



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

## Maestría y Doctorado en Ciencias Bioquímicas

Curva de estabilidad y diseño de proteínas: barriles TIM *de novo*

TESIS

QUE PARA OPTAR POR EL GRADO DE:  
Maestro en Ciencias

PRESENTA:  
Avilés Sánchez Octavio Martín

TUTOR PRINCIPAL  
Dr. Daniel Alejandro Fernández Velasco  
[Facultad de Medicina](#)

MIEMBROS DEL COMITÉ TUTOR

Dr. Miguel Antonio Costas Basín  
[Facultad de Química](#)

Dr. Rogelio Rodríguez Sotres  
[Facultad de Química](#)

Ciudad Universitaria, Ciudad de México. Agosto 2020



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

## RECONOCIMIENTOS INSTITUCIONALES

Este trabajo se llevó a cabo bajo la dirección del Dr. Daniel Alejandro Fernández Velasco en el Laboratorio de Fisicoquímica e Ingeniería de Proteínas (LFQIP) perteneciente a la Facultad de Medicina, UNAM.

Durante el desarrollo de este proyecto se contó con los siguientes apoyos financieros:

1. Programa de becas para estudios de posgrado (maestría) del Consejo Nacional de Ciencia y Tecnología (CONACyT) 254514 P con número de CVU/Becario: 815923.
2. Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), UNAM-DGAPA, con las claves de proyecto: IN220516 e IN219519.

## RECONOCIMIENTOS ACADÉMICOS

Se reconoce y agradece el trabajo del Comité Tutorial que asesoró el desarrollo de este proyecto, el cual se conformó por:

1. Dr. Miguel Antonio Costas Basín: Facultad de Química, UNAM.
2. Dr. Rogelio Rodríguez Sotres: Facultad de Química, UNAM.
3. Dr. Daniel Alejandro Fernández Velasco: Facultad de Medicina, UNAM.

Se agradece en demasía el apoyo para la revisión del manuscrito y la asistencia del jurado de examen conformado por:

1. Presidente: Dra. Gloria Saab Rincón: Instituto de Biotecnología, UNAM.
2. Secretario: Dr. Alfredo Torres Larios: Instituto de Fisiología Celular, UNAM.
3. Vocal: Enrique García Hernández: Instituto de Química, UNAM.
4. Vocal: Laura Domínguez Dueñas: Facultad de Química, UNAM.
5. Vocal: Marcelino Arciniega Castro: Instituto de Fisiología Celular, UNAM.

El trabajo se desarrolló con el valioso apoyo técnico de los miembros del laboratorio: M. en C. Laura Ileana Álvarez Añorve y la Dra. Isabel Velázquez López.

Los experimentos de Calorimetría Diferencial de Barrido se realizaron gracias al apoyo intelectual y académico del Dr. Miguel Costas Basín de la Facultad de Química, UNAM.

Se agradece a Julio Palacios y Lilia de la Rocha de la Coordinación de Posgrado por el apoyo administrativo brindado en todos los trámites de maestría.

## AGRADECIMIENTOS PERSONALES

**A mi madre Rosalinda.** Tu partida de este plano existencial lejos de ser la experiencia más dolorosa que he experimentado, ha sido el motor que me ha impulsado a convertirme en la mejor versión de mí mismo en todos los aspectos que he elegido como importantes. No tengo palabras para expresar toda la gratitud que tengo y tendré hacia ti. El amor y las enseñanzas que me diste nunca morirán; son y serán el estandarte que portaré con orgullo el resto de mi vida. Todos los logros que realice van dedicados hacia a ti, esperando que desde el plano de conciencia en el que te encuentres te dibujen una sonrisa. Infinita gratitud y amor hacia a ti mamá.

**A mi padre Martín.** Infinitas gracias por el apoyo y las enseñanzas que me has brindado a lo largo de toda mi vida. Gran parte de lo que soy y seré te lo debo a ti. Eres y serás un ejemplo a seguir en todos los aspectos de mi vida. Al igual que a mamá, todos mis logros van dedicados hacia a ti. Gracias por todo papá.

**A mis hermanos.** Soy muy afortunado de tener 3 hermanos y 3 hermanas. Cada uno me ha inspirado de distintas maneras. Mis logros son también suyos. Gracias por todo el apoyo que me han brindado. Ustedes son mi hogar, los amo.

**A Alejandro P. y Sergio R.** Muchas gracias a ambos por el apoyo moral y académico que me han brindado, además de la amistad que tanto aprecio y valoro de ustedes. La sangre nos hace hermanos, pero el fierro, parientes.

**A “The Craps”.** Son los mejores amigos que un sujeto puede tener. Hay tantas anécdotas de todo tipo que hemos vivido y que sería interesante narrar, pero serían más extensas que esta tesis y posiblemente ilegales en algunos países. Siempre pa’lante viejones.

**A las toneladas de música que he escuchado.** Gracias a todos los autores y músicos por plasmar su visión del mundo con el lenguaje universal llamado “música” y acompañar con su arte muchos momentos valiosos de nuestras vidas. *“Music is your only friend until the end...”*  
*Jim Morrison*

**A Carlos Castaneda y Jacobo Grinberg.** Sus libros cambiaron toda mi perspectiva y visión que tengo del mundo. Hubo un antes y un después de leer sus obras, que, para mí son una fuente de sabiduría e inspiración. Gracias por las enseñanzas tan profundas que me han transmitido a través de sus escritos.

*“Sólo como guerrero se puede sobrevivir en el camino del conocimiento” C. C.*

*“Desde el punto de vista lógico, todo puede llevarse al sin sentido. Por ello, la verdadera vida trasciende la razón” J. G.*

## ÍNDICE

RECONOCIMIENTOS INSTITUCIONALES .....	2
RECONOCIMIENTOS ACADÉMICOS .....	2
AGRADECIMIENTOS PERSONALES .....	3
Resumen .....	6
Introducción.....	7
<i>Ingeniería y diseño de proteínas</i> .....	7
<i>El uso de Rosetta en el diseño de proteínas</i> .....	8
<i>Barriles TIM: la historia de sus diseños de novo</i> .....	9
<i>Curvas de estabilidad de proteínas y el diseño de barriles TIM.</i> .....	11
Objetivo .....	13
Objetivos particulares.....	14
Materiales y métodos .....	14
<i>Métodos computacionales</i> .....	14
<i>Sobreexpresión de la proteína</i> .....	14
<i>Purificación de la proteína</i> .....	15
<i>Resuspensión en urea de las proteínas no solubles</i> .....	15
<i>Determinación de la concentración de la proteína</i> .....	16
<i>Mediciones espectroscópicas</i> .....	17
<i>Dicroísmo Circular</i> .....	17
<i>Determinación del Radio de Stokes y peso molecular</i> .....	17
<i>Desplegamiento proteico mediante temperatura</i> .....	17
<i>Dicroísmo Circular</i> .....	17
<i>Calorimetría Diferencial de Barrido</i> .....	18
<i>Desplegamiento proteico mediante urea y cloruro de guanidinio</i> .....	18
<i>Fluorescencia Intrínseca</i> .....	18
<i>Dicroísmo Circular</i> .....	18
<i>Cálculo robusto hecho a partir de los puntajes obtenidos con los softwares ROSETTA y MOPAC</i> .	19
Resultados.....	20
<i>Resultados computacionales</i> .....	20
<i>Sobreexpresión de las proteínas diseñadas</i> .....	26
<i>Radio de Stokes y peso molecular</i> .....	31
<i>Desplegamiento térmico seguido por Dicroísmo Circular</i> .....	34
<i>Desplegamiento térmico seguido por calorimetría diferencial de barrido</i> .....	39

<i>Desplegamiento químico seguido por Dicroísmo circular</i> .....	40
<i>Desplegamiento químico seguido por fluorescencia</i> .....	45
Discusión .....	48
Conclusiones .....	60
Perspectivas .....	60
Bibliografía .....	61
MATERIAL SUPLEMENTARIO.....	66
<i>Ecuaciones de Rosetta</i> .....	66
<i>Comandos del protocolo de diseño</i> .....	68
<i>Tablas de resultados del cálculo no robusto con ROSSETA y el cálculo robusto con ROSETTA y MOPAC</i> .....	69
APÉNDICE .....	76

## Resumen

El primer diseño exitoso de un barril TIM *de novo*, llamado sTIM11, muestra simetría de orden cuatro, contiene dos residuos de cisteína libres y es termoestable ( $T_m = 79.3$  °C); sin embargo, muestra una estabilidad a 25 °C relativamente baja ( $\Delta G_{Unf} = 4.2$  kcal/mol). Posteriormente se diseñó la colección de barriles ReTIM, con el objeto de aumentar la estabilidad y eliminar los grupos tiol de sTIM11. Una de estas proteínas (ReTIM18), muestra una mayor termoestabilidad ( $T_m = 92.9$  °C), sin embargo, su estabilidad a temperatura ambiente ( $\Delta G_{Unf} = 4.1$  kcal/mol) es semejante a la encontrada en sTIM11. La estructura cristalográfica de ReTIM18 muestra diferencias en comparación con el diseño *in silico*. El objetivo de este trabajo es modificar las regiones donde el diseño y la estructura experimental no coinciden. Para ello se diseñaron tres proteínas con simetría de orden 4: 1) ReTIM18 V2R con una mutación puntual en las asas superiores. 2) ReTIM30 con modificaciones en la interfaz entre hélices pares e impares, y 3) ReTIM29, en la cual se reconstruyó un cuarto del barril. Las propiedades hidrodinámicas de las proteínas diseñadas son similares a la forma monomérica plegada de ReTIM18, además el desplegamiento térmico fue irreversible con presencia de estructura residual. ReTIM30 soluble obtuvo una  $T_m = 105.2$  °C, superior a la encontrada en ReTIM18 ( $T_m = 92.9$  °C), a pesar de ello, el  $\Delta H = 0.26$  kcal/mol es muy pequeño comparado a ReTIM18 ( $\Delta H = 76.6$  kcal/mol). También mostró una  $C_m$  en GndCl similar a ReTIM18 (3.16 M y 3.0 M respectivamente) mostrando dos transiciones. El objetivo no se cumplió cabalmente, sin embargo, nos acerca a la comprensión de la correlación energético-estructural que nos permita modificar la curva de estabilidad de los barriles TIM *de novo*.

## Introducción

### *Ingeniería y diseño de proteínas*

La ingeniería de proteínas se define como la manipulación de la secuencia codificante, con la cual se consigue una propiedad funcional o estructural determinada [1]. La máxima expresión de la ingeniería de proteínas es el diseño de enzimas *de novo* y el diseño de proteínas con plegamientos no explorados por la naturaleza. Ambos casos consisten en el diseño y creación de una secuencia *de novo* cuya estructura la haga apta para catalizar una reacción novedosa, o bien adoptar un plegamiento de forma deliberada [1].

Uno de los objetivos de la ingeniería de proteínas es comprender el mecanismo de plegamiento de estos biopolímeros. Para ese propósito, el camino habitual de investigación es abordar el problema de las dos caras de plegamiento de proteínas. Una cara, es la predicción de la estructura de una proteína donde a partir de una secuencia fija de aminoácidos se busca predecir su estructura de mínima energía [3]. La otra cara, es el problema inverso, el cual se conoce como diseño de proteínas. En este caso, no se comienza con una secuencia de aminoácidos que ocurre naturalmente, sino con una estructura diseñada previamente; y tomando el camino inverso al problema de plegamiento, se busca una secuencia de aminoácidos que adoptará la conformación deseada [3]. La ingeniería de proteínas opera de forma iterativa, siguiendo un proceso cíclico que alterna etapas en las que se planean y ejecutan los cambios a realizar con otras en las que se evalúa el resultado de los mismos [1].

Dentro del diseño de proteínas se encuentra una aproximación metodológica llamada “diseño racional”, la cual hace hipótesis razonables sobre cambios concretos que pueden ocasionar un efecto funcional o estructural deseado en la proteína. A partir de la hipótesis, se proponen mutaciones en la secuencia, que dan lugar a sustituciones definidas de aminoácidos [1]. Para ello es necesario poseer gran cantidad de información estructural sobre la molécula de estudio y tomar en cuenta la relación que existe entre secuencia, estructura, estabilidad y función [2]. Una herramienta importante para esta tarea es la cristalografía, que brinda información sobre la ubicación de los aminoácidos en el espacio tridimensional mediante la difracción de rayos X de arreglos cristalinos de proteínas [2]. La determinación de la estructura cristalográfica de muchas proteínas ha permitido que las bases de datos en donde se deposita esta información estén en constante crecimiento [2]. Esta información es valiosa y es empleada en protocolos de diseño basados en inspección visual y datos bioinformáticos recopilados de un conjunto limitado de secuencias naturales. Desde entonces, han surgido poderosos métodos computacionales, que permiten el diseño automatizado del plegamiento de secuencias en una topología deseada [6]. Un ejemplo de ello son las técnicas computacionales para modelado molecular. El modelado molecular es un término general que engloba métodos teóricos y técnicas computacionales para modelar, imitar y predecir el comportamiento de las moléculas [4]. Estos métodos computacionales utilizan una función de búsqueda que puede muestrear rápidamente el espacio conformacional y el espacio de secuencia utilizando una función de energía que puede identificar la secuencia de mínima energía de una proteína [7, 8]. La complejidad del espacio de búsqueda conformacional se puede reducir muestreando solamente los rotámeros más comunes, esto es, conformaciones discretas de la cadena lateral observadas frecuentemente en las estructuras depositadas en las bases de datos [9, 10, 11]. Entre los programas que implementan este enfoque se encuentra Rosetta, que se ha aplicado con éxito a una amplia variedad de problemas de diseño [12].



Los métodos computacionales han madurado a tal punto que ahora es posible producir proteínas no existentes en la naturaleza (proteínas diseñadas *de novo*) con funciones novedosas [4], y predecir los efectos sobre la estructura y la estabilidad causados por mutaciones en sitios claves relacionados con la estructura o función de una proteína [5]. A pesar de estos logros, aún hay un largo camino por recorrer, particularmente en la optimización de diseños de enzimas *de novo* con funciones catalíticas altamente eficientes.

### *El uso de Rosetta en el diseño de proteínas*

ROSETTA es un paquete de software para predicción de estructuras de proteínas y diseño funcional [13]. Para cumplir con estas funciones, ROSETTA debe realizar dos tareas:

A) Muestrear el espacio conformacional relevante, y en el caso del diseño, el espacio de secuencia.

B) Evaluar la energía de los modelos estructurales resultantes [13].

Para muestrear el espacio conformacional, ROSETTA lleva a cabo dos estrategias:

1) Muestreo de los grados de libertad del esqueleto: En lugar de hacer un muestreo completo del esqueleto, lo divide en muestreos locales de nueve o tres residuos. Las conformaciones de péptidos se recolectan del PDB para fragmentos homólogos de la secuencia [14] que capturan la fracción estructural de la secuencia local [15]. Para el refinamiento local de los modelos, ROSETTA utiliza el muestreo Metrópolis Monte Carlo de ángulos  $\phi$  y  $\psi$  que se calculan para no perturbar el plegamiento global de la proteína. Rohl et al. [16] proporcionan una revisión de los algoritmos de selección de fragmentos y refinamiento del esqueleto en ROSETTA.

2) Muestreo de los grados de libertad de las cadenas laterales: ROSETTA reduce drásticamente el número de conformaciones muestreadas mediante el uso de conformaciones discretas de cadenas laterales observadas en el PDB [17, 11] Estos rotámeros obtenidos de la librería Dunbrack capturan combinaciones permitidas entre los ángulos de torsión de la cadena lateral, así como de los ángulos  $\phi$  y  $\psi$  del esqueleto, por lo tanto, se reduce la extensión del espacio conformacional [11]. Se utiliza una simulación de Metrópolis Monte Carlo para buscar la combinación de rotámeros que ocupan el mínimo global en la función de energía [17, 18].

Para evaluar la energía de los modelos resultantes, al utilizar la función de diseño, ROSETTA toma en cuenta la energía de interacción de todos los átomos [13], la cual se desglosa en los siguientes términos: preferencias de torsión de Ramachandran, interacciones de Lennard-Jones, puentes de hidrógeno, solvatación, interacciones de pares de residuos (electrostáticas, disulfuros), energía propia del rotámero y energía de referencia del estado desplegado (las ecuaciones que describen cada término se encuentran en el Material Suplementario).

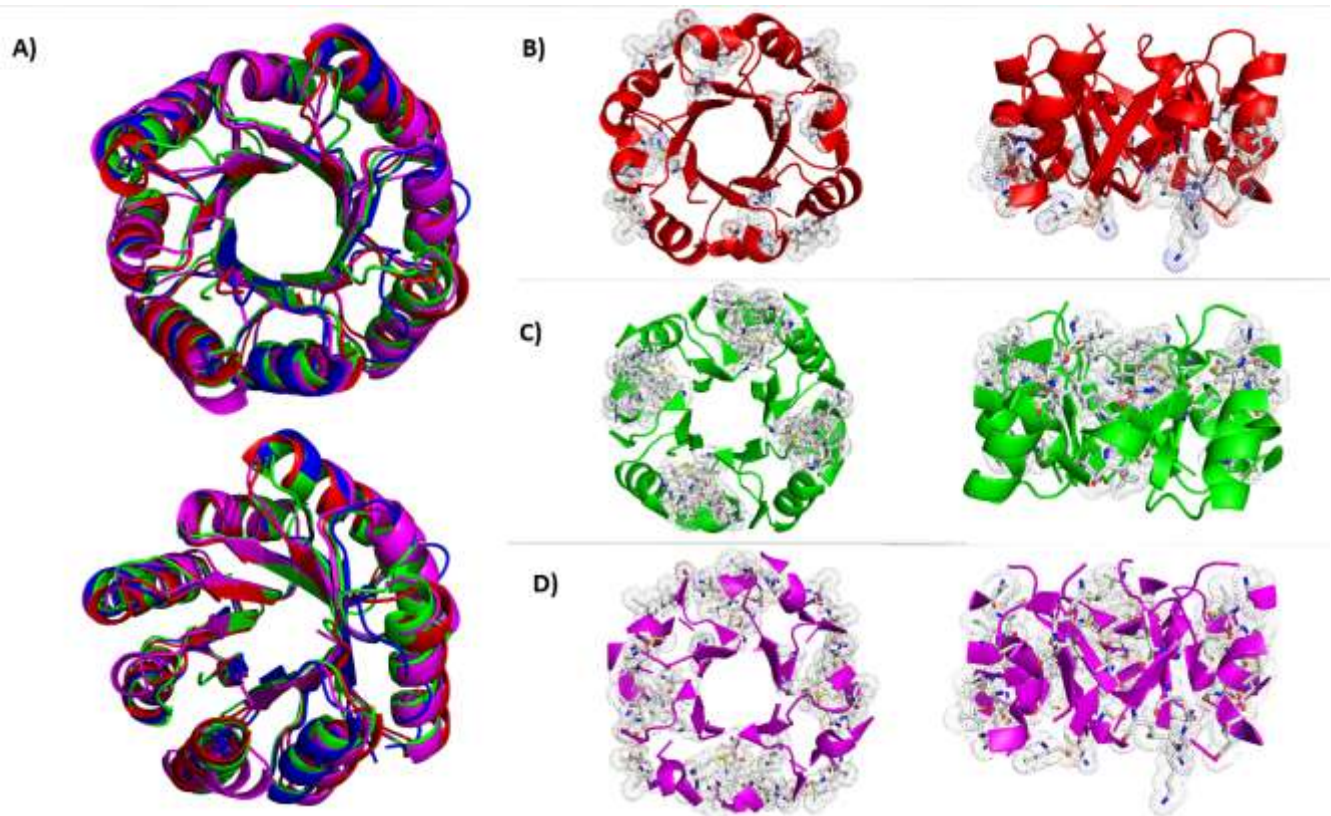
Tomando en cuenta todas las funciones antes mencionadas, el algoritmo ROSETTADesign (algoritmo de ROSETTA empleado para el diseño) realiza un proceso iterativo que optimiza energéticamente la estructura y la secuencia de una proteína, luego alterna rondas de optimización de secuencia de esqueleto fijo y minimización de la energía del esqueleto flexible [6]. Durante la etapa de optimización de secuencia, se utiliza una búsqueda por el método de Monte Carlo para muestrear el espacio de secuencia. Cada aminoácido se considera en cada posición en la secuencia, y los rotámeros están restringidos a la biblioteca Dunbrack [11]. Después de cada ronda de optimización de la secuencia de Monte Carlo, el esqueleto polipeptídico se relaja para acomodar los aminoácidos diseñados.

A pesar de que el diseño de proteínas *de novo* sigue siendo un problema desafiante, se han logrado crear diseños útiles para la obtención de información biológica, ya sea adquiriendo conocimiento de la ubicación del sitio funcional a partir de estructuras predichas de baja resolución o describiendo la estructura del sitio funcional de interés [16]. Entre los diseños que se han realizado con ROSETTA se encuentran: Motivos  $\beta/\alpha$ , proteínas con simetría interna y péptidos restringidos hiperestables *de novo* [19], proteínas de membrana [20] y nanopartículas [21]. En el 2018 en el laboratorio de D. Baker en la Universidad de Washington fueron diseñadas proteínas que se ensamblan en forma de filamentos de escala micrométrica con una amplia gama de geometrías *in vivo* e *in vitro*. Las estructuras de criomicroscopía electrónica de seis diseños son similares a los modelos de diseño computacional [35]. También fueron diseñadas las estructuras *de novo* de “jellyroll” con hélices  $\beta$  de doble cadena formadas por ocho cadenas  $\beta$  antiparalelas. La estructura determinada por resonancia magnética nuclear de uno de los diseños hipertermoestables es muy similar al modelo computacional [36]. Recientemente fueron publicados los diseños de moléculas miméticas de la citoquina inmune central interleucina-2 (IL-2). Los diseños son hiperestables, con estructuras cristalográficas muy parecidas a los modelos diseñados tanto solas como en complejo y con mayor afinidad que las citoquinas naturales [37].

### *Barriles TIM: la historia de sus diseños de novo*

La triosafosfato isomerasa (TIM) fue la primera proteína en la que se reportó la estructura de barril  $(\beta/\alpha)_8$ ; [22] esta estructura cuenta con ocho hebras  $\beta$  rodeadas por ocho  $\alpha$  hélices paralelas ordenadas en una conformación toroide. 33 superfamilias de barriles TIM cubren cinco de las siete clases de reacciones de la Comisión Enzimática. Este patrón de plegamiento ha sido un objeto atractivo para el diseño de catalizadores novedosos ya que posee la topología más común para las enzimas, además de ser estructural y funcionalmente diverso. Desde hace más de 25 años, se han realizado esfuerzos para lograr diseñar el plegamiento barril TIM  $(\beta/\alpha)_8$  *de novo*; sin embargo, este problema ha mostrado ser un reto de diseño muy desafiante. Entre los primeros intentos reportados están los barriles TIM's llamados “Heptarelina” [23], “Octarelina” [23] y “Nonarelina” [23] que contenían siete, ocho y nueve unidades repetitivas con el patrón de plegamiento “giro 1/hebra  $\beta$ /giro 2/hélice  $\alpha$ ”, es decir, mantenían una simetría de orden 7, 8 y 9 respectivamente. El diseño más “exitoso” fue la Octarelina, retomada en diseños posteriores, formada por ocho unidades con el patrón de plegamiento “giro 1/hebra  $\beta$ /giro 2/hélice  $\alpha$ ”, descartando así las simetrías de orden 7 y 9 para diseños futuros. Uno de los problemas de este diseño es que la proteína correspondiente no es soluble y se localiza en cuerpos de inclusión, lo cual dificultaba su caracterización fisicoquímica [23]. Años más tarde, fueron creadas las sucesoras de la Octarelina: la Octarelina II [24] que poseía una simetría de orden 8 (esto es, contiene ocho repeticiones del motivo fundamental en su secuencia) y la Octarelina III [24] que poseía una simetría de orden 4. La simetría de orden 4 fue introducida originalmente para aumentar la solubilidad y el empaquetamiento de la Octarelina, ya que, a pesar de revelar un alto grado de estructura secundaria y la capacidad de adoptar una estructura compacta tras la renaturalización, su solubilidad y empaquetamiento necesitaban ser mejorados [24]. La Octarelina III era más estable que la Octarelina II, además mostraba un desplegamiento cooperativo con urea, lo cual era un indicativo de que la simetría de orden 4 ayudaba a mantener la estabilidad de los diseños. A pesar de ello, estos diseños no adoptaban una estructura terciaria compacta, sino que se comportaban como glóbulos fundidos [24]. A partir de aquí, diseños posteriores adoptaron la simetría de orden 4, la cual ayudó al aumento de su estabilidad térmica, un ejemplo de ello fue el barril TIM creado por F. Offredi y su equipo de trabajo en 2003 que

poseía una  $T_m = 65^\circ\text{C}$  [25]. No fue sino hasta 2015 cuando se realizó el diseño computacional y la determinación estructural de sTIM11 [26], un barril TIM diseñado *de novo* con simetría de orden cuatro, que demostró ser termoestable ( $T_m = 79.3^\circ\text{C}$ ). El inconveniente de esta proteína es que su estabilidad a temperatura ambiente es relativamente baja ( $\Delta G_{Unf} = 4.2$  kcal/mol) y posee dos cisteínas libres que pueden formar interacciones no deseadas. Para solucionar este último problema, en el laboratorio de B. Höcker en la Universidad de Bayreuth fue creada la proteína sTIM11-SS (sTIM11 sin cisteínas, esto es sTIMC8Q/C18V) que, si bien resolvía el problema de las dos cisteínas, mostró una gran disminución en la estabilidad de la proteína ( $\Delta G_{Unf} = 3.2$  kcal/mol y  $T_m = 65.6^\circ\text{C}$ ). Con el fin de aumentar la estabilidad, se diseñaron la serie de barriles ReTIM (1 – 28) [27] entre las cuales se tomó como punto de partida para este trabajo la proteína con simetría de orden 4 ReTIM18. Al comparar con sTIM11-SS, ReTIM18 presenta un 76% de identidad, esto es, 44 diferencias, localizadas en dos regiones de la estructura tridimensional. Cabe mencionar que ReTIM18 se obtuvo de la combinación de las proteínas ReTIM6 y ReTIM8, con 24 y 28 mutaciones respectivamente, con respecto a sTIM11-SS (Figura 1).



**Figura 1. Diferencias estructurales entre sTIM11-SS y ReTIM18 (ReTIM6 + ReTIM8).** A) Alineamiento estructural múltiple de los modelos estructurales de sTIM11-SS (en color azul), ReTIM18 (magenta), ReTIM6 (rojo) y ReTIM8 (verde). En los incisos B) ReTIM6 (rojo), C) ReTIM8 (verde) y D) ReTIM18 (magenta), los residuos en color blanco y con el código de color HNOs representan a los residuos diferentes con respecto al modelo original (sTIM11-SS). En las imágenes de la izquierda, los barriles se observan desde los extremos carboxilo terminal de las hebras  $\beta$ . En las imágenes de la derecha, los barriles se observan de forma lateral, con los extremos carboxilos terminales de las hebras  $\beta$  apuntando hacia el norte. Ninguno de estos modelos tiene un código en el PDB.

ReTIM18 muestra un valor relativamente alto de  $T_m$  ( $92.9^\circ\text{C}$ ), pero una baja estabilidad a temperatura ambiente, similar a la encontrada en sTIM11 ( $\Delta G_{Unf} = 4.1$  kcal/mol). Esta proteína era la única mutante diseñada para la cual se poseía una estructura cristalográfica al iniciar este estudio, lo que proporciona valiosa información estructural, útil para generar nuevos diseños *in silico*. Es importante recalcar que ni ReTIM18, ni sus antecesoras (sTIM11-SS y

sTIM11), muestran coincidencia estructural absoluta entre el esqueleto del barril diseñado y el observado experimentalmente (en la sección de resultados se detallarán dichas diferencias), lo que da pie a diseñar cambios estructurales que disminuyan la diferencia entre el diseño y la estructura cristalográfica con el último fin de mejorar sus propiedades fisicoquímicas. Las desventajas que posee ReTIM18 son: una rápida agregación luego de ser purificada, es térmicamente irreversible, ya que al realizar un segundo barrido (“rescan”), no se obtiene una endoterma igual a la del primer barrido (datos no mostrados), además de que su  $\Delta H$  tiende a hacerse pequeño luego de un tiempo prolongado de almacenamiento, lo cual sugiere que la fracción de proteína plegada disminuye con el tiempo. En teoría, dichos problemas deben ser resueltos con modificaciones estructurales que favorezcan nuevas interacciones electrostáticas entre los residuos que conforman la proteína, estabilizando al estado nativo y/o desestabilizando conformaciones alternas.

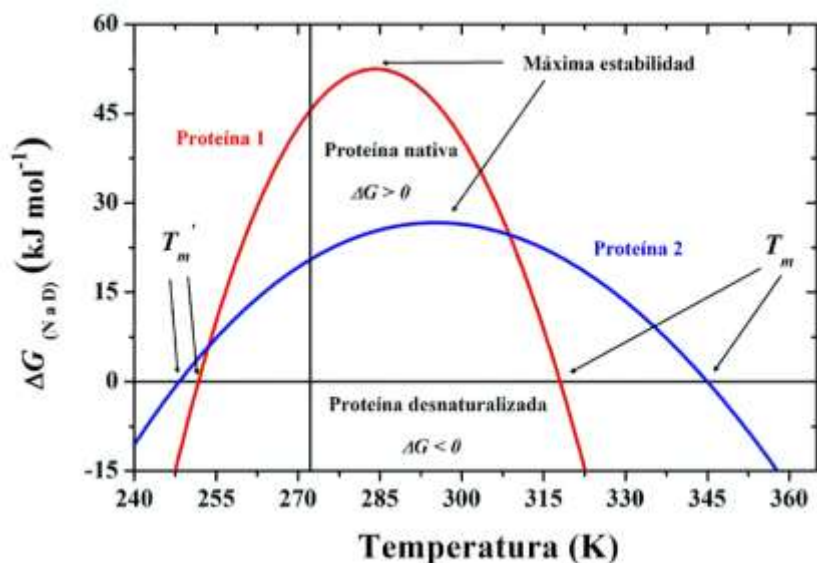
### *Curvas de estabilidad de proteínas y el diseño de barriles TIM.*

La termodinámica proporciona un criterio claro para determinar si un proceso es o no espontáneo: si el cambio de energía libre de Gibbs es negativo ( $\Delta G < 0$ ), el proceso ocurre de manera espontánea. Una proteína es termodinámicamente estable cuando el cambio de energía libre de Gibbs para el proceso de desnaturalización ( $\Delta G_{Unf}$ ) es positivo y, por tanto, desfavorable (cuando el estado inicial es el nativo,  $\Delta G_{Unf} = G_{Final} - G_{Inicial} = G_D - G_N$ ). Aquí se ha supuesto que la proteína sólo puede estar en dos estados (N y D), sin que existan estados intermediarios significativamente poblados. Cuando en el proceso de desnaturalización de una proteína se cumple que  $\Delta G_{Unf} > 0$  a ciertas condiciones de temperatura, presión, concentración y pH se dice que esta proteína es termodinámicamente estable. El andamiaje teórico de la termodinámica clásica proporciona una ecuación para calcular la dependencia del  $\Delta G$  con la temperatura ( $\Delta G(T)$ ) manteniendo la presión constante, dicha ecuación se conoce como la ecuación de Gibbs-Helmholtz y es la siguiente:

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

donde  $\Delta H$  y  $\Delta C_p$  son los cambios de entalpía y de capacidad calorífica a presión constante para el proceso de desnaturalización (N  $\rightarrow$  D) y  $T_m$  es la temperatura de referencia, en la cual la concentración de N y D es la misma (por lo tanto,  $\Delta G(T_m) = 0$ ). A la representación gráfica de  $\Delta G$  vs  $T$  empleando la ecuación de Gibbs-Helmholtz, se le conoce como curva de estabilidad termodinámica.





**Figura 2. Curvas de estabilidad para dos proteínas hipotéticas.** Esta curva se obtiene empleando la ecuación de Gibbs-Helmholtz con los valores de  $\Delta H$ ,  $\Delta C_p$  y  $T_m$  obtenidos de experimentos de Calorimetría Diferencial de Barrido. La curva de estabilidad es diferente para cada proteína, y puede considerarse como “una huella termodinámica” que la distingue de todas las demás [42].

La curva de estabilidad en la Figura 2 delimita el intervalo de temperatura en el cual el estado nativo es estable; esto ocurre a las temperaturas donde  $\Delta G > 0$  y por tanto la desnaturalización no es espontánea. La temperatura a la cual el valor de  $\Delta G$  es máximo se conoce como la temperatura de máxima estabilidad. En las regiones donde  $\Delta G < 0$ , la desnaturalización es espontánea y el estado estable es el desnaturado. La Figura 2 también muestra que hay dos temperaturas ( $T_m$  y  $T_m'$ ) a las cuales  $\Delta G = 0$ . Coloquialmente, la  $T_m$  se conoce como la temperatura de desnaturalización “por calor”, mientras que  $T_m'$  representa la temperatura de desnaturalización “por frío” [42]. La termoestabilidad está relacionada con el valor de  $T_m$ , mientras que la estabilidad a temperatura ambiente está dada por el valor de  $\Delta G$  a 25 °C. Hay varias formas en las que se puede alterar una curva de estabilidad de la proteína para dar como resultado una  $T_m$  más grande. Un aumento en el número de interacciones entálpicas elevará la curva y hará que la proteína sea más estable a cualquier temperatura. Alternativamente, una disminución del  $\Delta C_p^\circ$  produce una curva “más plana”, que resulta en una  $T_m$  mayor, tal como se observa en la curva de la proteína 2 de la Figura 2 [55].

Al comparar los valores de  $T_m$  y estabilidad a temperatura ambiente ( $\Delta G$ ) de los diseños ReTIM y sTIM11 contra algunos de los barriles TIM naturales como  $\alpha$  – amilasas y TIM bacterianas, encontramos que los diseños presentan valores altos de  $T_m$  y valores pequeños de  $\Delta G$  en comparación con los barriles TIM naturales. Dado que, en la naturaleza existen barriles TIM con valores altos de  $\Delta G$  a temperatura ambiente, hipotetizamos que es posible mejorar la estabilidad a temperatura ambiente de los barriles TIM diseñados *de novo* sin disminuir significativamente su termoestabilidad. De acuerdo a las ecuaciones paramétricas el  $\Delta H$  promedio calculado para una proteína de 184 residuos es de  $\Delta H = 128.4 \pm 3.5$  kcal/mol [28], mientras que ReTIM18 muestra un  $\Delta H = 76.6$  kcal/mol. Ya que la magnitud del  $\Delta H$  está determinada por las interacciones intramoleculares favorables que se rompen cuando la proteína se despliega, en principio, es posible obtener diseños que presenten un número mayor de estas interacciones y por lo tanto una mayor estabilidad. Tomando como base la proteína ReTIM18, analizamos las diferencias estructurales entre el diseño *in silico* y su estructura cristalográfica, para proponer cambios estructurales que minimicen estas

diferencias y aumenten el número de interacciones favorables que, en consecuencia, mejoren su estabilidad a temperatura ambiente.

Hasta ahora no existen referencias en las que, de forma directa, se haya buscado mejorar el  $\Delta G$  a temperatura ambiente sin alterar de forma significativa la termoestabilidad de una proteína, sin embargo, algunos autores han logrado resultados interesantes al respecto. Tal es el caso del trabajo de E. Lamazares [43], donde se estudió a una apoflavodoxina, la cual tiene un desplegamiento térmico de tres estados. Con anterioridad, se habían determinado experimentalmente dos zonas con una estabilidad distinta; una zona con estabilidad relevante ( $T_1$ ) y otra zona de estabilidad residual ( $T_2$ ). En ambas regiones se hicieron mutaciones combinando distintos enfoques de diseño racional. El resultado interesante fue que, en varias mutantes, se logró pasar de un desplegamiento térmico de tres estados a uno de dos estados, con un aumento en la termoestabilidad, sin embargo, la estabilidad química no aumentó, incluso, disminuyó [43]. Si bien, por lo general sabemos que los aumentos en  $T_m$  están asociados con aumentos en la energía libre de estabilidad máxima  $\Delta G(T^*)$  [57], el trabajo de E. Lamazares muestra que esta correlación no siempre es concordante, ya que los resultados que obtuvo señalan que una proteína térmicamente estable no necesariamente es químicamente estable; además ilustra que una proteína con un desplegamiento químico de dos estados, no necesariamente tiene un desplegamiento térmico de dos estados. Estas dos ideas ilustran la gran dificultad que hay en aumentar exitosamente el  $\Delta G$  a temperatura ambiente sin alterar de forma significativa la termoestabilidad y viceversa.

Otra de las grandes dificultades encontradas en el “diseño racional” es que a pesar de que el diseño computacional es muy adecuado para identificar posiciones con potencial de estabilización, sigue siendo difícil calcular las contribuciones de  $\Delta G$  debidas a cada una de las interacciones que estabilizan a las proteínas [44]. Además, aunque es claro que el desplegamiento de las proteínas naturales es cooperativo, no es claro de qué manera la autoorganización o cooperatividad en el plegamiento de la cadena polipeptídica puede elevar la temperatura de desnaturalización [45].

A pesar de las dificultades expuestas anteriormente, se han propuesto algunas reglas para la estabilización de proteínas, como la reducción de cavidades hidrofóbicas [58], [59], generación de puentes de hidrógeno [60], [61], [62], puentes salinos [63], [64] y puentes disulfuro [65], [66]. También existen estrategias más específicas como en el trabajo de Menéndez-Arias y Argos [46], donde al comparar seis familias de proteínas diferentes de microorganismos mesófilos y termófilos, se encontró que, para aumentar la estabilidad térmica, en las regiones  $\alpha$ -helicoidales, se debe disminuir la flexibilidad y aumentar la hidrofobicidad. Algunos ejemplos de sustituciones que presenta son las siguientes: Gly a Ala, Ser a Ala, Lys a Arg, Lys a Ala.

A pesar de los grandes avances que se han logrado, aún no son claras todas las pautas que guían la correlación energético-estructural en la estabilidad de un sistema proteico.

## **Objetivo**

Modificar la curva de estabilidad de los barriles TIM diseñados *de novo* y comprender la correlación energético-estructural de los cambios observados.

## Objetivos particulares

- Tomando como base la proteína ReTIM18; analizar las diferencias estructurales entre el diseño *in silico* y su estructura cristalográfica.
- Diseñar computacionalmente tres proteínas en las cuales se modifique la secuencia en los sitios donde existen diferencias estructurales significativas entre el diseño *in silico* y la estructura cristalográfica de ReTIM18, con el fin de aumentar el número de interacciones favorables que, en consecuencia, mejoren su estabilidad a temperatura ambiente.
- Purificar y caracterizar fisicoquímicamente las proteínas diseñadas con el objetivo de analizar el impacto en la estabilidad de las modificaciones propuestas.

## Materiales y métodos

### *Métodos computacionales*

En este trabajo fue empleada la versión Rosetta 2017.08.59291. La estructura de referencia fue la estructura cristalográfica de ReTIM18 determinada en el trabajo de Romero Romero (2017) [27], cabe destacar que la estructura aún no es pública, por lo que no tiene un código asignado en el PDB. Los datos de colecta y refinamiento de la estructura se encuentran en la Tabla 14 del material suplementario. El protocolo de diseño fue el siguiente:

- 1) Se realizó una ronda de limpieza de la estructura.
- 2) Se realizó una relajación a la estructura limpia.
- 3) Se aplicó el algoritmo de diseño a la estructura limpia y relajada. Los parámetros con los cuales se usó el algoritmo Rosetta Design son los siguientes: Para mutaciones puntuales y para explorar las posibilidades de encontrar un residuo de mínima energía, se pidieron 100 estructuras de salida, también se permitió el rearrreglo de toda la cadena, al igual que el uso de la conformación de la cadena lateral nativa, con una y dos muestras discretas adicionales de rotámeros de la librería Dunbrack. Para explorar las posibilidades de encontrar un residuo de mínima energía, se permitieron sólo aminoácidos naturales excluyendo a la cisteína.
- 4) Se hizo otra ronda de relajación con el fin de minimizar la estructura con las mutaciones hechas en el diseño.

Todos los comandos empleados en el protocolo de diseño se encuentran en el material suplementario. Los puntajes arrojados por Rosetta se tomaron como criterio de selección de modelos exitosos, es decir, los puntajes más negativos que la estructura de referencia (ReTIM18), fueron seleccionados como modelos exitosos. Las Tablas 11, 12 y 13 del material suplementario contienen los puntajes que sirvieron para este propósito.

### *Sobreexpresión de la proteína*

Las secuencias de los diseños seleccionados se sintetizaron comercialmente en la empresa Genscript, con uso de codones optimizados para *E. coli* y subclonadas en un vector pET-28b(+) con sitios de restricción NdeI (CATATG) en el extremo 5' y XhoI (TAACTCGAG) en el extremo 3' y cola de 6 histidinas en la región C-terminal. Los genes subclonados se transformaron en células calcio competente de la cepa de expresión *E. coli* BL21(DE3)pLysS mediante choque térmico. Las cepas se ensayaron con protocolos de sobreexpresión estándares que consistieron en crecer 10 mL de medio LB + 10 µL de kanamicina (a una concentración de 30 mg/mL) con el inóculo de colonia de la cepa transformada, se incubó una noche a 37 °C con agitación a 250 rpm. Después se diluyó el cultivo 100 veces en medio LB + 100 µL de

kanamicina (a una concentración de 30 mg/mL) y se creció a 37 °C con agitación a 250 rpm hasta alcanzar una  $DO_{600}$  de ~ 0.5 – 0.7. El cultivo resultante se indujo a una concentración de 1 mM de IPTG y se incubó durante 4 horas a 37 °C y 250 rpm. Se centrifugó el medio a 8000 rpm por 10 min a 4 °C y las células fueron resuspendidas en solución amortiguadora A ( $NaH_2PO_4$  35 mM, Imidazol 35 mM, NaCl 300 mM, pH =8) para sonicarlas por 3 ciclos de 30 segundos con descansos de 30 segundos y finalmente se centrifugaron a 10000 rpm por 10 min a 4 °C para recuperar la proteína ubicada en el sobrenadante o bien en el pellet.

La sobreexpresión se verificó en geles de poliacrilamida al 12% en condiciones desnaturalizantes teñidos con “mezcla mágica” (0.6 mL de SDS al 10%, 60  $\mu$ L de  $\beta$  mercaptoetanol, 48  $\mu$ L de EDTA 250 mM, 0.6 mL de glicerol, 0.3 mL de Tris HCl 1M a pH = 6.8, aforo de agua a un volumen de 1.5 mL y una pizca de azul de bromofenol). Se cargó el sobrenadante y la pastilla para verificar si la proteína se encontraba en la fracción soluble o insoluble. Al inicio se realizaron pruebas de sobreexpresión en volúmenes pequeños para verificar la sobreexpresión de la proteína. Se mostrará el gel de la primera sobreexpresión en la sección de resultados.

### *Purificación de la proteína*

Para la purificación de las proteínas se utilizó una columna de afinidad His Trap™ HP de 5mL (GE Healthcare Life Sciences ®) acoplada a un sistema FPLC (GE Healthcare Life Sciences ®); sin remoción de cola de histidinas. Cabe destacar que este protocolo se llevó a cabo sólo para las proteínas solubles.

Los amortiguadores empleados fueron los siguientes:

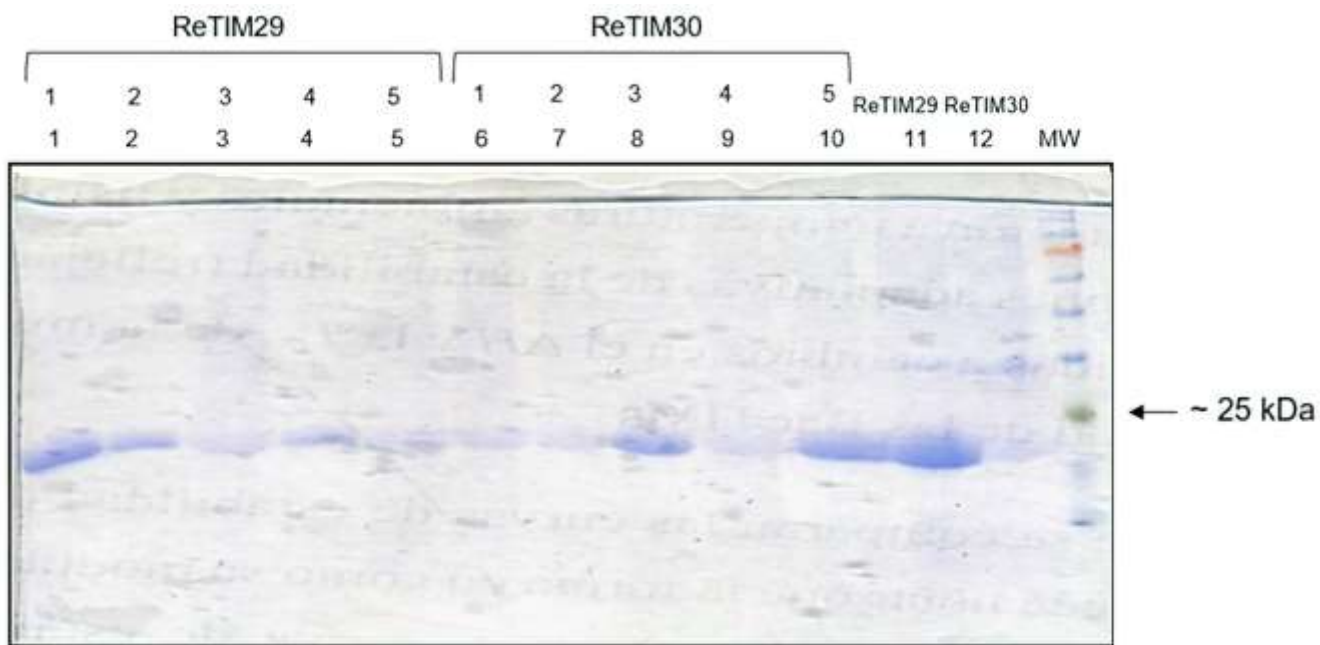
Amortiguador A:  $NaH_2PO_4$  35 mM, Imidazol 35 mM, NaCl 300 mM, pH 8.0

Amortiguador B:  $NaH_2PO_4$  35 mM, Imidazol 500 mM, NaCl 300 mM, pH 8.0

### *Resuspensión en urea de las proteínas no solubles*

Las proteínas localizadas en cuerpos de inclusión fueron lavadas con una solución de 50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA y pH 8.0, y centrifugadas a 8000 rpm por 10 min a 4 °C. Este procedimiento se realizó dos veces, posteriormente las proteínas fueron lavadas con una solución de 50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, 2% Tritón x 100 y pH 8.0, y centrifugadas a 8000 rpm por 10 min a 4 °C. Se retiró el sobrenadante y se hizo una resuspensión en 8 M de urea con agitación constante por 4 – 5 h. Finalmente se dializaron las proteínas en una relación 1:100 con concentraciones gradualmente inferiores de urea hasta retirarla en su totalidad. El amortiguador empleado para la diálisis fue: 50 mM  $NaH_2PO_4$ , 300 mM NaCl, 2 mM DDT y 0 M, 2 M, 4 M y 6 M de urea respectivamente.





**Figura 3. Gel de las proteínas ReTIM29 y ReTIM30 después de ser resuspendidas en urea y dializadas hasta 0 M de urea en las 5 condiciones.** Del carril 1 al 5, se presenta la proteína ReTIM29 en las 5 condiciones de diálisis; el carril 1 corresponde a la condición de diálisis 1 y así sucesivamente (ver condiciones de diálisis en la Figura 15). Del carril 6 al 10, se presenta la proteína ReTIM30 en las 5 condiciones de diálisis; el carril 6 corresponde a la condición de diálisis 1 y así sucesivamente (Figura 15). Los carriles 11 y 12 corresponden a las proteínas ReTIM29 y ReTIM30 después de los lavados con las soluciones 50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, pH 8.0 y 50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, 2% Tritón x 100, pH 8.0; y resuspendidas en 8 M de urea. El carril 13 corresponde a los marcadores de peso molecular.

#### *Determinación de la concentración de la proteína*

La concentración de proteína se determinó por su absorbancia a 280 nm. El coeficiente de extinción molar empleado para la medición fue un valor teórico obtenido a partir de la secuencia aminoacídica, mediante el servidor ProtParam [29]. Los valores teóricos de los pesos moleculares se obtuvieron mediante el servidor ExPASy [29].

Los coeficientes de extinción molar y pesos moleculares fueron los siguientes:

Coeficientes de extinción molar (a 280 nm calculados en agua):

$$\text{ReTIM18 V2R} = 22000 \text{ M}^{-1} \text{ cm}^{-1}$$

$$\text{ReTIM29} = 71960 \text{ M}^{-1} \text{ cm}^{-1}$$

$$\text{ReTIM30} = 22000 \text{ M}^{-1} \text{ cm}^{-1}$$

Pesos moleculares:

$$\text{ReTIM18 V2R} = 21.054 \text{ kDa}$$

$$\text{ReTIM29} = 21.086 \text{ kDa}$$

$$\text{ReTIM30} = 20.614 \text{ kDa}$$

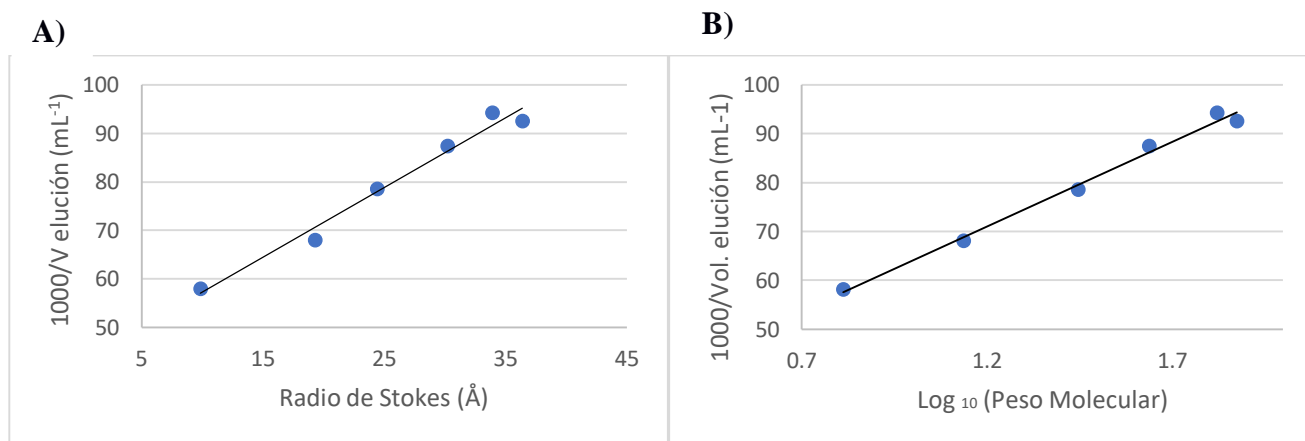
## Mediciones espectroscópicas

### Dicroísmo Circular

Los ensayos de Dicroísmo Circular se realizaron en un espectropolarímetro Chirascan™ (Applied Photophysics®) acoplado a un sistema Peltier (PolyScience®) para el control de temperatura. El espectro fue registrado de 190 a 260 nm. Previo al experimento, la proteína fue dializada en amortiguador NaH<sub>2</sub>PO<sub>4</sub> 10 mM, pH 8.0 en una relación 1:1000 durante 3 y 12 horas; y llevada a una concentración de 0.4 mg/mL. Los espectros fueron medidos a una temperatura de 25 °C.

### Determinación del Radio de Stokes y peso molecular

El Radio de Stokes y los pesos moleculares se determinaron mediante cromatografía de exclusión molecular con una columna analítica Superdex 75 HR 10/30 (GE Healthcare Life Sciences®) acoplada a un sistema FPLC (GE Healthcare Life Sciences®). Previamente se realizó una curva de calibración con las siguientes proteínas: albúmina bovina (Peso molecular (PM): 66.3 kDa, Radio de Stokes (R<sub>s</sub>): 33.9 Å), ovoalbúmina (PM: 43.5 kDa, R<sub>s</sub>: 30.2 Å), anhidrasa carbónica (PM: 28 kDa, R<sub>s</sub>: 24.4 Å), aprotitina (PM: 6.5 kDa, R<sub>s</sub>: 9.83 Å), ribonucleasa (PM: 13.7 kDa, R<sub>s</sub>: 19.3 Å) y conalbúmina (PM: 75 kDa, R<sub>s</sub>: 36.4 Å). Se graficó: 1000/volumen de elución contra el Radio de Stokes de las proteínas estándar de calibración (Figura 4A); los datos se ajustaron a una regresión lineal y a partir de éstas se obtuvieron los valores necesarios para la determinación del Radio de Stokes. El mismo procedimiento se realizó en la curva: 1000/volumen de elución contra el Log<sub>10</sub> (peso molecular) para determinar los pesos moleculares, como si se trataran de proteínas globulares (Figura 4B).



**Figura 4. Curvas de calibración de las proteínas estándar usadas para la determinación del Radio de Stokes y peso molecular de las proteínas diseñadas. A)** Dependencia de la tasa de migración (1000/V elución mL<sup>-1</sup>) de las proteínas estándares de calibración contra su Radio de Stokes (Å). Valor de R<sup>2</sup> = 0.978. Curva de ajuste:  $y = 1.4335x + 42.998$ . **B)** Dependencia de la tasa de migración (1000/V elución mL<sup>-1</sup>) de las proteínas estándares de calibración contra el logaritmo base 10 del peso molecular. Valor de R<sup>2</sup> = 0.991. Curva de ajuste:  $y = 34.709x + 29.285$ .

## Desplegamiento proteico mediante temperatura

### Dicroísmo Circular

La desnaturalización por temperatura se realizó en un espectropolarímetro Chirascan™ (Applied Photophysics®) acoplado a un sistema Peltier (PolyScience®) para el control de temperatura. Previo a la desnaturalización se obtuvo un espectro de la proteína de 190-260 nm de longitud de onda a 20 °C y una concentración de 0.4 mg/mL. Para minimizar la agregación se diluyó la proteína a 0.2 mg/mL y se monitoreó el cambio de señal de elipticidad molar a 208 y 220 nm en función del cambio de temperatura (de 20 °C a 90 °C) a una tasa de

1 °C/min. Después se mantuvo la temperatura a 90 °C y se tomó un espectro de 190 a 260 nm. Posteriormente se dejó enfriar la proteína a una tasa de 1 °C/min hasta alcanzar 20 °C y de nuevo se tomó un espectro de 190-260 nm de longitud de onda luego de dejarla 2 minutos en equilibrio. Los experimentos se hicieron en amortiguador NaH<sub>2</sub>PO<sub>4</sub> 10 mM, pH 8.0.

#### *Calorimetría Diferencial de Barrido*

También se siguió el proceso de desplegamiento térmico por DSC en el calorímetro capilar VP-Capillary DSC (Microcal<sup>®</sup>) del Laboratorio de Biofísicoquímica de la Facultad de Química, UNAM en colaboración con el Dr. Miguel Costas Basín. La concentración de proteína utilizada en los experimentos fue de 1 mg/mL, a menos que se indique lo contrario. Se usó como amortiguador NaH<sub>2</sub>PO<sub>4</sub> 10 mM, pH 8.0. La desnaturalización se monitoreó de 10 °C a 100 °C para ReTIM18 V2R y de 10 °C a 125 °C para ReTIM29 y ReTIM30, tomando un “rescan” (segundo barrido) para verificar su reversibilidad. La velocidad de barrido empleada en todos los casos fue de 90 K/h. El calorímetro capilar VP-Capillary DSC (Microcal<sup>®</sup>) mantiene una presurización constante de 3 atm lo que permite que la temperatura más elevada a la que puede calentarse la muestra (permaneciendo en el estado líquido) sea de 125 °C. Así, en el CDB se determina el  $C_p$  a una presión ligeramente mayor a la atmosférica. Sin embargo, como la variación del  $C_p$  con presión es despreciable ( $dC_p/dP \approx 0$ ) [56] los datos obtenidos de un CDB son prácticamente los correspondientes a presión atmosférica.

#### *Desplegamiento proteico mediante urea y cloruro de guanidinio*

##### *Fluorescencia Intrínseca*

Los experimentos fueron realizados a una concentración de proteína de 0.05 mg/mL en amortiguador NaH<sub>2</sub>PO<sub>4</sub> 10 mM, pH 8.0 a temperatura constante de 25 °C. Las muestras fueron desnaturalizadas a distintas concentraciones de urea (0 a 8 M) con un tiempo de incubación de 24 h y se midió el cambio de señal de la IF. Las mediciones se realizaron en un espectrofluorómetro PC1 (ISS<sup>®</sup>) acoplado a un sistema Peltier (PolyScience<sup>®</sup>) para el control de la temperatura. Las muestras se excitaron a 295 nm y se midió la fluorescencia de 310 - 410 nm. Para cada muestra se realizaron 3 repeticiones, los valores fueron promediados y se restó la señal del amortiguador en cada condición.

##### *Dicroísmo Circular*

La concentración de proteína empleada para los experimentos fue de 0.2 mg/mL en amortiguador NaH<sub>2</sub>PO<sub>4</sub> 10 mM, pH 8.0 a temperatura constante de 25 °C. Las muestras fueron desnaturalizadas a distintas concentraciones de urea (0 a ~7.5 M) y cloruro de guanidinio (0 a ~6 M) con un tiempo de incubación de 24 h. Las mediciones se realizaron en un espectropolarímetro Chirascan<sup>™</sup> (Applied Photophysics<sup>®</sup>) acoplado a un sistema Peltier (PolyScience<sup>®</sup>) para el control de temperatura. El espectro fue seguido de 190 - 260 nm de longitud de onda para la concentración de 0 M de desnaturalizante y de 220 - 260 nm para las muestras con distintas concentraciones de desnaturalizante. Para cada muestra se realizaron 5 repeticiones, los valores fueron promediados y se le restó la señal del amortiguador en cada condición.

## *Cálculo robusto hecho a partir de los puntajes obtenidos con los softwares ROSETTA y MOPAC*

El siguiente cálculo tiene como premisa que, al hacer mutaciones en las proteínas, el número de átomos cambia, por lo que la energía calculada de los distintos diseños, no es comparable entre sí. Por ello, tomando en cuenta el número de átomos de los diseños, se hizo cálculo robusto para los puntajes obtenidos de cada una de las mutaciones generadas por ROSETTA en todos los modelos donde se reconstruyó la estructura del asa superior, un cuarto de la proteína y la interfase alfa hélice 7 – alfa hélice 8; respetando la simetría de orden 4, es decir, cada mutación aparece cuatro veces a lo largo de la proteína. Para obtener mayor confiabilidad, también se realizó el mismo cálculo descrito usando el software MOPAC, el cual hace un cálculo del calor de formación considerando todos los efectos de solvatación. Los resultados se muestran en las Tablas 10, 11 y 12 en el material suplementario.

El protocolo utilizado para este tipo de cálculo se ejemplifica a continuación para una mutación de valina por arginina:

- 1 – Se generó un archivo GVG en conformación de lámina beta usando Rosetta.
- 2 – Se generó un archivo de la molécula GRG en conformación de lámina beta usando Rosetta.
- 3 – Se calcularon ambas energías con Rosetta.
- 4 – Con Rosetta, se calcularon las energías de la proteína con la mutación por arginina en el asa superior y la proteína con la valina nativa en el asa superior.
- 5 – Se calcularon las diferencias:  
 $\Delta H_f(\text{atm}_2 + \text{bnd\_gg}) = \Delta H_f(\text{GRG}) - \Delta H_f(\text{GVG})$   
 $\Delta H_f(\text{atm}_2 + \text{bnd\_loop}) = \Delta H_f(\text{NLoop-R-LoopC}) - \Delta H_f(\text{NLoop-V-LoopC})$

Donde:

$\Delta H_f(\text{GRG})$  = Energía de la lámina beta GRG.

$\Delta H_f(\text{GVG})$  = Energía de la lámina beta GVG.

$\Delta H_f(\text{NLoop-R-LoopC})$  = Energía de la proteína con la mutación por arginina en el asa superior.

$\Delta H_f(\text{NLoop-V-LoopC})$  = Energía de la proteína con la valina nativa en el asa superior.

- 6 – Se calculó la diferencia:

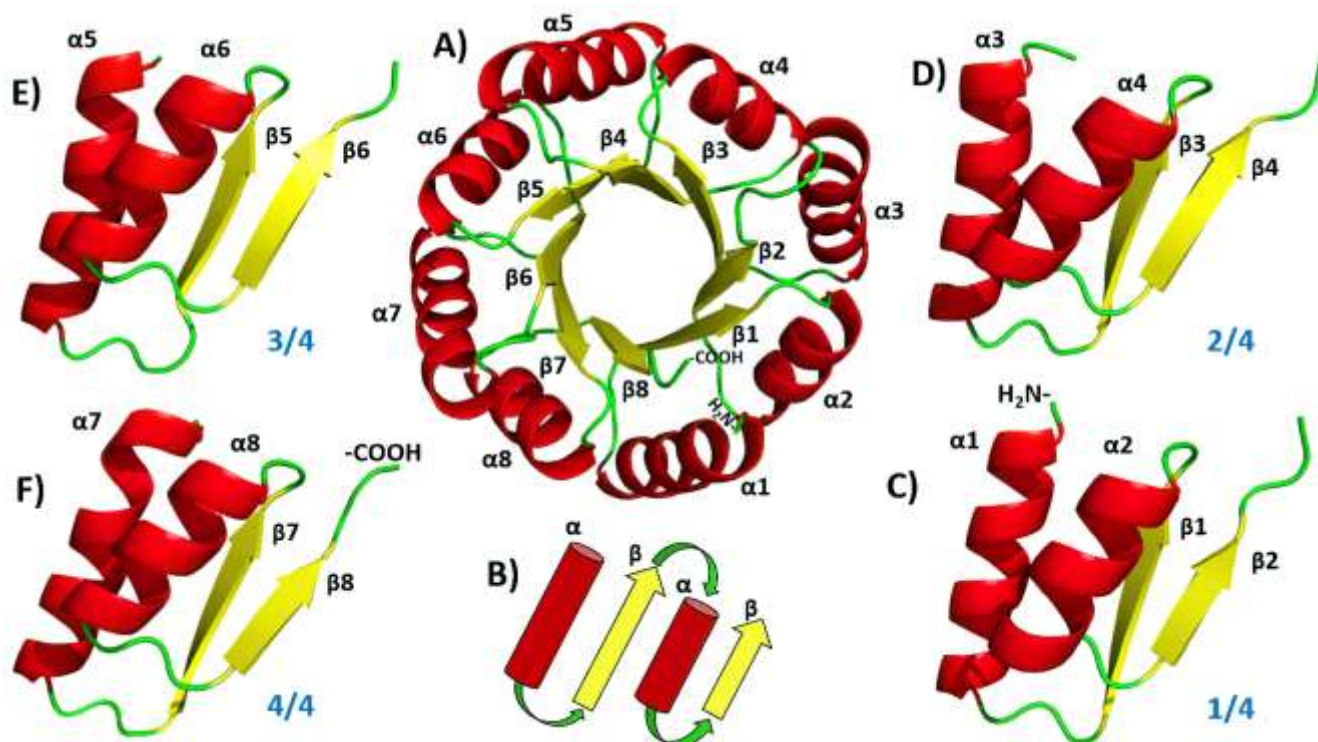
$\Delta H_f(\text{V} \rightarrow \text{R\_loop}) =$

$\Delta H_f(\text{atm}_2 + \text{bnd\_loop}) - \Delta H_f(\text{atm}_2 + \text{bnd\_gg}) =$

$\Delta H_f(\text{NLoop-R-LoopC}) + \Delta H_f(\text{GVG}) - \Delta H_f(\text{NLoop-V-LoopC}) - \Delta H_f(\text{GRG})$

## Resultados

### Resultados computacionales



**Figura 5.** ReTIM18 tiene una numeración de  $\alpha$  hélices y hebras  $\beta$  distinta a los barriles TIM naturales. El inciso A) muestra una imagen de ReTIM18 vista desde los extremos carboxilo terminal de las hebras  $\beta$  con la numeración correspondiente de las  $\alpha$  hélices y hebras  $\beta$  comenzando desde el extremo H<sub>2</sub>N- terminal hacia el -COOH terminal. En color rojo se muestran las  $\alpha$  hélices, en color amarillo las hebras  $\beta$  y en color verde las asas. El inciso B) muestra esquemáticamente cómo está compuesta una unidad repetitiva del barril ReTIM18. Los cilindros rojos simbolizan las  $\alpha$  hélices, las flechas amarillas a las hebras  $\beta$  y las flechas verdes a las asas que unen a las  $\alpha$  hélices con las hebras  $\beta$  y su dirección simboliza la orientación del extremo H<sub>2</sub>N- terminal hacia el -COOH terminal. Los incisos C), D), E) y F) son imágenes de las cuatro unidades repetitivas que componen a ReTIM18 y contienen la numeración de las  $\alpha$  hélices y hebras  $\beta$  comenzando desde el extremo H<sub>2</sub>N- terminal hacia el -COOH terminal. Cabe mencionar que el barril ReTIM18 tiene una numeración distinta a los barriles TIM naturales, ya que, por lo general, los barriles TIM naturales comienzan con una hebra  $\beta$  en el extremo H<sub>2</sub>N-, mientras que ReTIM18 comienza con una  $\alpha$  hélice.

La proteína que se tomó como base para realizar los diseños en este trabajo fue ReTIM18, ya que, al momento de iniciar este trabajo, era la única ReTIM para la cual se conocía la estructura cristalográfica. La Figura 5 muestra la numeración de las  $\alpha$  hélices y hebras  $\beta$  comenzando desde el extremo H<sub>2</sub>N- terminal hacia el -COOH terminal. La información estructural detallada proporciona una mayor confiabilidad al proponer cambios estructurales que modifiquen diversas características y/o funciones de la proteína. En principio, es mejor partir de una estructura cristalográfica que de un diseño *in silico* cuando se pretenden realizar nuevos modelos. Un modelo completamente exitoso mantiene una estrecha semejanza estructural entre el diseño *in silico* y su estructura cristalográfica a lo largo de toda la molécula. La premisa de este proyecto es: si se pudiera aumentar la semejanza estructural entre el diseño *in silico* y la estructura cristalográfica, la estabilidad del modelo sería optimizada; es por ello que los análisis estructurales tuvieron como objetivo aumentar esta semejanza.

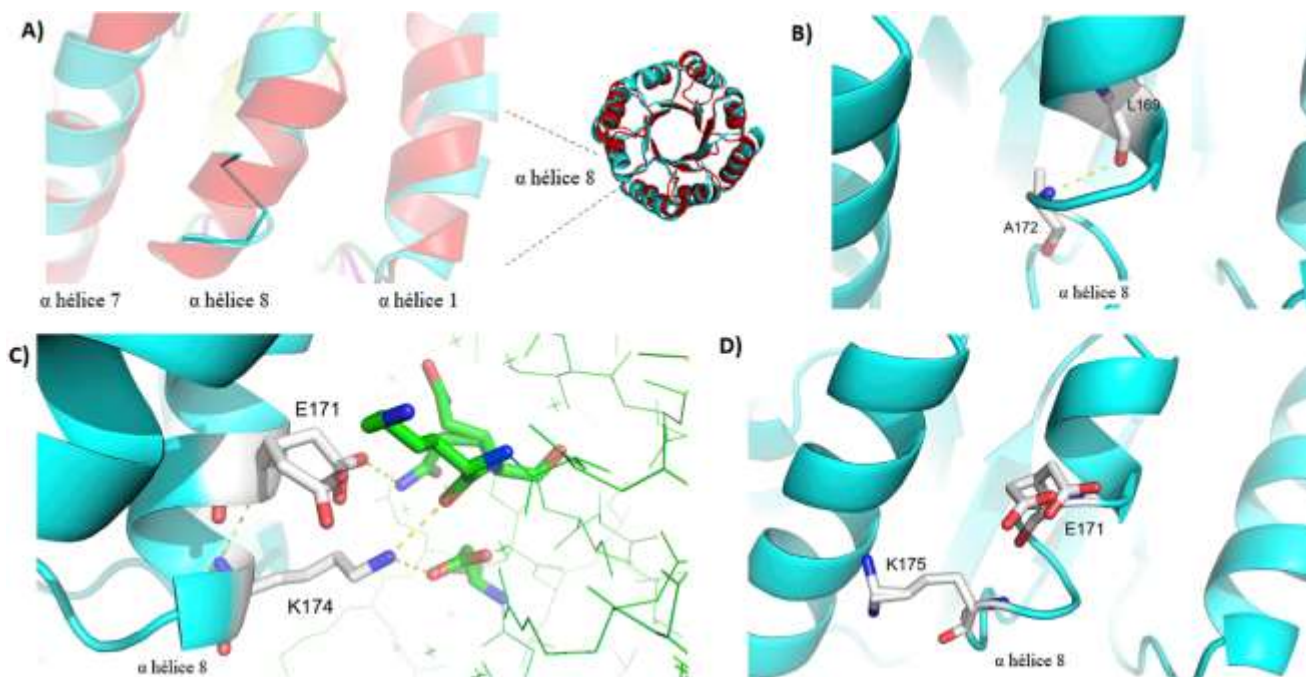




**Figura 6.** El gráfico de RMSD vs #residuo del alineamiento estructural del modelo de ReTIM18 contra la estructura experimental de ReTIM18 muestra las zonas de divergencia estructural. En la barra superior, las áreas azules indican conservación estructural, mientras que rojo significa que hay muy poca o ninguna conservación.

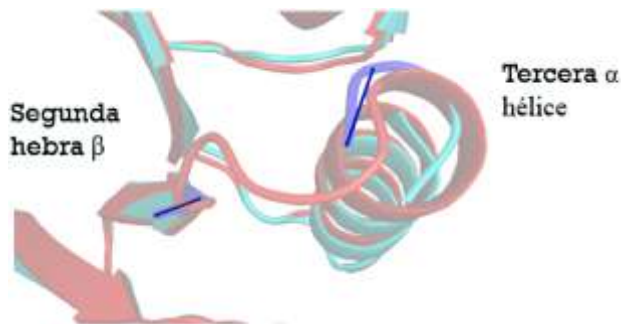
La Figura 6 muestra dos zonas donde las moléculas de ambas estructuras divergen notoriamente. Las dos zonas corresponden a la región del cierre del barril, donde se unen los extremos amino y carboxilo terminal; del residuo 1 a 3 (extremo amino terminal de la proteína) y el residuo 183 (cercano al extremo carboxilo terminal). Este resultado es típico de este tipo de análisis, por lo que extremos amino y carboxilo terminal fueron discriminados y nos enfocamos del residuo 10 al 180 para analizarlos. Los valores más grandes de RMSD se encuentran en los residuos ATDV que forman parte de las asas que conectan el amino terminal de las  $\alpha$  hélices impares con el carboxilo terminal las hebras  $\beta$  pares. Los residuos D140 – V141 pertenecientes al asa y al amino terminal de la séptima  $\alpha$  hélice son los que tienen el RMSD más grande en esta región (tercer cuarto de la proteína). Los residuos T93 y D94 pertenecientes al asa que une la quinta  $\alpha$  hélice con la cuarta hebra  $\beta$  también poseen valores altos de RMSD (cuarto cuarto de la proteína). A46 y T47 pertenecientes al asa que une la tercera  $\alpha$  hélice con la segunda hebra  $\beta$  no tienen un valor de RMSD asignado, dado que presentan una alta flexibilidad (segundo cuarto de la proteína). Los residuos A184, D2 y V3 que pertenecen al amino y carboxilo terminales de la proteína se excluyeron del análisis (primer cuarto de la proteína). Podemos ver que los residuos ATDV presentan muy poca conservación estructural en el tercer cuarto y en el cuarto cuarto de la proteína, mientras que, en el segundo cuarto hay una alta flexibilidad de esta región, por lo que podemos concluir que, esta región representa una zona potencial de mejora en la semejanza estructural.

Las demás zonas como el residuo W166 y K37 con divergencias estructurales relativamente altas fueron excluidas como regiones de mejora estructural, ya que sólo presentaban valores altos de RMSD en un solo cuarto de la proteína.

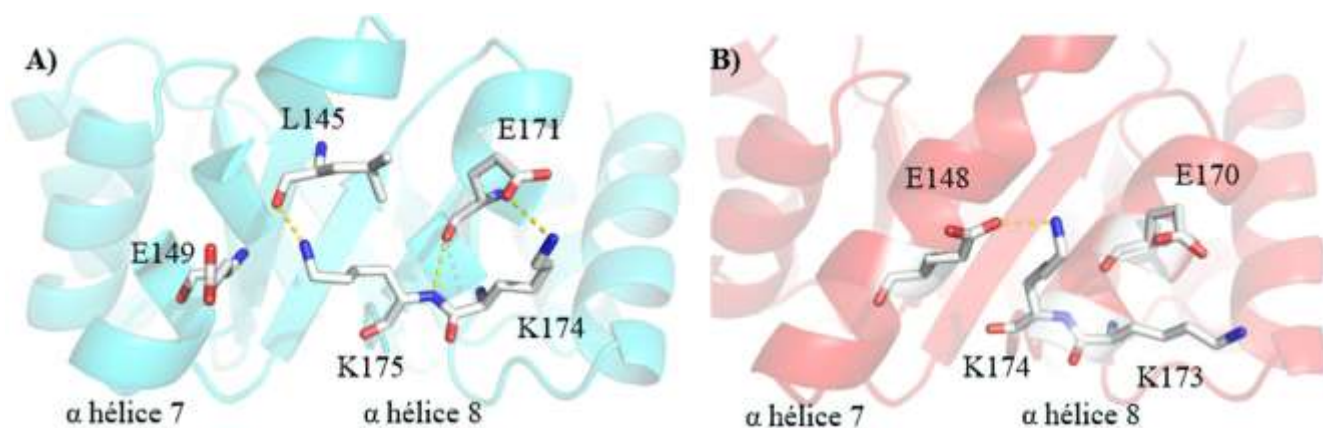


**Figura 7. La diferencia estructural de la alfa hélice 8 tiene distintas causas.** En color rojo está representado el diseño *in silico* de ReTIM18 en color azul la estructura cristalográfica y en color verde la red cristalina. A) Primera región de diferencia estructural en la  $\alpha$  hélice 8 (color azul). B) Puente de hidrógeno anómalo entre el grupo carbonilo del residuo L169 y el grupo amida del residuo A172. C) Contactos con la red cristalina (color verde) de la alfa hélice 8. D) Confórmeros alternativos de los residuos E171 y K175.

También se realizó una comparación visual del alineamiento estructural del modelo y la estructura cristalográfica para una mayor confiabilidad en la búsqueda de sitios de posible mejora estructural. Las diferencias significativas se encuentran en dos regiones: La primera es en la región del C-terminal de la  $\alpha$  hélice 8, donde cambios en la orientación de los ángulos diedros en esta  $\alpha$  hélice impiden que se formen las interacciones del diseño *in silico*. Esta diferencia estructural se ilustra en la Figura 7A. Las causas de los cambios son: 1) Presencia de un puente de hidrógeno anómalo en el esqueleto de la  $\alpha$  hélice 8 entre los grupos carbonilo y amida de los residuos L169 y A172 respectivamente. Este puente de hidrógeno no aparece en los demás cuartos equivalentes de la estructura cristalográfica ni en el modelo *in silico* (Figura 7B). 2) Torsión anómala de los residuos E171 y K174 debido a contactos con la red cristalina (Figura 7C). 3) Presencia de confórmeros alternativos en los residuos E171 y K175 (Figura 7D). Cabe señalar que la zona de diferencia estructural en la alfa hélice 8 aparece como una zona de conservación estructural media (Figura 6), (residuos 172 a 175).

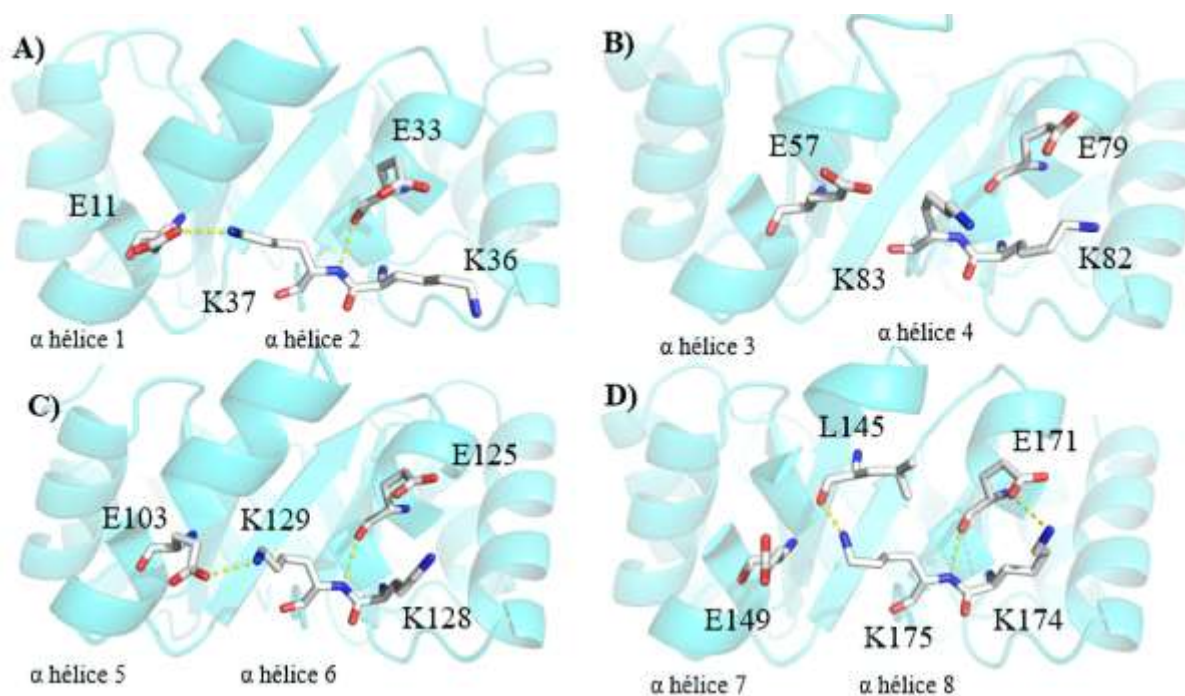


**Figura 8. Segunda región de diferencia estructural.** Dos residuos del asa superior entre la segunda hebra  $\beta$  y la tercera  $\alpha$  hélice, son demasiado flexibles para ser detectados en la estructura cristalográfica. En color rojo está el diseño *in silico* de ReTIM18, en color azul está la estructura cristalográfica de ReTIM18.



**Figura 9. Interacciones no diseñadas que aparecen en la estructura cristalográfica en la región C- terminal de la  $\alpha$  hélice 8.** A) Estructura cristalográfica de ReTIM18 en la región  $\alpha$  hélice 7 -  $\alpha$  hélice 8 con las interacciones no diseñadas (L145 – K175 y E171 – K174). B) Diseño *in silico* de ReTIM18 en la región  $\alpha$  hélice 7 -  $\alpha$  hélice 8 con la interacción diseñada (E148 – K174).

La segunda región se encuentra en el asa superior que une a la segunda hebra  $\beta$  con la tercera  $\alpha$  hélice (Figura 8). En esta zona no es posible cuantificar la divergencia estructural, ya que los residuos A46 y T47 son demasiado flexibles y no existe un RMSD asignado para ellos (Figura 6), además ya hemos mencionado que esta zona presenta una alta divergencia estructural en dos cuartos de la proteína. Es relevante mencionar que, los residuos A46 y T47 crean un desfase de una unidad en la numeración de la secuencia experimental respecto al diseño *in silico*. Por ejemplo: El residuo de la secuencia experimental E149, aparece como E148 en el diseño *in silico* (Figura 9A y 9B respectivamente).



**Figura 10. Interacciones no diseñadas a lo largo de las cuatro unidades repetitivas de ReTIM18.** A) Interacciones entre las  $\alpha$  hélices 1 – 2. B) Interacciones entre las  $\alpha$  hélices 3 – 4. C) Interacciones entre las  $\alpha$  hélices 5 – 6. D) Interacciones entre las  $\alpha$  hélices 7 – 8. Las cuatro imágenes muestran los cuartos equivalentes de la estructura cristalográfica de ReTIM18.

En la primera región, la interacción diseñada se da entre el grupo carboxilo de E148 de la  $\alpha$  hélice 7 y el grupo amino de la cadena lateral de K174 perteneciente a la  $\alpha$  hélice 8 (Figura

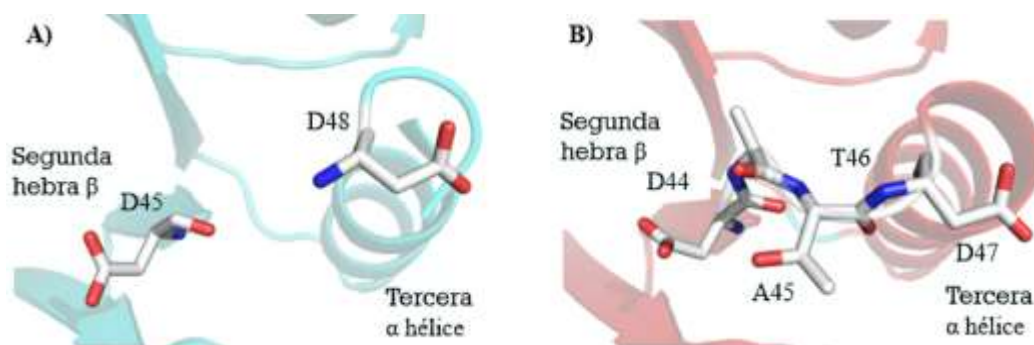


9B). Cabe mencionar que por la simetría de orden cuatro, esta interacción debería ser semejante en los cuatro cuartos del barril, sin embargo, aunque la interacción no se observa en las hélices 3-4 y 7-8 (Figura 10B y 10D), si está presente en las hélices 1-2 y 5-6 (Figura 10A y 10C).

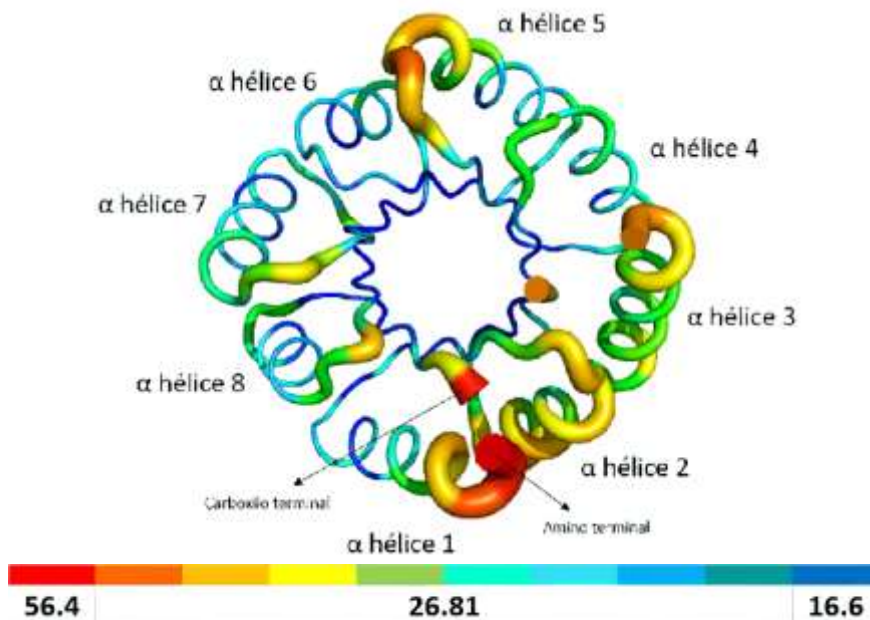
Por otra parte, la primera interacción no diseñada encontrada en la estructura cristalográfica de ReTIM18 es el puente de hidrógeno observado entre el grupo carbonilo de L145 de la  $\alpha$  hélice 7 y el grupo amino de la cadena lateral de K175 perteneciente a la  $\alpha$  hélice 8 (Figura 9A). Esta interacción crea una desviación del C-terminal del  $\alpha$  hélice 8 respecto al diseño *in silico* y no se observa en los otros cuartos de la proteína.

La segunda interacción no diseñada es entre E171 y K174, que forman un puente de hidrógeno entre el grupo carboxilo de E171 y el grupo amino de la cadena lateral de K174 (Figura 9A). El diseño *in silico* carece de esta interacción (Figura 9B), la cual no aparece en las  $\alpha$  hélices 2, 4 y 6, sólo se observa en la  $\alpha$  hélice 8 (Figura 10A, 10B y 10C).

Mediante el protocolo computacional (expuesto en la sección de materiales y métodos), se buscó la formación de un puente de sal entre los residuos 148 y 174 en el C-terminal de las  $\alpha$  hélices 7 y 8 respectivamente, para ello, se respetó la simetría de orden cuatro y se mantuvo el esqueleto fijo. Los resultados arrojaron que, al comparar con el modelo original, ninguna de las mutaciones propuestas mostró una mejor puntuación de acuerdo a Rosetta. También, buscando la formación de otro tipo de interacciones entre estas posiciones (como puentes de hidrógeno o interacciones hidrofóbicas) se intentó mutar estas posiciones permitiendo la presencia de cualquier aminoácido; sin embargo, Rosetta señaló a los aminoácidos nativos como los de menor energía. Análisis posteriores mostraron que la diferencia estructural se debía a que los aminoácidos de esta zona estaban involucrados en los contactos periódicos con la red cristalina. El desfase significativo que causan los contactos cristalinos, denota una alta flexibilidad en esta zona por lo que es candidata a ser estructuralmente optimizada. En el segundo cuarto de la proteína, entre las alfa hélices 3 y 4 no hay contactos cristalinos y a pesar de esto, las interacciones correctas no se forman, lo que nos indica que las interacciones no diseñadas son de mayor energía, o bien, que se genera una tensión a lo largo de toda la molécula debido a la formación de interacciones no diseñadas, y que éstas afectan las interacciones en sitios equivalentes de los cuatro cuartos de la proteína que no están involucradas en contactos cristalinos.



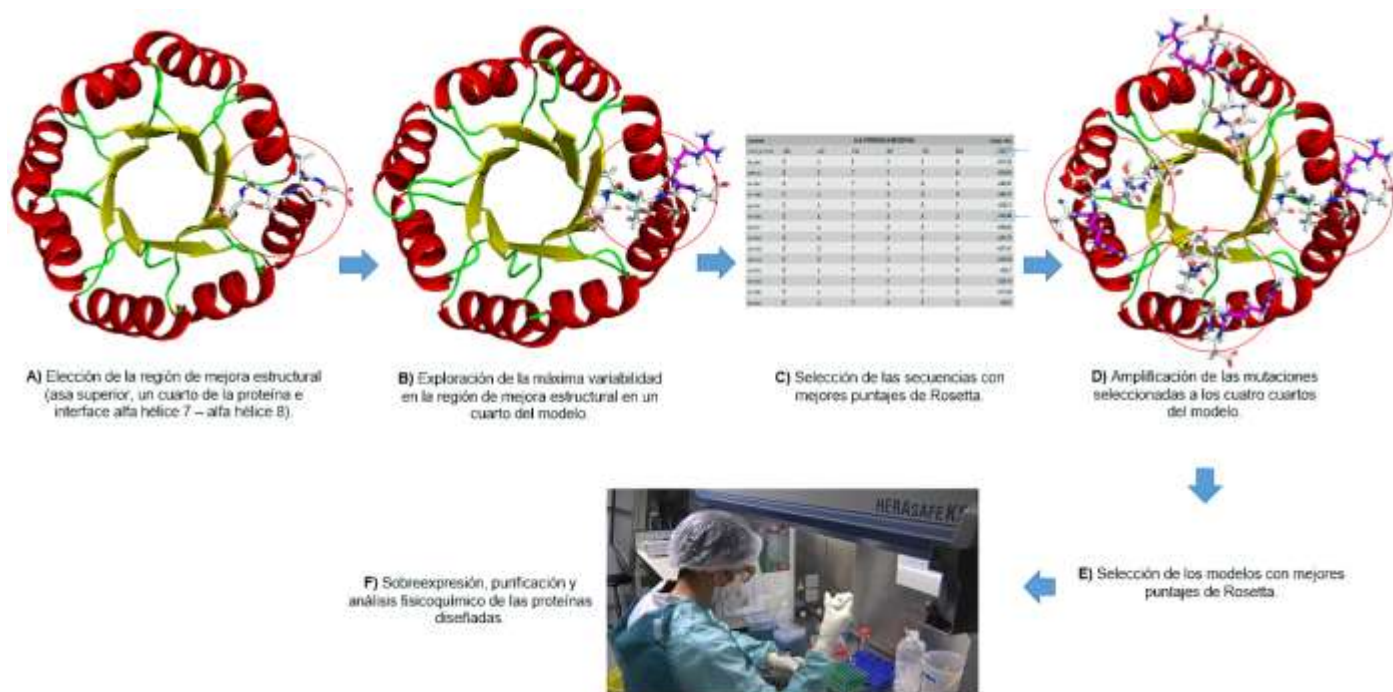
**Figura 11. Segunda región de diferencia estructural (residuos A46 y T47 demasiado móviles).** A) Estructura cristalográfica de ReTIM18 en la región de la segunda hebra  $\beta$  – tercera  $\alpha$  hélice B) Diseño *in silico* de ReTIM18 en la región de la segunda hebra  $\beta$  – tercera  $\alpha$  hélice.



**Figura 12.** El Factor B de la estructura cristalográfica de ReTIM18 muestra alta movilidad en la segunda región de diferencia estructural, en tres de las cuatro unidades repetitivas. Las  $\alpha$  hélices están enumeradas del amino terminal hacia el carboxilo terminal. El barril se observa desde la parte superior, donde se encuentran los extremos carboxilos terminales de las hebras  $\beta$ . Las tonalidades azules representan una baja movilidad en esas regiones de la proteína, mientras que las tonalidades naranja – rojo representan una alta movilidad. La escala mostrada en la parte inferior corresponde al factor R de los carbonos alfa de la cadena polipeptídica con código de colores. Los datos de colecta y afinamiento del modelo se encuentran en el material suplementario (Tabla 13) [27].

La segunda región de diferencia estructural significativa se encuentra en los residuos A46 y T47 pertenecientes al asa superior que une la hebra  $\beta$  2 con la  $\alpha$  hélice 3 (Figura 11) cuya ausencia en la estructura cristalográfica sugiere que son demasiado móviles para ser detectados. Luego de verificar la periodicidad de contactos cristalinos en esta zona, no se encontraron contactos cristalinos que interfirieran con los residuos A46 y T47, por lo que esta región es candidata para mejora potencial de la estabilidad de la proteína. Una evidencia de la movilidad de esta asa, son los valores altos de los factores B de tres de los cuatro cuartos de la estructura (en las asas que unen los extremos amino terminales de las  $\alpha$  hélices 1, 3 y 5 con las hebras  $\beta$  precedentes) (Figura 12).

De acuerdo al análisis recién expuesto, decidimos diseñar tres proteínas con las cuales se exploran niveles gradualmente mayores de complejidad al aumentar el número total de mutaciones. En la Figura 13 se muestra el esquema de trabajo que se siguió para el diseño.

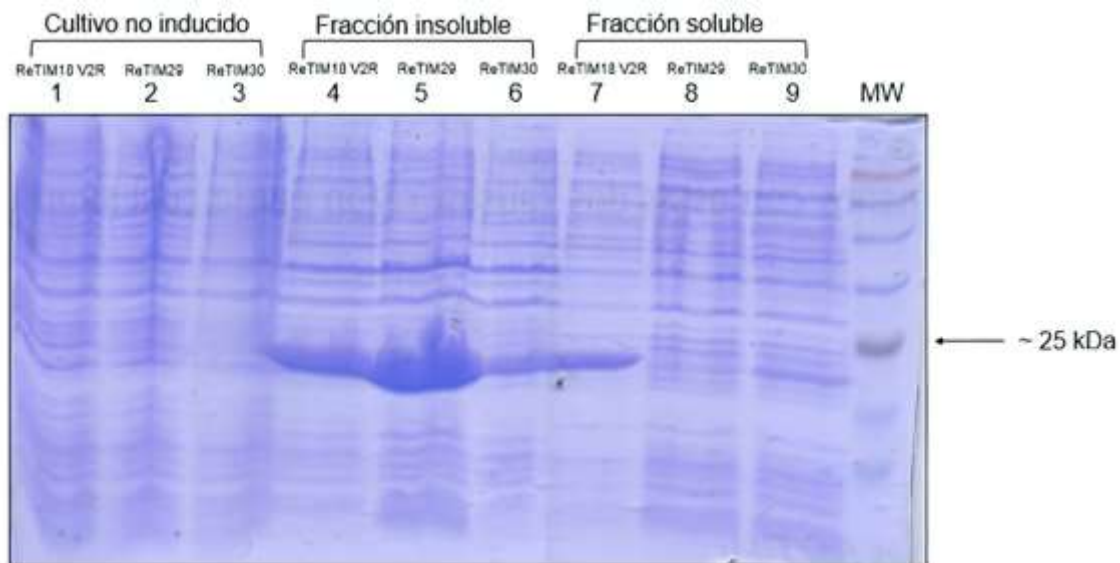


**Figura 13. Esquema general de trabajo para la obtención de los 3 modelos que se evaluaron experimentalmente.** Los aminoácidos en color blanco y con el código de color H<sub>N</sub>O<sub>S</sub>, representan las zonas de mejora estructural y están marcadas con un círculo rojo. Los aminoácidos en color magenta y con el código de color H<sub>N</sub>O<sub>S</sub>, son las mutaciones que propone Rosetta luego de explorar la máxima variabilidad de residuos en esa zona.

Los tres modelos con simetría de orden 4 que se obtuvieron al utilizar ROSETTA fueron los siguientes: a) ReTIM18 V2R, el cual se obtuvo modificando puntualmente el asa superior que une a las hebras  $\beta$  pares con las  $\alpha$  hélices impares ( $n+1$ ) mediante la mutación puntual V2R y sus equivalentes en los otros tres cuartos (V48R, V94R, V140R), b) ReTIM29 (obtenida al reconstruir un cuarto de la proteína mediante las mutaciones V2P, L6R, R13A, R14K, K18Q, Q19Y, S24A, Q31W, L34W, D39S, I40V, A45N y T46S; cuyo objetivo es lograr un cierre correcto del barril) y c) ReTIM30 cuyo propósito es reconstruir la interface entre  $\alpha$  hélices adyacentes, por ejemplo, en el caso de la  $\alpha$  hélices 7 y 8 con las mutaciones R151M, V160T, S162A y I167V. Todas estas mutaciones respetan una simetría de orden 4, por lo que cada mutación existe en las cuatro unidades repetitivas (Figura 43). En las Tablas 10, 11 y 12 del material suplementario se muestran los puntajes obtenidos mediante la función de energía de ROSETTA que sirvieron de criterio para la selección. Se eligieron los modelos con puntajes más negativos a excepción de ReTIM29, donde se hizo un análisis estructural y se encontró que había más puentes de hidrógeno formados en la secuencia 35 que en la secuencia 6 (dichas secuencias corresponden a modelos con un cuarto de la proteína reconstruida) (Tabla 10).

### Sobreexpresión de las proteínas diseñadas

La primera sobreexpresión de las tres proteínas modelo (ReTIM18 V2R, ReTIM29 y ReTIM30) se realizó con la cepa *E. coli* BL21(DE3)pLysS a una temperatura de inducción de 37 °C, una concentración de 1 mM de IPTG y siguiendo el protocolo descrito en materiales y métodos.



**Figura 14. ReTIM18 V2R se localiza en la fracción soluble, mientras que las proteínas ReTIM29 y ReTIM30 son insolubles.** En los primeros tres carriles se cargó una muestra del cultivo no inducido de las tres proteínas modelo (ReTIM18 V2R, ReTIM29 y ReTIM30). En los carriles 4, 5 y 6 se encuentran muestras de la fracción insoluble de las tres proteínas modelo. En los carriles 7, 8 y 9 se encuentran muestras de la fracción soluble de las tres proteínas modelo. En el carril 10 se encuentra el marcador de peso molecular.

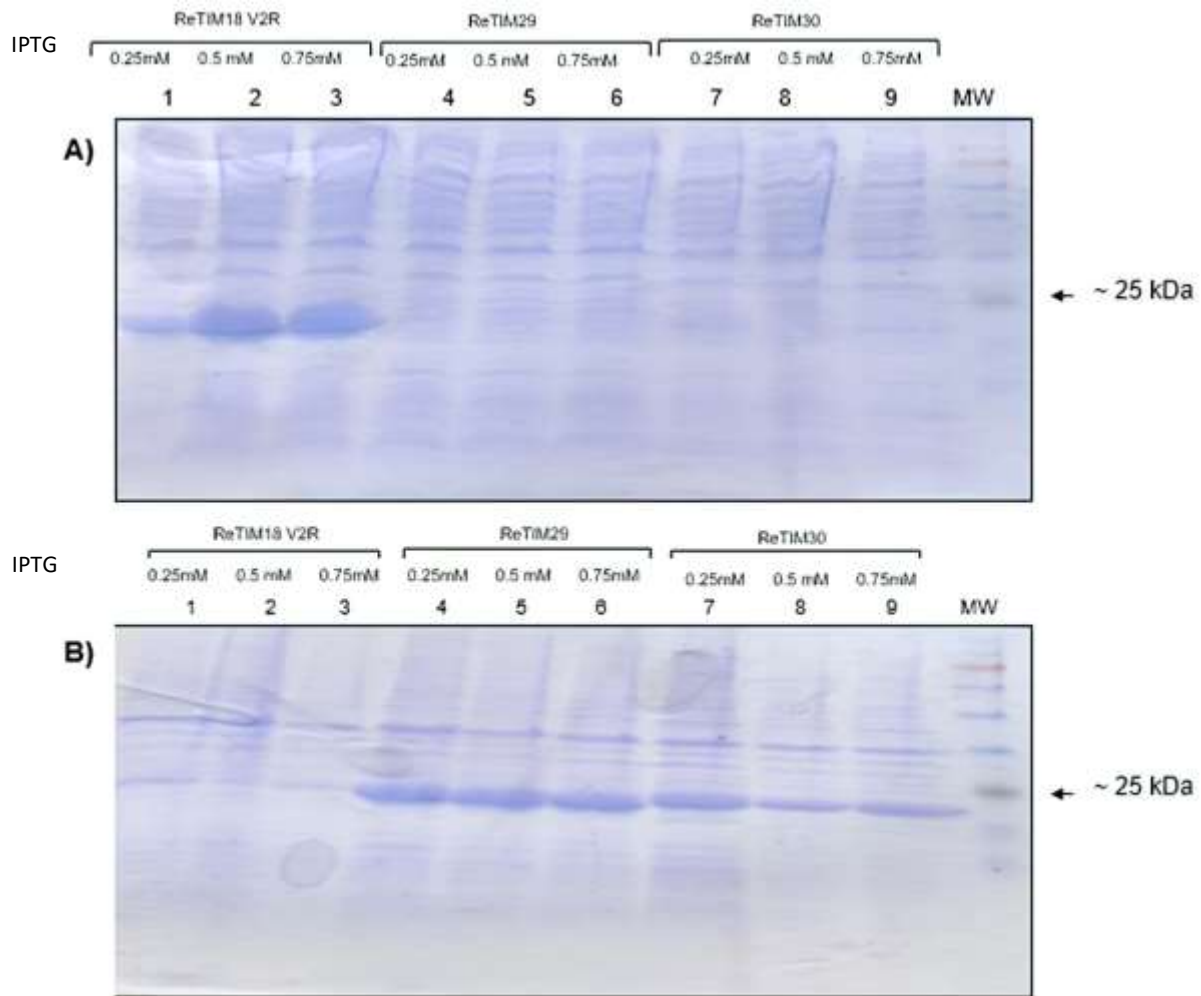
Se encontró que ReTIM29 y ReTIM30 se localizan en cuerpos de inclusión, mientras que ReTIM18 V2R se localiza tanto en la fracción soluble como en la insoluble (Figura 14), por lo que se optó por probar las siguientes condiciones de sobreexpresión y variedad de cepas:

Concentración de IPTG: 0.25mM, 0.5mM y 0.75mM

Temperatura de inducción: 20 °C, 25 °C y 30 °C

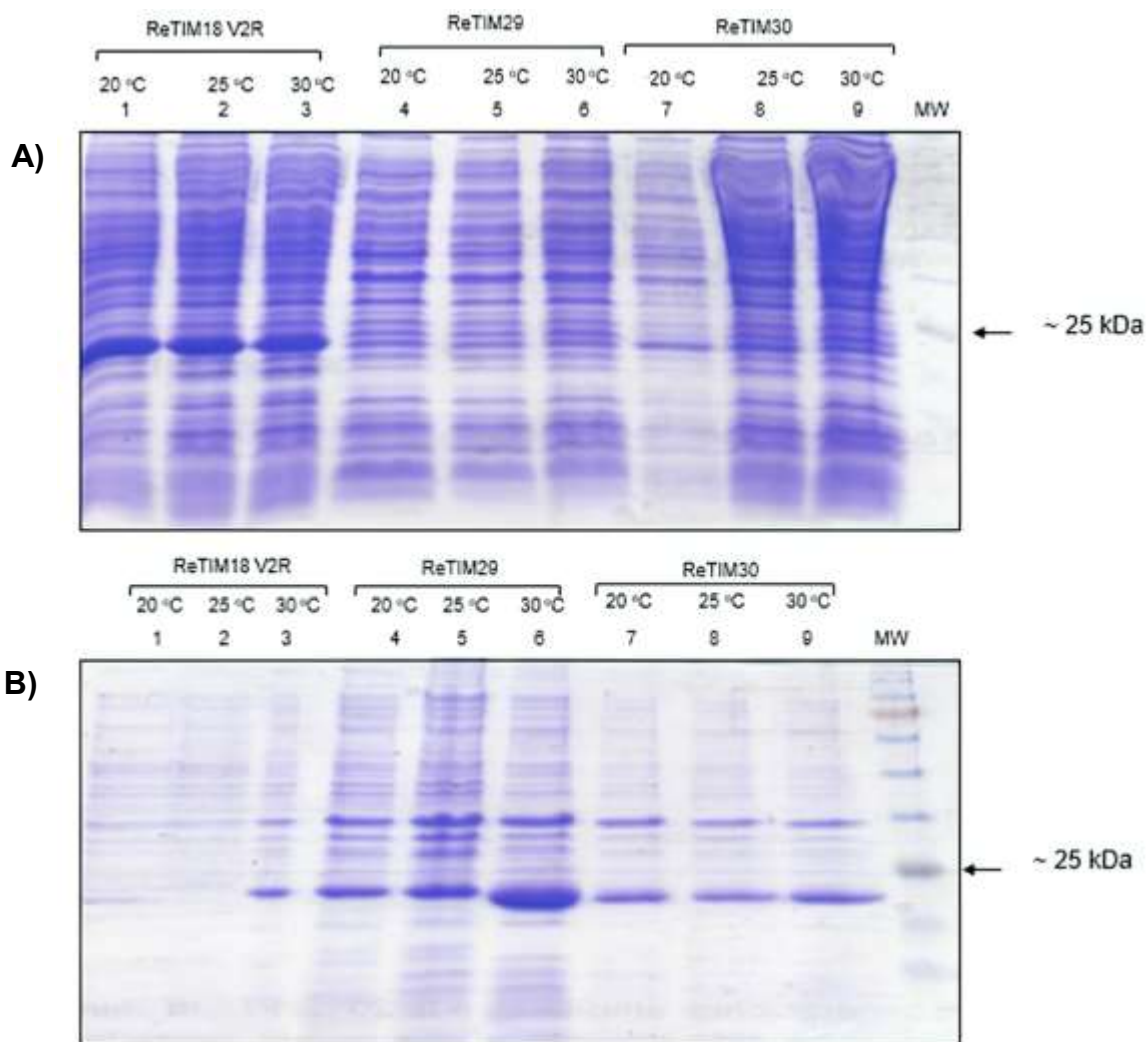
Variedad de cepas: Rosetta, Gold, LMG y ER

Los resultados de las distintas concentraciones de IPTG se ilustran en la Figura 15. Los resultados de las distintas temperaturas de inducción se muestran en la Figura 16 y los resultados de las distintas variedades de cepas se muestran en la Figura 17. La Tabla 1 resume estos resultados.

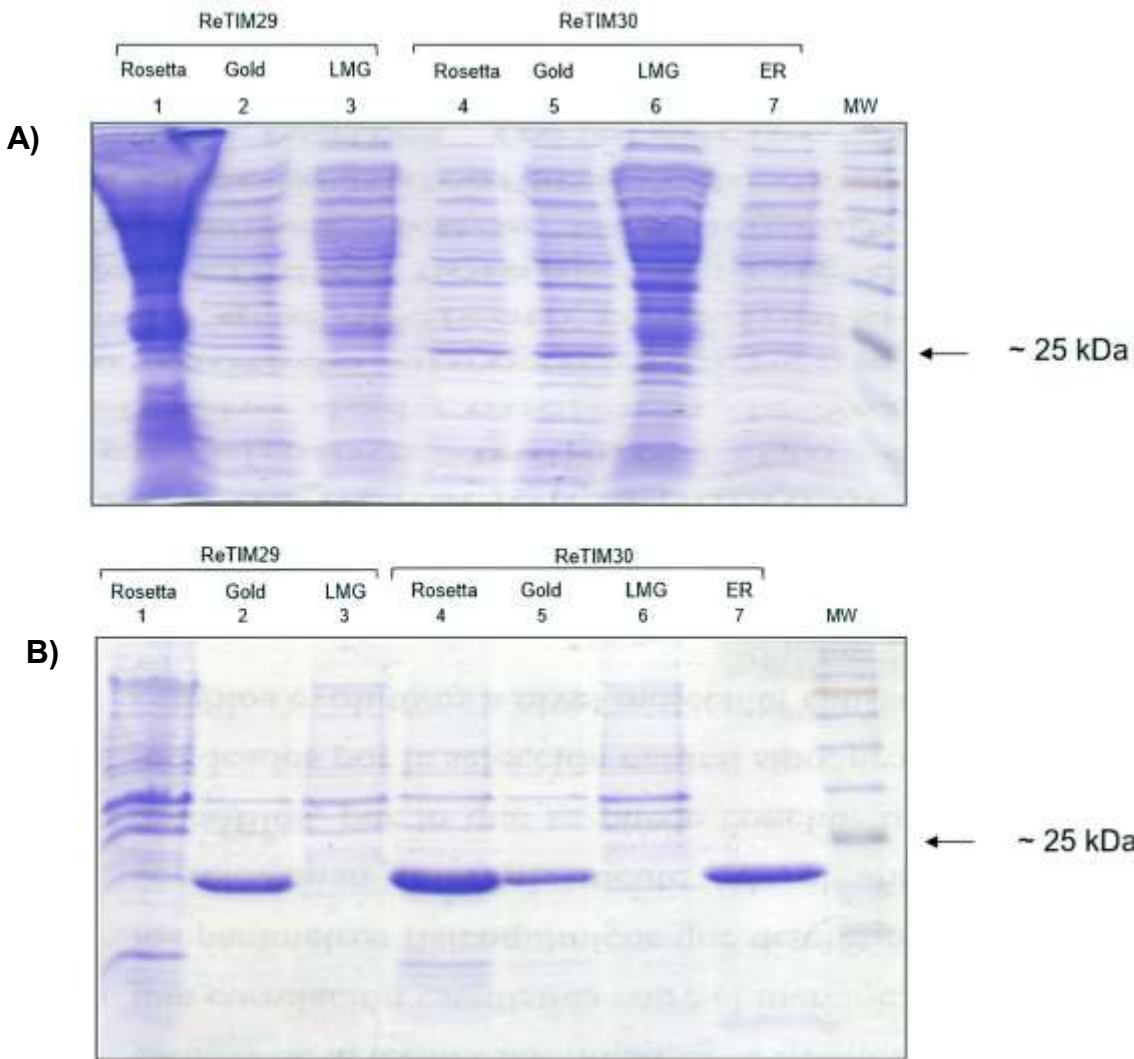


**Figura 15. La variación de las concentraciones de IPTG dieron como resultado una mayor expresión de ReTIM18 V2R soluble.** A) Gel de la fracción soluble. B) Gel de la fracción insoluble. Los primeros tres carriles corresponden a las concentraciones 0.25 mM, 0.5 mM y 0.75 mM de IPTG probadas para ReTIM18 V2R. Los carriles 4, 5 y 6 corresponden a concentraciones equivalentes de IPTG probadas para ReTIM29, mientras que los carriles 7, 8 y 9 corresponden a concentraciones de IPTG semejantes probadas para ReTIM30. En todos los ensayos, la temperatura de inducción fue de 37 °C. El carril 10 muestra el marcador de peso molecular.





**Figura 16. La disminución de la temperatura de inducción logró aumentar la fracción soluble de ReTIM18 V2R y una ligera sobreexpresión para ReTIM30 soluble.** A) Gel de la fracción soluble. B) Gel de la fracción insoluble. Los primeros tres carriles corresponden a las temperaturas de inducción de 20 °C, 25 °C y 30 °C probadas para ReTIM18 V2R. Los carriles 4, 5 y 6 corresponden a temperaturas de inducción equivalentes para ReTIM29, mientras que los carriles 7, 8 y 9 corresponden a las distintas temperaturas de inducción probadas para ReTIM30. Todos los ensayos se hicieron a una concentración de 1 mM de IPTG. El carril 10 muestra el marcador de peso molecular.



**Figura 17. ReTIM30 sobreexpresó una pequeña fracción de proteína soluble en dos de las cuatro cepas probadas.** A) Gel de la fracción soluble. B) Gel de la fracción insoluble. Los primeros tres carriles corresponden a las cepas Rosetta, Gold y LMG probadas en ReTIM29. Cabe mencionar que la cepa ER no creció para ReTIM29. Los carriles 4, 5, 6 y 7 corresponden a las cepas Rosetta, Gold, LMG y ER probadas en ReTIM30. Todos los ensayos se hicieron a 37 °C y una concentración de 1 mM de IPTG. El carril 8 muestra el marcador de peso molecular. Vale la pena mencionar que no se probó el cultivo con distintas cepas para ReTIM18 V2R, ya que, variando la temperatura de inducción y la concentración de IPTG, fue posible obtener proteína soluble.

	Temperatura (°C) (IPTG = 1 mM)			Concentración de IPTG (mM) (T = 37 °C)			Variedad de cepas (IPTG = 1 mM y T = 37 °C)			
	20 °C	25 °C	30 °C	0.25 mM	0.5 mM	0.75 mM	Rosetta	Gold	LMG	ER
ReTIM18 V2R	Soluble	Soluble	Soluble	Soluble	Soluble	Soluble	No probada	No probada	No probada	No probada
ReTIM29	Insoluble	Insoluble	Insoluble	Insoluble	Insoluble	Insoluble	Parcialmente insoluble	Insoluble	No definido	No sobreexpresó
ReTIM30	Parcialmente soluble	Insoluble	Insoluble	Insoluble	Insoluble	Insoluble	Parcialmente soluble	Parcialmente soluble	No definido	Insoluble

**Tabla 1. Resumen de los resultados al variar la concentración de IPTG y la temperatura de inducción.**

Como se puede apreciar, la proteína ReTIM18 V2R logró expresarse de manera soluble (Figuras 15A y 16A), mientras que las proteínas ReTIM29 y ReTIM30 no se obtuvieron abundantemente en la fracción soluble (Figuras 15B, 16B y 17B) por lo que se resuspendieron en urea 8 M y posteriormente se sometieron a cinco diferentes protocolos de diálisis para eliminar el desnaturizante. Las condiciones de diálisis se ilustran en la Figura 18.

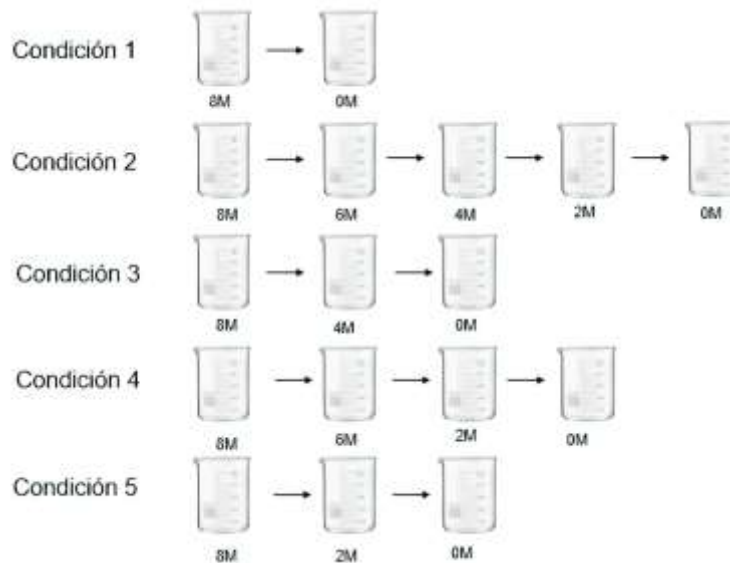


Figura 18. Las proteínas ReTIM29 y ReTIM30 fueron resuspendidas en urea y dializadas hasta remover la urea en su totalidad, probándose 5 condiciones distintas.

Debido a que los gels mostraron que las proteínas estaban prácticamente puras tras la resuspensión en urea y diálisis (Figura 3), las proteínas no pasaron por un proceso adicional de purificación.

Las concentraciones de proteína obtenidas luego de probar las 5 condiciones de resuspensión en urea y diálisis se muestran en la Tabla 2. Para ReTIM29 y ReTIM30, con la condición 1 se obtuvieron los mayores rendimientos:

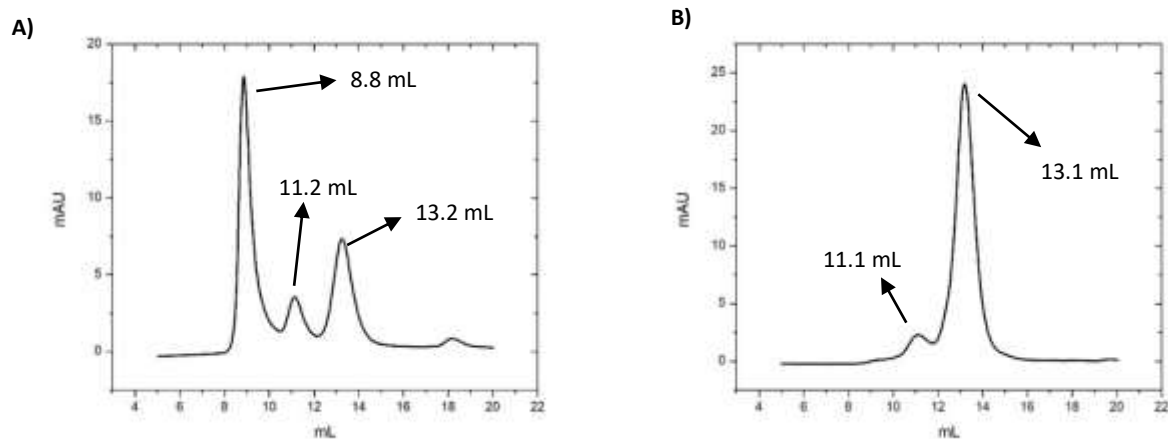
	Condición 1	Condición 2	Condición 3	Condición 4	Condición 5
ReTIM29 (mg totales finales)	3.25	1.15	1.95	0.75	2.4
ReTIM30 (mg totales finales)	4.3	2.8	3.8	3.2	3.45

Tabla 2. Recuperación de ReTIM29 y ReTIM30 en las diferentes condiciones de resuspensión en urea.

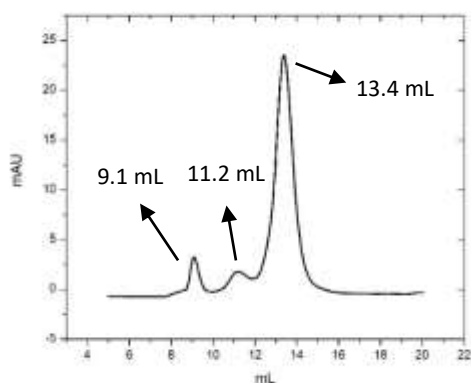
### *Radio de Stokes y peso molecular*

Con el fin de determinar el tamaño de las proteínas diseñadas, se determinaron las propiedades hidrodinámicas (Radio de Stokes y peso molecular), según el protocolo descrito en materiales y métodos. En las Figuras 19, 20 y 21 se pueden apreciar los volúmenes de elución. La Tabla 3 muestra el radio de Stokes (Rs) y los pesos moleculares de las tres proteínas respecto a ReTIM18 original.

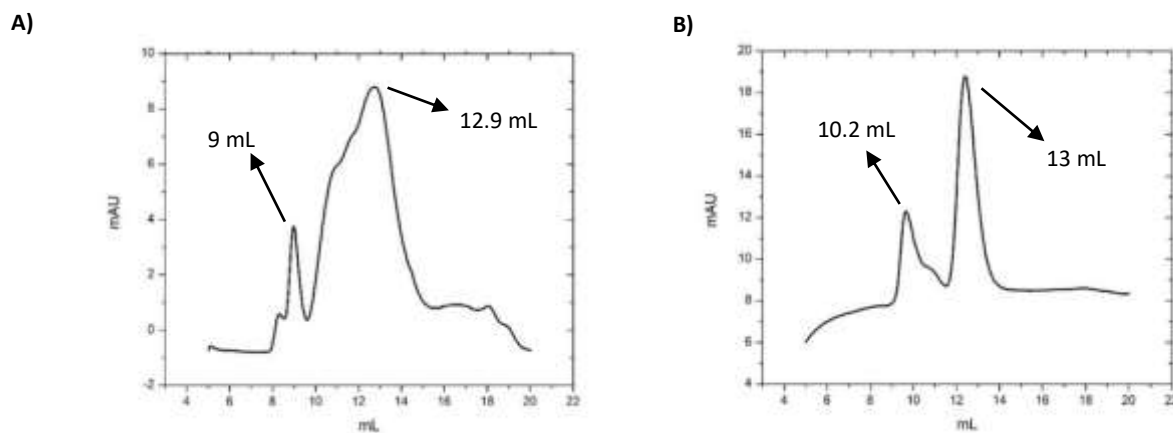




**Figura 19. El cromatograma de elución para ReTIM18 V2R muestra una reducción en la formación de agregados al minimizar el tiempo de almacenamiento.** A) En la primera repetición se pueden apreciar los picos correspondientes a tres volúmenes de elución (pico minoritario: 11.2 mL, pico intermedio: 13.2 y pico mayoritario: 8.8 mL). B) En la segunda repetición, se redujo el tiempo de almacenamiento a una noche. Se pueden apreciar los picos correspondientes a dos volúmenes de elución (Pico minoritario: 11.1 mL y pico mayoritario: 13.1 mL).



**Figura 20. Cromatograma de elución para ReTIM29.** Se pueden apreciar los picos correspondientes a tres poblaciones con diferentes volúmenes de elución (Pico minoritario: 11.2 mL, pico intermedio: 9.1 mL y pico mayoritario: 13.4 mL).



**Figura 21. El cromatograma de elución para ReTIM30 muestra como al reducir el tiempo de almacenamiento, se obtiene un pico mayoritario mejor definido.** A) En ReTIM30 no soluble se pueden apreciar los picos correspondientes a varias poblaciones con diferentes volúmenes de elución, es posible diferenciar un pico minoritario a los 9 mL y un “barrido” cuyo pico mayoritario se localiza a los 12.9 mL. B) En la segunda repetición, se utilizó proteína soluble y se redujo el tiempo de almacenamiento a una noche. Se pueden apreciar los picos correspondientes a dos poblaciones principales (pico minoritario: 10.2 mL y pico mayoritario: 13 mL).

Proteína	Volumen de elución	Rs (Å)	Peso Molecular (kDa)
ReTIM18 (original)	13.8	20.5	17.5
ReTIM18 V2R (pico minoritario)	11.2	32.2	53.5
ReTIM18 V2R (pico intermedio)	13.2	22.8	21.8
ReTIM18 V2R (pico mayoritario)	8.8	49.2	269.3
ReTIM18 V2R (repetición) (pico minoritario)	11.1	32.8	56.2
ReTIM18 V2R (repetición) (pico mayoritario)	13.1	22.8	21.9
ReTIM29 (pico mayoritario)	13.4	22	20.2
ReTIM29 (pico minoritario)	11.2	32.2	53.5
ReTIM29 (pico intermedio)	9.1	46.6	210
ReTIM30 (pico mayoritario)	12.9	24	24.5
ReTIM30 (pico minoritario)	9	47.5	227.7
ReTIM30 (soluble) (pico mayoritario)	13	23.6	23.4
ReTIM30 (soluble) (pico minoritario)	10.2	38	92.2

**Tabla 3. Volúmenes de elución, Radio de Stokes (Rs) y Peso Molecular (kDa) de ReTIM18, ReTIM18 V2R, ReTIM29 y ReTIM30**

En la proteína ReTIM18 V2R se encontraron tres picos, el pico que presenta el mayor volumen de elución muestra un Rs y peso molecular esperados para un monómero compacto, mientras que los otros picos corresponden a agregados (Figura 19A). Una de las desventajas que posee ReTIM18, la proteína usada como base para todos los diseños de este trabajo, es que tiende a agregarse en tiempos muy cortos. Ya que ReTIM18 V2R es muy parecida al modelo original, el experimento fue repetido a un tiempo corto de almacenamiento después de la purificación (1 noche); el resultado se expone en la Figura 19B. De acuerdo a los datos obtenidos, se lograron reducir significativamente los agregados de ReTIM18 V2R al reducir el tiempo de almacenamiento luego de ser purificada.

La proteína ReTIM29 obtenida mediante la resuspensión en urea y renaturalización mediante la condición 1 obtuvo tres picos, donde el pico con un mayor volumen de elución (13.4 mL), tiene un volumen similar a ReTIM18 original (13.8 mL), por lo que tiene un Rs y peso molecular esperados para un monómero compacto. Los dos picos restantes corresponden a agregados (Figura 20).

ReTIM30 obtenida por medio de la resuspensión en urea y renaturalización empleando la condición 5, mostró dos picos, de los cuales, el pico mayoritario (12.9 mL) tiene un volumen de elución esperado para un monómero compacto, pero es un pico amplio y con hombros, lo que denota que la muestra no es monodispersa y tiene una heterogeneidad significativa. El pico minoritario (9 mL) corresponde a la formación de agregados (Figura 21A).

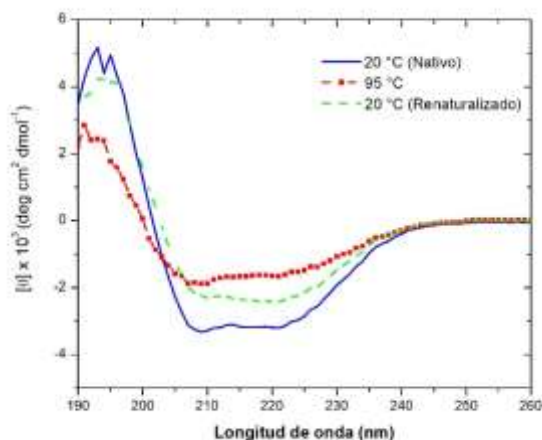
De acuerdo al gel mostrado en la Figura 16A, existe una pequeña porción de proteína ReTIM30 soluble a la temperatura de inducción de 20 °C, por lo que el experimento también fue repetido para ReTIM30 con proteína soluble y un tiempo corto después de la purificación (1 noche). El resultado se expone en la Figura 21B y muestra un pico mayoritario monodisperso y heterogéneo con un volumen de elución prácticamente igual al del pico mayoritario de ReTIM30 obtenida por medio de la resuspensión en urea y renaturalización empleando la condición 5, lo que indica que se logró reducir la heterogeneidad y cantidad de agregados al

emplear proteína ReTIM30 soluble a un tiempo corto de almacenamiento después de la purificación.

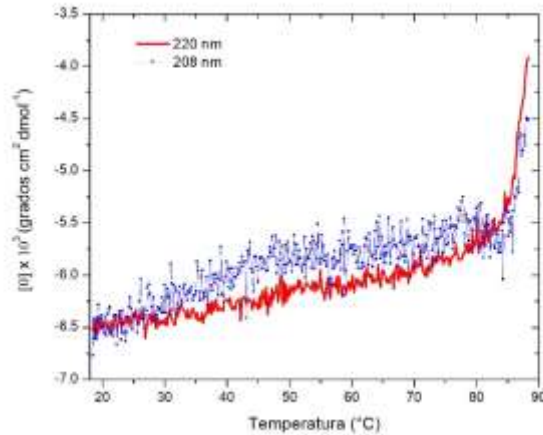
### *Desplegamiento térmico seguido por Dicroísmo Circular*

Se obtuvieron los diferentes espectros de las proteínas modelo a 20 °C, a los 95 °C y renaturalizando a 20 °C; siguiendo el protocolo de desplegamiento por temperatura descrito en materiales y métodos. ReTIM18 V2R tuvo 6 días de almacenamiento y un día de diálisis. ReTIM29 y ReTIM30 en las condiciones 1 y 5 respectivamente de renaturalización después de re-suspensión en urea, estuvieron en un día de almacenamiento y un día de diálisis. ReTIM30 soluble no tuvo tiempo de almacenamiento, se dializó durante una noche luego de ser purificada e inmediatamente se monitoreó su desplegamiento térmico por dicroísmo circular. La diálisis de todos los ensayos se llevó a cabo siguiendo el protocolo de mediciones espectroscópicas – dicroísmo circular, descrito en el apartado de materiales y métodos.

A 20°C, el espectro de DC en el UV lejano de la proteína ReTIM18 V2R muestra los dos mínimos de elipticidad a 208 y 222 nm correspondientes a estructuras de tipo  $\alpha/\beta$  (Figura 22). El desplegamiento por temperatura mostró el inicio de una transición cerca de los 85 °C (Figura 23), sin embargo, a temperaturas elevadas (95 °C) el espectro de DC (Figura 22) y las deconvoluciones hechas tanto con el software CDNN [48] como con el servidor en línea K2D2 [49] (Tabla 4 y 5), muestran la presencia de estructura residual, esto es, no se observa un desplegamiento total de la estructura secundaria. Después de regresar la muestra a temperatura ambiente, se encontró una recuperación parcial de la señal original (Figura 22). El desplegamiento por temperatura mostró un inicio de transición cerca de los 85 °C (Figura 23).

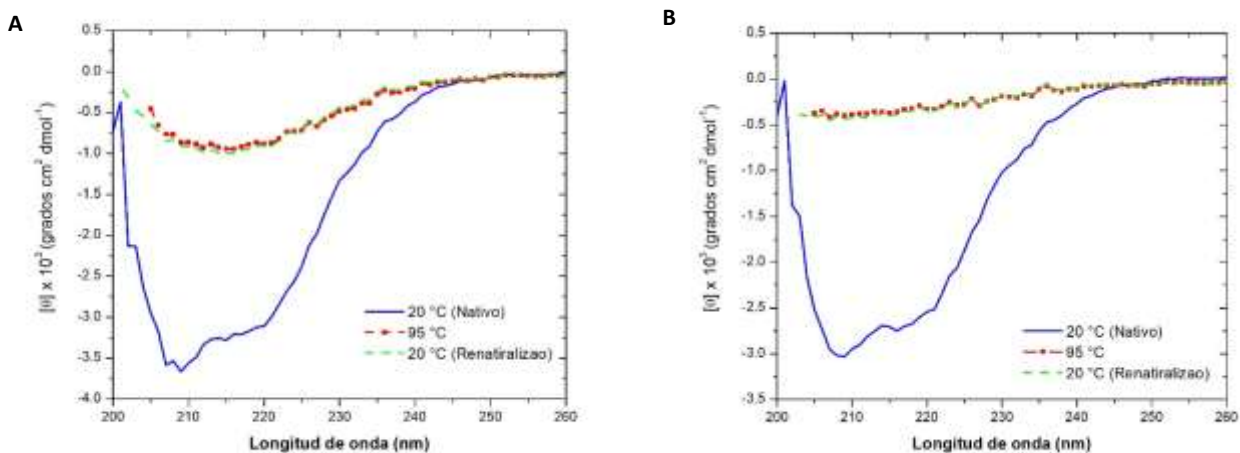


**Figura 22. Irreversibilidad del desplegamiento térmico de ReTIM18 V2R.** Espectro de DC de ReTIM18 V2R a 20 °C antes de calentar (línea azul), a 95 °C (línea roja) y renaturalizada a 20 °C (línea verde).

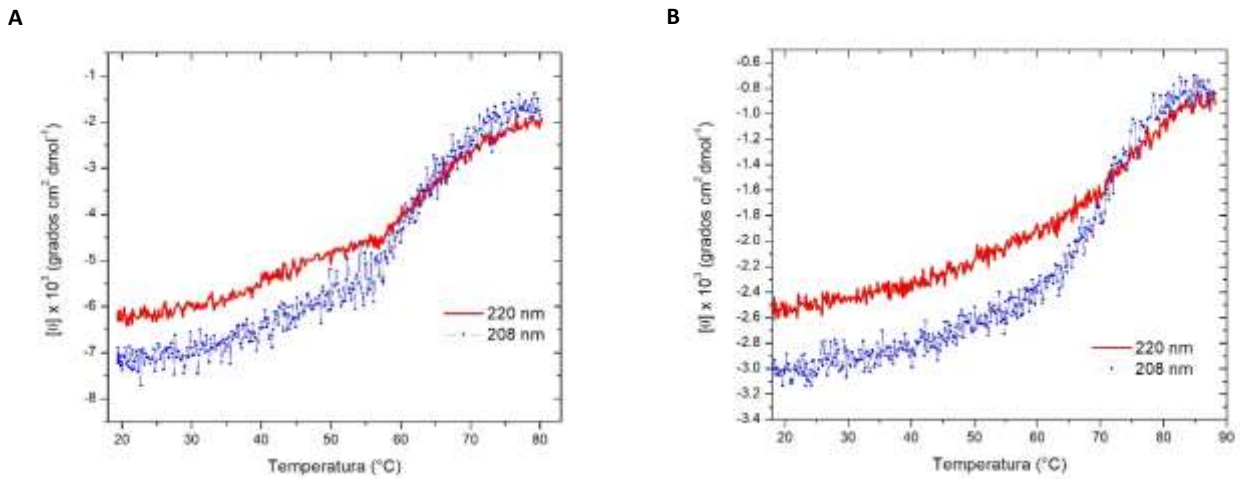


**Figura 23. Transición del desplegamiento térmico a los 89 °C para ReTIM V2R.** Desplegamiento térmico seguido a 208 y 220 nm.

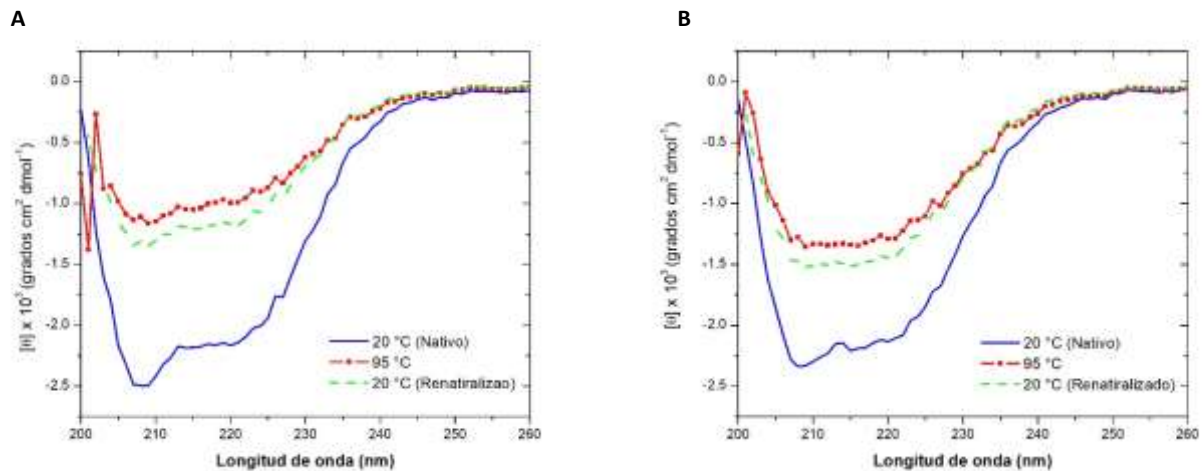
En las condiciones 1 y 5 de renaturalización después de re-suspensión en urea, la proteína ReTIM29 y ReTIM30 muestran dos mínimos de elipticidad molar a 208 y 222 nm correspondientes a estructuras de tipo  $\alpha/\beta$  (para ambos casos, más pronunciados en la condición de renaturalización 1) (Figuras 24A y 26A). En ambos casos se mostró una pérdida de estructura secundaria (alfa hélice), ganancia de hebras beta y de asas luego de ser calentada a 95 °C y renaturalizada a 20 °C (Tabla 4 y 5). En ambas proteínas el espectro a 95 °C es muy similar al espectro a 20 °C renaturalizado en ambas proteínas, lo que muestra que el desplegamiento térmico no es reversible (Figuras 24 y 26). Para ReTIM29 el desplegamiento por temperatura mostró una sola transición con  $T_m$  cercana a 65 °C para la proteína replegada mediante el protocolo indicado como condición 1 y de 75 °C para la condición 5 de replegamiento; ambas transiciones, son apreciables tanto a 208 como a 222 nm, sin embargo, son más claras a 208 nm (Figuras 25A y 25B). En el caso de ReTIM30 el desplegamiento por temperatura no mostró alguna transición cooperativa en proteínas obtenidas bajo ambas condiciones de replegamiento (Figuras 27A y 27B).



**Figura 24. Irreversibilidad del desplegamiento térmico de ReTIM29 a las condiciones 1 y 5 de resuspensión en urea.** **A)** Espectro de DC de ReTIM29 (resuspensión en urea y renaturalización mediante la condición 1) a 20 °C antes de calentar (línea azul), a 95 °C (línea roja) y renaturalizada a 20 °C (línea verde). **B)** Espectro de DC de ReTIM29 (resuspensión en urea y renaturalización mediante la condición 5) a 20 °C antes de calentar (línea azul), a 95 °C (línea roja) y renaturalizada a 20 °C (línea verde).

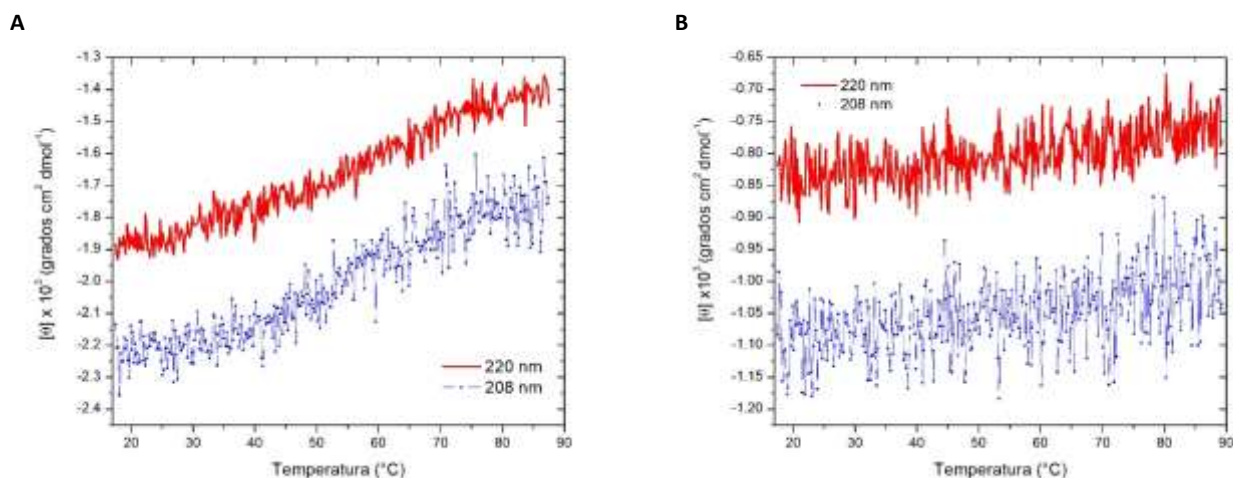


**Figura 25. Transición del desplegamiento térmico a los 65 ° C y 75 ° C para ReTIM29 a las condiciones 1 y 5 de resuspensión en urea.** **A)** Desplegamiento térmico seguido a 208 y 220 nm de ReTIM29 después de resuspensión en urea y renaturalización mediante la condición 1. **B)** Desplegamiento térmico seguido a 208 y 220 nm de ReTIM29 después de resuspensión en urea y renaturalización mediante la condición 5.



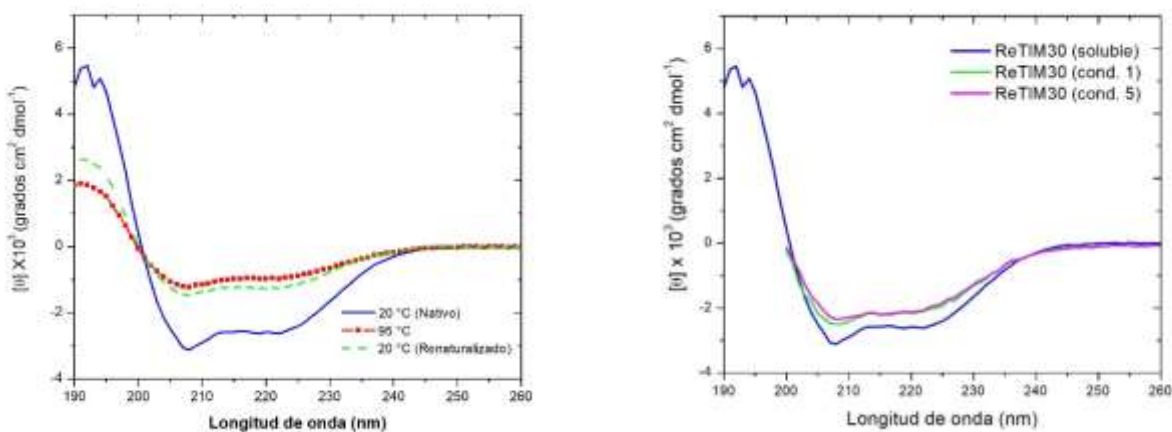
**Figura 26. Irreversibilidad del desplegamiento térmico de ReTIM30 a las condiciones 1 y 5 de resuspensión en urea.** **A)** Espectro de DC de ReTIM30 (resuspensión en urea y renaturalización mediante la condición 1) a 20 °C antes de calentar (línea azul), a 95 °C (línea roja) y renaturalizada a 20 °C (línea verde). **B)** Espectro de DC de ReTIM30 resuspensión en urea y renaturalización mediante la condición 5) a 20 °C antes de calentar (línea azul), a 95 °C (línea roja) y renaturalizada a 20 °C (línea verde).





**Figura 27. Las transiciones del desplegamiento térmico para ReTIM30 a las condiciones 1 y 5 de resuspensión en urea no fueron observables. A)** Desplegamiento térmico seguido a 208 y 220 nm de ReTIM30 replegada con la condición 1 de resuspensión en urea. **B)** Desplegamiento térmico seguido a 208 y 220 nm de ReTIM30 replegada con la condición 5 de resuspensión en urea.

El experimento también fue repetido para ReTIM30 con proteína soluble y un tiempo corto de almacenamiento (solo 1 noche de diálisis). La proteína ReTIM30 soluble muestra los dos mínimos de elipticidad molar a 208 y 222 nm (Figura 28) correspondientes a estructuras de tipo  $\alpha/\beta$  y fueron más pronunciados que en el primer experimento (Figura 28B), lo que quizá se deba a que, luego de ser replegada a las condiciones 1 y 5 de resuspensión en urea, ReTIM30 recuperada de cuerpos de inclusión tiene cierto grado inicial de desplegamiento. Por otra parte, ReTIM30 soluble mostró pérdida de estructura secundaria luego de ser calentada a 95 °C y renaturalizada a 20 °C, siguiendo el mismo patrón de las demás proteínas (pérdida de alfa hélices, ganancia de hebras beta y asas) (Tabla 4 y 5). El desplegamiento por temperatura no mostró alguna transición cooperativa (Figura 29).



**Figura 28. Irreversibilidad del desplegamiento térmico de ReTIM30 soluble y un espectro más definido. A)** Espectro de DC de ReTIM30 soluble a 20 °C antes de calentar (línea azul), a 90 °C (línea roja) y renaturalizada a 20 °C (línea verde). **B)** La línea azul muestra el espectro de DC de ReTIM30 soluble a 20 °C; las líneas verde y magenta muestran, respectivamente, los espectros de DC de ReTIM30 replegada con las condiciones 1 y 5 de resuspensión en urea a 20 °C.

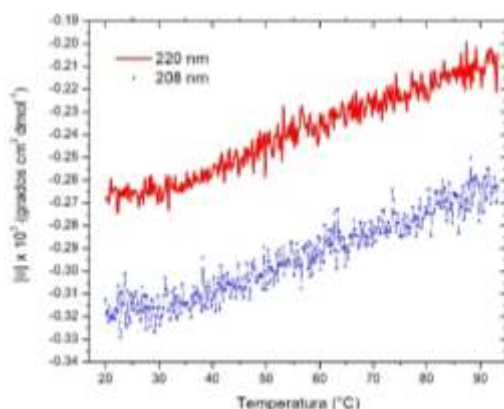


Figura 29. La transición del desplegamiento térmico para ReTIM30 soluble no fue observable. Desplegamiento térmico seguido a 208 y 220 nm de ReTIM30 soluble.

20 °C (Nativo)	ReTIM18 V2R	ReTIM29 (cond. 1)	ReTIM29 (cond. 5)	ReTIM30 (cond. 1)	ReTIM30 (cond. 5)	ReTIM30 soluble
Alfa hélice	13.78%	13.88%	12.90%	12.15%	12.04%	13.00%
Beta antiparalela	11.82%	11.84%	12.14%	12.31%	12.37%	12.08%
Beta paralela	14.71%	14.65%	15.08%	15.37%	15.42%	15.02%
Giro Beta	17.35%	17.38%	17.34%	17.27%	17.23%	17.28%
Asas	42.35%	42.25%	42.63%	42.89%	42.95%	42.62%
Suma total	100.00%	100.00%	100.08%	100.00%	100.00%	100.00%
<b>95 °C</b>						
Alfa hélice	11.29%	10.04%	9.34%	10.26%	10.67%	10.26%
Beta antiparalela	12.59%	12.96%	13.23%	12.89%	12.83%	12.89%
Beta paralela	15.76%	16.36%	16.73%	16.23%	16.04%	16.23%
Giro Beta	17.22%	17.08%	17.04%	17.10%	17.16%	17.10%
Asas	43.14%	43.48%	43.66%	43.44%	43.38%	43.44%
Suma total	100.00%	99.92%	100.00%	99.92%	100.08%	99.92%
<b>20 °C (Renaturalizado)</b>						
Alfa hélice	12.17%	10.06%	9.35%	10.56%	10.88%	10.67%
Beta antiparalela	12.33%	12.98%	13.24%	12.80%	12.73%	12.76%
Beta paralela	15.31%	16.31%	16.74%	16.08%	15.95%	16.05%
Giro Beta	17.22%	17.10%	17.06%	17.12%	17.16%	17.17%
Asas	42.96%	43.47%	43.61%	43.36%	43.27%	43.34%
Suma total	100.00%	99.92%	100.00%	99.92%	100.00%	100.00%

Tabla 4. Deconvolución de los espectros de DC para las proteínas diseñadas a 20 °C, calentada a 90 °C y renaturalizada a 20 °C, monitoreadas a 205 -260 nm mediante el programa CDNN (Circular Dichroism analysis using Neural Networks) [48].

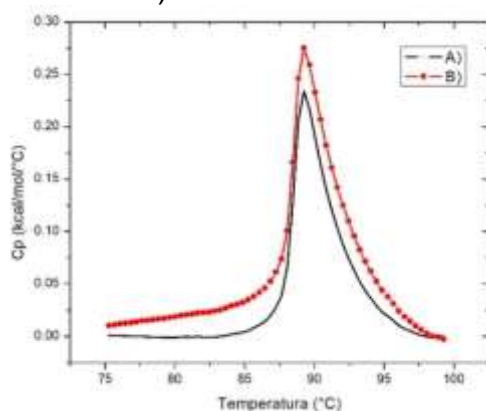
20 °C (Nativo)	ReTIM18 V2R	ReTIM29 (cond. 1)	ReTIM29 (cond. 5)	ReTIM30 (cond. 1)	ReTIM30 (cond. 5)	ReTIM30 soluble
% de alfa hélice predicha	35.15%	26.44%	14.98%	18.00%	9.89%	28.88%
% de hebras beta predicha	10.04%	17.40%	34.94%	31.28%	38.33%	16.41%
Asas	54.81%	56.16%	50.08%	50.72%	51.78%	54.71%
Suma total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Error máximo	0.23	0.4	0.4	0.38	0.38	0.23
<b>95 °C</b>						
% de alfa hélice predicha	10.56%	3.13%	4.44%	1.76%	1.94%	4.57%
% de hebras beta predicha	42.09%	49.32%	44.17%	47.42%	47.25%	46.69%
Asas	47.35%	47.55%	51.39%	50.82%	50.81%	48.74%
Suma total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Error máximo	0.32	0.4	0.4	0.4	0.4	0.32
<b>20 °C (Renaturalizado)</b>						
% de alfa hélice predicha	29.88%	3.13%	4.44%	1.76%	3.8%	10.56%
% de hebras beta predicha	16.04%	49.32%	44.17%	47.42%	49.02%	42.09%
Asas	54.08%	47.55%	51.39%	50.82%	47.18%	47.35%
Suma total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Error máximo	0.32		0.4	0.4	0.4	0.32

Tabla 5. Deconvolución de los espectros de DC para las proteínas diseñadas a 20 °C, calentada a 90 °C y renaturalizada a 20 °C, monitoreadas a 200 -240 nm mediante el servidor en línea K2D2 [49]. El valor de error máximo donde el espectro predicho por el servidor es poco parecido al espectro experimental es de 0.9, es decir, a un valor de error máximo de 0.9 la deconvolución es poco confiable. Los espectros experimentales y los espectros predichos se encuentran en la Figura 41 del material suplementario.

### Desplegamiento térmico seguido por calorimetría diferencial de barrido

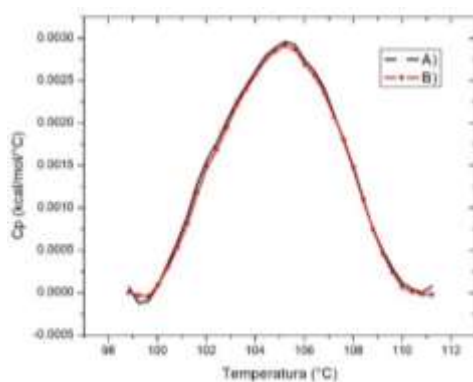
Se monitoreó el proceso de desnaturalización térmica de los tres diseños siguiendo el protocolo de desplegado por temperatura monitoreado por DSC, descrito en materiales y métodos. Las concentraciones fueron de 1.96 mg/mL y 6.28 mg/mL para ReTIM29 y ReTIM30 (recuperadas de cuerpos de inclusión) respectivamente, sin embargo, no mostraron algún pico en las trazas calorimétricas, lo que denota una falta de desplegado cooperativo en ambos modelos, o bien que las proteínas no se repliegan completamente después del tratamiento de resuspensión en urea, ya que los espectros de DC solo muestran presencia de estructura secundaria, pero no el grado de empaquetamiento de las proteínas.

La curva de la proteína ReTIM18 V2R mostró un proceso de desplegado irreversible (datos no mostrados) y presentó una  $T_m$  de 89.2 °C, además el ancho de la endoterma es estrecho (Figura 30), lo que indica un desplegado muy cooperativo; sin embargo, el  $\Delta H$  es extremadamente pequeño (1.9 kcal/mol).



**Figura 30.** La endoterma de ReTIM18 V2R muestra una transición muy cooperativa con un  $\Delta H$  muy pequeño. Línea A) Experimento de DSC para ReTIM18 V2R. Línea B) Ajuste de la endoterma de DSC de ReTIM18 V2R a un modelo irreversible de dos estados [34]. El experimento se realizó a una concentración de proteína de 1.3 mg/mL, el gráfico está normalizado.

La curva de la proteína ReTIM30 soluble fue irreversible (datos no mostrados), tiene una endoterma muy estrecha (Figura 31) que denota un desplegado altamente cooperativo, sin embargo, el  $\Delta H$  (0.26 kcal/mol) es muy pequeño y mostró una  $T_m = 105.2$  °C.



**Figura 31.** La endoterma de ReTIM30 soluble muestra una transición muy cooperativa con un  $\Delta H$  muy pequeño. Línea A) Experimento de DSC para ReTIM30 soluble. Línea B) Ajuste de la endoterma de DSC de ReTIM30 soluble a un modelo irreversible de dos estados [34]. El experimento se realizó a una concentración de proteína de 1.3 mg/mL, el gráfico está normalizado.



### *Desplegamiento químico seguido por Dicroísmo circular*

Se obtuvieron los espectros de las proteínas modelo ReTIM18 V2R, ReTIM29 y ReTIM30 en función de la concentración de agentes caotrópicos; ReTIM29 y ReTIM30 se obtuvieron de la resuspensión de cuerpos de inclusión con urea y su posterior renaturalización bajo la condición 1, ya que fue la mejor condición de recuperación para ambas proteínas. Se siguió el protocolo de desplegado proteico mediante urea y cloruro de guanidinio seguido por DC, descrito en materiales y métodos.

Es relevante mencionar las siguientes consideraciones:

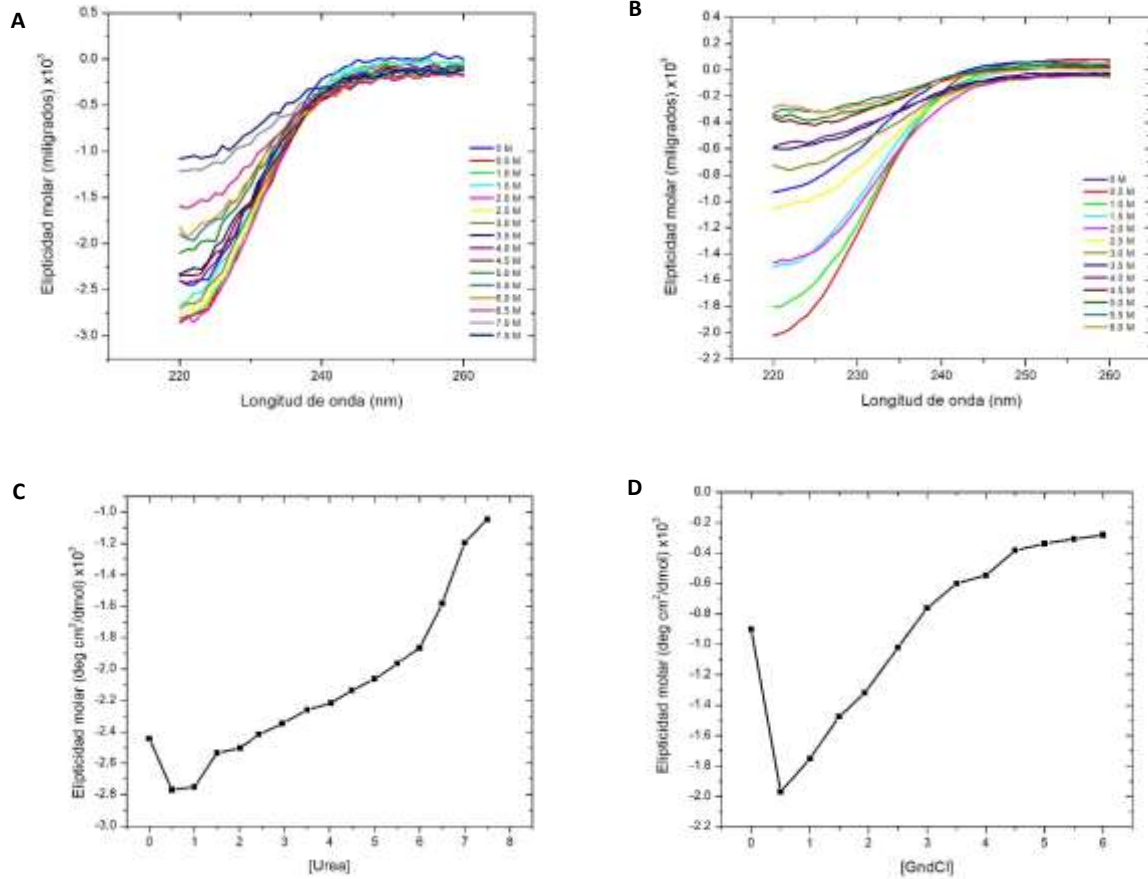
1 - Para el desplegado proteico de ReTIM18 V2R con urea y cloruro de guanidinio, ReTIM29 con urea y ReTIM30 con urea, el valor de 0 M fue anómalo, ya que mostró una mayor señal de elipticidad molar que la encontrada al agregar la primera concentración de desnaturizante, lo cual sugiere dos posibilidades: a) un error experimental, b) inicialmente hay presencia de agregados que al ser expuestos a la primera concentración de desnaturizante, logran disociarse, provocando una mejora en el contenido de estructura secundaria y en consecuencia, una menor señal de elipticidad molar.

2 - Para el desplegado proteico de ReTIM18 V2R, ReTIM29 y ReTIM30 con urea y cloruro de guanidinio, así como para el desplegado proteico de ReTIM30 soluble con urea, no fue posible encontrar una línea de pre-transición clara para el estado nativo ni una línea post-transición para el estado desnaturizado, por lo que no se conoce si la proteína está completamente plegada a 0 M del agente caotrópico o completamente desnaturizada a concentraciones altas de agente desnaturizante para poder hablar con certeza de fracción de proteína desnaturizada (Fu).

3 - En los casos de ReTIM18 V2R, ReTIM29 y ReTIM30 se observa que los valores de elipticidad molar en 0 M de desnaturizante cambiaron en función del tiempo de almacenamiento, por lo que estas proteínas no son completamente estables a las condiciones a las que se almacena.

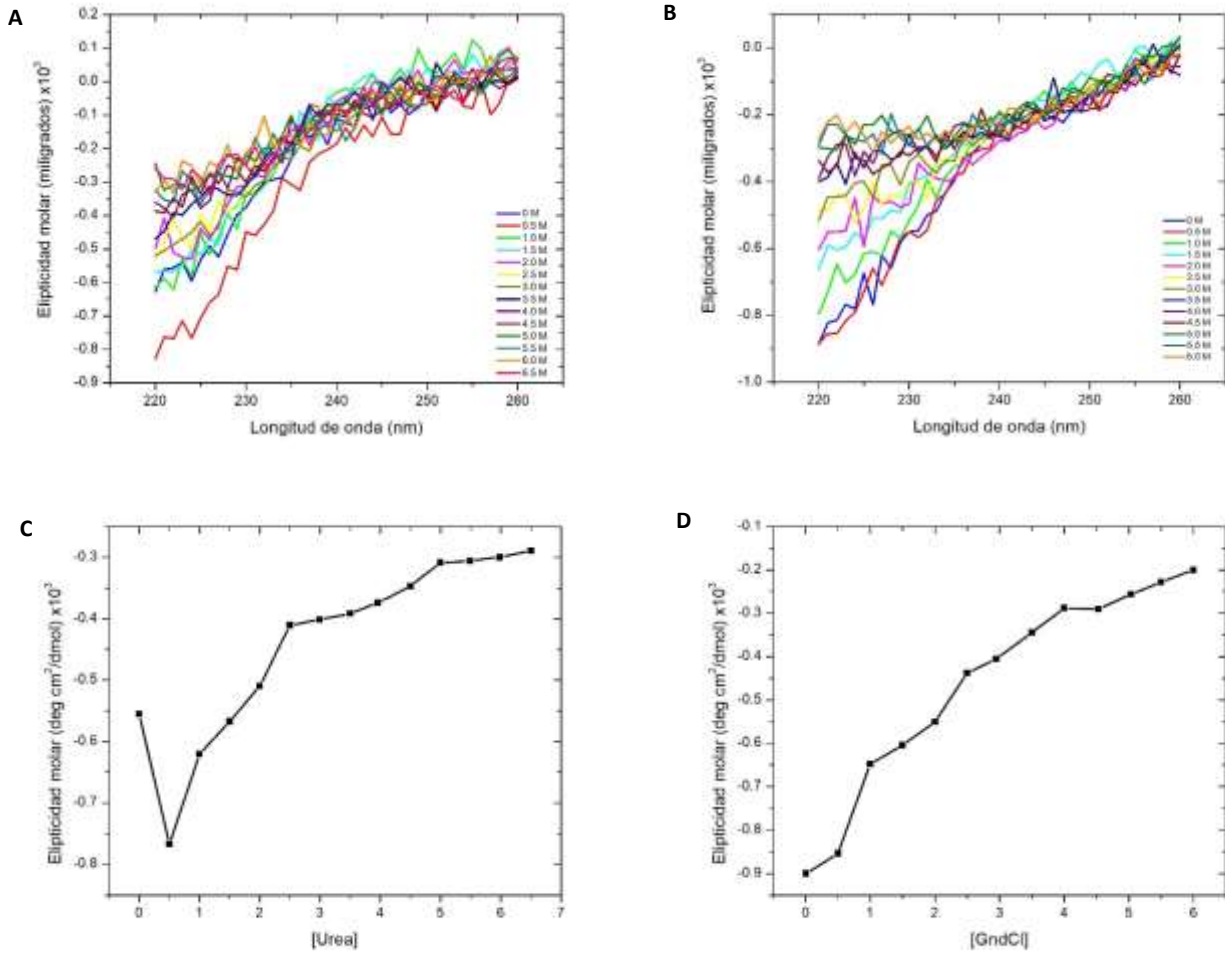
4 – Dadas las consideraciones anteriores, las transiciones encontradas no fueron analizadas de acuerdo a modelos de desplegado al equilibrio.

## ReTIM18 V2R



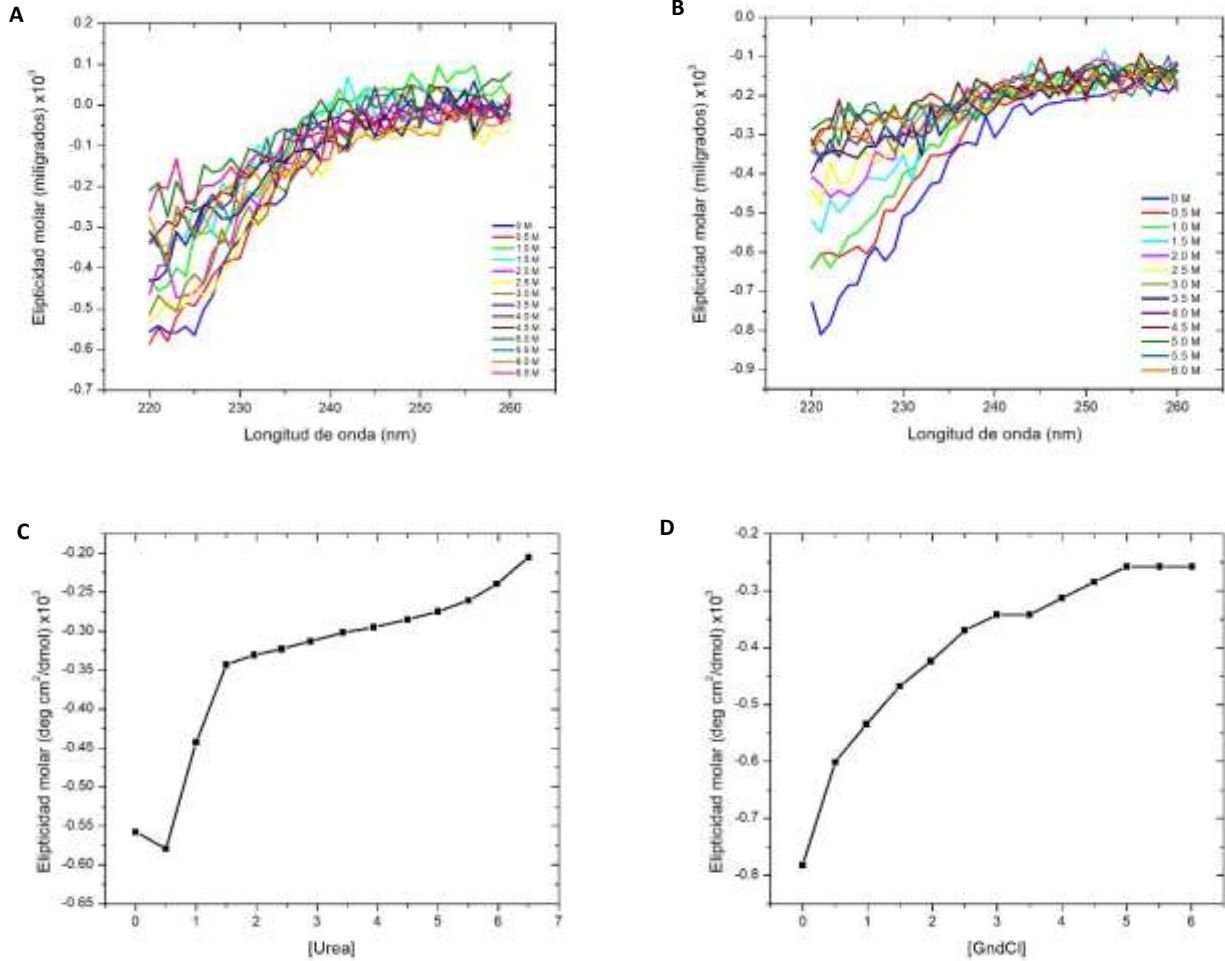
**Figura 32. El desplegamiento por agentes químicos de ReTIM18 V2R seguido por DC, no mostró una clara región de pretransición donde la proteína sea estable. A) Espectros de DC de ReTIM18 V2R incubada 24 h con urea. B) Espectros de DC de ReTIM18 V2R incubada 24 h con cloruro de guanidinio. C) Gráfico de elipticidad molar a 222 nm contra [Urea] para ReTIM18 V2R incubada con urea por 24 h. D) Gráfico de elipticidad molar a 222 nm contra [GndCl] para ReTIM18 V2R incubada con GndCl por 24 h.**

## ReTIM29



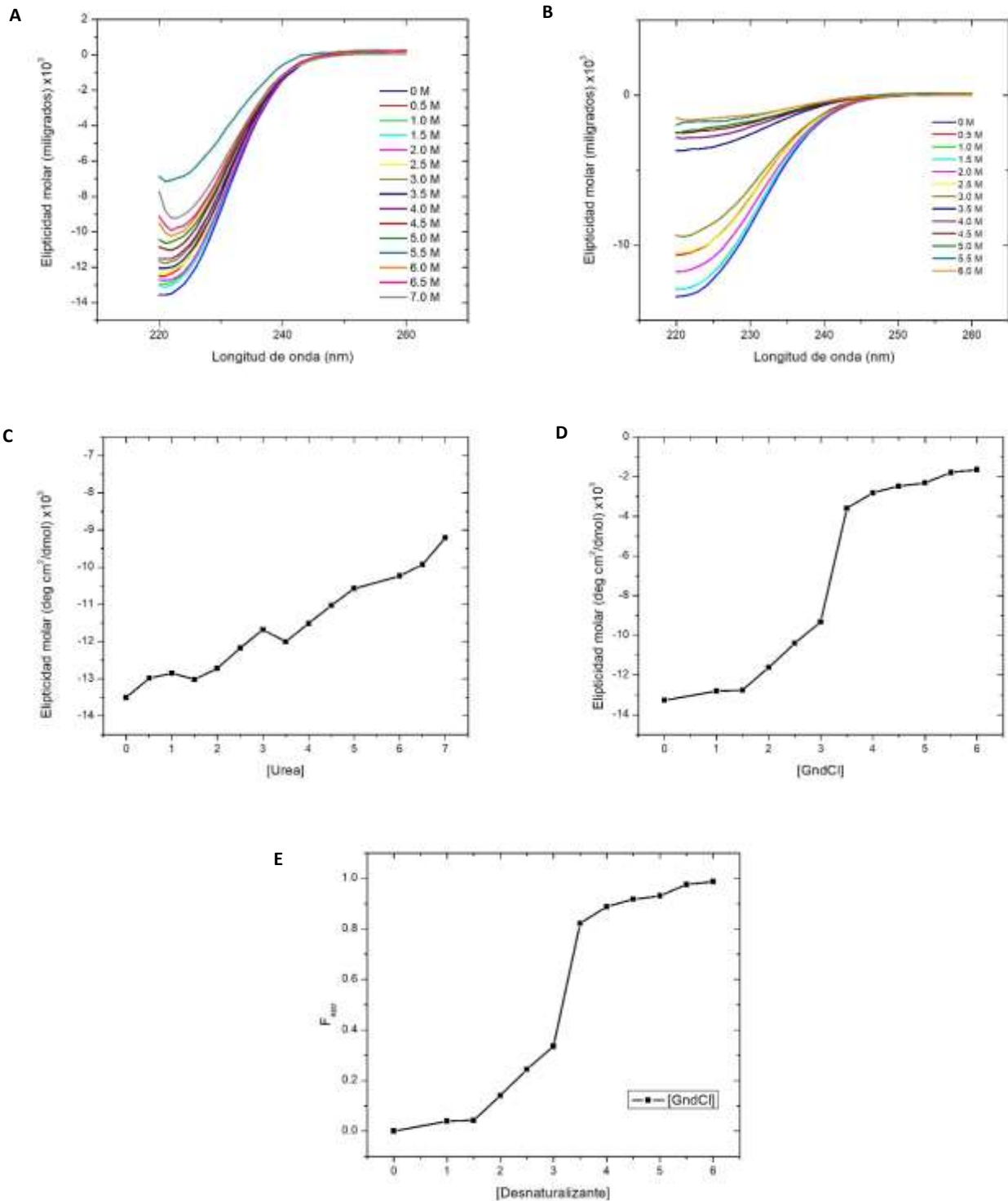
**Figura 33. El desplegamiento por agentes químicos de ReTIM29 seguido por DC, no mostró una clara región de pretransición donde la proteína sea estable. A)** Espectros de DC de ReTIM29 incubada 24 h con urea. **B)** Espectros de DC de ReTIM29 incubada 24 h con cloruro de guanidinio. **C)** Gráfico de elipticidad molar a 222 nm contra [Urea] para ReTIM29 incubada con urea por 24 h. **D)** Gráfico elipticidad molar a 222 nm contra [GndCl] para ReTIM29 incubada con urea por 24 h.

## ReTIM30



**Figura 34. El desplegamiento por agentes químicos de ReTIM30 seguido por DC, no mostró una clara región de pretransición donde la proteína sea estable. A) Espectros de DC de ReTIM30 incubada 24 h con urea. B) Espectros de DC de ReTIM30 incubada 24 h con cloruro de guanidinio. C) Gráfico de elipticidad molar a 222 nm contra [Urea] para ReTIM30 incubada con urea por 24 h. D) Gráfico de elipticidad molar a 222 nm contra [GndCl] para ReTIM30 incubada con urea por 24 h.**

## ReTIM30 Soluble



**Figura 35. El desplegamiento con GndCl de ReTIM30 soluble seguido por DC, mostró dos transiciones, una poco cooperativa y una muy cooperativa. A)** Espectros de DC de ReTIM30 soluble incubada 24 h con urea. **B)** Espectros de DC de ReTIM30 soluble incubada 24 h con cloruro de guanidinio. **C)** Gráfico de elipticidad molar a 222 nm contra [Urea] para ReTIM30 soluble incubada con urea por 24 h. **D)** Gráfico de elipticidad molar a 222 nm contra [GndCl] para ReTIM30 soluble incubada con urea por 24 h. **E)** Gráfico de cambio fraccional  $F_{app}$  contra [desnaturalizante].

Al graficar la señal de DC en el UV lejano (222 nm) contra la concentración de desnaturalizante, se encontró que ninguna de las tres proteínas muestra una clara región de pretransición en la



cual la proteína sea estable; desde las concentraciones más bajas de desnaturante, se observa una disminución gradual en señal de elipticidad. Las transiciones observadas son además muy poco cooperativas. La única excepción parece ser ReTIM30 soluble, que aparentemente tiene dos regiones de transición, una poco cooperativa (de 1.5 M a 3 M) y una muy cooperativa (de 3 M a 3.5 M), las características de esta última transición no pueden determinarse con precisión, ya que solo hay un punto en el que nos basamos para definir este comportamiento (Figura 35D). Suponiendo que la transición es verdadera, ya que este experimento se hizo con proteína y desnaturantes frescos, ReTIM30 soluble mantiene una  $C_m$  en GndCl parecida a la de ReTIM18 (Tabla 6). No se obtuvo una transición en el desplegamiento con urea de ReTIM30 soluble (Figura 35C), ya que aún a la concentración más alta de urea, todavía hay una cantidad significativa de estructura secundaria, lo que sugiere que se necesita un agente más fuerte (GndCl) para desnaturalizarla por completo. No fue posible hacer ajuste a un modelo de dos y tres estados para el desplegamiento con cloruro de guanidinio de ReTIM30 soluble, ya que el número de datos en la gráfica es muy reducido, por lo que solo podemos hablar de un  $C_m$  aparente en cloruro de guanidinio.

La Tabla 6 muestra los valores de  $C_m$  aparente de ReTIM30 soluble respecto a la proteína original:

	Urea (M)	GndCl (M)
ReTIM18	6.8	3.0
ReTIM30 (soluble)	No determinado	3.16 (aparente)

Tabla 6. Valores de  $C_m$  en urea y GndCl de ReTIM18 y de ReTIM30 soluble (valor de  $C_m$  en GndCl aparente).

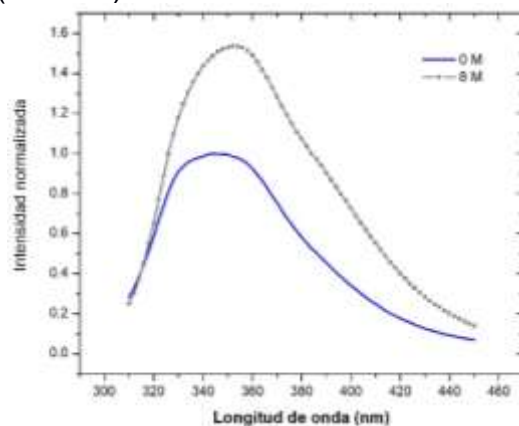
#### *Desplegamiento químico seguido por fluorescencia*

Se siguió la reacción de desplegamiento con urea mediante fluorescencia de las tres proteínas diseñadas. Típicamente cuando los triptófanos se encuentran formando parte del núcleo hidrofóbico de las proteínas en el estado nativo, el espectro de emisión de fluorescencia muestra un máximo de emisión generalmente cercano a 330 nm (corrimiento hacia el azul), en cambio, cuando se desnaturaliza la proteína, el espectro se corre hacia el rojo (cercano a 350 nm) [32]. Por otra parte, se espera que el desplegamiento se acompañe de una disminución en la intensidad de fluorescencia, ya que, la fluorescencia de los residuos aromáticos decae en ambientes acuosos, sin embargo, cuando existe apagamiento en el estado nativo (por interacción con cargas negativas, u otros aminoácidos aromáticos cercanos en la estructura), el despegamiento se acompaña de un aumento en la intensidad de fluorescencia [33]. Como ejemplo, tenemos el comportamiento de los espectros de fluorescencia para el desplegamiento inducido por urea de cuatro TIMs bacterianas. Todas ellas muestran un corrimiento al rojo en elevadas concentraciones de urea. Tres de ellas, *DrTIM*, *NpTIM* y *ScTIM*, (las TIM's de *Deinococcus radiodurans*, *Nostoc punctiforme* y *Streptomyces coelicolor*, respectivamente) muestran una disminución en la intensidad de fluorescencia, mientras que la cuarta *GoTIM* (la TIM de *Gemmata obscuriglobus*) presentó un aumento en la fluorescencia [27]. El comportamiento de *GoTIM* es similar al de ReTIM18 V2R (Figura 36) y ReTIM29 (Figura 37) que muestran un aumento de la fluorescencia de 0 M a 8 M de urea. Este comportamiento puede deberse a la transferencia de energía de residuos cargados que rodean a los triptófanos en el estado plegado, provocando un apagamiento de la fluorescencia y cuando las especies fluorescentes son expuestas al solvente en el estado desplegado, resultan en un aumento en la intensidad de fluorescencia global [50]. Además, si la energía de emisión de incluso un residuo de triptófano parcialmente enterrado en una proteína plegada generalmente cambia al rojo al desplegarse, no es posible predecir cuál será el efecto de la exposición al solvente sobre

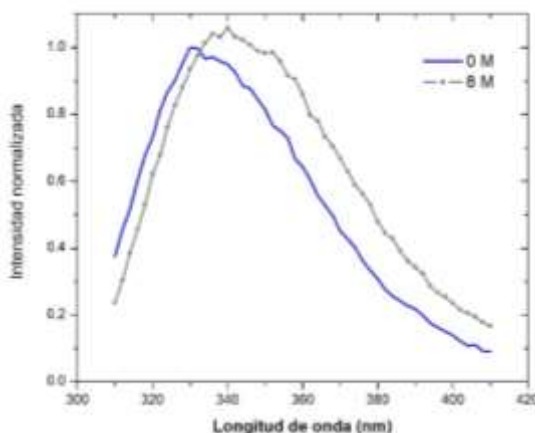
el rendimiento cuántico (intensidad total) de la emisión. Esto se debe al hecho de que el rendimiento cuántico de los triptófanos en las proteínas se ve afectado por apagamiento y fenómenos de transferencia de energía [33].

En ReTIM18 V2R y ReTIM29 hay un corrimiento hacia el rojo acompañado de un aumento en la intensidad de fluorescencia en el estado desplegado, este aumento es más pronunciado en ReTIM18 V2R que en ReTIM29. Debido a que es difícil conocer la contribución individual de cada residuo aromático sobre la intensidad total en el estado desplegado, no podemos dar una interpretación acertada del por qué ReTIM18 V2R a 8 M de urea tiene un aumento más grande en la intensidad total que ReTIM29 a 8 M de urea.

El centro espectral de masa o SCM por sus siglas en inglés (Central of Spectral Mass), es un parámetro que mide el promedio ponderado de la longitud de máxima emisión (ver materiales y métodos). Las SCM en el estado nativo de ReTIM18 V2R y ReTIM29 son de 360 nm y de 346.7 nm respectivamente, mientras que las SCM del estado desplegado son de 366 nm y de 352.1 nm respectivamente (Tabla 7).



**Figura 36.** El espectro de desplegamiento por urea de ReTIM18 V2R seguido por fluorescencia muestra un atípico aumento de la intensidad a 8 M de urea. En color azul se muestra el espectro a 0 M de urea y en gris el espectro a 8 M de urea. Ambos espectros están normalizados.

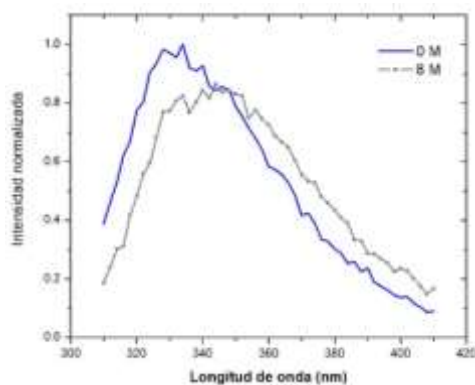


**Figura 37.** El espectro de desplegamiento por urea de ReTIM29 seguido por fluorescencia muestra un aumento de la intensidad a 8 M de urea similar a ReTIM18 V2R, pero de menor magnitud. En color azul se muestra el espectro a 0 M de urea y en gris el espectro a 8 M de urea. Ambos espectros están normalizados.

ReTIM30 mostró un decaimiento de la intensidad de 0 M a 8 M de urea y un corrimiento hacia el rojo en su longitud de onda máxima (Figura 38 y Tabla 7), comportamientos típicos del

despliegamiento proteico de las TIM's bacterianas *Dr*TIM, *Np*TIM, *Sc*TIM; su SCM en el estado nativo es diferente del SCM del estado desplegado (346.5 y 353.2 respectivamente) (Tabla 7).

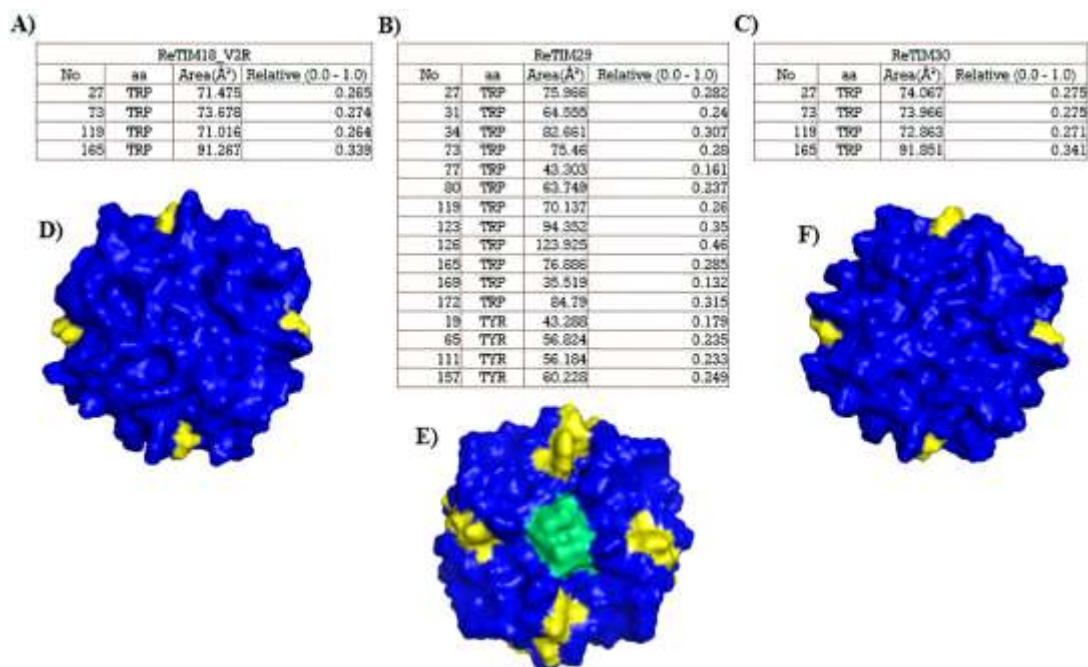
La Figura 39 muestra cualitativa (incisos D, E y F) y cuantitativamente (incisos A, B y C) el área de exposición al solvente de las tres proteínas diseñadas. Observamos que cada residuo aromático, a pesar de estar en posiciones simétricas y equivalentes a lo largo de cada cuarto de la proteína, presentan un área y porcentaje de exposición al solvente diferentes, lo que nos dice que cada residuo tiene una contribución distinta a la intensidad global de la proteína en el estado nativo. En principio, los residuos aromáticos más expuestos al solvente en el estado nativo tienen menor contribución a la intensidad global. El área máxima que puede exponer un triptófano es de 269.71 Å<sup>2</sup> y de 241.87 Å<sup>2</sup> para una tirosina. ReTIM18 V2R tiene 4 triptófanos de los cuales el W119 es el menos expuesto al solvente (71.016 Å<sup>2</sup>), por lo cual se esperaría que dicho residuo tuviera una contribución mayor en la intensidad global, sin embargo, solo podemos especularlo, ya que dicha hipótesis tendría que comprobarse haciendo mutantes que sustituyan cada especie fluorescente por residuos no fluorescentes de tamaño y propiedades químicas similares para conocer la contribución de cada residuo aromático a la intensidad total. Este mismo criterio podría aplicarse para conocer el comportamiento de cada residuo aromático cuando se encuentra completamente expuesto al solvente en el estado desnaturalizado, ya que, en las tres proteínas diseñadas, hay un corrimiento hacia el rojo, lo que indica un cambio en el ambiente de las especies fluorescentes, pero no podemos saber con certeza qué residuos aromáticos (específicamente los triptófanos, ya que las muestras se excitaron a 295 nm) son los responsables del aumento (en ReTIM18 V2R y ReTIM29) o disminución (en ReTIM30) de la intensidad de fluorescencia.



**Figura 38. El espectro de desplegamiento por urea de ReTIM30 seguido por fluorescencia muestra un decaimiento de la intensidad a 8 M de urea.** En color azul se muestra el espectro a 0 M de urea y en gris el espectro a 8 M de urea. Ambos espectros están normalizados.

Modelo	Número de W	Número de F	Número de Y	$\lambda_{max}$ Nativo (nm)	SCM Nativo (295 nm)	$\lambda_{max}$ Despl. (nm)	SCM Despl. (295 nm)	$\Delta \lambda_{max}$ (nm)	$\Delta$ SCM (295 nm)
ReTIM18 V2R	4	0	0	344	360	352	366	8	6
ReTIM29	12	0	4	330	346.7	340	352.1	10	5.4
ReTIM30	4	0	0	334	346.5	344	353.2	10	6.7

**Tabla 7. Número de triptófanos (W), número de fenilalaninas (F), número de tirosinas (Y),  $\lambda_{max}$  y SCM a 295 nm en los estados nativo y desplegado de las tres proteínas diseñadas.**



**Figura 39. Aminoácidos aromáticos de los tres modelos con área expuesta al solvente (Å<sup>2</sup>) y porcentaje de exposición (relativa).** A) Datos de los triptófanos del modelo ReTIM18 V2R. B) Datos de los triptófanos y tirosinas del modelo ReTIM29. C) Datos de los triptófanos del modelo ReTIM30. D) Imagen de ReTIM18 V2R se observa desde la parte superior, donde se encuentran los extremos carboxilos terminales de las hebras β (“vista desde arriba”). Los triptófanos están coloreados en amarillo. E) Imagen de ReTIM29 “vista desde arriba” con los triptófanos coloreados en amarillo y tirosinas en verde. F) Imagen de ReTIM30 “vista desde arriba” con los triptófanos coloreados en amarillo. El resto de la proteína está coloreado en azul [30].

En dos de los tres casos, la interpretación de los resultados obtenidos para el desplegamiento con urea seguido por fluorescencia muestra dificultades por el número elevado y la ubicación de los residuos de triptófano en las proteínas estudiadas, ya que Rosetta los diseñó para estar parcialmente expuestos. Sin embargo, la información es complementaria a la obtenida por DC.

## Discusión

Como se mencionó en la introducción, ReTIM18 se agrega en tiempos cortos de almacenamiento, lo cual fue heredado de forma notoria al diseño ReTIM18 V2R que presenta dos picos, uno de ellos en un volumen de elución que corresponde a un mayor radio de Stokes y un mayor peso molecular que el esperado para un monómero compacto y que muy posiblemente sean agregados (Figura 19A). El experimento de cromatografía de exclusión molecular fue repetido para ReTIM18 V2R con una noche de almacenamiento (requerida para la diálisis) con el objetivo de corroborar si podía disminuirse la formación de agregados. Los resultados indican que el número de agregados se redujo considerablemente (Figura 19B), con lo que se corrobora que ReTIM18 V2R heredó la agregación en tiempos cortos luego de ser purificada, tal como el diseño ReTIM18 original. Las propiedades hidrodinámicas de ReTIM29 (Figura 20) y ReTIM30 (Figura 21A) muestran proteínas con el tamaño y peso esperado, comparándola con ReTIM18 original, contra ReTIM18 V2R y entre sí mismas; sin embargo, hay presencia de agregados de tamaño similar a los agregados de ReTIM18 V2R (Tabla 3). Las propiedades hidrodinámicas fueron calculadas también para ReTIM30 soluble y el resultado fue la obtención de un cromatograma más definido en comparación con ReTIM30 recuperada de cuerpos de inclusión (Figura 21B). Los Radios de Stokes y pesos moleculares fueron muy similares en ambas proteínas: ReTIM30 soluble ( $R_s$ : 23.6 Å, PM: 23.4 kDa) ReTIM30 insoluble ( $R_s$ : 24 Å, PM: 24.5 kDa) (Tabla 3).



Mediante el servidor AGGREGSCAN 3D (A3D) [52], fue comprobada la propensión a la agregación de las proteínas diseñadas. Los resultados se muestran en los gráficos e imágenes de la Figura 42 del material suplementario. A3D predijo mediante una puntuación media (indicador normalizado que permite comparar la solubilidad de diferentes estructuras proteicas. Cuanto más negativo es el valor, mayor es la solubilidad normalizada), que ReTIM18 V2R es más soluble que ReTIM18 (puntuaciones medias = -1.31 y -1.27 respectivamente), lo cual concuerda con la obtención proteína soluble en la sobreexpresión, pero no logra explicar la alta propensión a la agregación después de ser purificada (la agregación observada es mayor que ReTIM18), tal como se observó en la experimentación. De hecho, los puntajes individuales para los residuos mutados en ReTIM18 V2R, resultaron tener una mayor solubilidad (puntajes más negativos) que los residuos originales (R2 = -3.8, R48 = -3.6, R94 = -2.8, R140 = -3.1 contra V2 = -2.4, V48 = 0, V94 = 1.6, V140 = -0.3), esto no explica por qué esta proteína se agrega en tiempos cortos tras su purificación, pero sí su solubilidad después de la sobreexpresión. ReTIM29 resultó ser, de las tres proteínas diseñadas, la más propensa a la agregación, ya que su puntuación media es de -1.07, que es mayor a -1.27 de ReTIM18. Este valor explica su baja solubilidad tras la sobreexpresión y su localización en cuerpos de inclusión. Las formaciones de agregados detectados por dos picos en la cromatografía de exclusión molecular de ReTIM29 (Figura 20) se pueden explicar mediante la detección de regiones con alta propensión a la agregación, que se muestran en color rojo en la Figura 42F del material suplementario. ReTIM30 obtuvo una puntuación media de -1.08 que es mayor a ReTIM18 (-1.27) y casi igual a ReTIM29 (1.07); esto explica la baja solubilidad que presenta tras la sobreexpresión. Aparentemente, por su puntuación media prácticamente igual a la de ReTIM29, se esperaría que ReTIM30 se encontrara totalmente en cuerpos de inclusión, pero a 20 °C de temperatura de inducción, se obtuvo proteína soluble (Figura 16A), lo cual se explica con una menor presencia de regiones propensas a la agregación (regiones en color rojo de la Figura 42H) en comparación con ReTIM29 (Figura 42F). Este dato también explica la presencia de picos correspondientes a agregados en sus cromatogramas (Figuras 21A y 21B).

Los resultados de los espectros de DC para todas las proteínas diseñadas, así como sus deconvoluciones, muestran que el desplegamiento térmico es irreversible, ya que sus espectros a 20 °C, después de ser calentadas a 95 °C, no tienen una forma semejante al espectro nativo a 20 °C. ReTIM18 V2R mostró una transición en el desplegamiento térmico monitoreado por DC, cerca de los 89 °C (Figura 23), ReTIM29 recuperada de cuerpos de inclusión bajo la condición 1 mostró una ligera transición a los 65 °C y cuando se recuperó bajo la condición 5, de 75 °C (Figura 25A y 25B); es probable que ReTIM29 recuperada de cuerpos de inclusión bajo la condición 1 no alcance un estado nativo luego de ser resuspendida con urea y posteriormente renaturalizada, ya que de ser así, no habría variaciones en la  $T_m$  respecto a ReTIM29 recuperada bajo la condición 5. Para las proteínas ReTIM30 en las condiciones 1 y 5 de resuspensión en urea (Figura 27A y 27B); y para ReTIM30 soluble (Figura 29), no hubo transición en el desplegamiento seguido por dicroísmo circular.

El desplegamiento térmico de ReTIM18 V2R seguido por DSC fue irreversible y mostró un  $\Delta H$  = 1.9 kcal/mol y una  $T_m$  de 89.2 °C, similar a la obtenida por DC (~ 89 °C). Comparando estos valores con los valores de ReTIM18 original ( $T_m$  = 92.9 °C,  $\Delta H$  = 76.6 kcal/mol), es notorio ver que este diseño no aumentó la estabilidad del diseño original. Cabe destacar que estos valores se obtuvieron con proteína a un mayor tiempo de almacenamiento (2 días) por lo que es posible que estos valores aumenten si se determinan con proteína recién purificada.



No se observó una transición de ReTIM29 monitoreando el desplegamiento térmico por DSC, sin embargo; lograron observarse transiciones seguidas por DC en proteínas recuperadas de cuerpos de inclusión y en las que la urea fue diluida en las condiciones 1 y 5; las transiciones fueron de 65 y 75 °C respectivamente.

En el caso de ReTIM30, no se observó transición alguna de desplegamiento por temperatura seguido por DSC, lo que, junto con los experimentos de desplegamiento térmico seguido por DC, nos dice que tiene un desplegamiento no cooperativo o bien que inicialmente tiene cierto grado de desplegamiento debido a que fue rescatada de cuerpos de inclusión y no se replegó completamente luego de ser resuspendida en urea.

Por su parte, la ReTIM30 soluble mostró una endoterma irreversible muy pequeña y una  $T_m = 105.2$  °C lo que explica que no se haya detectado alguna transición en el experimento de desplegamiento térmico por DC, ya que sólo se alcanza una temperatura de 100 °C. Su  $\Delta H$  fue de 0.26 kcal/mol que está muy por debajo de la  $\Delta H$  de la proteína original (76.6 kcal/mol). Lo interesante es saber por qué a pesar de tener una  $T_m$  mayor a la de la proteína original (105.2 °C vs 92.9 °C), su  $\Delta H$  es extremadamente bajo. Para intentar explicar este fenómeno tenemos que mencionar la naturaleza química de las mutaciones que Rosetta propuso para la reconstrucción de la interfase entre las  $\alpha$  hélices 7 y 8. Las siguientes son las mutaciones de dicha zona para un cuarto de la proteína: R13M, V22T, S24A y I29V. Como podemos observar, algunas de estas mutaciones favorecen interacciones de tipo hidrofóbico, y no interacciones de tipo electrostático, las cuales en principio son las principales responsables del valor del  $\Delta H$  en el desplegamiento de proteínas. En especial la mutación R13M quizá fue la que influyó mayoritariamente en la disminución drástica del  $\Delta H$ , ya que originalmente, la arginina forma dos puentes de hidrógeno con un ácido aspártico vecino, y si tomamos en cuenta la simetría de orden 4, en total hay 8 puentes de hidrógeno en la ReTIM18 original. Al hacer las mutaciones R13M, R59M, R105M y R151M, se eliminan 8 puentes de hidrógeno y la contribución del  $\Delta H$  que aportaban originalmente, desaparece. Es natural suponer que la principal contribución al efecto térmico de la desnaturalización se debe a la ruptura de los enlaces de hidrógeno intramoleculares, ya que, durante el proceso de reordenamiento conformacional de proteínas en solución, un cambio considerable de la energía total de las interacciones de van der Waals aparentemente no puede ocurrir. En cuanto a las interacciones de la proteína desplegada con el agua, se reducen al formarse nuevos enlaces de hidrógeno entre grupos polares de la proteína. Por lo tanto, resulta que la dependencia de la temperatura y la entalpía está relacionada principalmente con la capacidad de los grupos de proteínas para interactuar con el agua, pero principalmente a los grupos no polares de proteínas que entran en contacto con el agua en el desplegamiento del glóbulo [47]. Es posible que la baja entalpía de desplegamiento de ReTIM30 soluble se deba a que, a altas temperaturas, exista estructura residual donde halla la presencia de núcleos hidrofóbicos que no están totalmente expuestos al solvente. Según la evidencia recabada, ReTIM30 soluble está completamente plegada en el estado nativo, y si a ello sumamos las mutaciones de carácter polar sin carga e hidrofóbicas, notamos que, el carácter hidrofóbico de ReTIM30 soluble es mayor que ReTIM18, lo que puede estar favoreciendo la permanencia de núcleos hidrofóbicos en el estado desplegado y disminuir la entalpía de desplegamiento. Este fenómeno también se ha reportado con un barril TIM de la misma familia de nuestros diseños llamado NovoTIM14, que posee una entalpía baja ( $\Delta H = 14$  kcal/mol) comparada con ReTIM18 ( $\Delta H = 76.6$  kcal/mol), donde su baja entalpía se debe a que el estado desplegado tiene estructura residual [51]. Una forma de corroborar este comportamiento en NovoTIM14, fue agregar concentraciones crecientes de urea al desplegamiento por temperatura, ya que, en principio, de esta manera se puede desplazar el

estado desplegado a valores inferiores de temperatura. Lo natural es esperar que el valor de entalpía de desplegamiento decaiga a altas temperaturas, pero en vez de eso, los valores de entalpía de desplegamiento aumentaron y la  $T_m$  se mantuvo en un valor constante. Este resultado atípico probablemente esté relacionado con cambios en la estructura residual del estado desplegado, particularmente, con la exposición incrementada de residuos no polares en el estado desplegado en presencia de urea [51]. Estos resultados sugieren fuertemente que la razón de la baja entalpía de desplegamiento en NovoTIM14 es el alto contenido de estructura residual en el estado desplegado y podría estar sucediendo lo mismo con ReTIM30 soluble. Como perspectiva, puede sugerirse hacer este experimento con ReTIM30 soluble y corroborar esta hipótesis.

Los resultados del desplegamiento químico seguido por DC, revelan aparentemente dos transiciones para ReTIM30 soluble, una transición poco cooperativa entre 1.5 y 3.2 M de GndCl y una transición muy cooperativa entre 3.2 a 3.5 M de GndCl (Figura 35E), lo que indica que el desplegamiento aparentemente se rige por un modelo de tres estados con un intermediario, lo cual no pudo corroborarse, ya que el reducido número de datos es insuficiente para realizar un ajuste adecuado. De acuerdo a la  $C_m$  aparente en GndCl obtenida para ReTIM30 soluble es muy similar a la  $C_m$  de ReTIM18 en GndCl (3.2 M y 3 M respectivamente) (Tabla 6). Por otro lado, las proteínas ReTIM29 y ReTIM30 no mostraron regiones de pretransición donde las proteínas sean estables, ni una transición cooperativa en el desplegamiento químico con urea y GndCl (Figuras 33E y 34E).

Las evidencias de la falta de una transición en el desplegamiento térmico seguido por DC, la ausencia de un pico en las trazas calorimétricas obtenidas por DSC y de un desplegamiento a bajas concentraciones de urea y GndCl, indican que las proteínas ReTIM29 y ReTIM30 no se repliegan completamente luego de ser resuspendidas en urea.

De acuerdo con las ecuaciones paramétricas, una proteína de 184 residuos como las ReTIMs, puede alcanzar un  $\Delta H = 128.4 \pm 3.5$  kcal/mol y un  $\Delta C_p = 2.5 \pm 0.1$  kcal mol<sup>-1</sup> K<sup>-1</sup> [28]. Para ReTIM18 original,  $\Delta H = 76.6$  kcal/mol, lo que demuestra que aún existe posibilidad de aumentar estos parámetros fisicoquímicos. El inconveniente con esta proteína es que su desplegamiento es térmicamente irreversible, se agrega en tiempos cortos y la disminución en su  $\Delta H$  con tiempos prolongados de almacenamiento, sugiere su desnaturalización durante el mismo. Los modelos propuestos en teoría deberían de contrarrestar estos efectos, pero con la evidencia reunida, nos damos cuenta de que no fue así. ReTIM18 V2R continúa agregándose en tiempos cortos de almacenamiento a una concentración mayor de 1 mg/mL y pH 8.0; una posible explicación a este fenómeno es la siguiente: sabemos que la agregación de proteínas es un fenómeno que va más allá de la solvatación, ya que incluye no solo la energía de la interacción con el disolvente, sino también las interacciones no específicas de la proteína consigo misma, dicho lo anterior, los términos de energía en el potencial de Rosetta solo proporcionan una descripción precisa de los efectos de exposición al solvente una vez que la proteína se despliega, pero no describen las interacciones intermoleculares inespecíficas de la proteína consigo misma. Una versión más nueva del potencial de Rosetta que contemple las interacciones inespecíficas de la proteína consigo misma, podría ayudar a superar esta limitación. [31]. Una forma de tomar en cuenta estas interacciones, es empleando la estrategia que usa el servidor AGGRESAN3D (A3D), algoritmo que se basa en un conjunto de resultados obtenidos de la agregación de mutantes del péptido- $\beta$  amiloide. A3D explora la propensión a la agregación de cada residuo individual en la proteína de acuerdo a su posición relativa en la secuencia, además, toma en cuenta regiones predispuestas a la agregación que

están enterradas en el núcleo hidrofóbico o participan en las interacciones cooperativas no covalentes que sostienen la estructura proteica secundaria y terciaria en proteínas globulares plegadas de forma nativa. La propensión intrínseca de agregación para cada aminoácido en la proteína está modulada por su contexto estructural específico y se calcula para regiones esféricas centradas en el carbono alfa de cada residuo. Esto proporciona un valor único de agregación estructuralmente corregido (puntuación A3D) para cada aminoácido en la estructura, que se formula como:

$$A3D \text{ score} = Agg_i \times (\alpha \times e^{\beta \times RSA_i}) + \sum [Agg_e \times (\alpha \times e^{\beta \times RSA_e}) \times (\gamma \times e^{-\delta \times dist})]$$

donde  $Agg_i$  es la propensión a la agregación intrínseca del residuo en el centro de la esfera y  $RSA_i$  su superficie relativa expuesta a un solvente.  $Agg_e$  es la propensión de agregación intrínseca de cada residuo adicional incluido en la esfera,  $RSA_e$  es su área de superficie relativa expuesta al solvente y  $dist$  su distancia al residuo central  $i$  [52]. Rosetta podría implementar estos cálculos para tomar en cuenta la propensión a la agregación de los diseños que propone. Es importante resaltar que Rosetta solo optimiza la estabilidad termodinámica, sin tener en cuenta las vías y la cinética de plegado [31], lo que, de ser implementado, enriquecería aún más este algoritmo.

Como se observa en los resultados obtenidos, ningún modelo obtenido por ROSETTA pudo superar el  $\Delta H$  del modelo original, ni lograr un desplegamiento térmico reversible, lo que posiblemente se debió a que el puntaje arrojado por ROSETTA de los tres modelos seleccionados no fue el mejor criterio de selección. Cuando se generan mutaciones, hay un cambio en el número de átomos de la proteína, por lo que la energía no es completamente comparable entre los modelos generados y el diseño original. Para que las energías de los modelos sean comparables entre sí, se realizó un cálculo robusto a partir de los puntajes obtenidos con los softwares ROSETTA y MOPAC que están descritos en la sección de materiales y métodos. Los resultados de estos cálculos se encuentran en las Tablas 10, 11 y 12 del material suplementario.

De acuerdo a los resultados de la Tabla 10, no hay unanimidad para elegir el modelo con el asa superior reconstruida que tenga los mejores puntajes de ROSETTA y MOPAC, sin embargo, los tres modelos con puntajes favorables son: DATDKT, DATDRT y DATDRD (ReTIM18 V2R). De los tres diseños, sólo uno fue el que elegimos como el modelo con mejores puntajes (DATDRD), aunque en la práctica experimental no logró superar los parámetros fisicoquímicos de ReTIM18.

Los resultados de la Tabla 11 exponen varios modelos que poseen puntajes favorables de ROSETTA y MOPAC, pero el mejor modelo por unanimidad fue el modelo 36, mientras que elegimos la secuencia número 35 (ReTIM29) para su estudio, que a pesar de tener un buen resultado según MOPAC, fue el peor modelo según el cálculo robusto de ROSETTA. Los resultados obtenidos en la práctica experimental muestran que al menos, ReTIM29 recuperada de cuerpos de inclusión, fue el modelo con parámetros fisicoquímicos más desfavorables en cuanto a estabilidad se refiere.

Finalmente, los resultados de la Tabla 12, no muestran unanimidad para elegir un modelo que tenga puntajes favorables en ROSETTA y MOPAC, pero el modelo con la secuencia número 3 es el que tuvo el mejor resultado según MOPAC y el tercero mejor según ROSETTA, por lo que puede considerarse que, de esta serie de modelos, es el que tiene puntajes lo suficientemente favorables en los cálculos robustos con ambos softwares. Elegimos el modelo

con la secuencia 2 (ReTIM30) como modelo de estudio, que a pesar de tener el mejor resultado según Rosetta, su resultado en Mopac fue malo. En la práctica experimental fue el único modelo que superó la  $T_m$  de ReTIM18, sin embargo, su  $\Delta H$  fue muy bajo.

La Tabla 8 contiene información de los puentes de hidrógeno formados en los modelos con mutaciones en las tres regiones de mejora estructural y que poseen mejores evaluaciones. Debido a que el  $\Delta H$  se rige principalmente por interacciones de tipo electrostático, la búsqueda se enfocó en encontrar puentes de hidrógeno. Cabe mencionar que el número de puentes de hidrógeno formados es relativo a los modelos con los mejores puntajes de Rosetta y Mopac de cada región de mejora estructural, por ejemplo, en la región de la interfase alfa hélice 7 – alfa hélice 8 reconstruida, se comparan los puentes de hidrógeno formados solo entre los modelos con las secuencias 2 (ReTIM30) y 3.

En la serie de modelos expuestos en la Tabla 8, los modelos con las secuencias DATDRD (ReTIM18 V2R) y DATDRT formaron 6 puentes de hidrógeno. Se puede especular (ya que no existen pruebas experimentales), que el modelo con la secuencia DATDRT hubiera obtenido parámetros fisicoquímicos muy similares a ReTIM18 V2R, por lo que se podría especular que la zona del asa superior no era una buena candidata para mejorar el  $\Delta H$  del modelo original. Una posible explicación de este fenómeno es la ubicación espacial de los puentes de hidrógeno formados. En este caso los puentes de hidrógeno formados en el modelo ReTIM18 V2R están totalmente expuestos al solvente y es posible que, por fracciones de segundo, dichos puentes de hidrógeno se rompan para interactuar con el solvente, lo que hace que no aumente de forma significativa el  $\Delta H$  de la proteína. Quizá en un futuro se deban utilizar técnicas de dinámica molecular para obtener una mayor confiabilidad en la ubicación de zonas de mejora estructural.

En la Tabla 8 se puede observar que el modelo con la secuencia 36 supera a ReTIM29 por 7 puentes de hidrógeno. Este dato se corrobora con el análisis hecho con Mopac y Rosetta, donde observamos que el modelo con la secuencia 36 fue por unanimidad el mejor modelo de esta serie de diseños. Quizá el modelo con la secuencia 36 sea un buen candidato para mejorar el  $\Delta H$  de la proteína, ya que tiene una gran cantidad de puentes de hidrógeno formados que, aunque están expuestos al solvente, la suma de todos ellos puede repercutir favorablemente en el  $\Delta H$ . Cabe mencionar que la secuencia 36 se perfiló como el mejor diseño con un cuarto de la proteína reconstruida luego de realizar el cálculo robusto con Mopac y Rosetta, por lo que este diseño no fue construido, ya que los 3 diseños estudiados, fueron seleccionados con cálculos preliminares y no con cálculos robustos.

La Tabla 8, nos muestra que el modelo con la secuencia 3 tiene 3 puentes de hidrógeno más que ReTIM30 (secuencia 2), pero cabe mencionar que estos puentes de hidrógeno no se deben a nuevas interacciones diseñadas por Rosetta, sino al hecho de que no se eliminaron los puentes de hidrógeno ya existentes debido a mutaciones de residuos de tipo polar por residuos neutros o residuos polares sin carga como es el caso de ReTIM30. Analizando esta serie de diseños nos dimos cuenta que con Rosetta se optimizó la creación de interacciones de tipo hidrofóbico o polares sin carga, lo que no aumenta significativamente el  $\Delta H$  de la proteína, pero quizá sí su  $T_m$  como fue el caso de ReTIM30. Las siguientes son las mutaciones para un cuarto de la proteína de ReTIM30 y del modelo con la secuencia 3 respectivamente y se muestran para ejemplificar esta hipótesis: R13M, V22T, S24A y I29V (ReTIM30). E10D, V22L, S24A, I29V, A33G, K36Y (Secuencia 3).

<b>Modelos con asa superior reconstruida</b>		
Secuencia DATDKT		
Residuos involucrados	Tipo de puente de hidrógeno	Distancia (Angstroms)
K2 - D25	1	2.5
K2 - D25	1	2
K140 - D164	1	2
Total de puentes de hidrógeno		3
Secuencia DATDRD (ReTIM18 V2R)		
Residuos involucrados	Tipo de puente de hidrógeno	Distancia (Angstroms)
R2 - D25	2	1.8
R2 - D25	3	2.2
R140 - D163	2	1.7
R140 - D163	3	2.3
R166 - D141	2	1.9
R166 - D141	3	2.4
Total de puentes de hidrógeno		6
Secuencia DATDRT		
Residuos involucrados	Tipo de puente de hidrógeno	Distancia (Angstroms)
R2 - D25	2	1.8
R2 - D25	3	2.2
R94 - D124	2	1.8
R94 - D124	3	2.3
R140 - D163	2	1.7
R140 - D163	3	2.2
Total de puentes de hidrógeno		6
<b>Modelos con un cuarto de la proteína reconstruida</b>		
Secuencia 35 (ReTIM29)		
Residuos involucrados	Tipo de puente de hidrógeno	Distancia (Angstroms)
R6 - D3	2	2
R6 - D3	3	2.2
K14 - E10	4	1.9
S39 - I20	5	2.2
S30 - I20	6	2.3
S3 - Y19	7	2
R52 - D49	2	2
R52 - D49	8	2.5
R52 - E78	9	2.2
Q54 - D25	10	2
K60 - E56	4	2.6
Q64 - G62	11	2.3
S92 - D71	12	2
R98 - D95	2	1.9
K106 - E102	4	2.5
S131 - I112	5	2
R144 - D141	8	1.9

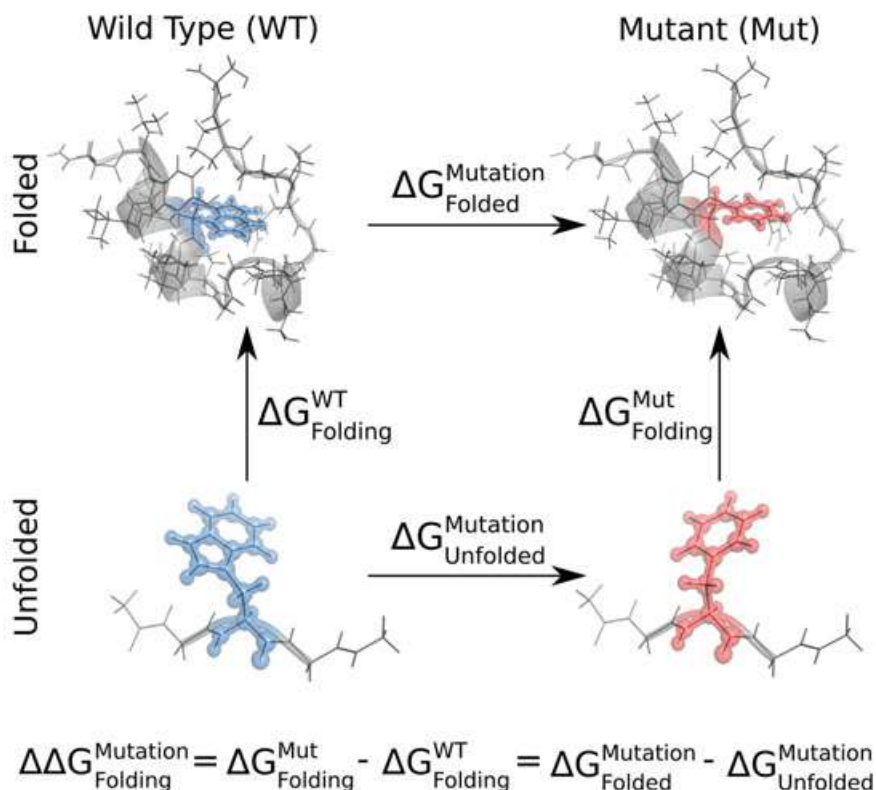


R152 - E148	13	2.6
Q156 - G154	11	2.1
Total de puentes de hidrógeno		19
Secuencia 36		
Residuos involucrados	Tipo de puente de hidrógeno	Distancia (Angstroms)
R4 - D1	2	2
R4 - D1	3	2.1
R6 - D3	2	2
R6 - D3	3	2.2
K14 - E10	4	2.3
R23 - D44	2	1.8
R23 - D44	8	1.9
R23 - D182	3	1.9
R50 - D25	2	2.9
R50 - Y54	14	2
R50 - Y54	15	2
R52 - D49	2	2.1
K60 - E56	4	2.5
S92 - D71	12	2
R96 - Y54	14	2
R96 - Y54	15	2
R98 - D95	2	1.8
K106 - E102	4	2.5
N137 - D139	16	2.5
R142 - D139	2	1.8
R144 - P140	17	2.1
R144 - P140	18	1.9
R144 - E170	13	2.2
K152 - E148	4	2.5
N183 - D163	19	1.9
S184 - D 182	12	1.8
Total de puentes de hidrógeno		26
<b>Modelos con la interfase alfa hélice 7 y 8 reconstruida</b>		
Secuencia 2 (ReTIM30)		
Residuos involucrados	Tipo de puente de hidrógeno	Distancia (Angstroms)
E10 – K36	20	1.9
K82 – E56	4	1.9
K128 – E102	4	1.9
K174 – E148	4	1.9
Total de puentes de hidrógeno		4
Secuencia 3		
Residuos involucrados	Tipo de puente de hidrógeno	Distancia (Angstroms)
R13 - D10	3	1.8
R13 - D10	8	2.4
D56 - R59	21	2.3

D56 - R60	21	1.9
R106 - D102	8	2.3
R106 - D102	3	1.8
D148 - R151	22	2.3
Total de puentes de hidrógeno		7

**Tabla 8.** Tabla de puentes de hidrógeno formados en los modelos con el asa superior reconstruida, con un cuarto de la proteína reconstruida y la interfase alfa hélice 7 y 8 reconstruida, respetando la simetría de orden 4. Las secuencias DATDRD, 35 y 2 fueron las elegidas en los análisis preliminares y se renombraron como ReTIM18 V2R, ReTIM29 y ReTIM30 respectivamente. Tipos de puentes de hidrógeno: 1 - K(Cε-NH3) - D(Cβ-COOH), 2 - R(Cδ-NH-C=(NH2+)(NH2)) - D(Cβ-COOH), 3 - R(Cδ-NH-C=(NH2+)(NH2)) - D(Cβ-COOH), 4 - K(Cε-NH3) - E(Cγ-COOH), 5 - S(Cβ-OH) - I(CO), 6 - S(NH) - I(CO), 7 - S(Cβ-OH) - Y(NH), 8 - R(Cδ-NH-C=(NH2+)(NH2)) - D(Cβ-COOH), 9 - R(Cδ-NH-C=(NH2+)(NH2)) - E(Cγ-COOH), 10 - Q(Cγ-CONH2) - D(CO), 11 - Q(Cγ-CONH2) - G(CO), 12 - S(Cβ-OH) - D(Cγ-COOH), 13 - R(Cδ-NH-C=(NH2+)(NH2)) - E(Cγ-COOH), 14 - R(Cδ-NH-C=(NH2+)(NH2)) - Y(Cβ-φOH), 15 - R(Cδ-NH-C=(NH2+)(NH2)) - Y(Cβ-φOH), 16 - N(Cβ-CONH2) - D(NH), 17 - R(Cδ-NH-C=(NH2+)(NH2)) - P(CO), 18 - R(Cδ-NH-C=(NH2+)(NH2)) - P(CO), 19 - N(Cβ-CONH2) - D(CO), 20 - E(Cγ-COOH) - K(Cε-NH3), 21 - D(Cβ-COOH) - R(Cδ-NH-C=(NH2+)(NH2)) y 22 - D(Cβ-COOH) - R(Cδ-NH-C=(NH2+)(NH2)). Los grupos funcionales marcados en negritas y/o entre paréntesis son los involucrados en los puentes de hidrógeno.

Una perspectiva interesante para este trabajo es emplear la metodología del cálculo alquímico de energía libre [53] como criterio de selección de modelos. Este método utiliza el paquete de software pmx que es una biblioteca de Python que permite la manipulación de las estructuras biomoleculares y sus archivos de topología. El software pmx utiliza Gromacs como motor de dinámica molecular para los cálculos de energía libre. Este cálculo define el estado inicial como el estado desplegado y el final como el estado plegado, donde la proteína silvestre se transforma en un mutante a través de una ruta no física (de ahí el nombre de cálculo alquímico). La diferencia de energía libre del plegamiento de proteínas en una mutación de aminoácidos ( $\Delta\Delta G_{\text{Mutation Folding}}$ ) se puede recuperar siguiendo dos caminos de plegamiento proteico, los caminos físicos de WT y proteína mutante ( $\Delta G_{\text{Mut Folding}} - \Delta G_{\text{WT Folding}}$ ), y los caminos alquímicos que transformar aminoácidos en los estados plegado y desplegado ( $\Delta G_{\text{Mutation Folded}} - \Delta G_{\text{Mutation Unfolded}}$ ). La Figura 40 ejemplifica esta metodología:



**Figura 40.** Representación esquemática de un ciclo termodinámico para calcular los cambios en la energía libre de plegamiento de proteínas tras la mutación ( $\Delta\Delta G_{\text{Mutation Folding}}$ ). La columna izquierda muestra el proceso de plegamiento de una proteína de tipo silvestre, con la energía libre  $\Delta G_{\text{WT Folding}}$

asociada al plegamiento; la columna de la derecha muestra la misma reacción de plegamiento, pero para una proteína mutada, resultando en la energía libre  $\Delta G_{\text{Mut Folding}}$  del plegamiento. El proceso representado en la fila inferior corresponde a la transformación química de la proteína desplegada de tipo silvestre en el mutante con la diferencia de energía libre asociada  $\Delta G_{\text{Mutation Unfolded}}$ . La reacción en la fila superior corresponde a la misma transformación química pero realizada en la proteína plegada, de modo que la diferencia de energía libre entre los dos mutantes es  $\Delta G_{\text{Mutation Folded}}$ . Las diferencias de energía libre para los procesos verticales son computacionalmente exigentes de calcular, pero las de las transformaciones horizontales son más accesibles. Por lo tanto,  $\Delta \Delta G_{\text{Mutation Folding}}$  se puede calcular a partir de la diferencia entre  $\Delta G_{\text{Mutation Folded}}$  y  $\Delta G_{\text{Mutation unfolded}}$  (vía química o no física).

Gracias al hecho de que computacionalmente tenemos control sobre la topología y la función de energía potencial que describe el sistema, podemos aprovechar al máximo las mejores propiedades de convergencia de la transformación no física a la física. Como la energía libre es una variable de estado, los cambios de energía libre obtenidos son independientes de la ruta. Por lo tanto, no es problemático elegir vías no físicas. Para simular el estado desplegado de la proteína, los tripéptidos independientes del contexto de secuencia (GXG, donde X es el residuo mutado) sirven como una buena aproximación del estado desplegado para estimar los cambios en la estabilidad de la proteína. Solo se necesita calcular  $\Delta G_{\text{Mutation Folded}}$ , mientras que  $\Delta G_{\text{Mutation Unfolded}}$  se puede encontrar en una tabla calculada previamente. Esta metodología ha probado tener éxito prediciendo valores experimentales de  $\Delta \Delta G$  con un error medio de 4 kJ / mol. Un ejemplo de ello, es la predicción de la estabilidad termodinámica ( $\Delta \Delta G$ ) de 22 variantes de la proteína  $\gamma$ D-Cristalina, la cual mostró una fuerte correlación ( $r = 0.97$ ) entre los valores de  $\Delta \Delta G$  experimental y los  $\Delta \Delta G$  calculados químicamente [54]. Como podemos ver, esta metodología provee un cálculo más exacto del  $\Delta \Delta G$  y puede ser una gran herramienta como criterio de selección de modelos con características termodinámicas favorables.

Merece la pena hablar de los barriles TIM *de novo* que se han diseñado a lo largo de la historia. Como podemos observar en la Tabla 9, los primeros diseños comparten varias cosas en común, la primera de ellas es que la mayoría carece de datos de desplegamiento térmico cooperativo. ¿Qué se requiere para diseñar barriles TIM *de novo* que muestren un desplegamiento cooperativo? Dicha incógnita es uno de los principales motores de este trabajo. El segundo aspecto que comparten en común los primeros diseños es que la mayoría de ellos son insolubles, lo que demuestra que obtener una proteína *de novo* soluble es un reto desafiante. Eso da cierto crédito al diseño ReTIM30, ya que, a pesar de ser parcialmente soluble, se pudo obtener una cantidad de proteína lo suficientemente grande para permitir caracterizarla fisicoquímicamente. Además, este diseño tiene 16 mutaciones totales lo cual significa que es un 8.7% diferente del diseño original. Otro de los grandes retos es obtener una proteína con estructura terciaria compacta, que, si bien no se corroboró en nuestros diseños con técnicas como ANS o  $^1\text{H}$  NMR, las propiedades hidrodinámicas muestran que, aunque ReTIM18 V2R y ReTIM30 se agregan, también muestran un pico que corresponde a una especie monomérica compacta. Algunos de los primeros diseños son resistentes al desplegamiento por agentes químicos, como es el caso de la proteína TIM-GO-TS-AMYL [3] que tiene un  $C_m$  en urea no visible aún a 7 M y un  $C_m$  en GndCl de 3.9 M, que supera incluso a ReTIM18. Lamentablemente no hay información del desplegamiento térmico de esta proteína. Un dato resaltable de esta proteína es que fue de las primeras en utilizar simetrías de órdenes menores (posee una simetría de orden 2), ya que los primeros diseños obtenidos, tenían simetrías de orden mayor. Al paso del tiempo la simetría de orden 4 se fue adoptando para favorecer los parámetros fisicoquímicos de las proteínas diseñadas. Quizá en diseños posteriores que busquen aumentar el  $\Delta G$  a temperatura ambiente, se deba adoptar la simetría de orden 2.

El primer dato resaltable de ReTIM30 es su  $C_m$  en GndCl muy similar a la de ReTIM18 (3.2 y 3.0 M respectivamente) y el segundo dato remarcable es su  $T_m$  mayor a ReTIM18 (105.2 y

92.9 °C respectivamente). Sin embargo, pese a estos parámetros mejorados, el  $\Delta H$  es significativamente menor al de ReTIM18 (1.9 y 76.6 kcal/mol respectivamente).

La Tabla 9 pone en contexto histórico este trabajo respecto a la historia de los barriles TIM *de novo*.

Nombre de la proteína	Año de publicación	Grado de simetría	Estado	¿Soluble?	¿Presencia de estructura secundaria?	¿Estructura terciaria compacta?	¿Glóbulo fundido?	Evidencias de estructura compacta / glóbulo fundido	Cm urea / Cm GndCl	¿Despliegue químico cooperativo?	Tm (°C)	¿Reversible?	¿Despliegue térmico cooperativo?	ΔH	ΔC <sub>p</sub>	ΔG	
Heptarelina [23]	1990	Orden 7	Monómero	No	Sí	No	No	Electroforesis en gradiente de urea	---	No	---	---	---	---	---	---	
Octarelina [23], [39]	1990, 1991	Orden 8	Monómero	No	Sí	Sí	No		---	Sí	---	No	---	---	---	---	---
Nonarelina [23]	1990	Orden 9	Monómero	No	Sí	No	No		---	No	---	---	---	---	---	---	---
TIM-GO-TS-AMYL [3]	1994	Orden 2	Monómero	Sí	Sí	No	Sí	ANS	Cm urea = No visible a 7M Cm GndCl = 3.9 M	Sí	---	---	---	---	---	---	
Sin nombre [40]	1994	Orden 2	Monómero	No	Sí	No	Sí	ANS / <sup>1</sup> H NMR	Cm GndCl = 2.0 M	---	---	---	Sí	2.0 kJ/mol	---	---	
L-βL-L-L [41]	1994	Orden 8	Monómero	No	Sí	No	Sí	ANS / <sup>1</sup> H NMR	Cm GndCl = 2.6 M	Sí	---	---	---	---	---	---	
Octarelina II [24]	1995	Orden 8	Monómero y dímero	No	Sí	No	Sí	ANS	Cm urea = 2-3 M	No	---	---	---	---	---	---	
Octarelina III [24]	1995	Orden 4	Monómero y dímero	No	Sí	No	Sí	ANS	Cm urea = 3.5 M	Sí	---	---	---	---	---	---	
Sin nombre [25]	2003	Orden 4	---	No	Sí	Sí	No	ANS / <sup>1</sup> H NMR	Cm GndCl = 3.9 M	Sí (reversible)	65	No	Sí	---	---	ΔG <sub>H2O</sub> = 35 kJ/mol	
Octarelina VI [31]	2013	No simétrico	Monómero	No	Sí	Sí	No	DLS / IF / CD	---	No	70	No	Sí	---	---	---	
ReTIM18 [27]	2017 (Tesis doctoral)	Orden 4	Monómero	Sí	Sí	Sí	No	Estructura cristalográfica	Cm Urea = 6.8 M Cm GndCl = 3.0 M	Sí, en GndCl	92.9	Parcialmente	Sí	76.6 kcal/mol	---	4.1 kcal/mol	
ReTIM18 V2R	---	Orden 4	Monómero	Sí	Sí	---	---	---	Cm Urea = 3.6 M Cm GndCl = 2.0 M	No	89.2	No	Sí	1.9 kcal/mol	0.06 kcal/mol	---	
ReTIM29	---	Orden 4	Monómero	No	Sí	---	---	---	Cm Urea = 1.7 M Cm GndCl = 1.6 M	No	---	No	No	---	---	---	
ReTIM30 soluble	---	Orden 4	Monómero	Parcialmente	Sí	---	---	---	Cm Urea = 4.3 M Cm GndCl = 3.2 M	Sí, en GndCl	105.2	No	Sí	0.26 kcal/mol	---	---	

Tabla 9. Tabla histórica de los barriles TIM *de novo*.



## Conclusiones

- La inspección visual debe acompañarse de otros análisis computacionales como la dinámica molecular para aumentar la confiabilidad en la selección de regiones de mejora estructural.
- Se pueden obtener modelos “computacionalmente exitosos” explorando la máxima variabilidad de residuos en regiones de mejora estructural potencial.
- Un modelo “computacionalmente exitoso” no necesariamente será un modelo con mejores características fisicoquímicas.
- Por sus características fisicoquímicas intrínsecas y a partir de los resultados encontrados, ReTIM18 no parece ser una buena candidata como modelo de mejora de estabilidad.
- Partir de un modelo con desplegamiento térmico irreversible, requiere la identificación de zonas con relevancia en la estabilidad, para poder sugerir mutaciones en dichas zonas y poderla hacer térmicamente reversible [43].
- Partir de un modelo con desplegamiento térmico y químico reversible, en principio, facilita la obtención de diseños con mejoría en su estabilidad.
- Las mutaciones propuestas en las dos zonas de mejora estructural (ReTIM18 V2R/asas superiores, ReTIM29/cierre del barril), no mejoraron la estabilidad.
- Las mutaciones propuestas en la región de la interfase entre las  $\alpha$  hélices 7 y 8 (ReTIM30) parecen estar ligadas al control de la estabilidad térmica y química, ya que fue la única proteína que mejoró su  $T_m$  y su estabilidad en GndCl.
- Hasta ahora, no es posible saber qué regiones de los barriles TIM *de novo* están relacionadas con el control de la estabilidad, ya que el cálculo refinado con Rosetta y Mopac propuso nuevas mutaciones que podrían repercutir favorablemente en la estabilidad.

## Perspectivas

- Solubilizar y caracterizar a ReTIM29, ya que ahora se cuenta con técnicas de expresión más sofisticadas.
- Caracterizar fisicoquímicamente el modelo con la secuencia 36 que se considera “computacionalmente exitoso” por el cálculo refinado de Rosetta y Mopac.
- Trabajar con un modelo que tenga inicialmente propiedades fisicoquímicas favorables, ya que ahora se cuenta con las estructuras experimentales de proteínas (resolución  $< 2\text{\AA}$ ) de 2 estados y con las cuales no se contaba antes (NovoTIM6, NovoTIM0, NovoTIM14) (Tabla14).
- Utilizar un cálculo refinado como criterio de selección de modelos.
- Utilizar técnicas de dinámica molecular para detectar zonas de alta tensión causadas por el desfase que crean las cuatro unidades repetitivas al unirse y no encajar perfectamente, ya que pueden ser zonas de mejora estructural en diseños futuros.
- Crear quimeras de unidades repetitivas de distintas proteínas ReTIM con alta estabilidad, ya que se ha demostrado [38] que la evolución utiliza esta estrategia para mejorar exitosamente las propiedades fisicoquímicas de las proteínas con topología de barril TIM.
- Favorecer el desplegamiento térmico de ReTIM30 soluble con concentraciones crecientes de urea y corroborar que el valor pequeño de entalpía de desplegamiento se debe a la presencia de estructura residual a temperaturas de desnaturalización.

- Utilizar la metodología del cálculo alquímico de energía libre [53] como criterio de selección de modelos.

## Bibliografía

- 1- Flores Herrera O, Riveros Rosas H, Sosa Peinado A, Vázquez Contreras E (eds). Mensaje Bioquímico, Vol XXVIII. Depto Bioquímica, Fac Medicina, Universidad Nacional Autónoma de México. Cd Universitaria, México, DF (2004), pp 70 – 71.
- 2- C. Martínez, J. F. García. (2014). Ingeniería de proteínas para el mejoramiento de enzimas. Revista Digital Universitaria | Publicación mensual Vol. 15 Num. 12 ISSN 1607 – 6079, pp. 6.
- 3- Tanaka Toshiki, Hayashi Mayumi, Kimura Hiromi, Oobatake Motohisa, Nakamura Haruki. (1994). *De novo* design and creation of a stable artificial protein. Biophysical Chemistry 50, 47 – 67.
- 4- Kiss Gert, Çelebi-Ölçüm Nihan, Moretti Rocco, Baker David, and Houk K. N. (2013). Computational Enzyme Design. Angew. Chem. 52, 2 – 28.
- 5- Dalby, P. A. (2007). Engineering Enzymes for Biocatalysis. Recent Patents on Biotechnology. Vol. 1, No. 1, pp. 1-9.
- 6- Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, et al. (2003). Design of a novel globular protein fold with atomic-level accuracy. Science 302: 1364– 1368.
- 7- Ambroggio XI, Kuhlman B. (2006). Computational design of a single amino acid sequence that can switch between two distinct protein folds. J Am Chem Soc 128: 1154–1161.
- 8- Kuhlman B, Baker D. (2000). Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci U S A 97: 10383–10388.
- 9- Dunbrack RL, Jr. (2002). Rotamer libraries in the 21st century. Curr Opin Struct Biol 12: 431–440.
- 10- Dunbrack RL, Jr., Cohen FE. (1997). Bayesian statistical analysis of protein sidechain rotamer preferences. Protein Sci 6: 1661–1681.
- 11- Dunbrack RL, Jr., Karplus M. (1993). Backbone-dependent rotamer library for proteins. Application to side-chain prediction. J Mol Biol 230: 543–574.
- 12- Das R, Baker D. (2008). Macromolecular modeling with rosetta. Annu Rev Biochem 77: 363–382.
- 13- Kaufmann Kristian W. Lemmon Gordon H. DeLuca Samuel L. Sheehan Jonathan H. and Meiler Jens. (2010). Practically Useful: What the ROSETTA Protein Modeling Suite Can Do for You. Biochemistry 2010, 49, 2987–2998.
- 14- Simons, K. T., Kooperberg, C., Huang, E., and Baker, D. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J. Mol. Biol. 268, 209–225.
- 15- Bystroff, C., Simons, K. T., Han, K. F., and Baker, D. (1996) Local sequence-structure correlations in proteins. Curr. Opin. Biotechnol. 7, 417–421.

- 16- Rohl, C. A., Strauss, C. E., Misura, K. M., and Baker, D. (2004) Protein structure prediction using Rosetta. *Methods Enzymol.* 383, 66–93.
- 17- Kuhlman B. and Baker D. (2000) Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10383–10388.
- 18- Leaver-Fay, A., Kuhlman, B., and Snoeyink, J. (2005) Rotamer-Pair Energy Calculations Using a Trie Data Structure. *Lect. Notes Comput. Sci.* 3692, 389–400.
- 19- Huang Po-Ssu, E. Boyken<sup>1</sup> Scott & Baker David (2016). The coming of age of de novo protein design. *537(7620):* 320-7.
- 20 - Duran AM & Meiler. (2018) Computational design of membrane proteins using RosettaMembrane. *Protein Science*, 27(1), 341–355.
- 21 – Rakesh Gupta & Beena Rai (2018). *In-silico* design of nanoparticles for transdermal drug delivery application.
- 22- D. W. Banner, G. A. Petsko, D. C. Phillips, C. I. Pogson & I. A. Wilson. (1975). Structure of triose phosphate isomerase determined crystallographically at 2.5 Å resolution. *Nature* vol. 255, pp. 609
- 23- Goraj Karine, Renard André and A. Martial Joseph. (1990). Synthesis, purification and initial structural characterization of octarellin, a de novo polypeptide modelled on the  $\alpha/\beta$ -barrel proteins - vol.3 no.4 pp.259-266.
- 24- Houbrechts Annick, Moreau Benoit, Abagyan Ruben, Mainfroid Véronique, Préaux Gisele, Lamproye Alain, Poncin Alain, Goormaghtigh Erik, Ruyschaert Jean-Marie, A. Martial Joseph and Goraj Karine. (1995). Second-generation octarellins: two new de novo ( $\beta/\alpha$ )<sub>8</sub> polypeptides designed for investigating the influence of  $\beta$ -residue packing on the  $\alpha/\beta$ -barrel structure stability - *Protein Engineering* vol.8 no.3 pp.249-259.
- 25- F. Offredi, F. Dubail, P. Kischel, K. Sarinski, A. S. Stern C. Van de Weerd, J. C. Hoch, C. Prospero, J. M. François S. L. Mayo and J. A. Martial. (2003). De novo Backbone and Sequence Design of an Idealized  $\alpha/\beta$ -barrel Protein: Evidence of Stable Tertiary Structure - *J. Mol. Biol.* 325, 163–174.
- 26- Huang Po-Ssu, Feldmeier Kaspar, Parmeggiani Fabio, Fernandez Velasco D. Alejandro, Höcker Birte & Baker David. (2015). De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chemical Biology*. Vol. 12, pp. 29 – 33.
- 27- Romero (2017). El paisaje conformacional de las proteínas: estudio de barriles TIM naturales y diseñados de novo, tesis doctoral posgrado en Ciencias Bioquímicas.
- 28- Andrew D. Robertson and Kenneth P. Murphy. (1997) Protein structure and the energetics of protein stability. *Chemical Reviews*, 97(5), pp. 1251 – 1267.
- 29- Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; (In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press (2005). Protein Identification and Analysis Tools on the ExPASy Server. pp. 571-607.

- 30- Accessible Surface Area and Accessibility Calculation for Protein. (2012) Center for Informational Biology, Ochanomizu University.
- 31- Figueroa Maximiliano, Oliveira Nicolas, Lejeune Annabelle, Kaufmann Kristian W, Dorr Brent M, Matagne André, Martial Joseph A., Meiler Jens, Van de Weerd Cécile. (2013) Octarellin VI: Using Rosetta to Design a Putative Artificial  $(\beta/\alpha)_8$  Protein. PLOS ONE, Vol. 8, pp. 1 – 15
- 32- Vivian, J. T., & Callis, P. R. (2001). Mechanisms of Tryptophan Fluorescence Shifts in Proteins. *Biophysical Journal*, 80(5), 2093–2109.
- 33- Royer Catherine A. (2006). Probing Protein Folding and Conformational Transitions with Fluorescence. *Chemical Reviews*, Vol. 106, No. 5. 1769 -1784.
- 34- Sanchez-Ruiz, J. M. (1992). Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry. *Biophysical Journal*, 61(4), 921–935.
- 35- Shen Hao; Fallas Jorge A; Lynch Eric; Sheffler William; Parry Bradley; Jannetty Nicholas; Decarreau Justin; Wagenbach Michael; Vicente Juan Jesus; Chen Jiajun; Wang Lei; Dowling Quinton; Oberdorfer Gustav; Stewart Lance; Wordeman Linda; De Yoreo James; Jacobs-Wagner Christine; Kollman Justin; Baker David (2018). De novo design of self-assembling helical protein filaments. *Science*, 362 (6415), pp. 705–709.
- 36- Marcos Enrique; Chidyausiku Tamuka M; McShan Andrew C; Evangelidis Thomas; Nerli Santrupti; Carter Lauren; Nivón Lucas G; Davis Audrey; Oberdorfer Gustav; Tripsianes Konstantinos; Sgourakis Nikolaos G; Baker David (2018). De novo design of a non-local  $\beta$ -sheet protein with high stability and accuracy. *Nature Structural & Molecular Biology*. ISSN: 1545-9985.
- 37- Silva Daniel-Adriano; Yu Shawn; Ulge Umut Y; Spangler Jamie B; Jude Kevin M; Labão-Almeida Carlos; Ali Lestat R; Quijano-Rubio Alfredo; Ruterbusch Mikel; Leung Isabel; Biary Tamara; Crowley Stephanie J; Marcos Enrique; Walkey Carl D; Weitzner Brian D; Pardo-Avila Fátima; Castellanos Javier; Carter Lauren; Stewart Lance; Riddell Stanley R; Pepper Marion; Bernardes Gonçalo J L; Dougan Michael; Garcia Christopher K; Baker David (2019). De novo design of potent and selective mimics of IL-2 and IL-15. *Nature*, Vol. 565, pp. 186 – 191.
- 38 - Höcker Birte, Claren Jörg & Sterner Reinhard (2004). Mimicking enzyme evolution by generating new  $(\beta\alpha)_8$ -barrels from  $(\beta\alpha)_4$ -half-barrels. *PNAS* Vol. 101, pp. 16448 – 16453.
- 39 - Marc Beauregard, Karine Goraj, Vincent Goffin, Karel Heremans, Eric Goormaghtigh, Jean-Marie Ruyschaert and Joseph A. Martial (1991). Spectroscopic investigation of structure in octarellin (a de novo protein designed to adopt the  $\alpha/3$ -barrel packing). *Protein Engineering* vol.4 no 7 pp 745-749.
- 40 - Toshiki Tanaka, Yutaka Kuroda, Hiromi Kimura, Shun-ichi Kidokoro & Haruki Nakamura (1994). Cooperative deformation of a de novo designed protein. *Protein Engineering* vol.7 no.8 pp.969-976.

- 41 - Toshiki Tanaka, Hiromi Kimura, Mayumi Hayashi, Yosinori Fujiyoshi, Ken-ichi Fukuhara, & Haruki Nakamura (1994). Characteristics of a de novo designed protein. *Protein Science*, 3:419-427.
- 42 - Romero-Romero Sergio, Fernández-Velasco D. Alejandro y Costas Miguel (2018). Estabilidad termodinámica de proteínas. *Educación Química* vol. 29, no. 3, pp. 3-17.
- 43 - Lamazares Emilio, Clemente Isabel, Bueno Marta, Velázquez-Campoy Adrián y Sancho Javier (2015) Rational stabilization of complex proteins: a divide and combine approach. *Scientific Reports*. 5(1).
- 44 - Bommarius, A. S., Broering, J. M., Chaparro-Riggers, J. F., & Polizzi, K. M. (2006). High-throughput screening for enhanced protein stability. *Current Opinion in Biotechnology*, 17(6), 606–610.
- 45 - Jaenicke, R. (2000). Stability and stabilization of globular proteins in solution. *Journal of Biotechnology*, 79(3), 193–203.
- 46 - Menéndez-Arias, L., & Argosf, P. (1989). Engineering protein thermal stability. *Journal of Molecular Biology*, 206(2), 397–406.
- 47- Privalov, P. L., & Khechinashvili, N. N. (1974). A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study. *Journal of Molecular Biology*, 86(3), 665–684.
- 48 – Böhm, G., Muhr, R., & Jaenicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. “*Protein Engineering, Design and Selection*”, 5(3), 191–195.
- 49 – Perez-Iratxeta, C., Andrade-Navarro, M. A. (2008). K2D2: Estimation of protein secondary structure from circular dichroism spectra. *BMC Structural Biology*, 8(1), 25.
- 50 - Hernández-Alcántara, G., Rodríguez-Romero, A., Reyes-Vivas, H., Peon, J., Cabrera, N., Ortiz, C., López-Velázquez, G. (2008). Unraveling the mechanisms of tryptophan fluorescence quenching in the triosephosphate isomerase from *Giardia lamblia*. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1784(11), 1493–1500.
- 51- Romero-Romero S., Costas M., Silva Manzano D., Kordes S., Rojas-Ortega E., Tapia C., Guerra Y., Shanmugaratnam S., Rodríguez-Romero A., Baker D., Höcker B., Fernández-Velasco D. (2019). Exploring the stability landscape of de novo TIM barrels by a modular design approach. En preparación.
- 52 - Zambrano, R., Jamroz, M., Szczasiuk, A., Pujols, J., Kmiecik, S., & Ventura, S. (2015). AGGREGSCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Research*, 43(W1), W306–W313.
- 53 - Sikosek, T. (Ed.). (2019). *Computational Methods in Protein Evolution*. *Methods in Molecular Biology*. doi:10.1007/978-1-4939-8736-8.



- 54 - Aguayo-Ortiz, R., González-Navejas, A., Palomino-Vizcaino, G., Rodríguez-Meza, O., Costas, M., Quintanar, L., & Dominguez, L. (2019). Thermodynamic Stability of Human  $\gamma$ D-crystallin Mutants Using Alchemical Free Energy Calculations. *The Journal of Physical Chemistry B*.
- 55 - Robic S., Guzman M., Sanchez R., and Marqusee S. Role of residual structure in the unfolded state of a thermophilic protein (2003). *Proc. Natl. Acad. Sci. USA*, 100: 11345-11349.
- 56 - Gyftopoulos, E. P. y Beretta, G. P. *Thermodynamics: Foundations and Applications* (2015). Mineola, New York: Dover Publications.
- 57 – C. Rees y D. Robertson A. Some thermodynamic implications for the thermostability of proteins (2001). *Protein Science*, 10(6), 1187–1194.
- 58 - Daopin S, Alber T, Baase WA, Wozniak JA, Matthews BW. Structural and thermodynamic analysis of the packing of two alpha-helices in bacteriophage T4 lysozyme. *J (1991). Mol Biol.* 20; 221(2): 647 - 67.
- 59 - Rashin AA, Rashin BH, Rashin A, Abagyan R. Evaluating the energetics of empty cavities and internal mutations in proteins (1997). *Protein Sci.* 6(10): 2143 - 58.
- 60 - Yamada H, Kanaya E, Ueno Y, Ikehara M, Nakamura H, Kikuchi M. Contribution of a hydrogen bond to the thermal stability of the mutant human lysozyme C77/95S (1994). *Biol Pharm Bull.* 17(5): 612 - 6.
- 61 - Chen YW, Fersht AR, Henrick K. Contribution of buried hydrogen bonds to protein stability. The crystal structures of two barnase mutants (1993). *J Mol Biol.* 234(4): 1158 - 70.
- 62 - Myers JK, Pace CN. Hydrogen bonding stabilizes globular proteins (1996). *Biophys J.* 71(4): 2033 - 9.
- 63 - Scholtz JM, Qian H, Robbins VH, Baldwin RL. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide (1993). *Biochemistry.* 32(37): 9668 - 76.
- 64 - Meeker AK, Garcia-Moreno B, Shortle D. Contributions of the ionizable amino acids to the stability of staphylococcal nuclease (1996). *Biochemistry.* 35(20): 6443 - 9.
- 65 - Zhou NE, Kay CM, Hodges RS. Disulfide bond contribution to protein stability: positional effects of substitution in the hydrophobic core of the two-stranded alpha-helical coiled-coil (1993). *Biochemistry.* 32(12): 3178 - 87.
- 66 - Wetzel R, Perry LJ, Baase WA, Becktel WJ. Disulfide bonds and thermal stability in T4 lysozyme (1988). *Proc Natl Acad Sci U S A.* 85(2): 401 – 5.

## MATERIAL SUPLEMENTARIO

### Ecuaciones de Rosetta

Las ecuaciones de los términos que toma en cuenta ROSETTA para calcular la energía de alta resolución de los átomos en las proteínas, son las siguientes:

#### Preferencias de torsión de Ramachandran:

$$\sum_i -\ln[P(\phi_i, \psi_i | aa_i, ss_i)]$$

Donde:

$i$  = índice de residuo

$\phi, \psi$  = ángulos de torsión del esqueleto ( $36^\circ$ )

$aa$  = tipo de aminoácido

$ss$  = tipo de estructura secundaria

$i, j$  = índices de residuo [16]

#### Interacciones de Lennard-Jones:

$$\sum_i \sum_{j>i} \left[ \left( \frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}}{d_{ij}} \right)^6 \right] e_{ij}$$

$$\text{si } \frac{d_{ij}}{r_{ij}} > 0.6$$

De lo contrario:

$$\sum_i \sum_{j>i} \left[ -8759.2 \left( \frac{d_{ij}}{r_{ij}} \right) + 5672.0 \right] e_{ij}$$

Donde:

$d$  = distancia interatómica

$e$  = media geométrica de las profundidades del pozo atómico

$r$  = sumatoria de radios de van der Waals [16]

#### Puentes de hidrógeno:

$$\sum_i \sum_j (-\ln[P(d_{ij} | h_j ss_{ij})] - \ln[P(\cos \theta_{ij} | d_{ij} h_j ss_{ij})] - \ln[P(\cos \psi_{ij} | d_{ij} h_j ss_{ij})])$$

Donde:

$i$  = índice del residuo donante

$j$  = índice del residuo aceptor

$d$  = distancia interatómica aceptor-protón

$h$  = hibridación ( $sp^2$ ,  $sp^3$ )

ss = tipo de estructura secundaria  
 $\theta$  = ángulo de unión de la base protón-aceptor-aceptor  
 $\psi$  = ángulo de unión protón-aceptor-aceptor  
 $i, j$  = índices atómicos [16]

**Solvatación:**

$$\sum_i [\Delta G_i^{\text{ref}} - \sum_i (\frac{2\Delta G_i^{\text{free}}}{4\pi^{3/2}\lambda_j r_{ij}^2} e^{-d_{ij}^2} V_i)]$$

Donde:

$d$  = distancia entre átomos  
 $r$  = sumatoria de radios de van der Waals  
 $\lambda$  = longitud de correlación  
 $V$  = volumen atómico  
 $\Delta G^{\text{ref}}, \Delta G^{\text{free}}$  = energía de un átomo completamente solvatado [16]

**Interacciones de pares de residuos (electrostáticas, disulfuros):**

$$\sum_i -\ln \left[ \frac{P(aa_i, aa_j, d_{ij})}{P(aa_i|d_{ij})P(aa_j|d_{ij})} \right]$$

Donde:

$i, j$  = índices del residuo  
 $aa$  = tipo de aminoácido  
 $d$  = distancia residuos [16]

**Energía propia del rotámero:**

$$\sum_i -\ln \left[ \frac{P(\text{rot}_i|\phi_i, \psi_i)P(aa_i|\phi_i, \psi_i)}{P(aa_i)} \right]$$

Donde:

$i, j$  = índices del residuo  
 $\text{rot}$  = rotámero Dunbrack dependiente del esqueleto  
 $aa$  = tipo de aminoácido  
 $\phi, \psi$  = ángulos de torsión del esqueleto [16]

**Energía de referencia del estado desplegado:**

$$\sum_{aa} n_{aa}$$

Donde:

$aa$  = tipo de aminoácido  
 $n$  = número de residuos [16]

## Comandos del protocolo de diseño

### 1- Limpieza de la estructura:

```
/home/usuario/rosetta_scr_2017.08.59219_bundle/tools/protein_tools/scripts/clean_pdb.py  
ReTIM18 A
```

### 2- Relajación (Minimización de la estructura)

```
/home/usuario/rosetta_scr_2017.08.59219_bundle/main/source/bin/relaz.linuxgccrelease -s  
ReTIM18_A.pdb -database ~/rosetta_scr_2017.08.59219_bundle/main/database -nstruct 3
```

### 3- Diseño

```
/home/usuario/rosetta_scr_2017.08.59219_bundle/main/source/bin/fixbb.linuxgccrelease -s  
ReTIM18_A.pdb -resfile resfile.txt -database  
~/rosetta_scr_2017.08.59219_bundle/main/database -nstruct 100
```

Resfile para cambiar un aminoácido puntual y permitir el rearrreglo de toda la cadena:

```
NATAA  
EX 1  
USE_INPUT_SC
```

```
start  
 1 A PIKAA E EX 1 EX 2
```

NATAA #permite sólo aminoácidos naturales

EX 1 #extra rotámero – muestra discretamente rotámeros de la librería Dunbrak. El número indica la inclusión de muestras discretas adicionales.

USE\_INPUT\_SC #Permite el uso de la conformación de la cadena lateral nativa (incluye el rotámero nativo)

```
start # inicio  
 1 A PIKAA E EX 1 EX 2
```

# 1 número de residuo

# A cadena

# PIKAA E indica que se quiere cambiar el residuo 1 por E

#EX 1 extra rotámeros

#EX 2 extra rotámeros con dos muestras discretas adicionales

Resfile para explorar las posibilidades de encontrar un residuo de mínima energía y permitir el rearrreglo de toda la cadena:

```
NATAA  
EX 1  
USE_INPUT_SC
```

```
start  
 1 A ALLAAxc EX 1 EX 2
```

#ALLAxc permite cambiar un aminoácido al azar de los 19 canónicos exceptuando la cisteína.

4- Relajación (Minimización de la estructura con las mutaciones hechas en el diseño)

/home/usuario/rosetta\_scr\_2017.08.59219\_bundle/main/source/bin/relax.linuxgccrelease -s ReTIM18\_A.pdb -database ~/rosetta\_scr\_2017.08.59219\_bundle/main/database -nstruct 3

Tablas de resultados del cálculo no robusto con ROSETTA y el cálculo robusto con ROSETTA y MOPAC

Secuencia	ROSETTA cálculo no robusto (Score)	ROSETTA (cálculo robusto) $\Delta H_f(X \rightarrow Y_{loop})$ (Score)	MOPAC (cálculo robusto) $\Delta H_f(X \rightarrow Y_{loop})$ Calor de formación (kJ/mol)
DASNP	-431.02	8.06	-507.51
DHTTV	-428.66	1.32	134.12
DATDK	-440.87	-7.07	-105.79
DATDK	-440.97	0.15	9.07
DATDR	-439.21	-5.54	-71.42
DATDR	-444.86	-3.87	-183.26
DATDE	-436.86	3.34	58.51
DATDE	-435.76	0.01	-80.07
DDTNT	-437.47	9.27	3107.02
DDTNV	-434.93	8.70	-440.12
DATNT	-436.7	0.43	38.04
DATNV	-436.19	-2.17	119.27
DATNP	-431.84	7.09	-651.15
DATDP	-430.5	10.19	-775.70

Tabla 10. Tabla de resultados obtenidos mediante el cálculo no robusto con ROSETTA y el cálculo refinado de ROSETTA y MOPAC para los diseños con el asa superior reconstruida, respetando la simetría de orden 4.

Secuencia	ROSETTA cálculo no robusto (Score)	ROSETTA (cálculo robusto) $\Delta H_f(X \rightarrow Y_{cc})$ (Score)	MOPAC (cálculo robusto) $\Delta H_f(X \rightarrow Y_{cc})$ Calor de formación (kJ/mol)
5	-444.8	-33.06	-580.38
6	-451.94	-31.52	-932.38
15	-446.75	-34.11	-722.79
35	-437.35	-11.93	-908.89
36	-439.52	-38.40	-1113.72
43	-448.8	-35.55	-594.59
56	-446.27	-35.39	-689.91
62	-444.68	-23.84	-789.07

Tabla 11. Tabla de resultados obtenidos mediante el cálculo no robusto con ROSETTA y el cálculo refinado de ROSETTA y MOPAC para los diseños con un cuarto de la proteína reconstruida, respetando la simetría de orden 4.

Secuencia	ROSETTA cálculo no robusto (Score)	ROSETTA (cálculo robusto) $\Delta H_f(X \rightarrow Y_{int.})$ (Score)	MOPAC (cálculo robusto) $\Delta H_f(X \rightarrow Y_{int.})$ Calor de formación (kJ/mol)
1	-436.58	-1.10	202.73
2	-439.66	-4.34	205.62



3	-435.19	-0.82	-177.14
4	-436.72	1.05	163.44

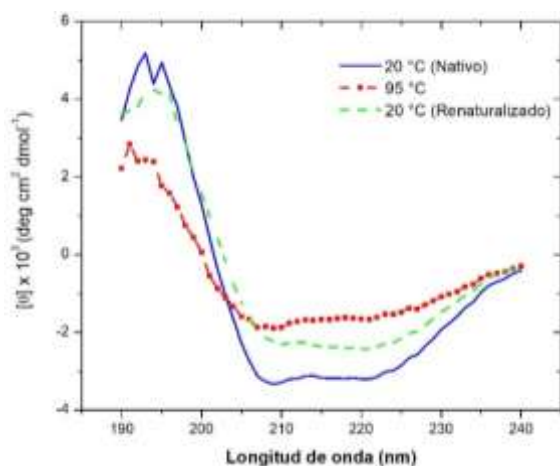
**Tabla 12.** Tabla de resultados obtenidos mediante el cálculo no robusto con ROSETTA y el cálculo refinado de ROSETTA y MOPAC para los diseños con la interface alfa hélice 7 – alfa hélice 8 reconstruida, respetando la simetría de orden 4.

<b>Parámetro</b>	<b>Valor</b>
Space group	P3 <sub>2</sub> (145)
Unit cell parameters	a= 51.72; b= 51.72; c= 63.97 $\alpha$ = 90.0°; $\beta$ = 90.0°; $\gamma$ = 120.0°
Resolution limit (Å)	44.79-1.64 (1.74-1.64)
Total reflections	79702 (3667)
Unique reflections	23430 (1128)
Multiplicity	3.4 (3.3)
Completeness (%)	99.5 (95.3)
Mean I / sig (I)	14.3 (1.9)
R-merge	0.039 (0.517)
CC1/2	0.998 (0.639)
Average mosaicity	0.116
R-work	0.186 (0.280)
R-free	0.214 (0.319)
Number of non-hydrogens atoms	1621
Number of macromolecules atoms	1464
Number of ligands atoms	7
Number of solvent atoms	150
Protein residues	181
RMS (bonds)	0.007
RMS (angles)	1.000
Ramachandran favored (%)	99.44
Ramachandran allowed (%)	0.56
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashcore	4.28
Overall B factor from Wilson Plot (Å <sup>2</sup> )	28.4
Average B-factor (Å <sup>2</sup> )	33.4
Macromolecules B-factor	32.2
Ligands B-factor	45.0
Solvent B-factor	44.9

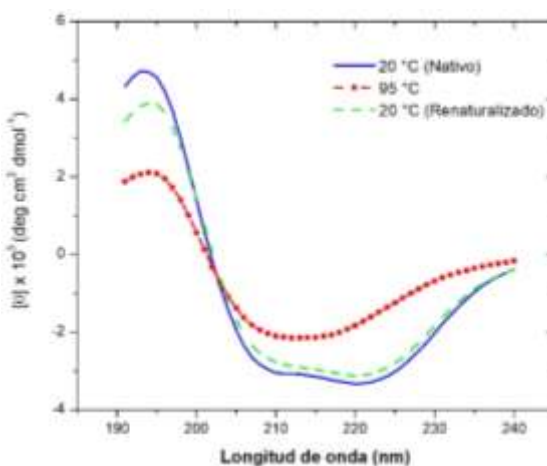
**Tabla 13.** Datos de colecta y refinamiento del modelo ReTIM18. La determinación de la estructura cristalográfica de la proteína ReTIM18, los datos de colecta y de refinamiento pertenecen a la tesis doctoral de Romero (2017) [27]

## ReTIM18 V2R

A

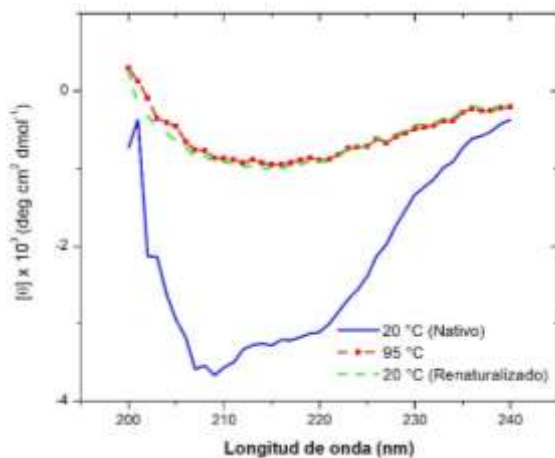


B

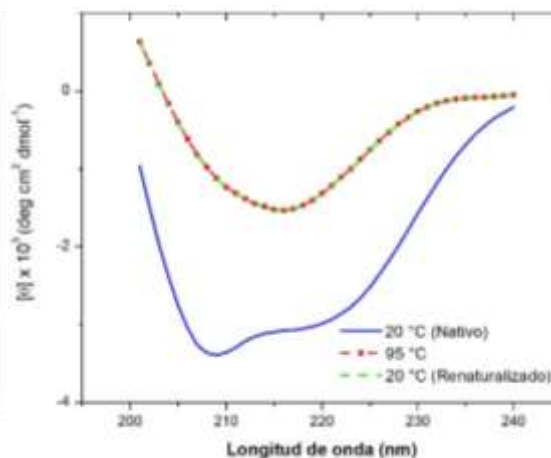


## ReTIM29 (resuspensión en urea y renaturalización mediante la condición 1)

C

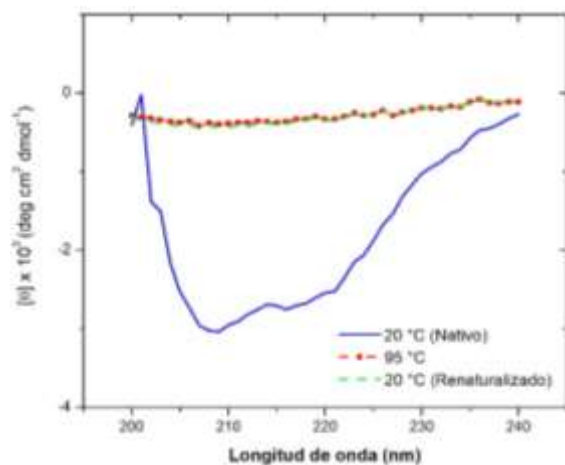


D

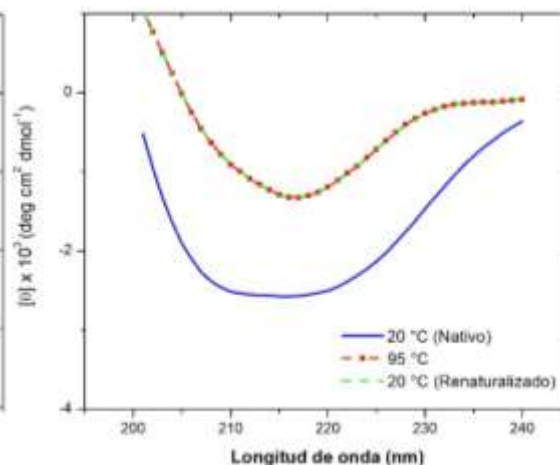


## ReTIM29 (resuspensión en urea y renaturalización mediante la condición 5)

E

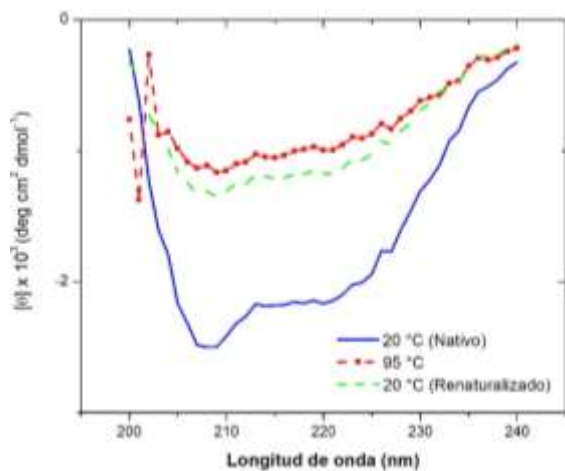


F

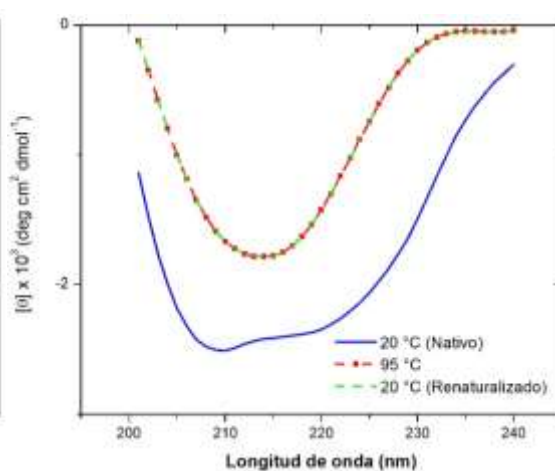


### ReTIM30 (resuspensión en urea y renaturalización mediante la condición 1)

G

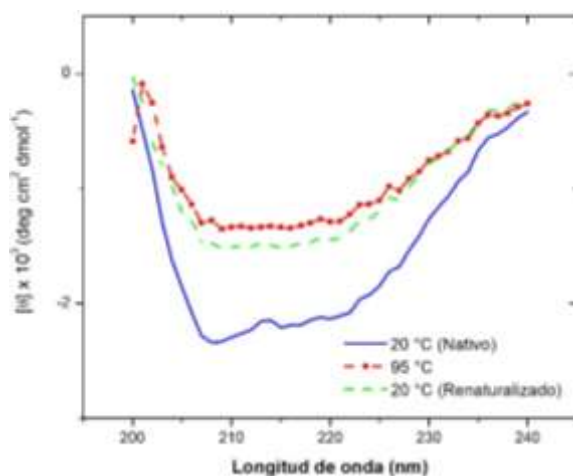


H

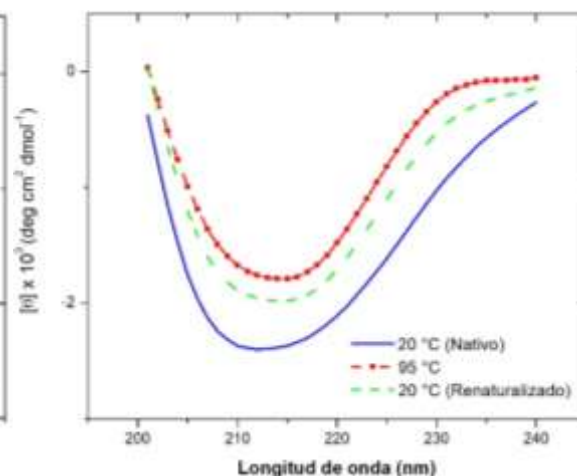


### ReTIM30 (resuspensión en urea y renaturalización mediante la condición 5)

I

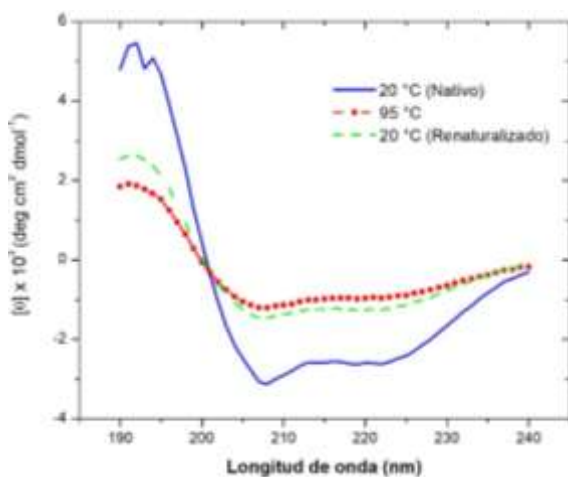


J



### ReTIM30 soluble

K



L

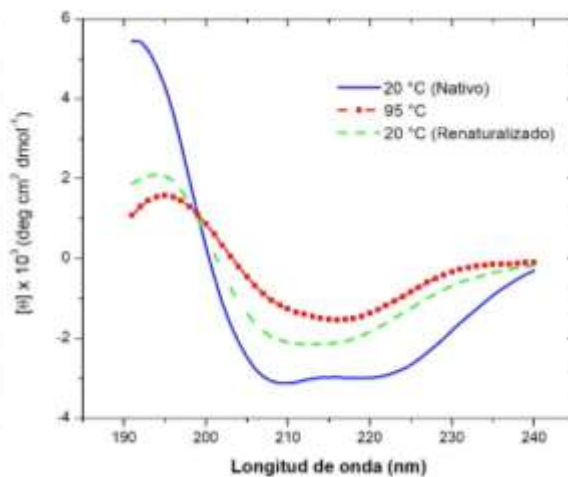
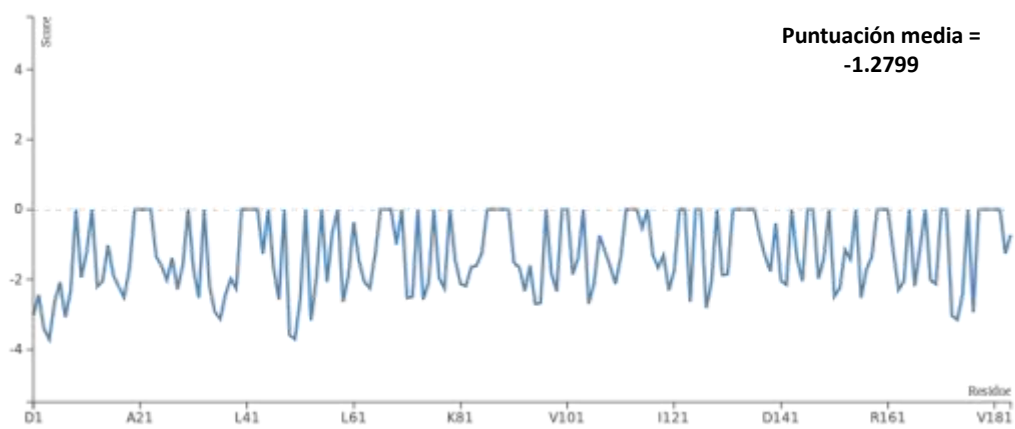


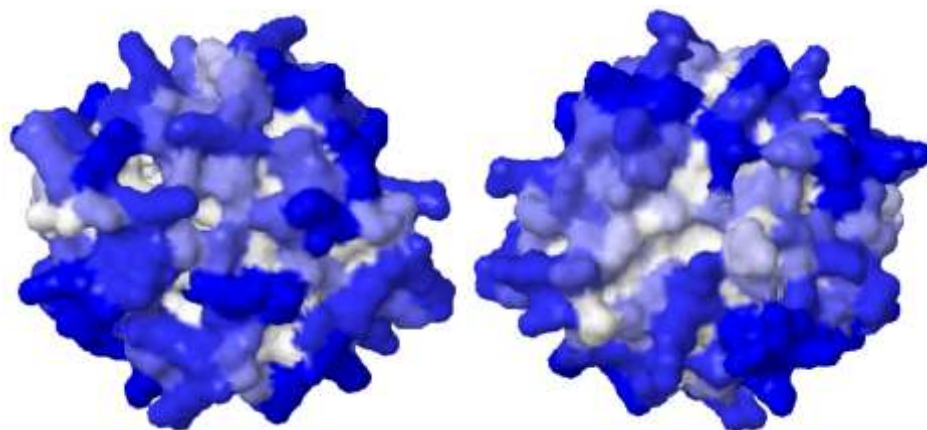
Figura 41. Espectros de DC experimentales y predichos mediante el servidor en línea K2D2. Los incisos correspondientes a los espectros experimentales son: A), C), E), G), I) y K). Los incisos de los espectros predichos son: B), D), F), H) y L).

### ReTIM18

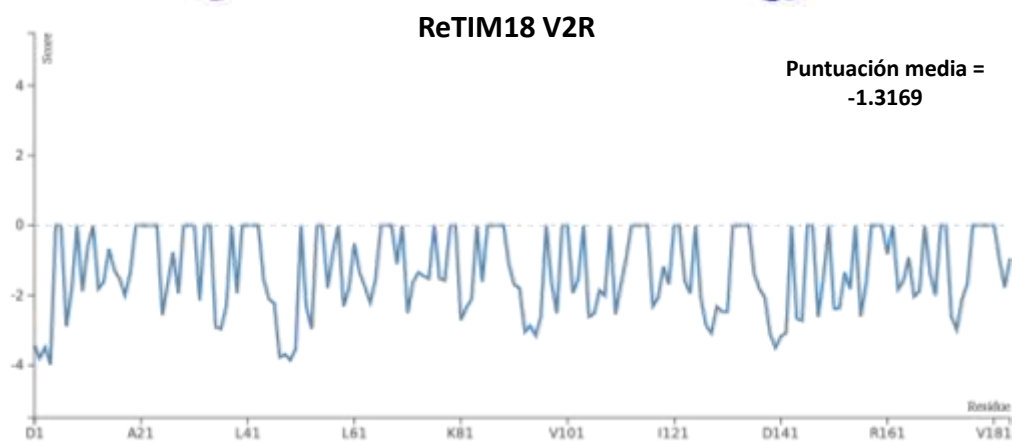
A)



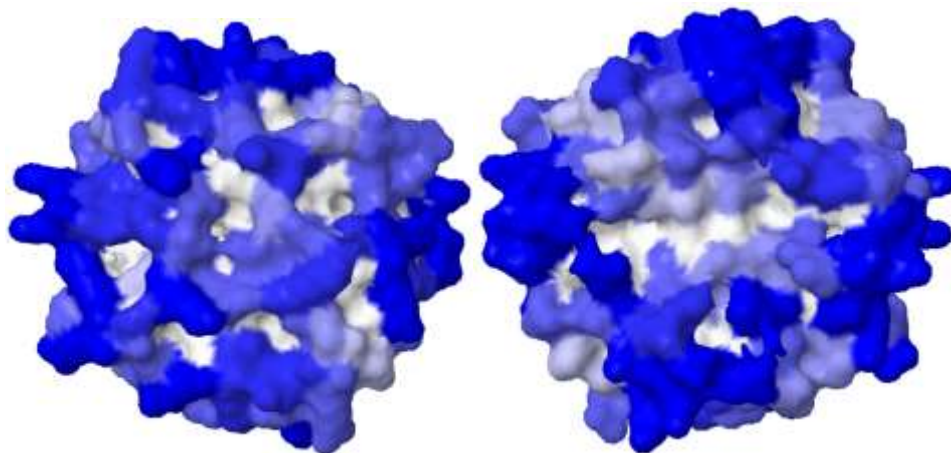
B)



C)



D)



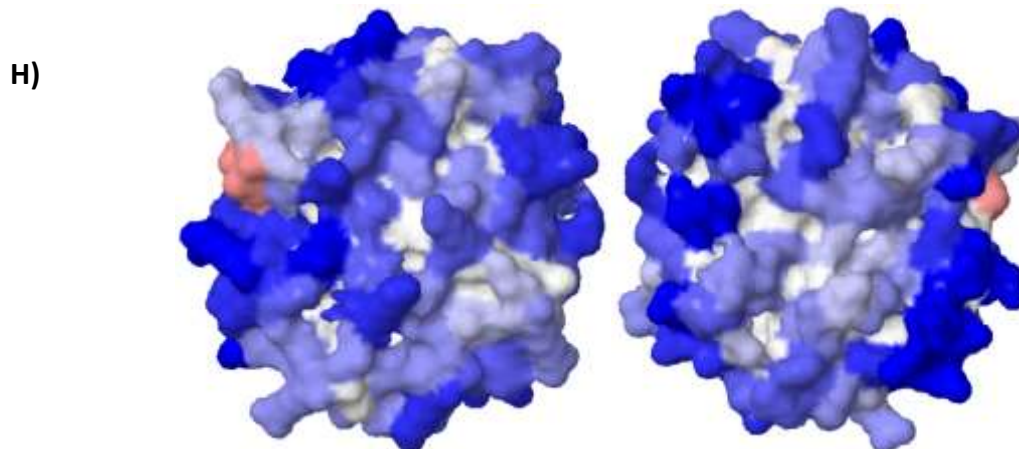
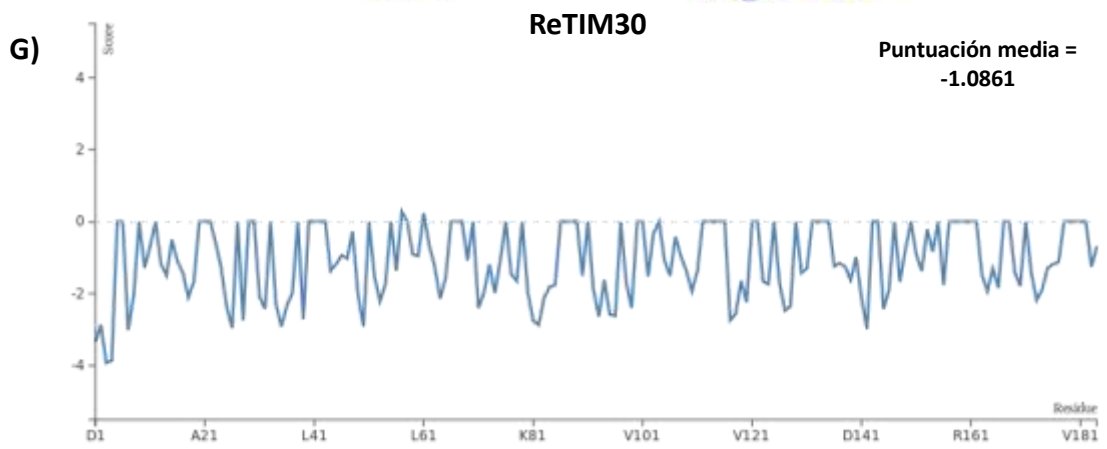
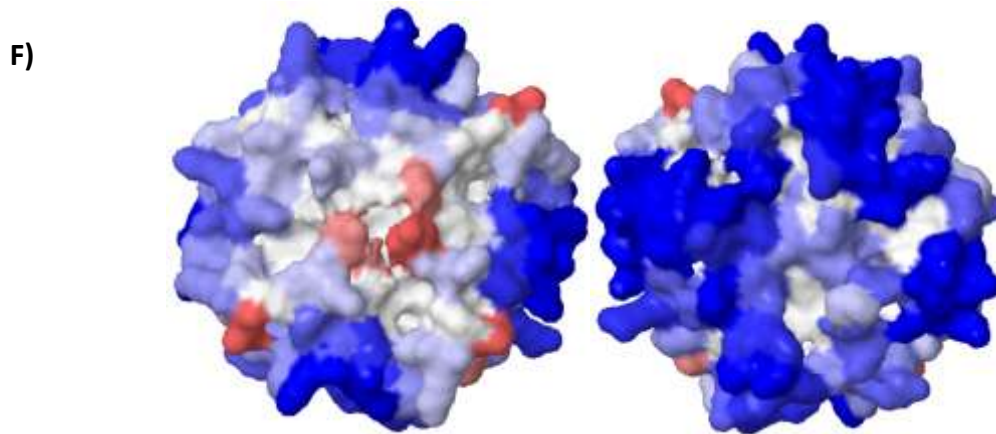
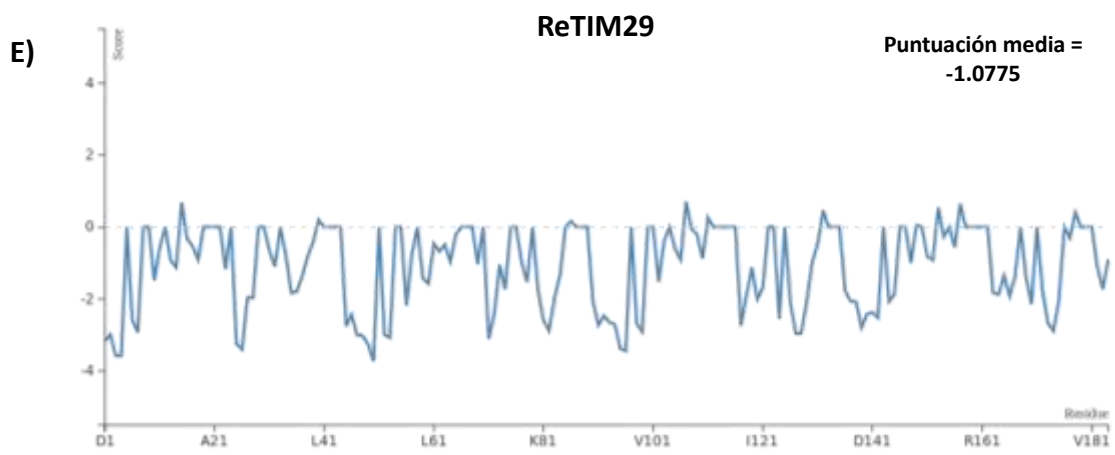




Figura 42. Resultados obtenidos mediante en servidor en línea AGGRESCAN 3D [52]. Los cálculos fueron hechos con el modo dinámico, lo que permite tener en cuenta las fluctuaciones dinámicas de la estructura proteica en solución, con una distancia de agregación de 10 Å. Los incisos A), C), E) y G) son gráficos de puntajes de solubilidad contra número de residuo; entre más negativo el puntaje, mayor es la solubilidad del residuo, por el contrario, residuos con puntajes positivos son propensos a la agregación. Si el residuo tiene un puntaje de cero, ese residuo se encuentra enterrado y no expuesto al solvente, por lo que no influye en la solubilidad / agregación. Los puntajes medios que se encuentran en la parte superior derecha de cada gráfico, son indicadores normalizados de la propensión a la agregación/solubilidad de la estructura proteica y sirven para comparar la solubilidad de diferentes estructuras proteicas. Los incisos B), D), F) y H) son imágenes de las proteínas diseñadas. La imagen derecha corresponde al barril TIM visto desde los extremos carboxilos terminales de las hebras  $\beta$  y la imagen izquierda corresponde al barril TIM visto desde los extremos amino terminales de las hebras  $\beta$ . Las zonas coloreadas en azul, corresponden a las regiones de la proteína que son solubles, entre más oscura la tonalidad, más soluble es. Las zonas rojas, corresponden a las regiones de la proteína propensas a la agregación, entre más oscura la tonalidad, es más propensa la agregación. Las zonas blancas son regiones poco expuestas al solvente y no influyen en la solubilidad/agregación.

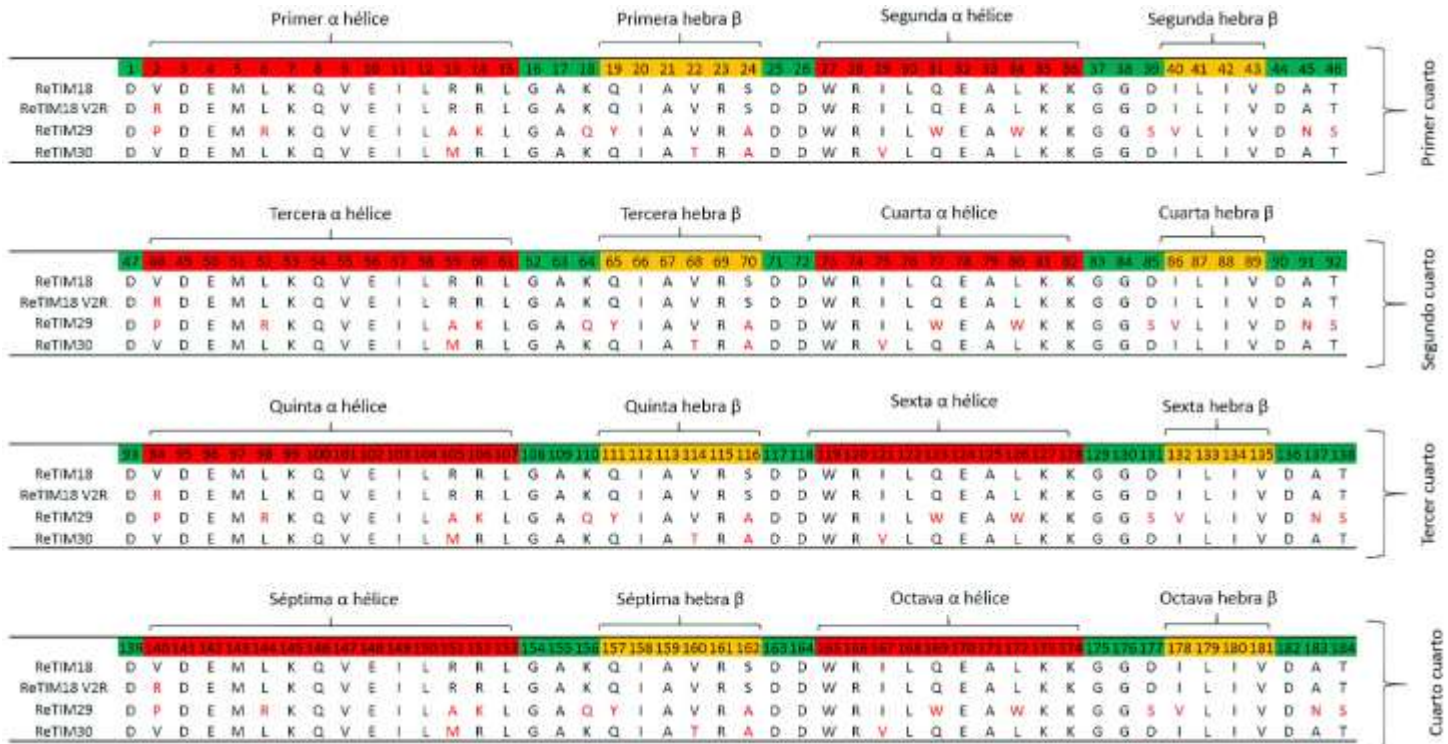


Figura 43. Alineamiento de las secuencias de los 3 modelos diseñados comparados con ReTIM18. Los residuos con color de fuente rojo son las mutaciones hechas por Rosetta. ReTIM18 V2R tiene 1 mutación por cuarto y 4 mutaciones totales. ReTIM29 tiene 13 mutaciones por cuarto y 52 mutaciones totales. ReTIM30 tiene 4 mutaciones por cuarto y 16 mutaciones totales. En la parte superior aparece el número de residuos con el texto resaltado en color verde, rojo y amarillo. El color verde representa a los residuos que forman parte de las asas, en color rojo aparecen los residuos que forman parte de las  $\alpha$  hélices y en color amarillo están los residuos que forman las hebras  $\beta$ .

## APÉNDICE

Nombre anterior	Nuevo nombre	Tipo de diseño modular
sTIM11	sTIM11	Primer diseño reportado
sTIM11-SS	NovoTIM0	sTIM11 sin puente de cisteínas
ReTIM1	NovoTIM1	Diseño en el núcleo interno
ReTIM2	NovoTIM2	Diseño en el núcleo interno
ReTIM3	NovoTIM3	Diseño en el núcleo interno
ReTIM4	NovoTIM4	Diseño en el núcleo interno
ReTIM5	NovoTIM5	Diseño en la cavidad inferior
ReTIM6	NovoTIM6	Diseño en la cavidad inferior
ReTIM7	NovoTIM7	Diseño en la cavidad inferior
ReTIM8	NovoTIM8	Diseño en la cavidad superior
ReTIM9	NovoTIM9	Diseño en la cavidad superior
ReTIM10	NovoTIM10	Diseño en la cavidad superior
ReTIM11	Proteína con mutación puntual que no sobreexpresó. Eliminada para estudios posteriores.	
ReTIM12	Proteína con mutación puntual que no sobreexpresó. Eliminada para estudios posteriores.	
ReTIM13	Proteína con mutación puntual. Eliminada para estudios posteriores.	
ReTIM14	Proteína con mutación puntual que no sobreexpresó. Eliminada para estudios posteriores.	
ReTIM15	Proteína con puente de cisteínas externo. Eliminada para estudios posteriores.	
ReTIM16	Proteína con puente de cisteínas externo. Eliminada para estudios posteriores.	
ReTIM17	NovoTIM12	Diseño en dos regiones: 1 + 8 (núcleo interno + cavidad superior)
ReTIM18	NovoTIM13	Diseño en dos regiones: 1 + 8 (cavidad inferior + cavidad superior)
ReTIM19	Proteína con mutación puntual que no sobreexpresó. Eliminada para estudios posteriores.	
ReTIM20	Proteína con mutación puntual que no sobreexpresó. Eliminada para estudios posteriores.	
ReTIM21	Proteína eliminada por combinación de mutaciones irrelevante.	
ReTIM22	NovoTIM15	Diseño en una sola región + Puente salino: 8 + PS (cavidad superior + puente salino superior)

ReTIM23	NovoTIM11	Diseño en dos regiones: 1 + 6 (núcleo interno + cavidad inferior)
ReTIM24	NovoTIM14	Diseño en tres regiones: 1 + 6 + 8 (núcleo interno + cavidad inferior + cavidad superior)
ReTIM25	NovoTIM19	Diseño en tres regiones + puente salino: 14 + PS (núcleo interno + cavidad inferior + cavidad superior + puente salino superior)
ReTIM26	NovoTIM17	Diseño en dos regiones + puente salino: 12 + PS (núcleo interno + cavidad superior + puente salino superior)
ReTIM27	NovoTIM18	Diseño en dos regiones + puente salino: 13 + PS (cavidad inferior + cavidad superior + puente salino superior)
ReTIM28	NovoTIM16	Diseño en dos regiones + puente salino: 11 + PS (núcleo interno + cavidad inferior + puente salino superior)

**Tabla 14. Nomenclatura de equivalencias.** Para estudios posteriores, se re bautizaron a las proteínas ReTIMs como NovoTIMs. Los detalles de los tipos de diseños modulares se pueden consultar las referencias [27] y [51].