



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE MEDICINA

Secuencias y estructuras de la Triosa fosfato Isomerasa en Arqueas: Análisis con enfoque evolutivo

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
LICENCIADA EN INVESTIGACIÓN BIOMÉDICA BÁSICA

P R E S E N T A:
ANDREA HERNÁNDEZ GARCÍA

DIRECTOR DE TESIS:
DR. DANIEL ALEJANDRO FERNÁNDEZ VELASCO

CIUDAD UNIVERSITARIA, CD. MX. SEPTIEMBRE DE 2022



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

In memoriam

Dedicada a mi padre Miguel Hernández Lecona

El entendimiento suscita aprecio y el aprecio suscita amor. El amor es lo único que se puede entender, ya que sólo el amor es real y, por lo tanto, sólo el amor tiene sentido.

~ UCDM

Tu eres una criatura del universo
No menos que las plantas y las estrellas
¡Tienes derecho a existir!
Sea que te resulte claro o no
Indudablemente el universo marcha como debería
-DESIDERATA-Max Ehrmann

Every day we live a miracle
Unpredictable
You don't need an upgrade anymore
Can't you see the link?
Don't worry
I will teach you how to take the pill to feel the thrill and touch it all!
-MIRACLE-Caravan palace

Que sepa el mundo que en marcha estoy
Que voy a cumplir mi misión
Los cielos azules por donde voy
Dan alas a mi corazón,
Diles todos que en marcha estoy
Que voy a volver a mi hogar
Con el sol más brillante mis pasos doy
Disfruto el camino al andar
- EN MARCHA ESTOY- Phill Collins

AGRADECIMIENTOS ACADÉMICOS.

A la Universidad Nacional Autónoma de México (UNAM) que me ha permitido cursar desde mis estudios de preparatoria en sus aulas, gracias por proporcionarme un espacio de alta calidad académica donde he podido crecer tanto académica como personalmente.

A la Licenciatura en Investigación Biomédica Básica que me permitió experimentar el día a día del quehacer científico y afirmar mi amor por el conocimiento. Me proporciono una excelente preparación en ciencias, además de permitirme conocer a personas extraordinarias, apasionadas con su quehacer científico y sobre todo buenos seres humanos.

A mi tutor de primer año Dr. Sebastián Poggio Ghilarducci, por su paciencia y dedicación en mi formación. A Ana Laura y Thelma que fueron el equipo perfecto para que confirmará que esta carrera era mi pasión.

A mi tutor de segunda rotación Dr. David Erasmo García que me permitió desarrollar de manera muy libre el proyecto de investigación, lo que me enseñó la gran responsabilidad y compromiso que requiere esa libertad.

A la Licenciada Cynthia Lima que se encargó de todo lo administrativo a lo largo de toda mi carrera en la LIBB y que me ha permitido llegar hasta este punto.

A mi tutor Dr. Alejandro Fernández Velasco por su gran apoyo y paciencia desde que empecé la licenciatura hace casi 5 años. También por su gran papel como asesor de tesis, proyecto que empezó hace casi 3 años.

Al Dr. Alejandro Zentella Dehesa quien también ha seguido mi formación académica desde inicios de la licenciatura, me ha acompañado y enseñado muchísimo, también le doy las gracias por permitirme enseñar con él, que ha sido una experiencia super enriquecedora.

Al M. C. Ricardo Muñoz Trejo por enseñarme cosas del laboratorio con gran amabilidad y entrega, espero haber sido una buena “hija” académica. También gracias por tu esfuerzo en trabajos previos para poder formular esta tesis

Al Dr. Sergio Romero Romero por la resolución del cristal, sus buenos comentarios en la discusión de este proyecto y por ser un gran ejemplo de vida en el que inspirarse.

A la Dra. Isabel Velázquez López que me ha enseñado las técnicas del laboratorio y ha sido un gran apoyo con todos los trámites administrativos.

A Erendira por ser una buena compañera del laboratorio que me escucho siempre y me enseñó la Técnica de Cytoscape y RING.

Al Laboratorio de Fisicoquímica e Ingeniería de Proteínas (LFQIP) que me ha proporcionado un lugar espléndido para el desarrollo de esta tesis, y todos sus integrantes han dejado una huella en mí.

El desarrollo de este proyecto fue financiado por la Facultad de Medicina de la UNAM y por los proyectos PAPIIT IN218222. Recibí beca para la realización de tesis de los proyectos de PAPIIT IN218222 e IN219519.

A los miembros de mi jurado por sus excelentes comentarios y sugerencias en el trabajo de esta tesis

AGRADECIMIENTOS PERSONALES

A mis padres Miguel Hernández Lecona y a Maria Antonieta García Morales gracias a ustedes soy lo que soy.

Gracias Miguel por inculcarme tu amor por el conocimiento. Por esas pláticas sobre los misterios del mundo que duraban horas. Por enseñarme que cuando uno está en el TODO no existe la tristeza ni la soledad. Por mostrarme la fortaleza de mi ser, en tu partida me has enseñado que soy una guerrera en todo el esplendor. Gracias por ser un padre amoroso y siempre hacer lo mejor que podías con las armas que tenías.

Gracias Maria Antonieta por inculcarme el orden. Por enseñarme que todos los actos, absolutamente todos, son actos de amor. Por mostrarme mi independencia tanto física como emocional. Gracias por ser una gran maestra en cada paso de mi vida, por ser una madre amorosa y siempre hacer lo mejor que puedes con las armas que tienes.

Gracias Jimena por ser siempre un espejo y un apoyo incondicional. Un cómplice, una amiga y una especie de pepe el grillo en mi vida. Gracias por crecer conmigo y acompañarme en tantas aventuras. Por ser una hermana amorosa y siempre hacer lo mejor que puedes con las armas que tienes.

A mis abuelos José García Villalva y María Antonieta Morales Hernández por ser un sostén en mi vida, un lugar de anclaje desde donde continuar. Por siempre proveer a nuestra familia con sus cuidados y amor. Por mostrarme el valor de la familia y sus sabios consejos.

תודה לבני אליאו על שעזר לי לרפא את חיי, שלימד אותי את אכזריות החושך שיש לי, על כך שקירב אותי לנשיות שלי ומצא את הכלים שחסרו לי כדי להיות בן אדם מאושר ומספק יותר.

To Elisha for my virtues uncounted, all my goodness is going with you now

A todos mis amigos, a Mike y Mario por los grandes momentos y risas que hemos tenido. A Vale, Luz, Sam y Adriana por ser grandes amigos a lo largo de toda la licenciatura, ser siempre confidentes y un gran apoyo durante todo este proceso. Gracias por esas tardes de baile, también por ser una fuente de inspiración y admiración. A Gerardo, Adrian y Sebastián por grandes compañeros y una fuente grande de inspiración y admiración.

A mi tutor Dr. D. Alejandro Fernández Velasco por siempre apoyarme en momentos difíciles y su gran paciencia y entendimiento.

Al Dr. Zentella por sus sabios consejos y ser un amigo incondicional.

A Octavio Avilés por las pláticas tan locas que tenemos, por ser un gran apoyo y enseñarme día a día tantas cosas. Gracias por tu amor.

A mis ancestros por permitirme estar aquí hoy y ser lo que soy

A la Akademia CDMX por permitirme conocer a personas tan maravillosas como Luis Fernando, Citlali y Emilio, por enseñarme también tantas herramientas de autoconocimiento que aplicar a mi vida diaria.

ÍNDICE

ÍNDICE DE FIGURAS Y TABLAS	10
LISTADO DE ABREVIATURAS	13
RESUMEN	15

INTRODUCCIÓN

1. Organismos extremófilos	17
2. APROXIMACIONES	
2.1 Aproximación bioquímica	20
2.2 Aproximación evolutiva / secuencial	25
▪ 2.2.1 Patrones evolutivos	26
▪ 2.2.2 Epistasis	28
2.3 Aproximación estructural	34
3. LIMITANTES	
3.1 Limitantes bioquímicas	38
3.2 Limitantes evolutivas	40
3.4 Limitantes estructurales	42
• Epílogo	43
4. MODELO DE ESTUDIO	
4.1 El plegamiento (β/α)8	45
4.2 Enzima Triosafosfato Isomerasa (TIM)	47
4.3 Las TIMs de Arqueas	51
5. ANTECEDENTES	
5.1 Antecedentes Bioquímicos	53
5.2 Antecedentes Evolutivos	63
5.3 Antecedentes Estructurales	67
HIPÓTESIS	71
OBJETIVOS	71

MÉTODOS

• Servidor PISA (Protein Interfaces Surfaces and Assemblies).....	72
1. RING	76
2. ALPHAFOLD 2	78
3. HSYMDOCK	81
4. UCSF CHIMERA	85
5. PDB2PQR & APBS	86

6. Análisis estadísticamente acoplado (SCA)	88
---	----

RESULTADOS

• Caracterización de los homotetrámeros	104
• Comparación entre los tetrámeros y el dímero	110
• Diferencias entre MjaTIM y MfoTIM	114
• Determinación de modelos mediante AlphaFold2	120
• Comparación de las superficies electrostáticas	125
• Comparación de las superficies hidrofóbicas	130
• Análisis estadísticamente acoplado (SCA)	134

DISCUSIÓN	150
CONCLUSIONES	155
PERSPECTIVAS	156
REFERENCIAS	157

ÍNDICE DE FIGURAS Y TABLAS

Figura	Tabla	Título	Página
1		Mecanismos utilizados por organismos extremófilos para sobrevivir en ambientes extremos.	17
2		Paisaje conformacional en forma de embudo.	21
3		Ejemplo de una curva de estabilidad.	22
4		Calorimetría diferencial de barrido	23
5		Estrategias fisicoquímicas que utilizan las proteínas para incrementar la termoestabilidad.	24
6		Espacio secuencial de una proteína	25
7		Técnicas para conocer los patrones evolutivos en las secuencias	27
8		Fenómeno de amurallamiento (entrenchment)	30
9		Fenómeno de Epistasia	32
10		Aumento de cargas electrostáticas en la superficie de proteínas termófilas.	35
11		Mecanismos <i>in vivo</i> para aumentar la termoestabilidad	39
12		Ilustración de la historia de los ciegos y el elefante	43
13		Plegamiento del Barril (β/α) ₈ TIM	45
14		Sustrato y producto de la enzima Triosafofato Isomerasa	47
15		Aminoácidos catalíticos de la TIM	48
16		Mecanismo de catálisis de la TIM	49
17		Conformación abierta y cerrada de la TIM	49
18		Conformación dimérica de la TIM	50
19		La Triosafofato Isomerasa de Arquea	52
20		Interfaz tetramérica de las TIMs de Arquea	52
21		Temperatura de crecimiento de las especies seleccionadas	53
	1	Caracterización espectroscópica de las TIMs de Arquea	54
22		Dicroísmo circular de las MesoTIMs, TermoTIMs, PfuTIM y TonTIM	55
	2	Estado oligomérico de las TIMs de Arquea	56-57
23		Ensayos catalíticos de las TIMs de Arquea	58
	3	Ensayos enzimáticos reportados para las TIMs de Arquea	59
24		Mecanismos extrínsecos de estabilización de las TIMs de Arquea	60
25		Calorimetría diferencial de barrido de las TermoTIMs	61
26		Características biofísicas de la TIM en relación con la filogenia de los Methanococcales	64
27		Relación entre la temperatura óptima de crecimiento y la filogenia de los Methanococcales	66
28		Cristal de la TIM de <i>Pyrococcus furiosus</i>	67
	4	Datos cristalográficos de las TIMs de Arqueas publicadas en el PDB.	69
29		Datos cristalográficos necesarios para el cálculo de PISA	73

	5	Criterios geométricos para la asignación de interacciones en RING 2.0	76-77
30		Flujo de trabajo de AlphaFold2	78
31		Flujo de trabajo de HSYMDOCK	81
32		Mediciones CAPRI para evaluar el docking de homoligomeros simétricos	82
33		Proceso iterativo del cálculo del puntaje de HSYMDOCK	83
34		Puntaje de HSYMDOCK	84
	6	Escala de hidropatía utilizada por UCSF Chimera	85
35		Flujo de trabajo de los servidores PDB2PQR & APBS	87
36		Divergencia de Kullback-Leibler	90
	7	Frecuencia de los aminoácidos proveniente de una base de datos no redundante	90
37		Relación entre la conservación, frecuencia y entropía relativa	91
38		Definición de Eigenvalores y Eigenvectores	93-94
39		Determinación de los Eigenvectores y Eigenvalores	97
40		Espectro de los eigenvectores de la matriz SCA	98
41		Gráfica de dispersión de los elementos de los eigenvectores	99
42		Identificación de los sectores por parte del SCA	101
43		Relación entre el alineamiento múltiple de secuencias y las matrices de covarianza	101-102
44		Cristal de MfoTIM	104
	8	Valores del volumen total y el área de las interfaces dimérica y tetramérica de los cristales de las ArqTIMs.	105
	9	Aminoácidos involucrados en la interfaz tetramérica de las TIMs de Arquea	106
45		Red de interacciones de la interfaz tetramérica	107-108
46		Diferencias entre TacTIM y MjaTIM	110-111
47		Diferencias en secuencias entre Methanococcales y Thermoplasmatales	112
48		Diferencias entre la superficie electrostática de MjaTIM y MfoTIM	115
	10	Diferencias entre los residuos de MjaTIM y MfoTIM	115
49		Diferencias entre los residuos de MjaTIM y MfoTIM	117-119
50		Calidad de los modelos obtenidos mediante AlphaFold2	120-121
	11	Resultados de HSYMDOCK	122
	12	Aminoácidos involucrados en la interfaz tetramérica de los modelos obtenidos por ALPHAFOLD2	123-124
51		Superficies electrostáticas de las ArqTIMs Hipertermófilas	125
52		Superficies electrostáticas de las ArqTIMs Termófilas	126
53		Superficies electrostáticas de las ArqTIMs Mesófilas	127
54		Gráfica de dispersión de las cargas netas de las ArqTIMs en relación a la Temperatura óptima de crecimiento	128

55		Superficies hidropáticas de las ArqTIMs Hipertermófilas	130
56		Superficies hidropáticas de las ArqTIMs Termófilas	131
57		Superficies hidropáticas de las ArqTIMs Mesófilas	132
	13	Aminoácidos con los valores Di más altos	134-135
58		Histogramas Di	135
59		Matrices de correlación SCA de los tres dominios de la vida	136
60		Eigenespectros de los tres dominios de la vida	137
61		Matrices de correlación limpias y determinación de los Sectores	138-139
	14	Aminoácidos involucrados en los Sectores de los tres dominios de la vida.	140
62		Sectores del dominio Arquea	141-142
63		Sectores del dominio Eucaria	144
64		Sectores del dominio Bacteria	145-146
	15	Código de colores para los sectores y para el orden filogenético de las Arqueas	147
65		Relación entre la filogenia del dominio Arquea y sus Sectores.	148-149
66		Resultados de la cromatografía de exclusión analítica de las TermoTIMs & MbuTIM.	152
67		Geles de purificación de la PKG de Pfu y las MesoTIMs	153

LISTADO DE ABREVIATURAS

Topt	Temperatura óptima de crecimiento
aa	Aminoácidos
ArqTIMs	TIMs de Arqueas
TIM	Triosafosfato Isomerasa
(β/α) ₈	Beta / alfa ocho
$\beta\alpha$	alfa / beta
$\alpha\beta$	beta / alfa
Tm	Temperatura media de desnaturalización
SCA	Análisis estadísticamente acoplado
pH	Potencial de hidrógeno
OH-	Hidróxido
ADP	Adenosín difosfato
ATP	Adenosín trifosfato
ADN	Ácido desoxirribonucleico
PCR	Reacción en cadena de la polimerasa
ΔG	Energía libre de Gibbs
Td	Temperatura de desnaturalización
Ts	Temperatura de máxima estabilidad
MSA	Alineamiento de secuencias múltiples
DCS	Calorimetría diferencial de barrido
CDB	Calorimetría diferencial de barrido
Cp	Capacidad calorífica
ΔH	Diferencia en entalpía
ΔC_p	Diferencia en la capacidad calorífica
DCA	Análisis de acoplamiento directo
GAPDH	Gliceraldehído-3-fosfato deshidrogenasa

Hsp	Proteína de choque térmico
G3P	Gliceraldehído 3 fosfato
DHAP	Dihidroxiacetona fosfato
Kcat	Constante catalítica
Km	Constante de Michaelis-Menten
Vmax	Velocidad máxima
NADH	Nicotamida adenina dinucleótido
GLPDH	Glutamato deshidrogenasa
PISA	Servidor de superficies, interfaces y ensamblajes de proteínas
PDB	Banco de datos de proteínas
MjaTIM	TIM de <i>Methanocaldococcus jannaschii</i>
MfoTIM	TIM de <i>Methanotorris formicicus</i>
MokTIM	TIM de <i>Mathanothermococcus okinawensis</i>
MaeTIM	TIM de <i>Methanococcus aelicus</i>
MmaTIM	TIM de <i>Methanococcus maripaludis</i>
MbuTIM	TIM de <i>Methanococcoides burtonii</i>
MfeTIM	TIM de <i>Methanothermus fervidus</i>
TtxTIM	TIM de <i>Thermoproteus tenax</i>
MbrTIM	TIM de <i>Methanococcoides burtonii</i>
TacTIM	TIM de <i>Thermoplasma acidophilum</i>

RESUMEN

En la tierra los seres vivos han logrado sobrevivir a temperaturas que desde nuestro punto de vista parecen extremas. Para la ciencia sigue siendo un misterio como estos organismos procariontes extremófilos han conquistado prácticamente todos los rincones del planeta (Coker J., 2019). Dado el papel central de las proteínas en la bioquímica de la vida, una de las aproximaciones clásicas para el estudio de la termoestabilidad es la comparación de estructuras proteicas provenientes de organismos termófilos contra sus homólogos mesófilos. Gracias a este método, se ha identificado que las proteínas termófilas tienden a tener más residuos cargados en su superficie y a aumentar su empaquetamiento, escondiendo más área hidrofóbica (Brininger, Cobani & Evilia 2018).

La triosafosfato isomerasa (TIM) tiene un plegamiento $(\beta/\alpha)_8$ y es un excelente modelo para el estudio de la evolucionabilidad y termoestabilidad debido a su robustez y a su conservación en los tres dominios de la vida. Las TIMs bacterianas y eucariontes son homodímeros con ~250 aminoácidos (aa) de longitud, mientras que las TIMs de arqueas (ArqTIMs) son dímeros, pero algunas pueden tetramerizar, además de ser 10% más pequeñas con ~225 aa (Sheramm, Kohlhoff & Hensel, 2001). La oligomerización tetramérica se ha propuesto como mecanismo adaptativo a la temperatura. Pero aún no existe evidencia que respalde dicha hipótesis, por lo que el objetivo de este trabajo es encontrar si realmente existe una relación entre la oligomerización y la termoestabilidad en las TIMs de Arqueas.

Al analizar las estructuras cristalográficas disponibles de las ArqTIMs tetraméricas observamos que la interfaz tetramérica se forma por interacciones hidrofóbicas de aminoácidos pequeños pertenecientes a las hélices 4, 5 y 6. Las principales diferencias entre los tetrámeros y el único cristal de una ArqTIM dimerica muestra una substitución de los residuos hidrofóbicos pequeños por residuos aromáticos y cargados en la principal zona de interacción (hélice 5). Nos preguntamos si estas características del dímero estaban presentes en otras secuencias de ArqTIMs, sorprendentemente encontramos su presencia en acidófilos y no en organismos mesófilos como esperaríamos si la hipótesis adaptativa fuera cierta.

En nuestro grupo de investigación se han caracterizado fisicoquímicamente dos TIMs de arqueas termófilas: la de *Methanocaldococcus jannashi* (MjaTIM) y la de *Methanotorris formicicus* (MfoTIM), ambas son tetrámeros en solución y tienen temperaturas medias de desnaturalización (T_m) de 111.6°C y 91.7°C, respectivamente (Muñiz, 2020). La diferencia en la T_m es de casi 20°C, obviamente la diferencia no se debe al estado de oligomérico, así que la diferencia debe ser estructural. Por esta razón, Sergio Romero resolvió la estructura de MfoTIM. Al comparar ambas estructuras logramos notar dos cambios importantes: el primero es que la

superficie electrostática de MjaTIM está cargada más negativamente, y la segunda es que MjaTIM esconde mayor área al formar la interfaz tetramérica.

Por último, realizamos un análisis estadísticamente acoplado (SCA en inglés) en alineamientos múltiples de ArqTIMs y TIMs de bacterias y eucariontes para encontrar aminoácidos que coevolucionan y tratar de rastrear los orígenes de la oligomerización. Resulta que la principal diferencia entre las secuencias de ArqTIMs con las TIMs bacterianas y eucariontes es una delección ~20aa en la hélice 5 de la TIM, que casualmente la principal zona de interacción del tetrámero, además, encontramos dos sectores (red de aa coevolutivos) asociados a la interfaz tetramérica en el alineamiento de Arqueas que no surgen en los alineamientos múltiples de Bacterias ni Eucariontes.

INTRODUCCIÓN

1. ORGANISMOS EXTREMÓFILOS

En la tierra los seres vivos sobreviven a temperaturas que desde nuestro punto de vista parecen extremas. Podemos mencionar el caso de la arqueas: *Pyrococcus furiosus* vive y se reproduce entre los 70°C a 103°C (Fiala, G. & Stetter, K. 1986) en contraste, *Methanococoides burtonii* puede crecer a 0°C y fue aislada del lago Ace de la Antártida (Frazmann, et. al. 1992). Para la ciencia sigue siendo un misterio como estos organismos extremófilos han conquistado prácticamente todos los rincones del planeta. Desde las montañas más altas, hasta los mares más profundos; desde el frío antártico, hasta los cráteres más calientes (Coker J., 2019). Aunque en su mayoría son procariontes, los organismos extremófilos se encuentran presentes en los tres dominios de la vida: eucaria bacteria y arquea.

Para sobrevivir algunos de los en ambientes extremos, los organismos han desarrollado mecanismos para mantener condiciones más neutras en su medio intracelular, por ejemplo: Los organismos halófilos que viven en altas concentraciones de sales, tienen bombas de iones para expulsar el exceso de sodio del medio intracelular (Rothschild & Mancinelli, 2001). Otro ejemplo son los organismos que viven en ambientes ácidos, estos tienen bombas de protones que expulsan el exceso de protones al medio extracelular. Por el contrario, los extremófilos que viven a un pH elevado tienen canales iónicos que permiten la salida de sodio y la entrada de protones a la célula, además de una capa polimérica que repele los iones OH⁻ del medio extracelular (Figura 1). De esta forma, los organismos que viven a extremos de la escala de pH mantienen un pH neutro dentro de sus células (Schmid, Allers & DiRuggiero, 2020).

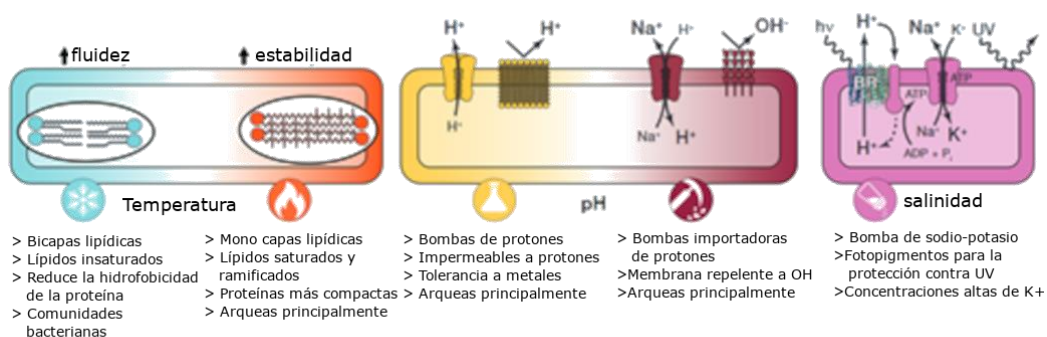


Figura 1. Mecanismos utilizados por organismos extremófilos para sobrevivir en ambientes extremos. Como altas y bajas temperaturas, extremos en la escala de pH y de alta salinidad. Modificado de (Schmid, Allers & DiRuggiero, 2020)

En cuanto a otras variables, existe un amplio rango de temperaturas óptimas de crecimiento (T_{opt}); a los organismos que crecen por debajo de 15°C se les denomina psicrófilos; a los que crecen entre 15°C y 60° , mesófilos; entre 60°C y 80°C , termófilos; y finalmente a los que crecen más allá de los 80°C se les llama hipertermófilos (Rothschild & Mancinelli, 2001). Los organismos psicrófilos tienen cambios en la estructura de su membrana lipídica con la mayoría de los fosfolípidos insaturados. La mayoría de estos organismos psicrófilos son bacterias, que se han aislado de lagos en la Antártica o suelos cerca del polo norte (Stetter, 1999). En cambio, la mayoría de los hábitats hipertermófilos son paredes de geiseres en el fondo del océano donde el agua llega a temperaturas de 200°C a 350°C , también se han encontrado en el subsuelo de depósitos de aceite caliente (60°C a 100°C) en Alaska, el Mar del Norte y Siberia. Por último, hay hipertermófilos en hábitats creados por el hombre, algunos bacilos termófilos han sido aislados de boilers de agua (Stetter, 1999) (Schmid, Allers & DiRuggiero, 2020).

La principal limitante para caracterizar a los organismos extremófilos es la dificultad de su cultivo en laboratorio. Porque desconocemos muchos detalles de su crecimiento como: requerimiento de oxígeno y concentración de nutrientes; así como la eliminación de productos tóxicos o dependencias sintróficas con otros organismos (Sterner & Liebl 2008). A pesar de su poca diversidad morfológica, los procariontes tienen una gran diversidad metabólica y proteica. La mayoría de los hipertermófilos cultivados son quimiolitóautótrofos, es decir, que pueden obtener su carbono a partir del CO_2 y tienen un metabolismo respiratorio que gana energía de la oxidación de sulfuros, hidrógeno y hierro. Por ejemplo, la fermentación de *Pyrococcus furiosus* procede de la vía modificada de Embden-Meyerhof-Parnas involucrando a la glucocinasa y la fosfofructocinasa que son dependientes de ADP, a comparación de la vía clásica en donde dependen de ATP. Los mecanismos moleculares que le permiten a los organismos vivir a estas altas temperaturas es diverso, incluyendo adaptaciones específicas a diferentes niveles, como la estructura y función de sus macromoléculas, su fisiología y metabolismo, así como la regulación y mantenimiento de su DNA (Sterner & Liebl 2008).

A pesar de que conocemos las estrategias que utilizan los psicrófilos y termófilos para cuidar sus ácidos nucleicos y membranas celulares, seguimos sin conocer exactamente cuáles son las características y mecanismos para adquirir termoestabilidad proteica. Las proteínas determinan como los seres vivos interactúan y se desarrollan en su ambiente, permitiéndoles sobrevivir en condiciones extremas. Entender este fenómeno es de gran interés industrial, pues las enzimas tienen una gran variedad de aplicaciones a altas temperaturas. Un claro ejemplo es la industria del almidón, que se gelatiniza a temperaturas elevadas para que sea accesible al ataque de las enzimas que lo transforman. Otro gran ejemplo es la industria del papel, que busca reducir el contenido de cloro para

blanquear la pulpa, una opción es remover el xilano residual de la pulpa con xilanasas, que reducirían el uso de blanqueadores químicos. El proceso se lleva a altas temperaturas y a pH alcalino. (Sterner & Liebl 2008). Un ejemplo más es el uso de la DNA polimerasas termoestables para la técnica de la reacción en cadena de la polimerasa (PCR en inglés) uno de los métodos clave para la ingeniería genética, secuenciación del ADN y diagnóstico molecular (Sterner & Liebl 2008).

El fracaso de aplicar las enzimas de organismos termófilos en la industria es multifactorial. Solo pocos organismos termófilos se han analizado a profundidad. Para encontrar biocatalizadores nuevos y estables, con las propiedades necesarias para aplicaciones biotecnológicas es necesario explorar la diversidad microbiana en los ambientes calientes. Esto requiere esfuerzos interdisciplinarios para tener un entendimiento profundo de cómo las proteínas hipertermófilas, se estabilizan tanto intrínsecamente como extrínsecamente (Sterner & Liebl 2008).

Se ha abordado el problema desde múltiples aproximaciones científicas incluyendo: microbiología, biología molecular, bioquímica, fisicoquímica, biología evolutiva y biología estructural. Esta tesis tiene como objetivo describir las características que proveen termoestabilidad a la Triosafosfato Isomerasa de Arqueas mediante una aproximación evolutiva, estructural y bioquímica.

2. APROXIMACIONES

2.1 APROXIMACIÓN BIOQUÍMICA

Las proteínas son biopolímeros de origen genético, es decir, se forman a partir de la información conservada en la doble hélice de ADN. Cuando esta información se traduce se generan arreglos específicos tridimensionales de cadenas polipeptídicas, es decir, proteínas en su estado nativo que son biológicamente funcionales. Resolviendo todas las necesidades celulares; -desde la movilidad y la comunicación hasta la alimentación y la reproducción-. Así, las proteínas son los elementos que definen la identidad de los seres vivos.

¿Cómo es posible que todos los aminoácidos que forman una proteína puedan encontrar su conformación nativa específica? En 1969 Cyrus Levinthal planteó su paradoja. Un experimento mental en el que imaginamos que existen 2 conformaciones por aminoácido, una proteína con 100 aminoácidos tendrá por tanto 2^{100} conformaciones posibles (su universo conformacional). Si le toma un picosegundo en ir de una conformación a otra. Muestrear todas las conformaciones le tomaría 10^{18} segundos, lo que equivale a 10^{10} años, tiempo mayor a la edad del universo. Como estas condiciones son incompatibles con la vida, podemos concluir dos cosas: la primera es que la búsqueda de la estructura nativa a través de todo el universo conformacional es imposible en un tiempo biológicamente relevante, y segundo que la búsqueda de la estructura nativa no es azarosa.

En 1961 Christian Boehmer Anfinsen trabajó con la ribonucleasa bovina, y estableció que la desnaturalización de las proteínas puede ser un proceso reversible dando origen a la hipótesis termodinámica del plegamiento; que se define como: -el proceso físico mediante el cual las cadenas polipeptídicas adquieren su conformación de mínima energía, también llamada conformación nativa-. sus experimentos también sugieren que la información necesaria para adquirir esta la conformación nativa se encuentra en la secuencia de aminoácidos.

De esta manera, cada conformación proteica tiene asociada una energía determinada, al conjunto de las energías correspondientes a todas las conformaciones posibles se les llama paisaje energético o paisaje conformacional. Dicho paisaje es único para cada proteína, pero generalmente se representa en forma de embudo en donde el fondo representa el estado nativo, y las rugosidades del entorno representan estados conformacionales parcialmente estructurados, llamados intermediarios (Figura 2). Conocer el paisaje conformacional de proteínas homólogas nos ayuda a estudiar cómo han evolucionado sus características moleculares (Romero, Fernández & Costas, 2018).

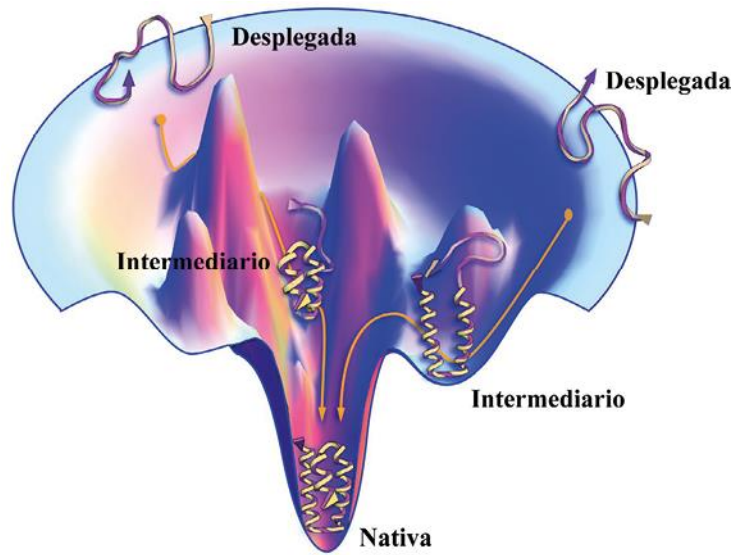


Figura 2. Paisaje conformacional en forma de embudo. Nos muestra como la proteína va del estado desplegado en la superficie, al estado nativo de menor energía en el fondo. Tomado de (Romero, Fernández y Costas 2018)

Los parámetros termodinámicos nos dan una descripción clara del plegamiento. Por ejemplo, el cambio en la energía libre de Gibbs (ΔG) describe la energía del sistema -proteína-solvente-, y determina si un proceso es espontáneo o no: si el cambio es negativo ($\Delta G < 0$) el proceso ocurre de manera espontánea, por el contrario, si es positivo ($\Delta G > 0$) nos indica que el estado inicial es el estado estable. La gráfica que ilustra el cambio del ΔG , en relación con la temperatura se le conoce como "curva de estabilidad" (Figura 3). Se obtiene experimentalmente por medio de la desnaturalización por temperatura y/o agentes químicos y solo se puede determinar cuando el desplegamiento es reversible. Cada proteína tiene su propia curva de estabilidad, la cual puede cambiar de acuerdo con las condiciones experimentales como: pH, amortiguador, solutos, etc. De dicha gráfica se pueden obtener otros valores importantes como las temperaturas de desnaturalización (T_d) por calor y por frío que son los puntos en donde interseca la curva con el eje de las X (Figura 3). También podemos observar en la gráfica que el ΔG alcanza un valor máximo, es un punto importante porque nos indica la temperatura de máxima estabilidad (T_s) (Romero, Fernández & Costas, 2018).

La expresión matemática que permite el cálculo de la curva de estabilidad ($\Delta G(T)$) es (ecuación 1):

$$\Delta G(T) = \Delta H \left(1 - \frac{T}{T_m} \right) - \Delta C_p \left[T_m - T + T \ln \left(\frac{T}{T_m} \right) \right] \quad (1)$$

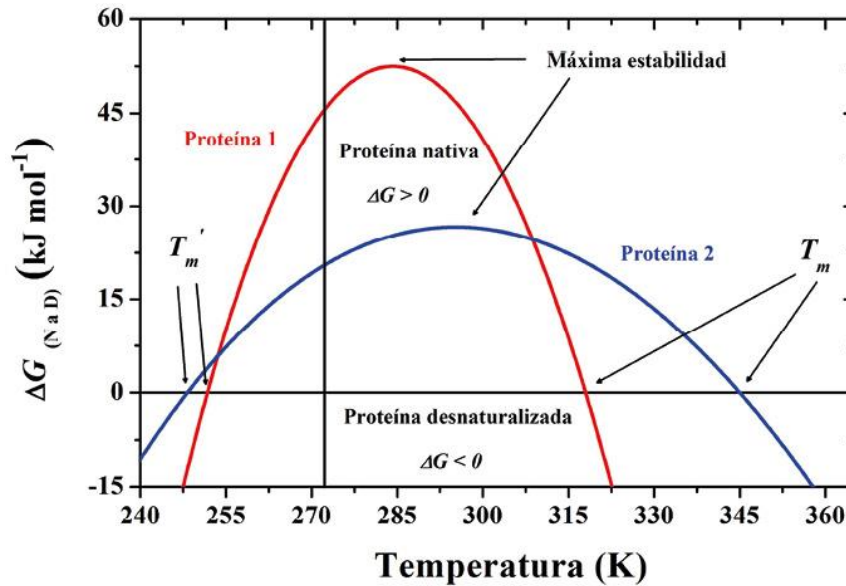


Figura 3. Ejemplo de una curva de estabilidad. Nos muestra la temperatura de desnaturalización por calor (derecha) y por frío (izquierda) como los puntos en los que la curva se interseca con el eje x. También nos muestra la temperatura de máxima estabilidad como el punto más alto de la curva. Tomado de (Romero, Fernández & Costas, 2018)

Otra aproximación experimental muy útil para evaluar la “huella termodinámica” de las proteínas es la Calorimetría Diferencial de Barrido (CDB o DSC en inglés). Que consiste en comparar dos celdas: una con la proteína de interés y la otra con el amortiguador. Ambas celdas se encuentran a presión constante, imbuídas en una chaqueta adiabática lo que impide el intercambio de energía con el exterior (Figura 4.A). Al calentar ambas celdas a velocidad constante, debido a que están aisladas, el instrumento determina la capacidad calorífica, esto es, la diferencia en la energía requerida para aumentar la temperatura de las celdas con proteína nativa y con solvente. La cual es constante (Figura 4.B) hasta que ocurre el desplegamiento de la proteína, este proceso requiere energía. Por lo que cuando la proteína se desnaturaliza y absorbe energía, el instrumento contrarresta este cambio devolviéndola al equilibrio con la otra celda, es decir, le suministra calor, hasta que la proteína se desnaturaliza completamente, a partir de esta temperatura se obtiene la capacidad calorífica del estado desnaturalizado, $C_{p,D}$. Al graficar la Capacidad calorífica a presión constante contra la temperatura (Figura 4.B), podemos obtener otros dos valores muy importantes: el primero es el cambio en la entalpía (ΔH) que se define como toda el área bajo la curva de desnaturalización, la cual nos indica el cambio de las interacciones químicas en el desplegamiento; el segundo es la temperatura media de desnaturalización (T_m) en la que el C_p alcanza su máximo, y nos indica que la concentración de las proteínas en el estado nativo es igual a la concentración de proteínas desplegadas (Romero, Fernández & Costas

2018). Este valor es particularmente importante en la investigación de la termoestabilidad, porque nos ofrece un factor de comparación entre la estabilidad térmica de las proteínas homólogas de organismos mesófilos y termófilos.

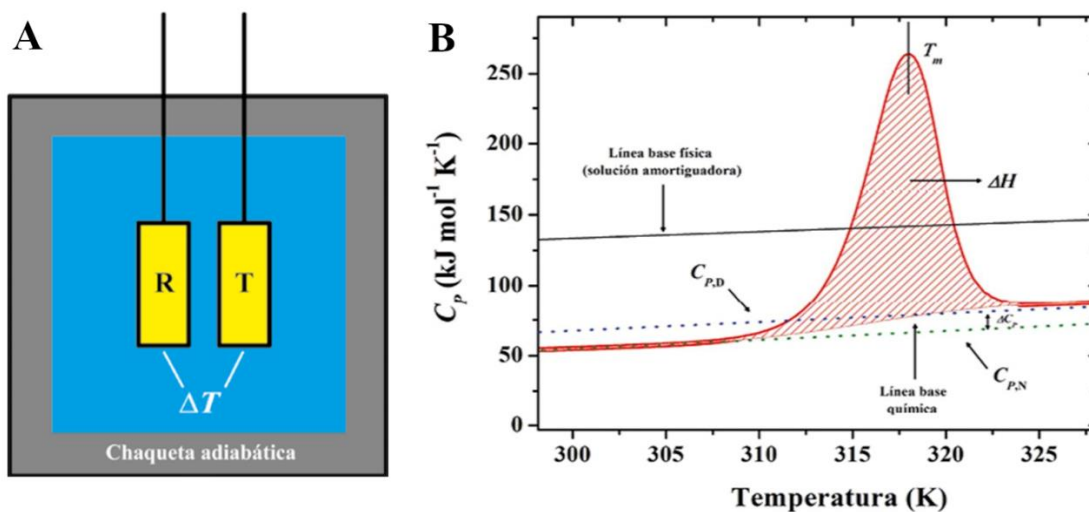


Figura 4. Calorimetría Diferencial de Barrido. A) Composición de las celdas experimentales dentro de la chaqueta adiabática. B) Gráfica del DSC que ilustra los valores de ΔH y T_m . Tomado de (Romero, Fernández & Costas, 2018).

La CDB resulta realmente útil para el estudio de la termoestabilidad, porque los contenidos de las celdas están presurizados con nitrógeno. Lo que aumenta el punto de ebullición de los líquidos, permitiendo estudiar proteínas en condiciones de hasta 125°C. Además, con los datos obtenidos del ΔH y del ΔC_p se puede construir la curva de estabilidad. A partir de la comparación de las curvas de estabilidad de proteínas de organismos termófilos y sus homólogos mesófilos, se han propuesto 3 estrategias mediante las cuales las proteínas son capaces de aumentar su termoestabilidad (Razvi & Scholtz, 2006). En la Figura 5 se observa la curva de estabilidad de un organismo mesófilo en una línea negra continua, con ella compararemos las otras 3 curvas que ilustran las estrategias que utilizan los organismos termófilos para aumentar su T_m .

La primera estrategia y la más usada por las proteínas, es aumentar su estabilidad general, es decir, incrementar su ΔG ; al hacer esto la curva de estabilidad gana altura y se amplía, permitiendo alcanzar valores más altos de T_m , como se muestra en la Figura 5 con rombos rosas. La segunda estrategia ilustrada con círculos amarillos en la Figura 5. es disminuir el cambio en la capacidad calorífica (ΔC_p), provocando un aplanamiento de la curva, alcanzando valores de T_m mayores. La tercer y ultima estrategia es aumentar la temperatura de máxima estabilidad (T_s)

recorriendo toda la curva hacia la derecha, permitiendo valores más altos de T_m , ilustrado en la Figura 5 con cuadrados azules (Razvi & Scholtz, 2006).

La mayoría de las proteínas termoestables utilizan dos o más estrategias dentro de las antes mencionadas para aumentar su termoestabilidad. Pese a que se conocen estas “estrategias” no es posible concluir que son estas “las características las que confieren la resistencia a la temperatura”, puesto que en la naturaleza no existe una característica que esté presente en todas las proteínas termoestables (Dhaunta, Arora, et al., 2013).

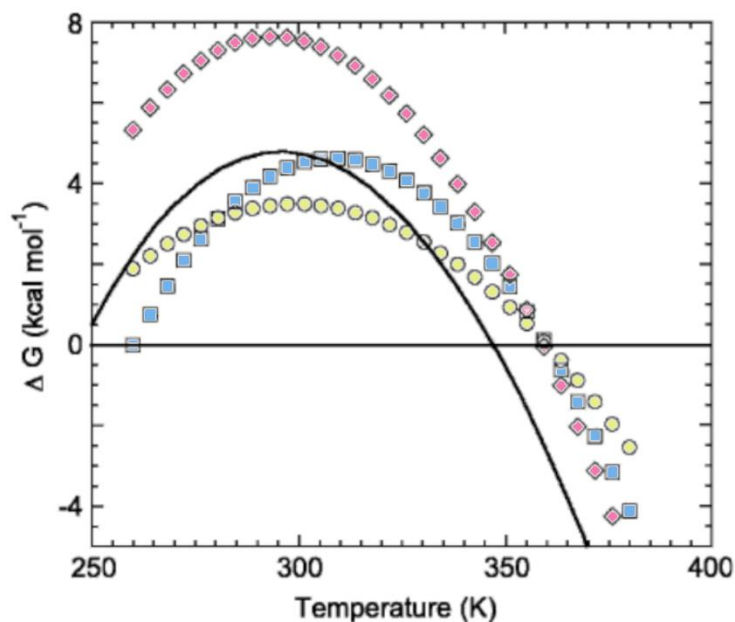


Figura 5. Estrategias fisicoquímicas que utilizan las proteínas para incrementar su termoestabilidad. Se muestra la comparación de una curva de estabilidad de una proteína mesófila en línea sólida negra, en rombos rosas se ilustra el aumento de la estabilidad en general, en cuadros azules el aumento de la temperatura de máxima estabilidad y por último en círculos amarillos la disminución del ΔC_p . Tomado y modificado de (Razvi & Scholtz, 2006)

2.2 APROXIMACIÓN EVOLUTIVA/SECUENCIAL

La distribución de los aminoácidos a lo largo de la secuencia no es aleatoria y contiene en sí misma la información necesaria para el plegamiento de la proteína y por lo tanto para su funcionalidad biológica. El espacio de secuencia es un término que representa todas las posibles mutaciones y sus conexiones entre ellas (Andreas Wagner, 2008). Para ejemplificarlo, supongamos que existe una proteína de longitud N y tenemos 20 aminoácidos disponibles para mutar, por ende, su espacio de secuencia tendría una magnitud de 20^N . Debido a esta enorme magnitud ($20^N > 2^{100}$) resulta imposible de caracterizar experimentalmente todo el espacio secuencial (Figura 6). Sin embargo, la naturaleza ha explorado secuencias, a lo largo de millones de años en millones de organismos. Dando como resultado una gran diversificación y optimización de secuencias, estructuras y funciones (Harms & Thornton 2013). El espacio de secuencia se convierte en un mapa del genotipo-fenotipo en el que cada nodo tiene información acerca de las propiedades (moleculares o adaptativas) de nuestro interés; -se ejemplifican como picos en el espacio de secuencia- (Figura 6). Esto nos ayuda a visualizar como estas propiedades se distribuyen a lo largo de las secuencias -parecido a un mapa- (Payner & Wagner 2019). Por ejemplo, en nuestro caso estamos interesados en la termoestabilidad, así que mientras más grande sea la altura del pico más grande es la termoestabilidad en esa secuencia específica (Figura 6). Lo maravilloso del concepto es que se puede aplicar a todos los rasgos que queramos, si nos interesa la actividad enzimática, los picos serán diferentes a los picos de termoestabilidad para un mismo espacio de secuencia. Permittiéndonos esquematizar la evolución de las moléculas biológicas (Harms & Thornton 2013).

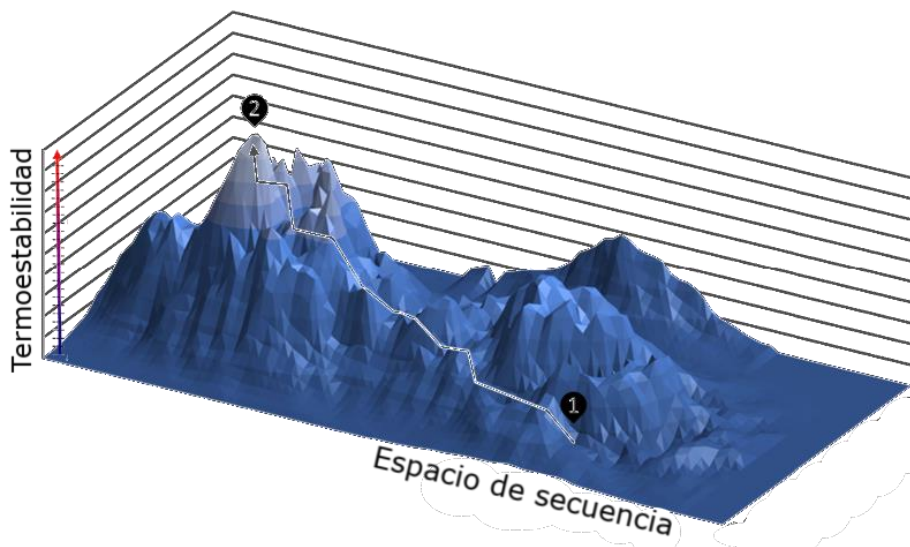


Figura 6. Espacio Secuencial de una proteína. En el eje y observamos el rasgo de nuestro interés. En los otros ejes el espacio de secuencia.

Los biólogos evolutivos estudian las trayectorias de las proteínas a lo largo de este mapa y qué fuerzas evolutivas dirigen el problema (Harms & Thornton 2013). Durante el estudio evolutivo se debe tener en cuenta que las proteínas no son objetos funcionales individuales. El proceso es mucho más complejo, porque la selección purificadora actúa a nivel del fenotipo. Sin embargo, las proteínas se encuentran justo en medio en el espacio genotipo-fenotipo. Porque su secuencia de aminoácidos se basa en la transcripción genética del DNA, y a la vez su estructura expresa el fenotipo celular. Lo que significa que las proteínas tienen información del sobre cómo fueron seleccionando (genotipo) y qué fuerzas evolutivas lo fueron guiando (fenotipo) (Harms & Thornton 2013).

2.2.1 PATRONES EVOLUTIVOS

La evidencia del experimento masivo de la naturaleza está en las secuencias de las proteínas, por lo que un alineamiento de secuencias múltiple (MSA en inglés) se convierte en un registro histórico (Valdar 2002). Los patrones de conservación y variación en dichas secuencias nos cuentan cómo fue la historia evolutiva, por medio de presión selectiva, recombinaciones, duplicaciones, deleciones o deriva génica (Harms, M. & Thornton, J. 2013). Gracias a la tecnología en secuenciación podemos acceder a una gran cantidad de secuencias de todos los dominios de la vida. Lo que nos permite tener más datos para buscar patrones estadísticamente significativos en los alineamientos. Sin embargo, no es una tarea sencilla, pues hay que distinguirlos de toda una amalgama de señales de fondo, producto de diferentes tipos de correlaciones entre aminoácidos y secuencias (Morcos, F. et.al. 2014).

¿Cómo se pueden reconocer dichos patrones en las secuencias? El primer rasgo que surge en los alineamientos múltiples es la existencia de ciertos aminoácidos cuya identidad se mantiene en la mayoría de las secuencias, es decir, están conservados. Como la mayoría de las sustituciones en la secuencia son neutras, la conservación nos habla de la existencia de presión selectiva y por lo tanto nos indica que ese aminoácido es importante o esencial para mantener la función de la proteína. Así que el grado de conservación indica la importancia funcional por lo que identificar las regiones conservadas de una proteína es tremendamente útil (Valdar, 2002).

Como segunda aproximación se han desarrollado técnicas como el Análisis directo de acoplamiento (DCA en inglés) y el Análisis estadístico de acoplamiento (SCA en inglés). Ambas técnicas utilizan alineamientos múltiples de secuencia (MSA) como datos iniciales; es necesario contar con al menos 1000 secuencias de proteínas homólogas, para el DCA y al menos 100 para el SCA. Lo que buscan ambas técnicas son aminoácidos que co-evolucionan, es decir, pares de aminoácidos cuya

conservación o variación dependan uno del otro. En el DCA se pretende predecir interacciones físicas en la estructura tridimensional utilizando únicamente la correlación de las frecuencias entre pares de aminoácidos (Figura 7A) (Morcos, F. et.al. 2011). En cambio, el SCA busca reconocer redes de aminoácidos co-evolutivos, (sectores) por medio de la perturbación del alineamiento y correlacionando los datos obtenidos en una matriz (Figura 7B). Ambas técnicas han tenido resultados sorprendentes para predecir aminoácidos funcionalmente relevantes y han sido aplicadas en diversas familias de proteínas como: señalizadores de dos componentes, serin-proteasas, proteínas G, etc.

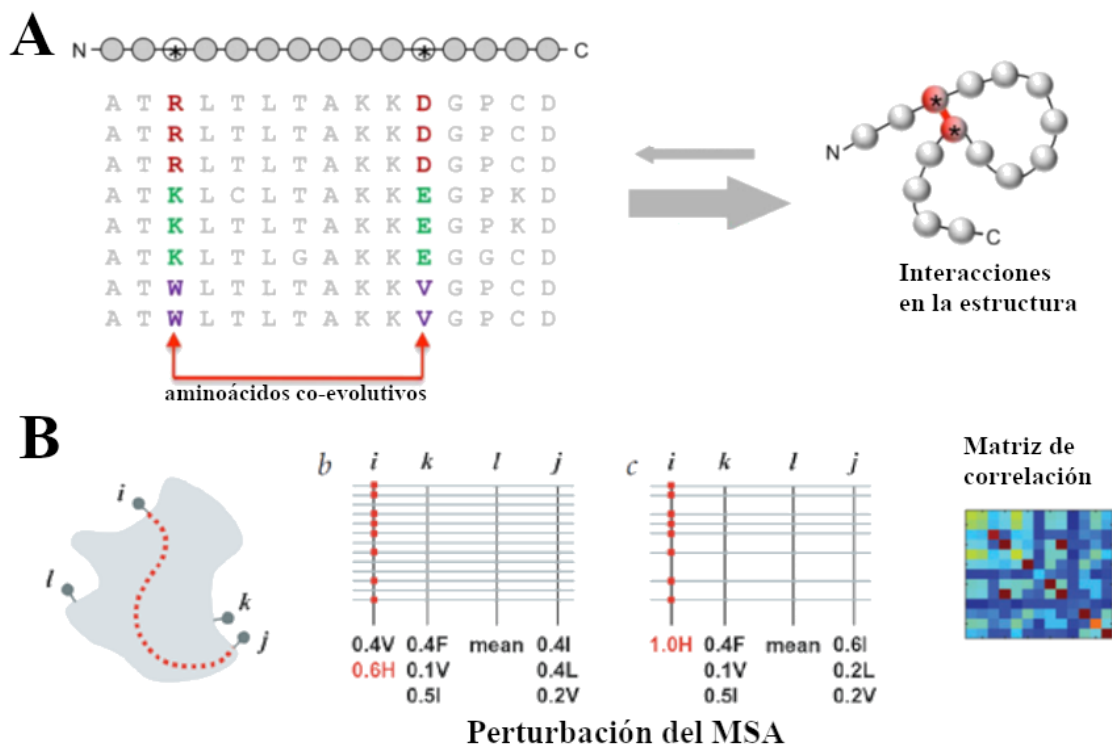


Figura 7. Técnicas para conocer los patrones evolutivos en las secuencias. A) Análisis de acoplamiento directo (DCA). B) Análisis estadístico de acoplamiento (SCA). Tomado y modificado de (Mark et. al. 2011) & (Ranganathan et. al. 2003)

La existencia de aminoácidos que co-evolucionan es evidencia circunstancial de un fenómeno llamado epistasis que ocurre cuando dos o más mutaciones no tienen un efecto aditivo predecible, causando así cambios en el fenotipo (Cooper & Santiago 2011). Sin embargo, se necesita evidencia experimental sobre los impactos funcionales o fenotípicos de las mutaciones ocurridas, que corroboren las firmas estadísticamente significativas (Starr & Thornton 2016).

2.2.2 EPISTASIS

El concepto de epistasis es complejo, pues trata de ilustrar el fenómeno de no-aditividad: cuando el resultado total es más que la suma de sus elementos (A-B, B-C $A \neq C$). Para ejemplificarlo, imaginemos que tenemos dos mutaciones que son neutras, es decir, no tienen cambios en el fenotipo. Uno esperaría que cuando se mutaran ambos sitios el resultado fuera neutro. Sin embargo, cuando suceden ambas mutaciones y se observa epistasis, si hay un cambio en el fenotipo, por ejemplo, en la conformación, estabilidad, actividad enzimática, etc. (Masel et. al. 2010). Aún se desconocen las causas de este efecto, pero se teoriza que, tras una mutación, aunque sea neutra, al final, se alteran las interacciones físicas entre los residuos de la proteína ya sea estabilizando o desestabilizando la red de interacciones. Pero, al cambiar estos contactos estructurales se obtienen efectos no lineales de las propiedades físicas. Que se tratan de compensar con la segunda mutación.

En 2012 el grupo de Richard Goldstein utilizó la evolución de la fosfatasa morada para investigar si el efecto de la sustitución de aminoácidos puede ser predecible. Encontraron que después de una mutación, aminoácidos similares tienden a ser gradualmente más aceptables a través del tiempo en esa posición estructural. Lo que nos indica que las proteínas tienden a compensar o equilibrar la presencia de un aminoácido en cierta posición a través de sustituciones en otras posiciones (Pollock, Thiltgen & Goldstein 2012) (Figura 8.A). Las relaciones epistáticas entre aminoácidos son muy importantes porque permiten fijar mutaciones nuevas para compensar estructuralmente a las mutaciones predecesoras. De otra manera, la mayoría de las sustituciones serían deletéreas (Figura 8.A) (Shah, McCandlish & Plotkin 2015). Al mismo tiempo cada mutación aleatoria se encuentra dentro de lo neutral y luego bajo cierto tiempo se fija por mutaciones subsecuentes, y esa misma mutación se vuelve deletérea en ausencia de las mutaciones que se predijeron, provocando así una irreversibilidad en el proceso. (Shah, McCandlish & Plotkin 2015) (Figura 8). A este fenómeno se le denomina amurallamiento o atrincheramiento (*entrenchment* en inglés). El amurallamiento es un término importante porque nos habla sobre dos cosas muy importantes; la primera es que la magnitud de los efectos de nuevas mutaciones depende del contexto secuencial (*background* en inglés), y la segunda es que los efectos en el fenotipo dependen de la trayectoria en la que surgieron las mutaciones (Figura 8.B) (Pollock, Thiltgen & Goldstein 2012).

En 2015 el grupo de investigación de Plotkin J. realizó un experimento muy interesante para conocer la relación entre epistasis, compensación, amurallamiento y selección purificadora. Consistía en realizar 10 mutaciones

aleatorias puntuales en la secuencia de la proteína de unión a periplasma lisina-arginina-ornitina (*argT*); para cada una de las mutantes se calculaba su estabilidad predicha (ΔG) utilizando FoldX. En cada paso se seleccionaban las proteínas que tenían mayor o igual estabilidad que la secuencia original, bajo un modelo de Moran. El proceso se repetía durante 30 substituciones consecutivas para simular una trayectoria evolutiva. Simularon 100 trayectorias replicadas, cada una iniciando con la secuencia original de (*argT*) (Plotkin J, et. al. 2015). A lo largo de toda la simulación de la trayectoria evolutiva calcularon el coeficiente de epistasis, que nos indica cuanto cambia el fenotipo con relación a dos mutaciones y si el fenotipo es muy diferente a la suma del fenotipo de las mutaciones independientes. Se calcula a partir de la desviación entre el efecto en la estabilidad de la doble mutante y la suma de los efectos de estabilidad de las mutantes individuales. Como se muestra en la siguiente ecuación. Donde ΔG_{ij} significa la estabilidad de la doble mutante, ΔG_0 la estabilidad de la proteína original, ΔG_i la estabilidad de mutante i y ΔG_j la estabilidad de la mutante j.

$$E = [\log(\Delta G_{ij}) - \log(\Delta G_0)] - ([\log(\Delta G_i) - \log(\Delta G_0)] + [\log(\Delta G_j) - \log(\Delta G_0)])$$

En la (Figura 8.C) se muestra como las mutaciones bajo selección purificadora mantienen la estabilidad de la proteína, mientras que las mutaciones aleatorias sin selección tienden a disminuir la estabilidad de la proteína. En la (Figura 8.D) se muestra una grafica con el coeficiente de epistasis en el eje "y" y la trayectoria evolutiva bajo selección purificadora en el eje "x", las líneas verticales representan ± 2 desviaciones estándar alrededor de la media del conjunto de las 100 trayectorias evolutivas simuladas. Se puede observar que bajo selección purificadora las mutaciones tienden primero a encontrar un equilibrio por medio de la compensación y despues incrementan sus interacciones epistaticas, generando el fenomeno de amurallamiento.

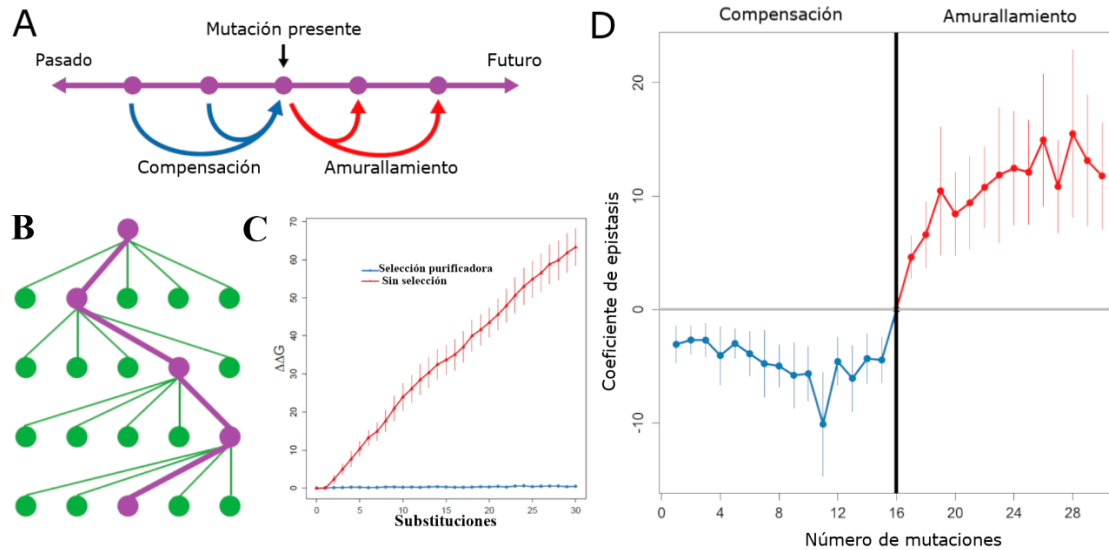


Figura 8. Fenómeno de amurallamiento (entrenchment). A) Es una línea del tiempo que representa cómo las mutaciones del pasado compensaron el ambiente químico para que la mutación del presente no sea deletérea, además muestra que en un futuro la mutación presente será necesaria para que nuevas mutaciones no sean deletéreas. B) Esquema que representa el concepto de amurallamiento, los círculos verdes representan todas las posibles sustituciones y el morado ilustra las sustituciones secuenciales. Nos muestra que este proceso depende de la trayectoria o de la historia de las mutaciones. C) Comparación de la estabilidad proteica entre las poblaciones con y sin selección purificadora. En el eje “y” se muestra el $\Delta\Delta G$ y en el eje “x” el número de sustituciones. D) Eje “y” muestra el coeficiente de epistasis, las líneas verticales representan ± 2 desviaciones estándar alrededor de la media del conjunto de las 100 trayectorias evolutivas simuladas, en el eje “x” se representa trayectoria evolutiva simulada. Tomado y modificado de (Plotkin, et. al. 2015)

Como ya mencionamos, la epistasis actúa a nivel del fenotipo. En términos del espacio de secuencia, la epistasis introduce un cambio de altitud que se refleja como rugosidades en el paisaje – pendientes altas o bajas – (Figura 9). La altura que se obtenga varía dramáticamente dependiendo de su punto de partida y de la dirección del paso (Starr & Thornton 2016) – pendientes altas o bajas – (Figura 9). La epistasis puede restringir las trayectorias disponibles en el proceso evolutivo por medio del atrincheramiento y a la vez puede abrir nuevos caminos al permitir alcanzar puntos altos en pasos cortos.

La robustez es el medio por el cual las poblaciones pueden explorar el espacio secuencial. Un sistema biológico es robusto si puede explorar muchas mutaciones sin alterar el plegamiento, es decir, que la mayoría de las sustituciones son neutras (Payner & Wagner 2019). A simple vista podríamos pensar que la robustez se opone a la búsqueda de nuevos fenotipos, pero no es así, las poblaciones que

evolucionan en redes neutrales pueden acceder a una mayor variación. Al tener mayor variación un fenotipo puede ser producido por diferentes secuencias o genotipos, aumentando así la variación fenotípica (Andreas Wagner 2018).

“Las mutaciones son eventos aleatorios. De aquí la necesidad de que las poblaciones posean en todo momento una reserva de variabilidad potencial oculta.” – Dobzhansky, 1937

La mayoría de las mutaciones son toleradas o neutras sin selección purificadora (Bershein & Tawfik 2006). Al contrario, las mutaciones bajo selección purificadora son epistáticas y tienden a amurallarse por sus substituciones anteriores (Shah, McCandlish & Plotkin 2015). En este caso la selección purificadora remueve las mutaciones deletéreas, así que se espera que las positivas son las que permanezcan a lo largo del tiempo (Starr & Thornton 2016), esto sugiere las mutaciones aleatorias bajo selección purificadora –como suele pasar en el proceso evolutivo- aumentan las substituciones con efectos epistáticos en la estabilidad (Shah, McCandlish & Plotkin 2015).

Para visualizar mejor estos conceptos propondremos un experimento; supongamos que tenemos un cultivo bacteriano creciendo en un medio rico y en algún punto colocamos un antibiótico. Observamos que la mayoría de las células muere, pero solo algunas sobreviven, las afortunadas tienen proteínas que les confieren resistencia a antibióticos. En este caso la robustez le permitió a la población de bacterias explorar varios genotipos y alguno(s) les confirió la resistencia antibióticos. En nuestras condiciones iniciales, esto era información encriptada porque estaba latente pero no estaba contribuyendo al fenotipo (Payne & Wagner 2019). Cuando agregamos el antibiótico generamos una selección negativa, pues sólo las bacterias resistentes a antibióticos sobrevivieron, si nosotros mantenemos la presión selectiva, las bacterias comenzaran a generar diversos mecanismos de supervivencia, es decir, generarán nuevos fenotipos. Aquí es donde la epistasis hace su magia, ya que nos permite explorar diversos fenotipos, por medio de pocas substituciones (Figura 9).

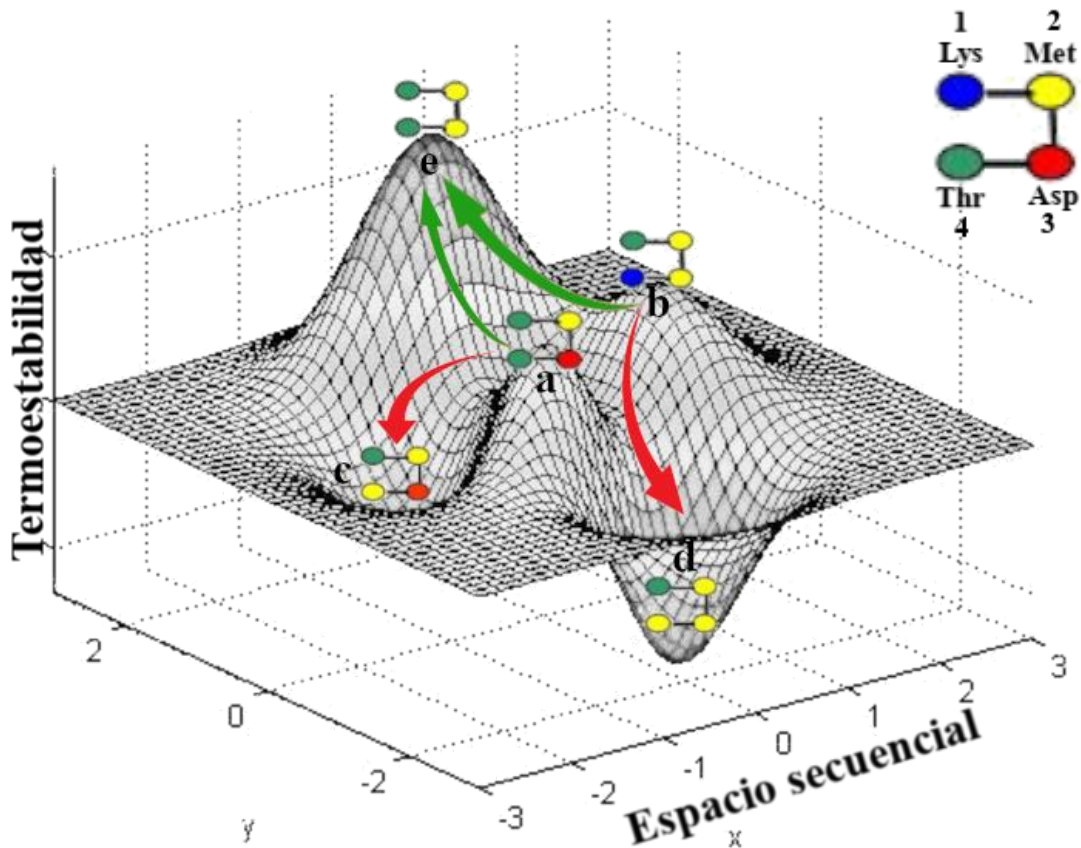


Figura 9. Fenómeno de Epistasia. Esta figura ilustra el espacio secuencial en la base de la gráfica, en el eje “z” se muestra la adaptabilidad de un rasgo (en este caso la termoestabilidad). Los puntos a, b, c, d & e representan secuencias de 4 residuos, y los colores representan la identidad del aminoácido (verde: treoninas, rojo: aspartatos, azul: lisinas y amarillo: metioninas). Observamos que la sustitución de una metionina en la posición 4 para a & b ocasiona la pérdida de termoestabilidad, resultando en las secuencias c & d respectivamente. Pero c tiene mayor termoestabilidad respecto de d, por lo que la misma sustitución no tuvo el mismo efecto en las secuencias, sino que dependió del fondo genético. En contraste podemos ver que el punto e tiene la mayor termoestabilidad, y los puntos a y b podrían llegar a adquirirla pero por medio de sustituciones diferentes; en el caso de a sería un cambio de aspartato por metionina en la posición 3; en el caso de b sería una mutación de la lisina en la posición 4 por una treonina. La conclusión es que la adquisición o pérdida de un rasgo depende de dos cosas; la primera: el punto de partida en el espacio de secuencia que ejemplifica el fondo genético; y la segunda: la dirección de la mutación, es decir, el tipo de sustitución.

En nuestro ejemplo podemos apreciar como la epistasia y la robustez incrementan la capacidad de una población para producir “variación fenotípica heredable”, a lo que llamamos evolucionabilidad (Masel et. al. 2010). Tanto la epistasia como la robustez permiten generar nuevas funciones en poco tiempo, por lo que son dos

fuerzas muy importantes en el proceso evolutivo, porque juntas exploran nuevos fenotipos. La robustez, se puede interpretar como un factor que afecta la divergencia de un fenotipo, sin embargo, la evidencia sugiere lo contrario. La capacidad de un fenotipo de explorar de manera neutral una amplia variedad de trasfondos genéticos aumenta su potencial adaptativo (evolucionabilidad) (Wagner Andreas, 2008). Por su parte, la epistasia tiene menores eventos históricos y deja marcas más fuertes en las secuencias que podemos reconocer por medio de los aminoácidos que co-evolucionan. El que las mutaciones sean heredables, ventajosas o deletéreas en un organismo, depende de su historia evolutiva y de su contexto secuencial. Sin embargo, una vez que se ha fijado la mutación en una población, ésta forma parte del contexto genético para modificaciones futuras (Shah, McCandlish & Plotkin 2015). Estos nuevos términos nos hacen replantearnos la idea tradicional que dice que el proceso evolutivo sucede lentamente, gracias a la baja probabilidad de acumular mutaciones poco a poco para derivar en la función (Harms & Thornton 2013)

“Las moléculas se están aboliendo constantemente, pero se preservan al mismo tiempo”

~ Zuckerman & Pauling 1965

2.3 APROXIMACIÓN ESTRUCTURAL

Una de las aproximaciones clásicas para el estudio de la termoestabilidad es la comparación de estructuras de proteínas provenientes de organismos termófilos contra sus homólogos mesófilos. Gracias a este método se han identificado potenciales características de las proteínas termoestables listadas a continuación:

1. Incremento de los puentes de hidrógeno.
2. Mejora de las interacciones electrostáticas o pares iónicos.
3. Optimización de las interacciones hidrofóbicas.
4. Aumento del empaquetamiento
5. Aumento del área polar respecto a la no polar
6. Aumento en el contenido y la estabilidad de las alfa hélices
7. Mejora en la unión de iones metal.
8. Eliminación de asas expuestas al solvente
9. Asociación en oligómeros
10. Reducción de residuos inestables como: asparagina, glutamina, cisteína y metionina.
11. Aumento de residuos cargados como: arginina, lisina, glutamato y aspartato.
12. Reducción de la flexibilidad, aumentando prolinas en las asas y disminuyendo la proporción de glicinas.

Se ha visto que en proteínas termófilas los enlaces iónicos aumentan no solo en cantidad sino en calidad, se vuelven más fuertes porque se encuentran a menor distancia (Szilagyi & Závodszky 2008). Específicamente estudios en la gliceraldehido-3-fosfato deshidrogenasa (GAPDH) de *Thermotoga marítima* (Jaenicke, et. al. 1996), en la Indol-3-fosfato sintasa de *Sulfolobus solfataricus* y la super-óxido dismutasa de *Aquifex pyrophilus*, confirman el papel crucial de los puentes salinos para la estabilización de las proteínas a altas temperaturas (Xiao & Honig 1999)

En 1999 Xiao & Honig realizaron un experimento de dinámica molecular para conocer la contribución de las interacciones electrostáticas en el plegamiento de cuatro familias proteicas: el glutamato deshidrogenas, GAPDH, ferredoxina y CheY. Posteriormente, compararon sus resultados con las contrapartes mesófilas. Encontraron que las proteínas termófilas tenían mayor cantidad de interacciones electrostáticas favorables. Incluso si tenían menos pares iónicos que los mesófilos, los hipertermófilos tienen más contribuciones electrostáticas favorables para la estabilidad. El papel individual de los pares iónicos y sus redes es variable. Los autores argumentan que quizá sea útil pensar en términos de que todos los grupos ionizables están contribuyendo a una sola red ionizable; cuya dimensión es la proteína en sí misma, incluyendo las interacciones relevantes y no solo las pocas interacciones representadas por pares iónicos (Xiao & Honig 1999).

Las moléculas de agua en contacto directo con la superficie de la proteína son parte de algo llamado “la primera capa de hidratación” (Levy & Onuchic 2006). Cuando se forma un puente salino, la capa de hidratación que es energéticamente favorable se elimina de las cadenas laterales de aminoácidos cargados involucrados. En 1998 Elcock realizó estudios teóricos donde calculaba la energía libre de hidratación de los 20 aminoácidos naturales y su dependencia con la temperatura (Elcock y McCammon, 1997). Los investigadores sugieren que esta penalización por la solvatación disminuye a medida que aumenta la temperatura. Concluyeron que las interacciones electrostáticas proteína-solvente son más favorables cuando aumenta la temperatura debido al aumento de la movilidad del agua. Resulta ser que la termoestabilidad no solo depende de la cantidad de interacciones electrostáticas sino también de su localización en la superficie de la proteína (Sterner & Liebl 2008). Un ejemplo claro es el caso de la proteína ribosomal L30e de *Thermococcus celer* que al cambiar los residuos cargados de la superficie por alaninas reduce drásticamente su termoestabilidad. (Brininger, Cobani & Evilia 2018). En la Figura 10 se ilustra claramente como las proteínas termófilas tienden a tener más residuos cargados en su superficie, y se muestran las estructuras cristalográficas de la malato deshidrogenasa (MDH) con sus mapas de Coulomb, en A tenemos la estructura de la MDH de un organismo mesófilo *Methylobacterium extorquens* (PDB: 5UJK), y en B se observa la estructura de la MDH de un organismo hipertermófilo *Haloferax volcanii* (PDB: 4BGU). Podemos observar claramente como la superficie de la estructura del organismo hipertermófilo está más cargada electrostáticamente ((Brininger, Cobani & Evilia 2018).

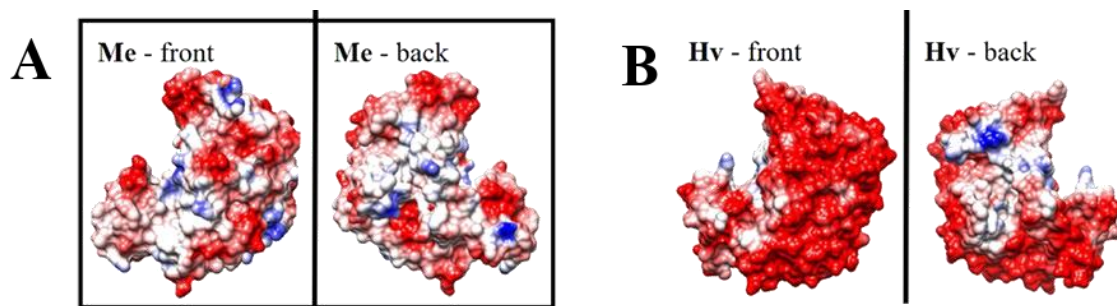


Figura 10. Aumento de cargas electrónicas en la superficie de proteínas termófilas. En A) se muestra la estructura de la Malato deshidrogenasa de Methylobacterium extorquens (Me) (PDB: 5UJK) un organismo mesófilo. En B) se observa la estructura de la Malato deshidrogenasa de Haloferax volcanii (Hv) (PDB: 4BGU) un organismo termófilo. En rojo se muestran las cargas negativas y en azul las positivas. Tomado y modificado de (Brininger, Cobani & Evilia 2018)

Otra consecuencia del aumento de temperatura es que algunas las interacciones hidrofóbicas se favorecen, la optimización de estas interacciones es otra característica reconocible en proteínas termófilas. Chakravarty y Varadarajan

notaron que, en promedio, las proteínas termófilas son más pequeñas en comparación con sus contrapartes mesófilas (Chakravarty S and Varadarajan R, 2002). Puede ser un hallazgo importante porque las proteínas pequeñas muestran una capacidad calorífica menor, lo cual resulta en una T_m más elevada (Figura 5). Se conocen casos como la N-(5'-fosforribosil) antranilato isomerasa (PRAI) de *Thermotoga maritima* que demuestra que al oligomerizar puede estabilizar dramáticamente a la proteína escondiendo superficie expuesta al solvente. (Sternner & Liebl 2008).

Otra característica importante de las proteínas termófilas es que reducen los elementos inestables de su estructura por ejemplo Thompson y Eisenberg encontraron en 1999 que las proteínas termófilas son más pequeñas debido a que tienen asas (regiones no estructuradas) más pequeños (Sternner & Liebl 2008). Además, a nivel genómico se nota una reducción significativa de residuos polares inestables (S, N, Q, G) (Szilagyi & Závodszky 2008). Tanto la asparagina como la glutamina se pueden desamidar, formando un anillo de succinimida que es hidrolizado en solución, produciendo residuos de aspartilo e isoaspartilo, que perturban la estructura de la proteína. La tasa de desamidación depende las condiciones del ambiente como el pH y la naturaleza de los residuos de aminoácidos vecinos (Sternner & Liebl 2008). Otro aminoácido que se observa en menor proporción en proteínas de organismos termófilos es la glicina, en cambio, también se ha notado un aumento en residuos de prolina que estabilizan la estructura y la vuelven más rígida porque las prolinas limitan la rotación del esqueleto de la proteína por la interacción de su anillo pirrolidina con el amino de la cadena principal (Bribibger, Cobani & Evilia 2018).

En algunos casos las proteínas hipertermófilas son poco activas a temperaturas “bajas”, pero tienen eficiencias catalíticas “normales” a altas temperaturas. Se ha postulado que la baja actividad de las enzimas termoestables se debe a la rigidez. De acuerdo con esta hipótesis se ha aumentado la actividad de la proteína sTrpC, aumentando la flexibilidad del loop catalítico. Por otro lado, hay proteínas termoestables que son catalíticamente eficientes a bajas temperaturas. Sugiriendo que la termoestabilidad y la flexibilidad no son características mutuamente excluyentes. En general, las enzimas hipertermófilas son mucho más eficientes catalíticamente que sus homólogas mesófilas a su correspondiente temperatura fisiológica, debido a la inevitable aceleración de las reacciones químicas con la temperatura (Sternner & Liebl 2008).

Todas estas características nos muestran que los cambios pueden ser a todos los niveles estructurales, desde la secuencia hasta la estructura cuaternaria. Algunas de estas características son mutuamente interdependientes. Cuando las familias de proteínas se analizan independientemente parece que solo algunas características son usadas solo en algunas familias proteicas. Cada familia de

proteínas usa una combinación específica para estabilizar las interacciones. Así que las proteínas de los hipertermófilos se estabilizan de varias formas, algunos mecanismos parecen no tener la mínima importancia en ciertas familias de proteínas, pero pueden ser cruciales en otra familia de proteínas (Sternner & Liebl 2008) (Bribibger, Cobani & Evilia)

3. LIMITANTES

3.1 LIMITANTES BIOQUIMICAS

La caracterización fisicoquímica de las proteínas nos provee de mucha más información, pero desafortunadamente es la más costosa y eso implica que podemos abarcar un menor número de secuencias. Además, que muchas veces las condiciones de laboratorio están muy alejadas de las condiciones en las que se encuentra la proteína en el organismo de origen, por lo que pueden adoptar una conformación fuera de la fisiología. Por ejemplo, al estudiar termoestabilidad muchos experimentos se realizan a 25°C, la dificultad surge de que los instrumentos de medición están limitados en los rangos de temperatura y pH que pueden soportar, así que muchas veces no podemos realizar experimentos en las mismas condiciones fisiológicas a las que viven los extremófilos (Park SH, et. al. 2015). Es importante realizar los experimentos a las temperaturas fisiológicas del organismo porque la temperatura estabiliza ciertas interacciones, así que mientras varias interacciones son entrópicas a temperatura ambiente se convierten en entálpicas a altas temperaturas (Sternier & Liebl 2008).

También nos podemos encontrar con mecanismos alternos que utiliza la célula para estabilizar sus proteínas, que perdemos de vista cuando se produce la proteína de interés mediante la expresión heteróloga en el laboratorio. Por ejemplo, los organismos que viven a altas temperaturas no solo aumentan la estabilidad intrínseca de sus proteínas, sino que tienen maquinarias moleculares responsables de reparar enzimas o bien altas concentraciones de solutos estabilizadores. No podemos acceder a los solutos compatibles a altas temperaturas porque dependen de varias vías enzimáticas celulares difíciles de replicar en el laboratorio (Figura 11). Los diferentes sistemas de chaperonas actúan sinérgicamente unos con otros para plegar, desplegar, desagregar, solubilizar y reactivar proteínas. Sin embargo, tenemos poca información acerca de su rol en vivo a altas temperaturas. La mayoría tiene chaperonas que se inducen por cambios en la temperatura llamadas Hsp (heat-shock-protein). Homólogos de la Hsp70 se encuentra en los 3 dominios de la vida, aunque no se ha encontrado presente en todos los clados de las arqueas. La Hsp70 tiene múltiples funciones, pero la principal es actuar como detergente, separando las regiones polipeptídicas hidrofóbicas (Figura 11) (Sternier & Liebl 2008).

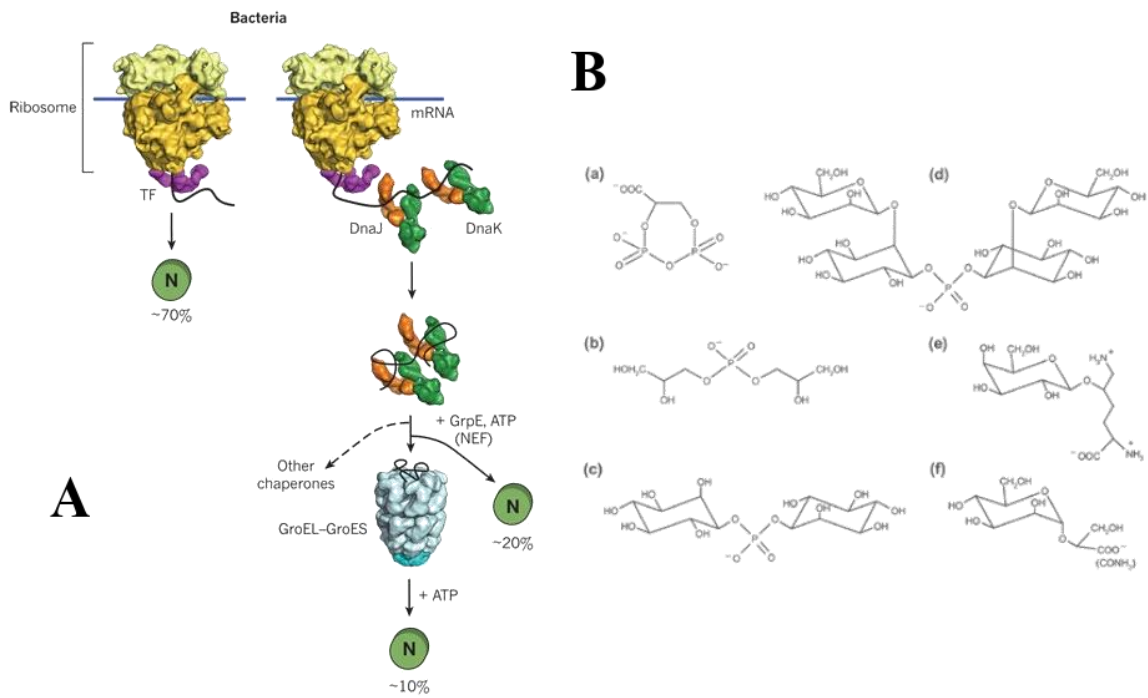


Figura 11. Mecanismos in vivo para aumentar la termoestabilidad. En A) Podemos observar el sistema de chaperonas en procariontes, DnaJ y DnaK corresponden a Hsp70 y Hsp60 respectivamente, se unen a la cadena polipeptidica que sale del ribosoma. En B) observamos solutos presentes en organismos hipertermófilos: a) 2,3-difosfoglicerato ciclico, b) diglicerol fosfato, c) di-mio-inositol-1,1'fosfato, d) di-2-O-β-manosil-di-mio-inositol-1,1'fosfato, e) β-galactopiranosil-5-hidroxisilina, f) α-manosilglicerato Tomado y modificado de (Sterner & Liebl 2008).

3.2 LIMINANTES EVOLUTIVAS

Una de las principales limitantes en esta aproximación es que tenemos acceso solo a una muestra de todas las secuencias compatibles con la vida y puede que dicha muestra no sea representativa. Así que estamos muy limitados por el número de secuencias disponibles de nuestra proteína de interés. Otra limitante es que no es tan sencillo inferir información de las secuencias, para eso se han realizado metodologías como el SCA, sin embargo, estos métodos estadísticos no tienen parámetros tan claros que definan los límites entre aquella información que es significativa de la que no lo es. Además, el resultado siempre cambia conforme cambia la muestra y corremos el riesgo de que la muestra no sea lo suficientemente representativa.

Hablando de evolución hay un tema con el cual tener especial cuidado, “La adaptación”. El que algo sea adaptativo significa que en alguna población ancestral hubo una variación que le permitió una supervivencia y reproducción diferenciada (Oslo & Arroyo 2015). Existe una tendencia de los científicos en enfocarse exclusiva o inmediatamente en la adaptación a condiciones locales y tienden a ignorar las restricciones del medio y solo hacen una invención del proceso (Gould & Lewontin 1979). En biología hay un “error para distinguir entre la utilidad actual y las razones de su origen” (Nielsen 2009). Tenemos que prestar especial atención a las llamadas historias ad-hoc cuya estructura es: –“este rasgo fue formado por selección de las poblaciones ancestrales invisibles y esta selección debió ocurrir porque se conserva ese rasgo en el presente”- La característica de estas historias es que son especulativas y no se pueden probar, se aceptan únicamente porque son razonables o plausibles. Es una historia especulativa porque la presencia de un rasgo en un organismo se utiliza como la única evidencia para inferir la variación heredable del rasgo en una población ancestral y de un régimen selectivo que favoreció esa variable sobre otras. Sin embargo, no se sustentan bajo ninguna evidencia, y puede haber mil hipótesis que sean igual de probables y no habría razón para creer en esa particular (Oslo & Arroyo 2015). El método comparativo ofrece una alternativa metodológica para probar o descartar la adaptación de un rasgo a una población, consiste en la observación cualitativa sobre las características de organismos similares en contextos similares de selección, así se obtiene información suficiente para apoyar o descartar la hipótesis adaptativa (Oslo & Arroyo 2015).

Otra dificultad tiene lugar cuando estudiamos una población con un rasgo específico y no toda la población está sometida a la presión selectiva que selecciona al rasgo. Por ejemplo, las proteínas son marginalmente estables, es decir, están ligeramente por encima del umbral de desplegamiento, este fenómeno nos habla de la inhabilidad de la selección entre distinguir entre hiperestabilidad y las proteínas suficientemente estables (Harms & Thornton 2013). Así mismo pasa

con la termoestabilidad, pues las proteínas de organismos psicrófilos y termófilos no están bajo presión selectiva. Por ejemplo, la Triosafosfato Isomerasa (TIM) de *Methanococoides burtonii*, que es un organismo psicrófilo aislado de un lago en la Antártica que crece naturalmente a 1°C, tiene una temperatura media de desnaturalización aproximadamente de 74°C (Dhaunta et. al. 2013). Este caso ejemplifica como organismos psicrófilos pueden tener proteínas termófilas, porque sus proteínas no están sujetas a la presión selectiva de altas temperaturas (Brininger, Cobani & Evilia 2018). Así que cuando nosotros analizamos y elegimos secuencias de organismos mesófilos y psicrófilos no podemos estar seguros de que no sean termoestables.

3.3 LIMITANTES ESTRUCTURALES

Es muy difícil llegar a una conclusión general acerca de las características estructurales responsables de la termoestabilidad, debido a que las estructuras tridimensionales de proteínas mesófilas y termófilas se parecen mucho, además que hay tantas mutaciones neutras que es muy difícil identificar cuáles son las cruciales en la diferencia de la termoestabilidad (Sterner & Liebl 2008). Debido a esto, se utilizan aproximaciones diferentes para abordar el problema lo cual dificulta un consenso (Szilagyí & Závodszyk 2008).

Otro problema es la cantidad limitada de estructuras disponibles en algunas familias proteicas, y la naturaleza de dichas estructuras, puesto que la mayoría provienen de la cristalografía de rayos X. Algunas desventajas de esta metodología es que los puentes disulfuro y los residuos ácidos son sensibles a radiación. También que solo podemos observar las moléculas de agua estructurales, es decir, moléculas de agua fijas en la estructura, pero no nos permite conocer el comportamiento de la proteína con el soluto (Weik 2003). En solución las proteínas poseen flexibilidad conformacional que engloba un gran rango de estados de hidratación que no se ven en el cristal (Levy & Onuchic 2006)

El soluto es importante para la estructura, estabilidad, dinámica y función de las moléculas biológicas. Así que no se debe tratar como un ambiente inerte sino como un componente integral y activo de los sistemas biomoleculares (Levy & Onuchic 2006). Por ejemplo, solutos de bajo peso molecular pueden tener varios efectos en la solubilidad y estabilidad de las proteínas, por ejemplo; a bajas concentraciones de sales las proteínas pueden aumentar su solubilidad, mientras que en altas concentraciones de sales las proteínas pueden precipitar. Los compuestos estabilizadores incluyen azúcares (sacarosa, trealosa, glicerol, manitol y sorbitol), aminoácidos como la prolina y glicina, metil aminas (sarcosina, trimetilamina-N-o, y betamina) y ATP. Así como numerosos componentes que se han encontrado específicamente solo en termófilos e hipertermófilos (Figura 11) (Sterner & Liebl 2008). También el agua actúa como lubricante facilitando los cambios de las parejas de puentes de hidrógeno responsables de las fluctuaciones conformacionales. La transición dinámica del agua en relación con la temperatura depende mucho de la viscosidad del solvente (Levy & Onuchic 2006)

Es difícil predecir la acción de los agentes estabilizantes a altas temperaturas al igual que el agua. Se necesitan más estudios del agua confinada por moléculas biológicas y también modelos no biológicos, los cuales son necesarios para esclarecer el papel del soluto en la dinámica y función de las proteínas en ambientes a altas temperaturas y tan complejos como los de una célula viva (Weik, 2003).

EPÍLOGO

Si buscamos entender a profundidad cualquier fenómeno biológico, es necesario tomar en cuenta que todas las aproximaciones tienen sus ventajas y limitaciones. Haciendo una analogía con la historia de "los ciegos y el elefante", nuestras aproximaciones serían los ciegos puesto que todas tienen limitaciones, y el elefante sería el objeto de estudio o fenómeno biológico que queremos conocer (Figura 12).

"Un grupo de ciegos escuchó que un extraño animal, llamado elefante, había sido traído al pueblo, pero ninguno de ellos era consciente de su figura y forma. Por curiosidad, dijeron: "Hay que inspeccionarlo y conocerlo al tacto, de lo que somos capaces". Entonces, lo buscaron, y cuando lo encontraron, lo buscaron a tientas. La primera persona, cuya mano se posó en la trompa, dijo: "Este ser es como una serpiente gruesa". Para otro cuya mano llegaba a su oreja, dijo que parecía una especie de abanico. En cuanto a otra persona, cuya mano estaba sobre su pata, dijo, el elefante es como el tronco de un árbol. El ciego que puso su mano en su costado dijo que el elefante "es una pared". Otro que sintió su cola, lo describió como una cuerda. El último sintió su colmillo, indicando que el elefante es lo que es duro, liso y como una lanza."



Figura 12. Ilustración de la historia de los ciegos y el elefante

Esta historia nos enseña que el fenómeno es mucho más que la suma de sus partes y que para dilucidar bien un fenómeno natural debemos mirar desde varios ángulos, mientras podamos reunir más puntos de vista nos acercaremos cada vez más a una descripción de la realidad, aunque es difícil conocer con certeza si la

descripción es correcta, tendremos mayor robustez para la explicación del fenómeno.

Esta tesis tratará de utilizar la aproximación evolutiva por medio de secuencias, la aproximación estructural por medio de estructuras cristalográficas, obtenidas por mi compañero de laboratorio Sergio Romero y algunos datos bioquímicos, obtenidos experimentalmente por mi compañero de laboratorio Ricardo Muñiz, para encontrar las características que le confieren a la Triosafosfato Isomerasa de arqueas su termoestabilidad. Resulta ser que la TIM es una excelente opción para investigar la termoestabilidad porque se encuentra ampliamente distribuido en el árbol de la vida; contamos con muchas secuencias, estructuras y datos bioquímicos disponibles, tanto de organismos termófilos como mesófilos. Acompáñeme a descubrir sus características.

4. MODELO DE ESTUDIO

4.1 EL PLEGAMIENTO $(\beta/\alpha)_8$

El plegamiento $(\beta/\alpha)_8$ es un dominio estructural conservado en los tres dominios de la vida, está formado por ocho hebras beta y ocho hélices alfa que se intercalan unas con otras. Las hebras betas se encuentran dispuestas en una lámina paralela formando una elipse cerrada; dando estructura al centro del barril. Mientras que las hélices alfa se encuentran rodeando esta lámina (Figura 13). Los elementos de estructura secundaria están unidos por asas, de manera que hay 8 asas conectando el carboxilo terminal de las hebras con las hélices, denominadas asas $\beta\alpha$; y hay 8 asas $\alpha\beta$ que conectan el carboxilo terminal de las hélices con las hebras (Figura 3) (Wierenga, 2001).

El plegamiento $(\beta/\alpha)_8$ también es conocido como “Barril TIM” porque la lámina de hebras beta en el centro asemeja un barril y TIM porque fue descrito por primera vez en la proteína Triosafosfato Isomerasa (TIM). Este plegamiento $(\beta/\alpha)_8$ está ampliamente distribuido en la naturaleza, se encuentra presente en los tres dominios de la vida y es el más común en las vías de metabolismo celular, se calcula que está presente aproximadamente en el 85% de todas las vías metabólicas. Además, de ser el más abundante de las estructuras conocidas, se estima que aproximadamente el 10% de las estructuras depositadas en el PDB tiene este tipo de dominio estructural. También posee una amplia diversidad de funciones, ya que es parte de 76 familias de proteínas y catalizan 6 de las 7 funciones catalíticas (Nagano & Orengo, 2002).

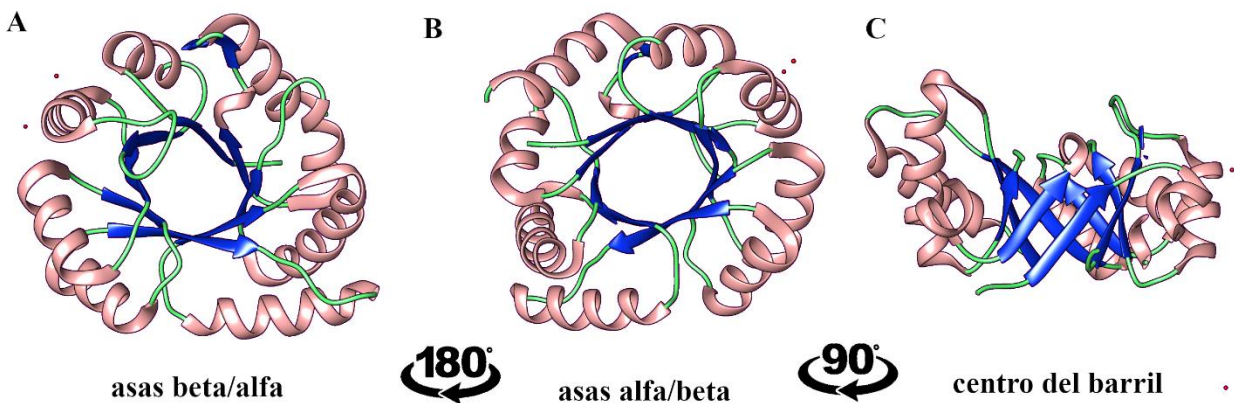


Figura 13. Plegamiento Barril TIM $(\beta/\alpha)_8$. Usando la enzima Fosforibosil Antranilato Isomerasa (PRAI) de *Thermococcus kodakáreaensis* como ejemplo. PDB: 5LHF. A) asas α/β , B) asas β/α , D) laminas beta y centro del barril.

Debido a su diversidad y abundancia, el Barril TIM ha llamado la atención como modelo de estudio en diversos campos como la biología sintética, el diseño de proteínas y la biología evolutiva. Resulta particularmente interesante en el campo

de la biología evolutiva porque existen evidencias que apoyan la hipótesis de que todas las proteínas con plegamiento $(\beta/\alpha)_8$ tienen un origen monofilético, es decir, que todas las estructuras provienen de un ancestro en común. Se hipotetiza que dicho ancestro debió estar involucrado en las vías metabólicas centrales. Los barriles TIM tienen dominios β/α internos repetitivos, por lo que se especula que el ancestro puede ser resultado de la duplicación de un gen y su fusión. La evidencia que apoya esta idea es que en algunos barriles TIM el amino (N-) y el carboxilo (C-) terminal se pueden reconocer como unidades evolutivas separadas (Goldman et al., 2016), además de que todos los barriles tienen el sitio catalítico en el extremo del barril formado por los extremos C-terminal de las hebras beta, la llamada "cara catalítica" (Wierenga, 2001).

Los barriles TIM son un buen modelo para la biología evolutiva ya que tienen un fenotipo robusto, es decir, los cambios en la secuencia suceden sin afectar el plegamiento. Como ya discutimos anteriormente la capacidad de explorar cambios neutros permite una gran variedad de trasfondos genéticos y por lo tanto una mayor capacidad adaptativa (Muñiz, 2020). Está claro como los pequeños cambios en la secuencia de los barriles TIM nos permiten cambiar de función catalítica sin perder su plegamiento. Lo que es experimentalmente conveniente porque nos ofrece mucha información acerca de los procesos adaptativos.

4.2 ENZIMA TRIOSAFOSFATO ISOMERASA

El miembro fundador de la superfamilia de los barriles TIM es la proteína Triosafosfato Isomerasa (5.3.1.1), la cual es una enzima altamente conservada en los tres dominios de la vida, que cataliza la interconversión del D-gliceraldehído-3-fosfato (G3P) a Dihidroxiacetona fosfato (DHAP) (Figura 14). Esta enzima participa en diferentes vías principales del metabolismo tales como: la glucólisis, gluconeogénesis, síntesis de triglicéridos y en el ciclo de Calvin, se localiza en el citoplasma, cloroplastos y mitocondrias. La TIM es muy importante para el correcto funcionamiento del organismo ya que existen enfermedades graves asociadas a mutaciones puntuales que provocan un descenso en la actividad de la enzima (Oroz, Oláh & Ovádi, 2009). Apoyando esta afirmación, algunos organismos anaerobios son dependientes de la actividad enzimática de la TIM, debido a que su principal fuente de energía es la glucólisis, por lo que la TIM ha sido blanco en el diseño de drogas dirigidas contra endoparásitos (Enríquez-Flores, Romero, Alcántara, et. al., 2011).

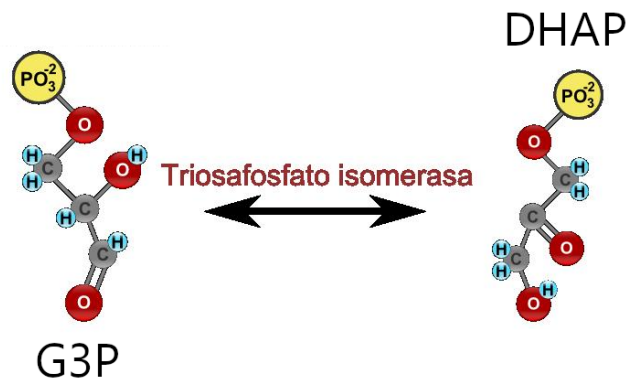
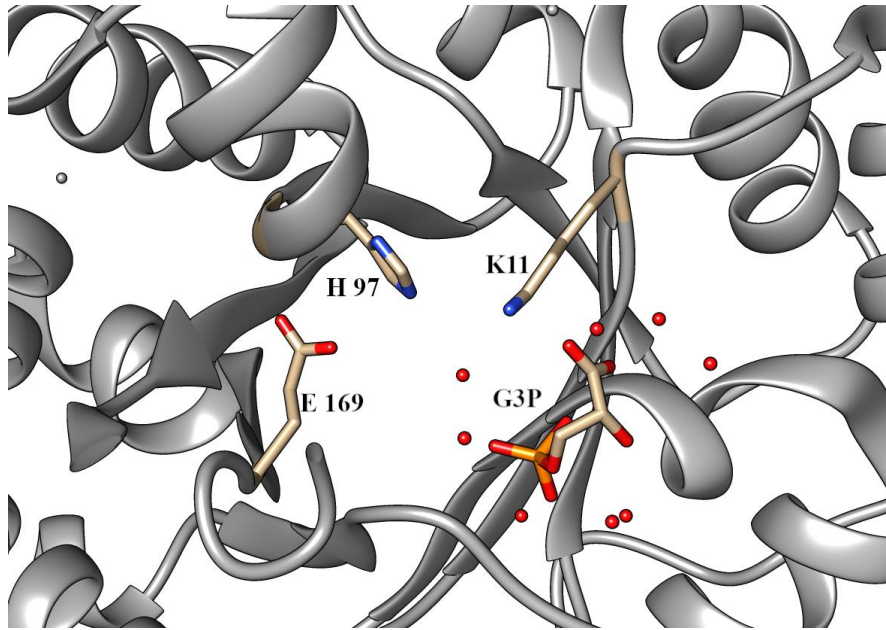


Figura 14. Sustrato y producto de la enzima Triosafosfato Isomerasa. Tomado y modificado de Universidad de Santiago Virtual.

La TIM se considera una enzima catalíticamente perfecta debido a que sus valores de k_{cat}/K_M están cerca del máximo valor posible $\sim 10^9 \text{ M}^{-1} \text{ s}^{-1}$, lo que significa que es una enzima limitada por difusión, es decir, la formación del producto ocurre inmediatamente cuando se encuentran la enzima con el sustrato, así que la limitante es cuánto tiempo tarda en desplazarse el sustrato a la enzima en solución (Blacklow, et. al., 1988). El sitio activo se encuentra en la cara de las asas $\beta\alpha$, en la catálisis participan 3 aminoácidos conocidos como la triada catalítica: la lisina 11, la histidina 97 y el glutámico 169 (según la numeración de la TIM de *Staphylococcus aureus* Figura 15, PDB: 3UWW) (Mukherjee, Roychowdhury, Dutta & Das, 2012). Estos aminoácidos están conservados en el 100% de las secuencias de la TIM reportadas.



*Figura 15. Aminoácidos catalíticos de la TIM. K11, H97 & E169 se muestran en líneas gruesas rosas. El análogo del sustrato Glicerol-3-fosfato (G3P), está entre estos aminoácidos y se muestra en líneas gruesas amarillas. Numeración de la TIM de *Staphylococcus aureus*, PDB: 3UWW.*

El mecanismo de catálisis inicia con la estabilización del sustrato en el sitio activo por medio de puentes de hidrógeno entre el fosfato del sustrato y la cadena principal de la proteína. El primer paso de la catálisis ocurre mediante un intercambio ácido-base entre los aminoácidos catalíticos y la dihidroxiacetona fosfato. El ácido glutámico 169 funciona como base extrayendo un protón del C1 de la DHAP, y la histidina 97 funciona como ácido al donar un protón al oxígeno carbonílico del DHAP; el intermediario formado en este primer paso es un enediol (Zhang Z., et. al., 1991); (Wierenga, et. al., 1992). Es de vital importancia estabilizar y aislar al enediol, puesto que al entrar en solución se separa rápidamente del fosfato, formando un compuesto llamado metilglioxal, altamente tóxico para la célula (Lolis y Petsko, 1990). En el segundo paso de la catálisis, la Histidina 97 funciona como base al extraer un protón del grupo hidroxilo del C1 del enediol, mientras que el glutámico 169 funciona esta vez como ácido al soltar el protón con el que se había quedado al C2 del enediol; formando así G3P (Zhang et. al., 1999) (Figura 16).

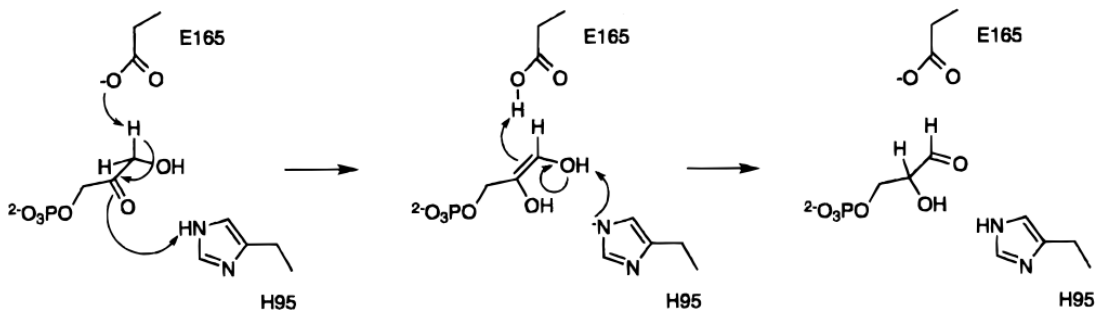


Figura 16. Mecanismo de catálisis de la TIM. Tomado de (Zang et. al. 1999)

La función de la lisina 11, al estar cargada positivamente, es estabilizar electrostáticamente el estado de transición (el enediol) que se encuentra cargado negativamente. Además, la proteína cuenta con un asa flexible (el asa $\beta\alpha 6$), que sirve como aislante del medio acuoso, previniendo la formación del metilglioxal (Wierenga et. al, 1991). El asa 6 se mueve aproximadamente 7 angstroms, adoptando una conformación cerrada cuando el sustrato se encuentra en el sitio activo. (Figura 17).

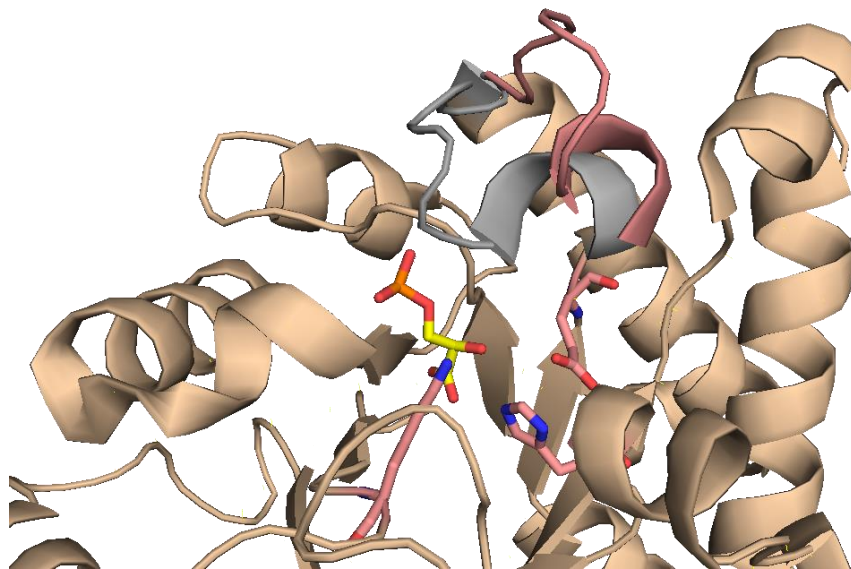


Figura 17. Conformación abierta y cerrada de la TIM. En rosa se encuentra la conformación abierta, en gris la conformación cerrada. El sustrato se encuentra en amarillo con líneas gruesas y los aminoácidos catalíticos en rosa con líneas gruesas. Ambas estructuras pertenecen a Staphylococcus aureus, conformación abierta PDB: 3M9Y, conformación cerrada 3UWW.

Las TIMs tienen en promedio 250 aminoácidos de longitud y se encuentran en la naturaleza como homodímeros obligados, es decir, que sólo las TIM diméricas son completamente activas; la interfase dimérica se forma gracias al asa $\beta\alpha 3$, conocida como el asa de interfase, dicha asa se extiende desde una subunidad monomérica

hacia la cavidad que forman las asas $\beta\alpha$ 1 y $\beta\alpha$ 4 del otro monómero (Lolis y Petsko, 1900) (Figura 18), el dímero se estabiliza por fuerzas de van der Waals e interacciones hidrofóbicas, entre las estructuras antes mencionadas. Actualmente, no existen reportes de TIMs monoméricas catalíticamente activas, este hecho sugiere que la estabilización del sitio activo depende de la oligomerización (Nájera et. al., 2003).

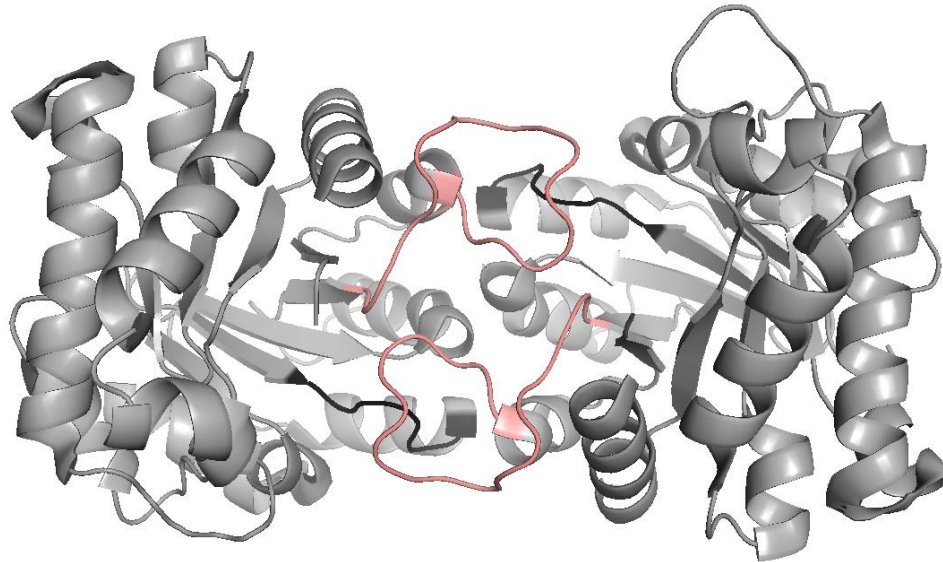


Figura 18. Conformación dimérica de la TIM. En rosa se muestra el asa 3 de la interfase dimérica y en negro el asa 1. Estructura perteneciente a Staphylococcus aureus PDB: 3UWW

La proteína TIM esta conservada en la mayoría de los seres vivos, sin embargo, existe un gran rango de porcentaje de identidad entre estas proteínas, el mayor porcentaje de identidad es de 100% entre la TIM de humano y la TIM de chimpancé, el más bajo es del 12% entre la TIM de *Plasmodium falciparum* y la arquea *Pyrococcus furiosus* (Peimbert et. al., 2008). Esto nos ejemplifica la robustez mutacional de la TIM, por esta razón y su amplia distribución en la naturaleza es un excelente modelo para estudiar biología evolutiva.

LAS TIMs DE ARQUEAS

La identidad entre bacterias y eucariontes es de aproximadamente 40%. El análisis secuencial revela que las TIMs provenientes de arqueas tienen 20% - 25% de identidad con sus contrapartes bacteriales y eucariontes. Además, las secuencias de las TIMs de Arqueas (ArqTIMs) son de un rango de 219 a 231 residuos, lo que las hace 10% más cortas que sus contrapartes, por lo que son más compactas en su estructura (Sheramm, Kohlhoff & Hensel, 2001). Las TIMs de arqueas son homodímeros obligados y se unen estructuralmente de la misma forma que las TIMs de bacteria y eucaria, por el asa 3 que se extiende desde una subunidad monomérica hacia la cavidad que forman las asas 1 y 4 de la otra subunidad (Figura 19). Una particularidad de las TIMs arqueas, es que algunas se oligomerizan como homotetrámeros, esto es, como dímeros de dímeros (Figura 19). Estas características permiten su separación en un grupo monofilético bien definido (Ronimus & Morgan, 2003). Sin embargo, se cuentan con pocas estructuras cristalográficas; hasta el momento (febrero 2022) se han reportado un total de 5 estructuras depositadas en el *Protein Data Bank* (PDB), procedentes de 4 especies de arqueas distintas. Estas estructuras han permitido determinar que la interfaz dimérica se forma de la misma manera que en otras TIMs y que la superficie de tetramerización no involucra ni puentes salinos ni puentes de hidrógeno, formándose principalmente por interacciones hidrofóbicas (Walden et. al., 2001). Las estructuras secundarias involucradas en la formación del tetrámero son la hélice 4 con algunos aminoácidos que la preceden del asa $\beta\alpha_4$; la hélice 5 que en su mayoría es rica en alaninas y los últimos aminoácidos de la hélice 6 (Gayathri et. al., 2006) (Figura 19).

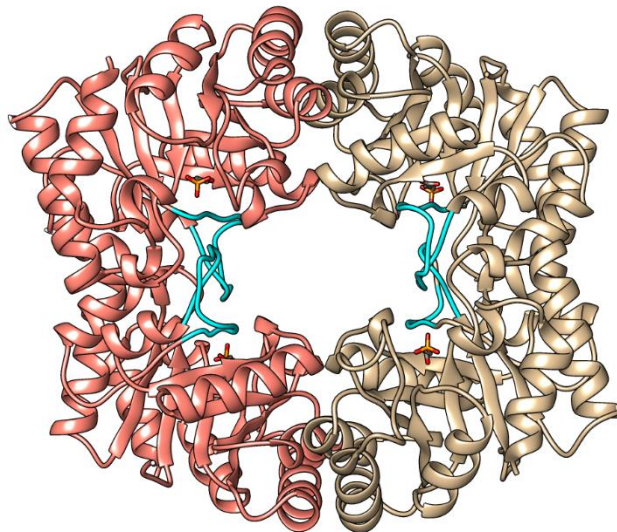


Figura 19. Triosafosfato Isomerasa de Arquea. Homotetrámero. Se observa la formación como dímero de dímeros. El asa 3 se encuentra en color cian. Estructura de Pyrococcus furiosus, PDB: 1HG3.

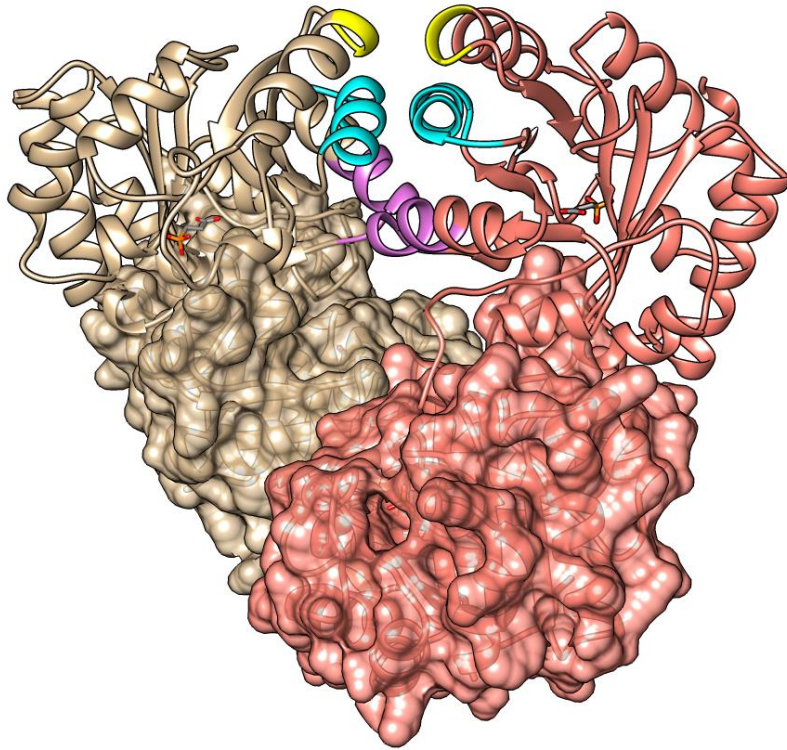


Figura 20. Interfaz tetramérica de las TIM de arqueas. En magenta se encuentra el asa beta/alfa4 y el amino terminal de la hélice 4, en cian la hélice 5 y en amarillo el carboxilo terminal de la hélice 6. Estructura de Pyrococcus furiosus PDB: 1HG3.

Las arqueas son el grupo dominante en organismos termófilos e hipertermófilos (Figura 1) (Schmid, Allers & DiRuggiero, 2020). Este hecho y su separación filogenética bien definida, hace que las TIMs de arqueas sean un buen modelo para estudiar los mecanismos moleculares por los que las proteínas se adaptan a temperaturas elevadas. La tetramerización se ha propuesto como el mecanismo adaptativo mediante el cual las TIM de arqueas se mantienen estables a temperaturas elevadas. Sin embargo, esta afirmación forma parte de los argumentos *ad hoc* antes mencionados y carece de pruebas experimentales para confirmar la veracidad del argumento.

5. ANTECEDENTES

5.1 ANTECEDENTES BIOQUÍMICOS

Hasta el momento existen pocos estudios fisicoquímicos sobre las TIMs de Arqueas, pero se han determinado algunas características biofísicas como la termoestabilidad, actividad enzimática y estado oligomérico para algunas especies. Un antecedente fundamental para este trabajo es la tesis de maestría de mi compañero Ricardo Muñiz, que propone un estudio comparativo filogenético entre diferentes TIMs de arqueas del grupo de los Methanococcales, con el fin de determinar si las características biofísicas de estas TIMs se seleccionaron de acuerdo con la temperatura óptima de crecimiento del organismo (Muñiz, 2020).

Con este fin, se seleccionaron cinco especies con el fin de abarcar cuatro géneros del clado Methanococcales y tener un amplio rango en las temperaturas de crecimiento óptimo de estos organismos (Figura 21). Las especies seleccionadas fueron: *Methanocaldococcus jannaschii*, *Methanotorris formicicus*, *Mathanothermococcus okinawensis*, *Methanococcus aelicus* y *Methanococcus maripaludis*. Nos referiremos a las TIMs de estas especies como: MjaTIM-85, MfoTIM-75, MokTIM-60, MaeTIM-46, MmaTIM-38; respectivamente. La nomenclatura está dada por M del orden Methanococcales, las siguientes dos letras minúsculas son las primeras dos letras del nombre específico de la especie y el número consecuente es la temperatura óptima de crecimiento. Para fines de análisis, MjaTIM-85 & MfoTIM-75 se clasificaron como TermoTIMs, y MokTIM-60, MaeTIM-46 y MmaTIM-38 se clasificaron como MesoTIMs (Muñiz, 2020).

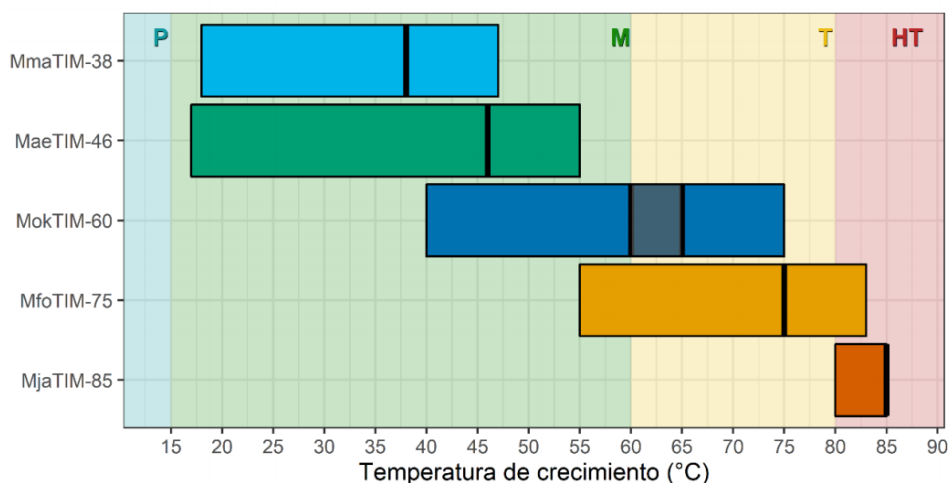


Figura 21. Temperaturas de crecimiento de las especies seleccionadas para el trabajo de maestría de Ricardo Muñiz. Tomado de Muñiz, 2020

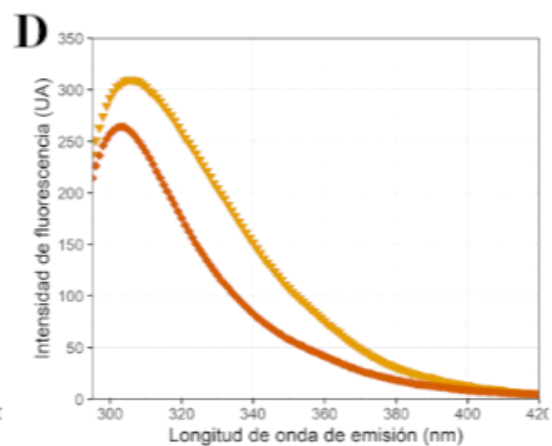
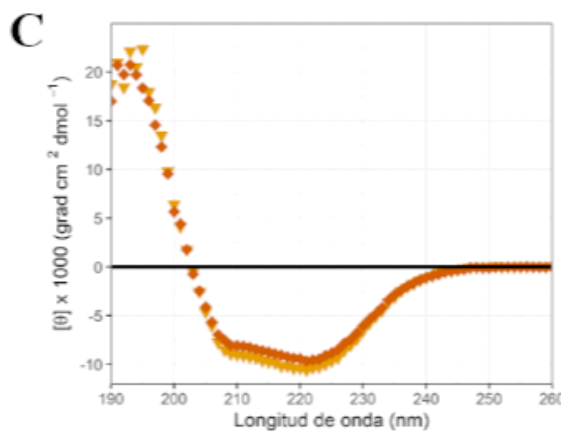
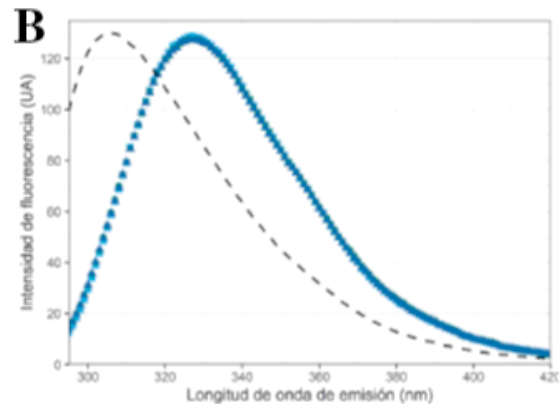
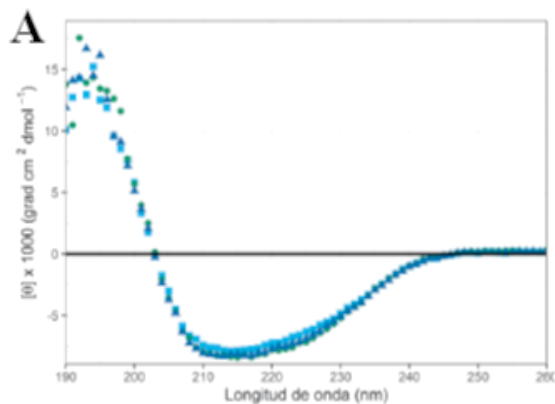
Las MesoTIMs no se pudieron purificar a homogeneidad, presentando algunos contaminantes en menor proporción (Figura 67). Sin embargo, se realizó una

caracterización espectroscópica para determinar el estado del plegamiento, estructura secundaria y terciaria a 25°C. Al realizar el dicroísmo circular se encontró que el espectro correspondía al plegamiento de un barril TIM, no obstante, los porcentajes de estructura no corresponden a este plegamiento (Tabla 1). También se encontró en el UV lejano que tenían un máximo de emisión a ~327 nm, característico de proteínas con triptófanos (Figura 22 A, B), sin embargo, estas MesoTIM no tienen triptófanos. Estos resultados anómalos son probablemente derivados de los contaminantes presentes. Por lo que no se pudo continuar con la caracterización de las MesoTIMs. El caso de las MesoTIMs es similar al de la TIM de *Methanococoides burtonii* (MbuTIM) que no se pudo seguir estudiando a profundidad porque la secuencia carece de triptófanos, por lo que no es posible monitorear las transiciones de la estructura terciaria, según los autores (Dhaunta, Guptasarma et. al. 2015).

Por el contrario, las TermoTIM se pudieron obtener con un alto grado de pureza e igualmente se caracterizaron espectroscópicamente. Resultó que los espectros son congruentes con un plegamiento de barril TIM y sus porcentajes de estructura corresponden perfectamente con lo observado en la estructura cristalográfica 2H6R de la MjaTIM-85 (Figura 22) (Tabla 1) (Muñiz, 2020). También se han caracterizado estructuralmente dos TIMs termófilas la TIM de *Pyrococcus furiosus* (PfuTIM) y la TIM de *Thermococcus onnurius* (TonTIM) (Sharma & Guptasarma 2015).

Parámetro	MesoTIMs			TermoTIMs	
	MmaTIM-38	MaeTIM-46	MokTIM-60	MfoTIM-75	MjaTIM-85
De convolución de los espectros de dicroísmo					
Hélices α	20%	22%	22%	33%	38%
Hebras β	30%	29%	30%	22%	18%
Estructura no repetitiva	49%	48%	48%	44%	45%
Espectros de fluorescencia					
Intensidad máx (UA)	128.8	127.6	127.4	309.8	264.1
máx (nm)	327	326	328	303	302

Tabla 1. Caracterización espectroscópica de las TIMs de Arqueas, tomado de Muñiz,2020.



Leyenda ▼ MfoTIM-75 ◆ MjaTIM-85

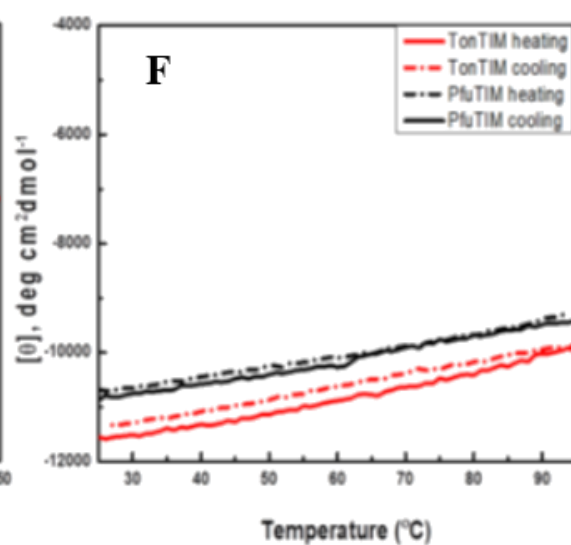
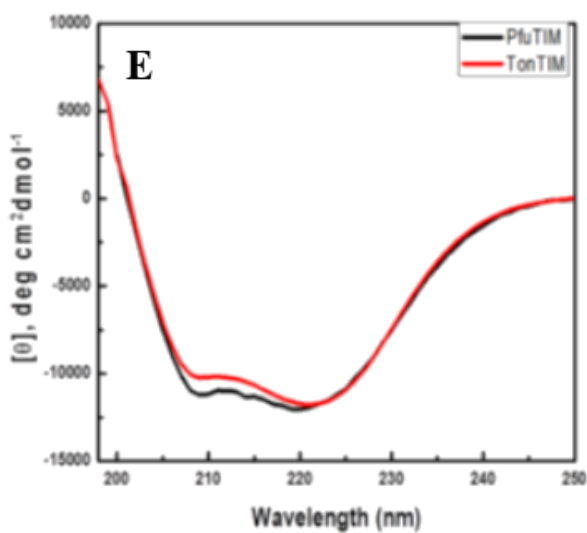


Figura 22. Dicroísmo circular de las MesoTIMs, TermoTIMs, PfuTIM y TonTIM. A) Muestra el UV cercano de las MesoTIMs, espectro característico de barril tipo TIM. B) UV lejano, espectro característico de proteínas con triptófano. C) UV cercano de TermoTIM, (MjaTIM línea roja) (MfoTIM línea naranja) espectro característico de Barril tipo TIM. D) UV lejano de las TermoTIM, espectro característico de proteínas sin triptófano. E) UV cercano de PfuTIM (línea roja) y TonTIM (línea negra). F) Estructura contra temperatura PfuTIM (línea roja) y TonTIM (línea negra). Tomado y modificado de (Muñiz, 2020) & (Sharma & Guptasarma 2015).

Continuando con la caracterización fisicoquímica de las TermoTIMs, se determinó su estado oligomérico por medio de cromatografía de exclusión molecular. Indicando que las TermoTIM son tetraméricas independiente de la concentración proteica (Tabla 2). Por medio de esta técnica y de dispersión dinámica de luz (DLS) se ha determinado el estado oligomérico de la TIM de diferentes organismos, la TIM de termófilos como *Pyrococcus furiosus*, *Thermococcus onnurineus* y *Methanothermus fervidus* (MfeTIM) son tetrámeros (Sharma & Guptasarma 2015) (Kohlhoff, Dahm & Hensel 1996). En cambio, las TIMs de *Methanobacterium bryantii* (MbrTIM) y *Methanococcoides burtonii* (MbuTIM) resultaron ser dímeros en solución, se determinó el estado oligomérico de MbuTIM por medio de DLS y ultra centrifugación analítica, las temperaturas óptimas de crecimiento de estos organismos son 37°C y 23°C respectivamente (Kohlhoff, Dahm & Hensel 1996) (Dhaunta, Guptasarma et. al. 2015). Otro dímero en solución es la TIM de *Thermoplasma acidophilum* (TacTIM) ésta es una arquea termófila, con temperatura óptima de crecimiento entre los 55°C y los 60°C que crece a valores de pH entre 0.5 y 2.0, (Park SH et. al. 2015)

TIM	Método experimental	Estado oligomérico	Referencia
MjaTIM	Cromatografía de exclusión analítica	Tetrámero	Muñiz 2020
MfoTIM	Cromatografía de exclusión analítica	Tetrámero	Muniz 2020
TonTIM	Cromatografía de exclusión analítica y DLS	Tetrámero	Guptasarma et. al. 2015
PfuTIM	Cromatografía de exclusión analítica y DLS	Tetrámero	Kohlhoff et.al. 1996
MfeTIM	Cromatografía de exclusión analítica	Tetrámero	Kohlhoff et. al. 1996
TtxTIM	Ultra centrifugación analítica	Dímero/Tetrámero	Walden et. al. 2004
MbrTIM	Cromatografía de exclusión analítica	Dímero	Kohlhoff et. al. 1996
MbuTIM	Cromatografía de exclusión analítica, DLS y ultra centrifugación analítica	Dímero	Guptasarma et. al. 2015

TacTIM	Cromatografía de exclusión analítica y ultra centrifugación analítica.	Dímero	Park SH et. al. 2015
---------------	--	--------	----------------------

Tabla 2. Estado oligomérico de las TIMs de Arquea.

El caso de la TIM de *Thermoproteus tenax* es interesante, ya que es un organismo hipertermófilo con una temperatura óptima de crecimiento de 86°C, que no forma tetrámeros estables; sin embargo, por medio de ultra centrifugación analítica se mostró que la enzima presenta un equilibrio dimero/tetrámero, la tetramerización depende de la concentración proteica, a 25°C tiene una constante de disociación $K_D = 7 \times 10^{-8} M$. Así mismo, la inactivación depende de la concentración enzimática, lo cual sugiere que los dímeros y los tetrámeros difieren en actividad enzimática, ya que la actividad se pierde completamente a concentraciones bajas de proteína, esto indica que los dímeros son prácticamente inactivos. Pero al aumentar la concentración a 500 µg/ml se reestablece el 100% de actividad catalítica, indicando que los tetrámeros son catalíticamente activos (Walden et. al. 2004).

Otra característica biofísica que se ha explorado de las TIMs de arqueas es la actividad catalítica, en los ensayos enzimáticos se determinan varios parámetros: la K_M nos habla de la afinidad del sustrato por la proteína, la V_{max} que es la velocidad límite que puede alcanzar la reacción, la k_{cat} que es la constante catalítica que nos habla del número de moléculas de sustrato convertidas en producto por molécula de enzima y unidad de tiempo. En los ensayos enzimáticos de las TIMs se analiza la conversión de gliceraldehido-3-fosfato a dihidroxiacetona fosfato (DHAP) acoplados con la enzima α -glicerol-3-fosfato deshidrogenasa, la cual reduce el DHAP α -glicerol-3-fosfato a través de la oxidación de la molécula NADH (Gomez-Puyou et. al. 1995) a partir de la cantidad de proteína en la celda de reacción y el coeficiente de absorción molar (ϵ) del NADH ($6220 M^{-1} cm^{-1}$) se puede calcular la actividad específica en moles de DHAP formados por minuto por miligramo de proteína. Con esta metodología se evaluaron en nuestro laboratorio las cinéticas de las Meso y las TermoTIMs a 25°C con diferentes concentraciones de G3P (Muñiz 2020). Se eligió dicha temperatura para poder comparar datos de otras TIM porque en el laboratorio no era posible realizar ensayos en la temperatura óptima de crecimiento de los organismos hipertermófilos, porque la enzima acoplante se inactiva por arriba de los 65°C (Merk, 2020).

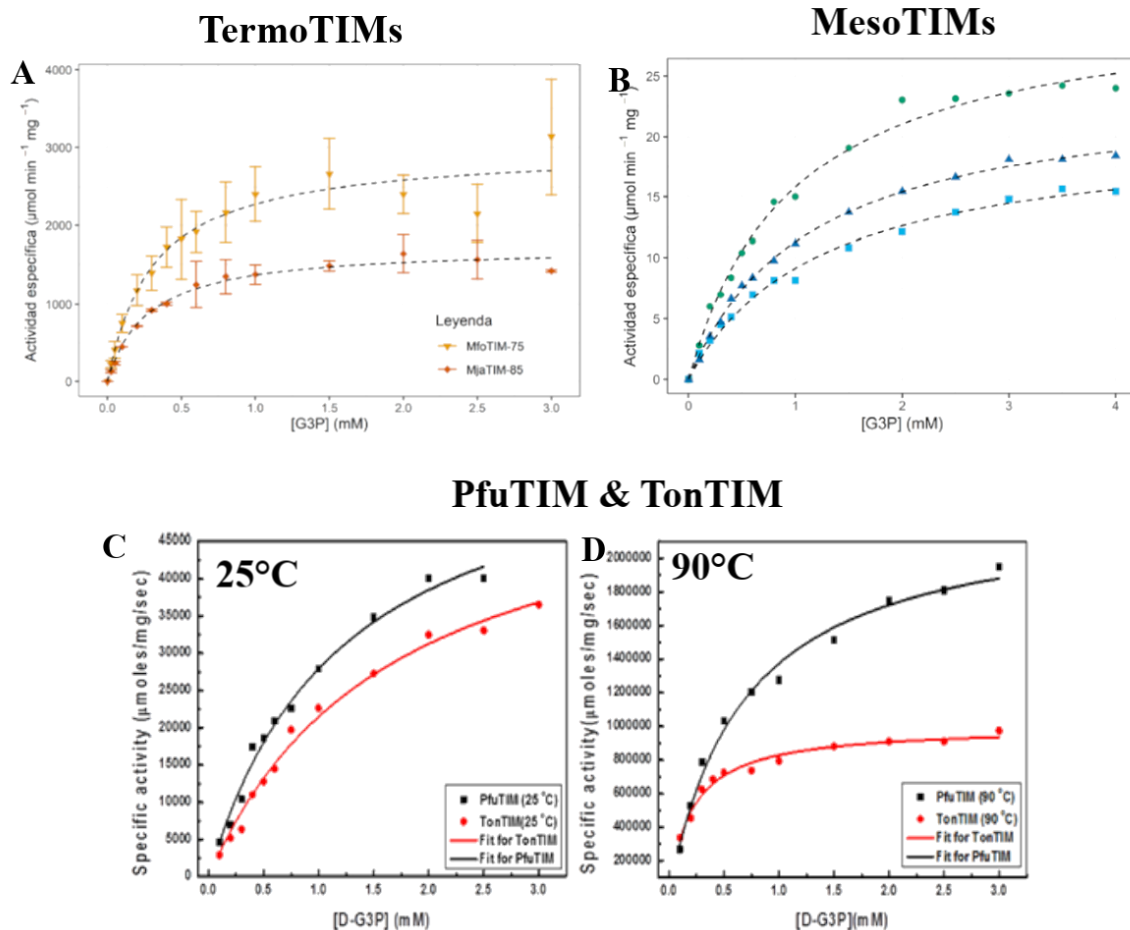


Figura 23. Ensayos catalíticos de las TIMs de Arqueas. A) TermoTIMs, Mja (rombos rojos) & Mfo (triángulos naranjas). B) MesoTIMs, MmaTIM-38 (cuadros azules), MaeTIM-46 (círculos verdes), MokTIM-60 (triángulos azules) C) PfuTIM (negro) & TonTIM (rojo) a 25°C. D) PfuTIM (negro) & TonTIM (rojo) a 90°C. Tomado y modificado de (Muñiz, 2020) & (Sharma & Guptasarma 2015).

Los datos nos muestran como las TermoTIMs tienen una eficiencia catalítica cerca o sobre el límite de difusión ($10^8 \text{ M}^{-1} \text{ s}^{-1}$) (Muñiz 2020) (Tabla 3). En 2015 el grupo de Guptasarma determinó por la misma técnica las eficiencias catalíticas otras proteínas TIM termófilas, PfuTIM y TonTIM, al igual que las TermoTIM muestran eficiencias catalíticas comparables con sus contrapartes mesófilas a temperatura ambiente (Tabla 3). Además, este grupo logró determinar sus eficiencias catalíticas (k_{cat}) a 90°C, por medio de un ensayo de actividad enzimática basado en dicroísmo circular que no necesita una enzima acoplante. Lo que observaron es que a 90°C las proteínas incrementan su k_{cat} 100 veces (Tabla 3) (Figura 23) (Sharma & Guptasarma, 2015). Otro grupo de investigación logró obtener los valores de K_M y V_{max} de la TIM de *Thermoproteus tenax* (TtxTIM), pero se midieron solo a 70°C para evitar la degradación de los sustratos. TtxTIM tiene una V_{max} de 6200

unidades/mg con una enzima auxiliar GLPDH de *Methanothermobacter thermatotrophicus*, pero cuando se cambia esta enzima por la GLPDH originaria de *Thermoproteus tenax* la Vmax aumenta a 9500 unidades/mg sugiriendo una interacción específica entre la TIM y la GLPDH de T. tenax (Walden et. al. 2004).

Parámetros	MmaTIM	MaeTIM	MokTIM	MjaTIM	MfoTIM	PfuTIM		TonTIM		TtxTIM
Temperatura	25°	25°	25°	25°	25°	25°	90°	25°	90°	70°
V _{max} (μmol min ⁻¹ mg ⁻¹)	20.41	31.47	24.15	1723	2990	ND	ND	ND	ND	9500
K _M (mM)	1.23	0.99	1.14	0.26	0.21	1.21	0.67	1.66	0.20	ND
k _{cat} (min ⁻¹)	4.67x10 ²	7.55x10 ²	5.67x10 ²	3.67x10 ⁴	7x10 ⁴	1.5x10 ⁶	0.57x10 ⁸	1.43x10 ⁶	0.25x10 ⁸	ND
k _{cat} /K _M (M ⁻¹ sec ⁻¹)	1.84x10 ⁵	3.69x10 ⁵	2.4x10 ⁵	1.6x10 ⁸	7.5x10 ⁷	1.26x10 ⁹	0.84x10 ¹¹	0.85x10 ⁹	1.24x10 ¹¹	ND
Actividad (U/mg)	ND	ND	ND	ND	ND	6.19x10 ⁴	2.3x10 ⁶	5.73x10 ⁴	1x10 ⁶	ND

Tabla 3. Ensayos enzimáticos reportados para las TIMs de Arqueas. Valores tomados de: (Muñiz, 2020), (Sharma & Gupta 2015) & (Walden et. al. 2004)

El grupo de investigación de Walden observaron que a 80°C la TtxTIM se inactiva rápidamente. Interesantemente la inactivación se puede prevenir al añadir la GLPDH. Utilizando la GLPDH de *M. thermatotrophicus* la actividad regresa al 95% a 80°C. Pero al utilizar la GLPDH originaria de *T. tenax*, la TtxTIM se estabiliza completamente, regresando su actividad al 100%. Se intentó la misma estabilización a base de solutos moleculares como: glicerol, inositol, trehalosa, KCl, K₂HPO₄, K₂SO₄ a concentraciones de 100mM, los cuales no previnieron la inactivación (Figura 24) (Walden et. al. 2004). A partir de estos experimentos, es claro que TtxTIM requiere interacciones proteína-proteína adicionales para su estabilización. Otro caso en donde podemos observar la estabilización por agentes externos a una TIM de arquea es la TIM de *Methanothermobacter fervidus* que se estabiliza con sales de potasio y con 2,3-difosfoglicerato, que se acumula naturalmente en *M. fervidus* en altas concentraciones ~300mM (Shramm, Kohlhoff & Hensel). También *T. tenax* acumula trehalosa a concentraciones aproximadamente de 60mM (Walden et. al. 2004). Así que es altamente probable que diferentes solutos estén estabilizando estas proteínas *in vivo*.

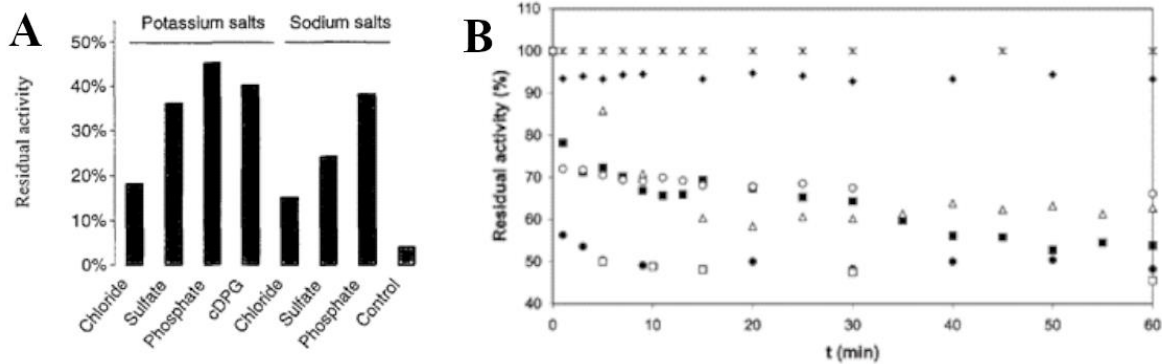


Figura 24. Mecanismos extrínsecos de estabilización de las TIMs de arqueas. A) Histograma de estabilización de la MfeTIM por sales de fosfato. B) Recuperación de la actividad de TtxTIM al añadir la GLPDH de *M. thermotrophicus* (rombos negros) y la GLPDH de *Thermoproteus tenax* (asteriscos negros). Tomado y modificado de (Shramm, Kohlhoff & Hensel) & (Walden et. al. 2004).

Al contrario de las TermoTIMs, los parámetros catalíticos de las MesoTIMs están por debajo del límite de difusión. Aunque es un comportamiento atípico, puede ser consecuencia de un proceso biológico o por una falta de estabilizante biológico ya sea interacciones proteína-proteína, chaperonas o solutos (Muñiz, 2020).

Para las TermoTIM se analizó el desplegamiento por temperatura, por medio de Calorimetría diferencial de barrido (DSC). La T_m para MfoTIM-75 es de $\sim 91.7 \pm 0.8^\circ\text{C}$. Al variar la velocidad de calentamiento, la diferencias entre las T_m s obtenidas fue de 2.15°C . Por otra parte, la T_m para la MjaTIM-85 fue de $\sim 111.6 \pm 0.6^\circ\text{C}$, al variar la velocidad de barrido la T_m cambia en 1.7°C . En ambos casos, las endotermas se analizaron bajo un modelo irreversible de dos estados. La línea de la postransición es la Capacidad calorífica del estado desnaturalizado mientras que la de la pretransición es la capacidad calorífica del estado nativo (Muñiz 2020). Regularmente la línea de postransición es más alta porque se ha roto el núcleo hidrofóbico de la proteína durante la desnaturalización, la exposición de grupos no polares hace que el agua se estructure alrededor de dichos grupos, incrementando la C_p del estado desplegado respecto al estado nativo. En el caso de la MfoTIM-75 se observa este comportamiento, indicando que el núcleo hidrofóbico de la proteína se rompió y se encuentra completamente desnaturalizada. En el caso de la MjaTIM-85 se observa que la línea de la postransición está por debajo lo que sugiere que la proteína mantiene estructura residual, manteniendo grupos hidrofóbicos escondidos o bien en forma de agregados. Debido al ΔC_p negativo podemos asumir que se trata de agregación (Jelesarvo & Bosshard, 1999). Mientras que MfoTIM-75 se despliega por completo MjaTIM-85 mantiene rastros de aminoácidos hidrofóbicos ocultos en su estructura desplegada. También por las altas temperaturas podemos sugerir que algunos residuos como asparagina, glutamina

o cisteínas sufren modificaciones químicas irreversibles. Puede ser que por esto ninguna de las dos proteínas presente reversibilidad en su proceso de desplegamiento (Muñiz 2020).

Se ha estudiado la termoestabilidad de diferentes TIMs de Arqueas. La PfuTIM es muy termoestable tiene una vida media de inactivación de 280 minutos a 104°C, el gráfico de Arrhenius muestra una dependencia lineal con la temperatura, la vida media de inactivación para la MfeTIM es de 60 minutos a 88.5° (Shramm, Kohlhoff & Hensel). La Tm aparente de la TIM de *Thermoplasma acidophilum* se midió por medio de dicroísmo circular es de ~74.6°C (Park SH, et. al. 2015).

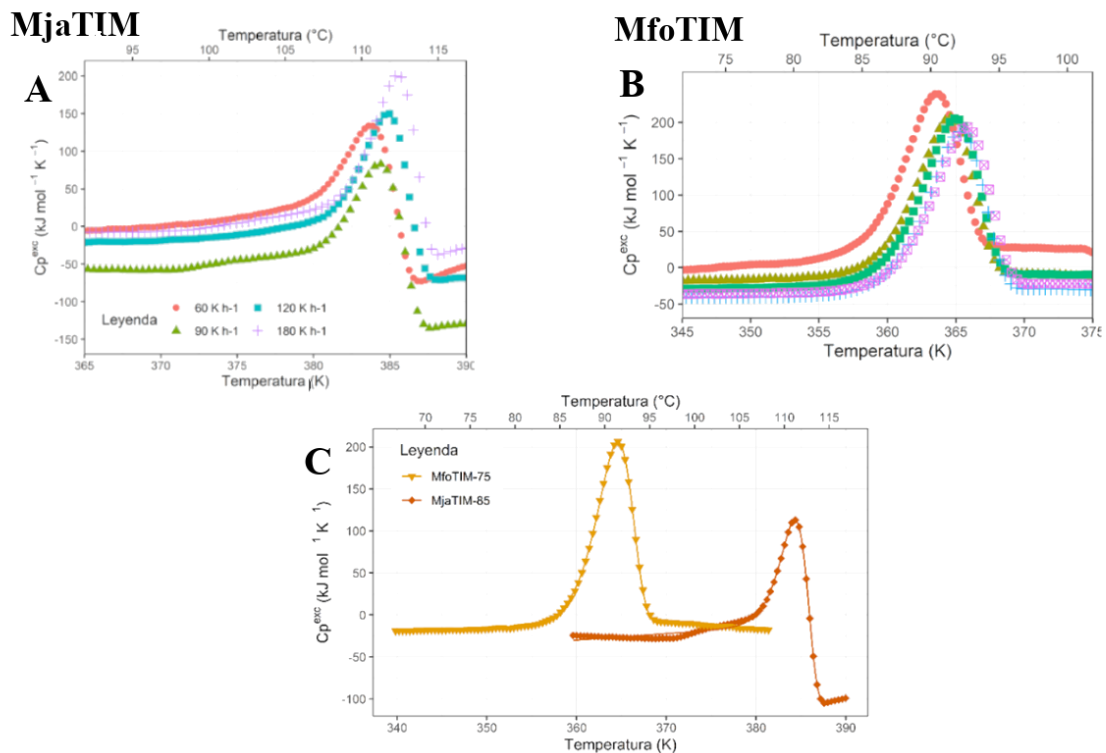


Figura 25. Calorimetría Diferencial de Barrido de las ThermoTIMs. A) MjaTIM. B) MfoTIM. C) Comparación de las isotermas entre MjaTIM (rombos rojos) MfoTIM (triángulos naranjas). Tomado y modificado de (Muñiz 2020)

El valor de la energía de activación es muy diferente entre las ThermoTIM (MjaTIM-85: 829.6 ± 131.7 kJ/mol; MfoTIM-75: 575.1 ± 5.3 kJ/mol), este valor nos indica la barrera energética que se tiene que pasar para llegar del estado nativo al desnaturalizado, cuya principal contribución está dada por la ruptura de los enlaces internos que se llenan de agua formando una red de enlaces insatisfechos. (Rodríguez-Larrea, Minning, Brochert, & Sanchez Ruiz, 2006); (Sanchez-Ruiz, 2010). La diferencia de casi 250 kJ/mol nos sugiere una modificación importante entre las redes de interacciones parcialmente rotas del estado de transición. Cabe destacar,

que ambas proteínas son tetraméricas, y la diferencia entre sus valores de T_m es de $\sim 20^\circ\text{C}$. Por lo que inferimos que la diferencia en termoestabilidad no se debe al estado oligomérico, sino a algún otro factor estructural (Muñiz 2020). Entre ambas proteínas existe un 80% de identidad por lo que es posible intentar rastrear los cambios en secuencia entre estas dos proteínas. Necesitamos información cristalográfica que nos ayude a determinar cuáles son las posibles diferencias para explicar el cambio en termoestabilidad y energía de activación (Muñiz 2020).

5.2 ANTECEDENTES EVOLUTIVOS

Se ha propuesto la oligomerización de las TIMs como una adaptación a la temperatura, a pesar de que algunos datos empíricos parecen contradecir la hipótesis, y aunque no exista evidencia experimental que lo respalde. Así que se toma a la selección natural como el agente causal más importante en la evolución de los seres vivos. La buena noticia es que podemos poner a prueba dicha hipótesis, si planteamos un experimento en donde una característica "X" es una adaptación de una variable ambiental "Z". Para fines de nuestro estudio; saber si la oligomerización es una adaptación de la temperatura ambiental. Podemos poner a prueba dicha hipótesis si comparamos el estado de oligomerización de organismos relacionados filogenéticamente que crezcan a diferentes temperaturas. Esta aproximación permite recopilar evidencia sobre cual podría ser el verdadero origen evolutivo de la oligomerización. A pesar de esto no podemos afirmar que la selección natural tenga una mayor importancia causal que otras fuerzas evolutivas. A esta perspectiva se le llama adaptacionismo metodológico que fue la base para realizar el trabajo de tesis de Ricardo Muñiz, quien propone un estudio comparativo filogenético entre diferentes TIMs de arqueas del grupo de los Methanococcales, con el fin de determinar si las características biofísicas de estas TIMs (estabilidad conformacional, termoestabilidad y estado oligomérico) se seleccionaron de acuerdo con la temperatura óptima de crecimiento del organismo (Muñiz, 2020).

Debido a que no se pudo concluir la caracterización de las MesoTIM por los problemas de purificación y la pandemia de 2020, este objetivo del trabajo de tesis de Muñiz 2020, quedó incompleto. Sin embargo, algunos de sus resultados permiten discutir sobre las ideas actuales sobre la evolución de las proteínas y como nos sirven para proponer futuras líneas de investigación. Para un estudio comparativo es importante determinar cuál fue la sucesión temporal con la que se adquirieron las características descritas experimentalmente durante la filogenia de los organismos estudiados. En la figura 27 se observa la relación propuesta por Muñiz en 2020 de las características biofísicas determinadas experimentalmente. En el caso de MfoTIM-75 y MjaTIM-85 tienen valores de T_m muy diferentes, aunque ambas son homotetrámeros así que la termoestabilidad de estas proteínas parece tener un origen estructural y no oligomérico. Obtener información cristalográfica de MfoTIM-75 será determinante para conocer con precisión cuáles factores se pueden asociar a este cambio.

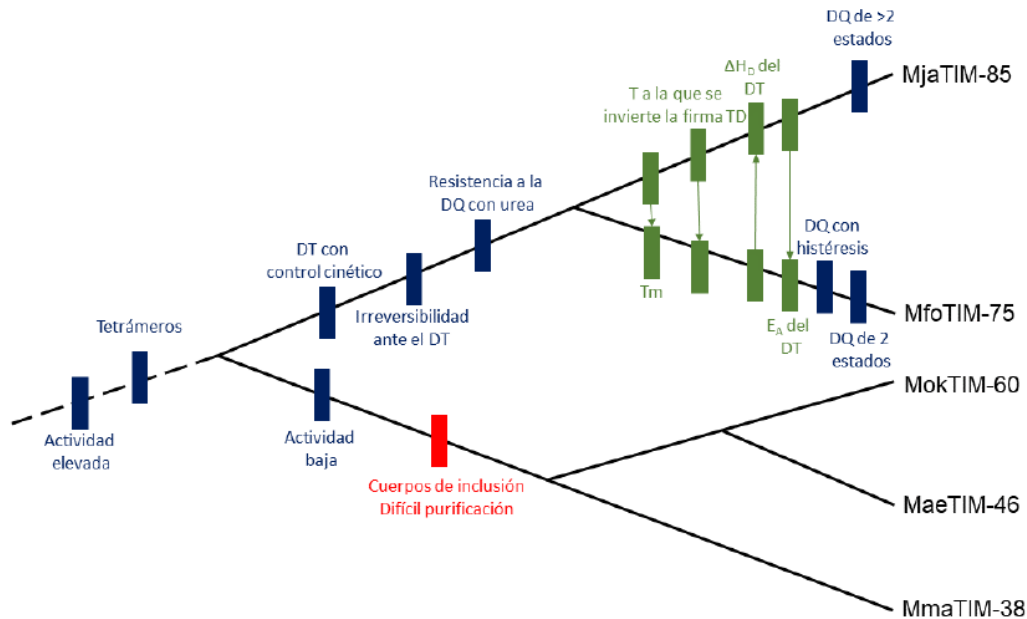


Figura 26. Características biofísicas de la TIM en relación con la filogenia de los Methanococcales. En azul se representan las variables categóricas, en verde las variables numéricas y en rojo el problema de la expresión heteróloga. *Methanocaldococcus jannaschii* (MjaTIM-85), *Methanoterris formicicus* (MfoTIM-75), *Mathanothermococcus okinawensis* (MokTIM-60), *Methanococcus aelicus* (MaeTIM-46) y *Methanococcus maripaludis* (MmaTIM-46). Tomado de (Muñiz, 2020)

Únicamente el caso de las TermoTIMs pareciera ser que si hay una relación entre la T_m y la temperatura óptima de crecimiento que podemos asociar con un proceso adaptativo. Porque, el intervalo de crecimiento de *Methanoterris formicicus* va de los (55°C a los 83°C) por lo que su proteína con una T_m de 91.7°C parece ser lo suficientemente elevada para mantenerse plegada a la temperatura que vive *M. formicicus* y *M. jannaschii*. Sin embargo, la transición de la desnaturalización comienza un poco antes de los 85°C por lo que su estabilidad cinética no podría ser suficiente para mantener la proteína funcionando por un tiempo fisiológicamente relevante ($t_{1/2}$ estimada de MfoTIM-75 a 85°C es de 24.5 minutos) en *M. jannaschii* que vive de 80°C a 85°C, en cambio la T_m de MjaTIM es de ~111.6°C y parece ser suficiente para el rango de temperatura al cual vive el organismo. Esta conclusión se ajusta adecuadamente para los datos obtenidos *in vitro*, pero hay que tener en cuenta que en condiciones celulares hay otros factores como las interacciones proteína-proteína o presencia de solutos en altas concentraciones que pueden incrementar la T_m (Muñiz, 2020) (Despa, Orgill & Lee 2005).

Aunque los parámetros enzimáticos de las MesoTIMs se obtuvieron de proteínas parcialmente puras, se puede argumentar que, de acuerdo con algunas hipótesis, el que una enzima en particular mantenga bajos niveles de actividad en condiciones

experimentales puede deberse a diferentes condiciones fisiológicas que no observamos, por ejemplo, puede existir presión selectiva que mantiene la actividad en cierto umbral donde no se afecta la adecuación de la especie. Se podría especular que el metabolismo de carbono de los methanococcales tuvo repercusiones diferentes en la evolución de las enzimas glucolíticas de estos organismos y que, por lo tanto, no sería del todo extraño encontrar que la actividad de la TIM sea distinta, al menos en este caso (Muñiz, 2020).

Otro antecedente evolutivo importante de este trabajo es un estudio evolutivo a nivel de genoma que se realizó en 2020 por el grupo de investigación de Armanet en donde utilizaron el linaje de Arqueas Methanococcales como modelo de estudio para determinar si la temperatura óptima de crecimiento (T_{opt}) afecta las variaciones de las frecuencias de aminoácidos en el proteoma o si está asociada con la transferencia horizontal de genes. Compararon 115 proteínas homólogas de 18 proteomas Methanococcales. El orden tiene cuatro géneros: Methanocaldococcus (termófilos e hipertermófilos), Methanoterris (termófilos e hipertermófilos), Methanothermococcus (termófilos) y Methanococcus (termófilos y mesófilos) (Lecocq, M. et. al. 2020).

También examinaron si la transferencia horizontal de genes había estado presente en la evolución del orden y si estaba asociado a la T_{opt} . Identificaron 2360 eventos de transferencia horizontal de genes que afectan a 1394 familias proteicas (65.9%). Pero no encontraron correlación de estos eventos con la T_{opt} , ya que la mayoría de los eventos ocurrieron en las ramas terminales del árbol y no en las internas como se esperaba por la naturaleza de estos eventos. Estos eventos son ampliamente extendidos entre los mesófilos y los termófilos del grupo, provenientes de donadores procariontes y en la mayoría de otras especies termófilas y mesófilas. El grupo tampoco encontró relación entre la T_{opt} y el contenido G+C (Lecocq, M. et. al. 2020).

Sus datos sugieren que la T_{opt} es el factor dominante en la composición de aminoácidos tanto en las proteínas de mantenimiento (house-keeping) como en las proteínas accesorias. Sin embargo, no hay correlación con la composición de aminoácidos y algún clado filogenético específico. Observaron una correlación positiva entre la arginina y la T_{opt} , y la serina muestra una correlación negativa con la T_{opt} . Además, estudiaron las substituciones de pares de aminoácidos en el genoma y su relación con la T_{opt} . Las substituciones correlacionadas con una disminución de la T_{opt} se muestran en azul en la (Figura 27.A) y las relacionadas positivamente con la T_{opt} , en flechas rojas (Figura 27.B) (Lecocq, M. et. al. 2020). Utilizando métodos de reconstrucción ancestral se predijo que el último ancestro en común del orden fue una arquea hipertermófila con una T_{opt} de 80°C (Figura 28) (Lecocq, M. et. al. 2020).

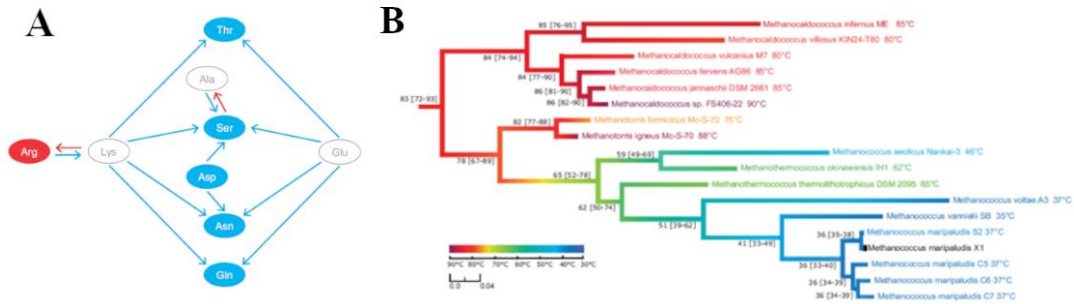


Figura 27. Relación de la Temperatura óptima de crecimiento y la filogenia de los Methanococcales. A) Representa las sustituciones correlacionadas positivamente con la T_{opt} en rojo y en azul las correlacionadas negativamente con la T_{opt} . B) Representación de la filogenia de los Methanococcales coloreada según la Temperatura óptima de crecimiento, en rojo los hipertermófilos, en verde los mesófilos y en azul los psicrófilos. Tomado y modificado de (Lecocq, M. et. al. 2020)

5.3 ANTECEDENTES ESTRUCTURALES

Hasta el momento solo se han reportado 5 cristales de TIMs de arqueas de 4 especies diferentes: La TIM de *Pyrococcus furiosus* PDB: 1HG3, *Thermoproteus tenax* PDB: 1W0M, *Methanocaldococcus jannaschii* PDB: 2H6R, *Thermoplasma acidophilum*, apo-enzima PDB: 5CSS, enzima con G3P PDB: 5CSR.

En 2001 el grupo de investigación de Garry Taylor resolvió la primera estructura cristalográfica de una TIM de arquea. La TIM de *Pyrococcus furiosus*, usando dispersión anómala múltiple, con una resolución de 2.7 Å. En la unidad asimétrica observaron 8 monómeros, los cuales formaban 2 tetrámeros, que conservan la formación clásica del dímero de las TIMs por medio de las asas 3, 1 y 4. En cambio, la interfaz tetramérica se forma por una gran red de interacciones hidrofóbicas en varias hélices (Walden et. al. 2001).

En comparación con otras TIMs bacterianas y eucariontes se encontró que el amino-terminal mira hacia dentro del barril y forma una interacción iónica con D139 que se encuentra en el asa después de la hélice 5, esta interacción fija el N-terminal al centro del barril. La hélice 4 es rica en alaninas (28.6%) con los residuos de alaninas orientadas hacia fuera del centro del monómero, participando en la interface del tetrámero, el otro lado de la hélice consiste en varios residuos cargados, que forman una red de pares iónicos (D105, E107, R111, R112, E114, E115, E107) que no están conservados en otras TIMs. La hélice 5 es mucho más pequeña, tiene una delección de ~20aa y en su mayoría está formado por la secuencia de residuos hidrofóbicos PAVSAAVA, esta área forma la interfaz tetramérica, mientras que las TIMs que no tetramerizan muestran en la hélice 5 bajas proporciones de alanina y aa no hidrofóbicos. También la hélice 6 es más pequeña y su C-terminal participa en la asociación del tetrámero, además, la mayoría de sus residuos expuestos están cargados (Walden et. al. 2001).

En la asociación dimérica de PfuTIM, no se aumentaron el número de pares iónicos, ni puentes de hidrógeno y el área de superficie es similar al de la interfaz dimérica de otras TIMs. El tetrámero es un dímero de dímeros, se establece por interacciones hidrofóbicas de las hélices 4 y 5. También existe un par iónico débil (5A) entre R100a y D105c. PfuTIM cuenta con una red de pares iónicos entre la K3 y D139, E6 y K178, E18 y K210, E147 y K159, E115 y R112 que no se observa en otras TIM (Walden et. al. 2001).

Como vimos anteriormente, una de las características estructurales propuestas para la termoestabilidad es el incremento de pares iónicos. Podemos sospechar que la presencia de los pares iónicos no conservados en PfuTIM ayuda a la termoestabilidad de la proteína. El grupo de Guptasarma demostró en 2009 que al mutar estos residuos por otros aminoácidos que no pudieran formar la red de interacciones iónicas, provocaba una disminución en la termoestabilidad. Dichas

mutaciones fueron: R111T, R112A, E114K y E115G que rompen la red de pares iónicos en el monómero de PfuTIM. Encontraron que la proteína con mutaciones no presenta cambios en la estructura secundaria, fluorescencia o volumen hidrodinámico en comparación con la silvestre. Sin embargo, las mutaciones vuelven a la proteína menos estable cinéticamente y termoestable (Chandrayan & Guptasarma 2009)

En 2015 el mismo grupo añadió esta red de interacciones, pero ahora a la TIM de *Methanococcoides burtonii* (MbuTIM), un organismo psicrófilo que no tiene analogía de esta red de pares iónicos en su estructura. Este organismo se ha adaptado a la vida en las frías aguas de los lagos de la Antártica, crece naturalmente a una temperatura de 1°C pero tiene un rango de crecimiento de los -2.5°C a los 28°C, con una temperatura óptima de 23°C. Así que se teoriza que la temperatura media de desnaturalización de sus proteínas se encuentre en un rango de 40°C 50°C. Dhaunta y colaboradores introdujeron los pares iónicos de PfuTIM a MbuTIM con el fin de observar si esta red aumenta la termoestabilidad. Las mutaciones introducidas fueron: T107R, A108R, K110E y G111E. El primer resultado sorprendente fue que MbuTIM silvestre mostró una curva de desplegamiento con una T_m aparente de ~73°C con una tasa de calentamiento de 1°C/min. Al aumentar esta tasa a 3°C/min la T_m aumenta a ~76°C. Estos valores altos de T_m indican que MbuTIM es termoestable a pesar de que proviene de un organismo psicrófilo. La curva de desplegamiento de la mutante MbuTIM muestra que a la tasa de 1°C/min la T_m es de ~78°C y a la tasa de 3°C/min es de ~80°C. Así que podemos concluir que la introducción de los pares iónicos aumenta la T_m en 6°C. Además, la velocidad de desnaturalización de la mutante MbuTIM es más lenta que la de MbuTIM en todas las temperaturas que examinaron. Por último, en presencia de sal (NaCl 2M) la T_m para MbuTIM y MutMbuTIM, de 5°C y 12°C, respectivamente. Indicando el cambio de las interacciones electrostáticas, entre la proteína y el solvente ayudan a la termoestabilidad (Dhaunta, Guptasarma et. al. 2015).

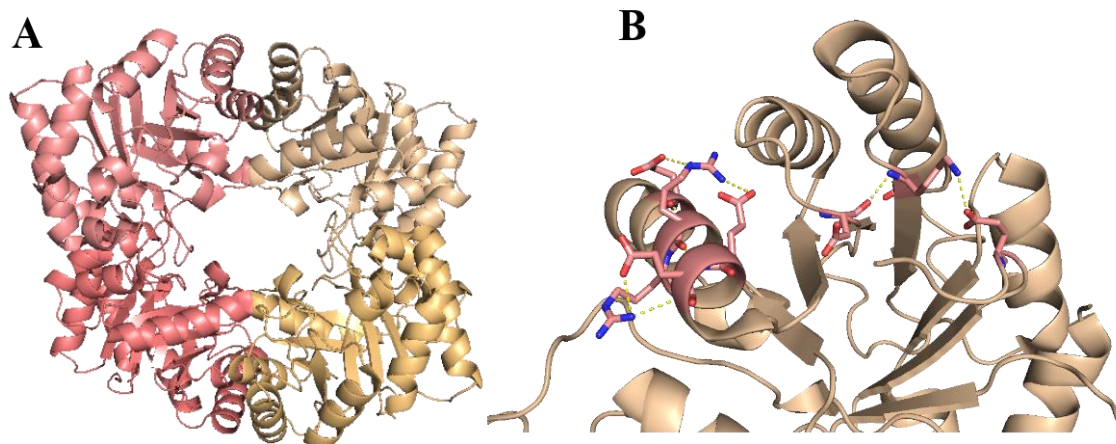


Figura 28. Cristal de la TIM de Pyrococcus furiosus. A) Se muestra la conformación de tetrámero del cristal como dímero de dímeros. B) Se muestran las redes de pares iónicos no conservados, marcados con líneas punteadas amarillas, del lado derecho se muestra el anclaje del N-t al centro del barril, y del lado izquierdo se ilustra la red de pares iónicos en el C-t de la hélice 4.

La segunda estructura resuelta para una Tim de arquea fue la TIM de *Thermoproteux tenax* en 2004. *Thermoproteux tenax* es un organismo hipertermófilo con una temperatura óptima de crecimiento de 86°C. Dentro de la unidad asimétrica hay 8 monómeros, dispuestos en 2 homotetrámeros que comprenden dímeros clásicos de la TIM. Cada subunidad de TxTIM tiene el mismo plegamiento que otras TIM y el cristal mostró una conformación cerrada en el asa catalítica. Tiene el 52% de identidad con PfuTIM y el rmsd es de 0.97 Å. El N-terminal no está fijo en el centro del barril por un par iónico como en PfuTIM, pero también apunta al centro del barril en TxTIM. El C-terminal de la hélice 4 está más cerca del centro del barril. La hélice 4 tampoco tiene la red de pares iónicos encontrados en PfuTIM (Walden et. al. 2004). El tetrámero se estabiliza principalmente por interacciones hidrofóbicas entre los residuos de las hélices 4,5 y 6 y las asas entre ellas. TxTIM tiene una red de pares iónicos entre R154, E183 y D187, & (E144, K156) que sirven como anclaje de la hélice 6 y la hélice 7. Otra característica interesante es que tiene más área expuesta cargada (Walden et. al. 2004).

La tercera estructura que se resolvió fue la de *Methanocaldococcus jannaschii* en 2007 a una resolución de 2.3°Å por medio de remplazo molecular con la TIM de *Pyrococcus furiosus*. No hay densidad electrónica en la zona de unión a ligando a pesar de que el cristal se obtuvo en presencia de un exceso de ácido 3-fosfoglicérico (Gayathri et. Al. 2007). Al igual que los cristales de PfuTIM y TtxTIM, en la unidad asimétrica se observan 8 monómeros con plegamiento de barril TIM, dispuestos en dos tetrámeros, formados por dímeros de dímeros con una interfaz dimerica muy similar a otras TIMs. Tiene 66% de identidad con PfuTIM y un rmsd de

0.75°Å. Sus aminoácidos catalíticos son K8, H90, y E138. MjaTIM tiene 2 puentes de sal, uno entre R46 y D44 (conservado en PfuTIM) y otro entre D80 y R17. La mayoría de los puentes de hidrógeno de la interfaz dimérica están conservados en PfuTIM y en MjaTIM. En la interfaz tetramérica los residuos involucrados son Met95, Leu96 y Leu97, así como E101, N120, N122 y T123 que forman una red de puentes de hidrógeno. La interfaz se forma mediante las interacciones hidrofóbicas entre los residuos de la hélice 5, A126, A129 y L130. En las TIMs diméricas la zona que correspondería a la interface tetramérica tiene muchos residuos con carga negativa, indicando que sería imposible la interacción tetramérica por fuerzas repulsivas. En cambio, la hélice 5 es rica en alaninas (A126, A128, A129) las cuales están conservadas en todas las TIMs tetraméricas. La N89 conservada participa en la interface tetramérica por medio de un puente de hidrógeno con la cadena principal de Met95 (Gayathri et. Al. 2007).

Los últimos dos cristales que se han resuelto más recientemente son la apoenzima y la enzima con glicerol-3-fosfato de la TIM de *Thermoplasma acidophilum* (TacTIM), este organismo es una arquea termófila, con temperatura óptima de crecimiento de 55°C a 60°C y crece a de pH0.5 hasta pH2; usa la vía Entner Doudoroff (ED) y la vía (EMP) para la degradación de glucosa. Ambos cristales tienen 4 monómeros en la unidad asimétrica incluyendo dos dímeros con plegamiento de barril TIM y el cristal con sustrato muestra que el glicero-3-fosfato está bien posicionado en el sitio catalítico. TacTIM Tiene un ~30% de identidad con las TIMs tetraméricas. Por medio de ultra centrifugación analítica y cromatografía de exclusión analítica se corroboró que TacTIM existe como dímero en solución. Sus aminoácidos catalíticos son K9, H89 y E137 (Park SH et. al. 2015).

Como mencionamos anteriormente, es necesario contar con datos cristalográficos que nos permitan conocer las diferencias estructurales entre las TIMs de arqueas. Específicamente, para dilucidar las determinantes estructurales responsables de la diferencia entre la Temperatura media de desnaturalización entre la MjaTIM y MfoTIM que es de ~20°C, mi compañero Sergio Romero cristalizó, difractó y resolvió la estructura de la TIM de *Methanothermus formicicus* a una resolución de 2.25 Å; con un R de 0.25 y un Rfree de 0.295 y sin "outliers" en el mapa de Ramachandran. El cristal tiene cuatro monómeros en la unidad asimétrica que forman un tetrámero, un dímero de dímeros que conserva tanto la forma clásica de asociación del dímero encontrado en todas las TIMs; como el modo de asociación tetramérica anteriormente observada con MjaTIM, PfuTIM y TtxTIM.

PDB	Organismo	Resolución (Å)	R work	R free	MTZ	Substrato	Referencia
2H6R	<i>Methanocaldococcus jannaschii</i>	2.3 Å	0.213	0.278	Si	No	Gayathri et. al., 2007
	<i>Methanothermus formicicus</i>	2.25 Å	0.244	0.284	Si	No	
1HG3	<i>Pyrococcus furiosus</i>	2.7 Å	0.254	0.286	No	No	Walden et. al., 2001
1W0M	<i>Thermoproteus tenax</i>	2.5 Å	0.197	0.226	No	No	Walden et. al., 2004
5CSS	<i>Thermoplasma acidophilum</i>	2.17 Å	0.183	0.225	Si	Glicerol-3-fosfato	Park et. al., 2016
5CSR	<i>Thermoplasma acidophilum</i>	1.94 Å	0.215	0.180	Si	No	Park et. al., 2016

Tabla 4. Cristales de ArqTIMs publicados hasta la fecha: mayo 2022.

HIPÓTESIS

La oligomerización no es una adaptación a la temperatura ambiental. La termoestabilidad debe relacionarse con las características estructurales de las ArqTIMs, mientras que la oligomerización debe estar relacionada con la trayectoria evolutiva del dominio Arquea.

OBJETIVOS

- Caracterizar las diferencias estructurales entre las ArqTIMs tetraméricas y el único cristal ArqTIM que es dímero en solución.
- Encontrar las diferencias estructurales responsables de la diferencia de termoestabilidad entre MfoTIM y MjaTIM.
- Encontrar los residuos co-evolutivos responsables de la oligomerización.

MÉTODOS

SERVIDOR PISA (Protein Interfaces Surfaces and Assemblies)

Las células vivas están repletas de biomoléculas, que interactúan constantemente (Dey S & Levy E, 2021). Fisiológicamente una proporción importante de las proteínas (30%-50%) adopta estructura cuaternaria en forma de homoligómeros (Dey S & Levy E, 2018). La funcionalidad de la mayoría de las proteínas depende completamente de su ensamblaje molecular, por ejemplo: los canales iónicos, polimerasas, microtúbulos, nucleosomas y muchas otras (Krissinel E & Henrick K 2007). La estructura cuaternaria no siempre está disponible en las estructuras cristalográficas porque aún no hay un método claro para distinguir los contactos biológicos de los cristalográficos (Dey S & Levy E, 2021).

La herramienta PISA (*Protein Interfaces Surfaces and Assemblies*) <https://www.ebi.ac.uk/pdbe/pisa/> del ePDB representa la primera aproximación sistemática para la identificación automática de las probables estructuras cuaternarias. Las predicciones de PISA están basadas en modelos fisicoquímicos de interacciones macromoleculares y parámetros termodinámicos. Krissinel y Henrick basaron su aproximación asumiendo que los ensamblajes macromoleculares mantienen sus estructuras en un estado cristalino. Así que la simetría del cristal compromete las interfaces que cumplan con los siguientes dos criterios:

- Si la interfaz es mantenida por otras interfaces del mismo tipo
- Una interfaz no puede ser mantenida si los resultados del ensamblaje equivalen a unidades monoméricas en orientaciones paralelas.

Para entender la condición 2 debemos tomar en cuenta que todas las unidades monoméricas paralelas se obtienen por translaciones repetidas de la subunidad original. Esto significa que ninguna de las dos unidades paralelas están conectadas por la misma configuración, lo que rompería con la primera condición. Así que se encuentran todos los grupos de posibles complejos en los cristales, cuyas interfaces satisfacen estas dos condiciones (Krissinel E & Henrick K 2007). PISA está diseñado para analizar contactos entre moléculas dentro de un cristal, una limitante importante es que no puede predecir la estructura cuaternaria de un modelo por homología que carece de datos cristalográficos dentro del archivo PDB. Para que PISA funcione son necesarios los siguientes comentarios, además de las coordenadas atómicas (PISA web).

```

HEADER      HYDROLASE (SERINE PROTEINASE)                22-JAN-85   5CHA
CRYST1      49.290   67.480   65.940   90.00 102.02  90.00 P 21
ORIGX1      1.000000   0.000000   .213065   0.000000
ORIGX2      0.000000   1.000000   0.000000   0.000000
ORIGX3      0.000000   0.000000   1.023106   0.000000
SCALE1      .020288   0.000000   .004323   0.000000
SCALE2      0.000000   .014819   0.000000   0.000000
SCALE3      0.000000   0.000000   .015516   0.000000
MTRIX1      1   .913870  -.006906  .405950  -9.94000  1
MTRIX2      1  -.006985  -.999918  -.012656  40.60000  1
MTRIX3      1   .406002   .010876  -.913807  47.60000  1
ATOM .....
```

Figura 29. Comentarios cristalográficos necesarios en el archivo PDB para los cálculos del servidor PISA. Tomado de la página web de PISA

El puntaje de energía que utiliza PISA modela la entropía de la proteína para poder determinar la calidad de su asociación (Krissinel E & Henrick K 2007). La disociación del complejo implica un cambio en la energía libre descrita como:

$$\Delta G_{\text{diss}}^0 = -\Delta G_{\text{int}} - T\Delta S \quad (1)$$

ΔG_{diss}^0 depende totalmente de las subunidades y lo llamaremos patrón de disociación, este término se puede expresar en términos de una constante de disociación. Supongamos que tenemos una proteína A, que forma interacciones con si misma: A₂, A₃, A₄, A₅, A_i

$$\Delta G_{\text{diss}}^0 = -RT \log K_d = -RT \log \frac{\prod_{i=1}^n [A_i^0]}{[A^0]} \quad (2)$$

Donde [A°] y [A°_i] son las concentraciones al equilibrio de las subunidades del complejo, podemos decir que el complejo es estable si $\Delta G_{\text{diss}} > 0$ o si $K_d < 1$.

Muchos factores contribuyen a la unión de una macromolécula, el segundo término de la ecuación (1). ΔG_{int} se origina del cambio en la energía de solvatación ΔG_{solv} , de las interacciones entre las subunidades, ΔG_{cont} que comprende puentes de hidrógeno, puentes salinos y enlaces disulfuro de la interfaz y de las interacciones electrostáticas entre las subunidades ΔG_{es} , que también toma en cuenta las moléculas de agua que estén a una distancia menor a 7Å.

$$\begin{aligned}
\Delta G_{\text{int}}(A) = & \Delta G_{\text{solv}}(A) - \sum_{i=1}^n \Delta G_{\text{solv}}(A_i) \\
& + \sum_{j>i} \Delta G_{\text{cont}}(A_i, A_j) \\
& + \sum_{j>i} \Delta G_{\text{es}}(A_i, A_j)
\end{aligned}
\tag{3}$$

La entropía de la subunidad A se puede representar como la siguiente suma:

$$\begin{aligned}
S(A) = & S_{\text{trans}}(A) + S_{\text{rot}}(A) + S_{\text{vib}}(A) \\
& + S_{\text{surf}}(A)
\end{aligned}
\tag{4}$$

Donde $S_{\text{trans}}(A)$ y $S_{\text{rot}}(A)$ representan los términos de energía traslacional y rotacional respectivamente, $S_{\text{vib}}(A)$ representa la entropía de los modos de vibración internos y $S_{\text{surf}}(A)$ cuenta la entropía de los átomos de la superficie con grados de libertad fraccionados. El cambio de entropía generado por la formación del complejo es el mayor factor que determina el tamaño del complejo y su simetría. Se observó, que esta entropía también dirige los complejos macromoleculares a estados simétricos (Krissinel E & Henrick K 2007).

Gracias al cálculo del ΔG podemos estimar otro parámetro que nos ofrece PISA: el CSS (Complexation Significance Score, en inglés) que indica la importancia de la interfaz para la formación del complejo. La puntuación se define como la razón entre la energía libre total de unión de la interfaz, entre la suma de la energía libre total de unión del complejo. Así el valor máximo de CSS es 1 y el mínimo es 0. Por ejemplo, si PISA nos muestra un tetrámero estable, con 2 interfaces, la primera con un $\Delta G = -10$ kcal/mol y la segunda interfaz $\Delta G = -5$ kcal/mol, entonces el CSS de la primera interfaz se calcularía como $2 \times 10 / (2 \times 10 + 2 \times 5) = 2/3$.

El servidor web de PISA también calcula las áreas de superficie al rodar una molécula de agua de 1.4Å de diámetro sobre la superficie y sumar todos los puntos muestreados en contacto con esta molécula de agua. Primero se calcula el área de cada monómero independientemente y después el complejo, el área de la interfaz se calcula mediante la suma del área de los dos monómeros menos el área del complejo (PISA web).

Para determinar los enlaces de hidrógeno, PISA considera una distancia entre los átomos donante y receptor menor o igual a 4Å y el ángulo debe ser entre 90° a 270°. Al igual para considerar presente un puente salino se debe tener una distancia máxima de 4Å entre los átomos aceptores y donadores (PISA web).

Nosotros utilizamos el servidor de PISA para calcular los aminoácidos, puentes de hidrógeno y puentes salinos involucrados en las interfaces. Además de calcular el área y los valores de CSS de las interfaces para los 5 cristales de las TIMs de Arqueas.

- Mja TIM (2H6R)
- Mfo TIM
- Ttx TIM (1W0M)
- Pfu TIM (1HG3)
- Tac TIM (5CSR)

También se calcularon las áreas, los aminoácidos, puentes de hidrógeno y puentes salinos involucrados en las interfaces de los modelos estructurales obtenidos por Alphafold 2 para las siguientes TIMs de arqueas:

- Mma TIM
- Mae TIM
- Mfe TIM
- Ton TIM
- Mok TIM
- Mbr TIM
- Mbu TIM

RING

Las redes de interacción para los cristales MjaTIM y MfoTIM se realizaron con el generador de redes de interacción de residuos RING 2.0 (Residue Interaction Network Generator en inglés) en donde se consideran a los aminoácidos como nodos y las interacciones fisicoquímicas entre ellos como las conexiones de la red. Gracias a esta representación se puede simplificar mucho el análisis estructural (Piovesan D, Minervini G & Tosatto S, 2016).

RING calcula puentes de hidrógeno, puentes salinos, interacciones de Van der Waals e interacciones π - π o catión- π . Genera la red en dos pasos, primero identifica los pares residuo-residuo que probablemente estén interactuando, basado en información cristalográfica de distancia entre átomos. A veces ocurren múltiples interacciones entre el mismo par de residuos y esas interacciones igual se organizan por energía y distancia, así el usuario puede elegir observar todas las interacciones o solo visualizar una interacción específica (Piovesan D, Minervini G & Tosatto S, 2016).

Posteriormente se calculan las energías asociadas a las interacciones entre aminoácidos, en base a criterios geométricos como distancia, ángulo y tipo de átomo, por ejemplo; los puentes de hidrógeno son calculados si los átomos donador-aceptor están a un ángulo menor a 63° . Para las interacciones de Van der Waals solamente se consideran los pares C-C y C-S con un radio atómico de 1.89Å y 1.77Å , respectivamente (Piovesan D, Minervini G & Tosatto S, 2016). En la (Tabla 5) se observan todos los criterios para asignar las diferentes interacciones entre los átomos.

Tipo de Interacción	Residuo	Átomo		Distancia Mínima (Å)	Distancia Máxima (Å)	Energía (kJ/mol)
Puente de hidrógeno	Todos	Aceptor C (cadena principal)	Donador NH (cadena principal)			
	R	OD1	NE, NH1, NH2			
	N	OD1, OD2	ND2			
	D		SG			115.0
	C	OE1	NE2			
	Q	OE1, OE2				
	E	ND1	NE2, ND1			40.0
	H		NZ			
	K	SD		3.5	5.5	
	M	OG	OG			
	S	OG1	OG1			17
	T		NE1			

	W Y	OH	OH			
Puentes salinos	D E R L H	Positivos CZ NZ <small>Masa del centro</small>	Negativos <small>Masa del centro</small> <small>Masa del centro</small>	4.0	5.0	20.0
Puentes disulfuro	C	SG		2.5	3.0	167.0
Van der Waals	Todos Q N	Cualquier C o S NE1, OE1 ND2, OD2		0.5	0.8	9.6
Catión- π	R K Y W F	Catión CZ NZ	π Átomo del anillo más cercano	5.0	7.0	9.6
π - π	H Y W F	Centro del anillo		6.5	7.0	7.0

Tabla 5. Criterios geométricos para la asignación de interacciones en RING 2.0.

Nosotros calculamos las redes de interacción para los cristales MjaTIM y MfoTIM para observar cambios en la red de interacción entre los residuos que no tienen la misma identidad en MjaTIM y MfoTIM (~20%). Dichas redes de interacción se visualizaron y procesaron en Cytoscape, que es un software de código abierto que permite la visualización y manejo de redes complejas como redes de interacción molecular y vías biológicas (<https://cytoscape.org/index.html>). Para complementar la visualización tridimensional de la red también utilizamos LIGPLOT + (Laskowski R & Swindells M, 2011) un sistema que genera automáticamente diagramas en dos dimensiones de interacciones proteína-ligando de coordenadas tridimensionales, también lo utilizamos para graficar las áreas de oligomerización o interfaces (Laskowski R & Swindells M, 2011).

ALPHAFOLD2

La red AlphaFold predice directamente las coordenadas tridimensionales de todos los átomos de una secuencia de aminoácidos. Es un método basado en redes neuronales que utiliza limitaciones evolutivas, físicas y geométricas observadas en las estructuras proteicas para mejorar la precisión de la predicción de estructuras tridimensionales (Jumper J, et. al. 2021).

La red neuronal está dividida en dos etapas principales, la primera es el bloque de red neuronal llamado Evoformer y la segunda es el bloque de estructura. Evoformer utiliza dos datos de entrada: un alineamiento múltiple de secuencias homólogas con la secuencia que proporciona el usuario; el segundo dato de entrada es una matriz de contactos entre pares de aminoácidos específica para la topología de la secuencia dada por el usuario, la matriz contempla distancias entre carbonos alfa, ángulos de torsión, etc., se obtiene de la base de datos entrenada de AlphaFold (Figura 30) (Jumper J, et. al. 2021). Básicamente lo que hace Evoformer es aprender de las limitantes evolutivas y espaciales, lo que delimita el espacio tridimensional donde puede existir la proteína.

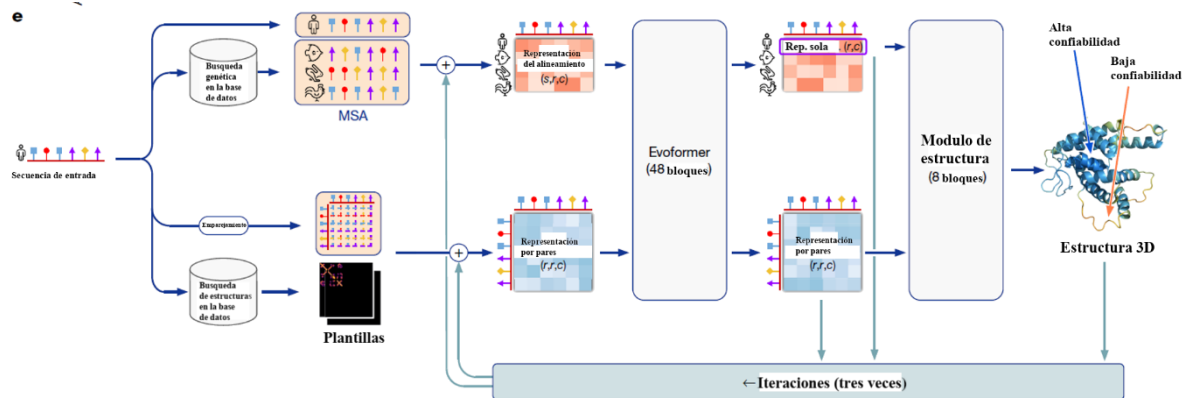


Figura 30. Flujo de trabajo de AlphaFold2. Se muestran los datos de entrada, los bloques de redes neuronales y el flujo del proceso de los datos. Tomado y modificado de (Jumper J, et. al. 2021)

El módulo siguiente es el bloque de estructura que utiliza como datos de entrada la secuencia proporcionada por el usuario y la matriz de contactos procesada por Evoformer. En este bloque se representa una estructura tridimensional explícita donde todos los residuos de la secuencia están en las posiciones inferidas por la matriz de contactos en forma de rotación y translación. Se refinan rápidamente a una estructura altamente precisa, gracias a que los enlaces peptídicos no tienen restricciones, así que se tolera romper la estructura de la cadena principal para permitir un refinamiento local y simultaneo de todas las partes de la estructura. Las predicciones de los ángulos χ de la cadena lateral, así como la precisión final

por residuo de la estructura se calculan con pequeñas redes por residuo en las activaciones finales de la red (Figura 30) (Jumper J, et. al. 2021).

Después se realiza un refinamiento iterativo al retroalimentar repetidas veces a la red con las salidas de los datos anteriores en los mismos módulos. La aplicación exacta de la geometría del enlace peptídico solo se logra en la relajación posterior a la predicción de la estructura mediante el descenso del gradiente en el campo de fuerza Amber32. Esta relajación no cambia la precisión del modelo, pero si remueve las violaciones estereoquímicas (Jumper J, et. al. 2021).

Nosotros utilizamos este método para realizar el modelo estructural de varias TIMs de arqueas porque se ha visto que Alphafold2 puede dar predicciones de alta precisión para los homoligómeros, además de su alta precisión en la predicción de toda la estructura. Utilizamos la versión avanzada de Alphafold2 (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/Alphafold2_advanced.ipynb#scrollTo=bQe3KeyTcv0n) no utilizamos plantillas con el fin de no sesgar los resultados a un determinado estado oligomérico. Utilizamos todos los parámetros por default en el cuadernillo de Alphafold2. Las secuencias utilizadas se obtuvieron de la base de datos del NCBI y se muestran a continuación:

Methanococcus maripaludis TIM

Dímero y Tetrámero

MKPNVIINYKTYTESVGKSGLAIAKAAEKVSEESGVTIGVAPQFLDLRMILDEVNIPVYAQH
MDAVTPGSSTGHILPDALKDAGVTGTLNHSERRMILADLEKCIENKLGKTVVCTNNIS
VSKAVSALNPTSIAVEPPELIGSGIPVSKANPEVVSQTVREVREINKNVEILCGAGISKGED
VKSALELGTNGVLLASGVVKAKNVEDAIRELIAEI

Methanococcus aeolicus TIM

Dímero y Tetrámero

MVIIIINSPIIIINYKTYSESVGKKGLNIAKACEKISEESGISIGVAPQYLDLKTVDNVNIPVYAQ
HFDAVGAGSNTGYVFAETIADCGLNGSLLNHSEKRMVLADLEEAINLANKYNFESVCTN
NIVVSKAVAVMNPTMLAIEPPELIGSGIPVSKANPEIVENTVNEVRKLNKTVTILCGAGISKG
EDVSAAMELGADGVLLASGVVKAKNVEESIRELLRNI

Methanothermococcus okinawensis TIM

Dímero y Tetrámero

MKPIIIINYKYAESVGEKGLKIAKAAEKVSEESGISIGVCPQFLDLRMISENVNIPVYAQHFD
AVSPGSHTGHVLAETLKDCGLNGSLLNHSEKRMILADLEKSIEIAKNYNLETIVCTNNIGVS
KAVAALNPNMIAIEPPELIGTGIPVSKANPEVVEGTVKEVRGINKDVKILCGAGISKGEDVS
SALELGACGVLLASGVVKSVDVEGSIQELIKHI

Methanobacterium bryantii TIM

Dímero y Tetrámero

MNEIKETPIVILNFKTYLESTGENALKLAKSSEMVAEETGVNIMVAPQYTDIYRIAREVDIPVL
AQHIDTIDAGGHTGSILPECVKEAGAVGTLNHSERRVELFEIDAAIKKADSLGLSTVVCTNN

IETSSAAATLNPDFVAIEPPELIGSGIPVSKAEPEIVEKTVESIHNINPKVRVLCGAGISTGDD
LKAAIDLGSEGVLLASGIILADDPKKALLDLVSKI

Methanococoides burtonii TIM

Dímero y Tetrámero

MKPLIVLNLKTYLEGTGEGAVRIARACKEVGEASGIEIAIAPQFCDIYRVASQVDVPVYSQHL
DVGAGSFTGHAFKCIKDAGAVGTLINHSECRLKLADIEASVTAAKGEGLRTIICTNNIATT
AAAAALGPDYVAVEPPELIGSGIPVSKADPEVVTGSVAAPERIDPAVKVLCGAGISKGEDL
KAAIELGSGVLLASGIVKAKDPKAALEDLVSLI

Methanothermus fervidus TIM

Tetrámero

MDRPIIVLNFKTYKESTGENALKLAKKCEQVSEEYGVKIIIVAPQHMDLRYVSENVNIPVIAQ
HIDPIDAGGHTGSVLLLECAKEAGAKGSLVNHSEKRMKLADISKVVKKLSENDMISIVCTNN
VETSAAAAALSPDFVAVEPPELIGSGIPVSKAKPEVVENTVEAVKTVNPDVRVLCGAGISS
GEDVKKAVELGTEGVLLASGVILAKDQKKALEELITEM

Thermococcus onnurineus TIM

Tetrámero

MTKLKEPIIAINFKTYIEATGERALEIAKAAEKVWKETGITIVVAPQLADLYRIAQVEVEIPVFAQ
HIDPIKPGSHTGHVLPEAVKEAGAVGTLLNHSENRMILADLEAAIRRAEEVGLMTMVCSN
NPAVSAAVAALGPDYVAVEPPELIGTGIPVSKAKPEVITDTVELVKRVNPEVKVLTGAGIST
GEDVKKALELGSGVLLASGVTKAKDPEKAIRDLVSLIV

HSYMDOCK

La mayor subclase de interacciones proteína-proteína está formada por los homoligómeros simétricos. Así que su determinación es esencial para entender funciones biológicas. En el PDB hay dos grandes grupos de simetría: la simetría cíclica (C_n) en el que la estructura del oligómero puede ser descrita como una rotación sobre un solo eje de cada unidad, la segunda es la simetría diédrica (D_n) que combina un eje rotacional con un eje perpendicular. Usualmente los algoritmos no son tan eficientes para predecir la estructura de homoligómeros simétricos. Sin limitaciones simétricas los algoritmos de *docking* generales tienen que buscar en un espacio mucho mayor, así que para mejorar la predicción son necesarios algoritmos especializados (Yumeng Y, Tao H & Huang S, 2018).

El algoritmo de HSYMDOCK se sostiene en un acoplamiento jerárquico limitado por simetría circular, desarrollado anteriormente por los mismos investigadores. El programa requiere que el usuario provea la estructura de la subunidad monomérica. Se usa el algoritmo para muestrear globalmente los modos de enlace probables con simetría cíclica (C_n) (Figura 31). Si se predice la estructura de un homo-oligómero simétrico C_n , solo se realizará una ronda del paso de acoplamiento simétrico C_n . Si no es así, el programa realiza un segundo acoplamiento simétrico C_2 con los 100 principales complejos C_n predichos en el primer paso, pero los busca en un eje perpendicular adicional, con el fin de buscar homo-oligómeros con simetría diédrica (D_n) (Yumeng Y, Tao H & Huang S, 2018).

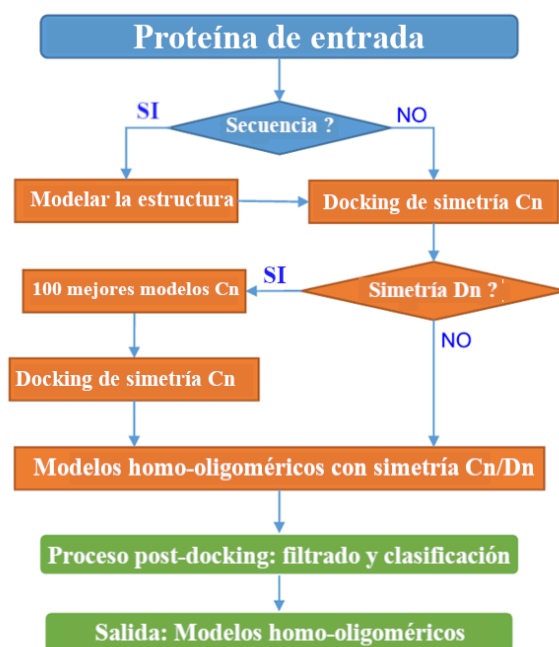


Figura 31. Flujo de trabajo de HSYMDOCK. Tomado y modificado de (Yumeng Y, Tao H & Huang S, 2018)

HSYMDOCK es el único programa de *docking* que toma en cuenta la simetría diédrica. Por su aproximación está considerado como el mejor predictor de homooligómeros simétricos por CAPRI (Critical Assessment of Predicted Interactions en inglés) (Figura 32) la comunidad especializada en evaluar los métodos de *docking* que predicen interacciones proteína-proteína (Tramontano A, 2017).

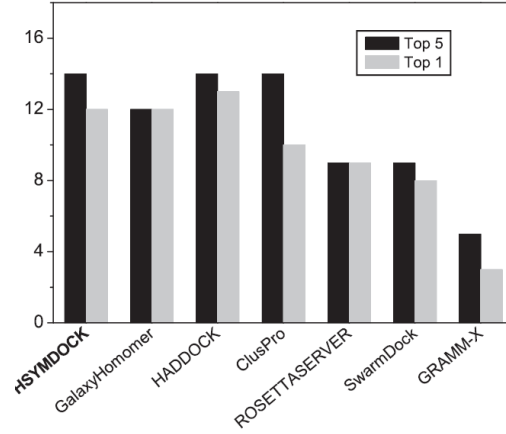


Figura 32. Mediciones CAPRI para evaluar el docking de homooligómeros simétricos que posicionan a HSYMDOCK como el mejor. Tomado de (Yumeng Y, Tao H & Huang S, 2018)

HSYMDOCK utiliza un intervalo de ángulo de 10° para el muestreo rotacional con un espaciado de 1.2 Å. El puntaje que utiliza está basado en los potenciales basados en conocimiento, utilizando parámetros conocidos de estructuras cristalográficas del PDB. Que se describen como una relación inversa de Boltzmann:

$$u_{ij}(r) = -k_B T \ln[g_{ij}(r)], \quad g_{ij}(r) = \rho_{ij}(r)/\rho_{ij}^*(r) \quad (5)$$

Donde k_b es la constante de Boltzmann, T es la temperatura del sistema, i & j son los pares de átomos en contacto en la formación del complejo (Huang S & Zou X, 2008). La función de distribución de los pares de átomos $g_{ij}(r)$ se define por la densidad de los átomos i & j a una distancia r , en las estructuras experimentales $\rho_{ij}(r)$ y en las estructuras predichas $\rho_{ij}^*(r)$. La suma de todas estas energías de interacción atómica nos da como resultado el puntaje de energía (Huang S & Zou X, 2008).

$$\text{energy score} = \sum_{\text{R-L atom pair}} u_{ij}(r) \quad (6)$$

El programa utiliza estos cálculos de forma iterativa, es decir, la ecuación (5) se alimenta conforme avanza el docking. Aquí k representa el paso iterativo, mientras que i & j representan los pares de átomos interactuando. En la Figura 33 se muestra gráficamente el proceso. Al inicio de la iteración $k=0$ se calcula con g_{ij}^{obs} , donde ρ_{ij}^{obs} son las densidades observadas experimentalmente. El método iterativo funciona muy bien con usualmente 60 pasos o iteraciones (Huang S & Zou X, 2008).

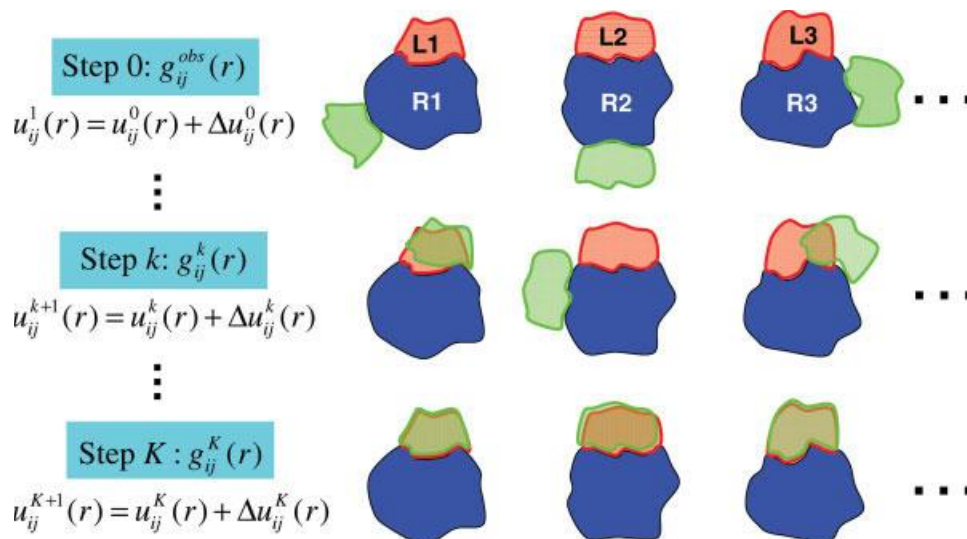


Figura 33. Proceso iterativo del cálculo del puntaje de HSYMDOCK. Tomado de (Huang S & Zou X, 2008).

Este puntaje también se verificó en términos de la predicción de afinidad, se probó con 74 complejos proteína-proteína. En la Figura 34 se muestra como el puntaje de unión de energía predicho en función de los datos experimentales de afinidad de unión. El coeficiente de correlación fue de 0.71 sugiriendo que el puntaje predice acertadamente la afinidad de unión (Huang S & Zou X, 2008).

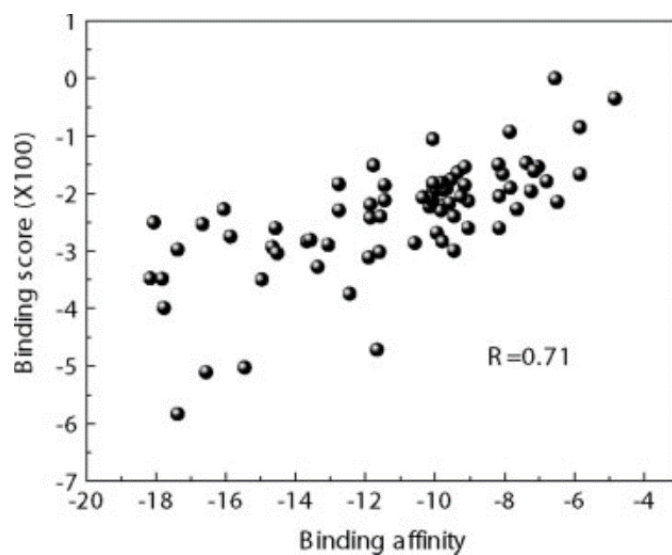


Figura 34. Puntaje de HSYMDOCK. Relación entre el puntaje (eje y) & la afinidad de unión (eje x) de 74 complejos proteína-proteína, la afinidad de unión se mide con el logaritmo natural de la constante de disociación ($\ln(K_D)$). Tomado de (Huang S & Zou X, 2008).

Se utilizó HSYMDOCK para encontrar homoligómeros simétricos, utilizando los monómeros estructurales, predichos con Alphfold de las siguientes TIMs:

- Mma TIM
- Mae TIM
- Mfe TIM
- Ton TIM
- Mok TIM
- Mbr TIM
- Mbu TIM

UCSF CHIMERA

Todas las figuras estructurales de esta tesis se realizaron con el software para visualización de estructuras UCSF Chimera (<https://www.cgl.ucsf.edu/chimera/>). Con Chimera también calculamos la hidrofobicidad de la superficie de todas las estructuras, de acuerdo con la escala de hidropatía de Kyte & Doolittle (1982). La escala de hidropatía refleja la hidrofobicidad (valor máximo 4.5) o hidrofilia (valor mínimo -4.5) relativa. Para obtener dicha escala Kyte & Doolittle recopilaron los coeficientes de partición y la energía libre de transferencia de cada aminoácido (Tabla 6) y lo correlacionaron con la accesibilidad para cada tipo de residuo calculada por Chothia en 12 proteínas globulares (Tabla 6). Para poder comparar ambos parámetros Kyte & Doolittle normalizaron ambos parámetros. Esta normalización es la escala que utiliza Chimera para asignar la hidrofobicidad a cada aminoácido en la superficie (Kyte J & Doolittle R, 1982).

Cadena lateral	Escala de hidropatía	ΔG de transferencia	Frecuencia de las cadenas laterales queden escondidas
ILE	4.5	4.4	4.5
VAL	4.2	4.2	4.3
LEU	3.8	4.5	3.2
PHE	2.8	2.5	2.5
CYS	2.5	1.9	6
MET	1.9	1.9	1
ALA	1.8	3.9	5.3
GLY	-0.4	-	4.2
THR	-0.7	-0.6	-0.5
TRP	-0.9	-0.9	-2.4
SER	-0.8	-0.8	-0.7
TYR	-1.3	-1.1	-3.3
PRO	-1.6	-	-2.4
HIS	-3.2	-4.2	-3.6
GLU	-3.5	-3.9	-2.8
GLN	-3.5	-3.5	-4.0
ASP	-3.5	-4.5	-2.5
ASN	-3.5	-3.8	-3.1
LYS	-3.9	-3.2	-
ARG	-4.5	-	-

Tabla 6. Escala de hidropatía utilizada por Chimera. Tomada de (Kyte and Doolittle 1982).

PDB2PQR & APBS

Para determinar la superficie electrostática en todas las estructuras utilizamos PDB2PQR & APBS que colaboran mutuamente, PDB2PQR es un servidor que automatiza la preparación de estructuras para cálculos electrostáticos y APBS (Adaptive Poisson-Boltzmann Solver en inglés) resuelve la ecuación de Poisson-Boltzmann para toda la estructura macromolecular. Los métodos electrostáticos computacionales se pueden dividir en dos categorías: Los de “solvente explícito” que tratan con todas las moléculas del solvente y los de “solvente implícito” que incluyen sólo las interacciones soluto-solvente. Los métodos de solvente implícito utilizan parámetros como la temperatura, fuerza iónica de solvente y las coordenadas atómicas, así como las cargas y los radios de los átomos (Dolinsky T, et. al. 2004).

En este caso se realiza el cálculo por medio de un método de solvente implícito. En el servidor el usuario proporciona un archivo PDB. En el primer paso el servidor identifica los átomos pesados (que no son hidrógenos) y reconstruye las cadenas laterales faltantes, asignándoles coordenadas a los átomos perdidos (Figura 35). En la parte de valoración de estado (titration-state) los ángulos χ de las cadenas laterales de las H N y Q se acomodan en una conformación óptima para los puentes de hidrógeno por medio de métodos Montecarlo y después se les asignan estados de protonación. Luego se añaden los átomos de hidrógeno, los cuales se revisan para evitar impedimentos estéricos y para optimizar toda la red de hidrógenos en la estructura. Posteriormente, el servidor asigna cargas y radios atómicos basados en el campo de fuerza que utilice el usuario “CHARMM22”, “AMBER99”, o “PARSE” (Figura 35) (Dolinsky T, et. al. 2004) (Jurrus E, et. al. 2018).

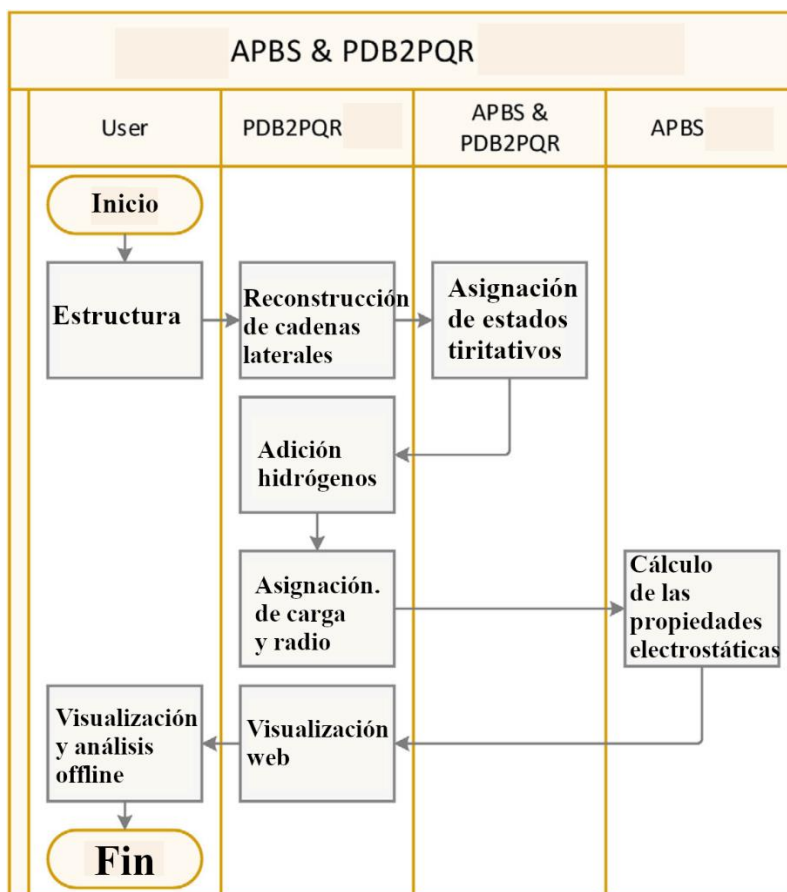


Figura 35. Flujo de trabajo de los servidores PDB32PQR & APBS que trabajan en conjunto. Tomado de (Jurrus E, et. al. 2018).

A partir de las cargas y radios atómicos, APBS realiza el cálculo de la ecuación Poisson-Boltzmann: resultando en la función de distribución de la carga ρ (Jurrus E, et. al. 2018).

$$-\nabla \cdot \epsilon \nabla \phi - \sum_i^M c_i q_i e^{-\beta(q_i \phi + V_i)} = \rho. \quad (7)$$

En donde ϕ es el potencial, ϵ_s es la constante dieléctrica del solvente. M representa las especies iónicas móviles, que se describen por una carga q_i , una concentración c_i en un potencial de interacción ion-soluto V_i . Finalmente, $\beta=(kT)^{-1}$ indica el inverso de la energía térmica, donde k es la constante de Boltzmann y T es la temperatura, la cual se mantiene constante en 298K (Jurrus E, et. al. 2018).

La visualización de la distribución de la carga (ρ) a lo largo de todas las estructuras se realizó en Chimera.

ANÁLISIS ACOPLADO ESTADÍSTICAMENTE (SCA)

El SCA (Statistical coupling analysis en inglés) es un método que utiliza el alineamiento de secuencias de proteínas homólogas para identificar redes de aminoácidos las cuales reflejan patrones de coevolución (Mira DL, et. al. 2018). El alineamiento múltiple de secuencias (MSA en inglés) es un registro histórico, si es lo suficientemente largo y diverso puede describir las limitaciones evolutivas de la familia (Gürol MS, etl. Al. 2002). Estas restricciones están involucradas en la oligomerización y en las interacciones proteína-proteína, o proteína-sustrato. La evolución de una proteína es el resultado de muchas mutaciones aleatorias sujetas a presión selectiva que al final resultan en una función (Lockless & Ranganathan, 1999). La presión selectiva hace que las substituciones de aminoácidos deletéreas se eliminen y algunas otras se conserven, asegurando la similitud del plegamiento entre familias para mantener interacciones favorables en la proximidad espacial (Marks, et. al. 2011).

Un ejemplo de la sinergia entre aminoácidos es que, aunque no estén acoplados tridimensionalmente, algunas veces, aminoácidos distantes pueden contribuir a la función del activo, o determinar la interacción proteína-proteína (Marks, et. al. 2011). Otra forma en la que se aprecia la correlación entre los pares de aminoácidos coevolutivos es por ejemplo, en una mutación que es deletérea en una posición, pero se convierta en neutra o se compense con otra mutación en un sitio diferente. Otro ejemplo es el escenario en el que dos mutaciones individuales son poco frecuentes pero la combinación de ambas es común (Tiberiu T. et. al. 2015). Se denominan “sectores” a estas redes de aminoácidos acoplados funcional y evolutivamente. Los sectores han mostrado ser interesantes porque están relacionados con las propiedades funcionales como la regulación alostérica y los sitios catalíticos, así como con la estabilidad del plegamiento (Rivoire O. et. al. 2016). El objetivo de los métodos basados en coevolución es proponer modelos de los patrones de aminoácidos que se puedan probar experimentalmente después, para corroborar si realmente tienen un significado bioquímico o evolutivo (Rivoire O. et. al. 2016).

El primer paso del SCA es preparar nuestro alineamiento de secuencias múltiples. Primero se alinea el MSA con una secuencia consenso elegida por el usuario, para que los resultados de la red de aminoácidos coincidan con la secuencia elegida. Después se eliminan todas las secuencias y columnas con más del 20% de *gaps*, así como las secuencias con menos del 20% y más del 80% de identidad con la secuencia consenso (Rivoire O. et. al. 2016).

Para este trabajo se realizaron tres alineamientos múltiples de secuencia, cada uno correspondiente con los tres dominios de la vida: Arquea, Bacteria y Eucaria. En la lista mostrada a continuación, se describen las secuencias consenso para

cada uno de los alineamientos, así como las secuencias resultantes después del proceso de limpieza.

ARQUEA

- Secuencia de referencia: TIM de *Pyrococcus furiosus*, PDB: 1HG3, p=0.965

“MAKLKEPIIAINFKTYIEATGKRALEIAKAAEKVYKETGVTIVVAPQLVDLRMIASVEIPVFA
QHIDPIKPGSHTGHVLP EAVKEAGAVG TLLNHSENRMILADLEAAIRRAEEVGLMTMVCS
NNPAVSAAVAALNP DYVAVEPPELIGTGIPVSKAKPEVITNTVELVKKVNPEVKVLCGAGIS
TGEDVKKAIELGTVGVLLASGVTKAKDPEKAIWDLVSGI”

Después del procesamiento la longitud del alineamiento es de 214 posiciones con 367 secuencias efectivas.

BACTERIA

- Secuencia de referencia: La TIM de *Geobacillus stearothermophilus*, PDB: 1BTM, p=0.965

“RKPIIAGNWKMHKTLAEAVQFVEDVKGHVPPADEVISVVCAPFLFLDRLVQAADGTDLKI
GAQTMHFADQGGAYTGEVSPVMLKDLGVTVILGHSERRQMFAETDETVNKKVLAAFRGLI
PIICCGESLEEREAGQTN AVVASQVEKALAGLTPEQVKQAVIAYEPIWAIGTGKSSTPEDAN
SVCGHIRSVSRLFGPEAAEAIRIQYGGSVKPDNIRDFLAQQQIDGPLVGGASLEPASFLQL
VEAGRHE”

Después del procesamiento la longitud del alineamiento es de 249 posiciones con 698 secuencias efectivas.

EUCARIA

- Secuencia de referencia: La TIM de *Oryctolagus cuniculus*, PDB: 4OWG p=0.965

“APSRKFFVGGNWKMN GRKKNL GELITTLNAAKVPADTEVVCAPPTAYIDFARQKLDPKIA
VAAQNCYKVTNGAFTGEISPGMIKDCGATWVVLGHSERRHVFGESDELIGQKVAHALSEG
LGVIACIGEKLDEREAGITEKVVFEQTKVIADNVKDW SKVVLAYEPVWAIGTGKTATPQQAQ
EVHEKLRGWLKSNVSDAVAQSTRIIYGGSVTGATCKELASQPDVDGFLVGGASLKPEFVDI
INAKQ”

Después del procesamiento la longitud del alineamiento es de 247 posiciones con 234 secuencias efectivas.

El segundo paso del método es calcular la conservación de los aminoácidos en el alineamiento por medio de un modelo llamado entropía relativa (D_i) o divergencia de Kullback-Leibler (ecuación 8). D_i es una medida de cómo una distribución de

probabilidad es diferente de una segunda distribución de probabilidad de referencia. La fórmula general se puede observar en la siguiente ecuación. Donde P es la distribución de nuestro interés y Q es nuestra distribución de referencia. En la (Figura 36.A) se muestran ambas distribuciones y en 36.B el resultado de la divergencia (Rivoire O. et. al. 2016).

$$D_{KL}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log \left(\frac{P(x)}{Q(x)} \right). \quad (8)$$

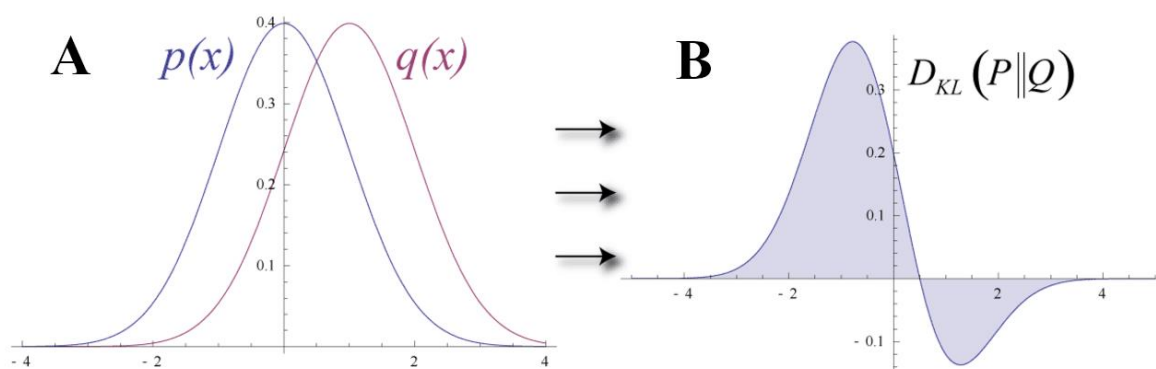


Figura 36. Divergencia de Kullback-Leibler. Tomada y modificada de (https://wblog.wiki/es/Relative_entropy)

El SCA utiliza como distribución de referencia la frecuencia de los aminoácidos de una base de datos no redundante de proteínas (q) (Tabla 7), la cual refleja la presencia de cada uno de los 20 aminoácidos en muchas secuencias y tipos de proteínas (Rivoire O. et. al. 2016).

$$D_i = \sum_{a=0}^{20} f_i^{(a)} \ln \left(\frac{f_i^{(a)}}{\bar{q}^{(a)}} \right), \quad (9)$$

A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
0.073	0.025	0.050	0.061	0.042	0.072	0.023	0.053	0.064	0.089	0.023	0.043	0.052	0.040	0.052	0.073	0.056	0.063	0.013	0.033

Tabla 7. Frecuencia de todos los aminoácidos en una base de datos no redundante de proteínas.

Si la frecuencia de nuestro interés es el aminoácido a en la posición i (f_i) es igual a q indicaría que su presencia en esa posición es completamente azarosa. Así que nuestro valor D_i es la diferencia entre estas dos distribuciones. El valor D_i crece mientras más se aleja del valor de q , hasta tomar valores de 3 o 3.5 cuando el aminoácido está 100% conservado (Figura 37) (Rivoire O. et. al. 2016). Se ha visto

que valores altos de este valor son capaces de predecir residuos ocultos en la estructura terciaria (Halabi et. al. 2009)

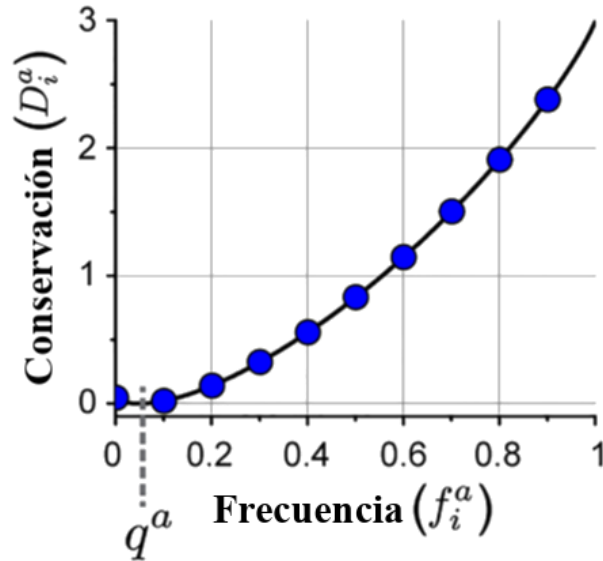


Figura 37. Relación entre la conservación, frecuencia y entropía relativa. Tomado y modificado de (Rivoire O. et. al. 2016).

El tercer paso del método es hacer una matriz de correlación entre las posiciones del alineamiento, la cual se obtiene por medio de la diferencia entre las frecuencias de los aminoácidos a y b en las posiciones i & j . Las frecuencias son básicamente promedios y se describen con las siguientes ecuaciones donde N es el número de secuencias totales en el alineamiento y f_i es la frecuencia de un aminoácido en la posición i y f_{ij} nos indica, que tan frecuente es que se encuentren ambos aminoácidos a y b en sus posiciones i & j respectivamente ($f_{ij}^{(ab)}$).

$$f_{ij} = \frac{1}{N} \sum_k X_{ki} X_{kj}, \quad f_i = \frac{1}{N} \sum_k X_{ki}, \quad (10)$$

La matriz de correlación es la diferencia de ($f_{ij}^{(ab)}$) menos la multiplicación de la frecuencia de que se encuentre el aminoácido a en la posición i solamente ($f_i^{(a)}$) por la frecuencia de que se encuentre el aminoácido b en la posición j únicamente ($f_j^{(b)}$). Esta matriz tiene dimensiones $L \times L \times 20 \times 20$, en donde L es la longitud del alineamiento, por 20×20 de todas las posibles combinaciones de aminoácidos (Reynolds, et. al. 2013).

$$C_{ij}^{(ab)} = f_{ij}^{(ab)} - f_i^{(a)} f_j^{(b)} \quad (11)$$

Para el tercer análisis necesitaremos un valor de ponderación ϕ . Dicho valor se utiliza para amortiguar los valores de las posiciones menos, evitando un sesgo a la hora de hacer la correlación entre residuos. Este valor está dado por el valor absoluto de la división entre la derivada parcial de la entropía relativa del aminoácido a en la posición i y la frecuencia de ese aminoácido en esa posición. De esta manera, se obtienen valores normalizados (Reynolds, et. al. 2013).

$$\phi_i^{(a)} = \left| \partial D_i^{(a)} / \partial f_i^{(a)} \right| \quad (12)$$

$$\phi = \frac{\partial D}{\partial f} = \ln \left[\frac{f(1-q)}{q(1-f)} \right] \quad (13)$$

A partir de estos valores es posible calcular una segunda matriz de correlación, esta vez ponderada, que elimina el sesgo de la primera matriz. En este paso se multiplica el valor de ponderación del aminoácido a en la posición i , el valor de ponderación del aminoácido b en la posición j y la matriz de correlación obtenida en el paso anterior. Esta matriz resulta ser de cuatro dimensiones $L \times L \times 20 \times 20$ (Reynolds, et. al. 2013).

$$\tilde{C}_{ij}^{(ab)} = \phi_i^{(a)} \phi_j^{(b)} C_{ij}^{(ab)}, \quad (13)$$

Con fines de análisis es necesario reducir la matriz por una de $L \times L$ por medio de una norma de Frobenius, el cual reduce dos dimensiones. Que reduce cada correlación de aminoácidos (dimensión 20×20) solo les asigna un valor por cada par de aminoácidos. La matriz reducida nos da el valor de las correlaciones de cada par de posiciones (i,j) de todas las posibles combinaciones de aminoácidos (20×20). Esta es la última matriz que utilizamos a la que llamaremos matriz-SCA (Reynolds, et. al. 2013).

$$\tilde{C}_{ij} = \sqrt{\sum_{(ab)} \left(\tilde{C}_{ij}^{(ab)} \right)^2}. \quad (14)$$

Las matrices de dos dimensiones se pueden analizar como un plano (x, y) (Figura 38.A) A los planos se les pueden aplicar transformaciones lineales, que modifican un espacio vectorial en otro (Figura 38.B). Cuando realizamos la transformación hay ciertos vectores que cambian completamente de orientación y dimensiones, como el que se muestra con línea rosa en la Figura 38.C.D. Existen además otros vectores especiales que conservan su estado inicial, a pesar del cambio en el espacio vectorial (Figura 38.E.F) a estos últimos se les conoce como eigenvectores. Estos conservan su orientación, pero pueden escalarse en su magnitud, en la Figura 38.G.H se puede observar cómo los eigenvectores marcados en líneas amarillas y verdes conservaron su orientación, pero se escalaron en magnitud, el amarillo por dos y el verde por tres. Los valores que escalan la magnitud de los eigenvectores se conocen como eigenvalores (Figura 38.H).

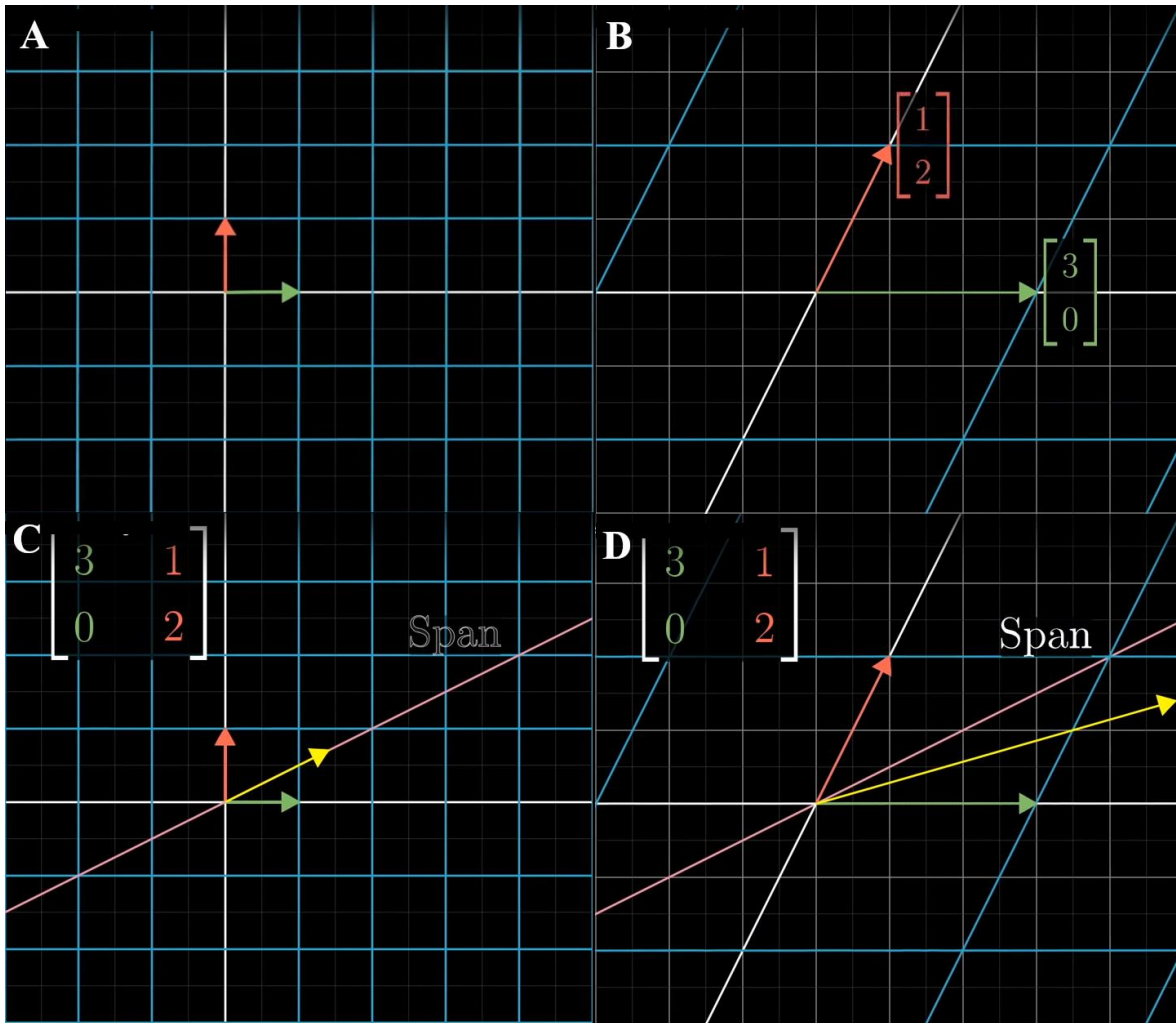
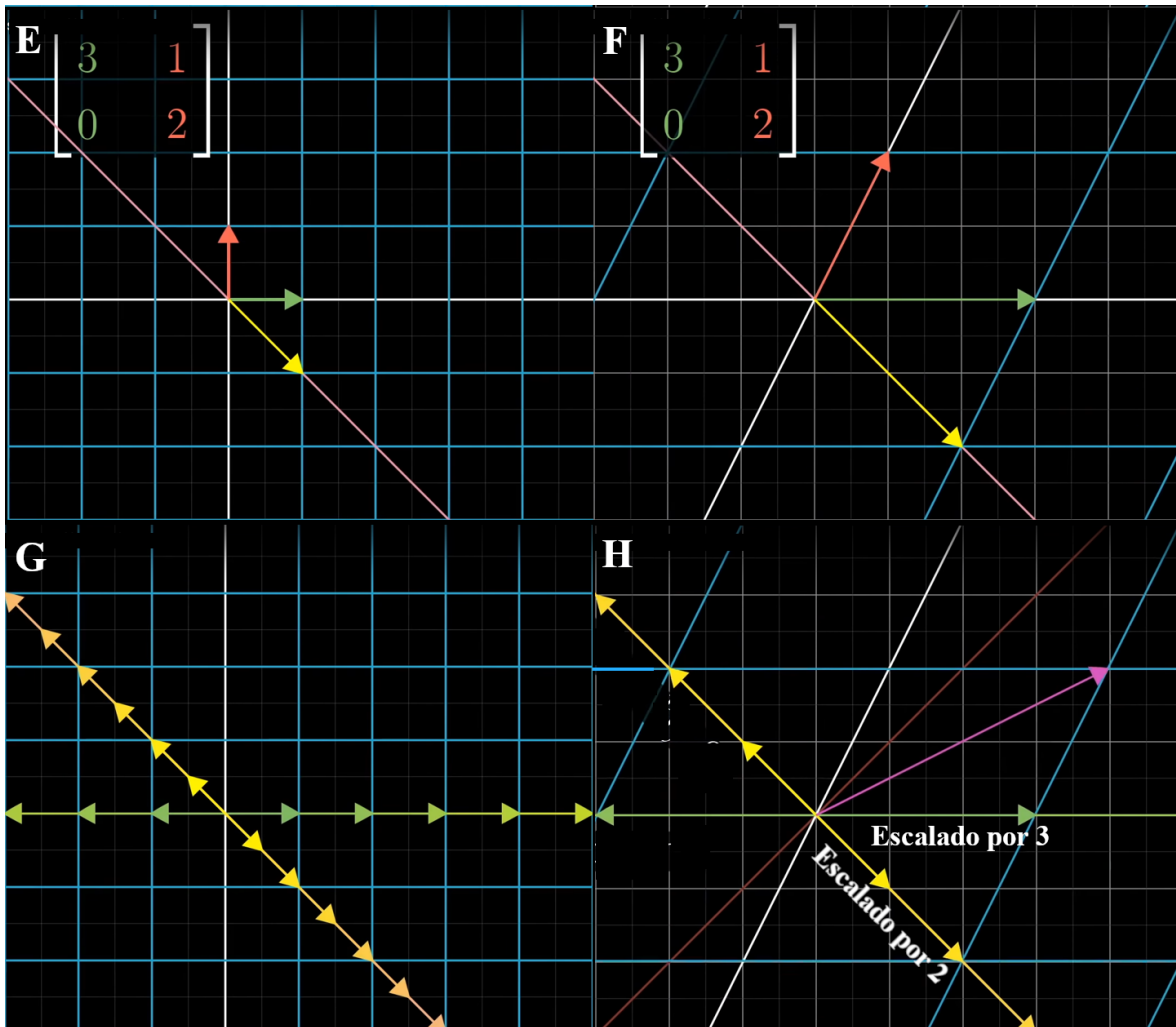


Figura 38. Definición de Eigenvalores y Eigenvectores. A) Plano de dimensiones $[1,1]$. B) Transformación del plano $[1,1]$ a un plano de dimensiones $[1,2][3,0]$. C) Vector rosa en el espacio $[1,1]$. D) Visualización gráfica de como no se conserva el vector rosa del plano original al plano de nuevas dimensiones. E, F) Conservación del vector rosa (eigenvector) en ambos planos sin importar el cambio en dimensiones. G) Muestra de los eigenvectores de nuestra transformación lineal H) Muestra gráfica de como los vectores no cambian de dirección (eigenvectores), pero cambian de magnitud (eigenvalores). Tomado y modificado de (<https://www.youtube.com/watch?v=PFDu9oVAE-g&t=614s>)



¿Cómo se calculan estos vectores en una matriz? Resulta ser la matriz multiplicada por los eigenvectores es igual al producto de los eigenvalores por los eigenvectores.

$$A\vec{v} = \lambda\vec{v} \quad (15)$$

Donde A es la matriz, v son los eigenvectores y λ son los eigenvalores. Primero, λ se convierte en una matriz diagonal (I). Posteriormente se iguala todo a 0 y se factoriza. Así que nos queda la matriz original y en la diagonal de la misma se resta el valor de λ . Lo que se hace es determinar el valor de λ para el cual la matriz se reduce a la dimensión 0 (Figura 41. A.B). Por último, para determinar los eigenvectores sólo se sustituye el valor de λ como resta en la diagonal de la matriz original.

$$A\vec{v} = (\lambda I)\vec{v} \quad \lambda \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_I$$

(16)

$$\begin{aligned} A\vec{v} - \lambda I\vec{v} &= 0 \\ (A - \lambda I)\vec{v} &= 0 \\ \det(A - \lambda I) &= 0 \end{aligned} \quad \begin{bmatrix} 3 - \lambda & 1 & 4 \\ 1 & 5 - \lambda & 9 \\ 2 & 6 & 5 - \lambda \end{bmatrix}$$

(17)

Realicemos estas operaciones para nuestro ejemplo (Figura 39. C.D) en este caso, los únicos valores de λ posibles son 2 y 3. Al sustituir el valor de λ en la matriz obtenemos como resultado el eigenvector [1,-1]. Todos los elementos que sean parte de esta línea (-1,1) (-2,2) (-3,3) etc. Se les conoce como eigenmodos.

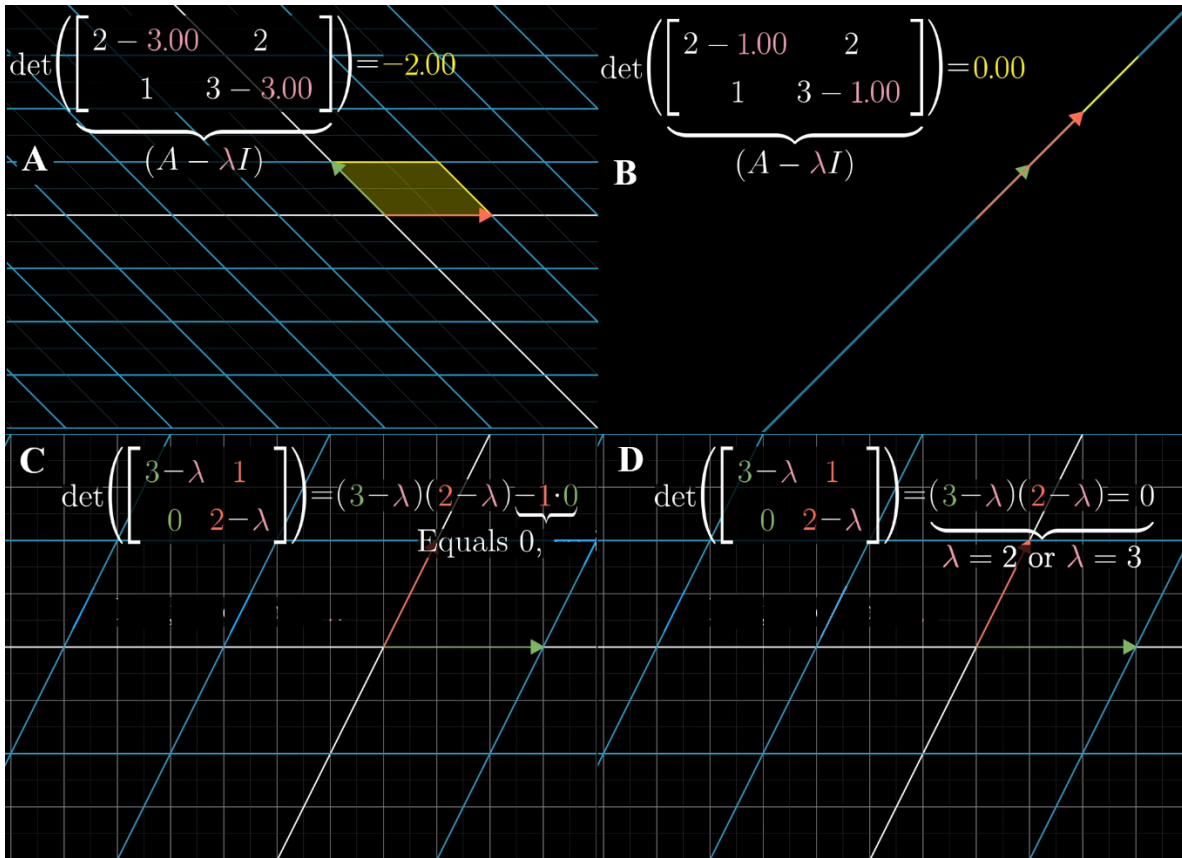


Figura 39. Determinación de los eigenvectores y eigenvalores. Para encontrar los eigenvectores se sustituye el valor λ buscando que la matriz se iguale a 0. A) en este caso el valor λ iguala la matriz a dos, por lo que se tiene que seguir buscando otros valores. B) El valor $\lambda=1$ iguala la matriz a 0 disminuyendo completamente sus dimensiones, por lo que al substituir λ en la matriz encontraremos los vectores. Para encontrar los eigenvalores consideremos nuestro ejemplo de la transformación lineal del espacio [1,1] a un plano de dimensiones [1,2][3,0]. C) Muestra la igualación de los elementos de la matriz a 0. D) Tenemos la ecuación resuelta, por lo que los únicos dos eigenvalores posibles en este ejemplo son $\lambda=2$ & $\lambda=3$. Tomado y modificado de (<https://www.youtube.com/watch?v=PFDu9oVAE-g&t=614s>)

El proceso para encontrar esta información es la descomposición espectral de la matriz SCA basada en los principios descritos arriba. La ecuación inferior (18) describe un producto de matrices, donde $\sim\Delta$ es una matriz diagonal de $L \times L$ con los eigenvalores y $\sim V$ son los eigenvectores. En nuestro caso los eigenvectores corresponden a los patrones de evolución encontrados en el alineamiento, los eigenmodos corresponden a los pares de aminoácidos coevolutivos asociados a cada vector o patrón de coevolución y finalmente los eigenvalores son la magnitud de varianza en la matriz de covarianza SCA (Rivoire O. et. al. 2016).

$$\tilde{C} = \tilde{V} \tilde{\Delta} \tilde{V}^T, \quad (18)$$

Uno de los grandes retos de la metodología es distinguir entre las parejas coevolutivas reales de todo el ruido estadístico. Con el fin de disminuir el ruido estadístico se realiza un alineamiento múltiple aleatorizado. Este se genera a partir del alineamiento que usamos inicialmente y revuelve todos los aminoácidos por columna, esto preserva las constricciones de las posiciones, mientras que remueve todo el ruido de muestreo. A este alineamiento aleatorio se le aplican los mismos cálculos hasta la descomposición espectral y se comparan sus resultados con el alineamiento inicial. Se compara con la siguiente gráfica (Figura 40) en donde las bandas negras son el espectro de los eigenvectores de la matriz SCA, y en línea roja se muestra el espectro de los eigenvectores de la matriz del alineamiento aleatorio. Al hacer la comparación podemos descartar todos los vectores negros que coincidan con la línea roja. Así obtenemos los eigenvectores con los eigenvalores más altos (ES) que no se forman por un proceso aleatorio (Rivoire O. et. al. 2016).

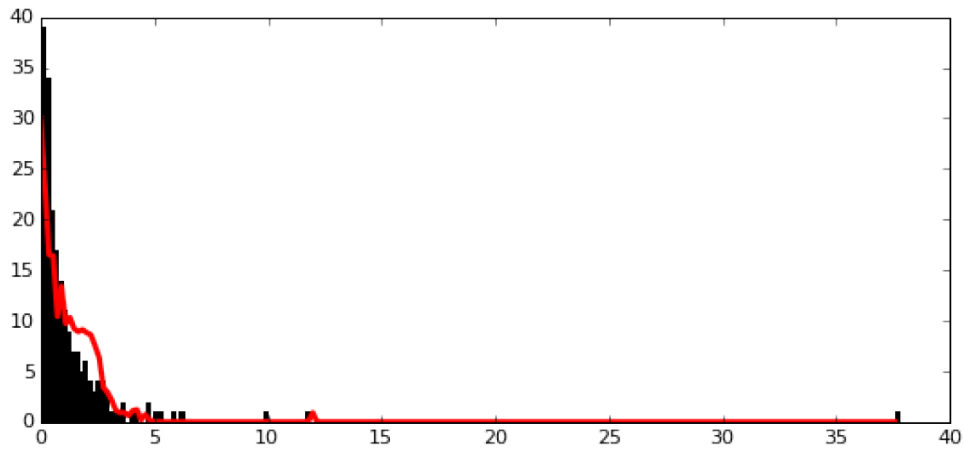


Figura 40. Espectro de los eigenvectores de la matriz SCA. El espectro de los eigenvectores de la matriz SCA se muestra en el histograma con barras negras, mientras que el espectro de los eigenvectores de la matriz del alineamiento aleatorio se muestra con línea roja.

Los elementos de los eigenvectores se pueden visualizar en gráficas de dispersión (Figura 41.A), en donde la distancia de una posición con relación al origen informa su contribución total a las correlaciones, mientras mayor sea dicha distancia al punto de origen = 0, mayor es su contribución. La distancia entre dos posiciones indica su grado de correlación, las posiciones fuertemente correlacionadas aparecen cercanas, mientras que las posiciones débilmente correlacionadas se

separan. Como podemos, ver la mayoría de los aminoácidos no hacen contribuciones sustanciales así que se agrupan cerca del origen. Por lo que se espera que conjuntos independientes de posiciones correlacionadas se agrupen en pequeños conjuntos de posiciones lejos del origen. La descomposición espectral nos ayuda con la reducción de dimensiones de la matriz, pero sus resultados (los eigenmodos) no son la mejor representación de los residuos coevolutivos. Ya que la simple diagonalización de la matriz no exige un acoplamiento estadísticamente significativo, como vemos en las gráficas de dispersión, muchas posiciones están débilmente correlacionadas (Rivoire O. et. al. 2016).

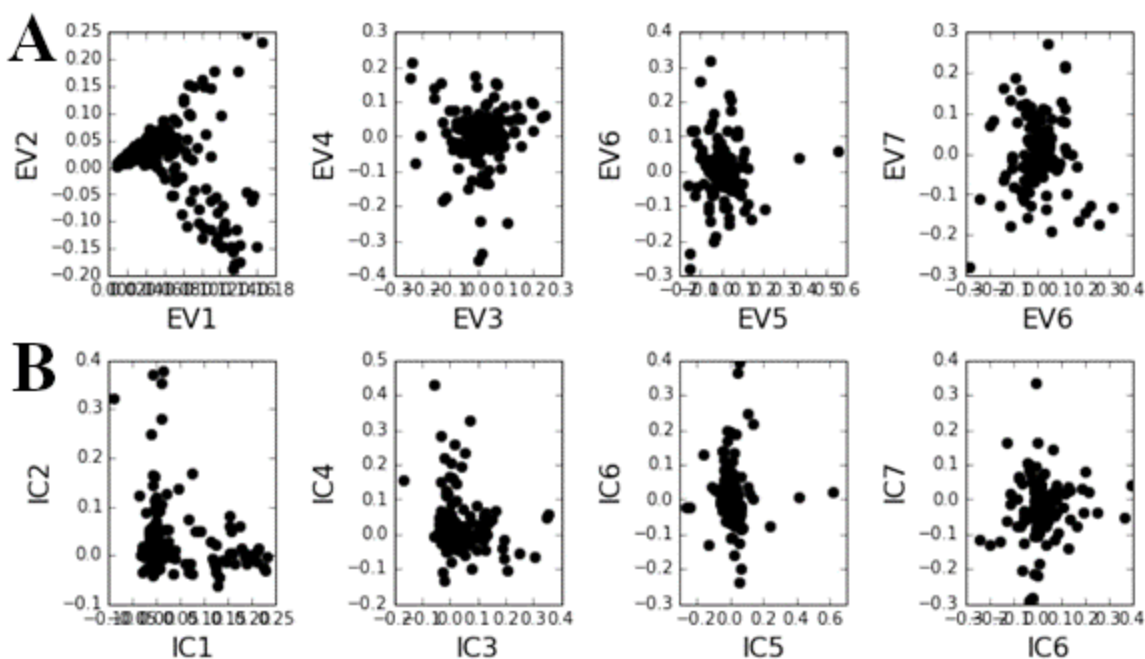


Figura 41. Gráficos de dispersión de los elementos de los eigenvectores. A) Elementos de los eigenvectores. B) Elementos de los eigenvectores bajo el análisis de componentes independientes (ICA).

Esto se solucionó con el análisis de componentes independientes (ICA en inglés). ICA produce una representación en la que la mayoría de las posiciones están débilmente correlacionadas y se agrupan cerca del origen y un subconjunto relativamente pequeño de posiciones comprende grupos casi independientes de aminoácidos (Figura 41.B), que representan modos de evolución colectiva de la proteína bajo estudio. Para realizar el ICA se toman los eigenvectores (k^*) y se concatenan en una matriz $k^* \times L$ (V) en la siguiente ecuación. Lo que hace el algoritmo es iterar la matriz W que empieza desde $k^* \times k^*$. El parámetro ρ es una tasa aprendida que es muy chica para que las iteraciones converjan (Rivoire O. et. al. 2016).

$$\Delta W = \rho \left(I_{k_{\text{top}}} + \left(1 - \frac{2}{1 + \exp(-WV)} \right) (WV)^{\top} \right) W. \quad (19)$$

Los componentes independientes $\sim V^p$ se obtienen al aplicar la matriz W a los eigenvectores $\sim V$.

$$\tilde{V}_{1\dots k^*}^p = W \tilde{V}_{1\dots k^*}. \quad (20)$$

Es importante aclarar que el resultado de los componentes independientes no necesariamente equivale a un sector, ya que éste es una red de aminoácidos coevolutivos con una propiedad funcional específica. Así que los componentes independientes pueden ser: 1) Sectores independientes asociados a una función específica, o bien, 2) La descomposición de un sector que se separó en varias subpartes. Para distinguir sistemáticamente la composición de los sectores primero se ajustan los datos de IC a una distribución normal para encontrar el 5% de los aminoácidos más significativos $p=0.95$ (Figura 42.A) (Rivoire O. et. al. 2016). Cabe destacar que p es un parámetro ajustable. Para fines de reducir el ruido de muestro en esta tesis se trabajaron los datos con un valor de $p=0.965$

Después se hace otra matriz que contiene solo los aminoácidos resultantes de la limpieza de los IC donde podemos visualizar las relaciones de estos aminoácidos entre cada IC y de todos los IC. Así que aquí podemos distinguir si los IC tienen dependencia estadística unos con otros y agruparlos para que sean un solo sector (Figura 42.B) (Rivoire O. et. al. 2016). De esta forma obtenemos el resultado final: la red de aminoácidos que son las parejas más significativas que el 95% de los componentes del IC.

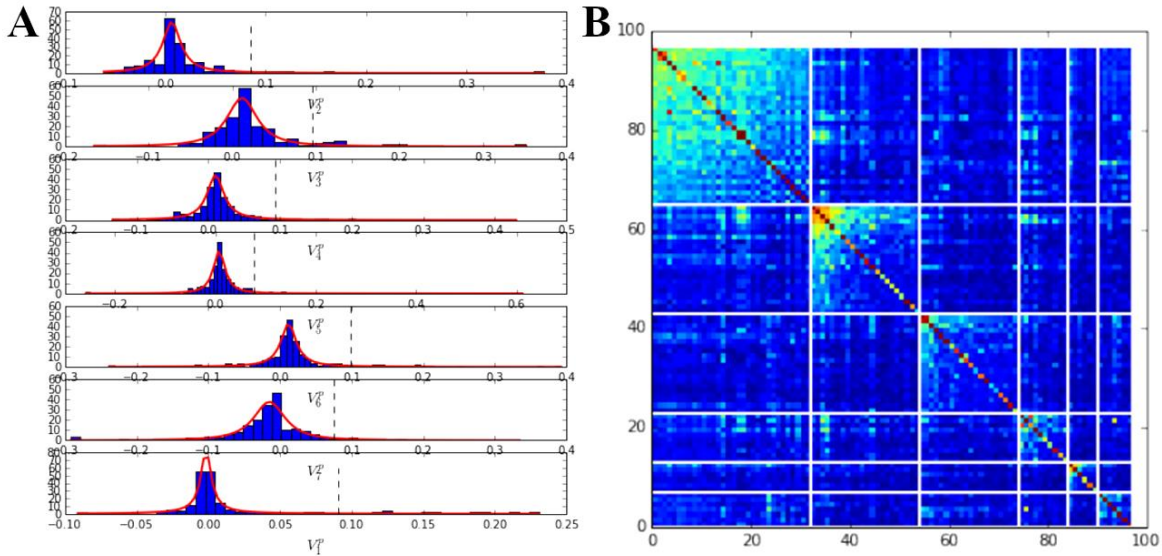


Figura 42. Identificación de los sectores por parte del SCA. A) Normalización de los datos de los componentes independientes (IC). B) Matriz de las interacciones entre los aminoácidos más significativos de los componentes independientes.

Un paso adicional que se añade en la última versión del SCA, es realizar los mismos cálculos de covarianza, pero en lugar de hacerlo por columnas (posiciones) hacerlo con las filas (secuencias). Esto nos daría como información la relación entre los patrones coevolutivos y el espacio de secuencia dado por las secuencias del alineamiento (Rivoire O. et. al. 2016).

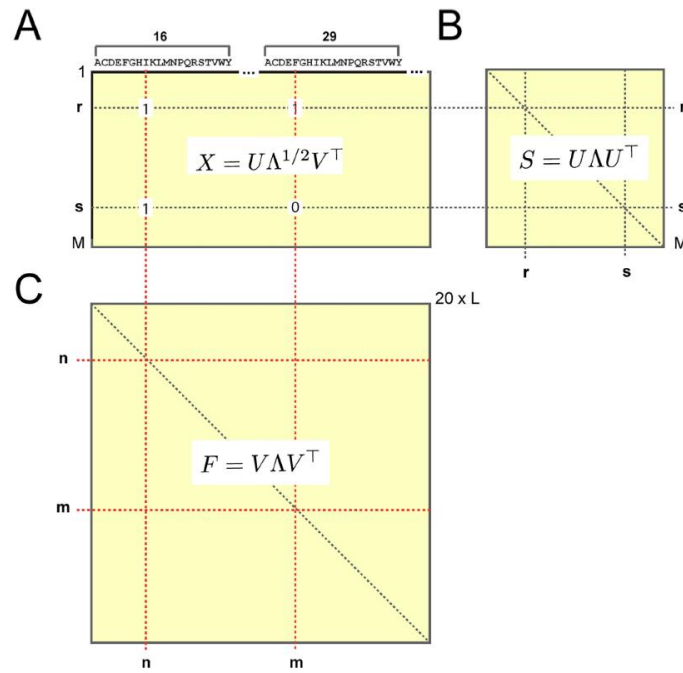


Figura 43. Relación entre el alineamiento múltiple de secuencias y las matrices de covarianza. A) Alineamiento múltiple de secuencias en donde las columnas son las posiciones y las filas son las secuencias. B) Matriz de correlación entre las secuencias con dimensiones $M \times M$. C) Matriz de correlación entre posiciones de dimensiones $20 \times L$.

Para los cálculos es necesario considerar una matriz binaria de un alineamiento X (Figura 43.A) de M secuencias. Para X podemos computar dos tipos de correlación:

1). La correlación entre las secuencias S_{rs} que representa la identidad de cada par de secuencias r y s (Figura 45.B).

$$S_{rs} = \frac{1}{L} \sum_n X_{rn} X_{sn}, \quad (21)$$

2). La matriz de correlación entre las columnas, que representa la frecuencia de un par de aminoácidos juntos (Figura 38.C).

$$F_{nm} = \frac{1}{M} \sum_s X_{sn} X_{sm}, \quad (22)$$

Ambas matrices (S & F) están íntimamente relacionadas por una propiedad conocida como la descomposición de un valor singular (SVD). Específicamente, si U representa los eigenvectores de la matriz de correlación de secuencia S ; y V representa los eigenvectores de la correlación de aminoácidos de la matriz de correlación F .

$$X = U \Lambda^{1/2} V^T, \quad (23)$$

Si Λ es una matriz diagonal en donde las entradas son los eigenvalores de S y F . Una clave del concepto SVD es que los eigenvectores de S están mapeados por los eigenvectores de F en donde el mapa es el alineamiento en sí mismo (X).

$$U = X V \Lambda^{-1/2}. \quad (24)$$

Esto introduce el principio del mapeo secuencia-posición en donde se usa la matriz del alineamiento completo X para relacionar los patrones de las correlaciones-aminoácido V con los patrones de divergencia U . Después se aplica la reducción de dimensiones realizada anteriormente con la matriz SCA. En donde los valores de los 20 aminoácidos pasan a un solo valor. La matriz reducida define el mapa entre los aminoácidos coevolutivos y su relación en el espacio de secuencia (Rivoire O. et. al. 2016).

$$\tilde{U} = x\tilde{V}\tilde{\Delta}^{-\frac{1}{2}} \quad (25)$$

Específicamente $\tilde{\Delta}$ son los eigenvalores y \tilde{V} son los eigenvectores de la matriz posicional del SCA Cij. Así que representan la estructura del espacio secuencial correspondiente a los patrones posicionales coevolutivos. Después se realiza el análisis de componentes independientes ICA.

$$\tilde{U}^P = W\tilde{U} \quad (26)$$

Así, podemos rastrear que relación filogenética o de secuencia tienen los sectores (Rivoire O. et. al. 2016).

RESULTADOS

CARACTERIZACIÓN DE LOS HOMOTETRÁMEROS

Hasta el momento (mayo de 2022) existen únicamente tres estructuras cristalográficas de TIMs de Arquea publicadas en el PDB, estas son: PfuTIM (1HG3) MjaTIM (2H6R) & TtxTIM (1W0M). Para añadir una nueva, nuestro grupo de investigación resolvió la estructura cristalográfica de MfoTIM a una resolución de 2.5 Å (Figura 44.A). El cristal resultó tener cuatro monómeros en la unidad asimétrica, ensamblados como un dímero de dímeros, es decir, dos pares de monómeros se asocian como un dímero clásico por medio de las asas 1 y 3, en la Figura 44.A se muestra el asa 3 en rojo. Este par de dímeros se asocian entre sí por medio de una nueva interfaz, que llamaremos interfaz tetramérica (Figura 44.A en café y con esferas). En la celda A de la (Figura 44.A), se aprecia como la interfaz tetramérica y la interfaz dimérica, quedan perpendiculares, es decir, a 90° una de la otra. Cuando observamos el tetrámero tridimensionalmente parece un toldo vela, en el que los monómeros que están a 180°, uno enfrente del otro, quedan arriba, mientras que los otros dos quedan abajo; se puede apreciar muy bien con una vista lateral del complejo (Figura 44.B.C). La interfaz tetramérica comprende “tres niveles” de interacción donde participan cinco elementos de estructura secundaria, el primer nivel se forma por un parche simétrico entre el asa $\beta\alpha$ 4 y el amino terminal de la hélice 4 (Figura 44.D), el segundo nivel es la interacción entre las hélices 5 de cada monómero y el asa $\beta\alpha$ 5 (Figura 44.E) y finalmente en el último nivel existe una pequeña interacción entre los últimos aminoácidos del carboxilo terminal de la hélice 6 (Figura 44.F). Todas estas características que encontramos en el cristal nuevo de MfoTIM se conservan para todos los cristales de ArqTIMs publicados en el PDB.

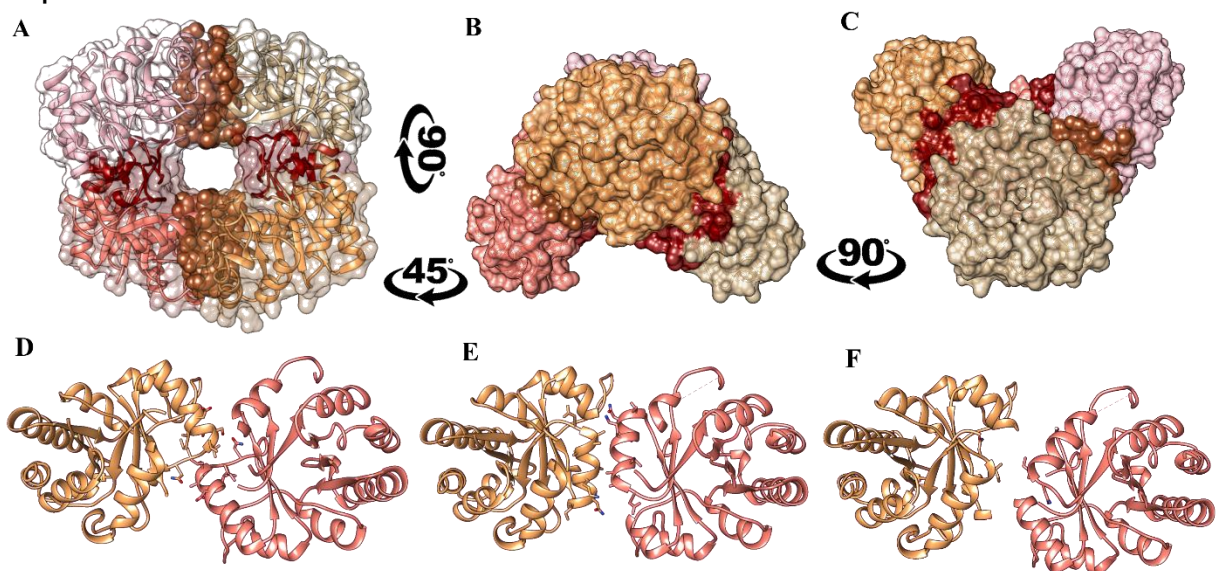


Figura 44. Cristal MfoTIM. A. Vista aérea del tetrámero. B Vista lateral del tetrámero. C Vista a 90°C lateral del tetrámero. D.E.F. Interfaz tetramérica. D. Primer nivel de interacción. E. Segundo nivel de interacción. F. Tercer nivel de interacción

El primer paso para analizar el cristal de Mfo y los cristales ya publicados; fue determinar: el área total, el volumen total y el área que se esconde al formar las interfaces. Incluimos en este análisis al único dímero reportado para una ArqTIM, TacTIM (5CSS) con el fin de contrastar con los tetrámeros. Para comparar la calidad de las interfaces utilizamos el valor de significancia del complejo (CSS) (Complexation Significance Score en inglés), que indica que tan energéticamente favorable es la formación de la interfaz; su valor máximo es de 1 y su valor mínimo de 0.

Proteína	PDB	Estado Oligomérico	Topt °C	Área total Å ²	Volumen Å ³	Área Dímero Å ²	CSS	Área Tetrám Å ²	CSS
MjaTIM	2H6R	Tetrámero	85	30706	73550	1460.5	1	799.1	1
MfoTIM		Tetrámero	75	29621	70383	1464.9	0.452	722.5	0.269
PfuTIM	1HG3	Tetrámero	100	30121	72871	1389.1	0.882	732.2	0.882
TtxTIM	1W0M	Dímero/Tetrámero	86	31286	72776	1480.6	0.524	944.5	0.174
TacTIM	5CSS	Dímero	60	20015	36891	1483.8	0.505	255.8	0.008

Tabla 8. Valores del volumen total y el área de las interfaces dimérica y tetramérica de los cristales ArqTIMs.

Los datos en la Tabla 8 indican que el área involucrada en la formación del dímero es prácticamente igual en todos los cristales, en promedio 1455.7 Å², sin representar una diferencia entre el homodímero y los homotetrámeros. Los valores de CSS para la interfaz dimérica son altos (~0.5) en la mayoría los casos, cabe destacar que los valores más altos los obtuvieron MjaTIM y PfuTIM con 1 y 0.882 respectivamente.

El área involucrada en la formación del tetrámero difiere más, va desde 940 Å² a 722 Å². Además, los valores de CSS varían bastante entre los homotetrámeros, pero concuerdan con lo citado en la literatura: MjaTIM y PfuTIM que viven como tetrámeros en solución tienen los valores de CSS más altos 1 y 0.882, respectivamente. Aunque MfoTIM también es homotetrámero en solución, su valor de CSS de 0.26 es más bajo. Mientras que TtxTIM que vive en un equilibrio dímero/tetrámero tiene el valor de CSS de 0.17 más pequeño, finalmente el dímero de TacTIM tiene muy poca área de asociación tetramérica 255.8 Å² y un valor CSS prácticamente de 0, lo que indica que la tetramerización no es posible en esta proteína. Las áreas totales y los volúmenes de los homotetrámeros casi no varían entre sí, pero si hay una notable diferencia con el dímero, obviamente.

Después analizamos el tipo de residuos y el tipo de interacciones involucrados en la interfaz tetramérica. En la (Tabla 9) se muestran los residuos que participan en esta interfaz conservados en todos los cristales, también se muestra en promedio cuánta de su área total esconden al formar la interfaz; con un código de colores se ejemplifica la naturaleza química de los aminoácidos, en color verde los hidrofílicos, en azul los cargados positivamente, en rojo los cargados

negativamente y en color crema se encuentran los hidrofóbicos. Como podemos observar, la mayoría de los residuos son pequeños y de naturaleza hidrofóbica, algunos aminoácidos ocultan prácticamente el 100% de su área al formar la interfaz, como es el caso de L99, I102, V125, A128 y V129 (tomando la secuencia de MfoTIM como referencia), en la tabla se muestra cómo se conserva la naturaleza hidrofóbica de estos aminoácidos en todos los cristales, a excepción de la V125 que en MjaTIM y TtxTIM es una treonina. Aunque esta sustitución no conserve la naturaleza química del aminoácido si conserva su tamaño. Debido a su nivel de conservación podemos inferir que dichos aminoácidos están sometidos a una presión selectiva que favorece la interacción tetramérica. En el caso de las interacciones hay algunas que se encuentran conservadas en todos los cristales a pesar de las sustituciones en la secuencia de cada proteína. Sin embargo, hay algunas otras interacciones que sólo se conservan en dos o tres cristales.

	MfoTIM	MjaTIM	PfuTIM	TtxTIM	% BSA
asa $\beta\alpha 4$	ASN 91	ASN 89	ASN 95	ASN 92	70.26
	ASN 96	ARG 94	ARG 100	PRO 97	48.6
hélice 4	ILE 97	MET 95	MET 101	LEU 98	54.4
	LEU 98	LEU 96	ILE 102	LYS 99	53
	LEU 99	LEU 97	LEU 103	LEU 100	99
	SER 100	ALA 98	ALA 104	ASN 101	87.1
	ASP 101	ASP 99	ASP 105	ASP 102	4.4
	ILE 102	ILE 100	LEU 106	LEU 103	100
	ASP 103	GLU 101	GLU 107	ALA 104	77.3
	GLU 107	ASN 105	ARG 111	ALA 108	13.12
asa $\beta\alpha 5$	ASN 121	ASN 199	ASN 125	PRO 122	50.3
	ASN 122	ASN 120	ASN 126	ASP 123	20.5
hélice 5	ASN 124	ASN 122	ALA 128	ARG 125	51.8
	VAL 125	THR 123	VAL 129	THR 126	98.8
	LYS 127	LYS 125	ALA 131	LEU 128	59.84
	ALA 128	ALA 126	ALA 132	ALA 129	98.98
	VAL 129	VAL 127	VAL 133	ALA 130	98.05
	ALA 131	ALA 129	ALA 135	ALA 132	85.85
	LEU 132	LEU 130	LEU 136	LEU 133	86.4
	ILE 169	ILE 167	VAL 173	HIS 170	45.9
hélice 6					

Tabla 9. Aminoácidos conservados en los cristales e involucrados en la interfaz tetramérica, en la primer columna se encuentra la estructura secundaria a la que pertenece cada aminoácido y en la última columna el promedio del porcentaje del área total que esconde cada residuo al formar la interfaz tetramérica.

En la (Figura 45) se muestra el mapa de interacción de la interfaz tetramérica, tanto estructural como esquemáticamente, se tomó como ejemplo la estructura de MfoTIM. Como podemos ver en las representaciones estructurales (Figura 45.A.B), los puentes de hidrógeno son siempre con la cadena principal de la proteína, a excepción del puente de hidrógeno entre la N96 y la S100; cabe destacar que esta interacción es propia de MfoTIM y no se presenta en los otros cristales. Así que la unión de esta interfaz es gracias a las interacciones hidrofóbicas de Van der Waals entre los diferentes monómeros.

Como se muestra en la Figura 45.G.H, la principal área de interacción entre los monómeros es la hélice 5, que se sitúa enfrente de la hélice 5 del dímero opuesto. En esta región participan principalmente alaninas, cabe destacar que en todos los tetrámeros se conserva la secuencia "AVAL" en esta zona. Es contraintuitivo pensar que los aminoácidos pequeños como la alanina favorecen nuevas superficies de interacción, sin embargo, esta superficie está muy bien empaquetada por interacciones hidrofóbicas. La segunda área más grande de interacción es el parche simétrico que se forma entre el asa $\beta\alpha_4$ de un monómero y el amino terminal de la hélice 4 del otro (Figura 45 A.B.C.D). Un poco más arriba en la hélice 4 sus cadenas laterales interactúan con las alaninas antes mencionadas de la hélice 5 del monómero contrario (Figura 45. C.D)

En el asa $\beta\alpha_5$ hay dos Asparaginas N121 y N122 que forman una red de puentes de hidrógeno conservados en todos los cristales (Figura 45 N.E.F). Finalmente, el último nivel de interacción se da entre los carboxilos terminales de la hélice 6 (Figura 45 N.I.J). Es importante tomar en cuenta que todas estas interacciones son simétricas, para fines de visualización la representación esquemática solo presenta las interacciones del monómero A hacia el monómero C. Pero no hay que perder de vista que las interacciones del monómero C al monómero A (que serían las interacciones simétricas) también están presentes en la interfaz.

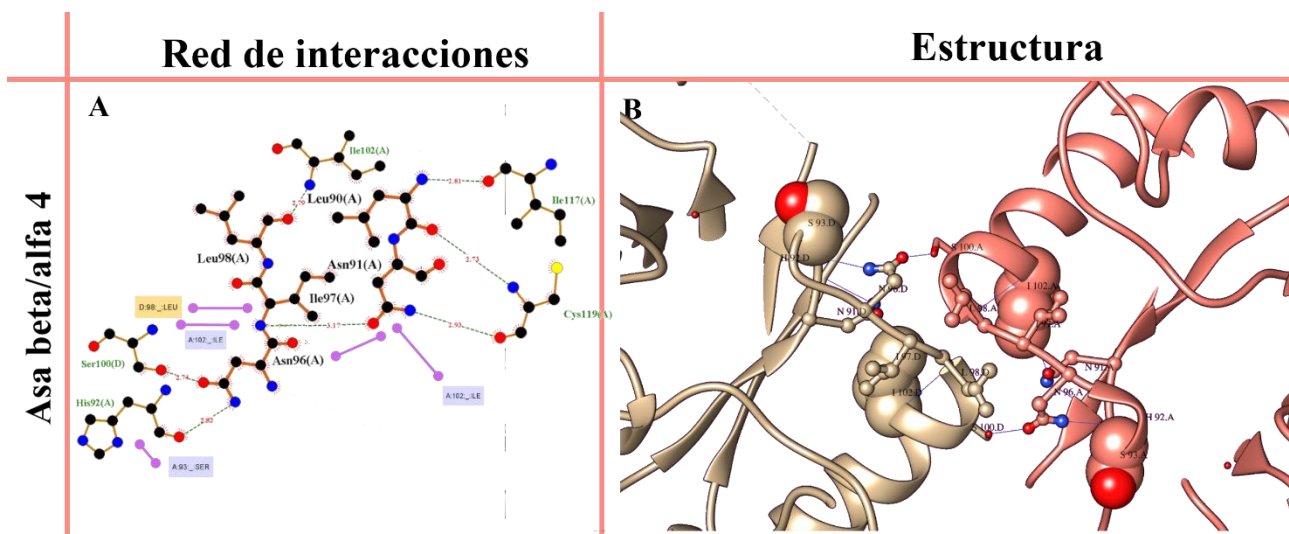


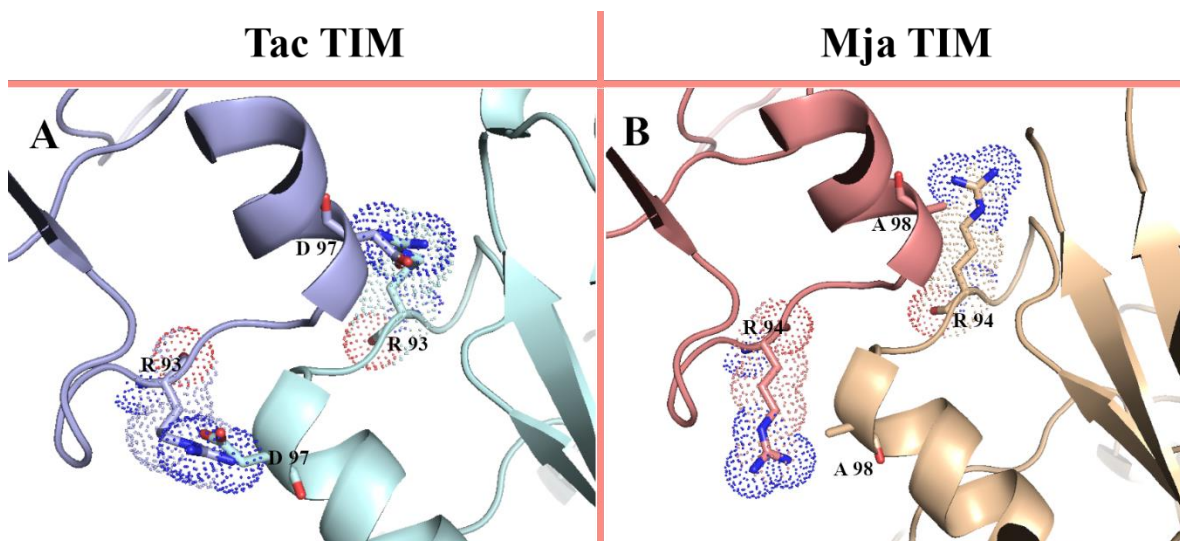
Figura 45. Interacciones de la interfaz tetramérica. La columna izquierda contiene los esquemas de la red de interacciones, con líneas punteadas verdes se representan los puentes de hidrógeno y con líneas moradas las interacciones de Van der Waals. En la columna derecha se representan estructuralmente las interacciones, con bastones los aminoácidos involucrados en la interfaz, con esferas los que interaccionan con fuerzas de Van der Waals y con líneas azules se muestran los puentes de hidrógeno.

COMPARACIÓN ENTRE LOS TETRÁMEROS Y EL DÍMERO

En la búsqueda de determinar las características que les permiten a las ArqTIMs tetramerizar, nuestro siguiente acercamiento fue comparar la secuencia y estructura del único cristal ArqTIM que existe como dímero en solución, la TIM *Thermoplasma acidophilum* (TacTIM), contra las secuencias y estructuras de las ArqTIMs tetraméricas.

Para observar estructuralmente estos cambios realizamos una alineación monómero por monómero de TacTIM con MjaTIM. Encontramos algunos residuos clave que provocan un choque estérico impidiendo la formación del tetrámero. Para fines de lectura en el texto, los aminoácidos correspondientes a TacTIM aparecerán en azul, mientras que los pertenecientes a MjaTIM aparecerán en rojo.

En la (Figura 46.A) observamos que R93-TacTIM se re-orienta provocando un impedimento estérico con D97, mientras que la misma arginina en MjaTIM - R94 se empaqueta perfectamente con L96, A98 y D99 (Figura 46.B). En el panel C de la Figura 46 se muestra el impedimento estérico de R96-TacTIM con I99, cabe destacar que esta arginina corresponde a una leucina en MjaTIM que es una pieza clave en la formación del tetrámero, ya que L97-MjaTIM esconde el total de su área y tiene múltiples interacciones con otros aminoácidos pertenecientes a diferentes estructuras secundarias: N89(asa β α 4), I100(h4), T123(asa β α 5), V127(h5) (Figura 46.D); por lo que esta única substitución evita todo el empaquetamiento del tetrámero. Otra diferencia importante es que los tetrámeros tienen conservados alaninas y leucinas en la hélice 5, así que A126 & L130-MjaTIM se empaquetan bien con A126, A129 y L130 (Figura 46.F). En cambio, TacTIM tiene tirosinas en lugar de alaninas en estas posiciones, lo que provoca que Y125 & Y129 causen un claro choque estérico con Y125, Y129 y E101 del otro monómero (Figura 46.E). Finalmente, encontramos la deleción de los dos últimos aminoácidos del carboxilo terminal de la hélice 6, esto es importante, porque provoca que la hélice 6 de TacTIM tenga un giro menos, perdiendo así la última superficie de interacción (Figura 46.G) presente en los tetrámeros (Figura 46.H)



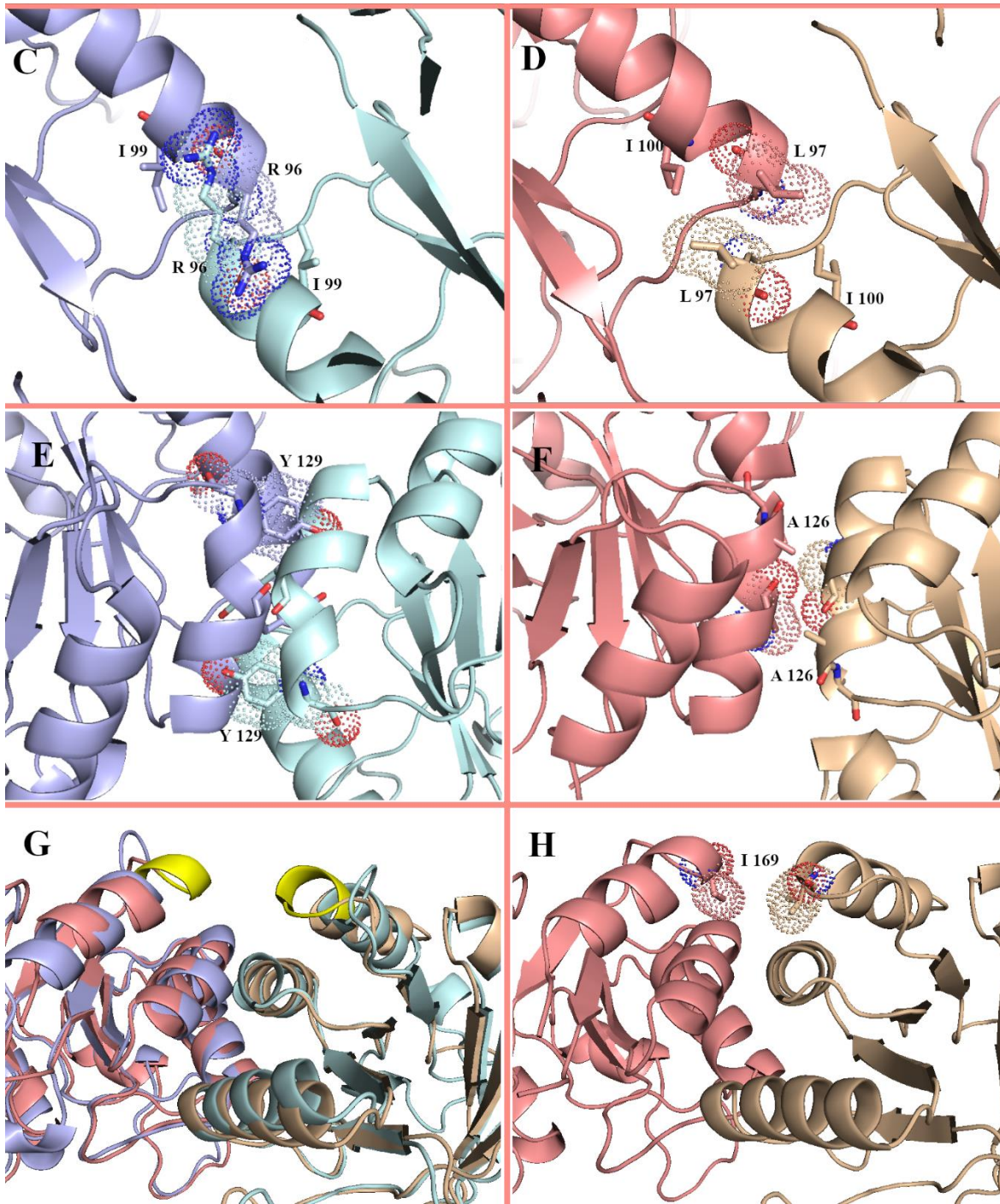


Figura 46. Diferencias entre TacTIM y MjaTIM. A) TacTIM-R93, B) MjaTIM-R94, C) TacTIM-R96, D) MjaTIM-L97, E) TacTIM-Y129, F) MjaTIM-A126, G) TacTIM-C-t hélice 6, B) MjaTIM-I169

A nivel secuencia las principales diferencias que encontramos fueron dos, la primera es que la hélice 5 de TacTIM tiene una secuencia de residuos aromáticos y aminoácidos cargados tanto positiva como negativamente (KYFREY) en lugar de la secuencia (AVAL) conservada por los tetrámeros en esta zona. La segunda diferencia significativa que encontramos fue una delección de los dos únicos

aminoácidos del carboxilo terminal de la hélice 6 que participan en la interfaz tetramérica, en los homotetrámeros (Figura 47).

El siguiente paso fue identificar si los residuos cargados y aromáticos en la hélice 5 y la delección del carboxilo terminal de la hélice 6 de TacTIM estaban presentes en secuencias de otros organismos. Si la hipótesis de oligomerización como una adaptación a la temperatura fuera cierta, esperaríamos que los cambios asociados al dímero TacTIM estuvieran presentes en secuencias de organismos mesófilos, mientras que las características de los tetrámeros deberían estar presentes en organismos termófilos e hipertermófilos. Sorprendentemente encontramos que las características encontradas en TacTIM si están presentes en otras secuencias, pero no de organismos mesófilos ni psicrófilos sino en organismos acidófilos (Figura 47), por lo que no concuerda con la hipótesis adaptativa a la temperatura, sino más bien parece ser una característica adquirida en función de ambientes con pH bajos, en donde viven los organismos acidófilos, o bien, pudo ser deriva génica de ese clado de los Thermoplasmatales en particular. Sin embargo, se necesitan experimentos que comprueben que dichas características son necesarias para resistir rangos bajos de pHs para considerarlo como una adaptación.

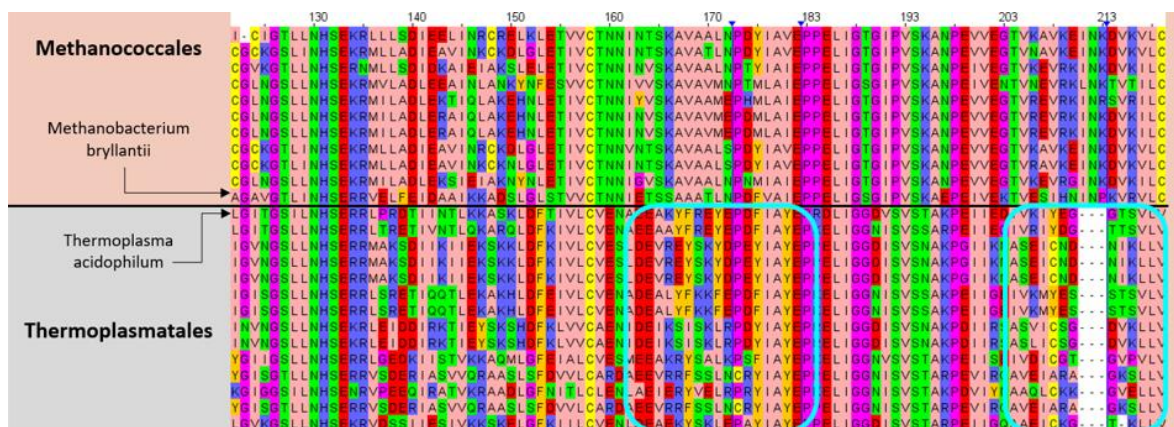


Figura 47. Diferencias en secuencia entre Methanococcales y Thermoplasmatales. En cuadros cian se denotan las diferencias en la hélice 5 y las dos delecciones del carboxilo terminal de la hélice 6.

En este caso, el estado oligomérico de TacTIM no se pudo correlacionar directamente con la adaptación a la temperatura, sino que resultó ser una característica asociada filogenéticamente al grupo de los Thermoplasmatales. Sin embargo, aún existen otros casos de ArqTIMs que son diméricas pero que no comparten las características en la secuencia de TacTIM, sino que parecen tener más características en común con las secuencias de los tetrámeros, como el caso de la TIM de Methanobacterium bryllantii (MbrTIM).

Así que todavía tenemos la pregunta sobre ¿Cuál es la relación entre el estado de oligomerización y la temperatura?

DIFERENCIAS ENTRE MjaTIM Y MfoTIM

Para responder esa pregunta tenemos otro ejemplo en el que existe una gran diferencia en termoestabilidad sin que necesariamente cambie el estado oligomérico. En los antecedentes mencionamos que nuestro grupo de investigación caracterizó fisicoquímicamente las ArqTIMs de *Methanocaldococcus jannashi* (MjaTIM) y *Methanotorrus formicicus* (MfoTIM). Se determinó mediante cromatografía de exclusión analítica que ambas proteínas eran homotetrámeros y mediante calorimetría diferencial de barrido se determinaron sus temperaturas medias de desnaturalización (T_m) de 111.6 °C para MjaTIM y de 91.7 °C para MfoTIM (Figura 25.C). La diferencia entre ambas T_m es muy grande, de casi 20°C, dado que ambas ArqTIMs son tetrámeros, el estado oligomérico no es responsable de dicha diferencia en la T_m , por lo que sospechamos que el origen debe ser estructural. Esta fue una de las principales razones que nos motivó a resolver la estructura de MfoTIM, además de que ambas proteínas tienen aproximadamente un 81% de identidad, de manera que es posible rastrear estructuralmente cada mutación.

Se ha visto que las proteínas termófilas tienden a tener mayor carga electrostática en su superficie. Por lo que nuestro primer acercamiento fue calcular la carga neta y las cargas electrostáticas en la superficie de ambas ArqTIMs por medio de los servidores PQR2PDB y APBS. Sorprendentemente, encontramos que la carga neta entre ambas proteínas difiere mucho para MjaTIM es de: -9 Coloumbs (C) y para MfoTIM es de: -1.6 C. En la (Figura 48) se muestra la comparación entre la superficie electrostática de ambas estructuras, en el primer panel se muestra la vista completa del tetrámero, en el segundo panel el dímero clásico de las TIMs y en el tercero un dímero formado por la interfaz tetramérica. En el segundo y tercer panel de la (Figura 48) observamos que la cara externa de MjaTIM tiene una superficie electrostática más negativamente cargada a comparación de MfoTIM, por lo que concluimos que la superficie que está en contacto con el solvente está más cargada negativamente en MjaTIM. Esto sin duda es un cambio estructural muy grande, lo que nos sugiere que los cambios electrostáticos de la superficie expuesta al solvente pueden ser parcialmente responsables de la gran diferencia en la T_m entre MjaTIM y MfoTIM. Ya que al aumentar la temperatura la velocidad y el movimiento del solvente aumentan, al tener una mejor distribución de cargas en la superficie, permite a las proteínas repeler mejor las cargas y el movimiento del solvente, por lo que es más difícil que penetre en la estructura, y por ende es más difícil que se desnaturalicen, esto les confiere estabilidad a las proteínas a altas temperaturas.

Como segunda aproximación comparamos las 42 mutaciones que difieren entre las secuencias y estructuras de MjaTIM y MfoTIM, con su red de interacciones. La red de interacciones la obtuvimos por medio de RING y su visualización se realizó con Cytoscape y Ligplot+. Las principales diferencias están en las sustituciones de los

residuos que participan en la interfaz tetramérica. Resulta que R94, M95, A98, E101, A102 y A109 de MjaTIM esconden mayor área al formar la interfaz (Tabla 10). Lo que aumenta el empaquetamiento de MjaTIM, una característica también observada en proteínas termoestables. Por lo que el aumento del empaquetamiento del tetrámero en MjaTIM también puede ser responsable del aumento en la termoestabilidad de MjaTIM.

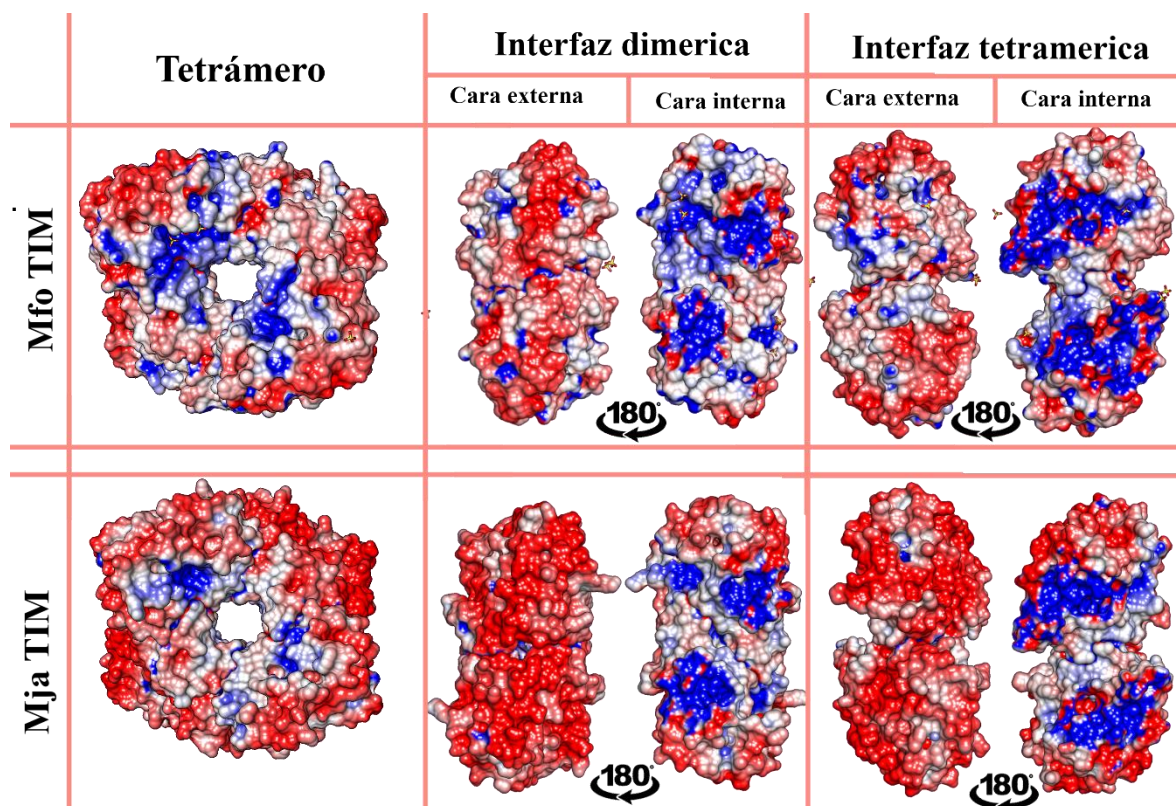


Figura 48. Diferencias entre la superficie electrostatica de MjaTIM y MfoTIM. Cargas negativas en rojo y cargas positivas en azul. Se muestra en la primer columna la vista aerea del tetrámero, en la segunda columna se encuentra el dímero clásico de la TIM, y en la tercer columna se muestra el dímero que se forma con la interfaz tetramérica.

Mfo TIM	BSA Å ²	Mja TIM	BSA Å ²
N 96	54.2	R 94	65.08
I 97	6.79	M 95	20.97
S 100	58.84	A 98	70.54
D 103	43.69	E101	68.15
A 109	Accesible a solvente	C 107	No-accesible a solvente

Tabla 10. Diferencias entre el área no accesible a solvente de los residuos que esconden mayor área en MjaTIM que en MfoTIM.

Las sustituciones de **R94 & M95** en **MjaTIM** provocan un aumento de las interacciones de Van der Waals con **I67, N89, H90, N119, S91 y A98** a comparación de la **N96 y I97** en **MfoTIM** que pierde dichas interacciones (Figura 49.A). Aunque **MfoTIM** gana un puente de hidrógeno entre la **S100** y la **N96** no es muy significativo para la tetramerización, pues recordemos que la interfaz se compone principalmente por interacciones hidrofóbicas, por lo que un solo puente de hidrógeno en esta zona no ayuda a aumentar la fuerza de la interacción tanto como la ganancia de todas las otras fuerzas de Van der Waals por parte de **MjaTIM**.

Otra sustitución que llamó nuestra atención fue la de **A98-MjaTIM** por la **S100-MfoTIM**, al igual que el caso anterior, **MjaTIM** aumenta las interacciones de Van der Waals pero ahora con residuos del monómero con quien forma la interfaz tetramérica **N119, R94 y I100**, permitiéndole esconder 12 \AA^2 más al formar el tetrámero a pesar de que la alanina sea más pequeña (Figura 49.B). Un caso similar sucede con la sustitución del **E101-MjaTIM** por **D103-MfoTIM** donde el glutamato gana interacciones de Van der Waals con el **N120** del otro monómero (Figura 49.C).

La siguiente pareja que llamó nuestra atención fue la sustitución de **K106 & C107** de **MjaTIM** contra **I108 & A109** de **MfoTIM**, en el caso de la **K106** gana dos interacciones de Van der Waals con **I64 & N65**, además de un puente de hidrógeno con la **N65**. Lo curioso con la sustitución de **C107-MjaTIM** es que esta no está accesible a solvente mientras que la **A109-MfoTIM** si es accesible a solvente a pesar de ser mas pequeña y tener naturaleza hidrofóbica (Figura 49.D). Por último, la sustitución de la **T123** en **MjaTIM** mejora la red de puentes de hidrógeno en esa zona, a comparación de la **V125** de **MfoTIM** que no tiene forma de realizar puentes de hidrógeno con su cadena lateral (Figura 49.E).

Todo esto nos indica una mejora en la red de interacciones en la interfaz tetramérica de **MjaTIM**, dichos cambios parecen aumentar el empaquetamiento y la fuerza de la interacción tetramérica en **MjaTIM**. Sin embargo, se requieren experimentos de mutagénesis dirigida para comprobar si existe una relación entre la mejora de interacciones en la interfaz tetramérica y el aumento en la termoestabilidad.

Para concluir encontramos que la principal diferencia entre **MjaTIM** y **MfoTIM** responsable del cambio en la termoestabilidad puede ser la distribución de cargas electrostáticas en su superficie accesible a solvente (Brininger, Cobani & Evilia 2018). La segunda diferencia que encontramos fueron las sustituciones de residuos que mejoran el empaquetamiento y la red de interacciones de la interfaz tetramérica. Aunque en este último caso son necesarios experimentos de mutaciones puntuales para determinar si realmente existe una correlación entre dichas sustituciones y el aumento de la termoestabilidad

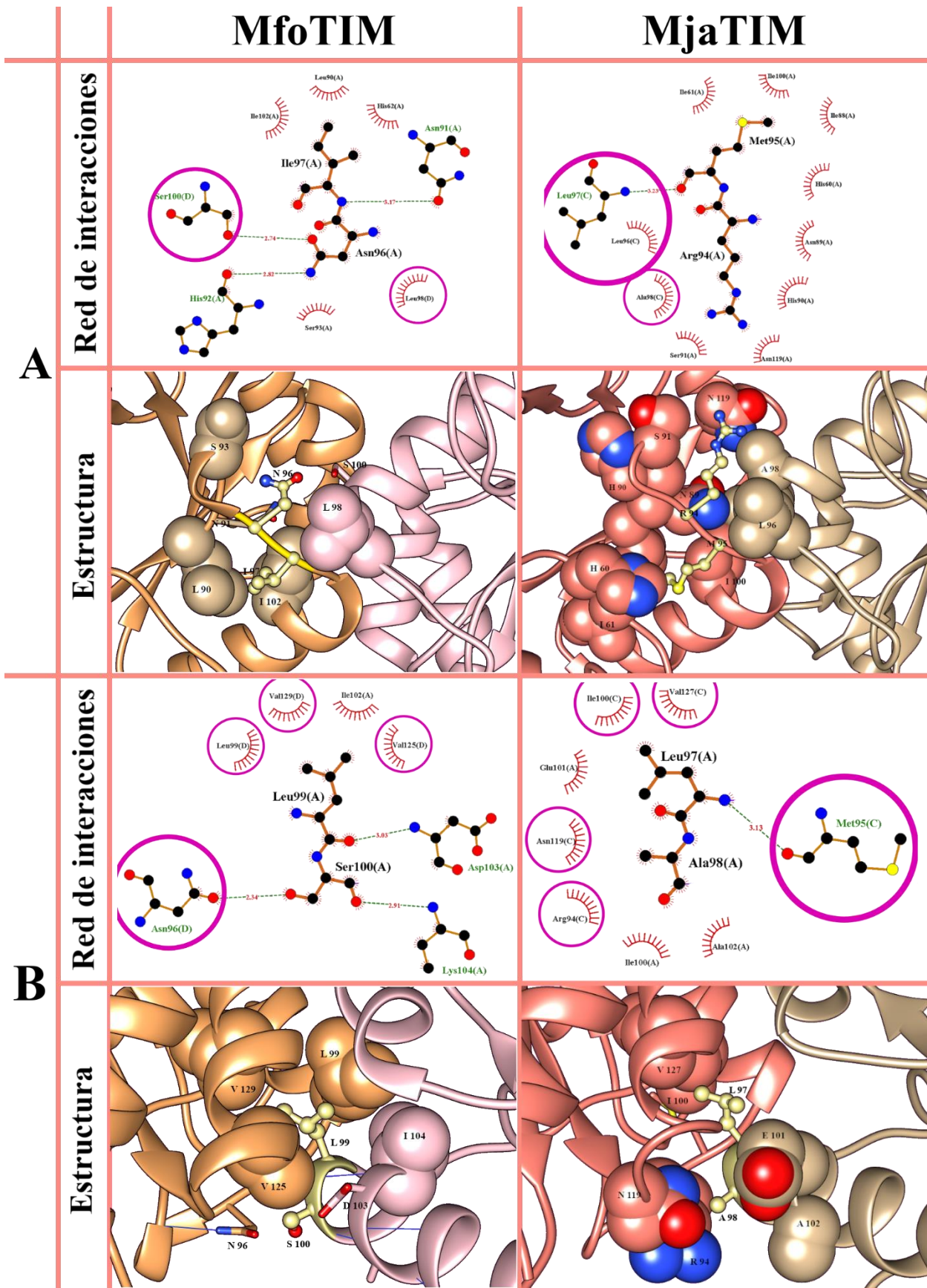


Figura 49. Diferencias entre los residuos de las interfaces de MjaTIM y MfoTIM

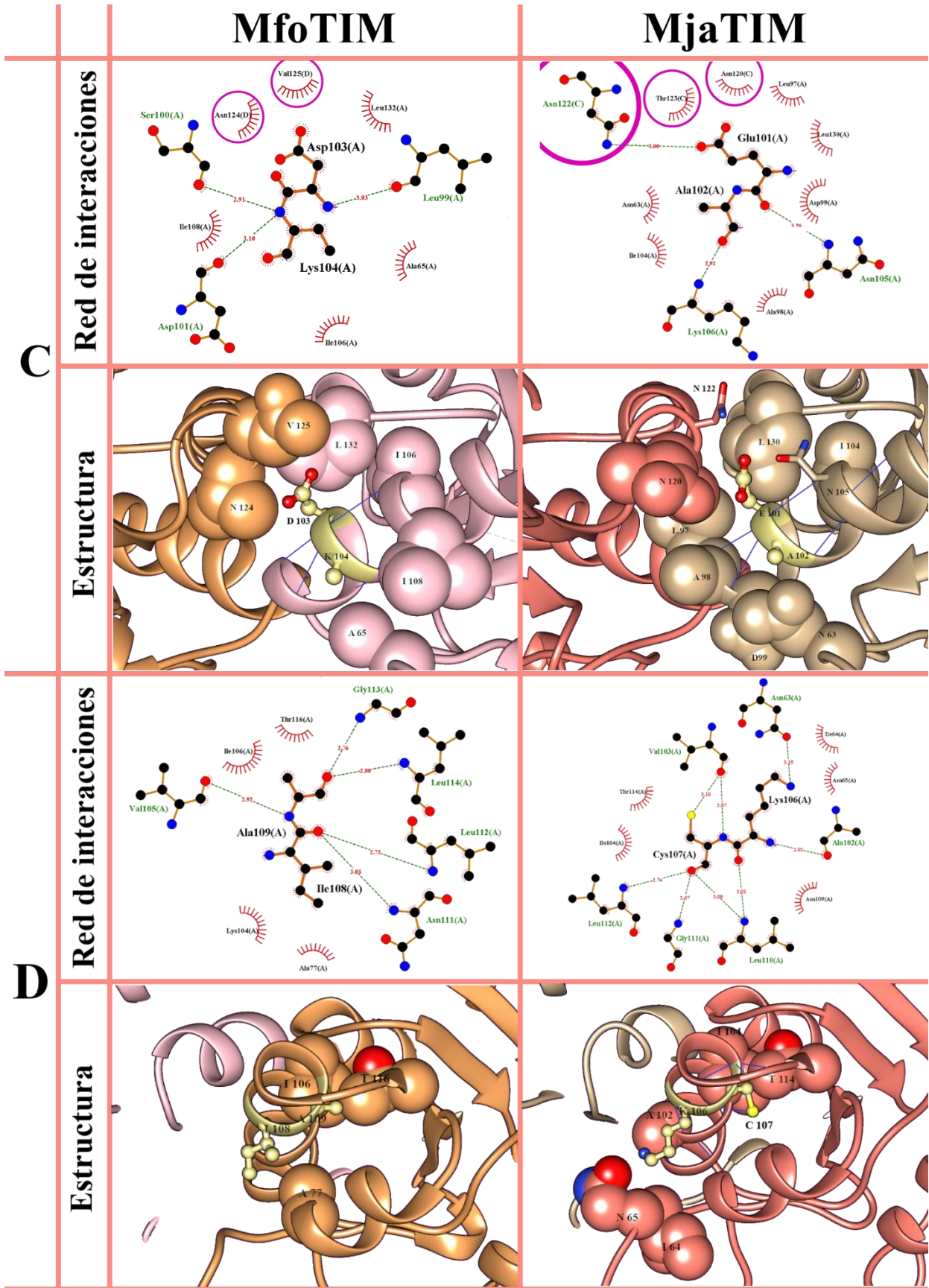


Figura 49. Diferencias entre los residuos de las interfaces de MjaTIM y MfoTIM

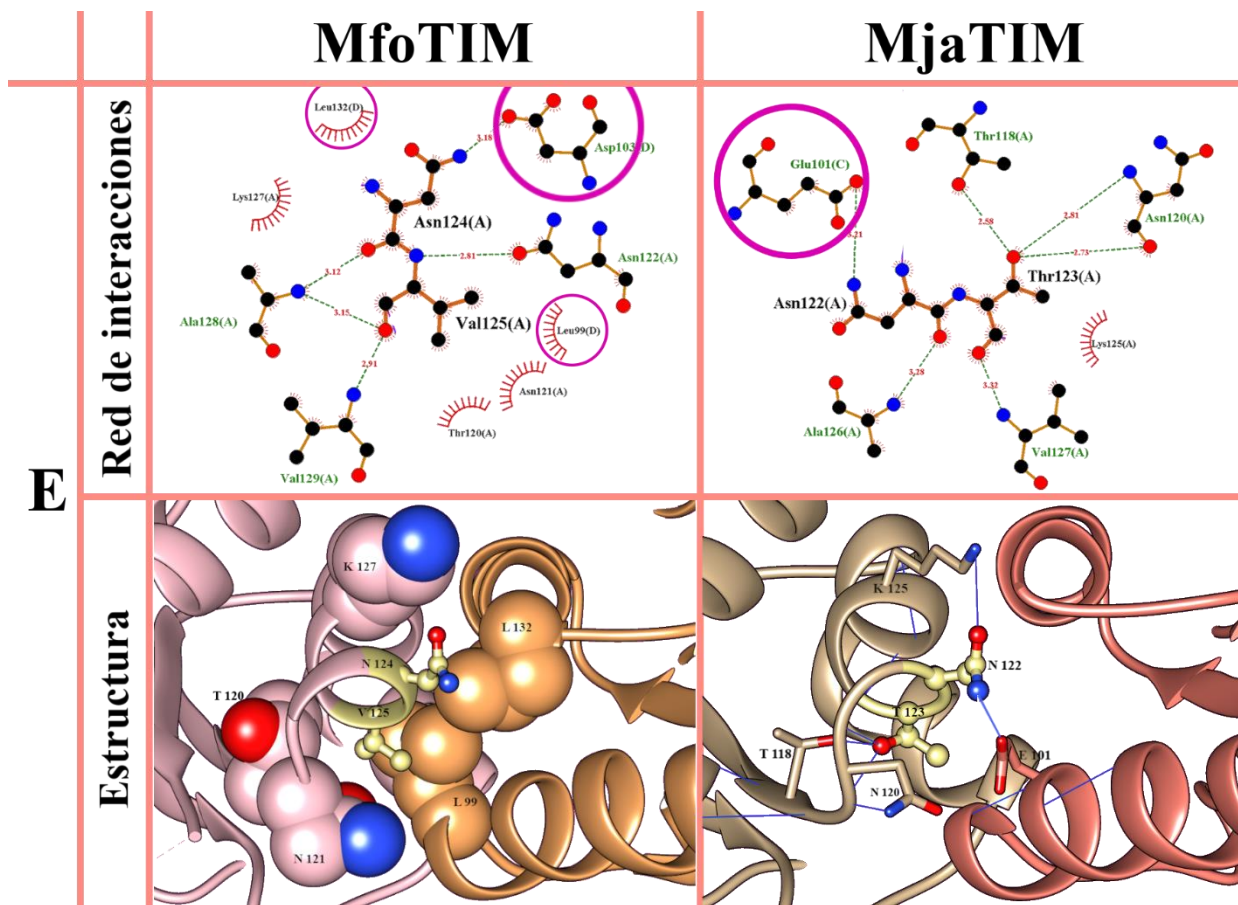
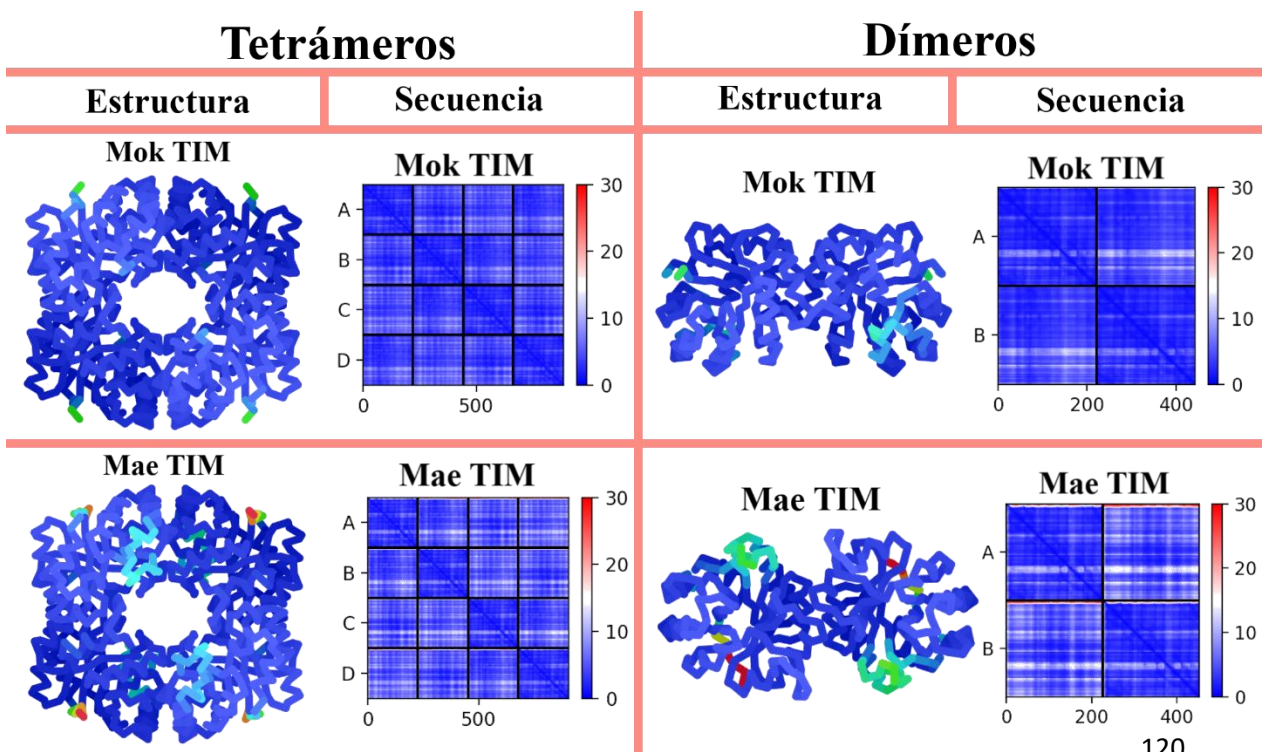


Figura 49. Diferencias entre los residuos de las interfaces de MjaTIM y MfoTIM. En la primer columna se encuentran los residuos de MfoTIM, en la segunda columna se encuentran los residuos de MjaTIM. En los esquemas de la red de interacciones, se representan los puentes de hidrógeno con líneas punteadas verdes, con semilunas rojas las interacciones de Van der Waals, y en círculos morados las interacciones con el otro monómero. En la representación estructural se representan los aminoácidos que cambian entre MfoTIM y MjaTIM en amarillo pálido y con bastones, con esferas se muestran los residuos que interactúan con fuerzas de Van der Waals y con líneas azules se muestran los puentes de hidrógeno. A) MfoTIM N96-I97 vs MjaTIM R94-M95, B) MfoTIM L99-S100 vs MjaTIM L97-A98, C) MfoTIM D103-K104 vs MjaTIM E101-A102, D) MfoTIM I108-A109 vs MjaTIM K106-C107, E) MfoTIM N124-V125 vs MjaTIM N122-T123.

DETERMINACIÓN DE MODELOS MEDIANTE ALPHAFOLD2

Debido a que no encontramos diferencias significativas entre las secuencias de TIMs reportadas como dímeros en solución MbrTIM y MbuTIM, decidimos hacer modelos tridimensionales con Alphafold2 para las ArqTIMs de las que se conoce su estado oligomérico en solución: MbrTIM, MbuTIM, MokTIM, MaeTIM, MmaTIM, TonTIM, MfeTIM. Con el fin de contar con más datos estructurales que nos permitan determinar las diferencias entre los dímeros y los tetrámeros. No utilizamos ninguna proteína como plantilla al generar los modelos en Alphafold2 con el fin de no sesgar los datos hacia un cierto estado oligomérico. Así TonTIM y MfeTIM se modelaron como tetrámeros y MaeTIM, MmaTIM y MokTIM se modelaron como dímeros y tetrámeros debido a que desconocemos su estado oligomérico en solución, en el caso de MbrTIM y MbuTIM se conoce que son dímeros en solución (Schramm A, Kohlhoff M, Hensel R. 2001) (Dhaunta N, et. al. 2013), pero al igual que las MesoTIMs se modelaron como dímeros y tetrámeros. AlphaFold genera un parámetro llamado pLDDT que mide la confianza de la predicción por residuo, en una escala de 0 a 100 donde el azul corresponde a los valores más altos mientras el rojo a los más bajos. La calidad de los modelos se muestra en la Figura 50, donde podemos observar que la mayoría de los modelos tienen altos grados de confianza, es decir, podemos confiar en que las coordenadas tridimensionales de los modelos muy probablemente corresponden con la estructura en solución de estas proteínas. La única excepción es el asa 6 en todos los dímeros y en algunos tetrámeros (MaeTIM, MbrTIM, MfeTIM y TonTIM). Esto no es de extrañar, ya que la estructura del asa 6 es difícil de observar debido a su función como tapa móvil en la catálisis de la enzima (Figura 17)



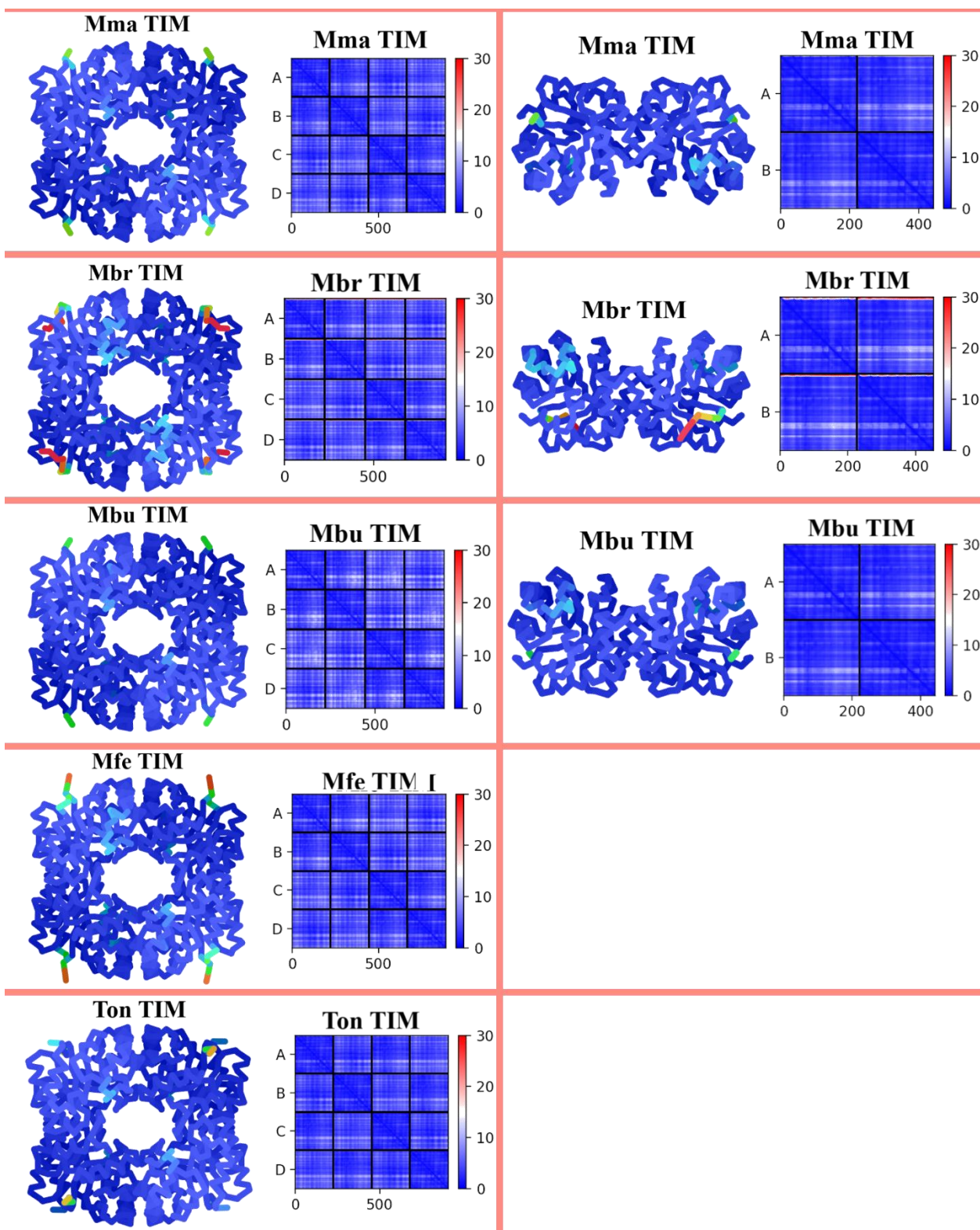


Figura 50. Calidad de los modelos mediante AlphaFold. Las estructuras y secuencias se colorearon en base al puntaje pLDDT que brinda AlphaFold2.

El primer acercamiento con estos modelos fue intentar determinar la calidad de las interfaces, para saber si la interacción tetramérica es estable o no, desafortunadamente no pudimos determinarlo a través del servidor de PISA como habíamos hecho anteriormente, ya que el servidor utiliza datos cristalográficos de los parámetros de la celda y el grupo de simetría para determinar posibles asociaciones. Para solucionar este problema y determinar si los tetrámeros son estables utilizamos el servidor HSYMDOCK que encuentra homoligómeros simétricos a partir de una estructura monomérica dada por el usuario.

Los monómeros de MfeTIM, TonTIM, MaeTIM, MmaTIM, MokTIM, MbuTIM y MbrTIM obtuvieron el puntaje más alto de HSYMDOCK en la asociación tetramérica. Como esperábamos para nuestro control positivo. MjaTIM, MfoTIM, PfuTIM y TtxTIM se asociaron como tetrámeros. Mientras que, nuestro control negativo TacTIM nunca formó tetrámeros. Este resultado sin duda es sorprendente, porque en la literatura MbrTIM y MbuTIM se reportan como dímeros en solución. En la (Tabla 11) se muestra el estado oligomérico predicho por HSYMDOCK y el puntaje asociado a cada modelo, además del rmsd de la alineación entre los modelos obtenidos mediante HSYMDOCK y los modelos generados por AlphaFold2.

El puntaje de HSYMDOCK se correlaciona linealmente con la afinidad de unión (Figura 34) y se le puede asignar un valor relativo del $\ln(K_D)$. En promedio el puntaje para los tetrámeros fue de: $\ln(-1252.3)$, que correspondería a un $\ln(K_D)$ calculado de aproximadamente -8. Mientras que el dímero TacTIM tuvo un puntaje de $\ln(-579.45)$ que correlaciona con un $\ln(K_D)$ de aproximadamente -5. Estos datos indican que la asociación tetramérica de las ArqTIMs puede ser estable. Además, todos los modelos de HSYMDOCK tienen desviaciones de rmsd menores a 1 Å, en relación con los modelos de Alphafold2. En base a estos dos argumentos decidimos continuar nuestro análisis de las ArqTIMs únicamente con los modelos tetraméricos.

Proteína	Estado oligomérico experimental	Estado Oligomérico predicho	HSYMDOCK SCORE	RSMD (HYSIMDOCK vs Alphafold2)	Topt °C	Área total Å ²	Volumen Å ³	Área Dímero Å ²	Área Tetrám Å ²
MjaTIM	Tetrámero	Tetrámero	-1505.6	0.76	85	30706	73550	1460.5	799.1
MfoTIM	Tetrámero	Tetrámero	-893.18	0.711	75	29621	70383	1464.9	722.5
PfuTIM	Tetrámero	Tetrámero	-1522.72	0.68	100	30121	72871	1389.1	732.2
TtxTIM	Dímero/ Tetrámero	Tetrámero	-1087.64	0.83	86	31286	72776	1480.6	944.5
MfeTIM	Tetrámero	Tetrámero	-1217	0.84	85	32367	78739	1377.1	731.4
TonTIM	Tetrámero	Tetrámero	-1147.3	0.75	80	31426	81208	1410.8	792.8
MaeTIM	-	Tetrámero	-1402.46	0.49	46	32393	80792	1483.2	808.1
MokTIM	-	Tetrámero	-1275.73	0.24	60	31016	78897	1468.9	765.5
MmaTIM	-	Tetrámero	-1359.17	0.26	38	30751	76519	1547.3	788.5
MbuTIM	Dímero	Tetrámero	-1189.18	0.35	23	29213	71097	1541.8	781
MbrTIM	Dímero	Tetrámero	-1175.36	0.19	37	17335	37141	1475.6	748.6
TacTIM	Dímero	Dímero	-579.45	0.32	60	20015	36891	1483.8	-

Tabla 11. Resultados de HSYMDOCK

El siguiente paso fue analizar el área y los residuos involucrados tanto en la interfaz dimérica como en la interfaz tetramérica (Tabla 11). Lo que observamos tanto en los modelos como en los cristales, es que el área total y el volumen casi no varían en los homotetrámeros. También que el área de la interfaz dimérica casi no varía entre todas las estructuras, siendo aproximadamente de $\sim 1463.5 \text{ \AA}^2$. A diferencia de los cristales observamos que el área de la interfaz tetramérica de los modelos no varía mucho, sino que es de aproximadamente de $\sim 787.1 \text{ \AA}^2$.

Se analizó si la interfaz tetramérica de los modelos, se asociaba de la misma manera que la interfaz tetramérica de las estructuras cristalográficas. Encontramos que, tanto los modelos como las estructuras determinadas experimentalmente, se asocian de la misma manera, por medio de tres niveles de interacción que involucran a las mismas 5 estructuras secundarias: el asa $\beta\alpha 4$, el N-t de la hélice 4, el asa $\beta\alpha 5$, la hélice 5 y el C-t de la hélice 6. En la (Tabla 12) se muestran los aminoácidos de los modelos involucrados en la interfaz tetramérica, podemos ver que siguen siendo en su mayoría aminoácidos hidrofóbicos y pequeños. También observamos que los residuos L99, I102, V125, A128 y V129 (tomando la secuencia de MmaTIM como referencia) conservan su naturaleza química en todas las secuencias, y esconden prácticamente el 100% de su área al formar la interfaz. Además, se conservan en todas las secuencias de los modelos, así como en los cristales las N91, N121 y N122. Así que podemos asegurar con mayor confianza que estos aminoácidos están bajo presión selectiva que ayuda a la formación del tetrámero. Cabe destacar que la secuencia con más diferencias en la zona de la tetramerización es la de MbrTIM, que es un dímero en solución.

	Mae TIM	Mfe TIM	Mbu TIM	MmaTIM	Mok TIM	Ton TIM	Mbr TIM	BSA %
asa $\beta\alpha 4$	ASN 96	ASN 92	ASN 91	ASN 91	ASN 91	ASN 95	ASN 96	48.1
	ARG 101	ARG 97	ARG 96	ARG 96	ARG 96	ARG 100	ARG 101	43.3
hélice 4	MET 102	MET 98	LEU 97	MET 97	MET 97	MET 101	VAL 102	56.2
	VAL 103	LYS 99	LYS 98	ILE 98	ILE 98	ILE 102	GLU 103	59.8
	LEU 104	LEU 100	LEU 99	LEU 99	LEU 99	LEU 103	LEU 104	99.5
	ALA 105	ALA 101	ALA 100	ALA 100	ALA 100	ALA 104	PHE 105	72.7
	ASP 106	ASP 102	ASP 101	ASP 101	ASP 101	ASP 105	GLU 106	28.33
	LEU 107	ILE 103	ILE 102	LEU 102	LEU 102	LEU 106	ILE 107	100
	GLU 108	SER 104	GLU 103	GLU 103	GLU 103	GLU 107	ASP 108	71.8
asa $\beta\alpha 5$	ASN 126	ASN 122	ASN 121	ASN 121	ASN 121	ASN 125	ASN 126	35.5
	ASN 127	ASN 123	ASN 122	ASN 122	ASN 122	ASN 126	ASN 127	27.44
hélice 5	VAL 129	GLU 125	ALA 124	SER 124	GLY 124	ALA 128	GLU 129	52.34
	VAL 130	THR 126	THR 125	VAL 125	VAL 125	VAL 129	THR 130	97.02
	LYS 132	ALA 128	ALA 127	LYS 127	LYS 127	ALA 131	SER 132	55.5

	ALA 133	ALA 129	ALA 128	ALA 128	ALA 128	ALA 132	ALA 133	99.5
	VAL 134	ALA 130	ALA 129	VAL 129	VAL 129	VAL 133	ALA 134	98.5
	VAL 136	ALA 132	ALA 131	ALA 131	ALA 131	ALA 135	THR 136	77.12
	MET 137	LEU 133	LEU 132	LEU 132	LEU 132	LEU 136	LEU 137	71.27
hélice 6	LEU 174	VAL 170	ILE 169	ILE 169	ILE 169	VAL 173	ILE 174	54.9

Tabla 12. Aminoácidos involucrados en la interfaz tetramérica de los tetrámeros modelos del AlphaFold2. En la primer columna se encuentra la estructura secundaria a la que pertenece cada aminoácido y en la última columna esta el promedio del porcentaje del área total que esconde cada residuo al formar la interfaz tetramérica. El color de los aminoácidos refleja su naturaleza química, en crema se encuentran los residuos hidrofóbicos, en verde los hidrofílicos, en azul los cargados positivamente, y en rojo los cargados negativamente.

Estos resultados nos hacen pensar que la asociación tetramérica es posible incluso para proteínas TIMs de arqueas mesófilas, debido a que conservan características estructurales simétricas, que les permiten homotetramerizar. Sin embargo, esta propuesta se tiene que analizar con detenimiento, pues los argumentos experimentales parecen contradictorios, en la sección de discusión el lector podrá encontrar una interesante exposición de argumentos entre los resultados experimentales y los resultados *in silico* encontrados en esta tesis.

COMPARACION DE LAS SUPERFICIES ELECTROSTÁTICAS

Hasta este punto seguimos sin encontrar una relación directa entre la temperatura óptima de crecimiento del organismo con su estado oligomérico. Así que nuestro siguiente acercamiento para analizar la termoestabilidad de los modelos fue comparar sus superficies electrostáticas, como habíamos hecho anteriormente entre MjaTIM y MfoTIM. En la Figura 51 se muestran las superficies electrostáticas de las ArqTIMs hipertermófilas y en la Figura 52 están las superficies electrostáticas de las ArqTIMs termófilas, incluyendo a TacTIM, y finalmente en la Figura 53 se aprecian las superficies electrostáticas de las ArqTIMs mesófilas.

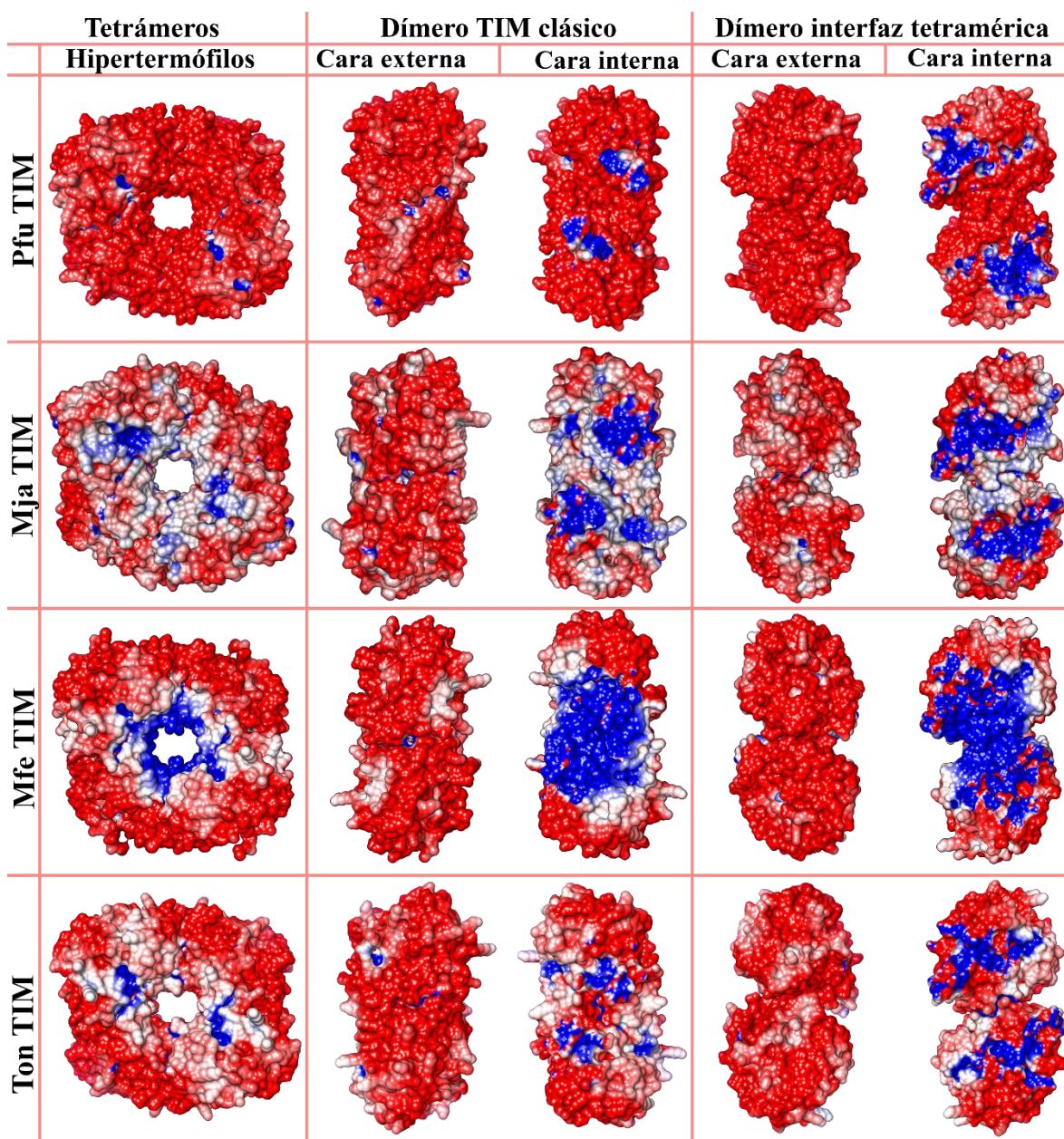


Figura 51. Superficies electrostáticas de las ArqTIMs Hipertermófilas. Las superficies están coloreadas por los siguientes valores: en rojo -10 C, en blanco 0 C y en azul 10 C.

Podemos observar en la (Figura 51) que las ArqTIMs hipertermófilas tienen superficies electrostáticas con alta densidad de carga negativa. En el caso de PfuTIM que es el organismo con la T_{opt} más grande (100°C) podemos ver que prácticamente toda su superficie es roja, aún las caras internas, que forman las interfaces tetramérica y dimerica, tiene la carga neta más negativa con -11.8 C. El valor de la carga neta nos indica el promedio de todas las cargas en la superficie. Para el caso de MjaTIM que crece a 85°C tiene una carga neta de -9 y podemos observar que su superficie es ligeramente menos negativa que PfuTIM, al menos en las caras internas tiene mayor cantidad de cargas neutras y positivas (Figura 51). En los siguientes casos que son MfeTIM y TonTIM con temperaturas óptimas de crecimiento de 85°C y 80°C respectivamente, y cargas netas de -9.4 C y -10.2 C respectivamente, igualmente se aprecia que la mayoría de su superficie está cargada negativamente.

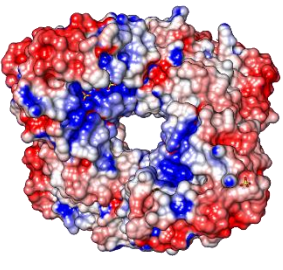
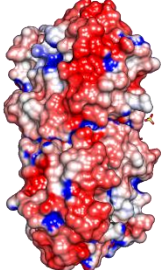
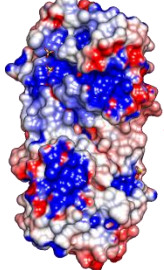
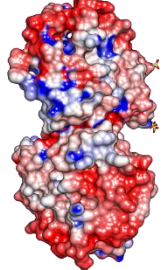
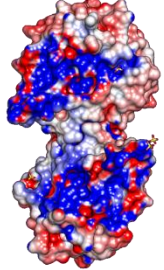
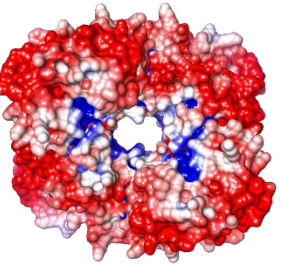
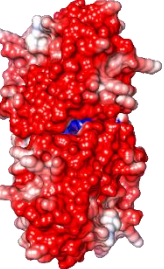
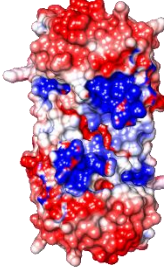
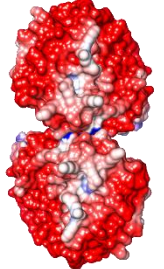
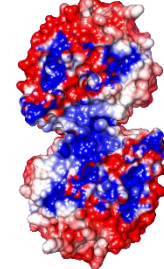
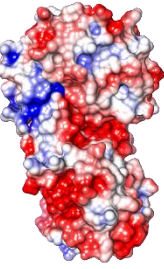
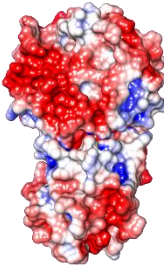
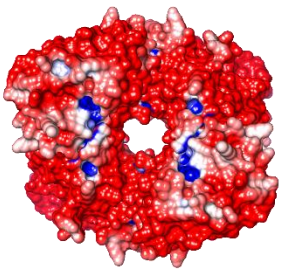
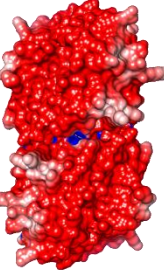
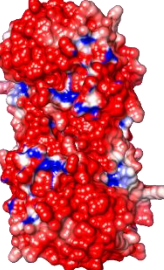
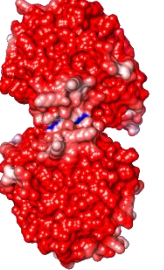
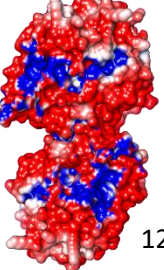
	Tetrámeros Termófilos	Dímero TIM clásico		Dímero interfaz tetramérica	
		Cara externa	Cara interna	Cara externa	Cara interna
Mfo TIM					
Mok TIM					
Tac TIM					
Mae TIM					

Figura 52. Superficies electrostaticas de las ArqTIMs Termófilas. Las superficies están coloreadas por los siguientes valores: en rojo -10 C, en blanco 0 C y en azul 10 C.

Entre las ArqTIMs Termófilas se encuentra TacTIM con una carga neta de -0.1 C aunque este caso es muy particular porque proviene de un organismo que vive a pH ácidos, y en ese medio la distribución de las cargas en la superficie cambia mucho por el intercambio de protones entre la superficie de la proteína y el solvente, así que, en este caso en particular, estos cálculos no son biológicamente significativos. Dentro de los termófilos también tenemos a MfoTIM con una carga neta de -1.6 C, que como vemos es un cambio muy grande respecto a las cargas netas de las ArqTIMs hipertermófilas (Figura 52) y que corresponde bien con el cambio en la termoestabilidad, pues el valor de su Tm es ~20°C menor que el valor de la Tm de MjaTIM. En el caso de MokTIM su carga neta es de -6.4 C mayor a MfoTIM pero menor a las ArqTIMs hipertermófilas, por lo que teorizamos que quizá sea un poco más termoestable que MfoTIM.

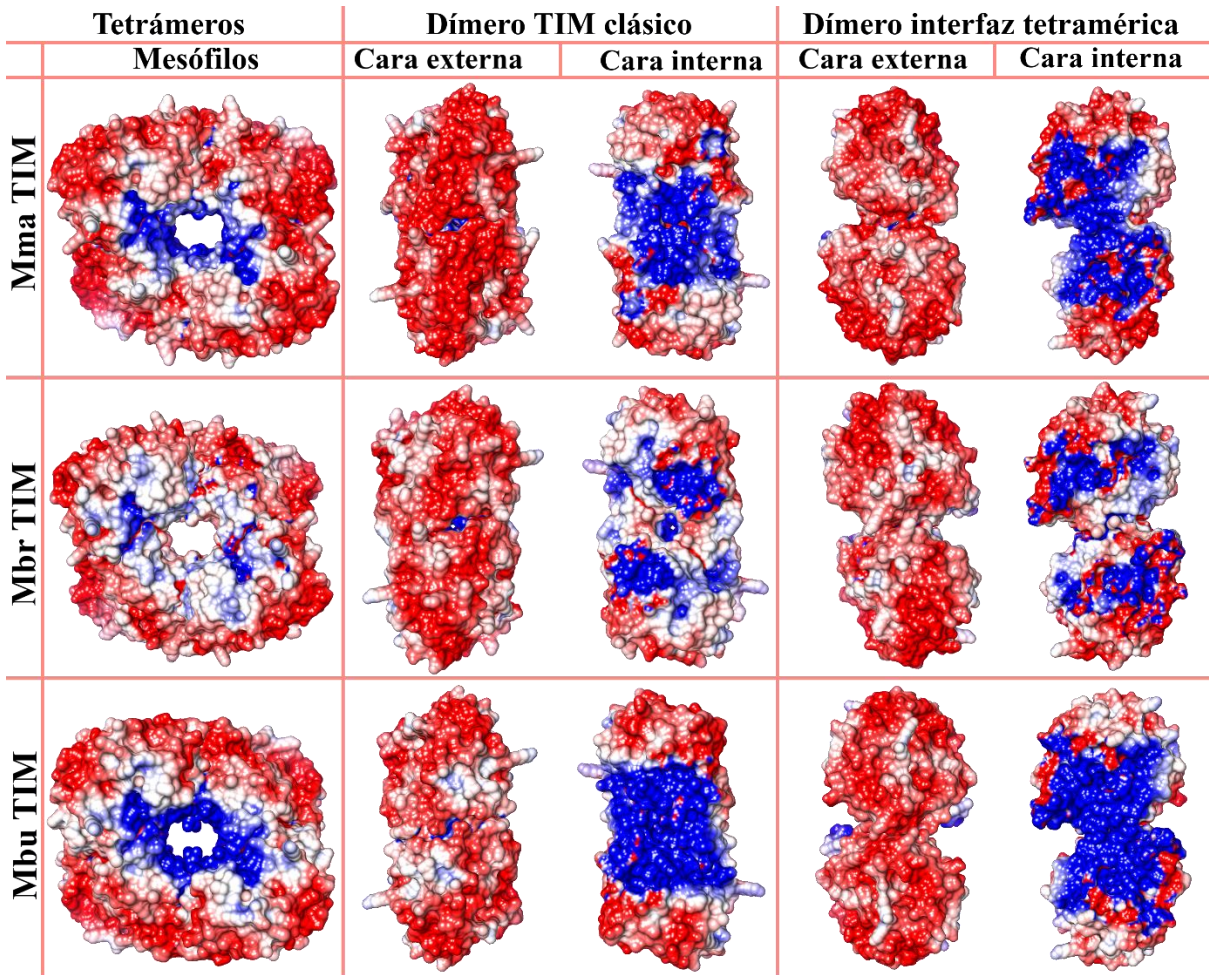


Figura 53. Superficies electrostaticas de las ArqTIMs Mesófilas. Las superficies están coloreadas por los siguientes valores: en rojo -10 C, en blanco 0 C y en azul 10 C.

El caso de MaeTIM es curioso, pues vive a una temperatura óptima de crecimiento de 46°C y tiene una carga neta de -10.6, similar a los valores de las ArqTIMs hipertermofilas. Como no conocemos los valores de Tm para MktTIM y MaeTIM no podríamos concluir si el valor de la carga neta se asocia directamente con la termoestabilidad, aunque que, debido a la distribución de cargas negativas en ambas proteínas, podrían tener valores de Tm elevados, sin embargo, necesitamos información experimental para corroborar o negar esta hipótesis.

La temperatura elevada ejerce una presión selectiva sobre los organismos, es decir, que los organismos que no son capaces de vivir a altas temperaturas se mueren. Cuando baja la temperatura del ambiente puede que los organismos conserven o no las características que les permitían vivir a altas temperaturas, porque ya no hay un factor externo que seleccione esas características. Así que la existencia de proteínas termoestables en organismos que viven a temperaturas óptimas de crecimiento más bajas es posible gracias a este fenómeno. Por lo que, no se puede inferir directamente la termoestabilidad de las proteínas de organismos psicrófilos o mesófilos, sino que se necesitan datos experimentales para conocer su Tm. La existencia de proteínas termoestables en organismos mesófilos es muy probable para el clado de los Methanococcales, si tomamos en cuenta que el ancestro del linaje era hipertermófilo (Lecocq M, et. al. 2020). Así tendría sentido que MbuTIM tenga una Tm de 74°C, relativamente elevada en comparación de su rango de crecimiento del organismo psicrófilo, que va de 1 a 28°C (Guptasarma, et. al. 2015). En la (Figura 53) se observa como las ArqTIMs mesófilas disminuyen la carga negativa electrostática en su superficie, MbrTIM tiene la carga neta más baja con -0.5 C, después sigue MbuTIM con una carga neta de -0.9 C y finalmente MmaTIM con una carga neta de -2.4 C, valores que son muy diferentes a los de las ArqTIMs termófilas e hipertermófilas.

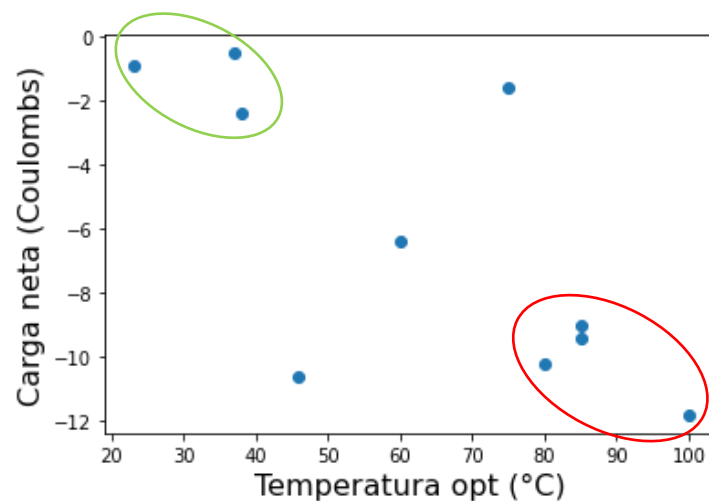


Figura 54. Gráfica de dispersión de el promedio de las cargas en la superficie de las ArqTIMs en relación a la Temperatura óptima de crecimiento. Las ArqTIMs hipertermófilas se muestran con un círculo rojo, mientras que las ArqTIMs mesófilas se señalan con un círculo verde.

En la (Figura 54) podemos observar que el valor de la carga neta se correlaciona bien con la temperatura de crecimiento óptima, vemos como todas las ArqTIMs hipertermófilas tienen valores muy negativos, mientras que las ArqTIMs mesófilas tienen valores de CN más positivos. Aunque dentro de la gráfica también se observan dos outliers, uno que corresponde a MfoTIM y a MaeTIM, que tienen cargas netas de -1.6 y -10.6, respectivamente. Sin embargo, es prometedor que la distribución de cargas electrostáticas en la superficie sea la característica adquirida en función de la temperatura óptima de crecimiento, ya que se ha visto en otras proteínas termófilas y porque aísla a la proteína del solvente a altas temperaturas previniendo su desnaturalización. Por lo que muy probablemente todas estas ArqTIMs tengan una termoestabilidad alta por dos razones, la primera es que descienden de un ancestro hipertermófilo y la segunda es que incluso para MbuTIM con uno de los valores más positivos para la carga neta, aún conserva una T_m de 74°C, relativamente elevada en relación con su rango de crecimiento. Sin embargo, para corroborar esta hipótesis es necesario determinar la T_m de todas las ArqTIMs aquí estudiadas.

COMPARACIÓN DE LAS SUPERFICIES HIDROFÓBICAS

Para explorar más a fondo el papel de la hidrofobicidad en el estado oligomérico de las ArqTIMs, coloreamos sus superficies mediante la escala de hidropatía de UCSF Chimera (Tabla 6), en donde los aminoácidos más hidrofóbicos se muestran en naranja mientras que los más hidrofílicos se muestran en azul. En la figura 55 se muestra la superficie hidropática de las ArqTIMs hipertermófilas, en la figura 56 la superficie hidropática de las ArqTIMs termófilas, y finalmente en la figura 57 se aprecian la superficie hidropática de las ArqTIMs mesófilas.

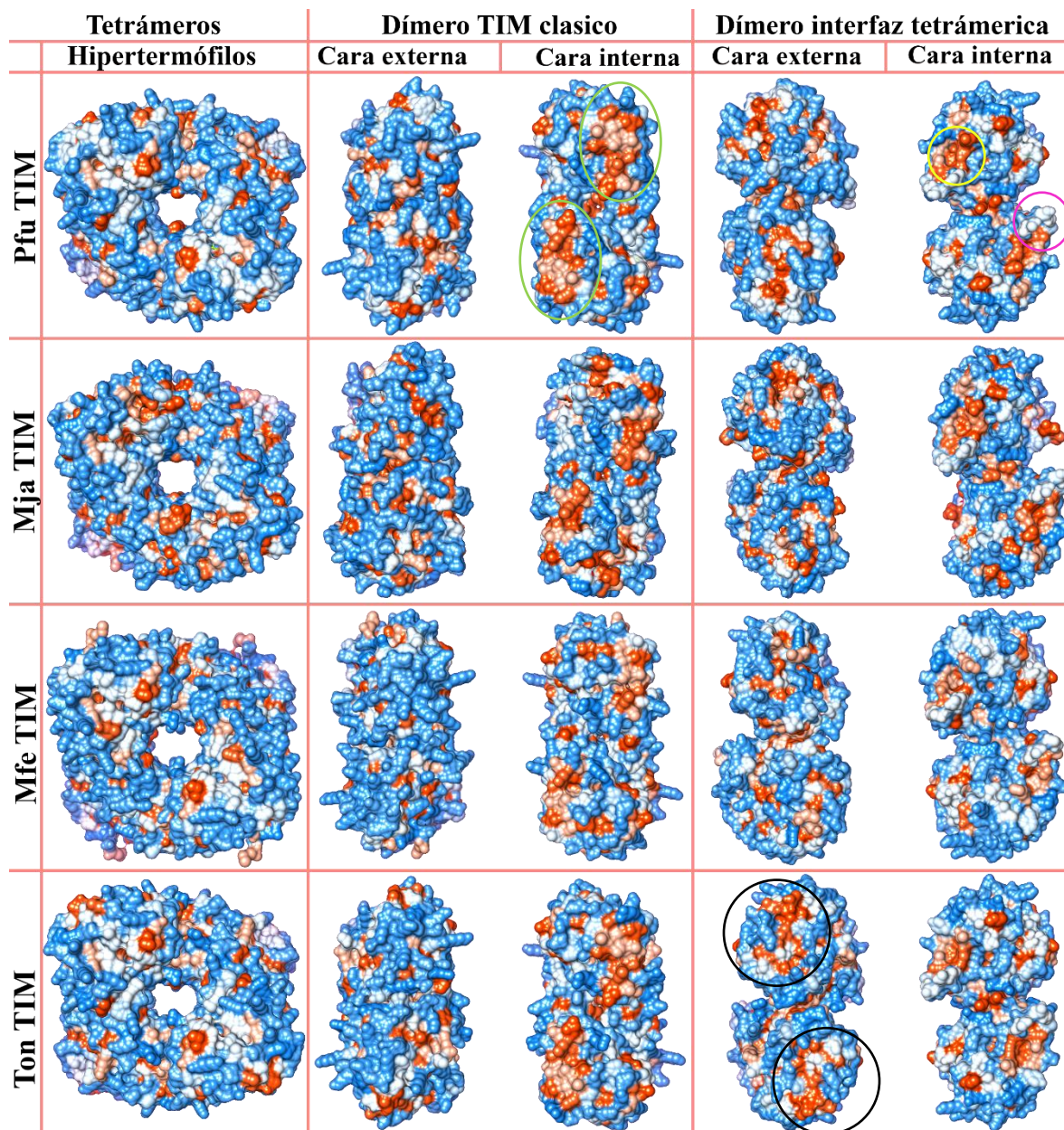


Figura 55. Superficies hidropáticas de las ArqTIMs Hipertermófilas.

Cuando comparamos las Figuras 55, 56 y 57 observamos que no existen cambios notables entre la distribución de la hidrofobicidad en la superficie y la temperatura óptima de crecimiento, por lo que analizaremos todas las ArqTIMs en conjunto. Para todas las ArqTIMs la mayoría de su superficie expuesta a solvente es hidrofílica, se puede apreciar bien en la vista del tetrámero y la vista de las caras externas. También en todos los casos, la vista de cara interna del dímero clásico muestra dos grandes parches hidrofóbicos que corresponden a la interfaz tetramérica, se marcan con círculos verdes en PfuTIM (Figura 55). Estos parches hidrofóbicos no se observan en TacTIM, porque carece del área de interacción de la interfaz tetramérica por su condición de dímero.

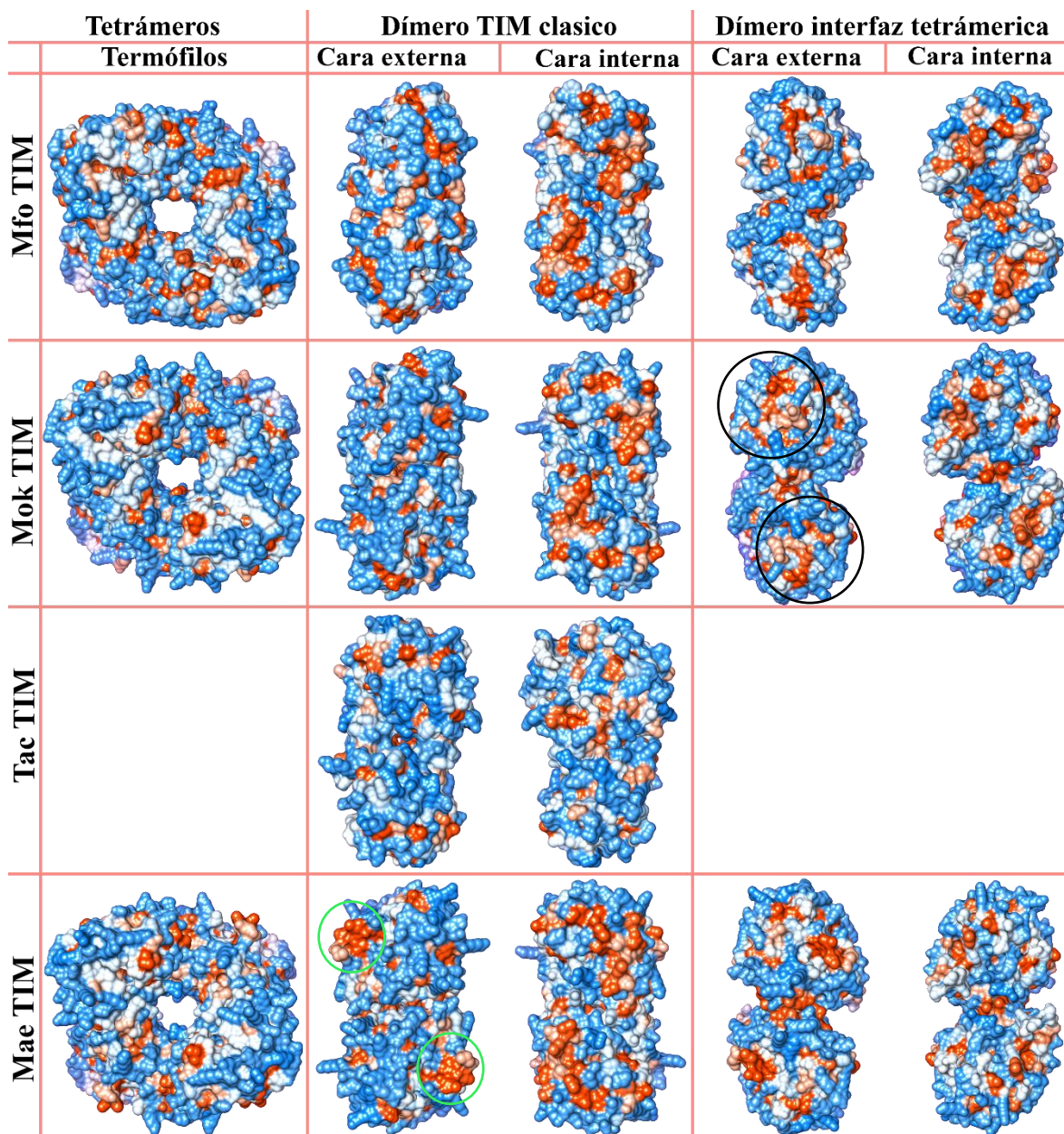


Figura 56. Superficies hidropáticas de las ArqTIMs Termófilas.

En el caso de la vista de la cara externa del dímero clásico observamos algunos puntos hidrofóbicos que se asoman en la superficie, sin embargo, éstos no se conservan en todas las ArqTIMs y no son lo suficientemente grandes para ser considerados una posible superficie de interacción. MaeTIM que muestra una gran zona hidrofóbica expuesta (marcada con círculos verdes en la Figura 55), al analizarla estructuralmente, encontramos que coincide con el amino terminal del modelo y AlphaFold lo marca con una baja confiabilidad (Figura 50), por lo que la exposición de esta zona hidrofóbica no es significativa.

En la vista de la cara interna del dímero tetramérico se puede ver la superficie correspondiente a la interfaz dimérica clásica, en PfuTIM el asa 3 se marca con un círculo magenta, mientras que el parche hidrofóbico donde encaja el asa 3 se marca con un círculo amarillo (Figura 55). Por último, en la vista de la cara externa del dímero tetramérico se aprecian dos superficies simétricas hidrofóbicas expuestas a solvente que se conservan en todas las ArqTIMs, se marcan con círculos negros en TonTIM, MokTIM y MbuTIM (Figuras: 55,56,57).

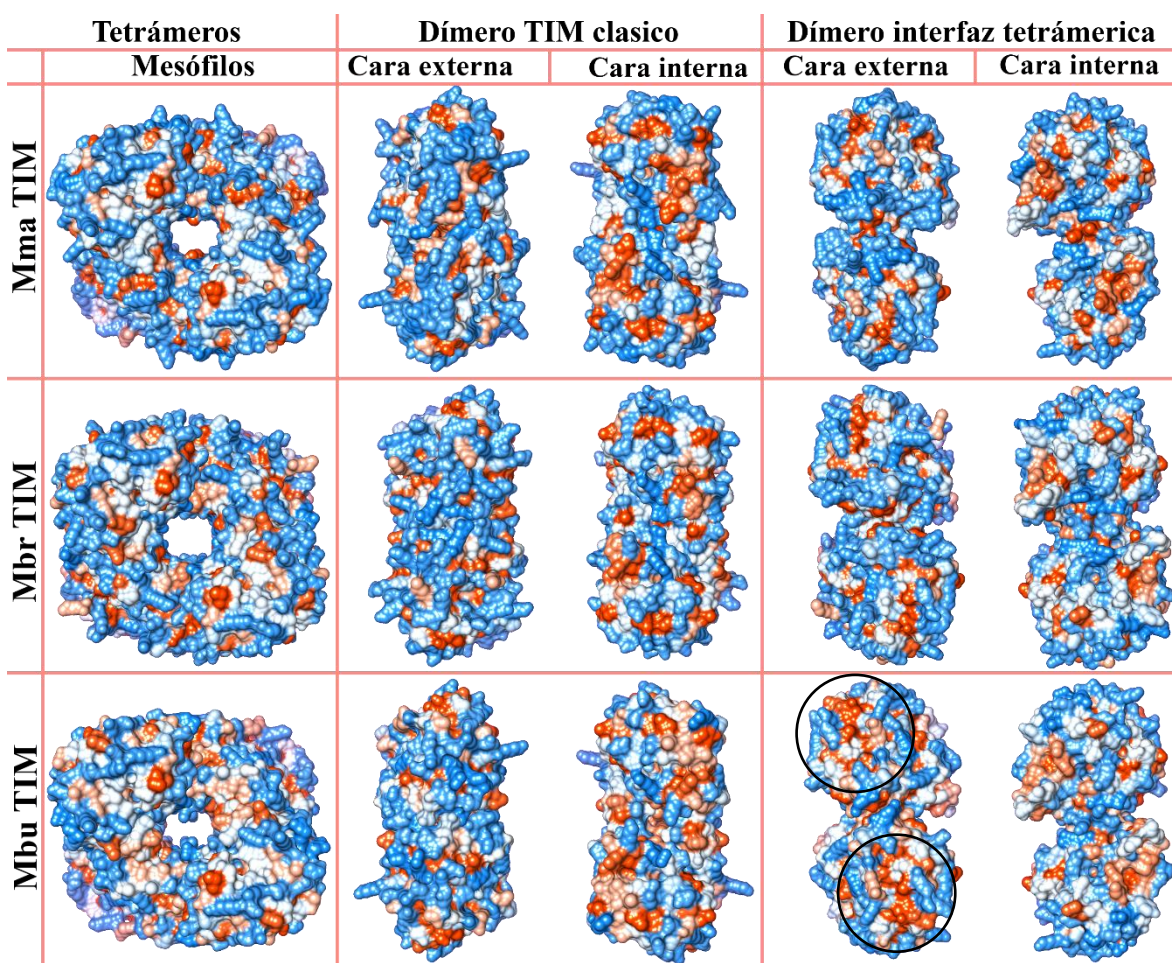


Figura 57. Superficies hidropáticas de las ArqTIMs Mesófilas.

Es raro que las proteínas tengan grandes áreas hidrofóbicas expuestas a solvente porque tienden a formar agregados, quizá esta sea una de las razones por las que las MesoTIMs no se pudieron purificar en el laboratorio. Debido a su simetría, conservación y exposición a solvente puede que estos parches hidrofóbicos sean biológicamente relevantes, pues dentro de las células las proteínas usualmente forman complejos moleculares simétricos con otras proteínas.

ANÁLISIS ACOPLADO ESTADÍSTICAMENTE (SCA)

Para determinar relaciones epistáticas y la coevolución entre los aminoácidos involucrados en la interfaz tetramérica, realizamos la metodología SCA con un alineamiento múltiple de secuencias de las ArqTIMs. También hicimos lo mismo con dos alineamientos múltiples de secuencias, uno del dominio bacteria y otro del dominio eukarya; con el fin de comparar si existen diferencias evolutivas con el dominio Arquea que nos puedan dar más pistas sobre la capacidad de tetramerización.

El primer resultado que analizaremos del SCA es la entropía relativa o valor D_i que nos indica que tan posible es la presencia de un aminoácido en esa posición, respecto a una base de proteínas no redundante (Figura 37). En la Figura 56 se muestran los histogramas de los valores D_i para cada posición de los alineamientos múltiples. En la Tabla 13 encontramos un resumen de los valores que obtuvieron el D_i más alto en sus respectivos alineamientos, las celdas en rosa visualizan los aminoácidos conservados en los tres alineamientos con valores D_i altos, que incluyen a los aminoácidos catalíticos (rosa más fuerte), así que sospechamos que todos estos residuos tienen una fuerte presión selectiva, probablemente relacionada con su papel en la catálisis, ya que se encuentran muy cerca de los aminoácidos catalíticos. En verde (Tabla 13) podemos ver los aminoácidos que únicamente tienen valores D_i altos en el alineamiento de las Arqueas, cabe destacar que la N95 está conservada en todas las ArqTIMs, este aminoácido se sustituye por una G93 que igualmente se conserva en el 100% de las secuencias de TIMs de bacterias y eucariontes. Claramente estos aminoácidos muestran una fuerte presión selectiva, por lo que suponemos que cumplen una función importante para el dominio Arquea, aunque desconocemos cuál es.

ARQUEA	VALOR D_i	BACTERIA	VALOR D_i	EUCARIA	VALOR D_i
N 12	2.89	N 8	2.72	N 11	2.43
		W 9	3.76	W 12	2.79
K 14	2.5	K 10	2.49	K 13	2.72
		M 11	3.13	M 14	2.55
Y 16	2.96				
Q 65	2.78			Q 64	2.8
H 66	2.84			N 65	2.54
N 95	2.73	G 93	2.43	G 94	2.44
H 96	3.53	H 94	3.54	H 95	3.53
E 98	2.55	E 96	2.59	E 97	2.58
		R 97	2.69	R 98	2.73
		R 98	2.7	R 99	2.75
C 123	3.39	C 125	3.43	C 126	3.43
		E 128	2.58	E 129	2.6
		Y 165	3.18	Y 164	3.19
E 144	2.59	E 166	2.59	E 165	2.6
P 145	2.48	P 167	2.75	P 166	2.75

P 146	2.54				
		W 169	3.98	W 168	4
I 149	2.67	I 171	2.69	I 170	2.73
		Y 208	3.19	Y 209	3.19
		F 242	2.62	F 240	2.56

Tabla 13. Aminoácidos con los valores Di más altos

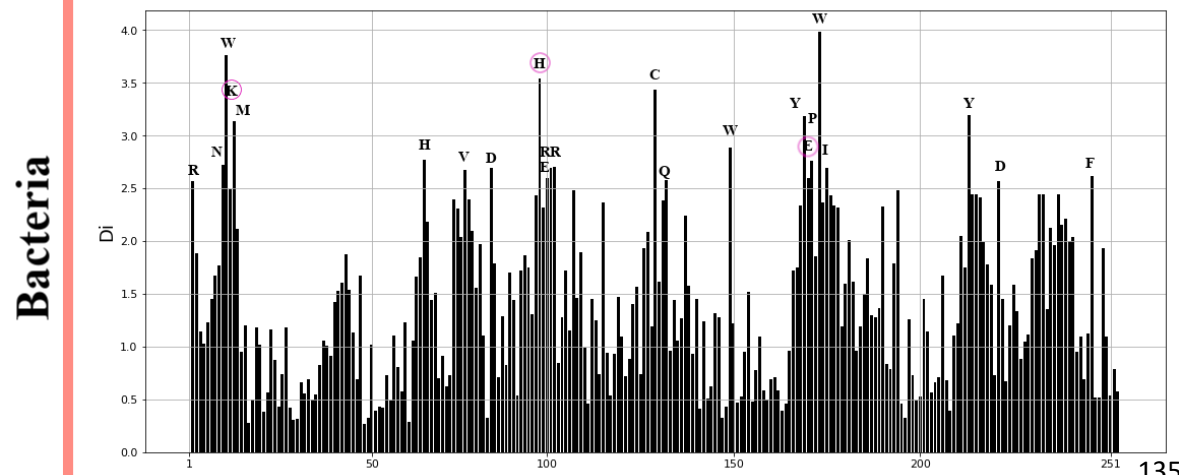
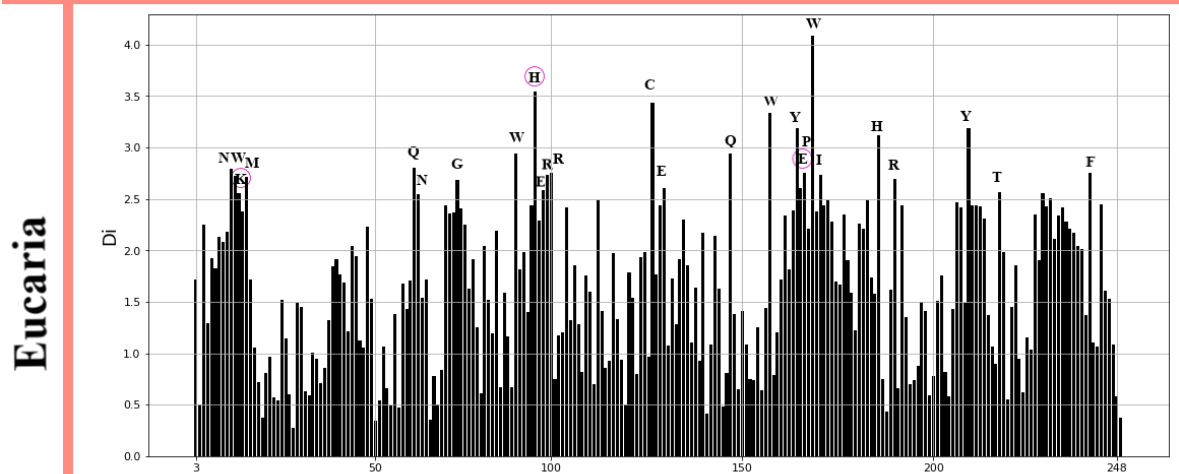
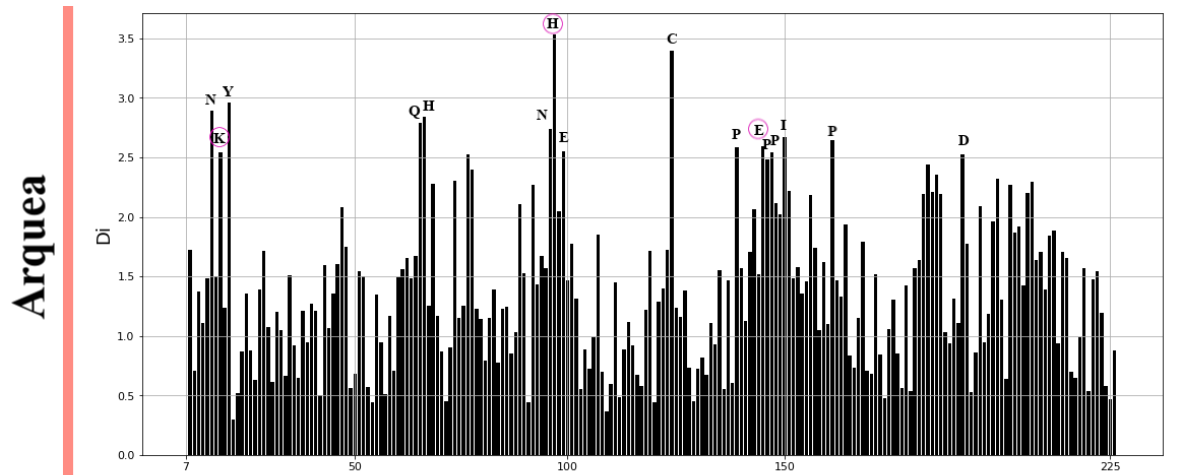


Figura 58. Histogramas D_i . En el eje "x" se muestra la secuencia de referencia para cada alineamiento, en el eje "y" el valor D_i asociado a cada posición. Las posiciones con los valores más altos se indican con el aminoácido correspondiente de la secuencia de referencia. Los aminoácidos catalíticos se marcan con un círculo magenta.

El segundo resultado de método es la generación de la matriz SCA para los 3 alineamientos múltiples (Figura 58). En especial observamos que la matriz de correlación SCA del dominio Eucaria parece tener más posiciones correlacionadas, se observa en la figura 58 con la presencia de mayor color verde y amarillo en la matriz. Esto se puede deber a que su alineamiento múltiple tiene menos secuencias. Aunque en este punto las matrices de todos los alineamientos tienen mucho ruido estadístico (Figura 59). Para quitar este ruido, se realiza un alineamiento aleatorio al que se le aplican los mismos cálculos, es decir, igualmente se hace una matriz de correlación y se encuentran sus eigenvalores y eigenvectores. Para finalmente restarlos a los datos originales y extraer el ruido estadístico.

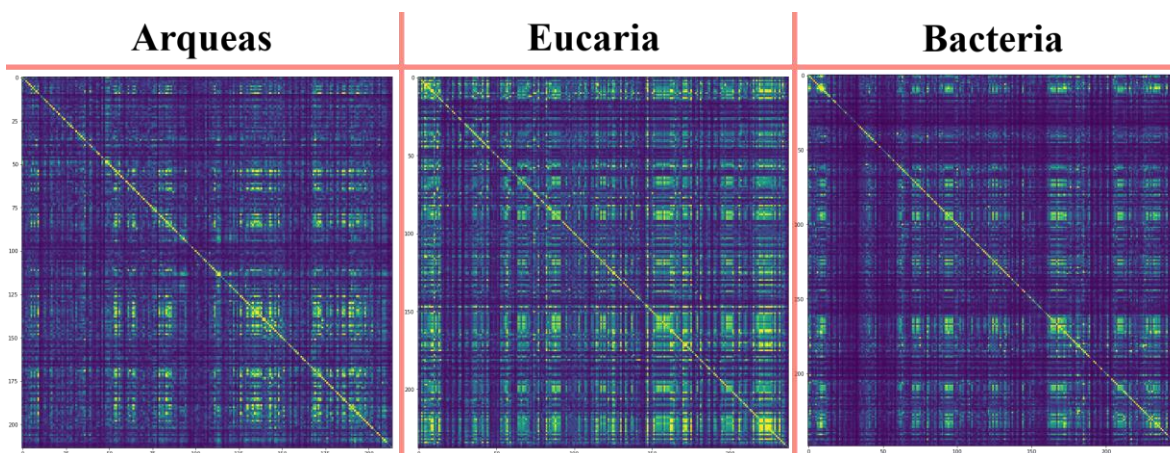


Figura 59. Matrices de correlación SCA de los tres dominios de la vida. La matriz tiene una dimensión $L \times L$, donde cada lado representa la secuencia completa del alineamiento, y cada punto representa una posición del alineamiento, el color verde y amarillo representa un valor de correlación alto entre esas posiciones y el color azul indica una falta de correlación entre ambas posiciones.

Esta resta se puede apreciar muy bien en la Figura 60, donde se muestran el eigenespectro de la matriz SCA, es decir, un histograma con los eigenvectores en el eje "y", y los eigenvalores en el eje "x". En barras negras se muestran los resultados de nuestros datos originales, mientras que con una línea roja se muestra los resultados del alineamiento múltiple. Así que todo lo que quede fuera de la línea roja son eigenvectores significativos. Para el caso de las arqueas hay 12

eigenvectores significativos (ES), para las bacterias hay 11 eigenvectores significativos, y finalmente las eucarias tienen 7 eigenvectores significativos.

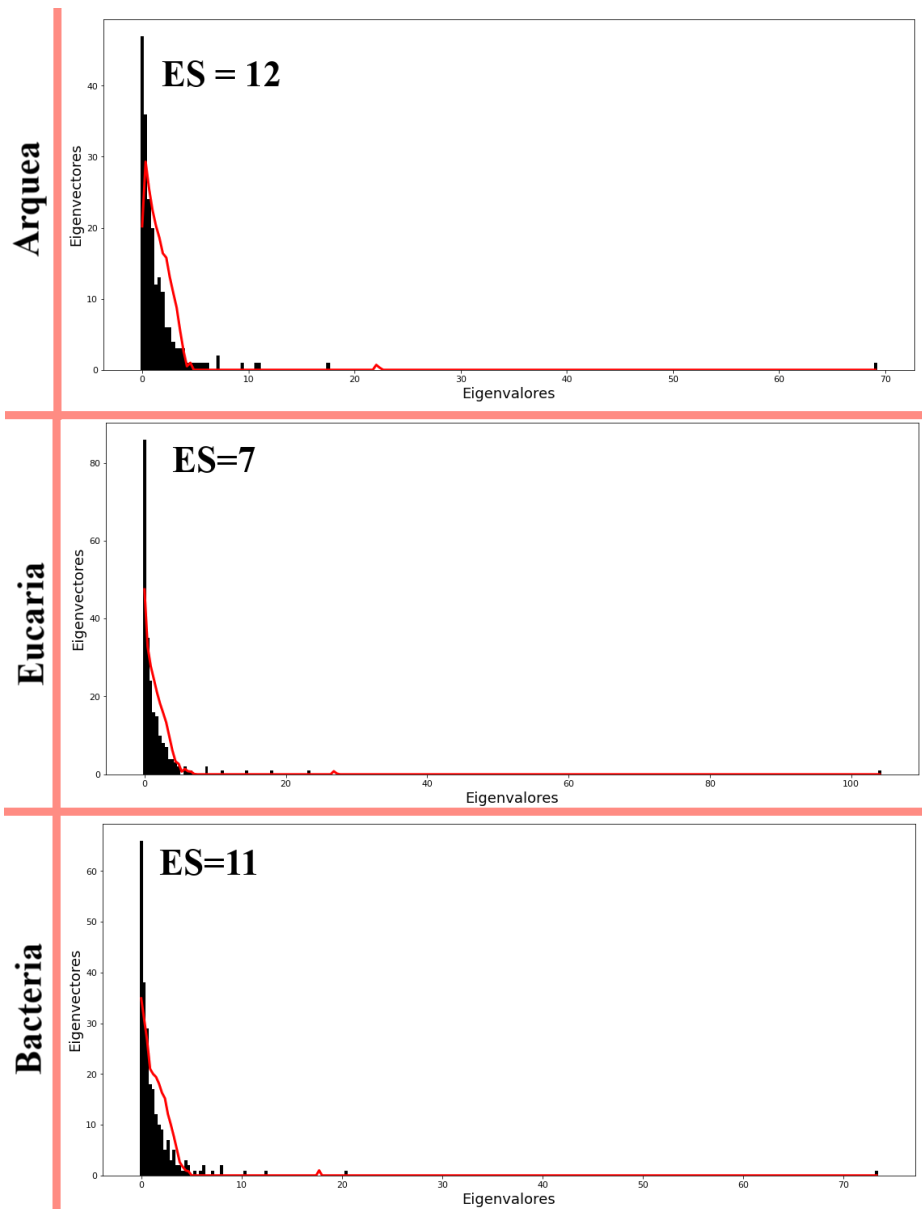


Figura 60. Eigenespectros de los tres dominios de la vida. En barras negras se muestran los eigenvectores de la matriz SCA, mientras que la línea roja representa los eigenvectores de los alineamientos múltiples.

Después de obtener los eigenvectores significativos se les aplica el análisis de componentes independientes (IC) para separar el contenido de los eigenvectores por pequeños grupos de elementos independientes (Figura 41). Para separar por completo estos grupos de elementos se ajustan todos los datos del vector IC a una

distribución normal (Figura 42), tomando únicamente el 3.5% ($p=0.965$) de los elementos más significativos de cada vector IC. Que equivalen a nuestros pares de aminoácidos coevolutivos.

Posteriormente, visualizamos este resultado en una matriz de correlación limpia (Figura 61) cada lado representa los residuos más significativos de cada vector IC, las líneas blancas dividen los resultados de los vectores IC, el color verde y amarillo representa un valor de correlación alto entre esos pares de residuos y el color azul en la matriz indica una falta de correlación entre ambos aminoácidos. Aquí debemos observar si los eigenvectores son independientes (que sean completamente azules) o si tienen algún tipo de correlación que los haga dependientes entre sí (que compartan color amarillo o verde). En caso de que los vectores IC sean dependientes entre sí se unen para formar un solo sector (Figura 61, columna "Sectores"); en caso de que sean independientes el resultado del vector IC forma un sector.

En el caso de las Arqueas, 3 eigenvectores resultaron dependientes entre sí por lo que se juntaron para formar el Sector 1, los demás eigenvectores fueron independientes entre sí, por lo que en total obtuvimos 10 Sectores. En el dominio Eucaria todos los eigenvectores muestran dependencia entre todos, ocasionado probablemente porque el alineamiento múltiple tiene un número menor de secuencias en el alineamiento múltiple. En este caso agrupamos únicamente 2 eigenvectores para formar el Sector 1, los sectores con menor dependencia los dejamos aparte, obteniendo al final 6 Sectores. Finalmente, el dominio Bacteria únicamente tuvo 2 eigenvectores claramente dependientes, que se juntaron para formar el Sector 1, por lo que en total resultaron 10 Sectores.

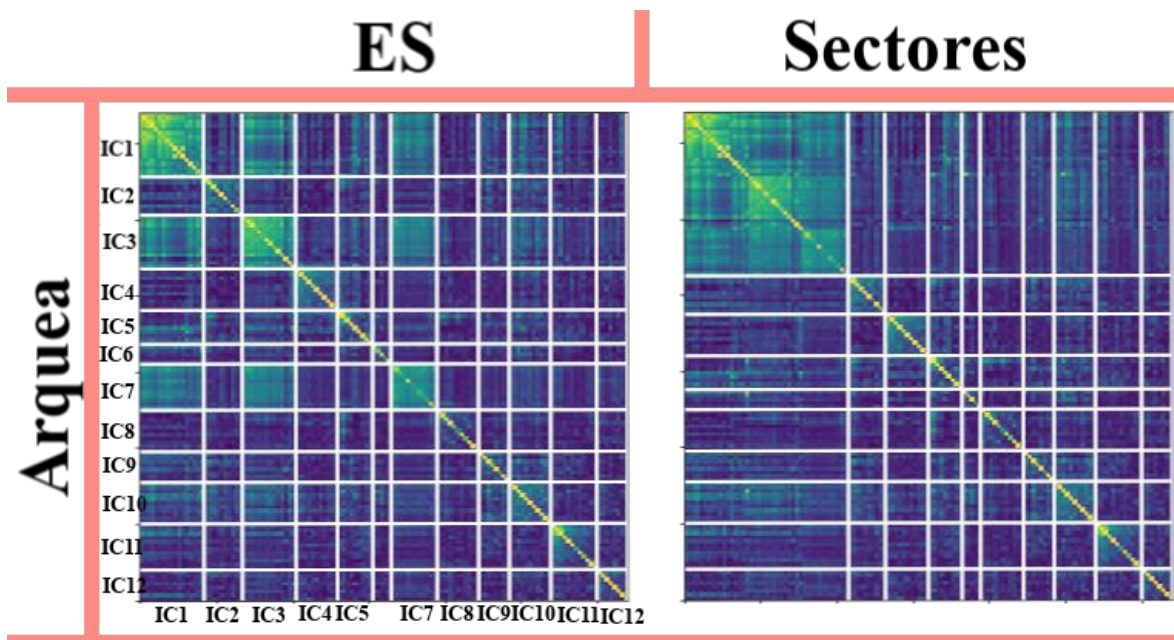


Figura 61. Matrices de correlación limpias y determinación de los Sectores.

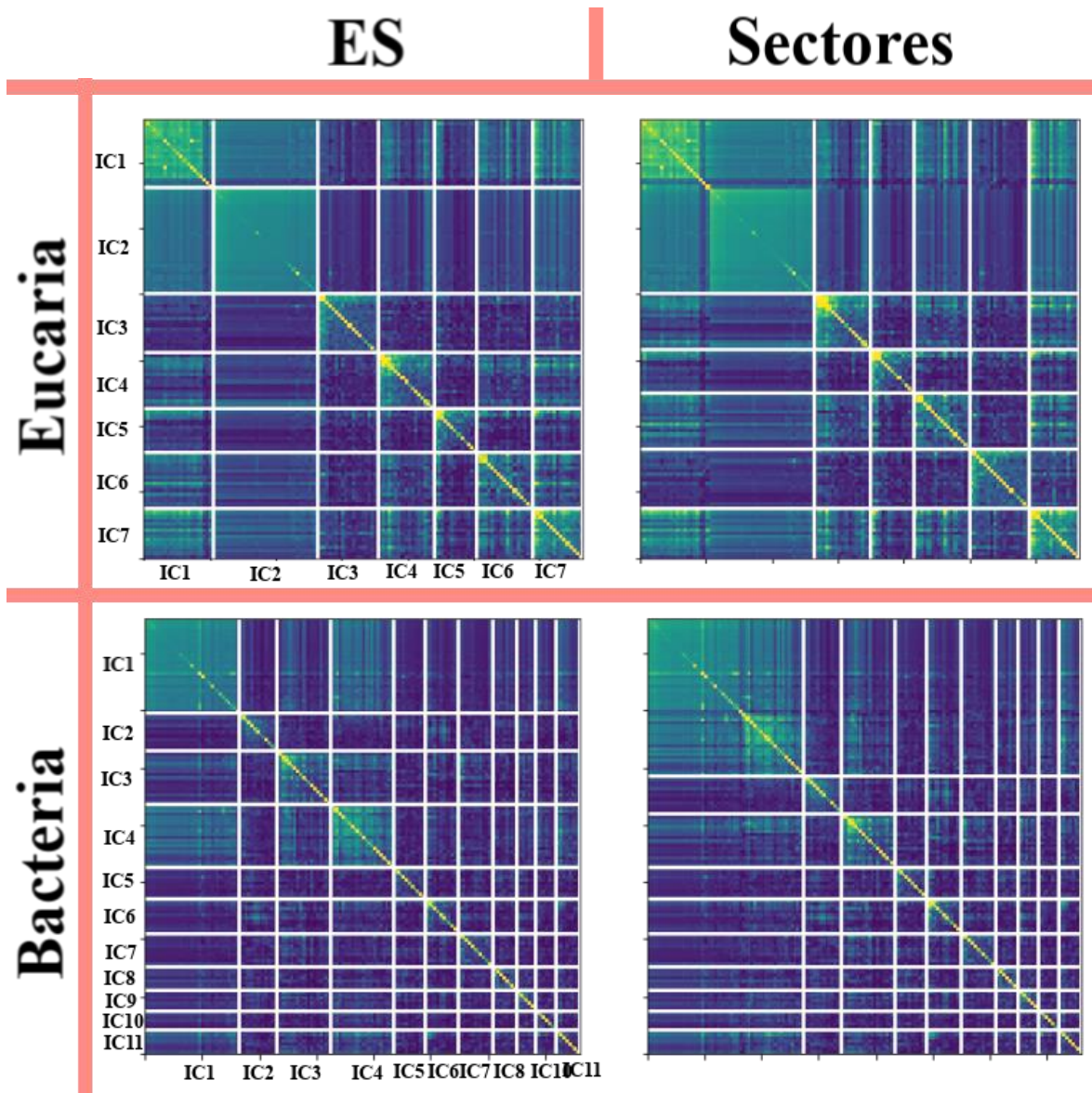


Figura 61. Matrices de correlación limpias y determinación de los Sectores.

Los residuos pertenecientes a todos los sectores se enlistan en la Tabla 14, según las secuencias de referencia para cada alineamiento de secuencia múltiple. Para el dominio Arquea se utilizó la secuencia y estructura de la TIM de *Pyrococcus furiosus* como referencia. Los aminoácidos que forman cada sector son los siguientes:

SECTORES	ARQUEA	EUCARIA	BACTERIA
Secuencia de Referencia	TIM de <i>Pyrococcus furiosus</i>	TIM de conejo "Oryctolagus cuniculus"	TIM de <i>Geobacillus stearothermophilus</i>
Sector 1	12+14+16+65+66+68+ 73+76+77+88+91+93+ 95+96+97+98+123+ 138+142+144+145+ 146+147+148+149+150+ 155+158+160+181+ 182+183+184+190+1 91+194+197+198+ 201+203+205+206+210	10+11+13+15+64+72+73+75+ 76+94+95+96+97+98+99+ 112+126+128+129+142+146+ 162+163+164+165+166+168+ 169+170+171+172+173+176+ 181+182+185+189+201+208+ 209+210+211+212+225+228+ 230+231+232+233+234+ 235+236+243	1+7+8+9+10+11+12+66+ 74+75+77+78+81+94+ 96+97+98+99+100+ 101+106+114+126+128+ 130+131+148+166+167+ 168+169+170+172+173+ 174+175+176+177+179+ 193+210+212+213+214+ 215+217+220+232+234+ 236+237+238+239+240+245
Sector 2	100+103+104+105+106+125+126+ 132+136+141	4+6+14+24+28+33+38+44+ 47+48+63+67+79+80+227+241+245	4+6+19+36+44+107+108+ 155+209+216+225+233+235
Sector 3	7+13+24+29+32+37+58+60+61+62+174	16+21+84+90+105+106+132+ 145+149+151+184+191+226	2+22+65+76+79+92+103+ 125+133+156+180+185+186+ 197+205+227+228+229+231
Sector 4	15+47+50+78+99+153+180+214+218	66+74+91+92+134+135+167+ 174+180+196+197+200+217+219+220+221+24	41+87+90+105+122+129+132+ 135+144+171+182
Sector 5	43+74+94+170+207	12+29+41+43+69+82+83+88+ 109+110+113+115+147+150+160+175+178+180	82+112+149+184+189+190+ 211+221+224+242+248+251
Sector 6	69+75+80+83+84+87+89+ 92+113+118+120	7+9+40+61+65+77+116+123+131+ 156+157+205+229+237	3+5+14+15+18+43+67+95+ 104+115+153+200
Sector 7	85+159+204+ 208+213+221+222		40+64+69+93+111+117+118+123
Sector 8	28+39+134+163+167+177+ 179+188+202+211+220		68+137+141+181+188+192+230
Sector 9	64+82+121+151+152+154+156+ 157+161+162+185+186		13+42+73+139+145+223+244
Sector 10	34+42+44+46+72+122+139+209		84+85+89+136+165+178+187+218

Tabla 14. Aminoácidos involucrados en los Sectores de los tres dominios de la vida.

El sector 1 contiene los 3 aminoácidos catalíticos que se muestran en amarillo en todas las figuras, todos los demás aminoácidos pertenecientes a este sector se encuentran espacialmente próximos a estos, como se puede apreciar en la vista monomérica (Figura 62). Además, este sector se conserva en los otros dos dominios bacteria y eucaria, que corresponde al Sector 1 en ambos dominios, cabe destacar que también contiene los aminoácidos catalíticos (Figura 63) (Figura 64). Esto indica que el Sector 1 está relacionado con la función catalítica.

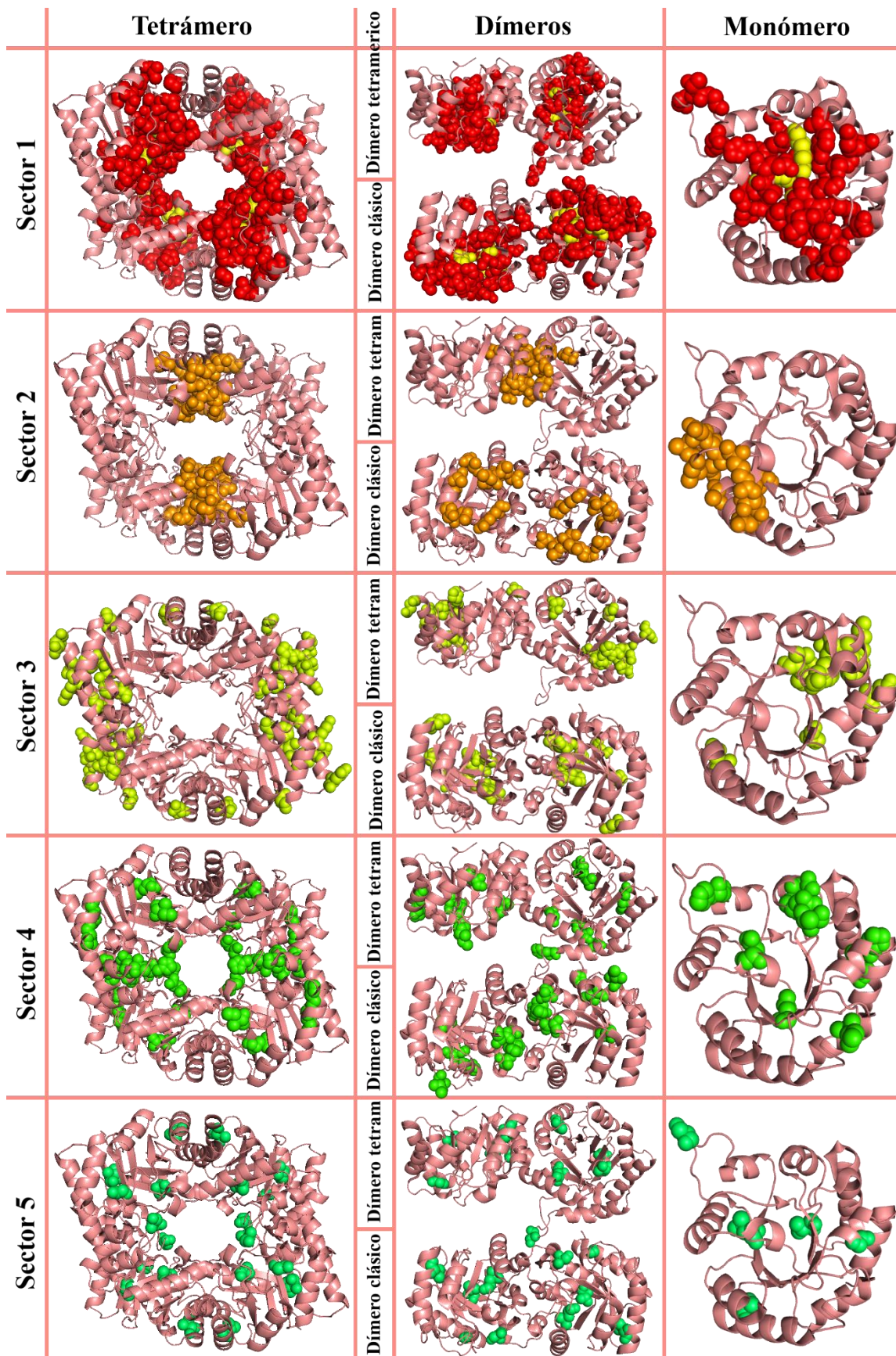


Figura 62. Sectores del dominio Arquea.

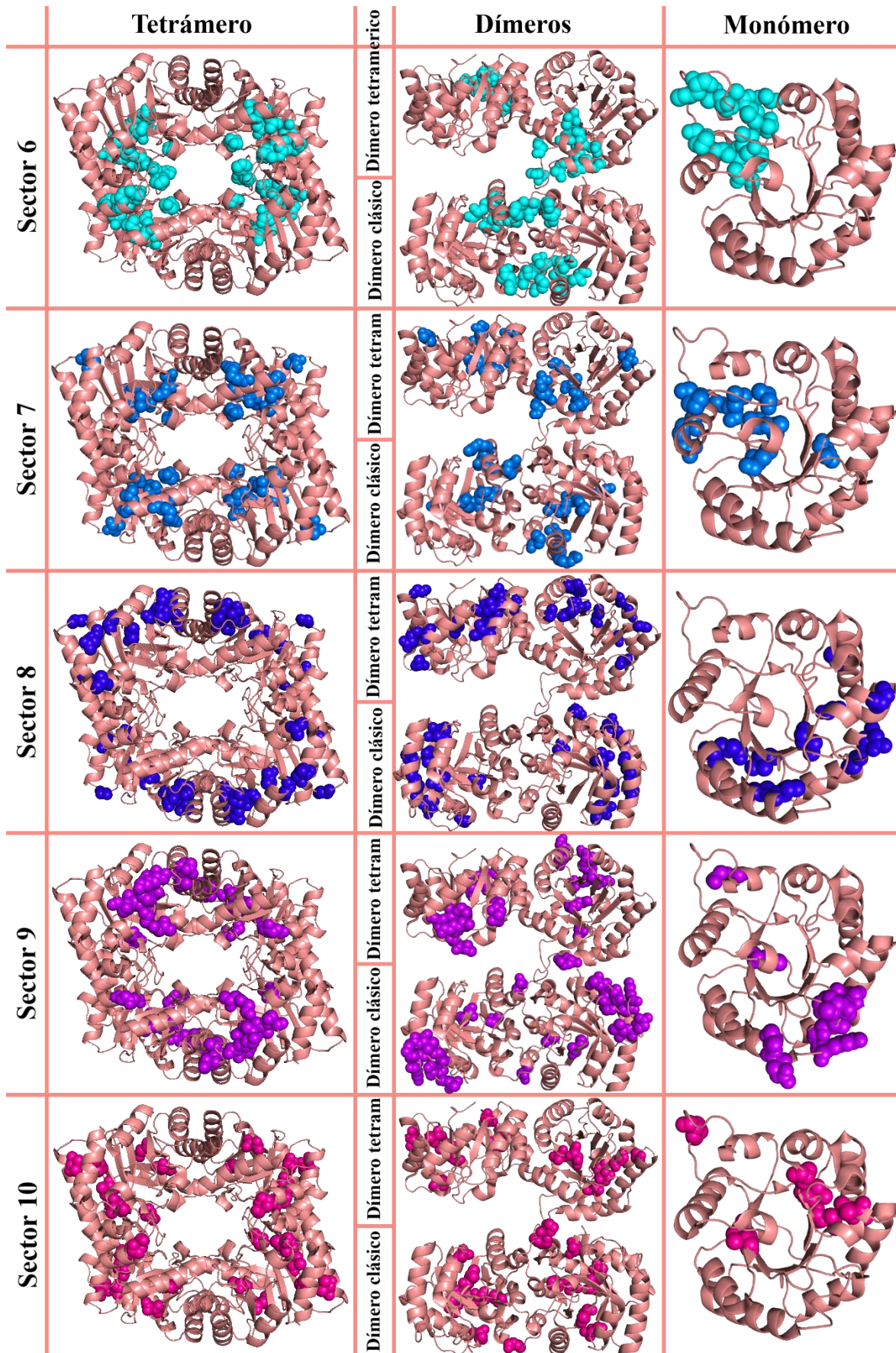


Figura 62. Sectores del dominio Arquea.

Los aminoácidos del sector 2 se agrupan perfectamente en la interfaz tetramérica, como se aprecia muy bien en la vista del tetrámero y del dímero tetramérico (Figura 62). Lo cual es sorprendente porque el único dato de entrada del SCA es un alineamiento múltiple de secuencias. Esto nos muestra que la metodología es capaz de distinguir grupos de aminoácidos asociados a una región de la estructura, en el caso del Sector 2: la interfaz tetramérica. Cabe destacar que el Sector 2 aparece únicamente en el dominio arquea.

También llamó nuestra atención el Sector 8, que tiene algunos aminoácidos de la hélice 5 y los aminoácidos del C-t de la hélice 6 que igualmente participan en la interfaz tetramérica. Parece ser que la disposición espacial de este sector ya no se encuentra tan concentrada en la interfaz tetramérica ya que tiene algunos otros residuos lejanos, pero eso no elimina la posibilidad de que tenga alguna relación epistática con los aminoácidos del Sector 2 y por lo tanto, también esté asociado a la interfaz tetramérica. Este sector tampoco está presente en los otros dominios.

Los residuos del Sector 4, se agrupan muy bien en la interfaz dimérica, como podemos ver en las vistas del tetrámero y del dímero clásico en la Figura 57. Además, existe un sector homólogo en los dominios bacteria y eucaria, que corresponden al Sector 3 en ambos dominios (Figura 63) (Figura 64). Por estas razones, teorizamos que el conjunto de residuos que lo forman juega un papel importante en la formación del dímero. Otro caso de homología se da entre los Sectores 4 del dominio Bacteria y Eucaria que tienen aminoácidos en el asa 3 y en la periferia del barril, por lo que parece tener un papel en la estabilidad del dímero (Figura 63) (Figura 64).

El caso del Sector 6 de las arqueas es interesante pues agrupa todos sus residuos en el asa 3 (Figura 62), tiene un sector homólogo en el dominio Eucaria que es el Sector 5, que igualmente agrupa la mayoría de sus residuos en el asa 3 por lo que suponemos que estos aminoácidos tienen un papel importante en la dimerización (Figura 63). El caso del Sector 6 de Eucaria, parece tener homología con los Sectores 7 y 9 del dominio Bacteria. Aunque se desconoce cuál podría ser su función asociada (Figura 63) (Figura 64).

Respecto a los Sectores 3, 5, 7, 9 y 10 de las Arqueas no tenemos pista de cual pueda ser la función relacionada a ellos, ya que no se agrupan en alguna zona con actividad específica conocida (Figura 62). Sucede lo mismo para los Sectores 2 y 6 del dominio Eucaria (Figura 63) y para los Sectores 2, 5, 6, 7, 8, 9 y 10 del dominio Bacteria (Figura 64).

El que no veamos una asociación a simple vista no significa que no exista. Sin embargo, siempre hacen falta datos experimentales que corroboren o nieguen las propuestas de las metodologías estadísticas como el SCA.

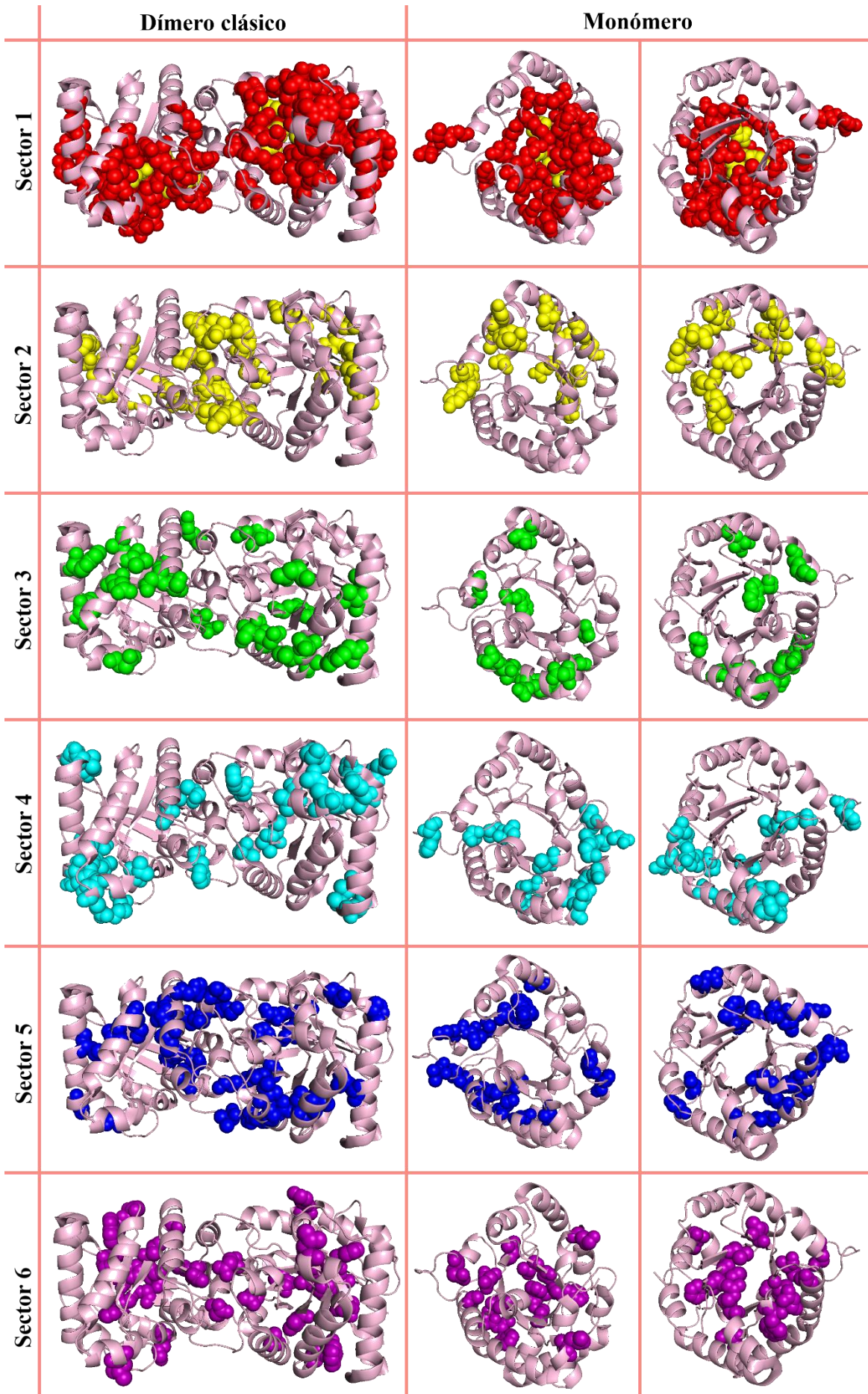


Figura 63. Sectores del dominio Eucaria.

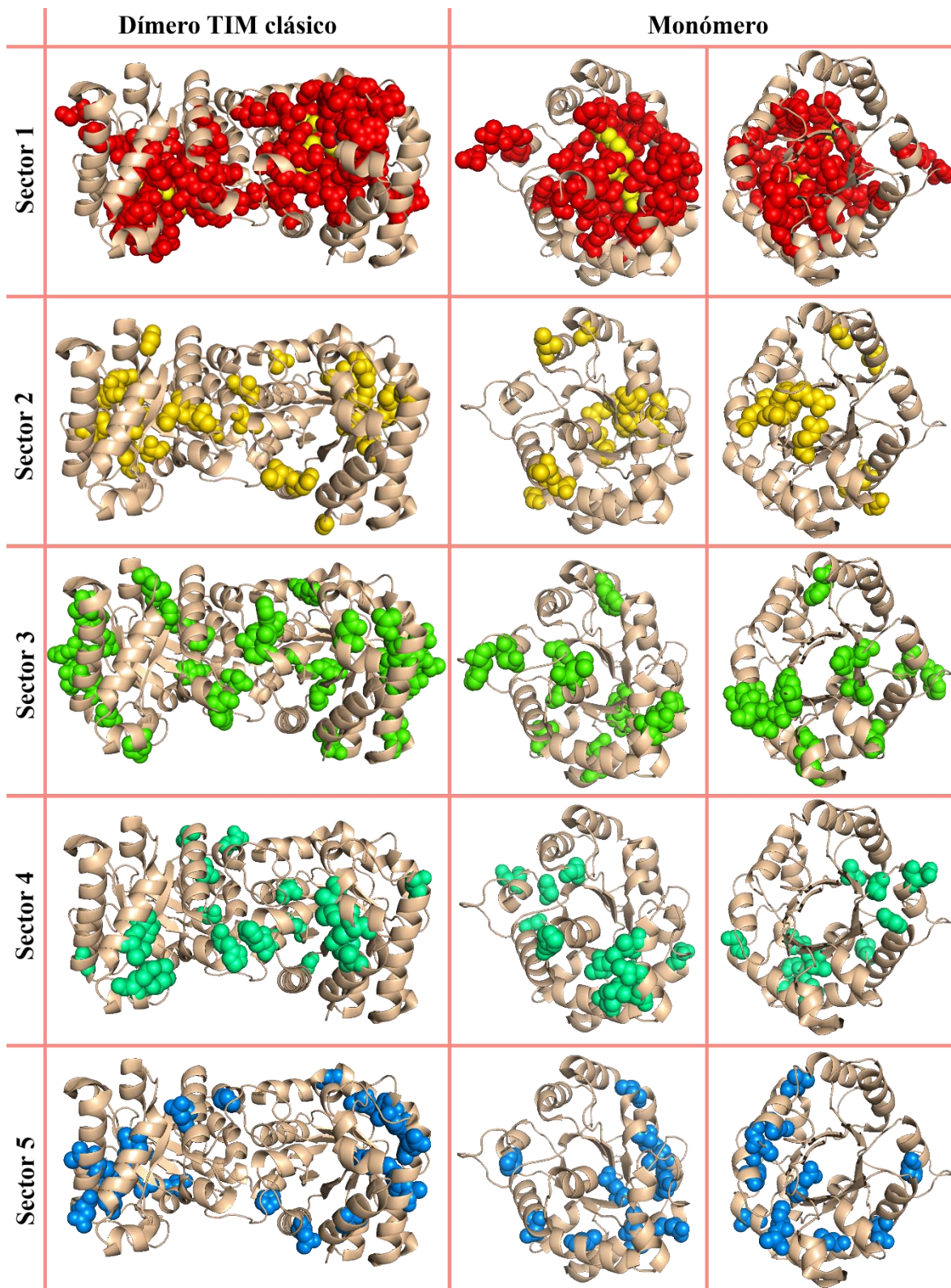


Figura 64. Sectores del dominio Bacteria.

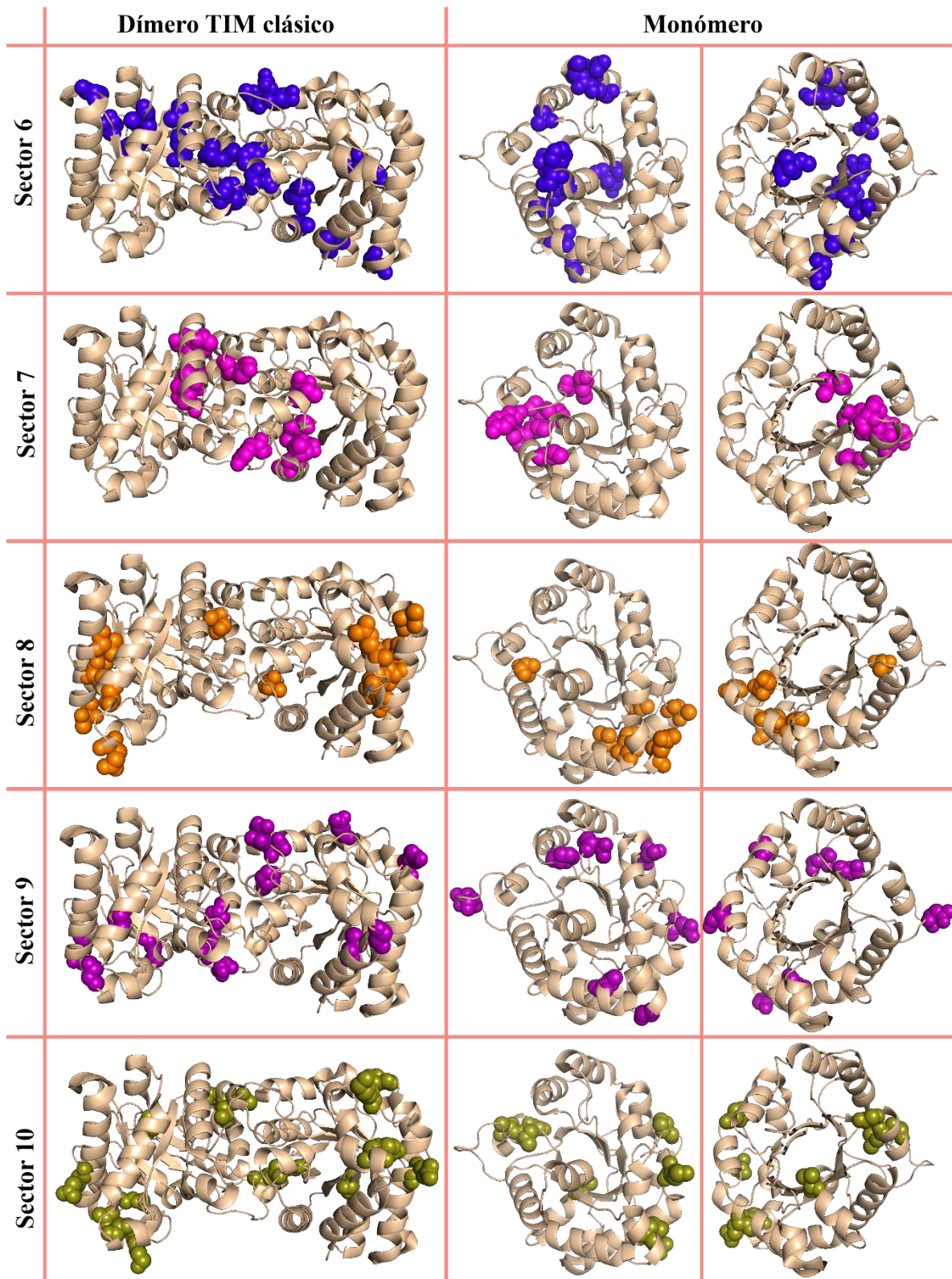


Figura 64. Sectores del dominio Bacteria.

La nueva versión del SCA (6.0) nos permite determinar si hay alguna correlación entre los sectores y los grupos filogenéticos de alineamiento de secuencias múltiple (Figura 43). Por fines prácticos, solo analizaremos este resultado para el dominio Arquea. En la Figura 60 se muestran dos tipos de gráficas de dispersión, ICs y Ui y un histograma Ui. En la gráfica de dispersión de componentes independientes (ICs) cada punto representa un aminoácido, los cuales son coloreados según el sector al que pertenecen, la Tabla 15 contiene el código de colores; en esta gráfica la distancia entre los puntos representa su correlación. En la gráfica de dispersión Ui cada punto representa una secuencia, coloreados según el orden filogenético de las Arqueas al que pertenecen, la Tabla 15 muestra el código de colores; en esta gráfica la distancia entre los puntos indica su porcentaje de identidad. Finalmente, el histograma Ui es una representación diferente de la misma información filogenética. Todas estas gráficas nos ayudan a identificar visualmente si algún grupo filogenético en particular es responsable de la aparición de algún sector específico.

CÓDIGO FILOGENÉTICO	CÓDIGO DE SECTORES
Archaeoglobales	1
Bathyarchaeota	2
Methanomicrobia	3
Halobacteria	4
Lokiarchaeota	5
Methanobacteria	6
Methanococci	7
Thermoprotei	8
Nitrosopumilales	9
Poseidoniales	10
Thaumarchaeota	
Thermococcales	
Thermoplasmatales	
Methanomassiliicoccales	

Tabla 15. Código de colores para los sectores y para el orden filogenético de las Arqueas

En la Figura 65, las gráficas IC1, IC2 e IC3 muestran como la mayoría de las secuencias del alineamiento contribuyen a los Sectores 1, 2 y 3, respectivamente. Mientras que las gráficas IC4 y U4 nos muestran que el Sector 5 tiene una gran contribución del grupo Halobacteria, esta contribución se marca en un círculo negro en el histograma Ui. Las gráficas IC6 e IC7 igualmente nos muestran que

todas las secuencias contribuyen a los Sectores 1 y 6. Las gráficas IC8 e IC9 nos muestran que todas las secuencias contribuyen al Sector 8. En cambio, la gráfica IC10 muestra que el Sector 9, se ve fuertemente influenciado por el grupo Thermoplasmatales. La influencia de los Thermoplasmatales en el Sector 9 también se observa en la gráfica IC11, y está marcado por un círculo negro en el histograma U11 (Figura 65). En la misma gráfica IC11 se observa que el grupo Thermoprotei contribuye a la formación del Sector 10.

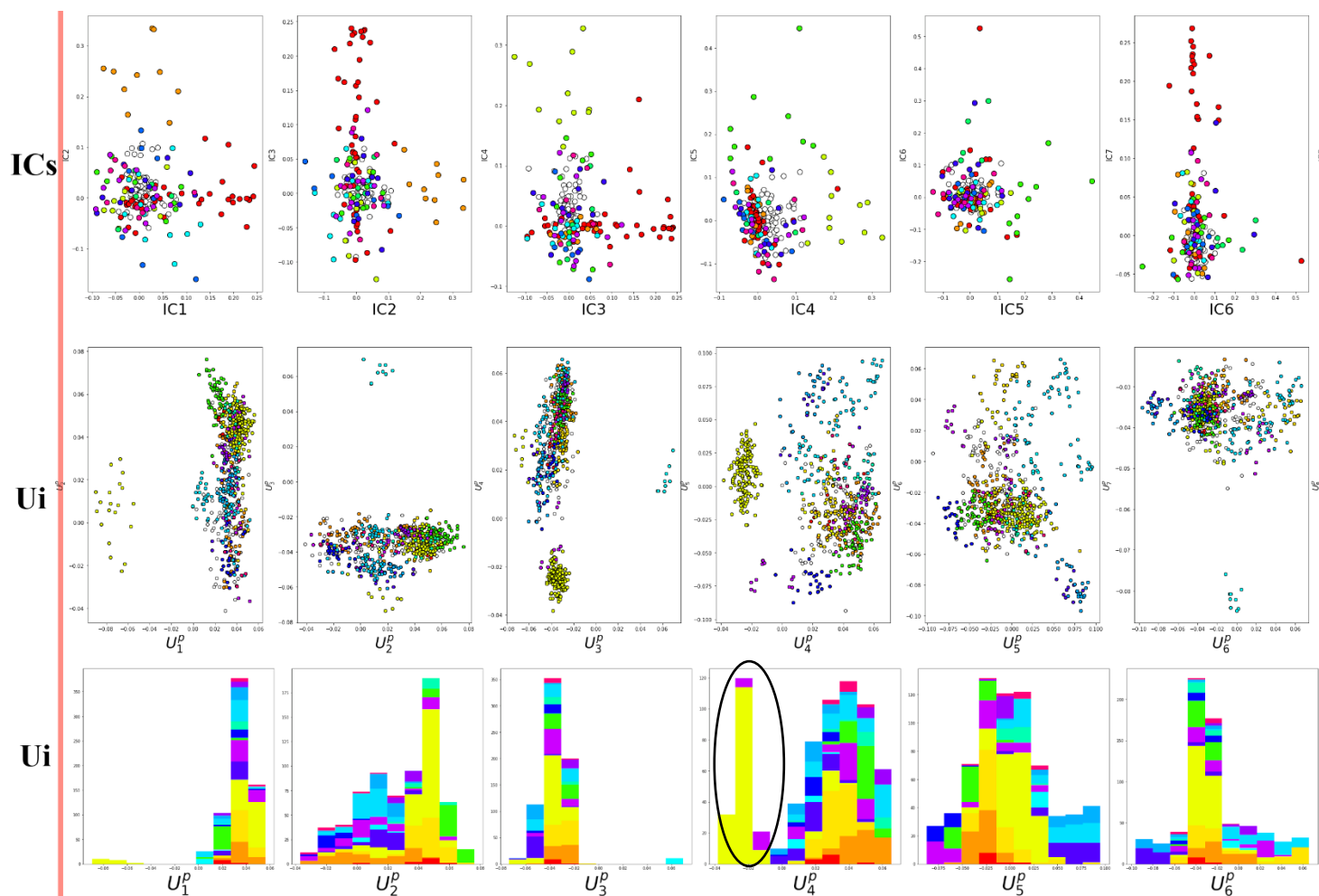


Figura 65. Relación entre la filogenia del dominio Arquea y sus Sectores.

Las relaciones entre los sectores y los órdenes filogenéticos de las Arqueas se resumen a continuación:

- Sector 5 ---- Halobacteria
- Sector 9 ---- Thermoplasmatales
- Sector 10 --- Thermoprotei

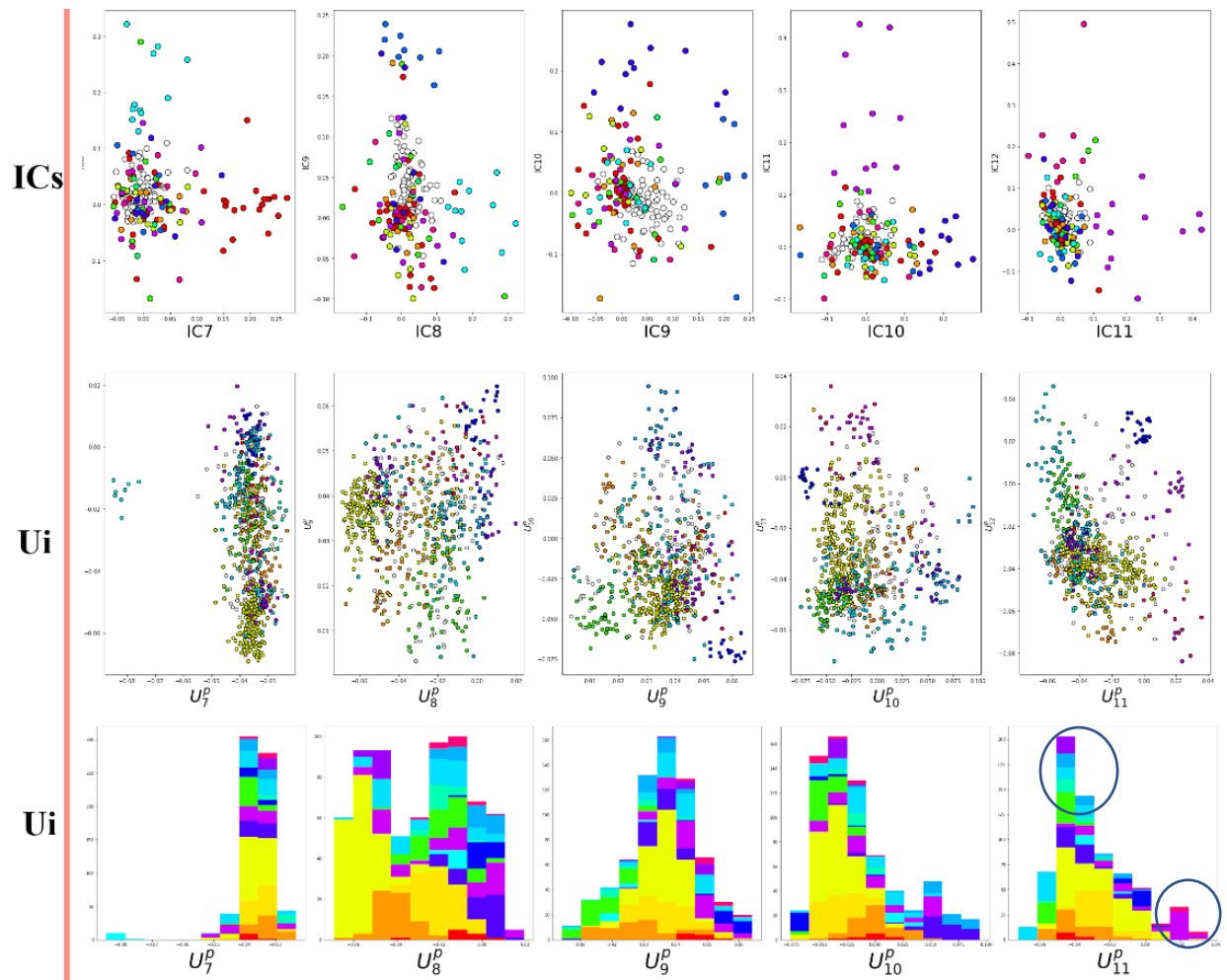


Figura 65. Relación entre la filogenia del dominio Arquea y sus Sectores.

DISCUSIÓN

A lo largo de este trabajo de tesis no encontramos una asociación directa entre la termoestabilidad y el estado de oligomerización en las ArqTIMs. El caso de TacTIM que es la única estructura de una ArqTIM que es dímero en solución, parece indicar que las características que no le permiten tetramerizar son compartidas únicamente por organismos que pertenecen al orden de los Thermoplasmatales. Además, los resultados del SCA (Figura 60) muestran que el grupo de los Thermoplasmatales está filogenéticamente bien definido y tiene bajos niveles de identidad con las otras secuencias de Arqueas. Por lo que concluimos que el estado oligomérico de TacTIM no está relacionado con la adaptación a altas temperaturas, más bien está relacionado filogenéticamente. Debido a que estas características están presentes únicamente en organismos acidófilos podríamos pensar que las características que les permiten ser dímeros se adquirieron en respuesta a la necesidad de supervivencia a ambientes con pH ácido. A pesar de esto, es necesario tener cuidado al realizar hipótesis adaptacionistas que carezcan de evidencia experimental. Es posible que estos rasgos de los Thermoplasmatales se hayan seleccionado en función del pH, sin embargo, necesitamos estudios experimentales, para afirmarlo. Una buena aproximación experimental sería la caracterización fisicoquímica de varias proteínas del grupo de los Thermoplasmatales y observar su comportamiento a diferentes pHs. Incluso, puede ser también que no sea un proceso adaptativo, sino que se trate de un caso de deriva génica en el orden de los Thermoplasmatales.

El segundo caso que investigamos fue la gran diferencia entre la termoestabilidad de MjaTIM y MfoTIM que comparten el mismo estado de oligomerización, por lo que esta condición no puede ser responsable de dicha diferencia. La estructura de MfoTIM que resolvimos en el grupo de investigación, conserva la manera en la que se asocia el tetrámero y los aminoácidos involucrados en la interfaz tetramérica observada en otros cristales. Las principales diferencias que encontramos fueron las substituciones de ciertos residuos que parecen empaquetar mejor el tetrámero en MjaTIM (Figura 49). Se ha visto que las proteínas termófilas tienden a ser más compactas (Szilagyí & Závodszky 2008). Por lo que quizá estas substituciones estén relacionadas con el aumento en la termoestabilidad de MjaTIM, aunque para comprobarlo es necesario realizar experimentos de mutagénesis dirigida sobre estos residuos que lo comprueben. Otro fenómeno que también se ha observado en las proteínas termófilas es que tienden a presentar más cargas electrostáticas en su superficie. Lo que concuerda con la gran diferencia en la distribución de cargas en la superficie que observamos entre ambas proteínas, la cual fue notablemente más negativa en MjaTIM (Figura 48).

Este fenómeno se observó nuevamente con los modelos obtenidos por AlphaFold2. Las TIMs hipertermófilas tuvieron en promedio valores de carga neta de -10, las TIMs termófilas de -6.2 y las TIMs mesófilas de -1.2. Visualmente se aprecia muy bien las diferencias en las distribuciones de carga en la superficie (Figura 51.52.53). Por lo que concluimos que esta característica parece correlacionar con la temperatura óptima de crecimiento y quizá es la responsable de las diferencias en termoestabilidad entre las ArqTIMs hipertermófilas, termófilas y mesófilas. Por otra parte, la distribución de cargas de la superficie no tuvo correlación con el estado oligomérico.

Los resultados de HSYMDOCK sugieren que la asociación tetramérica es posible incluso en ArqTIMs mesófilas, debido a que conservan características estructurales simétricas, que les permitirían asociarse como homotetrámeros. Esta es una propuesta atrevida, puesto que no se conoce el estado oligomérico de las MesoTIMs (R. Muñiz, 2020) y algunos datos experimentales indican que MbuTIM y MbrTIM son dímeros en solución (Schramm A, Kohlhoff M, Hensel R. 2001) (Dhaunta N, et. al. 2013). Por ello, propongo al lector que replantemos la evidencia. En la (Figura 62) se encuentra el experimento de cromatografía exclusión analítica de las TermoTIMs y de MbuTIM. Para ambos experimentos se utilizó la columna Superdex 200 10/300 GL. Se puede apreciar como en ambos experimentos el volumen de elución es de ~14mL, aunque tenemos que considerar varios aspectos:

1. El amortiguador que usaron para MbuTIM fue 20 mM Tris a pH 8.0, en cambio Muñiz R utilizó un amortiguador con 200 mM NaCl. La presencia de sal modifica el volumen de elución.
2. La cromatografía de exclusión analítica se acompaña siempre de una curva de calibración. La curva que presentan en el material suplementario de (Dhaunta N, et. al. 2013), el volumen de elución de MbuTIM corresponde a un hexámero, no tetrámero.
3. Dhaunta y colaboradores también realizaron ultracentrifugación analítica y dispersión de luz multi-ángulo, mostrando que en esas condiciones MbuTIM es dimérica.

Por lo que considero prudente repetir la caracterización fisicoquímica de esta proteína en condiciones iguales a las usadas por (Muñiz, 2020). Mi sana curiosidad científica me hace dudar si realmente MbuTIM es un dímero en solución o si su estado oligomérico depende de las condiciones del solvente. Por lo que considero prudente repetir la caracterización fisicoquímica de esta proteína.

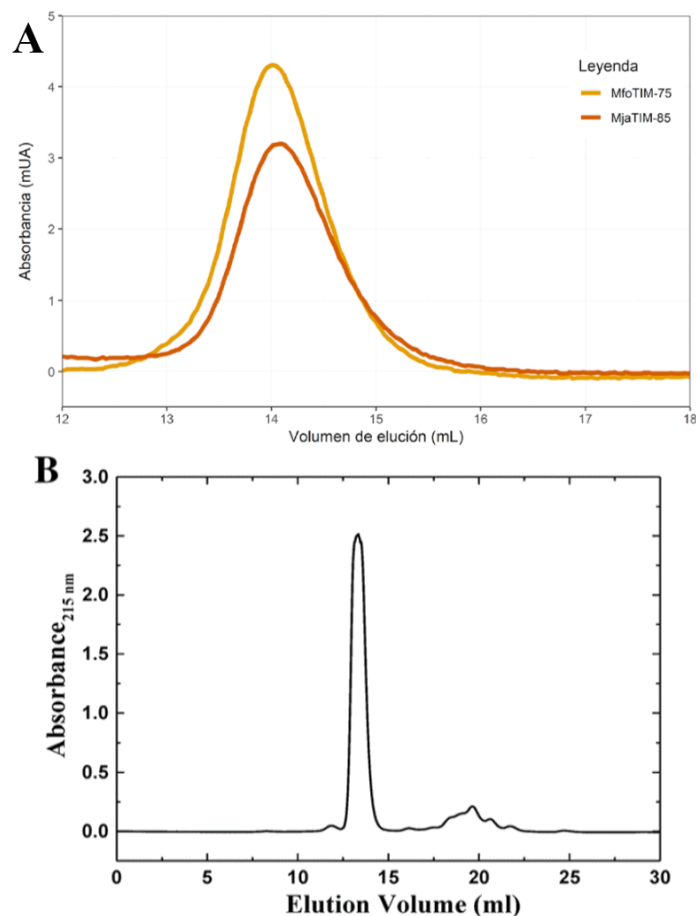


Figura 66. Resultados de la cromatografía de exclusión analítica. A. *TermoTIMs* B. *MbuTIM*. Tomado de (R. Muñiz, 2020) & (Dhaunta et. al. 2013)

En el caso de las *MesoTIMs* es probable que no se hayan podido purificar debido a una falta de un estabilizador externo que está presente en la expresión homóloga en Arqueas, pero ausente en la expresión heteróloga en *E. coli*. Esto podría deberse a la falta de un soluto, como el caso de *MfeTIM* que recobro ~50% de su actividad residual con sales fosfatos (Shramm, Kohlhoff & Hensel, 2001), o bien puede deberse a la falta de una proteína que estabilice a las *MesoTIMs*; como el caso de *TtxTIM* que recupero el 100% de su actividad a 80°C al incubarla en presencia de la enzima GLPDH (Walden et. al. 2004) (Figura 24). Los geles de purificación de las *MesoTIMs* muestran varias bandas de proteínas contaminantes. La banda de ~45kDa (Figura 62.B) podría corresponder a la enzima 3-fosfoglicerato cinasa (PKG). Al tratar de purificar la PKG de *Pyrococcus furiosus*, banda de ~45kDa, se encontraron con una banda contaminante de ~28kDa, lo que pesaría la TIM de *E. coli* (Hess D. and Hensel R. 1996). En la (Figura 62) en A se muestra el gel de la purificación de la PKG de *Pyrococcus furiosus* y en B el gel de la purificación de las

MesoTIMs, podemos ver como algunas bandas permanecen en ambos geles, se marcan con círculos verdes. Además, cabe mencionar que la única TIM bacteriana capaz de tetramerizar es la de *Thermotoga maritima* (TmaTIM) que se asocia con la PKG por medio de 4 puentes disulfuro (Schurig H. et. al. 1995).

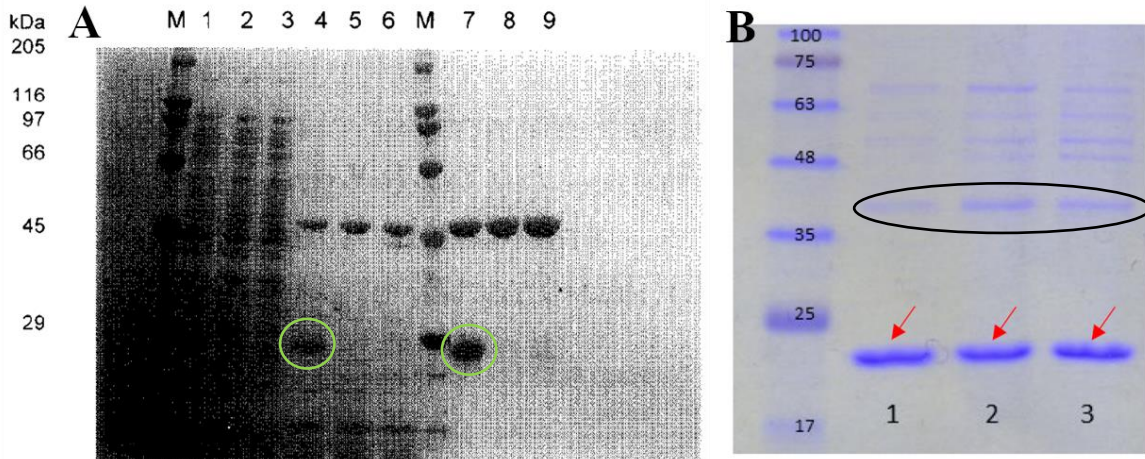


Figura 67. Geles de purificación de la PKG de Pfu (Hess D, Hensel 1996) y las MesoTIMs (Muñiz R, 2020). A. Gel de la purificación de la PKG de *Pyrococcus furiosus* en los carriles 1-3 se observan los extractos crudo de *E. coli*. En los carriles 4-6 se observan los precipitados calentados a 90° por 30min. M es el marcador de peso molecular y finalmente en los carriles 7-9 las fracciones después de pasar por una columna de Niquel +2. B. Gel de purificación de las MesoTIMs. En el primer carril vemos el marcador peso molecular. 1) *MmaTIM*, 2) *MaeTIM* y 3) *MokTIM*.

La asociación de las ArqTIMs con otras proteínas puede ser posible mediante el área hidrofóbica expuesta a solvente que encontramos conservada en todos los cristales y en todos los modelos de AlphaFold2 (Figura 54.55.56). Esta área es grande por lo que podría formar una nueva superficie de unión, para formar un hetero-oligómero estable. De hecho, quizá la exposición de esta área hidrofóbica hace inestables a las MesoTIMs lo cual podría originar su tendencia a la agregación, lo cual es una explicación probable al hecho de que no se pudieron purificar en el laboratorio.

En cuanto a los Sectores encontrados en las ArqTIMs, encontramos que algunos están conservados en todos los dominios de la vida, como el Sector 1, asociado a la función catalítica, así como los sectores 4 y 6 asociados a la dimerización de las TIMs. Por otra parte, también encontramos sectores únicos del dominio Arquea como el Sector 2 y 8, asociados a la interfaz tetramérica, también los Sectores 3, 5, 7, 9 y 10 son únicos de este dominio, aunque no tenemos todavía pistas sobre si estos sectores están relacionados a una función específica. Por último, encontramos correlación entre el Sector 5 y el orden Halobacteria, entre el Sector

9 y el orden Thermoplasmatales y entre el Sector 10 y el orden Thermoprotei (Figura 60).

El estado de oligomerización no depende únicamente de la adaptación a la temperatura, sino también puede incluir relaciones filogenéticas, casos quizá de deriva génica en el caso de los Thermoplasmatales, o bien relaciones proteína-proteína, como el caso de la PKG. La distribución electrostática de las cargas en las superficies de las ArqTIMs parece ser el único fenómeno asociado directamente con la adaptación a la temperatura. Aunque tendríamos que conocer la T_m de todos nuestros modelos para corroborarlo.

CONCLUSIONES

- La interfaz tetramérica de las ArqTIMs se mantiene gracias a interacciones hidrofóbicas
- Los aminoácidos involucrados en la interfaz tetramérica: L99, I102, V125, A128 y V129 (tomando la secuencia de MfoTIM como referencia) están sometidos a una presión selectiva
- Las características que le permiten a TacTIM ser un dímero están relacionadas filogenéticamente con el grupo de los Thermoplasmatales
- En las TIMs, El grupo de los Thermoplasmatales está filogenéticamente bien definido y tiene bajos niveles de identidad con las otras secuencias de Arqueas
- La característica responsable de la gran diferencia en la Tm entre MjaTIM y MfoTIM están probablemente relacionadas con la distribución de cargas electrostáticas en la superficie.
- Las sustituciones de MjaTIM empaquetan notablemente mejor al tetrámero que en MfoTIM.
- Es posible que las ArqTIMs mesófilas tengan Tm elevadas.
- La tetramerización es posible para todas las TIMs de arqueas incluso para las mesófilas
- La única característica que se relaciona directamente con la temperatura óptima de crecimiento es la distribución de cargas electrostáticas en la superficie de las ArqTIMs
- Existe un área hidrofóbica expuesta a solvente en la cara externa de la interfaz tetramérica que puede estar actuando como una nueva área de interacción con otras proteínas.
- Se propone a la PKG como una proteína que posiblemente sea capaz de interactuar con las ArqTIMs.
- Los aminoácidos Y16, N95 y P146 están conservados específicamente en el dominio Arquea
- Los sectores 2 y 8, relacionados con la interfaz tetramérica, solo se presentan en el dominio arquea
- El Sector 5 se relaciona con el orden de las Halobacterias
- El Sector 9 se relaciona con el orden de los Thermoplasmatales
- El Sector 10 se relación con el orden Thermoprotei.
- Aun hay muchas cosas por conocer de las ArqTIMs

PERSPECTIVAS

- Obtener experimentalmente la T_m y el estado de asociación de las siguientes ArqTIMs: MfeTIM, TonTIM, PfuTIM, TtxTIM, MokTIM, MmaTIM, MaeTIM y MbrTIM.
- Generar mutaciones puntuales que correspondan a las diferencias entre MfoTIM y MjaTIM para corroborar su efecto en la termoestabilidad.
- Sería interesante expresar otra TIM del orden de los Thermoplasmatales y estudiar cómo se comporta a diferentes pH y cuál es su estado oligomérico.
- Hacer mutaciones puntuales de los aminoácidos coevolutivos del sector 2 para verificar su papel en la formación del tetrámero.
- Coexpresar las MesoTIMs con la PKG para verificar o descartar la asociación TIM-PKG.

GRACIAS POR LEER ESTA TESIS...

REFERENCIAS

- Beaucamp N, Hofmann A, Kellerer B, Jaenicke R. Dissection of the gene of the bifunctional PGK-TIM fusion protein from the hyperthermophilic bacterium *Thermotoga maritima*: design and characterization of the separate triosephosphate isomerase. *Protein Sci.* 1997 Oct;6(10):2159-65.
- Beaucamp N, Schurig H, Jaenicke R. The PGK-TIM fusion protein from *Thermotoga maritima* and its constituent parts are intrinsically stable and fold independently. *Biol Chem.* 1997 Jul;378(7):679-85.
- Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature.* 2006 Dec 14;444(7121):929-32.
- Blacklow, C. Raines, T. et. Al. Triosephosphate isomerase catalysis is diffusion controlled. Appendix: Analysis of triose phosphate equilibria in aqueous solution by ³¹P NMR. *Biochemistry (1988) Feb 23;27(4): 1158-67.*
- Brininger C, Spradlin S, Cobani L, Evilia C. The more adaptive to change, the more likely you are to survive: Protein adaptation in extremophiles. *Semin Cell Dev Biol.* 2018 Dec;84:158-169.
- Chakravarty S, Varadarajan R. Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. *Biochemistry.* 2002 Jun 25;41(25):8152-61.
- Chandrayan SK, Guptasarma P. Attenuation of ionic interactions profoundly lowers the kinetic thermal stability of *Pyrococcus furiosus* triosephosphate isomerase. *Biochim Biophys Acta.* 2009 Jun;1794(6):905-12.
- Coker, James. Recent advances in understanding extremophiles. *Faculty Review.* (2019). 13; 8: F100.
- Cooper T. and Santiago E. The causes of epistasis. *Proceedings of the Royal Society* (2011).
- Despa F, Orgill DP, Lee RC. Effects of crowding on the thermal stability of heterogeneous protein solutions. *Ann Biomed Eng.* 2005 Aug;33(8):1125-31.
- Dey S, and Levy E. Inferring and Using Protein Quaternary Structure Information from Crystallographic Data. Chapter 23 Joseph A. Marsh (ed.), *Protein Complex Assembly: Methods and Protocols, Methods in Molecular Biology*, vol. 1764, 2018.
- Dey S, and Levy E. Inferring and Using Protein Quaternary Structure Information from Crystallographic Data. Chapter 23 Joseph A. Marsh (ed.), *Protein Complex Assembly: Methods and Protocols, Methods in Molecular Biology*, vol. 1764, 2018.
- Dey S, and Levy E. PDB-wide identification of physiological heterooligomeric assemblies based on conserved quaternary structure geometry. *Cell Press Structure* 29, 1303–1311, November 4, 2021.

- Dey S, and Levy E. PDB-wide identification of physiological heterooligomeric assemblies based on conserved quaternary structure geometry. *Cell Press Structure* 29, 1303–1311, November 4, 2021.
- Dhaunta N, Arora K, Chandrayan SK, Guptasarma P. Introduction of a thermophile-sourced ion pair network in the fourth beta/alpha unit of a psychrophile-derived triosephosphate isomerase from *Methanococcoides burtonii* significantly increases its kinetic thermal stability. *Biochim Biophys Acta*. 2013 Jun;1834(6):1023–33.
- Dhaunta, N. Arora, K. et. al. Introduction of a thermophile-sourced ion pair network in the fourth beta/alpha unit of a psychrophile-derived triosephosphate isomerase from *Methanococcoides burtonii* significantly increases its kinetic thermal stability. *Biochimica et Biophysica Acta*. 1834 (2013) 1023–1033.
- Dolinsky TJ, Nielsen JE, MacCammon AJ, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Research*, 2004, Vol. 32.
- Dolinsky TJ, Nielsen JE, MacCammon AJ, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Research*, 2004, Vol. 32.
- Elcock AH, Huber GA, McCammon JA. Electrostatic channeling of substrates between enzyme active sites: comparison of simulation and experiment. *Biochemistry*. 1997 Dec 23;36(51):16049–58.
- Fiala, G.; Stetter, K. O. (1986). «*Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100 °C». *Archives of Microbiology* 145: 56–61.
- Franzmann, P.D.; Springer, N.; Ludwig, W.; Conway De Macario, E.; Rohde, M. (1992). "A Methanogenic Archaeon from Ace Lake, Antarctica: *Methanococcoides burtonii* sp. nov". *Systematic and Applied Microbiology*. 15 (4): 573–581.
- Gayathri P, Banerjee M, Vijayalakshmi A, Azeez S, Balaram H, Balaram P, Murthy MR. Structure of triosephosphate isomerase (TIM) from *Methanocaldococcus jannaschii*. *Acta Crystallogr D Biol Crystallogr*. 2007 Feb;63(Pt 2):206–20.
- Goldman, A. et. al. The TIM Barrel Architecture Facilitated the Early Evolution of Proteins- Mediated Metabolism. *Journal Mol Evol*. (2016) 82:17–26.
- Gómez-Puyou A, Saavedra-Lira E, Becker I, Zubillaga RA, Rojo-Domínguez A, Pérez-Montfort R. Using evolutionary changes to achieve species-specific inhibition of enzyme action--studies with triosephosphate isomerase. *Chem Biol*. 1995 Dec;2(12):847–55.

- Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci.* 1979 Sep 21;205(1161):581-98.
- Gürol MS, Lockless SW, Wall MA, and Ranganathan R. Evolutionary conserved networks of residues mediate allosteric communication in proteins. *Nature structural biology* | Volume 10 number 1 | January 2003.
- Gürol MS, Lockless SW, Wall MA, and Ranganathan R. Evolutionary conserved networks of residues mediate allosteric communication in proteins. *Nature structural biology* | Volume 10 number 1 | January 2003.
- Halabi N, Rivoire O, Leibler S and Ranganathan R. Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138, 774–786, August 21, 2009.
- Halabi N, Rivoire O, Leibler S and Ranganathan R. Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138, 774–786, August 21, 2009.
- Harms J Michael and Thornton W. Joseph. *Evolutionary Biochemistry: Revealing the historical and physical causes of protein properties.* Nature Reviews, Volumen 14 (2013).
- Hess D, Hensel R. The 3-phosphoglycerate kinase of the hyperthermophilic archaeum *Pyrococcus woesei* produced in *Escherichia coli*: loss of heat resistance due to internal translation initiation and its restoration by site-directed mutagenesis. *Gene.* 1996 Jun 12;172(1):121-4.
- Huang SY, Zou X. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins.* 2008 Aug;72(2):557-79.
- Huang SY, Zou X. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins.* 2008 Aug;72(2):557-79.
- Isupov MN, Fleming TM, Dalby AR, Crowhurst GS, Bourne PC, Littlechild JA. Crystal structure of the glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J Mol Biol.* 1999 Aug 20;291(3):651-60.
- Jaenicke R, Kellerer B, Hoffman A, Beaucamp N. Dissection of the gene of the bifunctional PGK-TIM fusion protein from the hyperthermophilic bacterium *Thermotoga maritima*: Design and characterization of the separate triosephosphate isomerase. *Protein Science* (1997). 6:2159-2165.
- Jelesarov I, Bosshard HR. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J Mol Recognit.* 1999 Jan-Feb;12(1):3-18.
- Jumper J, et. al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021 Aug;596(7873):583-589.
- Jumper J, et. al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021 Aug;596(7873):583-589.

- Jurrus E. et al. Improvements to the APBS biomolecular solvation software suite. *PROTEIN SCIENCE* 2018 VOL 27:112–128.
- Jurrus E. et al. Improvements to the APBS biomolecular solvation software suite. *PROTEIN SCIENCE* 2018 VOL 27:112–128.
- Kohlhoff M, Dahm A, Hensel R. Tetrameric triosephosphate isomerase from hyperthermophilic Archaea. *FEBS Lett.* 1996 Apr 1;383(3):245-50.
- Krissinel E, and Henrick K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* (2007) 372, 774–797.
- Krissinel E, and Henrick K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* (2007) 372, 774–797.
- Kyte J, and Doolittle RF. A Simple Method for Displaying Hydrophobic Character of a Protein. *J. Mol Biol.* (1982) 157, 105-132.
- Kyte J, and Doolittle RF. A Simple Method for Displaying Hydrophobic Character of a Protein. *J. Mol Biol.* (1982) 157, 105-132.
- Laskowski R and Swindells M. LigPlot+: Multiple Ligand+ Protein Interaction Diagrams for Drug Discovery. *Journal of Chemical Information and Modeling* 2011, 51, 2778–2786.
- Laskowski R and Swindells M. LigPlot+: Multiple Ligand+ Protein Interaction Diagrams for Drug Discovery. *Journal of Chemical Information and Modeling* 2011, 51, 2778–2786.
- Lecocq M, Groussin M, Gouy M, Brochier-Armanet C. The Molecular Determinants of Thermoadaptation: Methanococcales as a Case Study. *Mol Biol Evol.* 2021 May 4;38(5):1761-1776.
- Levy Y, Onuchic JN. Water mediation in protein folding and molecular recognition. *Annu Rev Biophys Biomol Struct.* 2006;35:389-415.
- Lockless SW and Ranganathan R. Evolutionary conserved pathways of energetic connectivity in protein families. *SCIENCE* VOL 286 8 OCTOBER 1999.
- Lockless SW and Ranganathan R. Evolutionary conserved pathways of energetic connectivity in protein families. *SCIENCE* VOL 286 8 OCTOBER 1999.
- Lolis E, Petsko GA. Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5-Å resolution: implications for catalysis. *Biochemistry.* 1990 Jul 17;29(28):6619-25.
- Marks DS, et. al. Protein 3d structure computed from evolutionary sequence variation. *PLOS ONE*, December 2011, Volume 6, Issue 12.
- Marks DS, et. al. Protein 3d structure computed from evolutionary sequence variation. *PLOS ONE*, December 2011, Volume 6, Issue 12.
- Masel Joanna and Trotter Meredith. Robustness and Evolvability. *Cell press review. Trends in genetics* 26 (2010). P. 406-414.
- Mira DL, et al. Statistical coupling analysis-guided library design for discovery of mutant luciferases. *Biochemistry.* 2018 February 06; 57 (5): 663-671.

- Mira DL, et al. Statistical coupling analysis-guided library design for discovery of mutant luciferases. *Biochemistry*. 2018 February 06; 57 (5): 663-671.
- Morcos F. Hwa T. Onuchic N. J. and Weigt M. Direct coupling analysis for protein contact prediction. *Methods in Molecular Biology* (2014) Vol 1137.
- Morcos F., Paganani A., Lunt B. Bertolino A. Direct coupling analysis of residue coevolution captures native contacts across many protein families. *PNAS* (2011) Vol. 108 No. 49. E1293-E1301.
- Mukherjee S, Roychowdhury A, Dutta D, Das AK. Crystal structures of triosephosphate isomerase from methicillin resistant *Staphylococcus aureus* MRSA252 provide structural insights into novel modes of ligand binding and unique conformations of catalytic loop. *Biochimie*. 2012 Dec;94(12):2532-44.
- Muñiz, R. Mecanismos adaptativos en la evolución de la triosafosfato isomerasa. Tesis de maestría. UNAM (2020).
- Nagano, N. Orengo, C. One-Fold with Many Functions: The evolutionary Relationships between TIM Barrel Families Based on their sequences, Structures and Functions. *J. Mol. Biol.* (2002) 321, 741-765.
- Nájera H, Costas M, Fernández-Velasco DA. Thermodynamic characterization of yeast triosephosphate isomerase refolding: insights into the interplay between function and stability as reasons for the oligomeric nature of the enzyme. *Biochem J*. 2003 Mar 15;370(Pt 3):785-92.
- Nielsen Rasmus. Adaptationism - 30 years after Gould and Lewontin. *Evolution*. 2009 Oct;63(10):2487-90.
- Olson ME, Arroyo-Santos A. HOW TO STUDY ADAPTATION (AND WHY TO DO IT THAT WAY). *Q Rev Biol*. 2015 Jun;90(2):167-91.
- Orosz F, Oláh J, Ovádi J. Triosephosphate isomerase deficiency: new insights into an enigmatic disease. *Biochim Biophys Acta*. 2009 Dec;1792(12):1168-74.
- Park SH, Kim HS, Park MS, Moon S, Song MK, Park HS, Hahn H, Kim SJ, Bae E, Kim HJ, Han BW. Structure and Stability of the Dimeric Triosephosphate Isomerase from the Thermophilic Archaeon *Thermoplasma acidophilum*. *PLoS One*. 2015 Dec 28;10(12):e0145331.
- Payne L. Joshua and Wagner Andreas. The causes of evolvability and their evolution. *Nature Reviews*, Volumen 20 (2019)
- Peimbert M, Domínguez-Ramírez L, Fernández-Velasco DA. Hydrophobic repacking of the dimer interface of triosephosphate isomerase by in silico design and directed evolution. *Biochemistry*. 2008 May 20;47(20):5556-64.
- Piovesan D, Minervini G and Tosatto S. The ring 2.0 web server for high quality residue interaction networks. *Nucleic Acids Research*, 2016, Vol 44,
- Piovesan D, Minervini G and Tosatto S. The ring 2.0 web server for high quality residue interaction networks. *Nucleic Acids Research*, 2016, Vol 44.

- Pollock D., Thiltgen G. and Goldstein R. Amino acid coevolution an evolutionary Stokes shift. PNAS (2012) E1352-E1359.
- Razvi, A. Scholtz, M. Lessons in stability from thermophilic proteins. Protein Science (2006) 15:1569-1578.
- Reynolds KA, Pruss WP, Socolich M and Ranganathan R. Evolution-Based design of proteins. Methods in Enzymology, Volume 523.
- Reynolds KA, Pruss WP, Socolich M and Ranganathan R. Evolution-Based design of proteins. Methods in Enzymology, Volume 523.
- Rivoire O, Reynolds and Ranganathan R. Evolution-based functional decomposition of proteins. PLOS Computational Biology | June 2, 2016.
- Rivoire O, Reynolds and Ranganathan R. Evolution-based functional decomposition of proteins. PLOS Computational Biology | June 2, 2016.
- Rodriguez-Larrea D, Minning S, Borchert TV, Sanchez-Ruiz JM. Role of solvation barriers in protein kinetic stability. J Mol Biol. 2006 Jul 14;360(3):715-24.
- Romero R. Sergio, Fernández V. Daniel A., Costas M. Estabilidad Termodinámica de proteínas. Educación Química Vol 29, No 3 (2018).
- Ronimus, R. Morgan, W. Distribution and phylogenies of enzymes of the Embden-Meyerhof-Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. Archaea (2033) Oct; 1 (3): 199-221.
- Rothschild, J. and Mancinelli, L. Life in extreme environments. Nature (2001) Feb 22;409(6823).
- Sanchez-Ruiz JM. Protein kinetic stability. Biophys Chem. 2010 May;148(1-3):1-15.
- Schmid, K. Allers, T. and DiRuggiero, J. SanpShot: Microbial Extremophiles. Cell (2020) Volume 180, Issue 4, 818-818.
- Schramm A, Kohlhoff M, Hensel R. Triose-phosphate isomerase from Pyrococcus woesei and Methanothermus fervidus. Methods Enzymol. 2001; 331:62-77.
- Schramm A, Kohlhoff M, Hensel R. Triose-phosphate isomerase from Pyrococcus woesei and Methanothermus fervidus. Methods Enzymol. 2001;331:62-77.
- Schurig H, Beaucamp N, Ostendorp R, Jaenicke R, Adler E, Knowles JR. Phosphoglycerate kinase and triosephosphate isomerase from the hyperthermophilic bacterium Thermotoga maritima form a covalent bifunctional enzyme complex. EMBO J. 1995 Feb 1;14(3):442-51.
- Shah P. McCandlish M. D. and Plotkin J. Contingency and entrenchment in protein evolution under purifying selection. PNAS (2015). E3226-E3235.
- Sharma P, Guptasarma P. 'Super-perfect' enzymes: Structural stabilities and activities of recombinant triose phosphate isomerases from Pyrococcus

- furiosus and Thermococcus onnurineus produced in Escherichia coli. Biochem Biophys Res Commun. 2015 May 8;460(3):753-8.
- Starr N. Tayler and Thornton W. Joseph. Epistasis in protein evolution. The Protein Society (2016) VOL 25:1204–1218.
 - Sterner Rein H. and Liebl Wolfgang. Thermophilic Adaptation of Proteins. Critical Reviews in Biochemistry and Molecular Biology (2008) 36:1, 39-106.
 - Stetter Karl O. Extremophiles and their adaptation to hot environments. FEBS Letters Volume 452, Issue 1-2, p.22-25.
 - Szilágyi, A. Závodsky, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. Structure 2000, Vol 8 No. 5.
 - Tesileanu T, Colwell LJ and Leibler S. Protein sectors: statistical coupling analysis versus conservation. PLOS Computational Biology. 4 February 27, 2015.
 - Tesileanu T, Colwell LJ and Leibler S. Protein sectors: statistical coupling analysis versus conservation. PLOS Computational Biology. 4 February 27, 2015.
 - Tramontano Anna. The computational prediction of protein assemblies. Current Opinion in Structural Biology 2017, 46:170-175.
 - Tramontano Anna. The computational prediction of protein assemblies. Current Opinion in Structural Biology 2017, 46:170-175.
 - Valdar S. J. William. Scoring residue conservation. Proteins. 2002 Aug 1;48(2):227-41.
 - Wagner Andreas. Robustness and evolvability: a paradox resolved. Proceedings of the royal society (2008) 275, 91-100.
 - Walden H, Bell GS, Russell RJ, Siebers B, Hensel R, Taylor GL. Tiny TIM: a small, tetrameric, hyperthermostable triosephosphate isomerase. J Mol Biol. 2001 Mar 2;306(4):745-57.
 - Walden H, Taylor G, Lilie H, Knura T, Hensel R. Triosephosphate isomerase of the hyperthermophile Thermoproteus tenax: thermostability is not everything. Biochem Soc Trans. 2004 Apr;32(Pt 2):305.
 - Weik M. Low-temperature behavior of water confined by biological macromolecules and its relation to protein dynamics. Eur Phys J E Soft Matter. 2003 Sep;12(1):153-8.
 - Wierenga RK, Noble ME, Davenport RC. Comparison of the refined crystal structures of liganded and unliganded chicken, yeast and trypanosomal triosephosphate isomerase. J Mol Biol. 1992 Apr 20;224(4):1115-26.
 - Wierenga RK, Noble ME, Postma JP, Groendijk H, Kalk KH, Hol WG, Opperdoes FR. The crystal structure of the "open" and the "closed" conformation of the flexible loop of trypanosomal triosephosphate isomerase. Proteins. 1991;10(1):33-49.

- Wierenga, R. The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Letters* 492(2001) 193-198.
- Xiao L. and Honig B. Electrostatic contributions to the stability of hyperthermophilic proteins. *Journal of Molecular Biology* (1999) Jun 25; 289(5): 1435-1444.
- Yumeng Y, Tao H & Huang S. HSYMDOCK: a docking web server for predicting the structure of protein homo-oligomers with Cn or Dn. *Nucleic Acids Research*, 2018, Vol. 46.
- Yumeng Y, Tao H & Huang S. HSYMDOCK: a docking web server for predicting the structure of protein homo-oligomers with Cn or Dn. *Nucleic Acids Research*, 2018, Vol. 46.
- Zhang Z, Sugio S, Komives EA, Liu KD, Knowles JR, Petsko GA, Ringe D. Crystal structure of recombinant chicken triosephosphate isomerase-phosphoglycolohydroxamate complex at 1.8-Å resolution. *Biochemistry*. 1994 Mar 15;33(10):2830-7.