



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

## Maestría y Doctorado en Ciencias Bioquímicas

SÍNTESIS DE ÁCIDO CAFEICO USANDO VARIANTES DEL CITOCROMO BM3

TESIS

QUE PARA OPTAR POR EL GRADO DE:  
Maestro en Ciencias

PRESENTA:  
Q.F.B. JORGE LUIS JIMÉNEZ NIEBLA

TUTOR PRINCIPAL  
DRA. GLORIA SAAB RINCÓN  
[Instituto de Biotecnología-UNAM](#)

MIEMBROS DEL COMITÉ TUTOR

LORENZO PATRICK SEGOVIA FORCELLA  
[Instituto de Biotecnología-UNAM](#)

LILIANA QUINTANAR VERA  
[CINVESTAV-IPN](#)

Ciudad de México. Noviembre, 2019



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.

El presente proyecto de tesis de maestría fue desarrollado en el laboratorio de la Dra. Gloria Saab Rincón dentro del departamento de Ingeniería Celular y Biocatálisis del Instituto de Biotecnología de la UNAM, bajo el programa de Maestría y Doctorado en ciencias bioquímicas.

Para la realización del mismo se contó con el financiamiento del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT/UNAM a través del proyecto IT200617. Además se realiza un amplio reconocimiento al Programa de Apoyo a los Estudiantes del posgrado (PAEP) por todos los apoyos otorgados.

Este trabajo se realizó gracias a la beca de CONACyT para estudios de posgrado nivel maestría, con número de becario 482377.

## **AGRADECIMIENTO Y DEDICATORIAS**

Quiero empezar agradeciendo a la Dra. Gloria Saab por recibirme en su laboratorio, darme la oportunidad de realizar este proyecto y depositar su confianza en mis ideas y experimentos para sacar adelante el trabajo, su paciencia durante las etapas poco fructíferas y su apoyo moral fueron una herramienta indispensable en estos 2 años y medio.

Quiero dedicar y agradecer de igual manera el apoyo incondicional de mis padres, mi hermano y mi novia, que junto conmigo avanzaron y aguantaron durante este tiempo mi estrés y mis pláticas sobre mi proyecto que probablemente no tenían mucho sentido para ellos.

Junto con ellos, agradezco al extraordinario equipo de laboratorio con quienes compartí risas, mucha comida, sufrimiento, abrazos, llantos, muchísimo café y satisfacción cada día en el laboratorio y con cada uno de los experimentos.

No puedo irme del instituto sin agradecer a otra gran familia que es mi generación de maestría, 33 grandes personas que agradezco haber encontrado y compartido estudios, fiestas y actividades deportivas en Palomas ticas y EEEPAJA.

Mi maestría la terminé en la UNAM, pero muchas herramientas que me ayudaron a entrar en el programa y terminar las obtuve en UABC gracias al Dr. Marco Ramos, Dra. Rosi Mares y el apoyo de Ekaterina

No puedo mencionar todos los nombres de las grandes personas que me encontré aquí, pero puedo decirles que el proyecto no fue sencillo, pero ustedes hicieron que lo disfrutara al máximo....

Por último y no menos importante, quiero agradecer a Cory y Don Toño por su labor en mantener el laboratorio y los estantes limpios y llenos con material limpio y estéril. Una ardua labor que permite al laboratorio seguir adelante, de nuevo muchas gracias.

“All models are wrong, but some are useful”  
George E. P. Box.

# ÍNDICE

<b>AGRADECIMIENTO Y DEDICATORIAS</b> .....	<b>2</b>
<b>ÍNDICE</b> .....	<b>6</b>
<b>ÍNDICE DE FIGURAS</b> .....	<b>8</b>
<b>ÍNDICE DE TABLAS</b> .....	<b>9</b>
<b>NOMENCLATURA</b> .....	<b>10</b>
<b>RESUMEN</b> .....	<b>11</b>
<b>1 INTRODUCCION</b> .....	<b>12</b>
<b>2 MARCO TEORICO</b> .....	<b>13</b>
2.1 <i>Ácido cafeico</i> .....	13
2.2 <i>Biosíntesis de ácido cafeico</i> .....	14
2.3 <i>Alternativas de síntesis</i> .....	14
2.4 <i>Familia de citocromos P450</i> .....	15
2.5 <i>Citocromo P450 BM3</i> .....	18
2.6 <i>Evolución dirigida</i> .....	19
2.7 <i>Herramientas de acoplamiento moleculares</i> .....	20
<b>3 ANTECEDENTES</b> .....	<b>23</b>
<b>4 HIPÓTESIS</b> .....	<b>26</b>
<b>5 OBJETIVO</b> .....	<b>27</b>
5.1 <i>Objetivo general</i> .....	27
5.2 <i>Objetivos particulares</i> .....	27
<b>6 MATERIALES Y METODOS</b> .....	<b>28</b>
6.1 <i>Cepas y plásmido</i> .....	28
6.2 <i>Análisis de interacción proteína-sustrato acoplamiento molecular</i> .....	28
6.3 <i>Construcción de variantes</i> .....	31
6.4 <i>Expresión y purificación de P450 BM3</i> .....	33
6.5 <i>Ensayo de actividad</i> .....	33
6.6 <i>Determinación de productos de reacción por HPLC</i> . .....	34
6.7 <i>Protocolo de análisis de mutantes</i> .....	34
<b>7 RESULTADOS</b> .....	<b>36</b>
7.1 <i>Acoplamiento molecular</i> .....	36
7.2 <i>Selección de sitios a mutagenizar</i> .....	47
7.3 <i>Construcción de biblioteca de mutantes</i> .....	52
7.4 <i>Expresión y purificación de P450 BM3</i> .....	56

7.4.1 Protocolo de expresión en BL21.....	56
7.5 Ensayos de actividad.....	58
7.6 Método de tamizaje de alta eficiencia.....	61
7.7 HPLC.....	63
7.8 Análisis de variantes.....	66
7.9 Dinámica molecular de interacción proteína-ligando.....	66
<b>8 DISCUSIÓN.....</b>	<b>69</b>
<b>9 CONCLUSIONES.....</b>	<b>71</b>
<b>10 PERSPECTIVAS.....</b>	<b>71</b>
<b>11 ANEXOS.....</b>	<b>72</b>
<b>12 REFERENCIAS.....</b>	<b>84</b>

## ÍNDICE DE FIGURAS

Figura 1. Ciclo catalítico de citocromo P450. ....	17
Figura 2. Estructura del sitio activo.....	19
Figura 3. Ruta biosintética propuesta para la síntesis de ácido cafeico.....	23
Figura 4. Reacción propuesta para la catálisis.....	26
Figura 5. Protocolo de evaluación de mutantes .....	35
Figura 6. Delimitación del área de análisis para acoplamiento molecular	37
Figura 7. Protocolo de acoplamiento.....	38
Figura 8. Re-acoplamiento molecular CYTBM3 con N-palmitoilglicina.....	38
Figura 9. Protocolo de refinamiento de estructuras para acoplamiento molecular.....	39
Figura 10. Procesos de minimización y equilibrio en GOMACS.....	40
Figura 11. Re-acoplamiento molecular después de las estructuras seleccionadas. ....	42
Figura 12. Análisis del sitio activo.....	43
Figura 13. Representación tridimensional de la topología del CYT 199A .....	45
Figura 14. Análisis de modos normales.....	46
Figura 15. Acoplamiento molecular del dominio reductasa.....	45
Figura 16. Alineamiento de regiones seleccionadas .....	48
Figura 17. Análisis de potencial electrostático.....	49
Figura 18. Residuos seleccionados para mutagénesis.....	51
Figura 19. Metodología empleada para la construcción de la biblioteca de mutantes .....	52
Figura 20. Ensamble de combinatoria de variantes por el método de megaprimer. ....	53
Figura 21 Ensamble de fragmentos por OE-PCR.....	53
Figura 22. PCR de plásmido purificado de colonias con las variantes en 438.....	54
Figura 23. Construcción de variantes in silico e in vitro.....	55
Figura 24. SDS-PAGE de extractos purificados de citocromo. ....	56
Figura 25. Expresión de CYT BM3 en BL21.....	57
Figura 26. Espectro de reacción.....	60
Figura 27. Método de tamizaje en formato de alta eficiencia para selección de variantes.	62
Figura 28. Protocolo de evaluación de variantes.....	62
Figura 29. HPLC de componentes de reacción y reacción con la enzima silvestre .....	64
Figura 30. Perfil de productos obtenidos .....	65
Figura 31. Análisis de variantes.....	66



Figura 32. A) Minimización B) Termostato C) Baróstato .....	67
Figura 33. RMSD, RMSF y contactos observados en dinámicas moleculares. ....	68

## ÍNDICE DE TABLAS

Tabla 1. Contactos en diferentes campos de fuerza .....	41
Tabla 2. Validación con CYT199A2.....	44
Tabla 3.- Selección de residuos para realizar mutagénesis.....	51
Tabla 4. Oligonucleótidos sintetizados para construcción de biblioteca.....	52

## ÍNDICE DE ANEXOS

Anexo 1. Estructuras cristalizadas de CIT BM3 .....	72
Anexo 2. Alineamiento estructural de diferentes CIT BM3 cristalizados .....	72
Anexo 3. Mapas de densidad electrónica.....	72
Anexo 4. RMSD entre estructuras cristalográficas.....	73
Anexo 5. Estados de protonación a pH 7.4 .....	74
Anexo 6. Representación 2D de contactos proteína-ligando. ....	76
Anexo 7. Contactos, energías de unión y distancia de C3 de los conformeros al Fe. ....	77
Anexo 8. Contactos observados durante los protocolos de acoplamiento.....	78
Anexo 9. Protocolo de refinamiento y acoplamiento molecular.....	78
Anexo 10. Validación con CYT199A2.....	79
Anexo 11. Segunda replica de dinámica molecular.....	81

## NOMENCLATURA

<i>°C</i>	<i>Grados centígrados</i>
<i>4HPA3H</i>	<i>4-hidroxifenilacetato-3-hidroxilasa</i>
<i>APBS</i>	<i>Adaptive Poisson-Boltzmann solver</i>
<i>CA</i>	<i>Ácido cafeico</i>
<i>CIT</i>	<i>Citocromo</i>
<i>CPEC</i>	<i>Circular polymerase extensión cloning</i>
<i>DO</i>	<i>Densidad óptica</i>
<i>FAD</i>	<i>Flavín adenín dinucleótido</i>
<i>FMN</i>	<i>Flavín mononucleótido</i>
<i>GSH</i>	<i>Glutati3n</i>
<i>HCl</i>	<i>Ácido clorhídrico</i>
<i>HPLC</i>	<i>High pressure liquid chromatography</i>
<i>IPTG</i>	<i>isopropil-B-D-1 tiogalactopiran3sido</i>
<i>kDa</i>	<i>kilodalton</i>
<i>LB</i>	<i>Luria Bertani</i>
<i>mM</i>	<i>milimolar</i>
<i>mV</i>	<i>milivolts</i>
<i>NADPH</i>	<i>Nicotinamida adenina dinucle3tido fosfato</i>
<i>ng</i>	<i>nanogramo</i>
<i>nm</i>	<i>nanomolar</i>

Jorge Luis Jiménez Niebla. “**Síntesis de ácido cafeico usando variantes del citocromo BM3**” (Bajo la dirección de: la Dra. Gloria Saab Rincón).

*Palabras claves:* ácido cafeico, ácido *p*-cumárico, CYT BM3, HPLC.

## **RESUMEN**

El ácido cafeico es un metabolito importante en las plantas que forma parte de la vía de los fenilpropanoides. Sus aplicaciones en la industria farmacéutica y su bajo rendimiento en la extracción de tejidos vegetales ha impulsado la búsqueda de formas adecuadas para obtener esta molécula. Su precursor, el ácido *p*-cumárico es un compuesto ampliamente distribuido en tejidos vegetales como tal, o en su forma conjugada. Es un reactivo relativamente económico que está disponible comercialmente. La conversión de ácido *p*-cumárico a ácido cafeico depende de una hidroxilación del anillo aromático en C3; esta reacción es catalizada por citocromo P450 en plantas. Sin embargo, los citocromos eucariotas están anclados a membrana, por lo que su expresión en otros organismos es difícil, además de que dependen de la actividad de una reductasa para regenerar su sitio catalítico.

El citocromo P450 de *Bacillus megaterium* (CYT BM3) es un citocromo bacteriano soluble y autosuficiente que utiliza ácidos grasos de cadena lineal como sustratos. CYT BM3 ha sido ampliamente utilizado como un biocatalizador para una gran variedad de sustratos con diferentes propiedades químicas. Sin embargo, no es capaz de usar moléculas orgánicas pequeñas de forma natural. En el presente trabajo buscamos crear una biblioteca inteligente (utilizando estrategias combinadas de evolución dirigida y diseño racional) basada en análisis estructural y herramientas bioinformáticas para generar un número limitado de variantes, en conjunto con un método de detección de alto rendimiento que permita seleccionar variantes activas para su posterior análisis con técnicas de HPLC y RMN para caracterizar las variantes capaces de formar ácido cafeico a partir del ácido *p*-cumárico.

# 1 INTRODUCCIÓN

Hoy en día con una elevada tasa de crecimiento que genera un calentamiento global acelerado, la sobre explotación de las zonas agrícolas que compromete la calidad nutricional de los alimentos y el aumento en la tasa de enfermedades como diabetes y cáncer. Surge la necesidad de fármacos que ayuden a la protección del daño UV, que puedan ser utilizados como prebióticos y que ayuden a contrarrestar enfermedades crónicas.

Una molécula que ha llamado la atención de la industria farmacéutica para atender este tipo de situaciones es el ácido cafeico. El ácido cafeico es un polifenol derivado del ácido cinámico que se ha utilizado en la industria cosmética para la protección del daño solar, en la industria alimenticia como suplemento nutricional y estudios recientes reportan actividades como antiviral, antitumoral, prebiótico y una capacidad antioxidante elevada.

La principal fuente de obtención de esta molécula es mediante la extracción de material vegetal, sin embargo, su concentración es muy baja por lo que su implementación genera un producto de alto precio. Por esta razón, se han propuesto nuevas formas de obtención como la modificación de rutas metabólicas redirigiendo el flujo de carbono a la producción de ácido cafeico y cambiando la especificidad de enzimas para realizar las hidroxilaciones necesarias para obtener esta molécula.

En este trabajo, proponemos utilizar el citocromo P450 BM3 para acelerar el paso limitante de la síntesis de ácido cafeico, que es la hidroxilación en el carbono 3 del ácido p-cumárico. El CYT BM3 es la monooxigenasa con el mayor número de recambio reportado y con una eficiencia de alrededor del 90% del transporte de electrones, así que resulta un sistema eficiente para la hidroxilación de moléculas de interés farmacéutico. Sin embargo, no es capaz de utilizar al ácido p-cumárico como sustrato de manera natural, por lo que después de un análisis de refinamiento y acoplamiento molecular seleccionamos sitios a mutagenizar para generar variantes con la capacidad de utilizar ácido p-cumárico como sustrato.

## 2 MARCO TEORICO

### 2.1 Ácido cafeico

El ácido cafeico (ácido 3,4-dihidroxicinámico) es un compuesto fenólico ampliamente distribuido en tejidos vegetales, está presente en muchas fuentes de alimentos y es parte de varios medicamentos de uso popular [Spagnol *et al.*, 2015]. Los ácidos fenólicos se caracterizan por la presencia de un anillo de benceno, un grupo ácido carboxílico y uno o más grupos hidroxilo o metoxilo, lo que confiere propiedades antioxidantes [Magnani *et al.*, 2014].

El metabolismo celular normal produce radicales libres, incluidas especies reactivas de oxígeno y nitrógeno, y pueden derivarse de la exposición a factores ambientales (exógenos) como la contaminación, la radiación, los pesticidas, el tabaco, entre otros. Los radicales libres pueden causar efectos beneficiosos o nocivos para la salud [Schafer *et al.*, 2001]. Cuando existe un desequilibrio entre los sistemas pro- y antioxidante, con un predominio de pro-oxidantes, se produce el estrés oxidativo. Este estrés puede asociarse con daño a lípidos, proteínas o genes y está implicado también en una gran variedad de enfermedades humanas como insuficiencia renal, diabetes o cancer [Finker *et al.*, 2000].

Al realizar una comparación sobre la capacidad antioxidante de seis compuestos hidroxicinámicos: ácido cafeico (CA), éster fenético del ácido cafeico (CAPE), ácido ferúlico (FA), éster fenético del ácido ferúlico (FAPE), ácido rosmarínico (RA) y ácido clorogénico (CHA) por el método de Rancimat (tiempo de inducción de la oxidación de lípidos) y el método de DPPH (contenedor de radicales libres) el ácido cafeico destacó en ambos métodos compitiendo con el compuesto usado como control positivo, R-tocoferol en el método Rancimat y superándolo en el de DPPH [Chen *et al.*, 1997]. En tejido cardiaco de ratones diabéticos el ácido cafeico fue capaz de aumentar el contenido de GSH y optimizar la actividad catalasa y SOD mejorando su defensa antioxidante y de manera paralela en conjunto con ácido elágico lograron disminuir las citocinas proinflamatorias IL-6,

IL-1B, TNF-A y la hipocoagulabilidad asociada a la diabetes, aumentando la actividad de la proteína C [Chao *et al.*, 2009].

En humanos el ácido cafeico, que es absorbido en la corriente sanguínea, puede inhibir la oxidación *in vitro* de lipoproteínas de baja densidad. También es conocido como agente protector contra la oxidación catalizada por metales en aceite de oliva y muestra un potente efecto inhibitor sobre el daño oxidativo del ADN dependiente del hierro *in vitro* [Hynes *et al.*, 2004].

Estudios posteriores sobre sus actividades biológicas sugieren que el ácido cafeico posee actividad anti-cancerígena [Rojas *et al.*, 2016], propiedades antiinflamatorias [Lin *et al.*, 2012; Furuya *et al.*, 2012], anti-viral [Ikeda *et al.*, 2011] y exhibe una mayor actividad antioxidante para humanos, promueve una mayor protección ante la oxidación de lipoproteínas de baja densidad que el ácido p-cumárico y el ácido ferúlico [Furuya *et al.*, 2012]. Además, su derivado, el éster fenético (CAPE), ha atraído la atención por sus efectos terapéuticos, que incluyen su potencial como antidiabético y agente protector del hígado, así como actividad antitumoral para el tratamiento del cáncer de mama en humano [Lin *et al.*, 2012].

## **2.2 Biosíntesis de ácido cafeico**

El ácido cafeico es un metabolito importante para las plantas que lo utilizan como precursor en la ruta de fenilpropanoides. Su biosíntesis se lleva a cabo a partir de la desaminación de fenilalanina que genera ácido cinámico, subsecuentes hidroxilaciones en las posiciones 4- y 3- del anillo aromático dan lugar al ácido cinámico, el cual se convierte en ácido cafeico a través de ácido p-cumárico. Las enzimas involucradas, cinamato 4-hidroxilasa (C4H) y p-cumarato 3-hidroxilasa (C3H) son citocromos específicos de las plantas. Debido a su inestabilidad y asociación a la membrana, la purificación y la caracterización de estas enzimas es bastante desafiante, particularmente para C3H [Magnoni *et al.*, 2014].

## **2.3 Alternativas de síntesis**

El ácido cafeico ha demostrado tener aplicaciones de importancia para la industria farmacéutica como una potente herramienta terapéutica, por lo que se ha tratado de encontrar alternativas viables para la obtención de esta molécula. La hidroxilación directa de compuestos orgánicos por métodos de síntesis química suelen ser procesos intensivos de alta energía y a menudo no selectivos, que conducen a la formación de productos secundarios y desechos tóxicos [Munday *et al.*, 2016]. Actualmente, el ácido cafeico es comúnmente producido por extracción de fuentes vegetales, como granos de café, sin embargo, su contenido en esta fuente es bajo. La hidrólisis química o enzimática de derivados de ácido cafeoilquínico también ha sido ampliamente utilizada para este propósito [Magnoni *et al.*, 2014].

Su precursor, el ácido p-cumárico es un compuesto que se distribuye ampliamente en los tejidos vegetales como un producto o como formas conjugadas. Es un reactivo relativamente barato que está disponible comercialmente. La conversión de ácido p-cumárico a ácido cafeico depende de una hidroxilación en el C3 del anillo aromático; esta reacción es catalizada por citocromo P450 en plantas, los citocromos eucariontes suelen tener menor número de recambio, son de difícil expresión de manera heteróloga y en particular este citocromo suele utilizar como sustrato al ácido p-cumárico conjugado [Lin *et al.*, 2012].

Los avances en la ingeniería metabólica y la biología sintética en la última década han permitido la producción de varios metabolitos secundarios específicos de plantas en microorganismos recombinantes [Furuya *et al.*, 2012].

## **2.4 Familia de citocromos P450**

Los citocromos P450 son una superfamilia de enzimas (> 12 000 secuencias) dependientes del grupo hemo que se pueden encontrar en todos los reinos biológicos [Girvan *et al.*, 2016]. Esta familia de enzimas realiza la incorporación de un solo átomo de oxígeno molecular en un sustrato, con la reducción concomitante del otro átomo a agua [Bernhardt *et al.*, 2006]. Actúan principalmente como monooxigenasas externas ya que necesitan un donador

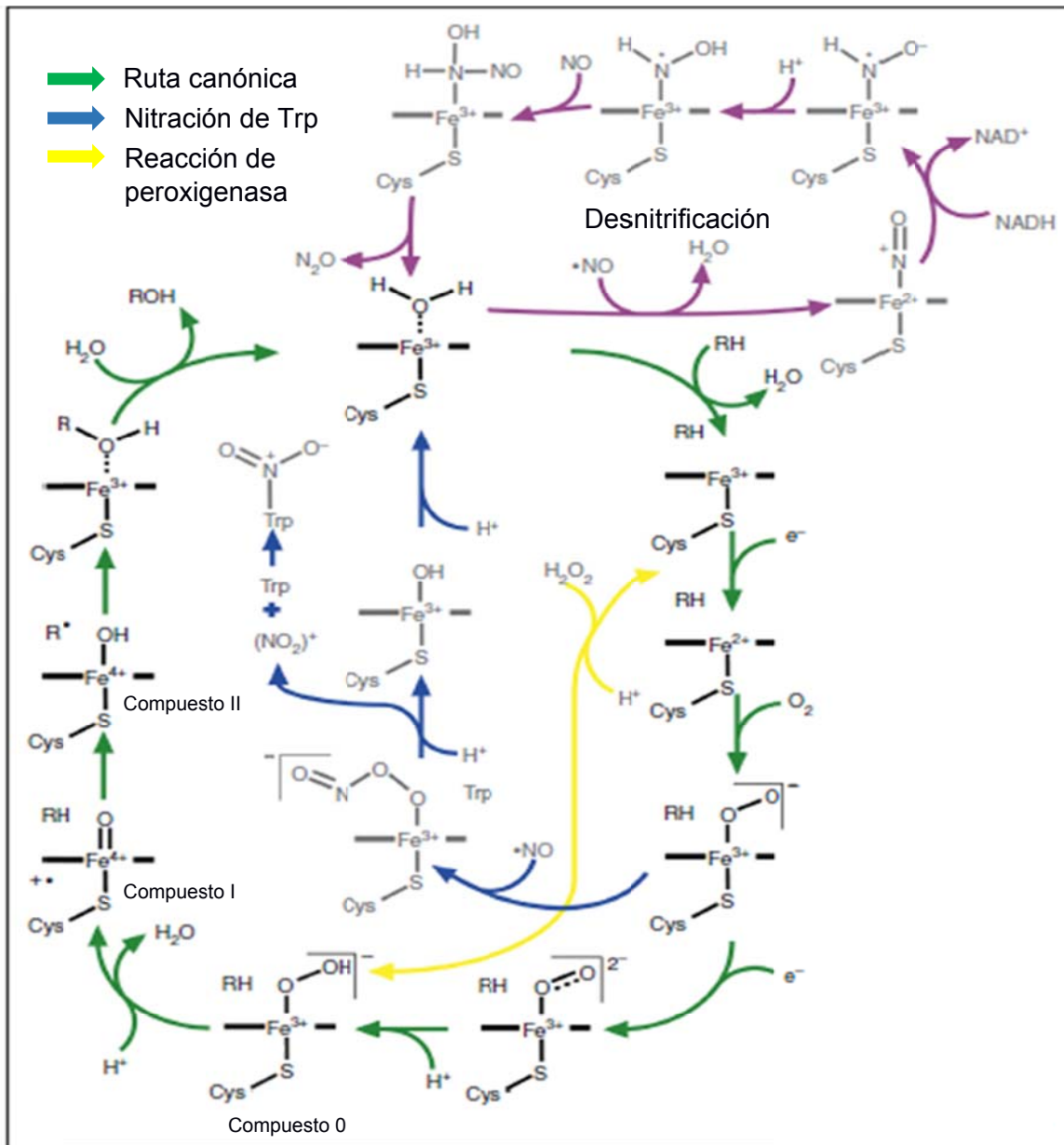
externo de electrones para regenerar su sitio catalítico. Se caracterizan por una diversidad extremadamente alta de las reacciones catalizadas y los sustratos convertidos [Brill *et al.*, 2014]. Están involucrados en reacciones tan diversas como, por ejemplo, hidroxilación, N-, O- y S-desalquilación, sulfoxidación, epoxidación, desaminación, desulfuración, deshalogenación, peroxidación y reducción de N-óxido [Bernhardt *et al.*, 2006].

La mayoría de los citocromos P450 son incapaces de interactuar directamente con el donante de electrones primario NADPH, por lo que requieren de una o dos proteínas adicionales para intercambiar los electrones del NADPH y transportarlos al sitio activo de P450 [Brill *et al.*, 2014], el ciclo de regeneración del hierro del grupo hemo y sus rutas alternas se encuentran descritos en la figura 1.

A través de la actividad de P450s se puede producir una gran variedad de sustancias cuya síntesis química resulta difícil y costosa [Cirino *et al.*, 2002], debido a la alta energía en el enlace C-H, a la presencia de múltiples enlaces con energías similares en compuestos orgánicos y a la mayor reactividad de los productos oxidados en comparación con la de los reactivos [Fasan *et al.*, 2012]. Sin embargo, se ha demostrado que las reacciones de oxidación mediadas por citocromos tienen un alto potencial ya que proporcionan rutas biocatalíticas por la activación sitio- y estereoespecífica de enlaces C-H [Roiban *et al.*, 2015], lo cual ha sido de interés para aplicaciones biocatalíticas.

Los sistemas bacterianos P450 normalmente exhiben mayores tasas catalíticas que P450s eucarióticos y son fáciles de manejar en el laboratorio debido a su solubilidad y alto nivel de expresión en huéspedes heterólogos. Estas propiedades hacen que los P450 bacterianos sean objetivos ideales para la ingeniería de proteínas [Bernhardt *et al.*, 2006].





**Figura 1. Ciclo catalítico de citocromo P450.**

La unión del sustrato (RH) desplaza la molécula de agua axial que se encontraba coordinada y cambia el equilibrio del hierro del hemo a la forma de alto spin, permitiendo que la primera transferencia de electrones reduzca al hierro al estado ferroso. La unión de  $O_2$  y la transferencia de un segundo electrón de un compañero redox, da lugar a la formación de especie radical de cationes reactivos de ferryl-oxo porfirina conocida como compuesto I. La abstracción de un protón del sustrato forma un intermediario ferryl protonado (compuesto II) que hidroxila el radical del sustrato de acuerdo con el mecanismo de rebote de radical. Imagen modificada de Girvan *et al.*, (2016).

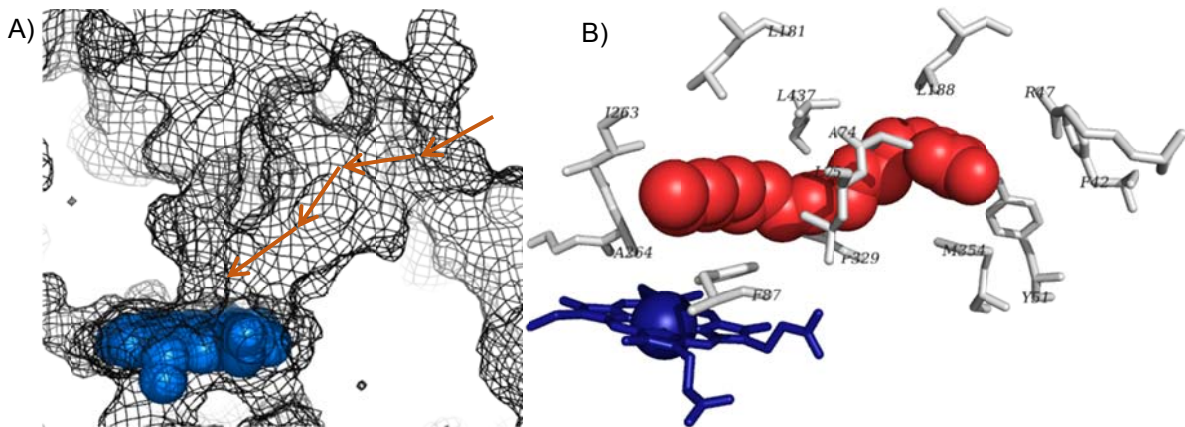
## 2.5 Citocromo P450 BM3

El citocromo P450 BM3 (Cyt BM3) de *Bacillus megaterium* es un citocromo bacteriano de 119 kDa de la familia CYP102 que cataliza la oxidación de los ácidos grasos de cadena larga. Hasta la fecha ha sido el más investigado y se ha demostrado que presenta tasas catalíticas muy altas con números de recambio de 5000 min<sup>-1</sup> o más para la oxidación de ácidos grasos [Carmichael *et al.*, 2001] con un acoplamiento de la oxidación de NADPH a formación de producto de 88-98% [Noble *et al.*, 1999], con la ventaja de que es catalíticamente autosuficiente, lo que implica que la transferencia de electrones del dominio FAD / FMN reductasa y el dominio monooxigenasa P450 están contenidos en un único péptido [Girvan *et al.*, 2016].

El P450 BM3 se ha utilizado y rediseñado para catalizar la oxidación de una gran variedad de sustancias de interés biotecnológico. Mientras que la enzima silvestre es poco selectiva, se han generado variantes con mayor regio- y estereoselectividad para diferentes sustratos [Seifert *et al.*, 2009]. La ingeniería de proteínas ha tenido éxito en la creación de numerosas variantes, que poseen especificidades de sustrato alteradas, incluida la conversión de indol [Li *et al.*, 2000], propranolol [Otey *et al.*, 2006], hidrocarburos aromáticos policíclicos [Carmichael *et al.*, 2001], alcoxiresulfuros fluorógenos [Lussenburg *et al.*, 2005], terpenos [Seifert *et al.*, 2009], entre otros componentes [Erkelenz *et al.*, 2011]. Las estrategias de mutagénesis continúan expandiendo el repertorio de sustratos, particularmente en el caso de los P450 microbianos solubles para los cuales los datos estructurales ayudan a guiar la ingeniería de proteínas [Girvan *et al.*, 2016].

Una particularidad importante del P450 BM3 es el canal de acceso del sustrato; es evidente en su estructura que es un canal largo que conecta la superficie de la proteína con el bolsillo catalítico donde se encuentra el grupo hemo. Más de una docena de aminoácidos extienden sus cadenas laterales hacia este espacio proporcionando múltiples objetivos para mutagénesis, (Figura 2) [Li *et al.*, 1997], los residuos P25, R47 y Y51 se encuentran en la entrada del sitio catalítico y están involucrados en el reconocimiento de los ácidos grasos [Ahmed *et al.*, 1999].

Simulaciones de dinámicas moleculares del P450 BM3 demostraron que los residuos F87 y A328 causan cambios en la regioselectividad mediante la estabilización del sustrato en diferentes conformaciones de unión [Branco *et al.*, 2008]; el residuo localizado en la posición 5 se encuentra en la vecindad inmediata del grupo hemo y su cadena lateral apunta hacia el centro del hemo, por lo tanto, se espera que interactúe con todos los sustratos durante la oxidación [Seifert *et al.*, 2009].



**Figura 2. Estructura del sitio activo.**

A) Entrada al sitio activo del hemo (PDB 2IJ2) B) Dominio hemo del CYP BM3 con unión a N-palmitoilglicina (PDB 1JPZ). El grupo hemo se encuentra en azul y el ligando en rojo.

## 2.6 Evolución dirigida

Los enfoques de la evolución dirigida se basan en la detección efectiva de una reacción o propiedad en la enzima a partir de enormes bibliotecas de variantes, de las cuales solo una pequeña parte del tamaño teórico puede ser evaluada. Información sobre la estructura de la proteína permiten disminuir el número de variantes a generar y hacer bibliotecas inteligentes [Cole *et al.*, 2011]

Resulta imposible cubrir todas las mutaciones posibles de una proteína típica: ya que si se realiza una aleatorización completa de un decapeptido por ejemplo, puede llegar a producir  $10^{20}$  combinaciones únicas de aminoácidos, lo cual excede el tamaño de las bibliotecas de casi todos los métodos conocidos de creación de bibliotecas de proteínas. Las estrategias de diversificación genética están diseñadas para realizar un muestreo disperso del espacio de secuencia dimensional. La contribución de cada variante crea un paisaje evolutivo en el cual se generan proteínas que tendrán un mejor desempeño en diferentes

condiciones, el objetivo de estudios de evolución dirigida es hacer un muestreo mutacional dentro de este paisaje y relacionarlo con niveles máximos de la propiedad buscada, por ejemplo actividad, estabilidad, etc.

La mutagénesis dirigida es una técnica que se utiliza para aumentar la probabilidad de que una biblioteca acumule variantes que favorezcan una condición deseada, siempre y cuando se tenga noción de las posiciones de aminoácidos que más probablemente sean determinantes de la función deseada. En ausencia de una relación estructura-función, la mutagénesis aleatoria puede proporcionar una mayor posibilidad de acceder a los miembros de la biblioteca funcional que la diversidad de la biblioteca en residuos elegidos incorrectamente [Packer *et al.*, 2015].

Los investigadores han desarrollado una amplia gama de métodos para realizar ambas formas de diversificación genética, y las estrategias más exitosas a menudo integran mutagénesis dirigida y aleatoria.

## **2.7 Herramientas de acoplamiento moleculares**

El acoplamiento molecular es un método que evalúa una energía de interacción del ligando sobre una superficie del receptor. Utiliza la orientación favorecida para predecir la fuerza de la conexión o la afinidad de unión entre el ligando y proteína mediante el uso de funciones de puntuación. El acoplamiento a menudo se aplica para predecir la afinidad y la actividad del fármaco. El objetivo principal de acoplamiento molecular es simular computacionalmente el proceso molecular de interacción y lograr una conformación optimizada para que la energía libre del sistema general se minimice.

En el descubrimiento de fármacos, el acoplamiento entre proteína-ligando o proteína-proteína juega un papel importante en la predicción de la orientación del ligando al unirse. Una de las limitaciones de este estudio es la función de puntuación para cada interacción, en donde las interacciones electrostáticas, interacciones de Van der Waals, interacciones coulómbicas y la formación de enlaces de hidrógeno tienen que ser medidas de manera eficiente por el programa de acoplamiento.

La suma de todas estas interacciones se aproxima a una puntuación, que representa la probabilidad de que la interacción se vea más o menos representada. En los sistemas de cuerpo rígido más simples, el ligando se busca en un espacio rotacional o translacional para encajar en el sitio de unión [Alberg *et al.*, 1993]. La precisión del acoplamiento en un enfoque de cuerpo rígido es mucho mayor para los complejos unidos que para las moléculas sin complejos. Aunque los cambios estructurales observados entre las formas ligada y libre son pequeños, la diferencia en la precisión implica que la suposición de rigidez no está plenamente justificada. Por lo tanto, varios grupos han dedicado sus esfuerzos para permitir la flexibilidad del receptor y el ligando [Pagadala *et al.*, 2017].

En las últimas dos décadas se han desarrollado más de 60 programas y herramientas diferentes para el acoplamiento molecular entre los cuales se encuentran DOCK, AutoDock, FlexX, Surflex, GOLD, ICM, GLIDE, Cdocker, LigandFit, MCDock, FRED, MOE-Dock, Autodock Vina, rDock, UCSF Dock, RosettaDock entre otros. Estos programas trabajan utilizando ya sea algoritmos genéticos, técnicas de búsqueda sistémica, simulaciones de Monte Carlo o algoritmos basados en la complementariedad de formas.

Después de realizar pruebas de validación de los acoplamientos moleculares se observó que entre estos programas, AutodockVina, GOLD y MOE-Dock son los más consistentes en sus resultados de reacoplamiento. Se ha observado que los resultados de los acoplamientos predicen las posiciones demostradas experimentalmente con un promedio de desviación cuadrática media (RMSD) de 1.5 a 2 Å [Bissantz *et al.*, 2000]. Cabe mencionar que de los tres programas seleccionados solo AutodockVina es de acceso libre.

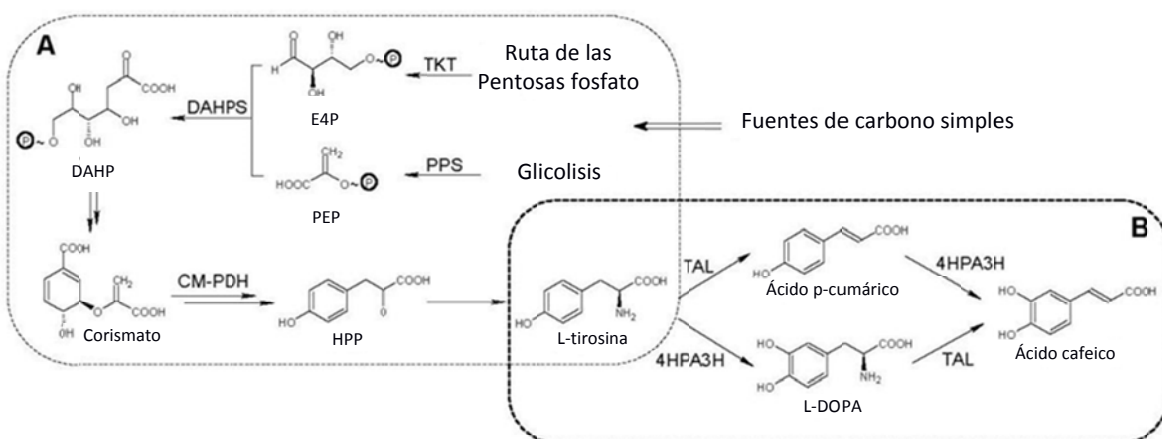
AutoDockVina es un programa de código abierto diseñado en *The Scripps Research Institute*. Se ejecuta especialmente para estudios de receptor-ligando y hay tres pasos principales involucrados en el acoplamiento: Preparar la proteína, definir el sitio activo y preparar el ligando. AutoDockVina es más rápido que otros programas de acoplamiento como Autodock4 y no requiere utilizar herramientas

como AutoGrid y AutoDock. Este programa se ha implementado con éxito, especialmente en el cribado virtual, el análisis de flexibilidad y el acoplamiento de iones metálicos.

Un programa que empieza a llamar la atención es RosettaDock, el cual ha sido mayormente utilizado para interacciones proteína-proteína; sin embargo, se ha adaptado para realizar análisis de interacción proteína-ligando. Este es un programa basado en predicción de estructura. Busca en el cuerpo rígido y lateralmente en el espacio conformacional de las cadenas de los aminoácidos de las dos proteínas que interactúan para encontrar una estructura compleja con un mínimo de energía. RosettaDock se basa principalmente en multi-inicio, utilizando un algoritmo Monte Carlo a escala múltiple [Chaudhary *et al.*, 2016].

### 3 ANTECEDENTES

Recientemente, se encontraron genes y enzimas involucradas en la biosíntesis de ácido cafeico en *Saccharothrix espanaensis* y *Rhodobacter capsulatus*, una tirosina amoniaco liasa (TAL) codificada por *sam8* y un C3H microbiano codificado por *sam5* son responsables para la conversión de tirosina en ácido p-cumárico y luego al ácido cafeico, respectivamente. La 4-hidroxifenilacetato 3-hidroxilasa (4HPA3H) de *E. coli* es capaz de hidroxilar ácido p-cumárico y tirosina además de su sustrato nativo, el ácido 4-hidroxifenilacético, para la producción de este compuesto. A partir del conocimiento de estas reacciones se plantea una ruta de biosíntesis dual, en la que se expresan la 4HPA3H y de manera heteróloga la enzima TAL para la producción de ácido cafeico a partir de tirosina como se observa en la figura 3 [Lin *et al.*, 2012].



**Figura 3. Ruta biosintética propuesta para la síntesis de ácido cafeico.**

(A) Vía nativa para la síntesis de tirosina en *E. coli*. (B) Vía artificial para la biosíntesis de ácido cafeico a partir de tirosina. PPS: fosfoenolpiruvato sintasa; TKT: transcetolasa; CM-PDH: corismato mutasa/prefenato deshidrogenasa; DAHPS: 3-deoxi-D-arabino-heptulosa-7-fosfato sintasa; 4HPA3H: 4-hidroxifenilacetato 3-hidroxilasa; TAL: Tirosina amoniaco liasa; E4P: D-eritrosa-4-fosfato; PEP: fosfoenolpiruvato; HPP: 4-hidroxifenilpiruvato. Imagen modificada de Lin, Y., & Yan, Y. (2012) *Microbial cell factories*.

Una vía alternativa y no muy explorada para la síntesis de ácido p-cumárico es a partir del uso de citocromos bacterianos como es el caso del citocromo 199A2 al cual tras someterlo a procesos de evolución dirigida se lograron identificar algunas variantes con la capacidad de realizar una primera hidroxilación sobre el ácido cinámico para la producción de ácido m-cumárico o ácido p-cumárico, y



otras variantes con la capacidad de tomar estos productos y formar ácido cafeico [Furuya *et al.*, 2012]

El citocromo P450 de *B. megaterium* (CYT BM3) es un citocromo bacteriano soluble y autosuficiente que utiliza ácidos grasos de cadena lineal como sustrato. CYT BM3 ha sido utilizado ampliamente como biocatalizador de una gran variedad de sustratos con propiedades químicas diferentes. El análisis de la estructura del sitio catalítico de CYT BM3 puede coadyuvar en el diseño de una biblioteca de variantes que puedan catalizar la reacción de producción de ácido cafeico.

La mutagénesis aleatoria ha resultado útil para mejorar la actividad de BM3 hacia moléculas de distintos tamaños y propiedades; así, las variantes P450BM3 (R47L / Y51F / F87A) o P450BM3 (A74G / F87V / L188Q) se identificaron utilizando un método de selección basado en la formación de indol y fueron reportados por hidroxilar hidrocarburos aromáticos policíclicos grandes a alta tasas ( $> 100 \text{ min}^{-1}$ ) aunque con baja eficiencia de acoplamiento (3- 10%) [Li *et al.*, 2001]. Durante las rondas de mutagénesis se han logrado encontrar aminoácidos requeridos para lograr cambios importantes en la función y que pueden afectar su termoestabilidad [Fasan *et al.*, 2008]. Cuando las sustituciones en BM3 dan como resultado una interacción enzima-sustrato subóptima promueven el desacoplamiento en la transferencia de electrones por un sellado imperfecto del sitio activo al solvente durante la catálisis y conlleva a la producción de especies reactivas de oxígeno [Li *et al.*, 1997; Kadkhodayan *et al.*, 1995; Loida *et al.*, 1993].

Recientemente se han utilizado moléculas análogas que actúan como coadyuvantes para realizar la oxidación de moléculas sencillas como benceno a fenol por parte del P450 BM3 silvestre [Shoji *et al.*, 2013], el análogo es una molécula inerte que llena parte del sitio activo y restringe la entrada de los sustratos permitiendo una interacción más cercana al hemo de lo que lo harían sin una molécula análoga [Munday *et al.*, 2016]. Estas moléculas análogas son ácidos perfluorocarboxílicos de cadena larga, que imitan a los sustratos preferidos por BM3 pero contienen enlaces C-F no oxidables en sus cadenas laterales. Este

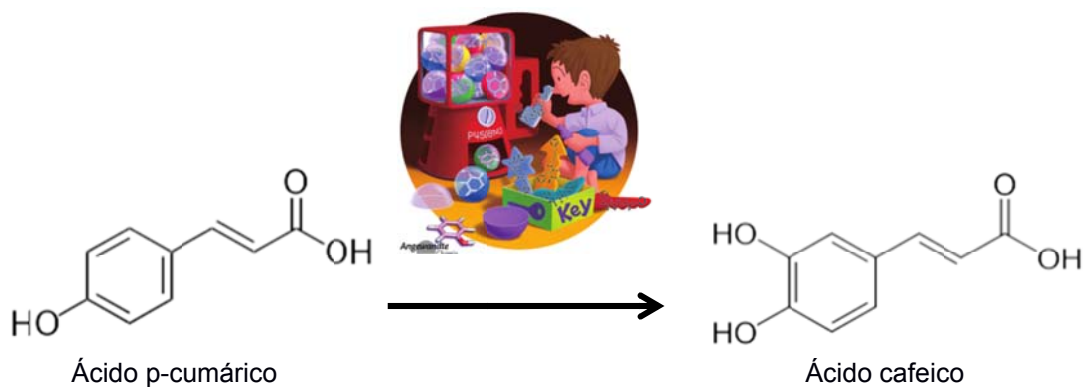


método podría proporcionar una herramienta valiosa, complementaria o sinérgica para la ingeniería de proteínas [Pochapsky *et al.*, 2010].

La relación entre la estructura, estabilidad y actividad de las proteínas ha sido un problema importante aún no resuelto en biología molecular. El pleno entendimiento de esta relación sería de gran utilidad para guiar la mutagénesis sitio dirigida y otras técnicas de ingeniería de proteínas. Los estudios de difracción de rayos X han proporcionado información estructural extensa para muchas proteínas, desafiando a los teóricos a desarrollar técnicas confiables capaces de utilizar dicho conocimiento como una base para la predicción de las mutaciones que permitan modificar diversos atributos en una enzima, como un cambio de especificidad, estabilidad, etc. [Lee *et al.*, 1991].

## 4 HIPÓTESIS

La promiscuidad de los citocromos puede ser aprovechada para moldear el sitio activo de la CYT P450 BM-3 para llevar a cabo la producción de ácido cafeico a partir de ácido p-cumárico.



**Figura 4. Reacción propuesta para la catálisis.**

En la imagen podemos observar el paso de hidroxilación de ácido p-cumárico a ácido cafeico con la imagen de un niño ejemplificando un modelo de llave cerradura con el CYTBM3 mientras juega en una máquina expendedora.

## **5 OBJETIVO**

### **5.1 Objetivo general**

Generar una variante de CYT P450 BM-3 con la capacidad de utilizar como sustrato ácido p-cumárico para la producción de ácido cafeico.

### **5.2 Objetivos particulares**

- Analizar si la enzima silvestre CYT BM3 es capaz de utilizar el ácido cumárico como sustrato
- Identificar blancos para modificar la especificidad mediante mutagénesis y favorecer la producción de ácido cafeico.
- Buscar e implementar un método de tamizaje en formato de alta eficiencia para evaluar la biblioteca de variantes.
- Generación de una biblioteca de variantes mediante mutagénesis combinatoria en los sitios identificados.
- Seleccionar y evaluar la capacidad de las variantes seleccionadas de reconocer ácido p-cumárico como sustrato.

## 6 MATERIALES Y METODOS

### 6.1 Cepas y plásmido

Para la construcción de la biblioteca de variantes se inició con el plásmido pCWori en el cual se encuentra insertado el gen de la proteína CYT-BM3 4E10 con mutaciones en las posiciones A82L and A328V, que le permiten hidroxilar análogos de sustratos pequeños poco polares (Chen *et al.*, 2012). Los fragmentos de las bibliotecas fueron generados a partir de este ADN molde e insertados en el vector pET28a en el cual se trabajará la biblioteca completa con la cepa de *E. coli* MC1061 (araD139  $\Delta$ (araA-leu)7697  $\Delta$ (lac)X74 galK16 galE15(GalS) lambda- e14-mcrA0 relA1 rpsL150(strR) spoT1 mcrB1 hsdR2). Una vez construida la biblioteca se utilizará la cepa *Escherichia coli* BL21 DE3 (*ompT gal dcm lon hsdS<sub>B</sub>(r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>)*  $\lambda$ (DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*])[*malB<sup>+</sup>*]<sub>K-12</sub>( $\lambda^S$ )) para la sobre expresión y evaluación de las variantes.

### 6.2 Análisis de interacción proteína-sustrato mediante acoplamiento molecular

Para la identificación de blancos de mutagénesis se realizó un análisis de acoplamiento sobre la estructura del sitio catalítico y se determinaron qué aminoácidos podrían beneficiar a la actividad de hidroxilación en el anillo aromático. Para los ensayos de acoplamiento se utilizó la herramienta de Autodock Vina (<http://vina.scripps.edu/>), en una computadora portátil Sony Vaio con Intel Core i5, 4GB de RAM y una tarjeta de video NVIDIA GeForce 410M 512MB con el sistema operativo Debian GNU/Linux. Las dinámicas moleculares se ejecutaron en la supercomputadora HP clúster platform 3000SL "Miztli".

Como parte de un ensayo preliminar se realizaron dos protocolos de tratamiento a tres estructuras cristalográficas, que corresponden a la enzima silvestre sin ligando (PDB 1BVY), la enzima silvestre unida a palmitoilglicina (PDB 1JPZ) y una variante que fue modificada y fue nombrada 4E10. Para el primer tratamiento se tomó a la estructura cristalográfica en formato PDB y se pasó directamente a un formato PDBQT siguiendo el protocolo recomendado para el análisis en autodock en el cual se agregaron todos los hidrógenos a la proteína. Se calcularon las

cargas y se conservaron sólo los hidrógenos polares (aquellos que se encuentren en un enlace polarizado) para al final obtener el formato PDBQT correspondiente, el segundo protocolo consistió en revisar la estructura cristalográfica en el servidor en línea WHATIF (<https://swift.cmbi.umcn.nl/servers/html/index.html>) y utilizar su tratamiento de preparación de estructuras para acoplamiento y posteriormente obtener el formato PDBQT para el análisis. Para realizar el acoplamiento se probaron 3 tamaños de caja para comprobar que no existieran sitios de unión diferentes a la hendidura catalítica en los cuales el sustrato interactúe de manera representativa, se hizo un análisis de acoplamiento ciego en el que se incluye todo el dominio hemo de la proteína, una caja de la mitad del dominio hemo en el que se incluía el sitio activo y un análisis solo del sitio activo. Para este primer acoplamiento se utilizó como ligando, palmitoilglicina, ácido p-cumárico y ácido cafeico.

Para validar el proceso de acoplamiento molecular y obtener resultados más cercanos a la realidad se realizaron tres protocolos adicionales diferentes, para el primer tratamiento del proceso de validación se realizó un análisis estructural del CYT BM3; se revisó la calidad de las estructuras cristalográficas con el programa COOT (<https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/>) tomando en cuenta los mapas de densidad electrónica, Ramachandran plot, factor B y moléculas cocrystalizadas; se realizaron alineamientos de secuencia y estructurales con la herramienta de PyMOL (<https://pymol.org/2/>) matchalign y el software profit (<http://www.bioinf.org.uk/software/profit/>) para seleccionar estructuras con buena calidad y que representen movimientos conformacionales representativos en el CYT BM3.

Una vez seleccionadas las estructuras se sometieron a un proceso de afinamiento para mejorar errores al momento de dilucidar los mapas de densidad electrónica, primero se eliminaron los choques entre cadenas laterales utilizando el software scrwl4 (<http://dunbrack.fccc.edu/SCWRL3.php/>) y después se agregaron hidrógenos a toda la proteína y se comprobó que los residuos como la Histidina (His), Asparagina (Asn) y Glutamina (Gln) se encontraran debidamente orientados

con respecto al mapa de densidad electrónica y los residuos de la vecindad. Al terminar este proceso se sometieron las estructuras a un proceso de minimización utilizando el campo de fuerza OPLS-AA con el modelo de agua SPC, agregando estados de protonación con ayuda de PROPKA ([http://nbcrc-222.ucsd.edu/pdb2pqr\\_2.1.1/](http://nbcrc-222.ucsd.edu/pdb2pqr_2.1.1/)) y sometiéndolo a 5000 pasos de descenso rápido para encontrar mínimos locales y 5000 adicionales de gradiente conjugado para explorar el mínimo local, una vez concluido el proceso de minimización se obtuvo el archivo PDBQT correspondiente para el acoplamiento. En los dos protocolos restantes se agregó el Hemo en su lugar correspondiente a las estructuras resultantes del protocolo uno y se sometieron a un proceso de minimización nuevamente, en este caso utilizando dos campos de fuerza diferentes, el campo de fuerza GROMOS53a6 para el segundo protocolo y el campo de fuerza CHARMM36 para el tercer protocolo, en todos los tratamientos se neutralizó el sistema utilizando sodio como catión y cloro como anión. En esta ocasión solamente se sometieron a 5000 pasos de descenso rápido y posteriormente se hizo un equilibrio nvt con el algoritmo de V-rescale de 100 ps y un equilibrio npt con el baróstato Parrinello-Rahman de 100 ps para llevar nuestra estructura a una temperatura de 25 °C y una presión de 1 atm; de igual manera se obtuvieron los archivos PDBQT correspondientes.

Para el acoplamiento simultáneo se realizó un pequeño script con el protocolo de acoplamiento molecular que consiste en definir el ligando y las proteínas receptoras, para someterlos a un proceso de acoplamiento con exhaustividades de entre 200-3000, se evaluaron los resultados con las 15 interacciones con energías de interacción más favorables calculados por el programa.

De manera alternativa se llevó a cabo un análisis de modos normales con el servidor en línea de El Nemo para observar los cambios conformacionales más representativos de la proteína y un análisis de potencial electrostático APBS (Adaptive Poisson-Boltzmann Solver) utilizando el servidor pdb2pqr para observar las cargas presentes en la hendidura catalítica.

### 6.3 Construcción de variantes

Para realizar la construcción de la librería de mutantes se utilizaron las siguientes estrategias:

#### Técnica de megaprimer

Mediante la síntesis de oligonucleótidos con los codones degenerados en los residuos seleccionados en el análisis "*in silico*", y un nucleótido que hibrida con alguno de los extremos del gene, se realizó una reacción de PCR. El producto de esta PCR se utiliza como un cebador (megaprimer) en amplificaciones futuras para ir aumentando el tamaño del producto e ir acumulando la combinatoria de variantes. Se realizaron una serie de PCRs en los que se fue extendiendo el tamaño del primer megaprimer de 150 hasta obtener un fragmento de aproximadamente 1100 pb correspondiente al dominio hemo de la proteína y donde se encuentran acumuladas la combinatoria de variantes seleccionadas.

#### Extension de sobrelapes mediante reacción en cadena de la polimerasa (por sus siglas en ingles OE PCR)

La extensión por traslapamiento permite generar un fragmento a partir de dos fragmentos que comparten una región de homología en alguno de sus extremos. Se tomó en cuenta el tamaño de cada fragmento para calcular la concentración final de cada uno y se procedió a realizar una amplificación por PCR. Esta amplificación se llevó a cabo con 50 ng de cada fragmento en 2 ciclos diferentes, de manera inicial se dieron 10x ciclos de amplificación sin añadir los oligos de cada extremo del fragmento, posterior a los 10 ciclos se agregaron los oligos y se amplificó por 25 ciclos adicionales. El producto de PCR fue purificado a partir de un gel de agarosa al 1% utilizando el kit de purificación en gel de Roche y cuantificado en Nanodrop.

#### Método de ensamble de Gibson

La estrategia metodológica de Gibson permite el ensamble de productos de PCR e incluso el vector sin la necesidad de digerir los fragmentos para la construcción final. Durante la reacción se utiliza una nucleasa que degrade los extremos 5', una polimerasa de alta fidelidad para rellenar los huecos en las regiones de homología después de la hibridación y una ligasa para cerrar todos los nick generados por la

amplificación. Al realizar el ensamble de Gibson se linealizó el vector digiriendo con las enzima de restricción NcoI/SacI, se utilizaron 200 ng de vector en una relación 1:3 con los fragmentos de la construcción en un volumen final de 5 µl, a los cuales se les agregó 15 µl del master mix de la reacción de gibson y se incubó durante 60 minutos a 50°C. Una vez terminada la reacción, ésta se utilizó para transformar células MC1061 por electroporación a 1.6mV, se recuperaron por una hora en medio SOC y se inocularon en cajas petri con kanamicina (30 mg/ml) para seleccionar las bacterias que fueron transformadas.

#### Clonación por extension circular de la polimerasa (siglas en ingles CPEC)

De manera alternativa al ensamble de gibson decidimos utilizar la técnica CPEC (circular polymerase extension cloning) en la que de igual manera se aprovechan las regiones de homología compartidas entre los fragmentos y el vector para realizar en este caso la amplificación del plásmido completo utilizando una polimerasa de alta fidelidad a través de ciclos de PCR. Se utilizaron concentraciones equimolares de vector e inserto, durante 15 ciclos de amplificación con 30s de extensión por Kb. Una vez terminada la reacción se usó para transformar células MC1061 por electroporación a 1.6mV, se recuperaron por una hora en medio SOC y se inocularon en cajas petri con kanamicina (30 mg/ml) para seleccionar las bacterias que fueron transformadas.

#### Método de mutagénesis Quick-change

Una propuesta ampliamente usada en la mutagénesis sitio dirigida ha sido la metodología desarrollada por Agilent denominada Quickchange™, en la que se utilizan un par de oligos complementarios que contienen en su secuencia los sitios a mutagenizar, de esta manera a partir del plásmido silvestre y con una polimerasa de alta eficiencia como la enzima Phusion es posible amplificar el vector completo introduciendo las mutaciones diseñadas en los oligos; el producto de la amplificación requiere un paso de digestión con la enzima *DpnI* para eliminar el DNA silvestre. Una vez terminada la reacción se transformaron células MC1061 por electroporación a 1.6mV, se recuperaron por una hora en medio SOC y se inocularon en cajas petri con kanamicina (30 mg/ml) para seleccionar las bacterias que fueron transformadas.



## 6.4 Expresión y purificación de P450 BM3

Un cultivo nocturno de *E. coli* BL21 DE3 que alberga el plásmido de expresión pET-28-4E10, que codifica para una variante de P450BM3 con las mutaciones A82L y A328V, se inoculó en 100 mL de medio LB suplementado con ampicilina. Las células se cultivaron a 37°C con agitación a 300 r.p.m hasta alcanzar una  $DO_{600nm}$  de 0.6, las células se indujeron añadiendo isopropil tio-b-d-galactopiranosido (IPTG) a una concentración final de 1 mM. Después de 12 horas de crecimiento a 30°C, las células se cosecharon por centrifugación a 4000 r.p.m durante 30 minutos. El pellet rojo marrón se volvió a suspender en amortiguador fosfatos de potasio pH 7.4 y se lisaron las células por sonicación. Los restos celulares se eliminaron por centrifugación a 10000 r.p.m durante 30 minutos a 4°C y el extracto libre de células se filtró a través de una membrana de 0.22  $\mu m$  antes de realizar la purificación por FPLC en una columna mono-Q, eluyendo con un gradiente lineal de cloruro de potasio de 250 a 400 mM en amortiguador  $PO_4$  pH 7.4.

## 6.5 Ensayo de actividad

Las reacciones se hicieron en volúmenes de 200  $\mu L$ , con concentración inicial de 0.7 mM de NADPH, 0.4 mM de sustrato (SDS como control positivo y ácido p-cumárico) con 40  $\mu L$  de extracto proteico en amortiguador de fosfatos sodio pH 7.4. Se realizó una medición del espectro de 230 a 500 nm midiendo cada 1 nm antes de iniciar la reacción con la adición de la enzima; seguida de una cinética a 340 nm durante 10-15 minutos aproximadamente y finalizando con una medición del espectro de igual manera que al inicio para observar diferencias en los picos de absorción observados.

### Ensayo por cromatografía en capa fina (TLC)

Una vez finalizada la reacción se tomaron los 200  $\mu L$  y se colocaron en un tubo eppendorf ámbar de 1.5 ml inmediatamente después de la reacción para evitar la oxidación de los productos de reacción y se le adicionaron 50  $\mu L$  de HCl 0.1M para obtener solo especies protonadas que puedan migrar correctamente por el cromatopla. Se cargaron 3  $\mu L$  de la reacción y se corrieron usando como fase móvil una mezcla de tolueno: acetato de etilo: ácido fórmico: agua (35:70:55:35)

#### Ensayo en formato de alta eficiencia

Una vez finalizada la reacción se toman 20  $\mu\text{L}$  y se mezclan con 200  $\mu\text{L}$  de la solución de revelado que contiene 20mM  $\text{FeCl}_3$ , 10 mM Ferrozina en amortiguador HEPES pH 3 y se mide la absorción a 593 nm. El cambio en la tonalidad incoloro, rosado a morado se observa en relación a la capacidad antioxidante del compuesto y en este caso a la concentración de ácido cafeico.

#### Ensayo de tinción Hemo con TMB

Con el propósito de verificar que la proteína de los extractos tanto silvestre como mutantes tuvieran su hemo activo, se utilizó TMB (3,3,5,5-Tetrametilbenzidina) 6.3mM para tratar geles SDS-PAGE en los que se corrieron los extractos. Después de correr los geles 12% SDS-PAGE en condiciones nativas se dejó incubando una hora en la oscuridad con el reactivo TMB, posterior a la incubación se agregó  $\text{H}_2\text{O}_2$  30mM para revelar los hemos activos.

### **6.6 Determinación de productos de reacción por HPLC.**

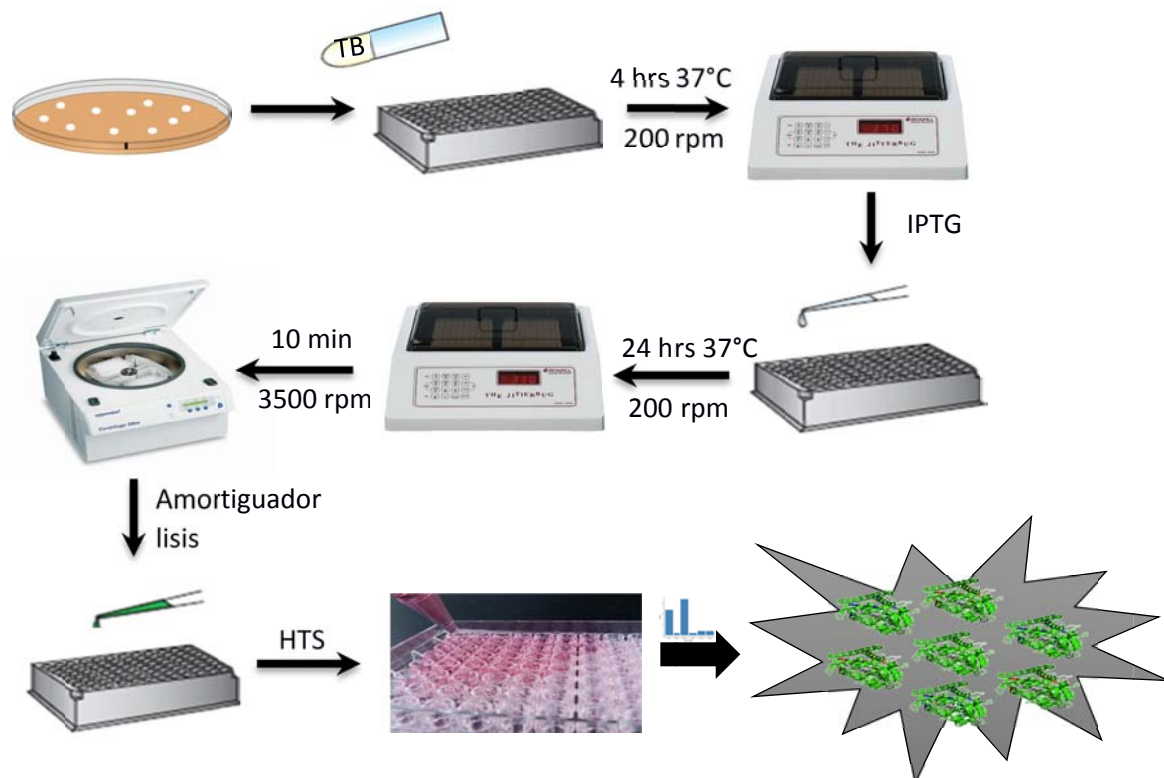
No existe un método para la determinación y cuantificación del ácido cafeico en los compendios oficiales, como las farmacopeas; sin embargo, la mayoría de los artículos actuales recomiendan la cuantificación de la sustancia mediante cromatografía líquida de alta resolución (HPLC).

Mezclas de ácido cafeico y ácido p-cumárico se pueden separar por HPLC utilizando una columna de fase reversa C18, con una fase agua:metanol:ac.ácetico (79:20:1) hasta 25 min a una velocidad de flujo de 1,0 ml  $\text{min}^{-1}$  y haciendo la detección a 330 nm. Los picos cromatográficos de ácido p-cumárico y ácido cafeico se confirmaron mediante la comparación de sus tiempos de retención y espectros UV con el de sus estándares de referencia.

Para demostrar que la hidroxilación se está realizando en el C3 del anillo aromático se realizará un análisis por resonancia magnética nuclear y una espectrometría de masas de los productos obtenidos, especialmente aquellos que eluyan en un perfil similar al del estándar de ácido cafeico por HPLC.

### **6.7 Protocolo de análisis de mutantes**

Para el análisis de la biblioteca de mutantes, se transformaron células de *E. coli* BL21 con la construcción de todas las variantes en el vector pET28 y se dejaron incubando durante 12 horas para obtener colonias aisladas y de buen tamaño. Actualmente se cuentan con un adaptador para la incubadora que tiene una capacidad aproximada de 24 cajas (con lo cual tenemos la posibilidad de trabajar con un máximo 2304 pocillos) la colonias se tomarán con palillos y se inocularon al medio terrific broth (TB) contenido en cada pocillo, tras 4 horas y media de incubación se inducen los cultivos con 0.5mM IPTG y después de 24 horas de incubación se centrifugan durante 5 minutos a 3500 rpm y 4°C, se descarta el sobrenadante. La pastilla celular se trata con amortiguador de lisis (Amortiguador fosfatos pH 8, con lisozima y DNAsa) durante 1 hora a 37°C y se realiza una centrifugación de 15 minutos a 3500 rpm y 4°C para obtener el extracto que será utilizado para hacer las reacciones y el tamizaje por el método de ferrozina. Durante cada proceso de evaluación de variantes se utiliza como control ácido p-cumárico, ácido cafeico y NADPH a las concentraciones finales de la reacción para evitar falsos positivos. La figura 5 muestra un esquema del protocolo.



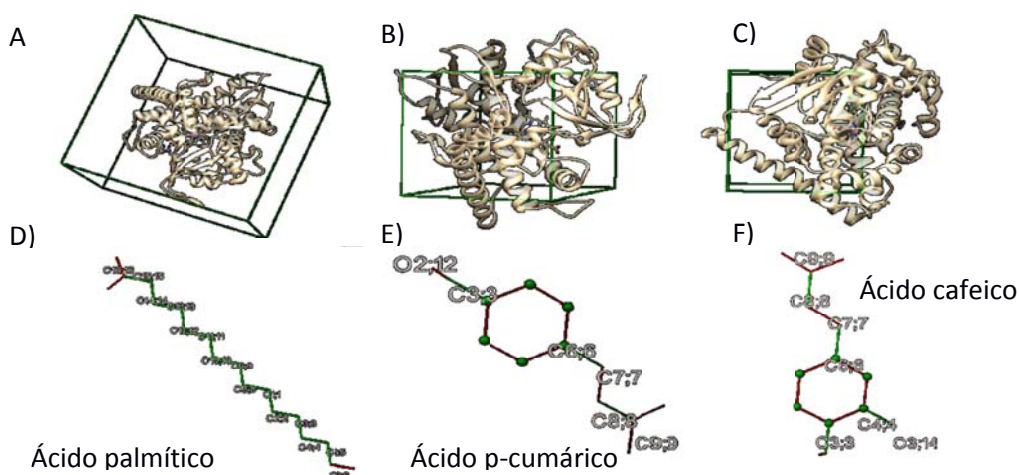
**Figura 5. Protocolo de evaluación de mutantes.** Una vez obtenidas las variantes se utilizó el siguiente protocolo de expresión y obtención de extractos para su evaluación y selección.-

## 7 RESULTADOS

### 7.1 Acoplamiento molecular

De manera inicial se realizaron 3 protocolos de tratamiento diferentes utilizando los programas ADT, Whatif y Molprobit (Anexo 8, Figura 9) con 4 estructuras cristalográficas para obtener un modelo de acoplamiento molecular adecuado y hacer un acoplamiento ciego. Al obtener los 12 formatos PDBQT se realizaron acoplamientos con tres tamaños diferentes de caja para buscar todos los sitios en los que se podría unir el ligando sobre la proteína (Figura 6). El tiempo de cómputo es proporcional al número de enlaces de rotación de cada molécula, en el caso de ácido p-cumárico este número es de 3, para el ácido cafeico de 4 y para el ácido palmítico de 16.

Se inició con una caja que cubre todo el dominio hemo de la proteína para hacer un acoplamiento ciego, en esta búsqueda solo se encontraron conformeros en el sitio activo de la proteína. Tomando en cuenta esto, se redujo a la mitad el sitio de la búsqueda, incluyendo el sitio activo para delimitar el análisis y explorar con mayor exhaustividad regiones cercanas al sitio activo, de igual manera



**Figura 6. Delimitación del área de análisis para acoplamiento molecular.**

En la parte superior se muestran las cajas utilizadas para el acoplamiento (A) (B) (C) y en la parte inferior tenemos a los ligandos utilizados, en verde se muestran los enlaces de rotación libre.

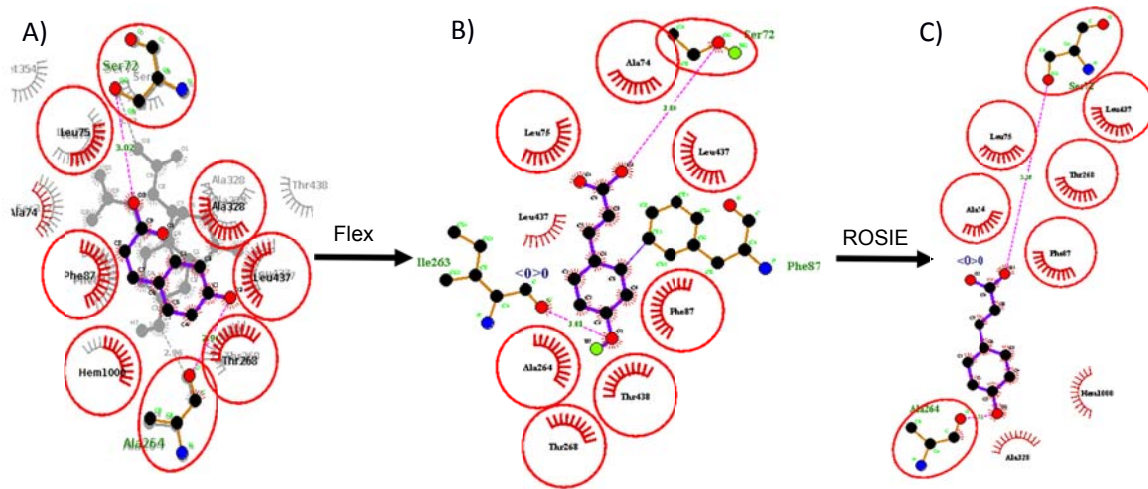
encontramos que nuestros ligandos solo interaccionaban con la proteína cerca del sitio activo, por lo que por último, sólo analizamos el sitio activo para ver los conformeros con menor energía. Al ir disminuyendo el tamaño de la caja nos permite hacer una búsqueda más exhaustiva en la que vamos limitando la

búsqueda y quitando regiones que sabemos que no interaccionan con el ligando, con la finalidad de buscar los confórmeros en el sitio activo con diferentes energías y poder hacer un análisis de las distancias y contactos. Al ver los confórmeros y energías de estos primeros análisis podemos ver que el resultado es reproducible y que a pesar de analizar diferentes regiones de la proteína obtenemos prácticamente el mismo resultado, por esto decidimos tomar la caja que corresponde al tamaño del sitio activo para ver los contactos entre proteína-ligando más importantes en los resultados del acoplamiento molecular.

Una vez que observamos que no existe una interacción fuera del sitio activo que nos pudiera generar ruido, se decidió tomar diferentes parámetros de acoplamiento, en esta ocasión disminuyendo la exhaustividad para ver cómo se comportaba el resultado y verificar que siguiera siendo reproducible, en busca de disminuir el tiempo de cómputo y tener buenos resultados. El valor de la exhaustividad nos permite decirle al programa cuantos ciclos hacer el análisis y evaluar la energía en los sitios y confórmeros seleccionados, en este caso se tomaron valores de exhaustividad de 200, 750, 1000, 2000 y 3000. En todos obtuvimos resultados consistentes y seleccionamos un valor de 2000 para análisis futuros.

Los resultados de las interacciones fueron analizados utilizando el programa LigPlot+ para una representación en 2D y el programa PyMOL para obtener una representación 3D. Analizando los resultados de los 12 modelos de acoplamiento se observó que los residuos que mayor contribución tenían eran Ser72, Leu75, Phe87, Ala264, Thr268, Ala328 y Leu437, ya que se encontraban representados en la mayoría de los resultados de los acoplamientos (Figura 7). Tomando esto en cuenta, se decidió hacer un acoplamiento molecular flexible, el cual permite dar libre movimiento a los residuos seleccionados durante el proceso de acoplamiento haciendo un muestreo mayor y permitiendo movimientos conformacionales dentro de la hendidura catalítica. Los resultados de este acoplamiento molecular flexible se utilizaron para someterlos a un programa de refinamiento de la interacción proteína-ligando llamado ROSIE (<https://rosie.graylab.jhu.edu/>) en el cual se observó que se conservaban los residuos mencionados previamente más una

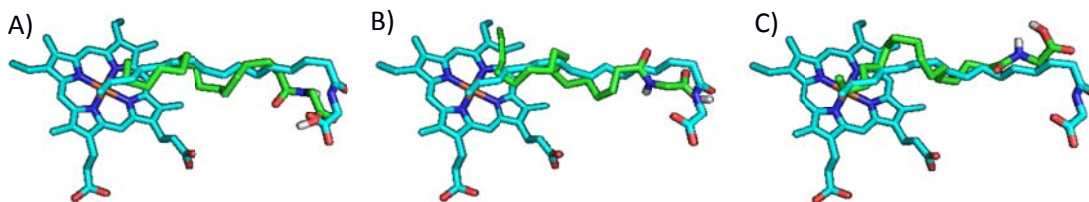
interacción, seleccionando de esta manera los residuos.- Ser72, Ala74, Leu75, Phe87, Ala264, Thr268, Ala328, Leu437



**Figura 7. Protocolo de acoplamiento.**

Se muestran mapas de contactos en 2D en (A) tenemos el resultado del acoplamiento con la proteína rígida (B) es el resultado del acoplamiento después de hacer flexibles los residuos que se vieron que interactuaban al estar rígidos (C) es el resultado después de un refinamiento hecho en el servidor en línea ROSIE.

Como parte de la validación de los protocolos utilizados en esta parte del análisis utilizamos la estructura 1JPZ, que fue cristalizada con una molécula de N-palmitoilglicina, como método de re-acoplamiento molecular y medir la desviación cuadrática media (RMSD por sus siglas en inglés) del conformero reportado en su estructura cristalográfica con respecto a los resultados del acoplamiento (Figura 8).



**Figura 8. Re-acoplamiento molecular CYTBM3 con N-palmitoilglicina.**

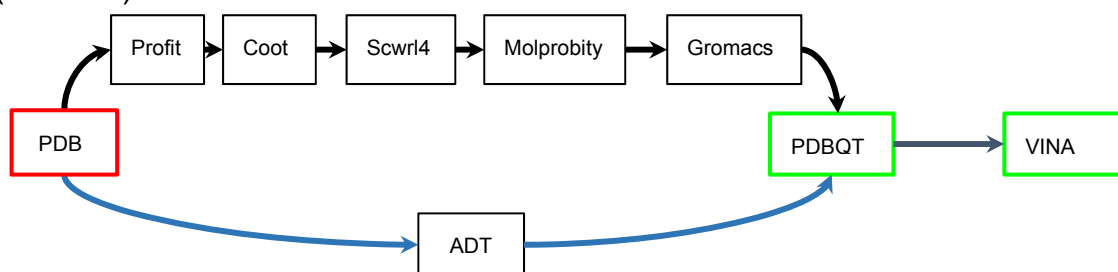
Resultado del proceso de validación por reacoplamiento en el que vemos los 3 conformeros de menor energía en verde y en azul la estructura cristalográfica con un RMSD de 2.1.

Debido a que las estructuras cristalográficas reportadas en el *Protein Data Bank* (PDB) para el dominio hemo del CYT BM3 son alrededor de 84, se decidió hacer una revisión para buscar las estructuras con buena resolución (1.2-2.8 ) y a partir de ellas realizar un alineamiento de secuencia y estructura que nos permita



detectar cambios representativos en la conformación de la proteína y tomar una muestra representativa de estas conformaciones para realizar un análisis más detallado de la interacción que nuestro sustrato puede tener con nuestra enzima (Anexo 1). Para esto se utilizó el programa profit y matchalign, con los cuales se observó que la proteína no tiene cambios conformacionales grandes y que las estructuras sin ligando son muy similares, mientras que las que se co-cristalizaron con un ligando tienden a tener una ligera diferencia no mayor a 0.5 Å con respecto a las que no tenían ligando (Anexo 2 y 4). En los alineamientos de 12 estructuras se observan cambios pequeños en las regiones de mayor nivel estructural mientras que en las asas se observan cambios más grandes como es esperado por su organización. El grupo hemo no presenta cambios representativos, por lo que asumimos que estos pequeños cambios pueden ser debido a un ajuste inducido sumamente fino que permite la llegada de los sustratos a distancias propicias para la hidroxilación. Se revisó parte del refinamiento que se realizó a los resultados de cristalografía para descartar que este cambio sea debido a uso de remplazo molecular para dilucidar la estructura.

Con base a estos resultados, se seleccionaron 10 estructuras de acuerdo con su resolución, valor de Rwork y fueron revisadas con el programa COOT para corroborar que el mapa de densidad electrónica estuviera completamente dilucidado para la hendidura catalítica y que no presentara huecos o residuos fuera de los ángulos permitidos en el gráfico de Ramachandran en esta región (Anexo 3).

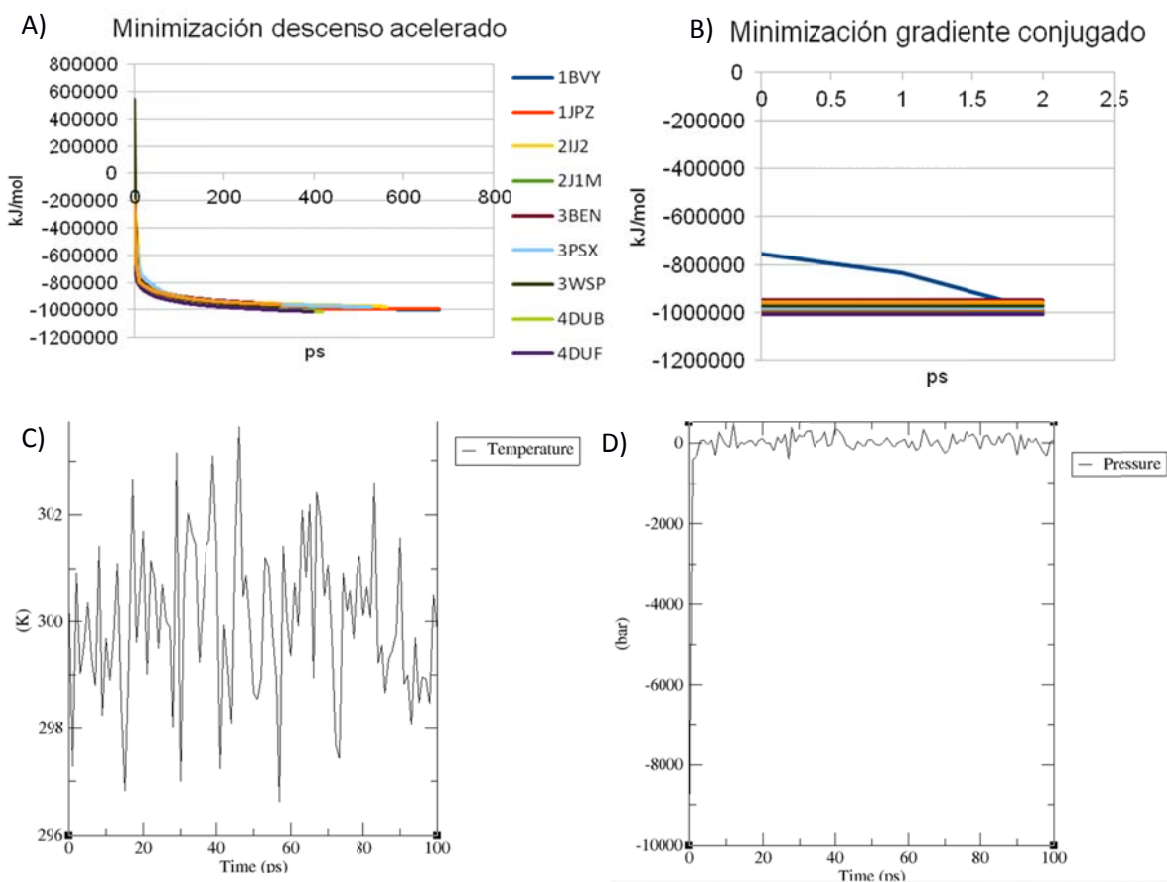


**Figura 9. Protocolo de selección y refinamiento de estructuras para acoplamiento molecular.**

El cuadro rojo representa las estructuras sin tratamiento y los cuadros verdes las estructuras listas para realizar el proceso de acoplamiento. Las flechas azules muestran el tratamiento cuando solo se utilizó autodock tools para convertir los archivos PDB a PDBQT y las flechas negras el proceso de refinamiento de la proteína para su análisis.

Después de este análisis se seleccionaron las estructuras 1BVY, 2IJ2, 2J1M (estructuras silvestres), 3PSX (variante), 1JPZ, 3BEN, 3WSP, 4DUB, 4DUF, 5XA3 (cocrystalizadas con un ligando). Se utilizó el programa Scwrl4 para revisar las cadenas laterales de los residuos y corregir su orientación, después se utilizó la plataforma en línea de molprobit para agregar hidrógenos y revisar que la orientación de las histidinas, asparaginas y glutaminas sean congruentes con los átomos vecinos y el mapa de densidad electrónica reportado.

Para buscar un nivel energético adecuado para nuestra enzima se utilizó el programa GROMACS 2018 para realizar un proceso de minimización de descenso acelerado inicial y uno de gradiente conjugado después utilizando el campo de fuerza OPLS-AA y el modelo de agua SPC.



**Figura 10. Procesos de minimización y equilibrio en GROMACS.**

Gráficos de proceso de minimización (A) Descenso acelerado, (B) Gradiente conjugado y equilibrio (C) nvt y (D) npt utilizando el programa GROMACS versión 2018.



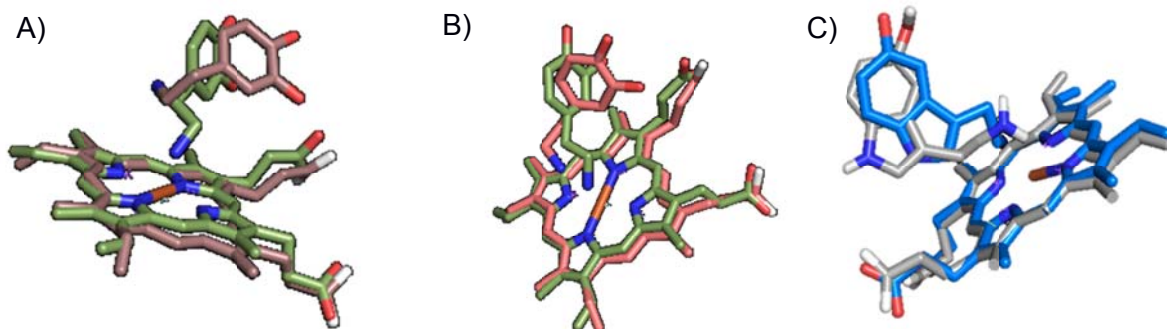
Todas las estructuras con excepción de 1BVY, encontraron un mínimo local después de los procesos de minimización (Figura 10) y llegaron a un mínimo local utilizando el método de descenso acelerado que convergió en energía con el de gradiente conjugado. Esto nos demuestra nuevamente que la proteína tiene un núcleo rígido y que no se modifica en gran medida. Durante este tratamiento la estructura perdió el grupo hemo por lo tanto este tiene que ser agregado de manera manual y realizar un proceso de equilibrio nvt para llevar a nuestra proteína a una temperatura ambiente (25°C) y posteriormente un npt para trabajar a una presión aproximada de 1 atm. Cada uno de estos procesos se llevó a cabo durante 100 ps, el termostato utilizado fue el V-rescale y el baróstato Parinello-Rahman. Para esta parte del análisis se realizó un análisis de los campos de fuerza disponibles para GROMACS que tuvieran parametrizado el grupo hemo y que tuvieran disponible el enlace cisteína-hierro presente en estos citocromos, por lo cual se decidió utilizar el campo de fuerza GROMOS53a6 y CHARMM36 para ver cuál era de más utilidad para mi sistema.

Una vez preparadas las estructuras se hicieron de nuevo los protocolos de acoplamiento revisando los confórmeros con LigPlot+ y analizando las distancias (Anexo 6). Tomando en cuenta estos resultados observamos que el campo de fuerza de CHARMM36 (Figura 11) nos daba resultados más reproducibles y un valor de RMSD menor al comparar con estructuras que ya habían sido cristalizadas (Tabla 1).

Contactos observados		
GROMOS 53A6	CHARMM 36	ADT
L75	F87	S72
A264	A264	L75
A268	A268	F87
A330	G328	I263
F331	P329	A264
L437	F331	T268
V438	L437	L437
	V438	V438

**Tabla 1. Contactos en diferentes campos de fuerza**

Comparación de contactos observados con ácido p-cumárico como ligando utilizando dos campos de fuerza y los resultados obtenidos antes del refinamiento.



**Figura 11. Re-acoplamiento molecular después de las estructuras seleccionadas.**

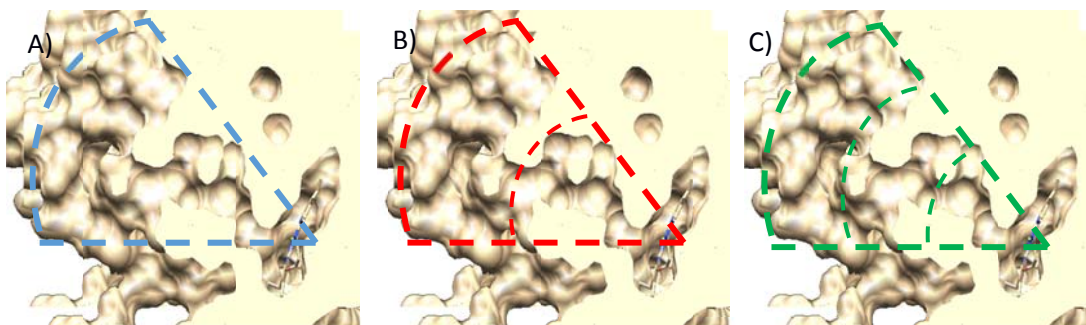
Resultado de proceso de validación por reacoplamiento con (A) dopamina utilizando el campo de fuerza GROMOS53a6 y (B) dopamina (C) serotonina utilizando el campo de fuerza CHARMM36. En verde (B, C) y blanco (D) se muestra la posición reportada en estructuras cristalográficas.

Tomando en cuenta los resultados del acoplamiento y los resultados de ensayos de actividad en el laboratorio en los que observamos que la enzima 4E10 no utilizaba al ácido p-cumárico como sustrato, llegamos a la conclusión de que probablemente lo que estábamos observando era que nuestro sustrato tenía un buen modelo de unión con probabilidad de ser utilizado como sustrato sin embargo estaba limitado en la entrada, y por alguno impedimento estérico dentro del sitio activo, no llegaba a las conformaciones encontradas en el acoplamiento.

Tomando esto en consideración decidimos hacer un análisis del sitio activo midiendo las distancias del C3 que es el que buscamos hidroxilar hacia el hierro y ver todos los modelos de unión que se podrían observar a lo largo de la hendidura catalítica. Para lograr esto, medimos la distancia de todo el sitio activo con una profundidad de 30 Å y de nuestro ligando el ácido p-cumárico 9 Å y diseñamos 6 tamaños de cajas que involucraran solo el sitio activo pero que posicionaran al ligando a diferentes distancias del hemo, hicimos tres cajas partiendo el sitio activo en tres regiones de 10 Å, dos cajas partiendo el sitio activo en regiones de 15 Å y una última tomando el sitio activo completo (Figura 12).

Cada procedimiento de acoplamiento se realizó por triplicado y usando cuatro proteínas (PDB 1BVY, 2IJ2, 4DUB y 4DUF), medimos las distancias de los carbonos con un script en Python y construimos una tabla en la que podíamos evaluar las energías, distancias y todos los contactos observados (Anexo 7 y 9). Con esta información buscamos todos los contactos que se encontraron en las

diferentes cajas donde observamos 62 residuos interaccionando, tomamos todos aquellos que se encontraran representados más del 5% en cada tratamiento y disminuimos este número a 35 residuos importantes en la interacción, para reducir este número comparamos la contribución total de cada residuo en todos los tratamientos sumando las contribuciones individuales observadas en cada tratamiento y seleccionamos 23 residuos. Analizando estos contactos conservados se organizaron en 8 grupos y se separaron por regiones para ser evaluados (Tabla 3).



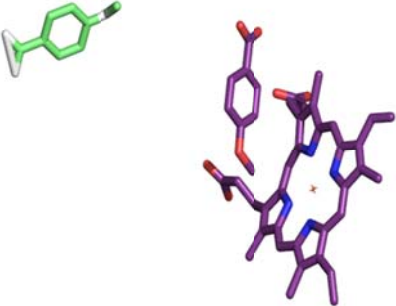
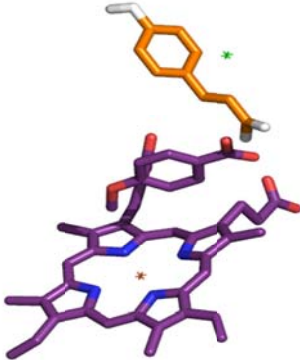
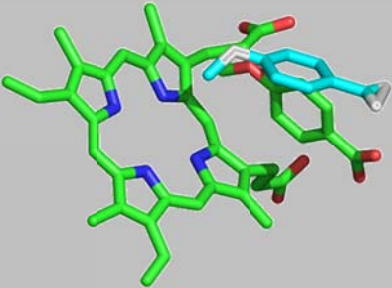
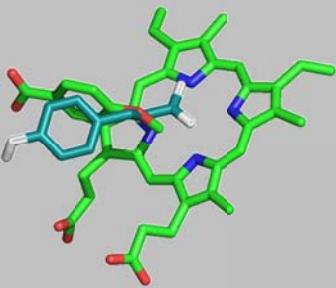
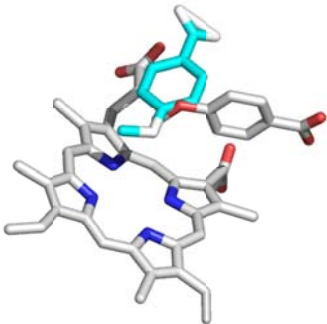
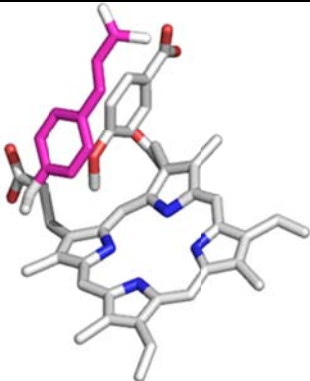
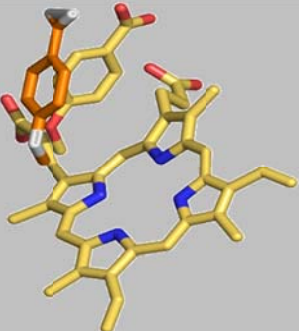
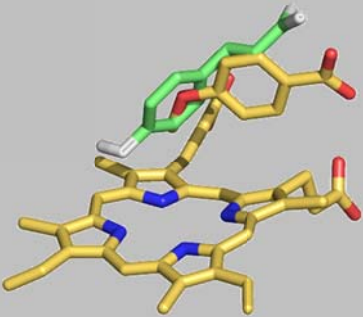
**Figura 12. Análisis del sitio activo.**

Corte de la hendidura catalítica en la que se diseñaron cajas para el acoplamiento a diferentes distancias (A) 30 Å (B) 15 Å (C) 10 Å.

A manera de control, sometimos la estructura del CYT 199A2 (Figura 13) que había sido mutado para realizar la síntesis de ácido cumárico y ácido cafeico a nuestro algoritmo de refinamiento y acoplamiento, para ver si observamos una mejora en las energías de unión con las mutaciones probadas experimentalmente. Para este análisis utilizamos una estructura sin ligando (PDB 2FR7), una estructura con ligando (PDB 4DNJ) y creamos la mutación F185L *in silico* en ambas estructuras ya que fue la mejor variante que encontraron después de hacer evolución dirigida.

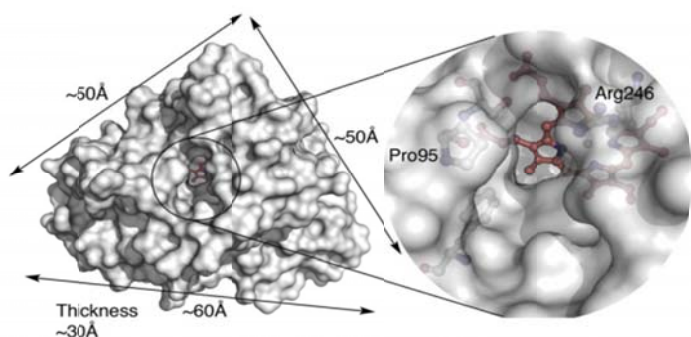
Analizando los resultados del acoplamiento molecular pudimos ver que en el análisis tenemos un confórmero muy similar al reportado en la estructura cristalográfica como se observa en la tabla 2 y anexo 10. Al evaluar los mejores confórmeros de la estructura sin ligando y con ligando podemos ver que esta estructura no permite la unión cercana al sitio activo en el caso de la estructura sin ligando lo que nos indica que la unión depende de un posible ajuste inducido al interactuar con el ligando, de esta manera pasamos a crear la mutación “*in*

Tabla 2.- Resultados de reacoplamiento en CYT199A2

PDB	Ácido 5-metoxibenzoico	Ácido p-cumárico
2FR7		
4DNJ		
2FR7_F185L		
4DNJ_F185L		

*Silico*” F185L para ver si podíamos demostrar la mejora en actividad que ya se había reportado.

El resultado de la mutante F185L nos permite concluir que este residuo modifica de manera directa la interacción con el sustrato y al parecer mantiene al sitio catalítico abierto para la entrada directa del sustrato permitiendo encontrar conformeros dentro del sitio activo tanto en la estructura con ligando como en la que no tiene ligando. De manera interesante, esta modificación permite la entrada al ácido p-cumárico orientado para una hidroxilación en el anillo aromático, caso contrario en la estructura 4DNJ que lo mantiene en otra orientación. Este resultado demuestra que nuestro protocolo de refinamiento y acoplamiento funciona y puede dar explicación a datos experimentales reportados. Por lo que hicimos un acoplamiento en el dominio reductasa de la proteína (Figura 15) en el que no observamos sitios de unión preferenciales por el ligando.

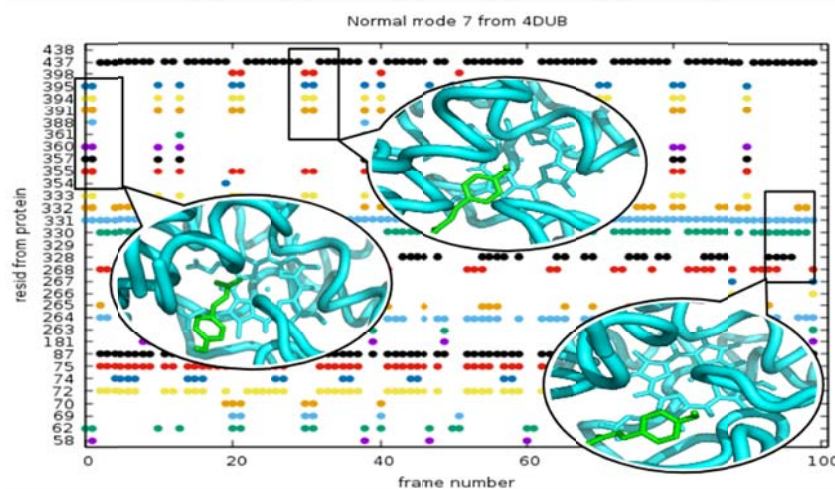
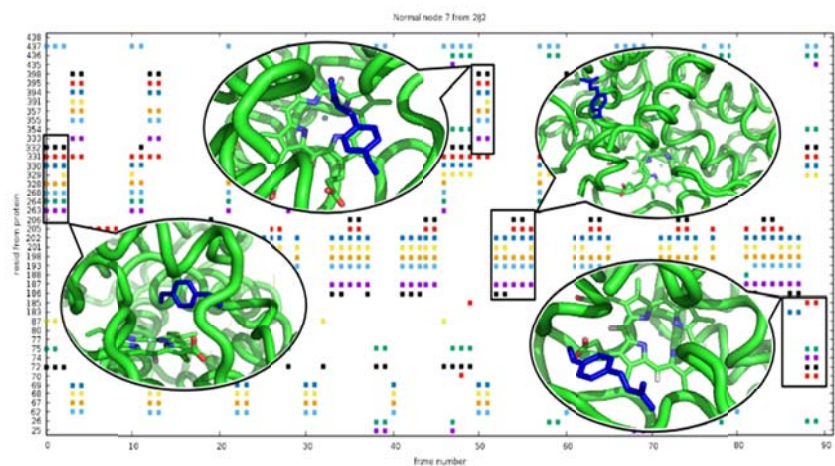


**Figura 13. Representación tridimensional de la topología del CYT 199A**

No todos los citocromos presentan una hendidura catalítica en la que se pueda ver el grupo hemo (en rojo) directamente, el 199A y el BM3 presentan esta cualidad.

Junto con estos resultados, surgió la duda de si se puede observar este ajuste inducido en los movimientos generales de la proteína. Para abordar esta pregunta realizamos un análisis de modos normales en el servidor en línea El nemo, descartamos los modos triviales (1-6) que corresponden a movimientos traslacionales y nos quedamos con los primeros 5 no triviales (7-11). Cada uno de estos 5 modos normales está compuesto por 11 estructuras que corresponden a cambios en la amplitud de 100 a -100 en intervalos de 20, tomamos todas las estructuras por separado y fueron analizadas utilizando el protocolo de acoplamiento y se seleccionaron los 10 mejores conformeros de cada estructura para hacer un análisis en el que se muestre la fluctuación de los residuos que se encuentran en contacto con el sustrato en los diferentes modos normales (Figura 14).



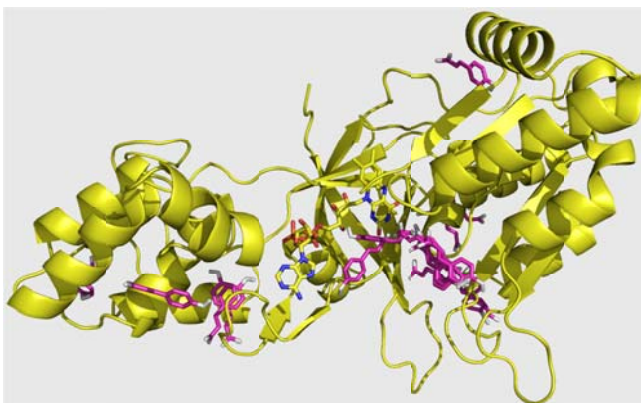
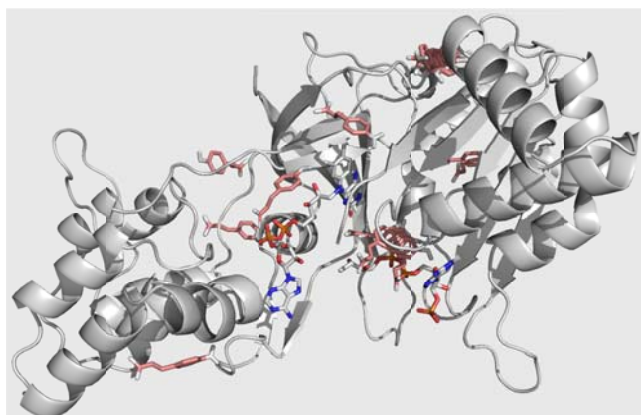


**Figura 14. Análisis de modos normales.**

A) Contactos del ligando en la proteína del resultado del acoplamiento molecular en el modo normal 7 de la estructura en conformación cerrada (PDB 1JPZ).

B) Contactos del ligando en la proteína del resultado del acoplamiento molecular en el modo normal 7 de la estructura en conformación abierta (PDB 4DUB).

La imagen en cada círculo corresponde a la posición tridimensional del ligando, se encuentran seleccionados con un cuadrado el grupo de contactos correspondientes a cada imagen.



**Figura 15. Acoplamiento molecular del dominio reductasa.**

- A) Reductasa sin NADPH
- B) Reductasa con NADPH co-cristalizado
- C) Dominio de unión FMN

## 7.2 Selección de sitios a mutagenizar

Para evaluar la contribución de los residuos seleccionados en el procedimiento de acoplamiento se decidió hacer una comparación de secuencia y estructura con citocromos que se ha reportado previamente que pueden utilizar ácido p-cumárico como sustrato. Con este propósito se seleccionó a la enzima de su ruta biosintética en plantas CYT98A3 (sin estructura cristalográfica disponible) y al citocromo bacteriano CYP199A1 (PDB 2FR7) que se ha utilizado antes para la síntesis de ácido p-cumárico y cafeico. Para lograr el análisis estructural realizamos un modelo de CYT98A3, tomando en cuenta que este tipo de enzimas tienen una similitud estructural alta a pesar de tener una identidad baja en secuencia. Dado que las estructuras de los citocromos se encuentran altamente conservadas y que se cuenta con moldes adecuados, se decidió utilizar modelado por homología utilizando I-TASSER (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) para hacer un modelo del CYT98A3.

Ya con las estructuras listas se realizaron alineamientos de secuencia y estructurales para evaluar las zonas en las que se localizaban los residuos del acoplamiento (Figura 16). Al revisar estas regiones observamos que el grupo A que comprende los residuos 20-25-26 se encontraba altamente conservado tanto en secuencia como estructura y era preferible no hacer modificaciones en esas regiones de la proteína.

El grupo B (residuos 42-47-49-51) se encuentra en un asa de la proteína en la parte externa del sitio activo y presenta una orientación diferente al de las proteínas tomadas como referencia, analizando la orientación de los residuos que se encontraban cercanos a esta región se puede observar que existe una distribución de puentes de hidrógeno diferentes y esto da como resultado una apertura diferente en esta región, apoyándonos con la información reportada sobre la importancia de estos residuos en la catálisis observamos que el R47 es el más importante de esta región ya que actúa como una puerta para la entrada de sustratos por lo cual fue seleccionado para la mutagénesis.

El grupo C (residuos 72-75) se encuentra, de igual manera en la entrada del sitio activo e interesantemente se encuentra interaccionando con el asa formada por el grupo B y se ha observado que la modificación del residuo R47 puede afectar esta región en gran magnitud, por lo tanto, no seleccionamos ninguno de estos residuos para modificar.

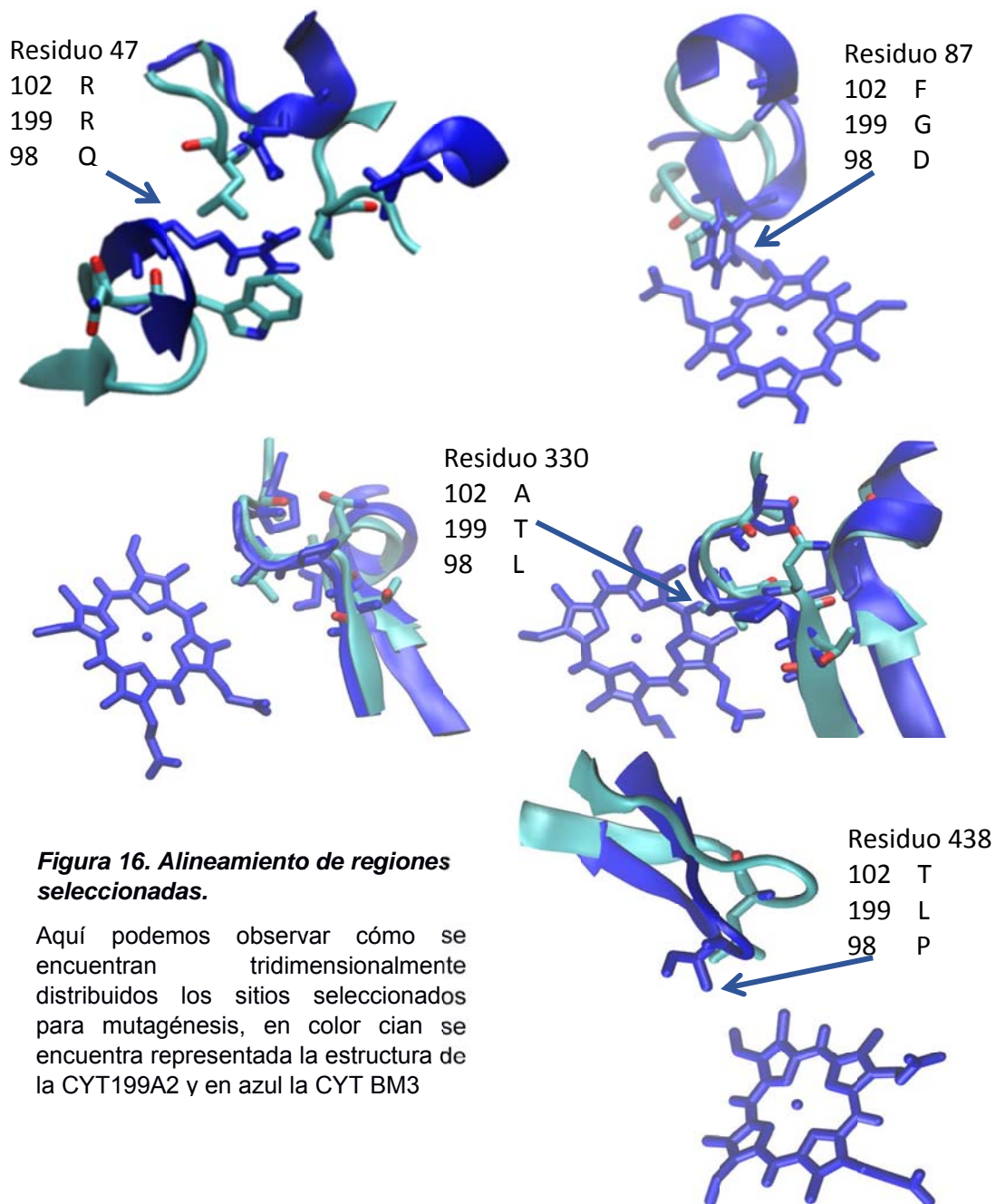
En el caso del grupo D que solo considera el residuo 87, observamos que tiene una gran variación en secuencia y que desempeña un papel muy importante para la catálisis; en la proteína silvestre este residuo se encarga de dar orientación y de estabilizar al sustrato cerca del grupo hemo, por esta razón este residuo fue seleccionado para probar diferentes sustituciones.

Para el grupo E (residuos 264-268) no observamos cambios en estructura ni secuencia por los que asumimos que esta región se ha conservado para mantener la catálisis de este tipo de moléculas y esta región no fue seleccionada. Los reportes sobre mutantes de esta región no presentan cambios representativos sobre la catálisis.

Los residuos del grupo F (328-329-330-331-332) se encuentran en una hoja beta de la proteína que se encuentra localizada en la misma región en las 3 proteínas. A nivel secuencia no hay una clara conservación y al analizar la orientación de las cadenas laterales podemos observar que se encuentran orientadas hacia afuera del sitio activo en el caso de BM3 mientras que para el 199A, hacia adentro. De acuerdo a los reportes en la literatura, el residuo A330 es un residuo clave para la especificidad por sustrato. Su sustitución por aminoácidos más grandes permite que la enzima acepte sustratos pequeños, así que mutantes en esta región podrían modificar la orientación de los residuos cercanos por lo que se seleccionó para modificarlo posteriormente. En el caso del Grupo G solo se observó el residuo 354 interactuando con el ligando, el cual no se encuentra conservado en secuencia y no se encuentra modificado a nivel de estructura. No se encontró información disponible sobre el efecto de la catálisis al modificar este sitio, por lo que no se seleccionó para su mutación.



Por último, el grupo H (residuos 435-436-437-438) se encuentra en un asa cercana al grupo hemo; en este grupo, el residuo A438 tiene una distancia de 10 Å con el Hierro del grupo hemo y se ha reportado que determina la orientación del sustrato a hidroxilar. En el alineamiento estructural del asa, el CYT-BM3 se encuentra en una orientación diferente al CYT98A3 y CYT199A1. Ya que el residuo A438 se encuentra en el inicio de esta asa, creemos que podría modificar la orientación del asa y al jugar con la densidad electrónica en esta región de igual manera modificar la orientación del sustrato.



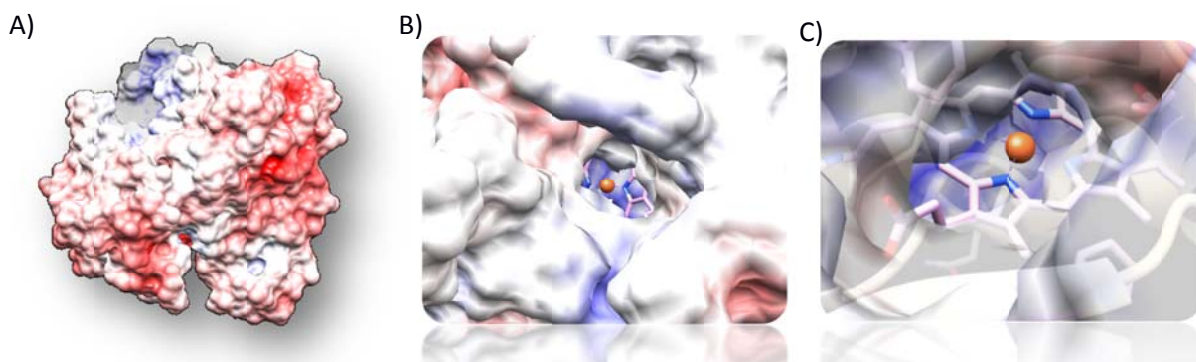
**Figura 16. Alineamiento de regiones seleccionadas.**

Aquí podemos observar cómo se encuentran tridimensionalmente distribuidos los sitios seleccionados para mutagénesis, en color cian se encuentra representada la estructura de la CYT199A2 y en azul la CYT BM3

Junto con estas modificaciones, el citocromo con el que trabajamos en el laboratorio tiene dos modificaciones en el sitio activo A328V y A82L que fueron realizadas con el propósito de utilizar una molécula pequeña como el dimetiléter. Dado que estas regiones se encuentran muy cerca de las regiones seleccionadas, se decidió seleccionarlas para su mutación.

Se realizó una medición de la distancia de cada uno de los residuos seleccionados al hierro del grupo hemo y en base a las observaciones en el alineamiento estructural y en conjunto con un análisis del potencial electrostático (Figura 17), el cual nos muestra la carga parcial presente a diferentes zonas del sitio activo, seleccionamos las variantes a explorar, en las que buscamos modificar la densidad electrónica en diferentes regiones y gracias a que tenemos residuos a diferentes orientaciones del hemo, hipotetizamos que la combinatoria de estos residuos permitirá dirigir la orientación del sustrato en el sitio activo (Figura 18).

Tomando en cuenta toda la gama de sustratos con los que se ha probado el CYTBM3 realizamos una tabla en la que comparamos las mutaciones realizadas a la proteína y como éstas han modificado su catálisis o especificidad, tomando en cuenta las características del ácido p-cumárico para ayudar a la selección de residuos para mutagénesis (Anexo 11).

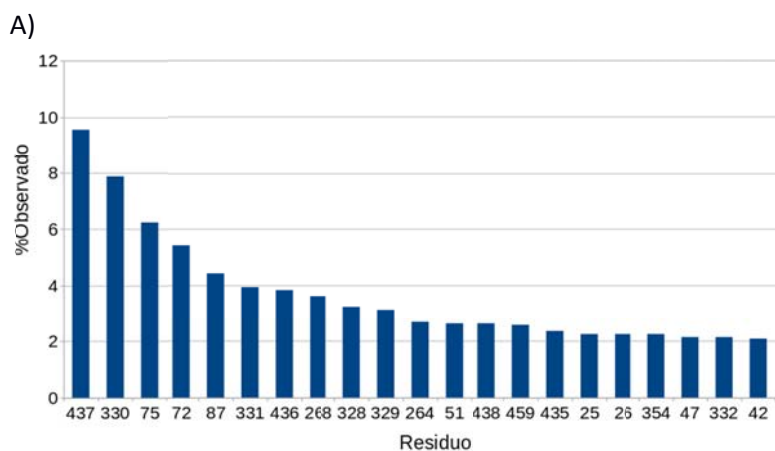


**Figura 17. Análisis de potencial electrostático.**

Análisis de APBS (A) proteína completa (B) entrada al sitio activo (C) región proximal al grupo hemo. Las cargas positivas están representadas con una superficie de color rojo, las cargas negativas de color azul y las superficies neutras con un color gris.

**Tabla 3.- Selección de residuos para realizar mutagénesis**

Grupo	Residuos	%	Alineamiento secuencia	Alineamiento estructural	Información
H	435/436/437/ 438	46	Semi conservados	Se encuentra en un loop con mayor ángulo que los otros	Dirección de sustrato
G	354	13	Variable	Ligero movimiento en la estructura	Sin reportes
F	328/329/331/ 330/329/328	35	Semi conservados	Se encuentra en una hoja beta, diferente orientación	Sustratos pequeños
E	264/268	19	Conservado	Sin cambios	No afectan la catálisis
D	87	23	Variable	Diferentes distancias del hemo	Interacción cerca del hemo
C	72/75	33	Variable	Ligero movimiento en la estructura	Sin reportes
B	42/47/49/51	17	Semi conservados	Se encuentra en un loop e interacciona con hélices	Entrada de sustratos
A	20/25/26	14	Conservado	Ligero movimiento en la estructura	Sin reportes

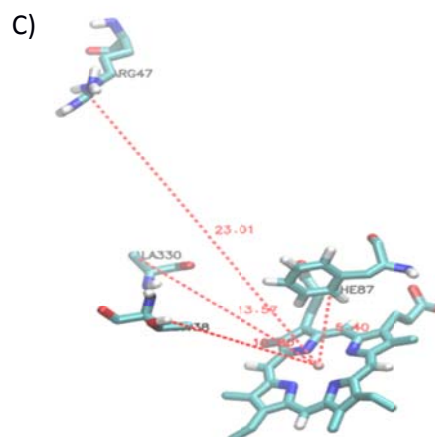


B)

Residuo	Variantes
47	R S
82	A S G
87	F A G W
328	A F V
330	A P T K
438	T Y N L V

**Figura 18. Residuos seleccionados para mutagénesis.**

(A) Gráfica de barras de los residuos más representados en el protocolo de acoplamiento (B) Tabla de los residuos seleccionados y las variantes a explorar en cada una de las posiciones (C) Distancias de los diferentes residuos seleccionados del átomo de hierro en el hemo.



### 7.3 Construcción de biblioteca de mutantes

Los residuos seleccionados con las variantes a explorar se presentan en la tabla 3 para crear la biblioteca se utilizó la técnica de megaprimer en la que se generaron oligonucleótidos degenerados para cada posición en los que incluimos los residuos de la proteína silvestre y los seleccionados para explorar (Tabla 4). La construcción se realizó siguiendo el siguiente protocolo (Figura 19).

Residuo		Primer
R47	←	agcgcgttac(c/g)(c)(t)accaggcgcc
F87	←	ccagcttgt(c/g)(a/g/c)(a/c)taaccctctcc(t)(g/a)(a/c)aaaatcacg
A328/A330	→	ggccaact(t/g)(t/c)ccct(a/c/g)(c/a)attttccct
A328/A330	←	agggaaaat(g/t)(t/g/c)aggg(a/g)(a/c)agttggcc
T438	→	aactacgagctcgatattaagaaactta(t/a/g)(c/a/t)(c/g)ttaaaccctg

Tabla 4. Oligonucleótidos sintetizados para construcción de biblioteca

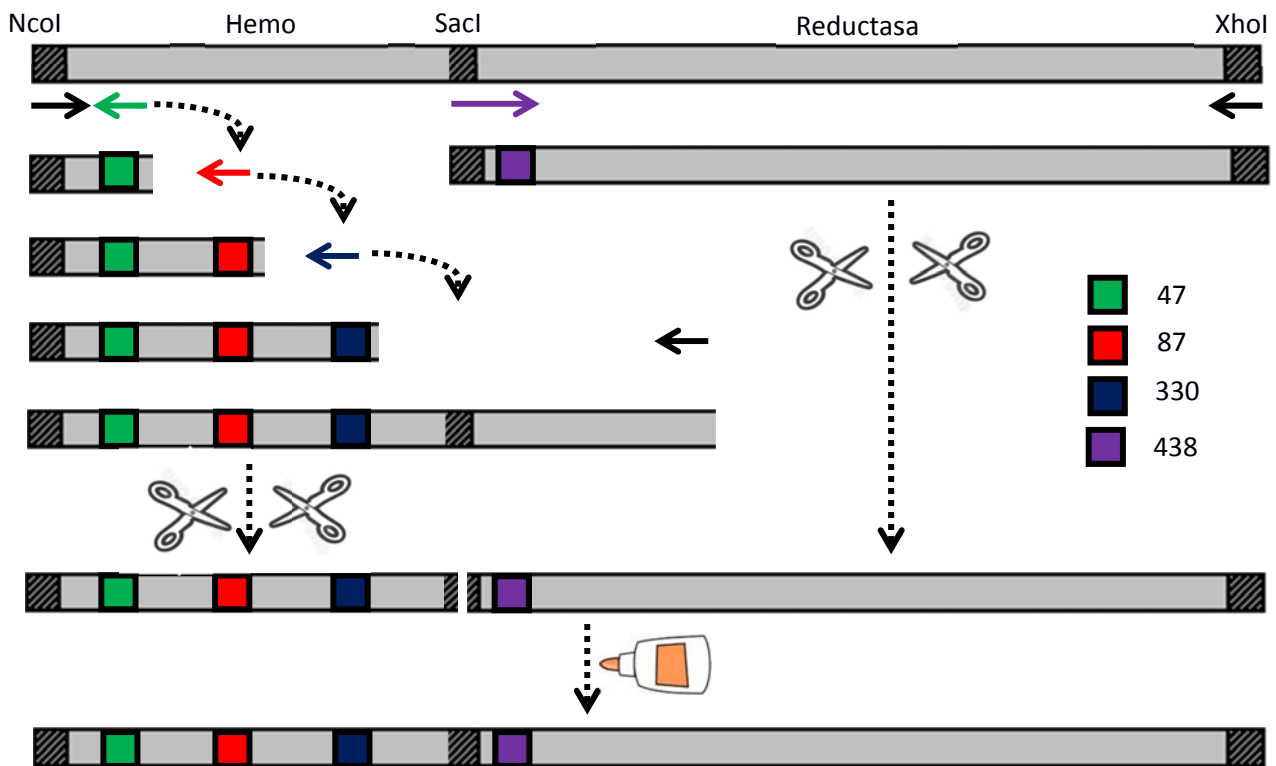
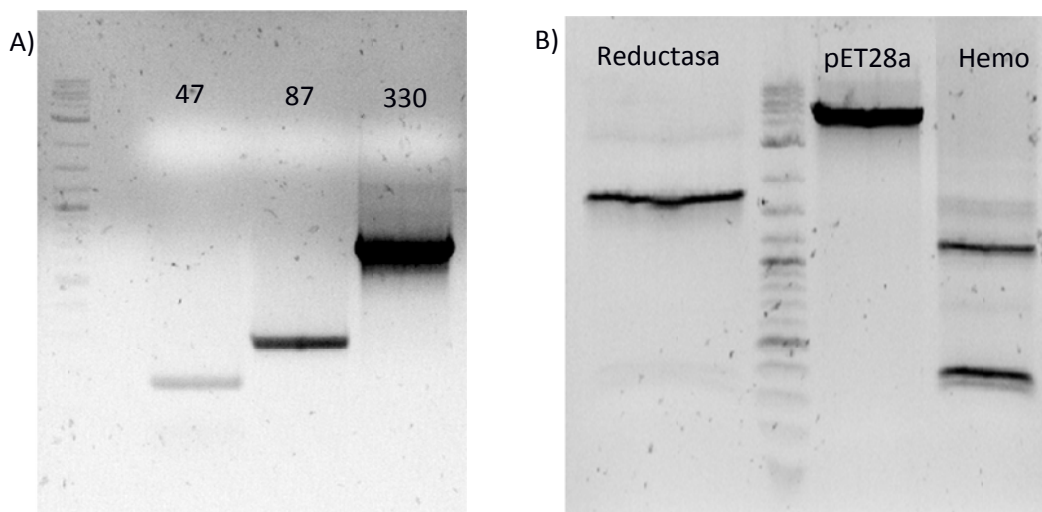


Figura 19. Metodología empleada para la construcción de la biblioteca de mutantes

Una vez construida la biblioteca se corroboró mediante secuenciación de DNA una mezcla de varias mutantes en la que pudimos observar diversidad en las regiones seleccionadas.

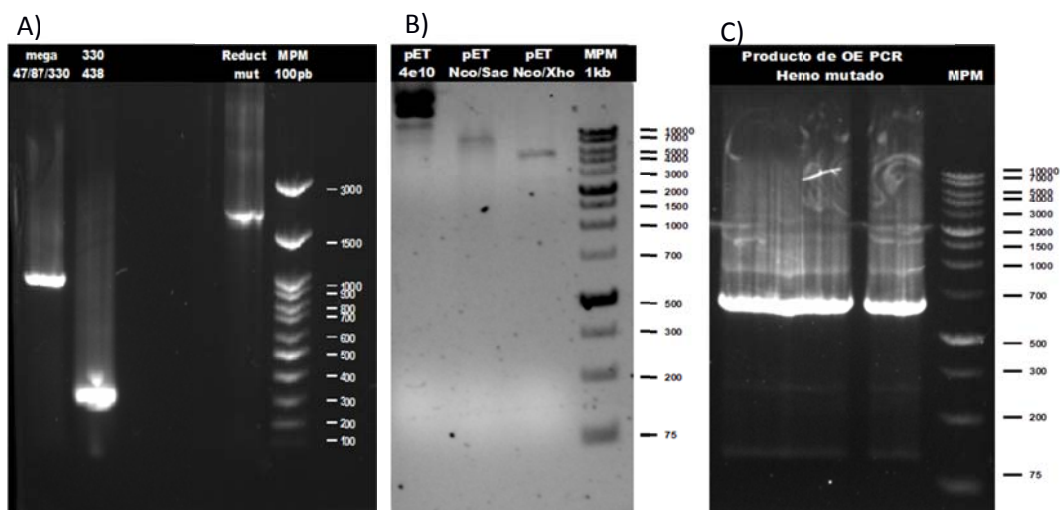
La primera estrategia que abordamos para realizar la construcción de la biblioteca que incluyera todas las mutaciones y sus combinatorias fue mediante megaprimer, sin embargo observamos que tras cada paso de amplificación se veía disminuido el rendimiento y nuestros intentos de enriquecer el fragmento final nos hacía perder diversidad. Por lo tanto, decidimos apoyarnos de otras técnicas en conjunto con el megaprimer para lograr una construcción con alta diversidad y buen rendimiento.

Para la construcción final decidimos trabajar el dominio hemo en dos regiones, utilizamos la técnica de megaprimer para generar una región en la que se encontraran las mutantes y se generó un fragmento pequeño para completar el dominio hemo y al mismo tiempo incluir una región de homología con el reductasa. Una vez obtenidos estos fragmentos se utilizó la técnica de OE PCR para generar un fragmento correspondiente al dominio hemo donde se incluyeran las mutaciones 47, 82, 87, 328, 330. La figura 20 muestra los productos de PCR generados a lo largo de la construcción de la biblioteca.



**Figura 20. Ensamble de combinatoria de variantes por el método de megaprimer.**

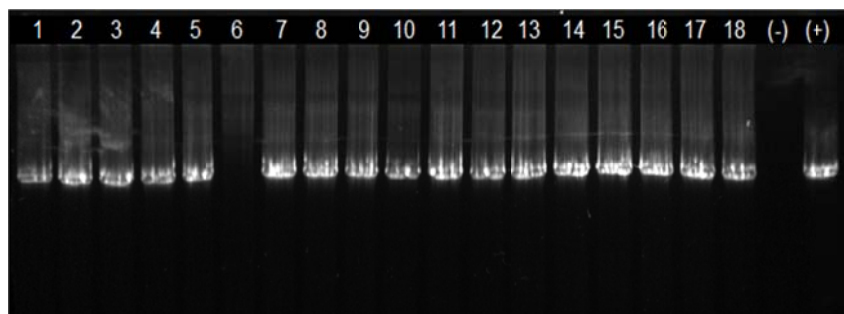
(A) Fragmentos 47, 87 y 330 (B) Digestión de fragmento de dominio reductasa *SacI/XhoI*, pET28 *NcoI/XhoI* y dominio hemo con codones degenerados *NcoI/SacI*.



**Figura 21** Ensamble de fragmentos por OE-PCR.

A) Productos de PCR de los fragmentos correspondientes al hemo y reductasa B) Plásmido sin digerir y digestión de plásmido con enzimas Nco/Sac y Nco/Xho C) Producto de OE PCR del fragmento hemo mutado.

A la par se trabajó con el fragmento del dominio reductasa en el cual se incluye la mutante 438. Para esta mutante se digirió el producto de su amplificación por PCR y el vector pET28 con las enzimas de restricción SacI y XhoI para proceder con una ligación y la transformación subsecuente de células MC1061 por electroporación. La figura 21 muestra los fragmentos obtenidos durante la construcción del gene completo. Se seleccionaron las colonias transformadas en cajas de medio LB con kanamicina y se purificaron plásmidos de 18 de ellas para utilizar en la construcción de la biblioteca completa. Se corroboró por PCR la presencia del gen correspondiente al dominio reductasa en 17 de las 18 colonias seleccionadas, como se muestra en la figura 22.



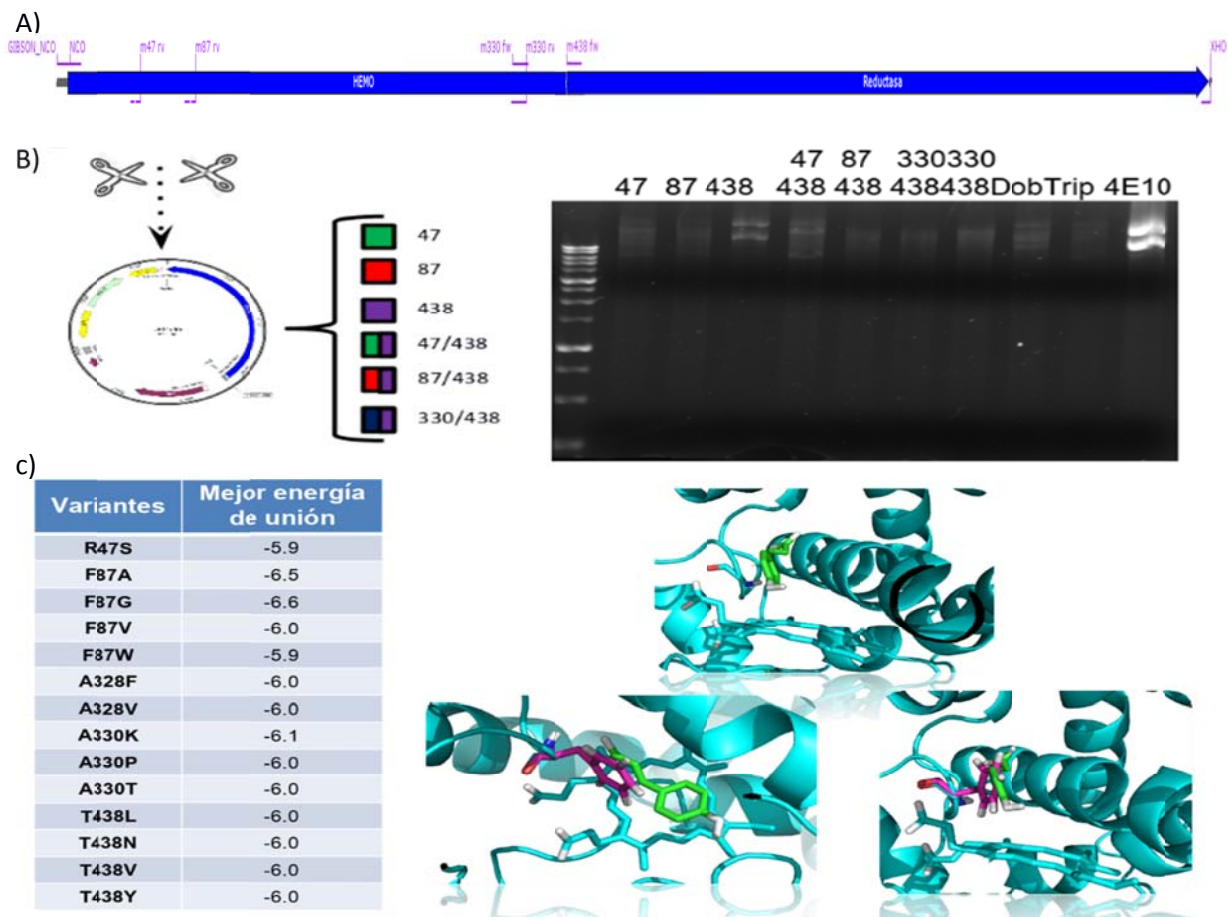
**Figura 22.** PCR de plásmido purificado de colonias transformadas con las variantes en 438.

Una vez obtenidos el fragmento mutado del dominio hemo (mHemo) y la construcción del fragmento reductasa pETmReduct, se utilizaron las técnicas de



Gibson assembly y CPEC para obtener la construcción completa con las mutaciones incluidas. Después de varios intentos con la técnica de Gibson assembly y tras meter una serie de controles nos dimos cuenta que esta tecnología no parece ser útil para la construcción de bibliotecas grandes, de igual manera se intentó construir las mutantes sencillas, dobles y triples utilizando la técnica convencional de digestión y ligación (Figura 23) en este proceso obtuvimos un número alto de religado pero tras evaluar las variantes pudimos identificar diferencias entre la silvestre con nuestro método de tamizaje.

Se diseñaron oligos para la amplificación del vector y crear homologías con nuestro fragmento utilizando las regiones T7 del vector pET28 y se procedió a realizar CPEC. Tras un par de intentos se obtuvieron librerías grandes ( $10^3$ - $10^4$ ), de las cuales se recogió el total de las colonias transformantes, se purificó el plásmido de la combinatoria y se mandó secuenciar. Como resultado de la



**Figura 23. Construcción de variantes *in silico* e *in vitro*.**

A) Fragmento hemo con los primers diseñados en esta región B) Construcción de variantes sencillas, dobles y triples con su plásmido purificado C) Análisis de acoplamiento de variantes *in silico*.

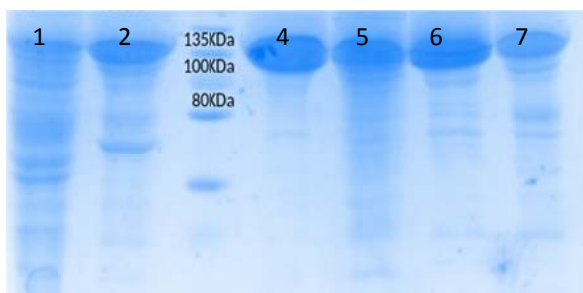
secuenciación se observaron mutagenizados los sitios 47, 82, 87, 328, 330 y 438. Se seleccionaron 400 colonias para su análisis.

De manera alternativa, se decidió construir librerías de las variantes sencillas propuestas *in silico*, para ver si se obtenían cambios en la afinidad por el sustrato, para lo que observamos que obtuvimos conformeros con mejores energías de interacción que en la silvestre subiendo el rango de entre 4.5-5.5 a 5.5-6.6 en el mejor de los casos, F87G.

#### 7.4 Expresión y purificación de P450 BM3

La expresión y purificación del citocromo P450 BM3 se realizó siguiendo el procedimiento previamente descrito. Las eluciones se recolectaron cada 3 mL y se seleccionaron aquellas en las que se observó un aumento en la absorción en el cromatograma, tomando preferencialmente aquellos con coloración marrón; se realizó una electroforesis en gel de poliacrilamida en condiciones desnaturizantes (SDS-PAGE 12%, figura 24) de las muestras seleccionadas para determinar la pureza de los extractos y se realizó una cuantificación de proteína de las fracciones consolidadas mediante el método de Bradford.

Para el ensayo de actividad se tomaron las fracciones en las que se observó la proteína y se dejaron dializando durante 12 hrs a 4°C en amortiguador PO<sub>4</sub> pH7.4, una vez transcurridas las 12 horas se siguió el protocolo previamente descrito.



**Figura 24. SDS-PAGE de extractos purificados de citocromo.**

1.- CYP A, 2.- CYP B, 4 y 5.- CYP MutM, 6 y 7 CYP 4E10. 1, 2, 4 y 5 son variantes que se han trabajado en el laboratorio y tienen mutaciones en el dominio reductasa.

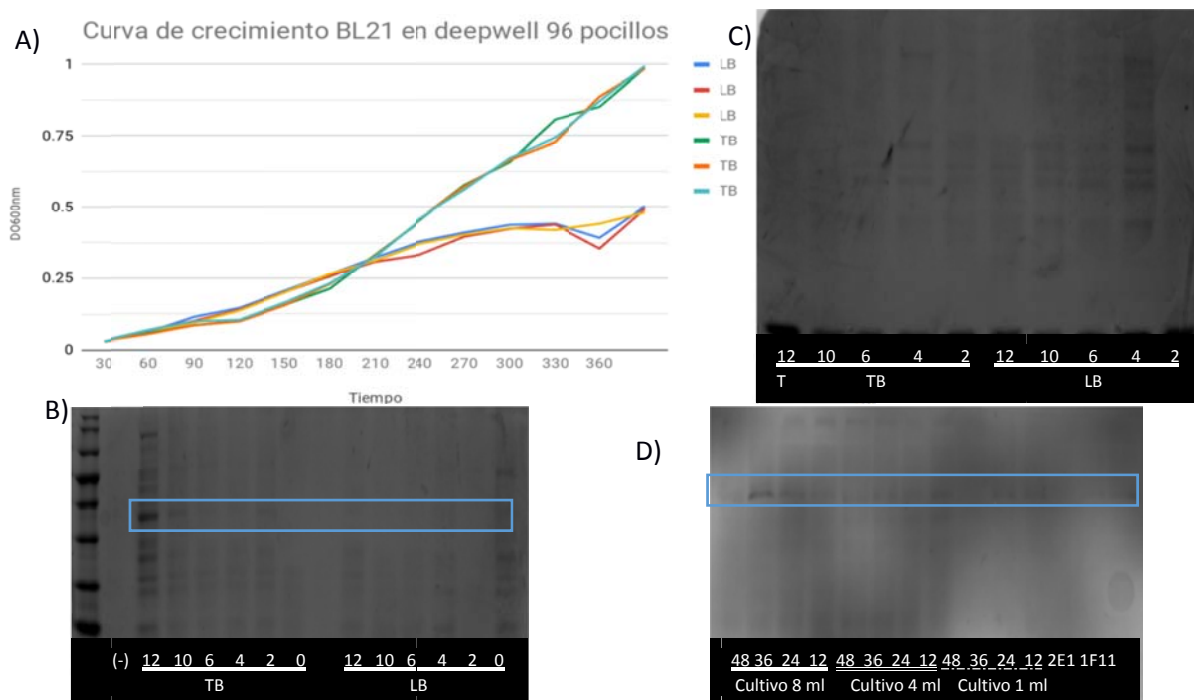
#### 7.5.1 Protocolo de expresión en BL21 DE3

Las condiciones de aireación y agitación en placas deep-well de 96 pocillos son diferentes a las obtenidas en tubo o matraz, por lo tanto realizamos curvas de crecimiento para ver cómo se comportaría la cepa de expresión *E. coli* BL21 DE3 y lograr estandarizar el protocolo de expresión de la proteína en medio LB y Terrific Broth. Como podemos observar en la figura 25 A, la cepa llega a



densidades de 0.4-0.6 óptimas para realizar la inducción en aproximadamente 4 horas, el crecimiento observado en medio TB fue mucho mayor al obtenido en LB probablemente debido a la diferencia en cantidad de fuente de carbono. Tomando en cuenta estos resultados pasamos a hacer cinéticas de expresión con la proteína silvestre en las que realizamos inducción a las 4 horas de cultivo y se tomaron muestras a las 2, 4, 6, 8, 10, 12 y 18 horas. Las muestras fueron centrifugadas y se trataron las pastillas con 500 mcl de amortiguador de lisis por una hora a 37°C, al finalizar la incubación se centrifugó y se utilizó el sobrenadante como extracto proteico.

Procedimos a realizar SDS-PAGE para ver la cantidad de proteína con respecto al tiempo de inducción en medio LB y TB. Los resultados indican que la proteína se encuentra sobre expresándose a las 12 horas en medio terrific broth bajo la inducción de 0.5 mM IPTG (Figura 25 B y C). Para verificar su actividad se realizó una tinción del grupo hemo en la que podemos observar los grupos hemo activos reaccionando con TMB (Figura 25 D).



**Figura 25. Expresión de CYT BM3 en BL21 DE3.**

A) Grafico del crecimiento de la cepa BL21 DE3 durante la expresión de la proteína B) SDS-PAGE de los diferentes tiempos de muestreo en medio TB y LB con inducción C) SDS-PAGE de los diferentes tiempos de muestreo en medio TB y LB sin inducción D) PAGE nativo de los diferentes tiempos de muestreo en medio TB con inducción, teñido con TMB y revelado con peróxido.

## 7.5 Ensayos de actividad

Para los ensayos de actividad se utilizaron las fracciones en las que se observó la presencia de una proteína de alrededor de 119 kDa a pesar de no estar puras, ya que al final se planea utilizar extractos de proteína. Para determinar la actividad del citocromo se midió el consumo de NADPH a 340 nm en la reacción de hidroxilación de SDS, una vez demostrada la actividad se utilizó como sustrato ácido p-cumárico midiendo a las longitudes de onda 230-500 nm al inicio de la reacción y al final; de igual manera, se realizó una cinética de la reacción midiendo a 340 nm el consumo de NADPH.

Para poder observar la contribución de cada uno de nuestros componentes de la reacción al momento de correr el espectro, se analizaron cada uno por separado y cómo podemos ver en el gráfico, el amortiguador y el SDS no contribuyen en nada a la absorción en el rango de 230-500 nm, mientras que para el NADPH podemos observar un primer pico correspondiente a su forma oxidada y un segundo pico alrededor de 340 nm que corresponde a su forma reducida, en el caso del ácido p-cumárico y ácido cafeico encontramos un solapamiento en prácticamente todo el espectro del NADPH con un pico cerca de los 230 nm que podría servir como un punto para diferenciar entre estas moléculas, el segundo pico del NADPH queda justo sobre un declive en la absorción del ácido p-cumárico y ácido cafeico lo que nos permite ver un pequeño cambio a esta región que podemos asociar a un consumo de NADPH (Figura 26).

Una vez descartadas las contribuciones del amortiguador y SDS, se procedió a realizar ensayos de actividad en los que utilizamos como control positivo a la reacción con SDS como sustrato natural de CYT BM3 y como blanco a la mezcla de reacción sin enzima, al realizar el espectro antes de la reacción se puede observar un espectro que representa el solapamiento de nuestros sustratos NADPH y ácido p-cumárico, con una contribución a 340 debido a la presencia de NADPH reducido, al correr la cinética durante 10 minutos podemos observar cómo se consume el NADPH con respecto al tiempo de reacción y que esta longitud de

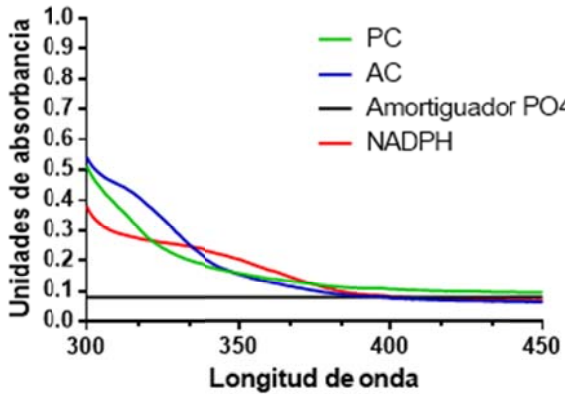
onda es apropiada para la medición de esta molécula al estar en combinación con ácido p-cumárico.

Al monitorear la reacción con la absorbancia a 340 nm podemos determinar el final de la reacción con el NADPH disponible para la reducción del dominio hemo, lo cual observamos que toma alrededor de 10 minutos. Al final se realizó un espectro nuevamente y se hizo la sustracción del espectro obtenido al inicio de la reacción, a manera de poder observar desplazamientos o cambios importantes en los picos resultantes. Al analizar los resultados no vimos cambios representativos en el espectro a pesar de ver un consumo de NADPH durante la reacción. Con estos datos nos dimos a la tarea de realizar una cromatografía en capa fina para ver si observábamos productos diferentes o cambio en la concentración de ácido p-cumárico que nos indicara que este estaba siendo utilizado como sustrato, sin embargo, no se observó una prueba evidente de que se estuviera oxidando el ácido p-cumárico por parte el CYT BM3.

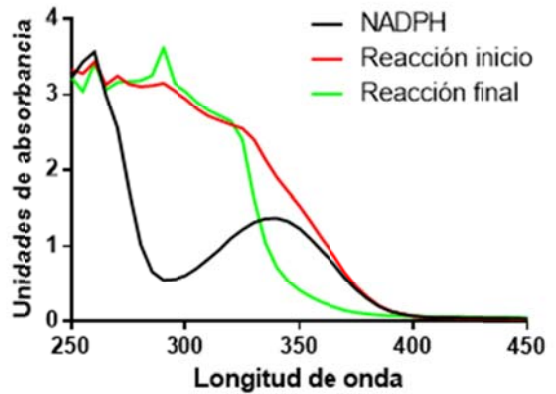
Al realizar la reacción por triplicado se puede observar que aparece un pico cerca de los 290-300 nm lo que podría indicar la formación de un producto por parte del CYT BM3, sin embargo, no se observaron productos diferentes al ácido p-cumárico cuando se realizó la cromatografía en capa fina.

Estos resultados nos pueden estar sugiriendo que la regeneración del sitio activo está siendo ineficiente y tal vez se estén generando especies reactivas de oxígeno como parte de un desacoplamiento del transporte de electrones hacia el hierro, lo que explicaría el consumo de NADPH sin el consumo aparente de ácido p-cumárico. Para verificar esto es necesario hacer un análisis por HPLC y corroborar que no se esté generando algún compuesto que no pueda ser resuelto por la cromatografía en capa fina, una alternativa podría ser buscar la presencia de especies reactivas de oxígeno que indiquen un proceso de desacoplamiento.

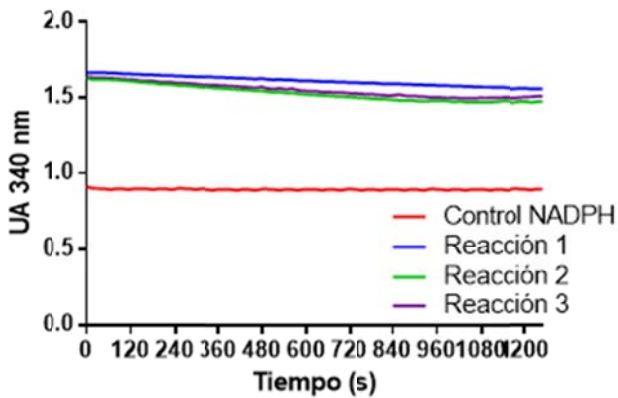
A) **Espectro de componentes de reacción**



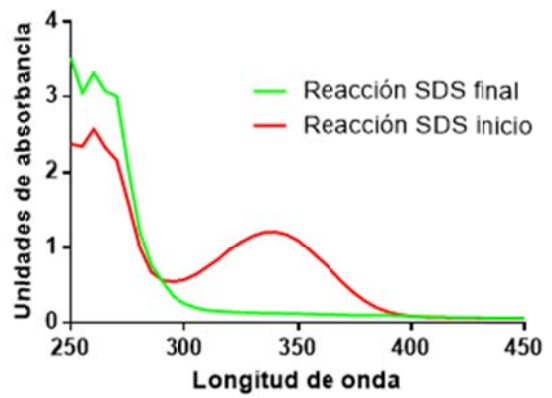
B) **Reacción con ácido p-cumárico**



C) **Cambio de absorbancia durante la reacción**

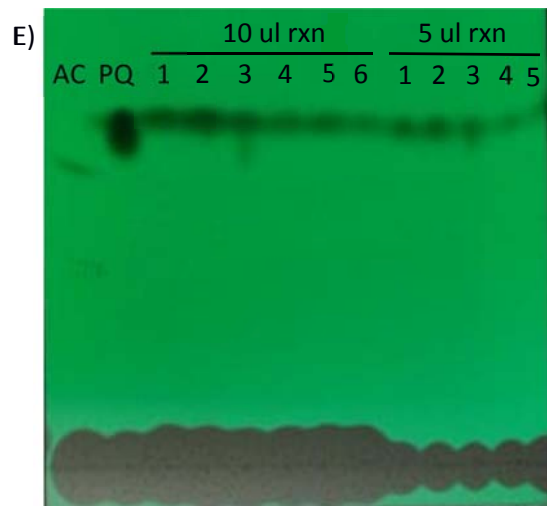


D) **Reacción con SDS**



**Figura 26. Espectro de reacción.**

(A) Espectro de absorción de los componentes de reacción (B) Espectro inicial y final de la reacción con ácido p-cumárico (C) Decremento de la absorción a 340 nm durante la reacción (D) Espectro al inicio y al final de la reacción con SDS (E) Cromatografía en capa fina de ácido cafeico (AC), ácido p-cumárico (PC) y dos concentraciones diferentes de reacciones.



## 7.6 Método de tamizaje de alta eficiencia

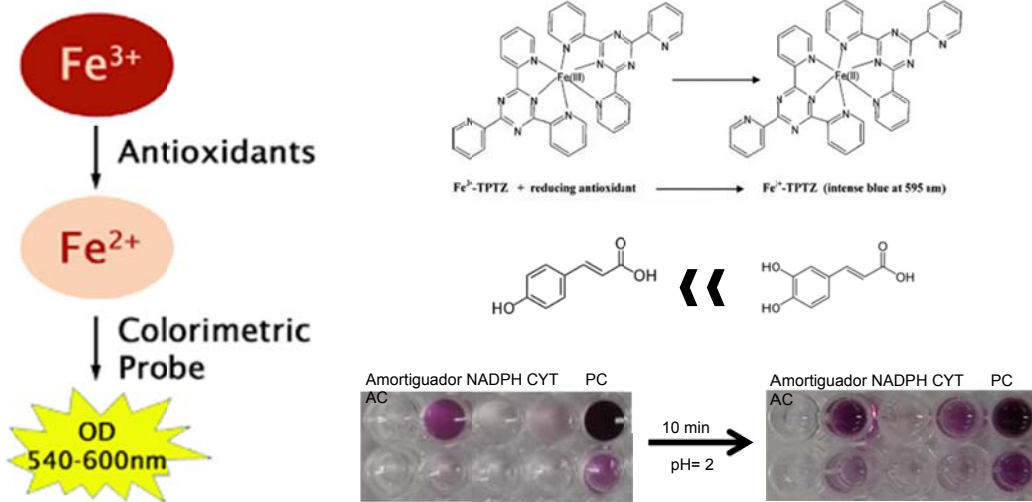
La determinación oportuna de la presencia de ácido cafeico en la presencia de ácido p-cumárico y NADPH se vuelve un proceso sumamente complicado, ya que sus espectros de absorción se sobrelapan en gran medida. Una alternativa es la derivatización con grupos protectores o que reaccionen sobre alguna parte de la molécula, sin embargo, la única diferencia es una hidroxilación en meta por lo que la derivatización suele complicarse y necesita de muchos pasos para obtener una razón de la cantidad de ácido cafeico que hay en la mezcla de reacción. Estas estrategias no resultan eficientes para el análisis sistemático de miles de variantes, por esta razón proponemos que el cambio en la capacidad antioxidante entre el ácido p-cumárico y el ácido cafeico podría ser un método eficiente y barato para el análisis de las variantes de CYT-BM3.

En esta reacción se muestra la capacidad del ácido cafeico de convertir  $\text{Fe}^{+3}$  a  $\text{Fe}^{+2}$  en medio ácido y la mínima contribución del ácido p-cumárico para catalizar esta conversión, esta prueba se puede hacer colorimétrica agregando un agente quelante del hierro en este caso la ferrozina y al estar presente el  $\text{Fe}^{+2}$  observamos la aparición de un color rosado a morado que puede ser monitoreado a 562 nm.

Como podemos observar en la figura 27, en la que evaluamos los componentes de reacción separados, el NADPH contribuye en una parte a la formación de color, sin embargo, al analizar una dilución de la mezcla de reacción a etapas tempranas de la reacción, se puede observar que prácticamente solo el ácido cafeico contribuye a la formación de color (parte inferior de la placa, figura 27). La velocidad de la reacción se puede modificar haciendo menos ácido el medio y permitirá ver el perfil de la producción de ácido cafeico a gran escala.

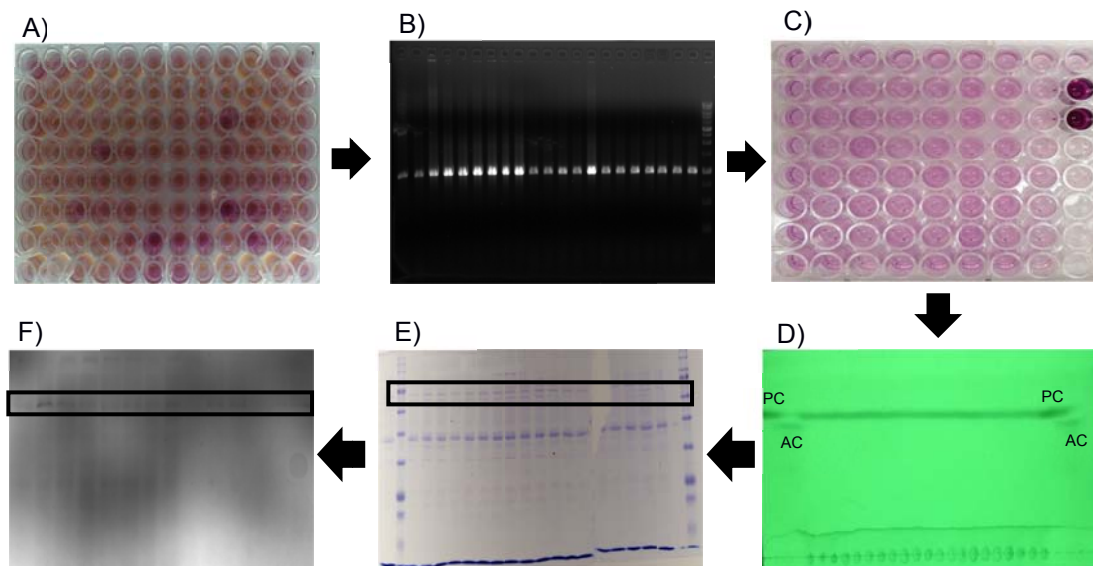
En la figura 28 se ilustra el proceso de evaluación del método de tamizaje en el que se prueba con variantes obtenidas, se seleccionan aquellas con mayor intensidad, se comprueba que contengan el gen y se repite el ensayo para verificar que si hay cambios con respecto a la silvestre. Se realizó una TLC para ver el perfil de productos, sin embargo no se observó producción evidente de

ácido cafeico por lo que corroboramos que la expresión de la proteína se llevara a cabo con un SDS-PAGE y se realizó la tinción del grupo hemo para verificar que fueran proteínas activas. De las 21 seleccionadas se tomaron 3 para analizar el producto de reacción por HPLC.



**Figura 27. Método de tamizaje en formato de alta eficiencia para la selección de las variantes.**

(A) y (B) Representación de la base de la metodología, (C) el mayor que representa la mejor capacidad antioxidante del ácido cafeico sobre el ácido p-cumárico en esta reacción y (D) componente de la reacción analizados por separado. Ácido p-cumárico (PC) y ácido cafeico (AC).



**Figura 28. Protocolo de evaluación de variantes.**

Se seleccionaron 21 variantes de 200 (A), se corroboró que tuvieran el gen de la proteína (B) y se repitió el ensayo de HTS (C). Se realizó TLC (D) y HPLC de las variantes para ver la presencia del producto, observamos la presencia de la proteína con un SDS-PAGE 12% (E) y comprobamos la actividad del grupo hemo con una tinción con TMB (F).

## 7.8 Análisis del perfil de productos por HPLC

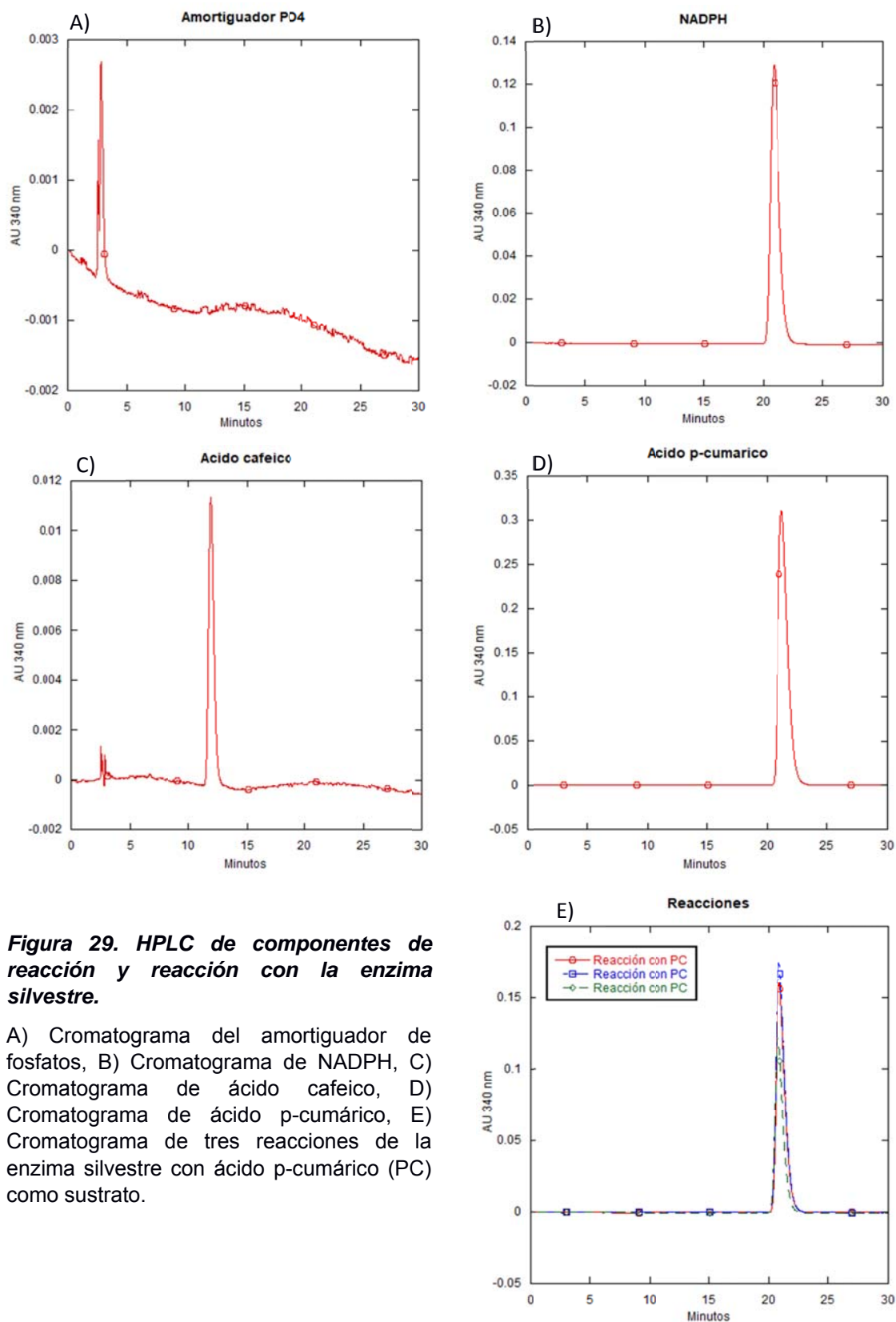
Para el análisis del perfil de productos se analizaron todos los componentes de la reacción por separado y las reacciones con p-cumárico como sustrato después de haberse consumido el NADPH. Las muestras se corrieron en un equipo Waters con un detector de diodos, utilizando la columna C18 Ultrasphere y una fase móvil de metanol: agua: ac. acético (20:79:1) con un flujo 1 mL/min y una temperatura de 30°C.

Los perfiles cromatográficos obtenidos de los componentes de reacción demuestran que el amortiguador, ácido cafeico, ácido p-cumárico tienen tiempos de elución diferentes que permiten su separación e identificación y que el NADPH tiene un tiempo de retención muy similar al ácido p-cumárico sin embargo al final de la reacción el NADPH ya se ha consumido casi por completo y no interfiere con la identificación de ácido cafeico. (Figura 29)

Al analizar las muestras de las reacciones observamos que no hay consumo de ácido p-cumárico y que no se detecta producción de algún otro compuesto producto de la hidroxilación en otra región del sustrato. Junto con los resultados de la cromatografía en capa fina podemos concluir que la enzima CYT-BM3 4E10, no utiliza el ácido p-cumárico como sustrato y que al explorar los residuos seleccionados estamos promoviendo un cambio de especificidad.

A partir del método de tamizaje de alta eficiencia se seleccionaron 3 variantes de 200 evaluadas inicialmente para verificar el perfil de producción mediante HPLC, como podemos observar en la figura 30, el inciso 1 nos muestra el control positivo de nuestro experimento en el que se produjo ácido cafeico durante la inducción por la 4-hidroxifenilacetato hidroxilasa; el inciso 2, 3 y 5 son las variantes seleccionadas y se puede ver un cambio en el perfil, donde vemos la aparición de un pico nuevo muy cercano al ácido p-cumárico, lo que indica que si bien no obtenemos la producción de ácido cafeico, estas variantes permiten al citocromo utilizar al ácido p-cumárico como sustrato. Queda por averiguar mediante IR, NMR y espectrometría de masas, la naturaleza del producto obtenido.

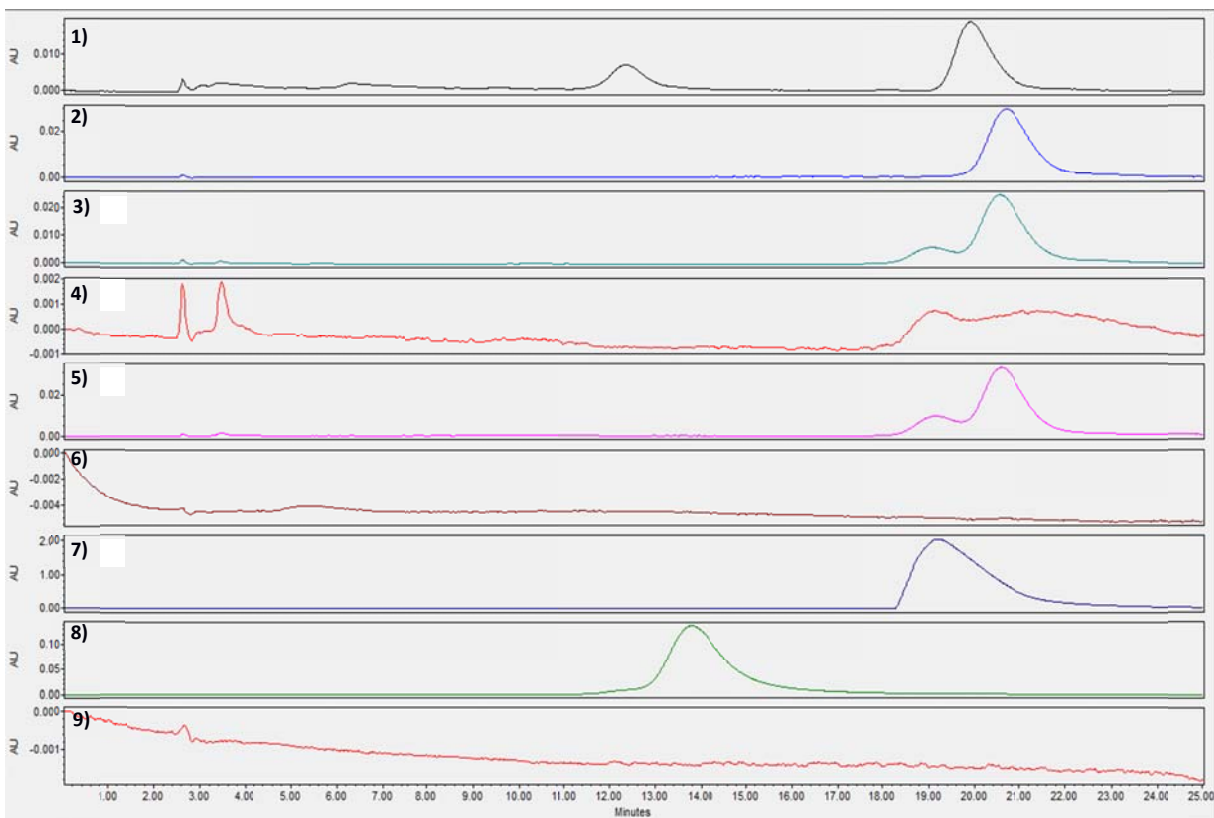




**Figura 29. HPLC de componentes de reacción y reacción con la enzima silvestre.**

A) Cromatograma del amortiguador de fosfatos, B) Cromatograma de NADPH, C) Cromatograma de ácido cafeico, D) Cromatograma de ácido p-cumárico, E) Cromatograma de tres reacciones de la enzima silvestre con ácido p-cumárico (PC) como sustrato.





**Figura 30. Perfil de productos obtenidos**

1) 4-hidroxifenilacetato hidroxilasa (4HFAH) suplementado con ácido p-cumárico durante la sobreexpresión en cultivo 2) Variante 2E1. 3) Variante 1F11. 4) NADPH. 5) Variante 2G8. 6) Actividad de 4HFAH en extracto. 7) Ácido p-cumárico 8) Ácido cafeico 9) Amortiguador de fosfatos 30 mM.

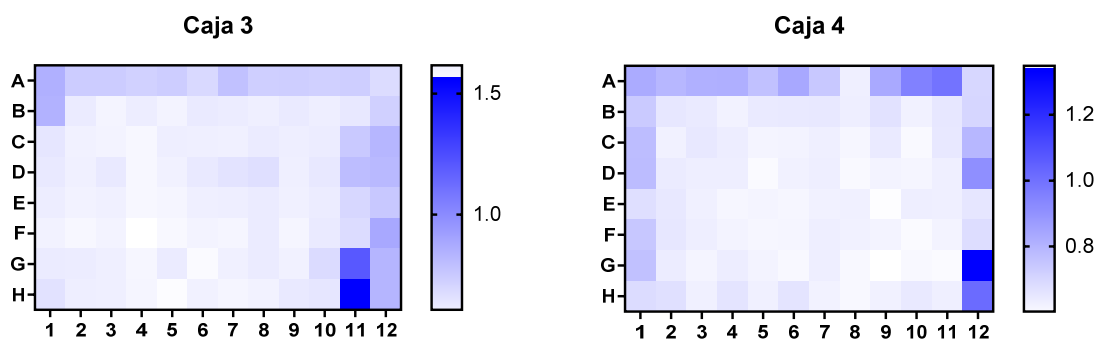
Es importante mencionar que el diseño propuesto establece la mutación de varios residuos simultáneamente. En el banco que dio origen a estas variantes analizadas sólo se exploraron mutantes en 1 o 2 posiciones, por lo tanto, esperamos que al explorar la combinatoria completa de variantes se puedan obtener variantes capaces de producir ácido cafeico.

Con este propósito se transformó la biblioteca ensamblada usando la técnica de CPEC en *E. coli* BL21, esta biblioteca fue secuenciada y se comprobó que los sitios estuvieran mutagenizados. Actualmente se está trabajando para evaluar un total de 10 000 variantes de esta librería.

## 7.9 Análisis de variantes

Utilizando la biblioteca construida mediante la técnica de CPEC se transformó en la cepa de *E.coli* BL21 y se procedió a evaluar las colonias obtenidas. En la figura 31 se muestran los resultados de este análisis, cada grafico corresponde a una caja en el que la columna 12 consta de puros controles. En las cajas con numeración del 3 al 12 se analizó solo una colonia por pocillo y en el caso del 13 al 22 se analizaron en conjunto 10 colonias en cada cultivo para poder explorar una mayor proporción de la librería, las dos primeras cajas que se evaluaron se encuentran entre los resultados previos.

Como se puede observar en los gráficos de calor, entre más azul se encuentre cada cuadrícula mayor el resultado al realizar la medición con el método de tamizaje propuesto, por lo tanto entre más claro o blanco se encuentre la cuadrícula menor la actividad. De esta manera es fácil observar aquellas variantes de interés, como por ejemplo si evaluamos en este caso la caja número 3 la cuadrícula correspondiente al H11 es una variante de interés para realizar su caracterización posteriormente y la llamaríamos CYT 3H11 durante su análisis de acuerdo a su posición en la cuadrícula y el número de su caja de procedencia. De esta manera se pueden analizar las demás cajas que se presentan en el anexo 13.



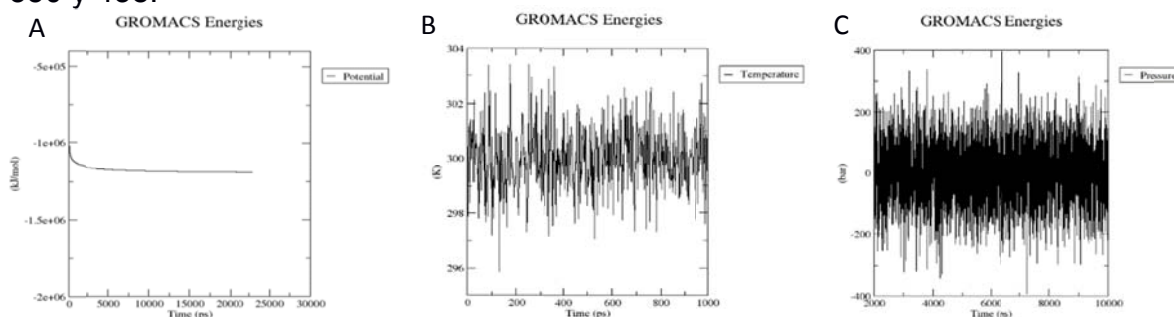
**Figura 31. Análisis de variantes.**

En estas cuadrículas se ejemplifica el análisis de las variantes con el método de tamizaje en formato de alta eficiencia, cada cuadro es una variante diferente y la intensidad del color azul nos muestra las variantes con una mayor actividad con respecto a las otras. Esta representación hace sencillo comparar la diferencia entre las variantes cuyos extractos se obtuvieron utilizando cajas de 96 pocillos de volumen de 1 ml. Las demás variantes analizadas se encuentran en el anexo 13.

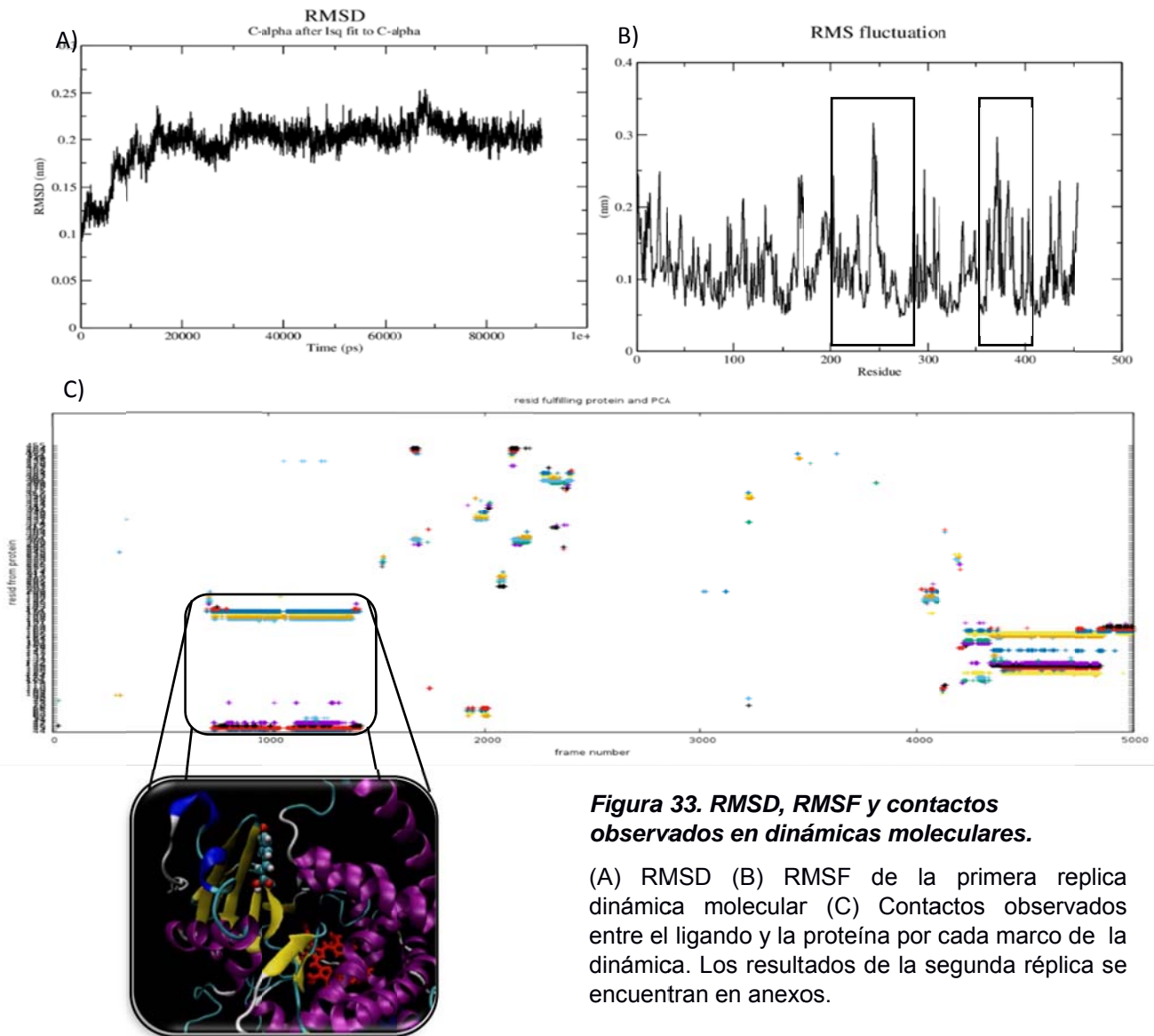
A manera de perspectiva, durante la construcción y evaluación de las mutantes propuestas mediante nuestro análisis inicial nos dimos a la tarea de realizar unas dinámicas moleculares para entender un poco más del proceso de interacción de la proteína con el ligando y de las conformaciones de la proteína como tal.

Para esto se utilizó el programa GROMACS 5.1.4 en el cual se preparó la estructura de la proteína con el campo de fuerza CHARMM36 y al ligando utilizando el mismo campo de fuerza apoyándonos con el servidor en línea CGenFF para generar la topología compatible. Se fusionaron los archivos de la proteína y ligando posicionando al ácido p-cumárico cerca del residuo 47 que se encuentra en la entrada del sitio activo, se minimizó el sistema utilizando descenso acelerado por 50000 pasos utilizando el algoritmo de LINCS para someter restricciones al sistema. Se equilibró el sistema utilizando el termostato de V-rescale durante 1000 ps y el baróstato de Berendsen durante 5 ns a 25°C y 1 atm (Figura 32), se utilizó PME (Particle Mesh Ewald) para calcular las interacciones electrostáticas de largo alcance y condiciones periódicas de frontera para conservar la ergodicidad del sistema.

Se realizó una dinámica de 100 ns por triplicado. En dos de tres de las dinámicas observamos una interacción del ligando en el sitio activo, para las cuales se presenta su RMSD, RMSF y un gráfico de los contactos por cuadro del ligando durante la dinámica con la proteína (Figura 33 y anexo 12). En esta dinámica podemos ver que la proteína se estabiliza en los primeros 50 ns y que las regiones de mayor fluctuación se encuentran entre los residuos 200, alrededor de 350 y 438.



**Figura 32. A) Minimización B) Termostato C) Baróstato**



**Figura 33. RMSD, RMSF y contactos observados en dinámicas moleculares.**

(A) RMSD (B) RMSF de la primera replica dinámica molecular (C) Contactos observados entre el ligando y la proteína por cada marco de la dinámica. Los resultados de la segunda réplica se encuentran en anexos.

Tres de nuestras mutaciones caen cerca de esta región de mayor movimiento, donde la variante 438 es la más cercana a esta región, modificar la flexibilidad de esta región puede ser importante para la especificidad y es importante recalcar que una de las tres mutantes seleccionadas vienen de la biblioteca sencilla de este residuo 438 que es la 1F11.

De igual manera, resulta interesante que durante la dinámica molecular se observa una interacción en la entrada del sitio activo que corresponde al residuo 47 responsable de anclar el carboxilo del sustrato natural y se encuentra interactuando de manera similar directamente con el p-cumárico (Figura 33).

## 8 DISCUSIÓN

Los resultados de la evaluación de la actividad de la enzima 4E10 demuestran que esta variante del CYT BM3 que fue modificada para tomar sustratos pequeños [Chen *et al.*, 2012] no puede utilizar como sustrato al ácido p-cumárico y por lo tanto no produce ácido cafeico o algún otro producto de hidroxilación, de igual manera Munday y colaboradores observan la incapacidad del sitio activo del hemo para hidroxilar ciclohexano y benceno. De esta manera, las mutaciones propuestas cambian la especificidad del sitio activo y permiten la entrada del ácido p-cumárico para su hidroxilación.

Hoy en día las técnicas de dinámica molecular utilizando cálculos cuánticos (QM/MM) [Cole *et al.*, 2019] o los métodos de teoría de función de densidades (DFT) [Vuong *et al.*, 2019] se han vuelto una herramienta de acceso para los centros de investigación y en conjunto con el trabajo de supercomputadoras permiten evaluar mecanismos de acción de forma muy fina, en enzimas poco estudiadas o en procesos como el transporte de electrones [Vidal Limón *et al.*, 2013]. Sin embargo, en enzimas con un trabajo extenso como el CYT BM3 se puede aprovechar todo el conocimiento reportado, en conjunto con análisis estructural para proponer mutaciones que propicien un cambio de especificidad.

El análisis del sitio activo mediante la implementación de un protocolo de refinamiento previo a realizar el acoplamiento molecular en las diferentes proteínas respaldado por información de variantes reportadas nos permitió identificar sitios de mutagénesis importantes para que esta proteína utilice el ácido p-cumárico como sustrato. Al aplicar este protocolo en un citocromo diferente pudimos identificar variantes *in silico* para cambiar su especificidad y obtener una validación de las predicciones que ya existían resultados experimentales de esta variante.

La construcción de las bibliotecas de variantes sencillas y dobles se logró mediante digestión y ligación del gen CYP BM3, mientras que para la combinatoria de variantes se probaron los métodos de Gibson assembly, Quickchange y CPEC, siendo la estrategia de CPEC con la que se obtuvo una librería de tamaño

adecuado para su evaluación, en la que se comprobaron los sitios mutagenizados por secuenciación.

Uno de los retos más grandes al trabajar con evolución dirigida es la disponibilidad de métodos robustos para evaluar las variantes propuestas de manera sencilla y reproducible, durante la búsqueda de metodologías para distinguir nuestro producto del sustrato que consta solo de una hidroxilación, nos apoyamos en el cambio en las propiedades fisicoquímicas que esta modificación otorga a la molécula y montamos un método en el que aprovechando el cambio en su capacidad antioxidante podemos medir la reducción del  $\text{Fe}^{+3}$  a  $\text{Fe}^{+2}$ , la cual se ve acelerada en presencia de ácido cafeico utilizando ferrozina, lo que nos da una tonalidad desde rosado hasta morado y permite distinguir las variantes con diferentes perfiles de actividad.

Si bien, queda pendiente la evaluación de la biblioteca de combinatorias completa (Anexo 13), las bibliotecas pequeñas analizadas con el método de tamizaje propuesto permitieron seleccionar variantes que al parecer utilizan al ácido p-cumárico como sustrato, dando como producto una molécula con tiempo de retención diferente al del cafeico pero que presentan un citocromo con la capacidad de tomar sustratos no naturales, cabe destacar que la hidroxilación en otras regiones de la molécula podrían generar metabolitos de interés como precursores de cumarinas y muy especialmente si presentan un poder reductor mayor que el del sustrato utilizado. En base a esto, creemos que la combinatoria permitirá la obtención de variantes que orienten adecuadamente al ácido p-cumárico en el sitio activo para la producción de ácido cafeico con el CYT BM3.

Las metodologías empleadas para analizar la estructura del sitio activo como el protocolo de refinamiento, acoplamiento molecular, análisis de modos normales y dinámica molecular sientan una base en el laboratorio para seguir trabajando con la modificación de la especificidad del CYT BM3 y lograr la conversión de metabolitos de interés.

## 9 CONCLUSIONES

- Los resultados de actividad de la enzima CYT BM3 4E10 muestran que esta variante no es capaz de hidroxilar al ácido p-cumárico.
- El protocolo de afinamiento y validación del proceso de acoplamiento molecular permitió seleccionar sitios a mutagenizar para cambiar la especificidad del CYT BM3.
- La medición de  $\text{Fe}^{+2}$  utilizando ferrozina en cajas de 96 pocillos como método de tamizaje de alta eficiencia, resultó en una estrategia sencilla y rápida para la selección de variantes con diferentes perfiles de actividad que utilicen el ácido p-cumárico como sustrato.
- Utilizando el método de CPEC se construyó la librería propuesta ( $10^3$ ), la cual contiene una combinatoria de mutaciones múltiples que se someterán a un exhaustivo escrutinio para identificar variantes capaces de transformar ácido p-cumárico en ácido cafeico.

## 10 PERSPECTIVAS

- Caracterizar los productos de la reacción de las variantes seleccionadas.
- Identificar mediante secuenciación de DNA los cambios que dieron lugar a cualquier cambio de especificidad en la enzima CYT BM3
- Continuar con la búsqueda de variantes activas en la biblioteca generada durante esta tesis.



# 11 ANEXOS

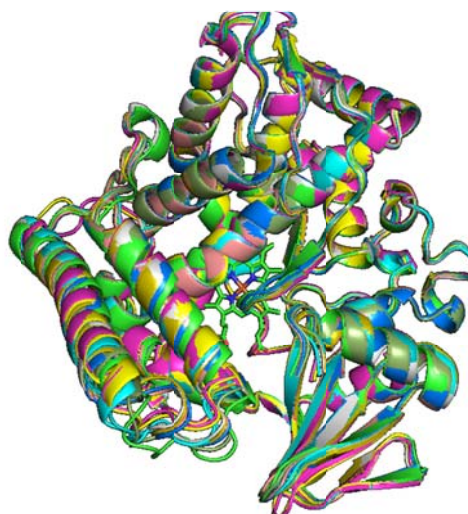
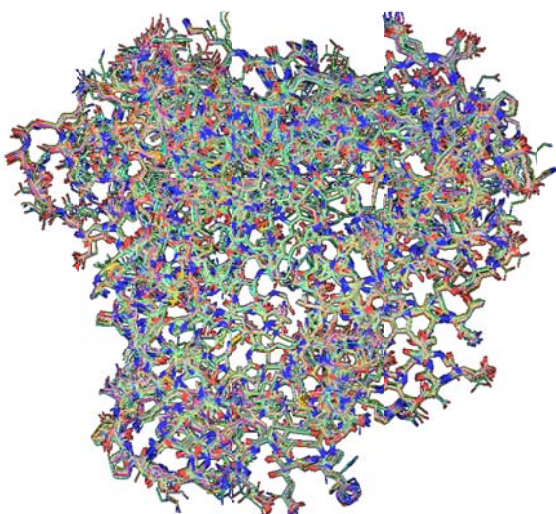
## Anexo 1. Estructuras cristalizadas de CIT BM3

Estructura sin mutaciones y sin ligando			
PDB	Res	Rfree	Rwork
1FAH	2.3		0.172
1BVY	2.03	0.275	
1BU7	1.65	0.251	
2IJ2	1.2	0.166	0.148
2J1M	1.7	0.199	0.169
2J4S	2.1	0.243	0.19
2BMH	2		0.184
2HPD	2		0.167

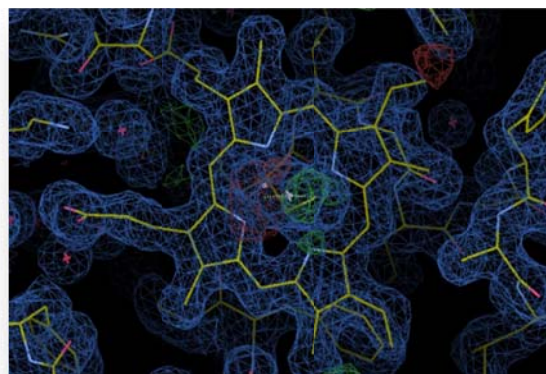
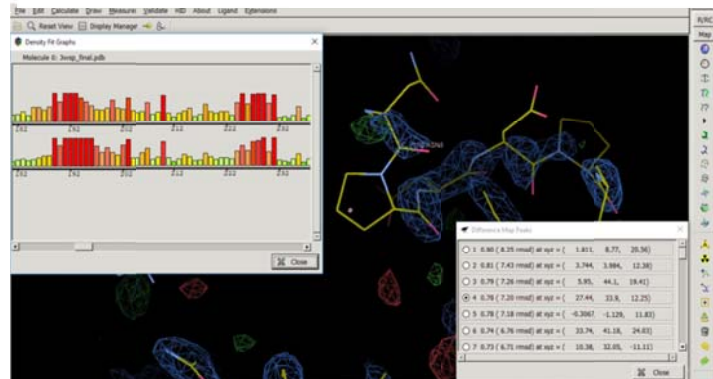
Estructura con mutaciones y sin ligando			
PDB	Res	Rfree	Rwork
1YQO	1.9	0.206	0.166
1SMI	2	0.288	0.224
1P0W	2	0.221	0.159
1YQP	1.8	0.21	0.174
2IJ3	1.9	0.218	0.177
2NNB	1.9	0.233	0.182
3M4V	1.9	0.232	0.187
3KX5	1.69	0.208	0.174
2IJ4	2.4	0.273	0.203
1P0X	2	0.237	0.172
4KF0	1.45	0.178	0.145
3PSX	1.9	0.222	0.175

Estructuras con ligando				
PDB	Res	Rfree	Rwork	Ligando
1FAG	2.7	0.353	0.246	PEGX5
4ZFA	2.7	0.302	0.244	PAM
1SMJ	2.7	0.338	0.253	PAM
1JPZ	1.65	0.193	0.177	N-GlyPAM
4KEW	1.89	0.237	0.188	Omeprazol
4ZFB	2.84	0.274	0.218	PAM
4ZF8	2.76	0.263	0.215	Metirapone
3BEN	1.65	0.191	0.161	Lys-HisPAM
5DYZ	1.96	0.219	0.179	N-GlyPAM
5B2V	2.3	0.255	0.215	PFCA
4DUF	1.8	0.203	0.163	Serotonina
4DUB	1.7	0.204	0.164	Dopamina
4KPA	2	0.23	0.194	N-GlyPAM
3WSP	1.8	0.212	0.174	Trp-PFCA

## Anexo 2. Alineamiento estructural de diferentes CIT BM3 cristalizados



## Anexo 3. Mapas de densidad electrónica.





**Anexo 4. RMSD entre estructuras cristalográficas.**

		Estructura sin ligando																
		1BU7		1BVY		1FAH		2BMH		2HPD		2IJ2		2J1M		2J4S		
		A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	
Estructuras sin ligando	1BU7	A	-	0.456	0.677	0.578	0.353	0.511	0.351	0.521	0.348	0.486	0.543	0.266	0.324	0.234	0.518	0.504
		B	0.456	-	0.318	0.259	0.425	0.305	0.428	0.313	0.415	0.277	0.215	0.501	0.503	0.57	0.466	0.453
	1BVY	A	0.677	0.318	-	0.232	0.625	0.452	0.642	0.452	0.632	0.435	0.31	0.719	0.717	0.804	0.554	0.553
		B	0.578	0.259	0.232	-	0.56	0.412	0.569	0.412	0.55	0.384	0.264	0.617	0.622	0.677	0.503	0.488
	1FAH	A	0.353	0.425	0.625	0.56	-	0.355	0.146	0.37	0.194	0.367	0.451	0.329	0.447	0.464	0.483	0.486
		B	0.511	0.305	0.452	0.412	0.355	-	0.354	0.147	0.359	0.217	0.333	0.476	0.546	0.64	0.509	0.482
	1P0X	A	0.377	0.368	0.562	0.502	0.331	0.396	0.318	0.398	0.294	0.385	0.377	0.367	0.48	0.489	0.402	0.41
		B	0.503	0.252	0.388	0.362	0.413	0.306	0.418	0.309	0.395	0.281	0.24	0.486	0.57	0.627	0.467	0.472
	1SMI	A	0.45	0.365	0.484	0.446	0.523	0.465	0.534	0.491	0.512	0.434	0.424	0.551	0.44	0.416	0.477	0.403
		B	0.305	0.52	0.634	0.569	0.423	0.555	0.43	0.587	0.407	0.54	0.537	0.361	0.372	0.303	0.461	0.47
	1YQO	A	0.272	0.331	0.552	0.464	0.323	0.379	0.331	0.4	0.295	0.376	0.323	0.305	0.434	0.413	0.324	0.325
		B	0.4	0.201	0.374	0.312	0.416	0.299	0.43	0.321	0.404	0.289	0.249	0.429	0.47	0.516	0.428	0.388
	1YQP	A	0.257	0.348	0.569	0.491	0.32	0.382	0.314	0.394	0.281	0.374	0.34	0.289	0.435	0.379	0.367	0.362
		B	0.434	0.182	0.354	0.303	0.402	0.276	0.408	0.289	0.384	0.253	0.199	0.432	0.527	0.567	0.418	0.424
	2BMH	A	0.351	0.428	0.642	0.569	0.146	0.354	-	0.367	0.192	0.367	0.452	0.326	0.442	0.456	0.485	0.485
		B	0.521	0.313	0.452	0.412	0.37	0.147	0.367	-	0.367	0.216	0.341	0.513	0.557	0.653	0.534	0.509
	2HPD	A	0.348	0.415	0.632	0.55	0.194	0.359	0.192	0.367	-	0.335	0.441	0.308	0.433	0.445	0.469	0.469
		B	0.486	0.277	0.435	0.384	0.367	0.217	0.367	0.216	0.335	-	0.312	0.476	0.548	0.606	0.522	0.501
	2IJ2	A	0.543	0.215	0.31	0.264	0.451	0.333	0.452	0.341	0.441	0.312	-	0.489	0.62	0.683	0.363	0.406
		B	0.266	0.502	0.719	0.617	0.329	0.476	0.326	0.513	0.308	0.476	0.489	-	0.427	0.387	0.429	0.413
	2IJ3	A	0.38	0.403	0.566	0.496	0.39	0.4	0.385	0.434	0.36	0.41	0.377	0.317	0.481	0.484	0.361	0.308
		B	0.416	0.321	0.491	0.405	0.439	0.348	0.436	0.379	0.407	0.353	0.337	0.375	0.477	0.543	0.412	0.345
	2IJ4	A	0.521	0.447	0.619	0.577	0.577	0.561	0.598	0.597	0.582	0.521	0.523	0.649	0.501	0.473	0.568	0.528
		B	0.384	0.593	0.716	0.673	0.485	0.692	0.498	0.717	0.487	0.661	0.656	0.446	0.423	0.373	0.596	0.607
	2J1M	A	0.324	0.503	0.717	0.622	0.447	0.546	0.442	0.557	0.433	0.548	0.62	0.427	-	0.293	0.646	0.588
		B	0.234	0.57	0.804	0.677	0.464	0.64	0.456	0.653	0.445	0.606	0.683	0.387	0.293	-	0.634	0.602
	2J4S	A	0.525	0.466	0.554	0.503	0.483	0.509	0.485	0.534	0.469	0.522	0.363	0.429	0.646	0.634	-	0.254
		B	0.504	0.453	0.553	0.488	0.486	0.482	0.485	0.509	0.469	0.501	0.406	0.413	0.588	0.602	0.254	-
	2NNB	A	0.341	0.357	0.524	0.454	0.375	0.401	0.374	0.422	0.334	0.399	0.306	0.312	0.507	0.482	0.292	0.282
		B	0.481	0.248	0.355	0.301	0.458	0.323	0.465	0.342	0.43	0.313	0.201	0.443	0.559	0.61	0.346	0.342
3KX5	A	0.359	0.447	0.663	0.58	0.302	0.43	0.296	0.454	0.296	0.426	0.445	0.194	0.428	0.473	0.468	0.477	
	B	0.556	0.222	0.328	0.297	0.437	0.287	0.451	0.297	0.443	0.28	0.167	0.495	0.572	0.687	0.49	0.399	
3M4V	A	0.362	0.471	0.702	0.624	0.376	0.49	0.379	0.518	0.36	0.492	0.427	0.232	0.501	0.48	0.378	0.343	
	B	0.347	0.455	0.666	0.591	0.393	0.482	0.388	0.511	0.359	0.471	0.393	0.27	0.528	0.456	0.356	0.871	
Estructuras con ligando	1FAG	A	0.618	0.829	0.936	0.879	0.707	0.911	0.719	0.947	0.74	0.912	0.946	0.725	0.607	0.62	0.876	0.894
		B	0.652	0.882	0.977	0.907	0.762	0.941	0.766	0.952	0.797	0.956	0.992	0.773	0.64	0.627	0.902	0.622
	1JPZ	A	0.382	0.633	0.833	0.754	0.576	0.73	0.581	0.761	0.573	0.698	0.691	0.494	0.426	0.358	0.661	0.606
		B	0.406	0.608	0.775	0.674	0.576	0.686	0.576	0.717	0.572	0.66	0.655	0.533	0.452	0.349	0.635	0.875
	1SMJ	A	0.644	0.846	0.939	0.889	0.732	0.923	0.734	0.954	0.752	0.914	0.959	0.757	0.656	0.669	0.87	0.887
		B	0.7	0.887	0.974	0.929	0.793	0.96	0.796	0.975	0.818	0.944	0.99	0.828	0.683	0.676	0.927	0.614
	4KEW	A	0.353	0.61	0.81	0.715	0.533	0.692	0.548	0.713	0.522	0.673	0.647	0.463	0.435	0.308	0.624	0.589
		B	0.366	0.617	0.818	0.729	0.546	0.696	0.556	0.724	0.53	0.67	0.656	0.476	0.453	0.331	0.617	0.614
	4ZFA	A	0.546	0.741	0.88	0.815	0.75	0.849	0.743	0.88	0.737	0.802	0.826	0.728	0.606	0.518	0.756	0.764

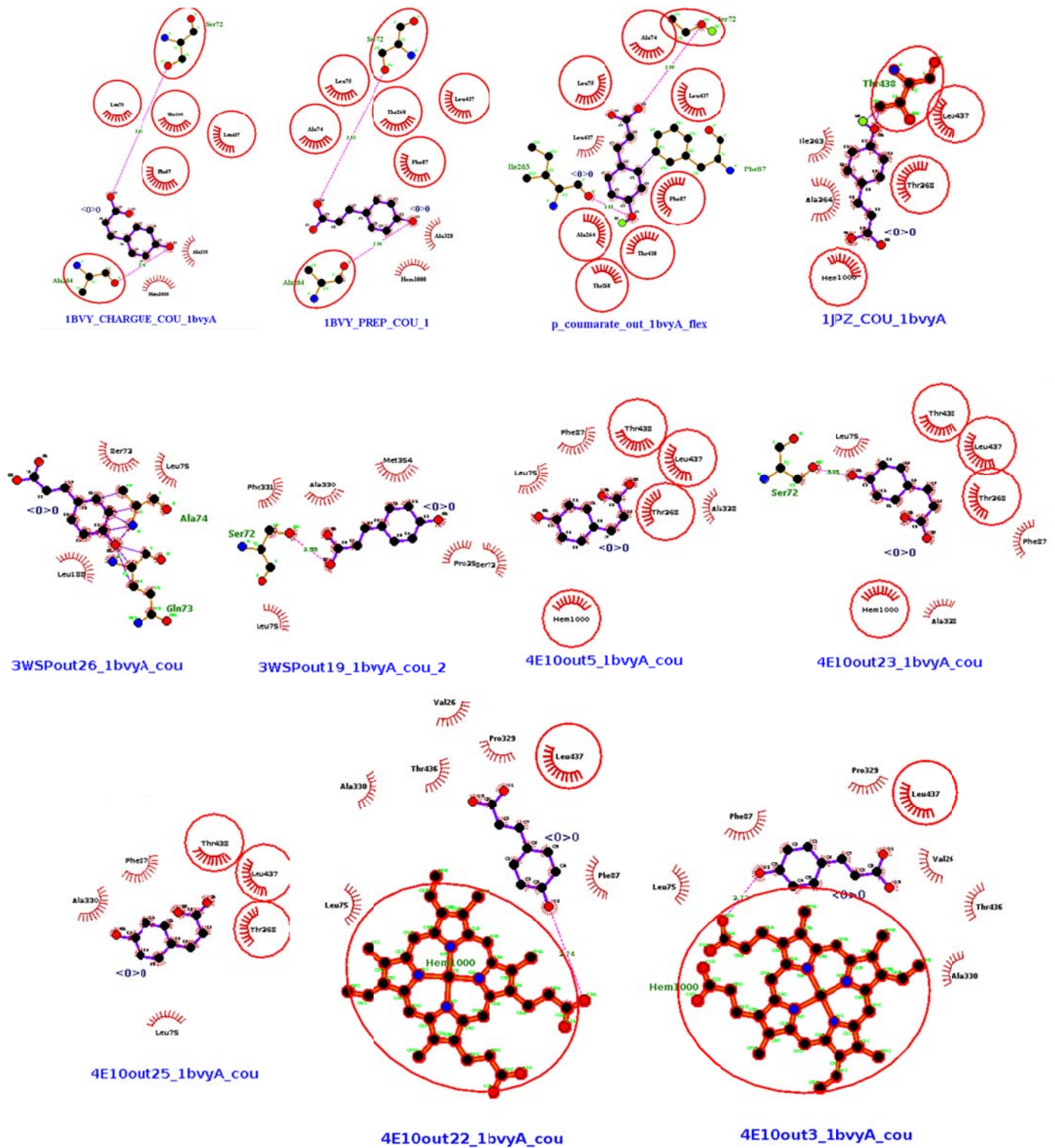
**Anexo 5. Estados de protonación a pH 7.4**

Residuo	pKa									
	1BVY	1JPZ	2IJ2	2J1M	3BEN	3PSX	3WSP	4DUB	4DUF	5XA3
ASP 23	3.160	2.650	3.620	3.920	3.890	3.840	3.920	3.460	2.300	2.650
ASP 34	4.690	4.060	3.860	3.900	3.820	3.900	4.210	3.990	3.790	3.840
ASP 63	2.620	1.410	1.620	2.110	1.840	3.890	3.020	1.810	1.550	2.920
ASP 68	3.240	3.010	2.490	2.790	3.050	3.020	2.880	2.590	3.340	3.390
ASP 80	4.780	5.010	4.180	4.720	4.780	4.010	4.680	4.610	4.380	4.700
ASP 84	4.380	3.940	4.600	4.400	4.030	4.480	4.270	3.990	4.350	4.520
ASP 121	2.460	3.750	3.530	4.510	3.550	4.580	4.480	3.540	3.510	3.990
ASP 136	3.920	3.340	3.720	3.340	3.320	3.640	3.760	3.800	3.810	3.650
ASP 144	3.650	4.440	4.570	4.170	4.550	4.110	4.390	4.170	4.030	4.280
ASP 151	5.610	5.960	5.430	5.870	5.770	4.760	5.000	5.750	5.930	5.800
ASP 168	3.880	3.740	3.830	3.890	3.730	3.780	3.550	3.840	3.760	3.740
ASP 182	4.530	4.820	5.330	5.130	5.000	-	4.750	4.850	5.120	5.170
ASP 194	3.290	2.790	2.780	2.810	2.760	-	2.700	3.880	2.760	3.380
ASP 195	3.370	3.310	3.210	3.380	3.260	-	3.290	3.270	3.260	3.320
ASP 199	3.780	3.990	3.950	3.950	3.960	-	4.060	4.010	3.850	3.790
ASP 208	4.530	4.370	4.990	5.090	4.760	-	4.490	5.240	5.120	5.220
ASP 214	3.870	3.030	2.990	3.040	3.010	3.950	4.070	3.990	3.030	2.970
ASP 217	4.560	4.080	4.680	4.530	4.600	4.160	4.930	4.830	4.810	4.690
ASP 222	3.720	3.990	2.980	3.000	2.920	3.050	2.950	2.960	3.010	2.970
ASP 231	3.970	3.780	4.120	3.590	4.070	3.890	3.630	4.070	4.050	3.090
ASP 232	2.990	3.890	3.090	3.790	2.950	3.630	3.440	3.010	4.960	4.660
ASP 242	2.330	2.520	1.950	2.930	2.490	3.530	2.040	1.910	1.860	2.690
ASP 250	3.420	3.210	3.390	3.840	3.370	3.700	3.820	3.930	3.830	3.350
ASP 251	3.310	3.960	3.220	3.130	3.490	3.880	4.040	2.970	4.150	4.170
ASP 300	3.440	3.850	4.050	4.080	3.850	3.710	3.990	3.970	4.070	3.950
ASP 338	3.160	3.210	3.310	3.240	3.260	3.170	3.240	3.340	3.240	3.340
ASP 351	3.720	3.880	3.480	4.160	3.120	5.160	3.880	2.740	4.130	3.020
ASP 363	3.270	2.870	3.400	3.280	3.010	3.050	3.280	3.480	4.090	3.630
ASP 369	3.880	3.750	3.840	3.760	3.790	3.800	3.770	3.800	3.790	3.770
ASP 370	1.610	3.380	3.060	2.670	3.160	2.470	3.890	2.770	2.830	2.830
ASP 422	3.170	3.550	3.700	3.810	3.500	3.330	3.820	3.890	3.710	3.540
ASP 425	2.860	2.930	2.640	2.680	2.420	3.180	3.030	2.760	2.270	2.550
ASP 432	3.080	3.980	2.940	2.890	2.890	2.880	2.910	2.900	3.940	4.110
GLU 4	-	4.570	4.540	4.510	4.170	4.350	4.520	4.560	4.560	4.520
GLU 13	-	4.560	4.550	4.560	4.560	4.550	4.550	4.550	4.520	4.550
GLU 35	4.580	4.620	4.700	4.730	4.660	4.620	4.600	4.720	4.670	4.570
GLU 38	2.890	4.580	4.120	3.620	3.940	4.500	3.870	4.700	4.540	4.450
GLU 43	3.470	3.450	4.380	2.900	3.460	3.410	3.550	4.240	4.430	3.530
GLU 60	3.550	4.180	4.070	3.880	3.750	3.270	3.700	3.750	4.470	3.910
GLU 64	4.760	4.070	4.190	4.170	4.240	4.430	4.420	4.060	4.160	3.880
GLU 93	2.850	3.400	3.210	3.480	4.870	2.720	2.950	3.320	3.310	2.800
GLU 131	4.660	2.950	3.000	3.680	2.920	2.970	4.240	5.560	4.200	2.890
GLU 137	3.530	4.080	4.000	4.130	4.100	3.910	3.900	3.920	4.060	3.960
GLU 140	4.570	4.580	5.530	5.100	5.400	4.670	4.860	5.900	5.720	4.950
GLU 143	5.650	5.340	3.690	4.500	4.070	5.390	5.350	4.580	4.530	3.970
GLU 183	4.290	6.680	5.230	4.440	3.920	-	4.630	4.600	3.890	4.430
GLU 200	4.640	3.750	3.680	3.720	3.910	-	3.690	3.710	4.270	4.120
GLU 207	4.610	4.740	3.550	3.490	4.530	-	3.460	4.480	4.710	3.720
GLU 228	4.880	-	4.560	4.560	4.570	-	4.530	3.810	4.490	4.530
GLU 244	4.760	3.800	3.850	4.740	4.640	4.690	4.680	4.700	3.910	3.800
GLU 247	3.410	3.370	2.520	2.430	3.410	3.270	2.560	2.510	2.540	2.480
GLU 252	5.120	4.150	4.010	4.780	4.500	3.730	4.560	4.300	3.980	4.970
GLU 267	4.490	4.980	3.950	3.380	3.560	3.920	5.670	2.420	4.630	4.600
GLU 292	4.790	3.710	3.950	4.560	3.990	3.610	4.340	4.050	3.900	3.920
GLU 293	5.330	4.920	4.900	5.310	4.830	4.920	5.050	5.410	5.160	5.120
GLU 320	5.170	6.870	6.650	6.540	6.650	6.880	6.660	6.350	6.130	6.050
GLU 337	4.530	3.740	3.710	4.650	4.660	4.110	4.650	4.680	3.730	4.050
GLU 344	4.770	4.390	3.340	4.360	4.560	4.610	4.410	3.390	3.480	3.400
GLU 348	4.990	4.950	4.900	4.950	4.910	4.710	4.980	4.900	4.950	4.920
GLU 352	4.300	5.260	5.440	3.830	5.000	4.780	5.050	5.300	5.060	4.450
GLU 372	3.530	4.400	4.260	4.510	4.240	3.750	4.020	3.660	3.710	4.260
GLU 373	4.810	4.570	4.630	3.850	4.210	3.780	4.660	4.710	4.670	3.790
GLU 377	2.960	3.080	3.370	4.220	3.200	4.180	3.290	3.320	3.180	3.650
GLU 380	4.710	4.660	4.680	4.820	4.700	4.430	4.450	4.870	4.610	4.740
GLU 409	8.530	8.840	7.590	7.800	8.140	8.680	8.420	7.710	8.340	8.050
GLU 424	3.340	4.330	4.450	4.380	4.410	3.440	4.400	4.340	4.390	3.580
GLU 430	4.620	4.650	4.610	4.610	4.610	4.660	4.630	4.630	4.650	4.600
GLU 435	5.220	6.720	5.810	6.090	5.880	5.490	6.030	5.960	6.150	6.260





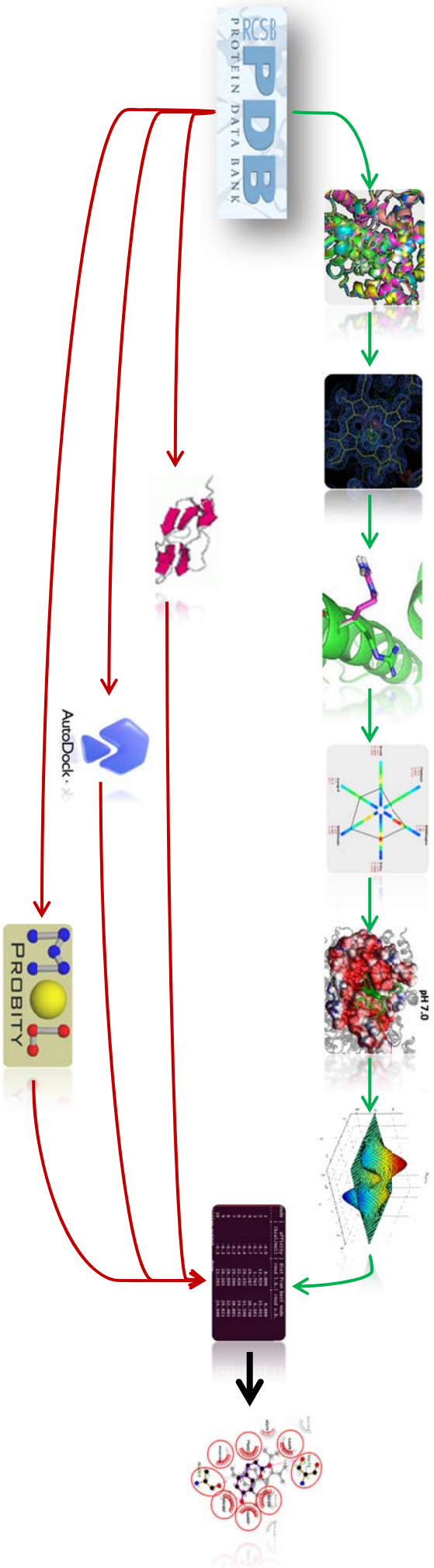
Anexo 6. Representación 2D de contactos proteína-ligando.



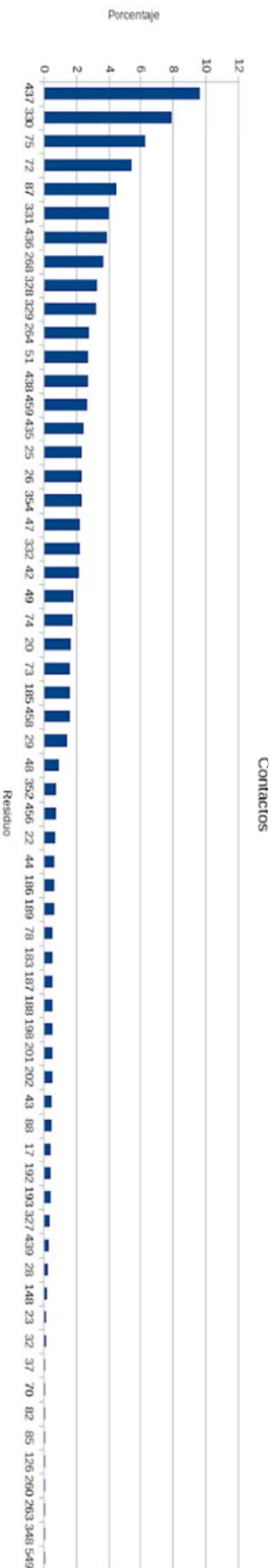




**Anexo 8. Protocolo de refinamiento y acoplamiento molecular.**

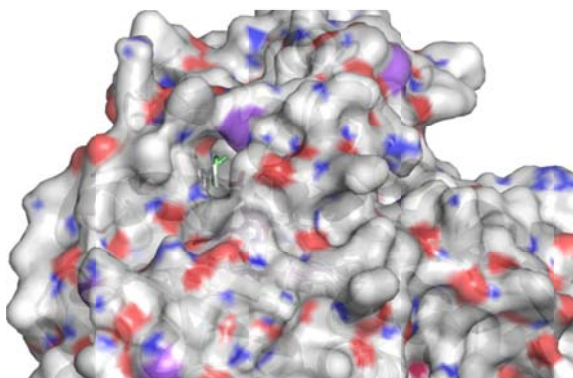


**Anexo 9. Contactos observados durante los protocolos de acoplamiento.**

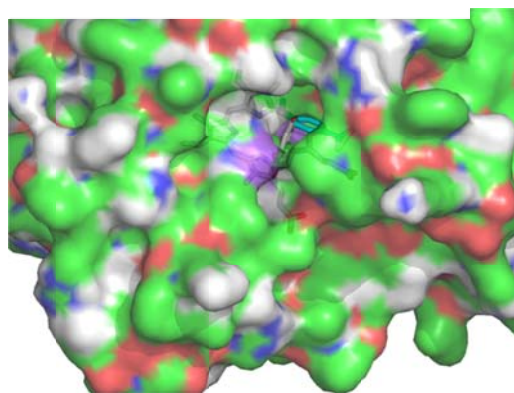




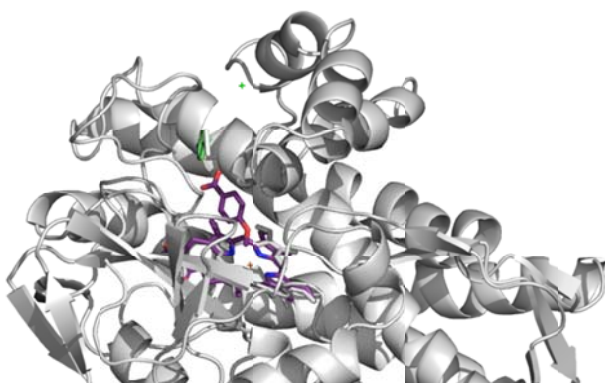
**Anexo 10. Validación con CYT199A2**



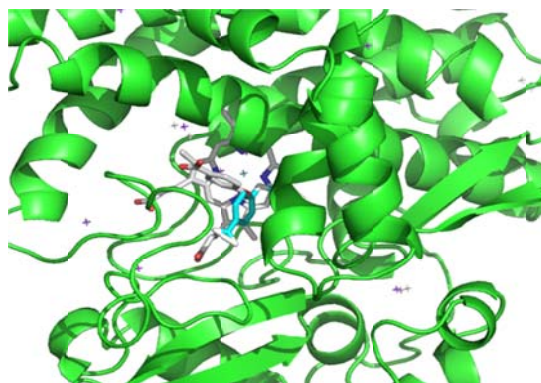
2FR7 Ácido 5-metoxibenzoico



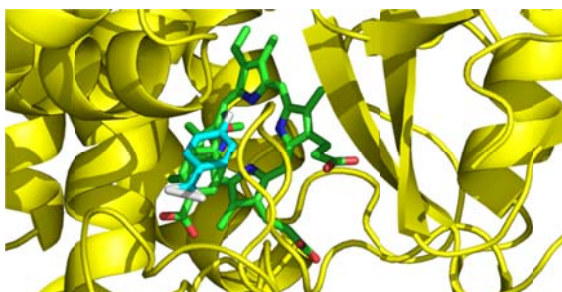
2FR7\_F185L Ácido 5-metoxibenzoico



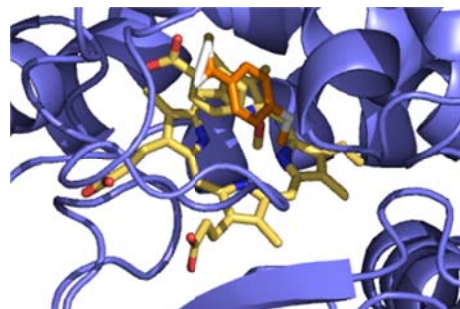
2FR7 Ácido 5-metoxibenzoico



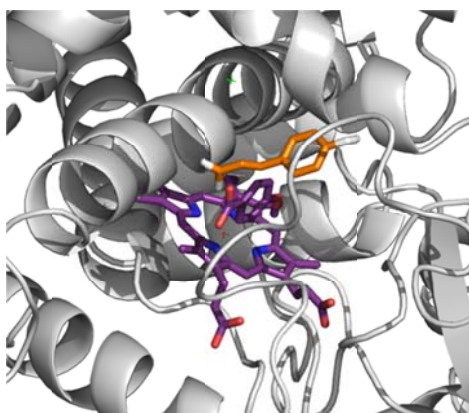
2FR7\_F185L Ácido 5-metoxibenzoico



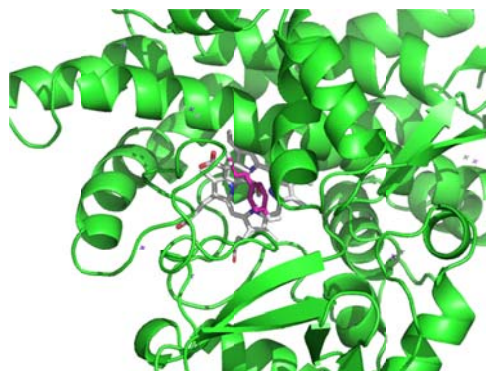
4DNJ Ácido 5-metoxibenzoico



4DNJ\_F185L Ácido 5-metoxibenzoico



2FR7 Ácido p-cumárico



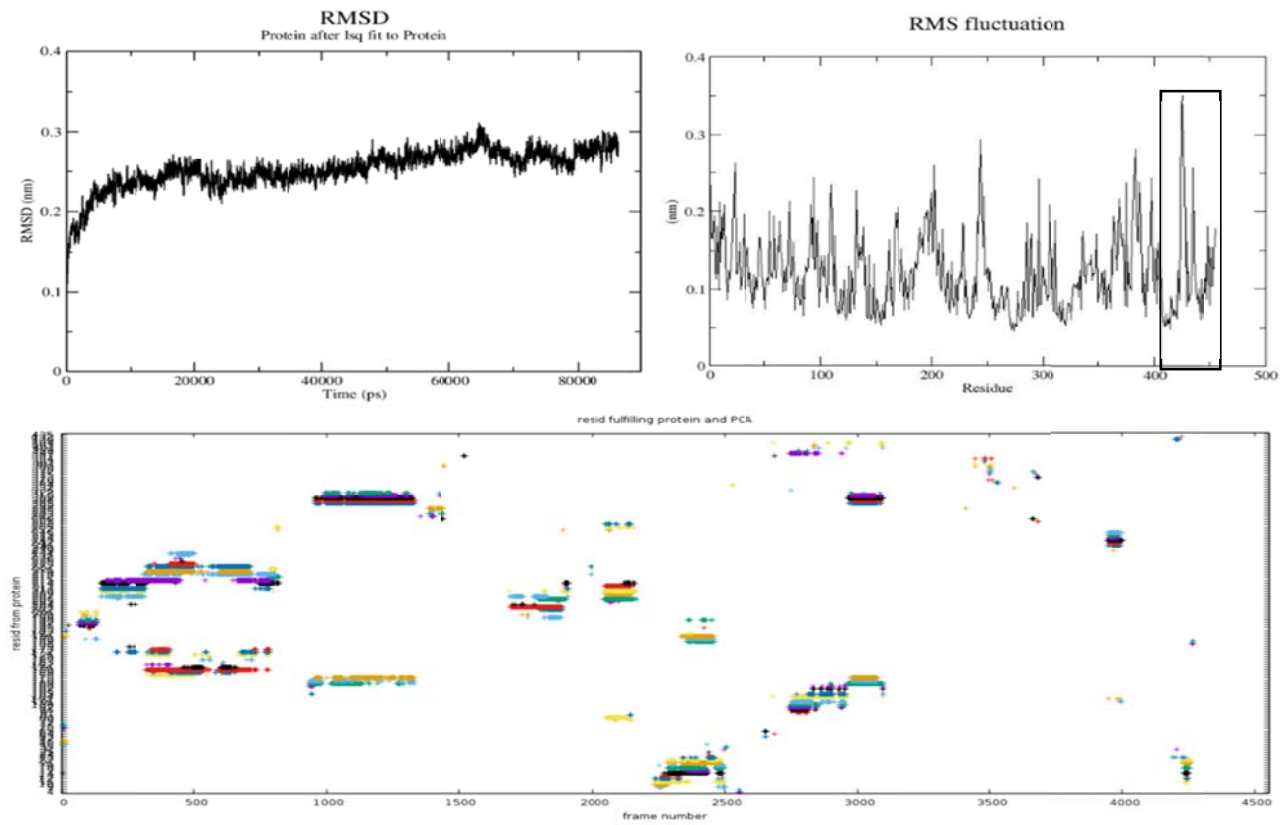
2FR7\_F185L Ácido p-cumárico



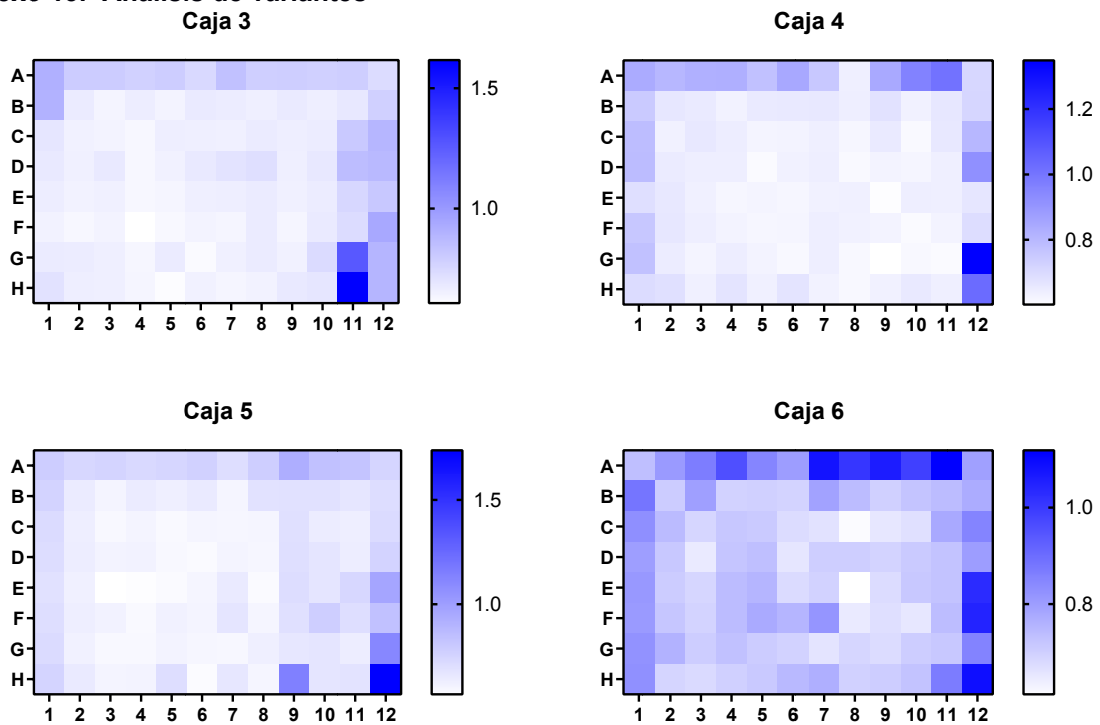
### Anexo 11.- Contribución de variantes en el dominio hemo a la catálisis

Mutación	Observaciones	Referencia
F42A	Aporta hidrofobicidad, esta mutante aumenta Kd y Km (hydrophobic lid)	[32]
R47E	Interactúa con el sustrato, si se muta por aa's pequeños se utilizan sustratos más grandes	PDB 4KPB
Y51F	Es importante para la entrada del sustrato, aumenta la oxidación de aromáticos policíclicos	[29]
L75T	Aumenta la catálisis con butirato y exanoato	[35]
A82F,W	Aumenta catálisis de indol	PDB 2VWH
F87	Interactúa directamente con el sustrato	PDB 2X80
W96	Unión al hemo y estabilización del estado del spin	[37]
L181K	Aumenta la catálisis con butirato y exanoato	[35]
D251G	Catálisis de ibuprofeno y diclofenaco	PDB 5DYP
A264	Cubre el pirrol C de la porfirina	[28]
A268	Indispensable para la catálisis	PDB 1FAH
A330P	Ayuda a utilizar moléculas de menor tamaño	PDB 3M4V
I366V	Estabiliza a la proteína	PDB 4KPB
F393A,H,Y,N	Modifican las propiedades electrónicas del hemo	PDB 1P0W
C400M	Cambio el rango de absorción, no es recomendable mutar	PDB 1FAH
I401P	Ayuda en el reconocimiento de sustratos no naturales y aumenta actividad de la proteína	PDB 3M4V
409	Indispensable para la catálisis	PDB 1FAH
R47L, F81I, F87V, L188Q, E267V	Desestabilizan a la proteína. L188Q pierde contacto con Q73, de igual manera E267V con K440; F81I afecta el empaquetamiento hidrofóbico	PDB 4ZF6
L188P,F162I,K187E	Desestabilizan a la proteína	[33]
F162I, K187E, M237I	Desestabilizan a la proteína	[36]
F162I, K187E, L188P, M237I	Desestabilizan a la proteína	[30]
L52I, A184V,L324I,V340M,I366V	Aumenta 15°C T50	[32]

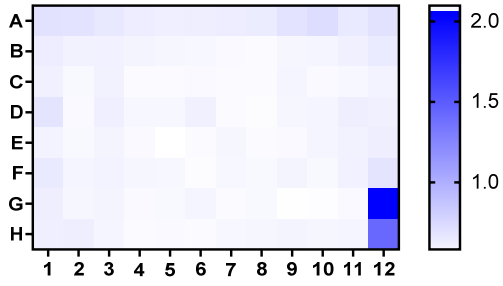
**Anexo 12. Segunda replica de dinámica molecular**



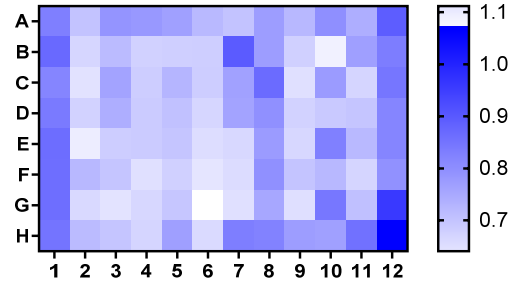
**Anexo 13.- Análisis de variantes**



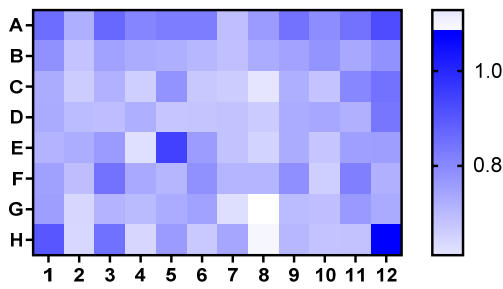
Caja 7



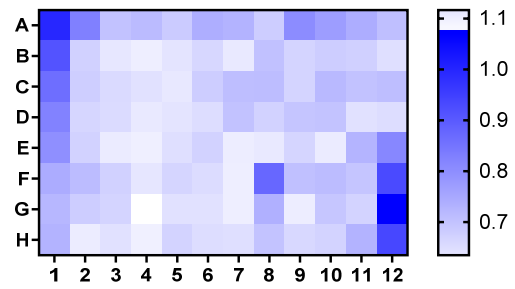
Caja 8



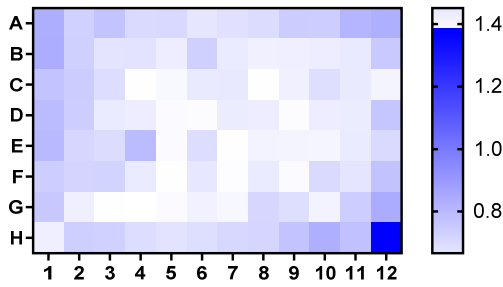
Caja 9



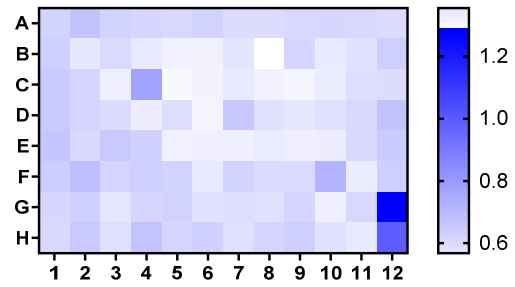
Caja 10



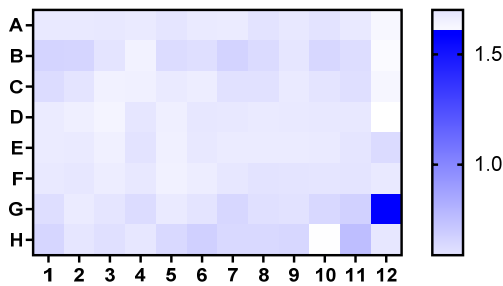
Caja 11



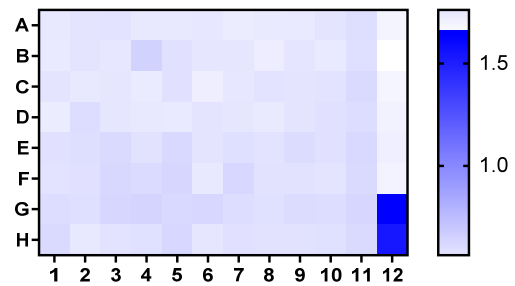
Caja 12



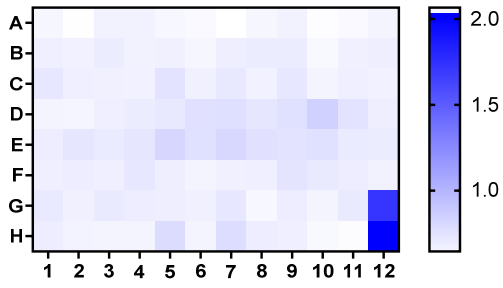
Caja 13



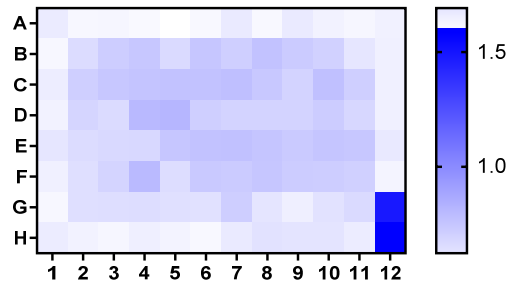
Caja 14



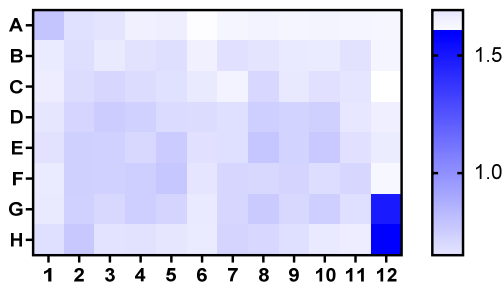
Caja 15



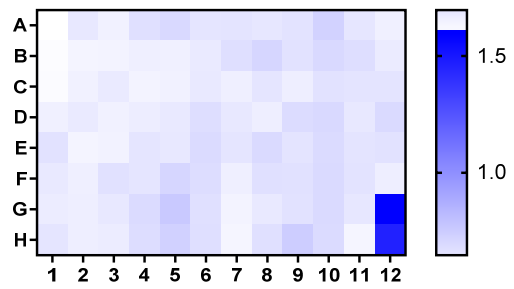
Caja 16



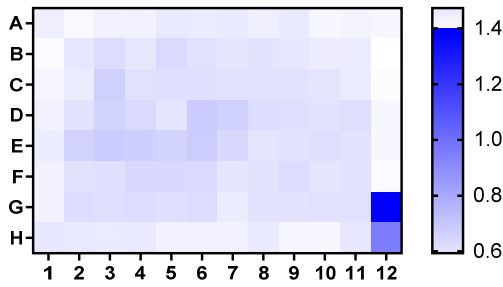
Caja 17



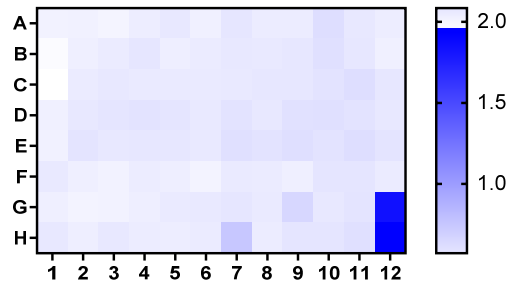
Caja 18



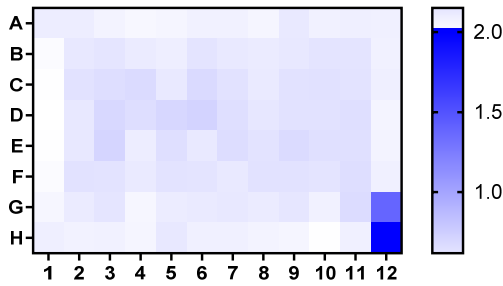
Caja 19



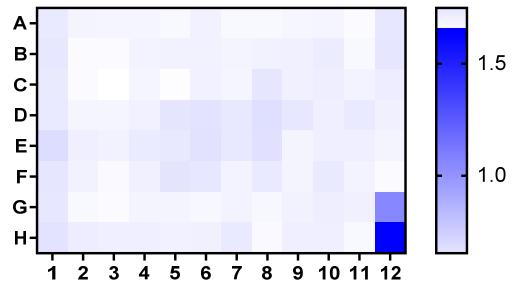
Caja 20



Caja 21



Caja 22



## 12 REFERENCIAS

1. Ahmed, F., Avery, K. L., Cullis, P. M., Primrose, W. U., Roberts, G. C., Al-Mutairi, E. H., & Willis, C. L. (1999). An unusual matrix of stereocomplementarity in the hydroxylation of monohydroxy fatty acids catalysed by cytochrome P 450 from *Bacillus megaterium* with potential application in biotransformations. *Chemical Communications*, (20), 2049-2050.
2. Bernhardt, R. (2006). Cytochromes P450 as versatile biocatalysts. *Journal of biotechnology*, 124(1), 128-145.
3. Branco, R. J., Seifert, A., Budde, M., Urlacher, V. B., Ramos, M. J., & Pleiss, J. (2008). Anchoring effects in a wide binding pocket: The molecular basis of regioselectivity in engineered cytochrome P450 monooxygenase from *B. megaterium*. *Proteins: Structure, Function, and Bioinformatics*, 73(3), 597-607.
4. Brill, E., Hannemann, F., Zapp, J., Brüning, G., Jauch, J., & Bernhardt, R. (2014). A new cytochrome P450 system from *Bacillus megaterium* DSM319 for the hydroxylation of 11-keto- $\beta$ -boswellic acid (KBA). *Applied microbiology and biotechnology*, 98(4), 1703-1717.
5. Carmichael, A. B., & Wong, L. L. (2001). Protein engineering of *Bacillus megaterium* CYP102. *The FEBS Journal*, 268(10), 3117-3125.
6. Chao, P. C., Hsu, C. C., & Yin, M. C. (2009). Anti-inflammatory and anti-coagulatory activities of caffeic acid and ellagic acid in cardiac tissue of diabetic mice. *Nutrition & metabolism*, 6(1), 33.
7. Chaudhary, K. K., & Mishra, N. (2016). A Review on Molecular Acoplamiento: Novel Tool for Drug Discovery. *databases*, 3, 4.
8. Chen, J. H., & Ho, C. T. (1997). Antioxidant activities of caffeic acid and its related hydroxycinnamic acid compounds. *Journal of Agricultural and Food Chemistry*, 45(7), 2374-2378.
9. Chen, M. M., Snow, C. D., Vizcarra, C. L., Mayo, S. L., & Arnold, F. H. (2012). Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. *Protein Engineering, Design & Selection*, 25(4), 171-178.
10. Cirino, P. C., & Arnold, F. H. (2002). Regioselectivity and activity of cytochrome P450 BM-3 and mutant F87A in reactions driven by hydrogen peroxide. *Advanced Synthesis & Catalysis*, 344(9), 932-937.

11. Cole, M. F., & Gaucher, E. A. (2011). Exploiting models of molecular evolution to efficiently direct protein engineering. *Journal of molecular evolution*, 72(2), 193-203.
12. Cole, Daniel J., Israel Cabeza de Vaca, and William L. Jorgensen. "Computation of protein–ligand binding free energies using quantum mechanical bespoke force fields." *MedChemComm* (2019).
13. Combs, S. A., DeLuca, S. L., DeLuca, S. H., Lemmon, G. H., Nannemann, D. P., Nguyen, E. D.,... & Meiler, J. (2013). Small-molecule ligand acoplamiento into comparative models with Rosetta. *Nature protocols*, 8(7), 1277-1298.
14. M. M. Y. Chen, C. D. Snow, C. L. Vizcarra, S. L. Mayo, F. H. Arnold, *Protein Eng. Des. Sel.* 2012, 25, 171 –178.
15. Erkelenz, M., Kuo, C. H., & Niemeyer, C. M. (2011). DNA-mediated assembly of cytochrome P450 BM3 subdomains. *Journal of the American Chemical Society*, 133(40), 16111-16118.
16. Fasan, R., Meharena, Y. T., Snow, C. D., Poulos, T. L., & Arnold, F. H. (2008). Evolutionary history of a specialized P450 propane monooxygenase. *Journal of molecular biology*, 383(5), 1069-1080.
17. Finkel, T., & Holbrook, N. J. (2000). Oxidants, oxidative stress and the biology of ageing. *Nature*, 408(6809), 239-247.
18. Furuya, T., Arai, Y., & Kino, K. (2012). Biotechnological production of caffeic acid by bacterial cytochrome P450 CYP199A2. *Applied and environmental microbiology*, 78(17), 6087-6094.
19. Girvan, H. M., & Munro, A. W. (2016). Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. *Current opinion in chemical biology*, 31, 136-145.
20. Hynes, M. J., & O'Coinceanainn, M. (2004). The kinetics and mechanisms of reactions of iron (III) with caffeic acid, chlorogenic acid, sinapic acid, ferulic acid and naringin. *Journal of inorganic biochemistry*, 98(8), 1457-1464.
21. Ikeda K, Tsujimoto K, Uozaki M, Nishide M, Suzuki Y, Koyama AH, Yamasaki H: Inhibition of multiplication of herpes simplex virus by caffeic acid. *Int J Mol Med* 2011, 28:595-598.
22. Kadkhodayan, S., Coulter, E. D., Maryniak, D. M., Bryson, T. A., & Dawson, J. H. (1995). Uncoupling Oxygen Transfer and Electron Transfer in the Oxygenation of Camphor Analogues by Cytochrome P450-Cam Direct Observation of an Intermolecular

Isotope Effect for Substrate Ch Activation. *Journal of Biological Chemistry*, 270(47), 28042-28048.

23. M.G.Joyce *et al.*, (2012). the crystal structure of the FAD/NADPH-binding domain of flavocytochrome P450 BM3. *Febs J*, **279**, 1694-1706

24. Lee, C., & Levitt, M. (1991). Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core. *Nature*, 352(6334), 448-451.

25. Li, H., & Poulos, T. L. (1997). The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nature Structural & Molecular Biology*, 4(2), 140-146.

26. Li, Q. S., Schwaneberg, U., Fischer, P., & Schmid, R. D. (2000). Directed Evolution of the Fatty-Acid Hydroxylase P450 BM-3 into an Indole-Hydroxylating Catalyst. *Chemistry-A European Journal*, 6(9), 1531-1536.

27. Lin, Y., & Yan, Y. (2012). Biosynthesis of caffeic acid in *Escherichia coli* using its endogenous hydroxylase complex. *Microbial cell factories*, 11(1), 42.

28. Loida, P. J., & Sligar, S. G. (1993). Molecular recognition in cytochrome P-450: mechanism for the control of uncoupling reactions. *Biochemistry*, 32(43), 11530-11538.

29. Lussenburg, B. M., Babel, L. C., Vermeulen, N. P., & Commandeur, J. N. (2005). Evaluation of alkoxyresorufins as fluorescent substrates for cytochrome P450 BM3 and site-directed mutants. *Analytical biochemistry*, 341(1), 148-155.

30. Magnani, C., Isaac, V. L. B., Correa, M. A., & Salgado, H. R. N. (2014). Caffeic acid: a review of its potential use in medications and cosmetics. *Analytical Methods*, 6(10), 3203-3210.

31. M.G.Joyce *et al.*, (2012). The crystal structure of the FAD/NADPH-binding domain of flavocytochrome P450 BM3. *Febs J*, 279, 1694-1706

32. Munday, S. D., Shoji, O., Watanabe, Y., Wong, L. L., & Bell, S. G. (2016). Improved oxidation of aromatic and aliphatic hydrocarbons using rate enhancing variants of P450Bm3 in combination with decoy molecules. *Chemical Communications*, 52(5), 1036-1039.

33. Noble, M. A., Miles, C. S., Chapman, S. K., Lysek, D. A., MacKay, A. C., HANZLIK, R. P., & MUNRO, A. W. (1999). Roles of key active-site residues in flavocytochrome P450 BM3. *Biochemical Journal*, 339(2), 371-379.



34. Otey, C. R., Bandara, G., Lalonde, J., Takahashi, K., & Arnold, F. H. (2006). Preparation of human metabolites of propranolol using laboratory-evolved bacterial cytochromes P450. *Biotechnology and bioengineering*, 93(3), 494-499.
35. Packer, M. S., & Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nature Reviews Genetics*, 16(7), 379.
36. Pochapsky, T. C., Kazanis, S., & Dang, M. (2010). Conformational plasticity and structure/function relationships in cytochromes P450. *Antioxidants & redox signaling*, 13(8), 1273-1296.
37. Roiban, G. D., & Reetz, M. T. (2015). Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chemical Communications*, 51(12), 2208-2224.
38. Rojas, J., Londoño, C., & Ciro, Y. (2016). The health benefits of natural skin UVA photoprotective compounds found in botanical sources. *International Journal of Pharmacy and Pharmaceutical Sciences*, 8(3), 13-23.
39. Schafer, F. Q., & Buettner, G. R. (2001). Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. *Free Radical Biology and Medicine*, 30(11), 1191-1212.
40. Seifert, A., Vomund, S., Grohmann, K., Kriening, S., Urlacher, V. B., Laschat, S., & Pleiss, J. (2009). Rational Design of a Minimal and Highly Enriched CYP102A1 Mutant Library with Improved Regio-, Stereo- and Chemoselectivity. *Chembiochem*, 10(5), 853-861.
41. Shoji, O., Kunimatsu, T., Kawakami, N., & Watanabe, Y. (2013). Highly Selective Hydroxylation of Benzene to Phenol by Wild-type Cytochrome P450BM3 Assisted by Decoy Molecules. *Angewandte Chemie International Edition*, 52(26), 6606-6610.
42. Spagnol, C. M., Oliveira, T. S., Isaac, V. L. B., Corrêa, M. A., & Salgado, H. R. N. (2015). Validation of Caffeic Acid in Emulsion by UV-Spectrophotometric Method. *Physical Chemistry*, 5(1), 16-22.
43. Vidal-Limón, A., Águila, S., Ayala, M., Batista, C. V., & Vazquez-Duhalt, R. (2013). Peroxidase activity stabilization of cytochrome P450BM3 by rational analysis of intramolecular electron transfer. *Journal of inorganic biochemistry*, 122, 18-26.
44. Vuong, V. Q., Nishimoto, Y., Fedorov, D. G., Sumpter, B. G., Niehaus, T. A., & Irlé, S. (2019). The Fragment Molecular Orbital Method Based on Long-Range Corrected Density-Functional Tight-Binding. *Journal of chemical theory and computation*.

45. Wang, H., Provan, G. J., & Helliwell, K. (2004). Determination of rosmarinic acid and caffeic acid in aromatic herbs by HPLC. *Food Chemistry*, 87(2), 307-311.