



CENTRO DE INVESTIGACIÓN Y ASISTENCIA EN  
TECNOLOGÍA Y  
DISEÑO DEL ESTADO DE JALISCO, A.C.



“ANÁLISIS *in silico* DE LOS POSIBLES  
DOMINIOS CONSERVADOS Y DE  
REGULACIÓN DE LA PROTEÍNA  
FLAVONOIDE-3',5'-HIDROXILASA  
(F3'5'H) EN *Petunia hybrida*”

# TESIS

QUE PARA OBTENER EL GRADO ACADÉMICO DE

MAESTRO EN CIENCIAS DE LA FLORICULTURA

PRESENTA

ADRIANA GONZÁLEZ DURÁN

GUADALAJARA, JALISCO, OCTUBRE 2013





Guadalajara, Jalisco a 27 de Noviembre de 2013

**Dr. Rodrigo Barba González**  
**Coordinador Académico de la Maestría**  
**en Ciencias de la Floricultura**  
**Guadalajara, Jalisco**

Los que suscriben integrantes del jurado de examen de grado del estudiante Adriana González Durán, una vez leída y revisada la Tesis titulada “**ANÁLISIS in silico DE LOS POSIBLES DOMINIOS CONSERVADOS Y DE REGULACIÓN DE LA PROTEÍNA FLAVONOIDE-3',5'-HIDROXILASA (F3'5'H) EN Petunia hybrida**” aceptamos que la referida tesis revisada y corregida sea presentada por el alumno para aspirar al grado de Maestro en Ciencias de la Floricultura durante el examen correspondiente.

Y para que así conste firmamos la presente a los 27 días del mes de Noviembre del año 2013.

\_\_\_\_\_  
*Presidente*

\_\_\_\_\_  
*Secretario*

\_\_\_\_\_  
*Vocal*

\_\_\_\_\_  
*Vocal*

\_\_\_\_\_  
*Vocal*

## **DEDICATORIA**

A Dios por iluminar mis pensamientos y mi camino pero sobre todo por permitirme lograrlo.

A mis chiquilines, Sophia y Ernesto, porque con su presencia vinieron a alegrar mi vida.

A mi esposo Gonzalo por alentarme y darme fortaleza para no darme por vencida y conseguir mis metas.

A mi padre por enseñarme a ser mejor cada día y apoyarme en todo momento con mi proceso de posgrado.

A mi madre por acompañarme en cada comenzar y ser mi inspiración de ser una mujer profesional.

A mi hermano Sergio por su apoyo durante la maestría.



## AGRADECIMIENTOS

Al fondo CONACYT, por el apoyo otorgado a través del proyecto: Estudios moleculares y bioquímicos para la modificación de la ruta de biosíntesis de las antocianinas en crisantemo (*Dendranthema grandiflorum* Tzvelev), con la clave: 104606, del cual es parte esta Tesis de maestría.

Al CONACYT por la beca otorgada con número 241852, para la realización de esta tesis.

Al Centro de Investigación y Asistencia En Tecnología y Diseño del Estado de Jalisco, A.C. Unidad Sureste por la beca complementaria otorgada durante los primeros 4 cuatrimestres de la maestría.

A la Unidad Sureste Centro de Investigación y Asistencia En Tecnología y Diseño del Estado de Jalisco, A.C. por facilitar el uso de las instalaciones y equipos para la realización de esta tesis.

A la Dra. Ana Luisa Ramos Díaz, por dirección y orientación para la realización de la presente, así como confianza y perseverancia en mi persona.

A los sínodos: Dra. Julia Cano Sosa, Dr. Alberto Uc Varguez, Dra. Nina Valadez González, M.C. Ángela Ku González y la Dra. Guadalupe López Puc, por sus comentarios que enriquecieron el presente documento.

A la Dra. Ingrid Mayanín Rodríguez por su profesionalismo y darme la oportunidad de obtener un éxito más en mi vida profesional...MI MAESTRÍA.

Así como a todos los investigadores que participaron en mi crecimiento profesional y personal.



## CONTENIDO

	PÁGINA
<b>1. ANTECEDENTES</b>	<b>1</b>
<b>1.1 IMPORTANCIA ECONÓMICA DE LA FLORICULTURA</b>	<b>1</b>
<b>1.2 PIGMENTOS RESPONSABLES DEL COLOR DE LAS FLORES</b>	<b>4</b>
1.2.1 Estructura y función de las antocianinas	9
1.2.2 Biosíntesis de antocianinas	10
1.2.3 El color de la flor y los Citocromos P450	12
<b>1.3 MEJORAMIENTO GENÉTICO EN LA MANIPULACIÓN DEL COLOR DE LA FLOR</b>	<b>15</b>
<b>1.4 LA BIOLOGIA DE SISTEMAS</b>	<b>17</b>
1.4.1 Bases de datos biológicas	21
1.4.2 La Bioinformática en plantas	25
<b>1.5 OBJETIVO GENERAL</b>	<b>27</b>
<b>1.6 OBJETIVOS ESPECÍFICOS</b>	<b>27</b>
<b>1.7 ESTRATEGIA EXPERIMENTAL INTEGRAL</b>	<b>28</b>
<b>1.8 REFERENCIAS BIBLIOGRÁFICAS</b>	<b>29</b>
<b>2. ANÁLISIS DE LA SECUENCIA DE LA PROTEÍNA F3'5'H E IDENTIFICACIÓN DE LOS DOMINIOS CONSERVADOS.</b>	<b>36</b>
2.1 Introducción	36
2.2 Estrategia experimental	37
2.3 Materiales y métodos	38
2.3.1. Búsqueda de la secuencia putativa del gen <i>F3'5'H</i> .	38
2.3.2. Búsqueda de los marcos de lectura abiertos (ORF) para el gen <i>F3'5'H</i> de <i>Petunia hybrida</i> .	39
2.3.3. Búsqueda de las secuencias ortólogas de la proteína F3'5'H de <i>Petunia hybrida</i> en las diferentes bases de datos.	39
2.3.4. Alineamiento de las secuencias obtenidas para la proteína F3'5'H de <i>Petunia hybrida</i> y búsqueda de los dominios conservados.	40

2.3.5. Elaboración del árbol filogenético molecular de la proteína F3'5'H.	40
2.4 Resultados y discusión	40
2.4.1 Selección de la secuencia del gen <i>F3'5'H</i> .	40
2.4.2 Análisis y propuesta de los Marcos de lectura abiertos (ORF) para el gen <i>F3'5'H</i> de <i>Petunia hybrida</i> y obtención de la secuencia de la proteína F3'5'H de <i>Petunia hybrida</i> .	41
2.4.3 Obtención de las secuencias ortólogas de la proteína F3'5'H de <i>Petunia hybrida</i> en las diferentes bases de datos.	43
2.4.4 Análisis de similitud e identidad de las secuencias obtenidas para la proteína F3'5'H de <i>Petunia hybrida</i> e identificación de los dominios conservados.	47
2.4.5. Árbol filogenético molecular de relación de función de la proteína F3'5'H y el Citocromo P450.	54
2.5 Referencias bibliográficas	58
<b>3. ANÁLISIS DE LA FUNCIÓN, REGULACIÓN Y CARACTERÍSTICAS DE LA PROTEÍNA F3'5'H DE <i>Petunia hybrida</i>.</b>	<b>60</b>
3.1 Introducción	60
3.2 Estrategia experimental	62
3.3 Materiales y métodos	62
3.3.1 Determinación del perfil hidrofóbico de la proteína F3'5'H de <i>Petunia hybrida</i> .	62
3.3.2 Obtención de la estructura secundaria de la proteína F3'5'H de <i>Petunia hybrida</i> y la identificación de los dominios conservados.	63
3.3.3 Obtención y análisis de la estructura terciaria de la proteína F3'5'H de <i>Petunia hybrida</i> .	63
3.4. Resultados y discusión	63
3.4.1 Perfil hidrofóbico de la proteína F3'5'H de <i>Petunia hybrida</i> .	63
3.4.2 Estructura secundaria y dominios conservados de la proteína F3'5'H.	65
3.4.3. Estructura terciaria de la proteína F3'5'H de <i>Petunia hybrida</i> .	68



3.5 Referencias bibliográficas	69
<b>4. PROPUESTA PARA EL AISLAMIENTO DEL GEN <i>F3'5'H</i> EN FLORES CON COLORACION AZUL</b>	<b>71</b>
4.1 Introducción	71
4.2 Estrategia experimental	72
4.3 Materiales y métodos	72
4.3.1. Condiciones del tejido floral para la extracción del ARN Total e flores con coloración azul.	72
4.3.2. Diseño de oligonucleótidos para amplificar el gen <i>F3'5'H</i> en plantas lílidas.	72
4.4 Resultados y discusión	73
4.4.1. Etapas ideales de desarrollo floral para la extracción del ARN Total	73
4.4.2. Oligonucleótidos para el gen <i>F3'5'H</i> en plantas lílidas.	73
4.5 Referencias bibliográficas	73
<b>5. DISCUSIÓN INTEGRAL</b>	<b>75</b>
5.1 Referencias bibliográficas	78
<b>6. CONCLUSIONES INTEGRALES</b>	<b>80</b>
<b>7. ANEXO 1</b>	<b>83</b>

## INDICE DE CUADROS

<b>CUADRO</b>	<b>PÁGINA</b>
<b>1.1.</b> SUPERFICIE CULTIVADA y valor de producción mundial DE FLORES DE CORTE Y DE MACETA.	<b>1</b>
<b>1.2.</b> CARACTERÍSTICAS DE LOS TIPOS DE ANTOCIANIDINAS	<b>10</b>
<b>2.1.</b> ESPECIES REPORTADAS PARA LA PROTEÍNA F3'5'H EN LAS DIFERENTES BASES DE DATOS.	<b>43</b>
<b>2.2.</b> CLASIFICACIÓN DE LAS PLANTAS REPORTADAS PARA LA PROTEÍNA F3'5'H.	<b>45</b>
<b>4.1</b> OLIGONUCLEÓTIDOS PARA AMPLIFICAR EL GEN <i>F3'5'H</i> EN LÍLIDEAS.	<b>73</b>

## ÍNDICE DE FIGURAS

<b>FIGURA</b>	<b>PÁGINA</b>
1.1. ESTRUCTURA DEL FENILBENZOPIRANO DE LOS COMPUESTOS FLAVONOIDES Y DE LAS SUSTITUCIONES Y MODIFICACIONES QUE ORIGINAN LOS DIFERENTES TIPOS DE FLAVONOIDES.	6
1.2. RUTA COMÚN FENILPROPANOIDE Y RUTA BIOSÍNTÉTICA DE LOS FLAVONOIDES.	8
1.3. REPRESENTACIÓN GRÁFICA DE LA ESTRUCTURA GENERAL DE UNA ANTOCIANIDINA Y LA POSICIÓN DE LAS SUSTITUCIONES DESCRITAS EN EL CUADRO 1.2	9
1.4. RUTA GENERAL DE LA BIOSINTESIS DE ANTOCIANINAS.	11
1.5. ESTRUCTURA PRIMARIA DE LA PROTEÍNA P450.	14
1.6. ENFOQUE DE LA INVESTIGACIÓN DE LAS REDES EN LA BIOLOGÍA DE SISTEMAS.	20
1.7. BASE DE DATOS NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (NCBI).	22
1.8. BASE DE DATOS PROTEIN INFORMATION RESOURCE (PIR).	23
1.9. BASE DE DATOS UNIVERSAL PROTEIN RESOURCE (UniProt).	24
1.10. BASE DE DATOS PROTEIN DATA BANK (PDB).	25
2.1. SECUENCIA DEL GEN FLAVONOIDE 3´5´HIDROXILASA DE <i>Petunia x hybrida</i> cultivar QL01.	41
2.2. MARCOS DE LECTURA PROPUESTOS PARA EL GEN <i>F3´5´H</i> (EF371021).	42
2.3. SECUENCIA DE LA PROTEÍNA FLAVONOIDE 3´5´HIDROXILASA DE <i>Petunia hybrida</i> .	43
2.4. ALINEAMIENTO MÚLTIPLE DE LAS SECUENCIAS ORTÓLOGAS PARA LA PROTEÍNA <i>F3´5´H</i> DE <i>Petunia hybrida</i> .	49
2.5. METODOLOGÍA DEL NEIGHBOR-JOINING TREES PARA OBTENER EL ÁRBOL FILOGENÉTICO MOLECULAR DE FUNCIÓN	56

DE LA PROTEÍNA F3'5'H.

**2.6. ÁRBOL FILOGENÉTICO MOLECULAR DE LA RELACIÓN DE FUNCIÓN DE LA PROTEÍNA F3'5'H. 57**

**3.1. PERFIL HIDRÓFICO DE LA PROTEÍNA F3'5'H EN *Petunia hybrida*. 64**

**3.2. ESTRUCTURA SECUNDARIA DE LA PROTEÍNA F3'5'H DE *Petunia hybrida*. 66**

**3.3. ESTRUCTURA TERCIARIA DE LA PROTEÍNA F3'5'H DE *Petunia hybrida*. 69**

## ANEXOS

	<b>PÁG.</b>
POSTER DE PARTICIPACIÓN EN EL CONGRESO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR DE PLANTAS “In silico analysis of possible conserved domains and regulation of Flavonoid-3',5'-hydroxylase”	<b>75</b>

## RESUMEN

El color de la flor es uno de los rasgos más importantes de las plantas ornamentales. De los pigmentos relacionados con la mayor gama de colores en las flores sobresalen los flavonoides, siendo las antocianinas de sus compuestos principales y de donde se han enfocado investigaciones para sobre expresar la enzima flavonoide-3',5'-hidroxilasa involucrada en la biosíntesis de delphinidina y responsable del color azul. Como resultado se ha visto, que existe una excelente correlación entre el pigmento y la presencia del color, dando la posibilidad de aumentar su valor comercial, no obstante, no han tenido los resultados esperados. La Bioinformática es una herramienta que utiliza los datos biológicos para estudiar los problemas biológicos en una amplia gama de escalas como la evolución, estructura, función y regulación de los ácidos nucleicos y proteínas con la ayuda de métodos informáticos. En el presente trabajo se realizó el análisis *in silico* centrado en los dominios conservados, estructura, función y características bioquímicas para entender el mecanismo de regulación de la actividad de la proteína F3'5'H. Este tipo de análisis, proporciona, no solo la clonación de este gen en diversos organismos, sino que facilita la obtención de flores de color azul. La investigación inició con la secuencia de nucleótidos del gen F3'5'H de *Petunia hybrida* (EF371021), se determinaron los marcos de lectura abiertos (ORF's), los cuales una vez analizados, resultó una proteína de 506 aa. Se hizo el análisis comparado de identidad con 40 secuencias de proteínas de plantas, por medio de la base de datos del NCBI (Ref Seq), EMBL (Uniprot y PDB) y PIR, resultando estar conservado con la familia del citocromo P450 presentando tres dominios conservados (ancla, bisagra y globular) con cinco motivos conservados, el de residuos de prolina (PPGP) en el dominio de bisagra y en el dominio globular los de unión al oxígeno (AGTDTS), EXXR, unión al hemo (FGAGRRICAG) y el representativo del P450 (PERFL). El perfil hidrofóbico ayudo a identificar la presencia de la membrana y a predecir la solubilidad, estructura secundaria con 16 alfa hélices y 9 beta plegadas y la estructura terciaria que presenta los dominios A y B. Para la futura clonación del gen F3'5'H, se recomiendan las etapas 3 y 4 de desarrollo floral ideales para la extracción del ARN total, diseñándose dos pares de oligonucleótidos para la amplificación del gen.



## ABSTRACT

The color of the flower is one of the most important traits of ornamental plants. In related pigments wider range of colors in the flowers stand flavonoids, anthocyanins being their main compounds of which have focused on research to express the enzyme flavonoid 3', 5'-hydroxylase involved in the biosynthesis of delphinidin and responsible the color blue. As a result we have seen there is an excellent correlation between the pigment and the presence of color, giving the possibility to increase its commercial value, however, have not had the expected results. Bioinformatics is a tool that uses biological data to study the biological problems in a wide range of scales like the evolution, structure, function and regulation of proteins and nucleic acids with the help of computational methods. In the present work we performed *in silico* analysis focused on the conserved domains, structure, function and biochemical characteristics to understand the mechanism of regulation of protein activity F3'5'H. This type of analysis provides not only the cloning of this gene in various organisms, but also facilitates obtaining blue flowers. Research began with the nucleotide sequence of the F3'5'H gene of *Petunia hybrida* (EF371021), identified open reading frames (ORF's), which when analyzed proved to be a protein of 506 aa. Comparative analysis was made identity with 40 protein sequences from plants, through the NCBI database (Ref Seq), EMBL (Uniprot and PDB) and PIR, resulting to be preserved with the cytochrome P450 family presenting three domains preserved (anchor, hinge and globular) with five conserved motifs, the proline residue (PPGP) in the hinge domain and the globular domain binding the oxygen (AGTDTS) EXXR, heme binding (FGAGRRICAG) and representative P450 (profile). Hydrophobic profile helped identify the presence of the membrane and to predict the solubility, secondary structure with 16 alpha helices and 9 beta pleated and tertiary structure having the domains A and B. For future F3'5'H gene cloning are recommended stages 3 and 4 of flower development ideal for total RNA extraction, being designed two pairs of oligonucleotides for amplification of the gene.



## CAPITULO I: ANTECEDENTES

### 1.1 IMPORTANCIA ECONÓMICA DE LA FLORICULTURA

La floricultura representa una actividad de gran importancia económica siendo una de las áreas que más destacan dentro del sector agropecuario internacional, debido al uso de este recurso natural que comprende el cultivo, producción y comercialización de flores (Ortiz y Larque, 1999; SAGARPA, 2008; AIPH, 2011).

El mayor porcentaje del comercio internacional de flores, se registra en Europa, siendo los países bajos los de mayor participación con 4, 110 billones de euros en valor de producción e Italia con la superficie cultivada abarcando 11, 318 ha. El valor total de la producción de flores se ubica en los 25, 918 millones de Euros, siendo la flor de corte el principal volumen de comercialización, seguido por las plantas en maceta, mientras que el total mundial de hectáreas cultivadas abarca las 564,755 ha (cuadro 1.1; SAGARPA, 2006; AIPH, 2011).

	<b>SUPERFICIE CULTIVADA</b>	<b>VALOR DE PRODUCCIÓN</b>
	<b>(Ha)</b>	
	<b>Total</b>	<b>Millones de Euros</b>
Austria	419	160
Bélgica	1 272	229
República Checa	612	110
Dinamarca	367	304
Finlandia	199	102
Francia	5 723	956
Alemania	7 167	1 652
Grecia	1 095	176
Guernsey	25	39
Hungría	640	33

Irlanda	61	18
Italia	11 318	1 467
Países bajos	7 664	4 110
Noruega	113	25
Polonia	4 593	101
Portugal	1 036	218
España	2 693	858
Suecia	135	65
Suiza	508	276
Reino Unido	5 246	441
<b>Europa</b>	<b>50 886</b>	<b>12 000</b>
Israel	2 700	214
Turquía	1 326	27
<b>Medio Este</b>	<b>4 026</b>	<b>250</b>
Etiopía	2 000	149
Kenya	2 180	344
Morocco	165	
África del Sur	11 461	73
Tanzania	120	23
Uganda	174	28
Zambia	125	18
África	16 300	600
Australia	5 335	175
China	133 767	3 406
China Taipéi	5 294	199
Hong Kong	153	
India	183 000	
Japón	18 800	2 512
Corea (República)	7 185	598
Malasia	2 000	20

Filipinas	670	3
Singapur	312	27
Sri Lanka		8
Tailandia	9 280	62
<b>Asia/Pacífico</b>	<b>398 321</b>	<b>8 245</b>
Canadá	838	657
Estados Unidos	17 537	3 115
<b>Norte América</b>	<b>18 375</b>	<b>3 772</b>
Brazil	51 437	454
Colombia	7 900	1 104
Costa Rica	4 500	31
Ecuador	5 893	450
México	23 417	281
<b>América Central/ Sur</b>	<b>93 147</b>	<b>2 320</b>
<b>MUNDO</b>	<b>564 755</b>	<b>25 918</b>

**CUADRO 1.1 SUPERFICIE CULTIVADA Y VALOR DE PRODUCCIÓN MUNDIAL DE FLORES DE CORTE Y DE MACETA. AIPH / UNION FLEURS: INTERNATIONAL STATISTICS FLOWERS AND PLANTS, 2011.**

La importancia de la demanda de flores a nivel mundial se puede ver reflejada en el valor del comercio de las mismas. En el mundo existen países con alta producción de flores, tanto a cielo abierto como en invernadero; sin embargo, no todos ellos tienen un peso significativo en el mercado mundial ya sea por baja calidad del producto o porque la mayoría produce para abastecer el mercado interno. Hay que destacar que el comercio mundial de flores no solo se limita a las transacciones de flores frescas, sino que también han tomado importancia otros aspectos, tal es el caso de la comercialización de semillas mejoradas genéticamente, la tecnología y el conocimiento para mejores técnicas de cultivo de los productores (*know-how*, por sus siglas en inglés), el material de propagación, etc., donde países como Holanda ocupan un lugar relevante en este comercio ya que se han

dirigido recursos para el desarrollo de tecnología de producción y obtención de nuevas especies de plantas ornamentales (SAGARPA, 2006).

Se estima que los grandes progresos obtenidos en la industria florícola han dependido principalmente de la generación de nuevas tecnologías de producción, manejo de numerosas especies y cultivos e introducción de otras nuevas al mercado internacional. Dentro de estas tecnologías encontramos los reguladores del crecimiento, desarrollo de las plantas y mejoramiento genético (Ortiz y Larque, 1999).

Para el mejoramiento genético en especies ornamentales existe un gran interés en la incorporación de características tales como resistencia a insectos, resistencias fúngicas, resistencias virales, tolerancia a herbicidas y androesterilidad, que son específicas de aplicación en la mayoría de los cultivos. A estas se suman los caracteres específicos para la floricultura, como son la modificación del color y la forma de las flores, el incremento de la vida post cosecha, la modificación de la fragancia, el control de la floración y el mejoramiento de la eficiencia de enraizamiento (Escandón, 2005).

## **1.2 PIGMENTOS RESPONSABLES DEL COLOR DE LAS FLORES**

Los colores de las plantas que son percibidos por los seres humanos se conocen como “pigmentos” (Tanaka et al., 2008), siendo el color de la flor, el elemento clave en la selección del consumidor entre las variedades ornamentales disponibles en el mercado, cuya demanda está dirigida a la novedad y singularidad con colores específicos como es el color azul (Davies y Schwinn, 1997).

Los colores de las flores se desarrollan a través de diversos mecanismos, sin embargo, la poliacetilación, la glicosilación y diversos patrones de acilación existen en la naturaleza y una amplia variedad de diversidad estructural se encuentra en los pétalos de la flor, dependiendo de las especies de plantas y cultivares. En especies ornamentales la acumulación de los pigmentos se ha relacionado principalmente a una respuesta de la presión de selección de los polinizadores y por lo general se correlaciona con el tipo de pigmentos presentes en la vacuola, que está influenciada por factores abióticos como la luz, la temperatura y la escasez de agua (Ben-Meir et al., 2002), además de factores endógenos

tales como el pH vacuolar, la presencia de iones metálicos, copigmentación y forma de la célula también ejercen influencia sobre el color (Griesbach, 1997; Ben-Meir et al.2002; Mol et al., 1989).

Uno de los pigmentos más importantes el cual está involucrado en la fotosíntesis, es la clorofila, que también ayuda atraer a los insectos y otros animales, primero para favorecer la polinización y después para ayudar a la distribución de semillas (Went, 1978). Como ejemplo de flores color verde está la *Strongylodon macrobotrys* o enredadera de Jade con un color verde azulado.

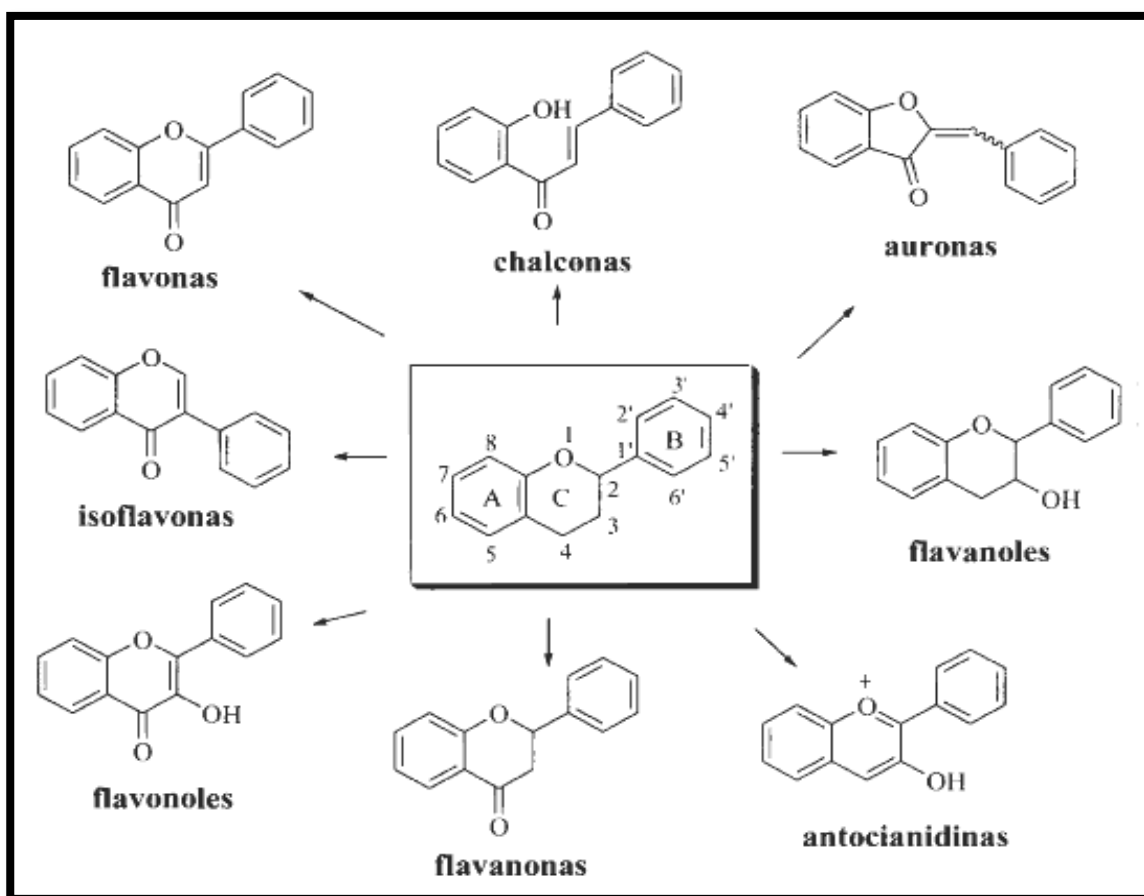
De los principales pigmentos relacionados con el color de la flor, las betalainas, son los pigmentos menos conocidos, probablemente por su relativa escasez natural encontrándose también en raíces y frutos. Entre las pocas fuentes ornamentales, están los del género de la *Opuntia*, *Hylocereus* y en especies de flores pertenecientes a las familias de Caryophyllales (Tanaka et al., 2008). Otros pigmentos que van del amarillo al rojo son los carotenoides, estos tienen una función esencial en la vida de las plantas, por ejemplo la provisión de sustratos para la biosíntesis del crecimiento de las plantas del regulador de crecimiento vegetal, el ácido abscísico ABA y la atracción de insectos que ayudan a la polinización (Villar- Martínez et al, 2007; Kishimoto y Ohmiya, 2009). También tienen una función importante en la nutrición humana y la salud, algunos son utilizados como colorantes de alimentos, productos cosméticos o farmacéuticos (Mayne, 1996; Kishimoto y Ohmiya, 2009). Dos especies de gran interés comercial, pertenecientes a esta clase de pigmentos son *Tagetes erecta* y *Chrysanthemum morifolium*.

Los responsables de la mayor gama de los colores en flores son los flavonoides, siendo también un punto clave en otros procesos fisiológicos, como la protección de la radiación UV-B, de defensa ante la infección microbiana y herbívora y la germinación de semillas (Holton y Cornish, 1995; Harborne y Williams, 2000; Winkel-Shirley, 2001; Ben-Meir et al., 2002). Actualmente representan el mayor grupo de metabolitos secundarios vegetales, este conocimiento ha permitido hacer un modelo para estudiar la expresión de genes y la regulación a nivel molecular (Cuevas et al., 2008; Tanaka et al., 2008). Pero su función más

significativa es la contribución a la planta específicamente con el color de la flor (Forkmann, 1991).

Esta clase de pigmentos son compuestos de bajo peso molecular (fig. 1.1), que comparten un esqueleto común de difenilpiranos (C6-C3-C6), compuesto por dos anillos de fenilos (A y B) ligados a través de un anillo C de pirano (heterocíclico). Los átomos de carbono en los anillos C y A se numeran del 1 al 8, y los del anillo B desde el 1' al 6'.

(Martínez-Flores et al, 2002; López et al, 2006; Russo y Sánchez, 2006).



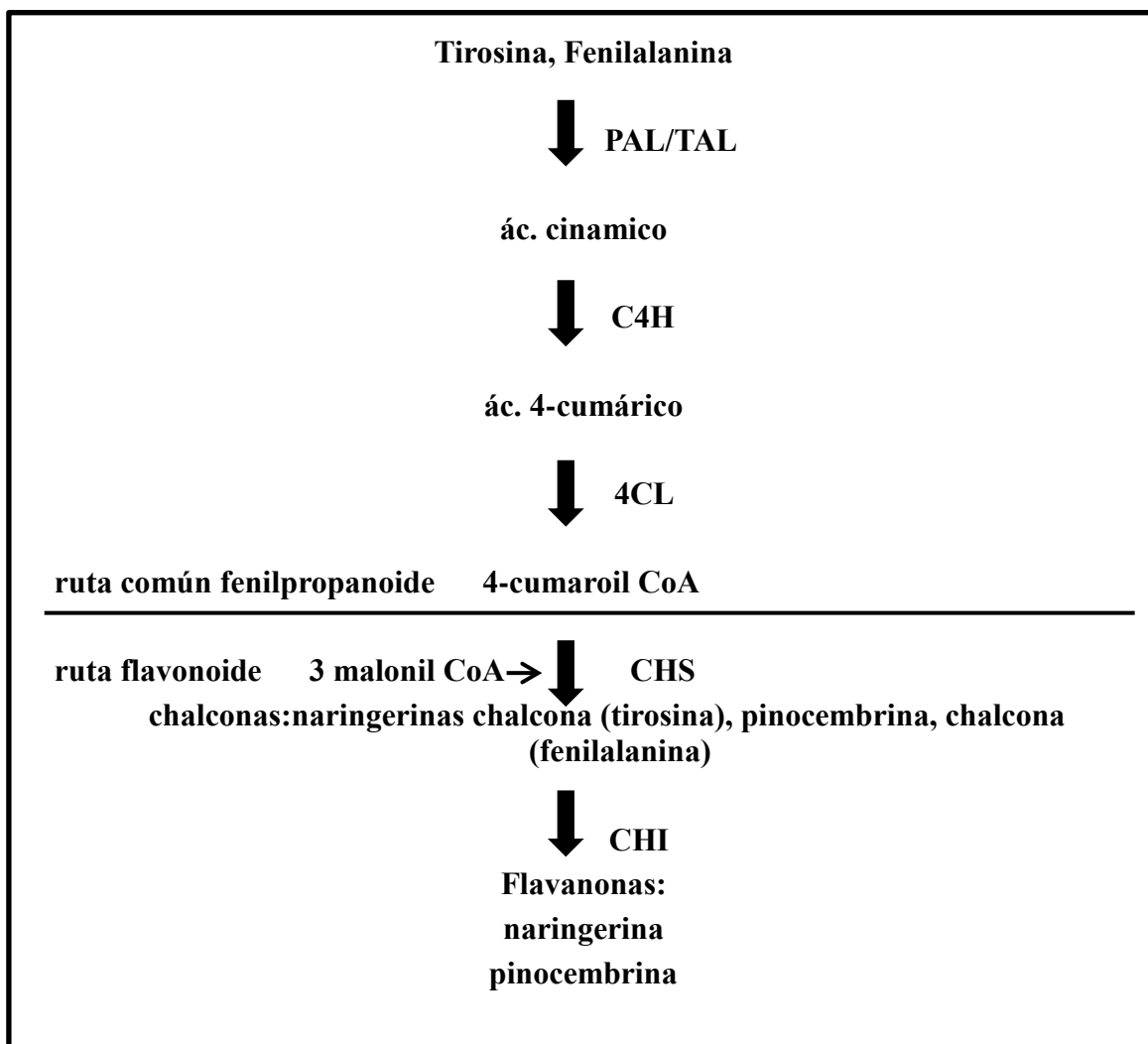
**FIGURA 1.1. ESTRUCTURA DEL FENILBENZOPIRANO DE LOS COMPUESTOS FLAVONOIDES Y DE LAS SUSTITUCIONES Y MODIFICACIONES QUE DAN ORIGEN LOS DIFERENTES TIPOS DE FLAVONOIDES (Puebla et al, 2004).**

La estructura básica de los flavonoides permite una multitud de patrones de modificación en el anillo C como se puede observar en la figura 1.1 y en función de sus características estructurales se pueden clasificar en (López *et al*, 2006; Russo y Sánchez, 2006):

- ❖ **Flavonas:** Poseen un grupo COOH en posición 4 y un doble enlace entre la posición 2 y 3. Ejemplo: la diosmetina.
- ❖ **Flavonoles:** Poseen un grupo COOH en posición 4 y un grupo -OH en posición 3. Ejemplo: quercitina.
- ❖ **Flavanonas:** Poseen un grupo COOH en posición 4. Ejemplo: la catequina.
- ❖ **Antocianidinas:** Tienen unido el grupo -OH en posición 3 pero además poseen un doble enlace entre los carbonos 3 y 4. Ejemplo: Delfinidina
- ❖ **Chalconas:** Difieren de todos los demás flavonoides por la ausencia del tercer anillo (anillo C). Ejemplo: 1-(4-metil-fenil)-3-(2,3-dimetoxifenil)-2-propenona (Mora, 2011).
- ❖ **Auronas:** Difiere un poco de la estructura general teniendo un heterociclo de cinco miembros en lugar de seis. Ejemplo: aureusidina (Puebla et al, 2004).

Como se observa en la figura 1.2, los precursores de los flavonoides proceden de la ruta de los fenilpropanoides, que inicia con la conversión de fenilalanina en ácido cinámico por la PAL (fenilalanina-amonio liasa). El ácido cinámico es convertido en ácido 4-cumárico (p-cumárico) por la C4H (cinamato-4 hidroxilasa). En algunas plantas, la PAL exhibe también actividad de TAL (tirosina-amonio liasa) sobre la tirosina para generar directamente ácido 4-cumárico. En el siguiente paso, el ácido 4-cumárico es transformado en 4-cumaroil-CoA por la 4CL (4-cumarato CoA ligasa). La 4-cumaroil-CoA, un producto central de la ruta fenil propanoide, es condensada con 3 moléculas de malonil CoA para

formar naringenina-chalcona o pinocembrina chalcona (de tirosina o fenilalanina respectivamente) por la enzima CHS (chalcona sintetasa). Después de esta reacción, la CHI (chalcona isomerasa) cicliza a la naringenina o pinocembrina chalcona mediante la isomerización estereoespecífica para formar las (2S) flavanonas naringenina o pinocembrina. Las flavanonas son ulteriormente modificadas por enzimas de la ruta flavonoide para generar una amplia diversidad de compuestos por medio de glicosilación, acilación o metilación del núcleo de derivados (Drago, 2007).



**FIGURA 1.2. RUTA COMÚN FENILPROPANOIDE Y RUTA BIOSÍNTÉTICA DE LOS FLAVONOIDES.** PAL (fenilalanina-amonio liasa), C4H (cinamato-4 hidroxilasa,



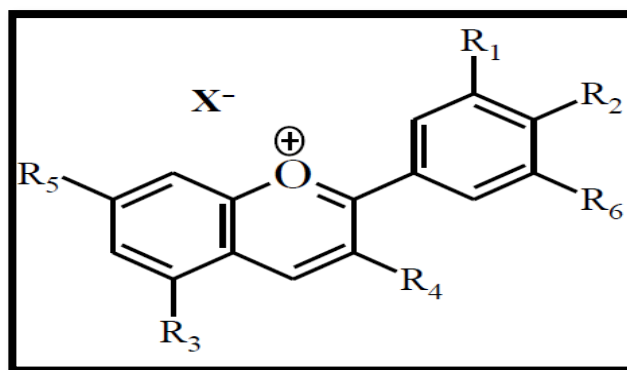
TAL (tirosina-amonio liasa), 4CL (4-cumarato CoA ligasa), CHS (chalcona sintetasa), CHI (chalcona isomerasa). Modificado de Drago, 2007.

De los compuestos más importantes con respecto al color de la flor están las antocianinas, flavonoles, flavones, chalconas y auronas (Forkmann, 1991).

### 1.2.1 ESTRUCTURA Y FUNCIÓN DE LAS ANTOCIANINAS

Dentro de las clases de flavonoides, los más importantes son las antocianinas (del griego *anthos* flor y *kyanos* azul). Estas, se acumulan en las vacuolas de la epidermis o células sub-epidérmicas, produciendo coloración en las flores como el rosado, magenta, rojo, violeta y el azul (Ben- Meir et al., 2002).

Químicamente se diferencian de otros polifenoles por presentar varios grupos OH. Su estructura es el 2- fenilbenzopirilio (fig. 1.3). Las diferencias individuales entre las antocianinas dependen del número de grupos hidroxilo que están unidos a la molécula y a la posición de esa unión (Leguizamón et al, 2005). Los compuestos más comunes de antocianinas son: pelargonidina, cianidina, delphinidina, peonidina, malvidina y petunidina. El color particular de cada antocianina depende del número y orientación de los grupos hidroxilos y metoxilos (fig. 1.4). Si ocurre un incremento en la hidroxilación produce un color azul, mientras un incremento en la metoxilación produce un color rojo (Cuevas et al, 2008; Garzón, 2008).



**FIGURA 1.3. REPRESENTACIÓN GRÁFICA DE LA ESTRUCTURA GENERAL DE UNA ANTOCIANINA DESCRITAS EN EL CUADRO 1.4 (Garzón, 2008).**

ANTOCIANIDINA	GRUPO RADICAL						FUENTE
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	
DELFINIDINA	OH	OH	OH	OH	OH	OH	<i>Delphinium</i>
PETUNIDINA	OMe	OH	OH	OH	OH	OH	<i>Petunia</i>
MALVINIDINA	OMe	OH	OMe	OH	OH	OH	<i>Malva</i>
PEONIDINA	OMe	OH	OH	OH	OH	--	<i>Paeonia</i>
CIANIDINA	OH	OH	OH	OH	OH	--	<i>Centaurea</i>
PELARGONIDINA	--	OH	OH	OH	OH	--	<i>Pelargonium</i>
LUTEOLINIDINA	OH	OH	OH	--	OH	--	<i>Rechsteineria</i>
APIGENIDINA	--	OH	OH	--	OH	--	<i>Rechsteineria</i>

**CUADRO 1.2. CARACTERÍSTICAS DE LOS TIPOS DE ANTOCIANIDINAS.**

Resultante según el número de hidroxilos que presentan y la especie de donde se extrajo, formando así su estructura particular (Garzón, 2008).

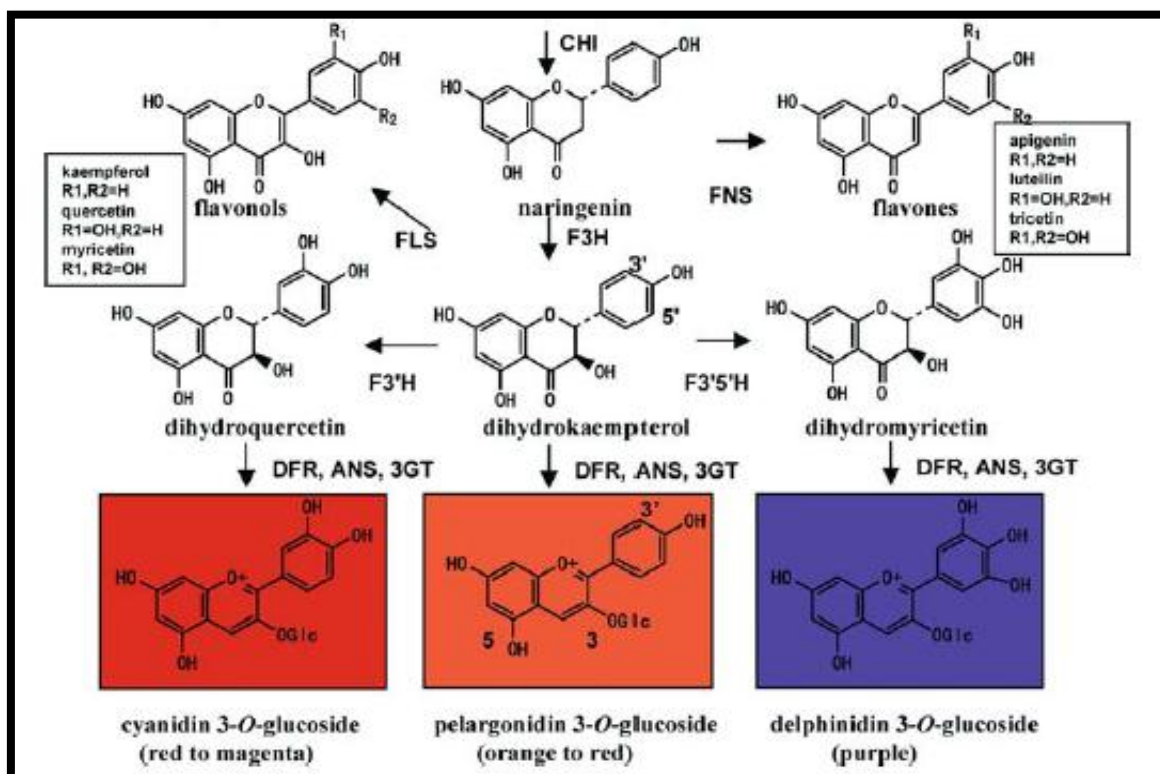
**1.2.2. BIOSÍNTESIS DE ANTOCIANINAS**

La biosíntesis de antocianinas es un aspecto clave del desarrollo de la flor en muchas angiospermas, proporcionando una de las mayores influencias sobre la elección de los polinizadores. En algunas especies la evolución ha dado lugar a estructuras antocianínicas complejas que proporcionan flores brillantes de colores, mientras que en otras especies sofisticadas combinaciones de patrones de pigmento y formas florales se han desarrollado para atraer a los polinizadores. Actualmente se conocen los genes que codifican las enzimas de la biosíntesis de antocianinas y copigmentos, como la regulación temporal y espacial de la producción de antocianinas (Tanaka et al, 2008; Holton y Cornish, 1995).

La ruta de biosíntesis de las antocianinas presentes en las plantas está completamente caracterizada (fig. 1.5) y sirve como modelo para el análisis de la regulación transcripcional de genes de plantas (Dooner et al., 1991; Harbourne 1967, Holton y Cornish, 1995; Van der Meer et al., 1992). Los análisis genéticos de la ruta de antocianinas en organismos como el maíz (*Zea mays*), boca de dragón (*Antirrhinum*), Petunia (*Petunia hybrida*) y Arabidopsis (*Arabidopsis thaliana*) indican que la ruta se compone de enzimas fundamentales:

Chalcona sintasa (CHS), Chalcona isomerasa (CHI), Flavona 3 Hidroxilasa (F3H), Flavonoide 3',5' Hidroxilasa (F3'5'H), Flavonoide 3' Hidroxilasa (F3'H), Glucosil transferasa (GT), Dididroflavonol reductasa (DFR), Antocianidina sintetasa (ANS), Antocianidina aciltransferasa (AAT) y antocianidina metiltransferasa (AMT) ( Dooner et al 1991; Holton y Cornish, 1995; Mol et al 1999).

La ruta biosintética de las antocianinas comienza con la catalización de la enzima CHI que lleva a cabo la isomerización de chalcona a naringerina, esta puede desviarse si es catalizada por la FLS originando los flavonoles o por la FNS que dirige la ruta a la síntesis de los flavones, pero si interviene la F3H se obtiene el dihidrokamferol que es catalizado por DFR, ANS y 3GT logrando directamente a la pelargonidina 3-O-glucosida responsable de los colores naranja a rojo. También el dihidrokamferol puede ser catalizado ya sea por la F3'H resultando la dihidroquercitina siendo catalizada por la DFR, ANS y 3GT como producto está la cianidina 3-O-glucosida responsable de los colores rojo a magenta; y por la catalización de la F3'5'H la dihidromirecitina catalizada a su vez por DFR, ANS y 3GT el compuesto resultante es la delphinidina 3-O-glucosida responsable de los colores morados a azules (Fig. 1.4) (Tanaka et al, 2005; Garzón, 2008).



**FIGURA 1.4. RUTA GENERAL DE LA BIOSINTESIS DE ANTOCIANINAS.** CHI (chalcona isomerasa), FLS (flavonol sintasa), FNS (flavonona sintasa), F3H (flavanona 3'-hidroxilasa), DFR (dihidroflavonol reductasa), F3'5'H (flavonoide 3'5' hidroxilasa), F3'H (flavonoide 3' hidroxilasa), 3GT (flavonoide glucosil-3- transferasa), ANS (antocianidina sintetasa), FGT (UDP glucosa flavonoide 3-O-glucosiltransferasa) Tanaka et al, 2005.

En 2010, Su y Hsu demostraron, que el gen P450 de *Phalaenopsis*, promueve eficazmente el contenido de antocianinas en pétalos, aplicando no solo para monocotiledóneas sino para dicotiledóneas también, estando implicada la acción de las enzimas P450 en la síntesis de antocianinas la cual es directamente visible a través del color de la flor (Kitada et al, 2001; Tanaka, 2006; Tanaka et al, 2009).

### **1.2.3. EL COLOR DE LA FLOR Y LOS CITOCROMOS P450**

De los compuestos derivados de las antocianinas, la delphinidina y cianidina, difieren en los patrones de hidroxilación del anillo B de los flavonoides, por el control de dos citocromos P450, la F3'5'H (CYP75A) y la F3'H (CYP75B) respectivamente (Kitada et al, 2001; Tanaka, 2006; Tanaka et al, 2009). En plantas, esta familia de enzimas Citocromo P450, participa en la biosíntesis de productos secundarios (flavonoides, alcaloides) y de hormonas así como la desintoxicación de herbicidas (Werck-Reichhart y Feyereisen, 2000; Kitada et al, 2001; Tanaka, 2006; González, 2007; Tanaka et al, 2009).

Las oxidaciones catalizadas por el P-450 son reacciones de monooxigenación dependientes de NADPH unidas a la membrana que contienen hemo-enzimas que están involucradas en una amplia gama de reacciones de hidroxilación y para las que utiliza oxígeno molecular (Werck-Reichhart y Feyereisen, 2000; Wang et al, 2006). Pueden dividirse en cuatro clases dependiendo de cómo los electrones del NAD(P)H pueden ser liberados al sitio catalítico:

- ❖ **Clase I:** Requieren tanto de la FAD-reductasa como de la redoxina de hierro-azufre.
- ❖ **Clase II:** Requieren solo de la FAD/FMN P450 reductasa para la transferencia de electrones.

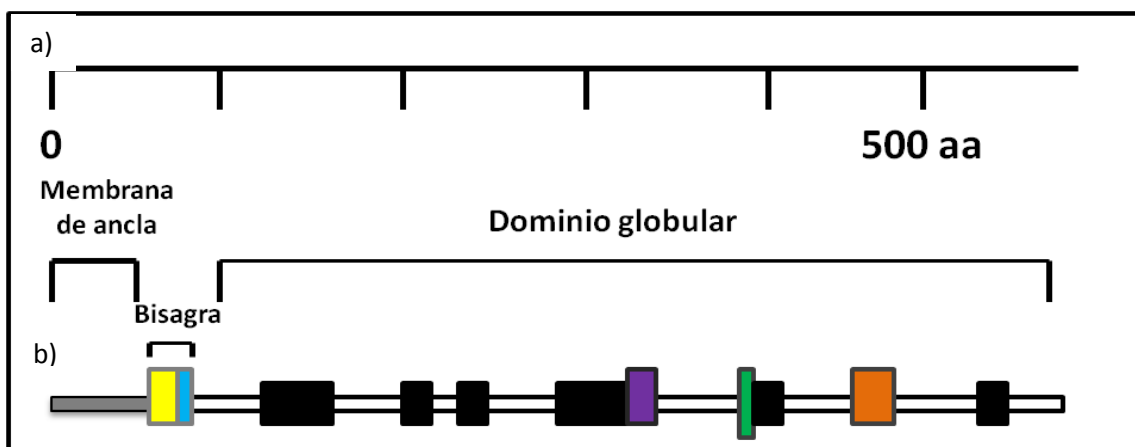
- ❖ **Clase III:** Son autosuficientes y no requieren de la donación de electrones.
- ❖ **Clase IV:** Reciben los electrones directamente del NAD(P)H.

Las enzimas clase II son las más comunes en los P450 eucariotas y en plantas están involucradas entre otras funciones en la síntesis de pigmentos de la flor. Las reductasas NADPH-P450, se disocian e independizan del anclaje en la cara exterior del retículo endoplasmático (RE) por las anclas amino-terminales hidrofóbicas. Un donante de electrones adicional de anclaje carboxi-terminal al RE, el citocromo  $b_5$ , transmite electrones desde el NAD(P)H, y se ha encontrado que aumenta la actividad de las enzimas P450 (Vetten et al, 1999; Werck-Reichhart y Feyereisen, 2000).

Las enzimas P450, comparten características estructurales o topológicas (fig. 1.5) que permiten su identificación como miembros de la familia aunque las secuencias solo presenten menos del 20% de similitud, teniendo una alta conservación en su estructura, situándose en el núcleo de la proteína alrededor del hemo y reflejando un mecanismo común de transferencia de electrones y protones, así como de activación de oxígeno. El núcleo conservado está formado por una agrupación de cuatro helicoides (D, E, I y L), dos conjuntos de  $\beta$  plegadas y una helicoide enrollada o “coil” llamada meandro. Estas regiones comprenden al hemo de unión al bucle que contiene la secuencia más conservada de los P450 Phe-X-X-Gly-X-Arg-X-Cys-X-Gly (donde X es un aminoácido no conservado), localizado en la cara proximal del hemo justo antes de la hélice L, en el dominio globular. Incluye un residuo de cisteína totalmente conservado que sirve para ligar al quinto hemo hierro. También el motivo Glu-X-X-Arg absolutamente conservado en la hélice K, en el lado proximal del hemo y probablemente necesario para estabilizar la estructura del núcleo. Por último, la parte central de la hélice I, que contiene otra secuencia consenso considerada como el sello del P450 (Ala/Gly-Gly-X-Asp/Glu-Thr-Thr/Ser), la cual corresponde a la ranura de transferencia de protones en el lado distal del hemo (Hasemann et al. 1995; Graham y Peterson, 1999; Werck-Reichhart y Feyereisen, 2000).

Las regiones más variables están asociadas con el lado amino-terminal de anclaje, la orientación de las proteínas unidas a la membrana o al sustrato de unión y reconocimiento; estas últimas regiones se encuentran cerca del canal de acceso-sustrato y el sitio catalítico y

a menudo se refieren, como sitios de reconocimiento de sustrato o SRSs. Se les describe como flexibles, moviéndose en el sustrato de unión con el fin de favorecer la reacción catalítica. Otras variaciones se reflejan en las diferencias de donación de electrones, la reacción catalizada o localización de la membrana. La mayoría de los P450s eucariotas se asocian con las membranas microsomales, y es muy frecuentemente que tengan un clúster de prolinas (Pro-X-Pro) que forman una bisagra, precedida por un grupo de residuos básicos (la señal de frenar la transferencia) entre la membrana amino-terminal hidrofóbica, segmento de anclaje y la parte globular de la proteína. La interacción de membrana adicional, parece estar mediado esencialmente por una región, localizada entre las hélices F y G, mostrando aumento de la hidrofobicidad (Gotoh, 1992; Hasemann et al. 1995; Graham y Peterson, 1999; Werck-Reichhart y Feyereisen, 2000; William et al, 2000).



**1.5. ESTRUCTURA PRIMARIA DE LA PROTEÍNA P450.** a) Tamaño de la proteína que abarca los 500 aminoácidos. b) La estructura consta de regiones hipervariables como son la membrana de ancla con un segmento hidrofóbico  y en el dominio globular los sitios de reconocimiento-sustrato (SRSs) . Las características conservadas las encontramos en el dominio de unión (bisagra) que consta de un motivo de residuos básicos  y otro de prolinas  (a menudo PPGP) y en el dominio globular el motivo de unión al oxígeno , motivo Glu-X-X-Arg  y el motivo de unión al hierro  (Werck-Reichhart y Feyereisen, 2000).

Sin embargo, el final de un color visible en la flor es generalmente una combinación de una serie de factores. Una forma es conociendo con más exactitud la función, estructura y regulación que la proteína realiza. La búsqueda del color azul es un reto que no ha sido esclarecido del todo y el estudio de la biosíntesis de delfinidina, es un punto clave, como también el conocer como la misma proteína se comporta en diferentes especies y tejidos.

### **1.3 MEJORAMIENTO GENÉTICO EN LA MANIPULACIÓN DEL COLOR DE LA FLOR**

La genética ha cautivado el interés del hombre a lo largo del tiempo y la agricultura ha utilizado el fitomejoramiento como una manera rentable para mejorar la productividad de la población cada vez más exigente. Una manera eficaz de crear variedades nuevas, es a través del mejoramiento genético tradicional. Este tipo de mejoramiento se ha utilizado con bastante éxito, dando lugar a una diversidad de colores para la mayoría de las variedades comerciales de flores, sin embargo, estas técnicas están limitadas por las restricciones de la reserva genética de una especie en particular y por esta razón una sola especie no tiene todo el rango de colores. Así, la introducción de nuevos colores a través del fitomejoramiento ha tenido un gran impacto en la industria atrayendo la atención de muchos investigadores, estableciendo una base fuerte en este tipo de investigación así como la extensa información disponible en la genética y la bioquímica de la biosíntesis de los pigmentos, (Davies y Schwinn, 1997; Tanaka y Brugliera, 2005). Las cruas seguirán siendo la forma más común de aumentar la variación genética, pero de particular interés está la capacidad de transferir, además de los genes importantes, caracteres cuantitativos tales como la productividad y otras características de interés para el ser humano (Baenziger y Al-Otayk, 2007).

La innovación en el color de las flores, en particular la generación de variedades azules, es objeto de estudio en el mejoramiento de plantas ornamentales. Uno de los más importantes es el Flavonoide 3',5' hidroxilasa, este gen esta involucrado en la biosíntesis de delfinidina, responsable de los colores morado-azul y ha sido clonado a partir de varias flores (*Petunia*, *Gentiana*, *Salvia*, *Nemophila*, *Centaurea*, *Phaleanopsis*, *Hydrangea*, *Nierembergia*, *Campanula*, entre muchas más). La introducción de este gen en pétalos de

color rojo produce un color azul, usándose como técnica para la obtención de nuevas variedades de plantas ornamentales (Yoshida et al, 2009).

Un ejemplo se da en flores de relevancia económica, tales como las rosas, crisantemos y claveles las cuales no producen delfinidina, debido a la falta de este gen *F3'5'H* en sus flores. Por lo tanto, no se puede producir una rosa o clavel azul por cultivo tradicional. La expresión heteróloga del *F3'5'H* en diversas especies ha dado como resultado diferentes niveles de éxito. En *tobacco*, la introducción de un gen *F3'5'H* (Ka1), aislado en *Campanula medium* resultó en flores con un mayor porcentaje de delfinidina que cuando los genes *F3'5'H* de *Petunia* o *Lisianthus* fueron introducidos (Okinaka et al., 2003). Esto es presumible debido a que la actividad de la enzima es más eficiente en *Campanula*. Sin embargo, cuando un gen de petunia del citocromo b5 junto con el *F3'5'H* de petunia se expresa en una línea de clavel (Fukui et al, 2003), el resultado fue una notable mejora en el nivel de producción de delfinidina y un cambio en el color de la flor de un color rosa y rojo para malva y morado. En abril de 2005, se anunció la producción de una rosa azul por la introducción de tres transformaciones que forman simultáneamente a las rosas: un gen de *F3'5'H* de pensamiento para producir delfinidina, un gen anti-rosa DFR que silencia el DFR endógeno para producir un fondo blanco, y un gen DFR de iris para hacer modificaciones específicas de delfinidina.

A pesar del éxito de los trabajos recientes, es muy difícil generar o suprimir compuestos flavonoides específicos de forma altamente controlada por ingeniería metabólica de la ruta de los flavonoides. El aislamiento de genes implicados en la ruta biosintética de los flavonoides y de secuencias promotoras permite conseguir grandes avances, además, un mejor conocimiento de los mecanismos implicados en la co-pigmentación y el pH vacuolar, así como la disponibilidad de los genes respectivos y la mejora de sistemas de transformación para plantas ornamentales, haciendo la modificación del color de las flores más factible (Forkmann y Martens, 2001; Koes et al, 2005; Tanaka et al, 2008).

Durante muchos años, el gen fue considerado como la unidad modular fundamental de la biología. Como tal, un gen es capaz de la transferencia de un fenotipo particular a otro



organismo. Sin embargo, ahora sabemos que los genes muestran una modularidad más compleja en forma de promotores, Marcos de Lectura Abiertos (ORF) y elementos reguladores; los ARNs mensajeros, contienen secuencias importantes para la orientación correcta intracelular y la degradación y las proteínas a menudo contienen secuencias de direccionamiento, centros activos y secuencias de degradación. La Biología Molecular y la Biotecnología constituyen dos áreas científicas prioritarias como desarrollo de Innovación Tecnológica. El desarrollo en ambas áreas está estrechamente relacionado, y potenciar la investigación genómica y postgenómica así como la bioinformática, herramienta imprescindible para el desarrollo de estas, es un objetivo primordial de esta era. Debido al extraordinario avance de la genética molecular y la genómica, la Biología Molecular se constituye como herramienta estratégica del futuro inmediato (Claverie y Notredame, 2007).

Actualmente, varias técnicas experimentales de alto rendimiento en biología molecular pueden proveer mediciones a escala-genoma a partir de moléculas biológicas que existen dentro de la célula tales como genes, proteínas, ARN, metabolitos y otras moléculas, que han tenido resultado en un enorme cantidad de datos acerca de los componentes (Chen et al, 2009). Las tecnologías de la información juegan un papel fundamental en la aplicación de los desarrollos tecnológicos en el campo de la genética (Claverie y Notredame, 2007).

#### **1.4 LA BIOLOGIA DE SISTEMAS**

La biología de las últimas décadas ha incorporado la informática a su plan de trabajo. Los genes, el ARNm y las proteínas, se conocen como las moléculas biológicas o componentes básicos y sus complejas relaciones e interacciones entre ellos, son responsables de una gran diversidad de funciones celulares. Estos componentes básicos en un sistema celular son moléculas individuales el cual, cada una afecta a la otra por sus interacciones en par. La cascada de estas interacciones forma una estructura local (ruta linear o una subred), el cual transforma las modificaciones locales en respuestas funcionales (Chen et al, 2009).

En términos de interacciones, cada tipo de red molecular, es reunido por las diferentes interacciones en par:

- ❖ De Regulación de la Transcripción: interacciones Factores de Transcripción-ADN
- ❖ De Regulación de Genes: interacciones gen-gen o interacciones génicas
- ❖ De Proteína: interacciones proteína-proteína
- ❖ Metabólicas: interacciones enzima-substrato
- ❖ Señalización: interacciones molécula-molécula

El resultado del conjunto de datos por estas técnicas experimentales, hace pasar un flujo de información centrados en el dogma de la biología molecular e incluyen el genoma, transcriptoma, proteoma, metaboloma, localizoma y los componentes interactoma que en conjunto se conocen como datos “oma” (“omic” por sus siglas en inglés) y proveen descripciones integrales de todos los componentes e interacciones dentro de la célula. En contraste la información de los componentes tales como los datos genómicos y proteómicos proporcionan contenidos moleculares específicos para un sistema celular, los datos de las interacciones en par incluyen interacciones ADN-proteína, proteína-proteína y ligación-proteína (substrato-enzima), los cuales determinan la conectividad local, que existe entre las especies moleculares facilitando una plataforma informática dentro del sistema celular (Chen et al, 2009).

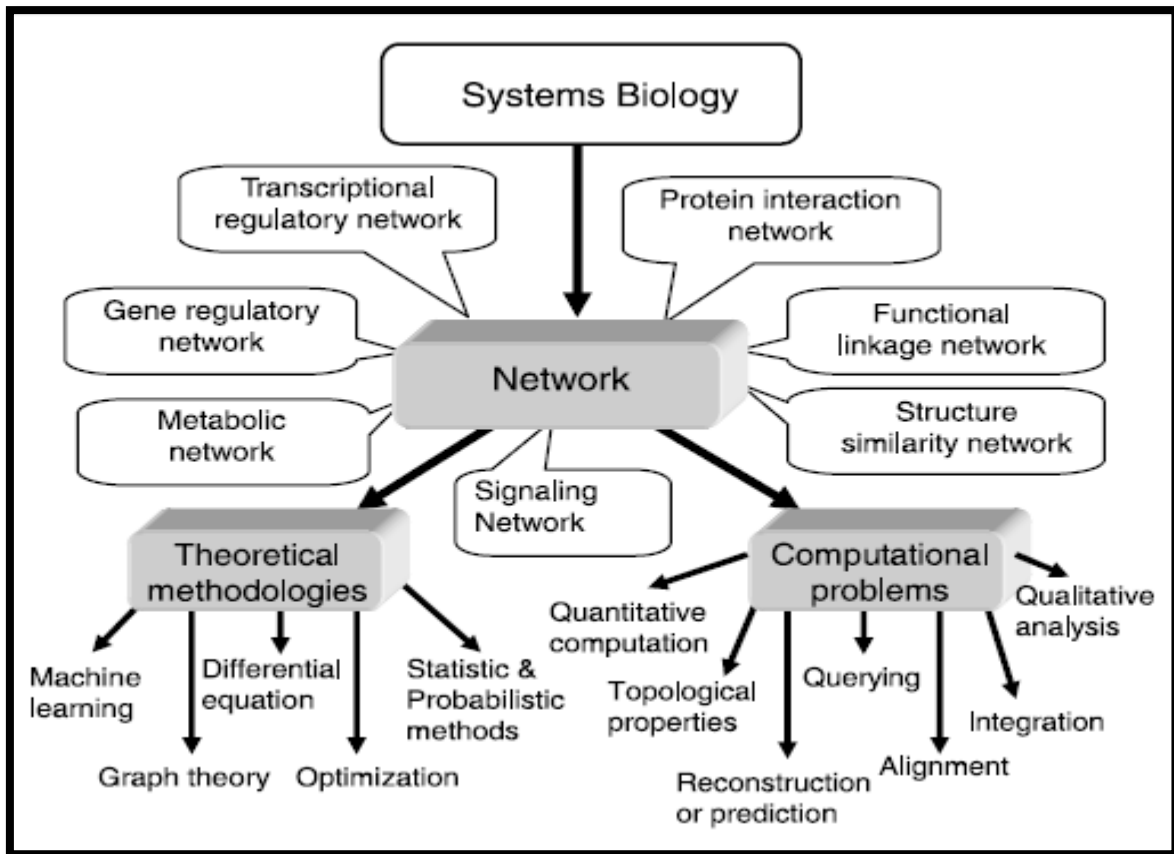
Los subsecuentes datos de la función están relativamente cerca de la interacción de datos desde muchos procesos biológicos en células que no son realizados por componentes individuales, pero que a través de regulaciones del gen, traducción de señales e interacciones entre biomoléculas, es que estas interacciones locales de esos componentes se unen dentro de una red informática y son responsables últimamente de las formas y funciones de un organismo. Con los diversos datos de interacción disponibles, la investigación biológica se enfoca en ser transformados de análisis de componentes individuales a estudiar la red informática global desde una perspectiva sistemática. En particular, un sistema celular puede verse como un trabajo informático-sistema biológico.

Con esto, un importante cambio en la biología, es entender la organización de la función celular por investigaciones dinámicas de la estructura y función de la compleja red informática biomolecular en sistemas celulares vivientes (Chen et al, 2009).

Los métodos experimentales de alto rendimiento en biología molecular resultan en una enorme cantidad de datos, incluyendo interacciones, redes informáticas y rutas. Por lo tanto, es crucial que los matemáticos y científicos informáticos proporcionen las herramientas computacionales para revelar los mecanismos biológicos esenciales desde una perspectiva del sistema. Para afrontar este reto, en lugar de analizar los componentes individuales o aspectos parciales del organismo, las redes de sistemas biológicos, son el estudio de un organismo visto como una interacción dinámica de redes de genes, proteínas y reacciones bioquímicas por el desarrollo de metodologías teóricas sofisticadas y herramientas computacionales. El objetivo de la red de la biología de sistemas es extraer el conocimiento a partir de datos en red o generados desde técnicas de alto rendimiento mediante la explotación de características especiales del sistema biológico, y obtener información biológica más que interpretar de manera sistemática. Como una disciplina estrechamente relacionada con la biología de sistemas, la red biológica enfatiza las interacciones locales y las redes mundiales de moléculas que caracterizan a diversos sistemas biológicos y los intentos por comprender la biología desde un punto de vista de las propiedades de los sistemas globales y locales de redes moleculares, ofreciendo una descripción cuantificable de las interacciones y redes. El objeto de estudio de la investigación de las redes de la biología de sistemas o del trabajo informático biológico es que sea rico y diverso (Chen et al, 2009).

Por otro lado, problemas como proporcionar una estimación robusta de los datos reportados de las redes biológicas para los biólogos; en las redes de alineamientos y consultas de datos, poder comparar con muchas especies y a la vez extraer el conocimiento importante relacionado con la evolución o poder encontrar la vía en cuestión en una red de interacción de proteínas y el orden para entender la estructura y función de la célula viva, conlleva a investigar la estructura y dinámica de estas redes biológicas. Este es el reto o análisis cualitativo de las redes biomoleculares. En contraste con este tipo de estudios, los

cuantitativos pueden predecir directamente una simulación de los comportamientos dinámicos de las células vivas concurrendo un tema importante, relacionado con el desarrollo de algoritmos de cálculo altamente eficientes, tanto en modelos dinámicos estocásticos como deterministas. En redes biomoleculares, el cómo se relacionan los patrones topológicos (motivos) a funciones biológicas específicas, o bien la reconstrucción de rutas metabólicas e identificación de subredes activas de una larga cadena de reacciones bioquímicas, el diseño o construcción de una red biológica sintética o representativo de un todo o una parte del sistema celular, todos ellos, son las dificultades que se esperan para resolver por medios computacionales, los cuales son la investigación principal que se centra en las redes de biología de sistemas (fig. 1.6).



**FIGURA 1.6. EL ENFOQUE DE LA INVESTGACIÓN EN LAS REDES DE LA BIOLOGÍA DE SISTEMAS (Chen et al, 2009).**

Con cada vez más datos acumulados de tecnologías de alto rendimiento, las redes biomoleculares y sus roles funcionales han sido extensamente estudiados a partir de varios

aspectos de los organismos vivos. Estas investigaciones, no solo ayudan a los científicos a entender lo complicado de los fenómenos bioquímicos sino que revela el mecanismo fundamental de los organismos vivos desde una perspectiva de un sistema. Por lo tanto, la biología de sistemas con el énfasis en redes que son anticipadas para intensificar el entendimiento de los sistemas celulares por la integración comprensiva de los datos de componentes moleculares en diferentes niveles y estudiando como las múltiples interacciones facilitan lo complicado de las funciones biológicas dentro de la célula (Chen et al, 2009).

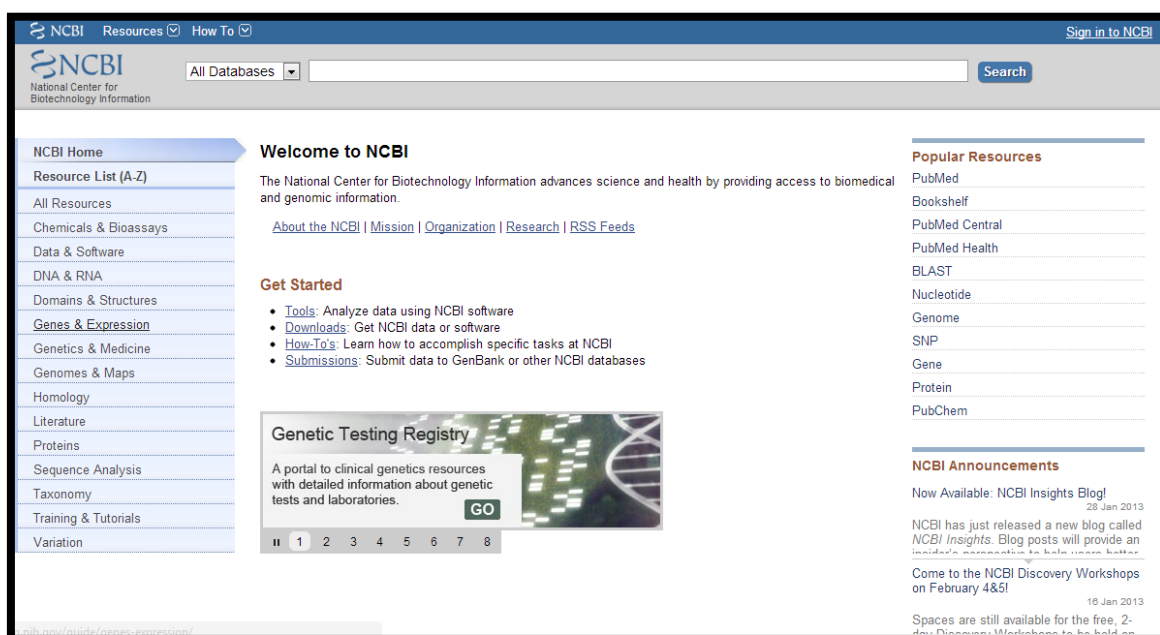
La integración de múltiples implicaciones, incluyen las de fuentes de diferentes datos, niveles del sistema, tecnologías, áreas disciplinarias, e incluso de investigaciones. Para hacer el uso de los datos de alta tecnología, claramente se necesita integrar no solo las fuentes de datos heterogéneos sino de otras metodologías y niveles de los sistemas.

#### **1.4.1 BASES DE DATOS BIOLÓGICAS**

Las bases de datos biológicas más importantes en las que podemos acceder para investigar, analizar y resolver los problemas biológicos acerca de una especie vegetal son The National Center for Biotechnology Information (NCBI), Protein Information Resource (PIR), The Universal Protein Resource (UniProt) y Protein Data Bank (PDB).

En el Instituto Nacional de la Salud fue creado en 1988 la base de datos del NCBI (fig. 1.7), cuyo objetivo primordial es desarrollar sistemas de información para la biología molecular, creando bases de datos públicos, investigación en biología computacional, desarrollo de herramientas de software para el análisis de datos del genoma y difusión de la información biomédica, disponibles a través de la página web NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Los servicios del NCBI incluyen herramientas como son el Entrez, MyNCBI, PubMed, PubMed Central (PMC), Gene, el Browser Taxonomía NCBI, BLAST, BLAST Link (Blink), Primer-BLAST, COBALT, Splign, RefSeq, UniGene, HomoloGene, ProtEST, dbMHC, dbSNP, dbVar, Epigenomics, Genome y herramientas relacionadas, el Map Viewer, Model Maker, Evidence Viewer, Trace Archive, Sequence

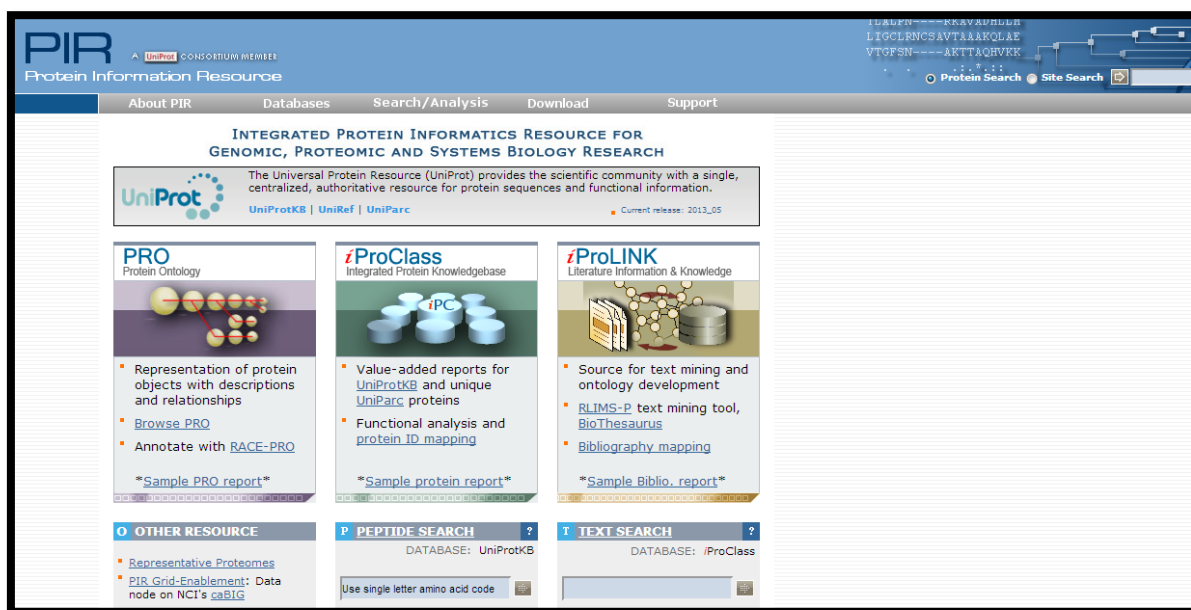
Read Archive, BioProject, BioSample, Retroviral Genotyping Tools, HIV-1/ Human Protein Interaction Database, Gene Expression Omnibus (GEO), Probe, Online Mendelian Inheritance in Animals (OMIA), the Molecular Modeling Database (MMDB), the Conserved Domain Database (CDD), the Conserved Domain Architecture Retrieval Tool (CDART), Biosystems, Protein Clusters PubChem una pequeña bases de datos de moléculas. Esta base de datos contiene unas 2800 familias de proteínas conservadas. (Sayers et al, 2012).



**FIGURA 1.7. BASE DE DATOS NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (NCBI).**

El Protein Information Resource (PIR), que se puede observar en la fig. 1.8, es un sistema integrado de los recursos públicos de la informática de proteínas que apoya el descubrimiento de la investigación científica de la genómica y proteómica. Mantiene la Protein Sequence DataBase (PSD), una base de datos que contiene proteínas anotando más de 283,000 secuencias que cubren la totalidad del rango taxonómico. La clasificación de familia se utiliza para una identificación sensible, anotación coherente, y detección de errores de anotación. El procesamiento de datos de la superfamilia define la firma de la arquitectura del dominio y categoriza para mejorar la clasificación automatizada. El PIR

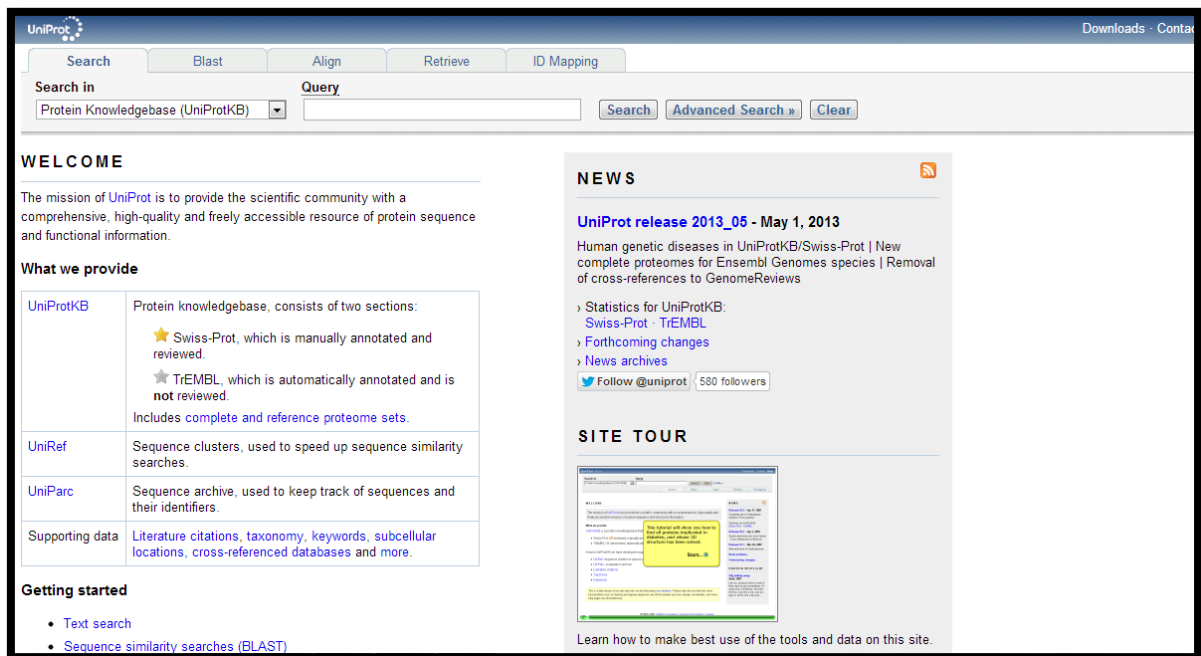
para aumentar la cantidad de anotación experimental, ha desarrollado un sistema de referencias bibliograficas para la búsqueda bibliográfica, la cartografía y la presentación del usuario, y se ha llevado a cabo la atribución retrospectiva de citas para funciones experimentales. También mantiene a NREF, una base de datos de referencia no redundante, y iProClass, una base de datos integrada de familias de proteínas, función y estructura de la información. PIR-NREF ofrece una colección amplia y oportuna de secuencias de proteína, en la actualidad consta de más de 1'000,000 de entradas de PIR-PSD, SWISS-PROT, TrEMBL, RefSeq, GenPept y PDB. El sitio web PIR (<http://pir.georgetown.edu>) conecta las herramientas de análisis de datos a bases de datos subyacentes de información recuperación y descubrimiento de conocimientos, con todas las funciones para consultas interactivas, combinaciones de secuencias y búsquedas de texto, y la clasificaciónasi como exploración visual de los resultados de búsqueda. El sitio FTP proporciona descarga gratuita de PSD y NREF comunicados quincenales y las bases de datos y los archivos auxiliares (Wu et al, 2003).



**FIGURA 1.8. BASE DE DATOS PROTEIN INFORMATION RESOURCE (PIR).**

El Universal Protein Resource (UniProt) es generada por el Consorcio de UniProt que consta de grupos como European Bioinformatics Institute (EBI), Swiss Institute of

Bioinformatics (SIB) y Protein Information Resource (PIR) y se actualiza y distribuye cada 3 semanas y se puede acceder en línea para la búsqueda o descarga en <http://www.uniprot.org> (fig. 1.9). La misión principal de UniProt es apoyar la investigación biológica, manteniendo estable y completa, totalmente clasificada, abundante y precisa la anotación de la información de las secuencias de proteínas, con referencias extensas cruzadas e interfaces de consulta de libre acceso para la comunidad científica. Se compone de cuatro grandes componentes, cada uno optimizado para diferentes usos: la Base de Conocimientos UniProt (UniProtKB) que es un punto de acceso central para el manejo integrado de la información de proteína con referencias cruzadas a múltiples fuentes. El Archivo UniProt (UniParc) es una integral repositorio de secuencia, lo que refleja la historia de todas las secuencias de proteínas. UniProt Clusters de referencia (UniRef) combina secuencias estrechamente relacionadas sobre la base de identidad de secuencia para acelerar las búsquedas, mientras que la bases de datos UniProt Secuencias Metagenomica y Ambiental (Unimes) fue creado para responder a la zona en expansión de los datos de la metagenómica. UniProt es libre y fácil acceso por los investigadores para llevar a cabo análisis interactivos de las proteínas de interés para facilitar la generación de hipótesis y descubrimiento de conocimiento (The UniProt Consortium, 2010).



## 1.9 BASE DE DATOS UNIVERSAL PROTEIN RESOURCE (UniProt).



El Protein Data Bank (PDB) es un repositorio para los datos estructurales 3-D de las moléculas biológicas grandes, tales como proteínas y ácidos nucleicos (fig.1.10). Los datos, generalmente obtenidos por cristalografía de rayos X o espectroscopia de RMN y presentado por biólogos y bioquímicos de todo el mundo, son de libre acceso en internet a través de los sitios web de sus organizaciones miembro PBDE, PDBj, y RCSB. PDB es un recurso clave en las áreas de la biología estructural, como la genómica estructural. La mayoría de las principales revistas científicas, y algunas agencias de financiación, como el NIH en los EE.UU., ahora requieren que los científicos presenten sus datos en el formato de la estructura del PDB la cual es supervisada por el PDB Worldwide y se puede acceder en la siguiente dirección <http://www wwpdb.org/> (Berman et al, 2000).

## 1.10. BASE DE DATOS PROTEIN DATA BANK (PDB).

## 1.4.2 LA BIOINFORMATICA EN PLANTAS

Si inicialmente la bioinformática se ocupó de moléculas y organismos unicelulares, a la vez que se descifraban los genomas de muchas bacterias, en la actualidad hay mucho interés en modelar sistemas biológicos y analizar especies multicelulares. Las últimas en llegar a esta “era genómica” han sido las plantas, y la resolución de los genomas de la

herbácea *Arabidopsis thaliana* y del arroz han servido para que la bioinformática empiece a ocuparse también de estos seres vivos (Contreras, 2009). La disposición de secuencias de estos genomas ha permitido dar un análisis genómico de cómo ocurre la regulación transcripcional en plantas, de proveer los recursos para entender procesos biológicos vegetales y favorecer el aislamiento de genes de importancia agronómica generando un impacto positivo en la producción de cultivos comerciales (genes controladores de resistencia a enfermedades, tolerancia a estrés abiótico, síntesis de vitaminas/alimentos nutricionalmente enriquecidos) (Xiong, 2006).

La enorme diversidad bioquímica que existe en las plantas hace que los distintos metabolitos que pueden encontrarse en plantas se estimen en más de 200,000. Esto hace que una aproximación a gran escala del estudio de los metabolitos presentes en las plantas sea un inmenso reto. La deducción de cómo es el contexto biológico de una planta dada a partir de datos obtenidos por la medida de las concentraciones de metabolitos requiere, por tanto, que la identificación y cuantificación de los mismos se realice con gran fiabilidad. El puente necesario que debe existir entre la metabolómica y las otras aproximaciones de la genética funcional: transcriptómica y proteómica, permitirá el desarrollo de bases de datos que almacenen, integren, relacionen y permitan establecer relaciones causales entre genes, transcritos, proteínas y metabolitos. Los metabolitos de las plantas son importantes para muchas de las respuestas de resistencia al ataque de patógenos, en condiciones de estrés y contribuyen al color, sabor y olor de flores y frutos. El fenotipo bioquímico de una planta es el resultado de la interacción entre el genotipo y el entorno. Por tanto, se necesita identificar y cuantificar los metabolitos para estudiar la dinámica del metaboloma y analizar los flujos metabólicos. El reto de la metabolómica reside en encontrar cambios en las cantidades de metabolitos que se puedan correlacionar con el estado fisiológico y de desarrollo de una célula, de un tejido o de un organismo (Claverie y Notredame, 2007).

El color azul en las flores aun es un enigma, ya se han reportado los genes involucrados en la síntesis de delphinidina, por lo que en este trabajo se pretende esclarecer la función en las diferentes especies reportadas, así como las características de la proteína F3'5'H, siendo la bioinformática la herramienta ideal para la comparación y búsqueda deseada.

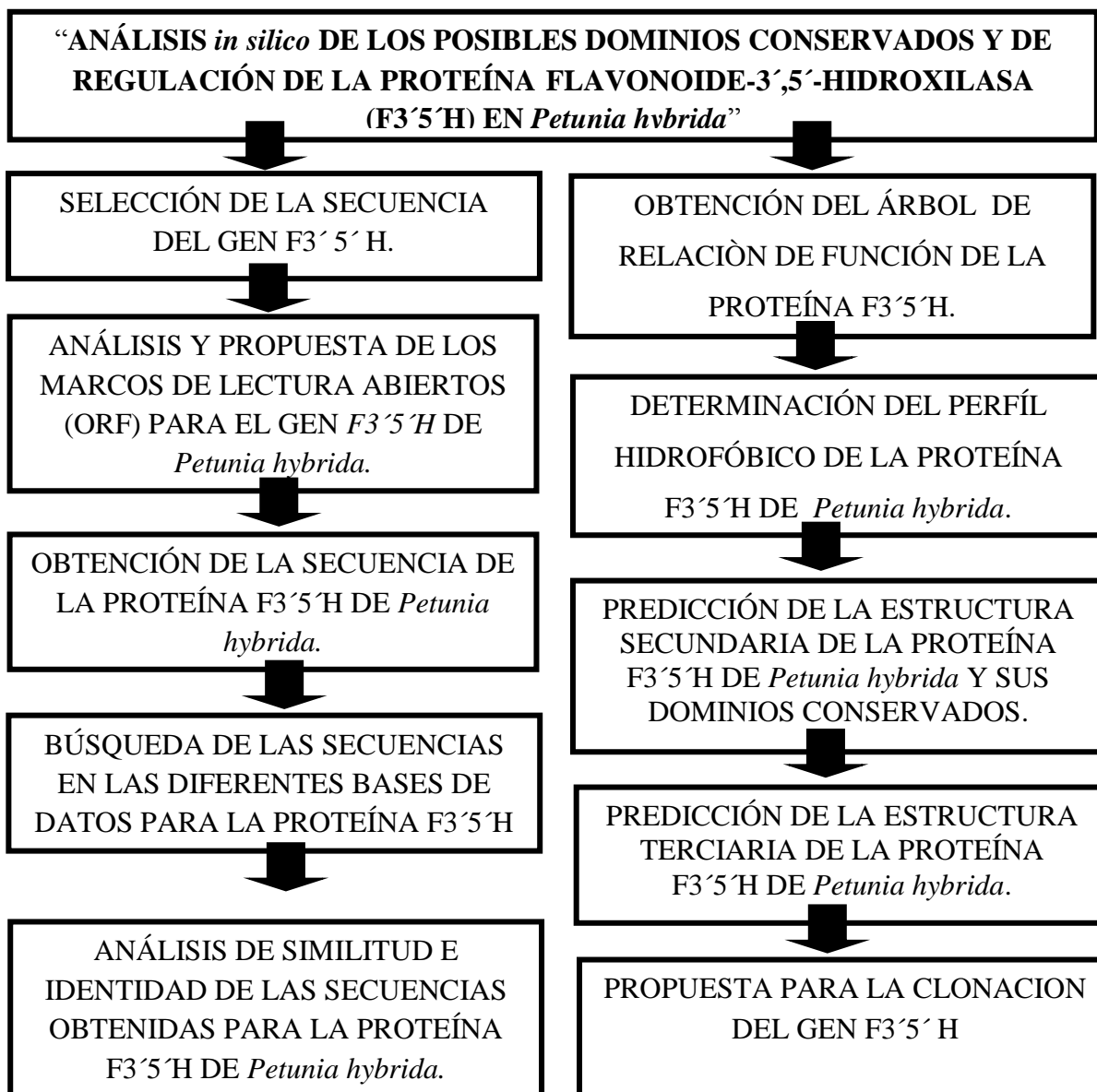
## 1.10 OBJETIVO GENERAL

Predecir la función, regulación y características estructurales de la proteína Flavonoide 3´5´Hidroxilasa a partir de un análisis *in silico*.

## 1.6 OBJETIVOS ESPECÍFICOS

- Realizar una búsqueda en bases de datos de las secuencias reportadas para el gen *F3´5´H*.
- Predecir la secuencia de la proteína F3´5´H.
- Determinar los dominios conservados por medio de alineamientos de las secuencias de ortólogos para la proteína F3´5´H de *Petunia hybrida* en relación a su función en diferentes tejidos, determinando por igual la conformación y el plegamiento de la proteína F3´5´H de *Petunia hybrida*.
- Predecir la regulación de la proteína F3´5´H con base a su posición en el clado del árbol filogenético..
- Proponer las condiciones para la clonación del gen *F3´5´H*.

## 1.8 ESTRATEGIA EXPERIMENTAL INTEGRAL



## 1.8 REFERENCIAS BIBLIOGRÁFICAS

Aiph/Union Fleurs: *International Statistics Flowers And Plants* (2011) National Agricultural Statistics Service.  
[http://www.aiph.org/site/index\\_en.cfm?act=teksten.tonen&parent=4685&varpag=3954](http://www.aiph.org/site/index_en.cfm?act=teksten.tonen&parent=4685&varpag=3954)

Baenziger P.S. y Al-Otayk S.M. (2007) *Plant Breeding in the 21<sup>st</sup> Century*. African Crop Science Conference Proceedings Vol. 8: 1-3.

Ben-Meir, H., A. Zucker, D. Weiss, y A. Vainstein (2002) *Molecular control of floral pigmentation: Anthocyanins..* Vaintein. Breeding for ornamentals: Classical and molecular approaches. Kluwer Academic Publishers, Dordrecht, the Netherlands. pp. 253-272

Berman H., Westbrook J., Feng Z., Gilliland C., Bhat T., Weissig H., Shindyalov I. y Bourne P. (2000) *The Protein Data Bank*. Nucleic Acids Research. Vol. 28:1 pp.235-242.

Chen L., Wang R., y Zhang X. (2009) *Biomolecular Networks*. Cap. 1 pp.1-22.

Claverie J. y Notredame C (2007) *Bioinformatics for dummies*. Wiley Publishing. Inc. 2da. Edición.

Cuevas E., Antezana A. y Winterhalter P (2008) *Análisis Y caracterización de antocianinas en diferentes variedades de maíz (Zea mays) boliviano*. Memorias RED-ALFA LAGROTECH, Comunidad Europea. pp. 79-95.

Contreras B. (2009) *La Bionformática Aplicada a las Plantas*. INNOVA-ARAGÓN No. 503 pp.1.

The UniProt Consortium (2010) *The Universal Protein Resource (UniProt) in 2010*. Nucleic Acids Research: Database issue. Vol. 38.pp. D142–D148.

Davies K. and Schwinn K. (1997) *Flower Colour*. Chapter 15.

Dooner, H.K., Robbins, T.P., and Jorgensen, R.A. (1991). Genetic and developmental control of anthocyanin biosynthesis. *Annu. Rev.Genet.* 25, 173-199.

Drago M (2007) *Flavonoides recombinantes de relevancia farmacéutica*. *Revista de Ciencias Farmacéuticas*. 38:4 pp.42-47.

Escandón, Alejandro (2005) *Biotecnología en el Cultivo de Especies Ornamentales*. pp 259-262.

Forkmann G. (1991) *Flavonoids as Flower Pigments: The formation of the natural spectrum and its extensión by genetic engineering*. *Review. Plant Breeding* Vol. 106 pp. 1-26.

Forkmann G. y Martens S. (2001) *Metabolic engineering and applications of flavonoids*. *Current Opinion in Biotechnology* 12:155-160. Elsevier Science.

Fukui Y., Tanaka Y., Kusumi T., Iwashita T., Nomoto K. (2003) *A rationale for the shift in colour towards blue in transgenic carnation flowers expressing the flavonoid 3',5'-hydroxylase gene*. *Phytochemistry* 63: 15–23.

Koes R., Verweij W. y Quattrocchio F. (2005) *Flavonoids: a colorful model for the regulation and evolution of biochemical pathways*. *TRENDS in Plant Science*. Vol. 10 No. 5: 236-242

Garzón, G. (2008) *Las antocianinas como colorantes naturales y compuestos bioactivos*. *Review. Acta biol. Colomb.*, Vol. 13 No. 3, 27 – 36.

González Mendoza, Daniel (2007) *El complejo enzimático citocromo P450 en las plantas*. Revista de Contaminación ambiental. Año 23 vol. 4. pp.177-183.

Graham S. y Peterson J. (1999) *How similar are P450s and what can their differences teach us*. Arch Biophys Biochem No. 369 pp. 24-29.

Griesbach, R.J. (1997) *The biochemical basis for the blue flower color mutations in Doritis pulcherrima and Phalaenopsis violacea*. Lindleyana 12(2):64-71.

Gotoh O. (1992) *Substrate recognition sites in cytochrome P450 family 2 (CYP2) protein inferred from comparative analyses of amino acid and coding nucleotide sequences*. J Biol Chem No. 267 pp. 83-90.

Harborne JB (1967) *Comparative biochemistry of the flavonoids-IV. Correlations between chemistry, pollen morphology and systematics in the family Plumbaginaceae*. Phytochemistry 6: 1415–1428

Harbone, J.B. and Williams, C.A. (2000) *Advances in research in flavonoid research since 1992*. Phytochemistry, 55, 481-504

Harding J., Singh F. and Mol (1991) *Genetics and Breeding of ornamentals species: Engenering of novel flower colours*. Klower Academic Publisers.

Hasemann C., Kurumbail R., Boddupalli S., Peterson J. y Deisenhofer J. (1995) *Structure and function of cytochromes P450: a comparative analysis of the crystal structures*. Structure Vol. 3 pp. 41-62.

Holton T. y Cornish E. (1995) *Genetics and Biochemistry of Anthocianin Biosynthesis*. The Plant Cell, pp. 1071-1083.

Kishimoto S. y Ohmiya A. (2009) *Studies on Carotenoids in the Petals of Compositae Plants. Review* . J. Japan. Soc. Hort. Sci. 78 (3): 263–272. 2009.

Kitada C., Gong Z., Tanaka Y, Yamazaki M, Saito K (2001) *Differential expression of two cytochrome P450s involved in the biosynthesis of flavones and anthocyanins in chemovarietal forms of Perilla frutescens*. Plant Cell Physiol. 42: 1338–1344.

López F., Del Valle Mondragón L. y Pastelín G (2006) *Los flavonoides y el sistema cardiovascular: ¿Pueden ser alternativa terapéutica?*. Vol. 76 Supl. 4, 33-45

Leguizamón G., González A. y Baéz R.(2005). *Antocianinas en uva (Vitis vinífera L.) y su relación con el color*. Revista Fitotecnia Mexicana 8: 359-368.

Martínez-Flórez S, González-Gallego J, Culebras JM, Tuñón MJ. (2002) *Los flavonoides: propiedades y acciones antioxidantes*. Nutr Hosp 17, 271-278.

Mayne, S. (1996) *Beta-carotene, carotenoids, and disease prevention in Humans*.FASEB Journal.Vol.10 pp. 690-701.

Mol, J., E. Cornish, J. Mason, y R. Koes (1999) *Novel coloured flowers*.Curr.Biotech. 10:198-201.

Mol, J.N.M., A.R. Stuitje y A. van der Krol (1989) *Genetic manipulation of floral*

Mora C. (2011) *Modo de acción de la chalcona 1-(4-metil-fenil)-3-(2,3-dimetoxifenil)-2-propenona como inhibidor de  $\beta$ -lactamasa: estudio cinético y caracterización inicial mediante espectrometría de masas*. Magíster en Ciencias Farmacéuticas. Grupo de Investigación en Sustancias Bioactivas (GISB).

Okinaka Y., Shimada Y., Nakano-Shimada R., Ohbayashi M., Kiyokawa S. y Kikuchi Y. (2003) *Selective Accumulation of Delphinidin Derivatives in Tobacco Using a Putative*



*Flavonoid 3',5'-Hydroxylase cDNA from Campanula médium*. Biosci. Biotechnol. Biochem., 67:1 pp. 161-165.

Ortiz Medina E. y Larque Saavedra A. (1999) *El uso de reguladores de crecimiento en la floricultura mexicana*. Ciencia y Desarrollo. 25:148, 26-39.

Puebla P., Guerrero M.F. y Correa S.X. (2004) *Flavonoides del género Crotón*. Rev. Col. Cienc. Quím. Farm. Vol. 33 (1), 77-85.

Russo R. y Sánchez M (2006) *Los flavonoides en la terapia cardiovascular*. Rev Costarr Cardiol, Vol 8, Revista #1. pp. 1-6.

SAGARPA (2006) *La floricultura mexicana, el gigante que está despertando: Biotecnología en la industria de flores de corte*. Claridades Agropecuarias. ASERCA. pp. 3-42. No. 154

SAGARPA (2008) Boletín ASERCA Regional Peninsular: *La Floricultura*. pp.3-8

Sayers E., Barrett T., Benson D., Bolton E., Bryant E., Canese K, Chetvernin V., Church D., DiCuccio M., Federhen S., Feolo M., Fingerman I, Geer L., Helmberg W., Kapustin Y., Krasnov S., Landsman D., Lipman D., Lu Z., Madden T., Madej T., Maglott D., Marchler-Bauer A., Miller V., Karsch-Mizrachi I., Ostell J., Panchenko A., Phan L., Pruitt P., Schuler G., Sequeira E., Sherry S., Shumway M., Sirotkin K., Slotta D., Souvorov A., Starchenko G., Tatusova T., Wagner L., Wang Y., Wilbur J., Yaschenko E. y Ye J. (2011) *Database resources of the National Center for Biotechnology Information*. Nucleic Acids Research: Database issue. Vol. 40. pp. D13–D25.

Tanaka Y., Katsumoto Y., Brugliera y Mason (2005) *Genetic engineering in floriculture*. Review of Plant Biotechnology and Applied Genetics Plant Cell, Tissue and Organ Culture. 80: 1–24. Springer.

Tanaka Y., Sasaki N. and Ohmiya A. (2008) *Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids*. The Plant Journal 54, pp. 733-749.

Tanaka Y. (2006) *Flower colour and cytochromes P450*. Phytochem Rev 5:283–291.

Tanaka Y., Brugliera F. y Chandler S. (2009) *Recent Progress of Flower Colour Modification by Biotechnology*. Review. Vol.10, pp. 5350-5369. International Journal of Molecular Sciences.

The UniProt Consortium (2010) *The Universal Protein Resource (UniProt) in 2010*. Nucleic Acids Research: Database issue. Vol. 38. pp. D142–D148.

Van der Meer I., Stam M., Van Tunen A., MOI J. y Stuitje A. (1992). *Antisense Inhibition of Flavonoid Biosynthesis in Petunia Anthers Results in Male Sterility*. The Plant Cell. American Society of Plant Physiologists. Vol. 4, pp. 253-262.

Vetten N., Horst ., Van schaik H., De Boer ., Mol . y Koes R. (1999). *A cytochrome b5 is required for full activity of flavonoid 3',5'-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors*. Plant Biology. Vol. 96, pp. 778–783.

Villar- Martínez A., Serrato-Cruz M., Solano-Navarro A., Arenas-Ocampo M., Quintero Gutiérrez A., Sánchez-Millán J., Evangelista-Lozano S., Jiménez-Aparicio A., García-Jiménez A. y Vanegas-Espinoza P. (2007) *Careotenoides en Tagetes erecta L. la modificación genética como alternativa*. Revista de Fitotecnia Mexicana 30:2, 109-118.

Yosida K., Mori M. y Kondo T. (2009) *Blue flower color development by anthocyanins: from chemical structure to cell physiology*. Review. Nat. Prod. Rep. Vol. 26 pp. 884-915.

Wang J., Ming F., Han Y. y Shen D. (2006) *Flavonoid-3',5'-hydroxylase from Phalaenopsis: a novel member of cytochrome P450s, its cDNA cloning, endogenous expression and molecular modeling*. Biotechnology Vol.28 pp. 327–334.

Went F (1978) *Las plantas: Maravillas de Color*. Enciclopedia Colección de la Naturaleza de TIME-LIFE. pp. 63-7.

Werck-Reichhart y Feyereisen (2000) *Cytochromes P450: a success story*. Genome Biology Vol 1 No 6.pp.1-9

Williams P., Cosme J., Sridhar V., Johnson E y McRee D. (2000) *Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity*. Mol Cell Vol. 5 pp. 121-131.

Winkel-Shirley, B. (2001) *Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology*. Plant Physiology, Vol. 126, pp. 485–493.

Wu C., Yeh L., Huang H., Arminski L., Castro-Alvear J., Chen J., Hu Z., KourtesisP., Ledley R., Suzek B., Vinayaka C., Zhang J. y Barker W (2003) *The Protein Information Resource*.Nucleic Acids Research. Vol. 31, No. 1 345–347.

Xiong, J., (2006) *Essential Bioinformatics*. Cambridge University Press.

**CAPITULO II: ANÁLISIS DE LA SECUENCIA DE LA PROTEÍNA  
FLAVONOIDE - 3', 5' - HIDROXILASA E IDENTIFICACIÓN DE LOS  
DOMINIOS CONSERVADOS.**

## **2.1 INTRODUCCIÓN**

La biotecnología moderna es una actividad científica multidisciplinaria e integradora que presenta una dependencia especial de los progresos tecnológicos relacionados con la velocidad de procesamiento de la información. La bioinformática es la herramienta por excelencia, que se encuentra en la intersección entre las ciencias de la vida y de la información. Proporciona herramientas y recursos para favorecer la investigación biológica. Trata de desarrollar sistemas que sirvan para entender el flujo de información desde los genes a las estructuras moleculares, su función bioquímica y conducta biológica. Si bien algunos restringen el rango de estudio de la bioinformática al manejo y análisis de bases de datos biológicas principalmente de secuencias, podría atribuírsele un sentido más amplio, como la fusión de las técnicas computacionales con el entendimiento y apreciación de datos biológicos, el almacenamiento, recuperación, manipulación y correlación de datos procedentes de distintas fuentes (Cañedo y Arencibia, 2004). Debido a esto, uno de los campos más comunes y lucrativos de la programación de computadoras es la gestión de base de datos, la cual consiste en el diseño y la programación de formas de almacenar y recuperar datos (EBI, 2006).

GenBank, PDB o cualquier otra base de datos biológicos, sirve para que los datos se puedan buscar fácilmente y las entradas recuperadas estén accesibles en un formato útil y significativo. De lo contrario, los esfuerzos de la secuenciación, tales como el genoma de *Arabidopsis thaliana*, no tienen ningún fin útil, ya que la comunidad biológica en su conjunto no puede hacer uso de la información oculta dentro de estos millones de bases y aminoácidos. Mucha energía se ha ido a hacer estos datos accesibles para el investigador, y los programas e interfaces resultantes de estos esfuerzos son el foco principal de la biología actual (Wiley y Sons, 2003).

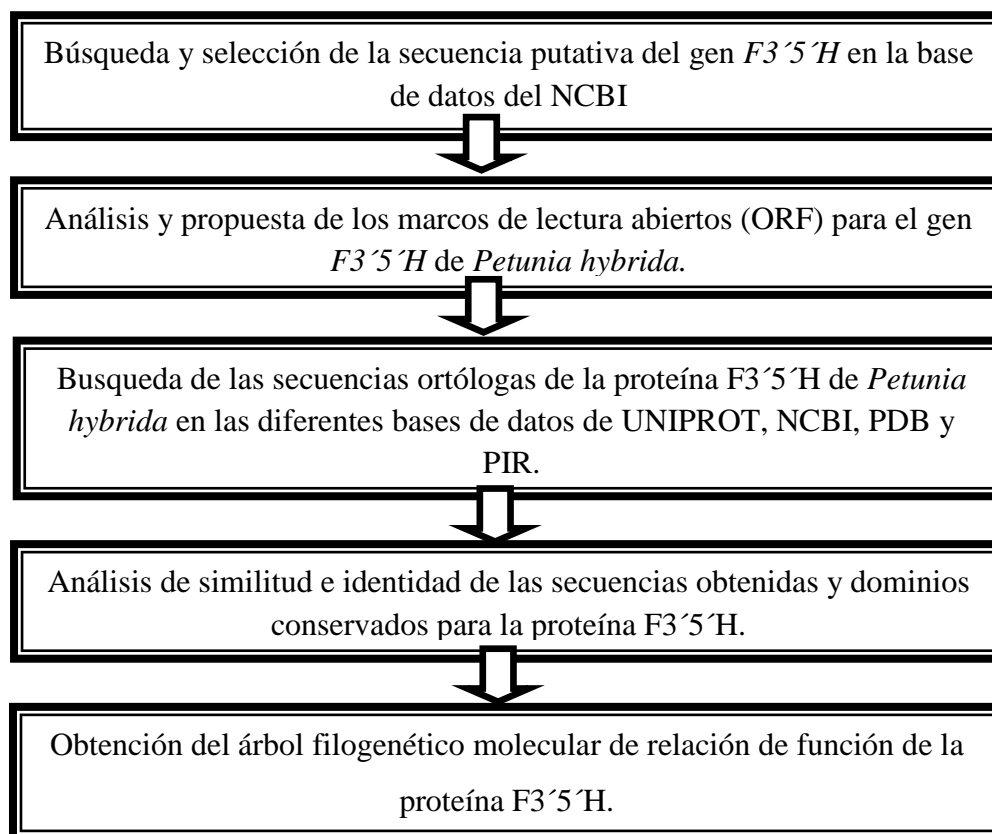
La libre disponibilidad de datos del genoma no sólo humanos, sino la variación de datos de la secuencia, el modelo de datos de la secuencia del organismo y la información sobre la estructura y función del gen o proteína, proporciona un terreno fértil para el biólogo a un mejor diseño e interpretación de sus experimentos en el laboratorio, cumpliendo la promesa de la bioinformática para avanzar y acelerar el descubrimiento biológico. La recuperación de información de bases de datos biológicas proporciona el primer paso para ser capaz de realizar los análisis bioinformáticos robustos y precisos (Wiley y Sons, 2003).

La comprensión de la evolución de las proteínas incluso a la secuencia primaria permanece aun en reto. Los paquetes de software actuales están mejorando rápidamente, y la alineación de la secuencia correcta puede ser incluso un desafío para los pares de proteínas. A bajos niveles de similitud, puede ser difícil de detectar homología. Para superar este problema, las alineaciones se realizan a veces mediante la búsqueda de similitud de secuencias, la cual es una tarea muy importante en la bioinformática. Instrumentos como estos, que utilizan métodos de comparación de secuencias para buscar homólogos con el fin de obtener pistas sobre la evolución funcional y las relaciones de las proteínas, lo tienen presente como una parte integral de la investigación del laboratorio de hoy (Lee et al, 2007).

En general las secuencias de nucleótidos pueden sufrir mutaciones, a no ser que exista presión de selección para que esto no ocurra, habiendo un posible cambio en la función de la proteína. Realizar una recopilación de esta información, observando si es conservada, incluye un análisis a fondo de todas las secuencias de las plantas reportadas hasta la fecha con esas características y de si la determinación de los dominios conservados y de las características son variables o no. El Entrez, PSIBLAST (una versión ampliada de BLAST), árbol de función y alineamiento son algunas de las herramientas ideales para el objetivo del capítulo, que es el de analizar las secuencias reportadas para la proteína F3'5'H, relación de función y su comparación con la de *Petunia hybrida*, así como la identificación de los dominios conservados.

## 2.2 ESTRATEGIA EXPERIMENTAL

El esquema general para el análisis *in silico* de la proteína F3'5'H de *Petunia hybrida* fue el siguiente:



## 2.3. MATERIALES Y MÉTODOS

Las actividades se realizaron utilizando una computadora Toshiba con una capacidad de 3 MB y el sistema operativo Windows 7.

### 2.3.1 Búsqueda de la secuencia del gen *F3'5'H*.

Para la búsqueda de la secuencia del gen *F3'5'H* se utilizó la técnica del Entrez en la base de datos del NCBI (<http://www.ncbi.nlm.nih.gov>), que consiste en la búsqueda de la secuencia problema donde el sistema de recuperación muestra los recuentos de registros coincidentes en cada componente de la base de datos en una línea de estadísticas y

vinculado al resumen del documento, que integra la base de datos de PubMed, con otras 39 publicaciones y bases de datos moleculares como ARNm y ADN (Romiti y Cooper, 2011).

### **2.3.2 Búsqueda de los marcos de lectura abiertos (ORF) para el gen *F3'5'H*.**

Una vez elegida la secuencia de nucleótidos de la *F3'5'H*, se utilizó el paquete computacional traductor, para la obtención de los posibles marcos de lectura abiertos (ORF=Open Reading Frame, por sus siglas en inglés). Que consistió en leer el mensaje del ARN mensajero obtenido de la transcripción por medio del código genético donde los nucleótidos codifican para un aminoácido tomando los codones respectivos, construyendolo en proteína. Los ORF representan los diferentes sitios de inicio de lectura en un fragmento de ADN dado. Como resultado existen seis secuencias de aminoácidos. Dado que cada codón toma tres nucleótidos, existen tres posibles lugares de inicio para tomar los nucleótidos de tres en tres, a lo que hay que sumar los otros tres posibles marcos de lectura abiertos tomando como molde la hebra complementaria, dando el sentido de lectura opuesto (Tatusov y Tatusov, 2011). La obtención de la proteína es importante debido a que es la que controla la estructura y función de la célula, utilizando el programa el análisis con el programa de la página <http://www.fr33.net/translator.php>

### **2.3.3 Búsqueda de las secuencias ortólogas de la proteína *F3'5'H* de *Petunia hybrida* en las diferentes bases de datos.**

Con la proteína obtenida se prosiguió a la búsqueda en las bases de datos de UNIPROT (<http://www.uniprot.org>), NCBI (<http://www.ncbi.nlm.nih.gov/protein>), PIR (<http://pir.georgetown.edu/>) y PDB (<http://www.wwpdb.org/>) de las secuencias ortólogas a la proteína *F3'5'H* de *Petunia hybrida*, realizándose posteriormente un PSIBLAST, el cual se utiliza para comparar dos secuencias nucleotídicas o proteicas y encontrar regiones de similitud local entre ellas, permitiendo localizar secuencias en una base de datos mediante el alineamiento de las mismas buscando regiones de similitud (Bhagwat y Aravind, 2007).

### **2.3.4 Alineamiento múltiple de las secuencias obtenidas y búsqueda de los dominios conservados para la proteína F3'5'H de *Petunia hybrida*.**

Para el análisis de identidad y similitud de las proteínas, se utilizó el ClustalW en el EMBL-EBI (European Molecular Biology Laboratory-, <http://www.ebi.ac.uk>) programa que toma secuencias homólogas (ADN, ARN o proteína) y realiza un alineamiento, encontrando regiones conservadas y no conservadas entre las secuencias. El EMBL-EBI proporciona los porcentajes de identidad entre las secuencias y el tamaño de cada secuencia proporcionada (Thompson et al, 2003). Para los dominios conservados se utilizó el programa Scan Prosite en línea <http://prosite.expasy.org/> (EXPASY, 2012).

### **2.3.5. Elaboración del árbol de la relación de función de la proteína F3'5'H.**

La reconstrucción de las relaciones filogenéticas es de suma importancia para descifrar patrones evolutivos y funcionales. Para este propósito existen diferentes paquetes bioinformáticos, eligiéndose el PHYLogeny Inference Package 3.69 (PHYMLIP), el cual contiene cerca de 30 programas que cubren los aspectos más importantes del análisis filogenético, incluyendo programas para hacer análisis de parsomonia, métodos de distancia, máxima verosimilitud y otros métodos sobre una gran variedad de tipos de datos incluyendo secuencias de ADN, ARN, proteicas, sitios de restricción, frecuencias génicas y matrices de distancia entre muchos, seleccionando la metodología del Neighbor-Joining trees, utilizando el protocolo 6.3 de Current Protocols in Bioinformatics (Gascuel, 2003).

## **2.4.RESULTADOS Y DISCUSIÓN**

### **2.4.1 Obtención de la secuencia putativa del gen F3'5'H.**

Se realizó una búsqueda de la secuencia del gen F3'5'H en la base de datos del NCBI (<http://www.ncbi.nlm.nih.gov/nucleotide>) seleccionando a la *Petunia x hybrida* cultivar QL01 con número de accesión EF371021 con un tamaño de 1773 pb (fig. 2.1). Debido a que es uno de los primeros genes que se clonaron de F3'5'H. Se observó que existen varias secuencias para petunia de las cuales se reportan diferentes clones, o genes con alelos



diferentes, que pertenecen a la misma familia: los citocromos P450. De todos esos se eligió al gen con mayor antigüedad.

```

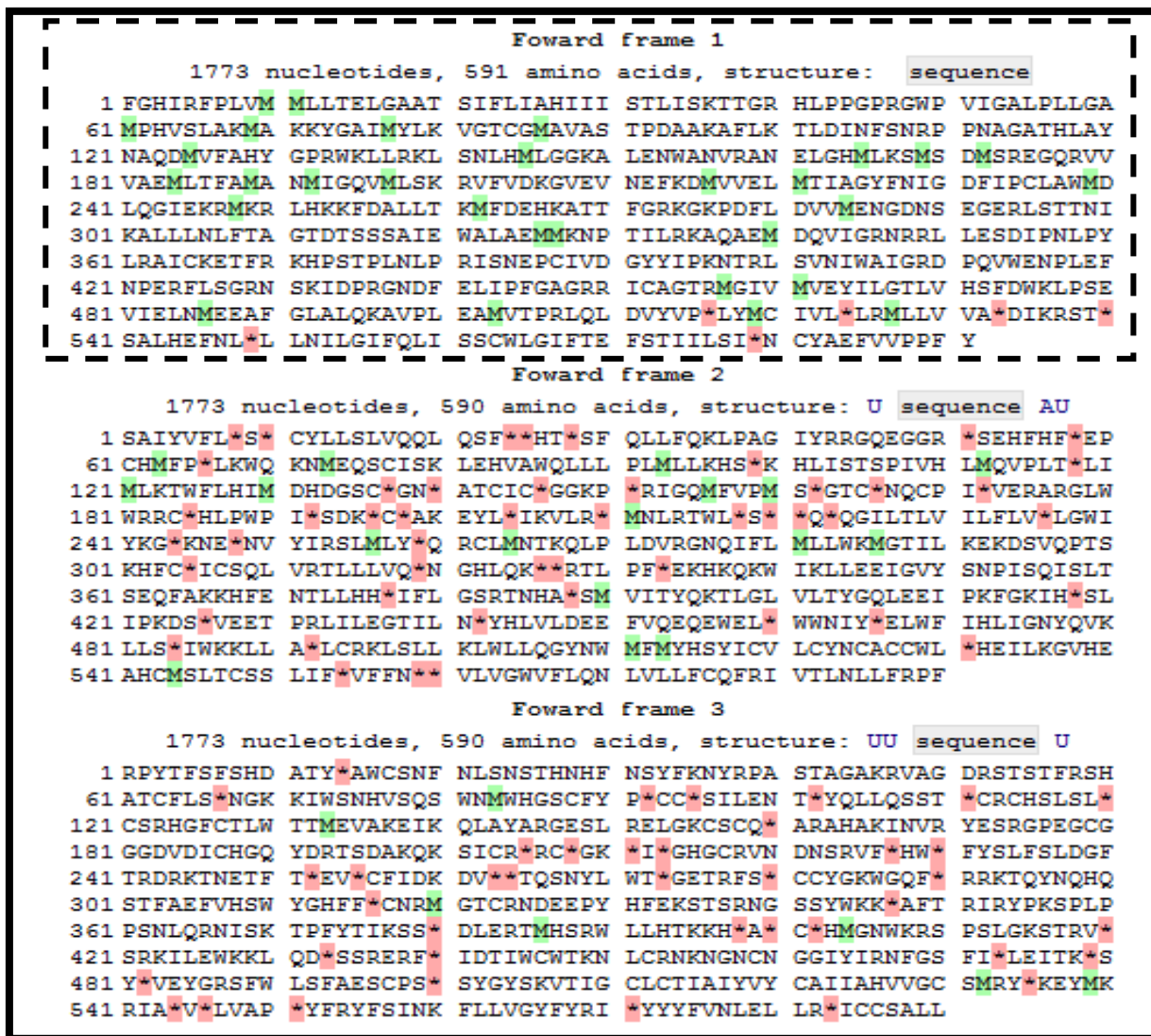
>EF103721_PH
TTCGGCCATATACGTTTTCTTTAGTATGATGCTACTTACTGAGCTTGGTGCAGCAACTTCAATCTTTC
TAATAGCACACATAATCATTTCAACTCTTATTTCAAAAAC TACCGCCGGCATCTACCGCCGGGGCCAAG
AGGGTGGCCGGTGATCGGAGCACTTCCACTTTTAGGAGCCATGCCACATGTTTCTTAGCTAAAATGGCA
AAAAAATATGGAGCAATCATGTATCTCAAAGTTGGAACATGTGGCATGGCAGTTGCTTCTACCCCTGATG
CTGCTAAAGCATTCTTGAAAACACTTGATATCAACTTCTCCAATCGTCCACCTAATGCAGGTGCCACTCA
CTTAGCTTATAATGCTCAAGACATGGTTTTTGCACATTATGGACCACGATGGAAGTTGCTAAGGAAATTA
AGCAACTTGCATATGCTAGGGGGGAAAAGCC TTAGAGAATTGGGCAAATGTTCTGTGCCAATGAGCTAGGGC
ACATGCTAAAATCAATGTCCGATATGAGTCTGAGAGGGCCAGAGGGTTGTGGTGGC GGAGATGTTGACATT
TGCCATGGCCAATATGATCGGACAAGTGATGCTAAGCAAAAGAGTATTTGTAGATAAAGGTGTTGAGGTA
AATGAATTTAAGGACATGGTTGTAGAGTTAATGACAATAGCAGGGTATTTTAAACATTGGTGATTTTATTC
CTTGTTTAGCTTGGATGGATTTACAAGGGATAGAAAAACGAATGAAACGTTTACATAAGAAGTTTGATGC
TTTATTGACAAAGATGTTTGATGAACACAAAGCAACTACCTTTGGACGTAAGGGGAAAACCAGATTTTCTT
GATGTTGTTATGGAAAATGGGGACAATTCGAAGGAGAAAGACTCAGTACAACCAACATCAAAGCATT
TGCTGAATTTGTTACAGCTGGTACGGACACTTCTTCTAGTGCAATAGAATGGGCACTTGCAGAAATGAT
GAAGAACCCTACCATTTTGAGAAAAGCACAAAGCAGAAATGGATCAAGTTATTGGAAGAAATAGGC GTTTA
CTCGAATCCGATATCCCAAATCTCCCTTACC TCCGAGCAATTTGCAAAGAAACATTTTGAAAACACCCTT
CTACACCATTAATCTTCC TAGGATCTCGAACGAACCATGCATAGTCGATGGTTATTACATACC AAAAAA
CACTAGGCTTAGTGTTAACATATGGGCAATTGGAAGAGATCCCCAAGTTTGGGAAAATCCACTAGAGTTT
AATCCC GAAAGATTC TTGAGTGGAAGAAAC TCCAAGATTGATCCTCGAGGGAACGATTTTGAATTGATAC
CATTTGGTGCTGGACGAAGAATTTGTGCAGGAACAAGAATGGGAATTGTAATGGTGAATATATATTAGG
AACTTTGGTTCATTCATTTGATTGGAAATTACCAAGTGAAGTTATTGAGTTGAATATGGAAGAAGCTTTT
GGCTTAGCTTTGCAGAAAGCTGTCCCTCTTGAAGCTATGGTTACTCCAAGGTTACAATTGGATGTTTATG
TACCATAGCTATATATGTGTATTGTGCTATAAATGCGCATGTTGTTGGTTGTAGCATGAGATATTTAAAAG
GAGTACATGAAGCGCATTGCATGAGTTTAACTTGTAGCTCCTTAATATTTTAGGTATTTTCAATTAATA
AGTTCCTTGTGGTTGGGTATTTTACAGAATTTAGTACTATTATTTGTCAATTTAGAATTGTTACGCTG
AATTTGTTGTTCCGCCCTT TAT
  
```

**FIGURA 2.1. SECUENCIA DEL GEN FLAVONOIDE 3´5´HIDROXILASA DE *Petunia x hybrida* CULTIVAR QL01.** El rectángulo verde indica el sitio de inicio y rojo de terminación de la traducción.

**2.4.2 Análisis y propuesta de los marcos de lectura abiertos (ORF) para el gen *F3´5´H* de *Petunia hybrida* y obtención de la secuencia putativa de la proteína *F3´5´H* de *Petunia hybrida*.**

Una vez seleccionada se ingresó a la página <http://www.fr33.net/translator.php> para proponer los posibles marcos de lectura abiertos del gen *F3´5´H* en donde se insertó la secuencia de nucleótidos dando como resultado la traducción con los tres posibles marcos de lectura abiertos, como se muestra en la figura 2.2, se eligió el primer marco de lectura abierto (forward frame 1) debido a que se muestra la proteína de una secuencia ininterrumpida de 506 residuos de aminoácidos (fig 2.3) considerando la primera metionina

como el inicio de la proteína, siendo una proteína con mayor número de residuos, a diferencia de los otros dos marcos, en donde los codones de terminación están muy cercanos al de inicio.



**FIGURA 2.2. MARCOS DE LECTURA PROPUESTOS PARA EL GEN *F3'5'H* (EF371021).** En cerrado en el cuadro punteado se muestra el marco de lectura 1 el cual es el seleccionado. Marcadas con verde podemos observar a las metioninas y los asteriscos en rojo para los codones de terminación. También se observa el número de nucleótidos por cada marco de lectura y el tamaño de la proteína de cada uno.

```

>F35H0179PH

MMLLTELGAATSIFLIAHIIISTLISKTTGRHLPPGPRGWPVIGALPLLGAMPHVSLAKMAKKGAIMYLVGTGMAVASTPDAKA
FLKTLDFNSNRPPNAGATHLAYNAQDMVFAHYGPRWKLLRKLNSLHMLGGKALENWANVRANELGHMLKMSDMSREGQRVV
VAEMLTAMANMIGQVMLSkrVFDKGVFVNEFKDMVVELMTIAGYFNIGDFIPCLAWMDLQIEKRMKRLHKKFDALLTKMFD
EHKATTFGRKGKPDFLDVVMENGDNSEGERLSTNIKALLNLFTAGTDTS SAIEWALAEMMKNPTILRKAQAEMDQVIGRNRRL
LESDIPNLPYLAICKETFRKHPSTPLNLRISNEPCIVDGYYPKNTRLVSNWAIGRDPQVWENPLEFNPERFLSGRNSKIDPRGNDFE
LIPFGAGRRICAGTRMGIVMVEYILGTLVHSFDWKLPSVIELNMEEAFGLALQKAVPLEAMVTPRLQLDVYVP

```

**FIGURA 2.3. SECUENCIA PUTATIVA DE AMINOÁCIDOS DE LA PROTEÍNA F3'5'H DE *Petunia hybrida*.** De color azul se muestra el motivo de residuos de prolina (PPGP), de morado el motivo de unión al oxígeno (AGTDTS), de verde el motivo EXXR, de fucsia el motivo representativo del P450 PERFL y de rojo el motivo de unión al hierro (FGAGRRICAG).

#### 2.4.3 Obtención de las secuencias ortólogas de la proteína F3'5'H de *Petunia hybrida* en las diferentes bases de datos.

Con el propósito de señalar los dominios conservados y de regulación de la familia F3'5'H, se prosiguió a la búsqueda de proteínas ortólogas en las bases de datos de UNIPROT (<http://www.uniprot.org>), NCBI (<http://www.ncbi.nlm.nih.gov/protein>), PIR (<http://pir.georgetown.edu/>) y PDB (<http://www.wwpdb.org/>) de las secuencias ortólogas a la proteína F3'5'H de *Petunia hybrida* (cuadro 2.1). En el cuadro 2.2 se muestran, las familias y la subclase a la que pertenece la especie, el tamaño de la proteína, el número de accesión y el porcentaje de identidad con respecto de cada una a *Petunia hybrida* (A3FEQ7).

BASE DE DATOS	ESPECIE	No. DE ACCESIÓN
NCBI ( REQSEQ)	<i>Arabidopsis thaliana</i>	CAG49299
	<i>Callistephus chinensis</i>	AAB45977
	<i>Glandularia x hybrida</i>	AAT34974

EMBL (PDB Y UNIPROT)	<i>Delphinium grandiflorum</i> <i>Eustoma grandiflorum</i> <i>Glycine max</i> <i>Gossypium hirsutum</i> <i>Osteospermum hybrid c.</i> <i>Pericallis cuneata</i> <i>Phalaenopsis hybrid c.</i> <i>Solanum lycopersicum</i> <i>Verbena hybrida</i>	Q52YL8 Q96418 D9N1C2 D0VBE7 Q304Q4 Q304Q5 Q3YAF0 Q6J210 D3W9H7
PIR	<i>Antirrhinum kelloggii</i> <i>Callistephus chinensis</i> <i>Camellia sinensis</i> <i>Campanula médium</i> <i>Catharanthus roseus</i> <i>Clitoria ternatea</i> <i>Cyclamen persicum</i> <i>Dendrobium hybrid</i> <i>Diospyros kaki</i> <i>Epimedium sagittatum</i> <i>Gentiana triflora</i> <i>Glycine max</i> <i>Lobelia erinus</i> <i>Lycianthes rantonnei</i> <i>Nierembergia sp</i> <i>Pericallis cuneata</i> <i>Pisum sativum</i> <i>Populus trichocarpa</i> <i>Rhododendron x pulchrum</i> <i>Ricinus communis</i> <i>Saintpaulia hybrid cultivar</i> <i>Solanum melongena</i>	E0D4W1 Q9FPN5 A3KLR7 O04773 Q9ZRY0 A4F1R0 D0EUZ8 A5XEH3 C6L1M3 D5M8Q4 Q96581 Q6YLS3 AF41T8 Q9FPN3 Q8LP20 Q304Q5 F2VPT1 B9HMH5 C7G325 B9SMA4 E3W760 P37120

	<i>Solanum tuberosum</i>	Q5EWY0
	<i>Torenia hybrid cultivar</i>	Q9FS35
	<i>Vanda coerulea</i>	B2LUN8
	<i>Vinca major</i>	Q76LL4
	<i>Viola x wittrockiana</i>	A9EEA6
	<i>Vitis amurensis</i>	COKY91

**CUADRO 2.1. ESPECIES REPORTADAS PARA LA PROTEÍNA F3'5'H EN LAS DIFERENTES BASES DE DATOS**

SUBCLASE	FAMILIA	ESPECIE	No. DE ACCESIÓN	TAMAÑO (aa)	% IDENTIDAD
Lílídea	Ranunculaceae	<i>Delphinium grandiflorum</i>	Q52YL8	502	70
Lílídea	Orchidaceae	<i>Phalaenopsis hybrid c.</i>	Q3YAF0	506	50
Lílídea	Orchidaceae	<i>Dendrobium hybrid</i>	A5XEH3	504	52
Lílídea	Orchidaceae	<i>Vanda coerulea</i>	B2LUN8	497	33
Magnolídea	Brassicaceae	<i>Arabidopsis taliana</i>	CAB45978	510	34
Magnolídea	Asteraceae	<i>Callistephus chinensis</i>	AAG49299	510	47
Magnolídea	Verbenaceae	<i>Glandularia x hybrida</i>	AAT34974	514	73
Magnolídea	Gentianaceae	<i>Eustoma grandiflorum</i>	Q96418	510	73
Magnolídea	Fabaceae	<i>Glycine max</i>	Q6YLS3	509	74
Magnolídea	Malvaceae	<i>Gossypium hirsutum</i>	D0VBE7	510	75
Magnolídea	Asteraceae	<i>Osteospermu</i>	Q304Q4	508	48

		<i>m hybrid c.</i>			
Magnolídea	Verbenaceae	<i>Verbena hybrida</i>	Q6J210	514	73
Magnolídea	Plantaginaceae	<i>Antirrhinum kelloggii</i>	E0D4W1	510	73
Magnolídea	Theaceae	<i>Camellia sinensis</i>	A3KLR7	510	75
Magnolídea	Campanulaceae	<i>Campanula medium</i>	O04773	523	67
Magnolídea	Apocynaceae	<i>Catharanthus roseus</i>	Q9ZRY0	512	76
Magnolídea	Fabaceae	<i>Clitoria ternatea</i>	A4F1R0	523	73
Magnolídea	Primulaceae	<i>Cyclamen persicum</i>	D0EUZ8	508	76
Magnolídea	Ebenaceae	<i>Diospyros kaki</i>	C6L1M3	503	69
Magnolídea	Berberidaceae	<i>Epimedium sagittatum</i>	D5M8Q4	508	72
Magnolídea	Gentianaceae	<i>Gentiana triflora</i>	Q96581	516	74
Magnolídea	Campanulaceae	<i>Lobelia erinus</i>	AF41T8	555	67
Magnolídea	Solanaceae	<i>Lycianthes rantonnei</i>	Q9FPN3	511	86
Magnolídea	Solanaceae	<i>Nierembergia sp</i>	Q8LP20	503	88
Magnolídea	Asteraceae	<i>Pericallis cuneata</i>	Q304Q5	504	46
Magnolídea	Fabaceae	<i>Pisum sativum</i>	F2VPT1	515	71

Magnolídea	Salicaceae	<i>Populus trichocarpa</i>	B9HMH5	509	76
Magnolídea	Ericaceae	<i>Rhododendron x pulchrum</i>	C7G325	516	72
Magnolídea	Euphorbiaceae	<i>Ricinus communis</i>	B9SMA4	537	29
Magnolídea	Gesneriaceae	<i>Saintpaulia hybrid cultivar</i>	E3W760	508	71
Magnolídea	Solanaceae	<i>Solanum melongena</i>	P37120	513	82
Magnolídea	Solanaceae	<i>Solanum tuberosum</i>	Q5EWY0	509	85
Magnolídea	Solanaceae	<i>Solanum lycopersicum</i>	D3W9H7	511	85
Magnolídea	Linderniaceae	<i>Torenia hybrid cultivar</i>	Q9FS35	511	70
Magnolídea	Apocynaceae	<i>Vinca major</i>	Q76LL4	505	76
Magnolídea	Violaceae	<i>Viola x wittrockiana</i>	A9EEA6	506	75
Magnolídea	Vitaceae	<i>Vitis amurensis</i>	COKY91	508	74

**CUADRO 2.2. CLASIFICACIÓN DE LAS PLANTAS REPORTADAS PARA LA PROTEÍNA F3'5'H.**

#### **2.4.4 Análisis de similitud e identidad de las secuencias obtenidas para la proteína F3'5'H e identificación de los dominios conservados.**

Fueron un total de 39 especies (cuadro 2.2) incluyendo a *P. hybrida* para el alineamiento completo. Un alineamiento de secuencias proteicas muestra la identificación de correspondencias residuo-residuo, y cualquier asignación de correspondencias que

preserve el orden de los residuos dentro de las secuencias es un alineamiento. Si los residuos de una secuencia tienen un alto grado de correspondencia con residuos de la otra secuencia, son similares y por tanto su cercanía funcional es probable, del alineamiento se muestra una cercanía con respecto a la función y regulación de las diferentes especies, considerando una proteína altamente conservada en especial de la posición 391 a la 553 y en la posición 486 a la 496 se muestra el dominio conservado (fig. 2.4), el cual resultó con la familia de hemoproteínas Citocromo P450, donde un residuo de cisteína conservado en la parte C-terminal de la P450 está implicado en la unión del hemo-hierro. En el alineamiento de la proteína F3'5'H de 10 especies de plantas se observa la región conservada característica de la familia del citocromo P450, "heme binding loop" (Phe-x-x-Gly-x-Arg-x-Cys-x-Gly) de la posición 486 a la 495 (fig. 2.4).

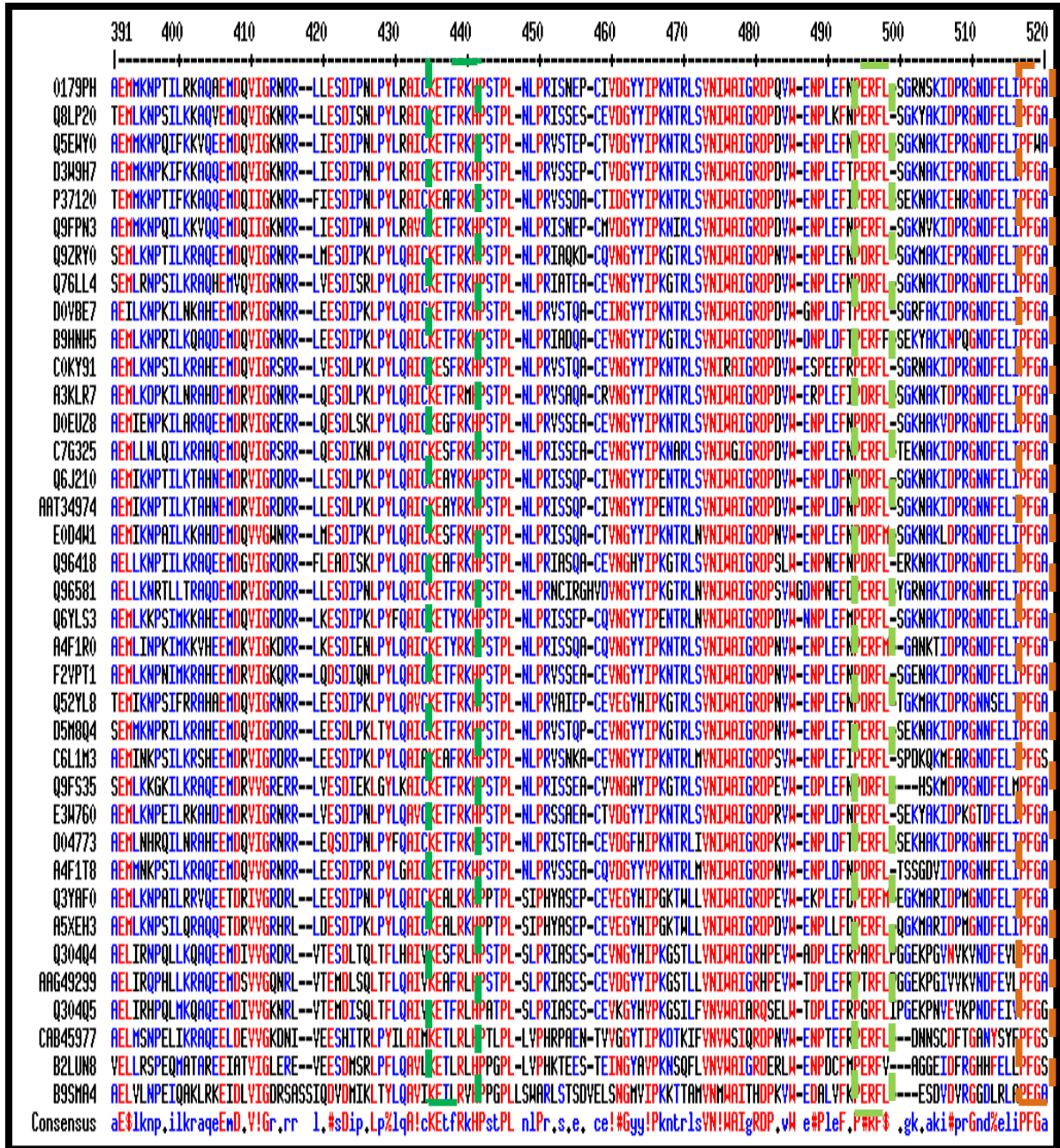


	1	10	20	30	40	50	60	70	80	90	100	110	120	130
Q179PH														
Q8LP20														
Q5E4Y0														
D349H7														
P37120														
Q9FPW3														
Q9ZRY0														
Q76LL4														
D0YBE7														
B9MHW5														
C0KY91														
A3KL7														
D0EUZ8														
C7G3Z5														
Q6J210														
AAT34974														
E0D4M1														
Q96418														
Q96581														
Q6YL53														
A4F1R0														
F2YPT1														
Q52YL8														
D5M8Q4														
C6L1M3														
Q9F535														
E34760														
Q04773														
A4F1T8														
Q3YAF0														
A5XEH3														
Q304Q4														
AAG49Z99														
Q304Q5														
CAB45977														
B2LUN8														
B95M44														
Consensus														

.....l.....lp..PGP.gwP..GALpL\$g.npHv.La.\$akkyGp..ny\$k.G...nvVastp..AraflkLl



	261	270	280	290	300	310	320	330	340	350	360	370	380	390
0179PH	---	EYNEFKDMYVELMTIAGYFNIGDFIPCLAAH	-DLQGI	EKRNLHKKFDALLTKMFE	HNKATTFGR	-K-GKPDFL	OVVME	----	NGDN-S	-EGERLST	TNIKALLNL	FTAGTDTSS	SAIEWAL	
Q8LP20	---	EYNEFKDMYVELMTIAGYFNIGDFIPCLAAH	-DLQGI	EKGKSLHKKFDVLLTKMFE	HNKATTFGR	-E-GKSDFL	OVVMA	----	NRDN-S	-EGERLST	TNIKALLNL	FTAGTDTSS	SVIEWAL	
Q5EMV0	---	EYNEFKDMYVELMTIAGYFNIGDFIPCLAAH	-DLQGI	EKGKSLHKKFDVLLTKMFE	HNKATTFGR	-K-ENPDFL	OVVMA	----	NKDN-S	-EGERLST	TNIKALLNL	FTAGTDTSS	SVIEWAL	
D3W9H7	---	EYNEFKDMYVELMTIAGYFNIGDFIPCLAAH	-DIQGI	EKGKSLHKKFDVLLTKMFE	HNKATTFGR	-K-ENPDFL	OVVMA	----	NRDN-S	-EGERLST	TNIKALLNL	FTAGTDTSS	SVIEWAL	
P37120	---	EYNEFKDMYVELMTIAGYFNIGDFIPCLAAH	-DLQGI	EKGKSLHKKFDVLLTKMFE	HNKATTFGR	-K-GKPDFL	OVVMA	----	NRDN-S	-EGERLST	TNIKALLNL	FTAGTDTSS	SVIEWAL	
Q9FPN3	---	EYNEFKDMYVELMTIAGYFNIGDFIPCLAAH	-DLQGI	EKGKSLHKKFDVLLTKMFE	HNKATTFGR	-K-GKPDFL	OVVMA	----	NRDN-S	-EGERLST	TNIKALLNL	FTAGTDTSS	SAIEWAL	
Q9ZRY0	---	ESNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DLQGI	ERGNKRLHKKFDALLTKMFE	HNKATTFGR	-K-EKPDFL	OYVLA	----	NRDN-S	-EGERLST	TNIKALLNL	FTAGTDTSS	SVIEWAI	
Q76LL4	---	DSNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DLQGI	EGGNKRLHKKFDVLLTKMFE	HNKATTFGR	-K-GKPDFL	OYVLA	----	NRDN-S	-EGERLST	TNIKALLNL	FTAGTDTSS	SVIEWAL	
D0VBE7	---	ESNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DLQGI	EGGNKRLHKKFDVLLTKMFE	HNKATTFGR	-K-GKPDFL	OITMD	----	NREN-S	-AGERLS	TNIKALLNL	FTAGTDTSS	SVIEWAL	
B9NHM5	---	ESNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DLQGI	ERGNKRLHKKFDVLLTKMFE	HNKATTFGR	-K-GKPDFL	OVVMA	----	NQEN-L	-DGEKLS	TNIKALLNL	FTAGTDTSS	SVIEWSL	
C0KY91	---	ESNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DIQGI	EGGNKRLHKKFDVLLTKMFE	HNKATTFGR	-K-GKPDFL	OVVMA	----	NQEN-S	-TGEKLS	TNIKALLNL	FTAGTDTSS	SVIEWSL	
A3KL97	---	ESNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DLQGI	ERGNKRLHKKFDVLLTKMFE	HNKATTFGR	-K-ETPDFL	OVVMA	----	QQEN-P	-GEEKLS	TNIKALLNL	FTAGTDTSS	SVIEWAI	
D0EUZ8	---	ESNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DLQGI	ERGNKRLHKKFDVLLTKMFE	HNKATTFGR	-R-GKPDFL	OVVMA	----	QQEN-S	-GEEKLS	TNIKALLNL	FTAGTDTSS	SVIEWSL	
C7G325	---	ESNEFKDMYVELMTIAGYFNIGDFIPALAAH	-DLQGI	EGGNKRLHKKFDVLLTKMFE	HNKATTFGR	-R-GKPDFL	OVVMA	----	NRET-S	-SQGGPL	SITNIKALLNL	FTAGTDTSS	SVIEWAL	
Q6J210	---	EYNEFKDMYVELMTIAGYFNIGDFIPALAAH	-DLQGI	EKGKSLHKKFDVLLTKMFE	HNKATTFGR	-N-AKPDFL	OVVLA	----	NRHN-S	-EGESL	TNIKALLNL	FTAGTDTSS	SVIEWAL	
AAT34974	---	EYNEFKDMYVELMTIAGYFNIGDFIPALAAH	-DLQGI	EKGKSLHKKFDVLLTKMFE	HNKATTFGR	-N-AKPDFL	OVVLA	----	NRHN-S	-EGESL	TNIKALLNL	FTAGTDTSS	SVIEWAL	
E0D4W1	---	ELNEFKDMYVELMTIAGYFNIGDFIPSAAMH	-DLQGI	EKGKSLHKKFDVLLTKMFE	HNKATTFGR	-K-EKPDFL	OVVLA	----	NRDT-L	-EGERLST	TNIKALLNL	FTAGTDTSS	SVIEWAL	
Q96418	---	ESNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DLQGI	EGGNKRLHKKFDVLLTKMFE	HNKATTFGR	-K-GSPDFL	OVVMA	----	NRDN-S	-EGERLST	TNIKALLNL	FTAGTDTSS	SVIEWAL	
Q96581	---	KSNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DLQGI	EGGNKRLHKKFDVLLTKMFE	HNKATTFGR	-K-QKPDFL	OVVMA	----	NGDN-S	-DGERLST	TNIKALLNL	FTAGTDTSS	SVIEWAL	
Q6YL53	---	ESNEFKDMYVELMTIAGYFNIGDFIPFLAKL	-DLQGI	ERGNKRLHKKFDVLLTKMFE	HNKATTFGR	-K-GKPDFL	OVVMA	----	NHSE	-SDG-EEL	SLTNIKALLNL	FTAGTDTSS	SVIEWSL	
A4F1R0	---	ESNEFKDMYVELMTIAGYFNIGDFIPFLARF	-DLQGI	ERGNKRLHKKFDVLLTKMFE	HNKATTFGR	-K-GKPDFL	OVVMA	----	NHTNE-S	-HEL	SLTNIKALLNL	FTAGTDTSS	SVIEWAL	
F2YPT1	---	ESNEFKDMYVELMTIAGYFNIGDFIPILAAH	-DLQGI	ERGNKSLHKKFDVLLTKMFE	HNKATTFGR	-K-GKPDFL	OVVMA	----	QSKQD	-SDG	GLTITNIKALLNL	FTAGTDTSS	SVIEWAL	
Q52YL8	---	ESNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DLQGI	VRGNKRLHKKFDVLLTKMFE	HNKATTFGR	-K-GKPDFL	OVVMA	----	NRDN-K	-SEGERL	STNIKALLNL	FTAGTDTSS	SVIEWAL	
D5M8Q4	---	ESDFEKDMYVELMTIAGYFNIGDFIPQIAAH	-DYHGI	VGGNKLHKKFDVLLTKMFE	HNKATTFGR	-K-DKPDFL	OVVMA	----	NRDD-KN	-REQL	SHTNIKALLNL	FTAGTDTSS	SVIEWAL	
C6L1M3	---	ESNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DLQGI	ERGNKRLHKKFDVLLTKMFE	HNKATTFGR	-Q-AKPDFL	OITLLT	----	ASRES-S	-GSL	SMINIKALLNL	FTAGTDTSS	SVIEWAL	
Q9FS35	---	EYNEFKDMYVELMTIAGYFNIGDFIPILAAH	-DLQGI	ERGNKSLHKKFDVLLTKMFE	HNKATTFGR	----	KPDFL	OAITA	----	ASL	DYSSEERL	STNIKALLNL	FTAGTDTSS	SVIEWAL
E3W760	---	ELNEFKDMYVELMTIAGYFNIGDFIPWGMH	-DLQGI	EKGKSLHKKFDVLLTKMFE	HNKATTFGR	-K-NRPDL	OVVMA	----	NRDDVDS	PEGERL	TTNIKALLNL	FTAGTDTSS	SVIEWAL	
004773	ASASYS	EYNEFKDMYVELMTIAGYFNIGDFIPYIAAH	-DLQGI	QRDMKVIQKFDVLLTKMFE	HNKATTFGR	-K-DNPDL	OVVMA	----	QENT	----	EGIQ	LNINIKALLNL	FTAGTDTSS	SVIEWAL
A4F1T8	TLKSAG	EYNEFKDMYVELMTIAGYFNIGDFIPSIAMH	-DLQGI	QRDMKVIQKFDVLLTKMFE	HNKATTFGR	-K-GKPDFL	OVVMA	----	NRDN-S	-EGERLST	TNIKALLNL	FTAGTDTSS	SVIEWAL	
Q3YAF0	---	EAKFEKIKELLVGGASNIGDLYPAMRHL	-DPQGI	ARKKLLGLNRFVYRHSKFLA	HNKATTFGR	-E-GKPDFL	OVVMA	----	DKIAG	DDGEG	SEENIKGFS	OVVMA	FTAGTDTSS	SVIEWAM
A5XEH3	---	EAKFEKIKELLVGGASNIGDLYPAMRHL	-DPMGI	VKKRLLGLNRFVYRHSKFLA	HNKATTFGR	-Q-GKPDFL	OVVMA	----	SEL	TGEDG	GLCEDNIKGFS	OVVMA	FTAGTDTSS	SVIEWAM
Q304Q4	GDPKAD	EYNEFKDMYVELMTIAGYFNIGDFIPMLDWH	-DLQGI	VTKNKVYHDFDFL	OVVMA	----	AHDM	STLIS	----	LKED	VNGEG	GLSDIEIKALLNL	FTAGTDTSS	SVIEWAM
AAG49239	DDPKAD	EYNEFKDMYVELMTIAGYFNIGDFIPMLDWH	-DLQGI	VTKNKVYHDFDFL	OVVMA	----	GHID	STLIS	----	LKOD	TSVDR	GRKSDIEIKALLNL	FTAGTDTSS	SVIEWAI
Q304Q5	GDPKAD	EYNEFKDMYVELMTIAGYFNIGDFIPMLDWH	-DLQGI	VTKNKVYHDFDFL	OVVMA	----	RHYD	STLIS	----	LKON	ADGEG	GLSFIEIKALLNL	FTAGTDTSS	SVIEWGI
CAB45977	---	SVGTEFGVISEITRLLSEPNVSDFFPFLARF	-DLQGI	VKRNVCARELDVLDRAIE	HNKATTFGR	-EYKDFL	OVVMA	----	LKQEG	DSVPT	INNIKALLNL	FTAGTDTSS	SVIEWAM	
B2LUN8	---	ESAQELKDLMSGINEEYGRPNVSDFFPFLAPI	-DLQGI	CRRFARYIKLSDFFVEIEN	HNKATTFGR	----	NKHD	OVVMA	----	QLS	RENSK	SRNTIIS	FTAGTDTSS	SVIEWAM
B9SMA4	---	EAKELKELVREGLELLGFNWSOYLPWLDYF	OPFCIKERCLALVPRVKLVKQID	ENQSNPKSVF	OVVMA	----	LDG	OVVMA	----	OVVMA	OVVMA	OVVMA	FTAGTDTSS	SVIEWAM
Consensus	...	e.nEFKd!vE1t.a6.fN!g0%IP.law	DLQGI	e.gmk.lhkkfd.ll.n..#h.....r..	.kPDFL#...	...	.....s.ege.Ls..n!kall!#ftaGDTSS.s.iEWa!							



	521	530	540	550	560	570	580	590	596
0179PH	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--SEVIELNMEEAFGLALQKAVPLEAMVTPRL	QLDYYVP					
Q8LP20	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--SDVIELNMEEVFGALALQKAVPLEAMVTPRL	PLNYYAP					
Q5EAY0	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--NDVIDINMEESEFGLALQKAVPLEAMVTPRL	SLDVYCC					
D3M9H7	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--NNVIDINMEESEFGLALQKAVPLEAMVTPRL	SLDVYRC					
P37120	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--NDVVDINMEESEFGLALQKAVPLEAMVTPRL	SFDIYQSSEPF					
Q9FPN3	GRRICAGT	RMGIYVVEYILGTLVHSFDWKFS	--NDVKEINMEESEFGLALQKAVPLEAMVTPRL	PFDDVYYTN					
Q9ZRY0	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--DDINELNMDESEFGLALQKAVPLAMVSPRL	PINAYSP					
Q76LL4	GRRICAGT	RMGILLVEYILGTLVHSFDWMLP	--SSVTKLNMDESEFGLALQKVVPLAALVTPRL	PINAYSP					
D0VBE7	GRRICAGT	RMGIYVVEYILGTLVHSFDWMLP	--PGTGELNMDEAFGLALQKAVPLSAMVTPRL	LAPTAYVS					
B9NNH5	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--KDY-ELNMDEVFGALALQKAVPLSAMVTPRL	EPNAYLA					
C0KY91	GRRICAGT	RMGIYVVEYILGTLVHSFDWKMP	--D-GVEINMDEAFGLALQKAVPLSAMVTPRL	HQSAYAY					
A3KLR7	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLK	--DDGVELNMDESEFGLALQKAVPLSAMVSPRL	PPNAYAA					
D0EUZ8	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--D-GVELNMDEAFGLALQKAVPLAAVTPRL	LYPSAYVA					
C7G325	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--DGMGELNMDESEFGLALQKAVPLAMVTPRL	QPSAYAMYCRLNG					
Q6J210	GRRICAGT	RMAMVVEYILGTLVHSFDWELP	--YGVDEMDEEAFGLALQKAVPLAMVSPRL	PPHCYAP					
AAT34974	GRRICAGT	RMAMVVEYILGTLVHSFDWELP	--YGVDEMDEEAFGLALQKAVPLAMVSPRL	PPHCYAP					
E0D4M1	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--EGYKEMNLDEAFGLALQKAVPLAMVTPRL	PSNICYAP					
Q96418	GRRICAGT	RLGILLVEYILGTLVHSFDWELP	--SSVIELNMDEPFGLALQKAVPLAMVTPRL	PLHIYCP					
Q96581	GRRICAGT	RMGILLVEYILGTLVHSFDWKLK	--FSEDELNMDETFGLALQKAVPLAMVTPRL	PLHYAP					
Q6YLS3	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--NGVRELNMDESEFGLALQKAVPLAALVTPRL	NPSAYIS					
A4F1R0	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--NGVVELNMDETFGLALQKAVPLAALVTPRL	PPTAYNVINS					
F2VPT1	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--NGVVALNMDESEFGLALQKAVPLAVVTPRL	SPSAYIP					
Q52YL8	GRRICAGT	RMGIYVVEYILGTLVHAFEWKL	---RDGEMLNMEETFGIALQKAVPLAAVTPRL	PPSAYVY					
D5M8Q4	GRRICAGT	RMGITLVEYILGTLVHSFDWKV	---CNGVEINMDEAFGLALQKAVPLAAVTPRL	SPKAYLY					
C6L1M3	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	--DGVVKLNMDETFGLALQKAVPLSAMATPRL	SPRAYVA					
Q9FS35	GRRICAGT	RMGIYVVEYILATLVHSFDWKAAD	--QDNNINMEEAFGIALQKATPLKALVTPRL	PLHCYLP					
E3M760	GRRICAGT	RMGVVVECLLGTLVHSFDWKF	----DGESHDMDETFGLALQKAVPLATVTPRL	PPSCYVGSN					
004773	GRRICAGT	RMGAASVEYILGTLVHSFDWKLP	--DGVVEYNMEESEFGLALQKAVPLAAVTPRL	PPSSYTY					
A4F1T8	GRRICAGT	RMGIYVVEYILGTLVHSFDWKVAAAEAAINVD	MESEFGLALQKAVPLAALVTPRL	PPSAYSPS					
Q3YAF0	GRRICAGT	LMGMVVEYILGTLVHAFDWSLP	--EGVVELDMEEGGLVLPKAVPLVYTPAR	PAAAYGVY					
A5XEH3	GRRICAGT	LAGMLVQYYLGTLVHAFDWSLP	--EGVVELDMEEGGLVLPKAVPLVYTPAR	PAAAYGLL					
Q304Q4	GRRICAGT	SLALRMVHLLIATLIQAFDWELANGLDPERLN	MEEFGISVQKAEPLVHPRPRL	APHTYI					
AAG49299	GRRICAGT	SLALRTVQLLMTLVQAFDWELANGIKPEKLN	MEAFGLSVQRAEPLVHPRPRL	PPHYKSG					
Q304Q5	GRRICAGT	SLGLRMVHLLIATLVQAFDWELANGLEPEKLN	MEVFGISLQVQPLVHPRPRL	LARHYGTG					
CAB45977	GRRICAGT	LAERMVLYTLATLLHSFDWKIP	---EGHVLDLKEKFGIYVLPKAVPLVPIPR	FSDSNLYL					
B2LUN8	GRRICPGM	PLGVRMVLMLASLLQSFENGLPDGMKPEDLDL	TEKHGLSTVLAAPLKAATPTKHN						
B9SMA4	GRRVCPG	NLGLVMVSLWAKLVQHFEWVQ	---DMANPYDLSEVLKLSCEKHPKCAVTPR	PRNNE					
Consensus	GRRICAGT	RMGIYVVEYILGTLVHSFDWKLP	.....L#\$.E.fGLalqavPL.a.vtPRL....y...						

## FIGURA 2.4. ALINEAMIENTO MÚLTIPLE Y DOMINIOS CONSERVADOS.

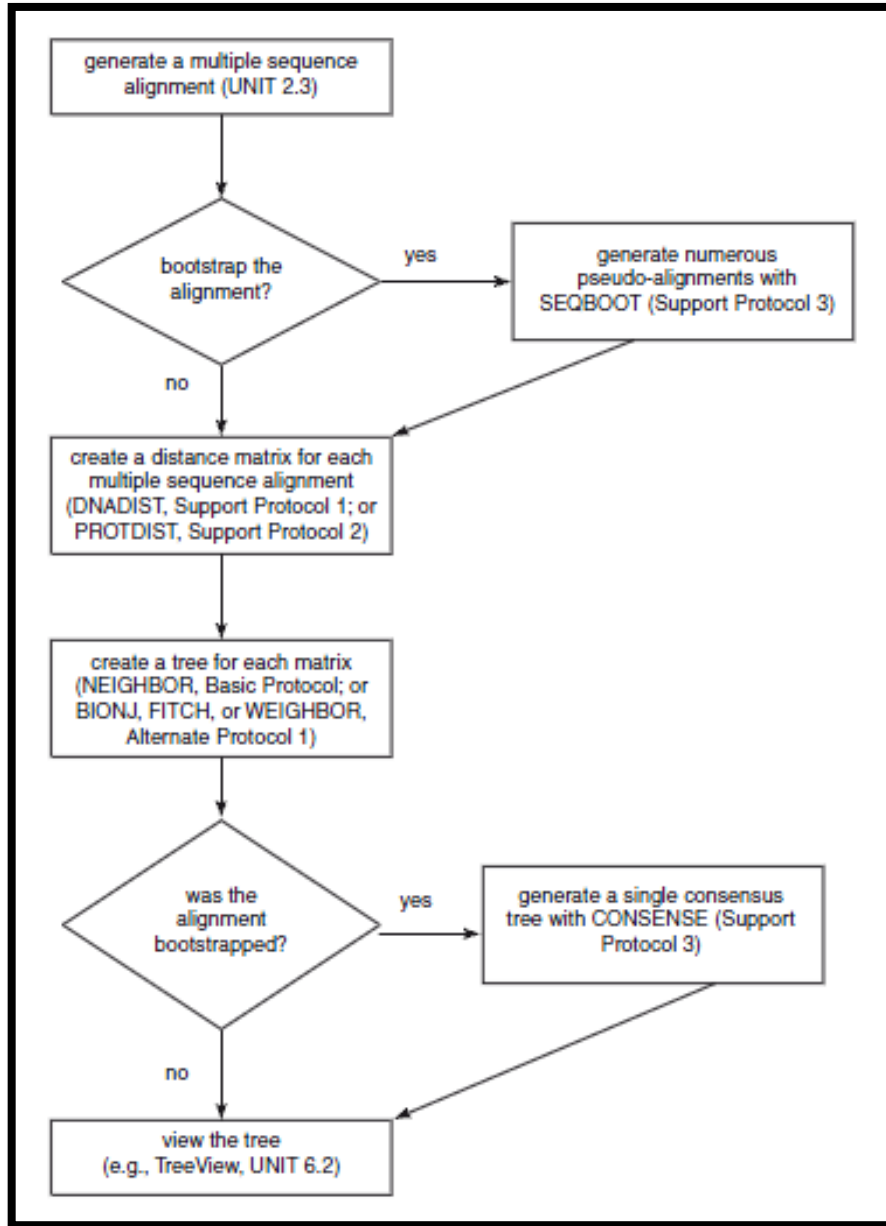
Se incluyen 39 secuencias de proteína considerando a las especies *Petunia hybrida* (F35H0179PH), *Arabidopsis thaliana* (CAB45977), *Callistephus chinensis* (AAG49299), *Cyclamen graecum* (BAJ08041), *Glandularia x hybrida* (AAT34974), *Delphinium grandiflorum* (Q52YL8), *Eustoma grandiflorum* (Q96418), *Glycine max* (Q6YLS3), *Glycine soja* (D9N1C3), *Gossypium hirsutum* (D0VBE7), *Osteospermum hybrid c.* (Q304Q4), *Phalaenopsis hybrid c.* (Q3YAF0), *Verbena hybrida* (Q6J210), *Vitis vinífera* (Q2UYU7), *Antirrhinum kelloggii* (E0D4W1), *Camellia sinensis* (A3KLR7), *Campanula medium* (O04773), *Catharanthus roseus* (Q9ZRY0), *Clitoria ternatea* (A4F1R0), *Cyclamen persicum* (D0EUZ8), *Dendrobium hybrid* (A5XEH3), *Diospyros kaki* (C6L1M3), *Epimedium sagittatum* (D5M8Q4), *Gentiana triflora* (Q96581), *Lobelia erinus* (AF41T8), *Lycianthes rantonnei* (Q9FPN3), *Nierembergia sp* (Q8LP20) *Pericallis cuneata* (Q304Q5), *Petunia hybrida* (A3FEQ7), *Pisum sativum* (F2VPT1), *Populus trichocarpa* (B9HNNH5), *Rhododendron x pulchrum* (C7G325), *Ricinus communis* (B9SMA4), *Saintpaulia hybrid cultivar* (E3W760), *Solanum melongena* (P37120), *Solanum tuberosum* (Q5EWY0), *Solanum lycopersicum* (D3W9H7), *Torenia hybrid cultivar* (Q9FS35), *Vanda coerulea* (B2LUN8), *Vinca major* (Q76LL4), *Viola x wittrockiana* (A9EEA6) y *Vitis amurensis* (COKY91). Los diferentes colores muestran las zonas conservadas (rojo) y las no conservadas (azul) entre las secuencias (aminoácidos) de las especies. El recuadro punteado de color azul representa al motivo de prolinas (PPGP), de color morado el motivo de unión al oxígeno (AGTDTS), en verde oscuro el motivo EXXR, en verde claro la firma del P450 PERFL y en color naranja se muestra la región del dominio conservado característica de la familia del citocromo P450, de unión al hierro (FGAGRRICAG).

### 2.4.5. Obtención y análisis del árbol filogenético molecular de la relación de función de la proteína F3'5'H.

Una vez obtenidas las secuencias de ADN para la proteína F3'5'H y hecho el alineamiento, se prosiguió a realizar la obtención del árbol para el análisis de la función de la proteína F3'5'H en las diferentes especies reportadas. Las 39 especies utilizadas fueron: *Arabidopsis thaliana* (CAB45977), *Callistephus chinensis* (AAG49299), *Cyclamen graecum* (BAJ08041), *Glandularia x hybrida* (AAT34974), *Delphinium grandiflorum* (Q52YL8),

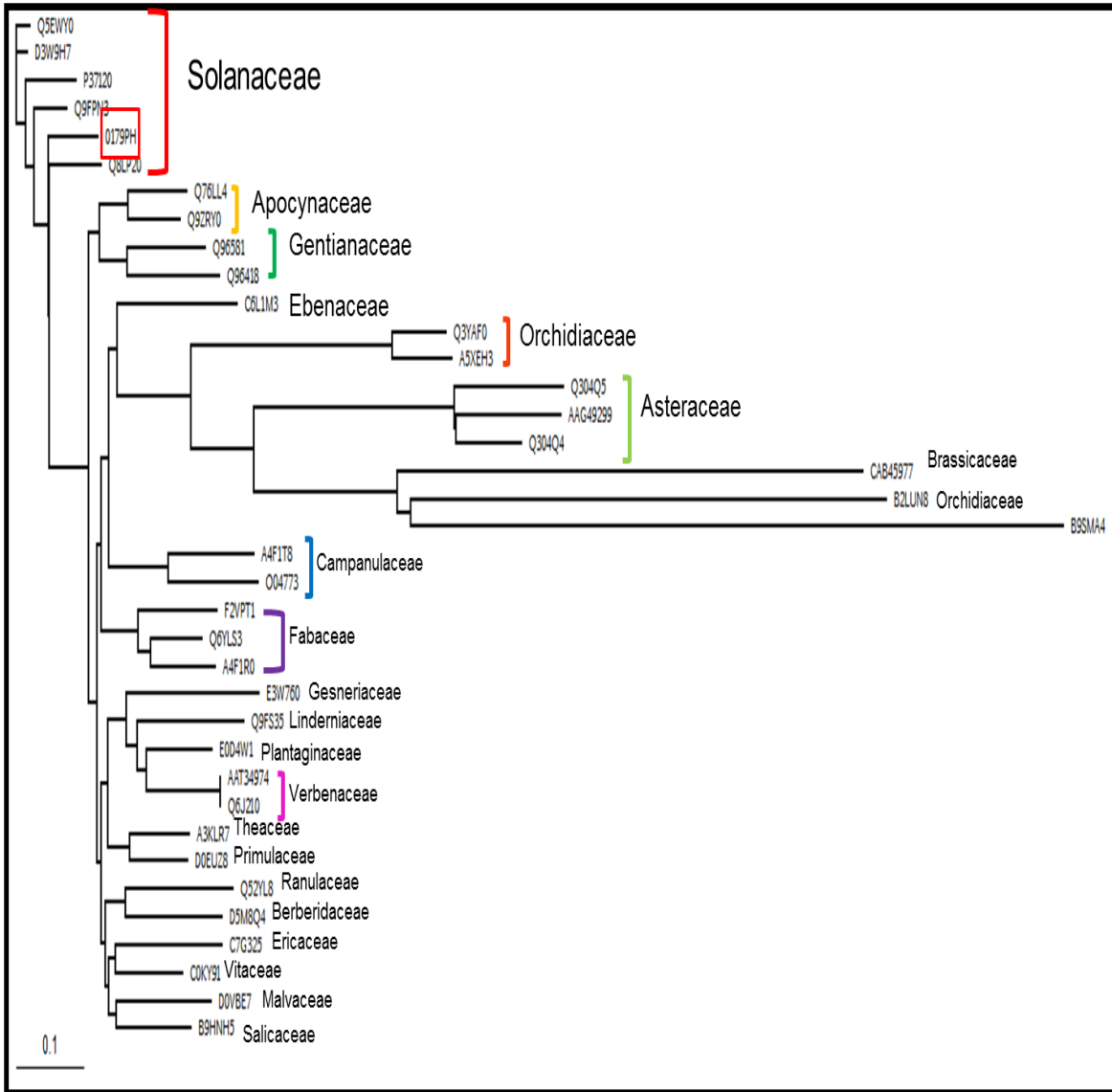
*Eustoma grandiflorum* (Q96418), *Glycine max* (Q6YLS3), *Glycine soja* (D9N1C3), *Gossypium hirsutum* (D0VBE7), *Osteospermum hybrid c.* (Q304Q4), *Phalaenopsis hybrid c.* (Q3YAF0), *Verbena hybrida* (Q6J210), *Vitis vinifera* (Q2UYU7), *Antirrhinum kelloggii* (E0D4W1), *Camellia sinensis* (A3KLR7), *Campanula medium* (O04773), *Catharanthus roseus* (Q9ZRY0), *Clitoria ternatea* (A4F1R0), *Cyclamen persicum* (D0EUZ8), *Dendrobium hybrid* (A5XEH3), *Diospyros kaki* (C6L1M3), *Epimedium sagittatum* (D5M8Q4), *Gentiana triflora* (Q96581), *Lobelia erinus* (AF41T8), *Lycianthes rantonnei* (Q9FPN3), *Nierembergia sp* (Q8LP20) *Pericallis cuneata* (Q304Q5), *Petunia hybrida* (A3FEQ7), *Pisum sativum* (F2VPT1), *Populus trichocarpa* (B9HMH5), *Rhododendron x pulchrum* (C7G325), *Ricinus communis* (B9SMA4), *Saintpaulia hybrid cultivar* (E3W760), *Solanum melongena* (P37120), *Solanum tuberosum* (Q5EWY0), *Solanum lycopersicum* (D3W9H7), *Torenia hybrid cultivar* (Q9FS35), *Vanda coerulea* (B2LUN8), *Vinca major* (Q76LL4), *Viola x wittrockiana* (A9EEA6) y *Vitis amurensis* (COKY91).

El procedimiento consistió en tomar el archivo obtenido del alineamiento por ClustalW y del esquema del paquete PHYLIP que se muestra en la figura 2.5 se realizó un bootstrap ya que este último permite obtener estimaciones de medidas de precisión así como la realización de contrastes de hipótesis en aquellas situaciones en las que no se dispone de información acerca de la distribución muestral de un estadístico o en casos en los que la distribución muestral es dependiente de parámetros desconocidos (López y Elosua, 2004), luego para generar otros alineamientos y dar más soporte y viabilidad se usó el SEQBOOT. Para crear una matriz de cada alineamiento múltiple de las secuencias se utilizó el PROTDIST, seguido de la creación del árbol para cada matriz con el NEIGHBOR, luego nuevamente un bootstrap finalizando con un consenso simple para el árbol con CONSENSE y para visualizar el árbol con TREEVIEW.



**FIGURA 2.5. METODOLOGÍA DEL NEIGHBOR-JOINING TREES DEL PAQUETE PHYLIP, PARA EL ÁRBOL DE FUNCIÓN DE LA PROTEÍNA F3'5'H (Gascuel, 2003).**





**FIGURA 2.6. ÁRBOL DE RELACION DE FUNCIÓN DE LA PROTEÍNA F3'5'H.**

*Arabidopsis thaliana* (CAB45977), *Callistephus chinensis* (AAG49299), *Cyclamen graecum* (BAJ08041), *Glandularia x hybrida* (AAT34974), *Delphinium grandiflorum* (Q52YL8), *Eustoma grandiflorum* (Q96418), *Glycine max* (Q6YLS3), *Glycine soja* (D9N1C3), *Gossypium hirsutum* (D0VBE7), *Osteospermum hybrid c.* (Q304Q4), *Phalaenopsis hybrid c.* (Q3YAF0), *Verbena hybrida* (Q6J210), *Vitis vinífera* (Q2UYU7), *Antirrhinum kelloggii* (E0D4W1), *Camellia sinensis* (A3KLR7), *Campanula medium* (O04773), *Catharanthus roseus* (Q9ZRY0), *Clitoria ternatea* (A4F1R0), *Cyclamen persicum* (D0EUZ8), *Dendrobium hybrid* (A5XEH3), *Diospyros kaki* (C6L1M3), *Epimedium sagittatum* (D5M8Q4), *Gentiana triflora* (Q96581), *Lobelia erinus* (AF41T8), *Lycianthes rantonnei*

(Q9FPN3), *Nierembergia sp* (Q8LP20) *Pericallis cuneata* (Q304Q5), *Petunia hybrida* (A3FEQ7), *Pisum sativum* (F2VPT1), *Populus trichocarpa* (B9HNN5), *Rhododendron x pulchrum* (C7G325), *Ricinus communis* (B9SMA4), *Saintpaulia hybrid* cultivar (E3W760), *Solanum melongena* (P37120), *Solanum tuberosum* (Q5EWY0), *Solanum lycopersicum* (D3W9H7), *Torenia hybrid* cultivar (Q9FS35), *Vanda coerulea* (B2LUN8), *Vinca major* (Q76LL4), *Viola x wittrockiana*, (A9EEA6) y *Vitis amurensis* (COKY91). El recuadro de color rojo indica a la especie con la que se trabaja *Petunia hybrida*. Perteneciendo todas las especies de la proteína F3'5'H al CYP75A, refiriéndose al citocromo P450 de la familia 75 que interviene en la ruta fenilpropanoide, específicamente en la catalización de compuestos flavonoides y la subfamilia A que cataliza el F3'5'H.

## 2.5.REFERENCIAS BIBLIOGRÁFICAS

Cañedo R. y Arencibia R. (2004) Bioinformática: en busca de los secretos moleculares de la vida. Acimed Vol.12:6.

European Bioinformatics Institute (2006) What is Bioinformatics?  
[http://www.ebi.ac.uk/2can/bioinformatics/bioinf\\_what\\_1.html](http://www.ebi.ac.uk/2can/bioinformatics/bioinf_what_1.html)

Gascuel O. (2003) *Getting a Tree Fast: Neighbor Joining and Distance-Based Methods*. Current Protocols in Bioinformatics. John Wiley & Sons, Inc 6.3.1-6.3.23.

Lee D., Redfern O. y Orengo C. (2007) *Predicting protein function from sequence and structure*. Nature <reviews. Molecular Cell Biology. Vol 8 (12).pp. 995-1005.

Romiti M. y Cooper P. (2011) Entrez Help.NCBI.pp. 1-19.

Tatusov T. y Tatusov R. (2011) ORF FINDER: Open Reading Frame Finder.  
<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>

Thompson J., Gibson T. y Higgins D. (2003) Multiple Sequence Alignment Using

ClustalW and ClustalX. Current Protocols in Bioinformatics. Unit 2.3.

Wiley J. y Sons, Inc. (2003) The Importance of Biological Databases in Biological Discovery. Current Protocols in Bioinformatics, 1.1.1-1.1.4

EXPASY (2012) Bioinformatics Resource Portal. <http://expasy.org/>

**CAPITULO III: ANÁLISIS DE LA FUNCIÓN, REGULACIÓN Y  
CARACTERÍSTICAS DE LA PROTEÍNA FLAVONOIDE - 3', 5' -  
HYDROXILASA DE *Petunia hybrida***

### **3.1. INTRODUCCIÓN**

Las proteínas tienen una gran influencia en la maquinaria molecular de la vida, llevando a cabo innumerables funciones en todas las células vivas. Conocer la función y la estructura de las proteínas es crucial para el desarrollo de mejores medicamentos, cultivos, incluso biocombustibles sintéticos. Por lo tanto, el conocimiento de la estructura y función de proteínas conduce a avances cruciales en la vida de la ciencia y la biología (Chen et al, 2009; Rangwala y Karypis, 2008).

El campo de la bioinformática, actualmente es el más popular, es el análisis e interpretación de varios tipos de datos, incluyendo dominios y estructura de proteínas. Objetivos tales como el desarrollo e implementación de herramientas que permitan el acceso, uso y manejo de varios tipos de información y desarrollo de nuevos algoritmos y estadísticos con los cuales se pueda relacionar partes de un conjunto enorme de datos, métodos para relacionar e identificar una secuencia de ADN con otra, predecir estructura o función de proteínas y poder agrupar secuencias de proteínas en familias relacionadas son puntos fundamentales para esta disciplina.

Las estructuras de las proteínas son determinadas principalmente con cristalografía de rayos X o la resonancia magnética nuclear (RMN), pero estos métodos llevan mucho tiempo, son costosos y no es factible para todas las proteínas. Por otro lado, los enfoques experimentales para determinar la función de la proteína (por ejemplo, eliminación de genes, mutaciones específicas e inhibiciones de los estudios de expresión génica) son bajos. Con los recientes avances en las tecnologías de secuenciación, hemos visto un crecimiento exponencial de la información de secuencias de proteínas. En consecuencia, los investigadores son cada vez más dependientes de los métodos de cómputo para extraer información útil de determinadas estructuras en tercera dimensión así como sus funciones (Chen et al, 2009; Hu et al, 2007; Pandey et al, 2006; Rangwala y Karypis, 2008).

Desentrañar la predicción de la relación entre la estructura de proteínas de la secuencia de la información pura y de la estructura en 3D y / o función sigue siendo uno de los retos fundamentales de la biología molecular. La predicción de la función suele ser abordado por uso de la herencia a través de homología, es decir, las proteínas con secuencias similares (ascendencia evolutiva común) con frecuencia llevan a cabo funciones similares. Sin embargo, varios estudios han demostrado que existe una fuerte correlación entre la conservación de la estructura y la función, es decir, la estructura implica la función, y existe una mayor correlación entre la conservación de la secuencia y la estructura, esto es, la secuencia implica la estructura (Olivares y García, 2004; Rangwala y Karypis, 2008).

Secuencia → Estructura → Función

Así la función de las macromoléculas biológicas está determinada tanto por su estructura como por la dinámica conformacional (3D). Estas moléculas son sistemas inherentemente flexibles, por lo que la evaluación de la calidad de las estructuras de la proteína es una importante parte de la validación experimental de la estructura y desempeña un objetivo crucial en la predicción de estructuras proteicas, donde los modelos predichos, puede contener errores sustanciales (Pandey et al, 2006; Rangwala y Karypis, 2008).

La proteína F3'5'H está involucrada en la biosíntesis de delphinidina que es la responsable del color azul de las flores. Se ha hecho una investigación amplia al respecto, demostrando la variabilidad de funciones según el tejido o el efecto de los genes en los diferentes colores de la flor. Su función es un punto clave para esclarecer más al respecto, ya que actualmente no se ha encontrado el punto clave en la obtención de este color, para poder ser introducido en otras especies que no lo poseen como son la rosa, el clavel y el crisantemo (Tanaka et al, 2005).

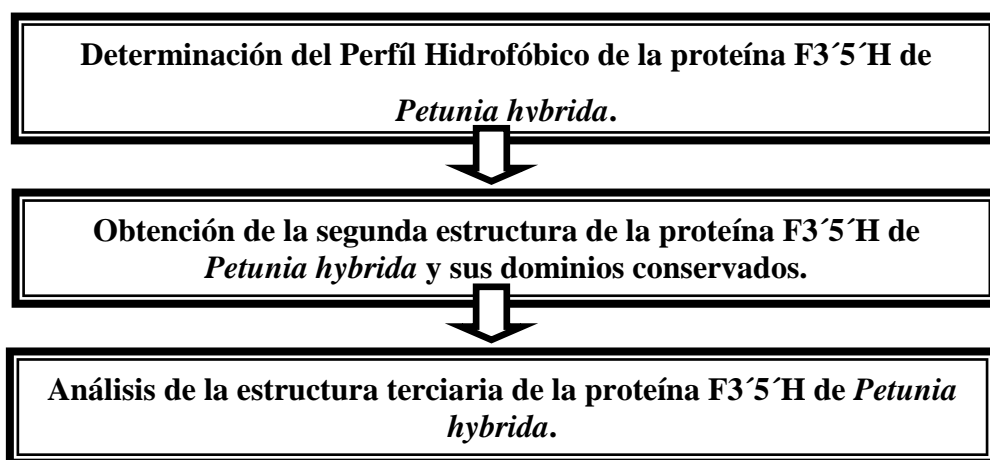
Por su parte, la petunia es una especie comercial muy demandada, que posee esta proteína, siendo utilizada como modelo de estudio para los fines antes mencionados. Dilucidar aún más esta tarea, reuniendo las herramientas de la bioinformática para llegar a solucionar, como es que se da la obtención de este color azul en las flores, por lo que este

capítulo tiene como objetivo el determinar los dominios conservados para la proteína F3'5'H de *Petunia hybrida* en relación a su función en diferentes tejidos, determinando por igual la conformación y el plegamiento de la proteína F3'5'H de *Petunia hybrida*.

### 3.2. ESTRATEGIA EXPERIMENTAL

Las actividades se realizaron en las instalaciones del Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco, A.C. (CIATEJ) Unidad Sureste ubicado en la ciudad de Mérida. Utilizando una computadora Toshiba con capacidad de 3GB y el sistema operativo Windows 7.

El esquema general para el análisis de la función y estructura de la proteína F3'5'H es:



### 3.3 MATERIALES Y MÉTODOS

#### 3.3.1. Determinación del Perfil Hidrofóbico.

Para la predicción de la solubilidad y estructura secundaria de la proteína F3'5'H, se determinó el perfil hidrofóbico con el bioedit V7 0.9 del modelo kyte y Doolittle.

### **3.3.2. Predicción de la estructura secundaria de la proteína F3'5'H, e identificación de los dominios conservados en *Petunia hybrida*.**

La obtención de la estructura secundaria de la proteína F3'5'H y los dominios conservados se analizaron con el programa psipred <http://bioinf.cs.ucl.ac.uk/psipred/>.

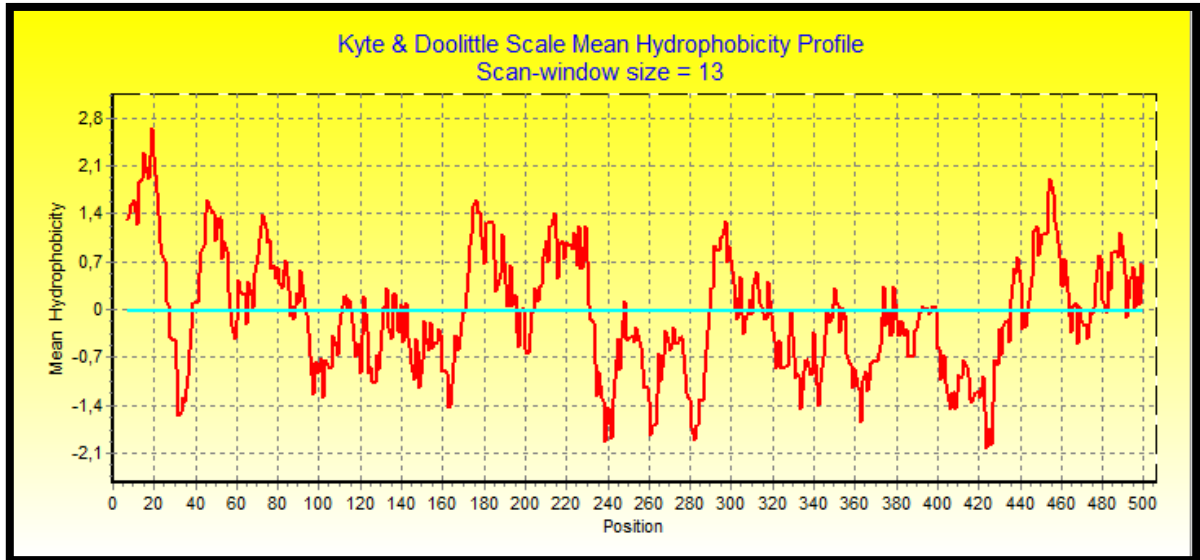
### **3.3.3 Análisis de la estructura terciaria de la proteína F3'5'H en *Petunia hybrida*.**

Por último se obtuvo la estructura terciaria de la proteína F3'5'H de *Petunia hybrida* para el estudio de la posible interacción de los residuos de los aminoácidos para la especulación del plegamiento de la proteína con el programa Rasmol.

## **3.4 RESULTADOS Y DISCUSIÓN**

### **3.4.1. Perfil Hidrofóbico de la proteína F3'5'H en *Petunia hybrida*.**

Hay un buen número de propiedades primarias de las proteínas que pueden ser inferidas directamente a partir de su secuencia como el punto isoeléctrico, la masa atómica, entre otros así también muchas otras características fisicoquímicas que se pueden derivar fundamentalmente a partir de conocer la composición en aminoácidos de una proteína. Una herramienta muy útil son los perfiles de hidrofobicidad. La técnica lo que hace es calcular en unas ventanas de residuos determinadas cual es la hidrofobicidad media en cada residuo a partir de los coeficientes de los aminoácidos. Este tipo de perfil es muy útil, para por ejemplo determinar regiones internas (típicamente hidrofóbicas) o externas (típicamente hidrofílicas) de proteínas. También es información que se puede procesar para determinar regiones transmembranas, regiones de interacción proteína proteína, etc. (Claverie y Notredame, 2007).



**FIGURA 3.1. PERFIL HIDRÓFÓBICO DE LA PROTEÍNA FLAVONOIDE 3-5-HIDROXILASA EN PETUNIA.** En el eje de abscisas esta la secuencia y en ordenadas el índice de hidrofobicidad. La línea negra marca el nivel establecido para el modelo de Kyte y Doolittle.

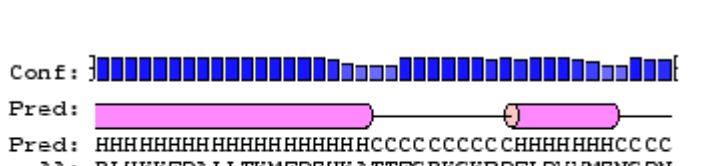
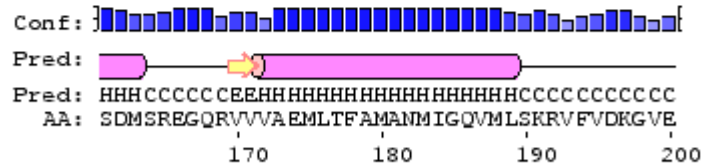
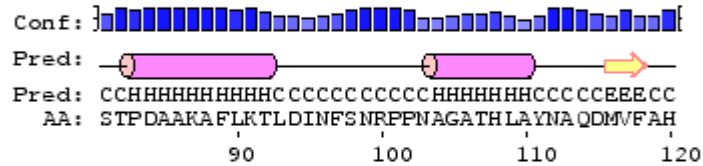
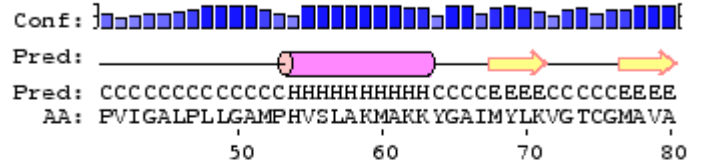
El perfil hidrofóbico es una estimación de las hidrofobicidades residuales de la secuencia problema, esto debido a que algunos segmentos solubles de proteínas globulares son más hidrofóbicos que otros transmembranales, de igual longitud. Por ellos se llega a la necesidad de incluir análisis de este tipo acerca de la lateralidad del patrón hidrofóbico, es como un modelo predictivo. Según el modelo de kyte y Doolittle, se estima un punto medio en la puntuación (score) de 1.6 para este modelo señalando que es poco soluble mostrando un pico sobre la línea marcada y apenas otro más, como puede observarse en la figura 3.1, estos son los segmentos transmembranales para esta proteína, y además confirma el hecho que se reporta para antocianinas que se consideran poco solubles, localizándose en organelos esféricos conocidos como antocianoplastos que se forman mientras la síntesis de pigmentos esta en operación. Un punto importante es su relación con la copigmentación, mecanismo de interacción molecular que previene el ataque nucleofílico del agua previniendo la decoloración del pigmento.



### **3.4.2 Análisis de la estructura secundaria de la proteína F3'5'H en *Petunia hybrida*.**

La estructura secundaria de una proteína es un nivel de organización que adquiere la molécula, dependiendo de cómo sea la secuencia de aminoácidos que la componen. La rigidez del enlace peptídico, la capacidad de giro de los enlaces establecidos con el carbono asimétrico y la interacción de los radicales de los aminoácidos con la disolución en la que se encuentra, lleva a plegar la molécula sobre sí misma. Las conformaciones resultantes pueden ser la estructura en  $\alpha$ -hélice, la  $\beta$ -plegada (Claverie y Notredame, 2007; Rangwala y Karypis, 2008).

Los dominios conservados de la proteína F3'5'H se encontraron relacionados con respecto a la familia de los Citocromos P450 ya que esta familia tiene que ver con las hidroxilaciones del anillo b de los flavonoides que a su vez tiene que ver con la proteína F3'5'H y la obtención del color azul de las flores (fig. 3.4).





### **3.4.3 Análisis de la estructura terciaria de la proteína F3'5'H en *Petunia hybrida*.**

La estructura primaria de una cadena determina su estructura terciaria, esta última se refiere al plegamiento de dominios, que son las unidades básicas de estructura y función y el arreglo final de los dominios en el polipéptido. La estructura de las proteínas globulares en solución acuosa es compacta, con una gran densidad (empaquetamiento) de los átomos en el núcleo de la molécula. Las cadenas hidrofóbicas se encuentran en el interior, por el contrario las hidrofílicas en la superficie de la molécula. La función de la estructura secundaria (hélices  $\alpha$  y hojas  $\beta$ ), es de proveer de la máxima posibilidad de hacer interacciones en el interior de la proteína, evitando así que las moléculas de agua queden atrapadas en su interior, interrumpiendo su integridad (Vazquez-Contreras, 2003).

La estructura terciaria de la proteína F3'5'H de *Petunia hybrida* que se obtuvo para el estudio de la posible interacción de los residuos de los aminoácidos para la especulación del plegamiento de la proteína que resultó ser un monómero, pudiéndose observar en la figura 3.5, con el programa Rasmol, con dos homeodominios. Cada uno con estructuras secundarias b-plegadas y a-hélices. Mostrando una gran capacidad de enrollamiento y plegamiento.



**FIGURA 3.3. ESTRUCTURA TERCIARIA DE LA PROTEÍNA F3'5'H DE PETUNIA HYBRIDA.** De color fucsia se pueden ver las estructuras  $\alpha$ -hélices. Las flechas amarillas indican las estructuras  $\beta$ -plegadas y las líneas grises con azules son las coils.

Se aprecia la membrana de ancla en la parte superior de la estructura con una alfa hélice de color fucsia, dentro de las beta plegadas están los motivos que pertenecen al dominio de bisagra, y en el complejo de las alfa hélices esta el dominio globular.

### 3.5. REFERENCIAS BLIBLIOGRÁFICAS

Claverie J.M. y Notredame C. (2007) *Bioinformatics For Dummies*®, 2nd Edición. Wiley Publishing, Inc.

Chen L., Wang R. y Zhang X. (2009) *Basic Concepts in molecular biology*. Biomolecular Networks. John Wiley & Sons, Inc. pp.1-22.

Hu H., Harrison R., Tai P. y Pan Y. (2007) *Current Methods for protein secondary-structure prediction based on support vector machines*. Knowledge Discovery in Bioinformatics: Techniques, Methods, and Applications.

Olivares L. y García S. (2004) *Plegamiento de las proteínas: Un problema interdisciplinario*. Revisión. Rev. Soc. Quím. Méx. 2004, 48, 95-105

Pandey G., Kumar V. y Steinbach M. (2006) *Computational Approaches for Protein Function prediction: A Survey*. Department of Computer Science and Engineering University of Minnesota. pp.1-142.

Rangwala y Karypis. (2010) *Introduction to Protein Structure Prediction: Methods and Algorithms*. John Wiley & Sons, Inc. pp.1-14.

Tanaka Y., Katsumoto Y., Brugliera y Mason. (2005) *Genetic engineering in floriculture*. Review of Plant Biotechnology and Applied Genetics Plant Cell, Tissue and Organ Culture. 80: 1–24. Springer.

**CAPITULO IV: PROPUESTA PARA EL AISLAMIENTO DEL GEN**  
**FLAVONOIDE - 3',5' - HIDROXILASA EN FLORES CON COLORACION AZUL**

#### **4.1 INTRODUCCIÓN**

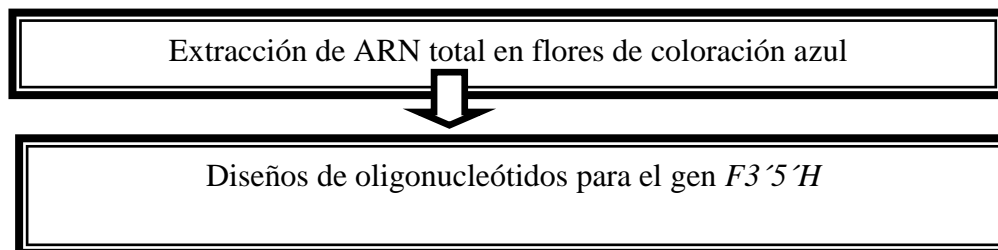
La industria de la floricultura ha centrado su atención principalmente en el desarrollo de nuevas variedades de colores en flores. La base para ello se estableció hace algunos años a través del aislamiento de genes "azules" y los genes de biosíntesis de etileno, lográndose importantes avances en la comercialización de variedades de flores modificadas genéticamente (Mol et al, 1999).

La F3'5'H tiene la función de catalizar la hidroxilación en dihidrokamferol en la posición 3'5' siendo responsable de la pigmentación azul/morado en los pétalos de las flores tales como *Gentiana (Gentiana scabra)* (Nakatsuka et al., 2008), *Verbena (Verbena hybrida)*, (Togami et al., 2006), *Nierembergia (Nierembergia sp.)*, (Ueyama et al, 2006) entre muchas otras. Varios de los genes que codifican para las enzimas de la biosíntesis de antocianinas ya han sido aislados y caracterizados de varias especies como maíz (*Zea mays*), petunia (*Petunia hybrida*) y dragonaria (*Antirrhinum majus*) (Holton y Cornish, 1995; Brugliera et al., 1999), pero en especies con pétalos azules en los cuales el color se debe a la presencia y acumulación de delphinidina (Okinaka et al., 2003) no se ha trabajado aún, por lo que aislar y caracterizar los genes que codifican las enzimas clave involucradas en la síntesis del pigmento de interés será la base para estudios posteriores que permitirán elucidar a detalle la vía de síntesis de este pigmento.

La síntesis de pigmentos de las antocianinas en el tejido floral proporciona un excelente marcador para estudiar la diferenciación de las células. La mayor parte de los genes estructurales que codifican las 15 enzimas requeridas para la síntesis de antocianinas y la modificación han sido aislados (Holton y Cornish, 1995; Winkel-Shirley, 2001). Por análisis mutacional en *Antirrhinum*, *Arabidopsis*, maíz y petunia, varios genes reguladores han sido identificados que se requieren para la transcripción específica del tejido de los genes estructurales de antocianinas (Mol et al, 1998; Winkel-Shirley, 2001).

## 4.2 ESTRATEGIA EXPERIMENTAL

A continuación se muestra el esquema general del diseño experimental:



## 4.3. MATERIALES Y MÉTODOS

### 4.3.1 Extracción de ARN total

Para la extracción de ARN total es necesario en primer lugar considerar los diferentes tipos de tejidos y el momento del desarrollo floral para su extracción. Para el ARN de *F3'5'H* se tomaron tejidos de hoja, y diferentes etapas de desarrollo floral para la identificación de donde se expresa el gen.

### 4.3.2 Diseño de oligonucleótidos

Para este fin, se hizo primeramente la búsqueda de las secuencias teniendo en la base de datos del ncbi o Genbank (<http://www.ncbi.nlm.nih.gov/sites/entrez>) del gen *F3'5'H*, luego se realizó un BLAST con el fin de encontrar las especies más cercanas. Una vez seleccionadas se alinearon con el programa clustawl W, se identificaron las regiones conservadas, y se diseñaron oligonucleótidos consenso a partir de estas regiones conservadas. Luego se consideraron las características como son el tamaño del oligonucleótido, la temperatura de fusión ( $T_m$ ), la Especificidad de Secuencias complementarias, el contenido en G/C y trectos de polipirimidinas (T, C) o polipurinas (A, G), secuencia 3' terminal, secuencia 5' terminal y regiones centrales (O'Connell, 2002).



## 4.4. RESULTADOS

### 4.4.1. EXTRACCIÓN DE ARN total

Según los reportes de las diferentes especies en donde se ha aislado el gen *F3'5'H*, particularmente un estudio en *Phaleanopsis* reporto la expresión del gen en las etapas de desarrollo floral de botón con sépalos con el color pálido y la apertura de la flor, habiendo mayor expresión en la de botón con sépalos (Wang et al, 2006).

### 4.4.2. DISEÑO DE OLIGONUCLEÓTIDOS

Se diseñaron oligonucleótidos a partir del ARNm de *Phaleanopsis hybrida* para lílidas.

Los oligos obtenidos para el gen *F3'5'H* se muestran en el siguiente cuadro:

OLIGONUCLEÓTIDO	TAMAÑO	Tm (°C)	DIRECCIÓN	REGIÓN (pb)
TCGATCCGGAGAGGTTTATG	20	60	FORWARD	1436
AGTCAAACCCTGCACCAAC	20	60	REVERSE	1600
GCGAATGTGATTGGGAGAAT	20	59	FORWARD	742
TGCTCCGCCAAGAACTTACT	20	60	REVERSE	971

**CUADRO 4.1. OLIGONUCLEÓTIDOS PARA LÍLIDAS.**

## 4.5 REFERENCIAS BIBLIOGRÁFICAS

Holton T. y Cornish Edwina (1995) *Genetics and Biochemistry of Anthocyanin Biosynthesis*. The Plant Cell, Vol. pp. 1071-1083.

Mol J., Cornish E., Mason J. y and Koes R.1999.*Novel coloured flowers*. Current Opinion in Biotechnology. Vol 10. pp198–201. Elsevier Science.

Okinaka Y., Shimada Y., Nakano-Shimada R., Ohbayashi M., Kiyokawa S y Kikuchi Y. (2003). *Selective Accumulation of Delphinidin derivatives in Tobacco Using a Putative*

*Flavonoid-3',5'-Hydroxylase cDNA from Campanula medium.* Biosci. Biotechnol. Biochem., 67 (1), 161-165.

Togami J., Tamura M., Ishiguro K., Hirose C., Okuhara H., Ueyama Y., Nakamura N., Yonekura-Sakakibara K., Fukuchi-Mizutani M., Suzuki K., Fukui Y., Kusumi T. y Tanaka Y. (2006) *Molecular characterization of the flavonoid biosynthesis of Verbena hybrida and the functional analysis of verbena and Clitoria ternatea F3'5'H genes in transgenic verbena.* Plant Biotechnology 23, 5–11.

Nakatsuka T., Sato K., Takahashi H., Yamamura y Nishihara M. (2008) *Cloning and characterization of the UDPglucose: anthocyanin 5-O-glucosyltransferase gene from blue-flowered gentian.* Journal of Experimental Botany, Vol. 59, No. 6, pp. 1241–1252.

Ueyama N., Katsumoto Y., Fukui Y., Fukuchi-Mizutani M., Ohkawa H., Kusumi T., Iwashita T., Tanaka Y. (2006) *Plant Molecular characterization of the flavonoid biosynthetic pathway and flower color modification of Nierembergia sp.* Biotechnology 23, 19–24.

Brugliera F, Barri-Rewell G, Holton TA, Mason JG (1999) *Isolation and characterization of a Flavonoid 3'5'-Hydroxylase cDNA clone corresponding to the Ht1 locus of Petunia hybrida.* Plant J 19: 506–602.

Wang J., Ming F., Han Y. y Shen D. (2006) *Flavonoid-3',5'-hydroxylase from Phalaenopsis: a novel member of cytochrome P450s, its cDNA cloning, endogenous expression and molecular modeling.* Biotechnology Vol.28 pp. 327–334.

Winkel-Shirley, B. (2001) *Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology.* Plant Physiology, Vol. 126, pp. 485–493.

## **CAPITULO V: DISCUSIÓN INTEGRAL**

A lo largo de este trabajo pudimos observar que la floricultura es un área económica fuerte que sin dudar no deja de tener impacto en el mercado, ya que se ha demostrado que a más novedad más demanda, y un ejemplo es la rosa azul, que claramente la gente está dispuesta a pagar. Por otro lado, es obvio que el color final de la flor es determinado por diferentes componentes y factores, siendo extremadamente complicado que los mejoradores clásicos definan un diseño para la estandarización de un color específico, entre las razones principales, debido a la limitación del acervo genético y la forma de controlar los cambios de color sin la influencia de otras características fenotípicas importantes y de producción. En años recientes los avances en biología molecular han venido a superar las barreras genéticas de las especies y con esto encontrar una alternativa eficaz para el cambio de color en flores. Como estrategias para este objetivo resaltan dos, la primera es la introducción de genes que no se encuentran presentes naturalmente en la planta y la segunda, la inhibición de los pasos específicos o de una parte de la ruta biosintética.

Los flavonoides están entre los metabolitos secundarios vegetales mejor caracterizados en términos de química, bioquímica, mecanismo de la coloración, genética y biología molecular (Grotewold, 2006; Harbone, 1998, 1994; Stafford, 1990). Este conocimiento se ha usado para mejorar las propiedades de flores, especialmente colores de la flor, con la aplicación de la biotecnología moderna de la floricultura. La generación de bibliotecas de pequeñas moléculas, información de la secuencia de la especie, bibliotecas específicas EST y consorcios destinados a gestionar y compartir esos recursos irán una manera larga en la obtener una comprensión completa de la biosíntesis de flavonoides en flor. La capacidad de considerar ahora flavonoides, enzimas para la primera vez, en tres dimensiones y para examinar la interdependencia de las vías de la secundaria metabolismo utilizando genómica, proteómica y metabólica métodos de perfiles es probable que nos mueve mucho más rápidamente hacia ese fin. (Winkel- Shirley, 2001).

La bioinformática, disciplina científica emergente que utiliza tecnología de la información para organizar, analizar y distribuir información biológica con la finalidad de responder preguntas complejas como esta, ha demostrado ser una vez más, una herramienta de gran utilidad para dislucidar aún más la compleja ruta de delphinidina, responsable del famoso color azul en las flores. El estudio de la ciencia de la proteína provee el conocimiento principal de las proteínas en cuestión de modificación genética, así también, de corrientes modernas de alta velocidad de datos como la genómica y la proteómica (Claverie y Notredame, 2007).

La estructura de la proteína está íntimamente ligada a su función y por lo tanto, predecir y estudiar la estructura básica de la proteína y la función con bioinformática, ha allanado el camino para diseñar nuevas proteínas con carácter específico deseado. La determinación de las funciones de una proteína, así como los mecanismos desencadenados para llevar a cabo tales funciones, implica años de experimentación así como una fuerte inversión económica, es por eso, que el empleo de programas de predicción y de comparación para el análisis de estructuras primarias, secundarias y terciarias de las proteínas, brinda la posibilidad de conocer sus funciones más probables al compararlas con otras biomoléculas proteicas homólogas cuyas funciones han sido descritas con gran detalle, lo que permite orientar con gran precisión los protocolos de experimentación antes y después de su realización para comprender con mayor visión los mecanismos por medio de los cuales una biomolécula puede realizar sus actividades ya sean catalíticas, de transporte, mensajeras, de reconocimiento del medio, etc. (Morales et al, 2008; Wild y Saqui, 2004).

La estructura primaria, secundaria y tridimensional que presentan las proteínas, ha sufrido evolución a lo largo de la vida con el objetivo de adquirir la conformación tridimensional más adecuada con la cual la proteína pueda realizar funciones especializadas y así, brindarles a las células a las cuales pertenecen una mejor oportunidad de sobrevivir, adaptarse, desarrollarse, multiplicarse y preservar su especie. Dando más razón a este tipo de estudio realizado, al esclarecer la diversidad de funciones que pueden darse con la proteína F3'5'H, proporcionando un sin numero de datos que se ha reportado con respecto

a esta proteína y mostrando la variabilidad de estructuras que puede haber, así como las regiones conservadas como lo son con el citocromo P450, presentándose siempre en las especies pertenecientes a la proteína F3'5'H. (Fester 2009; McGuffin et al, 2000; Tanaka, 2006; Vetten et al, 1999).

Para el perfil hidrofóbico según el modelo de kyte y Doolittle, señala que es poco soluble, confirmando el hecho que se reporta para antocianinas que se consideran poco solubles, y su relación con la copigmentación, mecanismo de interacción molecular que previene el ataque nucleofílico del agua previniendo la decoloración del pigmento.

La modificación genética de plantas ornamentales, especialmente de flores, se ha enfocado en la búsqueda de nuevos colores como lo son el azul, en especial en especies comerciales como la rosa, el crisantemo y los claveles que no poseen en su genética este color. Largo ha sido el camino de la rosa azul sin resultados contundentes, sin embargo, es un punto importante, que se ha desarrollado, para investigar varios temas que conciernen a este tema: el color azul de las flores. La biología molecular es una rama que se ha enfocado en estos estudios, lo cuales han crecido de forma exponencial, por lo que desarrollar técnicas que ayuden a esclarecer, recopilar y procesar esta información son de suma importancia, complementar este trabajo con el asilamiento del gen en una especie nativa le da más fuerza al trabajo (Tanaka, 2006; Vetten et al, 1999).

En el futuro, la generación de bibliotecas de pequeñas moléculas, información de la secuencia de flores específicas y consorcios destinados a gestionar y compartir recursos tales haga unir esfuerzos y lograr los objetivos deseados, compartiendo diferentes investigaciones y aclarando más el océano que es la biología de las plantas (Claverie y Notredame, 2007; Tanaka, 2006; Vetten et al, 1999).

Así podemos decir que la proteína F3'5H de petunia, ha demostrado con los resultados obtenidos, gran diversidad de funciones, especies y una estructura conservada y compleja. Con esto, se ayuda a esclarecer más, el entendimiento de como las flores toman el color azul y como con los diferentes tejidos o condiciones puede variar. Pero más

importante aún es el conocer la proteína, ya que teniendo su estructura podemos conocer su función, por lo que se observa que la familia a la que pertenecen las especies es un factor importante, pudiendo las estructuras ser comunes entre las especies o tener alguna característica en particular. Hacer la comparación de estas estructuras y funciones nos deja más claro el camino por recorrer, y poder así resolver el enigma de la rosa azul. Estudios como estos son parte de un marco científico a futuro que en materia de modificación genética abriría las puertas a dudas por resolver y a la obtención de productos innovadores.

## 5.1 REFERENCIAS BIBLIOGRÁFICAS:

Claverie, J. y Notredame, C. (2007) *Bioinformatics for Dummies*. 2ª edición, Wiley Publishing Inc.

Fester R. (2009) *Molecular and Cell Biology for Dummies*. Cap.6 y 7. pp 85-108  
Wiley Publishing Inc.

McGuffin L., Bryson K. y Jones D. (2000). *The PSIPRED protein structure prediction server*. *Bioinformatics*. Vol. 16 404:-405.

Morales L., Mena A., Castañeda E., Serkin V. y Peña R. (2007) *Estudio bioinformático de la molécula proteica UgpB de bacterias del género Brucella* Vol. 5, N° 1: 921-940.

Tanaka T. (2006) Flower color and cytochromes P450. *Phytochem* Vol. 5:283–291

Vetten N., Horst J., Van Schaik H., Boer A., Mol J. y Koes R. (1999) Cytochrome *b5* is required for full activity of flavonoid- 3',5'-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors. Vol. 96: 778–783. *Plant Biology*.

Wild D. y Saqui M. (2004) *Structural Proteomics: Inferring Function from Structure*. *Current Proteomics*, Vol. 1: 59-65.

Winkel-Shirley, B. (2001) *Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Physiology*, Vol. 126, pp. 485–493.

## **CAPITULO VI: CONCLUSIONES**

- Se reportaron 771 secuencias para el gen F3'5'H en la base de datos del NCBI, siendo la petunia quien presenta el mayor número de secuencias de diferentes especies, locus y clones, seleccionando a la primera secuencia del gen F3'5'H completa, reportada para petunia, con un tamaño de 1773 pb y número de accesoión EF371021.
- Para las proteínas P450 se reportan tamaños de 500 aa, eligiendo el primer ORF, tomando la primera metionina y el primer codón stop, con un tamaño final de 506 aa. El segundo y tercero ORF's se localizan los codones de stop muy cerca del inicio de la traducción, sugiriendo no ser la proteínas que se buscan.
- En la secuencia de proteína de la F3'5'H de *Petunia hybrida* nombrada como F35H0179PH, se pueden identificar los tres dominios del citocromo P450, incluyendo los motivos conservados y representativos de la familia.
- Se seleccionaron cuatro bases de datos para la búsqueda de las proteínas F3'5'H: PIR, NCBI, PDB y UniProt, reportándose 296 secuencias de proteínas para PIR; 213 para NCBI y 73 de UniProt-PDB. Seleccionando 27, 3 y 9 respectivamente, resultando un total de 39 secuencias de proteína F3'5'H en plantas.
- Se realizó un multialineamiento para la identificación de los dominios conservados y el análisis de identidad de las secuencias ortólogas a la proteína F3'5'H de *Petunia hybrida*, con 40 secuencias en total, mostrándose conservadas todas las proteínas en los dominios de bisagra y globular.
- El árbol de función mostró el agrupamiento de algunas especies por familias sugiriendo una estructura similar del F3'5'H, mostrando también una gran diversidad de tejidos.



- El árbol de relación de función mostró que todas las especies de la proteína F3'5'H, pertenecen al CYP75A, refiriéndose al citocromo P450 de la familia 75 que interviene en la ruta fenilpropanoide, específicamente en la catalización de compuestos flavonoides y la subfamilia A que cataliza el F3'5'H.
- El perfil hidrofóbico Kyte and Doolittle, mostró residuos de aa hidrofóbicos en la región N-terminal con una membrana, sugiriendo al dominio de ancla del P450; el motivo PPGP ubicado cerca de la región N-terminal y cerca del C-terminal el motivo de unión al hierro (PFGARRRICAG), altamente conservados en secuencias de proteínas F3'5'H.
- La predicción de la estructura secundaria reveló estructuras «coils», 9 alfa hélices largas y 7 pequeñas; 4 beta plegadas largas y 5 pequeñas. Los dominios y motivos conservados fueron ubicados en las regiones reportadas para la familia.
- La estructura terciaria es clasificada en un dominio A que contiene la mayoría de las alfa-hélices con dos beta-plegadas y coils a lo largo del dominio y un dominio B con 5 beta-plegadas y dos alfa hélices.
- Para el ARN total se recomienda realizar la extracción en las etapas 3 y 4 de desarrollo floral, debido a que en esta fases de desarrollo es cuando se da la expresión del gen F3'5'H.
- Se diseñaron dos pares de oligonucleótidos a partir de Petunia, Campanula, Phaleanopsis y Nierembergia, con una temperatura media de 60 grados centigrados y un tamaño de 20 nucleótidos para la amplificación del gen F3'5'H.

## **ANEXO 1**

## Introduction

Flavonoid-3',5'-hydroxylase (F3'5'H) is enzyme involved in delphinidin biosynthesis, chemical compound related with blue color in flowers, which had received importance seeking to modify the coloration of the petals of the flowers, to increase its commercial value (1). Nevertheless the experiments of synthetic biology to manipulate this enzyme have not obtained successful results. In our workgroup we realize *in silico* analysis focused to conserved domain, structure, function and biochemical characteristics, to understand mechanism of activity regulation. Also *in silico* analysis facilities not only the cloning of this gene in diverse organisms, but also would allow obtaining of real blue flowers. *In silico* analysis is a tool that performed using computers in conjunction with informatics software, to study evolution, structure, function and regulation of nucleic acids and proteins. Our research began with the sequence of nucleotides of F3'5'H of petunia (EF371021), the open reading frames (ORF) analyzed shows a 506 aa protein. We did identity analysis comparing with 78 protein sequences from 55 species of plants, reported by NCBI(2), PIR(3), Uniprot(4), PDB(5) and RefSeq(6) databases, with the aim to identify the protein conserved domains. The hydrophobic profile helps us to predict substrate.

## Methodology

Escribe un poco sobre la obtencion de la secuencia y el alineamiento

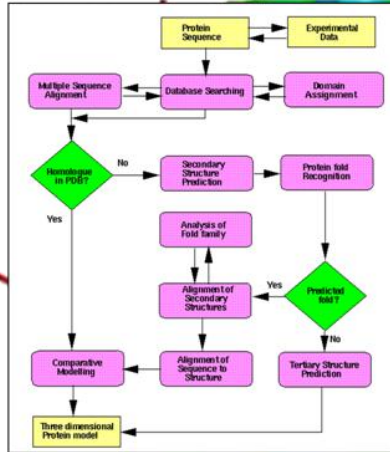


Figure 1. Diagram of

## Results

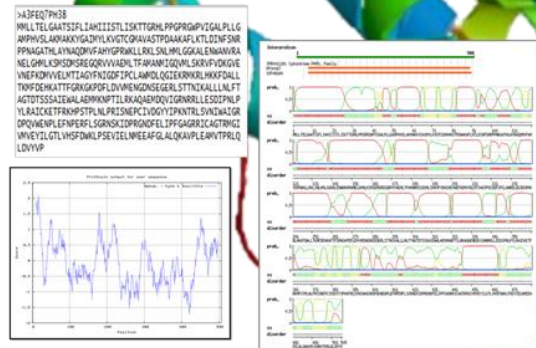


Figure 3

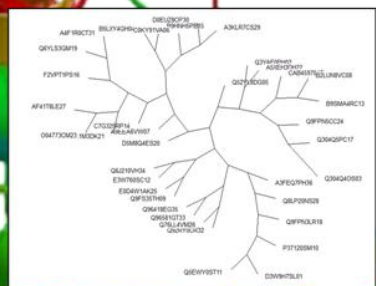


Figure 4

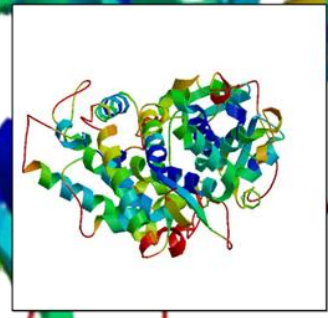


Figure 3

## Bibliography

1. ...
2. ...
3. ...
4. ...
5. ...
6. ...

