



UNIVERSIDAD AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Estrategias *in vitro* e *in silico* para la búsqueda de inhibidores contra el componente proteico de la ribonucleasa P bacteriana

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Maestro en Ciencias

PRESENTA:

Brian Saidel López Duran

Dr. Alfredo Torres Larios  
Instituto de Fisiología Celular

Dr. Ismael Bustos Jaimes  
Facultad de Medicina

Dra. Adela Rodríguez Romero  
Instituto de Química

Ciudad de México. Noviembre, 2020



Universidad Nacional  
Autónoma de México



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

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

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

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




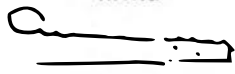
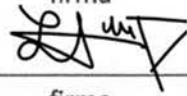
**SINODALES DESIGNADOS**

Presente

Estimado académico:

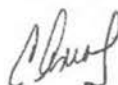
Los miembros del Subcomité Académico en reunión ordinaria del día 24 de agosto del presente año, conocieron la solicitud de asignación de **JURADO DE EXAMEN** para optar por el grado de **Maestro EN CIENCIAS**, con réplica de tesis "Estrategias in vitro e in silico para la búsqueda de inhibidores contra el componente proteico de la ribonucleasa P bacteriana", dirigida por el/la Dr(a). **Torres Larios Alfredo**.

De su análisis se acordó nombrar el siguiente jurado en el que se encuentra usted incluido:

		Acepto			
		Si <input checked="" type="checkbox"/>	No <input type="checkbox"/>		
PRESIDENTE	Sosa Peinado Alejandro			28/09/2020	
				fecha	firma
VOCAL	Arciniega Castro Marcelino	Si <input checked="" type="checkbox"/>	No <input type="checkbox"/>	28/09/2020	
				fecha	firma
VOCAL	Hernández Santoyo Alejandra	Si <input checked="" type="checkbox"/>	No <input type="checkbox"/>	04/10/20	
				fecha	firma
VOCAL	Arreguín Espinosa de los Monteros Roberto	Si <input type="checkbox"/>	No <input type="checkbox"/>	21/10/2020	
				fecha	firma
SECRETARIO	Domínguez Dueñas Laura	Si <input checked="" type="checkbox"/>	No <input type="checkbox"/>	28/9/2020	
				fecha	firma

Sin otro particular por el momento, aprovecho la ocasión para enviarle in cordial saludo

**Atentamente**  
**"POR MI RAZA, HABLARÁ EL ESPÍRITU"**  
 Cd. Universitaria, Cd. Mx., a 24 de agosto de 2020



**Dra. Ana Brigada Clorinda Aria Álvarez**  
**Coordinadora**

"Knowing your own ignorance is the first step to enlightenment."  
**-Patrick Rothfuss, The Wise Man's Fear**

# Índice

<b>1. Introducción</b>	<b>1</b>
1.1. Ribonucleasa P . . . . .	1
1.2. Proteína P . . . . .	1
1.3. Búsqueda de inhibidores contra la RNasa P . . . . .	2
1.4. Antecedentes . . . . .	3
1.5. Proteínas de fusión . . . . .	5
1.6. Pirrolisol tRNA sintetasa . . . . .	6
1.7. Proteína de fusión TmP-PylRS . . . . .	7
<b>2. Hipótesis</b>	<b>8</b>
<b>3. Objetivos</b>	<b>8</b>
<b>4. Metodología</b>	<b>9</b>
4.1. Expresión y purificación de la proteína P silvestre . . . . .	9
4.2. Búsqueda de nuevas condiciones de cristalización de la proteína P silvestre . . . . .	9
4.3. Colecta de datos . . . . .	10
4.4. Determinación de la estructura y comprobación del empaquetamiento . . . . .	10
4.5. Expresión y purificación de la proteína de fusión TmP-PylRS . . . . .	10
4.6. Cribado virtual de los compuestos provenientes de la biblioteca ZINC15 . . . . .	10
4.7. Análisis de similitud química . . . . .	11
4.8. Cribado virtual de compuestos provenientes de la biblioteca Sigma Aldrich . . . . .	11
4.9. Simulaciones de dinámica molecular . . . . .	11
<b>5. Resultados</b>	<b>12</b>
5.1. Proteína de fusión TmP-PylRS . . . . .	12
5.1.1. La proteína de fusión TmP-PylRS tiene bajos niveles de expresión . . . . .	12
5.1.2. Diferentes cepas de <i>Escherichia coli</i> expresan la proteína TmP-PylRS de manera diferencial . . . . .	13
5.1.3. La proteína TmP-PylRS se puede obtener pura y soluble, pero con bajo rendimiento	14
5.1.4. La proteína TmP-PylRS se puede obtener pura y con alto rendimiento, pero desnaturalizada . . . . .	14
5.1.5. No se logró la renaturalización de la proteína de fusión TmP-PylRS . . . . .	15
5.2. Proteína de fusión TmP-GST . . . . .	16
5.2.1. No se logró la renaturalización de la proteína de fusión TmP-GST . . . . .	16
5.3. Proteína P silvestre . . . . .	18
5.3.1. La proteína P cristaliza en nuevas condiciones utilizando la técnica de sembrado cruzado . . . . .	18
5.4. Búsqueda de inhibidores a través de métodos <i>in silico</i> . . . . .	21
5.4.1. El acoplamiento molecular de 700,000 compuestos resulta en 199 potenciales compuestos que interactúan con la proteína P en regiones relevantes . . . . .	22
5.4.2. La selección por acoplamiento molecular de los candidatos disponibles comercialmente resulta en 20 compuestos . . . . .	22
5.4.3. Simulaciones de dinámica molecular permiten reducir a cinco los candidatos para pruebas <i>in vitro</i> . . . . .	23
5.4.4. Ensayos control indican que los resultados del protocolo <i>in silico</i> deben tomarse con cautela . . . . .	25
5.4.5. El conjunto de datos obtenidos indica que es necesario afinar las distintas estrategias exploradas para mejorar la búsqueda de inhibidores. . . . .	26

6. Conclusiones	27
7. Perspectivas	27
8. Bibliografía	28
9. Anexos	I
9.1. Anexo 1. Formulación amortiguadores de lisis proteína TmP-PylRs . . . . .	I
9.2. Anexo 2. Condiciones de cristalización para proteína P con precipitante diferente a PEG.	II
9.3. Anexo 3. Sumario de valores obtenidos para la selección de inhibidores para la proteína P por métodos <i>in silico</i> . . . . .	III
9.4. Anexo 4. Lista de candidatos a pruebas <i>in vitro</i> . . . . .	IV

## Índice de figuras

<b>Figura 1.</b> La RNasa P realiza la maduración del extremo 5' de los pre-tRNAs. . . . .	1
<b>Figura 2.</b> La RNasa P es un complejo ribonucleoproteico con diferencias clave entre bacteria y eukarya. . . . .	1
<b>Figura 3.</b> Proteína P de <i>Thermotoga maritima</i> . . . . .	2
<b>Figura 4.</b> El sitio de unión al 5'-líder es altamente conservado. . . . .	2
<b>Figura 5.</b> El dominio bacteria posee una sola proteína sin homología de secuencia ni estructura con los dominios archaea o eukarya. . . . .	3
<b>Figura 6.</b> Plataforma para la búsqueda de inhibidores dirigidos contra la proteína P . . . . .	3
<b>Figura 7.</b> La proteína P puede unir otros ligandos además del 5'-líder . . . . .	4
<b>Figura 8.</b> La malla cristalina bloque el surco de interacción de la proteína P con el 5'-líder . . . . .	4
<b>Figura 9.</b> El cambio en las condiciones de cristalización puede modificar el arreglo de la malla cristalina . . . . .	5
<b>Figura 10.</b> La utilización de proteínas de fusión puede generar empaquetamientos cristalinos alternos	6
<b>Figura 11.</b> El dominio de unión a tRNA de la PylRS tiene un sitio diferente al de la proteína P .	6
<b>Figura 12.</b> La interacción de ambas proteínas podría aumentar la estabilidad del complejo con el pre-tRNA . . . . .	7
<b>Figura 13.</b> La búsqueda de empaquetamientos alternos de la proteína P se dividió en dos fases . .	12
<b>Figura 14.</b> La mayor cantidad de la proteína de fusión TmP-PylRS se encontró en la fracción insoluble de la cepa ArticExpress . . . . .	13
<b>Figura 15.</b> Es posible obtener pura y soluble a la proteína de fusión TmP-PylRS por afinidad a Ni-NTA y elución por gradiente escalonado . . . . .	14
<b>Figura 16.</b> La purificación por intercambio aniónico en condiciones desnaturizantes aumenta significativamente el rendimiento de la purificación de la proteína TmP-PylRS . . . . .	15
<b>Figura 17.</b> No fue posible obtener a la proteína TmP-PylRS soluble por los métodos de repliegamiento que exploramos . . . . .	15
<b>Figura 18.</b> La proteína de fusión TmP-PylRS se precipita en concentraciones menores a 6M de urea.	16
<b>Figura 19.</b> Es posible purificar la proteína de fusión TmP-GST en condiciones desnaturizantes .	17
<b>Figura 20.</b> Es posible obtener bajas cantidades de la proteína de fusión TmP-GST en condiciones óptimas para su cristalización . . . . .	17
<b>Figura 21.</b> La condición de cristalización optimizada favorece la formación de placas de mayor tamaño	18
<b>Figura 22.</b> Se encontraron 21 condiciones nuevas de cristalización con un diferente agente precipitante	19
<b>Figura 23.</b> Para la colecta de datos se seleccionaron los cristales con morfología diferente a los cristales semilla . . . . .	19
<b>Figura 24.</b> El cambio en el grupo espacial de la proteína no modifica el empaquetamiento cristalino	20
<b>Figura 25.</b> Los candidatos para pruebas <i>in vitro</i> fueron obtenidos a través de tres etapas de selección	21

<b>Figura 26.</b> De los ensayos de acoplamiento molecular con los compuestos de la biblioteca ZINC15, seleccionamos 199 compuestos . . . . .	22
<b>Figura 27.</b> De los 3 ensayos independientes de acoplamiento molecular se seleccionaron 20 candidatos para simulación por dinámica molecular . . . . .	23
<b>Figura 28</b> Los compuestos simulados pueden ser catalogados en 2 grupos de acuerdo con su $RMSD_{max}$	23
<b>Figura 29</b> Se seleccionaron 11 compuestos que forman complejos estables con la proteína P . . . . .	24
<b>Figura 30.</b> El $RMSD_{modal}$ nos ayudó a puntualizar la búsqueda a 5 candidatos para pruebas <i>in vitro</i>	24
<b>Figura 30.</b> El compuesto A se une de manera dosis-dependiente a la proteína P . . . . .	25

## Índice de cuadros

<b>Cuadro 1.</b> Se encontraron cuatro cepas con una inducción relativamente clara de la proteína de fusión TmP-PylRS . . . . .	13
<b>Cuadro 2.</b> Intervalos de titulación de los componentes de la solución madre para la cristalización de la proteína P silvestre . . . . .	18
<b>Cuadro 3.</b> El cambio de precipitante genera un cambio en el grupo espacial . . . . .	20
<b>Cuadro 4.</b> Formulación de amortiguadores de lisis para la proteína TmP-PylRS . . . . .	I
<b>Cuadro 5.</b> Estado de condiciones de cristalización para la proteína P obtenidas por sembrado cruzado	II
<b>Cuadro 6.</b> Datos obtenidos de la selección de inhibidores por métodos <i>in silico</i> . . . . .	III
<b>Cuadro 7.</b> Lista de candidatos a pruebas <i>in vitro</i> . . . . .	IV

## Lista de abreviaturas

<b>A<sub>600</sub></b>	Absorbancia óptica a 600 nanómetros
<b>BLI</b>	Interferometría de biocapa
<b>cLogP</b>	Coefficiente de partición octanol/agua
<b>DMSO</b>	Dimetilsulfóxido
<b>DTT</b>	Ditiotreitol
<b><i>E. coli</i></b>	<i>Escherichia coli</i>
<b>EDTA</b>	Ácido etilendiaminotetraacético
<b>EF</b>	Formula empírica
<b>FTU</b>	1-(4-Fluorophenyl)-2-tiourea
<b>GST</b>	Glutación S transferasa
<b>His-Tag</b>	Etiqueta de histidinas
<b>IC<sub>50</sub></b>	Concentración mínima inhibitoria al 50 %
<b>IUPAC</b>	Unidad internacional de química pura y aplicada
<b>LB</b>	Medio de cultivo bacteriano Luria-Bertani
<b>LIG</b>	Denominación a las moléculas candidatas a pruebas in silico
<b>Mh</b>	Mini hélice
<b>MW</b>	Peso molecular
<b>Ni-NTA</b>	Ácido Nitriloacético de níquel
<b>O.D.</b>	Densidad óptica
<b>P<sub>rha</sub>BAD</b>	Promotor inducible por L-ramnosa
<b>PDB</b>	Protein Data Bank
<b>PEG</b>	Polietilenglicol
<b>pH</b>	Potencial de hidrogeno
<b>Pre-tRNA</b>	Acido ribonucleico de transferencia inmaduro
<b>PyIRS</b>	Pirrolisil tRNA sintetasa
<b>PyMOL</b>	Molecular graphics system in Phyton
<b>RMSD</b>	Desviación de la media cuadrática entre posiciones atómicas
<b>RMSD<sub>max</sub></b>	Variación máxima de la desviación de la media cuadrática en la simulación.
<b>RMSD<sub>med</sub></b>	Mediana de la desviación de la media cuadrática en la simulación
<b>RMSD<sub>min</sub></b>	Variación mínima de la desviación de la media cuadrática en la simulación
<b>RMSD<sub>mod</sub></b>	Moda de la desviación de la media cuadrática en la simulación
<b>RMSD<sub>prom</sub></b>	Promedio de la desviación de la media cuadrática en la simulación
<b>RNA</b>	Acido ribonucleico
<b>RNA P</b>	Subunidad de RNA de la ribonucleasa P bacteriana
<b>RNasa P</b>	Ribonucleasa P
<b>RPM</b>	Revoluciones por minuto
<b>SMILE</b>	Simplified Molecular Input Line Entry System
<b><i>T. Maritima</i></b>	<i>Thermotoga marítima</i>
<b>TCEP</b>	tris(2-carboxyethyl)phosphine
<b>TEV</b>	Virus del grabado del tabaco
<b>TmP</b>	Proteína P de <i>Thermotoga marítima</i>
<b>tRNA</b>	Acido ribonucleico de transferencia
<b>WT</b>	Wild type, tipo silvestre



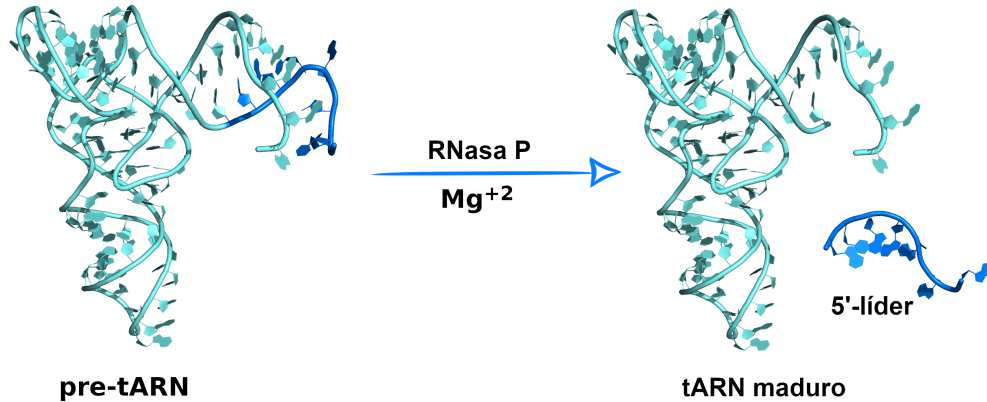
## Resumen

La ribonucleasa P (RNasa P) es un complejo ribonucleoproteico esencial que corta el extremo 5' de los pre-tRNAs en todos los dominios de la vida. En particular, la RNasa P bacteriana contiene una proteína (proteína P) que posee un surco que interacciona con el extremo 5' del pre-tRNA. Este surco representa un blanco para el desarrollo de nuevos inhibidores antibióticos, ya que no tiene un equivalente estructural ni en secuencia en otros dominios. A pesar de que existen avances para la búsqueda de inhibidores en formatos de alto rendimiento, existen aún limitantes técnicas en este sistema que impiden realizar búsquedas eficientes a nivel de la estructura tridimensional a través de cristalografía de rayos X. En este trabajo se exploraron algunas opciones para trascender este problema. Por un lado, se buscaron nuevas condiciones de cristalización de la proteína P, así como producirla fusionada con dos distintas proteínas, con objeto de crear espacios dentro del cristal que permitieran la difusión o cocrystalización con compuestos mayores a 250 Da. Debido a que el éxito de estas estrategias *in vitro* fue limitado, se desarrolló un esquema *in silico* en donde, por medio de bibliotecas de compuestos y su utilización en protocolos de acoplamiento y dinámica molecular, se obtuvieron cinco candidatos a inhibidores. A manera de control, se probó un análogo estérico de estos candidatos en ensayos de unión a la proteína P, obteniéndose una afinidad en el rango de micromolar alto. En los ensayos *in silico*, de manera sorpresiva, el compuesto es rechazado por acoplamiento molecular, pero de los ensayos de dinámica molecular se concluye que el compuesto debería ser estable en su interacción con la proteína P. En conjunto, los resultados sugieren que podrían buscarse otras estrategias, como mutagénesis dirigida acoplada a ensayos de unión, o NMR, que permitan inferir a nivel estructural el sitio de unión de posibles inhibidores.

# 1. Introducción

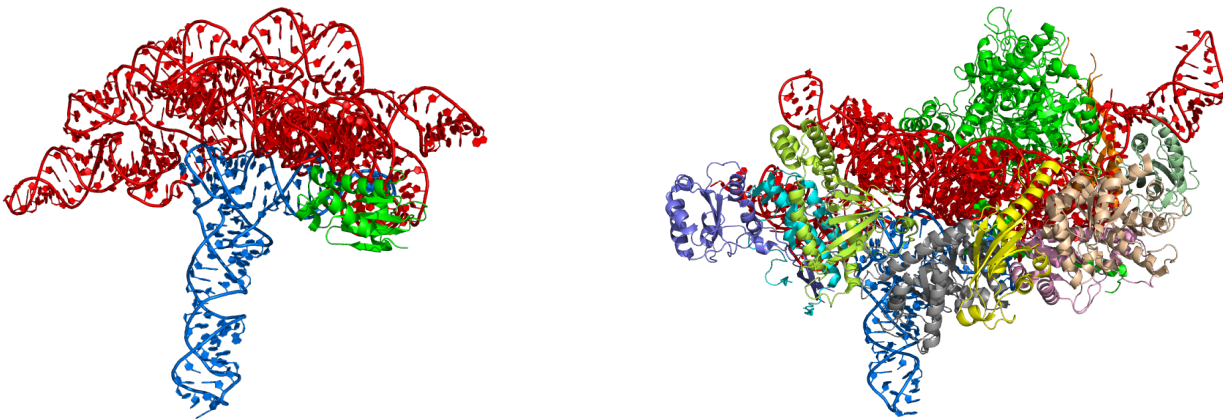
## 1.1. Ribonucleasa P

La ribonucleasa P es un complejo ribonucleoproteico con actividad endonucleasa que realiza la catálisis metalodependiente de la maduración del extremo 5' de los pre-tRNA's<sup>1,2</sup> (figura 1).



**Figura 1: La RNasa P realiza la maduración del extremo 5' de los pre-tRNAs.** Derecha, muestra el sitio de corte del extremo 5' de los pre-tRNA's. Izquierda, muestra los productos de la catálisis de la RNasa P. En cian se muestra el tRNA maduro. En azul, el extremo 5'-líder.

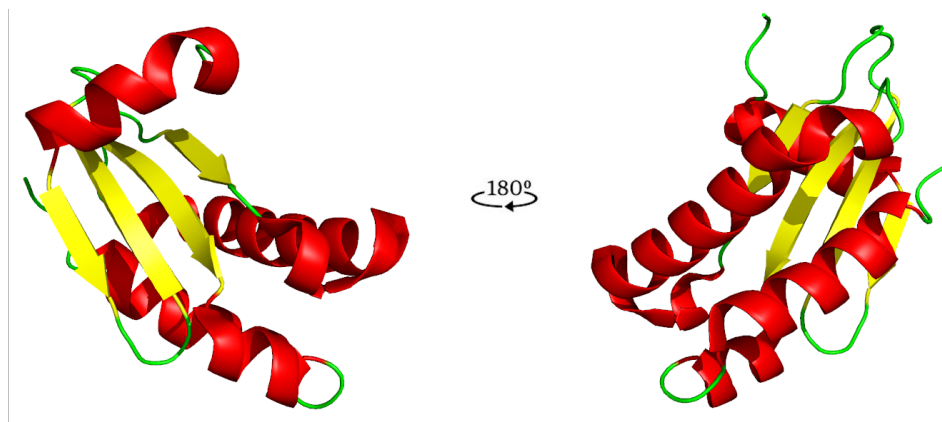
Este complejo posee un componente de RNA que lleva a cabo la catálisis<sup>1</sup> y uno o más componentes proteicos, que varían entre bacteria, eukarya y archaea<sup>3</sup>. En general, el componente de RNA es de mayor tamaño en el dominio bacteria (figura 2, izquierda) y el número de componentes proteicos aumenta en archaea y eukarya<sup>4,5</sup> (figura 2, derecha).



**Figura 2: La RNasa P es un complejo ribonucleoproteico con diferencias clave entre bacteria y eukarya.** Izquierda, Muestra la estructura tridimensional de la RNasa P bacteriana en complejo con su sustrato (código PDB 3Q1Q). La subunidad de RNA (400 nucleótidos, 125 kDa) y la proteína (120 aminoácidos, 12 kDa) se muestran en rojo y verde. Derecha, muestra la estructura tridimensional de la RNasa P humana en complejo con su sustrato (código PDB 6AH3). La subunidad de RNA (340 nucleótidos, 110 kDa) se muestra en rojo y las subunidades proteicas (10 subunidades) en diferentes colores. El pre-tRNA se muestra en azul.

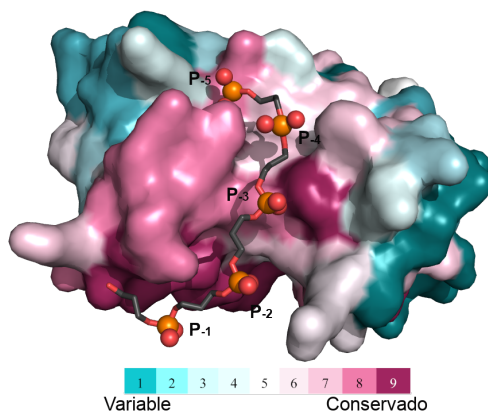
## 1.2. Proteína P

En particular, la RNasa P bacteriana contiene una sola unidad proteica (proteína P) de aproximadamente 120 aminoácidos<sup>6</sup>. La estructura cristalográfica de la proteína P de *Thermotoga maritima* contiene un solo dominio globular formado por 4 hebras  $\beta$  rodeadas de 4 hélices  $\alpha$ <sup>7</sup> (figura 3).



**Figura 3: Proteína P de *Thermotoga maritima*.** Se muestran los componentes de estructura secundaria de la proteína P, en rojo hélices  $\alpha$ , en amarillo hojas  $\beta$  y en verde loops.

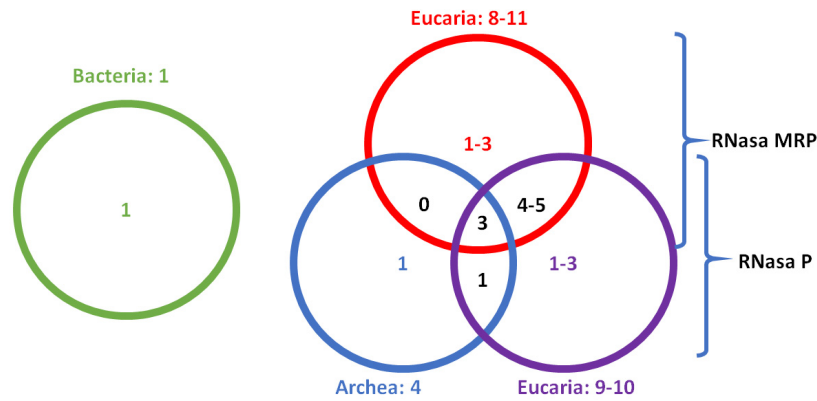
Se ha demostrado *in vitro* que el componente de RNA es catalíticamente activo en ausencia del componente proteico. Sin embargo, este último es esencial para la función *in vivo*. La eficiencia de la catálisis por parte de la holoenzima aumenta en dos órdenes de magnitud comparada con la del RNA solo<sup>8</sup>. Este aumento se atribuye en parte a la estabilización del extremo 5' por parte de la proteína P. La proteína P interacciona con el sustrato por medio de un surco con un alto grado de conservación<sup>9</sup> (figura 4).



**Figura 4: El sitio de unión al 5'-líder es altamente conservado.** Muestra los sitios conservados de la proteína en código de colores, en azul las regiones variables y en magenta las regiones conservadas (código PDB 1NZ0), los factores de conservación fueron obtenidos comparando 150 homólogos estructurales depositados en el PDB por medio del servidor Consurf. En color negro se muestra la posición del esqueleto de fosfatos al alinear la estructura de la holoenzima con la proteína.

### 1.3. Búsqueda de inhibidores contra la RNasa P

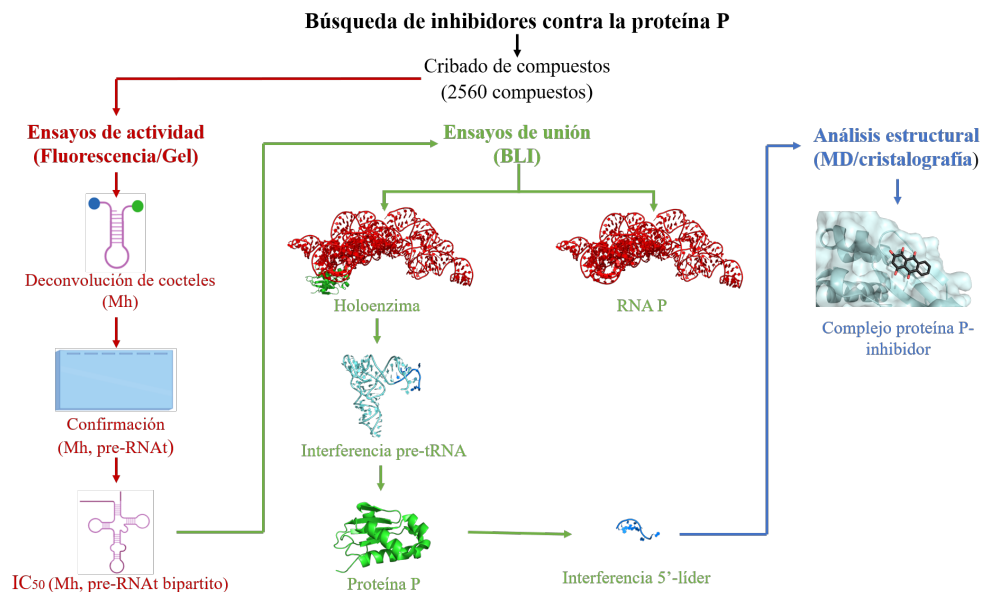
La subunidad de RNA ha sido el blanco predilecto para el desarrollo de inhibidores, debido a su función catalítica<sup>10-13</sup>. Sin embargo, se debe resaltar que este componente es común para todos los dominios de la vida. Por otro lado, la proteína P se encuentra sólo en el dominio bacteria, y no cuenta con homólogos en secuencia ni estructura con los otros dos dominios<sup>14,15</sup> (figura 5). Por ello, creemos que la proteína P representa un mejor blanco para el desarrollo de inhibidores específicos<sup>16-18</sup>.



**Figura 5:** El dominio bacteria posee una sola proteína sin homología de secuencia ni estructura con los dominios archaea o eukarya. Diagrama de Venn indicando la distribución general de las subunidades proteicas de la RNasa P en los tres dominios de la vida, cada círculo representa el total de componentes proteicos presentes de las ribonucleasas P de cada dominio y las intersecciones muestran los componentes homólogos compartidos (Tomado de Hernández *et. al*, 2012).

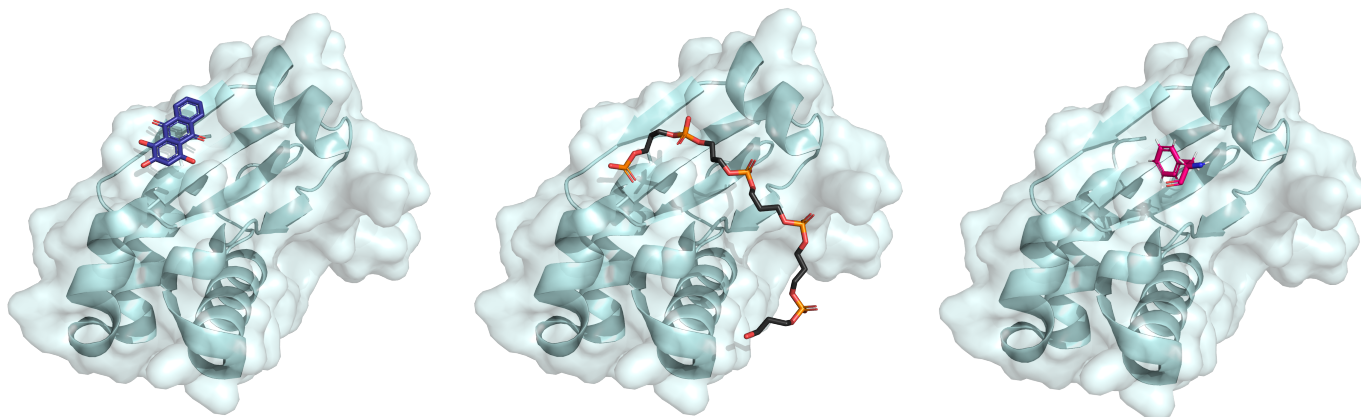
### 1.4. Antecedentes

En el laboratorio de adscripción se diseñó una plataforma de alto rendimiento para la búsqueda de inhibidores dirigidos contra el componente proteico de la ribonucleasa P bacteriana<sup>19</sup>. Durante estos ensayos se analizaron distintas bibliotecas de compuestos y fragmentos a través de ensayos de actividad, unión y análisis de estructuras tridimensionales<sup>20-23</sup>. La búsqueda inicio evaluando distintos cocteles de compuestos por ensayos de actividad de la holoenzima con un sustrato fluorescente (minihélice), se confirma la inhibición de la actividad por medio de gel y se obtiene la IC<sub>50</sub> con el sustrato canónico (figura 6, rojo). Los compuestos seleccionados de la deconvolución de los cocteles son descartados si muestran afinidad por el RNA P. Contrariamente, si muestran afinidad por la holoenzima y la proteína P se evalúa si interfieren la unión del pre-tRNA y el 5'-líder (figura 6, verde). Finalmente, determinó el sitio de unión de los compuestos por medio de cristalografía de rayos x y/o simulaciones de dinámica molecular (figura 6, azul).



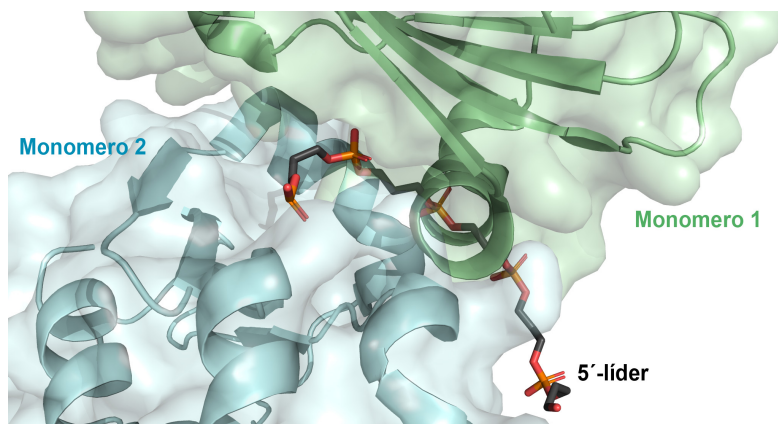
**Figura 6:** Plataforma para la búsqueda de inhibidores dirigidos contra la proteína P. Se muestra el esquema de la plataforma desarrollada para la búsqueda de inhibidores contra la proteína P bacteriana (Tomado de Madrigal *et. al*, 2012).

Mediante la plataforma desarrollada durante el trabajo del Dr. Alejandro Madrigal, se encontró el compuesto purpurina, que presenta una inhibición total de la unión del extremo 5'-líder en el rango micromolar. La estructura cristalográfica revela que la unión de la purpurina con el surco de interacción del 5'-líder (figura 7, izquierda) podría impedir estéricamente la unión del sustrato (figura 7, centro). En paralelo, se determinó la estructura en complejo con el compuesto 1-(4-Fluorophenyl)-2-thiourea, un fragmento cuyo sitio de unión es cercano al de la purpurina (figura 7, derecha). Estos dos compuestos representan una entidad atractiva para la obtención de un compuesto que sirva como escalafón para el desarrollo de inhibidores con mayor afinidad por el surco de unión al 5'-líder<sup>19</sup>.



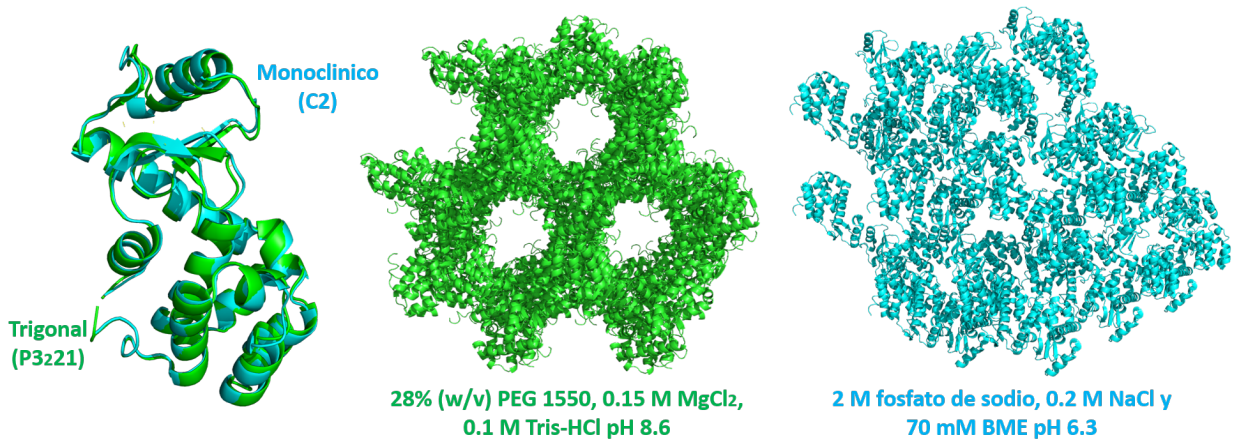
**Figura 7: La proteína P puede unir otros ligandos además del 5'-líder.** Se muestra la misma orientación de la proteína P en los 3 paneles. Izquierda, se muestra el sitio de unión del compuesto purpurina (morado) encontrado por medio de cristalografía de rayos X (código PDB 6MAX). Centro, se muestra la superposición del sitio de interacción de extremo 5'-líder. Derecha, se muestra la localización del compuesto (4-fluorophenyl)- 2-thiourea (magenta) encontrado en una búsqueda de fragmentos por medio de cristalografía de rayos X (codigo PDB 6CQC).

Dentro de la plataforma propuesta, la determinación de estructuras tridimensionales por cristalografía de rayos X representa un punto fundamental para entender la relación estructura-función del sistema y poder desarrollar mejores inhibidores. En este contexto, a pesar de que se logró obtener la estructura de la proteína P en complejo con purpurina (figura 7, código PDB 6MAX), la red cristalina o “empaquetamiento” impide la unión de compuestos de mayor tamaño, o de sustratos como el 5'-líder (figura 8). Por ello, trascender el problema del empaquetamiento resulta de gran importancia para el desarrollo de más y mejores inhibidores.



**Figura 8: La malla cristalina bloque el surco de interacción de la proteína P con el 5'-líder.** El sitio de unión con el 5'-líder (negro) de la proteína P (verde) tiene contactos con otra unidad proteica dentro de la malla cristalina.

Con el fin de obtener empaquetamientos alternos, en el presente trabajo se desarrollaron dos enfoques: la utilización de proteínas de fusión y la búsqueda de nuevas condiciones de cristalización. De manera alterna, se abordó el problema de la búsqueda de inhibidores a través de métodos *in silico*: acoplamiento y dinámica molecular. La formación de la malla cristalina consiste en regiones de contacto, donde grupos específicos toman un papel clave para la estabilización de un arreglo regular y periódico de las moléculas de proteína<sup>24</sup>. Este fenómeno es influenciado por las propiedades individuales de cada proteína y factores externos, incluyendo la composición de la solución de cristalización. En diferentes sistemas de estudio, se ha demostrado, que cambios en los componentes de la condición de cristalización (amortiguador, precipitante y aditivo) puede modificar el arreglo de la malla cristalina<sup>25-29</sup> (figura 9.)

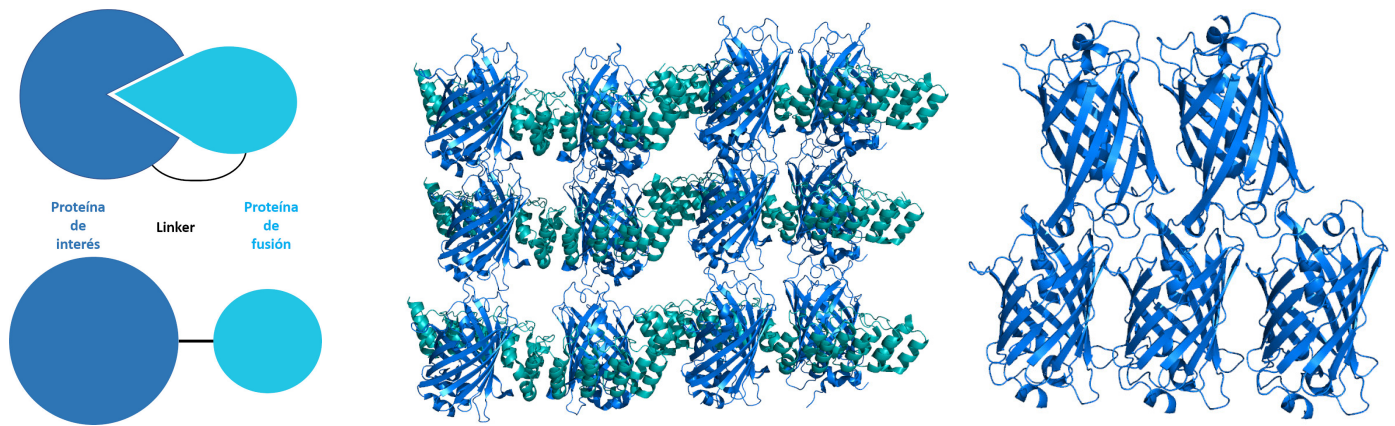


**Figura 9: El cambio en las condiciones de cristalización puede modificar el arreglo de la malla cristalina.** Izquierda, se muestra la sobreposición de dos estructuras de lisozima obtenidas en diferentes grupos espaciales, en verde se muestra la proteína lisozima cristalizada en el grupo espacial P3<sub>2</sub>21 (código PDB 2LZM) y en cian la proteína cristalizada en el grupo espacial C2 (código PDB 201L). En el centro se muestra el arreglo de la malla cristalina obtenida en la condición: 2M de fosfato de sodio, 0.2 M de cloruro de sodio y 70 mM BME pH 6.3. Derecha, se muestra el arreglo de la malla cristalina obtenida en la condición: 28 % (w/v) PEG 1550, 0.15 M cloruro de magnesio.

## 1.5. Proteínas de fusión

Se ha reportado que las proteínas de fusión pueden promover empaquetamientos alternos<sup>22</sup> (figura 10). En particular, la proteína P de *Thermotoga maritima* ha sido producida como una proteína de fusión unida a GST<sup>30</sup>, pero en nuestro grupo no se ha intentado purificar y cristalizar la construcción fusionada. Por las características individuales de ambas proteínas, creemos que es posible obtener la estructura cristalográfica de las proteínas fusionadas pero esta construcción no permitiría la cocrystalización con el pre-tRNA. Por lo que para el presente trabajo, también se realizó una búsqueda de candidatos a proteínas de fusión que cumplieran con las siguientes características:

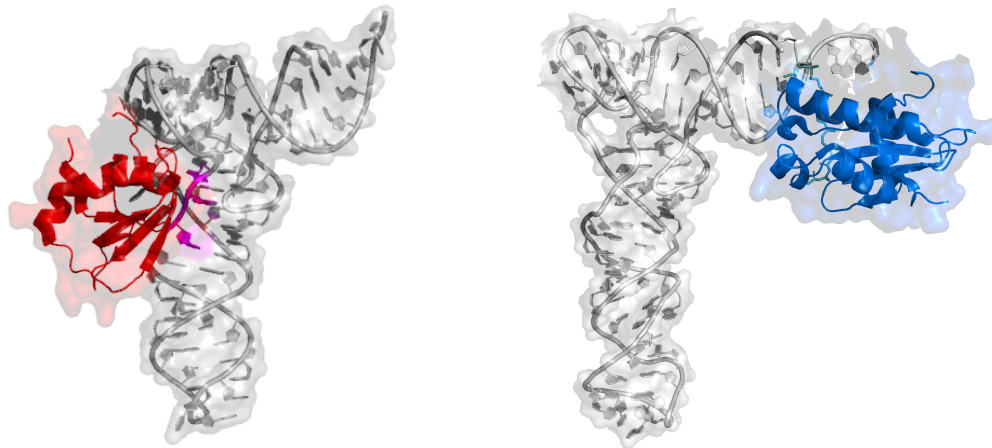
- Generalmente las proteínas de fusión son de menor tamaño que la proteína de interés. La proteína P tiene un peso de 12 kDa por lo que limitamos la búsqueda a proteínas monoméricas y con un peso menor a 25 kDa.
- Con el fin de obtener un complejo ternario de las dos proteínas y el pre-tRNA, nos centramos en proteínas con un sitio de interacción al tRNA diferente al de la proteína P,
- Que pudiera ser purificada en condiciones desnaturalizantes como ocurre con la proteína P.



**Figura 10: La utilización de proteínas de fusión puede generar empaquetamientos cristalinos alternos.** Izquierda, muestra el esquema de la construcción de proteínas de fusión heterólogas. Centro, muestra el empaquetamiento de la proteína GFP en azul como proteína de fusión con la proteína DARPin en cian (código PDB 5MAD). Derecha, muestra el empaquetamiento intrínseco de la proteína GFP (código PDB 6FWW). En conjunto, se puede apreciar que la proteína de fusión interviene en los contactos cristalinos, creando espacios más amplios en la red para la eventual unión de otros ligandos.

## 1.6. Pirrolisol tRNA sintetasa

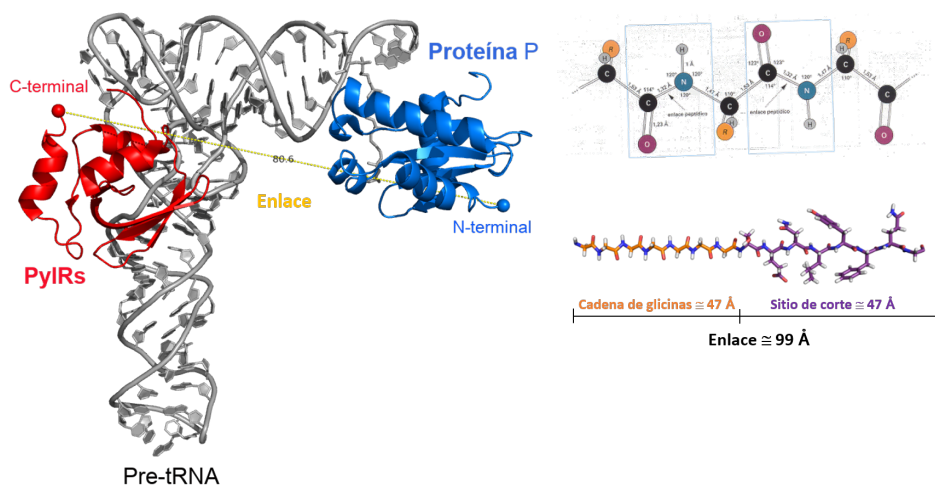
Como resultado de esta búsqueda, se seleccionó un dominio de unión a tRNA de la Pirrolisil tRNA sintetasa (PylRS)<sup>31-33</sup>. La estructura cristalográfica del complejo de dicho dominio con el tRNA<sup>Pyl</sup> muestra una proteína globular de 101 aminoácidos que interactúa con 3 bases del tRNA<sup>31</sup>. Este sitio de interacción (figura 11, izquierda) es distinto al surco de interacción del 5'-líder por parte de la proteína P<sup>34</sup> (figura 11, derecha).



**Figura 11: El dominio de unión a tRNA de la PylRS tiene un sitio diferente al de la proteína P.** Izquierda, se muestra el sitio de interacción (magenta) del dominio de unión a tRNA de la PylRS (rojo) en complejo con su sustrato (código PDB 5V6X). Derecha, se muestra el sitio de interacción del 5'-líder con la proteína P en complejo con su sustrato

## 1.7. Proteína de fusión TmP-PylRS

La proteína de fusión TmP-PylRS está constituida por la secuencia codificante a un fragmento cristalizante de la pirrolisil tRNA sintetasa (PylRS) en el extremo C-terminal, un enlace de 8 glicinas y la secuencia de corte para la proteasa TEV entre las dos proteínas y la secuencia de la proteína P (TmP) en el extremo N-terminal (Figura 12). Pensamos que la proteína de fusión TmP-PylRS podría generar una malla cristalina alterna para la obtención de complejos de la proteína con ligandos de diversos tamaños. Además, la interacción de ambas proteínas con el pre-tRNA podría aumentar la estabilidad del complejo, lo que nos permitiría obtener una estructura cristalina a una resolución suficiente para observar la interacción del extremo 5'-líder con la proteína P.



**Figura 12: La interacción de ambas proteínas podría aumentar la estabilidad del complejo con el pre-tRNA.** Izquierda, se muestra la construcción hipotética de la proteína de fusión entre la proteína P (azul), la distancia mínima del enlace entre las dos proteínas (amarillo) y PylRS (rojo) en complejo con el pre-tRNA. Izquierda, se muestra la longitud de los enlaces C-C, C-N, C-O y la composición del enlace entre las dos proteínas.



## 2. Hipótesis

Es posible cristalizar a la proteína P fusionada con otra proteína y/o encontrar nuevas condiciones de cristalización de la proteína P para promover empaquetamientos cristalinos alternos.

## 3. Objetivos

### Objetivo general

Purificar y cristalizar la proteína P de *Thermotoga marítima* y proteínas fusionadas con la misma.

### Objetivos particulares

- Sobreexpresar y purificar diferentes proteínas fusionadas con la proteína P de *Thermotoga marítima*.
- Cristalizar diferentes proteínas.
- Determinar de la estructura tridimension de las proteínas por cristalografía de rayos X y verificar los contactos cristalinos

## 4. Metodología

### 4.1. Expresión y purificación de la proteína P silvestre

El gen que codifica para la proteína P (síntesis del gen realizada por ATUM DNA 2.0) de *Thermotoga maritima* fue clonado en el vector pD861 (KanR, inducible a ramnosa en cualquier cepa de *Escherichia coli*). El plásmido fue contrasformado en la cepa de *E. coli* BL21 (DE3) pLyss con el plásmido pJ204 que codifica para la subunidad de RNA de RNAsa P del mismo organismo. Las células se crecieron en medio LB suplementado con 25 µg/ml de kanamicina, 100 µg/ml de ampicilina y 34 µg/ml de cloranfenicol a 37 °C hasta que la absorbancia a 600 nm ( $A_{600}$ ) alcanzara un valor de 0.6. La subunidad de RNA fue inducida por la adición de isopropil-β-D-1-tiogalactopiranosido (IPTG) a una concentración final de 1 mM, continuando el cultivo por 1h en las mismas condiciones. Posteriormente, se adicionó ramnosa a una concentración final de 3 mM para la inducción de la proteína P; el cultivo se mantuvo en las mismas condiciones durante 6-7 h antes de cosechar. El cultivo fue cosechado y resuspendido en un volumen adecuado de amortiguador de lisis (50 mM Tris-HCl pH 7.5, 4 mM EDTA, 5 % (v/v) glicerol y 0.1 % (v/v) NP40). La suspensión celular fue lisada por sonicación. El lisado clarificado fue centrifugado a 20,00 g; se recuperó el sobrenadante, el cual se llevó a una condición desnaturalizante con una concentración final de 5M de urea. Para la purificación, la proteína fue inmovilizada en una columna de intercambio catiónico SP Sepharose (Amersham Biosciences) preequilibrada en amortiguador A1 (50 mM Tris-HCl pH 7.5, 0.2 mM EDTA, 5M Urea) y eluida con un gradiente de 0.1M - 3M de NaCl en amortiguador B1 (amortiguador A1+3M NaCl). La proteína P eluye a una concentración de 1.3-1.8M NaCl. La fracción colectada del paso anterior fue sometida a un proceso de desalado para eliminar las condiciones desnaturalizantes y replegar la molécula a través de una columna Sephadex G-25 (GE Helthcare) equilibrada en amortiguador de desalado (50 mM Tris-HCl pH 7.5, 0.2 mM EDTA y 1 M NaCl); la fracción colectada de este paso de purificación fue diluida hasta una concentración de 200 mM NaCl en amortiguador A2 (20 mM Tris-HCl pH 7.5 0.2 mM EDTA) y cargada en una columna de intercambio catiónico fuerte de alta resolución Source 15 (Amersham Biosciences) preequilibrada en amortiguador A2; se eluyó con un gradiente de 0.2 M - 3M NaCl en amortiguador B2 (amortiguador A2+3M NaCl). La proteína eluyó en una concentración 1M - 1.2 M NaCl. La fracción colectada del último paso de purificación pasó por un proceso de desalado por intercambio de amortiguador en un tubo de Amicon Ultra 10,000 MWCO (Millipore) donde se concentró hasta 13 mg/ml en el amortiguador C (20 mM Tris-HCl pH 7.5, 0.2 mM EDTA y 50 mM NaCl). La proteína purificada fue almacenada a 4 °C.

### 4.2. Búsqueda de nuevas condiciones de cristalización de la proteína P silvestre

Los ensayos de cristalización fueron realizados por la técnica de difusión de vapor en su modalidad de gota sedente. Los cristales iniciales fueron obtenidos con una concentración de 3 mg/ml de proteína en una condición optimizada (100 mM acetato de sodio trihidratado pH 5.0, 200 mM sulfato de potasio, 18 % (m/v) PEG 1500) en una temperatura de 25°C, a partir de la reportada por Krivenko y colaboradores en el 2002. Se recuperaron 10 cristales de una longitud aproximada de 150 µm en 50 µl de solución madre; la suspensión fue tratada con perlas de vidrio y un vortex; a partir de la suspensión se prepararon tres diluciones (1:100,000, 1:1,000,000 y 1:10,000,000) en un volumen adecuado de solución madre. La búsqueda de nuevas condiciones de cristalización se realizó por la técnica de sembrado cruzado<sup>35</sup> en placas de 3 pozos; se colocaron gotas de 3 µl (1 µl de proteína (3 mg/ml) + 1 µl disolución de cristales + 1 µl de solución madre) con las diferentes disoluciones de cristales y en el reservorio una de las 96 diferentes condiciones de cristalización de los kits de cristalización de Hampton Reserch (Salt Rx, Crystal Screen, PEG/Ion, Natrix, Index). El crecimiento de los cristales fue realizado a 25°C y evaluado después de una semana. De las condiciones que promueven la cristalogénesis de la proteína, se descartaron todas aquellas cuyo precipitante fuera PEG (condición control ya conocida) y se evaluó el tamaño y morfología de los cristales obtenidos en condiciones con un agente precipitante diferente. Para la colecta de datos se seleccionaron los cristales obtenidos en la condición A1 (MPD) y E11 (1,6-Hexanediol) del Crystal Screen.

### 4.3. Colecta de datos

Los juegos de datos fueron colectados a partir de monocristales a temperatura criogénica en el sector LS-CAT, línea 21-ID-G (energía de 12 keV) en el sincrotrón Advanced Photon Source (APS) del Laboratorio Nacional Argonne en Illinois, U.S.A. Durante la colecta, la longitud de onda fue de 0.97854 Å, la distancia al detector varió entre 180-360 mm y se colectaron 180° con oscilaciones de 0.20° - 0.35° por imagen, el tiempo de exposición promedio por imagen fue 2.25 s. Las reflexiones se indexaron con el programa MOSFLM<sup>36</sup>, se integraron con el programa XDS<sup>37</sup> y se escalaron con SCALA<sup>38</sup>.

### 4.4. Determinación de la estructura y comprobación del empaquetamiento

La estructura de la proteína P fue determinada por el método de reemplazo molecular con el uso del programa PHASER<sup>39</sup> de CCP4. El modelo estructural de la subunidad proteica de la RNasa P de *T. maritima* fue utilizado como modelo de búsqueda (código PDB 1NZ0). El modelo obtenido fue sometido a una inspección manual en el programa Coot<sup>40</sup> donde se realizó la generación de la malla cristalina y la sobreposición del modelo de la holoenzima (código PDB 3Q1R) para comprobar el empaquetamiento del cristal.

### 4.5. Expresión y purificación de la proteína de fusión TmP-PylRS

La producción y purificación de la proteína de fusión TmP-GST se realizó con un protocolo modificado a partir del publicado por Krivenko y colaboradores (2002). La cepa de *E. coli* BL21(DE3)pLysS se transformó con el vector pGEX4Ta con el gen que codifica para la proteína P. Dicho vector tiene un promotor T7 lac, una etiqueta 6x His-Tag y una proteína de fusión GST en el N-terminal de la proteína y un corte de la proteasa de TEV. A una densidad óptica de 0.4, se induce la expresión de la proteína con la adición de 1mM de IPTG, continuando el cultivo por 6-7 hrs en las mismas condiciones. Se centrifugó el cultivo a 5,000g, se resuspendió en un volumen apropiado (20-30 ml/L) de amortiguador de lisis (50 mM fosfatos de sodio pH 8, 50 mM NaCl 20 mM Imidazol, 10 %<sup>(v/v)</sup> glicerol, 0.1 %<sup>(v/v)</sup> 1 mM TCEP). El sobrenadante se cargó en una columna de afinidad a iones de níquel preequilibrada en amortiguador A (50 mM fosfatos de sodio pH 8, 50 mM KCl, 20 mM imidazol); se realizó un lavado de 20 volúmenes de columna y se eluyó con amortiguador B (50 mM fosfatos de sodio pH 8, 50 mM KCl, 2M imidazol). La fracción colectada pasó por un proceso de desalado a través de una columna Sephadex G-25 para eliminar el imidazol contenido en el amortiguador; la elución se realizó con amortiguador C (50 mM Tris-HCl pH 7.5, 5M urea, 1.5M KCl). La fracción colectada del paso anterior fue dializada en un cassette Slide-A-Lyser G2 10,000 MWCO (Thermo Scientific) diluyendo la concentración de urea en intervalos de 0.5M en amortiguador de replegado (50 mM Tris-HCl pH 7.5, 1.5M KCl) cada 2 h para eliminar las condiciones desnaturizantes. La proteína dializada se intercambió de amortiguador en un tubo de Amicon Ultra 10,000 MWCO (Millipore) donde se concentró en el amortiguador C (50 mM Tris-HCl pH 7.5 y 50 mM NaCl). La proteína purificada fue almacenada a 4°C.

### 4.6. Cribado virtual de los compuestos provenientes de la biblioteca ZINC15

Para la evaluación de un amplio espacio químico, se construyó una biblioteca de fragmentos a partir de las estructuras depositadas en la base de datos ZINC15. Esta biblioteca consta de 700,000 compuestos que fueron evaluados por acoplamiento molecular con el programa AutoDock Vina<sup>41</sup>. Previamente, se procesó el modelo de la estructura de la RNasa P bacteriana (código PDB 3Q1R), eliminando el RNA P y el pre-tRNA. La proteína rasurada y los ligandos fueron preparados para el acoplamiento molecular con el programa AutoDock Tools. Se realizó el cribado virtual sobre la estructura de la proteína P considerando dos cajas de simulación diferentes en dos ensayos independientes. La primera caja de simulación, con dimensiones de 20Åx20Åx20Å, consideró el surco de interacción con el 5'-líder y la posición cristalográfica del compuesto 1-(4-Fluorofenil)-2-tiourea (código PDB 6CQC); la segunda caja de simulación consideró el resto de la proteína y tenía unas dimensiones de 22Åx26Åx24Å. Consideramos un acierto a aquellos

compuestos con un puntaje  $\leq 7.0$  Kcal/mol. Los aciertos provenientes de ambas cajas de simulación fueron unidos en un solo grupo de datos y evaluados en un análisis de agrupamiento jerárquico.

#### 4.7. Análisis de similitud química

Utilizamos las herramientas provenientes de la librería de Python RTKIT para generar un método de agrupamiento jerárquico que evaluara la similitud química<sup>42</sup> entre las moléculas candidatas provenientes del acoplamiento molecular. Previo al análisis, las estructuras de las poses del acoplamiento molecular fueron transformadas a formato SMILES con el programa Open Babel. Estas moléculas fueron transformadas en una matriz de “0s” y “1s” que denotaban ausencia y presencia respectivamente de subestructuras almacenadas en la librería FP4 de Open Babel<sup>43</sup>. Estas matrices representan la “huella digital” para cada molécula. Posteriormente se calculó la similitud química entre cada par de compuestos a través del cálculo del coeficiente de Tanimoto. El recíproco del coeficiente de Tanimoto<sup>44</sup> para cada par de compuestos fue denominado distancia “Z”, el cual fue representado en una matriz de diferencia. A su vez, la matriz de diferencia sirvió de base para el análisis de agrupamiento jerárquico. Los resultados del agrupamiento fueron graficados en un dendrograma que denotaba los grupos de similitud. La información de estos grupos fue extraída en forma de tabla para su posterior uso. Las instrucciones de este análisis fueron incluidas en un archivo texto.

#### 4.8. Cribado virtual de compuestos provenientes de la biblioteca Sigma Aldrich

Se utilizó un compuesto proveniente de cada grupo del análisis de similitud química de la biblioteca ZINC15 para la búsqueda de moléculas disponibles en la compañía Sigma Aldrich. De los resultados de la búsqueda se construyó una biblioteca de 599 compuestos que se evaluaron por acoplamiento molecular con los parámetros antes descritos. Adicionalmente se realizó un ensayo de acoplamiento molecular flexible en la caja de simulación 1. Este ensayo incorporaba flexibilidad a las cadenas laterales de los aminoácidos que conforman el surco de interacción al 5'-líder (Phe 17, Phe 21, Val 34, Leu 35, Arg 45, Lys 51, Arg 52, Lys 53, Arg 66, Ile 70, Arg 89 y Lys 90). Los resultados de los tres ensayos de acoplamiento molecular fueron promediados. Se tomaron como aciertos aquellos compuestos dentro de los primeros 20 lugares de la clasificación.

#### 4.9. Simulaciones de dinámica molecular

Con el fin de purgar falsos positivos, los “aciertos” seleccionados del acoplamiento molecular de la biblioteca Sigma Aldrich fueron evaluados a través de dinámicas moleculares. La generación de la topología para cada ligando fue realizada con la herramienta externa ACPYPE (Antechamber Python Paser Interface)<sup>45</sup>. Se construyó el complejo de la proteína con el ligando para cada “acierto” del acoplamiento molecular de la biblioteca Sigma Aldrich. Las simulaciones de dinámica molecular de cada complejo fueron realizadas con el programa GROMACS, con el campo de fuerza AMBER99SB-ILDN; este campo de fuerza fue seleccionado por: 1) Ser compatible con la parametrización del ligando, 2) se ha demostrado una buena correlación entre las simulaciones con este campo de fuerza y las mediciones experimentales<sup>46</sup> y 3) este campo de fuerza mejora en el cálculo del potencial de la rotación de cadenas laterales<sup>47</sup>. Las simulaciones se llevaron a cabo en una caja de  $\approx 6.6$  nm por lado, agregando moléculas de solvente TIP3P y 0.1 M de iones para equilibrar la carga neta del sistema. Para determinar el número de repeticiones se evaluó el RMSD de cada simulación. Solo se realizaron repeticiones de aquellos compuestos cuyos valores de RMSD se encontraban entre 0-2 Å a lo largo de la simulación.

## 5. Resultados

Con el fin de obtener empaquetamientos alternos de la proteína P, se abordaron 2 metodologías: la utilización de proteínas de fusión y la búsqueda de nuevas condiciones de cristalización. Durante el presente trabajo se evaluaron dos diferentes candidatos como chaperonas de cristalización: un dominio cristizable de la pirrolisil tRNA sintetasa (PylRS) (figura 35, rojo) y la glutation-S-transferasa (GST) (figura 13, verde). En los dos casos, se desarrolló un método de purificación en medio desnaturalizante y se evaluaron distintos medios de replegamiento. Por otro lado, obtuvimos cristales de proteína P silvestre utilizando diferentes agentes precipitantes y se resolvió la estructura de la proteína P en un grupo espacial diferente (figura 13, azul).

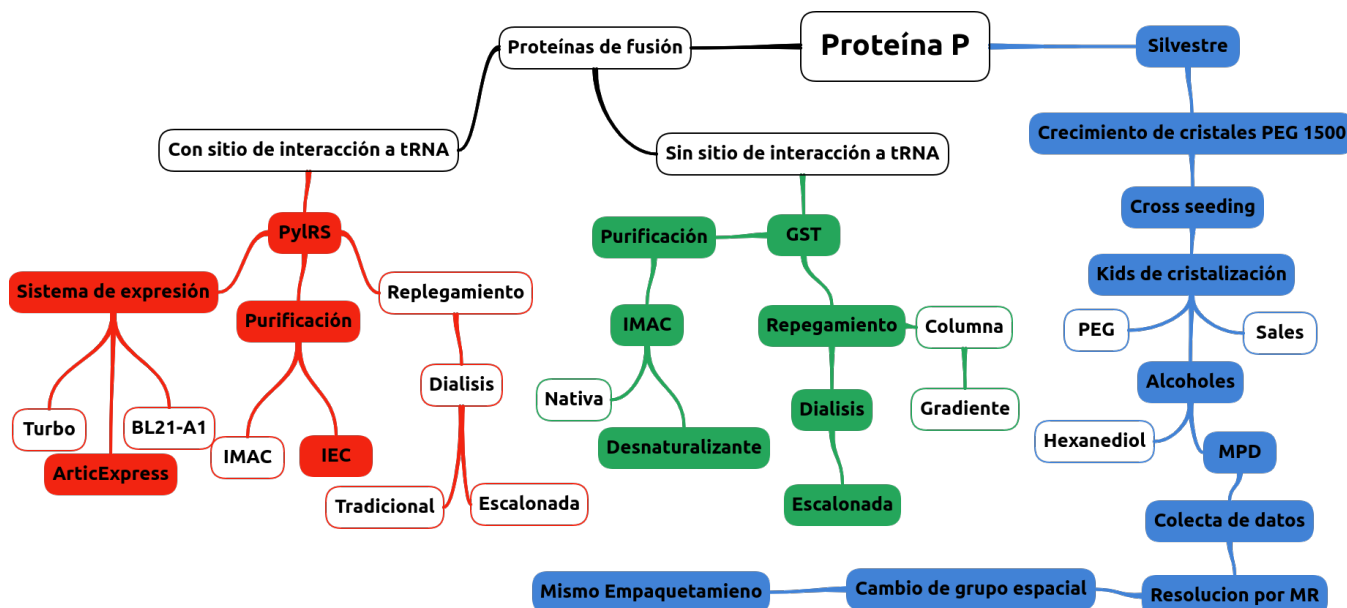


Figura 13: La búsqueda de empaquetamientos alternos de la proteína P se dividió en dos fases. Se optimizó el método de purificación en condiciones desnaturalizantes de la proteína de fusión TmP-PylRS (rojo). Se desarrolló un método para la obtención de la proteína de fusión TmP-GST en condiciones nativas (verde). Se obtuvo la estructura tridimensional de la proteína P silvestre en un diferente grupo espacial (azul).

### 5.1. Proteína de fusión TmP-PylRS

#### 5.1.1. La proteína de fusión TmP-PylRS tiene bajos niveles de expresión

Previamente, Krivenko y colaboradores desarrollaron un método para la expresión y purificación de la proteína P como proteína de fusión con GST. Este método se basa en el uso de condiciones desnaturalizantes para impedir que la proteína P se agregue al unirse inespecíficamente a ácidos nucleicos<sup>30</sup>. El lisado celular desnaturalizado se carga a una columna iónica con carga negativa a pH de 6.5, con lo que la mayoría de las proteínas y los ácidos nucleicos eluyen, uniéndose las proteínas cargadas positivamente. Debido a que se esperaba que la construcción TmP-PylRS tuviera esta misma característica, replicamos este tratamiento con ella. Sin embargo, el rendimiento de esta purificación fue muy bajo, en el orden de microgramos por litro de cultivo.

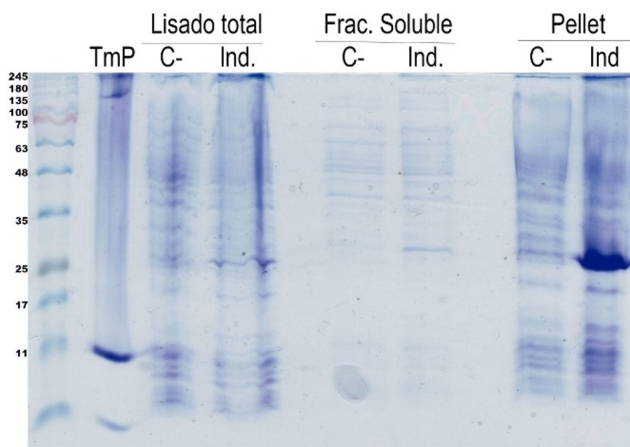
### 5.1.2. Diferentes cepas de *Escherichia coli* expresan la proteína TmP-PylRS de manera diferencial

Con el fin de aumentar el rendimiento de la purificación, realizamos pruebas de expresión utilizando distintas cepas. Al comparar cualitativamente los niveles de expresión de la proteína en la cepa BL21(DE3)pLyss, utilizada inicialmente, no encontramos diferencias visibles en el lisado celular entre el cultivo control (sin inducir) y el cultivo inducido (3mM ramnosa), indicando en efecto una concentración inicial muy baja de la proteína. Para la búsqueda de un sistema de expresión sin este problema, evaluamos seis diferentes cepas de *E. coli* cotrasformadas con la subunidad de RNA de la RNasa P de *Thermotoga maritima* y la proteína de fusión TmP-PylRS. Cuatro presentaron una expresión diferencial al comparar el cultivo control y el cultivo inducido (cuadro 1). Debido a que *a priori* la inducción por ramnosa puede ocurrir en cualquier cepa de *E. Coli*, posiblemente las diferencias observadas radican en inducciones diferenciales de la subunidad de RNA de la RNasaP.

Diferencia entre el nivel de expresión en ausencia y presencia del inductor de diferentes cepas de <i>E.coli</i>		
Cepa	Inducción idistinguible	Inducción distinguible
BL21(DE3)pLyss	X	
BL21(DE3)	X	
BL21-A1		X
Turbo		X
NiCo		X
ArticExpress		X

**Cuadro 1:** Se encontraron cuatro cepas con una inducción relativamente clara de la proteína de fusión TmP-PylRS. Resultado de experimentos de expresión comparando el cultivo control (sin inducir) y el cultivo inducido (IPTG+ramnosa). Recopilación de cuatro experimentos de expresión independientes y evaluados por SDS-PAGE.

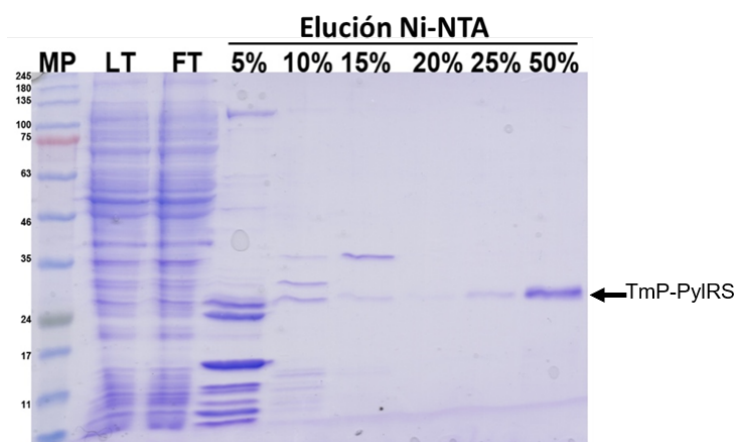
La siguiente etapa consistió en determinar la fracción del lisado celular (soluble o precipitada) en que se encontraba la proteína de interés. Para esta evaluación se seleccionaron las cepas ArticExpress, BL21-A1 y Turbo. La cepas Turbo y ArticExpress contienen una pequeña fracción soluble de una banda en el gel con un peso molecular que corresponde con la proteína de fusión, con el resto de esa banda en la parte insoluble. Comparando cualitativamente estas dos cepas, ArticExpress presentó mayor cantidad de proteína de fusión en la parte insoluble (figura 14).



**Figura 14:** La mayor cantidad de la proteína de fusión TmP-PylRS se encontró en la fracción insoluble de la cepa ArticExpress. Gel SDS-PAGE 15%. Expresión de una proteína con el peso molecular correspondiente a la fusión TmP-PylRS (26.8 KDa), en el lisado total, la fracción soluble y el pellet de un control sin inducir (C-) y la inducción de la proteína(Ind) en la cepa de *E. coli* ArticExpress.

### 5.1.3. La proteína TmP-PylRS se puede obtener pura y soluble, pero con bajo rendimiento

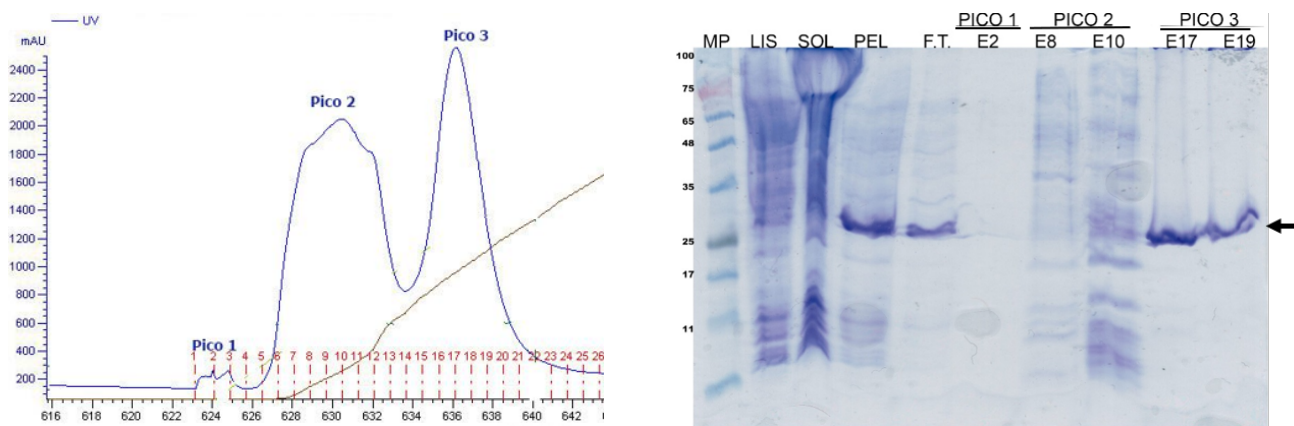
Para el primer esquema de purificación seleccionamos la cepa Turbo por presentar en apariencia mayor cantidad de la proteína en la fracción soluble. La proteína contenida en esta fracción se cargó en una columna de afinidad a iones metálicos en condiciones desnaturizantes. El principal problema a resolver con este esquema de purificación fue el grado de pureza con que eluye la proteína. Se evaluaron tres variantes para la elución en un amortiguador con 2M de imidazol: elución por golpe de imidazol, gradiente de 0-2M y gradiente escalonado (intervalos de 5% de amortiguador B). Las fracciones recuperadas de los 2 primeros métodos contenían la proteína de fusión y proteínas contaminantes. En cambio, con el último método, fue posible separar la proteína de los contaminantes. La fracciones que contenían a la proteína se encontraron en un intervalo 20-50% de amortiguador B (figura 15). El rendimiento de esta purificación fue de 200  $\mu\text{g/L}$  de cultivo, que en su momento consideramos insuficiente para realizar pruebas de cristalización. Aunado a ello, se tuvieron problemas en la reproducibilidad del rendimiento, teniendo en la mayor parte de los lotes mucho menos proteína que la obtenida inicialmente, por lo que asumimos que se estaba trabajando con condiciones de cultivo y de amortiguador al margen de lo que permitía obtener una proteína soluble y optamos por utilizar un esquema de purificación en condiciones desnaturizantes.



**Figura 15:** Es posible obtener pura y soluble a la proteína de fusión TmP-PylRS por afinidad a Ni-NTA y elución por gradiente escalonado. Gel SDS-PAGE 15%. Muestra los pasos de purificación de la proteína de fusión TmP-PylRS en una columna de afinidad Ni-NTA. De izquierda a derecha se muestra el marcador de pesos (MP), lisado total (LT), fracción sin afinidad por la columna (FT) y las fracciones obtenidas de la elución por gradiente escalonado (5-50% de amortiguador B).

### 5.1.4. La proteína TmP-PylRS se puede obtener pura y con alto rendimiento, pero desnaturizada

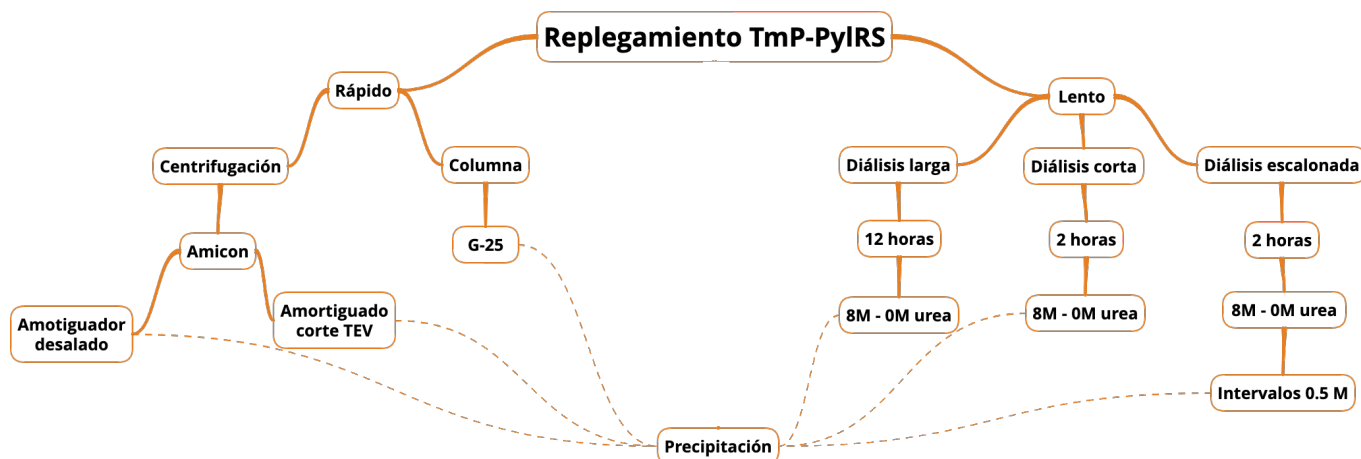
Diseñamos un esquema de purificación con base en que la cepa ArticExpress expresaba visualmente mayor cantidad de proteína en la fracción insoluble y a que los componentes de la proteína de fusión tenían afinidad por ácidos nucleicos. Así, recuperamos la proteína contenida en la fracción insoluble en un medio desnaturizante con 8M de urea. La purificación se realizó por cromatografía de intercambio catiónico, con un gradiente de 0.1-3M de NaCl. Las fracciones correspondientes al tercer pico de elución (figura 18, izquierda) contenían una banda con un peso molecular que corresponde a la proteína de fusión sin contaminantes (figura 16). Con este esquema aumentamos el rendimiento de la purificación a 7 mg/L de cultivo. Buscamos condiciones que nos permitieran trasladar la mayor cantidad de proteína posible a la fracción soluble. Construimos una batería de amortiguadores de lisis utilizando dos valores de pH (7.5 y 8.0) y 10 diferentes aditivos (Anexo 1). No encontramos ninguna condición que solubilizara a la proteína, por lo que procedimos a realizar pruebas de replegado con la proteína purificada, pero desnaturizada.



**Figura 16:** La purificación por intercambio aniónico en condiciones desnaturalizantes aumenta significativamente el rendimiento de la purificación de la proteína TmP-PylRS. Derecha, cromatograma de la purificación utilizando una columna SP Sepharosa. Se obtuvieron 3 picos y la proteína se encuentra pura en el pico 3 (fracciones 15-21). Izquierda, SDS-PAGE 15 %, mostrando el seguimiento de la purificación de la proteína de fusión TmP-PylRS. TmP (14.3 kDa) + PylRS (12.5 kDa) = 26.8 kDa. De izquierda a derecha se muestra: marcador de peso (MP), lisado total (LIS), fracción soluble (SOL), pastilla insoluble=pellet (PEL), fracción sin afinidad por la columna (FT), pico 1: fracción 2(E2), pico 2: fracción 8(E8) y fracción 10(E10), pico 3: fracción 17 (E17) y fracción 19 (E19).

### 5.1.5. No se logró la renaturalización de la proteína de fusión TmP-PylRS

A partir de las fracciones de la purificación por intercambio aniónico, seleccionamos la que contaba con mayor concentración de la proteína de fusión. Para cada ensayo de replegamiento utilizamos 3 mg totales de la proteína, que se transfirieron a un amortiguador sin urea y 1M de NaCl. Se exploraron 2 alternativas, cada una cambiando abruptamente las condiciones del medio: 1) desalado por columna y 2) cambio de amortiguador por centrifugación (Amicon). En ambos casos, se perdió el total de la muestra por precipitación. Intentamos evitar la precipitación de la proteína alargando el tiempo en el proceso de replegamiento. Exploramos 3 variantes de replegamiento por diálisis: larga, corta y escalonada. A pesar del cambio paulatino en las condiciones del amortiguador, no fue posible recuperar proteína en la fracción soluble (figura 17).

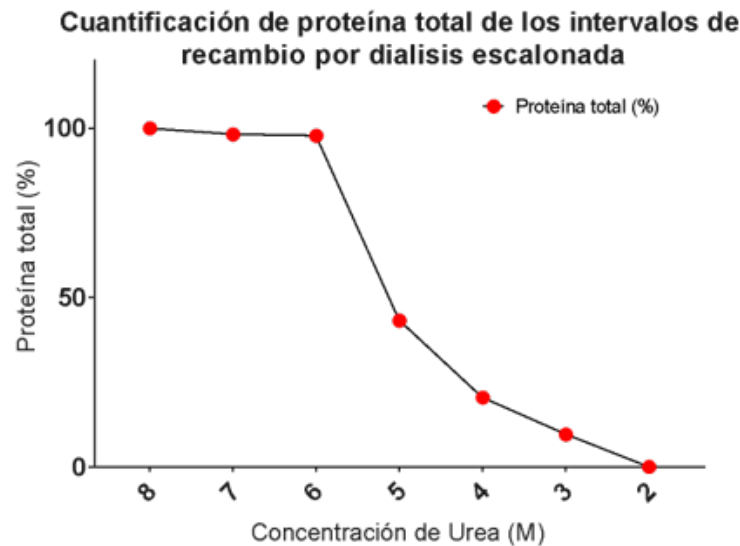


**Figura 17:** No fue posible obtener a la proteína TmP-PylRS soluble por los métodos de replegamiento que exploramos. Resumen de las diferentes estrategias para la eliminación de las condiciones desnaturalizantes.



A manera de ejemplo, por medio de diálisis escalonada, se diluyó la concentración de urea en intervalos de 0.5M cada 2 horas. En cada recambio de amortiguador se calculó el total de proteína soluble. El seguimiento del proceso de desalado (figura 18) indicó que:

- En un intervalo entre 6-8 M de urea, la proteína se mantiene soluble.
- En concentraciones <5M de urea, precipita >50 % de la proteína.
- A 3M urea, 10 % de la proteína se mantiene soluble.

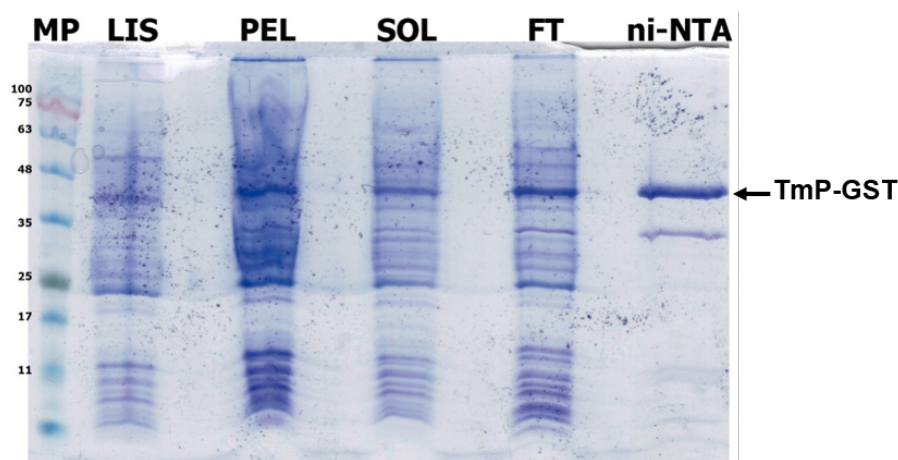


**Figura 18:** La proteína de fusión TmP-PylRS se precipita en concentraciones menores a 6M de urea. Seguimiento de la cantidad total de proteína en intervalos de 1M expresada en porcentaje del proceso de diálisis escalonada. El experimento se realizó con 3 ml con una concentración de 2.2 mg/ml de proteína (6.8 mg de proteína total).

## 5.2. Proteína de fusión TmP-GST

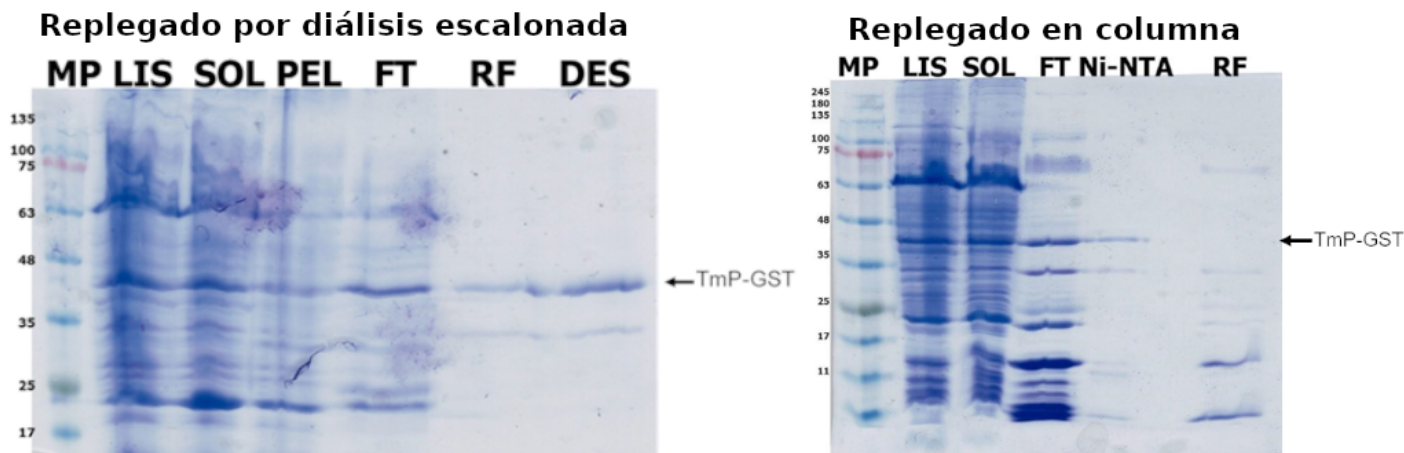
### 5.2.1. No se logró la renaturalización de la proteína de fusión TmP-GST

Para su purificación, la proteína P es expresada como proteína de fusión con GST. Se utilizan agentes caotrópicos para prevenir la agregación, posteriormente la proteína de fusión es cortada en el lisado celular y purificada en condiciones desnaturizantes en una columna de afinidad a iones metálicos. Se intentó este mismo esquema de purificación, pero sin cortar la proteína de fusión. La fracción colectada de la purificación contenía una banda con un peso correspondiente al de la proteína de fusión, así como otra banda de poco menos de 35kDa (figura 19).



**Figura 19: Es posible purificar la proteína de fusión TmP-GST en condiciones desnaturalizantes.** SDS- PAGE 15% que muestra el seguimiento de la purificación de la proteína de fusión de la proteína P (14.3 KDa) y la proteína GST (27.2kDa) como chaperona de cristalización. De izquierda a derecha: Marcador de peso (MP), lisado total (LIS), Pastilla=pellet (PEL), fracción soluble (SOL), fracción sin afinidad por la columna(FT), elución Ni-NTA (ni-NTA).

Para eliminar las condiciones desnaturalizantes, intercambiamos el amortiguador de la proteína por medio de diálisis escalonada. Para tratar de evitar que la proteína se precipitara, el amortiguador en cada recambio contenía 1M de sal monovalente. Posteriormente disminuimos la concentración de sal monovalente a 50 mM. Al término de este proceso, obtuvimos la proteína de fusión soluble (figura 20, izquierda), aunque en un rendimiento muy bajo. Intentamos también inmovilizar a la proteína desnaturalizada en una columna de Ni-NTA. Eliminamos lentamente las condiciones desnaturalizantes por medio de un gradiente de 16 horas. Finalmente, despegamos a la proteína de la columna y evaluamos la fracción colectada. Dicha fracción no contenía la proteína de fusión (figura 20, derecha) lo que señala, como para el caso de la fusión TmP-GST, que posiblemente nos encontramos en el margen de condiciones de solubilidad y/o estabilidad de la proteína.



**Figura 20: Es posible obtener bajas cantidades de la proteína de fusión TmP-GST en condiciones óptimas para su cristalización.** Derecha. Seguimiento de la purificación, replegamiento y desalado de la proteína de fusión de la proteína P (14.3 KDa) y la proteína GST (27.2kDa). Izquierda, seguimiento de la purificación a través de replegamiento por columna. Marcador de peso (MP), lisado total (LIS), pellet (PEL), fracción soluble (SOL), fracción sin afinidad por la columna (FT), elución Ni-NTA (Ni-NTA), replegado (RF), desalado (DES) y replegado (RF).

### 5.3. Proteína P silvestre

Como alternativa al trabajo con proteínas de fusión, buscamos nuevas condiciones de cristalización para la proteína P. Los primeros ensayos de cristalización se realizaron por el método de difusión de vapor en su modalidad de gota sedente, con distintos kits de la marca Hampton Research: Crystal Screen, Index, Natrix, PEG Rx, y Salt Rx. De los cristales obtenidos, no encontramos condiciones diferentes a las antes reportadas<sup>33</sup>, las cuales contienen PEG entre 400-6000. A partir de las condiciones en donde encontramos cristales de mayor tamaño, titulamos la concentración de precipitante (PEG 400, 550, 1000, 1500, 3350 y 6000), la concentración de sal (sulfato de potasio) y el pH del amortiguador (acetato de sodio trihidratado) (cuadro 2). Determinamos la condición de cristalización que inhibe el crecimiento de agujas y favoreció el crecimiento de placas de mayor tamaño (figura 21).

Componente	Intervalo	Condición optimizada
Acetato de sodio trihidratado	pH 4.2 - 5.8	5.0
Sulfato de potasio	0.2M - 0.35M	0.2M
PEG (550, 1000, 1500, 3550)	10 % - 18 %	18 %(1500)

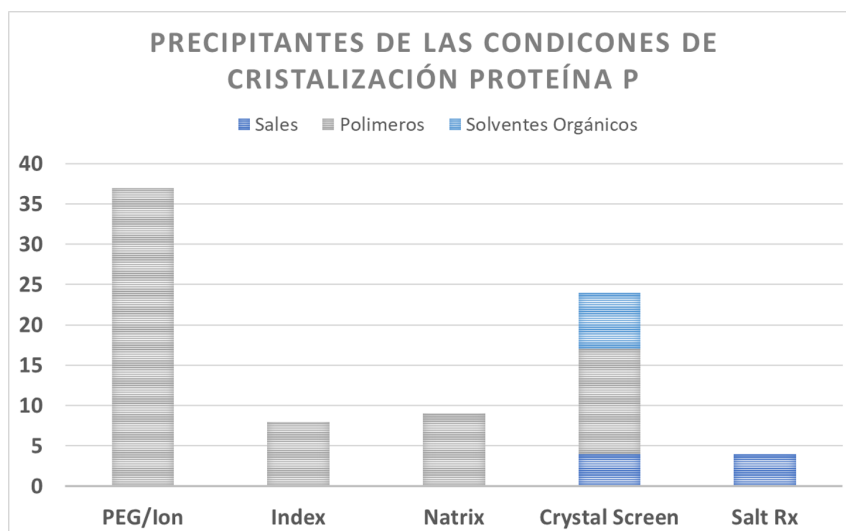
**Cuadro 2:** Intervalos de titulación de los componentes de la solución madre para la cristalización de la proteína P silvestre



**Figura 21:** La condición de cristalización optimizada favorece la formación de placas de mayor tamaño. Cristales obtenidos en la condición optimizada (100 mM de acetato de sodio trihidratado pH 5.0, 200 mM Sulfato de potasio, 18 % PEG 1500)

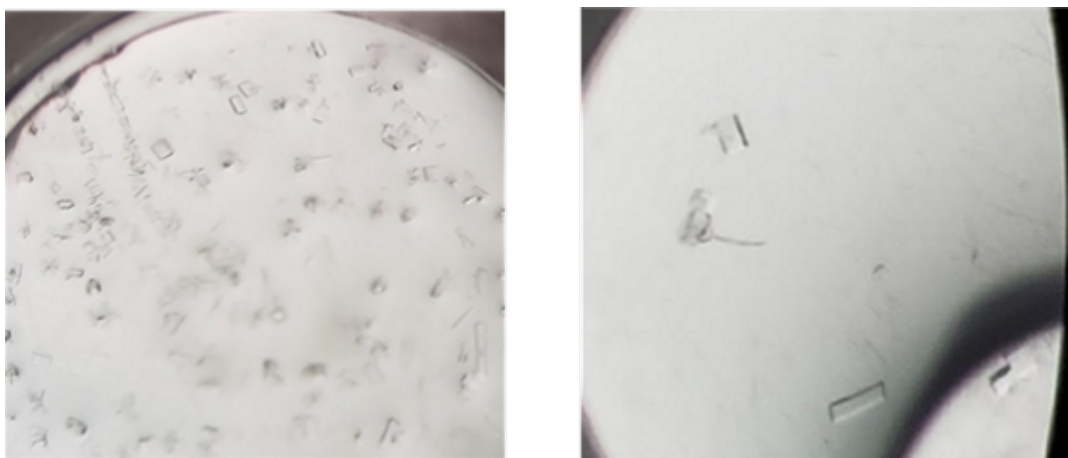
#### 5.3.1. La proteína P cristaliza en nuevas condiciones utilizando la técnica de sembrado cruzado

Colectamos cristales de la proteína P provenientes de la condición optimizada y los utilizamos como “semillas” para la búsqueda de nuevos cristales utilizando de nuevo los distintos kits disponibles y logramos obtener una gran variedad de cristales en esta ocasión (figura 22). Agrupamos las diferentes condiciones en tres grupos de acuerdo al agente precipitante: sales (10 agentes precipitantes), solventes orgánicos (4 agentes precipitantes) y polímeros (7 agentes precipitantes) (Anexo 2).



**Figura 22:** Se encontraron 21 condiciones nuevas de cristalización con un diferente agente precipitante. Muestra de manera gráfica el número de condiciones de cristalización de los diferentes kits. En azul se resalta las condiciones cuyo agente precipitante es una sal o un solvente orgánico.

Descartamos todas las condiciones cuyo agente precipitante fuera PEG ya que la morfología era muy parecida a la de los cristales semilla, que también crecen en PEG. De las condiciones restantes seleccionamos los cristales con morfología diferente a los cristales semilla y de mayor tamaño. Para la colecta de datos se seleccionaron cristales en condiciones de MPD (figura 23, izquierda) y Hexanediol (figura 23, derecha) como precipitante.



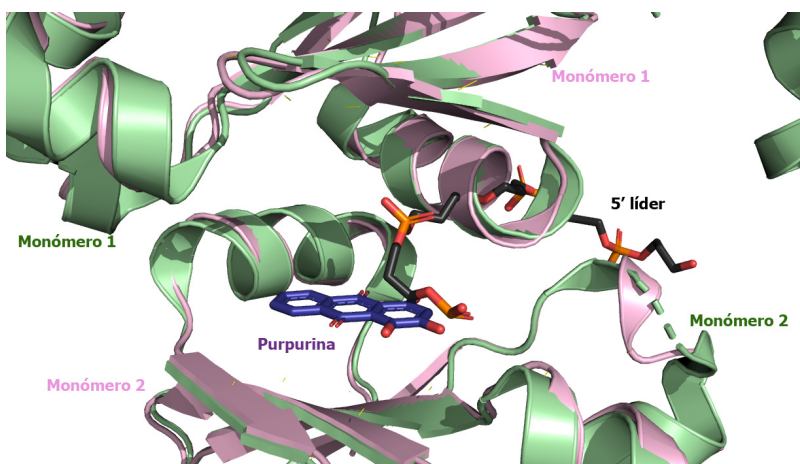
**Figura 23:** Para la colecta de datos se seleccionaron los cristales con morfología diferente a los cristales semilla. Izquierda, se muestran los cristales de la proteína P, obtenidos en las condición 0.02 M cloruro de calcio dihidratado, 0.1 M acetato de sodio trihidratado pH 4.6, 30 % (v/v) (+/-)-2-Methyl-2,4-pentanediol. Derecha, se muestran los cristales de la proteína P obtenidos en las condición 0.01 M cloruro de cobalto(II) hexahidratado, 0.1 M acetato de sodio trihidratado pH 4.6, 1.0 M 1,6-Hexanediol

Solo fue posible obtener datos suficientes para resolver la estructura de la condición de MPD pues el patrón de difracción de los cristales de hexanediol no pudo indexarse por presentarse demasiadas reflexiones “barridas” (smeared, no definidas). La estructura fue resuelta por reemplazo molecular a una resolución de 3.0 Å (cuadro 3). El cambio de precipitante cambió el grupo espacial a ortorrómbico y el número de moléculas en la unidad asimétrica se redujo de 4 a 2.

	TmP-MPD	TmP-PEG(1NZ0)
<b>Colecta de datos</b>		
Resolución(Å)	58.22(3.19)	30(1.2)
Grupo espacial	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub>
<b>Celda unitaria</b>		
a, b, c (Å)	46.7, 56.2, 103.8	56.2, 64.1, 68.3
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 102, 90
<b>Colecta de datos</b>		
Reflexiones totales	43319(6069)	774818 (67203)
Reflexiones únicas	5824(919)	118678 (16,982)
Rmerge	0.08(0.22)	0.06 (0.56)
Redundancia	7.4(6.6)	6.5 (4.0)
Integridad (%)	99.8(100)	80.0 (38.7)
I/ $\sigma$	16.5(8.0)	5.3 (1.1)
Factor B de Wilson (Å <sup>2</sup> )	42.1	13.6
<b>Afinamiento</b>		
Rwork	0.25 (0.19)	0.21
Rfree	0.39 (0.42)	0.23
R.M.S.D longitud de enlace (Å)	0.015	0.01
R.M.S.D angulos (°)	0.221	2.92

**Cuadro 3: El cambio de precipitante genera un cambio en el grupo espacial.** Comparación de la colecta y procesamiento de datos del modelo obtenido (TmP-MPD) contra el modelo reportado (código PDB 1NZ0)

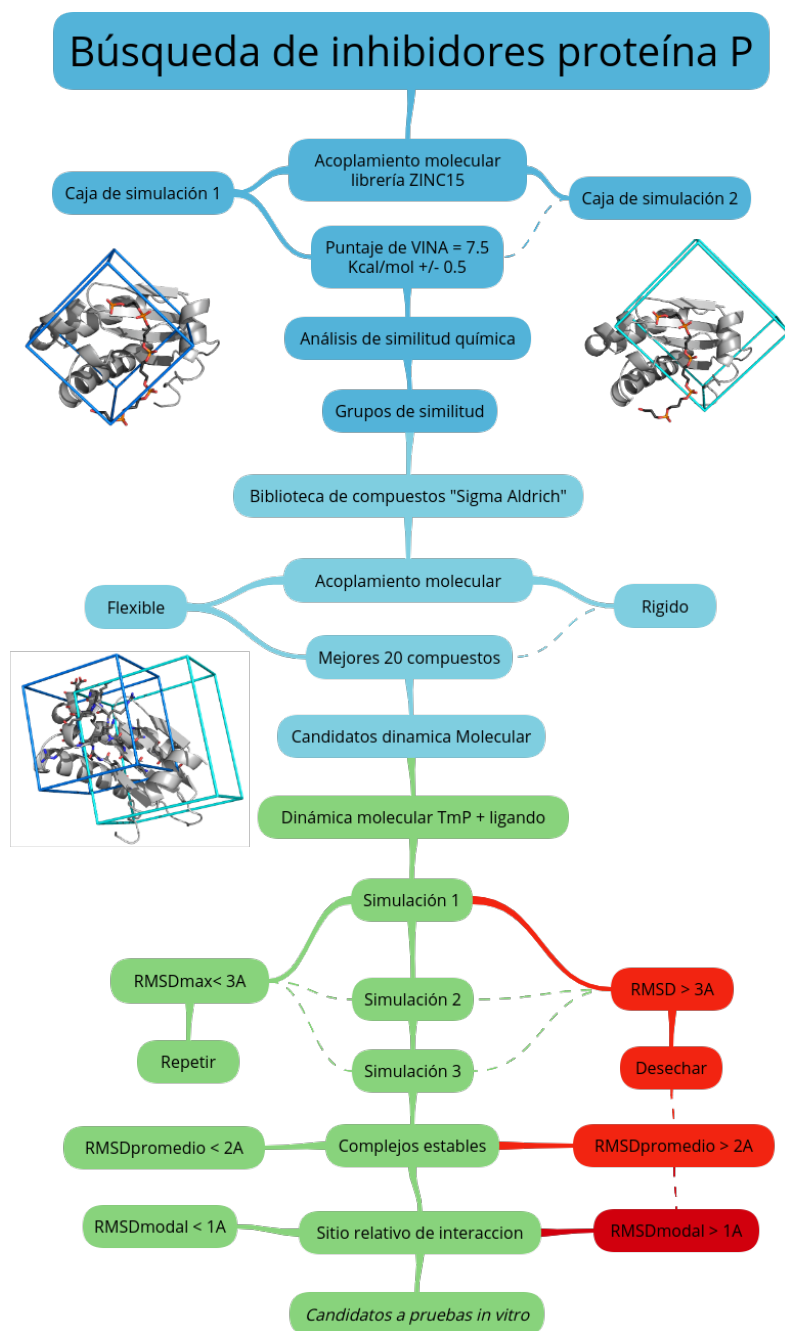
El modelo obtenido fue sometido a una inspección visual en el programa Coot. Para corroborar el empaquetamiento se generó la malla cristalina y se sobrepuso el modelo de la holoenzima (código PDB 3Q1R). Aunque el cambio del agente precipitante generó cambios en la unidad asimétrica y el grupo espacial, no modificó el ordenamiento intrínseco de la proteína en el cristal (figura 24).



**Figura 24: El cambio en el grupo espacial de la proteína no modifica el empaquetamiento cristalino.** En verde se observa el modelo obtenido de los cristales con PEG 1500 como agente precipitante (triclínico) y en rosa el nuevo modelo obtenido de los cristales con MPD como agente precipitante (ortorrómbico). Se muestra en negro la posición del 5'-líder y en morado el compuesto purpurina.

## 5.4. Búsqueda de inhibidores a través de métodos *in silico*

Debido a que no fue posible trascender el problema del empaquetamiento cristalino y a que la purificación de las proteínas de fusión tuvo un éxito limitado, buscamos rediseñar la plataforma ya desarrollada con anterioridad<sup>19</sup>, ahora iniciando la búsqueda por métodos *in silico* y terminando con el análisis *in vitro* (figura 25). Partimos de una biblioteca virtual de más de 700,000 compuesto y reducimos esta lista a 10 candidatos comercialmente disponibles.

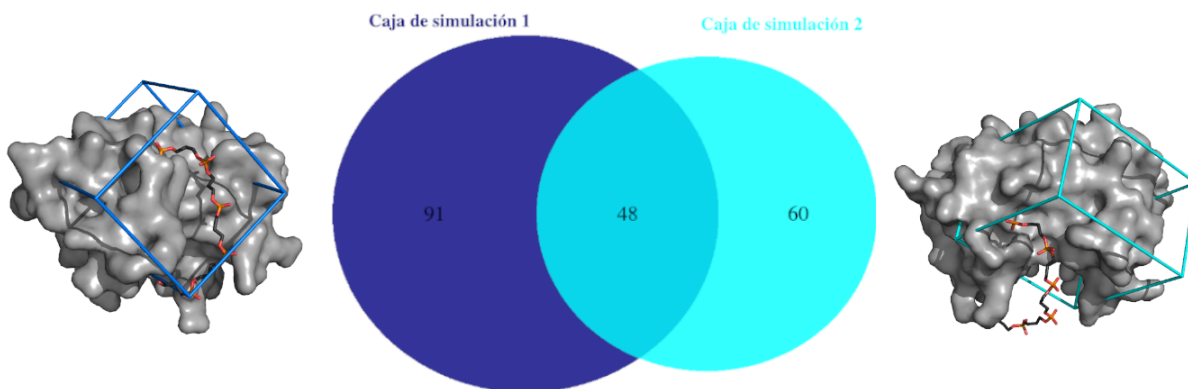


**Figura 25:** Los candidatos para pruebas *in vitro* fueron obtenidos a través de tres etapas de selección. Los 700,000 compuestos iniciales fueron evaluados por acoplamiento molecular en dos cajas de simulación diferentes (azul). Agrupamos los mejores 200 compuestos por su similitud química; a partir de estos grupos se construyó una biblioteca de compuestos comercialmente disponibles. Esta nueva biblioteca fue evaluada por acoplamiento molecular rígido y flexible (cian). Seleccionamos los mejores 20 compuestos para simulaciones de dinámica molecular. A partir del valor de RMSD seleccionamos los mejores candidatos a pruebas *in vitro* (verde) y descartamos el resto (rojo).

#### 5.4.1. El acoplamiento molecular de 700,000 compuestos resulta en 199 potenciales compuestos que interactúan con la proteína P en regiones relevantes

Previo a los ensayos de acoplamiento molecular, se redujo el total de compuestos disponibles en la biblioteca ZINC15 de acuerdo con su peso molecular (200-400Da), carga (neutra), coeficiente de partición ( $-0.4 < \text{Log } P < 5.6$ ) y su disponibilidad comercial. Con estos primeros condicionales, construimos un set de 700,000 que denominamos “biblioteca ZINC15”. Para esta biblioteca efectuamos ensayos de acoplamiento molecular en 2 cajas diferentes (véase material y métodos). De cada ensayo seleccionamos los compuestos con un puntaje igual a  $-7.0$  kJ/mol. La primera caja de simulación consideraba el surco de interacción del 5'-líder (figura 28, azul) y de ella obtuvimos 139 compuestos. La segunda caja consideraba la única cavidad de la proteína P y parte del surco de interacción del 5'-líder (figura 28, cian); de esta caja se obtuvieron 108 compuestos, de los cuales, 48 eran compartidos con la caja de simulación 1 (figura 26, diagrama de Venn). Los 199 ( $108+139-48$ ) compuestos obtenidos fueron agrupados por su similitud química en 48 conjuntos. De cada conjunto se seleccionó un elemento representativo para la búsqueda de moléculas comercialmente disponibles. Con este proceso se construyó una biblioteca de 599 compuestos que son distribuidos por la compañía Sigma Aldrich.

#### Compuestos provenientes del acoplamiento molecular de la biblioteca ZINC15

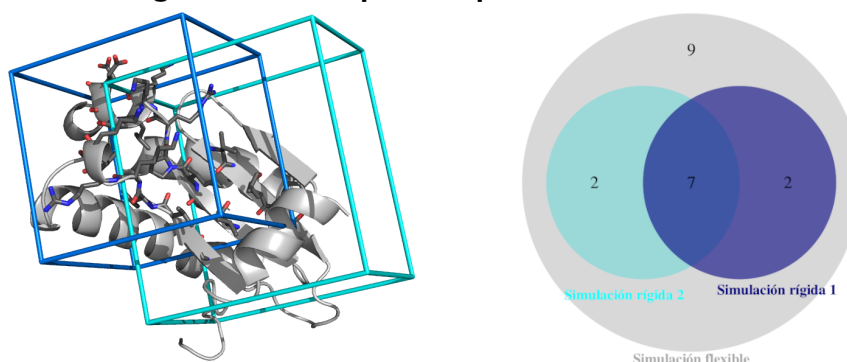


**Figura 26:** De los ensayos de acoplamiento molecular con los compuestos de la biblioteca ZINC15, seleccionamos 199 compuestos. Se realizaron ensayos en dos cajas de simulación diferentes; la primera consideró el surco de interacción del 5'-líder (azul) y la segunda consideró la única cavidad en la proteína P (cian). De los dos ensayos se seleccionaron 248 compuestos con una puntuación igual a  $-7.0$  Kcal/mol; 48 de estos compuestos son compartidos en ambas cajas (diagrama de Venn)

#### 5.4.2. La selección por acoplamiento molecular de los candidatos disponibles comercialmente resulta en 20 compuestos

Estos 599 compuestos fueron evaluados por acoplamiento molecular en las cajas de simulación previamente descritas. Adicionalmente, en la caja de simulación 1 se realizó un ensayo de acoplamiento molecular flexible, en donde se permitía el acomodo de distintos aminoácidos pertenecientes al surco de interacción del 5'-líder (figura 27, derecha). Se seleccionaron los 20 compuestos con menor puntaje de Autodock Vina como candidatos a simulación por dinámica molecular. Estos 20 compuestos estaban incluidos en los mejores del ensayo flexible y 7 en los mejores 20 de los 3 ensayos (figura 27, izquierda).

## Selección de los mejores 20 compuestos de la biblioteca "Sigma Aldrich" por acoplamiento molecular

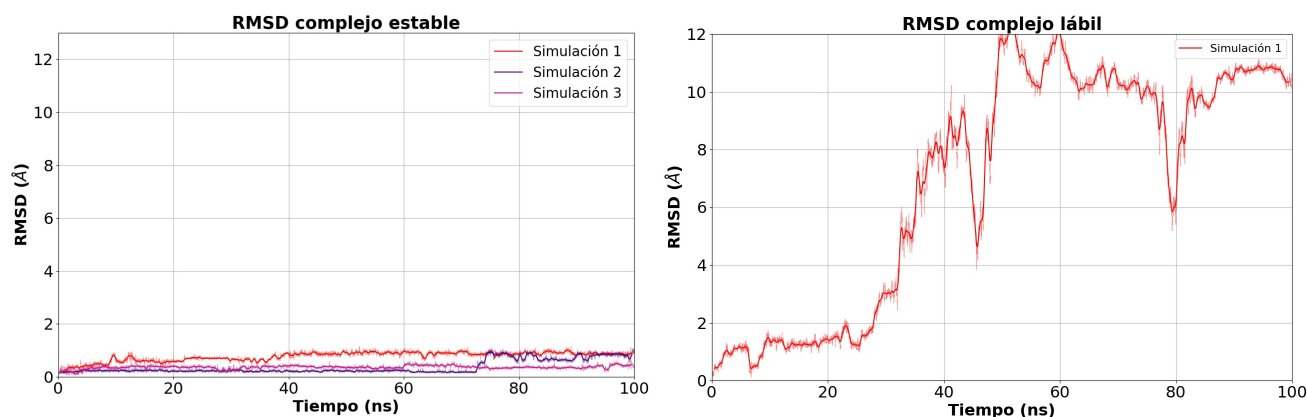


**Figura 27:** De los 3 ensayos independientes de acoplamiento molecular se seleccionaron 20 candidatos para simulación por dinámica molecular. Izquierda, muestra las 2 distintas cajas de simulación (azul y cian); en negro los aminoácidos seleccionados para el acoplamiento molecular flexible. Derecha, muestra la distribución de los compuestos con el menor puntaje de Autodock Vina.

### 5.4.3. Simulaciones de dinámica molecular permiten reducir a cinco los candidatos para pruebas *in vitro*

Para tratar de evitar falsos positivos, sometimos estas 20 moléculas a simulaciones de dinámica molecular. Para cada complejo de la proteína P con los diferentes ligandos se calculó el RMSD de la distancia entre los centros de masa de la proteína y el ligando. A través de este cálculo definimos 3 parámetros para la selección de los candidatos a pruebas *in vitro* (anexo 3):

1) Variación máxima del RMSD ( $\text{RMSD}_{\text{max}}$ ). Con este valor, definimos si el complejo proteína-ligando era estable ( $\text{RMSD}_{\text{max}} < 3\text{\AA}$ ) o no ( $\text{RMSD}_{\text{max}} > 3\text{\AA}$ ). Aquellos compuestos que formaban complejos lábiles fueron descartados (figura 28, derecha). Los complejos estables fueron corroborados con dos repeticiones independientes (figura 28, izquierda).

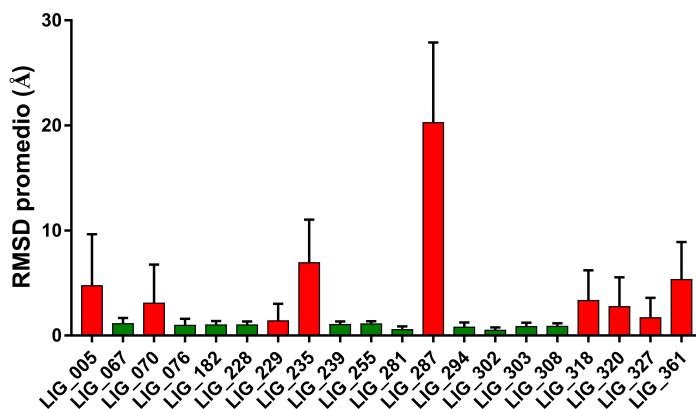


**Figura 28:** Los compuestos simulados pueden ser catalogados en 2 grupos de acuerdo con su  $\text{RMSD}_{\text{max}}$ . Izquierda, muestra el comportamiento típico de un complejo estable ( $\text{RMSD}_{\text{max}} < 3\text{\AA}$ ). Derecha, muestra el comportamiento típico de un complejo lábil ( $\text{RMSD}_{\text{max}} > 3\text{\AA}$ ).

2) Valor promedio de RMSD en las tres simulaciones ( $\text{RMSD}_{\text{prom}}$ ). Con este valor definimos si el comportamiento de cada complejo era constante y se redujo el límite para definir un complejo estable ( $\text{RMSD}_{\text{prom}} < 2\text{\AA}$ ). De la selección con los primeros dos filtros, descartamos 9 de los 20 compuestos iniciales (figura 29).



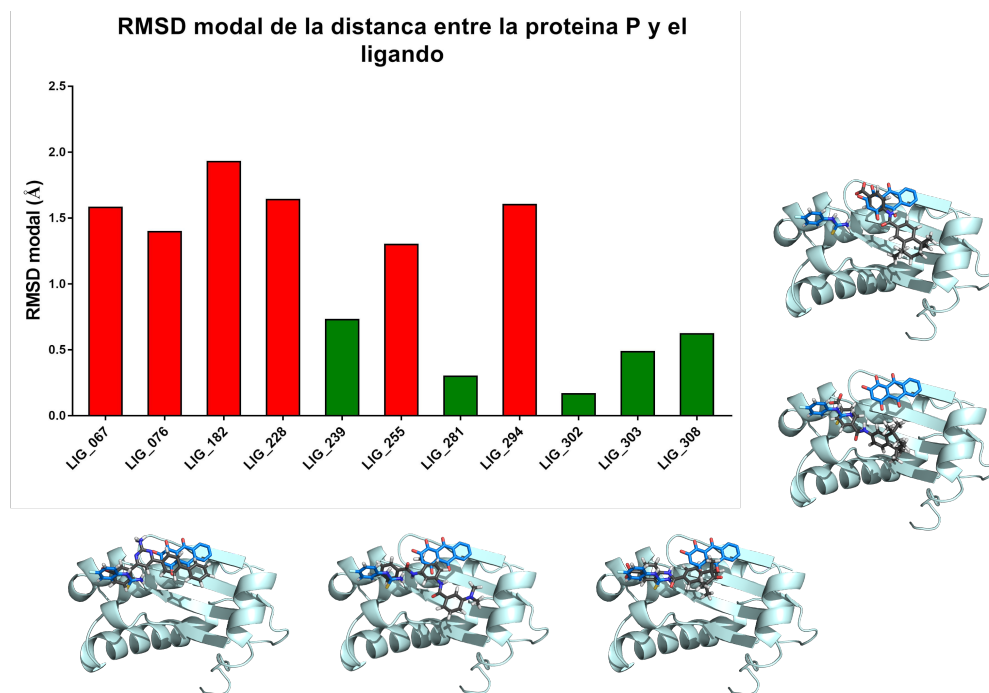
### RMSD promedio de la distancia entre proteína P y el ligando



**Figura 29:** Se seleccionaron 11 compuestos que forman complejos estables con la proteína P. Se graficó el promedio del RMSD para cada complejo proteína-ligando junto con su desviación estándar. Se indica en rojo los compuestos que forman complejos lábiles y en verde los complejos estables.

El análisis  $RMSD_{promedio}$  nos ayudó a determinar si un ligando se encontraba en contacto con la proteína P. Sin embargo, este parámetro no es útil para definir el sitio de interacción de la proteína y el ligando, por lo que para cada grupo de datos se calculó:

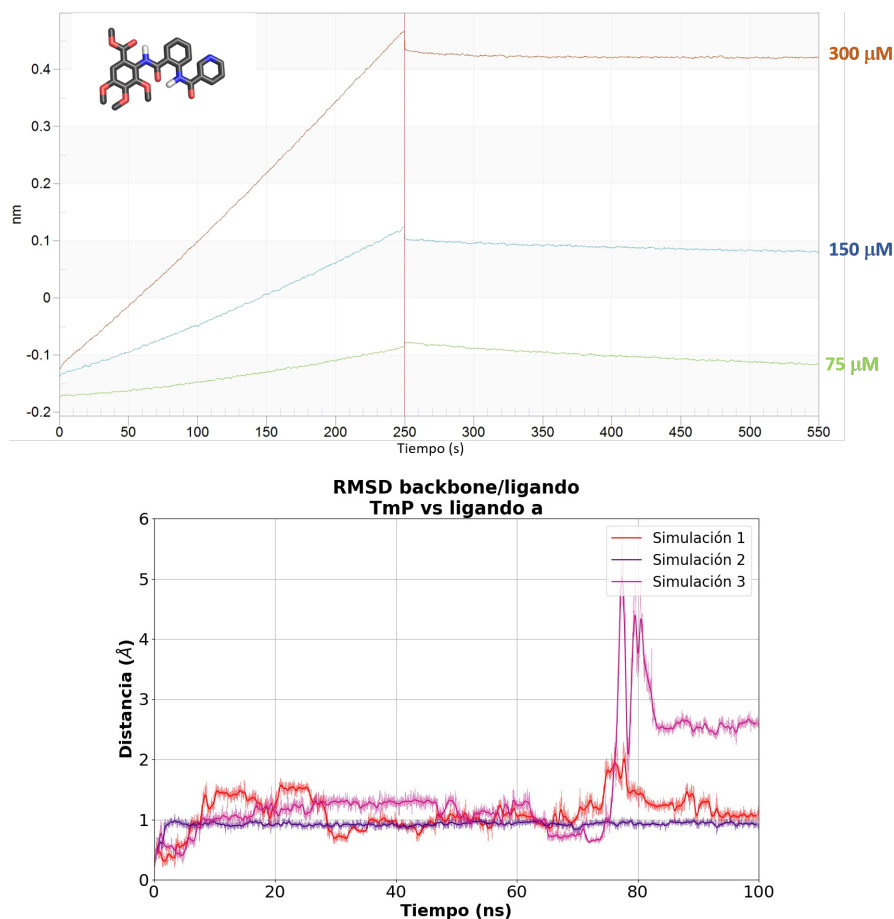
3) Valor de RMSD más frecuente ( $RMSD_{modal}$ ) y asociamos este valor al área de contacto entre la proteína y el ligando (figura 30). Por último, emparentamos el  $RMSD_{modal}$  con la posición del ligando dentro de la simulación. Para puntualizar la búsqueda, descartamos aquellos compuestos con un área de contacto amplia ( $RMSD_{modal} > 1 \text{ \AA}$ ) (anexo 4).



**Figura 30:** El  $RMSD_{modal}$  nos ayudó a puntualizar la búsqueda a 5 candidatos para pruebas *in vitro*. Valor de RMSD más frecuente durante las 3 simulaciones de los ligandos que forman complejos estables con la proteína P. Este valor fue asociado a que la posición del ligando era sobre la proteína P (cian), utilizando como referencia los compuestos purpurina y FTU (azul). En rojo se muestran los compuestos descartados por tener una amplia área de contacto ( $RMSD_{modal} > 1 \text{ \AA}$ ) y en verde los compuesto seleccionados para pruebas *in vitro*.

#### 5.4.4. Ensayos control indican que los resultados del protocolo *in silico* deben tomarse con cautela

Por razones presupuestarias y logísticas, no fue posible adquirir ninguno de los 5 candidatos propuestos por el protocolo *in silico*. Sin embargo, gracias a la colaboración del Instituto de Química de la UNAM (a través de la Dra. Adela Rodríguez Romero), obtuvimos una molécula similar a los compuestos 201 y 239 de análisis *in silico* (véase anexo 4). Este compuesto es un derivado del ácido antralínico con un peso mayor a 250 Da, que denominamos compuesto A. Con este compuesto probamos la interacción con la proteína P, titulando la concentración del compuesto desde 300  $\mu\text{M}$  a 75  $\mu\text{M}$ . Si bien se observó un cierto comportamiento dosis-respuesta (figura 35, izquierda), la curva de asociación no llega a saturarse, lo que impide calcular la afinidad de la proteína por el compuesto, de cualquier manera se estima que estaría en el rango de micromolar alto. Dado este resultado realizamos las pruebas *in silico* de este compuesto y lo contrastamos con nuestros candidatos. En cuanto al puntaje obtenido por acoplamiento molecular, el compuesto A debería ser descartado como candidato a simulación por dinámica molecular (puntaje de Vina  $< -7.0$  KJ/mol). Sin embargo, al realizar la simulación, este compuesto se comporta como un candidato a pruebas *in vitro* (figura 31, derecha). Este resultado pone de manifiesto que el protocolo de discriminación *in silico* debe ajustarse para las simulaciones por dinámica molecular a condiciones más astringentes, como mayor cantidad de sal monovalente o tal vez mayor tiempo de simulación, siendo el fin último que representen un “espejo” de las mediciones de afinidad experimentales.



**Figura 31:** El compuesto A se une de manera dosis-dependiente a la proteína P. Izquierda, muestra el sensograma de las curvas de titulación del compuesto A con un de concentración de 75  $\mu\text{M}$  – 300  $\mu\text{M}$ . La concentración de la proteína P con colas de histidinas se mantuvo en 0.05 mg/mL (3.6  $\mu\text{M}$ ). Derecha, muestra la variación de RMSD de la distancia entre los centros de masa de la proteína P y el compuesto A.

#### 5.4.5. El conjunto de datos obtenidos indica que es necesario afinar las distintas estrategias exploradas para mejorar la búsqueda de inhibidores.

El conjunto de los resultados obtenidos podrían indicar que, para este sistema en particular, la búsqueda de inhibidores a través de cristalografía de rayos X pareciera ser incompatible con la implicación de tener que trascender el empaquetamiento cristalino. Es desafortunado que los contactos entre moléculas coincidan con la zona donde interacciona el sustrato, independientemente del contexto del solvente (uso de distintos agentes precipitantes). La estrategia de utilizar proteínas de fusión parece también demandar la búsqueda de más candidatos con diferentes longitudes de “linkers” o purificados por separado y luego cocristalizados, así como esquemas de purificación relativamente complejos. Por otro lado, la estrategia *in silico* debe ser afinada también para este sistema, al menos en lo que respecta a las simulaciones de dinámica molecular. Una alternativa a estas aproximaciones para mapear los sitios de unión de manera indirecta podría ser el uso de mutagénesis dirigida en combinación con interferometría de biocapas. Previamente se ha explorado un poco esta posibilidad (Ezequiel Alejandro Madrigal Carrillo, Tesis de Doctorado), pero las mutaciones también se han correlacionado con una menor estabilidad de la proteína. Si se logra trascender este problema con otra batería o tipo de mutaciones, tal vez este será el mejor camino de exploración.

## 6. Conclusiones

- Las fusiones de la proteína P no parecen ser promisorias por la necesidad intrínseca de utilizar condiciones desnaturalizantes en la purificación.
- Un cambio en condiciones de cristalización y parámetros cristalinos no implica un cambio en los contactos moleculares de la red.
- Los protocolos de simulación de dinámica molecular para evaluar la unión de ligandos a la proteína P deben ser calibrados, de ser posible, con un sistema experimental.

## 7. Perspectivas

- Cocrystalizar la proteína P con el dominio de la PylRS, producido de manera independiente, en presencia de pre-tRNA. La obtención del complejo ternario de las dos proteínas y el pre-tRNA podría aportar información de la interacción del 5'-líder con la proteína P
- Realizar diversos controles con compuestos en los que se haya probado experimentalmente que se unen a la proteína P, a pesar de que no sean inhibidores y verificar sus parámetros de salida en diversos programas de acoplamiento molecular y tratar de ajustar los parámetros de las dinámicas moleculares para que correspondan relativamente a las distintas afinidades registradas. La inclusión de suficientes controles experimentales podría generar una plataforma confiable para la búsqueda de inhibidores contra la proteína P.

## 8. Bibliografía

1. Gobringer M, Kretschmer-Kazemi Far R, Hartmann RK. Analysis of RNase P Protein (rnpA) Expression in *Bacillus subtilis* Utilizing Strains with Suppressible rnpA Expression. *J Bacteriol.* 2006;188(19):6816-23.
2. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell.* 1983;35(3 Pt 2):849-57.
3. Kazantsev AV, Pace NR. Bacterial RNase P: a new view of an ancient enzyme. *Nat Rev Microbiol.* 2006;4(10):729-40.
4. Mondragón A. Structural studies of RNase P. *Annu Rev Biophys.* 2013;42:537-57.
5. Esakova O, Krasilnikov AS. Of proteins and RNA: The RNase P/MRP family. *RNA.* 2010;16(9):1725-47.
6. Stams T, Niranjanakumari S, Fierke CA, Christianson DW. Ribonuclease P protein structure: evolutionary origins in the translational apparatus. *Science.* 1998;280(5364):752-5.
7. Kazantsev AV, Krivenko AA, Harrington DJ, Carter RJ, Holbrook SR, Adams PD, et al. High-resolution structure of RNase P protein from *Thermotoga maritima*. *PNAS.* 2003;100(13):7497-502.
8. Reich C, Olsen GJ, Pace B, Pace NR. Role of the protein moiety of ribonuclease P, a ribonucleoprotein enzyme. *Science.* 1988;239(4836):178-81.
9. Reiter NJ, Osterman A, Torres-Larios A, Swinger KK, Pan T, Mondragón A. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature.* 2010;468(7325):784-9
10. Mikkelsen NE, Brannvall M, Virtanen A, Kirsebom LA. Inhibition of RNase P RNA cleavage by aminoglycosides. *PNAS.* 1999;96(11):6155-60.
11. Eubank TD, Biswas R, Jovanovic M, Litovchick A, Lapidot A, Gopalan V. Inhibition of bacterial RNase P by aminoglycoside-arginine conjugates. *FEBS Lett.* 30 de enero de 2002;511(1- 3):107-12.
12. Kawamoto SA, Sudhakar CG, Hatfield CL, Sun J, Behrman EJ, Gopalan V. Studies on the mechanism of inhibition of bacterial ribonuclease P by aminoglycoside derivatives. *Nucleic Acids Res.* 2008;36(2):697-704.
13. Wu S, Mao G, Kirsebom LA. Inhibition of Bacterial RNase P RNA by Phenothiazine Derivatives. *Biomolecules.* 2016;6(3).
14. Hernandez-Cid A, Aguirre-Sampieri S, Diaz-Vilchis A, Torres-Larios A. Ribonucleases P/MRP and the expanding ribonucleoprotein world. *IUBMB Life.* junio de 2012;64(6):521-8.
15. Woodhams MD, Stadler PF, Penny D, Collins LJ. RNase MRP and the RNA processing cascade in the eukaryotic ancestor. *BMC Evol Biol.* 2007;7(Suppl 1):S13.
16. Willkomm DK, Gruegelsiepe H, Goudinakis O, Kretschmer-Kazemi Far R, Bald R, Erdmann VA, et al. Evaluation of bacterial RNase P RNA as a drug target. *Chembiochem.* 2003;4(10):1041-8.
17. Henkels CH, Kurz JC, Fierke CA, Oas TG. Linked Folding and Anion Binding of the *Bacillus subtilis* Ribonuclease P Protein. *Biochemistry.* 2001;40(9):2777-8
18. Nienaber VL, Richardson PL, Klighofer V, Bouska JJ, Giranda VL, Greer J. Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat Biotechnol.* 2000;18(10):1105-8.
19. Madrigal-Carrillo E-A, Díaz-Tufinio C-A, Santamaría-Suárez H-A, Arciniega M, Torres-Larios A. A screening platform to monitor RNA processing and protein-RNA interactions in ribonuclease P uncovers a small molecule inhibitor. *Nucleic Acids Res.* 2019;47(12):6425-38.

20. Concepcion J, Witte K, Wartchow C, Choo S, Yao D, Persson H, et al. Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization. *Comb Chem High Throughput Screen*. 2009;12(8):791-800.
21. Wartchow CA, Podlaski F, Li S, Rowan K, Zhang X, Mark D, et al. Biosensor-based small molecule fragment screening with biolayer interferometry. *J Comput Aided Mol Des*. 2011;25(7):669- 76.
22. Kobe B, Ve T, Williams SJ. Fusion-protein-assisted protein crystallization. *Acta Crystallogr F Struct Biol Commun*. 2015;71(Pt 7):861-9.
23. Price AJ, Howard S, Cons BD. Fragment-based drug discovery and its application to challenging drug targets. *Essays Biochem*. 2017;61(5):475-84.
24. Yamada H, Tamada T, Kosaka M, Miyata K, Fujiki S, Tano M, et al. 'Crystal lattice engineering,' an approach to engineer protein crystal contacts by creating intermolecular symmetry: Crystallization and structure determination of a mutant human RNase 1 with a hydrophobic interface of leucines. *Protein Sci*. 2007;16(7):1389-97.
25. Blaise M, Thirup S. Crystallization of *Escherichia coli* maltoporin in the trigonal space group R3. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2010;67(Pt 1):114-6.
26. Eyal E, Gerzon S, Potapov V, Edelman M, Sobolev V. The limit of accuracy of protein modeling: influence of crystal packing on protein structure. *J Mol Biol*. 2005;351(2):431-42.
27. El Omari K, Dhaliwal B, Ren J, Abrescia NGA, Lockyer M, Powell KL, et al. Structures of respiratory syncytial virus nucleocapsid protein from two crystal forms: details of potential packing interactions in the native helical form. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2011;67(Pt 10):1179-83.
28. Rajavel M, Gopal B. Analysis of multiple crystal forms of *Bacillus subtilis* BacB suggests a role for a metal ion as a nucleant for crystallization. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 5):635-9.
29. Zhang J, Wang X, Zhou K, Chen G, Wang Q. Self-Assembly of Protein Crystals with Different Crystal Structures Using Tobacco Mosaic Virus Coat Protein as a Building Block. *ACS Nano*. 2018;12(2):1673-9.
30. Krivenko AA, Kazantsev AV, Adamidi C, Harrington DJ, Pace NR. Expression, purification, crystallization and preliminary diffraction analysis of RNase P protein from *Thermotoga maritima*. *Acta Crystallogr D Biol Crystallogr*. 2002;58(Pt 7):1234-6.
31. Suzuki T, Miller C, Guo L-T, Ho JML, Bryson DI, Wang Y-S, et al. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat Chem Biol*. 2017;13(12):1261- 6.
32. Mukai T, Crnković A, Umehara T, Ivanova NN, Kyrpides NC, Soll D. RNA-Dependent Cysteine Biosynthesis in Bacteria and Archaea. *mBio*. 2017;8(3).
33. Wan W, Tharp JM, Liu WR. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim Biophys Acta*. 2014;1844(6):1059-70.
34. Torres-Larios A, Swinger KK, Krasilnikov AS, Pan T, Mondragón A. Crystal structure of the RNA component of bacterial ribonuclease P. *Nature*. 2005;437(7058):584-7.
35. Benvenuti M, Mangani S. Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nat Protoc*. 2007;2(7):1633-51.
36. Leslie A, Powell H. Evolving Methods for Macromolecular Crystallography . *NATO Science*; vol. 245: 41-51.
37. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr*. 1 de febrero de 2010;66(Pt 2):125-32.
38. Evans P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*. 2006;62(Pt 1):72-82.

39. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr.* 2007;40(Pt 4):658-74.
40. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004;60:2126-32.
41. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem.* 2010;31(2):455- 61.
42. Chen Y, Torres J. Special topics in drug discovery .IntechOpen; 53-70 p.
43. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *J Cheminform.* 2011;3:33.
44. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint- based similarity calculations. *J Cheminform.* 2015;7:20.
45. Sousa da Silva AW, Vranken WF. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes.* 23 de julio de 2012;5(1):367.
46. Martín-García F, Papaleo E, Gomez-Puertas P, Boomsma W, Lindorff-Larsen K. Comparing Molecular Dynamics Force Fields in the Essential Subspace. *PLoS One* ;10(3).
47. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins.* 2010;78(8):1950-8.

## 9. Anexos

### 9.1. Anexo 1. Formulación amortiguadores de lisis proteína TmP-PylRS

	<b>Amortiguador</b>	<b>Aditivo</b>
1	50 mM Tris-HCl pH 7.5	500 mM Cloruro de Sodio
2	50 mM Tris-HCl pH 7.5	500 mM Sulfato de Potasio
3	50 mM Tris-HCl pH 7.5	250 mM N-óxido de trimetilamina
4	50 mM Tris-HCl pH 7.5	50 mM Arginina
5	50 mM Tris-HCl pH 7.5	50 mM Ac. Glutámico
6	50 mM Tris-HCl pH 7.5	50 mM Arginina/Ac.Glutámico
7	50 mM Tris-HCl pH 7.5	20 % w/v PEG 3350
8	50 mM Tris-HCl pH 7.5	1 % v/v Tween 20
9	50 mM Tris-HCl pH 7.5	10 % v/v Glicerol
10	50 mM Tris-HCl pH 7.5	500 mM Sacarosa
11	50 MM Fosfatos de Na pH 8.0	500 mM Cloruro de Sodio
12	50 MM Fosfatos de Na pH 8.0	500 mM Sulfato de Potasio
13	50 MM Fosfatos de Na pH 8.0	500 mM N-óxido de trimetilamina
14	50 MM Fosfatos de Na pH 8.0	50 mM Arginina
15	50 MM Fosfatos de Na pH 8.0	50 mM Ac. Glutámico
16	50 MM Fosfatos de Na pH 8.0	50 mM Arginina/Ac.Glutámico
17	50 MM Fosfatos de Na pH 8.0	20 % w/v PEG 3350
18	50 MM Fosfatos de Na pH 8.0	1 % v/v Tween 20
19	50 MM Fosfatos de Na pH 8.0	10 % v/v Glicerol
20	50 MM Fosfatos de Na pH 8.0	500 mM Sacarosa

**Cuadro 4: Formulación de amortiguadores de lisis para la proteína TmP-PylRS.** Para la formulación de la batería de amortiguadores se tomaron las 20 combinaciones posibles de 2 valores de pH con 10 aditivos diferentes.



## 9.2. Anexo 2. Condiciones de cristalización para proteína P con precipitante diferente a PEG.

Kit	Posición	Precipitante	Sal	Amortiguador	Aditivo
CS	A1	30 % (v/v) MPD	0.02 M cloruro de calcio dihidratado	0.1 M Acetato de sodio pH 4.6	
CS	A2		0.4 Tartrato de sodio tetrahidratado		
CS	A3		0.4 Fosfato de amonio monobásico		
CS	A11		1 M Fosfato de amonio monobásico	0.1 M Citrato de Sodio dihidratado pH 5.6	
CS	C1		1 M Acetato de Sodio trihidratado	0.1 M Imidazol pH 6.5	
CS	D8		0.2 M Formato de magnesio dihidratado		
CS	E2	0.5 M Cloruro de Sodio	0.01 M Cloruro de Magnesio hexahidratado		0.001M HDTAB
CS	E3	25 % (v/v) Etilenglicol			
CS	E4	35 % (v/v) Dioxano			
CS	E9		2 M Cloruro de Sodio	0.1 M Acetato de sodio trihidratado pH 4.6	
CS	E11	1 M Hexanediol	0.01 M Cloruro de cobalto(II) hexahidratado	0.1 M Acetato de sodio trihidratado pH 4.6	
CS	G2	1.8 M Sulfato de Amonio	0.01 M Cloruro de cobalto(II) hexahidratado	0.1 M MES pH 6.5	
Nx	D6	5 % (v/v) Propanol	0.01 M Cloruro de magnesio hexahidratado	0.05 M Tris-HCl pH 7.5	
SRx	D7		1.5 M Nitrato de Sodio	0.1 M Acetato de sodio trihidratado pH 4.6	
SRx	E1		1.0 M Fosfato de amonio monobásico	0.1 M Acetato de sodio trihidratado pH 4.6	
SRx	F7		0.8 M Sulfato de litio monohidratado	0.1 M Acetato de sodio trihidratado pH 4.6	
SRx	G1		1M Sulfato de magnesio hidratado	0.1 M Acetato de sodio trihidratado pH 4.6	

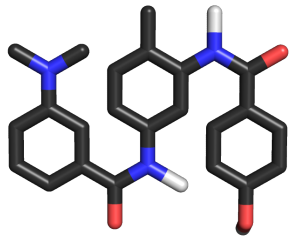
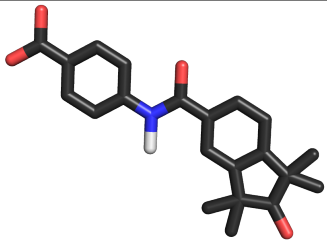
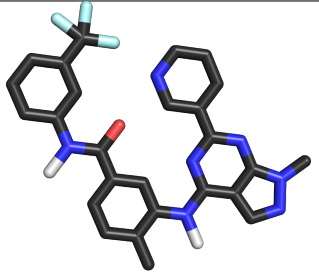
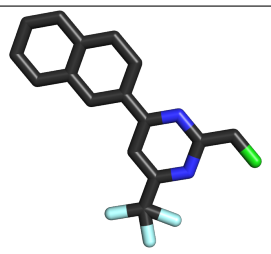
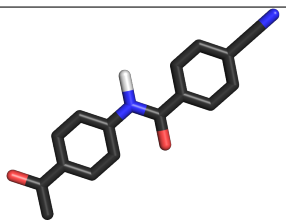
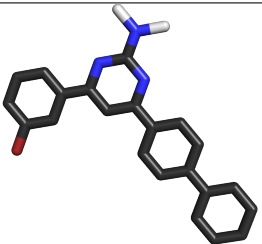
**Cuadro 5: Listado de condiciones de cristalización para la proteína P obtenidas por sembrado cruzado.** Se enlista las 20 condiciones de cristalización obtenidas cuyo agente precipitante es diferente a PEG. Para aquellas condiciones donde no se especifica el agente precipitante, se toma considera el componente con la mayor concentración de la formulación. Cada condición esta asociada a un kit de cristalización de la compañía Hampton Reserch: Crystal Screen (CS), Natrix(Nx), Salt Rx(SRx).

**9.3. Anexo 3. Sumario de valores obtenidos para la selección de inhibidores para la proteína P por métodos *in silico*.**

Acoplamiento Molecular (KJ/Mol)				Simulación de dinámica molecular (Å)				
ID	Caja 1	Caja 2	Flexible	RMSDmin	RMSDmed	RMSDmax	RMSDprom	RMSDmod
005	-6.3	-6.1	-7.5	0.0004	1.394	16.82	4.74±4.9	1.03
067	-5.9	-6.2	-7.0	0.0004	0.85	1.62	1.11±0.5	0.08
070	-5.7	-7.1	-9.3	0.0005	1.16	14.02	3.08±3.6	0.70
076	-6.2	-6.8	-10.0	0.0004	1.13	3.30	0.97±0.6	0.19
182	-7.5	-7.7	-9.6	0.0005	0.92	2.28	1.00±0.3	0.65
228	-6.8	-7.5	-9.2	0.0004	0.99	1.09	3.01±0.3	0.63
229	-6.7	-7.1	-8.5	0.0005	0.98	11.12	1.38±1.6	0.23
235	-6.8	-7.1	-9.1	0.0004	8.69	12.77	6.93±4.1	10.17
239	-6.9	-7.2	-8.8	0.0004	0.87	2.28	0.79±0.4	0.28
255	-6.4	-6.9	-7.6	0.0005	1.11	2.67	1.11±0.2	1.29
281	-6.3	-6.5	-8.6	0.0004	0.64	1.73	0.56±0.3	0.30
287	-7.1	-5.5	-6.7	0.0004	23.32	27.61	20.26±7.6	19.07
294	-5.6	-6.1	-7.9	0.0004	0.63	2.37	0.78±0.4	0.16
302	-6.5	-6.8	-5.9	0.0004	0.40	1.15	0.49±0.2	0.16
303	-6.1	-6.6	-8.1	0.0005	0.74	2.81	0.84±0.3	0.48
308	-6.5	-6.7	-8.4	0.0004	0.80	1.88	0.87±0.2	0.62
318	-7.1	-6.9	-8.2	0.0004	2.37	8.60	3.33±2.8	0.61
320	-7.7	-7.5	-10.4	0.0004	1.34	11.64	2.73±2.7	1.20
327	-7.1	-7.1	-8.5	0.0004	0.73	8.25	1.68±1.9	0.71
361	-6.3	-6.3	-6.9	0.0005	7.8	10.27	5.31±3.5	1.11
A	-5.1	-5.7	-4.7	0.0005	0.99	5.82	1.75±0.5	0.80

**Cuadro 6: Datos obtenidos de la selección de inhibidores por métodos *in silico*.**

#### 9.4. Anexo 4. Lista de candidatos a pruebas *in vitro*.

<b>ID Docking</b>	281	
<b>Sigma MDL</b>	MFCD02683971	
<b>Nombre</b>	ZM 336372	
<b>MW</b>	389.45	
<b>EF</b>	C23H23N3O3	
<b>IUPAC</b>	3-(Dimethylamino)-N-3-(4-hydroxybenzoyl)amino-4-methylphenyl-benzamide	
<b>ID Docking</b>	302	
<b>Sigma MDL</b>	MFCD18384955	
<b>Nombre</b>	BMS 753	
<b>MW</b>	351.4	
<b>EF</b>	C21H21NO4	
<b>IUPAC</b>	4-(2,3-Dihydro-1,1,3,3-tetramethyl-2-oxo-1H-inden-5-yl)carbonylamino-benzoic acid	
<b>ID Docking</b>	182	
<b>Sigma MDL</b>	MFCD20272929	
<b>Nombre</b>	NVP-BHG712	
<b>MW</b>	503.48	
<b>EF</b>	C26H20F3N7O	
<b>IUPAC</b>	4-Methyl-3-(1-methyl-6-(3-pyridinyl)-1H-pyrazolo[3,4-d]pyrimidin-4-ylamino)-N-(3-(trifluoromethyl)phenyl)-benzamide	
<b>ID Docking</b>	255	
<b>Sigma MDL</b>	MFCD14585229	
<b>Nombre</b>	-	
<b>MW</b>	322.71	
<b>EF</b>	C16H10ClF3N2	
<b>IUPAC</b>	2-(Chloromethyl)-4-(2-naphthyl)-6-(trifluoromethyl)pyrimidine	
<b>ID Docking</b>	294	
<b>Sigma MDL</b>	MFCD04544404	
<b>Nombre</b>	-	
<b>MW</b>	264.28	
<b>EF</b>	C16H12N2O2	
<b>IUPAC</b>	4-Cyano-N-(4-acetylphenyl)benzamide	
<b>ID Docking</b>	239	
<b>Sigma MDL</b>	MFCD21335911	
<b>Nombre</b>	-	
<b>MW</b>	402.29	
<b>EF</b>	C22H16BrN3	
<b>IUPAC</b>	4-(1,1-Biphenyl-4-yl)-6-(3-bromophenyl)-2-pyrimidinamine	

Cuadro 7: Lista de candidatos a pruebas *in vitro*