

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**PROGRAMA DE MAESTRÍA Y
DOCTORADO EN CIENCIAS
BIOQUÍMICAS**

INSTITUTO DE BIOTECNOLOGÍA

**DEPARTAMENTO DE INGENIERÍA CELULAR Y
BIOCATÁLISIS**

**" SITIOS CON ACOPLAMIENTO ESTADÍSTICO EN LA FAMILIA
CHEY"**

**TESIS
QUE PARA OBTENER EL GRADO ACADÉMICO
DE
MAESTRO EN CIENCIAS**

**PRESENTA:
JOSÉ FERNANDO GARCÍA GUEVARA**

DIRECTOR DE TESIS: DR. LORENZO PATRICK SEGOVIA FORCELLA



CUERNAVACA, ABRIL 2012



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.

El presente trabajo se realizó bajo la dirección del Dr. Lorenzo Segovia en el Instituto de Biotecnología de la Universidad Nacional Autónoma de México.

AGRADECIMIENTOS

Al comité tutorial:

Dr. Enrique Merino, Dr. Georges Dreyfus, Dr. Lorenzo Segovia

A los miembros del Jurado:

Dra. Rosa Gutiérrez, Dra. Gloria Saab, Dr. Alejandro Sosa,

Dr. Baltazar Becerril, Dr. Guillermo Gosset

A mis compañeros de laboratorio

A la Universidad Nacional Autónoma de México

A mi familia y amigos

INDICE	1
RESUMEN	3
1 Introducción	4
2 Antecedentes	5
2.1 CHEY	5
2.2 PLEGAMIENTO ROSSMANN	6
2.3 EVOLUCIÓN: FUSIÓN DE GENES Y FUNCIONES DIVERGENTES	7
3 Hipótesis	9
4 Objetivos	9
5 Metodología	10
6 Resultados y discusión	12
6.1 ANÁLISIS ESTADÍSTICO DE ACOPLAMIENTO DE LA FAMILIA CHEY	12
6.1.1 Recolección de secuencias, eliminar redundancia y alineamiento ...	12
6.1.2 Correlaciones entre las secuencias	12
6.1.3 Conservación por posiciones y correlaciones entre posiciones.	14
6.1.4 Análisis espectral de la matriz SCA y definición de sectores	16
6.1.5 Identificación de sectores sin aproximación binomial.	20
6.2 SUBLINEAMIENTO CON 120 SITIOS E IDENTIFICACIÓN DE SECTORES	20
6.3 RELACIÓN ENTRE LOS SITIOS IDENTIFICADOS Y LA INFORMACIÓN EXPERIMENTAL DE ESTRUCTURA Y FUNCIÓN.	24
6.3.1 Sitios involucrados en vía de plegamiento.	24
6.3.2 Sitios involucrados en la unión al ión magnesio y la fosforilación.	25
6.3.3 Sitios involucrados en interacción con otras proteínas.	26
6.3.4 Mutaciones puntuales que afectan la función.	29
6.3.5 Sitios involucrados en la acetilación.	29

6.3.6 Sitios Involucrados en el mecanismo de activación.	30
6.4 SITIOS CORRELACIONADOS EN OTRAS FAMILIAS CON PLEGAMIENTO ROSSMANN	31
6.4.1 Alineamiento Estructural Entre CheY de <i>Escherichia coli</i> contra Spo0F de <i>E. coli</i> y Flavodoxina de <i>A. vinelandii</i>	33
6.5 CORRELACIONES UTILIZANDO ALINEAMIENTOS CON DIFERENTES COMPOSICIONES	35
6.6 CORRELACIONES ENTRE PAREJAS DE AMINOACIDOS EN TENSOR 4D SCA	38
7 Conclusiones	40
8 Perspectivas	41
8.1 LOS RESIDUOS M60 y M63 TIENEN VALORES DE CORRELACIÓN SIGNIFICATIVOS PERO HAN SIDO POCO ESTUDIADOS.	41
8.2 ANALISIS SCA EN ALINEAMIENTO CONCATENADO	41
8.3 DISEÑO DE SECUENCIAS CON FUCIÓN CHEY	42
8.4 INTERCAMBIO DE FUNCIONES ENTRE PROTEINAS CON TOPOLOGÍA SIMILAR	42
8.5 ANÁLISIS DE REDES, PROPIEDADES DE LAS REDES DE INTERACCIÓN ENTRE AMINOACIDOS CORRELACIONADOS	42
8.6 ESTABILIDAD DEL GRUPO FOSFATO	42
9 Bibliografía	43

RESUMEN

Las bacterias son capaces de moverse para acercarse a fuentes de alimento o se alejan de factores de estrés, a este proceso se le conoce como quimiotaxis. Dicho proceso se lleva a cabo por una maquinaria especializada formada por un conjunto de proteínas, una de estas es la proteína CheY. El papel de CheY en esta maquinaria consiste en transmitir la señal recibida por las proteínas receptoras hacia las proteínas que conforman el flagelo provocando un cambio en la dirección de su rotación.

Esta proteína tiene un plegamiento conocido como Rossmann, está presente en una gran diversidad de bacterias, incluso pueden existir varias proteínas homólogas a CheY dentro de un mismo organismo. Gracias a los avances recientes en las tecnologías de secuenciación en la actualidad se cuenta con una enorme cantidad de secuencias homólogas a CheY.

El Análisis Estadístico de Acoplamiento (SCA por sus siglas en inglés) es una técnica desarrollada recientemente que aprovecha la información contenida en un conjunto de secuencias homólogas para caracterizar las codependencias evolutivas en dicha familia de proteínas. A través de un alineamiento múltiple de secuencias homólogas es posible conocer cuáles son los residuos de aminoácidos que están evolucionando de manera conjunta, para ello se calcula una matriz de correlaciones entre las posibles parejas residuos de aminoácidos y posteriormente se buscan grupos de aminoácidos que tengan patrones semejantes de correlación. Cuando se observan dichos grupos en la estructura tridimensional de las proteínas estos forman redes conectadas físicamente y usualmente son residuos de gran importancia para la estructura y/o función de las proteínas.

Al aplicar esta técnica a distintos conjuntos de proteínas homólogas a CheY se encontraron dos grupos de aminoácidos, uno localizado mayormente en la región amino terminal de la proteína y que está involucrado en el proceso de plegamiento de la proteína y la unión a un ión de magnesio, y otro grupo en la región carboxilo terminal que está relacionado con la interacción física con las proteínas receptoras y efectoras en el proceso de quimiotaxis, mediante el cual se transmite la señal.

Además cuando se aplicó la técnica en otras familias con la misma topología pero con funciones y secuencias distintas se encontró que los sitios con correlaciones significativas se encontraban en regiones equivalentes en el plegamiento.

1 Introducción

La evolución molecular representa un laboratorio natural invaluable para probar y desarrollar nuestro entendimiento acerca de la relación entre la estructura, la dinámica y la función de las proteínas. De igual forma, nuestro entendimiento acerca del plegamiento y la función de las proteínas es clave para el análisis racional de las señales enviadas a través de la evolución de proteínas.

Estudios clásicos mostraron que la secuencia de aminoácidos contiene toda la información necesaria para determinar el plegamiento y la función de la proteína [1]. Cuando se realizan Alineamientos Múltiples con las Secuencias de proteínas homólogas (MSA por sus siglas en inglés), y éstos son lo suficientemente extensos y diversos, pueden proporcionar información acerca de sitios importantes para la estructura y la función de las proteínas [2].

En los alineamientos múltiples hay sitios con conservación particularmente alta, esto es resultado de algún tipo de presión selectiva, dichos sitios pueden determinar ciertas características para la estabilidad (posiciones de anclaje, señales de inicio o término de hélices alfa y láminas beta), el plegamiento (núcleo de plegamiento, cambios de conformación) o la función (sitio activo, sitio de reconocimiento), o bien, contribuyen a varias características a la vez [3] aunque el papel de los sitios identificados aún no es claro.

En principio la información contenida en los MSA puede ser mantenida por interacciones complejas entre posiciones de aminoácidos que no son necesariamente evidentes en modelos simples de conservación de secuencia que tratan cada sitio como si fueran estadísticamente independientes uno de otro. Para desarrollar un modelo físicamente más consistente con la conservación de secuencia, que tome en cuenta las interacciones cooperativas de los residuos, varios grupos han desarrollado nuevos algoritmos para capturar la evolución correlacionada de residuos a partir de MSA que muestran gran consistencia con datos experimentales [4].

CheY es una proteína citoplasmática bacteriana que posee un plegamiento Rossmann. Esta proteína ha sido estudiada experimentalmente con gran detalle y funciona como un regulador de respuesta activado por fosforilación que controla la quimiotaxis bacteriana. La estructura tridimensional de rayos X y de Resonancia Magnética Nuclear se encuentran disponibles [5-14]. También se ha hecho la caracterización termodinámica y cinética del estado de transición [15]. El análisis de la familia CheY haciendo uso de los algoritmos mencionados, en conjunto con información de plegamiento, estudios bioquímicos de unión y comunicación molecular son fundamentales para entender las relaciones entre la estructura y función.

2 Antecedentes

2.1 CHEY

CheY es una proteína implicada en la quimiotaxis, es una proteína relativamente pequeña. En *Escherichia coli* posee 129 aminoácidos y tiene un peso molecular de 14000 Daltons. CheY tiene un plegamiento Rossmann y es miembro de una gran familia de proteínas involucrada en la transducción de señales tanto en procariotas como en eucariotas, conocidos como reguladores de respuesta (RR) [16].

En la quimiotaxis una señal, producida por un estímulo químico, se transmite al flagelo de la bacteria a través de una cascada de señales. El estímulo es captado por un receptor que activa a CheA, que es una histidina quinasa. CheA fosforila a CheY en un residuo de aspartato conservado, la presencia de magnesio (Mg^{2+}) es necesaria para la transferencia del grupo fosfato [17]. El movimiento rotatorio del flagelo es por “default” contrario a las manecillas del reloj, y por la interacción de la proteína citoplasmática CheY con el interruptor FliM en la base del motor flagelar hay un cambio en la dirección de la rotación[18]. CheY en su forma no fosforilada puede interactuar con el interruptor flagelar y causar movimiento en dirección de las manecillas del reloj [19], pero su actividad se incrementa por lo menos dos órdenes de magnitud por la fosforilación [20].

CheY ha servido como proteína modelo para comprender la activación inducida por fosforilación de los RR [21-23]. Estudios genéticos, bioquímicos y estructurales en CheY han indicado que la fosforilación induce un cambio en la conformación, de la forma inactiva hacia la forma activa. Se cuenta con la estructura de CheY en su conformación activa [11] Asimismo, se ha reportado la estructura cristalográfica de CheY unida al dominio de reconocimiento de su quinasa, CheA [24] y de CheY unida a los dominios de reconocimiento de otras dos proteínas con las que interactúa. En un mismo organismo pueden haber varios reguladores de respuesta que participan en diferentes vías de señalización, a pesar de que el mecanismo de activación es común a todos los RR, el entrecruzamiento de la señal entre proteínas de diferentes vías es minimizado mediante el uso de dominios específicos de reconocimiento [25], [26]. En *E. coli* la proteína CheY es modificada por acetilaciones en 6 residuos de lisina localizados en la parte carboxilo terminal, el papel que desempeñan está relacionado con la unión a las proteínas con las que interactúa [27].

La ruta de plegamiento de CheY en *E. coli* ha sido estudiada [28] y ésta involucra un intermediario cinético. Con base en los resultados experimentales y la estructura de CheY, la proteína puede ser dividida en dos subdominios, que van de la hebra beta 1 al principio de la hebra beta 3 y de la hebra beta 3 hasta el carboxilo terminal (figura 1a). En el primer subdominio el empacamiento de la primera hélice alfa contra las hebras beta 1 y 2 parece ser el núcleo a

través del cual el resto de la proteína se pliega, el segundo subdominio es compacto pero no está estructurado en el estado de transición. CheY parece plegarse mediante un mecanismo de nucleación/condensación tal como se ha encontrado en otras proteínas con plegamiento distinto [15].

2.2 PLEGAMIENTO ROSSMANN

Como se mencionó al principio del apartado anterior CheY tiene plegamiento Rossmann, el cual es un motivo en la estructura de proteínas que comprende dos unidades beta-alfa-beta. La unidad beta-alfa-beta es una unidad muy importante de estructura súper-secundaria, con sus hebras paralelas y unidas por puentes de hidrógeno forma la base estructural de muchas proteínas. Con frecuencia varias unidades se combinan para formar una lámina beta de hebras paralelas.

El plegamiento Rossmann crea un dominio característico beta-alfa-beta-alfa-beta y frecuentemente contiene un núcleo hidrofóbico entre la lámina beta y las hélices alfa. El plegamiento Rossmann ocurre en una gran variedad de proteínas pero fue descrito primero en una familia de deshidrogenasas con el dominio de unión a NAD(P) (figura 1b) [29].

Este plegamiento crea una cavidad adyacente al enlace que une las dos hebras no contiguas de la lámina beta. El asa larga formada entre dichas hebras tiende a formar un hueco para formar un sitio de unión. Para el caso particular de las deshidrogenasas la cavidad formada es el sitio de unión a NAD, en tanto que los grupos NH de la primera hélice alfa se encargan de estabilizar el grupo fosfato cuando el cofactor es NADP [29].

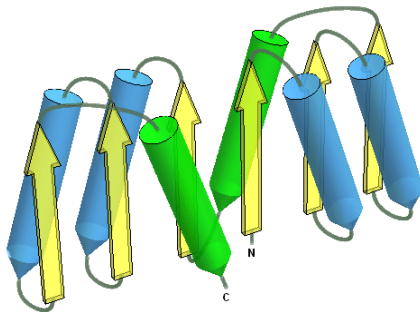
Otras proteínas con plegamiento Rossmann, diferentes a las deshidrogenasas, unen cofactores en forma similar a la unión entre NAD y las deshidrogenasas. Estas proteínas crean sitios de unión con estructura secundaria y terciaria similar. Como ejemplo se puede mencionar al FMN (Flavin Mono-Nucleótido) en la Flavodoxina, y al FAD (Flavin Adenin Dinucleótido) en la enzima piruvato oxidasa [29].

Existen ciertas variaciones en la topología del plegamiento Rossmann. Por ejemplo en la Flavodoxina cinco hebras forman la lámina beta principal, en lugar de seis que es el caso en el dominio de unión a NAD(P) (figura1c).

a)



b)



c)

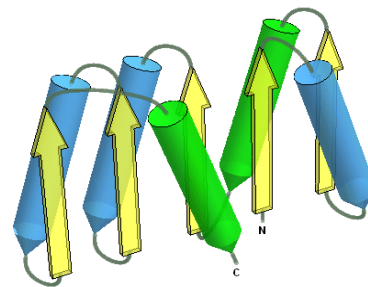


Figura 1. a) Subdominios de la proteína CheY; en verde primer subdominio y en rojo segundo subdominio. b) Plegamiento Rossmann con seis hebras paralelas en la lámina beta principal, presente en los dominios de unión a NAD(P). c) Plegamiento Rossmann con cinco hebras paralelas en la lámina beta, las Flavodoxinas y la CheY son ejemplos.

CheY, de forma similar a las Flavodoxinas, contiene cinco hebras en su lámina beta, aunque su secuencia y funciones no están relacionadas con las Flavodoxinas. Un alineamiento estructural de 184 átomos de la cadena principal de CheY de *E. coli* con la apoFlavodoxina de *A. vinelandii* tiene una desviación (r.m.s.d.) de sólo 1.62 Å, a pesar de que presenta sólo un 4 % de identidad de secuencia y 20% de similitud [30].

2.3 EVOLUCIÓN: FUNCIONES DIVERGENTES

En contraste a muchos plegamientos proteicos, el plegamiento Rossmann es compartido por un gran número de superfamilias de proteínas. Según la clasificación de CATH [31] el plegamiento Rossmann contiene 120 superfamilias que exhiben poca o ninguna similitud en la secuencia de aminoácidos y abarcan un amplio rango de proteínas no relacionadas con diferentes funciones

como unión a cofactores, catalasas, proteínas quimiotácticas, lipasas, estererasas, Flavodoxinas, entre otras.

Dada la gran versatilidad funcional que existe dentro de las superfamilias es útil considerar las posibles rutas hacia nuevas funciones. Las nuevas funciones evolucionan con frecuencia a partir de una combinación de mecanismos, en especial a través de fusión de genes, duplicación de genes, acumulación de mutaciones y oligomerización [32].

La fusión de genes o quimerismo es un fenómeno que involucra la combinación de múltiples secuencias ancestrales en una única proteína multi-dominio mediante la evolución. La duplicación de dominios ha sido extensiva a través de la evolución, esto es ilustrado en las familias de proteínas con plegamiento Rossmann. Como ejemplo se tiene al dominio CheY fusionado a diversos dominios de unión a DNA en los Reguladores de Respuesta (RR) de los sistemas de señalización de dos componentes [33].

Existen proteínas con funciones bioquímicas muy diferentes pero que, a pesar de ello, tienen estructuras tridimensionales muy similares. Dichos casos indican que la estructura diverge mucho más lentamente que la función durante la evolución. Otros tres ejemplos, además de CheY, de proteínas con diferentes funciones pero que comparten plegamiento Rossmann, y que además se cuenta con su estructura cristalográfica, son Spo0F de *Bacillus subtilis*, Flavodoxina de *Azotobacter vinelandii* y el dominio de unión a NADP de la shikimato deshidrogenasa de *E. coli*.

Bacillus subtilis responde a la privación de nutrientes y a la alta densidad celular expresando un grupo de genes que transforman a la bacteria en una espora. La vía de señalización involucra varias quinasas, reguladores de respuesta, fosfatasas y fosfotransferasas que son responsables de integrar las señales externas para los genes de esporulación se activen solo cuando las condiciones ambientales lo requieran. Spo0F juega un papel central en la integración de dichas señales, Spo0F pertenece a la gran clase de proteínas reguladores de respuesta, los cuales participan en diversas vías de transducción de señales en bacterias, los reguladores de respuesta poseen un dominio Rossmann de alrededor de 120 residuos que es fosforilado en una reacción dependiente de magnesio.

Las Flavodoxina de *Azotobacter vinelandii* tiene 179 residuos y es una proteína para transferencia de electrones, las Flavodoxinas contienen flavin mono nucleótido (FMN) y actúan como donadores de electrones para un gran número de enzimas rédox. La topología de esta proteína es similar a la de los reguladores de respuesta, con cinco hebras beta en la lámina central.

La shikimato deshidrogenasa (SDH) AroE participa en la vía de síntesis de aminoácidos aromáticos y cataliza la reacción reversible dependiente de NADPH de 3-dehidroshikimato a shikimato. La molécula contiene 2 dominios, uno de ellos es el catalítico y el otro es el dominio de unión a NADP. En nuestro grupo se está explorando la generación de quimeras de proteínas homólogas de la shikimato deshidrogenasa donde se intercambien los dominios Rossmann de esta proteína. El análisis realizado con SCA — un método estadístico que detecta los residuos que covarían en alineamientos múltiples de familias de proteínas homólogas [34] — en la familia SDH detectó sitios con acoplamiento estadístico. Al comparar los sitios con acoplamiento del dominio Rossmann de la SDH con datos experimentales se encontró que están estrechamente relacionados con la actividad catalítica y la unión al cofactor.

Es posible que al contrastar los sitios conservados y los sitios con acoplamiento estadístico entre familias con función y secuencia tan diferentes aunque con plegamientos similares se puedan revelar algunas pistas para comprender mejor los principios que dictan la estructura y la función en este plegamiento tan versátil y evolutivamente exitoso.

3 Hipótesis

Mediante el Análisis Estadístico de Acoplamiento en alineamientos de secuencias homólogas a CheY es posible encontrar sitios que han coevolucionado para mantener la estructura y función.

4 Objetivos

Realizar un análisis de la secuencia para identificar las posiciones con acoplamiento estadístico en la familia CheY e identificar posiciones relevantes para la estructura o la función en familias con plegamiento Rossmann.

Objetivos particulares:

- Realizar el Análisis Estadístico de Acoplamiento:
 - Recolección de secuencias homólogas a CheY, eliminar redundancia en las secuencias y realizar un alineamiento múltiple.
 - Calcular las correlaciones entre las secuencias del alineamiento
 - Medir la conservación por cada posición y el acoplamiento entre posiciones, encontrar los sitios con valores significativos de correlación.
 - Análisis espectral de la matriz SCA y definición de sectores, formar grupos de sitios que posean patrones de correlación semejantes.
- Buscar relaciones entre los sitios identificados y la información experimental disponible:
 - Identificar los sitios involucrados en vía de plegamiento; en la unión al ión magnesio y la fosforilación; en interacción con otras proteínas; mutaciones

puntuales que afectan la función; sitios involucrados en la acetilación; sitios involucrados en el mecanismo de activación.

- Identificar los sitios correlacionados en otras familias con plegamiento Rossmann y comparar la localización de los sitios con acoplamiento entre CheY y estas familias.
- Observar lo que ocurre cuando se calculan correlaciones utilizando alineamientos con diferentes composiciones.
- Identificar las correlaciones ente parejas de aminoácidos en la matriz de cuatro dimensiones de correlaciones.

5 Metodología

-Recolección de secuencias homólogas a CheY disponibles en bases de datos.

Con la secuencia de CheY de *E coli* se realizó una búsqueda de BLASTp [35] contra la base de datos no redundante de proteínas, se modificaron los parámetros e-valor en $1e^{-3}$, el número máximo de secuencias objetivo de 20 000 y la matriz de pesos, Blosum 65. El mismo procedimiento se realizó para las otras familias de proteínas con plegamiento Rossmann, usando como pregunta la secuencia de Spo0F de *Bacillus subtilis*, Flavodoxina de *Azotobacter vinelandii* y el dominio de unión a dinucleótido de AroE de *E. coli*. Para recolectar secuencias de un solo dominio homólogas a CheY se buscó en la base de datos de NCBI las secuencias que en su anotación contaran con la palabra CheY, de este conjunto se tomaron las que tuvieran una longitud entre 100 y 140 aminoácidos. Para recolectar secuencias con evidencia de contexto genómico se tomaron las secuencias anotadas como CheY en la base de datos Genecont [36] que tienen alguna otra proteína involucrada en el proceso de quimiotaxis cercana en el genoma del organismo.

-Eliminar redundancia.

Con las secuencias recolectadas y alineadas se generaron bases de datos con las secuencias representativas del alineamiento original usando el programa CD-hit [37]. Para hacer esto se utilizaron diferentes porcentaje de identidad (40%, 50%, 60%, 70%, 80% y 90%) como punto de corte.

-Alineamiento múltiple.

Debido a la gran cantidad de secuencias los alineamientos no se pueden realizar con programas convencionales, por ello para construir los alineamientos se utilizan como referencia los resultados de alineamientos pareados de la búsqueda hecha con BLAST. Cuando era posible debido a que el número de secuencias a alinear era pequeño (menor que 500) se utilizaron herramientas convencionales de alineamiento múltiple (MAFFT, CLUSTAL).

Al final se trabajó con 6 alineamientos distintos, tres de ellos con secuencias homólogas a CheY y los otros tres con secuencias homólogas a proteínas con topologías similares a CheY (Spo0F, Flavodovina y AroE). El primero de los alineamientos está formado por homólogos a CheY encontrados con la herramienta BLASTp y con un punto de corte del 70% de identidad en el paso para eliminar la redundancia (llamado chy70). El segundo alineamiento con secuencias homólogas a CheY posee secuencias con una longitud entre 110 y 140 aminoácidos, y tiene punto de corte del 40% (llamado monochy). En el último de los alineamiento con homólogos de CheY las secuencias que lo conforman poseen evidencia de contexto genómico de participar en la quimiotaxis y tiene punto de corte del 90% (chygenecont). Los alineamientos con homólogos a Spo0F de *Bacillus subtilis*, Flavodoxina de *Azotobacter vinelandii* y al dominio Rossmann de AroE de *E. coli*, tienen puntos de corte de porcentaje de identidad del 70%, 80% y 70%, respectivamente.

-Análisis de la diversidad del alineamiento.

Un alineamiento es diverso si las secuencias que lo componen son diferentes entre sí, es decir hay poca redundancia. Usando Matlab y el paquete de funciones SCA40, se realizó un Análisis de componentes principales de la matriz de identidad entre parejas de secuencias con el propósito de observar la diversidad del alineamiento.

-Análisis Estadístico de Acoplamiento (SCA)

Utilizando el algoritmo reportado en la versión de SCA 4.0 [4] [34] se calculó la matriz SCA de correlaciones para cada alineamiento. El principio básico para definir la matriz de correlaciones es ponderar las correlaciones entre parejas de sitios por una función de las conservación de los sitios.

-Análisis espectral de la matriz SCA de correlaciones. Identificación de sectores y sitios con correlaciones significativas en familias con plegamiento Rossmann.

Usando el paquete de funciones SCA 4.0 se realizó el análisis espectral, para poder distinguir las correlaciones significativas de las correlaciones generadas al azar debido al tamaño finito del alineamiento. Y también separar las correlaciones con significado funcional de las correlaciones con significado histórico [4]. El análisis espectral consiste en comparar los eigenvectores y eigenvalores de la matriz de correlaciones obtenido del alineamiento original contra el promedio de eigenvectores y eigenvalores de alineamientos obtenidos a partir del muestreo sin reemplazo de cada columna del alineamiento original.

-Análisis de clustering de la matriz de correlaciones limpia (donde se eliminan las correlaciones que son indistinguibles de las generadas por el azar) y reducida a los sitios con correlaciones

significativas (se consideran únicamente los sitios que tienen valores de correlación mayores a un punto de corte).

Utilizando el programa Mev4 se realizó un análisis de clustering jerárquico utilizando correlación de Pearson sobre las matrices de correlación obtenidas del análisis de correlación.

-Interpretación de datos obtenidos mediante: la comparación con datos experimentales disponibles en la literatura, el uso de herramientas bioinformáticas disponibles.

6 Resultados y discusión

6.1 ANÁLISIS ESTADÍSTICO DE ACOPLAMIENTO DE LA FAMILIA CHEY

6.1.1 Recolección de secuencias, eliminar redundancia y alineamiento.

Con la búsqueda de BLAST usando como consulta la secuencia de CheY de *E. coli* se obtuvieron 20,000 secuencias, que es el número máximo de secuencias que se pueden obtener en el servidor de NCBI, el *e-valor* más pequeño de las secuencias obtenidas fue de 1×10^{-6} .

Usando, CD-Hit que es un programa que genera una bases de datos sólo con las secuencias representativas removiendo secuencias redundantes, se obtuvo un conjunto de secuencias con identidad entre ellas inferior al 70%. Posteriormente se alinearon todas las secuencias de este conjunto para obtener un alineamiento múltiple para calcular las correlaciones entre los sitios de éste, a este alineamiento se le nombrará de aquí en adelante Chy70, este alineamiento tiene 129 posiciones y cuenta con 7451 secuencias.

Una estrategia útil para calcular las correlaciones es quitar del alineamiento los sitios con una frecuencia de gaps mayor a un punto de corte determinado, en este caso fue del 20%, como resultado se obtiene un alineamiento con menos posiciones que el inicial, en este paso se eliminaron 23 posiciones quedando ahora sólo 106. En este caso la mayor parte de los sitios con alta frecuencia de gaps se encuentra en los extremos amino y carboxilo terminal, en consecuencia no será posible analizar las correlaciones de los sitios eliminados en este paso, los sitios eliminados fueron del 1 al 7, el 31,76,77 y del 117 al 129, la numeración es respecto a la secuencia de CheY de *E. coli*.

6.1.2 Correlaciones entre las secuencias

Con el alineamiento Chy70 se construyó una matriz de identidad entre parejas de secuencias con el objetivo de determinar el grado en que el alineamiento representa una muestra representativa del espacio de secuencias, idealmente no debería existir un grupo de secuencias correlacionadas entre sí, más bien todas ellas deberían ser igualmente distintas.

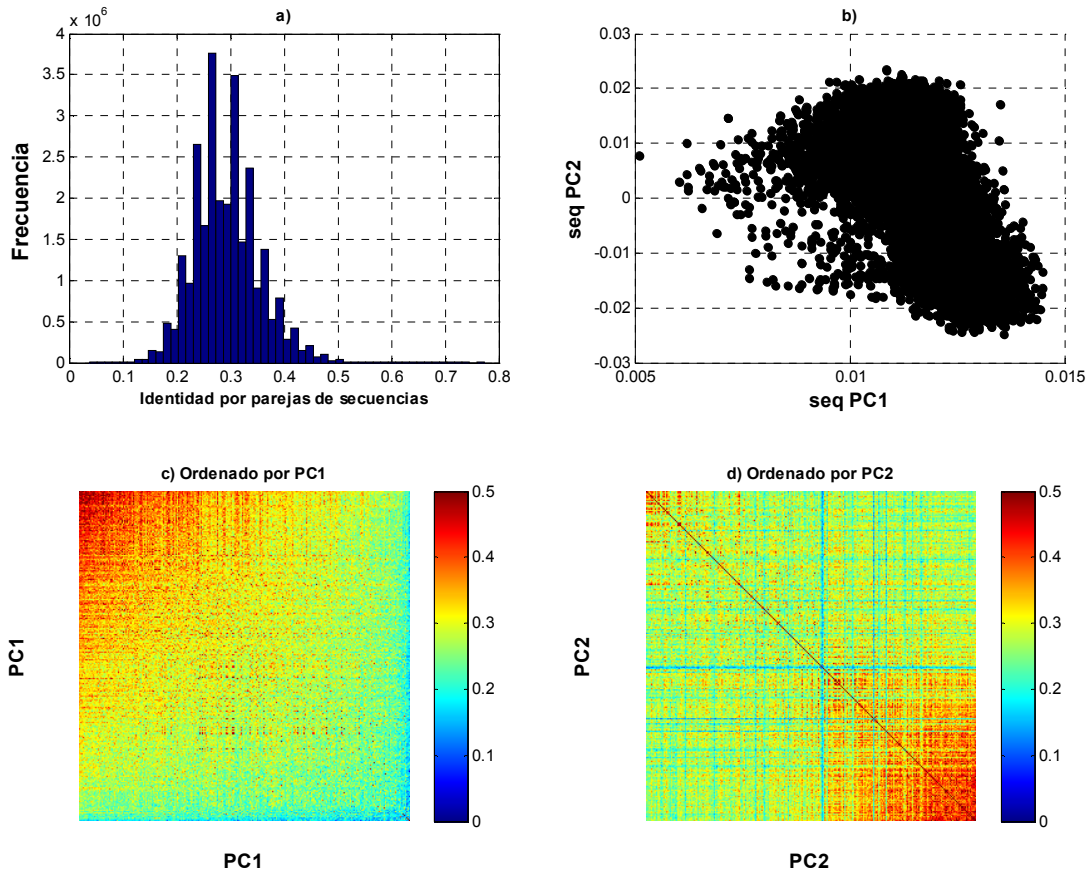


Figura 2. Correlaciones entre las secuencias. a) Histograma de la frecuencia de identidad por pares en la secuencia. b) Grafica de dispersión de los dos primeros componentes principales de la matriz de identidad por pares de secuencia. c) Matriz de identidad ordenada por el primer componente principal. d) Matriz de identidad ordenada por el segundo componente principal.

El histograma de arriba (figura 2a) presenta una distribución poco dispersa con una media alrededor del 30% y con valores que van del 10 al 50% de identidad, esto sugiere que la mayoría de las secuencias son igualmente distintas entre sí.

Si el alineamiento tiende a una divergencia homogénea de las secuencias entonces éstas deberían aparecer uniformemente distribuidas en el espacio abarcado por los primeros componentes principales. Al observar los diagramas dispersión de los dos primeros componentes principales (figura 2b) parece ser el caso de que las secuencias están distribuidas de manera uniforme ya que no se observan grupos claros. La matriz de identidad de secuencias ordenada por el primer componente tiene una estructura jerárquica, la misma matriz ordenada por el segundo componente muestra aparentemente la formación de un grupo (figura 2c y 2d).

Con el objetivo de construir un alineamiento más homogéneo se intentó construir uno con secuencias más parecidas entre sí. Para ello en la búsqueda inicial de secuencias homólogas con BLASTp el valor de expectancia se disminuyó a $1e-10$ y a $1e-15$, en intentos independientes. Sin embargo la cantidad inicial de secuencias fue muy baja (menos de 100 secuencias) y la distribución en los dos primeros componentes principales presentó grupos, así que se decidió trabajar con este primer alineamiento.

6.1.3 Conservación por posiciones y correlaciones entre posiciones.

En el Análisis de Acoplamiento Estadístico el grado de conservación de cada posición se mide con una cantidad llamada entropía relativa. En esencia esta medida captura la divergencia entre las frecuencias observadas en cada posición contra la frecuencia de fondo en la base de datos no redundante de proteínas [4]. En la figura 3 se muestran los valores de conservación en cada sitio del alineamiento, se pueden distinguir tres zonas de conservación al principio, al medio y al final del alineamiento, separadas entre sí por zonas menos conservadas.

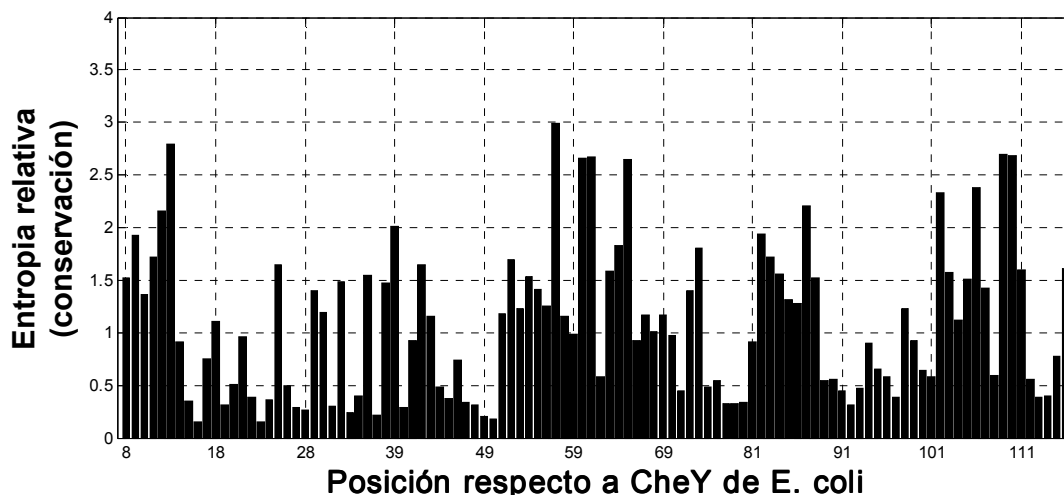


Figura 3. Grafica de barras indicando la Entropía relativa o conservación por sitio en el alineamiento. La numeración va de acuerdo a la secuencia CheY de *E. coli*.

La matriz de correlaciones en su forma completa es un tensor de 4 dimensiones (posición X posición X aminoácido X aminoácido) donde cada elemento del tensor es un valor de correlación para un par de aminoácidos entre una pareja de sitios, para identificar los grupos de sitios correlacionados se requiere colapsar el tensor de cuatro dimensiones en un tensor de sólo dos dimensiones (posición X posición). Existen dos estrategias para lograr la reducción de dimensiones, la primera es calcular la raíz del valor cuadrado medio entre todos los pares de aminoácidos para un par de sitios produciendo la magnitud total de la correlación entre dos

sitios. La segunda estrategia para reducir las dimensiones viene de la observación de que con frecuencia la conservación global en un sitio se aproxima bien a una representación binaria del alineamiento donde sólo se considera el aminoácido más frecuente en cada posición del alineamiento (figura 4a). La validez de la aproximación es evidente cuando se observa la similitud entre la matriz SCA de correlación calculada con cualquiera de las dos estrategias (figura 4b).

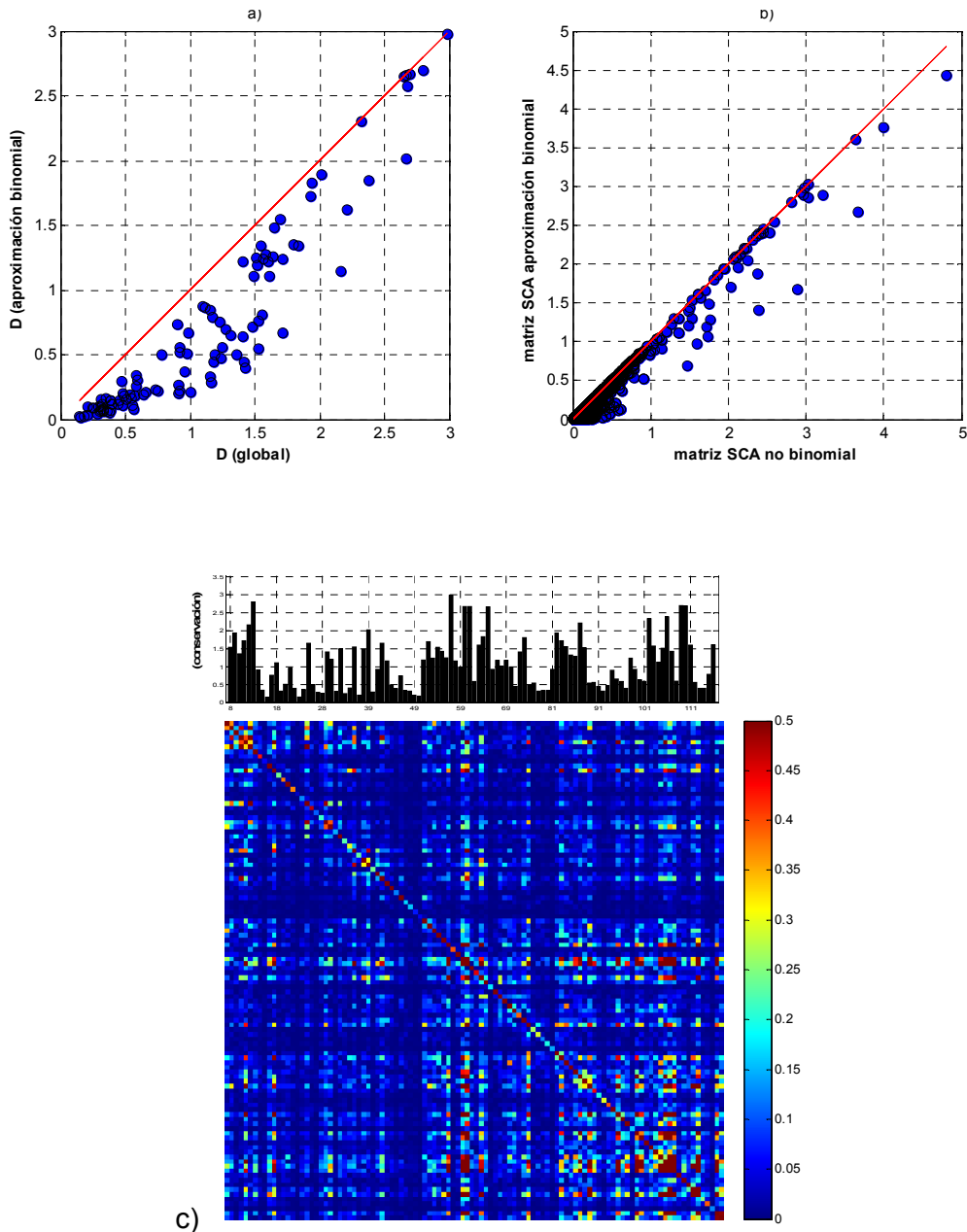


Figura 4. a) Diagrama de dispersión de los valores de la conservación global calculada con el alineamiento completo contra los valores de conservación calculados con la aproximación binomial del alineamiento. b) Diagrama de dispersión de los valores de la matriz SCA de correlación, aproximación binaria contra no binaria. c) Matriz SCA de correlaciones, el orden de la matriz es igual que el orden del alineamiento.

En la matriz SCA hay posiciones que muestran correlaciones débiles con la mayor parte de las otras posiciones y correlaciones fuertes con otras pocas posiciones. También hay algunas posiciones que muestran correlaciones débiles con todos los demás sitios, éstos sitios corresponden a las zonas donde hay menor conservación (figura 4c).

6.1.4 Análisis espectral de la matriz SCA y definición de sectores

El análisis espectral de una matriz de correlación consiste en comparar los valores de dicha matriz de correlación obtenida con datos experimentales contra el promedio de los valores de matrices de correlación obtenidos con los datos aleatorizados, con el fin de distinguir cuáles son los valores que no se explican por el azar. El primer paso para realizar el análisis espectral consiste en hacer un análisis de componente principales de la matriz. Posteriormente se repite el análisis de la matriz SCA para alineamientos obtenidos después de revolver aleatoria e independientemente los aminoácidos de las columnas del alineamiento. Esta manipulación conserva las frecuencias de aminoácidos en cada posición pero remueve las correlaciones entre parejas de posiciones. De este modo se definen los componentes que pueden ser explicados simplemente por limitaciones en la muestra debido al tamaño finito del alineamiento.

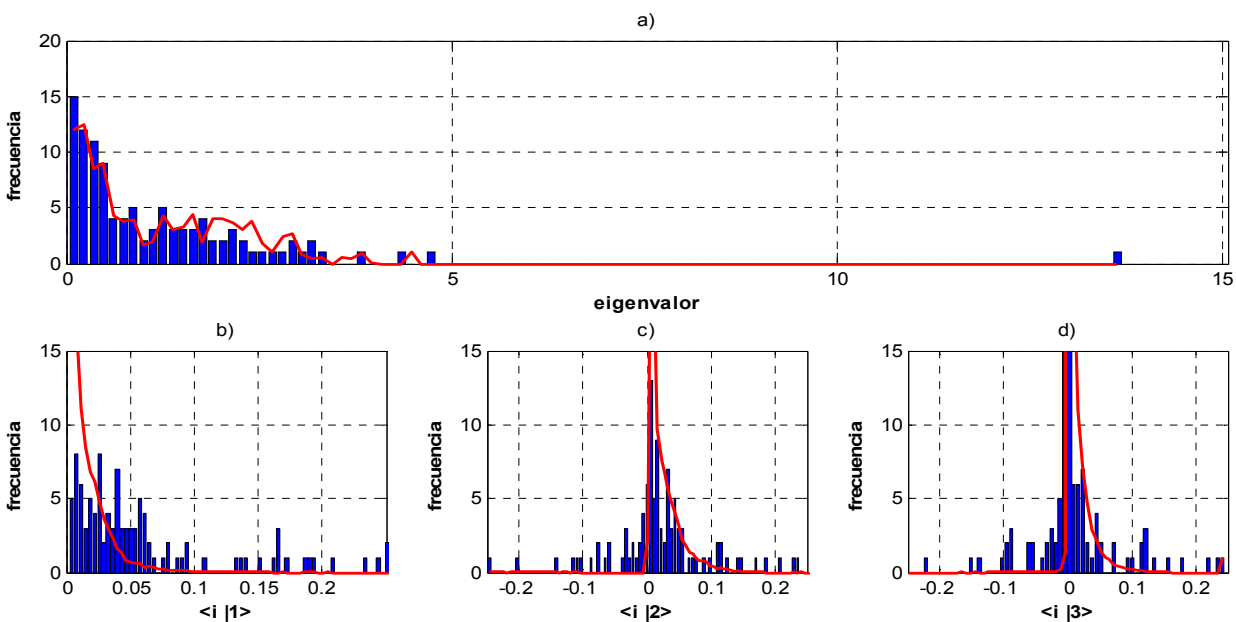


Figura 5. Eigenmodos de la matriz SCA. 1) Comparación entre el espectro de eigenvalores del alineamiento real (barras azules) y los ensayos al azar (línea roja). b), c) y d) comparación de las contribuciones del primer, segundo y tercer componente del alineamiento real (barras azules) contra las contribuciones del primer, segundo y tercer componente de los ensayos al azar (línea roja).

Basados en la figura anterior los eigenvalores de los tres primeros componentes no son explicables por limitaciones en el tamaño del alineamiento (figura 5a), por ello éstos componentes se eligen para identificar los grupos de sitios correlacionados o sectores.

Los sectores están definidos por las contribuciones de cada sitio a los primeros componentes. Para determinar los puntos de corte se comparan los pesos de los componentes principales obtenidos del alineamiento real con los pesos de los componentes principales de los ensayos al azar (figura 5b, 5c y 5d). Basados en estos resultados los puntos de corte del segundo y tercer componente serían -0.05 a la izquierda y 0.1 a la derecha. Siguiendo la definición anterior el primer sector (sector azul) está compuesto por los sitios con valores inferiores a -0.05 y el segundo (sector rojo), por las posiciones con valores superiores a 0.1 a través del segundo componente (figura 6a).

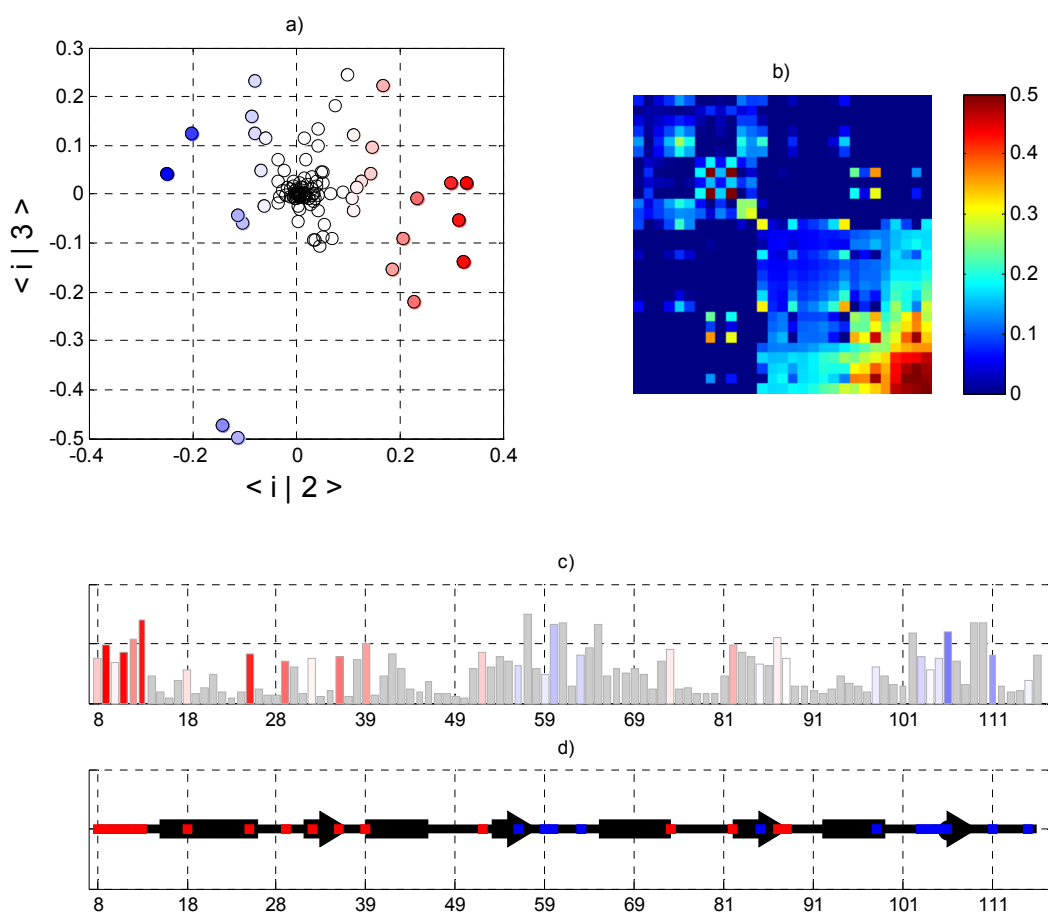


Figura 6. Sectores, en azul se muestra el primer sector y en rojo se muestra el segundo sector a) Gráfica de dispersión de la contribución de cada sitio en el segundo y tercer componente. b) Matriz limpia de SCA, sólo se muestran las posiciones identificadas en los sectores, reteniendo únicamente las contribuciones de cada posición al segundo y tercer componente, ordenada de acuerdo a la contribución de cada sitio en los sectores. c) Localización de los sectores en la gráfica de conservación de secuencia y d) en la estructura secundaria.

Sector	Sitios
sector azul	56, 59, 60, 63, 85, 98, 103, 104, 105, 106, 111, 115
sector rojo	8, 9, 10, 11, 12, 13, 18, 25, 29, 33, 36, 39, 52, 73, 82, 87, 88
Conservados	13, 57, 60, 61, 65, 109, 110

Tabla 1. Sitios incluidos en los sectores de acuerdo a la numeración de la secuencia CheY de *E. coli*. Identificados a partir del alineamiento chy70. Y sitios con conservación mayor al 90%.

Como está descrito por Halabi, Rivoire, Leibler, & Ranganathan, 2009, en el caso de secuencias homogéneamente distribuidas, el primer componente principal representa la correlación global entre todos los sitios de las secuencias, esta característica se explica simplemente por las relaciones filogenéticas entre las secuencias, por ello este vector no se considera para la identificación de sectores. El "ruido histórico" se elimina de la matriz de correlación cuando este primer componente es removido. En la matriz limpia de SCA se pueden apreciar los dos sectores, y que además son casi independientes uno de otro.

A pesar de que los sectores no están relacionados de manera evidente con la conservación de los sitios cuando se ilustran en la estructura primaria se observa que la mayor parte del sector azul está localizada en la región carboxilo terminal, mientras que el sector rojo se encuentra principalmente en la zona amino terminal. El sector rojo incluye casi completamente la primera hebra beta (figura 6d).

Al observar los sectores en la estructura tridimensional de CheY se aprecia que los sitios de cada sector están conectados físicamente. El sector azul (figura 7a) se localiza sólo en la región carboxilo terminal de la proteína, en esta parte de la proteína se encuentra también el sitio de reconocimiento de las proteínas con las que interactúa. El sector rojo (figura 7b) está ubicado en la región amino terminal de la proteína, en esta región también se encuentra la mayor parte de los residuos identificados como núcleo de plegamiento [15]. Ambos sectores rodean el residuo de ácido aspártico conservado que se fosforila para activar a la proteína.

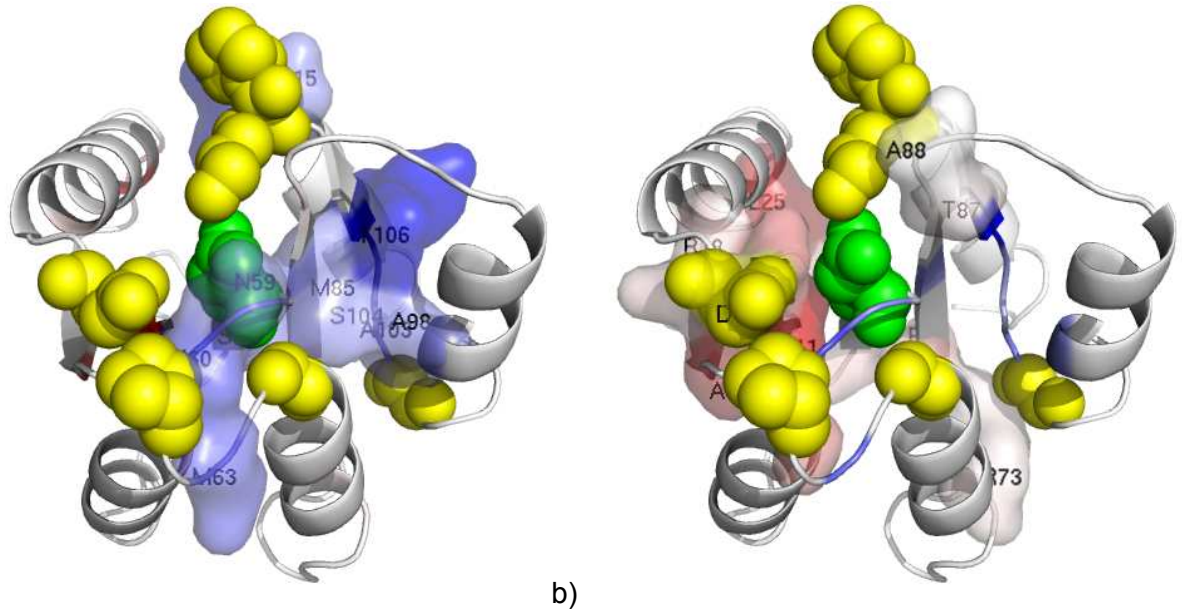


Figura 7. Sectores en la estructura tridimensional. En esferas amarillas se muestran los residuos con una conservación mayor al 90%, en esferas verdes se muestra el residuo aspártico 57. a) Superficie del sector azul. b) Superficie del sector rojo. (La intensidad del color es proporcional a la contribución de ese sitio sobre el segundo componente)

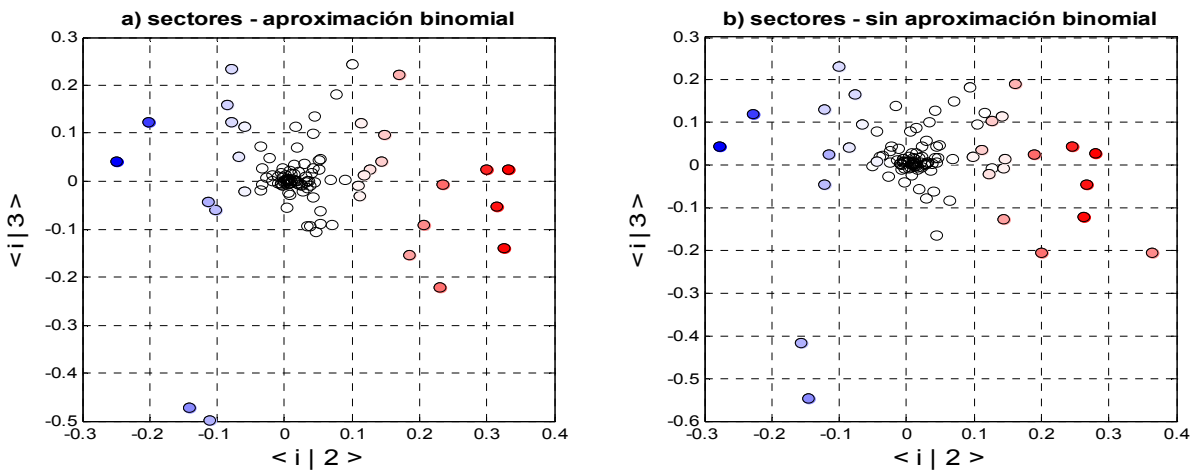


Figura 8. Identificación de sectores, se muestra el diagrama de dispersión del segundo y tercer componente de la matriz de correlaciones SCA calculada con a) con aproximación binomial y b) sin aproximación binomial. En rojo y azul se muestran los sitios que conforman los sectores identificados utilizando la aproximación binomial.

6.1.5 Identificación de sectores sin aproximación binomial.

Como se mencionó la matriz SCA también se puede calcular utilizando la información para todos los aminoácidos por cada par de posiciones. La composición de los sectores no debería depender mucho de la forma en que se calcula la matriz SCA de correlación. Para mostrar esto se identifican los sectores con y sin la aproximación binomial. En este caso la identificación de sectores es consistente en usando cualquiera de las estrategias.

6.2 SUBALINEAMIENTO CON 120 SITIOS E IDENTIFICACIÓN DE SECTORES

Una estrategia práctica para el Análisis de Acoplamiento es eliminar los sitios con una fracción de gaps mayor al 20%. En el alineamiento utilizado arriba el número de posiciones total fue de 109 después de realizar este corte, mientras que el número total sitios en la secuencia de aminoácidos de CheY es de 129. La mayor parte de los sitios no incluidos en el análisis se localizan en la región amino terminal, carboxilo terminal y en algunas asas. Con el fin de incluir más posiciones en el análisis, del alineamiento anterior se eligieron sólo las secuencias con una longitud mayor o igual a 120 aminoácidos. El sub-alineamiento generado fue utilizado para identificar de nuevo sitios con acoplamiento, siguiendo el mismo procedimiento que en el caso anterior. El nuevo alineamiento (Chy70I120) ahora tiene 1109 secuencias y 120 posiciones. En la tabla siguiente se muestran las posiciones identificadas con correlaciones significativas

Sector	Sitios
Sector azul	52, 59, 69, 82, 86, 88, 94, 98, 104, 106, 111, 116, 119
Sector rojo	11, 25, 29, 36, 38, 39, 41, 43, 60, 63, 64, 67, 72, 73, 81, 84, 85, 107

Tabla 2. Posiciones en sectores identificados en subalineamiento chy70I120.

Nuevamente uno de los sectores se localiza en la región carboxilo terminal y el otro en la amino terminal (Figura 9a). Se puede observar que una parte de los sitios identificados aquí son iguales que para el alineamiento anterior (59, 98, 104, 106, 111 del sector azul; 11, 25, 29, 36, 39, 73 del sector rojo), sin embargo algunos sitios cambiaron de sector (60, 63, 85 del azul al rojo; 52, 82, 88 del rojo al azul), además hay sitios que ahora no se identifican en los nuevos sectores (56, 103, 105, 115 antes estaban en el sector azul; 8, 9, 10, 12, 13, 18, 33, 87 antes estaban en el sector rojo) y sitios que ahora sí se identifican y en el otro alineamiento aparecían (52, 69, 82, 86, 88, 94, 116, 119 ahora están en el sector azul; 38, 41, 43, 64, 67, 72, 81, 84, 107 ahora están en el sector rojo).

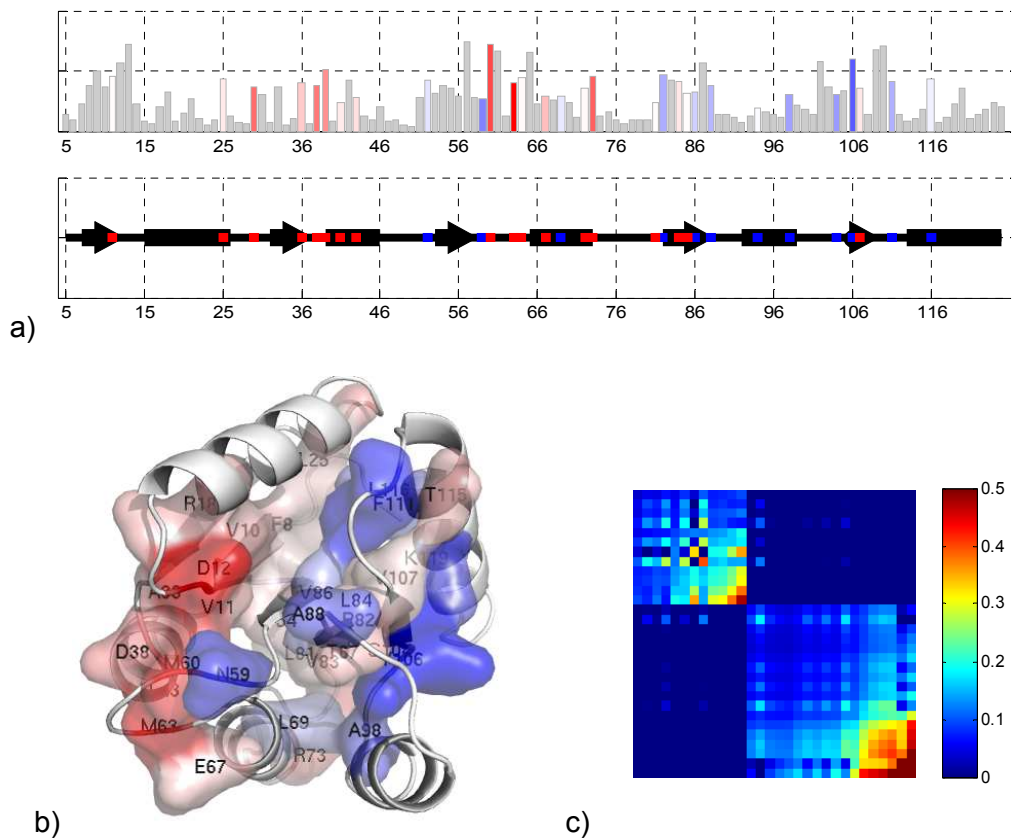


Figura 9. Sectores identificados en el sub-alineamiento Chy701120, a) sitios mapeados en estructura primaria y secundaria, b) sitios mapeados en estructura tridimensional c) Matriz limpia y reducida SCA

Las diferencias observadas se deben a la distinta composición de secuencias de los alineamientos utilizados y es difícil saber si los valores de correlación para dichos sitios tienen implicaciones funcionales, históricas o son correlaciones debidas únicamente al azar ya que la mayoría de dichos sitios se encuentran cerca de los límites del nivel de detección del método utilizado, abajo del cual no es posible distinguir con confianza correlaciones verdaderas de las correlaciones creadas únicamente por el azar. Este punto se trata de analizar más adelante.

Al aplicar el Análisis de Componentes Principales sobre la matriz SCA de arriba y realizar el análisis espectral se obtuvo que solo hasta el tercer componente principal no era explicado por el azar. De estos tres componentes el primero se eliminó para eliminar con él el ruido evolutivo. Esto nos dejó con 2 componentes sin embargo los sitios incluidos en sectores fueron definidos únicamente tomando en cuenta sus valores sobre el segundo componente principal por lo cual los sitios con valores significativos sobre el tercer componente principal no fueron considerados.

Con el fin de incluir dichos sitios se calculó la magnitud de cada sitio tomando en cuenta sus valores de correlación sobre el segundo y tercer componente principal y se eligieron únicamente

los sitios con una magnitud resultante mayor a un punto de corte, lo que también se puede interpretar como trazar un círculo alrededor del origen del diagrama de dispersión del segundo y tercer componente principal con un radio igual al punto de corte establecido (Figura 10b), los sitios que se encuentren por afuera del círculo son considerados como significativos.

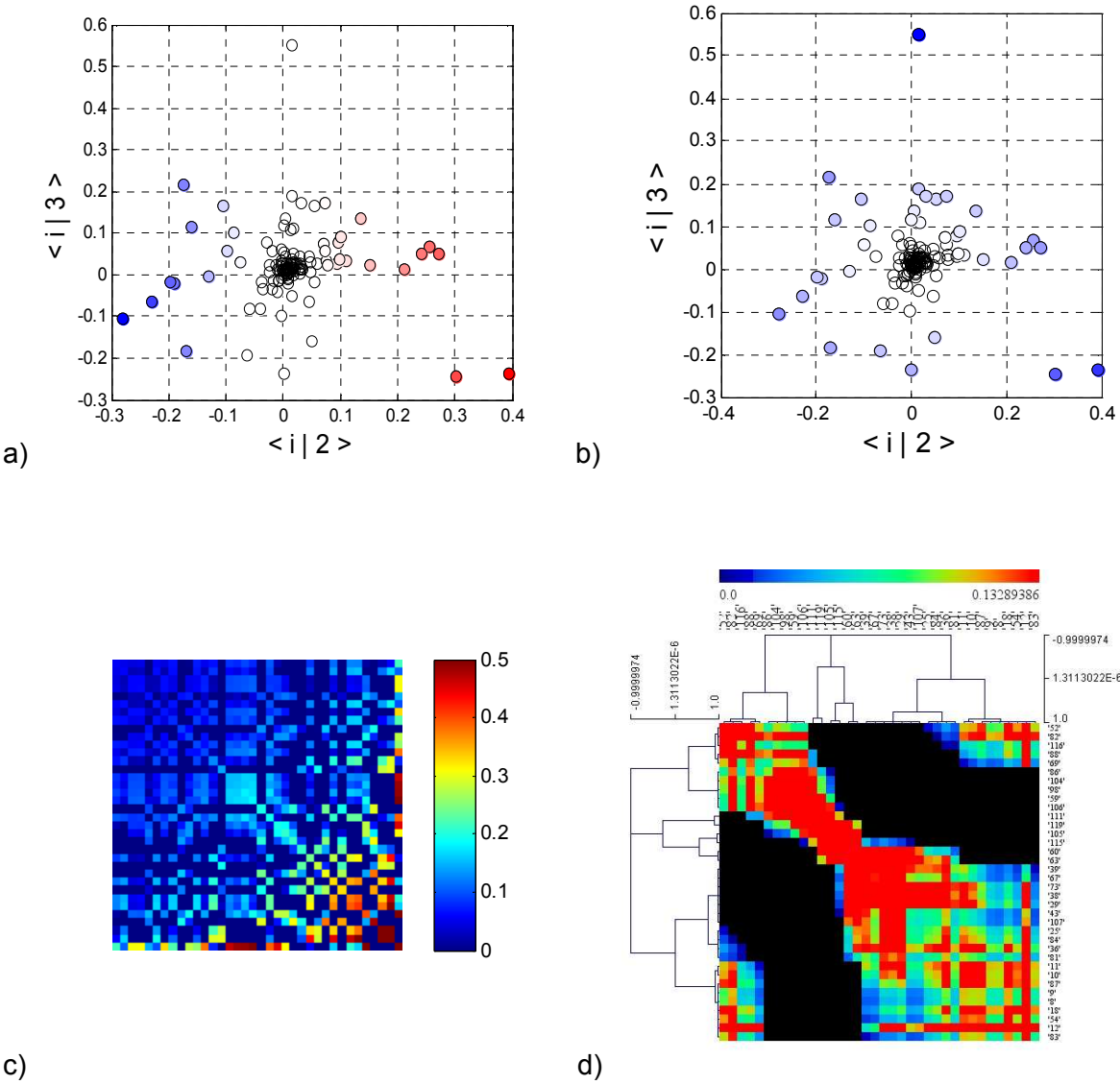


Figura 10. Componentes principales significativos definición de sectores a) usando valor en segundo componente principal b) trazando un círculo para considerar valor de sitios en ambos componentes principales. c) Matriz SCA del círculo y d) matriz sca agrupada con clustering jerárquico.

Con los sitios elegidos se construyó la matriz SCA libre de ruido evolutivo y estadístico (Figura 10c). Posteriormente la matriz reducida y limpia se analizó con métodos de clustering jerárquico, lo que permite encontrar conjuntos de sitios con patrones de correlación semejantes (Figura 10d). En el análisis de clustering se identificaron dos grupos principales, que corresponden a los sectores definidos con el alineamiento Chy70, en la tabla siguiente se enlistan los sitios de los dos grupos, azul y rojo. Hay un subconjunto de sitios en cada grupo que también tienen valores significativos de correlación con el otro grupo, estos se enlistan también y se representan en morado (figura 11 y tabla 3). Al observar la estructura de tres dimensiones de la proteína CheY con los sectores representados se observa que los aminoácidos del grupo rojo se encuentran desplazados hacia un lado, los del grupo azul hacia el lado opuesto y los de morado se encuentran en medio de ambos

Grupo	Sitios
Azul	52, 59, 60, 63, 69, 82, 86, 88, 98, 104, 105, 106, 111, 115, 116, 119
Rojo	8, 9, 10, 11, 12, 18, 25, 29, 36, 38, 39, 43, 54, 67, 73, 81, 83, 84, 87, 107
Morado	8, 9, 10, 11, 12, 18, 52, 54, 60, 63, 69, 81, 82, 83, 84, 87, 88

Tabla 3. Sitios que conforman los grupos identificados, combinando técnicas de clustering y de componentes principales.

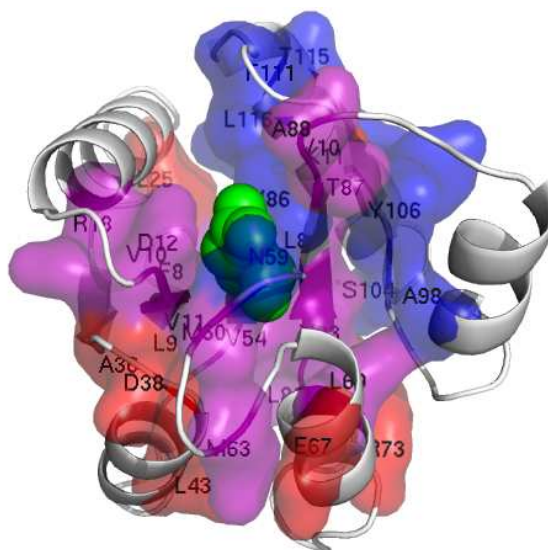


Figura 11. Estructura terciaria de CheY y de los grupos identificados combinando métodos de clustering y componentes principales. Se muestra la superficie de cada grupo, en esferas verdes se muestra el residuo de aspártico 57.

6.3 RELACIÓN ENTRE LOS SITIOS IDENTIFICADOS Y LA INFORMACIÓN EXPERIMENTAL DE ESTRUCTURA Y FUNCIÓN.

Existe gran cantidad de información experimental disponible en la literatura, relacionada con la función y estructura de CheY. Por ejemplo, podemos encontrar mutantes puntuales que producen una proteína constitutivamente activa o hacen que pierda la función, y otras que desestabilizan la estructura. En seguida se compara esta información con la información obtenida del análisis de acoplamiento estadístico para establecer relaciones entre los resultados de ambas estrategias, las comparaciones se realizan utilizando los sectores rojo y azul identificados con el alineamiento Chy70 a menos que se indique algo diferente.

6.3.1 Sitios involucrados en vía de plegamiento.

Para estudiar la ruta de plegamiento en CheY se realizó un estudio de valor Phi [15]. El análisis de valor Phi es un método experimental de ingeniería de proteínas usado para estudiar la estructura del estado de transición en la vía de plegamiento en proteínas pequeñas que se pliegan en una forma de dos estados. Típicamente una alta fracción de los residuos en una proteína son mutados uno por uno para identificar grupos de residuos que están bien ordenados en el estado de transición. Un valor phi de uno indica que en el estado de transición dicho sitio está bien ordenado, mientras que un valor phi cercano a cero indica que dicho aminoácido no se encuentra estructurado en el estado de transición.

Las siguientes mutaciones puntuales tiene un valor Phi menor a 0.3: F14A, P61G, A74G, A99G, A101G, A113G, A114G de estas mutantes ninguna de ellas se encuentra en los sitios con valores significativos de correlación. Las mutantes puntuales con valores mayores a 0.3 y menores que 0.8 son: D12A, D13A, D57A. Dos de estos sitios se localizan en el sector rojo (D12 y D13) mientras que el otro sitio se encuentra son una conservación mayor al 95% en el alineamiento chy70.

Las mutantes, que usan como fondo F14N, con valor de Phi menor o igual a 0.3 son G39A, V40T, I55V, D64A, I72V, G76A, V83T, A97G, A98G, A103G, V108T, T112A, T112G, I123V; de éstos solamente G39 se encuentra en el sector rojo, mientras que A98 y A103 se localizan en el sector azul. En el mismo estudio, las mutantes (que usan como fondo F14N) con valores phi mayores que 0.3 son V10T, V11T, V33T, A36G, D38G, A42G, V54T. De estas mutantes los sitios D38, A42, V54 no están en ningún grupo de sitios correlacionados, todos las demás están en el grupo rojo.

El grupo de aminoácidos con valores Phi mayores que 0.3 se localiza en la región amino terminal de la proteína y se propone como núcleo de plegamiento de CheY, alrededor del cual

se pliega la proteína y coincide ampliamente con la localización de los sitios correlacionados que pertenecen al sector rojo identificado en el alineamiento Chy70. En el sector rojo existe un elevado número de residuos que está implicado en el plegamiento de esta familia de proteínas, un poco menos de la mitad de los sitios del sector forma parte del núcleo de plegamiento de la proteína (figura 12a).

6.3.2 Sitios involucrados en la unión al ión magnesio y la fosforilación

Estudios de resonancia magnética nuclear han mostrado desplazamientos químicos después de la unión al ión magnesio o después de la fosforilación en varias zonas de la proteína [6], [38]. Estas zonas coinciden con las zonas donde se observan cambios en la orientación cuando se observan las estructuras cristalográficas [11], [39].

Los residuos que cambian su orientación en la conformación sin magnesio respecto a la conformación inactiva con magnesio [39], [40] son D12, D13, F14 y D57 del sitio activo; los residuos que van de T87 hasta E92 que forman el asa beta 4 alfa 4 y el inicio del hélice alfa 4; K109 y P110 del asa beta 5 alfa 5; W58, N59 y 106. Los residuos D12, D13, D57 y N59 se encuentran coordinando al ión magnesio en CheY. Se ha mostrado que K109 rompe su puente salino con D57 cuando el magnesio se une a los residuos que lo coordinan y en su lugar K109 forma un puente de hidrógeno con D12. En la forma sin magnesio apo-CheY, el oxígeno del grupo carbonilo de A88 está apuntando hacia el solvente, en tanto que la forma con el magnesio unido apunta hacia la dirección contraria formando un puente de hidrógeno con el grupo amida de K91. De los residuos mencionados D12, D13, T87 y A88 están presentes en el sector rojo, D57, K109 y P110 están altamente conservados, N59 está en el sector azul (figura 12b).

Cuando se superponen las estructuras de CheY unida al ión magnesio sobre la estructura activa con BeF-3(trifluorido de berilio, un análogo del ión fosfato) de CheY [10] los cambios más grandes observados se localizan en el asa beta 4 alfa 4 y el asa beta 5 alfa 5. El loop beta 4 alfa 4 adopta diferentes configuraciones en las estructuras inactivas de CheY. El cambio de la configuración inactiva a la activa involucra la formación de un puente de hidrógeno entre el grupo hidroxilo de T87 y el grupo BeF-3 en D57. La hebra beta 4 se desplaza y el anillo aromático de Y106 se entierra en la cavidad hidrofóbica entre la hélice alfa 4 y la hebra beta 5, formando un enlace de hidrógeno entre su grupo hidroxilo y el grupo carbonilo de la cadena lateral de E89. Finalmente K109 forma un puente salino con BeF-3, lo que sugiere que juega un papel importante en la nueva orientación del asa beta 5 alfa 5. Nuevamente T87 del sector rojo participa en los cambios de conformación después de la fosforilación. Y106 de la hebra beta 5 y F111 del asa beta5-alfa5 son los sitios encontrados del sector azul. Es notorio que a pesar de haber cambios significativos en la estructura del asa beta 4 alfa 4 no se haya detectado ningún sitio de ellos en alguno de los sectores.

6.3.3 Sitios involucrados en interacción con otras proteínas.

En el proceso de quimiotaxis CheY de *E. coli* interactúa físicamente con tres proteínas, CheA, FliM y CheZ, en tres etapas diferentes del evento. CheA es una histidina quinasa que al recibir la señal de los receptores en la capa exterior de la célula fosforila a CheY. Después de la fosforilación la afinidad de CheY por FliM aumenta [41], FliM es el interruptor del motor flagelar [42] y cuando interactúa con CheY provoca un cambio en la dirección de la rotación del flagelo. Por último, CheZ es una fosfatasa con afinidad a CheY-P [43], que cuando actúa regresa a CheY a su estado inactivo.

Análisis mutagénicos [44] indican que los residuos N23, K26, E27, F111, T112 y E117, localizados en la hélice alfa 1 y 5 participan en la unión a CheZ, de estos residuos únicamente F111 está incluido en el grupo azul, en el mismo trabajo se encontró que mutaciones en los residuos A88, A90 y E93 causan deficiencias en la quimiotaxis pero sin alterar la afinidad por CheZ, lo que sugiere que estos residuos participan en la actividad catalítica de CheZ y no en la afinidad. En un estudio posterior donde se utilizó Resonancia Magnética Nuclear (NMR) para encontrar los sitios con un elevado desplazamiento químico en CheY durante la interacción con el péptido carboxilo terminal de CheZ [45] K119 fue el único residuo de la hélice alfa 5 que mostró un desplazamiento químico elevado. Los residuos fuera de la hélice alfa 5 que también mostraron un desplazamiento químico elevado son Y106, V107, T87, A99, V86, M85, K91, A98, S104, V108, N94, I96, D57 y N59. De estos residuos V107, A99, V86, K91, V108 N94, I96 no están incluidos en ningún sector, el resto está incluido en el sector azul (figura 13a), menos T87 (D57 no se considera porque está altamente conservado y K119 tampoco debido a que no está incluido en el alineamiento inicial). CheY activa es más afín a CheZ que cuando no está activa [46], cuando T87 se reorienta el asa beta 4 alfa 4 ofrece contactos adicionales donde más residuos del extremo amino de CheZ pueden interactuar. Esto apunta a que el sector azul está involucrado en el reconocimiento, la afinidad y la actividad de CheZ.

La estructura cristalográfica de CheY activada con BeF-3, en complejo con la región N-terminal de la proteína FliM [11] revela que los primeros siete residuos del péptido de FliM se empaquetan en forma extendida en contra del asa beta 4 alfa 4 y la hélice 4 de CheY, mientras que los residuos restantes forman dos vueltas de hélice y se empaquetan contra la cara alfa 4 beta 5 alfa 5, esta zona de interacción en CheY es la única que sufre cambios notables en seguida de la activación. Esta zona también es justo donde se localiza el sector azul que además conecta dicha superficie con el sitio de fosforilación. En un estudio de resonancia magnética nuclear [45], los residuos que mostraron un desplazamiento químico elevado después de la adición del péptido FliM fueron V107, K109, Y106, A99, M85, V86, D57, L116, V108, N121, T87, E118, Y21,

K91 y K126. Estos residuos están localizados en la hélice alfa 4 y 5 y en las hebras beta 4 y 5 aunque sólo los residuos 87, 85 y 106 están en alguno de los dos sectores (figura 13b). Los residuos que se ven altamente afectados por la adición de cualquiera de los péptidos de FliM o CheZ son D57, M85, V86, T87, K91, A99, Y106, V107 y V108, nuevamente encontramos en sectores a los sitios 87, 85 y 106.

Otros estudios de NMR [38], [47] enfocado en encontrar los sitios con desplazamiento químico elevado al agregar el péptido de CheA con el que interactúa CheY mostraron que dichos sitios son 85, 88, 94, 96, 103, 105, 107, 122, 123 y 126, de los cuales únicamente los residuos 85, 103, 105 pertenecen al sector azul y el residuo 88 al sector rojo. En el mismo trabajo [38] se menciona que la mutación A103V causa defectos en la quimiotaxis debido a que CheA se une con menor afinidad (figura 13c).

Solo algunos residuos incluidos en el sector azul y T87 del sector rojo participan en la interacción con las proteínas CheA, CheZ y FliM, se sabe que los reguladores de respuesta son un mecanismo recurrente en bacterias para la transmisión de señales, puede haber varios reguladores de respuesta expresados al mismo tiempo por ello requieren que no haya entrecruzamiento en la señal. Debido a que la mayoría de los sitios del sector azul se encuentran muy cercanos a los residuos en la superficie de interacción y forman una conexión física entre dicha superficie y el sitio activo es muy probable que dicho sector esté involucrado en la transmisión de la señal una vez reconocidas las proteínas y los sitios encargados del reconocimiento específico lo formen otro conjunto independiente de residuos.

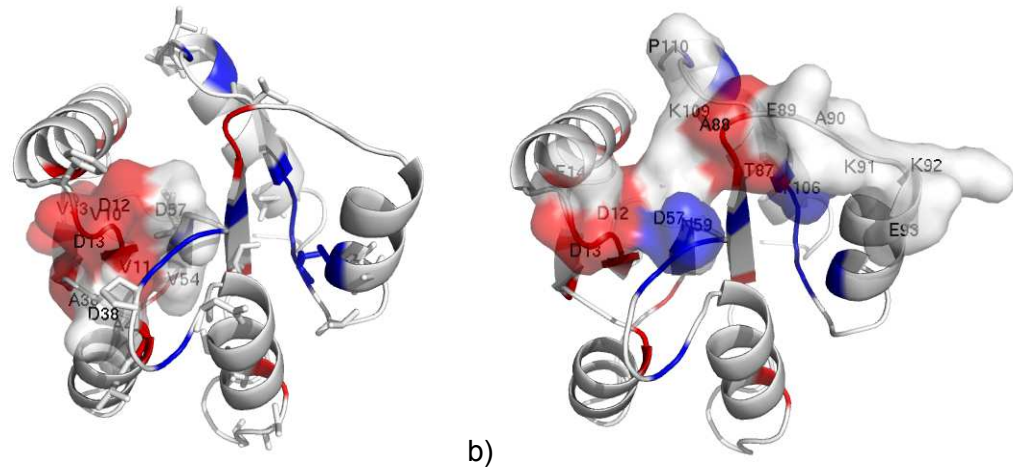


Figura 12. Sitios involucrados en a) el núcleo de plegamiento (se muestra la superficie de residuos con valor phi superior a 0.3) b) la fosforilación y unión al ión magnesio (se muestra la superficie de los sitios que tienen contactos con el fosfato o magnesio, o que cambian de conformación después de la fosforilación o la unión al ión magnesio). Los sitios del sector azul y rojo se marcan con sus respectivos colores..

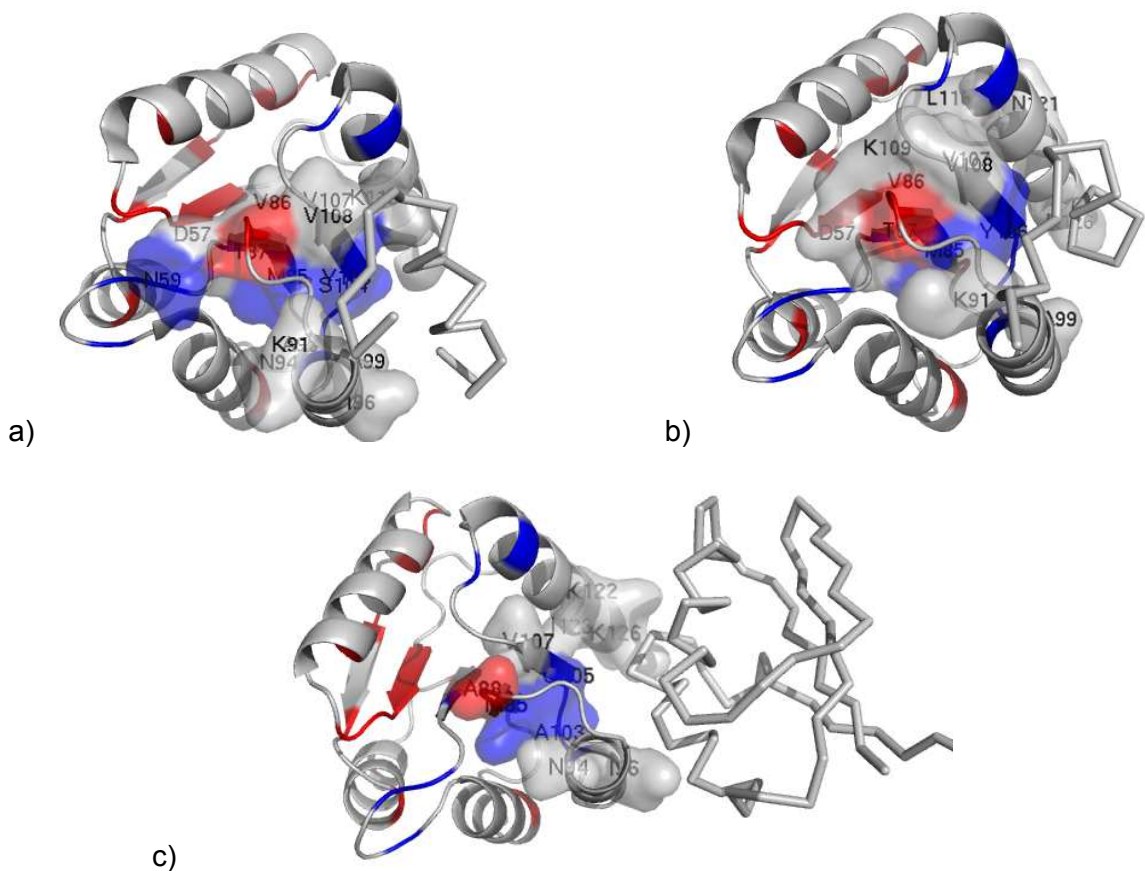


Figura 13. Estructura de CheY unida a a) el péptido carboxilo terminal de CheZ, b) el péptido amino terminal de FlIM y c) el dominio de unión HPT de CheA. En color azul y rojo se muestran los dos sectores identificados. Se muestra también la superficie de los residuos con alto desplazamiento químico después de la adición del respectivo péptido en estudios de RMN.

6.3.4 Mutaciones puntuales que afectan la función.

La mutante puntual T87I y la doble mutante T87I/Y106W en CheY son fosforilables; sin embargo son incapaces de generar el cambio en la dirección de la rotación del flagelo, ambas mutantes provocan una disminución en la afinidad de CheY fosforilada por CheZ y FliM [12]. La mutación T87I que causa que Y106 adopte la conformación hacia afuera, lo que provoca un fenotipo de nado liso y elimina la quimiotaxis. El fenotipo de nado liso implica que el interruptor FliM no esté recibiendo la señal por parte de CheY. Tanto T87 como Y106 forman parte de los sectores pero pertenecen a diferentes grupos, sin embargo existe correlación entre T87 y varios residuos del sector azul.

Mutaciones en I95 [48], [49] tienen efectos diversos sobre el cambio de dirección en la rotación del flagelo. I95 interactúa directamente con el interruptor FliM, cuando es mutado por alanina o valina aumenta su capacidad para causar el cambio en la dirección de la rotación del flagelo, pero cuando se muta por glicina o un residuo cargado o más grande se elimina la capacidad para causar el cambio en la dirección de la rotación, debido a que estas mutaciones modifican su capacidad para interactuar con FliM o CheZ. Especialmente, I95 se encuentra en contacto con el grupo azul aunque no está incluido en los sitios con correlaciones significativas y sin embargo tiene efectos considerables e inesperados en la función de CheY al ser mutado.

Las mutantes puntuales CheYD13R, CheYD13K y CheYD13K/Y106W [17], [50] son mutantes permanentemente activas, cuando se revisó su estructura tridimensional se observó una conformación similar a la estructura silvestre en su estado inactivo [8].

6.3.5 Sitios involucrados en la acetilación.

Uno de los procesos por los cuales CheY de *Escherichia coli* se puede activar para cambiar la dirección en la rotación del flagelo es mediante la acetilación, ya sea auto catalizada o mediante la acetil co-A sintetasa [51], [52], [53]. Se conocen 6 sitios de auto acetilación, éstos son: K91 y K92, que se encuentran en el asa beta alfa 4; K109, que tiene un porcentaje de conservación superior al 90%; K119 que se localiza en el grupo azul (grupo del alineamiento chy2000I120, tabla 3) y por último K122 y K125 que están alineados con K119 en el alfa hélice 5. Varios de estos sitios de acetilación están involucrados con la unión a FliM, CheZ y CheA, además se encontró que la unión a dichas proteínas se ve reprimida cuando hay acetilación [27]. Los sitios con mayor nivel de acetilación son K122 seguido de K92, ninguno de ellos está incluido en un grupo de sitios correlacionados. Probablemente este mecanismo de regulación no esté sujeto a una gran presión de selección, se haya adquirido recientemente o simplemente esos sitios no se

encuentran en el alineamiento, y no sea posible detectar los sitios involucrados con la diversidad de secuencias de los alineamientos disponibles.

Mutantes en K92 alteran las capacidades quimiotácticas de *E. coli*; cuando se muta por metionina reduce la tasa de movimiento al 66% de la del tipo silvestre, cuando se sustituye por glutamina la tasa de movimiento baja hasta el 42% respecto al tipo silvestre y cuando se sustituye por glutamato o aspartato no funciona la quimiotaxis [54]. Una vez más se encuentra un sitio que tiene efectos considerables en la función de CheY al ser mutado y sin embargo no fue detectado como significativo en el análisis de acoplamiento.

6.3.6 Sitios Involucrados en el mecanismo de activación.

La estructura de CheY después de la fosforilación sufre diversas modificaciones. El asa beta 4, alfa 4, que abarca de los residuos A88 hasta K91, tiene un movimiento considerable, ya que la desviación para la cadena principal es de 1.9Å y considerando las cadenas laterales la RMSD es de 4.9 Å respecto a la configuración inactiva del asa [10], [11], [22]. La cadena lateral de Y106 cambia de estar expuesta al solvente a estar enterrada en la cavidad formada bajo el asa beta 4 alfa 4; comúnmente se considera que la isomerización de Y106 modula la afinidad de CheY por FliM, aunque estudios mutacionales [55] interesantes muestran que Y106 no está involucrado en la unión a FliM, sin embargo es importante para su activación después de la unión. La cadena lateral de T87, K109 y la cadena principal de A88 forman puentes de hidrógeno con el grupo fosforilo D57.

En un estudio previo de dinámica molecular se pretendió explorar el acoplamiento entre el cambio conformacional y la fosforilación en CheY [56]. Los resultados apoyan un mecanismo de activación en el cual el asa beta 4 alfa 4 promueve el cambio en la conformación de Y106. Los puentes de hidrógeno del grupo fosforilo en D57 con T87 y la cadena principal de A88 estabilizan la configuración activa del asa beta 4 alfa 4. Se encontró también que los residuos A90, I95 y V108 forman un portal para la entrada de Y106. Cuando el asa se va desplazando hacia la configuración activa A90 se quita del camino, entonces los residuos I95 y V108 forman interacciones hidrofóbicas entre ellos, luego cuando el asa se desplaza más los residuos I95 y V108 dejan de interactuar para permitirle el paso a Y106, una vez que estos residuos permiten el paso de Y106 la transición ocurre rápidamente. El papel que juega K109 no fue evidente en las simulaciones pero se pudo observar que el puente de hidrógeno era más estable en la conformación activa y fosforilada, en los demás casos el puente se rompía con frecuencia, es posible que K109 participe en orientar al grupo fosfato en la conformación activa de la proteína, jugando un papel indirecto en el proceso de activación. Los únicos sitios que se encuentran en

sectores involucrados en el mecanismo de activación arriba descrito son T87y A88 del sector rojo y Y106 del sector azul.

6.4 SITIOS CORRELACIONADOS EN OTRAS FAMILIAS CON PLEGAMIENTO ROSSMANN

Hay familias de proteínas con plegamiento semejante al de CheY, como los dominios de unión a cofactor en las deshidrogenasas. A continuación se muestran los sitios correlacionados en estas familias con estructura similar pero con funciones distintas y se compara la localización de estos sitios respecto a la localización de los sitios correlacionados en la familia de CheY.

Se eligieron tres proteínas y sus respectivas familias para realizar el análisis de SCA. Estas son Spo0F de *Bacillus subtilis*, un regulador de respuesta que participa en el proceso de esporulación; Flavodoxina de *Azotobacter vinelandii*, proteína que participa como donador de electrones en un número considerable de enzimas rédox y utiliza Flavin Mono Nucleótido como cofactor; por último la Shikimato Deshidrogenasa AroE de *E. coli*, enzima involucrada en la vía metabólica de síntesis de aminoácidos aromáticos. Se cuenta con la estructura cristalográfica de las tres proteínas [57], [58] y [59].

Al analizar y comparar la localización de los sitios con valores elevados de correlación en la estructura tridimensional de estas tres proteínas (figura 14), se observa que la mayor parte de los sitios con valores elevados de correlación se encuentran rodeando la zona correspondiente a región carboxilo terminal de la primera hebra beta del dominio Rossmann. En la proteína Spo0F (figura 14a) esta zona está muy cercana al sitio de fosforilación, mientras que en la Flavodoxina (figura 14b) y el dominio Rossmann de AroE (figura 14c) esta zona es el sitio de unión a cofactor.

Otra zona donde también se comparte la localización de sitios con correlación significativa es en la cara formada por la hélice 4 y 5 y la hebra beta 5, en CheY. Dicha región corresponde al sitio de interacción con otras proteínas. En todos los casos este conjunto de sitios está conectado físicamente a los sitios que rodean los de unión a cofactor o fosforilación mencionados en el párrafo anterior. En CheY se propone que mediante estos sitios se propaga la señal, es muy probable que pase lo mismo con Spo0F y sería interesante saber cuál es el papel que desempeñan dichos residuos en la Flavodoxina y el dominio Rossmann de la Shikimato Deshidrogenasa

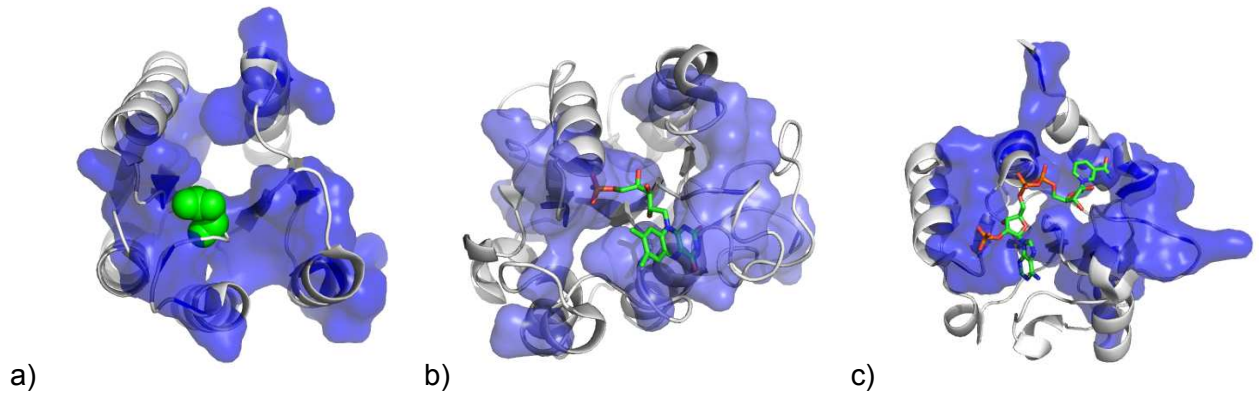


Figura 14. estructura terciaria mostrando los sitios con acoplamiento estadístico significativo. a) Spo0F, b) Flavodoxina y c) dominio Rossmann de SDH En color verde se muestra el residuo de aspártico conservado en Spo0F y el cofactor en el caso de la Flavodoxina y la SDH.

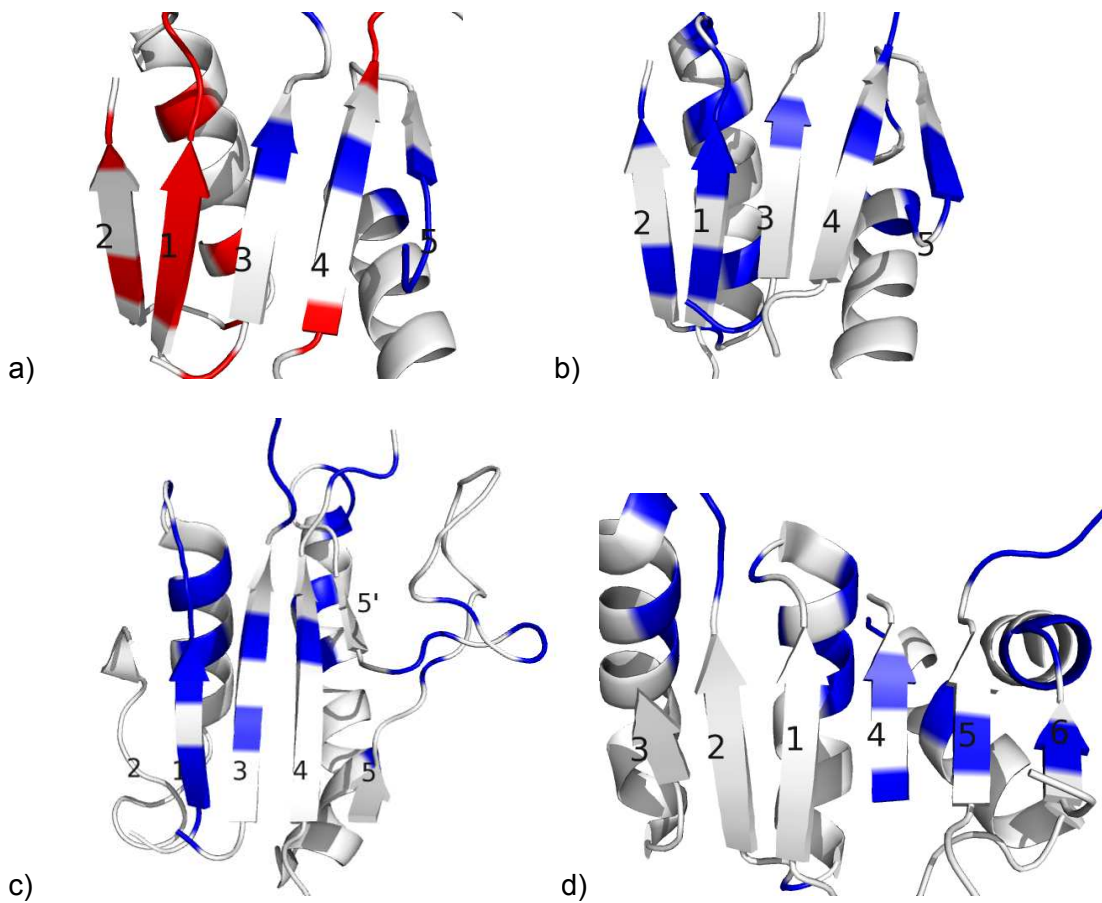


Figura 15. Sitios con correlación sobre la lámina beta del dominio Rossmann. a) lamina beta central de CheY, b) lámina beta central de Spo0F, c) lámina beta de Flavodoxina y d) lámina beta de dominio Rossmann de AroE.

6.4.1 Alineamiento Estructural Entre CheY de *E. coli* contra Spo0F de *E. coli* y Flavodoxina de *A. vinelandii*

Asimismo en las cuatro estructuras se observan correlaciones significativas sobre los sitios de la lámina beta central del plegamiento Rossmann y dichos sitios están conectados espacialmente. (figura 15), sin embargo la localización de dichos sitios no es similar para las cuatro familias. En el caso de los dos reguladores de respuesta se aprecia que algunos de los sitios de la lámina beta con acoplamiento están en la región que corresponde al núcleo de plegamiento en CheY en la hebra beta 1 y 2 (figura 15a y 15b), lo que sugiere que el núcleo de plegamiento está conservado entre estos reguladores de respuesta [60], así como sitios en la parte superior de las hebras beta 3 y 4 (que conectan el sitio activo con la cara que forman la hélice 4 y 5 y la hebra beta 5). En el caso de la Flavodoxina los residuos identificados sobre la lámina beta se concentran en la hebra beta 1 y la hebra beta 3 (figura 15c). Un estudio previo [30] encontró que la ruta de plegamiento para esta proteína es distinta a la ruta que sigue CheY y aquí al observar la hebra beta 2 de la Flavodoxina no encontramos conservación de los sitios con correlación respecto a CheY y Spo0F. Para el caso del dominio Rossmann de la Shikimato deshidrogenasa se aprecian más diferencias, casi todos los sitios significativos en la lámina beta se encuentran en las hebras beta 4, 5 y 6 (figura 15d) .

En la sección de arriba sólo se compara a grandes rasgos la localización de los sitios correlacionados entre las diferentes familias de proteínas con plegamiento Rossmann, pero es pertinente realizar una comparación más fina para mapear los sitios con valores significativos de correlación entre las diferentes familias.

Cuando se alinea la estructura tridimensional y la estructura secundaria de CheY contra la de Spo0F (figura 16a) se encuentra una gran cantidad de coincidencias en cuanto a la localización de los sitios con valores significativos de correlación. Esto es de esperarse debido a que ambas proteínas forman parte de los reguladores de respuesta de dos componentes, incluso es posible realizar un alineamiento confiable a nivel de secuencia entre los representantes de ambas familias, en donde se obtiene que la identidad entre esta pareja de secuencias es del 26% y la similitud es del 52%.

El alineamiento entre la Flavodoxina y CheY (figura 16b) es más interesante debido a que estas proteínas presentan un bajo porcentaje de identidad y similitud, 5% y 20% respectivamente, sin embargo la topología general es la misma. No existe una correspondencia uno a uno entre sitios de las dos proteínas debido a que la longitud de la Flavodoxina es mayor, sin embargo es posible hacer una descripción general tomando como referencia la estructura secundaria y terciaria. Las zonas que presentan sitios correlacionados en común entre CheY y Flavodoxina son: la hebra beta 1, zona donde se localiza el núcleo de plegamiento de CheY . El asa beta 1

alfa 1 cercano al sitio de fosforilación o unión a cofactor en CheY y Flavodoxina respectivamente, los sitios en la zona carboxilo terminal de la hebra beta 4 y el asa beta 5 alfa 5. Hay que notar el núcleo estable en la Flavodoxina se localizó en la región carboxilo terminal de la proteína , mientras que en CheY este núcleo estable está en la zona amino terminal [61].

El dominio Rossmann de AroE presenta una hélice alfa y una hoja beta más que el dominio Rossmann de las proteínas anteriores, por ello no es posible alinear su estructura con certeza contra CheY (figura 16b). La única zona donde se aprecia algo de conservación de sitios correlacionados alineando su estructura tridimensional es en la mitad de la parte superior de la lámina beta, donde hay un conjunto de sitios correlacionados, justo en la zona que corresponde al sitio de unión a NADH y que en CheY equivale al sitio de fosforilación, como se mencionó .

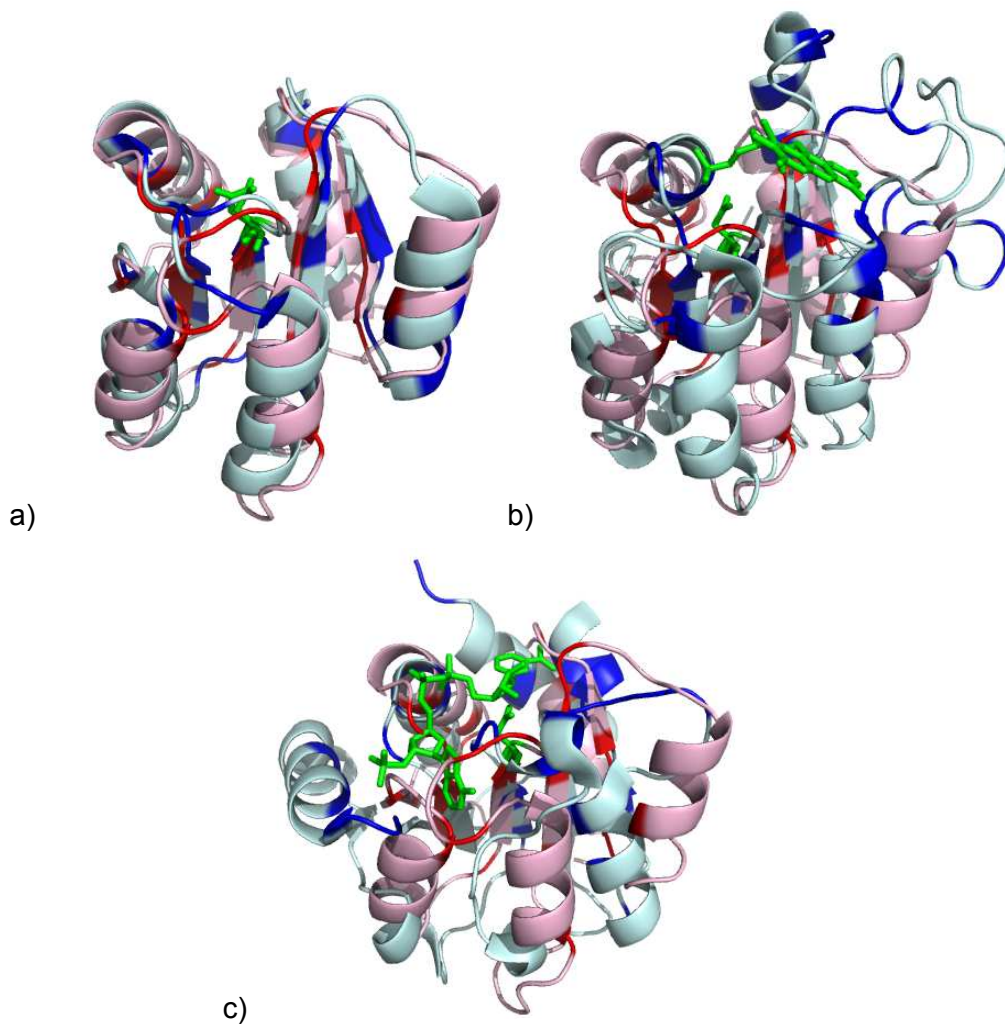


Figura 16. Alineamiento estructura terciaria entre CheY de *E. coli* (en color rosa) y a) Spo0F de *B. subtilis* (en color azul) RMSD b) Flavodoxina de *A. vinelandii* (en color azul) RMSD y c) el dominio Rossmann de AroE de *E. coli* RMSD (en color azul). Los sitios acoplados se remarcan en colores oscuros.

6.5 CORRELACIONES UTILIZANDO ALINEAMIENTOS CON DIFERENTES COMPOSICIONES

La cantidad de secuencias reportadas homólogas a CheY es muy grande y hasta el momento ha sido difícil encontrar secuencias distribuidas de forma más homogénea. En un intento por encontrar un alineamiento con una distribución más homogénea se recolectó de la base de datos de proteínas no redundantes de NCBI las secuencias de todas las proteínas que en su nombre contuvieran la palabra "CheY". El conjunto inicial de secuencias fue de 277,256 secuencias, de esas secuencias sólo se eligieron las que tuvieran en tamaño entre 100 y 140 residuos de aminoácidos, para quedarse únicamente con Reguladores de Respuesta de un solo dominio [62], este filtro redujo drásticamente el número de secuencias a tan solo 7780. Posteriormente se alinearon las secuencias con respecto a CheY de *E. coli* y se quitaron las secuencias con identidad entre ellas mayor al 40%. El conjunto final contenía 528 secuencias, con 23% de promedio de identidad entre parejas de secuencias (figura 17a).

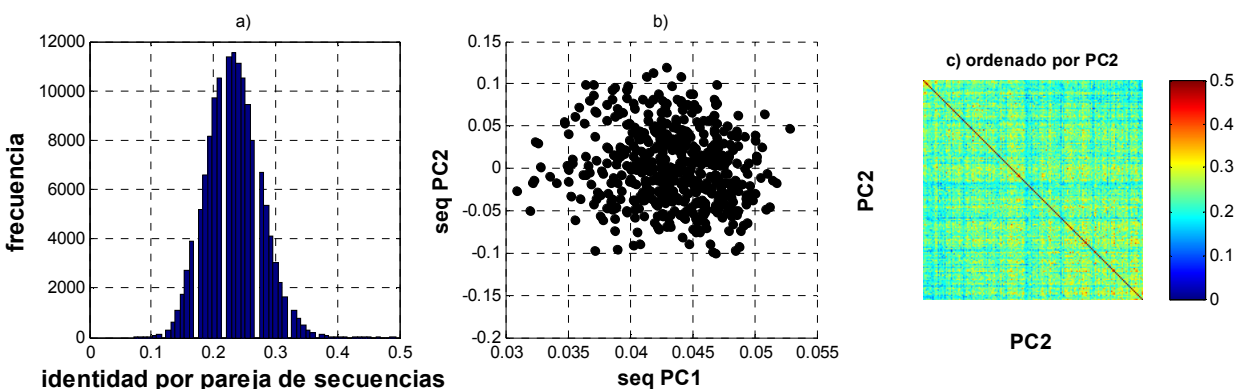


Figura 17. Estructura del alineamiento con Reguladores de Respuesta de un solo dominio. Se observa una distribución uniforme tanto en a) el histograma de identidad entre parejas de secuencias, b) en el diagrama de dispersión de los 2 primeros componentes principales de la matriz de identidad entre parejas de secuencias y c) en la matriz de identidad entre parejas de secuencias.

El conjunto de secuencias tiene una distribución mucho más uniforme que en los alineamientos anteriores según se puede apreciar en la en matriz de identidad entre parejas de secuencias y el diagrama de dispersión de los primeros dos componentes principales de dicha matriz (figura 17b y 17c).

Al identificar sectores se encontró que en el sector rojo una parte coincide con los residuos identificados como núcleo de plegamiento, y otro conjunto de sitios localizados en el mismo sector (los residuos F8, L9, V33, G52, F53 y P82) conectan la base de la lámina beta de la hebra 1 a la 4, y el residuo L25 de la hélice alfa se conecta con L116 del hélice alfa 5 (figura

19a). En el sector azul se observan residuos que rodean el sitio de unión al ión magnesio. En este caso los residuos involucrados en la transmisión de la señal y el reconocimiento de proteínas no fueron detectados.

Sector	Sitios
Sector azul	12, 18, 39, 41, 60, 61, 63, 65,
Sector rojo	8, 9, 25, 29, 33, 36, 42, 52, 53, 82, 102, 103, 110, 111, 116,

Tabla 4 Composición de los sectores identificados utilizando alineamiento que contiene proteínas de un solo dominio y secuencias homólogas a CheY

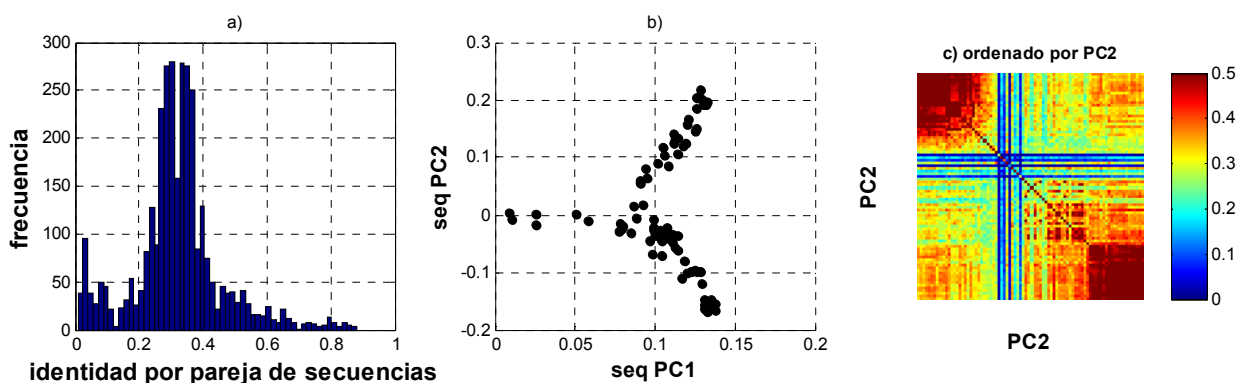


Figura 18. Estructura del alineamiento de CheY *bona fide* a) se puede observar en el histograma, b) el diagrama de dispersión de los primeros dos componentes principales y c) la matriz de identidad de secuencias que se trata de un alineamiento con dos grupos divergentes.

En la quimiotaxis CheY interactúa con otras proteínas. En las bacterias usualmente las proteínas que están involucradas en el mismo proceso suelen estar cercanas en el genoma. Con el mismo objetivo de encontrar un alineamiento homogéneo se buscaron secuencias homólogas a CheY donde existiera evidencia en el contexto genómico de que se tratara de una CheY auténtica, es decir, involucrada en la quimiotaxis. En la base de datos GeConT [36] se recolectaron las secuencias homólogas a CheY que además estuvieran cerca de alguna otra proteína involucrada en la quimiotaxis. Luego se filtraron las secuencias con una identidad mayor al 90%. El alineamiento resultante está claramente estructurado, al parecer hay 2 grupos divergentes separados por el segundo componente principal (figura 18b). El porcentaje de identidad promedio entre parejas de secuencias en este alineamiento es del 33%, mayor que en el alineamiento anterior. En este caso los residuos identificados también fueron diferentes que en el caso anterior. Los residuos de uno de los sectores identificados (sector azul) aquí están localizados en la cara de reconocimiento de las proteínas de interacción e incluyen a muchos de

los elementos que cambian de configuración después del proceso de activación entre ellos T87, Y106, los residuos 89, 90 y 91 del asa beta 4 alfa 4 (figura 19b). Sliversmith reportó que el residuo 89 juega un papel muy importante en la rápida tasa de defosforilación en CheY de *E. coli* [63]. Los residuos W58 y N59, que conectan físicamente la cara de reconocimiento con el sitio de fosforilación también se identificaron. Los residuos T16, I20, L24, L116 de este sector conectan la hélice alfa 1 con la hélice alfa 5, análisis mutagénicos previos [44] habían identificado sitios involucrados en la interacción de CheY con CheZ en estas dos hélices, a pesar de ello ninguno de los sitios identificados en ese estudio coincide con estos.

Los residuos del otro sector (sector rojo) están rodeando el sitio de fosforilación y el de coordinación del ión magnesio (sector rojo, figura 19b) pero ahora no se encuentran sitios sobre la lámina beta, diferente a los casos anteriores. Durante la identificación de sectores se detectó otro, aunque éste no tiene relación evidente con la información experimental disponible, es muy probable que este grupo se forme debido a lo estructurado del alineamiento pero no se puede excluir la posibilidad de que sea algo no reportado [4].

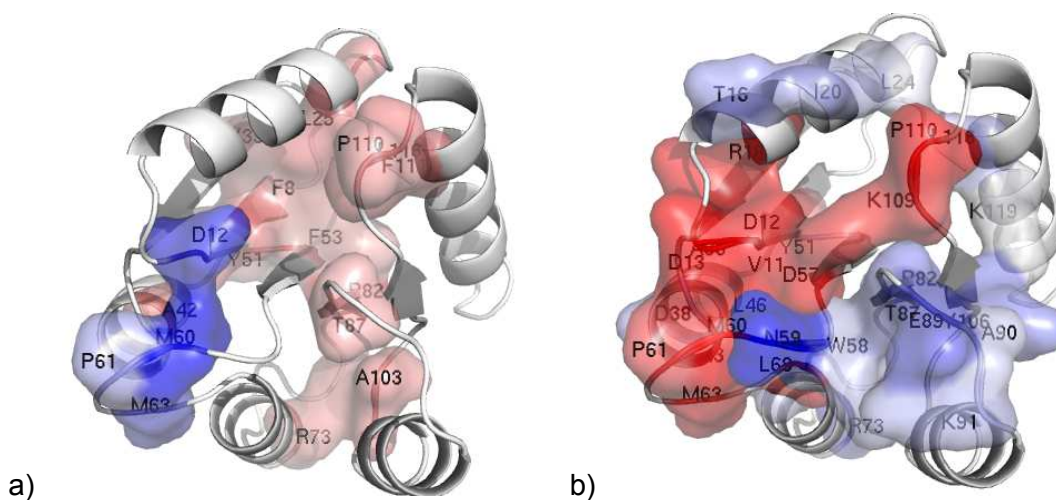


Figura 19. Sectores identificados utilizando alineamientos con distintos tipos de información. a) Sectores identificados con el alineamiento de reguladores de respuesta de un solo dominio porcentaje de identidad promedio de 22% , en rojo se muestran los sitios involucrados con la estructura y la base de la lámina beta b) Sectores identificados con el alineamiento de CheY, con evidencia de contexto genómico, en azul se muestran los sitios involucrados en el reconocimiento, interacción y el mecanismo de activación. En rojo se muestran los sitios que rodean el sitio de fosforilación y coordinación del ión magnesio.

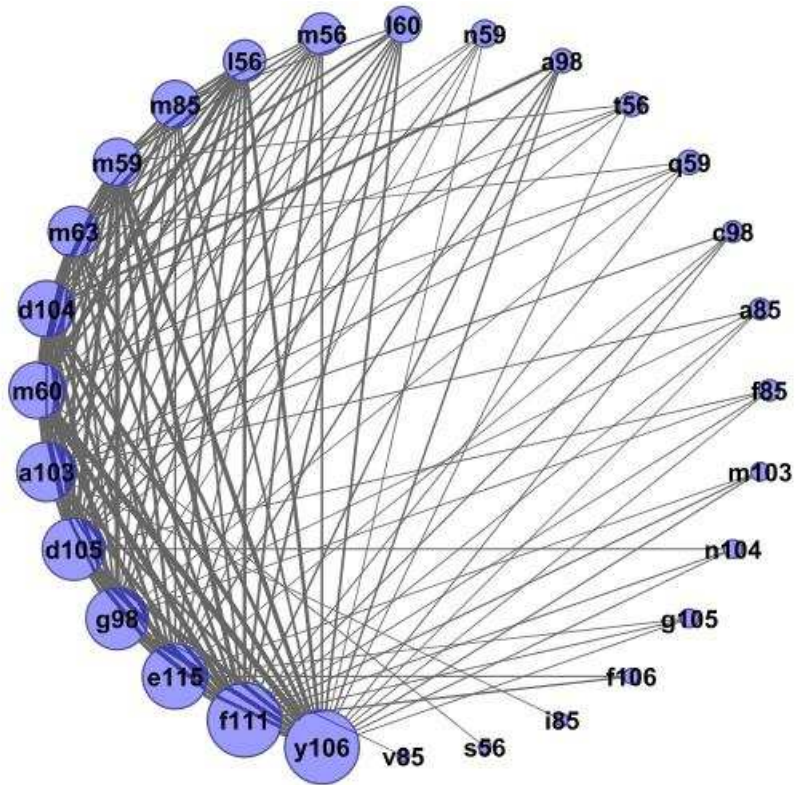
Sector	Sitios
Sector azul	16, 20, 24, 43, 46, 51, 58, 59, 68, 73, 82, 87, 89, 90, 91, 106, 116,
Sector rojo	12, 13, 17, 18, 36, 39, 57, 60, 65, 109, 110, 111

Tabla 5. Composición de sectores utilizando alineamiento con proteínas ortólogas a CheY

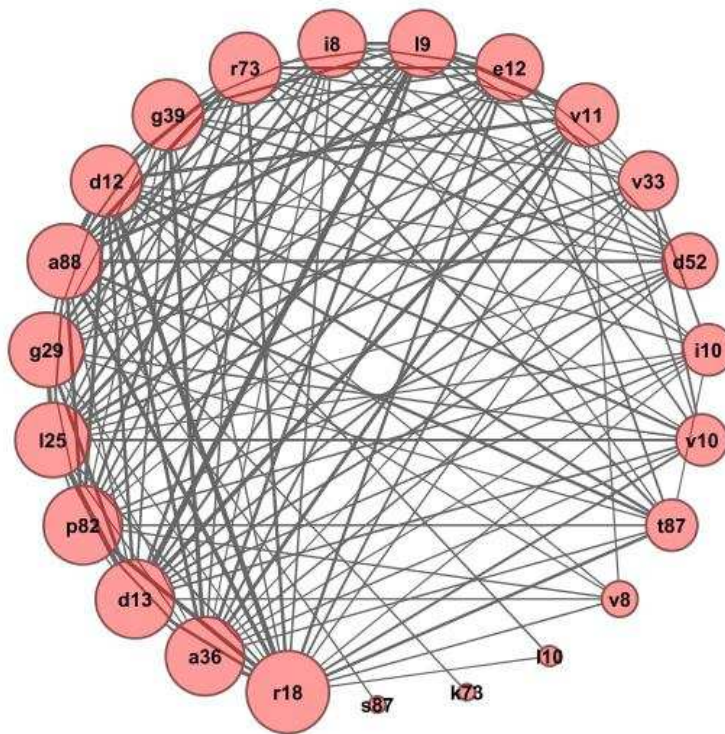
Al utilizar un alineamiento donde las secuencias involucradas tienen la misma función, los sectores identificados contienen sitios implicados en la función de la proteína (transmisión de la señal e interacción con otras proteínas). Cuando se utiliza un alineamiento con secuencias que participan en procesos diferentes pero que tienen la misma estructura, entonces al identificar sectores se detectan sitios involucrados con el plegamiento y la estructura. Es atractivo observar que al cambiar la composición del alineamiento cambia también la composición de los sectores en esta forma.

6.6 CORRELACIONES ENTRE PAREJAS DE AMINOACIDOS EN EL TENSOR 4-D SCA

Al realizar el análisis de acoplamiento estadístico se calculan las correlaciones entre las 400 posibles parejas de aminoácidos para cada una de las posibles parejas de sitios del alineamiento, esta información está contenida en una matriz de 4 dimensiones (N sitios x N sitios x 20 aminoácidos x 20 aminoácidos) a partir de la cual es posible extraer la composición de aminoácidos de los sectores. Una forma práctica de representar este tipo de información es utilizando redes. A continuación se presenta una figura en donde cada nodo representa un aminoácido en un sitio en particular (representado con el código de una letra y el sitio en la secuencia de *E. coli*) y cada arista une dicho nodo con los demás aminoácidos del sector con los que está más correlacionado (figura 20). En ambas redes hay ciertos aminoácidos que poseen únicamente una conexión con algún aminoácido en particular de otro sitio, esto se podría interpretar de dos formas, la primera es que sea un caso particular del alineamiento donde existe un grupo filogenéticamente relacionado que causa dicha observación y la otra es que cierto aminoácido en un sitio en particular prefiere estar con únicamente con esa pareja debido a razones funcionales o estructurales.



a)



b)

Figura 20. Grafo representando correlaciones entre los aminoácidos de los sectores. a) Sector azul b) sector rojo. Las letras en los nodos representan el aminoácido y su posición en el alineamiento inicial, su radio es proporcional al número de nodos con que se conecta. Una arista indica que hay correlación, el grosor de la arista es proporcional al valor de correlación.

A partir de dicha información sería posible reconstruir secuencias silvestres de los sitios correspondientes a los sectores y sería muy interesante proponer nuevas secuencias no vistas en la naturaleza.

7 Conclusiones

Se obtuvo un alineamiento con más de 7000 secuencias homólogas a CheY de *E. Coli*. A partir del alineamiento se calculó la matriz SCA de correlaciones entre sitios. Con la matriz de correlaciones se identificaron dos grupos de sitios (sectores) que evolucionan de manera casi independiente. Uno de los sectores coincide con el sitio de reconocimiento de otras proteínas, en el otro de los grupos se encuentra el núcleo de plegamiento. En la interfaz entre ambos grupos están casi todos los sitios con más del noventa por ciento de conservación, entre ellos el residuo de aspártico que se fosforila.

En el sub-alineamiento obtenido a partir del alineamiento original se pudo extender la identificación de sitios correlacionados hasta la región amino y carboxilo terminal de la proteína. La mayoría de los sitios identificados en el alineamiento original se identificaron en el nuevo alineamiento, también hubieron sitios que no se volvieron a identificar y nuevos sitios identificados. Estas diferencias se originan por la nueva composición del alineamiento y es difícil determinar si las causas son funcionales o históricas.

Se relacionaron residuos identificados experimentalmente como importantes para la estructura, la función o ambas con los sitios identificados con el Análisis Estadístico de Acoplamiento. Uno de los grupos está implicado de forma evidente con la estructura, ya que presenta valores phi elevados, indicio de estar estructurados en el estado de transición durante la ruta de plegamiento. El otro grupo está implicado en la interacción específica con otras proteínas y la transmisión de la señal. La interfaz entre ambos grupos se encuentra rodeando al sitio de fosforilación D57 y conforman también el sitio de unión a magnesio.

Se lograron identificar sitios correlacionados en tres familias de proteínas con plegamiento semejante a CheY. Gran parte de estos sitios se localizan de manera burda en regiones equivalentes. La familia Spo0F que también es un regulador de respuesta es la que más se parece en la composición de sitios correlacionados, después se encuentra la familia de Flavodoxinas y por último se encuentra el dominio Rossmann de unión a dinucleótido.

En el alineamiento estructural, CheY y Spo0F comparten la localización de la mayoría de los sitios correlacionados. En el alineamiento de CheY con Flavodoxina a pesar de que no exista una correspondencia uno a uno entre los sitios de ambas proteínas es posible observar cierta conservación usando criterios más generales como la estructura secundaria. En las tres familias

la zona superior de la lámina beta, que corresponde a los sitios alrededor del sitio de fosforilación o unión al cofactor, contiene gran cantidad de sitios con valores significativos de correlación. En la cara que forma la hélice alfa 4 y 5 y la hebra beta 5 también se encuentran muchos sitios con valores significativos de correlación en todas las familias analizadas.

Al realizar SCA en un alineamiento con secuencias mono-dominio, alejadas evolutivamente entre sí, se identifican sitios importantes para la estructura y la estabilidad, al hacer el mismo análisis de acoplamiento usando secuencias donde se tiene evidencia que cumplen funciones semejantes se identifican sitios involucrados en la función y el mecanismo de activación.

Por último, una vez que se han identificado los sitios con valores significativos de correlación, es posible regresarse a la matriz de correlación e identificar cuáles son las parejas de aminoácidos que están correlacionadas, con esta estrategia es posible conocer cuáles son las combinaciones de aminoácidos que han sido exitosas y permite proponer nuevas combinaciones posibles aun no exploradas en la naturaleza.

8 Perspectivas

8.1 LOS RESIDUOS M60 y M63 TIENEN VALORES DE CORRELACIÓN SIGNIFICATIVOS PERO HAN SIDO POCO ESTUDIADOS.

En los 3 alineamientos de secuencias homólogas a CheY analizados, los residuos M60 y M63 siempre han resultado estar entre los sitios con los valores de correlación más elevados, sin embargo existe poca información acerca de la función de dichos residuos en CheY, estos sitios se localizan en el asa beta 3 alfa 3 cerca del sitio de fosforilación y el de unión al ión magnesio. Estudios experimentales podrían contribuir a conocer más acerca del papel de dichos sitios en los reguladores de respuesta.

8.2 ANALISIS SCA EN ALINEAMIENTO CONCATENADO

CheY interactúa físicamente con CheA, FliM y CheZ, encontrar cuáles son los determinantes de dichas interacciones mediante análisis de correlaciones ha sido objeto de estudio de algunos grupos [64]. Sería interesante hacer un análisis de SCA utilizando un alineamiento concatenando parejas de secuencias de distintas proteínas que interactúan para tratar de identificar sitios importantes para el reconocimiento y especificidad, y comparar los resultados con estudios previos [65], [66], o realizar nuevos estudios. Recordando que no siempre los residuos importantes son sólo los que están en contacto directo según las estructuras cristalográficas [67] y también que no siempre el modo de reconocimiento está conservado entre proteínas de señalización [68].

8.3 DISEÑO DE SECUENCIAS CON FUNCIÓN CHEY

Un área interesante que se ha abordado con éxito usando SCA es el diseño de secuencias no reportadas en la naturaleza que recuperan la estructura y función de secuencias silvestres [69], [70]. A partir del análisis realizado en esta familia sería posible también diseñar nuevas secuencias y observar si se recupera la estructura y la función. Se sabe que la función de esta proteína en la célula depende completamente de la interacción específica con otro conjunto definido de proteínas, por lo cual sería pertinente considerar también las interacciones con las otras proteínas durante el diseño.

8.4 INTERCAMBIO DE FUNCIONES ENTRE PROTEÍNAS CON TOPOLOGÍA SIMILAR

Como se observó en los análisis de SCA realizados en proteínas con plegamiento Rossmann todos ellos presentan sitios con valores significativos de correlación en la zona de asas que unen el extremo carboxilo de las hebras beta con el extremo amino de las hélices alfa (que corresponde al sitio de fosforilación o de unión al cofactor) y en la superficie que forma la última hebra beta y las dos hélices alfa que la rodean. Se propone que sería posible intercambiar funciones entre proteínas si se conserva la zona identificada como importante para la estabilidad en el plegamiento Rossmann y se manipulan las regiones que participan en la función.

8.5 ANÁLISIS DE REDES, PROPIEDADES DE LAS REDES DE INTERACCIÓN ENTRE AMINOACIDOS CORRELACIONADOS

Las redes utilizadas para representar fenómenos biológicos usualmente siguen ciertas propiedades en donde unos pocos nodos están conectados con muchos otros, mientras que muchos nodos están conectados con pocos [71]. Por las figuras mostradas, todo apunta a pensar que las redes aquí encontradas no poseen dichas propiedades, un análisis más detallado de las propiedades de estas redes, ayudará a entender un poco más cómo son las interacciones entre los residuos de esta familia de proteínas.

8.6 ESTABILIDAD DEL GRUPO FOSFATO

La estabilidad del grupo fosfato en los reguladores de respuesta es relevante fisiológicamente porque éstos participan en procesos que tienen escalas de tiempo diferentes, por ejemplo la quimiotaxis requiere que el regulador de respuesta se renueve constantemente debido a que el medio en que se encuentra la bacteria se modifica a medida que la bacteria cambia de posición, en cambio en la esporulación el regulador de respuesta una vez activado por fosforilación tarda

muhco más tiempo en regresar a su estado inactivo. Existen estudios que han tratado de relacionar la secuencia con la estabilidad del grupo fosfato [72], [73]; utilizando análisis estadísticos de acoplamiento en secuencias de reguladores de respuesta donde se conozca la estabilidad del grupo fosfato de las secuencias, sería posible relacionar la secuencia con la estabilidad de dicho grupo.

9 Bibliografía

- [1] C. B. Afinsen, «Principles that govern the folding of protein chains,» *Science* 181, pp. 223-230, 1973.
- [2] A. M. Poole y R. Ranganathan, «Knowledge-based potentials in protein design,» *Current Opinion in Structural Biology*, 16, pp. 508-513, 2006.
- [3] L. A. Mirny, V. I. Abkevich y E. I. Shakhnovich, «How evolution makes proteins fold quickly,» *PNAS*, 95, pp. 4976-4981, 1998.
- [4] N. Halabi, O. Rivoire, S. Leibler y R. Ranganathan, «Protein Sectors: Evolutionary Units of Three-Dimensional Structure,» *Cell*, 138, pp. 774-786, 2009.
- [5] L. Bellsollell, P. Cronet, M. Majolero, L. Serrano y M. Coll, «The three-dimensional structure of two mutants of the signal transduction protein CheY suggest its molecular activation mechanism,» *J. Mol. Biol.*, 257, pp. 116-128, 1996.
- [6] H. S. Cho, S.-Y. Lee, D. Yan, X. Pan, J. S. Parkinson, S. Kustu, D. E. Wemmer y J. G. Pelton, «NMR structure of activated CheY,» *J. Mol. Biol.* 297, pp. 543-551, 2000.
- [7] H. Cho, W. Wang, R. Kim, H. Yokota, S. Damo, S.-H. Kim, D. Wemmer, S. Kustu y D. Yan, «BeF-3 acts as a phosphate analog in proteins phosphorylated on aspartate: structure of a BeF₃⁻ complex with phosphoserine,» *PNAS*, 98, pp. 8525-8530, 2001.
- [8] C. M. Dyer, M. L. Quillin, A. Campos, J. Lu, M. M. McEcoy, A. C. Hausrath, E. M. Westbrook, P. Matusmura, B. W. Matthews y F. W. Dahlquist, «Structure of the constitutively active double mutant CheY D13K Y106W alone and in complex with a FliM peptide,» *J. Mol. Biol.* (2004), 342, pp. 1325-1335, 2004.
- [9] C. M. Dyer y F. W. Dahlquist, «Switched or not?: The structure of unphosphorylated CheY bound to the N terminus of FliM,» *Journal of Bacteriology*, pp. 7354-7363, 2006.
- [10] S.-Y. Lee, H. S. Cho, J. G. Pelton, D. Yan, E. A. Berry y D. E. Wemmer, «Crystal structure of activated CheY, comparison with other activated receiver domains,» *The Journal of Biological Chemistry*, 276, pp. 16425-16431, 2001.
- [11] S. Y. Lee, H. S. Cho, J. G. Pelton, D. Yan, R. K. Henderson, D. S. King, L.-s. Huang, S. Kustu, E. A. Berry y D. E. Wemmer, «Crystal structure of and activated response regulator bound to its target,» *Nature Structural Biology*, 8, pp. 52-56, 2001.
- [12] K. McAdams, E. S. Casper, R. M. Haas, B. D. Sintasiero, A. L. Egger, A. Mesecar y C. J. Halkides, «The structures of the T87I phospho-CheY and T87I/Y106W phospho-CheY help to explain their binding affinities to the FliM and CheZ peptides,» *Archives of Biochemistry and Biophysics*, pp. 105-113, 2008.
- [13] A. M. Stock, E. Martinez-Hackert, B. F. Rasmussen, A. H. West, J. B. Stock, D. Ringe y G. A. Petsko, «Structure of the Mg²⁺-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis,» *Biochemistry*, pp. 13376-13380, 1993.
- [14] X. Zhu, J. Rebello, P. Matsumura y K. Volz, «Crystal structures of CheY mutants Y106W and T87I/Y106W, CheY activation correlates with movement of residue 106,» *The Journal of Biological*

Chemistry, pp. 5000-5007, 1997.

- [15] E. López-Hernández y L. Serrano, «Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, Cl-2,» *Folding and Design*, 1, pp. 43-55, 1996.
- [16] P. Matsumura, J. J. Rydel, R. Linzmeier y D. Vacante, «Overexpression and sequence of the *Escherichia coli* cheY gene and biochemical activities of the CheY protein,» *Journal of bacteriology* 160, p. 3641, 1984.
- [17] R. B. J. F. H. Bourret y M. I. Simon, «Conserved aspartate residues and phosphorylation in signal transduction by the chemotaxis protein CheY,» *PNAS*, 87, pp. 41-45, 1990.
- [18] N. J. Delalez, G. H. Wadhams, G. Rosser, Q. Xeu, M. T. Brown, R. M. B. Ian M. Dobbie, M. C. Leake y J. P. Armitage, «Signal-dependent turnover of the bacterial flagellar switch protein FlIM,» *PNAS*, 107, pp. 11347-11351, 2010.
- [19] B. E. Scharf, K. A. Fahrner, L. Turner y H. C. Berg, «Control of direction of flagellar rotation in bacterial chemotaxis,» *PNAS*, 95, pp. 201-206, 1998.
- [20] R. Barak y M. Eisenbach, «Correlation between phosphorylation of the chemotaxis protein CheY and its activity at the flagellar motor,» *Biochemistry*, 31, pp. 1821-1826, 1992.
- [21] R. B. Bourret y A. M. Stock, «Minireview, Molecular Information Processing: Lessons from bacterial Chemotaxis,» *The Journal of Biological Chemistry*, 277, pp. 9625-9628, 2002.
- [22] D. F. Lowry, A. F. Roth, P. B. Rupert y F. W. Dahlquist, «Signal transduction in Chemotaxis. A propagating conformation change upon phosphorylation of CheY,» *The Journal of Biological Chemistry*, 269, pp. 26358- 26362, 1994.
- [23] M. Solà, E. Lopez-Hernandez, P. Cronet, E. Lacroix, L. Serrano, M. Coll y A. Parraga, «Towards understanding a molecular switch mechanism: thermodynamic and crystallographic studies of the signal transduction protein CheY,» *J. Mol. Biol.* 303, pp. 213-225, 2000.
- [24] M. Welch, N. Chinardet, L. Mourey, C. Birck y J.-P. Samama, «Structure of the CheY-binding domain of histidine kinase CheA in complex with CheY,» *Nature Structural Biology*, 5, pp. 25-29, 1998.
- [25] J. A. Hoch y K. I. Varughese, «Minireview, Keeping signals straight in phosphorelay signal transduction,» *Journal of Bacteriology*, 183, pp. 4941-4949, 2001.
- [26] J. M. Skerker, B. S. Perchuk, A. Siryaporn, E. A. Lubin, O. Ashenberg, M. Goulian y M. T. Laub, «Rewiring the specificity of Two-Component signal transduction systems,» *Cell*, 133, pp. 1043-1054, 2008.
- [27] O. Liarzi, R. Barak, V. Bronner, M. Dines, Y. Sagi, A. Shainskaya y M. Eisenbach, «Acetylation represses the binding of CheY to its target proteins,» *Molecular Microbiology*, 76(4), pp. 932-943, 2010.
- [28] E. Lacroix, M. Bruix, E. Lopez Hernandez, L. Serrano y M. Rico, «Amide hydrogen exchange and internal dynamics the chemotactic protein CheY from *Escherichia coli*,» *J. Mol. Biol.* , 271, pp. 472-487, 1997.
- [29] A. M. Lesk, *An introduction to protein architecture*, Oxford: Oxford Press University , 2001.
- [30] Y. J. M. Bollen, I. E. Sanchez y C. P. M. van Mierlo, «Formation of On and Off- pathway intermediates in the folding kinetics of *Azotobacter vinelandii* Apoflavodoxin,» *Biochemistry*, 43, pp. 10475-10489, 2004.
- [31] C. A. Orengo, A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells y J. M. Thornton, «CATH — a hierarchic classification of protein domain structures,» *Structure* 5, pp. 1093-1108, 1997.
- [32] A. E. T. Todd, C. A. Orengo y J. M. Thornton, «Evolution of function in protein superfamilies, from a structural perspective,» *J. Mol. Biol.*, pp. 1113-1143, 2001.

- [33] M. Y. Galperin, «Structural classification of bacterial response regulators: diversity of output domains and domain combinations,» *Journal of Bacteriology*, 188, pp. 4169-4182, 2006.
- [34] S. W. Lockless y R. Ranganathan, «Evolutionary conserved pathways of energetic connectivity in protein families,» *Science* 286, pp. 295-299, 1999.
- [35] S. Altschul, W. Gish, W. Miller, E. Myers y D. Lipman, «Basic Local Alingment Search Tool,» *Journal of Molecular Biology* 215, pp. 403-410, 1990.
- [36] R. Ciria, C. Abreu-Goodger y E. & M. E. Morett, «GeConT: gene context analysis,» *Bioinformatics*, 20, pp. 2307-2308, 2004.
- [37] W. Li y A. Godzik, «Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences,» *Bioinformatics*, 22, pp. 1658-1659, 2006.
- [38] R. V. Swanson, D. F. Lowry, P. Matsumura, M. M. McEvoy, M. I. Simon y F. W. Dahlquist, «Localized perturbations in CheY structure monitored by NMR identify a CheA binding interface,» *Nature Structural Biology*, 2, pp. 906-910, 1995.
- [39] L. Bellsollell, J. Prieto, L. Serrano y M. Coll, «Magnesium binding to the bacterial chemotaxis protein CheY results in large comformational changes involving its functional surface,» *J. Mol. Biol.*, pp. 489-495, 1994.
- [40] L. Kar, P. Matsumura y M. E. Johnson, «Bivalent-metal binding to CheY protein, Effecto on protein conformation,» *Biochem. J.* 287, pp. 521-531, 1992.
- [41] M. Welch, K. Oosawa y S.-I. M. E. Aizawa, «Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria,» *PNAS*, 90, pp. 8787-8791, 1993.
- [42] P. Cluzel, M. Surette y S. Leibler, «An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells,» *Science*, 287, pp. 1652-1655, 2000.
- [43] Y. Blat, B. Gillespie, A. Bren, F. W. Dahlquist y M. Eisenbach, «Regulation or phosphatase activity in Bacterial Chemotaxis,» *J. mol. Biol.* 284, pp. 1192-1199, 1998.
- [44] X. Zhu, K. Volz y P. Matsumura, «The CheZ-binding surface of CheY overlaps the CheA and FliM-binding surfaces,» *The Journal of Biological Chemistry*, 272, pp. 23758-23763, 1997.
- [45] M. M. McEvoy, A. Bren, M. Eisenbach y F. W. Dahlquist, «Identification of the binding interfaces on CheY for two of its targets, the phosphatase CheZ and the flagellar switch protein FliM,» *J. Mol. Biol.* 289, pp. 1423-1433, 1999.
- [46] J. Guhaniyogi, T. Wu, S. S. Patel y A. M. Stock, «Interaction of CheY with the C-terminal peptide of Che-Z,» *Journal of Bacteriology*, 190, p. 14191428, 2008.
- [47] M. M. McEvoy, D. R. Muhandiram, L. E. Kay y F. W. Dahlquist, «Structure and Dynamics of a CheY-binding domain of the chemotaxis kinase CheA determined by nuclear magnetic resonance spectroscopy,» *Boichemistry*, 35, pp. 5633-5640, 1996.
- [48] M. Schuster, W. N. Abounhamad, R. E. Silversmith y R. B. Bourret, «Chemotactic response regulator mutant CheY95IV exhibits enhanced binding to the flagelar switch and phosphorylation dependent constitutive signaling,» *Molecular Microbiology*, pp. 1065-1075, 1998.
- [49] M. Schuster, R. Zhao, R. B. Bourret y E. J. Collins, «Correlated switch binding and signaling in bacterial chemotaxis,» *The journal of biological chemistry*, pp. 19752-19758, 2000.
- [50] R. B. Bourret, S. K. Drake, S. A. Chervitz, M. I. Simon y J. J. Falke, «Activation of the phosphosignaling protein CheY,» *The Journal of Biological Chemistry*, 268, pp. 13089-13096, 1993.
- [51] R. Barak, K. Prasad, A. Shainskaya, A. J. Wolfe y M. Eisenbach, «Acetylation of the chemotaxis response regulator CheY by acetyl- CoA Synthetase Purified from Escherichia coli,» *J. Mol. Biol*, 342, pp. 383-401, 2004.

- [52] R. Barak, J. Yan, A. Shainskaya y M. Eisenbach, «The chemotaxis response regulator CheY can catalyze its own acetylation,» *J. Mol. Biol.*, 359, pp. 251-265, 2006.
- [53] J. Yan, R. Barak, A. S. Oma Liarzi y M. Eisenbach, «In vivo acetylation of CheY, a response regulator in chemotaxis of *Escherichia coli*,» *J. mol. Biol.*, 376, pp. 1260-1271, 2008.
- [54] R. Ramankrishnan, M. Schuster y R. B. Bourret, «Acetylation at Lys-92 enhances signaling by the chemotaxis response regulator protein CheY,» *PNAS*, 95, pp. 4918-4923, 1998.
- [55] X. Zhu, C. D. Amsler, K. Volz y P. Matsumura, «Tyrosine 106 of CheY plays an important role in chemotaxis signal transduction in *Escherichia coli*,» *Journal of Bacteriology*, pp. 4208-4215, 1996.
- [56] M. S. Formanek, L. Ma y Q. Cui, «Reconciling the "old" and "new" views of protein allostery: A molecular simulation study of Chemotaxis Y protein (CheY),» *Proteins: Structure, Functions and Bioinformatics* 63, pp. 846-867, 2006.
- [57] M. Madhusudan, J. Zapf, J. Hoch, J. Whiteley, N. Xuong y K. Varughese, «A response regulatory protein with the site of phosphorylation blocked by an arginine interaction: crystal structure of Spo0F from *Bacillus subtilis*,» *Biochemistry*, 36, pp. 12739-12745, 1997.
- [58] A. Roszak y A. Laphorn, «Structures of shikimate dehydrogenase AroE and its Paralog YdiB. A common structural framework for different activities,» *J. Biol. Chem.*, pp. 19463-19472, 2003.
- [59] S. Alagaratnam, G. van Pouderooyen, T. Pijning, B. Dijkstra, D. Cavazzini, G. Rossi, W. Van Dongen, C. van Mierlo, W. van Berkel y G. Canters, «A crystallographic study of Cys69Ala flavodoxin II from *Azotobacter vinelandii*: structural determinants of redox potential,» *Protein Sci.* 14, pp. 2284-2295, 2005.
- [60] Y. Y. Tseng y J. Liang, «Are Residues in a Protein Folding Nucleus Evolutionarily Conserved?,» *J. Mol. Biol.*, 335, pp. 869-880, 2004.
- [61] E. Steensma y C. P. M. van Mierlo, «Structural Characterization of Apoflavodoxin shows that the location of the stable nucleus differs among proteins with a flavodoxin-like topology,» *J. Mol. Biol.*, 282, pp. 653-666, 1998.
- [62] U. Jenal y M. Y. Galperin, «Single domain response regulators: molecular switches with emerging roles in cell organization and dynamics,» *Current Opinion in Microbiology*, 12, pp. 152-160, 2009.
- [63] R. E. Silversmith, G. P. Guanga, L. Betts, C. Chu, Z. Rui y B. B. Bourret, «Che-Z mediated dephosphorylation of *Escherichia coli* Chemotaxis Response Regulator CheY: Role for CheY glutamate 89,» *Journal of Bacteriology*, 185, pp. 1495-1502, 2003.
- [64] R. B. Bourret, «Signal transduction meets systems biology: deciphering specificity determinants for protein-protein interactions,» *Mol Microbiol*, 69, pp. 1336-1340, 2008.
- [65] R. Zhao, E. J. Collins, R. B. Bourret y R. E. Silversmith, «Structure and catalytic mechanism of the *E. coli* chemotaxis phosphatase CheZ,» *Nature Structural Biology*, 9, pp. 570-575, 2002.
- [66] H. Thakor, N. Sarah, I. M. Porter, N. Hand y R. C. Stewart, «Identification of an anchor residue for CheA-CheY interactions in the chemotaxis System of *Escherichia coli*,» *Journal of Bacteriology* 193, pp. 3894-3903, 2011.
- [67] A. M. Freeman, B. M. Mole, R. E. Silversmith y R. B. Bourret, «Action at a distance: Amino acid substitutions that affect binding of the phosphorylated CheY Response Regulator and Catalysis of Dephosphorylation can be far from the CheZ phosphatase active site,» *Journal of Bacteriology*, Sept. 2011, pp. 4709-4718, 2011.
- [68] S.-Y. Park, B. D. Beel, M. I. Simon, A. M. Bilwes y B. R. Crane, «In different organisms, the mode of interaction between two signaling proteins is not necessarily conserved,» *PNAS*, 101, pp. 11646-11651, 2004.
- [69] W. P. Russ, D. M. Lowery, P. Mishra, M. B. Yaffe y R. Ranganathan, «Natural-like function in artificial

WW domains,» Nature, 437, pp. 579-583, 2005.

- [70] M. Socolich, S. W. Lockless, W. P. Russ, H. Lee, K. H. Gardner y R. Ranganathan, «Evolutionary information for specifying a protein fold,» Nature, 437, pp. 512-518, 2005.
- [71] A.-L. Barabási y Z. N. Oltvai, «Network Biology: Understanding the cell's functional organization,» Nature Reviews | Genetics February 2004, pp. 101-113, 2004.
- [72] Y. Pazy, A. C. Wollish, S. A. Thomas, P. J. Miller, E. J. Collins, R. B. Bourret y R. E. Silversmith, «Matching biochemical reactions studies to the timescales of life: Structural determinants that influence the autodephosphorylation rate of Response Regulator proteins,» J. Mol. Biol. (2009) 392, pp. 1205-1220, 2009.
- [73] S. A. Thomas, J. A. Brewster y R. B. Bourret, «Two variable active site residues modulate response regulator phosphoryl group stability,» Molecular Microbiology, 69, pp. 453-465, 2008.

Material suplementario

alineamiento Chy70 y sub alineamiento Chy70I120

alineamiento de homólogos a CheY mono dominio

alineamiento ortólogos a CheY contexto genómico.

alineamiento de regulador de respuesta Spo0F

alineamiento de Flavodoxinas

alineamiento de dominio rossman de AroE

script de matlab para cada alineamiento