

UCH-FC
Biotecnología
E. 74
Co 1



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS
ESCUELA DE PREGRADO



ANÁLISIS DE LOS RESIDUOS DE AMINOÁCIDOS INVOLUCRADOS EN EL RECONOCIMIENTO DE β - Y γ -TUBULINA POR LA CHAPERONINA CCT

Seminario de Título entregado a la Universidad de Chile en cumplimiento parcial
para optar al título de Ingeniero en Biotecnología Molecular

YERKO IGNACIO ESCALONA BALBOA

Director del Seminario de Título: Dr. Octavio Monasterio

Co-Director del Seminario de Título: Dr. Luis Pouchucq

SANTIAGO DE CHILE

ENERO 2015



INFORME DE APROBACIÓN SEMINARIO DE TÍTULO

Se informa a la Escuela de Pregrado de la Facultad de Ciencias, de la Universidad de Chile que el Seminario de Título, presentado por el Sr. **Yerko Ignacio Escalona Balboa**. “**Análisis de los residuos de aminoácidos involucrados en el reconocimiento de β - y γ -tubulina por la chaperonina CCT**” ha sido aprobado por la Comisión de la Evaluación, en cumplimiento parcial de los requisitos para optar al Título de **Ingeniero en Biotecnología Molecular**.

Dr. Octavio Monasterio Opazo
Director Seminario de Título

Dr. Luis Pouchucq Marinkovic
Co-Director Seminario de Título

Comisión Revisora y Evaluadora

Dr Ricardo Cabrera Paucar
Presidente Comisión

Dr. Juan Fernández Hidalgo
Evaluador

Santiago de Chile, Enero de 2015

DEDICATORIA

A mis padres, por su apoyo incondicional

AGRADECIMIENTOS

Me gustaría agradecer a todas las personas que me han apoyado en el proceso de esta tesis. Fueron innumerables las experiencias que tuve que pasar para lograr esto.

Agradezco en primera instancia al Dr. Octavio Monasterio por darme soporte en el laboratorio. Su apoyo y conocimiento me permitieron sumergirme en distintas áreas del laboratorio, lo que me permitió descubrir que la biología estructural a través de la bioinformática o biología computacional era lo mío. Agradezco al Dr. José Jaime Arbildua, por enseñarme desde cero a como programar y manejarme con el sistema operativo GNU/Linux. Su forma de ver los problemas biológicos a través de la bioinformática me impresionaba, al igual que sus ganas de ir un paso más adelante que la literatura. Agradezco al Dr. Luis Pouchucq que me enseñó lo sacrificado que es el trabajo en el mesón, la importancia de entender el protocolo más allá del volumen y concentración precisos, y también por las múltiples conversaciones que me motivaba a confiar en mi trabajo y en mí mismo. Agradezco al Dr. Ricardo Cabrera por las correcciones y su dedicación a entender mi trabajo para discutirlo. Agradezco a Felipe Hurtado por las conversaciones donde compartimos nuestra pasión por la ciencia y sacar adelante nuestras ideas en lo que investigamos. Agradezco a Felipe Villanelo por compartir conmigo la aventura de ser bioinformático y estar constantemente buscando nuevos métodos o algoritmos para solucionar los problemas. Agradezco también otras personas que están o pasaron por el laboratorio: a Andrea Poch, Bernardita Poblete, Gisella Araya, Daniel Maturana, Pamela Lisboa y Alexis Ordenes por su gran compañerismo.

Parte del proceso de titulación que no se encuentra en el escrito es mi familia y amigos, que con su apoyo y cariño esto no hubiera sido posible. Primero a mis padres, a mi mamá por abrirme las puertas a descubrir nuevas cosas y hacer del estudio algo entretenido desde pequeño y a mi papá por sacrificarse trabajando para poder tener muchas de las herramientas que necesité para desempeñare estudiando. A mi polola Belén por tener un abrazo cariñoso incondicional en los diferentes momentos por lo que pasé, tanto como compañeros de universidad y ahora conviviendo juntos. A mis compañeros de carrera: Valentina Verdejo, Diego Ampuero, Ignacio Cabrera y Agustín Cabrera por compartir juntos el "stress" de pasar los años de carrera con alegría. Al Dr. Tomás Pérez Aclé por forjarme como presentador. A Sebastian Gutierrez por sus palabras de apoyo. Agradezco también a las personas que me ayudaron a reevaluar mi vida a como la tenía. De partida a Don Marcos que con su cariño encontré como a un segundo padre que me permitió sacar adelante mi escrito. A Emilia Almarza por hacerme entender que los juicios se encontraban en mi cabeza. Y en la línea recta para presentar mi tesis agradezco a Luis Muñoz, que me permitió a enfrentarme a la vida desde una perspectiva diferente, reconciliándome conmigo mismo y sacar lo mejor de mí.

Junto a la dedicación y cariño de todas estas personas, esto no hubiera sido posible. Los quiero mucho.

ÍNDICE DE CONTENIDOS

Dedicatoria.....	iv
Agradecimientos.....	v
Índice de Tablas.....	vii
Índice de Figuras.....	vii
Índice de Anexos.....	viii
Lista de Abreviaturas.....	ix
Resumen.....	x
Abstract.....	xii
1 Introducción.....	1
2 Hipótesis.....	14
3 Objetivos Generales.....	14
4 Objetivos Específicos.....	14
5 Materiales y Métodos.....	15
5.1 Análisis de mutaciones correlacionadas aplicadas a alineamientos de secuencia múltiples.	15
5.2 Análisis de mutaciones correlacionadas entre tubulinas y subunidades de CCT.....	16
5.2.1 Alineamientos múltiples.....	16
5.2.2 Removiendo secuencias redundantes.....	19
5.2.3 Análisis de mutaciones correlacionadas en proteínas de interés.....	19
5.2.4 Ideogramas circulares.....	20
5.2.5 Modelos estructurales.....	21
5.2.6 Experimento de armonía de secuencia para determinar residuos firma.....	21
5.2.7 Determinación de tipos de residuos en β - y γ -tubulina no encontrados en BtubB mediante análisis de armonía de secuencias.....	23
5.2.8 Complementar los resultados de CMA y SH.....	24
6 Resultados.....	25
6.1 Comportamiento de MIp y PCCa con alineamientos con bajo número de secuencias.....	25
6.2 CMA de β - y γ -tubulina con las subunidades de CCT.....	27
6.2.1 Pares de residuos correlacionados de β - y γ -tubulina / CCT visualizadas mediante CIRCOS.....	28
6.2.2 Pares de residuos correlacionados con alta probabilidad de contacto físico	30
6.3 Análisis de residuos armónicos.....	32
6.3.1 Diferencias de β - y γ -tubulina con la tubulina bacteriana BtubB.....	32
6.3.2 Residuos-firma de los dominios apicales de CCT.....	34
6.4 Análisis de mutaciones correlacionadas (CMA) y armonía de secuencias (SH)	34
6.4.1 Nexo entre el análisis de mutaciones correlacionadas y análisis de armonía de secuencias.....	34
6.4.2 Pares de residuos correlacionados con alta probabilidad de contacto.....	38
7 Discusión	41

8 Conclusiones.....	50
9 Bibliografía.....	52
10 Anexos.....	58

ÍNDICE DE TABLAS

Tabla 1. Secuencias de referencia para analizar las proteínas de interés y sus respectivos moldes para construir modelos mediante MODELLER.....	19
---	----

ÍNDICE DE FIGURAS

Figura 1. Homología estructural entre β -tubulina, γ -tubulina y tubulina bacteriana BtubB, y asignación de nombre a las estructuras secundarias de γ -tubulina.....	3
Figura 2. Diagrama de flujo de análisis de mutaciones correlacionadas (CMA) para un set de datos compuesto de diferentes alineamientos de secuencias múltiples con bajo número de secuencias.....	17
Figura 3. Diagrama de flujo para los procedimientos relacionados con el análisis de mutaciones correlacionadas de CCT y β - o γ -tubulina.....	22
Figura 4. Gráficos de distancia promedio entre residuos correlacionados versus el número de secuencias según su puntaje Z-score, y gráficos de número de pares individuales y fracción en contacto según Z-score.....	26
Figura 5. Ideogramas de CIRCOS mostrando los pares de residuos correlacionados para cada tubulina con las subunidades de la chaperonina CCT.....	29
Figura 6: Ideogramas de CIRCOS que representan las interacciones entre residuos correlacionados con alta probabilidad de contacto entre tubulina y las subunidades de la chaperonina CCT.....	31
Figura 7. Residuos obtenidos por el análisis de armonía de secuencias.....	33
Figura 8. Residuos de la chaperonina CCT involucrados en la interacción con β -tubulina.....	36
Figura 9. Residuos de la chaperonina CCT posiblemente involucrados con la interacción con γ -tubulina.....	37
Figura 10. Residuos de β - y γ -tubulina posiblemente reconocidos por las subunidades CCT α y	

CCT δ	39
Figura 11. Residuos de la chaperonina CCT con alta probabilidad de contacto con β - y γ - tubulina.....	40

ÍNDICE DE ANEXOS

Anexo 1: Secuencias usadas para generar el alineamiento β -tubulina/CCT.....	59
Anexo 2: Secuencias usadas para generar el alineamiento γ -tubulina/CCT.....	60
Anexo 3: Gráficos de la cantidad de mutaciones correlacionadas que poseen las tubulinas investigadas con las subunidades de la chaperonina.....	61
Anexo 4: Residuos armónicos que comparten β -tubulina y γ -tubulina que se diferencian de la tubulina Bacteriana BtubB.....	62
Anexo 5: Residuos firma de las diferentes subunidades de CCT clasificados por el dominio al cual pertenecen.....	63

LISTA DE ABREVIATURAS

- CCT: Chaperonina conteniendo TCP1, del ingles "chaperonin containing TCP1"
- CDD: Base de datos de dominios conservados, del ingles "Conserved Domains Database"
- CMA: Análisis de mutaciones correlacionadas, del ingles "Correlated Mutation Analysis"
- MI: Información mutua, del ingles "Mutual Information"
- MIp: Algoritmo derivado de MI
- MSA: Alineamiento múltiple de secuencias, del ingles "Multiple Sequence Alignment"
- MTOC: Centros organizadores de microtúbulos
- PCC: Coeficiente de correlación de pearson, del ingles "Pearson Coeficient Correlation"
- PCCa: Algoritmo derivado de PCC
- PDB: Base de datos de proteínas, del ingles "Protein Data Bank"
- PSSM: Matriz de puntaje de posición específica, del ingles "Position-Specific Scoring Matrix"
- SH: Análisis de armonía de secuencia, del ingles "Sequence Harmony"
- TCP1: del ingles "T-Complex protein 1"

RESUMEN

Durante las últimas décadas se ha descubierto un número creciente de proteínas cuyo plegamiento es mediado por un complejo compuesto por ocho subunidades llamado chaperonina. Los sustratos de este complejo comprenden una amplia diversidad de proteínas, abarcando diferentes funciones, muchas veces esenciales para la mantención de la fisiología celular. En eucariontes, entre los sustratos plegados por la chaperonina CCT (Chaperonin Containing TCP1), se encuentra una proteína que pertenece al citoesqueleto de la célula llamada tubulina. Experimentos de inmunoprecipitación de péptidos de tubulina indican que ciertas regiones tienen una mayor afinidad con CCT, y mediante experimentos de microscopía electrónica se ha podido determinar que existe un patrón de interacción entre tubulina y CCT. Lamentablemente, ninguno de estos métodos ha permitido dilucidar los residuos involucrados en las interacciones entre estas proteínas. Sin embargo, se conoce que la superficie electrostática e hidrofóbica involucrada en la interacción con los sustratos de la chaperonina CCT es heterogénea a diferencia de otras chaperoninas en procariontes o arqueas en que es principalmente hidrofóbica y homogénea. Esta evidencia demuestra que podría existir una especificidad de reconocimiento de las tubulinas por la CCT. Lo cual debería verse reflejado por mutaciones o conservación de residuos específicos en ambas proteínas, a través de un análisis de sus alineamientos múltiples de secuencias (MSA).

En este trabajo se usaron dos métodos bioinformáticos para lograr identificar los

residuos involucrados en la interacción entre tubulina y CCT, específicamente en las tubulinas β y γ . El primer método corresponde al análisis de mutaciones correlacionadas (Correlated Mutations Analysis, CMA) con el cual se obtuvieron los residuos mutados que se correlacionan en tubulina y CCT, para obtener pares de residuos relacionados con su estructura y funcionalidad. El segundo método es el análisis de armonía de secuencias (Sequence Harmony, SH), con el cual se determinan residuos conservados que caracterizan a un grupo de proteínas dentro de una familia de proteínas, para obtener los residuos conservados de β - y γ -tubulina que no se encuentran en la tubulina bacteriana BtubB (cuyo plegamiento no es mediado por chaperonina). De este modo se obtuvo los residuos que estarían implicados en el reconocimiento de tubulina por CCT.

Se utilizaron las herramientas bioinformáticas: BLAST+, MAFFT y Multi-Harmony; los algoritmos para calcular mutaciones correlacionadas: MIp y PCCa; y las bases de datos: CDD y UniprotKB (Swiss-Prot y TrEMBL). Se usaron programas para manejar estas herramientas escritos con el lenguaje de programación Python. De esta forma se logró comprobar que el algoritmo MIp posee una mayor precisión para determinar pares de residuos correlacionados en contacto físico al usar diferentes MSAs de las bases de datos con bajo número de secuencias. Esto permitió investigar las mutaciones correlacionadas entre tubulinas y CCT, con lo cual se pudo identificar que el reconocimiento de tubulina por CCT se explicaría por la presencia de residuos hidrofóbicos y electrostáticos en el dominio intermedio de tubulina, los cuales serían reconocidos por las subunidades CCT δ y CCT α .

ABSTRACT

During the last decades has been found a growing number of proteins whose folding is mediated by a complex of eight subunits called chaperonin. The substrates of this complex include a wide range of proteins, covering different functions and often essential for the maintenance of cell physiology. In eukaryotes, among the substrates folded by the chaperonin CCT (chaperonin Containing TCP1), there are proteins that belong to the cell cytoskeleton called tubulin. Through immunoprecipitation experiments of tubulin peptide it has been determined that tubulin has regions with a higher affinity for CCT, and through electron microscopy experiments a pattern of interaction between tubulin and CCT has been proposed. Unfortunately, none of these techniques can identify the residues involved in the interactions between CCT and tubulin. However, it is known that the electrostatic and hydrophobic surface involved in the interaction between CCT chaperonin and substrate is heterogeneous unlike the interaction of other chaperonins with the substrate in prokaryotes or archaea. This fact suggest that, there should be a specific recognition of tubulin by CCT. Which should be reflected by mutation or conservation of specific residues in both proteins, that could be analyzed by multiple sequence alignments (MSA).

In this work we used two bioinformatics methods in order to identify residues involved in the interaction between tubulin and CCT, specifically β - and γ tubulin. The first method was the analysis of correlated mutations (Correlated Mutations Analysis, CMA) that allowed to obtain mutated residues correlated into tubulin and CCT, to get

pairs of residues associated with its structure and function. The second method was the analysis of harmonic sequences (Harmony Sequence, SH), which identify conserved residues to characterize a group of proteins within a protein family, to find the conserved residues of β - and γ -tubulin that are not found in bacterial tubulin BtubB (chaperonin not required). These tools allowed to identify the residues involved in the recognition of tubulin by CCT.

We used bioinformatics tools: BLAST+, MAFFT and Multi-Harmony; algorithms to calculate correlated mutations: MIP and PCCa; and databases: CDD and UniProtKB (Swiss-Prot and TrEMBL). We used programs to managed all these tools written in Python programming language. Comparison of these tools allowed us to verify that the algorithm MIP had a greater precision to determine that correlated residues pairs involved in physical contact by using different MSAs of the databases with low number of sequences. This permitted us to investigate the correlated mutations between tubulins and CCT, founding that the recognition of tubulin by CCT was mediated by the presence of electrostatic and hydrophobic residues in the intermediate domain of tubulin, which was recognized by the CCT δ and CCT α subunit.

1 INTRODUCCIÓN

La organización interna, forma, motilidad y ciclo de vida de las células eucariontes es manejada por una red compleja de polímeros de proteínas llamado citoesqueleto, el cual incluye los filamentos de actina, filamentos intermedios y microtúbulos (Nogales, 2001).

Los microtúbulos son polímeros cilíndricos de 25 nm de ancho contruidos por el ensamblaje de dos proteínas, la α -tubulina y β -tubulina. Estas proteínas forman heterodímeros permanentes (heterodímero α/β), los cuales actúan como monómeros interactuando longitudinalmente, definido por Nogales et al. 1998 como aquella interacción que forma polímeros longitudinales llamados protofilamentos y lateralmente para formar las paredes del microtúbulo. Los microtúbulos están formados por 13 protofilamentos ensamblados lateralmente formando un cilindro. Debido a que el heterodímero α/β se ensambla longitudinalmente, la estructura completa presenta una polaridad, donde los dos extremos del polímero tienen cinéticas de polimerización distintas: el extremo (-), coronado por α -tubulina, presenta una cinética más lenta que el extremo (+), coronado por β -tubulina. La dinámica de polimerización de este polímero esta mediada por la presencia de GTP, Mg^{2+} y temperaturas fisiológicas. Esta ocurre en dos etapas, una fase lenta, de nucleación, que consiste en la formación de pequeños protofilamentos a modos de núcleos, y la segunda fase mucho más rápida, que corresponde a la elongación lateral o longitudinal a partir de estos núcleos (Desai & Mitchison, 1997). Se conoce que la primera es una fase limitante *in vitro*, en cambio *in*

in vivo, la nucleación de los microtúbulos es asistida en los centros organizadores de microtúbulos (MTOCs). En células animales el centrosoma es el MTOC primario, organizado en torno a un par de centriolos, que sirve como un punto de anclaje central de los microtúbulos en la célula. Su función depende de una proteína llamada γ -tubulina, un homólogo de α -tubulina y β -tubulina. Esta proteína también se encuentra involucrada en la nucleación de microtúbulos en sitios donde no se encuentra los MTOCs, lo que sugiere que es un componente crítico para la nucleación de los microtúbulos *in vivo* (Kollman et al., 2011).

La familia de las tubulinas se encuentra en una superfamilia conocida como FtsZ/Tubulina que poseen representantes tanto en organismos eucariontes (α , β , γ , δ , ϵ tubulinas) como en procariontes (FtsZ, BtubA/B, TubZ) (Pilhofer et al., 2011). Todas las proteínas de la familia de tubulinas presentan una estructura conservada común, que puede ser dividida en tres dominios (ver Fig. 1): un dominio amino-terminal que posee un plegamiento tipo Rossman, que consta de la alternación de hebras betas paralelas (S1-S6) y hélices (H1-H6), donde se encuentra el sitio de unión a GTP, formada por cada "loop" que conecta las hélices y las hebras beta ("loops" T1-T6) y el término del dominio amino-terminal con la hélice central (H7); un dominio intermedio formado por tres hélices (H8-H10) y hebras beta mezcladas (S7-10); y un dominio carboxilo terminal, conformado por dos hélices antiparalelas (H11 y H12), involucrada en la interacción de la tubulina con otras proteínas, como proteínas motoras de microtúbulos y proteínas asociadas a microtúbulos (MAPs) (Nogales, 2001).

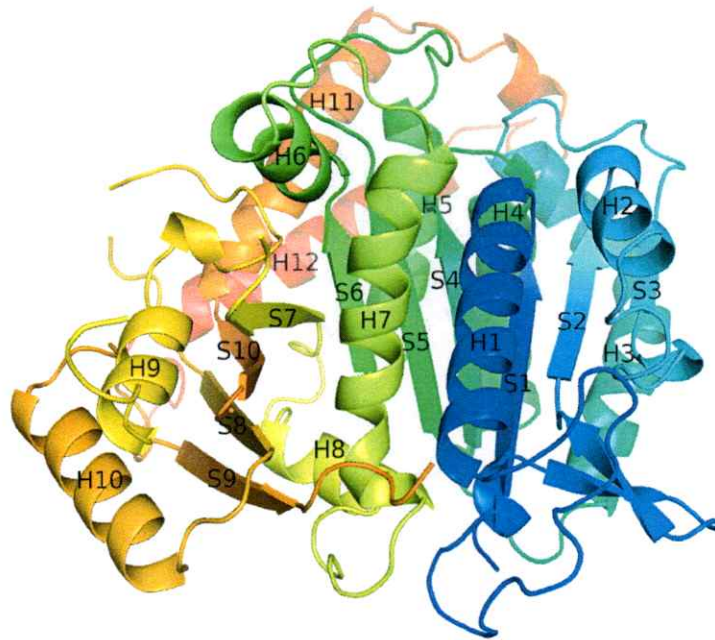
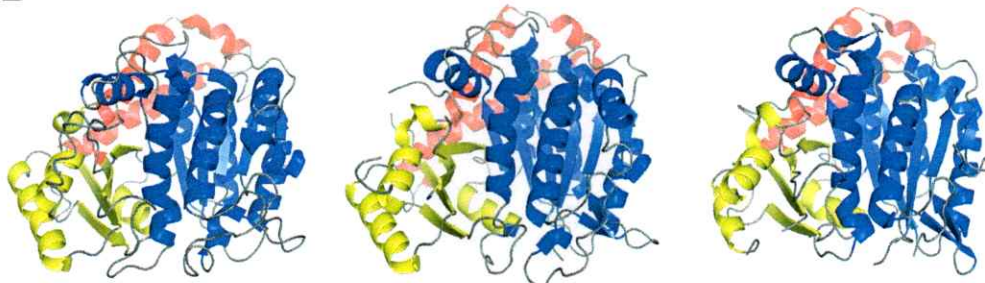
A**B** **β -tubulina** **γ -tubulina****BtubB**

Figura 1. Homología estructural entre β -tubulina, γ -tubulina y tubulina bacteriana BtubB, y asignación de nombre a las estructuras secundarias de γ -tubulina.

A. Representación en cinta de la proteína γ -tubulina con colores del arcoíris desde el carboxilo-terminal en azul, hasta el amino-terminal en rojo, identificando a las diferentes estructuras secundarias según Nogales et al. 1998. **B.** Representación en cinta de las proteínas β -tubulina (1jffB), γ -tubulina (3cb2A) y BtubB (2btqB), con colores según sus dominios: amino-terminal (azul), dominio intermedio (amarillo) y dominio carboxilo-terminal (rojo). Para demostrar la homología se alinearon las estructuras γ -tubulina y BtubB con β -tubulina mediante PyMol, obteniéndose un RMSD para la superposición entre 1jffB y 3cb2A de 1.646 Å, y entre 1jffB y 2btqB de 1.688 Å.

Las proteínas del citoesqueleto, tubulinas y actinas además de ser un extenso y heterogéneo grupo de proteínas, tienen la característica de que son incapaces de plegarse espontáneamente, requiriendo ser asistidas por otros factores. Dichos factores inducen el plegamiento de proteínas desplegadas o a medio plegar, mediante interacciones proteína-proteína. Las células han evolucionado con esta maquinaria de plegamiento que asiste a proteínas en el proceso de síntesis y que además está involucrada en el rescate de aquellas proteínas que han perdido su conformación nativa inducida por algún tipo de estrés. Las proteínas que cumplen la función de asistir el plegamiento de otras proteínas se conocen como chaperonas, existiendo dos grandes clases: las Hsp70 y las chaperoninas (Frydman, 2001).

Frydman (2001) propone que esta maquinaria aísla a la proteína sustrato de la solución acuosa, reconociendo los residuos hidrofóbicos que quedan expuestos, y en el caso de las chaperoninas introduciéndola en la cavidad central de su estructura. Los residuos hidrofóbicos de la proteína sustrato van a permitir luego la formación del "core" o núcleo hidrofóbico, el cual, requiere una gran fuerza motriz para su formación. Así, luego de transitar por distintos estados intermedios, la proteína alcanza su estado final nativo (Frydman, 2001).

Las chaperoninas son complejos de proteínas ATPasas formadas por varias subunidades, que poseen una forma cilíndrica toroidal. Los complejos de chaperonina consisten en dos anillos que interaccionan espalda con espalda. Cada anillo consta de 7 a 9 subunidades, dejando una cavidad interna en donde el sustrato polipeptídico denaturado se pliega. La unión del ATP y su hidrólisis produce cambios

conformacionales que conducen a ciclos de unión y liberación de su sustrato (Spiess et al., 2004).

Se han clasificado a las chaperoninas en dos grandes grupos; el primero corresponde a chaperoninas presentes en procariontes y organelos eucarióticos de origen endosimbiótico como GroEL (que necesita de un cofactor llamado GroES). El segundo grupo, está presente en arqueas, llamado termosomas y en el citosol eucariótico llamado CCT (Chaperonin containing TCP1) o también conocidos como TRiC (TCP1-Ring Complex) (Frydman 2001).

La estructura de cada subunidad de CCT consiste de tres dominios: Un dominio ecuatorial de unión a ATP, un dominio apical involucrado en la unión al sustrato, y un dominio intermedio que conecta los dos dominios anteriores funcionando como bisagra (Spiess et al. 2004).

La chaperonina de eucarionte, a diferencia del termosoma y GroEL-GroES, posee anillos con ocho subunidades diferentes, compartiendo entre ellas un 27 a 39% de identidad de secuencia. La mayor diferencia ocurre en los dominios apicales, donde se observa una alta variabilidad de sus residuos y también una mayor cantidad de residuos conservados por subunidad que no son encontrados en las otras subunidades (Archibald et al., 2001; Pappenberger et al., 2002). En contraste, las chaperoninas bacterianas poseen un solo tipo de subunidad, y las de arqueas, hasta tres diferentes tipos de subunidades. Otra diferencia que distingue a CCT de los termosomas, es su arreglo único de subunidades para formar los anillos, mientras que en los termosomas cada subunidad es dispensable, pudiendo cada una de ellas formar el complejo entero,

mientras que GroEL-GroES está formada por varias repeticiones de la misma subunidad (Archibald et al., 2001; Kapatai et al., 2006). Estas diferencias de CCT con el termosoma y GroEL-GroES se reflejan con mayor fuerza en la diversidad de las propiedades fisicoquímicas de la superficie interior de las chaperoninas. CCT, al igual que en el termosoma, tiene un menor porcentaje de residuos hidrofóbicos en la superficie interna que GroEL-GroES, pero la distribución de los residuos hidrofóbicos en la cara interna de las subunidades de CCT es más asimétrica que en el termosoma y en GroEL-GroES. También la superficie electrostática de CCT posee una distribución de cargas asimétricas en la cavidad interior (Cong et al., 2010).

Dado que las interacciones del termosoma y GroES-GroEL son inespecíficas, lo cual es atribuido por su superficie interior hidrofóbica y la poca variabilidad de sus subunidades. Se deduce por las propiedades de CCT, que la chaperonina pudiera tener interacciones específicas con sus sustratos gracias a su asimétrica superficie electrostática e hidrofóbica. A favor de esta idea se puede mencionar los experimentos de crio-microscopía electrónica que han permitido observar que las proteínas actina y tubulina interactúan con los dominios apicales de CCT con una geometría determinada (Llorca et al., 1999; Llorca et al., 2000). Estos resultados dan indicios que la interacción de actina y tubulina con CCT pudiera estar mediada por interacciones por cargas. No obstante, se ha observado que la proteína G β y el supresor tumoral de von Hippel Lindau (VHL) son reconocidas por CCT mediante interacciones hidrofóbicas (Etchells et al., 2005; Valpuesta et al., 2002; Tam et al., 2006; Spiess et al., 2006). Experimentos de proteómica indican que CCT puede interactuar con múltiples

sustratos de forma similar que GroEL *in vitro*, pero *in vivo* su especificidad es mediada por factores presentes en la biogénesis de las proteínas. Un análisis bioinformático de las proteínas reconocidas por CCT indican que estas corresponden a proteínas principalmente hidrofóbicas, que poseen múltiples dominios y están enriquecidas con sabana β , indicando que CCT podría tener preferencia por evitar el desplegamiento o agregación de esta estructura secundaria (Yam et al., 2008). Aún así, esto no descarta la idea que exista una interacción específica mediada por este tipo de contactos (Muñoz et al., 2011).

Se ha definido que las proteínas que interaccionan entre sí están sometidas a un proceso de coevolución relacionada con sustituciones de aminoácidos involucrados en la interfase proteína-proteína que promueven la selección natural (Pazos et al., 1997). La coevolución entre CCT y sus sustratos se apoya por el hecho de que los dominios putativos de actina y tubulina que están involucrados en la unión con CCT están ausentes o modificados en sus correspondientes homólogos procariontes FtsA/MreB y FtsZ o tubulina bacteriana (BtubA/B) que se pliegan sin necesidad de una chaperonina, y a su vez están implicados en el proceso de polimerización (Llorca et al., 2001). En el caso de la tubulina la mayoría de estas regiones estarían relacionados con las interacciones laterales responsables de la extensión bidimensional de la polimerización en microtúbulos y también los residuos que permiten la dimerización entre α - y β -tubulina, que constituirían a dos “loops” relacionados con la interacción lateral y un dominio conservado del loop T7, respectivamente (Andreu et al., 2002; Pilhofer et al., 2011). Esto se respalda con experimentos de quimeras de FtsZ/tubulina que consta de

una proteína de FtsZ con “loops” de tubulina, lo cual induce el desplegamiento de la quimera y la habilidad de interactuar con CCT (Bertrand et al., 2005). También es sabido que actina y tubulina poseen una arquitectura común, teniendo dos dominios topológicos con una región intermedia que los une, y un nucleótido ubicado en la vecindad. Se cree que la chaperonina ha evolucionado de manera que supera una barrera cinética específica, mediante la estabilización de una conformación abierta que permita el plegamiento de sus dos dominios interactuando por lados opuestos del anillo de la chaperonina. Entonces la evolución de actina y tubulina con CCT tendría relación con las estructuras implicadas en la unión intramolecular de los dos dominios (Llorca et al., 2000). Esto es respaldado por observaciones que demuestran que en las primeras etapas del desplegamiento de la tubulina existe una liberación del nucleótido unida a esta (Andreu et al., 2002).

En nuestro laboratorio (Tesis Pouchucq, 2009; Tesis Araya, 2010 y resultados aún no publicados) se ha demostrado la especificidad de la interacción de CCT con γ -tubulina tanto *in vivo* como *in vitro*. Se introdujo mutaciones sitio específicas en la secuencia de γ -tubulina en dos regiones que según los trabajos de Llorca et al. (2001) podrían estar involucradas en la interacción. Estas regiones corresponden a los “loops” T7 y M. Estos trabajos muestran que mutaciones en la región del “loop T7” afectan en gran medida la interacción, mientras que mutaciones en el “loop M” afectan en menor medida, o probablemente no tengan efecto detectable. Por ello, nuevos estudios en el laboratorio quieren indagar en el proceso de plegamiento de la γ -tubulina por la chaperonina CCT, explicando los fenómenos observados en el laboratorio.

A pesar de toda la información que existe acerca de la unión de CCT a sus sustratos, el mecanismo de cómo ocurre el plegamiento es desconocido. Por ende, un paso en esta dirección sería identificar los residuos o regiones de residuos esenciales en el proceso, en específico aquellos residuos involucrados en el reconocimiento de tubulina por CCT. Para ello, en el presente seminario de título se propone identificar los residuos esenciales en la interacción de CCT con β - y γ -tubulina, utilizando técnicas bioinformáticas llamadas análisis de mutaciones correlacionadas (CMA, del inglés “Correlated Mutations Analysis”) y análisis de armonía de secuencias. La razón de incluir la proteína β -tubulina al análisis, nos permite comparar resultados con los obtenidos para γ -tubulina en busca de patrones similares, pues ambas tubulinas poseen similares geometrías de interacción con CCT (Tesis Pouchucq, 2009). También ya que ambas se encuentran en un mismo nodo junto con BtubB en un árbol filogenético (Pilhofer et al., 2011), una tubulina bacteriana que no interacciona con CCT y se ocupó para diferenciar los residuos conservados entre estas proteínas en un alineamiento de secuencias múltiple.

CMA se basa en el hecho que las proteínas normalmente sufren mutaciones. Estas mutaciones no son toleradas cuando ocurren en sitios conservados relacionados con su funcionalidad o estructura. Pero mutaciones que ocurren en sitios variables pueden afectar en algún grado a la proteína y a la vez ser compensadas por otra mutación, ocurriendo el fenómeno de mutaciones correlacionadas. Estas podrían estar relacionadas en mantener la funcionalidad o estructura de la proteína (Dunn et al., 2008).

Los algoritmos encargados de determinar los residuos implicados en mutaciones

correlacionadas, obtienen esta información mediante alineamientos múltiples de secuencias (MSA, del inglés "Multiple Sequence Alignment"). Entre mayor cantidad de secuencias que sean ingresadas, más información tiene el algoritmo para determinar si un par de mutaciones están correlacionados verdaderamente o si se trata de simples cambios de aminoácidos en las secuencias.

Otro requisito de los algoritmos de CMA, es un alineamiento múltiple con baja influencia de la filogenia en los cálculos. Este corresponde a un problema inherente al analizar alineamientos de secuencias biológicas, dado que dos posiciones del alineamiento no pueden ser independientes ya que los residuos que la componen pertenecen a una misma unidad. Entonces para contrarrestar este efecto se utilizan alineamientos con secuencias no redundantes (Martin et al., 2005). Hecho importante puesto que las proteínas de nuestro interés son muy conservadas en el dominio de los eucariontes. Debido a esto, se pueden obtener muy pocas secuencias no redundantes a las cuales aplicar el algoritmo de análisis de mutaciones correlacionadas.

Debido a que hay pocas investigaciones a causa de familias de proteínas con baja diversidad se decidió evaluar dos algoritmos especializados en el análisis de mutaciones correlacionadas con MSAs con baja cantidad de secuencias para conocer cual es el mejor en la correcta predicción de pares de residuos correlacionados. El primer método esta basado en el algoritmo de información mutua (MI), el cual estima la dependencia mutua que existe entre dos variables azarosas, que en nuestro caso corresponden a los residuos ubicados en una columna de un alineamiento de secuencias con otra columna del mismo alineamiento. Específicamente se utilizó un algoritmo denominado MIp, el

cual se caracteriza por abstraer información de mutaciones correlacionadas ocurridas por simple azar, dándole prioridad a aquellas relacionadas con la estructura y funcionalidad (Dunn et al., 2008) . El segundo método ocupado se basa en el algoritmo que calcula el coeficiente de correlación de Pearson (PCC), que estima los cambios de residuos mediante una matriz de intercambio (basada en propiedades fisicoquímicas) de dos posiciones en el alineamiento y luego calcula el coeficiente de correlación de Pearson. De forma similar a MIp, PCCa subtrae la señal de ruido del “background” para obtener mejores predicciones (Ashkenazy & Kliger, 2010).

Se ha comprobado que los residuos que coevolucionan, son mayoritariamente pares de residuos que se encuentran cercanos unos a otros en la estructura de la proteína (Choi et al., 2005). Mediante este criterio podemos evaluar el comportamiento de los dos métodos utilizando alineamientos obtenidos de bases de datos que poseyeran una secuencia con una estructura conocida, para observar la relación entre los pares de residuos correlacionados con su distancia física en la proteína.

De esta forma se observó el comportamiento de las mutaciones correlacionadas y también de un subgrupo de estos, llamado pares individuales, definido como mutaciones correlacionadas entre dos residuos exclusivamente y que se caracterizan por tener una alta probabilidad de participar en interacciones físicas (Gloor et al., 2005).

Mediante esta metodología se comprobó que el algoritmo MIp es más preciso que PCCa en la búsqueda de mutaciones correlacionadas involucradas en interacciones estructurales. Esto permitió el cálculo de mutaciones correlacionadas de cada tubulina con las diferentes subunidades de CCT, en busca de patrones en común entre β - y γ -

tubulina.

Debido a que el análisis de mutaciones correlacionadas solamente abarca aquellos residuos variables en los alineamientos, se utilizó el método de análisis de armonía de secuencias (SH) para determinar residuos conservados que tuvieran cercanía física con los residuos involucradas con el reconocimiento de tubulina por CCT obtenidos por mutaciones correlacionadas. SH consiste en la comparación de la composición de aminoácidos en una posición determinada de un MSA hecho con diferentes grupos de proteínas de una familia, con el fin de conocer los sitios relacionados con las diferencias funcionales de los grupos (Pirovano et al., 2006).

Así, se identificaron los residuos conservados en tubulina que se diferencian de los residuos de tubulina bacteriana (BtubB) en una misma posición de un alineamiento. Se eligió esta proteína como control negativo debido a que posee similares estructuras secundarias con α - y β -tubulina (ver Fig. 1B), y también se agrupa en el clado de β - y γ -tubulina en un análisis filogenético de la superfamilia de FtsZ/Tubulina (Pilhofer et al., 2011). De esta forma se determinarían cuales son los residuos de tubulina que tuvieran relación con la dependencia de tubulina a ser plegada por CCT, sabiendo que la tubulina bacteriana no necesita de esta maquinaria, ya que no se encuentran codificados los genes de chaperonina en las bacterias con la tubulina bacteriana BtubA/B (Pilhofer et al., 2011).

Con el mismo método, se identificaron los residuos conservados de cada subunidad de CCT que no se encuentran en otras subunidades en una misma posición en la estructura o residuos “firma” de CCT (Pappenberger et al., 2002). Así, poder determinar

los residuos conservados que podrían tener implicancias en el reconocimiento específico de CCT. Estos análisis fueron complementados, lo que dio como resultado en un conjunto de posibles interacciones tanto de β -tubulina como de γ -tubulina con los dominios apicales de CCT.

2 HIPÓTESIS

Debido a que datos experimentales indican un patrón común de interacción entre tubulinas y CCT. El análisis de mutaciones correlacionadas entre tubulinas beta y gamma con las diferentes subunidades de CCT mostrará un patrón común de posibles interacciones.

3 OBJETIVOS GENERALES

Determinar qué residuos o regiones están involucradas en la interacción de CCT y β - o γ -tubulina, mediante técnicas bioinformáticas de análisis de mutaciones correlacionadas en conjunto con el análisis de residuos armónicos.

4 OBJETIVOS ESPECÍFICOS

- Implementación de los algoritmos (MIP y PCCa) para buscar mutaciones correlacionadas en bases de datos con un número de secuencias menor a 125.
- De alineamientos con al menos una secuencia con estructura conocida. Evaluar los pares de residuos correlacionados del análisis de mutaciones correlacionadas mediante su cercanía física.
- Analizar mutaciones correlacionadas de CCT y β - o γ -tubulina, y analizar los residuos armónicos posiblemente involucrados en la dependencia de plegamiento de tubulinas por CCT.
- Determinar la existencia de patrones en común de la interacción entre cada tubulina y los dominios apicales de CCT.

5 MATERIALES Y MÉTODOS

El manejo de programas, datos y análisis de estos, se realizó usando principalmente el lenguaje de programación Python en un computador con el sistema operativo GNU/Linux (Arch Linux y Debian).

5.1 Análisis de mutaciones correlacionadas aplicadas a alineamientos de secuencia múltiples.

Para evaluar la predicción de mutaciones correlacionadas con posible interacción entre dos residuos de una proteína, se ocuparon dos algoritmos denominados MIp y PCCa. Para el análisis, se usaron alineamientos de secuencia múltiples de la base de datos de dominios conservados (CDD) de NCBI (Marchler-Bauer et al., 2011). De este conjunto de datos se eligieron los alineamientos de ortólogos (PFAM, COG y CDD) que poseyeran entre 8 a 125 secuencias y también que poseyeran al menos una secuencia con una estructura conocida por NMR o cristalografía de rayos X, con una resolución menor a 3,5 Å. La primera secuencia con estructura conocida de cada alineamiento se tomo como referencia. Se eliminaron de los alineamientos aquellas secuencias que tuvieran 20% o más de “gaps” con respecto a la secuencia de referencia. Se aplicó MIp y PCCa a los alineamientos y los puntajes obtenidos para cada par de residuos de cada alineamiento fue normalizado obteniéndose los valores de Z-score.

Para cada secuencia de referencia, se obtuvo su correspondiente información estructural en formato PDB de la base de datos PDB (Protein Data Bank), y se corrigió la numeración de los residuos en el PDB según la información FASTA que provee la misma base de datos, usando el programa PDB-FASTA Alignment Tool (P-FAT,

<http://www1.jcsg.org/prod/scripts/pfat/>). En esta parte de la metodología se descartaron del análisis aquellas proteínas con archivos PDB con problemas en la renumeración de sus residuos, debido a duplicación de residuos o desorden en la numeración inicial de los residuos. Lo cual provocó que el programa no los reconociera como válidos.

Luego, para comparar el puntaje *Z*-score obtenido por el análisis de mutaciones correlacionadas con su proximidad física, se calculó la distancia en Angstroms (Å) a cada par de residuos correlacionados con un *Z*-score mayor a 4. Esta distancia, fue calculada entre los centros geométricos de las cadenas laterales de cada residuo, puesto que las cadenas laterales están mayormente implicadas en la interacción entre residuos coevolucionados (Choi et al., 2005). La metodología utilizada se resume en el diagrama de la Figura 2.

Se seleccionaron los resultados de acuerdo con su número de secuencias por alineamiento y valores de *Z*-score mayores a diferentes puntos de corte. Lo mismo se realizó para el subconjunto de mutaciones correlacionadas que forman pares individuales (mutaciones correlacionadas entre dos residuos exclusivamente).

5.2 Análisis de mutaciones correlacionadas entre tubulinas y subunidades de CCT

5.2.1 Alineamientos múltiples

Los alineamientos múltiples de secuencia de las proteínas β -tubulina, γ -tubulina y de las subunidades α , β , γ , δ , ϵ , ζ , η y θ de la chaperonina CCT, se construyeron con secuencias obtenidas de la base de datos Swiss-Prot y TrEMBL de UniProtKB, usando matrices de superposición específica (PSSM) de la base de datos de dominios

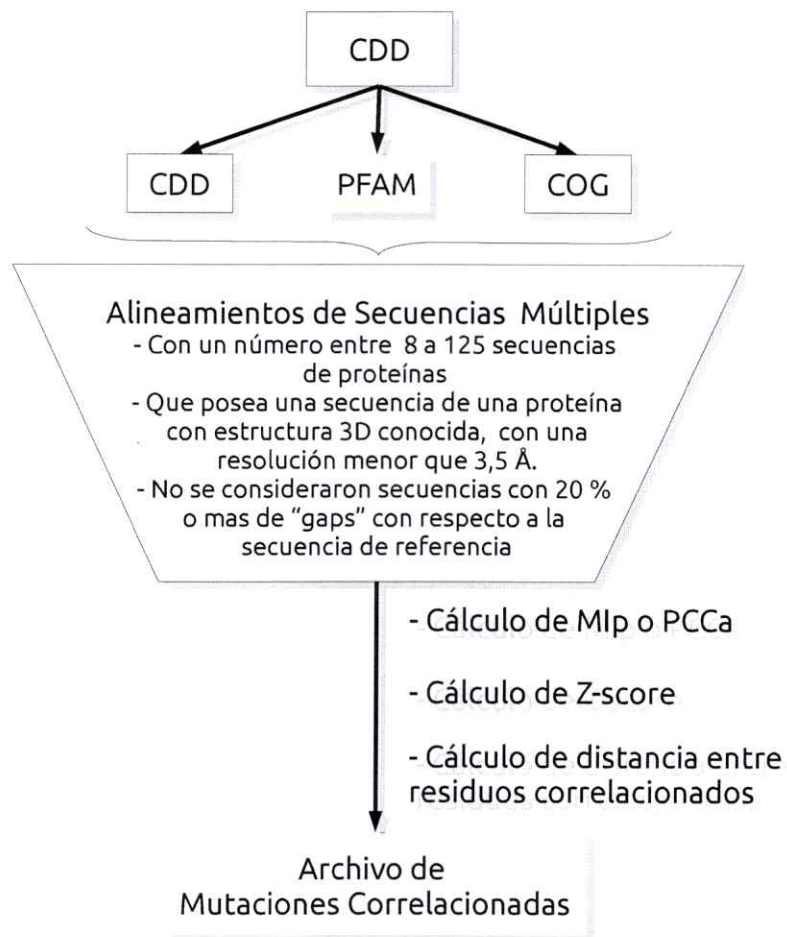


Figura 2. Diagrama de flujo de análisis de mutaciones correlacionadas (CMA) para un set de datos compuesto de diferentes alineamientos de secuencias múltiples con bajo número de secuencias.



conservados (CDD) de cada una estas proteínas como entrada para el programa BLAST+ v2.24.

Cada secuencia obtenida se adquirió su secuencia completa desde su correspondiente base de datos. En el proceso se eliminaron aquellas secuencias que no pertenecen a una especie conocida determinada, las obtenidas mediante la técnica “shotgun sequencing” y las incompletas. Se filtraron secuencias que tenían una diferencia mayor al 10% en el largo con respecto a la secuencia que posee estructura conocida tomada como referencia, la cual pertenece a la especie humana (*Homo Sapiens*) (ver Tabla 1). Se alinearon las secuencias con el programa MAFFT v6.851 usando la estrategia LINSI (Kato et al., 2002). Se estimó como secuencia ortóloga a aquella que apareciera primero en el orden dispuesto por la búsqueda con BLAST, mientras que las secuencias parálogas a esta se eliminaron para evitar la redundancia de una misma secuencia de proteína.

Puesto que el programa de análisis de mutaciones correlacionadas puede recibir como entrada sólo un alineamiento múltiple de secuencia. Para analizar la interacción entre las diferentes proteínas, se necesitó concatenar el MSA de β -tubulina con las diferentes subunidad de la chaperonina (Anexo 1). El mismo procedimiento se realizó con γ -tubulina (Anexo 2). Las secuencias de cada alineamiento se concatenaron según su especie y se filtraron aquellas especies a las cuales les faltara más de una secuencia de cada alineamiento.

Tabla 1. Secuencias de referencia para analizar las proteínas de interés y sus respectivos moldes para construir modelos mediante MODELLER.

Proteína	Secuencia de Referencia (Uniprot)	Molde (PDB)
β -Tubulina	TBB2B_HUMAN (Q9BVA1)	1tubB
γ -Tubulina	TBG1_HUMAN (P23258)	3cb2A
CCT α	TCPA_HUMAN (P17987)	3p9dA
CCT β	TCPB_HUMAN (P78371)	3p9dB
CCT γ	TCPG_HUMAN (P49368)	3p9dC
CCT δ	TCPD_HUMAN (P50991)	3p9dD
CCT ϵ	TCPE_HUMAN (P48643)	3p9dE
CCT ζ	TCPZ_HUMAN (P40227)	3p9dF
CCT η	TCPH_HUMAN (Q99832)	3p9dG
CCT θ	TCPQ_HUMAN (P50990)	3p9dH

5.2.2 Removiendo secuencias redundantes.

Mediante el programa de “clustering” USEARCH (Edgar, 2010) se removieron secuencias de los alineamientos que compartían más de un 95% de identidad entre ellas y se dejó una representante de cada grupo, para disminuir el efecto de la filogenia en los cálculos de mutaciones correlacionadas (Pollock & Taylor, 1997). Luego se usó el programa TrimAl para eliminar aquellas columnas sin residuos o “all gaps” en el alineamiento (Capella-Gutiérrez et al., 2009).

5.2.3 Análisis de mutaciones correlacionadas en proteínas de interés

Se aplicó el algoritmo MIP para el análisis de mutaciones correlacionadas de β - y γ -tubulina con CCT. No se tomaron en cuenta en el análisis las columnas de los alineamientos con más de un 10 % de gaps en cada columna. Los puntajes entregados por el programa se normalizaron obteniendo su puntaje estándar (Z-score).

Con el fin de obtener puntajes de mutaciones correlacionadas que fueran significativos, se generaron diferentes alineamientos a partir del alineamiento inicial (21

en total), en el cual se permutaron al azar los residuos ubicados en una misma posición del alineamiento, de esta forma, sólo se cambia la probabilidad del evento de mutación correlacionada. Se aplicó el CMA a los MSAs con residuos permutados. Se obtuvieron para cada uno sus máximos puntajes de MIp, y estos se representaron como una curva normal, desde donde se calculó su promedio y desviación estándar. Con esto se determinó un valor crítico mayor al promedio con un “p-value” menor a 0,001 en una cola según la distribución t-Student con 20 grados de libertad. Este puntaje crítico de MIp se convirtió a Z-score según el promedio y la desviación estándar de los valores MIp obtenidos sin permutar.

Se tomaron en cuenta en los resultados aquellas mutaciones correlacionadas en posiciones localizadas en el dominio apical de las diferentes subunidades de CCT, ya que este dominio es el principal implicado en el reconocimiento de las tubulinas.

5.2.4 Ideogramas circulares

Se usó el programa CIRCOS para realizar representaciones de interacciones entre los residuos de las diferentes proteínas. De esta forma se crearon ideogramas circulares donde se representan las diferentes proteínas y permite la visualización de pares de residuos correlacionados mediante enlaces (Krzywinski et al., 2009).

Debido a la gran cantidad de enlaces, se realizó un ideograma con el fin de mostrar las regiones con mayor densidad de estos. Para ello se agruparon los enlaces cercanos que se encuentran a 30 residuos entre ellos en la secuencia. También que poseyeran al menos 2 enlaces por grupo, en el caso de γ -tubulina y 3 enlaces por grupo, en el caso β -tubulina. Esto último debido a la disparidad de cantidad de enlaces encontrados de β -

tubulina y γ -tubulina con las diferentes subunidades de CCT.

5.2.5 Modelos estructurales

Con el fin de poder representar gráficamente los resultados obtenidos por mutaciones correlacionadas en la estructura tridimensional de las proteínas de interés, se modelaron las proteínas β - y γ -tubulina, así como también las 8 subunidades de la CCT humana con el programa MODELLER (Sali & Blundell, 1993), usando diferentes moldes (Tabla 1). Los alineamientos utilizados para el modelaje se realizaron mediante HHsearch (Söding, 2005). Se generaron 1000 modelos para cada proteína y se seleccionó el mejor modelo mediante el programa de evaluación de calidad de modelos DOPE de Modeller (Shen & Sali, 2006). También se generó un modelo de β - y γ -tubulina basado en la estructura parcialmente desplegada de α -tubulina que se encuentra dentro de la chaperonina CCT (Llorca et al. 2001). Las figuras de estructuras moleculares fueron hechas por PyMOL (<http://www.pymol.org/>) y se procesaron con PovRay.

Procedimiento del análisis de mutaciones correlacionadas de las tubulinas con las subunidades de CCT se resume en el diagrama de la Figura 3.

5.2.6 Experimento de armonía de secuencia para determinar residuos firma.

Los residuos-firma son aquellos residuos conservados de cada subunidad de CCT que no pertenecen a residuos conservados a todas las subunidades (Pappenberger et al., 2002). A diferencia de la metodología usada por Pappenberger (2002) para encontrar este tipo de residuos, en este trabajo se usó la técnica de armonía de secuencias,

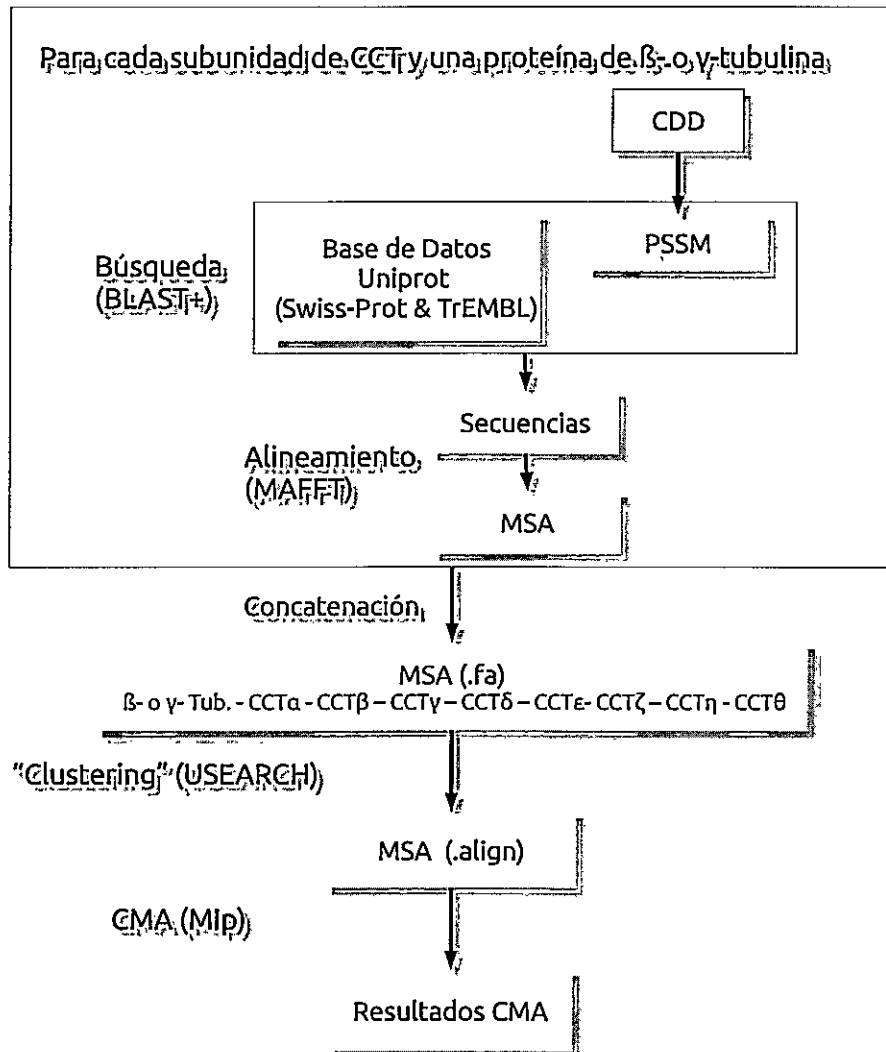


Figura 3. Diagrama de flujo para los procedimientos relacionados con el análisis de mutaciones correlacionadas de CCT y β -o γ -tubulina.

permitiendo considerar más secuencias en el proceso y también clasificar según características fisicoquímicas (por carga positiva, negativa, polaridad e hidrofobicidad). De esta forma los residuos-firma, deben ser residuos conservados, pero que tengan propiedades fisicoquímicas distintas a otras subunidades en la misma posición del alineamiento.

Se colectaron los perfiles PSSM de la base de datos de dominios conservados (CDD) de cada subunidad, para obtener sus secuencias de la base de datos Swiss-Prot y TrEMBL de Uniprot mediante BLAST+. Un conjunto no redundante de datos se obtuvo usando “clustering” mediante el programa USEARCH (Edgar, 2010) obteniendo secuencias con un 90% de identidad de secuencia entre ellas. Se alinearon las secuencias mediante el programa MAFFT usando la metodología LINSI (Kato et al., 2002). Se usó el servidor web multi-Harmony para calcular la armonía de secuencias de múltiples grupos de alineamientos (Brandt et al., 2010). Se tomaron en cuenta los residuos armónicos que poseen un Z-score menor a -6. Luego, por cada posición del alineamiento se seleccionaron los residuos de cada subunidad que se diferenciaron de otros residuos de las demás subunidades por sus propiedades fisicoquímicas para distinguir los residuos firmas.

5.2.7 Determinación de tipos de residuos en β - y γ -tubulina no encontrados en BtubB mediante análisis de armonía de secuencias

Se utilizó la armonía de secuencias para determinar los residuos que podrían estar implicados en la dependencia de las tubulinas a plegarse mediante el uso de chaperoninas. Se usó BtubB como control negativo ya que no necesita de esta

maquinaria, plegándose por si misma.

Se obtuvieron los perfiles PSSM de la base de datos CDD de β - y γ -tubulina, para obtener sus secuencias de la base de datos Uniprot mediante BLAST+. En el caso de las secuencias BtubB solo se colectaron las secuencias que pertenecían al genero *Prostheco bacter* en Uniprot (3 en total). Se alinearon las secuencias, se eliminó las redundancias y se calculó la armonía de secuencias entre estas diferentes proteínas de la misma familia. Luego se seleccionaron los residuos en β - y γ -tubulina no encontrados en BtubB por sus propiedades fisicoquímicas.

5.2.8 Complementar los resultados de CMA y SH

Se seleccionaron los residuos armónicos que se encuentra en un radio de 6 Å de los residuos correlacionados de las regiones más densas, en las estructuras de tubulina y CCT por separado mediante el programa PyMOL.

6 RESULTADOS

6.1 Comportamiento de MIP y PCCa con alineamientos con bajo número de secuencias.

Se estudió el comportamiento de la predicción de mutaciones correlacionadas en relación a su cercanía física. Para ello se usó el procedimiento utilizado por Martin et al. (2005), actualizando las bases de datos que describen los autores, las cuales contienen una mayor cantidad de MSAs con proteínas de estructura conocida. Se obtuvo un total de 5223 alineamientos a los cuales se les aplicó CMA con los algoritmos MIP y PCCa, para comparar sus puntajes normalizados de los pares de residuos correlacionados obtenidos, con la distancia física en la estructura conocida.

De los resultados obtenidos (Fig. 4), se observó una tendencia al clasificar por cantidad de secuencia en el alineamiento aquellos pares de residuos correlacionados con diferentes puntos de corte de Z-score: pues entre mayor es el puntaje de corte, Z-score, menor es el promedio de distancias entre los pares encontrados. Sólo el grupo de alineamientos que posee un rango entre 8 y 24 secuencias no manifiesta esta tendencia (Figura 4A y B). Al realizar este mismo tipo de clasificación en el subgrupo de pares de residuos correlacionados que conforman pares individuales (definido en la introducción), la tendencia es la misma pero los valores promedio de distancia encontrados son menores a los anteriores. En la figura 4 C y D se observa que estos valores alcanzan distancias promedio menores a 10 Å con el algoritmo MIP, en el caso de pares con puntajes Z-score mayor a 5 y 6 y alineamientos con al menos 75 y 50 secuencias, respectivamente. Mientras que al usar el algoritmo PCCa, el promedio de la

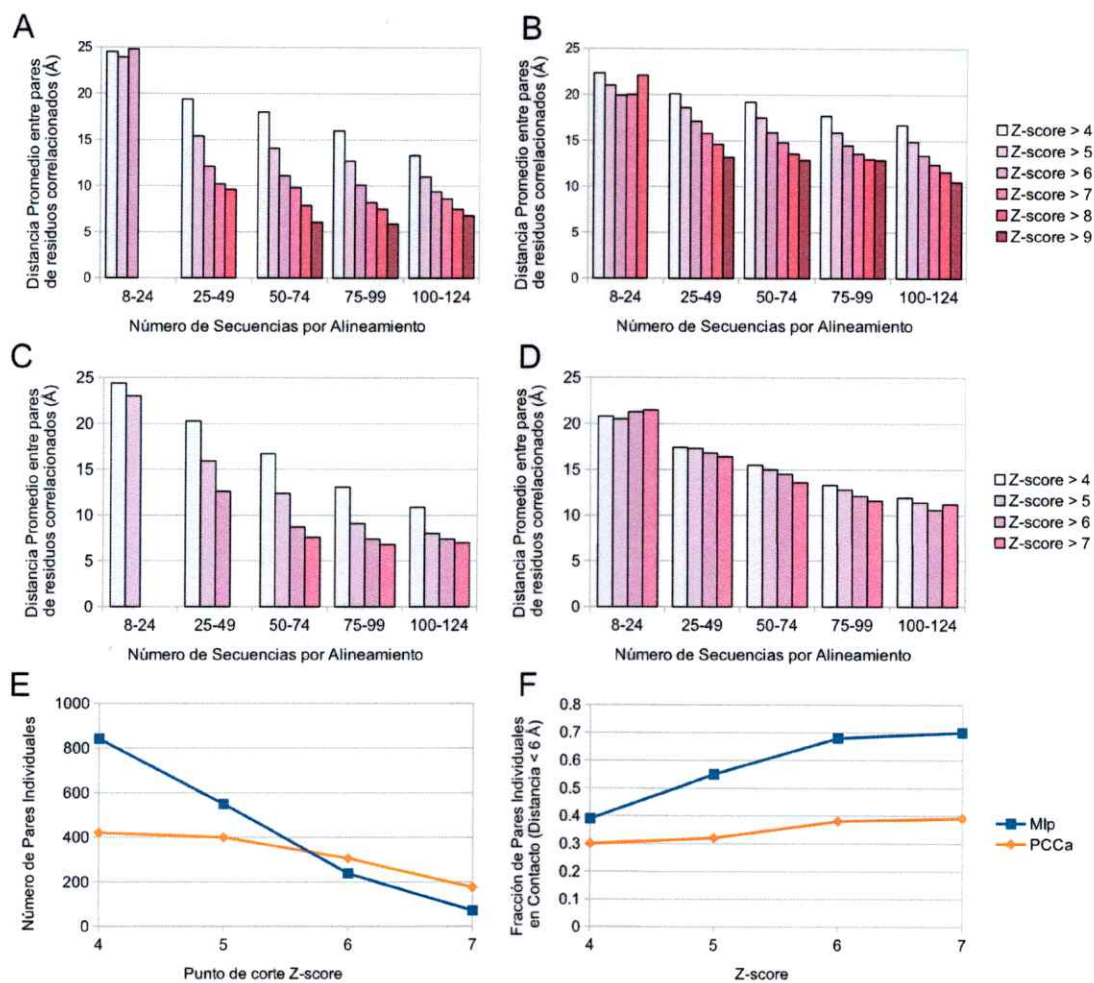


Figura 4. Gráficos de distancia promedio entre residuos correlacionados versus el número de secuencias según su puntaje Z-score, y gráficos de número de pares individuales y fracción en contacto según Z-score.

A y B, corresponde a histogramas que relacionan el número de secuencias de los alineamiento a los cuales se aplicó CMA, versus la distancia promedio (Å) entre residuos correlacionados según su puntaje Z-score, mediante diferentes algoritmos MIP (A) y PCCa (B). C y D, similar a los gráficos anteriores, pero en este caso solamente se utilizan pares individuales. En E y F se grafica la dependencia de los alineamientos que contienen entre 75 a 99 secuencias en función del punto de corte Z-score, con la cantidad y la precisión de los pares individuales. E, gráfico de número de pares individuales de los diferentes alineamientos versus punto de corte de Z-score. F, gráfico de fracción de pares individuales en contacto, definido como aquellos con una distancia menor a 6 Å, versus el Z-score, con el fin de mostrar la precisión de los diferentes algoritmos ocupados.

distancia de los residuos correlacionados predichos son mayores a 10 Å en los diferentes grupos de alineamientos.

Puesto que los alineamientos de las proteínas de interés poseen alrededor de 90 secuencias, el análisis se acotó al rango de alineamientos con 75 a 99 secuencias. En este rango se observa que mientras mayor es el punto de corte de Z-score, menor es el número de pares individuales encontrados (Figura 4E). Siendo una predicción correcta, aquella cuyos pares individuales se encuentran en contacto ($< 6\text{Å}$). Se observó que la precisión, definida como la fracción de predicciones correctas sobre el total de predicciones, con MIP llega a un límite de alrededor de un 70% al seleccionar pares correlacionados con un Z-score mayor a 6 y 7. Mientras que la precisión al usar PCCa, en general es menor a un 40% (Figura 4F).

Cabe considerar que la precisión obtenida de pares individuales en contacto se puede obtener al seleccionar aquellas mutaciones correlacionadas con un Z-score mayor a 9. Información útil para determinar pares de residuos correlacionados con alta probabilidad de contacto.

Por lo tanto, según el puntaje normalizado por cada algoritmo, MIP resulta más preciso que PCCa al predecir mutaciones correlacionadas en contactos al usar los pares individuales.

6.2 CMA de β - y γ -tubulina con las subunidades de CCT

Sabiendo que MIP posee un mejor comportamiento, se calculó mediante este algoritmo, las mutaciones correlacionadas existentes de β -tubulina y γ -tubulina, con el conjunto de subunidades de la chaperonina CCT.

Al determinar un punto donde los puntajes de mutaciones correlacionadas fueran significativos, usando la metodología de analizar alineamientos con residuos permutados de posición, se obtuvo un punto de corte de Z-score de 6,917 para el alineamiento del complejo β -tubulina/CCT y de 6,463 para el de γ -tubulina/CCT. Obteniendo así una cantidad total de pares de residuos correlacionados de: 1287 para el alineamiento β -tubulina/CCT y 1394 para γ -tubulina/CCT, que comprende pares de residuos de una proteína consigo misma o con otras proteínas en el alineamiento. De estos, se obtuvieron 414 mutaciones correlacionadas entre β -tubulina y las diferentes subunidades de CCT, mientras que 181 para γ -tubulina y las diferentes subunidades de CCT.

6.2.1 Pares de residuos correlacionados de β - y γ -tubulina / CCT visualizadas mediante CIRCOS

Los pares de residuos correlacionados pueden ser visualizados mediante los ideogramas de CIRCOS (Fig. 5), en los cuales se puede apreciar líneas que unen un residuo de cada tubulina con un residuo del dominio apical de las diferentes subunidades de CCT. Para β -tubulina se obtuvo un total de 71 pares de residuos correlacionados y su contraparte γ -tubulina con CCT se obtuvo un total de 22 pares de residuos correlacionados. En cuanto a la cantidad de pares de residuos correlacionados por subunidad de CCT: las subunidades que presentan la mayor cantidad de pares de residuos correlacionados entre β -tubulina y CCT son las subunidades CCT α (24), CCT δ (17) y CCT ϵ (14). Por otro lado, las subunidad que presentan mayor número de pares de residuos correlacionados entre γ -tubulina y CCT son las subunidades CCT α (3) y CCT δ (9) (Figura 5 A y B, Anexo 3).

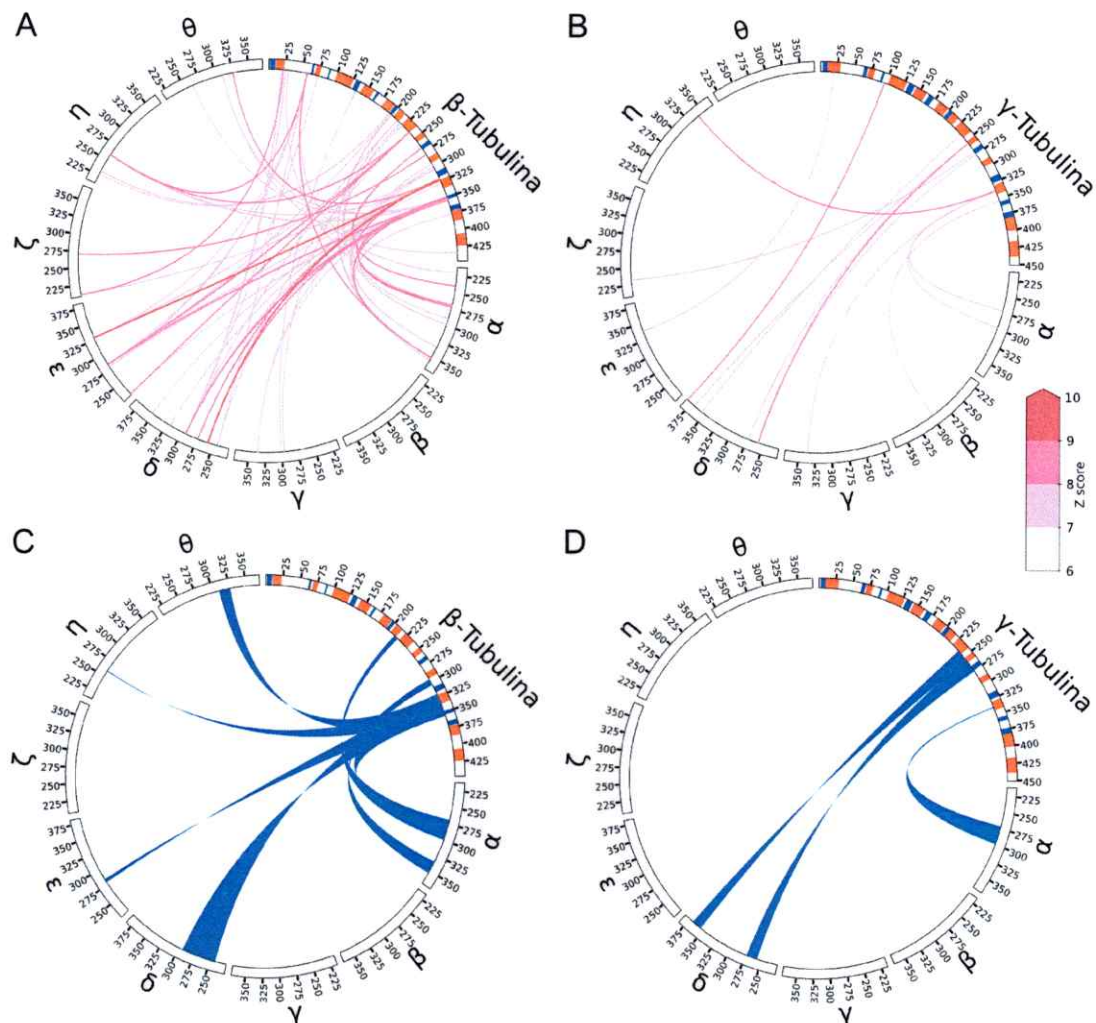


Figura 5. Ideogramas de CIRCOS mostrando los pares de residuos correlacionados para cada tubulina con las subunidades de la chaperonina CCT.

Cada círculo contiene una proteína tubulina y las diferentes subunidades de la chaperonina CCT designadas por letras griegas. También se incluye la estructura secundaria correspondiente para la secuencia de tubulina que se muestra en colores, en rojo (hélice α) y azul (hebra β), sin color (sin estructura). **A**, pares de residuos correlacionados significativos entre β -tubulina y CCT, y **B**, entre γ -tubulina y CCT. Ambos pares identificados por colores según Z-score. **C**, cintas que representan zonas más densas de pares de residuos correlacionados entre β -tubulina y CCT, y **D**, entre γ -tubulina y CCT.

Para simplificar el análisis, se agruparon los pares de residuos correlacionados según las zonas más densas. Se puede observar que en el análisis de mutaciones correlacionadas de β -tubulina/CCT destacan las subunidades α , δ , ϵ , η y θ . Mientras que en γ -tubulina/CCT destacan las subunidades α y δ solamente. Sin embargo, resulta muy interesante observar las regiones donde se agrupan los pares de residuos correlacionados en la estructura de las tubulinas. Estas regiones se encuentran entre la hélice H6 y la hebra beta S7 y también entre la hélice H9 y la hebra beta S9, siendo la primera región la que interactúa primordialmente con la subunidad α , y la segunda con la subunidad δ en β -tubulina/CCT, mientras que en γ -tubulina/CCT es al revés (Fig. 5 C y D).

6.2.2 Pares de residuos correlacionados con alta probabilidad de contacto físico

Se utilizaron los pares individuales con un valor Z-score mayor a 7 y también se incluyeron los residuos que no forman pares individuales pero poseen un Z-score mayor a 9, pues la precisión de encontrarse en contacto es similar. Este tipo de pares de residuos correlacionados con alta probabilidad de estar en contacto directo se visualizan mediante un ideograma de CIRCOS en la Figura 6. Se puede observar que estos residuos corresponden a los de las subunidades CCT γ , CCT δ y CCT ϵ en el complejo β -tubulina/CCT y CCT γ y CCT δ en el complejo γ -tubulina/CCT.

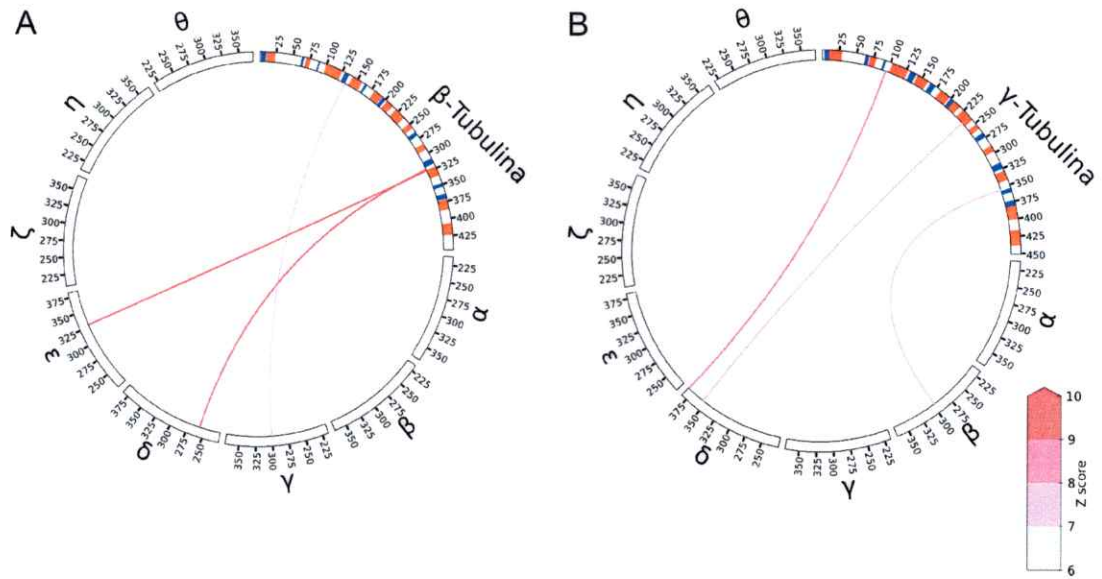


Figura 6: Ideogramas de CIRCOS que representan las interacciones entre residuos correlacionados con alta probabilidad de contacto entre tubulina y las subunidades de la chaperonina CCT.

Cada círculo contiene una proteína tubulina y las diferentes subunidades de la chaperonina CCT designadas por letras griegas. También se incluye la estructura secundaria correspondiente para la secuencia de tubulina que se muestra en colores, en rojo (hélice α) y azul (hebra β), sin color (sin estructura). **A**, pares de residuos correlacionados significativos entre β -tubulina y CCT, y **B**, entre γ -tubulina y CCT. Ambos pares identificados por colores según Z-score.

6.3 Análisis de residuos armónicos

6.3.1 Diferencias de β - y γ -tubulina con la tubulina bacteriana BtubB

Para complementar los resultados obtenidos mediante el análisis de mutaciones correlacionadas se identificaron los residuos conservados según sus propiedades fisicoquímicas, de β - y γ -tubulina que no se encuentran en la tubulina bacteriana BtubB.

La diferencia existente entre las tubulinas eucariontes y la tubulina bacteriana BtubB es la presencia de aminoácidos cargados o hidrofóbicos conservados en las tubulinas eucariontes, que en la tubulina bacteriana corresponden en su mayoría a aminoácidos polares (Anexo 4) . Estos residuos se representaron gráficamente en la estructura de γ -tubulina con el fin de distinguir en que regiones estructurales se encuentran. Para ello se construyeron dos modelos de la proteína γ -tubulina de humano, uno basado en la estructura cristalina de γ -tubulina humana unida a GDP (3cb2 PDB), y el otro basado en la estructura parcialmente desplegada de α -tubulina obtenida a partir de su interacción con la chaperonina CCT en su interior desde datos de microscopía electrónica (Llorca et. al., 2001) (Fig. 7). Con respecto al primer modelo, la localización de estos residuos en la γ -tubulina nativa muestra una distribución de los residuos con carga concentrados principalmente en los dominios amino y carboxilo terminal. Estos mismos residuos cargados, como se observa en el segundo modelo con la estructura desplegada de la proteína, se encuentran distribuidos de una manera no uniforme en los extremos de la proteína, así la razón negativos:positivos es de 7:1 en el dominio amino-terminal, mientras que en el dominio carboxilo esta razón es de 3:3. También existe una concentración de grupos hidrofóbicos principalmente en el dominio intermedio de las

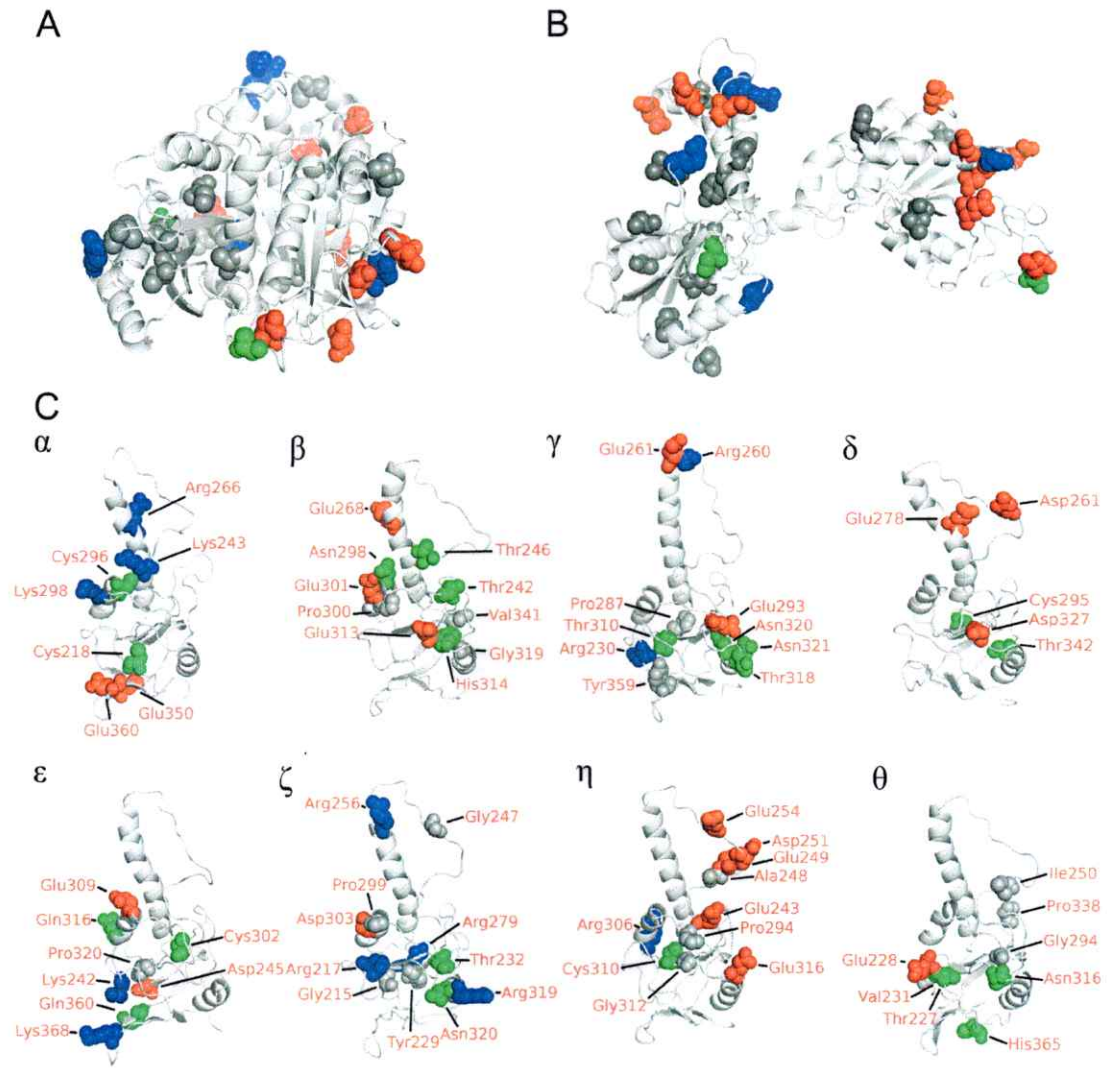


Figura 7. Residuos obtenidos por el análisis de armonía de secuencias.

Los residuos del análisis de armonía de secuencias se representan como esferas y con colores según sus propiedades fisicoquímicas, con residuos de color rojo (carga negativa), color azul (carga positiva), color verde (polares) y gris (apolares y aromáticos). **A**, residuos de β y γ tubulina que se diferencian de la tubulina bacteriana BtubB, mapeados en el modelo plegado y **B**, parcialmente desplegado de γ -tubulina. **C**, residuos-firma del dominio apical de las diferentes subunidades de CCT, identificadas con letras griegas.

tubulinas, ubicada en el "core" del modelo de la proteína plegada, mientras que en el modelo de estructura parcialmente desplegada, este grupo de residuos se encontraría expuesto al medio acuoso.

6.3.2 Residuos-firma de los dominios apicales de CCT

Al identificar los residuos-firma que determinan las características exclusivas que posee cada subunidad, se determinó que cada dominio apical de las subunidades posee entre 5 a 11 residuos-firma, siendo las subunidades β , γ , δ y ζ las que poseen una mayor cantidad de residuos-firma (Anexo 5). Al representar gráficamente estos residuos en la estructura modelada, los residuos-firma con carga negativa predominan y se encuentran principalmente en las subunidades β , γ , δ y η , y los residuos con carga positiva predominan en las subunidades α , γ , ε y ζ , los residuos polares predominan en las subunidades β , γ y η , y finalmente los residuos hidrofóbicos predominan en las subunidades η y θ (Fig. 7).

6.4 Análisis de mutaciones correlacionadas (CMA) y armonía de secuencias (SH)

Para el análisis en conjunto de CMA y SH se representaron gráficamente las proteínas. Se utilizó el arreglo de las subunidades recientemente caracterizada por Kalisman et al. (2012) como referencia, debido a que la disposición de las subunidades en este modelo tiene un mayor respaldo experimental que otros arreglos determinados previamente (Kalisman et al., 2012).

6.4.1 Nexos entre el análisis de mutaciones correlacionadas y análisis de armonía de secuencias

Para complementar la información obtenida por el análisis de mutaciones

correlacionadas con la de armonía de secuencias, se procedió a representar gráficamente los resultados de mutaciones correlacionadas en los modelos estructurales de las diferentes proteínas de interés. Se tomaron en cuenta los pares correlacionados de las regiones más densas para simplificar el análisis. Estos gráficos muestran solamente los residuos de los análisis de armonía de secuencias cercanos a los residuos de mutaciones correlacionadas.

Al tomar en conjunto las figuras 8 y 9 se pueden observar las diferencias entre los patrones de residuos correlacionados de β - y γ - tubulina. En relación con γ -tubulina, β -tubulina posee además, mutaciones correlacionadas con las subunidades ϵ , η y θ , en las cuales no existe una alta cantidad de residuos firma cercanos a los residuos de mutaciones correlacionadas comparado con las subunidades CCT δ y CCT α (Fig. 8). También se observa que las mutaciones correlacionadas entre β - y γ tubulina tienen un patrón común con los dominios apicales de las subunidades de CCT, en donde se destacan las subunidades CCT δ y CCT α en los dos casos. En específico los residuos Asp261 y Glu278 de CCT δ , y Cys296 y Lys298 de CCT α (Fig. 8 y 9).

A partir de estos resultados se puede inferir que las cargas negativas de CCT δ podrían estar involucradas en el reconocimiento de tubulinas, al igual que un residuo polar y otro con carga positiva de CCT α .

De la misma forma que se analizaron los residuos-firma cercanos a los residuos involucrados en mutaciones correlacionadas, se analizaron los residuos armónicos de tubulina que son diferentes a los de la tubulina bacteriana BtubB cercanos a los pares correlacionados. De este análisis se determinó que el dominio intermedio de las

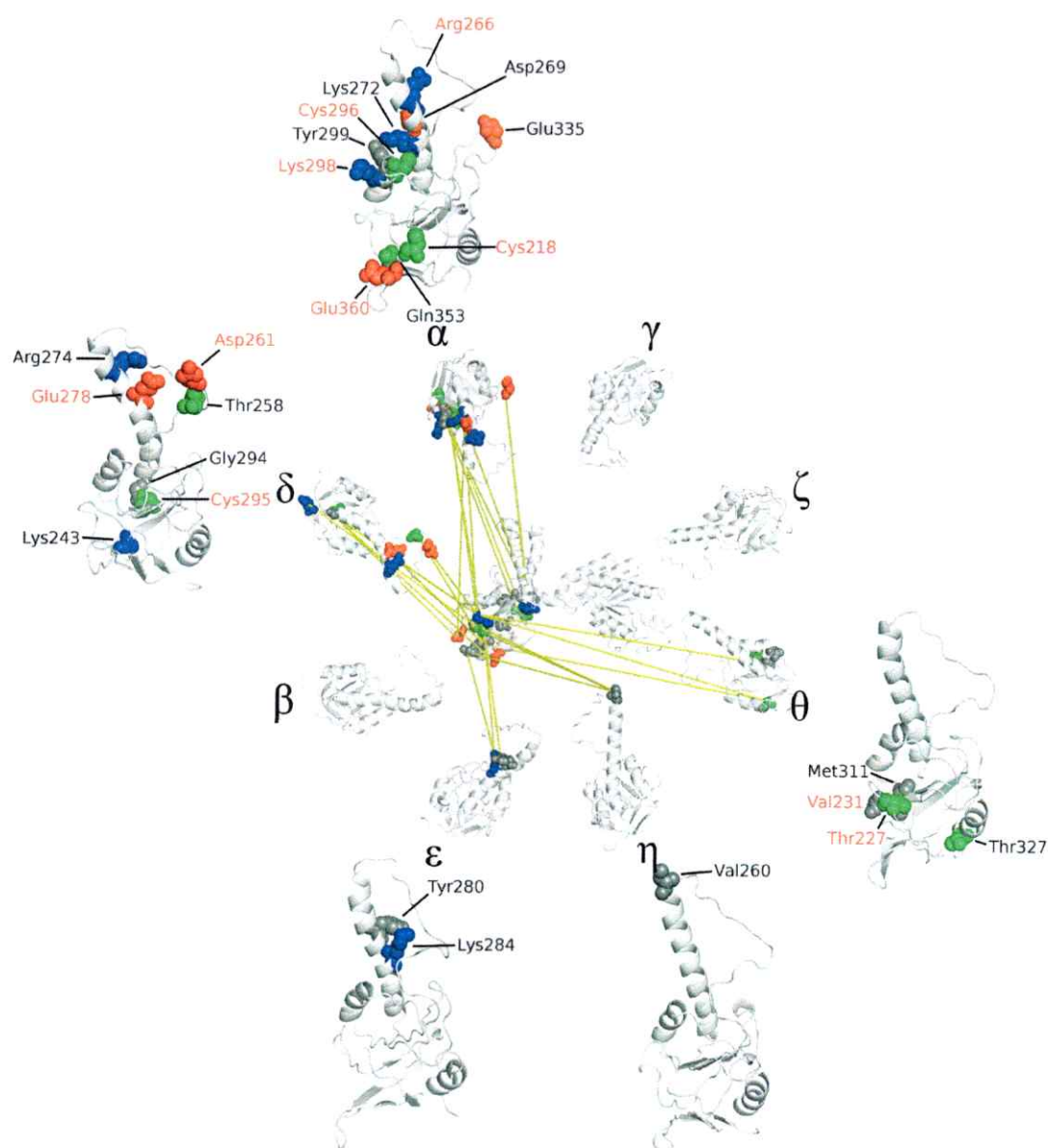


Figura 8. Residuos de la chaperonina CCT involucrados en la interacción con β -tubulina. Representación de cintas de las proteínas β -tubulina parcialmente desplegada (en el centro) y los dominios apicales de las diferentes subunidades de CCT (alrededor de la tubulina), con residuos provenientes de los análisis CMA y SH que se destacan en forma esférica con colores según propiedades fisicoquímicas con residuos de color rojo (carga negativa), color azul (carga positiva), color verde (polares) y gris (apolares y aromáticos) y líneas amarillas que unen los residuos correlacionados. Las subunidades se encuentran ordenadas según Kalisman et al. (2012). Para aquellas subunidades que poseen residuos correlacionados con tubulina, se adiciona una representación en cinta de su dominio apical desde una vista interior de la chaperonina. Se destacan los residuos correlacionados CMA mediante letras con color negro, mientras que los residuos-firma cercanos a menos de 6Å de los residuos de CMA con letras en color rojo.

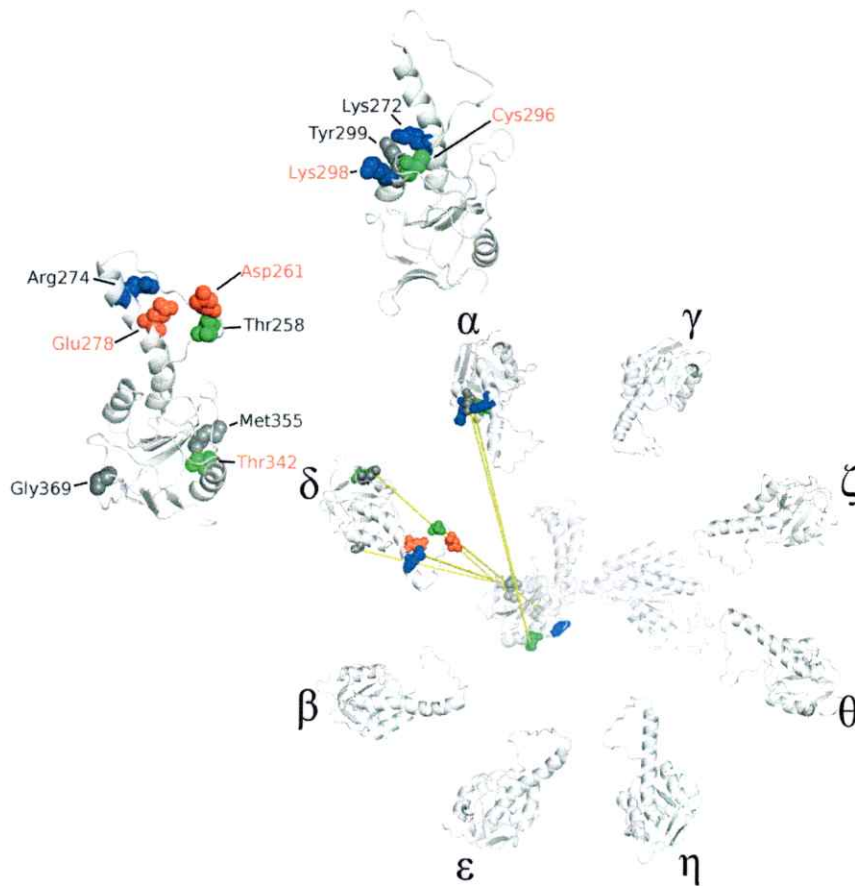


Figura 9. Residuos de la chaperonina CCT posiblemente involucrados con la interacción con γ -tubulina.

Representación de cintas de las proteínas γ -tubulina parcialmente desplegada (en el centro) y los dominios apicales de las diferentes subunidades de CCT (alrededor de la tubulina), con residuos provenientes de los análisis CMA y SH que se destacan en forma esférica con colores según propiedades fisicoquímicas con residuos de color rojo (carga negativa), color azul (carga positiva), color verde (polares) y gris (apolares y aromáticos) y líneas amarillas que representan los residuos correlacionados. Las subunidades se encuentran ordenadas según Kalisman et al. (2012). Para aquellas subunidades que poseen residuos correlacionados con tubulina, se adiciona una representación en cinta de su dominio apical desde una vista interior de la chaperonina. Se destacan los residuos correlacionados CMA mediante letras con color negro, mientras que los residuos-firma cercanos a menos de 6Å de los residuos de CMA con letras en color rojo.

tubulinas es importante en la interacción con CCT. Este dominio intermedio contiene residuos hidrofóbicos y también residuos cargados que podrían ser reconocidos por las subunidades CCT δ y CCT α (Fig. 10).

El análisis en su conjunto mostró que no existen residuos armónicos en común que se encuentren en una misma posición en β - y γ -tubulina. Sin embargo, ellos comparten un patrón común de propiedades fisicoquímicas por residuo, que comprende residuos hidrofóbicos y un residuo con carga positiva. En β -tubulina estos corresponden a Met293, Met300, Met363 y Arg309, mientras que en γ -tubulina los residuos Ala258 y Arg343. La disposición de estos residuos armónicos es distinta en el dominio intermedio de las tubulinas, lo cual implicaría una diferencia en el reconocimiento de este dominio por parte de la chaperonina, aún más porque entre los residuos correlacionados de β -tubulina, se encuentran residuos con carga negativa como Asp355 y Glu328, y con carga positiva Lys350.

6.4.2 Pares de residuos correlacionados con alta probabilidad de contacto

Los pares de residuos correlacionados con alta probabilidad de estar en contacto (ver Figura 11), muestran que β -tubulina poseería interacciones de su dominio intermedio con las subunidades CCT δ y CCT ϵ además de una interacción de su dominio amino-terminal con la subunidad CCT γ . Por su parte γ -tubulina poseería interacciones del dominio intermedio con las subunidades CCT δ y CCT β , además de una interacción de su dominio amino-terminal con la subunidad CCT δ . Por lo tanto la interacción con la subunidad CCT δ es un patrón común que se presenta tanto en el complejo de β -tubulina/CCT como γ -tubulina/CCT, en el cual la mayoría son residuos hidrofóbicos.

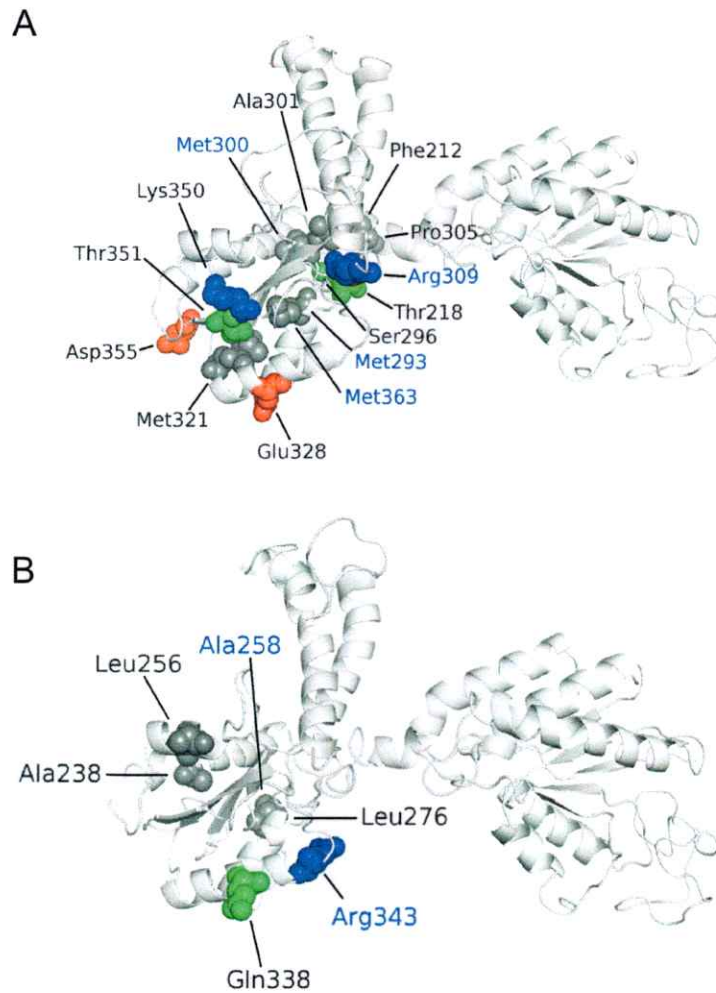


Figura 10. Residuos de β - y γ -tubulina posiblemente reconocidos por las subunidades CCT α y CCT δ .

Representación de cintas de las proteínas (A) β - y (B) γ tubulina parcialmente desplegadas en el interior de CCT con residuos determinados por CMA y SH con forma esférica por su radio de VDW en colores según propiedades fisicoquímicas con residuos de color rojo (carga negativa), color azul (carga positiva), color verde (polares) y gris (apolares y aromáticos). Se muestran los nombres de residuos correlacionados (letras en color negro) y residuos provenientes del análisis de SH, que se diferencia de BtubB, con probable cercanía física ($< 6\text{\AA}$) de los residuos de correlacionados (letras en color azul).

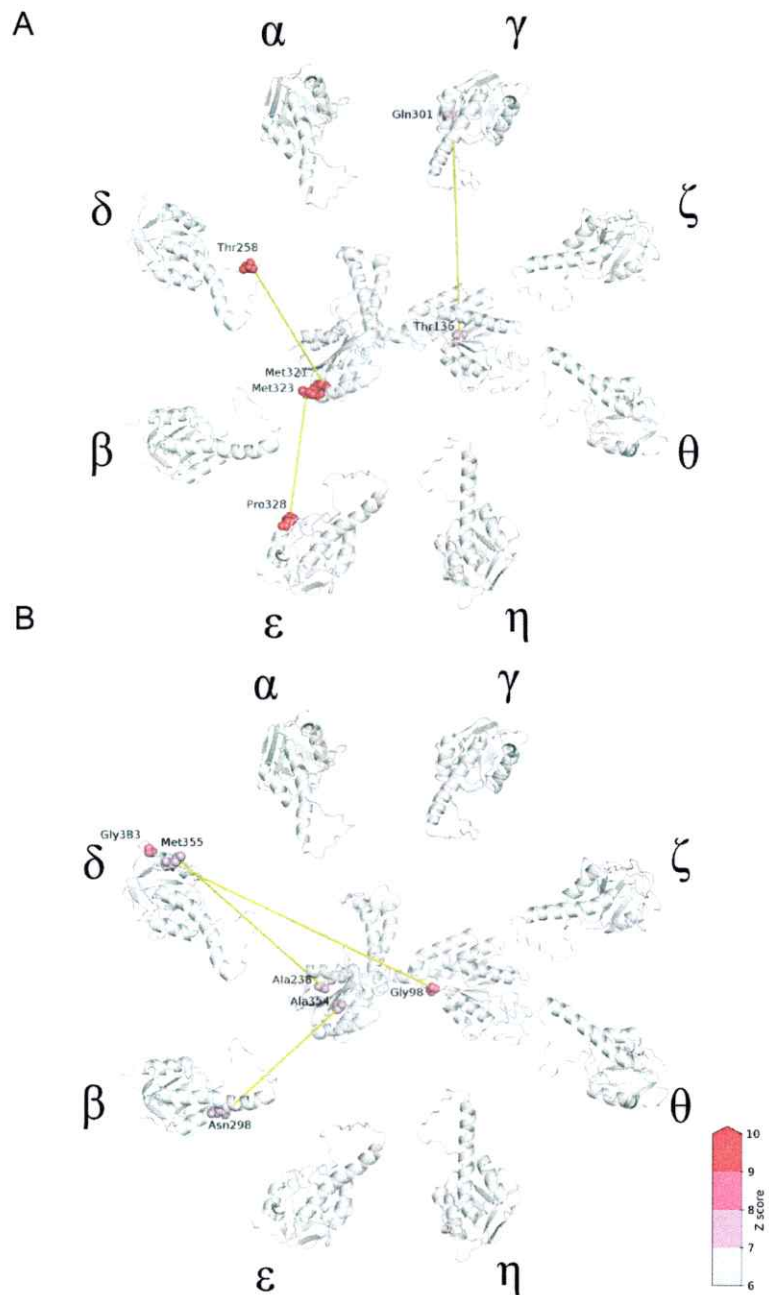


Figura 11. Residuos de la chaperonina CCT con alta probabilidad de contacto con β - y γ -tubulina.

Representación de cintas de las proteínas (A) β - y (B) γ -tubulina parcialmente desplegada y los dominios apicales de las diferentes subunidades de CCT. Se muestran los residuos del CMA con alta probabilidad de contacto (menor a 6 Å de distancia), destacados en forma esférica con colores según Z-score y las líneas amarillas representan los residuos correlacionados.

7 DISCUSIÓN

La predicción de mutaciones correlacionadas mediante el análisis de mutaciones correlacionadas (CMA) depende de la cantidad de secuencias que posea el alineamiento múltiple (MSA). Estos algoritmos necesitan extraer la información de los alineamientos y distinguir los residuos involucrados en las mutaciones correlacionadas de aquellos residuos que no lo están. Entre mayor es la cantidad de secuencias en el alineamiento, mayor será la probabilidad de encontrar mutaciones correlacionadas que den cuenta de un contacto físico real entre dos residuos.

Muchos de estos análisis se han hecho con alineamientos de proteínas que poseen sobre 125 secuencias ortólogas no redundantes (Dunn et al., 2008). Esto significa un problema en el uso de esta técnica en la mayoría de las familias de proteínas, debido a que generalmente sus alineamientos muestran una baja cantidad de secuencias no redundantes (Ashkenazy & Kliger, 2010). Entre estas proteínas se encuentran las tubulinas y las chaperoninas, las cuales son cruciales en los procesos de estructura celular y plegamiento de proteínas respectivamente, y poseen una alta conservación de sus residuos, lo que conlleva a obtener una baja cantidad de secuencias no redundantes en las bases de datos existentes.

Dado que los alineamientos múltiples de secuencia de las proteínas a estudiar poseen alrededor de 90 secuencias no redundantes, se hizo necesario estudiar el comportamiento del CMA como herramienta para encontrar residuos que coevolucionan y se relacionan por contacto físico a partir de MSAs con bajo número de secuencias.

Esto debido a que hay pocas investigaciones que hayan estudiado esta situación, donde se hayan usado los algoritmos MIp y PCCa (Dunn et al., 2008; Ashkenazy & Kligler, 2010).

Para evaluar la precisión de MIp o PCCa en la determinación de los pares de residuos identificados por mutaciones correlacionadas que tuvieran una cercanía física, se procedió a evaluarlos con alineamientos múltiples de secuencia de proteínas de la base de datos CDD con estructura conocida.

En otros trabajos se ha estimado que al usar el algoritmo MI, los contactos predichos para pares individuales con alineamientos de al menos 125 secuencias y con un Z-score mayor a 4, muestran un valor de distancia física promedio de los pares de residuos menor a 7 Å. Mientras que en la presente tesis, al usar el algoritmo MIp con un alineamiento que comprenden un rango de 75-99 o 50-74 secuencias, al seleccionar aquellos pares de residuos correlacionados con puntaje de Z-score mayores a 6 o 7, se obtuvo un promedio de distancia similar o menor a 10 Å, respectivamente. Por otra parte, en el caso de PCCa, solamente se obtuvieron promedios de distancia entre pares individuales similar y menor a 15 Å en todos los rangos de número de secuencias por alineamiento y valores Z-score estudiados.

Al usar el algoritmo MIp (Dunn et al., 2008) con alineamientos con al menos 125 secuencias, se obtuvo una fracción en contacto de 0.8 para los pares individuales. En la presente tesis, con alineamientos de entre 75-99 secuencias y seleccionando pares de residuos correlacionados con un Z-score mayor a 6 y 7, obtuvimos una fracción de contacto de alrededor de 0,7. En el caso de los pares individuales encontrados mediante

PCCa el valor obtenido es de 0,4.

Al considerar estos métodos para el análisis de los resultados obtenidos se llegó a la conclusión de que MIp posee un mejor comportamiento que PCCa al predecir pares correlacionados que están en cercanía física. También se concluye que al utilizar MIp, con menos secuencias por alineamiento se necesita aumentar el valor de Z-score para seleccionar pares de residuos correlacionados con similares distancias promedio y fracción de contacto que en estudios anteriores donde se usó MI y MIp (Martin et al., 2005; Dunn et al., 2008).

Esto demuestra que el uso del algoritmo MIp es robusto para calcular las mutaciones correlacionadas de los alineamientos de β - y γ -tubulina con el conjunto de subunidades de CCT.

Para dar mayor solidez a los resultados al realizar los análisis, se filtraron aquellas mutaciones correlacionadas que no tenían relación con los dominios apicales de las distintas subunidades de CCT. Esto se hizo debido a que el reconocimiento de las proteínas sustrato es principalmente mediado por el dominio apical de las diferentes subunidades de CCT, dado que resultados experimentales obtenidos mediante microscopía electrónica se observa un intermediario de tubulina parcialmente desplegado que interacciona con los dominios apicales de CCT, específicamente con la región más amplia de este dominio y con parte de la protuberancia helicoidal (Llorca et al., 2000). Esto también se apoya en el análisis de secuencia de CCT donde se observa que el domino apical posee la mayor cantidad de residuos-firma (Pappenberger et. al., 2002).

Con respecto a los resultados obtenidos, se puede observar que el punto de corte de Z-score para obtener mutaciones correlacionadas significativas es mayor en el caso de β -tubulina que en γ -tubulina, esto podría deberse al hecho que β -tubulina posee secuencias más conservadas, que en el cálculo con MIp produce que sus valores aumenten. Aun así, existe una mayor cantidad de pares correlacionados entre β -tubulina con los dominios apicales de las subunidades de CCT que las mostradas con γ -tubulina, lo que daría cuenta de una mayor especificidad de la interacción del complejo β -tubulina/CCT. Esto puede resultar del mayor número de copias de β -tubulina que γ -tubulina que hay en la célula (Shu & Joshi, 1995), por lo tanto habría un mayor frecuencia de interacciones de β -tubulina con CCT. Esto también se apoya por el hecho que γ -tubulina no necesita de un cofactor llamado prefoldina en la unión con CCT (Melki et al., 1993), indicando que tendría una interacción menos directa que con β -tubulina. Pero a pesar de estas diferencias, las dos tubulinas presentan un mismo patrón geométrico de unión con CCT observado por microscopía electrónica (Tesis Pouchucq, 2009).

De las múltiples mutaciones correlacionadas encontradas, se observó que los lugares donde existe una mayor densidad de pares correlacionados se concentran en el dominio intermedio de tubulina y en las subunidades δ y α de CCT. Este dominio intermedio se destaca por pertenecer a residuos involucrados tanto en las interacciones longitudinales como laterales, en la que se incluyen los “loops” T7 y M, respectivamente. Esto concuerda con el análisis de los “loops” ausentes en FtsZ que posee tubulina que podrían estar relacionados con su interacción con CCT (Andreu et al., 2002). Este permite suponer que los residuos involucrados en interacciones laterales

de β -tubulina se encontrarían correlacionados con residuos de la subunidad CCT δ , mientras que los residuos involucrados en interacciones longitudinales se encontrarían correlacionados con los residuos de subunidad CCT α . En cambio, las interacciones de γ -tubulina con CCT ocurrirían del modo contrario. Por lo tanto, esto entrega indicios de diferencias en el reconocimiento del dominio intermedio por CCT.

Ya que el análisis de mutaciones correlacionadas solamente nos permite observar correlaciones de mutaciones en residuos que son más variables en el alineamiento. Para investigar aquellos residuos más conservados que tendrían importancia en la funcionalidad de la proteína, considerando que la dependencia de la tubulina a la chaperonina CCT no se encuentra en sus homólogos en procariontes, se procedió a utilizar el análisis de armonía de secuencias, el cual se caracteriza por determinar residuos conservados entre grupos de proteínas dentro de una misma familia de proteínas.

Utilizando esta metodología de análisis de armonía de secuencias, se determinaron los residuos que comparten iguales propiedades fisicoquímicas entre β -tubulina y γ -tubulina y que se diferencian de la proteína homóloga BtubB, que es una de las tubulinas bacterianas que se conoce que no interacciona con algún homólogo de CCT (Pilhofer et al., 2011). Los resultados obtenidos serían reflejo de la incapacidad de las tubulinas eucariontes de plegarse espontáneamente, puesto que existe la presencia de residuos cargados en el dominio amino y carboxilo terminal, y la presencia de residuos hidrofóbicos en el dominio intermedio, los cuales son de características polares en la tubulina bacteriana BtubB. Ya que en el dominio intermedio se observan la mayor

cantidad de residuos correlacionados. Esto nos sugiere que estos residuos apolares de la región intermedia (compuesta principalmente de sabanas β) serían los que CCT estaría estabilizando para obtener intermediarios de plegamiento, sin que la tubulina se agregue en la célula. Lo cual estaría en concordancia con los experimentos de proteómica que indican que CCT reconoce proteínas enriquecidas en sábana β y con alta hidrofobicidad (Yam et al. 2008).

También se ocupó el análisis de armonía de secuencia para determinar los residuos firma, que corresponden a residuos conservados según propiedades fisicoquímicas que caracterizan a cada subunidad, confiriéndole heterogeneidad en la superficie de CCT (Pappenberger et al., 2002). En el presente trabajo se obtuvo una menor cantidad de residuos firma, pero similares en proporción a los obtenidos originalmente por Pappenberger et al. (2002). Esto era esperado debido a que se tomó en consideración el tipo de residuo según sus propiedades fisicoquímicas en el procedimiento de asignación de residuos-firma.

Así, se procedió a ocupar los residuos armónicos que comparten igual propiedades fisicoquímicas entre β -tubulina y γ -tubulina que se diferencian de la proteína homóloga BtubB y también los residuos armónicos que caracterizan a cada subunidad o residuos-firma de CCT, para representarlos gráficamente en sus respectivos modelos estructurales de tubulinas y subunidades de CCT, mostrándose sólo aquellos que se encuentran cercanos a los residuos del análisis de mutaciones correlacionadas. De esta forma se seleccionaron los residuos más conservados que tendrían relación con el reconocimiento de las tubulinas por CCT.

Mediante esta estrategia se observó un patrón común de residuos armónicos cercanos a pares correlacionados en las subunidades CCT δ y CCT α . En CCT δ , estos residuos tienen carga negativa, por lo tanto podrían interactuar con residuos con carga positiva del dominio intermedio de las tubulinas. Mientras que en la subunidad CCT α , estos corresponden a un residuo polar y uno con carga positiva, los cuales podrían interactuar con parte del dominio carboxilo terminal cercano al dominio intermedio, que se caracteriza por poseer residuos con carga negativa. Al observar patrones de residuos armónicos en tubulinas cercanos a pares correlacionados, no existe un patrón en común. Sin embargo comparten una misma característica, que comprenden residuos hidrofóbicos y un residuo con carga positiva. Su carga positiva podría tener relación con los residuos armónicos con carga negativa en la subunidad CCT δ , pero los residuos hidrofóbicos no poseen una contraparte en las subunidades de CCT. Esto quizás indicaría que existe una especificidad mediada por interacciones electrostáticas en una región de la proteína tubulina que posee alta hidrofobicidad comparado con la tubulina bacteriana BtubB.

La interacción del dominio intermedio con CCT δ se respalda aún más con los resultados obtenidos por el análisis de mutaciones correlacionadas con pares de residuos correlacionados con alta probabilidad de contacto físico, pues tanto en el complejo de β -y γ -tubulina con CCT existe un interacción con CCT δ .

Los residuos de la subunidad delta podrían explicar los resultados de los experimentos de unión entre la mutante de γ -tubulina, que posee la mutación M249E, con CCT, pues corresponde a un cambio de un residuo hidrofóbico por uno cargado

negativo en el “loop” T7, que podría generar un efecto de repulsión con la subunidad delta, explicando la disminución en la afinidad por CCT que disminuye un 70% la interacción con CCT (Tesis Pouchucq, 2009).

En el caso de la triple mutante llamada T7 β (M249L, Y245F y Y248Q) que consiste en cambio de estos residuos en el “loop” T7 de γ -tubulina por los residuos presentes en la misma localización en β -tubulina. La mutación Y248Q podría afectar la interacción con delta por la presencia del grupo polar pero no de una forma tan drástica como un cambio por un residuo con carga negativa. Esto es justamente lo que se observa experimentalmente, al disminuir un 40 % la interacción de la mutante de γ -tubulina T7 β con CCT con respecto a la interacción de la proteína WT con CCT.

Así, los resultados muestran patrones que indican que el reconocimiento del dominio intermedio de tubulina y el dominio apical de CCT, ocurrirían mediante interacciones electrostáticas tanto como hidrofóbicas, y muestra que existen otros factores, aparte de los “loops” que se encuentran en tubulinas eucariontes y que están ausentes en BtubB, que influyen en el plegamiento de la tubulina.

Una proyección de este trabajo sería estudiar las mutaciones correlacionadas que se obtuvieron del dominio ecuatorial e intermedio de la chaperonina con las tubulinas, para determinar los residuos importantes en el reconocimiento de los sustratos por el dominio ATPasa, que gatillan el cierre de la chaperonina y de esta forma comprender aún más el funcionamiento de esta maquinaria de plegamiento.

Los residuos encontrados *in silico*, que pueden tener una función en el plegamiento de tubulinas, debieran servir de base para diseñar experimentos que permitan corroborar

su participación en el plegamiento de estas proteínas, en especial mediante la construcción de nuevas mutantes de los residuos que han sido localizados durante la presente tesis.

8 CONCLUSIONES

Las conclusiones obtenidas de este trabajo son las siguientes:

- MIp posee una mayor precisión que PCCa para determinar pares de residuos relacionados con su cercanía física.
- MIp es un algoritmo el cual permite realizar el análisis de mutaciones correlacionadas con al menos 75 secuencias, lo cual define un nuevo punto de corte para la investigación de proteínas mediante este método.
- El análisis de mutaciones correlacionadas de las tubulinas con el dominio apical de las diferentes subunidades de CCT demuestran un claro patrón de mutaciones correlacionadas que se caracteriza por interacciones del dominio intermedio de las tubulinas con CCT δ y CCT α .
- El análisis de armonía de secuencias, da cuenta que los residuos hidrofóbicos del dominio intermedio de β - y γ -tubulina podrían impedir la estabilización de complejos desplegados que necesitarían ser estabilizados por CCT.
- Al observar los residuos armónicos cercanos a residuos involucrados en mutaciones correlacionadas, se puede observar patrones de tipos de residuos que indicarían un reconocimiento específico de la tubulina con la chaperonina mediante interacciones electrostáticas e hidrofóbicas.
- Se espera que la identificación de los residuos involucrados en el reconocimiento de tubulina por CCT sean de ayuda en la investigación de futuros experimentos, tanto del reconocimiento de los sustratos por el dominio ecuatorial de CCT *in*

silico y también mediante la construcción de nuevas mutantes en residuos que afecten la interacción entre tubulina y CCT *in vitro*.

9 BIBLIOGRAFÍA

Andreu, J. M., Oliva, M. A., & Monasterio, O. (2002). Reversible unfolding of FtsZ cell division proteins from archaea and bacteria. Comparison with eukaryotic tubulin folding and assembly. *The Journal of Biological Chemistry*, 277(45), 43262–70.

Araya, G. (2010). Caracterización de las mutantes γ TUB D279N y γ TUB Δ A283 en el “loop” M de γ -tubulina.

Archibald, J. M., Blouin, C., & Doolittle, W. F. (2001). Gene duplication and the evolution of group II chaperonins: implications for structure and function. *Journal of Structural Biology*, 135(2), 157–69.

Ashkenazy, H., & Kliger, Y. (2010). Reducing phylogenetic bias in correlated mutation analysis. *Access*, 12(12), 1–6.

Bertrand, S., Barthelemy, I., Oliva, M. A., Carrascosa, J. L., Andreu, J. M., & Valpuesta, J. M. (2005). Folding, Stability and Polymerization Properties of FtsZ Chimeras with Inserted Tubulin Loops Involved in the Interaction with the Cytosolic Chaperonin CCT and in Microtubule Formation. *J. Mol. Biol.*, 346, 319–330.

Brandt, B. W., Feenstra, K. A., & Heringa, J. (2010). Multi-Harmony: detecting functional specificity from sequence alignment. *Nucleic Acids Research*, 38, 35–40.

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)*, 25(15), 1972–3.

Choi, S. S., Li, W., & Lahn, B. T. (2005). Robust signals of coevolution of interacting residues in mammalian proteomes identified by phylogeny-aided structural analysis. *Nature Genetics*, 37(12), 1367–71.

Cong, Y., Baker, M. L., Jakana, J., Woolford, D., Miller, E. J., & Reissmann, S. (2010). mammalian chaperonin TRiC / CCT reveals its unique subunit arrangement. *October*, 4(4), 4967–4972.

Desai, A., & Mitchison, T. J. (1997). Microtubule Polymerization Dynamics. *Annual Review of Cell and Developmental Biology*, 13, 83–117.

Dunn, S. D., Wahl, L. M., & Gloor, G. B. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics*, 24(3), 333–40.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)*, 26(19), 2460–1.

Etchells, S. a, Meyer, A. S., Yam, A. Y., Roobol, A., Miao, Y., Shao, Y., Carden, M. J., Skach, W. R., Frydman, J., & Johnson, A. E. (2005). The cotranslational contacts between ribosome-bound nascent polypeptides and the subunits of the hetero-oligomeric chaperonin TRiC probed by photocross-linking. *The Journal of Biological Chemistry*, 280(30), 28118–26.

Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annual Review of Biochemistry*, 70, 603–47.

Gloor, G. B., Martin, L. C., Wahl, L. M., & Dunn, S. D. (2005). Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions. *Biochemistry*, 44(19), 7156–65.

Kalisman, N., Adams, C. M., & Levitt, M. (2012). Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8), 2884–9.

Kapatai, G., Large, A., Benesch, J. L. P., Robinson, C. V., Carrascosa, J. L., Valpuesta, J. M., Gowrinathan, P., & Lund, P. a. (2006). All three chaperonin genes in the archaeon *Haloferax volcanii* are individually dispensable. *Molecular Microbiology*, 61(6), 1583–97.

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–66.

Kollman, J. M., Merdes, A., Mourey, L., & Agard, D. a. (2011). Microtubule nucleation by γ -tubulin complexes. *Nature Reviews. Molecular Cell Biology*, 12(11), 709–721.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, 19, 1639–1645.

Llorca, O., Martín-Benito, J., Gómez-Puertas, P., Ritco-Vonsovici, M., Willison, K. R., Carrascosa, J. L., & Valpuesta, J. M. (2001). Analysis of the Interaction between the Eukaryotic Chaperonin CCT and its Substrates Actin and Tubulin. *Journal of Structural Biology*, 135(2), 205–18.

Llorca, O., Martín-Benito, J., Ritco-Vonsovici, M., Grantham, J., Hynes, G. M., Willison, K. R., Carrascosa, J. L., & Valpuesta, J. M. (2000). Eukaryotic chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations. *EMBO Journal*, 19(22), 5971–5979.

Llorca, O., McCormack, E. a, Hynes, G., Grantham, J., Cordell, J., Carrascosa, J. L., Willison, K. R., Fernandez, J. J., & Valpuesta, J. M. (1999). Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. *Nature*, 402(6762), 693–6.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., ... Bryant, S. H. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, 39(Database issue), D225–9.

Martin, L. C., Gloor, G. B., Dunn, S. D., & Wahl, L. M. (2005). Using information theory to search for co-evolving residues in proteins. *Bioinformatics (Oxford, England)*, 21(22), 4116–24.

Melki, R., Vainberg, I. E., Chow, R. L., & Cowan, N. J. (1993). Chaperonin-mediated folding of vertebrate actin-related protein and gamma-tubulin. *The Journal of Cell Biology*, 122(6), 1301–10.

Muñoz, I. G., Yébenes, H., Zhou, M., Mesa, P., Serna, M., Park, A. Y., Bragado- Nilsson, E., Beloso, A., de Cárcer, G., Malumbres, M., Robinson, C. V, Valpuesta, J. M., & Montoya, G. (2011). Crystal structure of the open conformation of the mammalian chaperonin CCT in complex with tubulin. *Nature Structural & Molecular Biology*, 18(1), 14–9.

Nogales, E. (2001). Structural Insights into Microtubule Function. *Annual Review of Biophysics and Biomolecular Structure*, 30, 397–420.

Nogales, E., Downing, K. H., Amos, L. A., & Löwe, J. (1998). Tubulin and FtsZ form a distinct family of GTPases. *Nature*, 5(6), 451–8.

Pappenberger, G., Wilsher, J. A., Roe, S. M., Counsell, D. J., Willison, K. R., & Pearl, L. H. (2002). Crystal Structure of the CCTg Apical Domain: Implications for Substrate Binding to the Eukaryotic Cytosolic Chaperonin. *Journal of Molecular Biology*, 318(02), 1367–1379.

Pazos, F., Helmer-Citterich, M., Ausiello, G., & Valencia, a. (1997). Correlated mutations contain information about protein-protein interaction. *Journal of Molecular*

Biology, 271(4), 511–23.

Pilhofer, M., Ladinsky, M. S., McDowall, A. W., Petroni, G., & Jensen, G. J. (2011). Microtubules in Bacteria: Ancient Tubulins Build a Five-Protofilament Homolog of the Eukaryotic Cytoskeleton. *PLoS Biology*, 9(12), e1001213.

Pirovano, W., Feenstra, K. A., & Heringa, J. (2006). Sequence comparison by sequence harmony identifies subtype-specific functional sites. *Nucleic Acids Research*, 34(22), 6540–8.

Pollock, D., & Taylor, W. (1997). Effectiveness of correlation analysis in identifying protein residues undergoing correlated evolution. *Protein Engineering*, 10(6), 647–657.

Pouchucq, L. (2009). Análisis estructural y funcional de las regiones de interacción de γ -tubulina en la nucleación de microtúbulos y el plegamiento asistido por la chaperonina CCT.

Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3), 779–815.

Shen, M.-Y., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science : A Publication of the Protein Society*, 15(11), 2507–24.

Shu, H. B., & Joshi, H. C. (1995). Gamma-tubulin can both nucleate microtubule assembly and self-assemble into novel tubular structures in mammalian cells. *The Journal of Cell Biology*, 130(5), 1137–47.

Söding, J. (2005). Sequence analysis Protein homology detection by HMM – HMM comparison. *Bioinformatics*, 21(7), 951–60.

Spiess, C., Meyer, A. S., Reissmann, S., & Frydman, J. (2004). Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends in Cell Biology*,

14(11), 598–604.

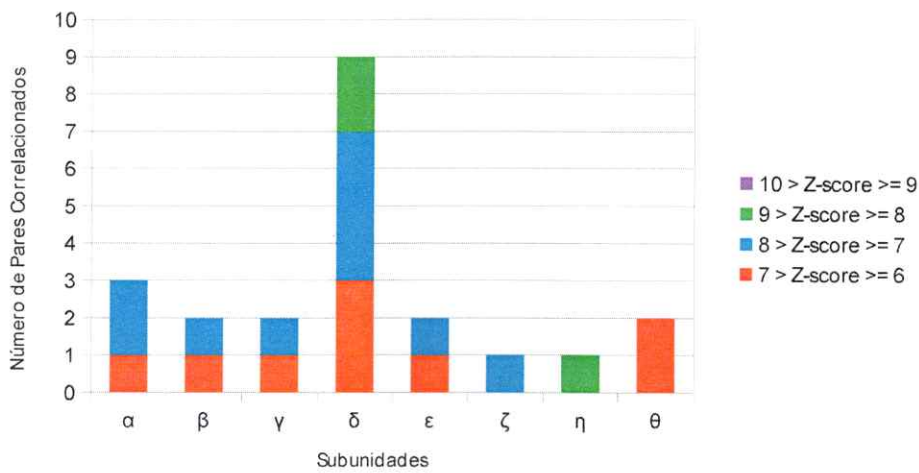
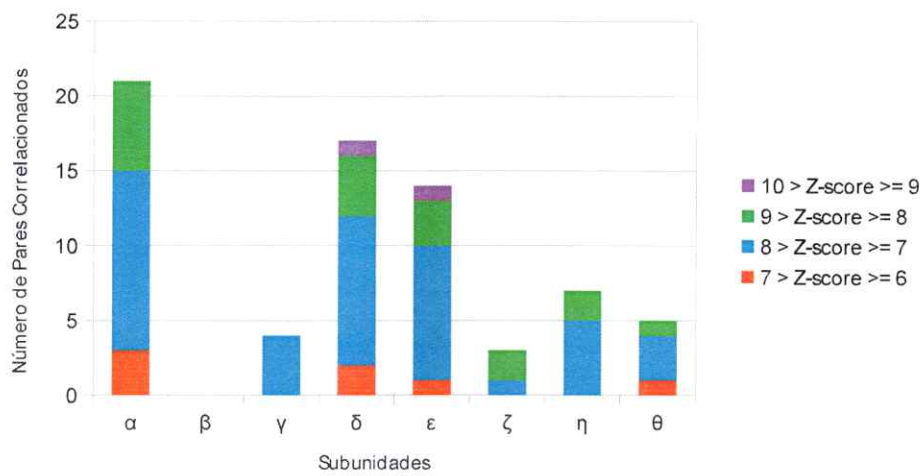
Spiess, C., Miller, E. J., McClellan, A. J., & Frydman, J. (2006). Identification of the TRiC/CCT substrate binding sites uncovers the function of subunit diversity in eukaryotic chaperonins. *Molecular Cell*, 24(1), 25–37.

Tam, S., Geller, R., Spiess, C., & Frydman, J. (2006). The chaperonin TRiC controls polyglutamine aggregation and toxicity through subunit-specific interactions. *Nature Cell Biology*, 8(10), 1155–62.

Valpuesta, J. M., Martín-Benito, J., Gómez-Puertas, P., Carrascosa, J. L., & Willison, K. R. (2002). Structure and function of a protein folding machine: the eukaryotic cytosolic chaperonin CCT. *FEBS Letters*, 529(1), 11–6.

Yam, A. Y., Xia, Y., Lin, H. J., Burlingame, A., Gerstein, M., & Frydman, J. (2008). Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nature Structural & Molecular Biology*, 15(12), 1255–62.

10 ANEXOS



Anexo 3: Gráficos de la cantidad de mutaciones correlacionadas que poseen las tubulinas investigadas con las subunidades de la chaperonina.

Gráficos que muestran la cantidad de mutaciones correlacionadas que posee β -tubulina (A) y γ -tubulina (B) con las diferentes subunidades de CCT, distinguiéndose el valor de Z-score que poseen estas mutaciones correlacionadas.

Armonía de Secuencia	Z-score	Consenso			Referencias		
		β -Tubulina	γ -Tubulina	BtubB	β -Tubulina Humana	γ -Tubulina Humana	BtubB <i>Prostheco bacter dejongeii</i>
0.216	-6.859	L	EG	N	L44	T45	N45
0.155	-8.142	E	D	AS	E45	D46	S46
0.115	-10.505	D	E	S	D88	E89	S89
0.076	-9.505	F	LIV	R	F92	L93	R93
0.444	-8.426	Y	Y	HN	Y106	F108	N107
0.552	-6.306	D	D	N	D118	D120	N119
0.325	-8.473	RK	R	AS	K122	R124	S123
0.414	-11.510	E	E	A	E123	E125	A124
0.410	-12.906	D	D	K	D128	D130	K129
0.231	-10.098	E	D	Q	E157	D159	Q158
0.177	-7.281	IM	L	R	I163	L165	R164
0.195	-8.648	G	A	S	G235	A238	S235
0.091	-9.163	V	A	T	V255	A258	T255
0.321	-8.870	R	R	G	R262	R265	G262
0.259	-12.680	L	LC	N	L263	L266	N263
0.108	-8.038	Y	V	Q	Y281	V285	Q281
0.149	-9.746	M	L	T	M293	L297	T293
0.464	-9.185	M	M	T	M300	M304	T300
0.254	-6.260	R	KC	V	R309	C316	V309
0.264	-7.639	K	R	ES	K336	R343	S336
0.172	-7.279	M	V	S	M363	V373	S363
0.263	-7.705	R	K	D	R390	K400	D390
0.387	-7.237	R	R	N	R391	R401	N391
0.432	-9.897	L	L	T	L395	L405	T395
0.318	-6.429	E	E	NA	E401	E411	A401
0.477	-7.935	E	E	QM	E407	E420	M407
0.415	-12.404	E	E	S	E421	E434	S421
0.258	-6.461	A	SADEP	-	A428	P441	-427
0.053	-8.663	VA	Y	-	A430	Y443	-427

Anexo 4: Residuos armónicos que comparten β -tubulina y γ -tubulina que se diferencian de la tubulina Bacteriana BtubB.

La primera columna corresponde al puntaje obtenido de los residuos armónicos del alineamiento hecho con las secuencias β -tubulina, γ -tubulina y BtubB, donde β -tubulina y γ -tubulina poseen residuos con propiedades fisicoquímicas (columnas 3 y 4) diferentes que BtubB (columna 5) indicadas con colores: rojo (carga negativa), azul (carga positiva), verde (polar), hidrofóbico (negro). La segunda columna corresponde al valor de Z-score que posee el puntaje de residuos correlacionados. Las tres últimas columnas corresponden al nombre y número de residuo al cual corresponde cada residuo armónico en β - y γ -tubulina Humana y BtubB de *Prostheco bacter dejongeii*, respectivamente.

Dominio	TCPA	TCPB	TCPG	TCPD	TCPE	TCPH	TCPQ	TCPZ
Equatorial	V39	D17	K21	P30	D65	Q21	R31	T43
	K84	E18	G56	S136	Q86	K109	Q34	K55
	D105	T103	Q116	Q138	R132	K135	E38	K58
					D153		N50	T73
							N53	S75
							M88	K79
							I121	I87
							E127	T88
								R117
								K127
Intermedio	P191	D215	C173	S181	G213	E156	Q172	E192
			K181	D207		C158	C187	H198
			K197			G196		
			I208			K198		
Apical	C218	T242	R230	D261	K242	E243	T227	G215
	K243	T246	R260	E278	D245	A248	E228	R217
	R266	E268	E261	C295	C302	E249	V231	Y229
	C296	N298	P287	D327	E309	D251	I250	T232
	K298	P300	E293	T342	Q316	E254	G294	G247
	E350	E301	T310		P320	P294	N316	R256
	E360	E313	T318		Q360	R306	P338	R279
	T371	H314	N320		K368	C310	H365	P299
		G319	N321			G312		D303
		V341	Y359			E316		R319
Intermedio		E369	L403	C414		A378	D393	Q386
		A398		K418			V401	D389
							T403	R392
								R396
Equatorial	E440	C412	K425	P428	C493	T519	K440	K426
	Q458	E432	P436	C450	Q522	P523	Y468	A433
	K512	E437	C455	N494			T500	Q434
	H534	R444		L507				Q467
		E512		Q510				A468
				S515				E469
								P486
								A489
							V510	
							L517	

Anexo 5: Residuos firma de las diferentes subunidades de CCT clasificados por el dominio al cual pertenecen.