) fueron el punto de partida para una nueva línea de investigación sobre la correlación entre cambios coordinados en alineamientos de secuencias primarias y contactos físicos en la estructura tridimensional de una proteína (10,11). Otras de las contribuciones más destacadas fue el estudio de Göbel *et al.* quien propuso la existencia de una relación directa entre coevolución y contactos en estructuras, y su potencial para mejorar las predicciones de estructura de proteínas (10) como se demostró posteriormente (12). A estos cambios coordinados se les llamó mutaciones correlacionadas o coordinadas. Numerosos estudios realizados para explicar la relación existente entre la covariación en una secuencia y sus contactos físicos han observado que mutaciones deletéreas pueden ser compensadas por otras mutaciones en su entorno espacial próximo evitando su efecto perjudicial en la función biológica (13); y comúnmente es utilizado el término coadaptación para referirse a adaptaciones específicas de residuos (14).

Las variaciones en las secuencias de una proteína están restringidas tanto por motivos estructurales (estabilidad, plegamiento, interacciones) como por motivos funcionales (sitios catalíticos, sitios de unión a ligandos). Algunas de las aportaciones más relevantes en este campo son la observación de que las posiciones más conservadas en alineamientos múltiples de secuencias homólogas se encuentran asociadas a roles funcionales o estructurales importantes (15). La conservación de las posiciones en la secuencia de las familias de proteínas sigue siendo una fuente inagotable de información evolutiva y funcional (16).

1.1.2 Coevolución a nivel de proteínas

La coevolución entre proteínas sugiere interacción entre ellas. En el año 2000 Goh *et al.* (17) propuso que la coevolución entre proteínas que interactúan se podía detectar a través de la similitud entre los árboles filogenéticos de ambas familias de proteínas. Para ello emparejaron las secuencias de las proteínas que interactuaban en las familias estudiadas y evaluaron la correlación en los grados de divergencia de ambas familias. Pese al valor de esta idea, no permitía ni pretendía predecir interacciones usando coevolución

sino detectarla entre proteínas de interacción conocida. El uso de este concepto para predecir interacción entre proteínas presentaba un concepto que parecía paradójico, pues era necesario el conocimiento previo de que dos proteínas interactuaban para evaluar su coevolución. Este problema fue resuelto por el grupo de investigación de Pazos y colaboradores quienes propusieron el uso de relaciones de ortología biunívocas (18). Esta aproximación denominada Árbol filogenético espejo, o *MirrorTree* por su nombre en inglés, permitió por primera vez calcular la coevolución entre proteínas a gran escala y demostrar que las altas correlaciones en las divergencias entre secuencias ortólogas de proteínas diferentes contenían información significativa sobre las interacciones entre proteínas.

La coevolución entre proteínas implica una interdependencia funcional entre ellas, por ello se han desarrollado métodos capaces de predecir interacciones funcionales mediante la detección de señales empleando la información proveniente de los genomas. Estos métodos pueden dividirse en tres grandes familias: los basados en perfiles filogenéticos, los basados en similitud de árboles filogenéticos y los que se basan en fusión de genes (19).

1.1.3 Predicción de contactos entre residuos mediante coevolución

Existen dos tipos de métodos para predecir contactos físicos entre residuos de una misma proteína o de dos proteínas: los métodos locales y los métodos globales. Los métodos locales integran tres grandes familias de métodos, aquellos basados en correlación, los basados en información mutua y los que utilizan árboles filogenéticos. En todos, la información básica es la misma, un alineamiento múltiple de secuencias homólogas a la proteína objetivo y medir la correlación de los cambios observados. Sin embargo, los modelos locales aplicados a secuencias de proteínas no detectan consistentemente residuos que están cercanos en la estructura tridimensional (10,11). A pesar de que en la literatura existen algunas aplicaciones exitosas que muestran interacción entre residuos o la identificación de contactos entre proteínas utilizando información biológica adicional (20). La baja habilidad de estos métodos para la predicción de residuos en contacto ha sido atribuida a diferentes razones como la pérdida de señal debido a las dependencias filogenéticas, el limitado número de secuencias disponibles y la idea de que no se debería esperar que los residuos que coevolucionan deben estar, en su mayoría, cercanos en la estructura (21). La solución para este problema es utilizar una clase de modelos de probabilidad global conocido como modelo Potts (un modelo de máxima entropía) en física estadística (21). El uso de este modelo de dependencias de tipos de aminoácidos en pares de posiciones permite que sean computados simultánea y consistentemente, en vez de

analizar pares de posiciones independientemente una de la otra. La aplicación de estos modelos estadísticos globales fue la clave en la identificación de acoplamientos evolutivos (*evolutionary couplings*, ECs) entre pares de posiciones en alineamientos múltiples de secuencias que corresponden a residuos en contacto (22,23). Los contactos predichos mediante acoplamientos evolutivos han permitido la predicción *de novo* de estructuras tridimensionales de proteínas, incluso para moléculas grandes más allá del alcance de enfoques anteriores (24,25), sus complejos (26), contactos multiméricos (27), conformaciones alternativas (28), e incluso la habilidad de predecir estados estructurados de proteínas aparentemente desordenadas (29). Además, se han utilizado para identificar residuos funcionalmente restringidos y más recientemente, para hacer predicciones cuantitativas de cambios mutuos en las proteínas (30). En el presente trabajo fueron utilizados los métodos globales para el cálculo de coevolución.

1.2 El ribosoma

Los ribosomas son las máquinas moleculares que permiten la traducción del ARN mensajero (mRNA) a proteína en las células. En eucariontes, estos complejos están formados por las subunidades 40S (subunidad menor) y 60S (subunidad mayor) las cuales albergan la función de decodificación y la actividad catalítica, respectivamente. Los ribosomas consisten de un centro universalmente conservado, el cual fue descrito por primera vez gracias a las estructuras cristalinas de los ribosomas procariontes (31-33). Las estructuras que conforman este ribosoma ancestral son el centro peptidil transferasa (PTC), que cataliza la formación de enlaces peptídicos, el túnel de salida de los péptidos recién sintetizados y el centro de decodificación que descifra la información contenida en el mRNA. Los ribosomas de eucariotas son al menos un 40% más grandes que los ribosomas de bacterias debido a segmentos de expansión en los ARN ribosomales (rRNA), proteínas adicionales e inserciones dentro de las proteínas existentes (34-36). En conjunto el ribosoma está formado por cuatro rRNAs y alrededor de 80 proteínas ribosomales (37). En eucariotas, la subunidad 40S está conformada por el rRNA 18S y 33 diferentes proteínas ribosomales, mientras que la subunidad 60S consiste de los rRNAs 25S, 5.8S y 5S junto con 47 r-proteínas (38). En cambio los ribosomas de bacteria, la subunidad menor 30S la componen 21 r-proteínas y el rRNA 16S y la subunidad mayor 50S, por 33 r-proteínas y los rRNAs 23S y 5S (39).

1.2.1 Biogénesis ribosomal

La síntesis de ribosomas es uno de los procesos más demandantes energéticamente en la célula. En eucariotas, el proceso de biogénesis ribosomal requiere de la participación de ~80 ribonucleoproteínas nucleolares pequeños (snoRNPs), las tres ARN polimerasas y del trabajo conjunto de alrededor de 200 factores de ensamblaje del ribosoma (40-42). La RNA polimerasa II sintetiza los mRNA que codifican para las proteínas ribosomales y los factores de ensamblajes, mientras que las RNA polimerasas I y III transcriben los precursores de los rRNAs. Durante el complicado proceso de biogénesis ribosomal en eucariotas, estos son plegados, procesados y ensamblados con las r-proteínas para formar las dos subunidades ribosomales. Este proceso tiene lugar mayormente en el nucléolo y puede dividirse en cinco etapas: síntesis de RNA pre-ribosomal (pre-rRNA) y de las r-proteínas, modificación de bases, plegamiento, ensamblaje con las r-proteínas y digestión endo- y exo-nucleolítico de los pre-rRNA (38,43).

El ensamblaje del ribosoma eucariota comienza con la participación de RNA polimerasa I, para producir pre-rRNA 35S, el precursor de los rRNA maduros 18S, 5.8S y 25S, como se puede ver en la Figura 1. El pre-rRNA 35S se asocia a varios snoRNPs, r-proteínas y factores de ensamblaje para formar el pre-ribosoma 90S. Luego, una escisión cotranscripcional en el sitio A2 separa la subunidad menor de la subunidad mayor 60S, las cuales experimentan procesos de maduración independientes. Factores de ensamblaje que se asociación transitoriamente promueven la maduración de las subunidades pre-ribosomales, mientras estas viajan a través del núcleo para llegar al citoplasma. Antes de incorporarse a la reserva de ribosomas capaces de realizar el proceso de traducción, las subunidades pre-ribosomales son sometidas a una revisión funcional que está asociada con la liberación de factores ensamblaje y de transporte, antes de su ensamblaje en la unidad madura 80S del ribosoma (44,45).

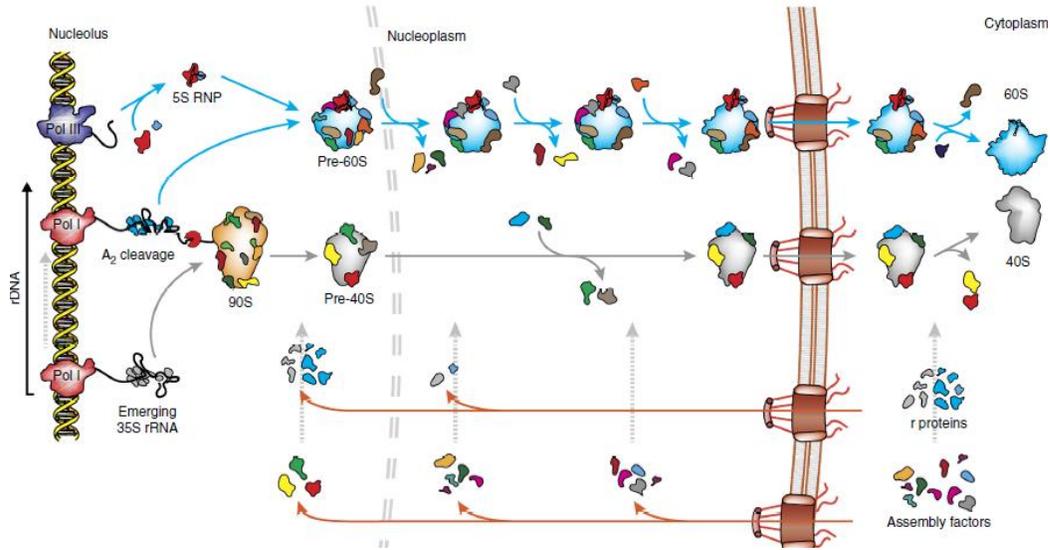


Figura 1. Proceso de ensamblaje del ribosoma en eucariotas. Los ARNr se asocian a proteínas ribosomales para formar la unidad prerribosomal 90S en el nucleólo, la cual posteriormente da lugar a las pre-subunidades 40S y 60S. Estas pre-subunidades son transportadas del nucleoplasma al citoplasma donde sufren las últimas etapas de maduración para posteriormente ensamblar la unidad funcional 80S del ribosoma. Tomado de (44).

El proceso de maduración de la subunidad 60S es más complejo que el de la subunidad 40S (45,46). Involucra el procesamiento de rRNA y la disociación de diversos factores accesorios y asociación de diversas proteínas ribosomales (46-48). Varios factores ayudan en el transporte de la pre-subunidad ribosomal 60S a citoplasma tales como Npl3, Mex67, Mtr2, Ecm1, Gle2, Bud20, Arx1 y Abl1, sin embargo, la proteína que tiene un papel protagónico en este proceso es Nmd3 (45,49). En el último paso de maduración de la subunidad 60S intervienen los factores EFL1 y SBDS que participan en la liberación del factor de iniciación eIF6 para permitir la unión de esta subunidad a la subunidad 40S (44). EFL1 se une al centro asociados de GTPasas y está en contacto directo con SBDS y eIF6 en un estado inactivo unido a GTP. Debido a que SBDS y EFL1 compiten por un sitio común de unión al ribosoma, el dominio V de EFL1 promueve un desplazamiento rotacional de 180° del dominio III del SBDS de la base del tallo P en la conformación cerrada (complejo SBDS-EFL1) y hacia la hélice 69 en la conformación abierta (complejo SBDS-EFL1- eIF6). En la conformación abierta, SBDS conduce el equilibrio de la GTPasa EFL1 hacia una conformación activa unida al asa sarcina-ricina (SRL), promoviendo el desplazamiento de eIF6 de la subunidad 60S. En el paso final del ciclo catalítico, algunos estudios proponen que la interacción de EFL1 con SRL contribuye a la hidrólisis de GTP, quedando unida a GDP cambiando la conformación de equilibrio de EFL1 a un estado de alta a baja afinidad

de unión al ribosoma y promoviendo la disociación de ambas proteínas EFL1 y SBDS de la subunidad 60S (50,51).

1.2.2 Las proteínas SBDS y EFL1

La proteína SBDS (Shwachman-Bodian-Diamond Syndrome protein) se encuentra conservada en los dominios Eukarya y Archaea (52,53). Estudios de cristalografía de rayos X, Resonancia Magnética Nuclear y análisis de secuencias de la familia de SBDS han demostrado que esta familia de proteínas está conformada por tres dominios (52,54-57). El dominio 1 (residuos S2-S96) conocido también como dominio FYSH, interacciona en el sitio P de la subunidad ribosomal 60S entre el asa P (hélice 80), el asa interna formada por la proteína uL16 y la hélice 69 del ARNr 25S (50). El dominio 2 (residuos D97-A170) promueve un cambio conformacional en EFL1 y media la unión a la subunidad 60S junto con el dominio 1. El dominio 3 (residuos H71-E250) está en contacto con la hélice 95 del SRL y la base del tallo P (uL11, hélices 43 y 44), y también contacta con la proteína EFL1 (58,59).

EFL1 (Elongation Factor Like-1) es una GTPasa homóloga a la translocasa ribosomal EF-G de procariontes y al factor de elongación 2 (EF-2) en eucariotas (60). Al igual que EF-2, EFL1 está constituida por cinco dominios de los cuales el dominio 1 contiene los motivos G1-G5 que unen e hidrolizan GTP. La presencia de una inserción de longitud variable en el dominio 2 distingue a EFL1 de las translocasas ribosomales (60-62). Para liberar eIF6 de la subunidad ribosomal mayor, EFL1 transita a través de varios estados conformacionales (50,63). EFL1 se une a la subunidad 60S en el centro asociado a GTPasa en una conformación unida a GTP y sufre un cambio conformacional dependiente a SBDS en el cual su dominio GTPasa junto con la inserción en el asa del dominio 2 desplaza estéricamente a eIF6. Subsecuentemente, la hidrólisis de GTP mediada por la subunidad ribosomal lleva a EFL1 a la conformación unida a GDP que desencadena la liberación de la GTPasa y su cofactor SBDS (50).

- *El Síndrome Shawchman-Diamond (SDS)*

El síndrome Shawchman-Diamond (OMIM #260400) es un desorden autosómico recesivo pleiotrópico. Este síndrome está asociado a mutaciones en los genes *SBDS* (Gene ID:51119), *EFL1* (Gene ID:79631) (64-67), *DNAJC21* (Gene ID:134218) (68) y *SRP54* (Gene ID: 6729) (69) que afectan la liberación de eIF6 de la subunidad 60S (70,71). Los pacientes sufren trastornos tanto hematológicos (neutropenia, anemia, trombocitopenia, hipoplasia de la médula ósea) como no hematológicos (anormalidades esqueléticas, defectos pancreáticos) (72). En el 90% de los casos, mutaciones en el gen *SBDS* que

codifica para la proteína SBDS son las responsables de SDS (73). SBDS interactúa con EFL1 para promover la liberación del factor de iniciación 6 (eIF6), el cual es un paso necesario para la unión de las subunidades ribosomales 40S y 60S, y formar la unidad 80S funcional del ribosoma. La pérdida de expresión y/o función de SBDS y EFL1 trae como consecuencia la retención de eIF6 en la subunidad 60S ocasionando afectaciones en el ensamblaje del ribosoma 80S (50,55).

2 Planteamiento del problema

Estudios realizados en nuestro grupo de investigación de complementación genética entre proteínas ortólogas de SBDS y EFL1, mostraron que ninguno fue capaz de reemplazar la función de los correspondientes genes de levadura de forma individual o conjunta. Lo que sugiere que el complejo SBDS-EFL1 reconoce otro componente celular de manera especie específica, y que, además, la maduración de la subunidad mayor del ribosoma puede estar conservada evolutivamente. A su vez, lo anterior nos llevó a sugerir una posible coevolución a nivel intra- e inter-proteínas en las familias de SBDS y EFL1.

Por otra parte, aún no se cuenta con estructuras a alta resolución que expliquen con exactitud cuáles aminoácidos de SBDS y EFL1 estabilizan las conformaciones abierta SBDS-EFL1-eIF6 y cerrada SBDS-EFL1 de estos complejos, y con ello entender mejor su accionar durante la liberación de eIF6. Lo que nos llevó a hacernos las siguientes preguntas: ¿Cuáles son las bases estructurales y de secuencia que han dictado la coevolución en las familias de SBDS y EFL1?

2.1 Hipótesis

Si las proteínas EFL1 y SBDS poseen dominios estructurales que se sabe están conservados en el dominio Eukarya, y juegan un rol fundamental en la última etapa de la maduración de la subunidad 60S pues se requiere su acción conjunta para liberar al factor de iniciación eIF6, entonces estas proteínas han coevolucionado con el objetivo de trabajar como un binomio funcional.

2.2 Objetivos

Objetivo general

Predecir la coevolución a nivel intra- e inter-proteína en las familias de las proteínas SBDS y EFL1 con la ayuda de herramientas Bioinformáticas.

Objetivos específicos

- Encontrar residuos dentro de la familia de proteínas SBDS y EFL1 (nivel intra-proteína) que han cambiado conjuntamente para mantener su estructura y plegamiento.
- Evaluar la evolución entre las familias de las proteínas de SBDS y EFL1 y predecir posibles contactos de interacción entre ambas proteínas.

3 Materiales y métodos

3.1 Cálculos para inferir coevolución a nivel intra e inter-proteína

En este trabajo se realizó un análisis evolutivo de las familias de proteínas SBDS y EFL1. En el caso de la familia de las SBDS se trabajó en los dominios Archaea y Eukarya. En la familia de EFL1, solo se restringió al dominio Eukarya, pues esta proteína es exclusiva de este. En ambos casos se incluyeron representantes de todos los subgrupos de dichos dominios. Para el análisis inter-proteína, se limitó al dominio Eukarya.

Para darle solución a los problemas planteados en la sección 1.1.3, se empleó el algoritmo del servidor EVcouplings como método global (74). Se usaron las secuencias en formato FASTA de las proteínas SBDS y EFL1 de humano como entrada. Se tomó la base de datos UniRef100 como referencia para la búsqueda de secuencias similares. El método estadístico empleado fue Maximización de Pseudo-verosimilitud (Pseudo-likelihood maximization) (23), pues este es el método establecido en EVcouplings para inferir los acoplamientos evolutivos (ECs).

En el análisis intra-proteína, se analizó cada una por separado, y se recolectaron los valores estadísticos más importantes para garantizar una mejor predicción en el modelo del complejo proteico. Para la familia de la proteína SBDS el mejor rango de bitscore fue 0.1 (el recomendado por el servidor, teniendo en cuenta otros parámetros como cantidad de gaps en los alineamientos, calidad de los resultados), con un total de 1338 de secuencias alineadas por el servidor, en las regiones correspondientes a las familias Pfam: PF01172 <https://pfam.xfam.org/family/PF01172> correspondiente a los residuos 14-101 de la secuencia primaria de SBDS humana y Pfam: PF09377 <http://pfam.xfam.org/family/PF09377>, residuos 107-225. Para la familia de la proteína EFL1 el análisis de la cadena peptídica fue dividido en dos partes debido a su tamaño. En la primera parte se alinearon 1387 secuencias de las regiones de la familia [Pfam: PF00009](https://pfam.xfam.org/family/GTP_EFTU) https://pfam.xfam.org/family/GTP_EFTU, GTP_EFTU (amino ácidos 18-268) y parte de la familia Pfam: PF14492 <http://pfam.xfam.org/family/PF14492>, EFG_II (615-678), con un

bitscore de 0.1 y un total de 1387 secuencias alineadas. En la segunda parte se alinearon 689 secuencias de las regiones [Pfam: PF14492](http://pfam.xfam.org/family/PF14492) <http://pfam.xfam.org/family/PF14492>, EFG_II (615-678) y [Pfam: PF00679](http://pfam.xfam.org/family/PF00679) <http://pfam.xfam.org/family/PF00679>, EFG_C (980-1064) con un bitscore de 0.5.

La coevolución entre residuos que interactúan es necesaria no solo para mantener la estructura tridimensional de las proteínas, sino también para mantener las interacciones entre proteínas. Para realizar el análisis coevolutivo entre proteínas que interactúan utilizando acoplamientos evolutivos, el servidor EVcouplings calculó los acoplamientos evolutivos dentro de cada monómero (intra-ECs), que están definiendo la interacción de cada subunidad y luego los acoplamientos evolutivos entre proteínas que determinan la interacción entre diferentes cadenas (inter-ECs). Se utilizó EVcomplex (26) para predecir posibles contactos proteína-proteína. Fueron seleccionados los dominios I,II y III de la proteína SBDS y los dominios II, III, IV y V de EFL1, teniendo en cuenta los resultados obtenidos por (58,59). Se utilizaron como entradas los alineamientos de las proteínas SBDS y EFL1 de humano y se hicieron combinaciones con los dominios de SBDS y EFL1 como se muestra en la sección 4.2. Los mejores bitscores para el complejo proteico, se seleccionaron de los resultados obtenidos anteriormente de coevolución a nivel intra-proteína.

4 Resultados

4.1 Coevolución a nivel intra-proteína

- **Pares de residuos con relación coevolutiva (Evolutionary Couplings, ECs)**

Para encontrar coevolución a nivel intra-proteína en las familias de las proteínas SBDS y EFL1 se utilizaron los alineamientos de las proteínas HsSBDS y HsEFL1 como entrada. EVcouplings predijo 225 pares de residuos que están relacionados evolutivamente en el caso de la familia de la proteína SBDS. Las estructuras experimentales utilizadas para el cálculo fueron PDBs: 2kdo:A (54); 2l9n:A (55); 5an9:J; 5anb:J; 5anc:J; 6qkl:J (50). La Tabla suplementaria 1 muestra los pares de residuos con mayor probabilidad de estar relacionados coevolutivamente. Se obtuvieron 535 ECs para la primera parte de HsEFL1 y 445 ECs para la segunda parte de HsEFL1. Las estructuras experimentales utilizadas para el cálculo fueron los PDBs: 5ANB:k, 5ANC:k (50). En la Tabla suplementaria 2 se presentan en detalle los pares de aminoácidos relacionados evolutivamente, la distancia entre ellos en la estructura 3D y la probabilidad de que ocurran tales acoplamientos.

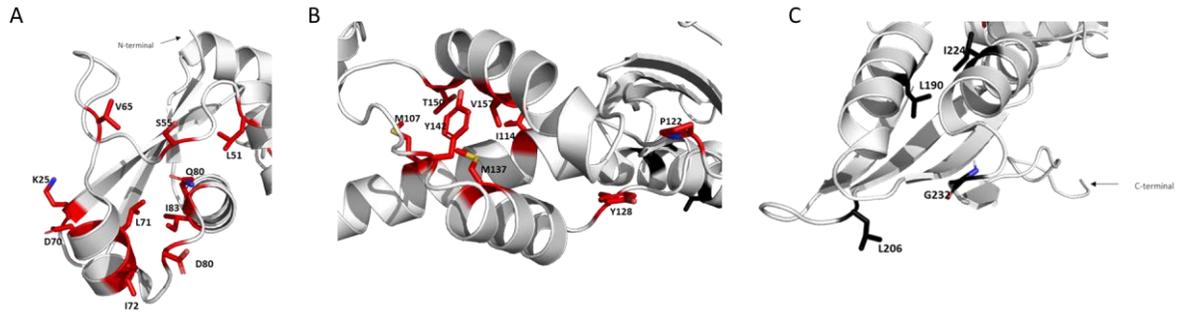


Figura 2. Interacciones a nivel intra-proteína en la familia de proteínas SBDS. Representación de las parejas de residuos con mayor relación coevolutiva en la estructura tridimensional de SBDS humana (PDB, 2L9N: A). **A:** dominio I, **B:** dominio II, **C:** dominio III. Los residuos representados en rojo son aquellos que están en contacto en la estructura terciaria de la proteína y los residuos en negro tienen menor probabilidad de estar en contacto en la estructura 3D, sin embargo, sí están relacionados evolutivamente.

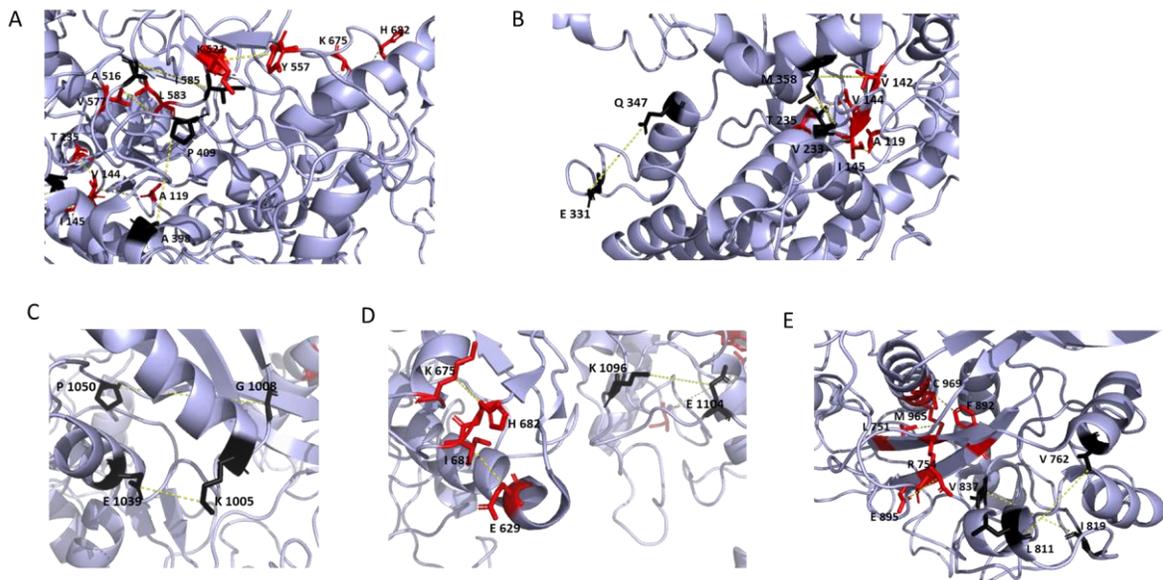


Figura 3. Interacciones a nivel intra-proteína en la familia de proteínas EFL1. Representación de las parejas de residuos con mayor probabilidad de coevolución en la estructura tridimensional de EFL1 humana (PDB: 5ANC: K). **A,B,C,D,E:** Vista más cercana de los dominios de EFL1 con pares acoplados evolutivamente. Los residuos representados en rojo son aquellos que están en contacto en la estructura terciaria de la proteína y los residuos en negro tienen menor probabilidad de estar en contacto en la estructura 3D.

En las Figura 2 y Figura 2 se pueden observar las estructuras tridimensionales de las proteínas humanas SBDS (PDB: 2L9N) y EFL1 (PDB:5ANC), representadas con la ayuda del visualizador molecular Pymol 2.5. Como se aprecia, para la familia de SBDS existe coevolución en los tres dominios estructurales de la proteína, y se predijeron contactos intra-proteína en los dominios I y II, principalmente. Los residuos M137 y L51 coincidieron con posiciones conservadas en la familia de SBDS (53). Para la familia de la proteína EFL1, se encontró coevolución y contactos entre residuos a nivel intra-proteína en los 5 dominios

estructurales con un mayor número en los dominios I, II y IV; con una probabilidad más baja en el dominio V. En ambas familias de proteínas se obtuvieron predicciones de aminoácidos relacionados evolutivamente y que, además están en contacto directo, como se esperaba. Aún en la literatura no hay evidencia experimental para corroborar dichos resultados, por lo que es necesario realizar experimentos en el laboratorio que ayuden a reforzar estas predicciones. Además, se encontraron otros pares de aminoácidos que están más lejanos en la estructura, pero que sí están relacionados evolutivamente.

- **Sitios acoplados fuertemente**

Adicionalmente a estos cálculos para encontrar coevolución dentro de las familias de SBDS y EFL1, se obtuvieron los sitios o posiciones fuertemente acopladas dentro de cada proteína. En el experimento anterior (coevolución a nivel intra-proteína en las familias SBDS y EFL1) se obtuvieron pares de aminoácidos relacionados evolutivamente, mientras que en los resultados que se presentan en esta sección se muestran cuáles son los aminoácidos que tienen mayor cantidad de aminoácidos asociados evolutivamente en sus alrededores. El cálculo de estas posiciones se basa en la suposición de que residuos con un fuerte acoplamiento evolutivo con respecto a otras posiciones, están bajo una fuerza de selección particular, las cuales puede que no sean visibles desde la conservación de un solo sitio por sí solo y por tanto, se cree que son funcionalmente importantes (75).

El cálculo arrojó un total de 192 sitios fuertemente acoplados en el caso de la familia de SBDS. En la

Figura 4A se presentan los 15 sitios con mayor puntaje de acoplamiento evolutivo (suma de puntajes EC a todas las demás posiciones a las que está acoplada ese sitio) en la estructura tridimensional de SBDS humana (PDB: 29LN: A). Los colores varían de azul fuerte a azul claro, indicando, el azul oscuro mayor número de residuos acoplados evolutivamente a ese sitio y el azul claro, menor cantidad de residuos acoplados a ese sitio. Los residuos más acoplados para la familia de SBDS resultaron ser K25, W39, V50, D70, D78, A74, Q80, L88, A112, I114, V115, M137, T150, I220. Los residuos con mayor número de acoplamientos pertenecen en su gran mayoría a los dominios I y II de SBDS. Los residuos L88, W39, V50, V115 y M137 en la familia de SBDS coinciden con posiciones conservadas según los estudios realizados por Boocock et. al. 2006(107), lo que pudiese indicar que estos desempeñan un rol relevante en el plegamiento y función de la proteína SBDS. Las posiciones restantes altamente acopladas evolutivamente resultaron ser posiciones

adyacentes o muy cercanas a residuos muy conservados e invariantes en la familia de esta proteína.

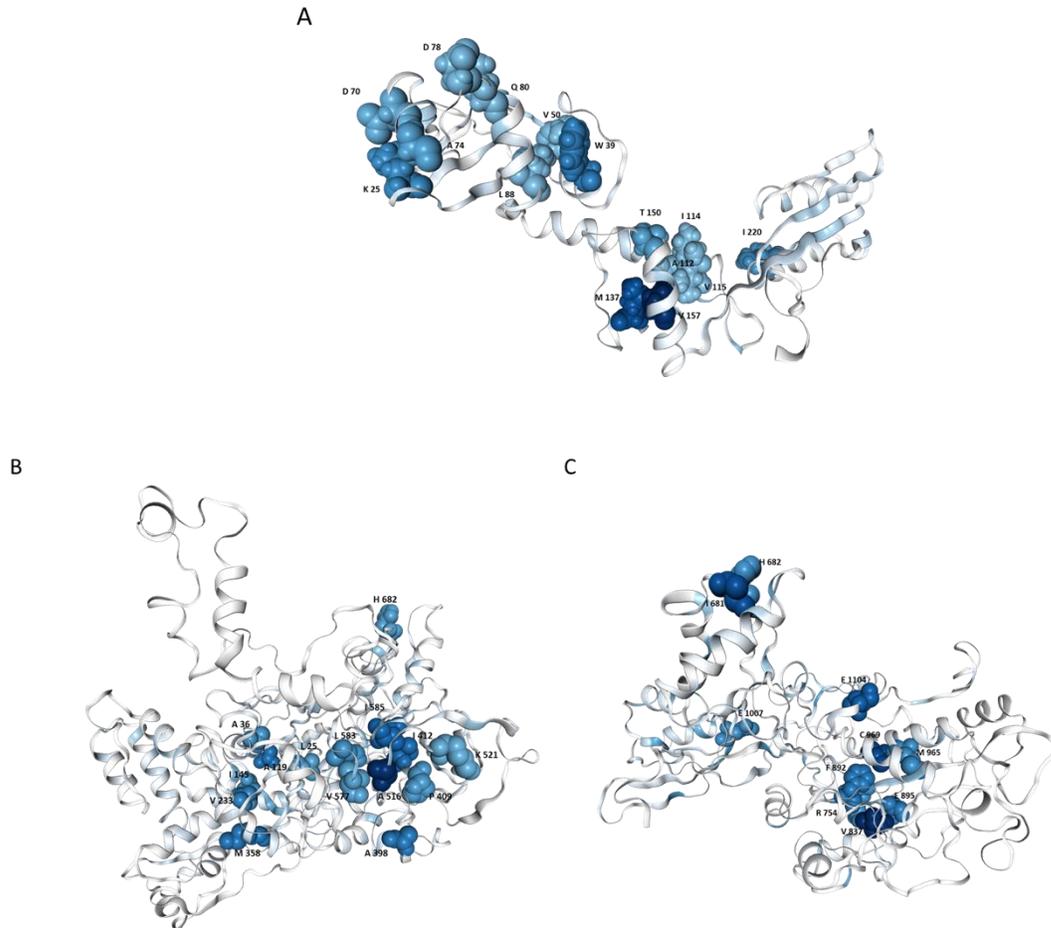


Figura 4. Sitios acoplados evolutivamente en la familia de proteínas SBDS y EFL1. A. Representación de los 15 residuos con mayor puntaje evolutivo en la estructura tridimensional de SBDS humana (PDB:2L9N: A), entre los que destacan M137, V157 y W39. **B y C.** Representación de los residuos con mayor puntaje evolutivo en la estructura de EFL1 humana (PDB: 5ANC: k). **B:** primera parte de EFL1; representación de los 15 sitios con mayor puntaje evolutivo. **C:** segunda parte de EFL1; representación de los 10 sitios de mayor puntaje evolutivo. Residuos más fuertemente acoplados: A516, V837, I681.

En la familia de EFL1 se obtuvieron un total de 980 sitios fuertemente acoplados evolutivamente, 535 para la primera parte de EFL1 y 445 para la segunda parte. Estos sitios se representaron en la estructura tridimensional de la proteína HsEFL1 (PDB: 5ANC, Figura 4B-C). Para la primera mitad de EFL1, los sitios más fuertemente acoplados fueron L25, A36, A119, I145, V233, M358, V577, A398, L583, I585, I412, K521, P409, A516, H682, mientras que para la segunda parte fueron H682, I681, E1007, E1104, C969, F892, R754, V837, E895, M965. Los sitios más acoplados evolutivamente en EFL1 fueron A516, V837, I681. Los residuos L25, A36, I412 coincidieron con posiciones invariantes según el estudio

realizado en nuestro laboratorio (76) para la familia de la proteína EFL1, y los residuos A119, V233, M358, P409, L583, I585, R754 coinciden con sitios conservados y solo dos residuos I145 y V577 correspondientes a la región de inserción característica de las EFL1 en levadura. La mayor cantidad de sitios con acoplamientos evolutivos se agrupa en los dominios II y IV de la familia de EFL1, lo cual tiene sentido teniendo en cuenta que las proteínas EFL1 presentan una inserción en el dominio II que las distingue de las EF-2 y pudiese indicar que existe especificidad en esta inserción para cada especie, por lo que ha tenido que ocurrir coevolución para mantener la funcionalidad.

4.2 Coevolución a nivel inter-proteína

EVcomplex (26) es una de las herramientas del servidor EVcouplings que sirve para predecir aminoácidos que interactúan entre proteínas a partir de la covariación de sus secuencias para un complejo proteico de interés. Primeramente, hace una búsqueda de secuencias homólogas a las secuencias de trabajo y crea sus propios alineamientos emparejados. Al igual que en el caso de monómeros, calcula acoplamientos evolutivos entre pares de residuos (ECs) pero en esta ocasión, lo hace a nivel intra-proteína e inter-proteína y seguidamente realiza un acoplamiento molecular o docking entre las proteínas de interés (25).

Con el objetivo de encontrar sitios que han coevolucionado y puntos de interacción entre las familias de las proteínas SBDS y EFL1, se hicieron cálculos con los tres dominios de SBDS y con los dominios II, III, IV y V de EFL1 basándonos en lo reportado en la literatura hasta el momento (58,59,76,77). La Tabla 1 muestra las combinaciones entre los dominios de las dos familias de proteínas que se evaluaron.

Tabla 1. Combinaciones entre los dominios de las familias SBDS y EFL1 evaluados para encontrar coevolución a nivel inter-proteína.

Dominio en SBDS	Dominio en EFL1	Probabilidad de interacción
I	II	no
I y II	II	no
I y II	III y IV	sí
I y II	III	no
II	II	no
II	III	no
II	IV	sí
II	III y IV	sí
II y III	II	sí

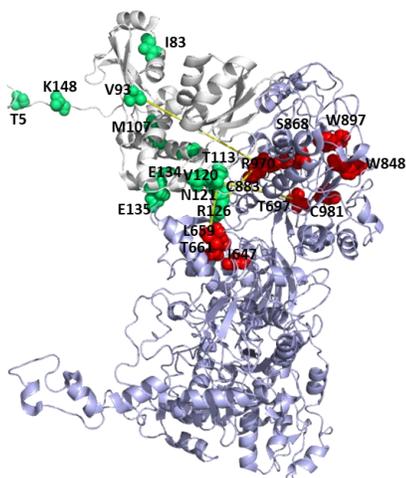
II y III	III y IV	sí
II y III	V	no

El resultado de combinar los dominios de SBDS y EFL1 fue positivo para cinco de esos casos como se muestra en la Tabla 1.

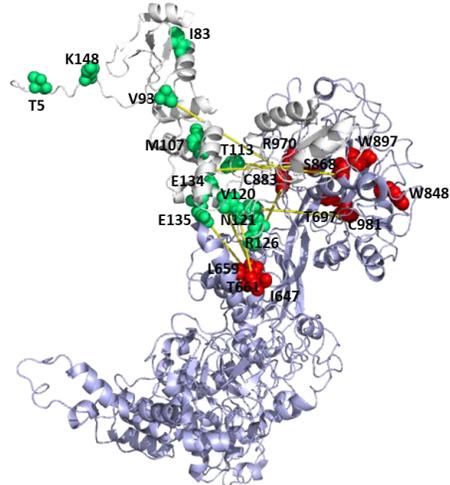
Se obtuvieron un total de 44 contactos con la probabilidad más alta de coevolución interacción inter-proteína, 12 entre los dominios I, II de SBDS y los dominios III y IV de EFL1, 4 entre el dominio II de SBDS y el dominio IV de EFL1, 4 entre el dominio II de SBDS y los dominios III y IV de EFL1, 9 entre los dominios II y III de SBDS y el dominio II de EFL1 y 15 entre los dominios II y III de SBDS los dominios III y IV de EFL1. En la Figura 5 se muestran las predicciones de coevolución y contactos entre las familias de proteínas de SBDS y EFL1. Las representaciones se hicieron utilizando el visualizador molecular Pymol (78) y los PDBs: 5ANB y 5ANC (50). Se alinearon 541 secuencias efectivas del dominio I, II de SBDS y los dominios III, IV de EFL1 (

Tabla suplementaria 3); 445 para el dominio II de SBDS y los dominios III, IV de EFL1 (Tabla suplementaria 4); 509 para el dominio II de SBDS y el dominio IV de EFL1(Tabla suplementaria 5); 414 para los dominios II,III de SBDS y IV de EFL1 (Tabla suplementaria 6), y 616 para los dominios II, III de SBDS y III,IV de EFL1(Tabla suplementaria 7).

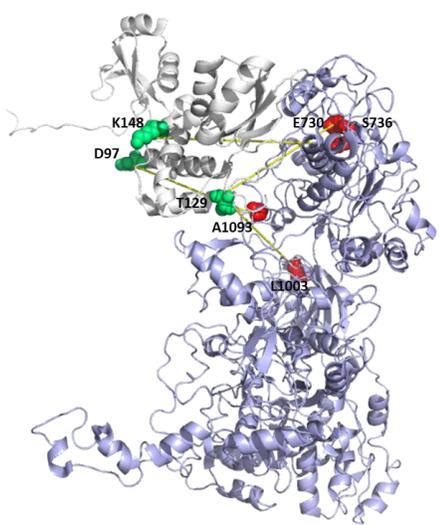
A



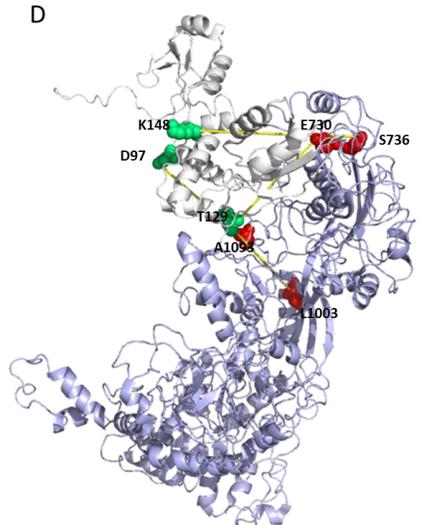
B



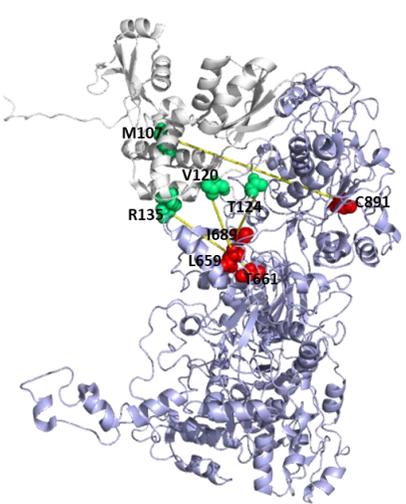
C



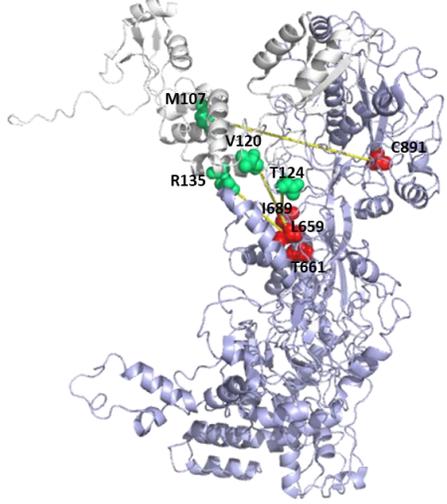
D



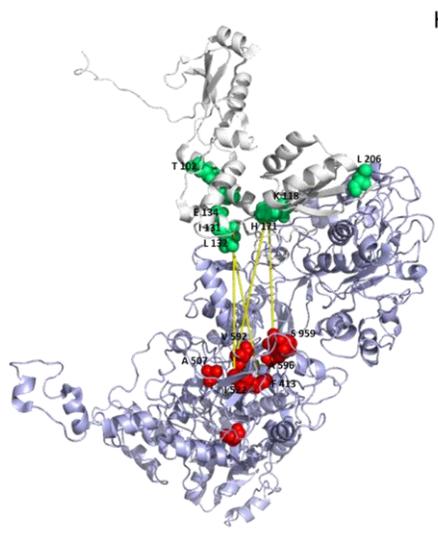
E



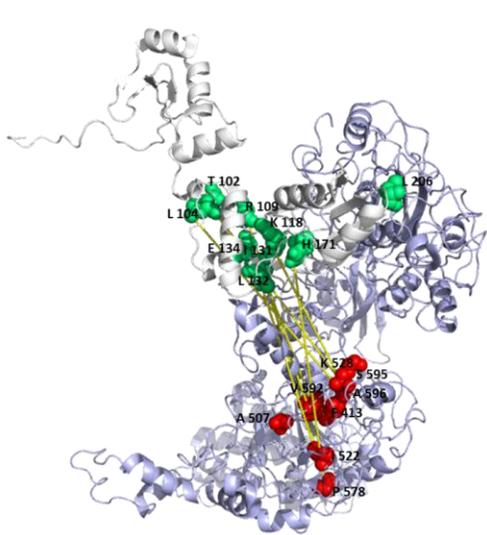
F



G



H



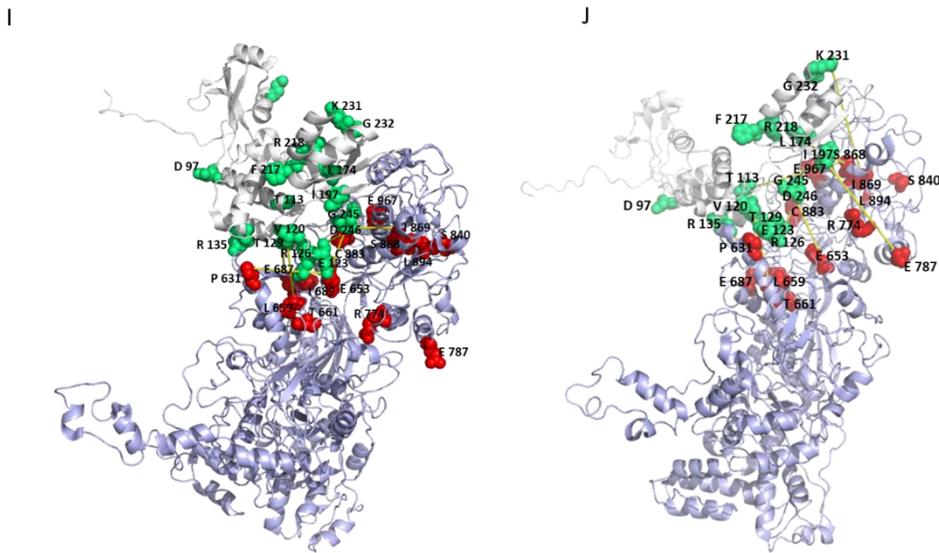


Figura 5. Interacciones inter-proteína entre las familias de proteínas SBDS y EFL1. Representación en las estructuras tridimensionales de los complejos proteico HsSBDS y HsEFL1, PDBs:5ANB (figuras de la izquierda) y 5ANC (figuras de la derecha). SBDS: Cadena de color gris. EFL1: cadena de color violeta. Residuos de color verde pertenecen a SBDS y residuos en color rojo pertenecen a EFL1. **A y B:** dominio I y II de SBDS con dominio III y IV de EFL1. **C y D:** dominio II de SBDS y dominio IV de EFL1. **E y F:** dominio II de SBDS y III y IV de EFL1. **G y H:** dominios II y III de SBDS y dominio II de EFL1. **I y J:** dominios II y III de SBDS y dominios III y IV de EFL1.

Los resultados obtenidos según los cálculos de coevolución inter-proteína mostraron que existe coevolución entre los tres dominios de SBDS y los dominios II, III, IV de EFL1, lo cual también pudiese estar relacionado con posible interacción física entre estos dominios en ambas familias de proteínas. Y esto puede ser corroborado según los resultados de criomicroscopía electrónica obtenida por *Weis et. al* (50). En los modelos de los complejos 60S-eIF6-SBDS-EFL1 y 60S-SBDS-EFL1, presentados con resoluciones que van desde 8 a 3.3 Å, se puede observar la cercanía que existe entre los dominios II y III de SBDS y los dominios III y IV de EFL1. En el caso de la proteína SBDS, algunos de los residuos que están relacionados evolutivamente y/o en contacto con EFL1 coinciden con mutaciones encontradas en pacientes con el Síndrome Shwachman-Diamond como son: R126 (73), N121 (79), T129 (55), K118 (52), R218 (55). En EFL1 también se encontraron residuos que coinciden con mutaciones que ocasionan este síndrome, como fueron C883 (65) y R970 (65). Y es de señalar que el residuo C883 de EFL1 parece estar relacionado evolutivamente con el residuo R126 de SBDS, según lo obtenido para los dominios I, II de SBDS con los dominios III, IV de EFL1 (Tabla suplementaria 4) y II,III de SBDS con III,IV de EFL1 (Tabla suplementaria 7).

5 Discusión

La coevolución se puede entender como la combinación de dos componentes esenciales y aparentemente contradictorios: la conservación y el cambio operando a diferentes niveles (7,8,74). La conservación se entiende, frecuentemente, como la ausencia de cambio, especialmente a nivel molecular. Esta resistencia al cambio es una señal evidente de una fuerte selección negativa, la cual se ve reflejada en la conservación de la interacción. La ruptura de dicha interacción ha de ser perjudicial para al menos una de las especies y la presión de selección se impone al efecto perjudicial sobre las especies beneficiarias que puedan tener otras especies en la interacción (8).

La introducción de nuevos métodos globales al cálculo de coevolución ha permitido una mejora notable en cuanto a la predicción de contactos (80) . Por ejemplo, con proteínas de membrana (22), además, en el estudio de conformaciones alternativas (81), estados multiméricos (82), procesos de plegamientos (83), distinción entre plegamientos nativos e incorrectos (84), proteínas desordenadas (29), estructuras de RNAs (85), interacciones entre proteínas (26,86) y predicción del efecto de mutaciones (30). La predicción de contactos mediante coevolución ha tenido un papel fundamental en la predicción de estructura en los últimos años. Como se puede ver en algunas ediciones de CASP, en particular la celebrada en 2018, el método AlphaFold logró un avance notorio en la categoría de modelado libre de plantillas basándose en el uso de predicción de distancias empleando coevolución (87). Aunque AlphaFold y EVcouplings usan el concepto de coevolución para hacer sus cálculos, la principal diferencia entre ambos métodos es que el primero se utiliza solo para predecir la estructura tridimensional y el plegamiento de proteínas, incluso de proteínas de las que aún no se ha reportado información estructural en el PDB. Mientras que el segundo además de emplearse para lo anterior, es más frecuentemente utilizado para predecir coevolución entre familias de proteínas, el efecto de mutaciones y la interacción entre residuos tanto dentro de una misma proteína como entre dos proteínas. Otros de los servidores disponibles para predecir coevolución son iCOMS, MISTIC, MirrorTree, DCA, entre otros. La principal causa de no haber seleccionado ninguno de los anteriores fue porque algunos de estos sí fueron probados por nosotros, pero al comparar con los resultados experimentales reportados hasta el momento estaban muy alejados de la evidencia experimental hasta ahora descrita, además, muchos de ellos eran solo para ser usados por usuarios expertos en programación, lo que evidentemente limita su uso. La selección de EVcouplings se hizo basada en las principales ventajas que ofrece este

servidor. EVcouplings es un programa escrito en Python para el cálculo de coevolución de moléculas de proteínas y de ARN, así como su interacción. Se basa en el cálculo de acoplamientos evolutivos (ECs) a partir de la covariación de sus secuencias de nucleótidos y de aminoácidos (74). Dentro de las herramientas de este servidor se encuentran *EVcouplings sever*, que predice la estructura 3D de monómeros. *EVcomplex server* predice residuos que están interactuando en complejos proteicos. *EVmutation*, mutaciones predichas para miles de proteínas humanas, entre ellas podemos encontrar HsSBDS y HsEFL1. *Protein plasticity* que predice conformaciones de regiones desordenadas en el proteoma humano. *RNA structure* predice la estructura 3D de moléculas de ARN e interacciones ARN-proteína. *E. coli complexes* que cuenta con predicciones pre-calculadas para el proteoma de *Escherichia coli*, incluido todo el interactoma de las proteínas de membrana. Las principales ventajas que ofrece este servidor son una interfaz gráfica agradable para usuarios principiantes en Bioinformática, la rapidez de los cálculos, hace sus propias búsquedas de secuencias en bases de datos de como el UniProt, realiza los alineamientos de estas secuencias utilizando jackhmmer y hmmsearch, compara los cálculos con información experimental depositada en la Base de Datos de Proteínas (PDB), lo cual proporciona resultados muy cercanos a la realidad, y es libre para ser utilizado desde su código Python. Algunas de las principales desventajas son que al tener que usar el servidor por ser un usuario principiante, esto te limita a poder usar tus propios alineamientos curados, lo que quiere decir que los alineamientos que genera el servidor pueden tener secuencias no deseadas que de alguna forma pudiesen estar afectando la calidad de las predicciones, y además, en el cálculo inter-proteína, limita el largo de las secuencias a solo 500 aminoácidos, lo que significa, en el caso del presente trabajo, que se tuvo que dividir la proteína EFL en dos partes dado su tamaño.

La Maximización de Pseudo-verosimilitud (Pseudo-likelihood maximization) es el método implementado en EVcouplings para inferir acoplamientos evolutivos a partir de alineamientos, utilizando un modelo gráfico indirecto de pares de secuencias (23). Es un método global para calcular coevolución por lo que al emplearlo se evitan las correlaciones indirectas de pares de residuos comentado anteriormente en el epígrafe 1.1.3. Los métodos globales infieren una descripción probabilística de los alineamientos de secuencias explicando las correlaciones observadas, utilizando acoplamientos causales entre posiciones (75).

Los resultados de coevolución a nivel intra-proteína para la familia de proteínas SBDS y EFL1 predicen co-evolución a nivel de residuos en varias regiones de estas proteínas. Por

ejemplo, para la proteína SBDS, como se observa en la tabla 2 y figura 1, se encontró coevolución entre residuos en sus tres dominios estructurales, con un mayor número en los dominios I y II, lo cual se puede explicar teniendo en cuenta la importancia funcional de estos dos dominios. Como demuestran los ensayos de complementación genética realizados por *Boocock et. al* en 2006, el dominio I de SBDS juega un papel fundamental en la viabilidad celular y el dominio II es indispensable para la función de la proteína y además es especie específico, mientras que el dominio III es dispensable (53). En el caso de la proteína EFL1, los dominios con mayor número de pares evolutivos fueron el I, II y IV. Esto es de esperarse ya que los dominios II de la familia de esta GTPasa contienen inserciones que varían en longitud de aminoácidos, además de contactar a la proteína SBDS; el dominio IV conduce al reconocimiento de la subunidad 60S, lo que parece sugerir que este fenómeno pudiera darse de forma especie-específico(76) y por último, se puede explicar la coevolución en el dominio I, ya que estudios por *Méndez-Godoy (76)* mostraron que en algunos organismos como *S. cerevisiae*, *S. pombe*, *H. sapiens*, *A. thaliana* y *P. falciparum* existen inserciones en el dominio G además de los dominios II, IV y V. Los sitios evolutivos de mayor acoplamiento se hallaron en los dominios I y II de SBDS y en los dominios II y IV de EFL1. Algunos de los sitios con mayores valores de acoplamientos evolutivos en SBDS coincidieron con posiciones muy conservadas (53). Para EFL1 algunos de estos sitios coincidieron con posiciones invariantes y otros con posiciones conservadas dentro de la familia (76). El hecho de que coincidan con sitios conservadas dentro de cada familia puede estar indicando que dichos aminoácidos son imprescindibles para mantener la función y la estructura de estas proteínas.

La herramienta EVcomplex predijo residuos que han coevolucionado conjuntamente entre las dos familias de proteínas SBDS y EFL1. Según los resultados obtenidos, utilizando Maximización de Pseudo-Verosimilitud, se predijo coevolución/interacción, entre los dominios I, II y III de SBDS y los dominios II, III y IV de EFL1, con una mayor abundancia entre los dominios II y III de SBDS y los dominios III y IV de EFL1. Esto coincide con estudios que se han realizado con anterioridad, donde se muestra que en la última etapa de maduración de la subunidad 60S el dominio II de la proteína SBDS interactúa con EFL1 para ayudar a esta proteína en su cambio conformacional. El dominio III de SBDS previene la unión de la GTPasa a la base del tallo P, y posteriormente, es desplazado de este sitio por el dominio V de EFL1. Lo que explica la homología estructural entre dichos dominios de estas proteínas. Es decir, SBDS y EFL1 comparten un sitio de unión común en la subunidad 60S pero no pueden unirse simultáneamente a este (50). Otros estudios realizados en

nuestro grupo de investigación (59,76) donde se hicieron ensayos de doble híbrido y anisotropía de fluorescencia entre las proteínas SBDS y EFL1, mostraron que el dominio II de la proteína ScSdo1 estaba implicado en la interacción con el dominio II de ScEfl1 (58). Y que los dominios II y III de SBDS eran suficientes en la interacción con EFL1, pero la unión se veía más favorecida en la presencia del dominio I de SBDS mediante cooperatividad positiva con este dominio (59).

Asimismo, nuestros resultados indicaron que los residuos R126 de SBDS y el C883 de EFL1, están relacionados evolutivamente y pudiesen ser residuos que están en contacto. Como muestran estudios anteriores, obtenidos por *Finch et. al 2011* (55), el aminoácido R126 está mutado en pacientes con el Síndrome Shwachman Diamond. En ese estudio se demostró que las mutaciones en R126 no afectaban el plegamiento de la proteína SBDS con respecto a la proteína silvestre. Sin embargo, a pesar de que sí se unía eficientemente a la pre-subunidad 60S no mejoraba la actividad de la GTPasa EFL1 en comparación con la SBDS silvestre. Además, la actividad de la SBDS mutada en R126 es deficiente para desencadenar la liberación de eIF6, en comparación a la SBDS silvestre (55). Otros resultados obtenidos en nuestro laboratorio por *Gijsbers et. al* (59) demostraron que los residuos K126 y S143 del ortólogo de levadura de SBDS están en contacto directo con EFL1. Más aún, estos residuos son equivalentes a aquellos descritos en pacientes con Síndrome Shwachman Diamond y su mutación afecta la interacción con EFL1. Por otra parte, las mutaciones observadas en C883, según *Tan et. al 2019* (65) fueron asociadas a la reducción en la expresión de EFL1, y al igual que el residuo R970, C883 se encuentra en la interfaz entre el dominio IV de EFL1 y el dominio C-terminal de SBDS lo que pudiese interrumpir la activación alostérica que ejerce SBDS sobre EFL1.

6 Conclusiones

- Los resultados a nivel intra-proteína mostraron que los dominios I y II de la familia de SBDS son los que más han sufrido evolución recíproca, y para la familia de EFL1 esto se observa en los dominios I, II y IV.
- Los residuos de mayor importancia evolutiva en SBDS resultaron ser K25, W39, V50, D70, D78, A74, Q80, L88, A112, I114, V115, M137, T150, I220.
- Los aminoácidos de mayor importancia evolutiva en EFL1 fueron L25, A36, A119, I145, V233, M358, V577, A398, L583, I585, I412, K521, P409, A516, H682, H682, I681, E1007, E1104, C969, F892, R754, V837, E895, M965.

- Los estudios a nivel inter-proteína revelaron coevolución/interacción entre los dominios I, II y III de SBDS y los dominios II, III y IV de EFL1.
- Los aminoácidos R126, N121, T129, K118, R218 de SBDS para los cuales se encontró coevolución/interacción con la familia de EFL1 son sitios en los que los que se han descrito mutaciones en pacientes con el Síndrome Shwachman Diamond.
- Se encontró coevolución/interacción entre los aminoácidos C883 y R970 de EFL1 que corresponden a posiciones mutadas en los pacientes con Síndrome Shwachman Diamond, y, además, están relacionados evolutivamente con el residuo R126 de SBDS.

7 Perspectivas

- Realizar mutaciones en pares de residuos en las proteínas SBDS y EFL1 para los que se predijo coevolución.
- Evaluar experimentalmente la co-evolución dentro de cada familia y entre las dos familias.

8 Bibliografía

1. Darwin, C. (1930) *The Origin of Species by Means of Natural Selection: Or, The Preservation of Favoured Races in the Struggle for Life*, Odhams
2. Ehrlich, P. R., and Raven, P. H. (1964) Butterflies and plants: a study in coevolution. *Evolution*, 586-608
3. Thompson, J. (1994) *The Coevolutionary Process*. Chicago: Univ. Chicago Press
4. Van, V., and Van Valen, L. (1973) A new evolutionary law.
5. Cott, H. B. (1940) Adaptive coloration in animals.
6. Brown, J. K., and Tellier, A. (2011) Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual review of phytopathology* **49**, 345-367
7. De Juan, D., Pazos, F., and Valencia, A. (2013) Emerging methods in protein co-evolution. *Nature Reviews Genetics* **14**, 249-261
8. Juan Sopeña, D. A. d. (2016) Desarrollo de métodos computacionales basados en co-evolución para la predicción de interacciones entre proteínas.
9. Altschuh, D., Lesk, A., Bloomer, A., and Klug, A. (1987) Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. *Journal of molecular biology* **193**, 693-707
10. Göbel, U., Sander, C., Schneider, R., and Valencia, A. (1994) Correlated mutations and residue contacts in proteins. *Proteins: Structure, Function, and Bioinformatics* **18**, 309-317
11. Shindyalov, I., Kolchanov, N., and Sander, C. (1994) Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? *Protein Engineering, Design and Selection* **7**, 349-358
12. Olmea, O., Rost, B., and Valencia, A. (1999) Effective use of sequence correlation and conservation in fold recognition. *Journal of molecular biology* **293**, 1221-1239
13. Ivankov, D. N., Finkelstein, A. V., and Kondrashov, F. A. (2014) A structural perspective of compensatory evolution. *Current opinion in structural biology* **26**, 104-112
14. Pazos, F., and Valencia, A. (2008) Protein co-evolution, co-adaptation and interactions. *The EMBO journal* **27**, 2648-2655
15. Eck, R. V., and Dayhoff, M. O. (1966) Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science* **152**, 363-366

16. Tang, H., and Thomas, P. D. (2016) Tools for predicting the functional impact of nonsynonymous genetic variation. *Genetics* **203**, 635-647
17. Goh, C.-S., Bogan, A. A., Joachimiak, M., Walther, D., and Cohen, F. E. (2000) Co-evolution of proteins with their interaction partners. *Journal of molecular biology* **299**, 283-293
18. Pazos, F., and Valencia, A. (2001) Similarity of phylogenetic trees as indicator of protein–protein interaction. *Protein engineering* **14**, 609-614
19. Rodriguez Rivas, J. (2020) La coevolución en regiones de interacción entre proteínas: estudio y desarrollo de métodos computacionales.
20. Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M., and Laub, M. T. (2008) Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043-1054
21. Lapedes, A., Giraud, B., and Jarzynski, C. (2012) Using sequence alignments to predict protein structure and stability with high accuracy. *arXiv preprint arXiv:1207.2484*
22. Hopf, T. A., Colwell, L. J., Sheridan, R., Rost, B., Sander, C., and Marks, D. S. (2012) Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* **149**, 1607-1621
23. Ekeberg, M., Lökvist, C., Lan, Y., Weigt, M., and Aurell, E. (2013) Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. *Physical Review E* **87**, 012707
24. Marks, D. S., Hopf, T. A., and Sander, C. (2012) Protein structure prediction from sequence variation. *Nature biotechnology* **30**, 1072-1080
25. Hopf, T. (2015) *Phenotype prediction from evolutionary sequence covariation* Dr. rer. nat, PhD thesis, TUM
26. Hopf, T. A., Schärfe, C. P., Rodrigues, J. P., Green, A. G., Kohlbacher, O., Sander, C., Bonvin, A. M., and Marks, D. S. (2014) Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife* **3**, e03430
27. Dos Santos, R. N., Morcos, F., Jana, B., Andricopulo, A. D., and Onuchic, J. N. (2015) Dimeric interactions and complex formation using direct coevolutionary couplings. *Scientific reports* **5**, 1-10
28. Morcos, F., Jana, B., Hwa, T., and Onuchic, J. N. (2013) Coevolutionary signals across protein lineages help capture multiple protein conformations. *Proceedings of the National Academy of Sciences* **110**, 20533-20538
29. Toth-Petroczy, A., Palmedo, P., Ingraham, J., Hopf, T. A., Berger, B., Sander, C., and Marks, D. S. (2016) Structured states of disordered proteins from genomic sequences. *Cell* **167**, 158-170. e112
30. Hopf, T. A., Ingraham, J. B., Poelwijk, F. J., Schärfe, C. P., Springer, M., Sander, C., and Marks, D. S. (2017) Mutation effects predicted from sequence co-variation. *Nature biotechnology* **35**, 128-135
31. Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., and Franceschi, F. (2000) Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *cell* **102**, 615-623
32. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. D. (2005) Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827-834
33. Selmer, M., Dunham, C. M., Murphy, F. V., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R., and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935-1942
34. Ben-Shem, A., Jenner, L., Yusupova, G., and Yusupov, M. (2010) Crystal structure of the eukaryotic ribosome. *Science* **330**, 1203-1209
35. Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., and Ban, N. (2011) Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* **334**, 941-948
36. Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A., and Ban, N. (2011) Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* **331**, 730-736

37. Ameismeier, M., Zemp, I., van den Heuvel, J., Thoms, M., Berninghausen, O., Kutay, U., and Beckmann, R. (2020) Structural basis for the final steps of human 40S ribosome maturation. *Nature* **587**, 683-687
38. Baßler, J., and Hurt, E. (2019) Eukaryotic ribosome assembly. *Annual review of biochemistry* **88**, 281-306
39. Klug, W. S., Cummings, M. R., Spencer, C. A., Palladino, M. A., and Killian, D. (2019) *Concepts of genetics*,
40. Warner, J. R. (1999) The economics of ribosome biosynthesis in yeast. *Trends in biochemical sciences* **24**, 437-440
41. Fromont-Racine, M., Senger, B., Saveanu, C., and Fasiolo, F. (2003) Ribosome assembly in eukaryotes. *Gene* **313**, 17-42
42. Kressler, D., Hurt, E., and Baßler, J. (2010) Driving ribosome assembly. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1803**, 673-683
43. Thomson, E., Ferreira-Cerca, S., and Hurt, E. (2013) Eukaryotic ribosome biogenesis at a glance. *Journal of cell science* **126**, 4815-4821
44. Pena, C., Hurt, E., and Panse, V. G. (2017) Eukaryotic ribosome assembly, transport and quality control. *Nature structural & molecular biology* **24**, 689-699
45. Woolford Jr, J. L., and Baserga, S. J. (2013) Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* **195**, 643-681
46. Panse, V. G., and Johnson, A. W. (2010) Maturation of eukaryotic ribosomes: acquisition of functionality. *Trends in biochemical sciences* **35**, 260-266
47. Nissan, T. A., Baßler, J., Petfalski, E., Tollervey, D., and Hurt, E. (2002) 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm. *The EMBO journal* **21**, 5539-5547
48. Harnpicharnchai, P., Jakovljevic, J., Horsey, E., Miles, T., Roman, J., Rout, M., Meagher, D., Imai, B., Guo, Y., and Brame, C. J. (2001) Composition and functional characterization of yeast 66S ribosome assembly intermediates. *Molecular cell* **8**, 505-515
49. Nerurkar, P., Altvater, M., Gerhardy, S., Schütz, S., Fischer, U., Weirich, C., and Panse, V. G. (2015) Eukaryotic ribosome assembly and nuclear export. *International review of cell and molecular biology* **319**, 107-140
50. Weis, F., Giudice, E., Churcher, M., Jin, L., Hilcenko, C., Wong, C. C., Traynor, D., Kay, R. R., and Warren, A. J. (2015) Mechanism of eIF6 release from the nascent 60S ribosomal subunit. *Nature structural & molecular biology* **22**, 914-919
51. Menne, T. F., Goyenechea, B., Sánchez-Puig, N., Wong, C. C., Tonkin, L. M., Ancliff, P. J., Brost, R. L., Costanzo, M., Boone, C., and Warren, A. J. (2007) The Shwachman-Bodian-Diamond syndrome protein mediates translational activation of ribosomes in yeast. *Nature genetics* **39**, 486-495
52. Shammas, C., Menne, T. F., Hilcenko, C., Michell, S. R., Goyenechea, B., Boocock, G. R., Durie, P. R., Rommens, J. M., and Warren, A. J. (2005) Structural and mutational analysis of the SBDS protein family: insight into the leukemia-associated Shwachman-Diamond syndrome. *Journal of Biological Chemistry* **280**, 19221-19229
53. Boocock, G., Marit, M., and Rommens, J. (2006) Phylogeny, sequence conservation, and functional complementation of the SBDS protein family. *Genomics* **87**, 758-771
54. de Oliveira, J. F., Sforça, M. L., Blumenschein, T. M., Goldfeder, M. B., Guimarães, B. G., Oliveira, C. C., Zanchin, N. I., and Zeri, A.-C. (2010) Structure, dynamics, and RNA interaction analysis of the human SBDS protein. *Journal of molecular biology* **396**, 1053-1069
55. Finch, A. J., Hilcenko, C., Basse, N., Drynan, L. F., Goyenechea, B., Menne, T. F., Fernández, Á. G., Simpson, P., D'Santos, C. S., and Arends, M. J. (2011) Uncoupling of GTP hydrolysis from eIF6 release on the ribosome causes Shwachman-Diamond syndrome. *Genes & development* **25**, 917-929
56. Savchenko, A., Krogan, N., Cort, J. R., Evdokimova, E., Lew, J. M., Yee, A. A., Sánchez-Pulido, L., Andrade, M. A., Bochkarev, A., and Watson, J. D. (2005) The Shwachman-Bodian-Diamond syndrome protein family is involved in RNA metabolism. *Journal of Biological Chemistry* **280**, 19213-19220
57. Ng, C. L., Waterman, D. G., Koonin, E. V., Walters, A. D., Chong, J. P., Isupov, M. N., Lebedev, A. A., Bunka, D. H., Stockley, P. G., and Ortiz-Lombardía, M. (2009)

- Conformational flexibility and molecular interactions of an archaeal homologue of the Shwachman-Bodian-Diamond syndrome protein. *BMC structural biology* **9**, 1-15
58. Asano, N., Atsuumi, H., Nakamura, A., Tanaka, Y., Tanaka, I., and Yao, M. (2014) Direct interaction between EFL1 and SBDS is mediated by an intrinsically disordered insertion domain. *Biochemical and biophysical research communications* **443**, 1251-1256
 59. Gijssbers, A., Montagut, D. C., Méndez-Godoy, A., Altamura, D., Saviano, M., Siliqi, D., and Sánchez-Puig, N. (2018) Interaction of the GTPase Elongation Factor Like-1 with the Shwachman-Diamond syndrome protein and its missense mutations. *International journal of molecular sciences* **19**, 4012
 60. Senger, B., Lafontaine, D. L., Graindorge, J.-S., Gadal, O., Camasses, A., Sanni, A., Garnier, J.-M., Breitenbach, M., Hurt, E., and Fasiolo, F. (2001) The nucle (ol) ar Tif6p and Efl1p are required for a late cytoplasmic step of ribosome synthesis. *Molecular cell* **8**, 1363-1373
 61. Bécam, A.-M., Nasr, F., Racki, W., Zagulski, M., and Herbert, C. (2001) Ria1p (Ynl163c), a protein similar to elongation factors 2, is involved in the biogenesis of the 60S subunit of the ribosome in *Saccharomyces cerevisiae*. *Molecular Genetics and Genomics* **266**, 454-462
 62. Jørgensen, R., Ortiz, P. A., Carr-Schmid, A., Nissen, P., Kinzy, T. G., and Andersen, G. R. (2003) Two crystal structures demonstrate large conformational changes in the eukaryotic ribosomal translocase. *Nature Structural & Molecular Biology* **10**, 379-385
 63. Luviano, A., Cruz-Castañeda, R., Sánchez-Puig, N., and García-Hernández, E. (2019) Cooperative energetic effects elicited by the yeast Shwachman-Diamond syndrome protein (Sdo1) and guanine nucleotides modulate the complex conformational landscape of the elongation factor-like 1 (Efl1) GTPase. *Biophysical chemistry* **247**, 13-24
 64. Stepensky, P., Chacón-Flores, M., Kim, K. H., Abuzaitoun, O., Bautista-Santos, A., Simanovsky, N., Siliqi, D., Altamura, D., Méndez-Godoy, A., and Gijssbers, A. (2017) Mutations in EFL1, an SBDS partner, are associated with infantile pancytopenia, exocrine pancreatic insufficiency and skeletal anomalies in a Shwachman-Diamond like syndrome. *Journal of Medical Genetics* **54**, 558-566
 65. Tan, S., Kermasson, L., Hoslin, A., Jaako, P., Faille, A., Acevedo-Arozena, A., Lengline, E., Ranta, D., Poirée, M., and Fenneteau, O. (2019) EFL1 mutations impair eIF6 release to cause Shwachman-Diamond syndrome. *Blood, The Journal of the American Society of Hematology* **134**, 277-290
 66. Tan, Q. K.-G., Cope, H., Spillmann, R. C., Stong, N., Jiang, Y.-H., McDonald, M. T., Rothman, J. A., Butler, M. W., Frush, D. P., and Lachman, R. S. (2018) Further evidence for the involvement of EFL1 in a Shwachman–Diamond-like syndrome and expansion of the phenotypic features. *Molecular Case Studies* **4**, a003046
 67. Lee, S., Shin, C. H., Lee, J., Jeong, S. D., Hong, C. R., Kim, J.-D., Kim, A.-R., Park, B., Son, S. J., and Kokhan, O. (2021) Somatic uniparental disomy mitigates the most damaging EFL1 allele combination in Shwachman-Diamond syndrome. *Blood*
 68. Tummalam, H., Walne, A. J., Williams, M., Bockett, N., Collopy, L., Cardoso, S., Ellison, A., Wynn, R., Leblanc, T., and Fitzgibbon, J. (2016) DNAJC21 mutations link a cancer-prone bone marrow failure syndrome to corruption in 60S ribosome subunit maturation. *The American Journal of Human Genetics* **99**, 115-124
 69. Carapito, R., Konantz, M., Paillard, C., Miao, Z., Pichot, A., Leduc, M. S., Yang, Y., Bergstrom, K. L., Mahoney, D. H., and Shardy, D. L. (2017) Mutations in signal recognition particle SRP54 cause syndromic neutropenia with Shwachman-Diamond-like features. *The Journal of clinical investigation* **127**, 4090-4103
 70. Bezzerri, V., Bardelli, D., Morini, J., Vella, A., Cesaro, S., Sorio, C., Biondi, A., Danesino, C., Farruggia, P., and Assael, B. M. (2018) Ataluren-driven restoration of Shwachman-Bodian-Diamond syndrome protein function in Shwachman-Diamond syndrome bone marrow cells. *American journal of hematology* **93**, 527-536
 71. Orgebin, E., Lamoureux, F., Isidor, B., Charrier, C., Ory, B., Lézot, F., and Baud'huin, M. (2020) Ribosomopathies: New Therapeutic Perspectives. *Cells* **9**, 2080
 72. Rothbaum, R., Perrault, J., Vlachos, A., Cipolli, M., Alter, B. P., Burroughs, S., Durie, P., Elghetany, M. T., Grand, R., and Hubbard, V. (2002) Shwachman-Diamond syndrome: report from an international conference. *The Journal of pediatrics* **141**, 266-270

73. Boocock, G. R., Morrison, J. A., Popovic, M., Richards, N., Ellis, L., Durie, P. R., and Rommens, J. M. (2003) Mutations in SBDS are associated with Shwachman–Diamond syndrome. *Nature genetics* **33**, 97-101
74. Hopf, T. A., Green, A. G., Schubert, B., Mersmann, S., Schärfe, C. P., Ingraham, J. B., Toth-Petroczy, A., Brock, K., Riesselman, A. J., and Palmado, P. (2019) The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics* **35**, 1582-1584
75. Hopf, T. (2016) *Phenotype prediction from evolutionary sequence covariation*, Technische Universität München
76. Méndez-Godoy, A., García-Montalvo, D., Martínez-Castilla, L. P., and Sánchez-Puig, N. (2021) Evolutionary and functional relationships in the ribosome biogenesis SBDS and EFL1 protein families. *Molecular Genetics and Genomics*, 1-16
77. Holding, A. N. (2015) XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods* **89**, 54-63
78. System, P. M. G. (2017) Version 2.0 Schrödinger, LLC.
79. Erdős, M., Alapi, K., Balogh, I., Oroszlán, G., Rákóczi, É., Sümegi, J., and Maródi, L. (2006) Severe Shwachman-Diamond syndrome phenotype caused by compound heterozygous missense mutations in the SBDS gene. *Experimental hematology* **34**, 1517-1521
80. Weigt, M., White, R. A., Szurmant, H., Hoch, J. A., and Hwa, T. (2009) Identification of direct residue contacts in protein–protein interaction by message passing. *Proceedings of the National Academy of Sciences* **106**, 67-72
81. Sfriso, P., Duran-Frigola, M., Mosca, R., Emperador, A., Aloy, P., and Orozco, M. (2016) Residues coevolution guides the systematic identification of alternative functional conformations in proteins. *Structure* **24**, 116-126
82. Uguzzoni, G., Lovis, S. J., Oteri, F., Schug, A., Szurmant, H., and Weigt, M. (2017) Large-scale identification of coevolution signals across homo-oligomeric protein interfaces by direct coupling analysis. *Proceedings of the National Academy of Sciences* **114**, E2662-E2671
83. Morcos, F., Schafer, N. P., Cheng, R. R., Onuchic, J. N., and Wolynes, P. G. (2014) Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. *Proceedings of the National Academy of Sciences* **111**, 12408-12413
84. Wozniak, P. P., Vriend, G., and Kotulska, M. (2017) Correlated mutations select misfolded from properly folded proteins. *Bioinformatics* **33**, 1497-1504
85. Taylor, W. R., and Hamilton, R. S. (2017) Exploring RNA conformational space under sparse distance restraints. *Scientific reports* **7**, 1-15
86. Rodríguez-Rivas, J., Marsili, S., Juan, D., and Valencia, A. (2016) Conservation of coevolving protein interfaces bridges prokaryote–eukaryote homologies in the twilight zone. *Proceedings of the National Academy of Sciences* **113**, 15018-15023
87. AlQuraishi, M. (2019) AlphaFold at CASP13. *Bioinformatics* **35**, 4862-4865

9 Anexos

- Coevolución intra-proteína

Tabla suplementaria 1. Mejores predicciones de residuos relacionados de forma evolutiva en la familia de proteínas SBDS especificados de acuerdo con la secuencia primaria de la proteína humana.

Pares de residuos relacionados evolutivamente	Puntaje	Probabilidad	Distancia [Å]
M137 - V157	10.96	1.00	2.47
K25 - D70	9.33	1.00	3.07
P122 - Y128	7.74	1.00	4.19
S55 - V65	7.44	1.00	2.87
Y142 - V157	7.02	1.00	2.17
I72 - D78	6.97	1.00	2.00
I114 - T150	6.73	1.00	3.49
L71 - I83	5.53	1.00	1.98
L51 - Q80	5.48	1.00	4.35

M107 - T150	5.31	0.99	2.17
L190 - L206	5.06	0.99	5.60

Puntaje de acoplamiento entre las posiciones.

Probabilidad de que las posiciones de aminoácidos que estén acopladas.

Distancia atómica mínima (pares de átomos más cercanos en angstrom [Å]) en la estructura tridimensional.

Tabla suplementaria 2. Mejores predicciones de residuos relacionados de forma evolutiva en la familia de proteínas EFL1 especificados de acuerdo con la secuencia primaria de la proteína humana.

Pares de residuos relacionados evolutivamente	Puntaje	Probabilidad	Distancia [Å]
A119 - I145	13.66	1.00	4.08
A398 - P409	12.99	1.00	8.99
V144 - T235	11.41	1.00	3.17
A516 - I585	11.18	1.00	9.53
V233 - M358	11.13	1.00	6.63
V577 - L583	11.11	1.00	3.39
K521 - Y557	9.94	1.00	4.12
V142 - M358	9.00	1.00	3.55
K675 - H682	8.83	1.00	3.99
E331 - Q347	8.80	1.00	8.32
R754 - E895	6.33	1.00	3.01
E629 - I681	5.07	0.99	3.76
G1008 - P1050	4.78	0.99	13.16
F892 - C969	4.68	0.99	3.92
K1096 - E1104	4.42	0.99	8.17
I819 - V837	4.32	0.99	11.21
V762 - L811	4.10	0.98	7.37
K1005 - E1039	3.93	0.98	5.84
K675 - H682	3.90	0.98	3.99
L751 - M965	3.80	0.98	3.78

Puntaje de acoplamiento entre las posiciones.

Probabilidad de que las posiciones de aminoácidos que estén acopladas.

Distancia atómica mínima (pares de átomos más cercanos en angstrom [Å]) en la estructura tridimensional.

- Coevolución inter-proteína

Dominio I y II de SBDS con dominio III y IV de EFL1

Tabla suplementaria 3. Predicciones de residuos relacionados evolutivamente entre las familias de SBDS y EFL1. Dominios I y II de SBDS con dominio III y IV de EFL1.

Residuo/posición en SBDS	Residuo/posición en EFL1	Probabilidad
V120	L659	0.99908836
T113	S868	0.98924632
I83	W897	0.98223344
M107	C891	0.98215651
R135	T661	0.97566452
R126	C883	0.97505931
V120	T697	0.93640399
E134	R970	0.92509448
I10	T661	0.92452555
N121	I649	0.91764748
T5	W848	0.91668356
V93	C883	0.91373468

Dominio II de SBDS con dominio IV de EFL1

Tabla suplementaria 4. Predicciones de residuos relacionados evolutivamente entre las familias de SBDS y EFL1. Dominios II de SBDS con dominio IV de EFL1.

Residuo/posición en SBDS	Residuo/posición en EFL1	Probabilidad
T129	E730	0.94981464
T129	L1003	0.94538656
K148	S736	0.9119614
D97	A1093	0.90020419

Dominio II de SBDS con dominio III y IV de EFL1

Tabla suplementaria 5. Predicciones de residuos relacionados evolutivamente entre las familias de SBDS y EFL1. Dominios II de SBDS con dominio III y IV de EFL1.

Residuo/posición en SBDS	Residuo/posición en EFL1	Probabilidad
V120	L659	0.97407817
M107	C891	0.96285854
T124	I689	0.94270464
R135	T661	0.93064205

Dominios II y III de SBDS (95-250) con dominio II de EFL1 (361-612)

Tabla suplementaria 6. Predicciones de residuos relacionados evolutivamente entre las familias de SBDS y EFL1. Dominios II y III de SBDS con dominio II de EFL1.

Residuo/posición en SBDS	Residuo/posición en EFL1	Probabilidad
K118	S595	0.99999614
L132	F413	0.9980778
E134	I522	0.99719798
T102	S595	0.99312276
H171	V592	0.99222283
I131	A596	0.98299004
L206	A507	0.97894698
R109	P578	0.95392546
L104	K528	0.90904471

Dominios II y III de SBDS con dominio III y IV de EFL1 (610-980)

Tabla suplementaria 7. Predicciones de residuos relacionados evolutivamente entre las familias de SBDS y EFL1. Dominios II y III de SBDS con dominios III y IV de EFL1.

Residuo/posición en SBDS	Residuo/posición en EFL1	Probabilidad
K231	L659	0.99250471
V120	L659	0.98936486
T129	E687	0.98851767
R126	C883	0.98641426
L174	I689	0.98481231
R135	T661	0.97449051
G232	L894	0.97423316
T113	S868	0.96333182
E123	P631	0.95851831
G245	E653	0.95222994
I197	E787	0.92729448
D97	R774	0.92508754
R218	S840	0.91287467
F217	I689	0.9026858

D246	E967	0.90206736
------	------	------------