

**FACULTAD DE
CIENCIAS EXACTAS**
DEPARTAMENTO DE CIENCIAS
BIOLÓGICAS



**UNIVERSIDAD
NACIONAL
DE LA PLATA**

PÉPTIDOS BIOACTIVOS DE GLOBULINAS DE AMARANTO. UNA APROXIMACIÓN COMPUTACIONAL Y EXPERIMENTAL

Tesis Doctoral | Lic. Agustina E. Nardo

Director Dr. Gustavo D. Parisi – Codirector Dra. M. Cristina Añón

2017



UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE CIENCIAS EXACTAS
DEPARTAMENTO DE CIENCIAS BIOLÓGICAS

Trabajo de Tesis Doctoral

**PÉPTIDOS BIOACTIVOS DE
GLOBULINAS DE AMARANTO. UNA
APROXIMACIÓN COMPUTACIONAL Y
EXPERIMENTAL**

Tesista: Lic. Agustina E. Nardo

Director: Dr. Gustavo D. Parisi

Codirectora: Dra. M. Cristina Añón

2017

El presente trabajo de tesis ha sido desarrollado para acceder al grado académico de Doctor de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata. El mismo se realizó en el Centro de Investigación y Desarrollo en Criotecnología de los Alimentos (CIDCA CONICET-CIC-UNLP) y el Departamento de Ciencia y Tecnología de la Universidad Nacional de Quilmes, con la dirección del Dr. Gustavo D. Parisi y la codirección de la Dra. M. Cristina Añón.

Su realización fue posible gracias a las becas otorgadas por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

Dedicado a Pedro Jacinto y Sofía Catalina

AGRADECIMIENTOS

A la Universidad Nacional de La Plata que me permitió formarme.

A la Facultad de Ciencias Exactas, por ser el lugar donde encontré mi vocación.

A CONICET por las becas otorgadas, que han hecho posible la realización de este trabajo.

A mis directores, Gustavo y Cristina, por guiarme durante todo el trabajo de investigación y el proceso de escritura, por sus importantes aportes, comentarios y sugerencias. Sin su colaboración no hubiera sido posible la realización y escritura de este trabajo.

A todos mis compañeros del laboratorio y de cada oficina que compartí.

A mi familia, quienes me acompañaron desde el primer día y me alentaron a continuar, aun sin entender de qué se trata realizar un doctorado. Gracias por estar siempre presentes y por acobijar a mis hijos durante el proceso de escritura. Sin ustedes la presentación de este trabajo no hubiera sido posible.

A mis papás, por todo. Por educarme y jamás condicionarme en la elección de mi carrera profesional.

A Agustín, por todo, por acompañarme estos años, por el día a día, por la paciencia, y mucho más.

Finalmente, a mis hijos, porque en el deseo de educar con el ejemplo, me impulsaron a finalizar.

ÍNDICE

<i>ABREVIATURAS</i>	1
<i>CÓDIGO DE AMINOÁCIDOS</i>	2
INTRODUCCIÓN GENERAL	3
<hr/>	
I. PEPTIDOS BIOACTIVOS DE PROTEÍNAS ALIMENTARIAS	4
I.1 ALIMENTOS FUNCIONALES	4
I.2 PÉPTIDOS BIOACTIVOS	5
I.3 ESTRATEGIAS PARA EL ESTUDIO DE PÉPTIDOS BIOACTIVOS	7
I.4 HERRAMIENTAS BIOINFORMATICAS PARA EL ESTUDIO DE PÉPTIDOS BIOACTIVOS	9
II. ENFERMEDADES CARDIOVASCULARES Y COLESTEROL	14
II.1 ENFERMEDADES CARDIOVASCULARES	14
II.2 COLESTEROL	16
III. AMARANTO	18
III.1 GENERALIDADES DEL AMARANTO	18
III.2 PROTEÍNAS DE RESERVA DE LAS SEMILLAS DE AMARANTO	20
<i>ALBÚMINAS</i>	20
<i>GLOBULINAS</i>	21
<i>PROLAMINAS</i>	23
<i>GLUTELINAS</i>	23
III.3 PÉPTIDOS BIOACTIVOS IDENTIFICADOS EN AMARANTO	23
<i>ACTIVIDAD ANTIHIPERTENSIVA</i>	24
<i>ACTIVIDAD ANTIPROLIFERATIVA</i>	26
<i>ACTIVIDAD ANTIOXIDANTE</i>	28
<i>ACTIVIDAD INMUNOMODULADORA</i>	30
<i>ACTIVIDAD ANTITROMBÓTICA</i>	31
OBJETIVOS	33
<hr/>	
OBJETIVOS GENERALES	34
OBJETIVOS ESPECÍFICOS	34
CAPÍTULO 1 ESTUDIO DE LA OCURRENCIA DE PÉPTIDOS BIOACTIVOS UTILIZANDO MÉTODOS BIOINFORMÁTICOS EN PROTEÍNAS DE SECUENCIA CONOCIDA	35
<hr/>	
1.1 INTRODUCCIÓN	36
<i>NATURALEZA DEL PROBLEMA</i>	36
1.2 OBJETIVOS ESPECÍFICOS	39
1.3 MATERIALES Y MÉTODOS	40
1.3.1 ESQUEMA GENERAL	40
1.3.2 BREVE RESUMEN DE NIVELES ESTRUCTURALES EN PROTEÍNAS	42
1.3.3 PÉPTIDOS BIOACTIVOS Y BASES DE DATOS SECUENCIALES	44

1.3.4 BÚSQUEDA SECUENCIAL DE PB EN LA BASE DE DATOS <i>PROTEIN NOT-REDUNDAT (NR)</i>	44
1.3.5 ASIGNACIÓN DE DOMINIOS CATH (ASIGNACIÓN DE PLEGAMIENTO)	45
1.3.6 LOCALIZACIÓN ESTRUCTURAL DE PB _s	47
1.3.7 OTRAS HERRAMIENTAS	48
1.4 RESULTADOS Y DISCUSIÓN	49
1.4.1 DISTRIBUCIÓN DE PÉPTIDOS EN BIOPEP	49
1.4.2 DISTRIBUCIÓN DE PÉPTIDOS EN EL ESPACIO SECUENCIAL DE PROTEÍNAS	53
1.4.3 ASIGNACIÓN ESTRUCTURAL Y DISTRIBUCIÓN DE SECUENCIAS	53
1.4.4 DISTRIBUCIÓN DE PB _s EN LAS SUPERFAMILIAS ESTRUCTURALES	60
1.4.5 LOCALIZACIÓN ESTRUCTURAL DE PB	62
1.5 CONCLUSIONES CAPÍTULO 1	69
CAPÍTULO 2 INHIBIDORES PEPTÍDICOS DE HMGR DE AMARANTO	71
<hr/>	
2.1 INTRODUCCIÓN	72
2.1.1 LA VÍA DEL MEVALONATO	72
2.1.2 HMGR	74
2.1.3 REGULACIÓN HMGR	78
2.1.4 INHIBIDORES DE LA HMGR: ESTATINAS	79
2.2 OBJETIVOS ESPECÍFICOS	81
2.3 MATERIALES Y MÉTODOS	82
2.3.1 OBTENCIÓN DE LA HARINA DE AMARANTO	82
2.3.2 PREPARACIÓN DE LOS AISLADOS PROTEICOS	82
2.3.3 PREPARACIÓN DE LOS HIDROLIZADOS PROTEICOS CON ALCALASA	83
2.3.4 DIGESTIÓN GASTROINTESTINAL SIMULADA	84
2.3.5 CARACTERIZACIÓN DE LAS MUESTRAS	85
2.3.5.1 DETERMINACIÓN DE PROTEÍNAS	85
<i>MÉTODO DE KJELDAHL</i>	85
<i>MÉTODO DE MICROKJELDAHL</i>	86
<i>DETERMINACIÓN DE PROTEÍNA SOLUBLE</i>	87
2.3.5.2 DETERMINACIÓN DE CENIZAS	88
2.3.5.3 DETERMINACIÓN DEL CONTENIDO DE HUMEDAD	88
2.3.5.4 DETERMINACIÓN DEL CONTENIDO DE HIDRATOS DE CARBONO	88
2.3.5.5 DETERMINACIÓN DEL GRADO DE HIDRÓLISIS (GH%) POR OPA	89
2.3.5.6 ELECTROFORESIS	90
2.3.6 ANÁLISIS POR CROMATOGRAFÍA LÍQUIDA RÁPIDA PARA PROTEÍNAS (FPLC)	92
2.3.6.1 CROMATOGRAFÍA FPLC DE EXCLUSIÓN MOLECULAR	92

2.3.6.2 CROMATOGRAFÍA FPLC DE INTERACCIÓN HIDROFÓBICA	92
2.3.7 ENSAYO DE INHIBICIÓN DE HMGR IN VITRO	94
2.3.8 ANÁLISIS ESTADÍSTICO DE LOS RESULTADOS	95
2.4 RESULTADOS Y DISCUSIÓN	97
2.4.1 OBTENCIÓN Y CARACTERIZACIÓN DEL AISLADO PROTEICO DE AMARANTO Y SUS HIDROLIZADOS	98
2.4.2 DETERMINACIÓN DE LA INHIBICIÓN DE HMGR	104
2.4.3 FRACCIONAMIENTO DEL AISLADO SOMETIDO A DIGESTIÓN GASTROINTESTINAL SIMULADA	108
2.4.4 BIBLIOTECAS DE INHIBIDORES DE HMGR	110
2.4.5 IDENTIFICACIÓN DE PÉPTIDOS INHIBIDORES DE HMGR EN SECUENCIAS DE AMARANTO	113
2.5 CONCLUSIONES CAPÍTULO 2	116
CAPÍTULO 3 APLICACIÓN DE TÉCNICAS BIOINFORMÁTICAS AL ESTUDIO DE OTRAS ACTIVIDADES BIOLÓGICAS DE LA FRACCIÓN PROTEÍCA DE AMARANTO	119
<hr/>	
3.1 INTRODUCCIÓN	120
3.1.1 ACTIVIDAD ANTITROMBÓTICA	121
3.1.2 ACTIVIDAD ANTIOXIDANTE	122
3.2 APLICACIÓN DE TÉCNICAS BIOINFORMÁTICAS AL ESTUDIO DE LA ACTIVIDAD ANTITROMBÁTICA	125
3.2.1 OBTENCIÓN DE LA FRACCIÓN ACTIVA	125
3.2.2 IDENTIFICACIÓN DE PÉPTIDOS PRESENTES EN LA FRACCIÓN ACTIVA Y EN LA CÁMARA BASOLATERAL	127
3.3 APLICACIÓN DE TÉCNICAS COMPUTACIONALES AL ESTUDIO DE ACTIVIDADES ANTIOXIDANTE	136
3.3.1 OBTENCIÓN DE PÉPTIDOS ANTIOXIDANTES DE AMARANTHUS MANTEGASIANUS	136
3.3.2 RELACIÓN ENTRE LA ESTRUCTURA DE PÉPTIDOS Y ACTIVIDAD ANTIOXIDANTE	138
3.3.3 SELECCIÓN DE PÉPTIDOS POTENCIALMENTE ANTIOXIDANTES PARA SU SÍNTESIS Y EVALUACIÓN DE SU ACTIVIDAD	139
3.4 CUPINAS Y LECTINAS	144
3.5 CONCLUSIONES CAPÍTULO 3	147
CONCLUSIONES GENERALES	149
<hr/>	
BIBLIOGRAFÍA	153
<hr/>	
ANEXO	171
<hr/>	

Abreviaturas

ABTS	Ácido 2,2'-azino-bis (3-etilbenzotiazolina-6-sulfónico) Método de determinación de actividad antioxidante
ACE	Enzima convertidora de angiotensina
Ais	Aislado proteico de amaranto
BP	Péptidos bioactivos
COX-1	Ciclooxigena isoforma I
Dig	Aislado proteico de amaranto sometido a digestión gastrointestinal simulada
DPP4	Dipeptidil peptidasa IV
ECV	Enfermedades cardiovasculares
FRAP	Ferric ion reducing antioxidant power - Método de determinación de actividad antioxidante
Hid	Aislado proteico de amaranto hidrolizado con alcalasa
HMGR	3-hidroxi-3-metil glutaril CoA reductasa
IC₅₀	Concentración de una sustancia activa que produce un 50% de inhibición de un proceso biológico dado (o componente de un proceso, es decir, una enzima o receptor celular, etc.)
IgE	Inmunoglobulina E
LOX-1	Receptor 1 de lipoproteína de baja densidad (LDL) oxidado tipo lectina
MC3T3E1, UMR106, Caco-2, y TC7	Líneas celulares de diferente origen y tumorigenicidad
NF-κB	Vía clave en la regulación de la respuesta inmune
ORAC	Capacidad de absorción de radicales de oxígeno
ROS	Especies reactivas del oxígeno
RSN	Especies reactivas al nitrógeno
Sistema RAS	Sistema renina-angiotensina
TBARS	Sustancias reactivas al ácido tiobarbiturico
TGF-β	Factor de crecimiento transformante β
TNF-α	Factor de necrosis tumoral α

Aminoácido	Código de tres letras	Código de una letra
Alanina	Ala	A
Arginina	Arg	R
Asparagina	Asn	N
Ácido aspártico	Asp	D
Cisteína	Cys	C
Glutamina	Gln	Q
Ácido glutámico	Glu	E
Glicina	Gly	G
Histidina	His	H
Isoleucina	Ile	I
Leucina	Leu	L
Lisina	Lys	K
Metionina	Met	M
Fenilalanina	Phe	F
Prolina	Pro	P
Serina	Ser	S
Treonina	Thr	T
Triptófano	Trp	W
Tirosina	Tyr	Y
Valina	Val	V

INTRODUCCIÓN GENERAL

I. PEPTIDOS BIOACTIVOS DE PROTEÍNAS ALIMENTARIAS

I.1 Alimentos funcionales

En los últimos años uno de los temas más importantes de investigación en Ciencia y Tecnología de Alimentos es el desarrollo de alimentos que promuevan la salud. El concepto de estos alimentos, conocidos como alimentos funcionales fue desarrollado por primera vez en Japón en la década de 1980. Internacionalmente, se definen como alimentos capaces de aportar sustancias con funciones fisiológicas definidas, brindando beneficios para la salud de quien los consume (Doyon y Labrecque, 2008; Martirosyan y Singh, 2015). Si bien no existe una definición única de alimentos funcionales la mayoría de ellas coinciden en cuatro aspectos básicos:

- *Beneficios para la salud:* Este concepto es el eje central en la mayoría de las definiciones, aunque no en todos los casos se indica explícitamente que el mismo debe ser probado experimentalmente. En general los beneficios para la salud se refieren a la reducción del riesgo de padecer una determinada enfermedad o al aumento de una dada función.
- *Naturaleza del alimento:* Este también es un concepto relevante en la mayoría de las definiciones y en general se acuerda que un alimento funcional debería ser un alimento tradicional o ser similar a ellos. En algunas definiciones se indica que los alimentos funcionales deben ser enriquecidos, fortificados o contener un ingrediente adicionado mientras que en otras se indica que deben estar libres de alérgenos o de componentes que puedan afectar la salud del consumidor.

- *Nivel de función:* En este caso existe acuerdo que los alimentos funcionales deben tener un beneficio más allá del meramente nutricional
- *Patrón de consumo:* Este es un concepto que se pone de manifiesto en algunas definiciones y se refiere al hecho que un alimento funcional debe formar parte de una dieta normal o patrón de consumo de una región geográfica y/o contexto cultural. Esto significa que lo que es considerado un alimento funcional en un dado país no necesariamente lo es en otro.

Los alimentos son sistemas complejos formados por una gran variedad de sustancias que interaccionan entre sí. Si bien es difícil identificar a los componentes responsables de un determinado efecto biológico, si se desea desarrollar alimentos funcionales resulta indispensable lograrlo. Estos componentes se conocen como bioactivos. Para que un compuesto sea considerada bioactivo debe ser capaz de generar un efecto fisiológico medible y afectar de manera positiva la salud del ser humano, excluyendo posibles efectos negativos (como toxicidad, alergenicidad y mutagenicidad) (Schrezenmeir y col., 2000). En los últimos años se ha puesto un gran énfasis en la identificación y purificación de dichos componentes, ya no sólo para catalogar a un alimento como funcional si no para poder usarlos como ingredientes en nuevos alimentos, en suplementos dietarios y/o en la industria farmacéutica.

I.2 Péptidos Bioactivos

En las últimas dos décadas varios investigadores han detectado la generación de péptidos con acción fisiológica sobre los sistemas cardiovascular, nervioso, gastrointestinal e inmune a partir de proteínas alimentarias (Erdmann y col., 2008; Möller y col., 2008). Desde los primeros reportes de péptidos con actividad opioide –comportamiento similar a la morfina *in vitro*– derivados de proteínas lácteas (Zioudrou y col., 1979) la obtención, caracterización e identificación de péptidos bioactivos se ha convertido en una de las temáticas más importantes en ciencia y tecnología de los alimentos (**FIGURA 1**).

Los péptidos bioactivos (PB) han sido aislados y caracterizados a partir de distintas fuentes alimentarias, tales como proteínas de huevo, pescado y cereales (arroz, trigo, cebada, maíz, entre otros), pero hasta el momento los más estudiados son los provenientes de la leche y otros productos lácteos (Choi y

col., 2012; Floris y col., 2003; Korhonen y Pihlanto, 2006; Madureira y col., 2007; Nagpal y col., 2011).

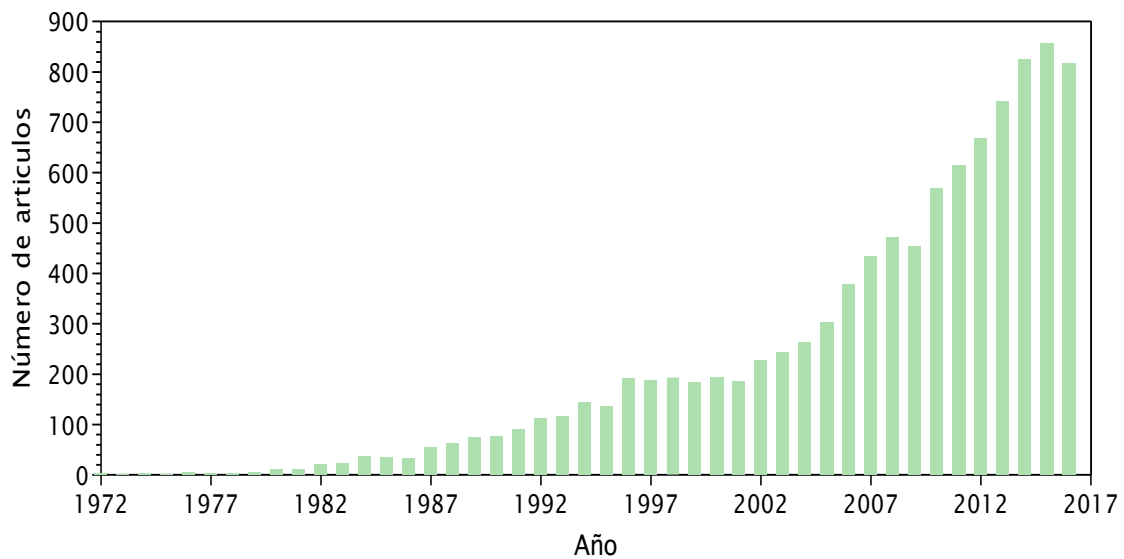


FIGURA 1: Evolución del número de publicaciones indexadas hasta fines de 2016 en Scopus que se obtienen introduciendo “*Biactive Peptide*” como palabra clave para la búsqueda.

Los PBs pueden encontrarse como entidades independientes o encriptados en la secuencia de proteínas madres. Usualmente consisten en secuencias de 2 a 20 aminoácidos que son liberados de la proteína original luego de un proceso de proteólisis que puede ocurrir *in vivo* durante la digestión por la acción de las enzimas digestivas, como la pepsina y la tripsina entre otras y/o por la acción de enzimas microbianas de la flora intestinal e *in vitro* durante el procesamiento del alimento o de las diferentes materias primas que pueden formar parte de la formulación de los alimentos (Mulero Cánovas y col., 2011; Möller y col., 2008; Shahidi y Zhong, 2008; Wang y González De Mejía, 2005).

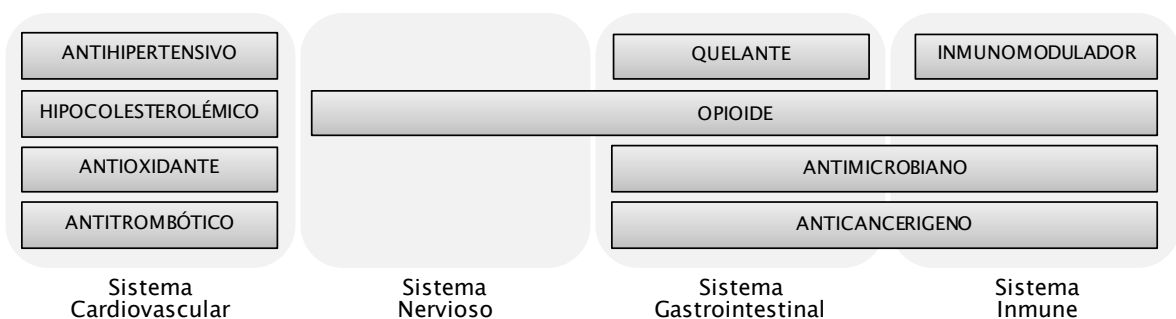


FIGURA 2: Efectos de los péptidos bioactivos en los principales sistemas del organismo.

Se han descrito una gran variedad de actividades biológicas atribuidas a péptidos provenientes de proteínas alimentarias. Entre ellas podemos mencionar la actividad antihipertensiva, antitrombótica, anticolesterolémica, antioxidante, antimicrobiana, anticancerígena y actividad opioide. Las mismas se pueden agrupar de acuerdo al sistema fisiológico sobre el cual ejercen su efecto, tal como se muestra en la **FIGURA 2**.

I.3 Estrategias para el estudio de péptidos bioactivos

Un estudio integral para determinar las potenciales actividades fisiológicas de un determinado compuesto bioactivo se puede dividir en tres etapas jerarquizadas.

La primera etapa corresponde a determinaciones *in vitro*, donde son incluidos los ensayos químicos y determinaciones con líneas celulares. También se pueden realizar ensayos previos o en simultáneo a las determinaciones tradicionales de modelado *in silico* utilizando técnicas computacionales. La segunda etapa de ensayos corresponde a las determinaciones *in vivo* en sistemas modelo en animales, mientras que la etapa final corresponde a estudios epidemiológicos y ensayos clínicos con pacientes.

En particular, los estudios aplicados al descubrimiento y producción de péptidos bioactivos de proteínas alimentarias implican identificar una fuente de proteína adecuada para su posterior liberación mediante la acción proteolítica de enzimas endógenas o exógenas (preparaciones enzimáticas comerciales) o mediante fermentación (mediante la adición de cultivos iniciadores). El hidrolizado resultante usualmente es sometido a procesos de fraccionamiento para producir una preparación enriquecida en péptidos bioactivos y/o a etapas de purificación adicionales para aislar idealmente péptidos individuales e identificar sus secuencias (Li-Chan, 2015) (**FIGURA 3, Aproximación Clásica**).

La identificación y caracterización de péptidos bioactivos se ha visto favorecida por la aplicación de técnicas utilizadas en otros campos de la Ciencia como la Farmacología y Toxicología, así como por el posible abordaje con herramientas propia de la bioinformática (Iwaniak y col., 2015). El enfoque *in silico* (**FIGURA 3, Aproximación Bioinformática**) se refiere a la aplicación de métodos computacionales para manejar, curar e interpretar información sobre sistemas biológicos, en este caso, los péptidos bioactivos derivados de los alimentos. Basándose en el conocimiento sobre la estructura y actividad de los péptidos reportados en la bibliografía y depositados en las bases de datos pertinentes, se

pueden aplicar aproximaciones computacionales para dilucidar las relaciones estructura-función, predecir secuencias peptídicas susceptibles de exhibir actividades específicas, localizar péptidos encriptados en fuentes proteicas, prever la liberación de esos fragmentos por la acción de enzimas específicas, y proponer un supuesto mecanismo de acción a través de técnicas de acoplamiento molecular (*docking*) a sitios de unión (Holton y col., 2013; Li-Chan, 2015; Udenigwe, 2014).

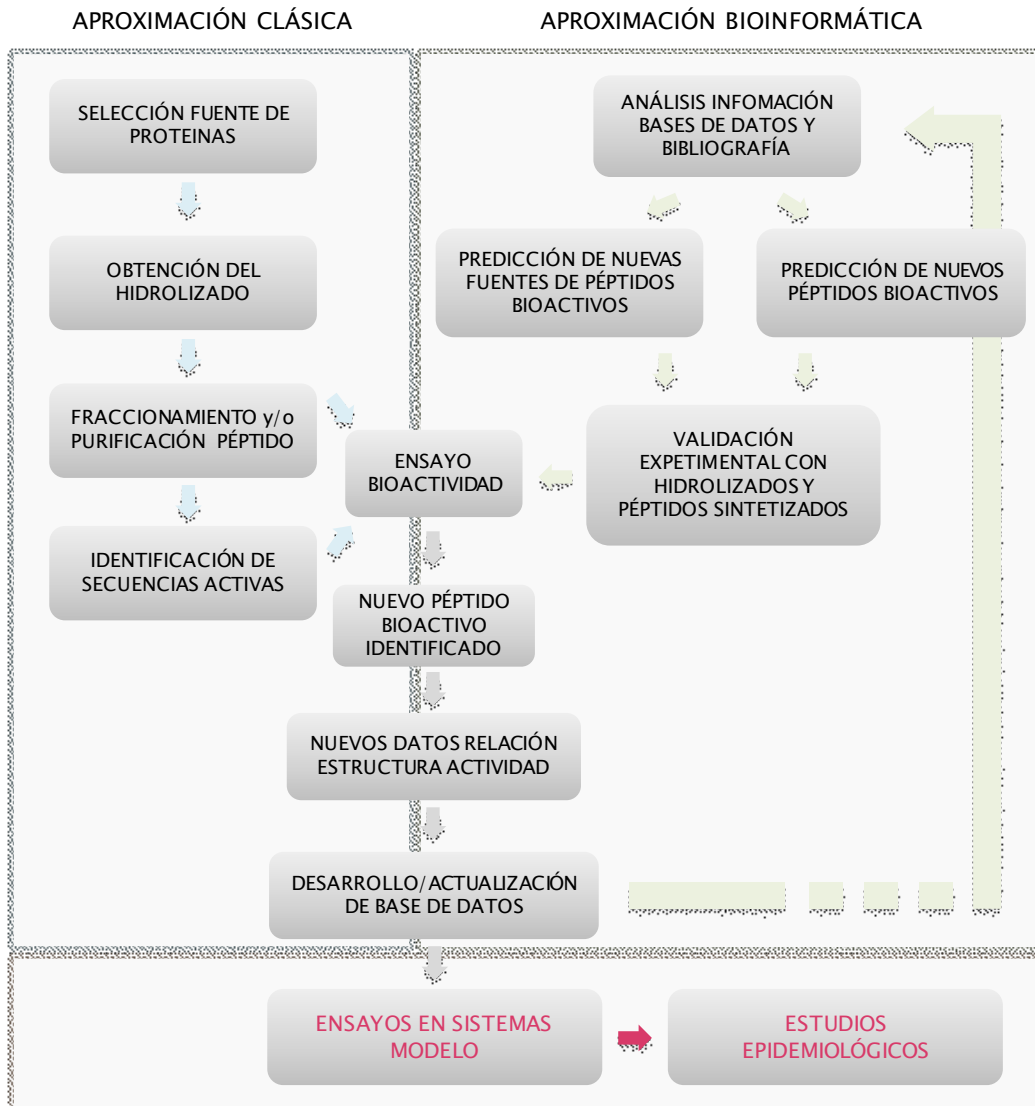


FIGURA 3: Estrategias para el estudio de péptidos bioactivos.

Los ensayos clínicos y estudios epidemiológicos son fundamentales para verificar la eficacia de los péptidos bioactivos en humanos. Corresponden a la etapa final en la caracterización de compuestos bioactivos, y si bien actualmente en bibliografía el número de trabajos referido a la determinaciones *in vitro* y *en sistemas modelo* supera ampliamente al número de trabajos de ensayos clínicos,

podemos encontrar péptidos ya caracterizados que se encuentran en este estadio (Bouglé y Bouhallab, 2015; Nongonierma y FitzGerald, 2015). Fundamentalmente se han llevado a cabo estudios de péptidos antihipertensivos –por ser la actividad más ampliamente estudiada– (Kwak y col., 2013; Majumder y Wu, 2014), pero también hay estudios de péptidos derivados de proteínas lácteas –por ser las proteínas alimentarias más estudiadas– con efecto hipotensivo (Chanson–Rolle y col., 2015; Cicero y col., 2016), antidiabético y antiobesidad (Akhavan y col., 2010; Geerts y col., 2011).

I.4 Herramientas bioinformáticas para el estudio de péptidos bioactivos

Las **simulaciones de acoplamiento molecular o *docking* molecular** (por su nombre en inglés) son un procedimiento de simulación computacional que es utilizado para predecir la conformación de un complejo receptor–ligando. En el estudio de péptidos bioactivos se emplean típicamente para proponer un mecanismo de interacción entre el péptido y el sitio de unión de la enzima diana. También se han aplicado para dilucidar qué secuencias de péptidos –ya sea identificados experimentalmente o predichos a partir de investigación bioinformática– son en realidad las capaces de interactuar con las proteínas diana de la actividad biológica (Li–Chan, 2015).

Los protocolos de docking son una combinación de un algoritmo de búsqueda y una función de puntaje. El algoritmo de búsqueda es el conjunto de reglas y parámetros aplicados para definir una determinada conformación. A grandes rasgos existen dos tipos de algoritmos: basados en la complementariedad de forma y basados en simulaciones. La función de puntaje o scoring, si bien poseen parámetros y términos diferentes en los diferentes programas disponibles (London y col., 2013), buscan representar un modelo termodinámico de interacción del sistema proteína–ligando adecuado, para así distinguir los verdaderos modos de unión de todo el conjunto explorado y generar un ranking adecuado.

En la actualidad, la cantidad de protocolos y programas de docking, producto de la combinación de diferentes algoritmos de búsqueda y funciones de scoring es extensa (Fischer y col., 2014; London y col., 2013). La elección de un protocolo en particular debe considerar aspectos tales como la velocidad y la efectividad del algoritmo de búsqueda en cubrir el espacio conformacional relevante. El desempeño de las distintas herramientas de docking es altamente

dependiente de las características específicas de la proteína, sitios de unión y del ligando, de manera que resulta imposible establecer cuál de ellas es más apropiada en un determinado contexto (Dias y de Azevedo, 2008; Rosenfeld y col., 1995).

En el estudio de péptidos bioactivos derivados de proteínas alimentarias los estudios de docking se emplean para seleccionar de una lista posible de péptidos bioactivos cuáles serán sintetizados para la evaluación de la actividad en cuestión y proponer un probable mecanismo de acción de la secuencia activa. La lista de posibles ligandos usualmente proviene del análisis *in silico* de las secuencias de proteínas precursoras (Nongonierma y Fitzgerald, 2014; Vecchi y Añón, 2009; Velarde-Salcedo y col., 2013) o corresponde a secuencias identificadas mediante espectrometría de masas de diferentes fracciones que resultaron activas (García-Mora y col., 2017; Martínez-Villaluenga y col., 2010; Pan y col., 2012; Wang y col., 2017). Este tipo de estudios se ha aplicado exitosamente con péptidos derivados de proteínas alimentarias en actividades relacionadas con la enzima convertidora de angiotensina I (ACE, EC 3.4.15.1) (García-Mora y col., 2017; He y col., 2014; Jimsheena y Gowda, 2010; Norris y col., 2012; Vecchi y Añón, 2009; Wang y col., 2017), renina (EC 3.4.23.15) (He y col., 2014), ácido graso sintasa (EC 3.2.1.85) (Martínez-Villaluenga y col., 2010), ciclooxigenasa 1 (COX-1, EC 1.14.99.1) (Yu y col., 2016) y dipeptidil peptidasa IV (DPP4, EC 3.4.14.5) (Nongonierma y col., 2014).

Otra de las metodologías aplicadas en la identificación de nuevos péptidos bioactivos es la **relación cuantitativa estructura-actividad** (QSAR, *Quantitative Structure Activity Relationship*). Basándose en el principio que estructuras moleculares diferentes presentarán propiedades moleculares diferentes y estructuras moleculares similares presentarán propiedades similares (Tropsha, 2010) se aplican aproximaciones estadísticas para obtener correlaciones significativas entre una variable dependiente (la propiedad o actividad de un conjunto de compuestos de interés) y uno o más descriptores moleculares. El descriptor molecular "es el resultado final de un procedimiento lógico y matemático que transforma la información química codificada dentro de una representación simbólica de una molécula en un número útil o el resultado de algún experimento estandarizado" (Todeschini y Consonni, 2008). Existen miles de descriptores y son generados mediante diferentes herramientas computacionales (Dudek y col., 2006). Los descriptores moleculares pueden clasificarse en dos categorías principales: descriptores empíricos, derivados de

mediciones experimentales y descriptores teóricos, derivados de una representación simbólica de la molécula. En particular los descriptores teóricos pueden ser descriptores independientes de la conformación de un compuesto (descriptores bidimensionales) o bien dependientes (descriptores 3D) (Dudek y col., 2006).

La construcción de un modelo QSAR implica como primera actividad una selección minuciosa de compuestos de interés, con y sin actividad, de los cuales se seleccionará un subconjunto para el cálculo de las variables para construir el modelo, y se reserva otro subconjunto para la posterior validación. La selección de cuales descriptores son los adecuados en el contexto de la actividad estudiada y que serán incluidos en el modelo es una de las etapas claves que determina la calidad y validez del modelo. Una vez definidos y obtenidos los descriptores se procede a la búsqueda y construcción del modelo matemático óptimo que correlacione la actividad biológica o propiedad específica con la estructura molecular.

En el caso particular de péptidos bioactivos los modelos generados por análisis QSAR se aplican para predecir la actividad antes de (o incluso en lugar de) sintetizarlos y/o medir su actividad experimentalmente, lo cual suele ser muy costoso (Iwaniak y col., 2015). A diferencia del docking puede ser aplicado en el estudio de actividades que no deriven de la inhibición enzimática, como es la actividad antioxidante. Para péptidos se utilizan fundamentalmente descriptores físico-químicos que tienen en cuenta tamaño, carga, polaridad, secuencia y topología de los péptidos. La potencia de una actividad en particular se expresa con valores de IC_{50} . Esta medida cuantitativa representa la concentración de proteínas/péptido en la mezcla de reacción necesaria para producir un 50 % de inhibición de un proceso biológico dado (o componente de un proceso, es decir, una enzima o receptor celular, etc.). Los valores se expresan típicamente como concentración molar. Para la construcción de un modelo QSAR en la mayoría de casos se emplean valores que proviene de bibliografía y/o bases de datos y no de experimentos internos (Pripp y col., 2005). Esto se debe tener en cuenta por qué puede derivar de experimentos llevados a cabo en diferentes condiciones lo que finalmente puede afectar la previsibilidad del modelo QSAR.

Zhou y Col., (2013) combinando el análisis QSAR junto con métodos de mecánica cuántica y posterior análisis mecánico molecular de la estructura y el perfil energético de complejos de péptidos-enzima, correlacionaron la actividad inhibidora de ACE y sabor amargo con las propiedades estructurales del péptido

y los perfiles de interacción. La correlación entre la inhibición de la ACE y el amargor fue más estricta para los dipéptidos y disminuyó notablemente para tri y tetrapéptidos, los autores consideraron que este comportamiento era debido al aumento exponencial de la diversidad conformacional con cada aminoácido adicional en la longitud del péptido. Además, el análisis estructural y energético de los complejos enzima-péptido indicaron que la potencia inhibidora de ACE (inferida de la energía de unión) aumentó desde el dipéptido al tripéptido y tetrapéptidos, pero no así para péptidos más largos. En estos casos los residuos terminales residen fuera del sitio activo de la enzima y por tanto tienen poca influencia en la unión. Utilizando un enfoque similar, Wang y col., (2013) informaron una correlación positiva significativa entre la potencia de inhibición de ACE y la actividad antioxidante para tripéptidos, y encontraron sólo una correlación moderada con el sabor amargo de los péptidos -una característica típica de los di y tripéptidos- lo que sugiere el potencial para desarrollar productos de péptidos funcionales no amargos con bioactividad múltiple.

Actualmente hay disponibles diversas bases de datos de péptidos bioactivos como PepBank (Shtatland y col., 2007), Antimicrobial Peptide Database (APD) (Wang y col., 2016), cancerPPD (Tyagi y col., 2015) y BIOPEP (Minkiewicz y col., 2008) entre otras. La mayoría reúnen información de péptidos endógenos o de una actividad en particular. BIOPEP sigue siendo la única base especializada en péptidos derivados de proteínas alimentarias y le ofrece al usuario la capacidad de generar perfiles de potencial actividad biológica de la proteína de interés, así como la frecuencia de aparición de fragmentos bioactivos en la proteína. Otro recurso disponible son los repositorios de enzimas proteolíticas y sus inhibidores como MEROPS y BRENDA (Rawlings y col., 2014; Schomburg y col., 2002). Estos recursos permiten predecir la mejor combinación enzimática para obtener productos enriquecidos en péptidos con determinada actividad. Sin embargo, este tipo de análisis no puede predecir fácilmente la bioactividad de las combinaciones de péptidos que están presentes en los hidrolizados o fracciones de proteínas. Esta metodología encuentra limitaciones que deben tenerse en cuenta. Una de ellas es que la especificidad de los preparados enzimáticos comerciales empleados usualmente en alimentos para obtener hidrolizados difiere considerablemente de la especificidad de las enzimas aisladas, hecho que limita la predicción del perfil de péptidos que se obtendrá. Segundo esta metodología solo es aplicable en proteínas que ya han sido secuenciadas, lo cual no es común en proteínas alimentarias y mucho menos

aun en los casos en los que se busca identificar péptidos de nuevas fuentes, con el objetivo de dar valor agregado a un material o fuente de proteínas que de otra manera se descartaría. Finalmente, el hecho de que: una mayor frecuencia de aparición de secuencias bioactivas en una molécula de proteína no se correlaciona necesariamente con el potencial de esa proteína para servir como una buena fuente de péptidos bioactivos, deberían por lo tanto considerarse la potencia de cada fragmento bioactivos y los solapamientos de secuencias bioactivas.

Es importante resaltar que la fiabilidad y utilidad de la bioinformática depende en gran medida de los datos utilizados para el análisis *in silico*. Holton y col. (2013) señalaron que los enormes avances en las herramientas bioinformáticas aplicadas en varias disciplinas incluyendo la Biotecnología, el descubrimiento de fármacos, la Genómica comparada y la Genómica microbiana, no han sido paralelos en la investigación en Ciencias y Tecnología de los Alimentos y Nutrición, en las cuales el uso de bioinformática se encuentra aún "en su infancia". Una de las principales falencias es la ausencia de un lenguaje o léxico de consenso. Esto resulta evidente al momento de realizar la determinación *in vitro* de una determinada actividad. No se emplea un único protocolo estandarizados para la determinar si un péptido presenta actividad inhibitoria de una enzima en particular, y pequeñas variaciones en las condiciones en las cuales se ensayan las actividades deberían estar disponibles en las bases de datos. Dichos protocolos pueden ser muy diferentes entre los distintos laboratorios de investigación, y también pueden variar mucho de un alimento a otro. Fundamentalmente los valores de IC_{50} , no siempre son informados en la bibliografía y, además, pueden variar para secuencias idénticas si se ensayan en diferentes condiciones y pueden estar influenciados por condiciones de ensayo incluyendo las concentraciones de enzima y sustrato. De este modo, a menos que la actividad inhibitoria se informe como la constante de afinidad del inhibidor (K_i), la potencia de los diferentes péptidos reportados por diferentes investigadores puede no ser siempre comparable (Holton y col., 2013).

Como resulta evidente a partir de la discusión precedente, aun reconociendo sus limitaciones, la incorporación de herramientas bioinformáticas al estudio de péptidos bioactivos resulta necesaria. Las aproximaciones conducidas por la bioinformática pueden conducir al descubrimiento de nuevos péptidos en menores tiempos y con menores costos que los enfoques exclusivamente experimentales.

II. ENFERMEDADES CARDIOVASCULARES Y COLESTEROL

A continuación, se describirán aspectos generales de las enfermedades cardiovasculares, el colesterol y de la actividad hipocolesterolémica, por estar relacionadas en forma directa con el objetivo del presente trabajo de tesis y posteriormente se detallará la posible acción de los péptidos bioactivos en estas enfermedades.

II.1 Enfermedades Cardiovasculares

Las enfermedades cardiovasculares (ECV) como la arterioesclerosis, apoplejías e infartos de miocardio constituyen hoy en día uno de los problemas más relevantes de la salud a nivel mundial. Según datos de la Organización Mundial de la Salud (OMS) más de 17 millones de personas mueren por año a causa de este tipo de afecciones y se prevé que para el año 2020 las enfermedades asociadas al sistema cardiovascular se convertirán en la primera causa de mortalidad y discapacidad a nivel mundial (López y Murray, 1998).

Actualmente en los países desarrollados este tipo de afecciones se encuentra entre las primeras cinco causas de muerte siendo, de acuerdo con datos oficiales del Ministerio de Salud de la Nación, la principal causa en la Argentina (Indicadores Básicos, Ministerio de Salud, 2016). De acuerdo a información, a nivel país en el 2014 la tasa de mortalidad bruta correspondiente a enfermedades cardiovasculares fue de 215,59 cada 100.000 habitantes, seguida por cáncer y en tercer lugar por enfermedades infecciosas.

Las ECV son patologías que afectan tanto al corazón como a los vasos sanguíneos. Entre ellas podemos mencionar las enfermedades coronarias, las

cerebro-vasculares, las enfermedades arteriales periféricas, la enfermedad cardíaca reumática, las cardiopatías congénitas, la trombosis venosa profunda y la embolia pulmonar.

Se han descrito una gran variedad de actividades biológicas atribuidas a péptidos provenientes de proteínas alimentarias. Entre ellas podemos mencionar la actividad antihipertensiva, antitrombótica, anticolesterolémica, antioxidante, antimicrobiana, anticancerígena y actividad opioide. Las mismas se pueden agrupar de acuerdo al sistema fisiológico sobre el cual ejercen su efecto, tal como se muestra en la **FIGURA 2**.

El estrés oxidativo es considerado como factor de la iniciación y/o progresión de las ECV, debido a que las moléculas de LDL oxidadas actúan en el endotelio como precursores de la formación de placas (Steinbrecher y col., 1990; Witztum y Steinberg, 1991).

La obesidad, en particular, genera una carga socioeconómica y de enfermedad muy alta, que compromete tanto la salud de la población como las finanzas públicas y privadas actuales y futuras. Su causa está relacionada a los cambios de hábitos de la población, en particular los cambios en la dieta y el sedentarismo y es una enfermedad que prevalece en todas las clases sociales.

El último mapa mundial sobre gordura en América del Sur dado a conocer por la Organización Mundial de la Salud da cuenta que en la región Argentina, Chile y Uruguay presentan la mayor proporción de individuos con sobrepeso (63,9; 63,2 y 62,4%, respectivamente), en tanto que la población de obesos alcanza el 23,6; 23,3 y 22%, respectivamente. La obesidad en Argentina se incrementó más de un 60% (el valor correspondiente al 2005 fue de 14,6 %). Según la encuesta mundial de salud escolar realizada por el Ministerio de Salud de la Nación en 2012, uno de cada tres adolescentes entre 13 y 15 años padecía sobrepeso u obesidad. Según la OMS Argentina presenta un alto porcentaje de obesidad en niños menores de 5 años, con una tasa de prevalencia del 7,3%.

El Ministerio de Salud de la Nación cuenta con un Programa Nacional de Alimentación Saludable y Prevención de la Obesidad, cuyo objetivo es “Ofrecer un enfoque integral de la alimentación saludable y la problemática del sobrepeso y la obesidad abarcando los aspectos principales que determinan su incremento”. En febrero 2016 fue publicado el Boletín de Vigilancia N° 8 sobre Enfermedades No Transmisibles y Factores de Riesgo el cual fue dedicado a la obesidad; en dicho boletín se puede encontrar información sobre determinantes de la obesidad, epidemiología de la obesidad y el sobrepeso en la Argentina,

intervenciones poblacionales para la prevención y control de la obesidad y sobrepeso y ejes de acción para abordar la obesidad en niños y adolescentes. La prevención y el control de las ECV debe abordarse mediante una acción global e integrada que requiere la combinación de medidas que tiendan a reducir los riesgos en la totalidad de la población y el establecimiento de estrategias dirigidas hacia los individuos con alto riesgo o que ya padecen la enfermedad. Aun cuando las enfermedades cardiovasculares son, como se indicó previamente, la primera causa de muerte y discapacidad en Argentina, los programas nacionales, provinciales y municipales continúan centrando sus intervenciones casi exclusivamente en el control de las enfermedades transmisibles o perinatales. Es imperativo emprender acciones y establecer programas dirigidos a cambiar los hábitos y estilos de vida de la población, e implementar simultáneamente intervenciones clínicas que lleven a reducir la incidencia.

Debido a la problemática que representan las ECV y la relación directa entre la dieta y el estilo de vida sobre la incidencia de estas enfermedades y sus factores de riesgo, la ciencia ha puesto un gran interés en dilucidar de qué manera se puede emplear la dieta como estrategia para disminuir el riesgo de padecer este tipo de enfermedades.

II.2 Colesterol

El colesterol es un constituyente de las membranas celulares, que modula la fluidez y permeabilidad de las mismas. El colesterol, por otra parte, es necesario para la formación de estructuras de transporte cumpliendo un rol fundamental en el tráfico vesicular, en procesos de señalización transmembrana y en la viabilidad y proliferación celular (Goedeke y Fernández-Hernando, 2012; Van der Wulp y col., 2013). Su distribución en membranas a nivel subcelular es muy heterogénea. Su principal destino es la membrana plasmática (60–80%) mientras que en el retículo endoplásmico solo se encuentra el 0,5–1% del colesterol total. Además, el colesterol es precursor de todas las hormonas esteroideas, vitamina D, oxisteroles y ácidos biliares que cumplen funciones fisiológicas importantes. A pesar de su importancia crítica, niveles anormales de colesterol pueden causar serias consecuencias a nivel celular provocando enfermedades como aterosclerosis y diabetes tipo II. Por esta razón, las células han desarrollado mecanismos muy complejos de regulación relacionados con los niveles y

distribución del colesterol (Goedeke y Fernández-Hernando, 2012; Ikonen, 2006).

Los seres humanos, y mamíferos en general, obtienen el colesterol a partir de la síntesis endógena y de la dieta en una relación aproximada de 70/30. En una dieta occidental típica, la biosíntesis a partir de acetil-CoA oscila entre 800–1000 mg/día, mientras que el colesterol proveniente de la ingesta ronda los 400–500 mg/día. Este último valor resulta del aporte de colesterol presente tanto en los alimentos ingeridos (300 mg/día) como del que se excreta en sales biliares (1g/día), teniendo en cuenta que solo se absorbe un 40% del colesterol disponible a nivel intestinal (Ikonen, 2006; Ros, 2003; Van der Wulp y col., 2013). El hígado es el órgano central en el metabolismo del colesterol ya que no solo es el principal sitio de síntesis endógena (50% del total en el organismo), sino que además controla su secreción en sangre y su eliminación (Zhao y Dahlman-Wright, 2010). Esto lo hace a través de la producción de lipoproteínas y su captación mediada por receptores, de la formación de ésteres de colesterol y su almacenamiento y de su excreción del organismo a través de la bilis. El intestino también cumple un papel determinante regulando su absorción y excreción en las heces (Fisher y col., 2012; Zhao y Dahlman-Wright, 2010). Dado que el colesterol es una molécula muy hidrofóbica, su absorción intestinal requiere previa emulsificación, hidrólisis del colesterol esterificado de la dieta, solubilización micelar por sales biliares y posterior captación por parte de los enterocitos (Van der Wulp y col., 2013).

III. AMARANTO

III.1 Generalidades del amaranto

El amaranto es considerado un pseudocereal porque sus semillas contienen cantidades importantes de almidón y son ricos en materia harinosa, al igual que las semillas de los cereales. Pertenece a la familia de las dicotiledóneas, familia *Amaranthaceae*, que son plantas con dos hojas embrionarias en sus semillas, diferentes a las monocotiledóneas gramíneas, familia a la que pertenecen los cereales como el arroz, el maíz y el trigo (Kiegel, 1994). Las dicotiledóneas no contienen proteínas generadoras de gluten y son fácilmente digeribles, razón por la cual su consumo, sobre todo en países europeos donde la incidencia de la enfermedad celíaca es alta, ha aumentado y se ha promovido en los últimos años. Otra ventaja de los pseudocereales es que crecen de forma rústica y se adaptan a variados ambientes, resistiendo sequías, altas temperaturas y suelos salinos (Lozoya–Gloria, 1994). La planta puede alcanzar entre 2 y 2,5 m de altura en la madurez, presenta ramas que se ubican en el tallo principal pudiendo comenzar desde su base. La raíz principal es corta y las raíces secundarias se dirigen hacia el interior del suelo. Sus flores brotan del tallo principal y en algunos casos las inflorescencias llegan a medir 90 cm de largo, pudiendo ser rectas, semirectas o sueltas, de coloración roja, rosada, anaranjada o amarilla. Las semillas, contenidas en los frutos, son levemente lenticulares y de tamaño muy pequeño, aproximadamente 2 mm de diámetro. Se pueden encontrar semillas de diversos colores, desde el negro pasando por el rojo, hasta el marfil y el blanco.

El amaranto es una planta de origen americano que formaba parte de la dieta habitual de los pueblos originarios. En esos tiempos y actualmente,

principalmente en los países andinos, se aprovechan tanto las semillas como las hojas dado su alto valor nutritivo.

El amaranto es capaz de realizar fotosíntesis con alta eficiencia, ya que al ser una planta C4 su mecanismo de fijación de carbono supera al resto de las plantas. Las plantas tipo C4 crecen, en general, más rápidamente, responden bien a condiciones ambientales adversas y utilizan más eficientemente el agua consumida para formar biomasa en comparación con las plantas C3. Estas características hacen que las plantas C4 sean más amigables con el medio ambiente. La semilla, presenta un elevado contenido de proteínas (13 al 19% p/p), con un buen balance aminoacídico, lo que las hace inusualmente apropiadas para el consumo humano (Bale y Kauffman, 1992). El valor nutricional de las proteínas de la semilla de amaranto es cercano al óptimo requerido para la dieta humana, según lo detallan la Organización Mundial de la Salud y la Organización de Alimentos y Agricultura (FAO). La cantidad de lisina que contiene es dos a tres veces superior a la de los cereales, y la cantidad de aminoácidos azufrados es superior al de las legumbres (Segura-Nieto y col., 1994). El contenido de lípidos varía de 7 a 8% p/p. Rodas y Bressani (2009) han encontrado en esta fracción una cantidad relativamente alta de escualeno (aproximadamente 8% del total), el cual presenta diversos beneficios para nuestro organismo: protege de los radicales libres debido a su capacidad antioxidante, es precursor del colesterol fortaleciendo a la membrana celular y presenta acción bactericida y antifúngica. Las semillas de amaranto también son una buena fuente de fibra dietaria, cuya composición depende de la especie y variedad y del procesamiento realizado (Venskutonis y Kraujalis, 2013). Entre los componentes de la fibra, mayoritariamente insolubles, pueden mencionarse homogalacturonano y ramnogalacturonano-I con cadenas laterales de arabinanos, xiloglucanos altamente ramificados y celulosa. La fracción de fibra soluble se encuentra compuesta por xiloglucanos y polisacáridos pécticos ricos en arabinosa (Lamothe y col., 2015).

Las partes verdes de la planta, especialmente las hojas, son ricas en minerales como K, Fe, Mg y Ca y en vitaminas A y C. La apariencia, textura y calidad nutricional de las hojas de amaranto son comparables a las de la espinaca (Segura-Nieto y col., 1994).

El Código Alimentario Argentino ha incorporado las especificaciones relacionadas con el amaranto hace pocos años atrás, en 1994. En el Capítulo XI, Alimentos vegetales, encontramos la siguiente especificación: “con el nombre

de Amaranto se entienden las semillas sanas, limpias y bien conservadas de las siguientes especies de este pseudocereal: *Amaranthus cruentus* L., *Amaranthus hypochondriacus* L., *Amaranthus caudatus* L. y *Amaranthus mantegazzianus* Passer.” Además, especifica los siguientes valores:

El contenido de proteínas no debe ser menor de 12,5%.

La humedad no debe ser mayor de 12,0%.

El contenido de cenizas debe ser menor de 3,5%.

El contenido de almidón no debe ser menor de 60%.

El peso hectolítrico mínimo será de 77 kg.

Los granos de amaranto, que respondan a las especies mencionadas, serán de color blanco, ámbar pálido, amarillo o castaño muy pálido, opacos o translúcidos" (CAA, Artículo 660 – Res MSyAS 80, 13.01.94).

III.2 Proteínas de reserva de las semillas de amaranto

Las proteínas de almacenamiento de semillas pueden clasificarse utilizando el criterio de Osborne (Osborne, 1924), que se basa en la extracción de fracciones proteicas de las harinas obtenidas a partir de las semillas, según su solubilidad diferencial en diferentes medios. Así, las proteínas de semillas se pueden clasificar en cuatro fracciones. La fracción proteica que se extrae por solubilización en agua destilada se conoce como *albúminas*; en general corresponde a proteínas involucradas en las rutas metabólicas. Si sobre el residuo de la extracción con agua se utiliza una solución salina diluida, se extraerán las proteínas que corresponden a la fracción *globulinas*. Una vez separadas las globulinas y utilizando sobre el residuo una solución hidroalcohólica se obtendrá la fracción conocida como *prolaminas*. Por último, utilizando soluciones de pH extremo o incorporando detergentes en el solvente de extracción, obtendremos la fracción *glutelinas*. Las globulinas, prolaminas y glutelinas generalmente corresponden a proteínas de reserva. Estas proteínas se localizan en las semillas a fin de ser utilizadas como fuente de nitrógeno en el momento de la germinación para originar una planta. En el caso particular del amaranto, existe una fracción proteica cuantitativamente importante que se obtiene por extracción con agua destilada luego de extraer las albúminas y las globulinas, denominada *globulina-P*.

Albúminas de amaranto

Es la fracción que tradicionalmente se extrae de la harina debido a su capacidad de solubilizarse en agua. Está compuesta por polipéptidos de diversos tamaños

la mayoría entre 10 y 43 kDa (Marccone, 2000; Segura-Nieto y col., 1992) encontrándose también algunos polipéptidos de mayor masa molecular, hasta 94 kDa (Barba de la Rosa y col., 1992; Konishi y col., 1991; Martínez y Añón, 1996). La mayoría de los polipéptidos son monoméricos, no dissociables por 2-mercaptoetanol. Las albúminas presentan una baja estabilidad térmica, siendo su temperatura de desnaturalización de aproximadamente 64°C (Martínez y Añón, 1996). En su composición poseen principalmente aminoácidos azufrados, glutamina, ácido glutámico y glicina (Segura-Nieto y col., 1992). La fracción de albúminas contiene proteínas con diferente función biológica que no son de reserva. Entre estas últimas, se encuentra un inhibidor de tripsina y uno de α -amilasa que han sido caracterizados y secuenciados por Valdes-Rodriguez y col., (1993) y Lu y col., (1999) respectivamente. También se encuentran la lectina, una proteína de unión carbohidratos que juegan un rol importante en la defensa de la planta. Presenta una estructura homodimérica con masas moleculares en el rango de 54–70 kDa. Es además, una de pocas proteínas de género *Amaranthus* de la cual su estructura ha sido determinada experimentalmente (PDB 1JLY) (Transue y col., 1997).

Globulinas de amaranto

Se extraen del residuo de harina resultante de la extracción de las albúminas. Para extraer globulinas se utilizan soluciones salinas de pH cercano a la neutralidad. En el amaranto existen tres tipos de globulinas mayoritarias: 7S, 11S y globulina-P.

Globulina 7S: Es la que se encuentra en menor proporción. Los estudios realizados por Marccone y Kakuda (1999) sobre esta fracción indican que se organiza estructuralmente formando trímeros, cuya masa molecular varía entre 150 y 200 kDa. El heteroligómero se encuentra constituido por tres subunidades de diferentes tamaños que varían entre 15 y 90 kDa, unidas de modo no covalente. Sus subunidades mayoritarias son las de 66 y 52 kDa, mientras que los polipéptidos menores resultan de la hidrólisis de la subunidad de 52 kDa (Quiroga y col., 2012). Posee baja proporción de estructura en hélice α (12,6%) y elevada proporción de estructura secundaria en hoja plegada β (50%). Su punto isoeléctrico (pI) se encuentra entre 5,2 y 5,8.

Globulina 11S: Esta globulina está organizada en hexámeros con masas moleculares aparentes entre 300 y 360 kDa (Marccone y col., 1998). Los hexámeros están formados por subunidades polipeptídicas de diferentes tamaños que se mantienen unidas por interacciones no covalentes (Salnikow y

col., 1998). Como otras globulinas de almacenamiento de semillas de la familia de las 11S, cada una de las subunidades están formadas por un polipéptido ácido (A) y uno básico (B), unidos mediante enlaces disulfuro. La composición aminoacídica de globulina 11S parcialmente purificada fue determinada por Marcone y col. (1998) y se puede observar que los aminoácidos más abundantes son glutamina/ácido glutámico y asparagina/ácido aspártico. Contiene relativamente alto porcentaje de estructura hélice α (31,4%) y desordenada (41,6%), siendo el resto hoja plegada β (Gorinstein y col., 1996). Es la única proteína del grupo de las globulinas cuya secuencia completa y estructura ha sido determinada mediante cristalografía de rayos X (PDB 3QAC) (Tandang-Silvas y col., 2012).

Globulina P: Es una proteína perteneciente a la familia de las 11S. Esta fracción es soluble en agua y soluciones salinas de baja fuerza iónica, pero insoluble en concentraciones de NaCl mayores que 0,1 M. Para extraerla, se utiliza agua destilada sobre el residuo que queda luego de extraer las globulinas 7S y 11S. Su pI aparente se encuentra entre 5 y 6 y su composición aminoacídica, comparada con la de albúminas, presenta mayor proporción de prolina, serina, arginina, histidina y fenilalanina, y menor de lisina, metionina, cisteína y valina. Contiene varios polipéptidos; entre ellos, algunos de masas moleculares 56, 36, 31, 26 y 22 kDa (Konishi y col., 1991). Los polipéptidos de aproximadamente 56 kDa corresponden a los precursores de polipéptidos A y B que, como en la mayoría de las globulinas 11S se sintetizan como una única cadena polipeptídica. En la globulina P, la modificación postraduccional que escinde la cadena para producir polipéptidos ácidos y básicos no se produce en todas las moléculas (Quiroga y col., 2007) lo que conduce a unidades moleculares constituidas por los precursores además de los polipéptidos A y B. Una característica particular de esta proteína es su alto grado de polimerización, presentando aproximadamente sólo un tercio de sus moléculas no agregadas. Las unidades moleculares presentan una masa molecular cercana a 300 kDa (Castellani y col., 1998 y 2000; Martínez y col., 1997). Su nombre, globulina-P (GP), deriva de globulina polimerizada, debido a la tendencia que presenta esta fracción. La GP presenta elevada estabilidad térmica; su temperatura de desnaturalización determinada por calorimetría diferencial de barrido fue cercana a 100°C, según describen Martínez y col., (1997).

Prolaminas de amaranto

En este pseudocereal, las prolaminas se encuentran en muy baja proporción (1 al 3,2% del total de las proteínas de amaranto, Segura–Nieto y col., 1994). Respecto a su composición aminoacídica, la leucina y la treonina son los aminoácidos esenciales más abundantes (Barba de la Rosa y col., 1992). Se ha descrito que esta fracción está constituida por polipéptidos de baja masa molecular (entre 10 y 22 kDa), aunque existen algunas discrepancias entre los estudios realizados que abordan cuáles son los componentes mayoritarios (Gorinstein y col., 1991; Segura–Nieto y col., 1992).

Glutelinas de amaranto

Las proteínas pertenecientes a esta fracción son solubles únicamente en soluciones ácidas o alcalinas. Presentan una composición polipeptídica y características estructurales similares a la globulina–P (Abugoch, 2006). Se han descrito polipéptidos de masas moleculares de entre 50–67 kDa, 30–38 kDa y 20–28 kDa (Barba de la Rosa y col., 1992; Gorinstein y col., 1998). De sus aminoácidos constituyentes, el más abundante es el ácido glutámico, después del cual se encuentran en abundancia decreciente arginina, leucina, fenilalanina y tirosina (Bressani y Garcia–Vela, 1990).

III.3 Péptidos bioactivos identificados en amaranto

Diferentes estudios han demostrado que las hojas y las semillas de amaranto presentan características nutricionales y promotoras de la salud relevantes (Huerta–Ocampo y Barba de la Rosa, 2011; Montoya–Rodríguez y col., 2015). Entre sus componentes bioactivos pueden mencionarse polifenoles con efectos antioxidantes; extractos metanólicos con actividades antihiperlipidémica, antidiabética y antihelmíntica; extractos acuosos con actividades antidiarreica, antifúngica y antimalaria; hidrolizados proteicos que inhiben la enzima convertidora de angiotensina (ACE) (efecto antihipertensivo) y el péptido lunasina con propiedades antitumorales (Huerta–Ocampo y Barba De La Rosa, 2011; Venskutonis y Kraujalis, 2013).

Por otra parte, haciendo uso de herramientas bioinformáticas se determinó que las proteínas de amaranto contienen, con alta frecuencia, secuencias peptídicas con potencial actividad inhibidora de la ACE y de la dipeptidil peptidasa IV

(DPP4), así como actividades antioxidante, estimulante de la captación de glucosa, antitrombótica y anticancerígena (Montoya-Rodríguez y col., 2015). En nuestro laboratorio, haciendo uso de diferentes tipos de ensayos, se ha demostrado la existencia de actividades antihipertensiva (Fritz y col., 2011; Quiroga y col., 2012; Vecchi y Añón, 2009), antioxidante (Lado y col., 2015; Orsini Delgado y col., 2016a,b; 2015; 2011; Tironi y Añón, 2014), antiproliferativa (Barrio y Añón, 2010; Mengoni, 2015; Quiroga y col., 2015) e inmunomoduladora (Moronta y col., 2016a,b), antitrombótica (Sabbione y col., 2016a,b; 2015) e hipocolesterolemica (Lado y col., 2015).

Actividad antihipertensiva

Sin lugar a dudas esta es la actividad fisiológica más estudiada de los péptidos bioactivos de diferentes proteínas alimentarias, no siendo ajeno a esta tendencia el amaranto.

Como es sabido, la hipertensión arterial es una afección multifactorial que puede dañar órganos vitales, por lo que está considerada como un problema importante de salud pública. Normalmente es tratada con diversos medicamentos sintéticos que afectan sitios claves de las vías metabólicas relacionadas particularmente con el sistema RAS. Entre este tipo de medicamentos se encuentran los inhibidores de la ACE, los β -bloqueantes, diuréticos y bloqueantes de los receptores de angiotensina-II. Si bien estos medicamentos son eficaces y bien tolerados por los pacientes, son caros y pueden producir ciertas alteraciones secundarias a corto y largo plazo como producción de tos (Atkinson y Robertson, 1979). Una alternativa a este tipo de medicamentos la ofrecen los péptidos bioactivos los cuales se sabe pueden ejercer diferentes acciones fisiológicas. Los más estudiados, hasta el presente, son aquellos inhibidores de la ACE. En la mayoría de los casos estos péptidos han sido identificados y caracterizados haciendo uso de ensayo *in vitro* y, en menor medida, se han realizado ensayos *in vivo*, tanto con animales modelo como estudios clínicos en humanos (Fritz y col., 2011; Martínez-Maqueda, Miralles y col., 2012; Quiroga y col., 2012).

La actividad antihipertensiva potencial de proteínas y péptidos de amaranto fue mayormente estudiada *in vitro* como capacidad inhibitoria de ACE. Así se demostró esta actividad en un hidrolizado triptico de glutelinas con un $IC_{50} = 0,20$ mg/ml (Barba de la Rosa y col., 2010) y en fracciones peptídicas aisladas por HPLC en fase reversa (RP-HPLC) de hidrolizados por alcalasa de

albúminas y globulinas, con IC_{50} de 0,35 y 0,15 mg/ml respectivamente (Tovar-Pérez y col., 2009). En nuestro laboratorio se ha demostrado que hidrolizados de aislado proteico de *A. mantegazzianus* de diferente grado de hidrólisis, obtenidos por acción de alcalasa, son potencialmente más activos que los aislados sin hidrolizar, alcanzando un IC_{50} de $0,12 \pm 0,02$ mg/ml (Fritz y col., 2011). Por otra parte, también hemos encontrado que las fracciones proteicas de globulinas 7S y 11S presentan actividad antihipertensiva luego de ser sometidas a digestión gastrointestinal simulada presentando un IC_{50} de 0,22mg/ml (Quiroga y col., 2012). Por su parte, Tiengo y col., (2009) analizaron la capacidad inhibitoria de ACE de un concentrado proteico de amaranto sometido a varios tratamientos, digestión gastrointestinal *in vitro*, hidrólisis con alcalasa y calentamiento. Los resultados por ellos obtenidos indican que la mayor actividad correspondió al hidrolizado con alcalasa ($IC_{50}=0,118$ mg/ml), la cual no se modificó significativamente luego del tratamiento térmico y la digestión gastrointestinal simulada. Fracciones peptídicas provenientes de estas muestras presentaron valores de IC_{50} levemente superiores a las correspondientes muestras originales (Tiengo y col., 2009). Luna-Suárez y col., (2010) insertaron en la subunidad ácida de la globulina 11S cuatro copias en tándem de un dipéptido de probada actividad inhibitoria (VY) con el objetivo de incrementar la capacidad inhibitoria de ACE de las proteínas de amaranto. Esta modificación de la secuencia aminoacídica de 11S mostró una actividad superior a la presentada por los hidrolizados ($IC_{50}=0,064$ mg/ml).

En relación a la actividad inhibitoria de ACE se realizaron también estudios *in silico* a partir de secuencias conocidas de proteínas de amaranto que mostraron poseer secuencias bioactivas encriptadas (Barba de la Rosa y col., 2010; Vecchi y Añón, 2009). Estos últimos autores, a partir de un estudio *in silico* sobre la secuencia de la globulina 11S, predijeron la actividad inhibitoria de dos tetrapéptidos, VIKP y ALEP, ubicados en la superficie molecular; actividad que fue confirmada mediante experiencias *in vitro* (Vecchi y Añón, 2009).

Unas pocas experiencias *ex vivo* permitieron proponer posibles mecanismos de acción. Mediante experiencias con músculo papilar de corazón y músculo liso de aorta aislados de ratas hipertensas se propuso que los hidrolizados de amaranto estarían actuando a nivel local o autocrino del sistema renina-angiotensina (Fritz y col., 2011). Por otro lado, Barba de la Rosa y col., (2010) estudiaron el efecto de un hidrolizado trípico de glutelina sobre la producción de NO de un cultivo de células endoteliales coronarias obteniendo resultados

que les permitieron proponer que el hidrolizado induce la producción endotelial de NO y la consecuente vasodilatación a través de su actividad inhibitoria de ACE. Más allá de la capacidad de los péptidos de modular la actividad de ACE, se han demostrado otros mecanismos de acción que conllevan a una reducción de la hipertensión tales como inhibición de la renina, efectos beneficiosos directos en la vasculatura (Hirota y col., 2007; Sipola y col., 2002), incremento de la expresión de genes involucrados en la regulación de la presión sanguínea como la óxido nítrico sintasa (De Gennaro Colonna y col., 2002; Rummery y col., 2005; Yamaguchi y col., 2009) y estimulación de los receptores periféricos de opioides (Fitzgerald y col., 2011). Teniendo en cuenta al estrés oxidativo como una causa de hipertensión, deben incluirse también aquellos péptidos con demostrada capacidad antioxidante (Martínez–Maqueda y col., 2012)

Estudios llevados a cabo por Udenigwe y col., (2009) y Li y Aluko (2010) demostraron que péptidos derivados de proteínas alimentarias liberados por hidrólisis enzimáticas son capaces de inhibir la enzima renina *in vitro*. Mediante la aplicación de modelos QSAR, Udenigwe, Li, y Aluko, (2012) arribaron a que los dipéptidos IW y LW eran capaces de inhibir el 70% (IC₅₀ 2,3 mM) y 37% (IC₅₀ 3,2 mM) de la actividad de dicha enzima. Estos estudios sugieren que los aminoácidos de baja masa molecular y naturaleza hidrofóbica ubicados en la posición N-terminal y los aminoácidos voluminosos en posición C-terminal son importantes en la determinación de la potencia de inhibición. Estudios recientes realizados en nuestro laboratorio (Quiroga y col., 2017) han demostrado por primera vez, la existencia de péptidos encriptados en las proteínas de almacenamiento de amaranto capaces de inhibir a la enzima renina. Dichos péptidos actúan como inhibidores competitivos, estando su potencia inhibitoria relacionada con la hidrofobicidad de los mismos.

Sólo se ha publicado, hasta el presente, una experiencia *in vivo* mostrando la capacidad antihipertensiva de péptidos de amaranto (Fritz y col., 2011) mediante la administración intragástrica de hidrolizado a ratones hipertensos.

Actividad antiproliferativa

En amaranto, la primera información sobre actividad inhibitoria de proliferación de células tumorales ha sido atribuida a compuestos de hojas de *A. tricolor* (Jayaprakasam y col., 2004) y más recientemente se ha demostrado la actividad antiproliferativa de proteínas y péptidos. En *A. hypochondriacus* Silva–Sánchez y col., (2008) detectaron un péptido de ~20 kDa que presentaba homología con

la lunasina de soja. Se demostró que este péptido, mayoritariamente presente en la fracción glutelinas, es resistente a tratamientos térmicos extremos y presenta actividad antiproliferativa selectiva sobre células neoplásicas. También se demostró, mediante experiencias *in vitro*, que este péptido es capaz de internalizarse en las células y localizarse en el núcleo y, que también presenta actividad inhibitoria de la acetilación de histonas en células HeLa (Maldonado-Cervantes y col., 2010).

Barrio y Añón (2010) describieron la actividad antiproliferativa de un aislado proteico de *A. mantegazzianus* y su hidrolizado con alcalasa, sobre cuatro líneas celulares de diferente origen y tumorigenicidad (MC3T3E1, UMR106, Caco-2, y TC7). Estos autores demostraron que la actividad dependía de la dosis y se debía a las proteínas, no a moléculas pequeñas como flavonas ni a la presencia de un inhibidor de proteasas tipo Bowman-Birk. Se estableció la concentración de 0,1 mg/ml con actividad antiproliferativa reversible tornándose irreversible a partir de 0,3 mg/ml. El mecanismo de acción propuesto para estas proteínas involucra la inhibición de la adhesión celular y proliferación junto con la producción de daño celular que resulta en una pérdida permanente de viabilidad celular. Con posterioridad, en nuestro laboratorio, se procedió a fraccionar la harina de amaranto en sus componentes proteicos más representativos: albuminas, globulinas 7 y 11S, globulina-P, gluteninas y una fracción de naturaleza hidrofóbica y se pudo demostrar que la última de las fracciones mencionadas presentaba la mayor actividad antiproliferativa sobre la línea celular UMR106. El análisis posterior de esta fracción por LC-MS/MS permitió identificar a la lectina como responsable de esta actividad (Quiroga y col., 2015). Los resultados encontrados en este trabajo sugieren que la inhibición de la proliferación celular producida por la fracción hidrofóbica se debe a un efecto citotóxico mostrado por una alteración de la función lisosomal y mitocondrial. Esta toxicidad promovería la apoptosis celular que se expresa en la modificación de la morfología celular y en la desorganización de los filamentos de actina (Quiroga, y col., 2015). Vistos estos resultados se llevó a cabo un proceso de purificación obteniéndose una lectina parcialmente purificada la cual fue sometida a un proceso simulado de digestión gastrointestinal. La lectina fue sensible a este proceso perdiendo parte de su actividad hemaglutinante y antiproliferativa (Mengoni, 2015). Se encuentran en desarrollo dos sistemas modelo animal, de ratas Wistar y peces Zebra, para ser utilizados en el estudio

de compuestos antitumorales derivados de amaranto, entre otro tipo de compuestos (Barrio, comunicación personal, 2012).

Actividad antioxidante

En cuanto al estudio de las propiedades antioxidantes en nuestro laboratorio se han realizado diversos ensayos *in vitro*, así como una primera aproximación de estudios *in vivo* en ratas. Se ha podido demostrar que existen en la semilla de amaranto (*A. mantegazzianus*) componentes peptídicos/polipeptídicos con moderada actividad antioxidante *in vitro*. Dicha actividad es fuertemente incrementada luego de la hidrólisis extensiva con alcalasa, así como también mediante la digestión gastrointestinal simulada sugiriendo que luego de la ingesta se producirá la liberación *in vivo* de péptidos antioxidantes (Orsini Delgado y col., 2011; Tironi y Añón, 2014).

Los principales resultados obtenidos indican que los digeridos obtenidos por simulación gastrointestinal de aislado e hidrolizado de amaranto con alcalasa, así como algunas fracciones obtenidas a partir de ellos, presentan diversas actividades *in vitro* frente a especies reactivas del oxígeno (ROS) y de nitrógeno (RNS). El aislado digerido fue el más activo, presentando una interesante capacidad para neutralizar radicales peroxilos, hidroxilos y anión peroxinitrito, así como para quelar metales inhibiendo la formación de radicales hidroxilos (Orsini Delgado y col. 2015). Algunas fracciones activas separadas a partir de los digeridos fueron capaces de atravesar una membrana epitelial de mucosa intestinal simulada mediante una monocapa Caco-2, manteniendo y en algunos casos incrementando su actividad (Orsini Delgado y col., 2016b)

Se han podido, además, identificar mediante RP-HPLC seguida de espectrometría MS/MS varios péptidos potencialmente antioxidantes liberados por digestión gastrointestinal simulada, cuya actividad fue confirmada sobre los péptidos sintetizados. Se analizó la relación estructura-actividad de los mismos mediante la predicción de las estructuras utilizando herramientas bioinformáticas. Los péptidos más activos detectados (AWEEREQGSR > YLAGKPQQEH ~ IYIEQGNGITGM ~ TEVWDSNEQ) presentaron al menos un aminoácido aromático voluminoso. Todos fueron provenientes de regiones poco estructuradas expuestas o internas de la subunidad ácida de la globulina 11S (Orsini Delgado y col., 2016a).

Recientemente se ha registrado un efecto *in vitro* de prevención de la oxidación inducida por Cu^{+2} y H_2O_2 de LDL humanas por parte del digerido gastrointestinal

del aislado, así como de fracciones separadas del mismo mediante FPLC de exclusión molecular.

Los ensayos *in vivo* realizados han demostrado que la incorporación de aislado proteico (2,5 % p/p) en dietas consumidas por ratas Wistar –adicionadas también con colesterol– produjo mejoras en parámetros relacionados con el estatus antioxidante del plasma: aumento de la capacidad reductora (FRAP), descenso del contenido de productos secundarios de oxidación (TBARS) y de la actividad superóxido dismutasa, y del hígado (descenso de TBARS). También se observaron efectos positivos sobre el perfil lipídico en hígado y sobre el descenso de la presión arterial (Lado y col., 2015).

Por otra parte, se ha estudiado la actividad antioxidante de aislados proteicos y sus hidrolizados con alcalasa incorporados a distintas matrices alimentarias tales como aceites vegetales y productos cárneos reestructurados, registrándose actividad antioxidante con el agregado del hidrolizado (García Fillería y Tironi, 2015; Tironi y Añón, 2014). Se ha comenzado a sí mismo, el estudio de otros posibles componentes antioxidantes. Se están analizando los compuestos fenólicos de la harina de amaranto y productos derivados tales como aislado, hidrolizado con alcalasa y digerido gastrointestinal de aislado, utilizando diferentes protocolos de extracción a fin de evaluar su interacción con otros componentes. La harina y el aislado contienen compuestos polifenólicos que se encuentran en alta proporción interaccionando con las proteínas u otros componentes. La hidrólisis enzimática de las proteínas y especialmente la digestión gastrointestinal simulada produciría una liberación de los mismos.

Estas conclusiones posicionan a la harina y al aislado proteico de amaranto como fuentes de componentes antioxidantes que pueden ser tanto beneficiosos para la salud del consumidor como para la estabilidad oxidativa de un alimento.

El amaranto contiene también polifenoles, como nicotiflorina, rutina e isoquercitina, y algunos ácidos fenólicos y amidas que exhiben propiedades antioxidantes. En este sentido, Alvarez–Jubete y col., (2010); Amin y col., (2006); Repo–Carrasco–Valencia y col., (2010); Yawadio Nsimba, y col., (2008) han demostrado la existencia de actividad antioxidante de componentes fenólicos en diferentes especies de amaranto. Por otra parte, también se ha asociado esta bioactividad con el escualeno, un triterpeno presente en la fracción lipídica del amaranto (Conforti y col., 2005). Este compuesto muestra actividad de secuestro de radicales libres y efecto protector frente a algunos tipos celulares (Spanova y Daum, 2011). Dada su naturaleza química, el escualeno no es atacado por

enzimas del sistema digestivo y sería absorbido en alta proporción por la mucosa intestinal mediante un mecanismo pasivo.

Actividad inmunomoduladora

El estrés y estilos de vida insalubre provocan un efecto acumulativo sobre el sistema inmune que aumenta el riesgo de muchas enfermedades entre las que se cuentan las cardiovasculares y el cáncer (Nakano y col., 2001). Experiencias *in vitro* e *in vivo* han demostrado la existencia de péptidos bioactivos con efectos inmunomoduladores que incluye la regulación de la expresión de citoquinas, de la producción de anticuerpos y de las funciones inmunes inducidas por las especies reactivas al oxígeno (ROS) (Agyei y Danquah, 2012; Udenigwe y col., 2012). Se han descrito péptidos con propiedades inmunoregulatoras de distintos orígenes como hidrolizados de soja, suero de leche, caseína, huevo, arroz, trigo, arveja y otros, los cuales presentaron diferentes acciones, ya sea regulando la respuesta inmune por estimulación humoral, mediada por células ó por ambas vías (Ndiaye y col., 2012; Phelan y col., 2009).

Lin, Chiang, y Lin (2005) mostraron que extractos de hojas de Amaranto ejercen un efecto estimulante inmunomodulatorio a través de la estimulación de células B. Por el contrario, en un estudio realizado hace un par de años, Montoya-Rodríguez y col., (2014) mostraron que hidrolizados de *A. hypochondriacus* obtenidos por acción de pepsina y pancreatina inhibían el proceso inflamatorio inducido por LPS en macrófagos de mono y humanos. Los resultados obtenidos sugieren que los péptidos generados reducen la fosforilación citosólica de I κ B- α previniendo la translocación de p50 y p65, subunidades de NF- κ B, hacia el núcleo y reduciendo la expresión de PEG2 y COX-1 y la secreción de TNF- α . Además, estos autores indicaron que los hidrolizados podrían ejercer una acción antiateroesclerótica y antiinflamatoria a través del control de señalización de LOX-1, que conduce a una disfunción endotelial y de la formación de plaquetas.

En nuestro grupo también se han realizado estudios relacionados con la actividad antiinflamatoria de péptidos de amaranto. Se ha demostrado que hidrolizados extensivos del aislado proteico con alcalasa redujeron la emisión de luz por parte de células Caco-2 transfectadas con reportero luciferasa bajo el control del promotor CCL20 (Caco-2 CCL20:luc) activadas con flagelina bacteriana, sugiriendo la presencia de péptidos con actividad antiinflamatoria. La separación de las fracciones más activas mediante RP-HPLC, su secuenciación

y el uso de herramientas bioinformáticas posibilitó la identificación de posibles péptidos activos. Se comprobó que el péptido SSEDIKE posee la capacidad de modular células activadas suprimiendo la expresión del ARNm que codifica para CCL20, sugiriendo una modulación de la vía NF- κ B (Moronta y col., 2016a). Posteriormente, se estudió la actividad *in vivo* de dicho péptido en un modelo de alergia alimentaria mediada por IgE en ratón. La administración oral del péptido indujo un alivio en los síntomas de hipersensibilidad ocasionados por proteínas de leche, negativizando el test de piel, suprimiendo la secreción de IgE y controlando la inflamación intestinal (supresión de la expresión de genes de marcadores inflamatorios tales como CCL20, gata-3 y NF- κ B, y aumento de transcritos codificando para $\text{tgf-}\beta$ y foxp3) mediante prevención de la activación de NF- κ B. Los resultados mostraron además que este pequeño péptido cargado resistió las condiciones ácidas del estómago pudiendo alcanzar la mucosa intestinal y probablemente moduló la respuesta inmune local por un mecanismo aún no dilucidado (Moronta y col., 2016 b).

Actividad antitrombótica

La trombosis es un proceso patológico asociado al inicio de los mecanismos de la hemostasia ante estímulos no fisiológicos, tiene como consecuencia la formación de coágulos que obstruyen vasos sanguíneos y no una finalidad reparadora.

De todas las actividades fisiológicas estudiadas en péptidos derivados de las proteínas de almacenamiento de amaranto, esta sin lugar a dudas prácticamente no ha sido abordada.

En nuestro laboratorio se han realizados estudios *in silico*, *in vitro*, *in vivo* y *ex vivo* relacionados con esta actividad. Los resultados obtenidos muestran que el proceso gastrointestinal simulado es el procedimiento que libera péptidos con mayor actividad antitrombótica, superior a aquellos péptidos liberados por acción de enzimas exógenas como alcalasa y tripsina o por acción de una proteasa endógena presente en granos de amaranto que se activa a pHs del orden de 2-4. Los IC_{50} calculados a partir de los valores obtenidos en los ensayos *in vitro*, haciendo uso del método de microplaca, fueron: $10,9 \pm 0,1$; $5,9 \pm 0,1$ y $0,23 \pm 0,02$ mg de proteína/ml para los hidrolizados obtenidos por acción de enzimas exógenas, la proteasa endógena y la digestión gastrointestinal simulada, respectivamente (Sabbione y col., 2016a,b,c; Sabbione y col., 2015). Durante los estudios realizados con animales se determinaron diferentes

parámetros asociados con distintas etapas de la hemostasia, los resultados obtenidos sugieren que el aislado proteico de amaranto incrementa el tiempo de sangría lo que indicaría una acción antitrombótica. Por otra parte los ensayos *ex vivo* mostraron que los péptidos liberados por las enzimas gastrointestinales serían capaces de ejercer un efecto antiplaquetario y/o anticoagulante (Sabbione y col., 2016a,b,c; Sabbione y col., 2015). Por último estudios realizados con las diferentes fracciones proteicas que forman parte de los aislados proteicos indicaron que la fracción albúminas es la que más prolonga el tiempo de formación de tromboplastina, en tanto que las fracciones de glutelinas y glutelinas hidrolizadas son las que mostraron los mayores tiempos de coagulación (Sabbione, 2015).

OBJETIVOS

La explotación óptima de los péptidos bioactivos para la nutrición humana y la salud es un desafío científico y tecnológico, a la vez que ofrece una posibilidad cierta de transferencia al sector comercial. Estos péptidos pueden ser incorporados como ingredientes en nuevos alimentos funcionales, ser utilizados en la formulación de suplementos dietarios y/o ser empleados por la industria farmacéutica.

Actualmente los métodos computacionales constituyen una herramienta fundamental para un abordaje integral en la búsqueda de nuevas fuentes de péptidos bioactivos y el estudio de los mecanismos moleculares que fundamentan su acción.

Teniendo en consideración lo antedicho nos planteamos como **Objetivo General** del presente trabajo de Tesis:

- ❖ Evaluar la potencialidad de la fracción proteica de globulinas de amaranto como fuente de péptidos bioactivos para su utilización como ingredientes alimentarios y/o desarrollo de alimentos funcionales.
- ❖ Comenzar con una nueva línea de trabajo relacionada con el uso de herramientas bionfórmicas en el análisis de las propiedades bioactivas de péptidos.

Siendo nuestros **objetivos específicos**:

- Estudiar la ocurrencia de péptidos bioactivos en el universo de proteínas conocidas.
- Analizar la ocurrencia de péptidos bioactivos en familias proteicas.
- Estimar mediante estudios computacionales la presencia de péptidos bioactivos en la fracción globulinas de amaranto.
- Obtener, caracterizar y purificar los péptidos que provienen de la hidrólisis de la fracción globulinas con actividad hipocolesterolemiantes mediante la inhibición de la enzima 3-hidroxi-metil-CoA reductasa.
- Caracterizar posibles mecanismos de acción de péptidos de amaranto con diferentes actividades utilizando técnicas *in silico*.

CAPÍTULO 1

**ESTUDIO DE LA OCURRENCIA DE
PÉPTIDOS BIOACTIVOS UTILIZANDO
MÉTODOS BIOINFORMÁTICOS EN
PROTEÍNAS DE SECUENCIA CONOCIDA**

1.1 INTRODUCCIÓN

Naturaleza del problema

En la introducción general se han descrito las propiedades más importantes de los péptidos bioactivos y la importancia biotecnológica de su detección. Evidentemente la predicción de la ocurrencia de péptidos bioactivos posee la ventaja de simplificar enormemente los protocolos experimentales reduciéndolos a aquéllos necesarios para poner de manifiesto la correspondiente actividad biológica del presunto péptido bioactivo. De esta forma, tanto los tiempos de los procedimientos experimentales como los costos asociados se reducen notablemente.

En los últimos años la bioinformática ha jugado un rol central en la predicción de numerosas propiedades en una gran diversidad de sistemas biológicos. En general, los métodos bioinformáticos detectan una determinada “señal” o más correctamente el contenido informativo de una secuencia, estructura, imagen, etc. (Luscombe y col., 2001). En el análisis secuencial, muchas veces este contenido informativo se correlaciona con regiones más conservadas evolutivamente que otras en un mismo gen o genoma. Según la teoría neutral de Kimura (Kimura y Ota, 1974) las regiones más conservadas evolucionan bajo una determinada presión selectiva, generalmente asociada a la conservación de la función biológica. Sin embargo, el término “función biológica” es un término muy amplio, ya que el mismo en este contexto puede implicar regiones conservadas debidas a restricciones estructurales (Chothia y Lesk, 1986), dinámicas (Maguid y col., 2006), provenientes de interacciones proteína-proteína (Chelliah y Lovell, 2004), funcionales (Worthy y Blundell, 2009), por exposición al solvente (Franzosa y Xia, 2006), por nombrar sólo algunos pocos

entre los muchos ejemplos descritos en la bibliografía. De esta forma la teoría neutral, la detección de conservación secuencial y su correlación con la función biológica dio lugar a la generación de centenas de métodos para predecir muy diversas cuestiones de interés biológico (Claverie, 2000; Higgins y Taylor, 2001). Sin embargo, no todas las regiones conservadas en una secuencia o genoma necesariamente estarán sujetas a un proceso evolutivo condicionado por la conservación de la función biológica. Esto hace necesario la estrecha interacción entre predicción bioinformática y experimentación. Este inter-juego entre predicción teórica y comprobación experimental implica una retroalimentación entre las dos áreas, una, la experimental, ahorrando tiempo y costos para evidenciar regiones realmente importantes en un sistema biológico, la otra, la predictiva y teórica, para controlar y mejorar sus métodos predictivos.

Entonces el principal resultado entre el inter-juego entre predicción y comprobación experimental es un conjunto de datos biológicos que contienen información biológica (“conservación”) y que además poseen una determinada actividad/función/importancia biológica comprobada experimentalmente. Esta información en Bioinformática se guarda ordenadamente, en forma sistemática y anotada en las denominadas bases de datos, las cuales pueden ser muy diversas (secuenciales, de motivos, estructurales, de sitios activos, de rutas metabólicas, etc). Las bases de datos, entonces, pueden contener datos crudos (a modo de repositorio de archivos) o información que ha sido depurada (curada) por expertos. Una curación cuidadosa agrega valor a la información y resulta en un incremento de la precisión de la información biológica depositada.

En cuanto al problema que nos concierne, la predicción de péptidos bioactivos en secuencias proteicas, es sumamente difícil. Como mencionamos en el capítulo anterior, los péptidos bioactivos son péptidos crípticos, que, liberados en el proceso de digestión, tienen alguna actividad biológica. Ya explicamos que la conformación que adquieren estos péptidos una vez liberados de la proteína, semeja la conformación de un sustrato biológico o de un regulador positivo de una proteína de interés en la salud humana. La dificultad en la predicción mediante herramientas bioinformáticas radica, centralmente, en el hecho que las regiones que contienen péptidos bioactivos en una determinada proteína no evolucionaron para mantener la función biológica propia de esos péptidos y en la enorme mayoría de los casos tampoco están asociados con la función biológica de la proteína que los contiene. Quizás, estas regiones están asociadas con cuestiones de estabilidad energética, estructural, cinética, etc. de la

proteína, pero difícilmente cumplan un rol biológico en la proteína que la contiene. A modo de ejemplo, el pentapéptido IIAEK (Naogaka y col., 2001) se encuentra encriptado en la β -lactoglobulina (proteína de suero de leche) en una región que adopta una conformación de cadena beta y, a la cual no se le asigna ninguna función biológica en su proteína de origen (FIGURA 1.1).

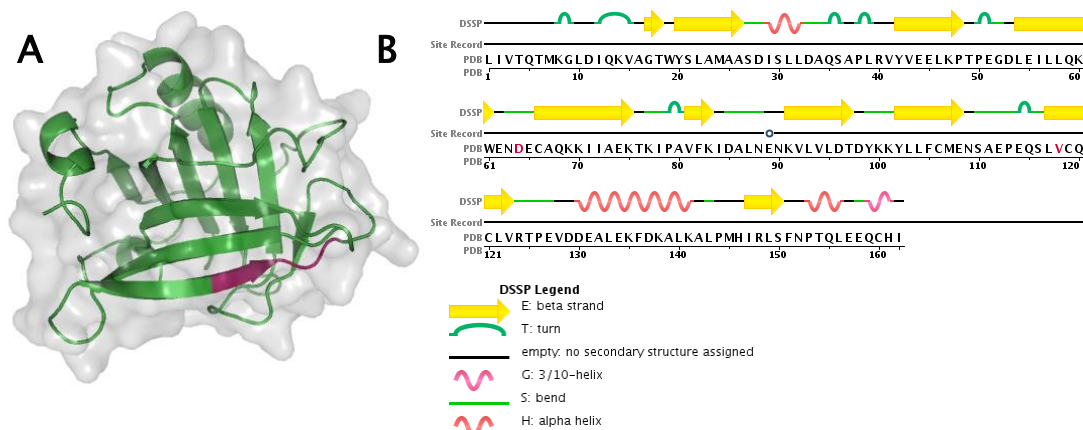


FIGURA 1.1: Localización del péptido IIAEK en la estructura (A) y secuencia (B) de la β -lactoglobulina (PDB 3BLG).

Esta aparente aleatoriedad de la ocurrencia de un péptido bioactivo en una determinada proteína y su desvinculación de los condicionamientos selectivos que mantienen regiones con “información” biológica para mantener la función biológica, hacen de los métodos bioinformáticos más comúnmente usados para predicción, prácticamente inútiles a la hora de predecir la ocurrencia de péptidos bioactivos. Por lo tanto, la única forma viable de predecir la ocurrencia de ellos es aplicar métodos de comparación secuencial, haciendo uso de bases de datos existentes de péptidos que hayan tenido su debida comprobación experimental (ver más adelante y para un review sobre métodos de predicción ver Udenigwe, (2014)).

Ante esta situación nos proponemos los siguientes objetivos:

1.2 OBJETIVOS ESPECÍFICOS

La información disponible sobre inhibidores enzimáticos encriptados en proteínas es conocida en el área de la Tecnología de los Alimentos, Nutrición y Ciencias médicas hace algunos años. Cualquier secuencia proteica, en principio, puede ser fuente potencial de péptidos bioactivos (PB). Hemos mencionado la dificultad de utilizar las técnicas bioinformáticas disponibles para la predicción de la ocurrencia de nuevos PB. Como los PB deben adoptar determinadas conformaciones para llevar a cabo su actividad biológica, y como aún no se han indagado las relaciones que existen entre las proteínas que contienen péptidos de una determinada actividad y sus propiedades estructurales y/o conformacionales, proponemos los siguientes **objetivos específicos**:

- Caracterizar las proteínas que contienen PB y explorar la relación estructural entre ellas.
- Estudiar la relación evolutiva entre las proteínas que contienen un mismo PB.
- Indagar si existen dominios que presenten regiones conservadas con mayor ocurrencia de péptidos bioactivos (**hot spots**).
- Caracterizar propiedades generales de proteínas que contengan PBs para utilizarlas con fines predictivos.

1.3 MATERIALES Y MÉTODOS

1.3.1 Esquema general

Para estudiar las relaciones estructurales de los PB en las proteínas que le dan origen y derivar propiedades generales que puedan ser utilizadas para predecir la ocurrencia de PB en proteínas, ya sea en forma específica o sólo indicando la posibilidad de que una determinada proteína contenga PBs, desarrollamos un pipeline (conjunto de herramientas bioinformáticas) para cumplir con estos objetivos. El mismo se resume en la **FIGURA I.2**.

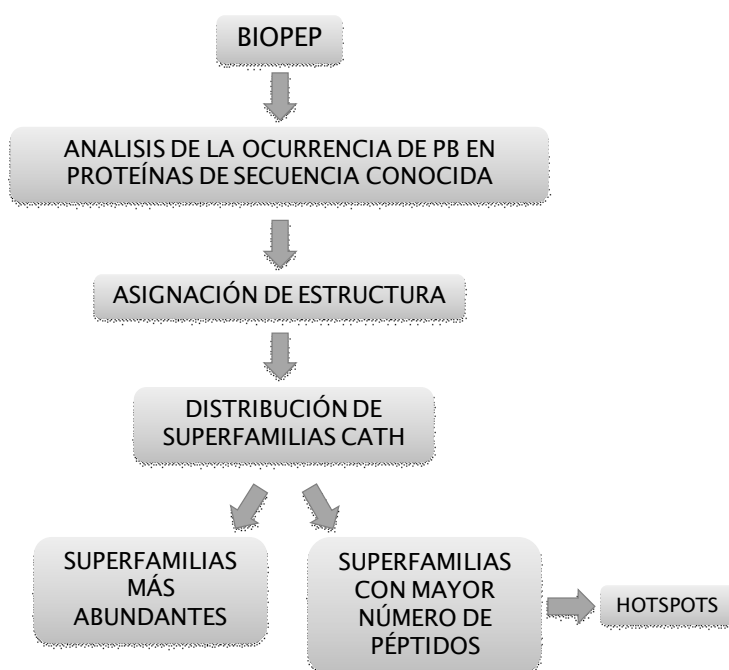


FIGURA I.2: Pipeline estudio de la ocurrencia de péptidos bioactivos utilizando métodos bioinformáticos en proteínas de secuencia conocida.

Utilizando bases de datos conteniendo PB conocidos y caracterizados experimentalmente, deseamos encontrar todas las proteínas conocidas que contengan al menos una ocurrencia exacta del péptido. En este primer estudio utilizamos la ocurrencia exacta de aminoácidos entre PB y secuencias proteicas y no la similitud entre aminoácidos. Creemos que esta última podría introducir imprecisiones, ya que no tendríamos la certeza que los reemplazos sinónimos o neutros considerados en la similitud entre aminoácidos (ponderadas en las denominadas matrices de sustitución por ejemplo JTT (Jones y col., 1992) fueran efectivamente neutros desde el punto de vista funcional del PB.

Esta primera etapa se llevó a cabo utilizando la base de datos no-redundante de NCBI utilizando un programa ad hoc escrito en el lenguaje de programación C. Una vez identificadas las secuencias con ocurrencias exactas de PBs, procedimos a clasificar estructuralmente a cada una de las proteínas que contenían al menos una ocurrencia de un PB conocido. Para esta segunda etapa utilizamos la base de datos de clasificación estructural CATH (Greene y col., 2007). El procedimiento que seguimos para esta asignación de plegamiento fue realizar búsquedas de similitud secuencial utilizando el programa BLAST (Altschul y col., 1997) para cada secuencia con un PB sobre las secuencias CATH. Como CATH es una base de datos de dominios estructurales (ver más abajo) una misma secuencia proteica se separó en sus correspondientes dominios estructurales según la clasificación de CATH. De tal forma que para cada secuencia conteniendo un PB se registró la posición del mismo y su ubicación en un dominio particular de la secuencia correspondiente. Además, la clasificación CATH permite identificar familias de proteínas homólogas, por lo que en la anotación de cada secuencia se incluyó la familia de homólogos estructurales a la cual pertenece.

Finalmente, para estudiar si la ubicación del PB tenía alguna ocurrencia particular dentro del dominio estructural detectado, utilizamos alineamientos estructurales entre los dominios CATH (proteínas de estructura conocida) y las secuencias conteniendo el PB. Estos alineamientos estructurales se llevaron a cabo con el programa ProFit.

La integración, exploración y análisis de la ocurrencia de los PB en determinadas regiones estructurales requirió de la escritura de numerosos programas computacionales ad hoc, que se realizaron en el lenguaje de programación C. A continuación, detallamos cada una de las metodologías empleadas.

1.3.2 Breve resumen de niveles estructurales en proteínas

Recordemos brevemente, para facilitar el posterior análisis, los distintos niveles estructurales de las proteínas **FIGURA I.3**:

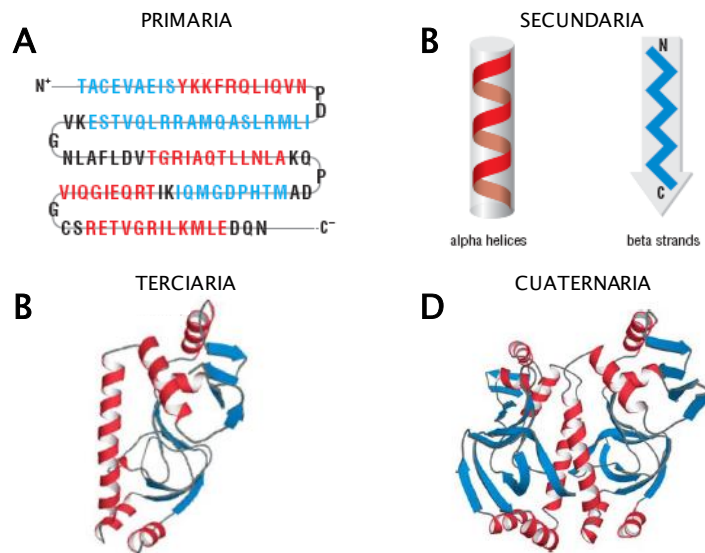


FIGURA I.3: Niveles estructurales en proteínas: (A) estructura primaria, (B) estructura secundaria, (C) estructura terciaria y (D) estructura cuaternaria.

La estructura primaria de una proteína implica la secuencia ordenada de aminoácidos, desde el extremo amino al carboxilo terminal. La estructura secundaria es un ordenamiento local estabilizado por interacciones de puentes de hidrógeno. La estructura terciaria es el ordenamiento espacial de los elementos de estructura secundaria, mientras que finalmente la estructura cuaternaria resulta de la interacción y disposición espacial de distintas cadenas proteicas (secuencialmente iguales o distintas).

En este trabajo utilizamos dominios proteicos. Un dominio estructural es un elemento de la estructura de las proteínas que se autoestabiliza y a menudo estabiliza a los motivos conformacionales independientemente del resto de la cadena de proteína. Muchos dominios son únicos y proceden de una secuencia única de un gen o una familia génica, pero en cambio otros aparecen en una variedad de proteínas. Los dominios son, a menudo, seleccionados evolutivamente porque poseen una función prominente en la biología de la proteína a la que pertenecen; por ejemplo, "el dominio de unión a calcio de calmodulina". Un motivo en este sentido se refiere a una combinación específica de elementos estructurales secundarios (como los hélice-giro-hélice). Estos elementos son llamados a menudo superestructuras secundarias.

Suele denominarse motivo conformacional de forma global a un tipo de motivo, como los barriles-beta. La estructura de los motivos a menudo consiste en solo unos pocos elementos, por ejemplo, las hélice-giro-hélice, que sólo tienen tres elementos. Se denota que la “secuencia espacial” es la misma en todas las instancias del motivo. Su orden es bastante irregular dentro del gen subyacente. Los motivos estructurales de la proteína a menudo incluyen giros de longitud variable en estructuras indeterminadas, lo que en efecto crea la plasticidad necesaria para unir dos elementos en el espacio que no están codificados por una secuencia de ADN inmediatamente adyacente en un gen. Se indica también que incluso cuando están codificados los elementos estructurales secundarios de un motivo en el mismo orden en dos genes, la composición cuantitativa de aminoácidos puede variar. Esto no sólo es cierto debido a las complicadas relaciones entre la estructura terciaria y primaria, sino por cuestiones relativas al tamaño. Si bien en la base de datos de levadura hay descritas unas 6.000 proteínas hay muchos menos dominios, motivos estructurales y pliegues. Esto es, en parte, consecuencia de la evolución. Esto significa, por ejemplo, que un dominio de una proteína puede ser trasladado de una a otra, dando así una nueva función a las proteínas. Debido a estos mecanismos, los dominios o motivos estructurales pueden ser comunes a varias familias de proteínas. En la **FIGURA I.3** se muestra un ejemplo de una proteína conteniendo cuatro dominios distintos.

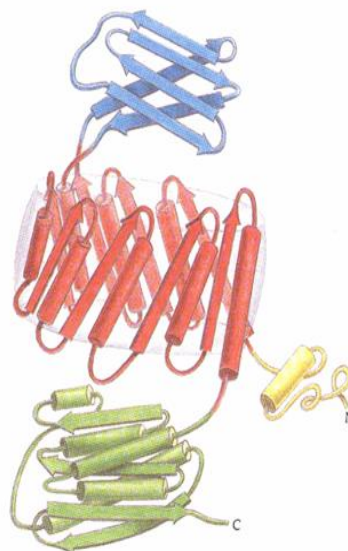


FIGURA I.3: Estructura esquemática de una proteína conteniendo 4 dominios distintos, cada uno representado con un color diferente.

Los dominios son de suma importancia en el estudio de proteínas. Distintas proteínas con distintas funciones biológicas pueden contener el mismo dominio con una determinada actividad biológica en combinación con otros dominios. Por ej. el dominio de unión a la molécula ATP está muy conservado en miles de proteínas con funciones completamente distintas, ya que no es lo mismo tener el dominio de unión de ATP en una proteína quinasa (por ej. para la fosforilación de un determinado sustrato) que en un motor molecular (por ej. para generar trabajo mecánico).

1.3.3 Péptidos bioactivos y bases de datos secuenciales

Mencionamos anteriormente que básicamente la única forma de predecir PB nuevos es por el uso de bases de datos. Estas bases de datos son mayormente secuenciales y contienen los PBs bien caracterizados, por ejemplo, PepBank (<http://pepbank.mgh.harvard.edu/>; Shtatland y col., 2007), Antimicrobial Peptide Database (APD) (<http://aps.unmc.edu/AP/main.php>; Wang y col., 2016) y BIOPEP (<http://www.uwm.edu.pl/biochemia/index.php/pl/biopep>; Minkiewicz y col., 2008). Particularmente, BIOPEP contiene aproximadamente 3000 péptidos y se focaliza en aquellos derivados de proteínas dietarias. Esta base contiene información sobre las proteínas precursoras de los péptidos incluidos en la misma. La ocurrencia de péptidos con diferentes funciones a las de la proteína donde se encuentran encriptados (crypeptides) no es un tópico de interés exclusivo en proteínas alimentarias (Pimenta y Lebrun, 2007). Moughan y col. (2013) han sugerido que la presencia de PB en proteínas alimentarias juega un rol secundario en la regulación de la función gastrointestinal y otras funciones fisiológicas y que los péptidos endógenos generados a partir de proteínas secretadas en el tracto digestivo jugarían un rol importante como fuente de péptidos bioactivos.

En particular en este trabajo se utilizó la base de datos BIOPEP (Minkiewicz y col., 2008) ya que se encuentra focalizada en la presencia de PB provenientes de proteínas con importancia alimentaria. Utilizando esta base de datos se obtuvieron ID, nombre, masa, IC₅₀, actividad y la secuencia de todos los péptidos bioactivos almacenados en ella.

1.3.4 Búsqueda secuencial de PB en la base de datos *protein not-redundat*

Se realizó una búsqueda secuencial de los PB mayores con más de 5 residuos mediante programas propios en lenguaje C en las secuencias incluidas en la versión descargable de la base de datos de proteínas no redundante (nr) de NCBI en mayo de 2011. La base de datos de proteínas de NCBI contiene registros de texto para las secuencias de proteínas individuales derivadas de una variedad de fuentes, incluyendo GenBank, el proyecto de Referencia NCBI de Referencia (RefSeq) y varias bases de datos externas incluyendo UniProtKB/SWISS-Prot y Protein Data Bank (PDB). Se debe aclarar que las secuencias contenidas en muchos de estos registros de proteínas (con la excepción de los registros de PDB) son traducciones conceptuales de una secuencia codificadora de ARN, lo que significa que nadie determinó experimentalmente la secuencia proteica, sino que fue inferida de la secuencia del ARN correspondiente.

El programa de búsqueda registró solo las secuencias que presentaron ocurrencias exactas y completas de al menos un péptido en su secuencia.

A efectos de reducir la redundancia de las secuencias obtenidas se usó el Servidor CD-HIT (Huan y col., 2010) con un valor de corte de 100% de identidad. De esta forma las múltiples secuencias idénticas fueron descartadas. Esto se llevó a cabo para bajar los costos computacionales en la realización de la asignación de plegamiento para cada una de las proteínas encontradas.

1.3.5 Asignación de dominios CATH (asignación de plegamiento)

La asignación de plegamiento (de dominios CATH) para cada una de las secuencias que contienen al menos un PB se realizó mediante el algoritmo BLAST (Altschul y col., 1997)(Basic Local Alignment Search Tool) contra la base de datos CATH (Orengo y col., 2003). Se empleó el programa blastp en su versión ejecutable de línea de comando (versión 2.2.18) empleando la base de datos CathDomainSeqs.ATOM con un valor de $E=1 \times 10^{-9}$ y el resto de las opciones en su valor predeterminado.

Para cada secuencia se seleccionaron los mejores dominios CATH (P-values $< 1 \times 10^{-3}$) que corresponden a segmentos no superpuestos. En el caso que la proteína presentara más de un dominio estructural solo se registraron aquellos dominios asignados a la región que contiene el péptido bioactivo. La base de datos CATH clasifica a las distintas estructuras de las proteínas utilizando 4 categorías Clase, Arquitectura, Topología y Homólogos que caracterizan una superfamilia estructural dada como se muestra en la **FIGURA I.5**.

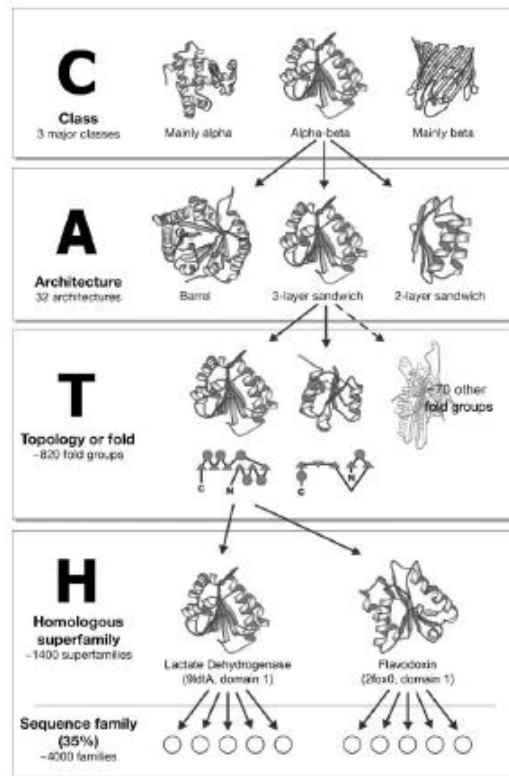


FIGURA 1.5: Esquema de la clasificación CATH (adaptado de Orengo y col., 2003)

Por ejemplo, CATH posee en la actualidad 4 clases identificadas con números: 1 dominios que contienen todas hélices alfa, 2, dominios que contienen beta plegadas, 3 para dominios que combinan hélices alfa con beta plegadas y finalmente la clase 4, para dominios con escaso contenido de estructuras secundarias.

Cada superfamilia de homólogos, va a estar caracterizadas por 4 números idénticos, indicando que esa superfamilia comparte la clase, la arquitectura y la topología, siendo el cuarto número el que identifica a esa determinada superfamilia. Por ejemplo 1.10.150.110 es el código para la superfamilia de la DNA polimerasa. Todos los dominios que compartan estos 4 números (clase, arquitectura, topología y familia de homólogos) pertenece a la superfamilia de la DNA polimerasa. Los dominios contenidos en esta superfamilia se pueden clasificar por su similitud secuencial. Si los dominios se parecen en más del 35, 60, 95 o 100% de identidad secuencial se agrupan en familias de proteínas homólogas dando lugar a la asignación de 4 nuevos números en la clasificación jerárquica. Por ej. la superfamilia de la DNA polimerasa contiene 7 clusters (1.10.150.110.1 al 1.10.150.110.7) de familias de homólogas que comparten más del 35% identidad. Este clustering llega hasta la identificación de cada dominio en particular, de una determinada estructura PDB, por ejemplo, el

número 1.10.150.110.2.1.1.1.42 indica al dominio PDB 4DOC que representa a la estructura de la subunidad beta de la DNA polimerasa humana.

Finalmente, CATH otorga una determinada nomenclatura para los dominios basándose en el código PDB. Así a la cadena A de la DNA polimerasa con código PDB 4DOC se la puede dividir en 4 dominios estructurales 4DOCA01, 4DOCA02, 4DOCA03 y 4DOCA04. Cada uno de estos dominios tendrá una combinación de 9 números correspondiente a su clasificación estructural y secuencial como vimos antes.

1.3.6 Localización estructural de PBs

Para aquellas secuencias en las que fue posible asignar un plegamiento dado, se realizó un mapeo estructural de los péptidos sobre la estructura representativa de la superfamilia de CATH asignada, haciendo uso de un procedimiento de dos pasos (FIGURA I.6).

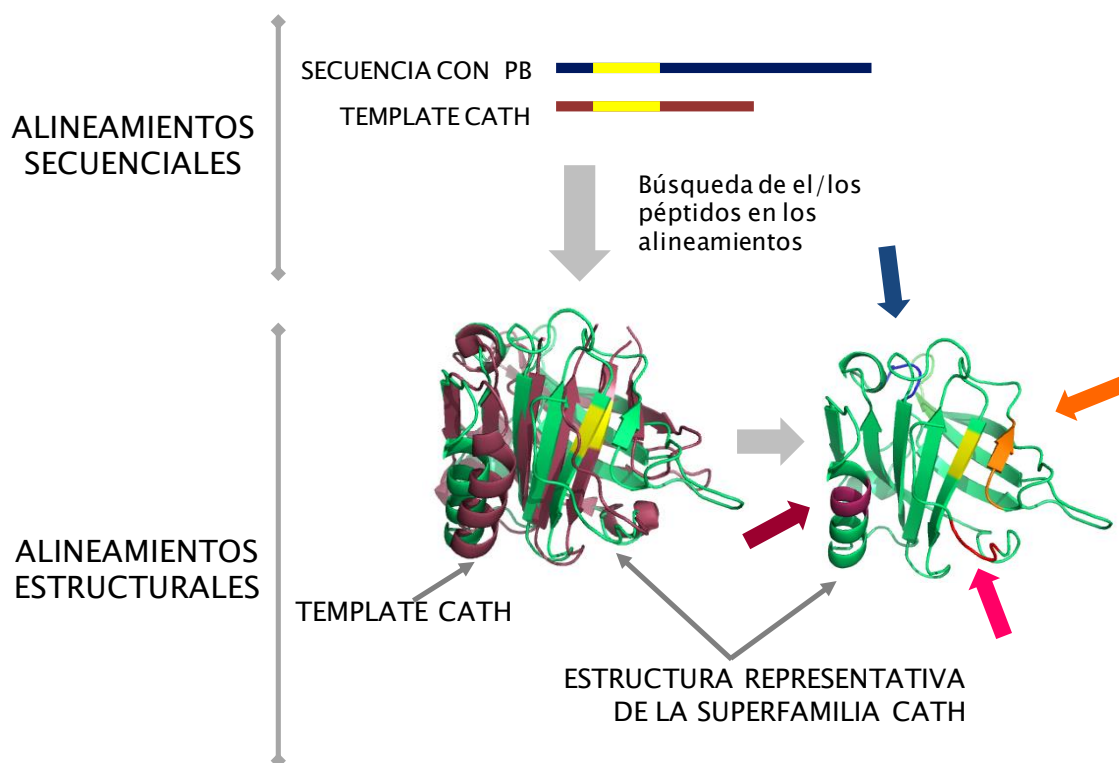


FIGURA I.6: Esquema de la metodología de aplicada para el mapeo de péptidos en las estructuras representativas de las superfamilias CATH.

En primer lugar, se alineó cada secuencia con la secuencia del modelo (denominadas template) asignado en la búsqueda de plegamientos efectuada (dominio CATH), mediante el programa ClustalX (Higgins, y col., 1992). Este paso se realizó para ubicar los posibles elementos estructurales provenientes

del template CATH en la secuencia problema. Posteriormente se obtuvieron los alineamientos estructurales del dominio CATH con el plegamiento de la proteína que se toma como representante de la superfamilia estructural en cuestión, haciendo uso del algoritmo de McLachlan (McLachlan, 1982) como se implementa en el programa ProFit (Martin y Porter, n.d.; versión 3.1)

Haciendo uso de este segundo paso, se pudieron asignar los péptidos en un único plegamiento proteico que representa a diferentes proteínas de la misma familia homóloga. Este segundo paso nos permitió acumular información en un determinado plegamiento que es compartido por toda una superfamilia estructural.

1.3.7 Otras herramientas

Hemos utilizado los términos GO para clasificar las funciones biológicas de los dominios encontrados en el análisis secuencial de las proteínas conteniendo PBs. El proyecto GO es un consorcio que se ocupa de establecer ontologías para describir los productos génicos en términos de tres componentes: funciones moleculares, procesos biológicos en los cuales participan y los componentes celulares en los que ocurren. El consorcio se encuentra disponible en <http://www.geneontology.org/> y la gran mayoría de las bases moleculares, secuenciales y estructurales han asociado sus entradas con los respectivos términos GO.

Hemos utilizado también, la base de datos FunTree (<http://www.funtree.info/FunTree/>; Furnham y col., 2012) que utiliza la clasificación estructural de CATH para proveer información secuencial, estructural y filogenética sobre la función biológica

Finalmente, el análisis y exploración de los datos se realizaron utilizando programas ad hoc escritos en lenguaje de programación C. También se utilizaron scripts escritos en bash. Los análisis estadísticos se realizaron utilizando el programa R.

1.4 RESULTADOS Y DISCUSIÓN

1.4.1 Distribución de péptidos en BIOPEP

Los avances en la ciencia de péptidos bioactivos derivados de proteínas alimentarias han requerido de la creación de bases de datos donde agrupar la información asociada a secuencias bioactivas. Para la realización del análisis de ocurrencias de péptidos bioactivos (PB) en el universo de proteínas de secuencias conocidas se trabajó con las secuencias de PB almacenadas en bases de datos, en particular BIOPEP. En mayo de 2011 se encontraron registrados 2.595 péptidos agrupados en 40 actividades fisiológicas diferentes (detalle en **TABLA I.1**). En BIOPEP la clasificación de las actividades es provista por los autores, y resulta en algunos casos muy general o ambigua. De esta forma, se reportan péptidos como inhibidores, reactivos o acelerantes sin una etiqueta rápida del detalle del proceso biológico del cual participan. En algunos casos la actividad esta detallada en el nombre del péptido y en otros se debe recurrir directamente a la cita bibliográfica. Por ejemplo, la actividad estimulante agrupa péptidos estimulantes de la captación de glucosa y estimulación de la liberación de sustancias vasoactivas. Sin embargo, con fines prácticos se respetó la clasificación de actividades utilizada en BIOPEP. La lista completa de péptidos se muestra en el **ANEXO I**.

Las actividades con mayor número de representantes resultaron ser el inhibidor de la enzima convertidora de angiotensina (ACE) (556), con actividad antibacteriana (449), antioxidante (392), toxina celíaca (247), opioide (139) y neuropéptido (108). Esta distribución refleja las líneas de investigación principales en PBs. Los primeros reportes de péptidos encriptados en proteínas alimentarias corresponden a péptidos opioides (presentan *in vitro* actividad

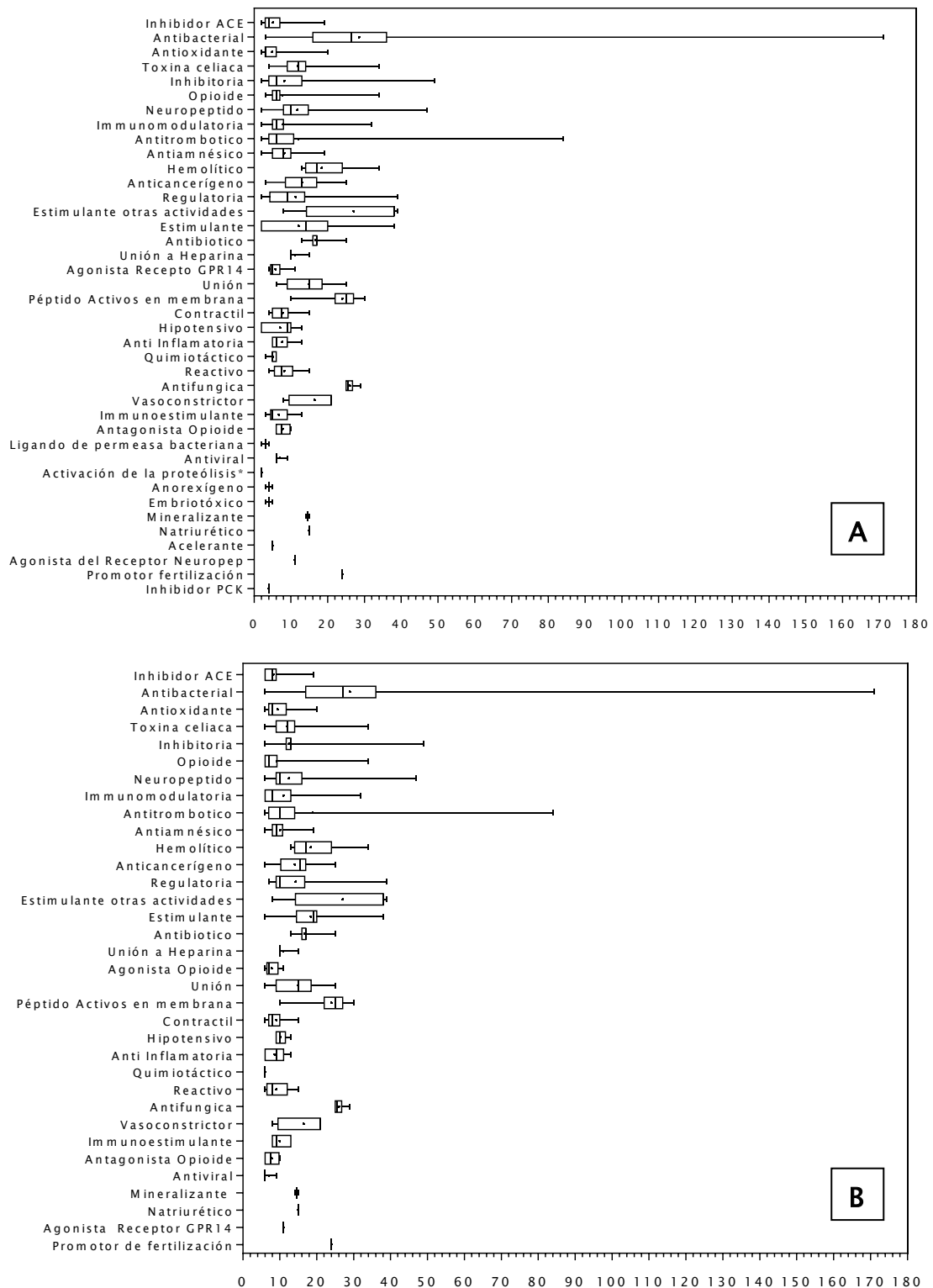
similar a la morfina) derivados de proteínas de leche vacuna (Brantly col., 1979). Resulta entonces esperable que se trate de una de las actividades con un mayor número de secuencias identificadas. Actualmente los péptidos inhibidores de ACE son los más estudiados (Puchalska y col., 2015). La longitud promedio (expresada en número de residuos) depende fuertemente de la actividad fisiológica con la cual se lo asocie (**FIGURA I.7**). La distribución está enriquecida en di y tripéptidos.

Una de las conclusiones que salta a la vista al evaluar las distribuciones de longitudes de PBs y de las distintas actividades fisiológicas que llevan a cabo, es que evidentemente la actividad biológica puede residir en un número de aminoácidos muchas veces menor que la correspondiente a la longitud del péptido informado en BIOPEP. Esto se hace evidente, por ejemplo, en péptidos con actividad antibacteriana, donde la longitud mínima es de 5 aminoácidos y llega hasta 171 residuos. Creemos que esta diversidad de longitudes se debe a la aún escasa información sobre el conjunto mínimo de aminoácidos para sustentar la actividad biológica de los PBs.

Por otra parte, por tratarse de una base de datos de péptidos derivados de proteínas alimentarias, en la mayoría de los casos se detalla el alimento del cual se purificó el PB (debiendo el usuario recurrir a el trabajo/cita del correspondiente PB) y no así a la identidad de la proteína que lo contiene. Esto se debe a que los alimentos son matrices complejas, pudiendo ser tanto de origen animal como vegetal sumado al hecho de que no se han secuenciado aún muchas de las proteínas presentes en ellos.

TABLA I.1 Distribución según actividad de (A) PB totales y (B) mayores a cinco aminoácidos detallada en BIOPEP (<http://www.uwm.edu.pl/biochemia>, mayo de 2011).

Actividad	A	B
	N Total	N Total
Antibacteriano	449	439
Toxina celíaca	247	242
Inhibidor de ACE	556	209
Antioxidante	392	108
Neuropéptido	108	100
Opioide	139	98
Inhibitoria	147	74
Hemolítica	47	47
Anticancerígena	45	40
Immunomoduladora	75	39
Antiamnésico	52	36
Estimulante de otras actividades	36	36
Regulatoria	44	32
Antitrombótica	56	30
Antibiótico	25	25
Unión a Heparina	20	20
Estimulante	33	20
Unión	13	13
Péptidos activos en membrana	12	12
Contráctil	10	7
Antifúngica	6	6
Immunoestimulante	14	6
Antiinflamatoria	7	5
Hipotensiva	8	5
Agonista Opioide	13	5
Reactivo	6	5
Vasoconstrictor	5	5
Antagonista Opioide	4	4
Antiviral	3	3
Quimiotáctica	7	3
Mineralización de tejidos y fluidos corporales (mineralizante)	2	2
Péptido promotor de fertilización	1	1
Natriurética	1	1
Agonista del Receptor Neuropeptídico GPR14	1	1
Acelerante	1	
Activación de la proteólisis mediada por ubiquitina	3	
Anorexígeno	2	
Ligando de permeasa bacteriana	3	
Embriotóxica	2	
Inhibidor de Proteína Kinasa C (CPK)	1	
Total	2596	1679



Teniendo en cuenta este marco y con el objetivo de indagar en que proteínas se encuentran estos péptidos y poder estudiar la relación estructural y/o evolutiva de las proteínas que los contienen y con el objetivo final de indagar el porqué de esta ocurrencia, se seleccionaron los péptidos de más de 5 residuos para realizar una búsqueda secuencial en la base de datos de proteínas no redundante de NCBI. Los péptidos con menor tamaño se descartaron para la búsqueda debido a la elevada probabilidad de ocurrencia al azar en el conjunto de proteínas de secuencia conocida. El nuevo conjunto de PB quedó constituido entonces por 1679 secuencias de PBs agrupadas en 34 actividades diferentes (TABLA I.1 B y FIGURA I.7 B)

1.4.2 Distribución de péptidos en el espacio secuencial de proteínas (NR)

El análisis de ocurrencias exactas de péptidos con más de 5 residuos encriptados se realizó utilizando las secuencias almacenadas en la base de datos de proteínas no redundante de NCBI (<https://www.ncbi.nlm.nih.gov/>). Se registraron 88.909 secuencias que contienen encriptados al menos un PB, con longitudes que abarcan desde 6 a 13.256 residuos y con 1 a 65 ocurrencias exactas de PB en una misma secuencia. El 85,6% (77.003) de las secuencias contiene solo un PB, el 7,9% (6.374) dos, mientras que el 2,0% (1.549), el 1,4% (1.277) y el 0,4% (368) contienen 3, 4 y 5 ocurrencias de péptidos encriptados respectivamente. El número de ocurrencias, no es mayor para secuencias de mayor longitud (FIGURA I.8), resultando de interés aquellas secuencias que contienen el mayor número de ocurrencias de PB.

1.4.3 Asignación estructural y distribución de secuencias

Con el objeto de caracterizar las secuencias que contienen al menos un péptido y estudiar si existe correlación entre la estructura, tipo de plegamiento o superfamilia estructural y la ocurrencia y la actividad del PB se realizó la asignación de dominios estructurales mediante búsquedas de similitud secuencial con el algoritmo BLAST contra la base de datos de dominios estructurales de CATH como se detalló anteriormente (Sección 1.3.5 de Materiales y Métodos). Mediante este procedimiento fue posible asignar estructura a 58.167 secuencias (72,2% de la cantidad inicial). En este nuevo *dataset* sólo se encontraron 260 de los PB (15,5% de los PB mayores a 5).

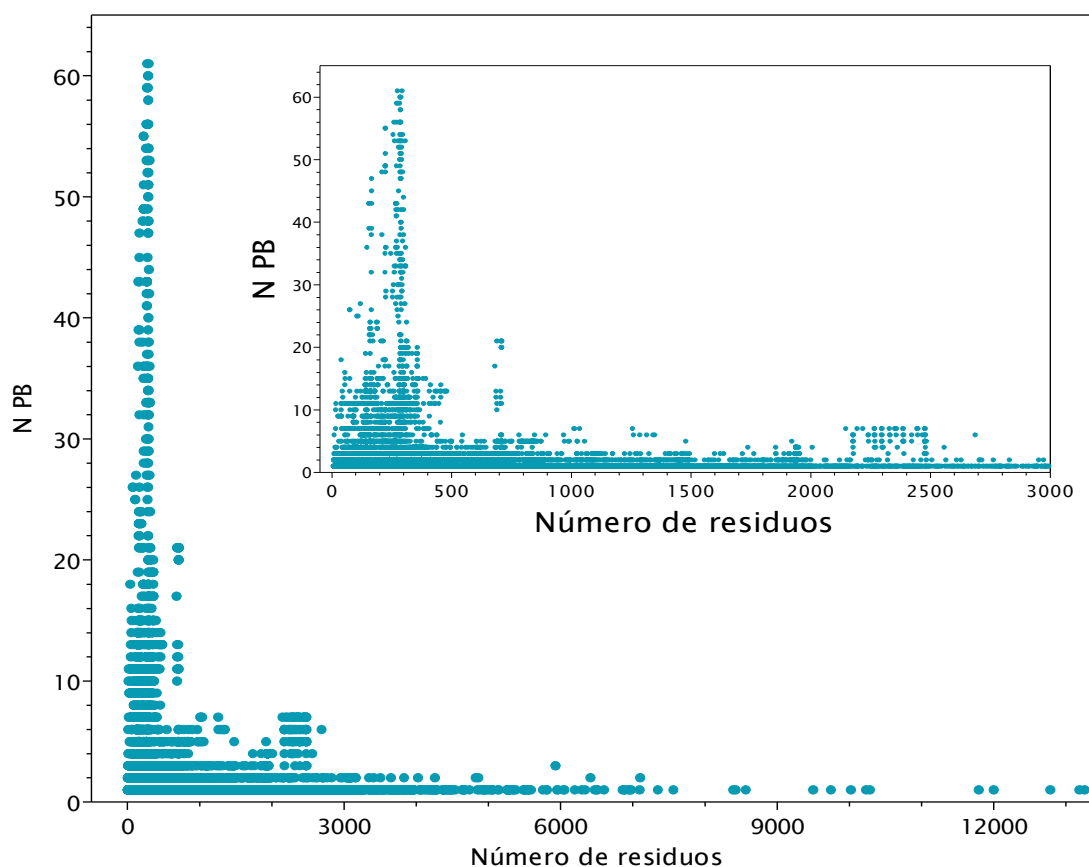


FIGURA I.8 Número de ocurrencias de PB (N PB) en una misma secuencia en función de su longitud (número de residuos). En la parte superior se amplía la región correspondiente a secuencias con longitud entre 6 y 3000 residuos.

Para cada secuencia se seleccionó el dominio CATH asignado (*template*) de mejor score. En el caso de las proteínas a las cuales les fuera asignado más de un dominio CATH, se verificó en que dominio se asignó la región de la proteína que contiene la ocurrencia exacta del/los PB y se descartaron los templates que no cumplieron con esta condición. Las secuencias a las cuales fue posible asignar estructura se agruparon en 333 superfamilias de dominios según CATH (FIGURA I.9 y TABLA I.2).

Resulta interesante destacar que cuatro plegamientos agrupan el 87,54% de las secuencias. La familia más abundante agrupó el 33,6% de las secuencias y se identifica con el CATH Id 3.30.70.150, que representa el dominio “ $\alpha\beta$ -2-layers sándwich”. Este plegamiento es representado por la subunidad grande de la ribulosa 1,5-bifosfato carboxilasa oxigenasa (RuBisCO) (EC 4.1.1.39) y la mayoría de las secuencias pertenecientes a este grupo resultaron proteínas homólogas cercanas de la RuBisCO. Un porcentaje menor de proteínas de este

grupo es representado por un dominio cercano a la RuBisCO, el 2,3-diceto-5-metiltiofentil-1-fosfato enolasa (DK-MTP-1-P enolasa) (EC 5.3.2.5) también llamado proteínas tipo RuBisCO. El 96% de las proteínas de este grupo pertenecen al reino de las plantas y el resto al de las bacterias principalmente cianobacterias. El cluster que sigue en abundancia es el correspondiente a “mainly alpha, orthogonal bundle” (CATH Id 1.10.287.210) el cual agrupó el 27,5 % de las secuencias. La mayoría de las secuencias de este grupo pertenecen a proteínas de virus y son representadas por la glucoproteína de envoltura gp160 (GP160). El plegamiento “mainly beta with a complex beta structure” (CATH Id 2.170.40.20) agrupó el 21,3% de las secuencias del dataset. Estas proteínas son también pertenecientes a la familia de las proteínas llamadas “envelope glycoprotein gp160” (envpolyprotein). Finalmente, la familia representada por la beta-actina y proteínas relacionadas con la actina (CATH Id 3.30.420.40) fue asignada al 14 % de las secuencias del a las cuales se asignó estructura. En la **TABLA I.3** se muestra la clasificación de las primeras 10 superfamilias con mayor número de secuencias.

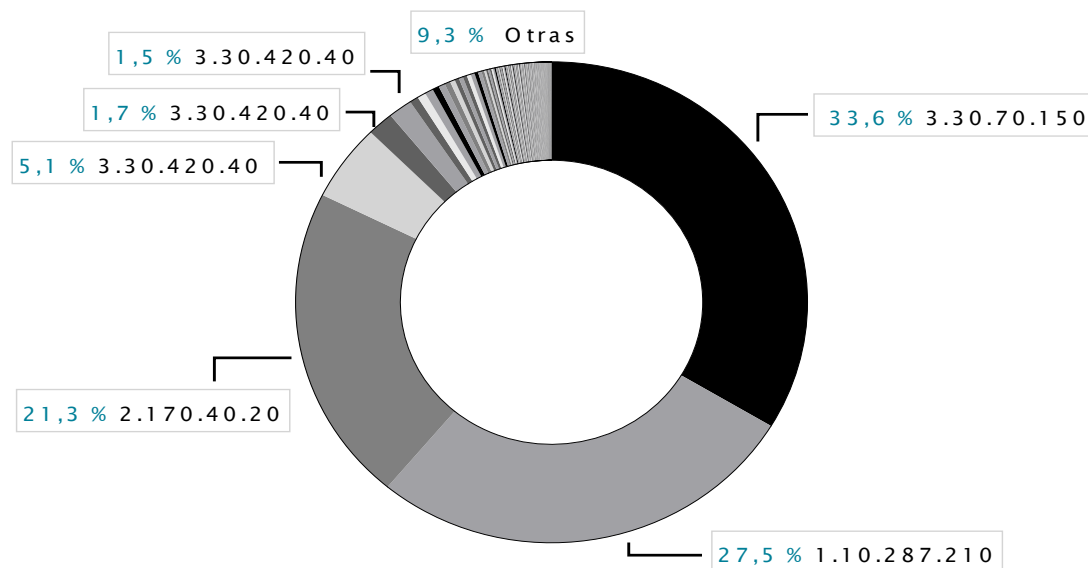


FIGURA I.9. Distribución de los templates en superfamilias.

TABLA I.2. Descripción de las primeras diez superfamilias con mayor cantidad de representantes en distribución de templates de CATH.

%	CATH id	Dominio representativo	C	A	T	Proteína representativa
33,6	3.30.70.150	1wddA01	Alfa Beta	2-Layer Sandwich	Alfa-Beta Plaits	Rubisco subunidad grande
27,5	1.10.287.210	1qbzA00	Principalmente Alfa	OrthogonalBundle	HelixHairpins	Envelope glicoproteína gp160
21,3	2.170.40.20	1kmoA02	Principalmente Beta	Beta Barrel	Maltoporina; Cadena A	HIV Envelope Proteina Gp120; Cadena G
5,1	3.30.420.40	3i33A01	Alfa Beta	2-Layer Sandwich	Nucleotidyl transferasa; dominio 5	
1,5	1.10.490.10	2nrIA00	Principalmente Alfa	OrthogonalBundle	Tipo Globina	
1,7	2.30.36.70	1s22A02	Principalmente Beta	Roll	Actina; Cadena A, dominio 2	Actina; Cadena A, dominio 3
0,6	3.40.190.10	1ixhA01	Alfa Beta	3-Layer(aba) Sandwich	Proteína de unión D-Maltodextrina; dominio 2	Poteína de unión Periplasmatica-tipo II
0,5	2.60.40.720	2ioiA00	Principalmente Beta	Sandwich	Tipo Immunoglobulina	
0,5	3.40.50.720	1c0pA01	Alfa Beta	3-Layer(aba) Sandwich	Plegamiento de Rossmann	Dominio de unión NAD(P)-tipo Rossmann
0,4	2.60.120.10	1juhA02	Principalmente Beta	Sandwich	JellyRolls	JellyRolls

Tabla I.3 Distribución de los péptidos ≥ 5 aminoácidos en superfamilias según la clasificación de CATH. No se muestran las superfamilias que solo presentan un péptido.

CATH id	Proteínas	%	PB
3.40.50.300	215	0,370	35
3.40.50.720	309	0,531	25
3.20.20.70	35	0,060	18
2.60.120.10	246	0,423	17
3.40.50.1820	22	0,038	17
3.20.20.80	126	0,217	13
3.40.50.150	144	0,248	12
3.50.50.60	56	0,096	10
3.90.226.10	237	0,407	9
1.10.630.10	18	0,031	9
3.40.190.10	322	0,554	8
2.60.40.10	161	0,277	8
1.10.530.10	151	0,260	8
3.40.50.200	53	0,091	8
3.40.50.2300	41	0,070	8
3.40.640.10	28	0,048	8
3.40.50.620	12	0,021	8
3.30.420.40	2989	5,139	7
3.40.605.10	12	0,021	7
2.40.50.140	8	0,014	7
3.20.20.30	6	0,010	7
3.80.10.10	39	0,067	6
1.10.510.10	33	0,057	6

3.30.930.10	20	0,034	6
3.40.30.10	12	0,021	6
1.10.490.10	884	1,520	5
2.80.10.50	148	0,254	5
3.30.200.20	36	0,062	5
2.40.128.20	30	0,052	5
3.20.20.140	30	0,052	5
3.40.50.880	13	0,022	5
3.40.1190.20	10	0,017	5
3.40.630.10	10	0,017	5
3.40.50.80	8	0,014	5
3.40.710.10	7	0,012	5
3.30.420.10	5	0,009	5
1.10.565.10	39	0,067	4
2.30.39.10	32	0,055	4
1.20.1070.10	31	0,053	4
3.60.21.10	29	0,050	4
3.40.50.1970	13	0,022	4
2.40.10.10	11	0,019	4
3.40.309.10	11	0,019	4
3.30.565.10	10	0,017	4
3.40.50.970	10	0,017	4
1.10.3720.10	9	0,015	4
3.10.310.10	9	0,015	4

3.90.1300.10	9	0,015	4
1.20.1560.10	7	0,012	4
2.40.170.20	7	0,012	4
3.20.20.60	5	0,009	4
1.10.8.60	4	0,007	4
3.40.47.10	4	0,007	4
3.60.15.10	4	0,007	4
4.10.230.10	63	0,108	3
1.10.238.10	55	0,095	3
3.10.250.10	39	0,067	3
3.10.50.40	31	0,053	3
3.30.365.10	29	0,050	3
3.40.50.2020	26	0,045	3
3.40.980.10	17	0,029	3
3.90.550.10	15	0,026	3
1.10.10.10	14	0,024	3
3.10.450.10	12	0,021	3
3.40.50.1240	12	0,021	3
3.20.20.300	8	0,014	3
3.40.1190.10	8	0,014	3
3.30.450.20	7	0,012	3
3.30.470.20	7	0,012	3
3.90.1150.10	7	0,012	3
2.130.10.10	6	0,010	3

4.10.70.10	6	0,010	3
3.40.109.10	5	0,009	3
3.90.180.10	5	0,009	3
3.40.50.170	4	0,007	3
3.60.20.10	4	0,007	3
3.90.190.10	4	0,007	3
1.10.760.10	3	0,005	3
1.25.40.10	3	0,005	3
3.40.1280.10	3	0,005	3
3.40.50.1360	3	0,005	3
3.40.50.1980	3	0,005	3
2.60.120.20	1	0,002	3
3.20.20.10	1	0,002	3
2.170.40.20	12388	21,297	2
2.60.40.720	313	0,538	2
3.30.1340.10	76	0,131	2
3.50.7.10	65	0,112	2
1.10.246.10	46	0,079	2
3.30.1330.10	32	0,055	2
3.30.360.10	25	0,043	2
1.20.5.110	24	0,041	2
1.20.58.70	24	0,041	2

1.50.40.10	21	0,036	2
2.60.120.200	19	0,033	2
3.20.20.120	17	0,029	2
3.40.50.1100	16	0,028	2
3.40.810.20	14	0,024	2
3.30.70.270	13	0,022	2
3.10.105.10	11	0,019	2
3.90.870.10	11	0,019	2
3.30.450.90	10	0,017	2
1.25.10.10	9	0,015	2
3.40.850.10	9	0,015	2
3.30.497.10	8	0,014	2
1.10.3080.10	6	0,010	2
1.20.1050.10	6	0,010	2
1.10.560.10	5	0,009	2
1.10.600.10	4	0,007	2
3.20.20.100	4	0,007	2
3.30.1060.10	4	0,007	2
3.40.50.10540	4	0,007	2
3.90.245.10	4	0,007	2
2.30.110.10	3	0,005	2
3.10.129.10	3	0,005	2

3.20.20.330	3	0,005	2
3.30.300.20	3	0,005	2
3.40.1030.10	3	0,005	2
3.40.50.2000	3	0,005	2
3.40.50.980	3	0,005	2
3.40.630.20	3	0,005	2
3.60.40.10	3	0,005	2
1.10.290.10	2	0,003	2
1.10.3470.10	2	0,003	2
1.10.520.10	2	0,003	2
1.20.120.80	2	0,003	2
1.20.210.10	2	0,003	2
1.20.810.10	2	0,003	2
2.40.30.10	2	0,003	2
2.60.40.820	2	0,003	2
2.70.98.10	2	0,003	2
3.20.20.240	2	0,003	2
3.20.20.370	2	0,003	2
3.30.1360.30	2	0,003	2
3.30.300.30	2	0,003	2
2.50.20.10	1	0,002	2

TABLA I.4 Actividades y secuencia de los péptidos ≥ 5 aminoácidos en las primeras cuatro superfamilias con mayor cantidad de PBs. P indica el número de proteínas distintas en cada superfamilia y N el número de PBs encontrados en cada superfamilia. Los PB son clasificados de acuerdo a la actividad detallada en BIOPEP.

CATH	P	N	PÉPTIDOS IDENTIFICADOS	ACTIVIDAD
3.40.50.300	215	35	GKKVLQ; DYGLYP; PQEVLP; KVLILA; RADHPF; LAHKAL; KVLAGM; GLDIQK; VTSTAV; IKPLNY; FQKVVA; VPQPIP; VIEKYP; KVREGT; VHLPPP;	Inhibidor de ACE
			LKKISQ	Inhibidor de ACE /Antibacteriano
			RRPYIL; KIPYIL	Anti Inflamatoria
			HLPLPL	Anti amnésico
			STVATL; ALCSEK	Antibacteriano
			IEAEGE; LLPHHH; QYDQGV; GALAAH	Antioxidante
			LGTIPG; VGVAPG	Quimiotáctico
			SIKVAV	Immunomodulador/ Anticancerígeno
			VGGIPY	Immunomodulador
			FFGLMG; KRQHPG	Neuropéptido
			YLGYLE; RYLGYL; GGFLTRH	Opioide
SFLLRN	Estimulante			
3.40.50.720	309	25	VTSTAV; PANIKWGD; PSKIKWGD; QSLVYP; PQEVLP; PANLPWGSSNV; VLAQYK; VIEKYP; LAHKAL; EPKAIP; KVLAGM; GLDIQK	Inhibidor de ACE
			LKKISQ	Inhibidor de ACE /Antibacteriano
			ALCSEK	Antibacteriano
			GALAAH; LGFEYY; PKAVHE; ISELGW	Antioxidante
			RGDSPA	Antitrombótica
			VGVAPG; LGTIPG; PGAIPG	Quimiotáctico
			LVCYPQ	Immunomodulador
			SIKVAV	Immunomodulador/ Anticancerígeno
EVQKQLQ	Neuropéptido			
3.20.20.70	35	18	KVLILA; VLPYPV; PQEVLP; GKKVLQ; LAHKAL; KVLVVP; YLYEIA; VTSTAV	Inhibidor de ACE
			LKKISQ	Inhibidor de ACE /Antibacteriano
			STVATL	Antibacteriano
			YFYPEL; GALAAH; PKAVHE; EELDNALN; YGYTGA	Antioxidante
			PGAIPG; VGVAPG	Quimiotáctico
			SIKVAV	Immunomodulador/ Anticancerígeno
2.60.120.10	246	17	KVLILA; NWGPLV; YLAGNQ; YQEPVL	Inhibidor de ACE
			LSPFWNINA	Anti amnésico
			STVATL	Antibacteriano
			AIRQGDVF; KHNRGDEF; LLPHHADADY; LVNPHDHQN; VIPAGYP; VLEANPRSF; YFPVGGDRPESF	Antioxidante
			RGDSPA	Antitrombótico
			HCQRPR	Immunomodulador
			EITPEKNPQLR; VAWWMY	Inhibidor

1.4.4. Distribución de PBs en las superfamilias estructurales

Las distribuciones de tipos estructurales y abundancia relativa de secuencias por superfamilia estructural, cambian completamente cuando estudiamos el número de PBs por cada superfamilia estructural. Las superfamilias estructurales con CATH Id 3.30.70.150, 2.170.40.20 y 1.10.287.210 mencionadas arriba como las superfamilias estructurales con mayor número de representantes secuenciales en nuestro dataset, solo poseen en sus secuencias 1 PB, mientras que la superfamilia 3.30.420.40 posee 7 PBs. Al estudiar esta nueva distribución de los distintos tipos de PBs por superfamilia estructural, nos encontramos que las distintas superfamilias pueden tener varios PBs iguales o con distintas propiedades biológicas. En la **TABLA I.4** se detalla la información del número de proteínas y PBs encontrados para las superfamilias con mayor número de ocurrencias de PB. Así podemos ver por ejemplo que la superfamilia 3.40.50.300 (P-loop containing nucleotide triphosphate hydrolases superfamily), contiene 35 PBs distintos con 12 actividades biológicas diferentes para la misma superfamilia (o plegamiento). Simplemente por poner un límite, separamos al total de superfamilias en dos, considerando arbitrariamente el número 5 PBs como límite. Así encontramos que las superfamilias con más de 5 PBs sólo representan el 11,5% del total de proteínas estudiadas. En este punto nos preguntamos si las proteínas con más de 5 PBs tienen alguna particularidad funcional. Utilizando el número de términos GO asociados a cada superfamilia (ver sección 1.3.7 de Materiales y Métodos), que describen la función molecular, el proceso biológico en donde participa la proteína y el componente celular al que pertenece, encontramos que el grupo de superfamilias con más de 5 PBs son funcionalmente más diversas que las que tienen menos de 5 PBs (**FIGURA I.10, A-C**). El mismo tipo de resultado se obtiene al analizar la base de datos FunTree (Furnham y col., 2012). Esta base de datos deriva de la anotación funcional de CATH (<http://cpmb.lshhtm.ac.uk/FunTree/index.php>). Al utilizar el número de clusters funcionales en FunTree mapeando las superfamilias con más de 5 PBs encontramos los mismos resultados que utilizando GO terms (**FIGURA I.10, D**). Finalmente encontramos también que las superfamilias con más de 5 PBs son secuencialmente y estructuralmente más diversas que el grupo con menor número de PB. Pero interesantemente, además son secuencialmente y estructuralmente más diversas que todas las superfamilias de CATH tengan o no PBs.

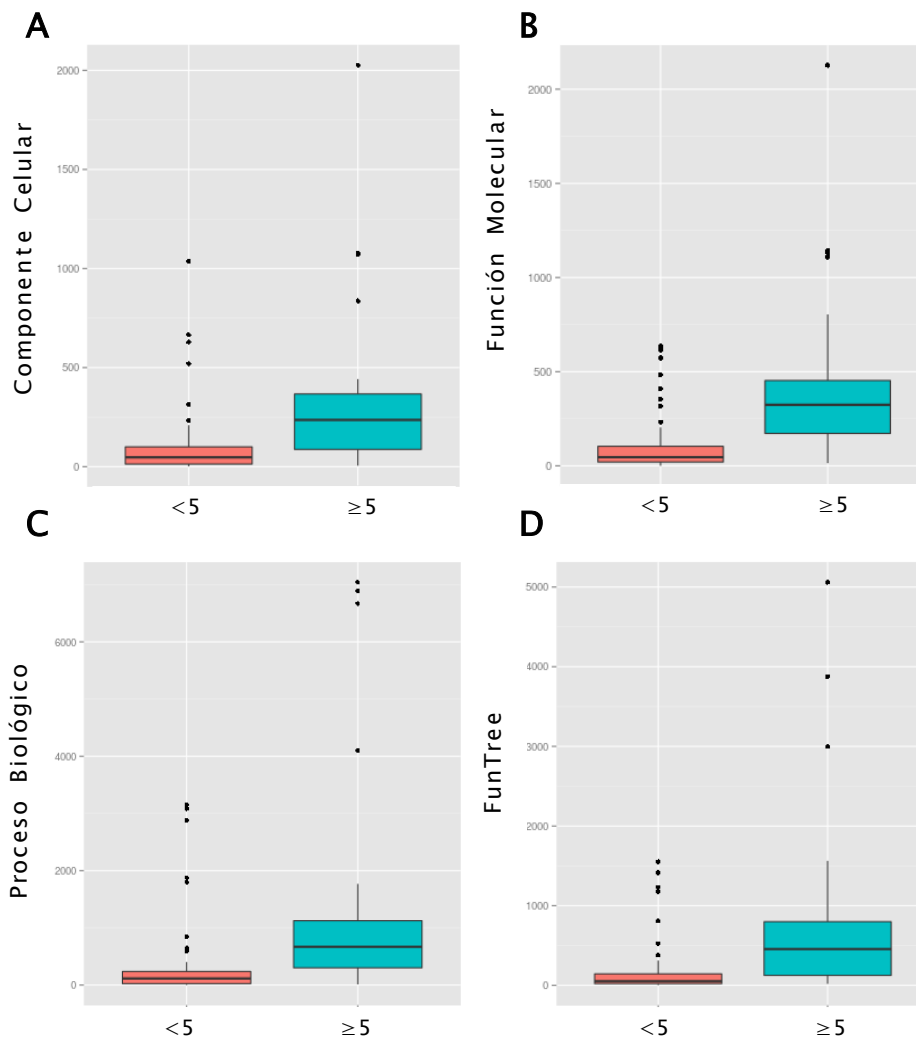


FIGURA I.10 Distribuciones de términos GO para las tres categorías: **(A)** componente celular, **(B)** función molecular y **(C)** proceso biológico. En el panel **(D)** se muestran las distribuciones utilizando el número de clusters funcionales derivado de la base de datos FunTree. Todas las distribuciones son estadísticamente significativas con un P-value < 0.005.

En la **FIGURA I.11** se muestra el número de clusters al 35% identidad vs el número de clusters estructurales para cada superfamilia. En puntos rojos están indicadas las superfamilias con más de 5 PBs. En general se puede apreciar que las superfamilias con más de 5 PBs tienden a ser mucho más diversas. Sólo unas pocas familias (como las superfamilias 3.40.50.720 y la 2.60.40.10) tienen una enorme cantidad de familias de homólogas a nivel del 35%, pero sólo contienen 1 cluster estructural. Se trata de los denominados superplegamientos (Orengo y col., 1994), plegamientos con una enorme distribución funcional y filogenética pero con una alta conservación estructural. Un ejemplo de estos tipos de superplegamientos lo constituye el plegamiento de las inmunoglobulinas o el TIM barrel (triosa fosfato isomerasa).

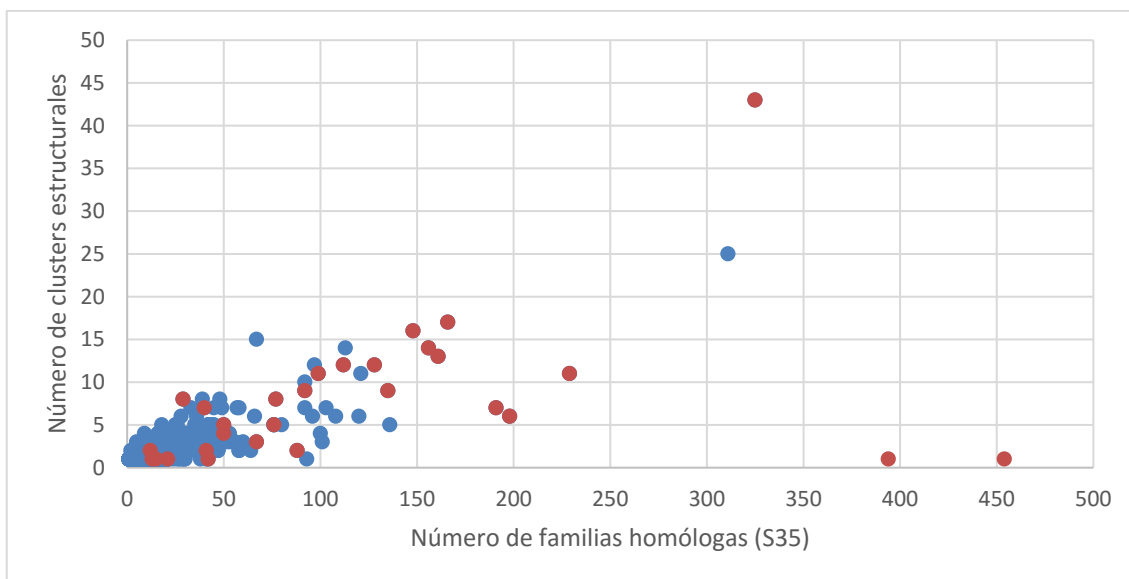


FIGURA I.11: Número de familias de homólogos al nivel 35% identidad (S35) (diversidad secuencial) vs número de clusters estructurales (diversidad estructural) para todas las superfamilias contenidas en CATH (puntos azules) mientras que las superfamilias que tienen más de 5 PBs se encuentran coloreados en rojo.

1.4.5 Localización estructural de PB

Para continuar con los análisis se seleccionaron las proteínas de las superfamilias que presentaron mayor ocurrencia de PB. A fin de conocer la localización de los PB en la estructura de las proteínas que los contienen se realizó un procedimiento de dos etapas tal como se detalló en Materiales y Métodos. Para cada superfamilia se realizó el alineamiento secuencial entre cada una de las proteínas del dataset 2 con el dominio representativo (template) de CATH asignado en la asignación de estructura, y, posteriormente se localizaron/mapearon los péptidos en el alineamiento obtenido. En una segunda etapa se alineó estructuralmente cada template con la estructura de referencia de su correspondiente superfamilia. De esta forma se logró conocer la localización en la estructura representativa de la superfamilia de todos los péptidos presentes en esa superfamilia. La lógica de este procedimiento es la siguiente: como las estructuras proteicas están más conservadas que las secuencias (Chothia y Lesk, 1986; Illergård y col., 2009) y como todas las proteínas dentro de una superfamilia son homólogas, se espera que todas estas proteínas tengan el mismo plegamiento. Para estudiar cómo se disponen los PBs pertenecientes a una determinada superfamilia, elegimos la estructura representativa (dominio de referencia) para acumular los resultados de la

co-localización de cada proteína en la superfamilia (**FIGURA I.6** de Materiales y Métodos). En la **TABLA I.5** se resumen los resultados obtenidos.

TABLA I.5 Detalle de los péptidos encontrados en los alineamientos estructurales entre los templates de CATH y la estructura del dominio representativo de cada superfamilia.

Superfamilia	Dominio de referencia	Longitud del dominio de referencia	Número de templates	Péptidos encontrados originalmente en la superfamilia	Péptidos encontrados en los alineamientos template- dominio representativo
3.40.50.300	1BYIA00	244	215	35	21
3.40.50.720	1C0PA01	199	309	25	15
3.20.20.70	2VXNA00	249	35	18	15
2.60.120.10	1JUHA02	145	246	17	15
3.40.50.1820	1YS1X00	320	22	17	9
3.20.20.80	1ITXA01	345	126	13	7
3.50.50.60	1D5TA01	175	56	10	3
3.90.226.10	1SG4A01	203	237	9	8
1.10.630.10	1JFBA00	399	18	9	7

En las **FIGURA 1.12** se muestra la co-localización de distintos PBs en la secuencia de la estructura de referencia de cada superfamilia. El eje horizontal corresponde a la posición en la secuencia de la referencia y el eje vertical corresponde a la cantidad de veces que esa posición se co-localiza un PB. Se puede observar que en las superfamilias 3.40.50.300 (21 péptidos), 3.40.50.720 (15 péptidos) y 2.60.120.10 (15 péptidos) las posiciones que se solapan con péptidos se encuentran distribuidas a lo largo de toda la secuencia de la estructura representativa pero llamativamente en algunas superfamilias los distintos PBs se encuentran mapeados en la misma región de la estructura (en el eje y se puede llegar a ver algunos casos con hasta 6 PBs en la misma región estructural).

En la **FIGURA I.13–I.16** se muestran algunos ejemplos coloreados diferencialmente por actividad biológica y a su vez mapeados en la estructura terciaria de la estructura que representa a la superfamilia.

Estos resultados sugieren que en algunas familias encontramos lo que denominamos hot-spots, esto es la ocurrencia de distintos PBs en la misma región estructural.

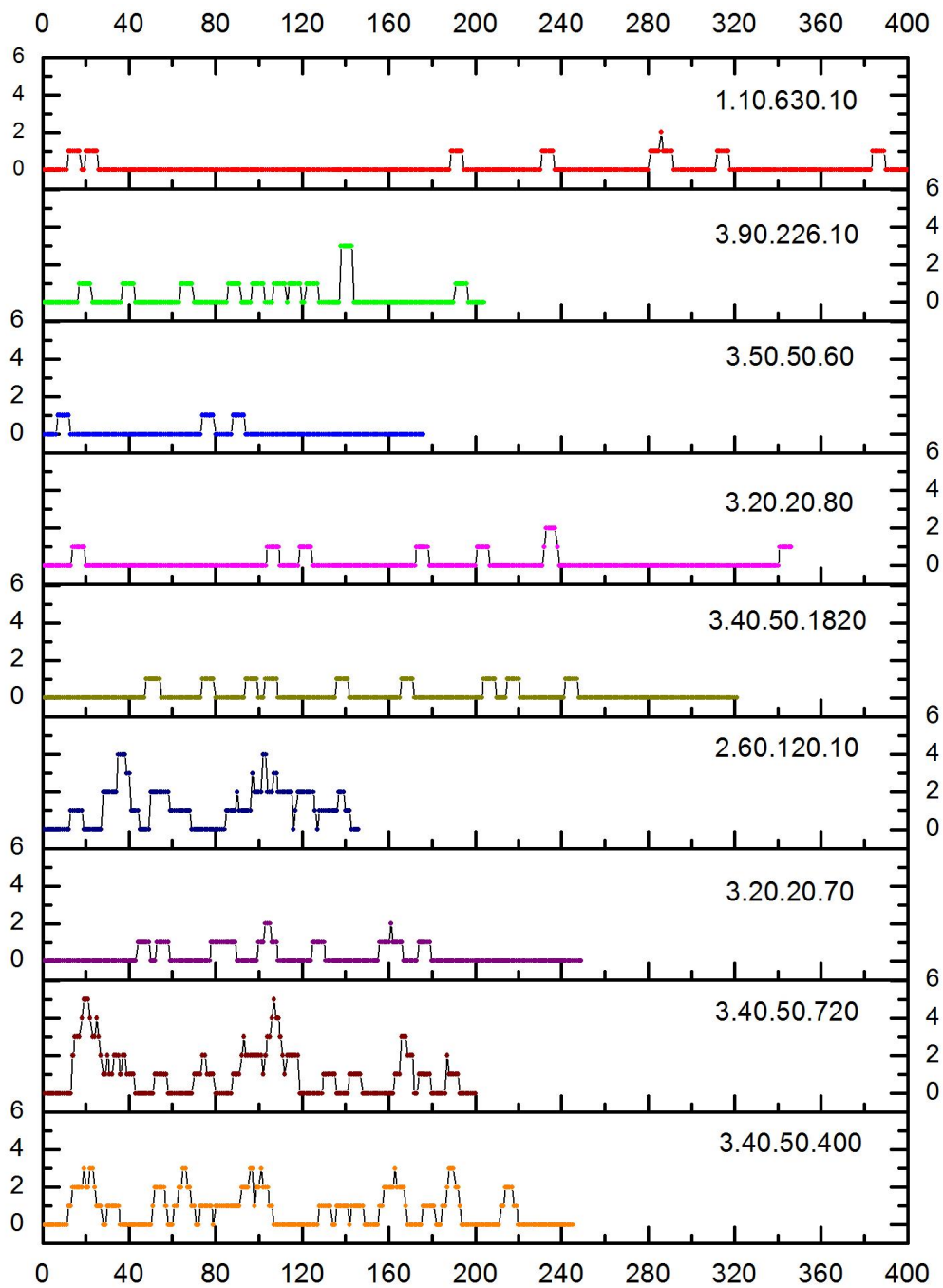


FIGURA 1.12. Hots spots en la secuencia de la estructura de referencia de cada superfamilia. El eje horizontal corresponde a la posición en la secuencia de la referencia. El eje vertical corresponde a la cantidad de veces que esa posición machea con un péptido.

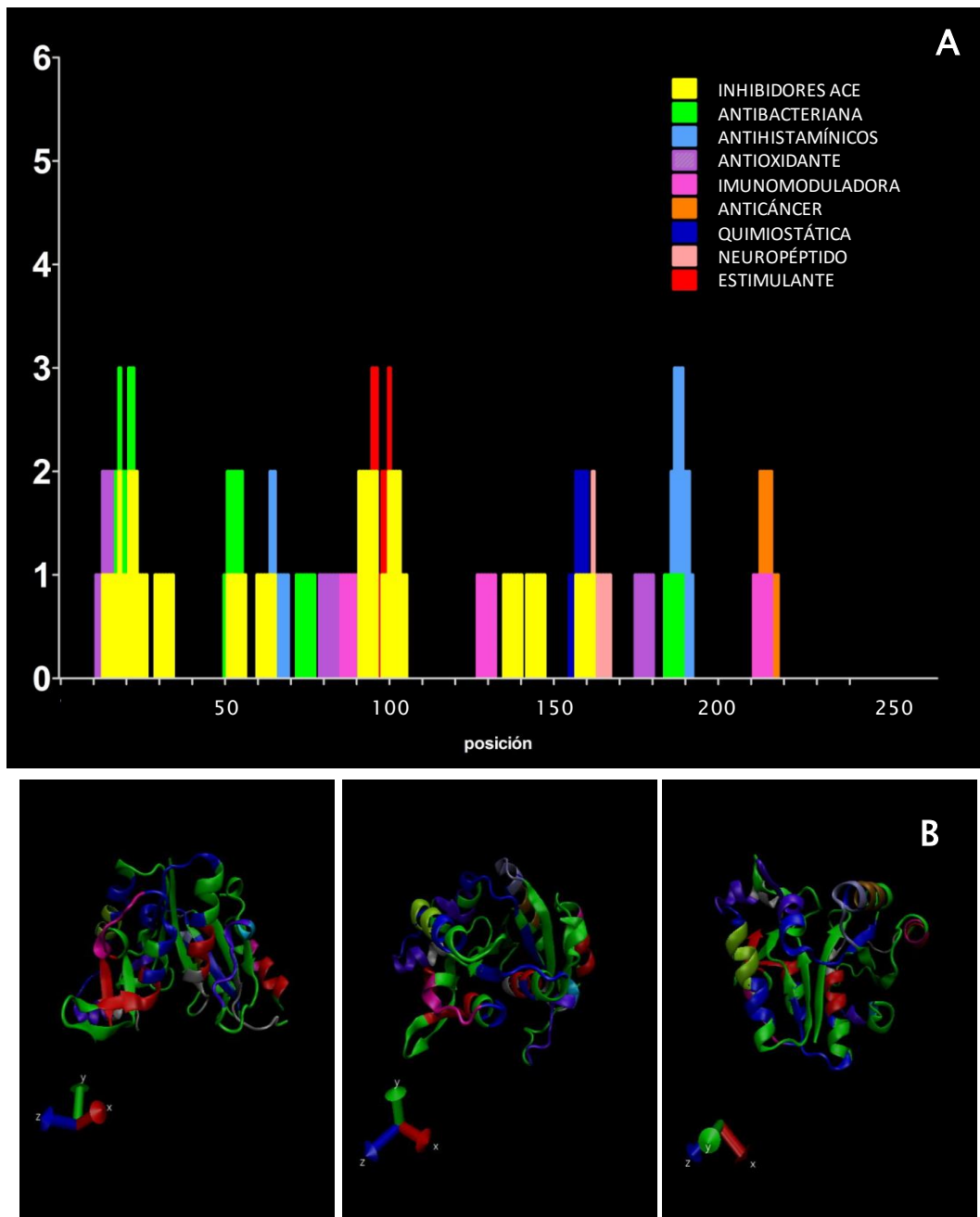


FIGURA I.13. (A) Hots spots en la secuencia de la estructura de referencia de la superfamilia 3.40.50.300. por actividad. **(B)** Dominio 1byia00, representante superfamilia 3.40.50.300 (verde). Hot-spots según actividad detallada en BIOPEP: Inhibidores ACE (azul); antihistamínicos (magenta); antibacteriana (rojo); anticáncer (cyan); antioxidante (gris); quimiostática (ocre); neuropéptido (ice-blue); estimulante (amarillo lima); inmunomoduladora (violeta).

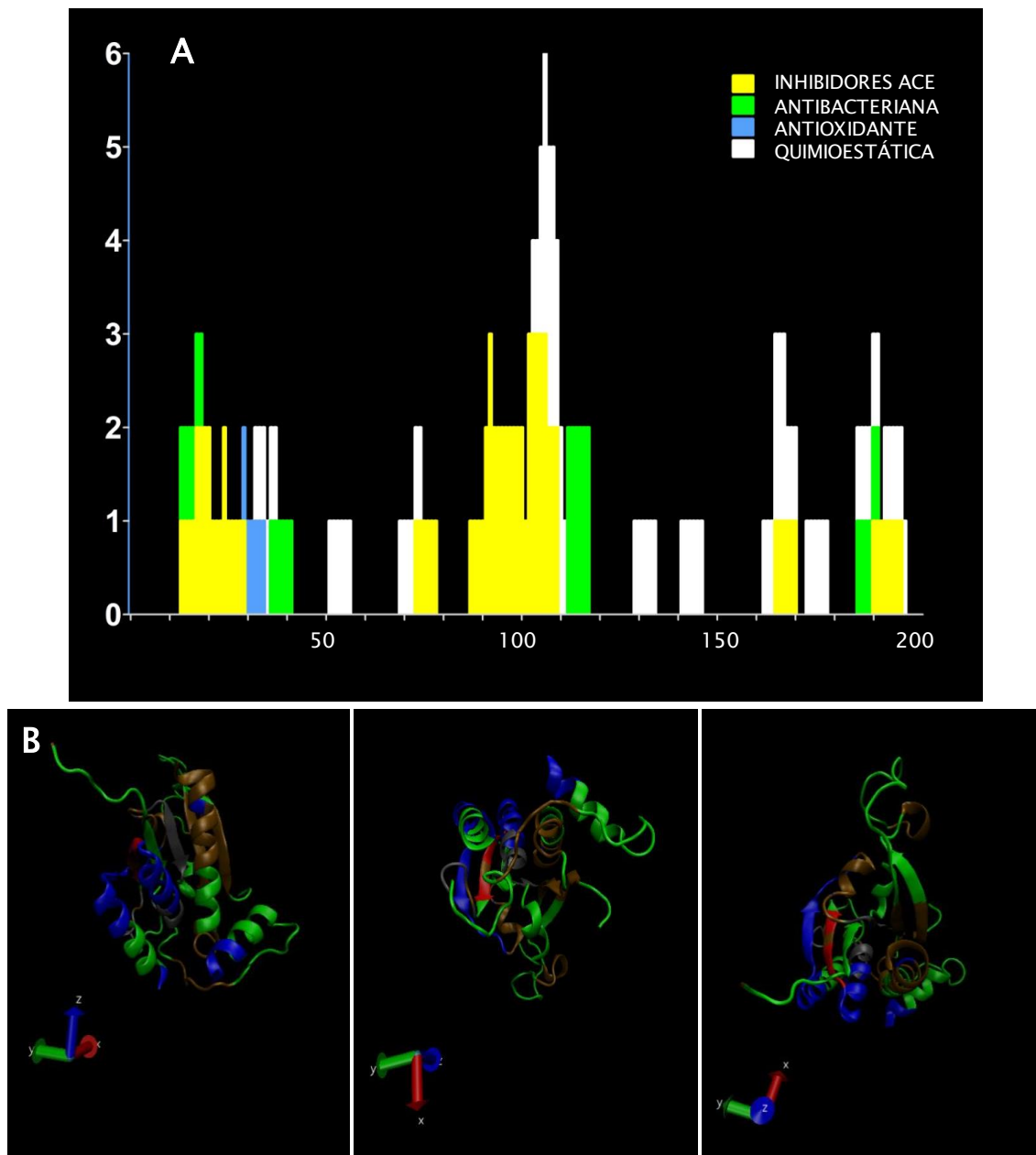


FIGURA I.14 (A) Hots spots en la secuencia de la estructura de referencia de la superfamilia 3.40.50.720 por actividad. **(B)** Dominio 1c0pa01, representante superfamilia 3.40.50.720 (verde). Hot-spots según actividad detallada en BIOPEP: Inhibidores ACE (azul); antibacteriana (rojo); antioxidante (gris); quimiostática (ocre).

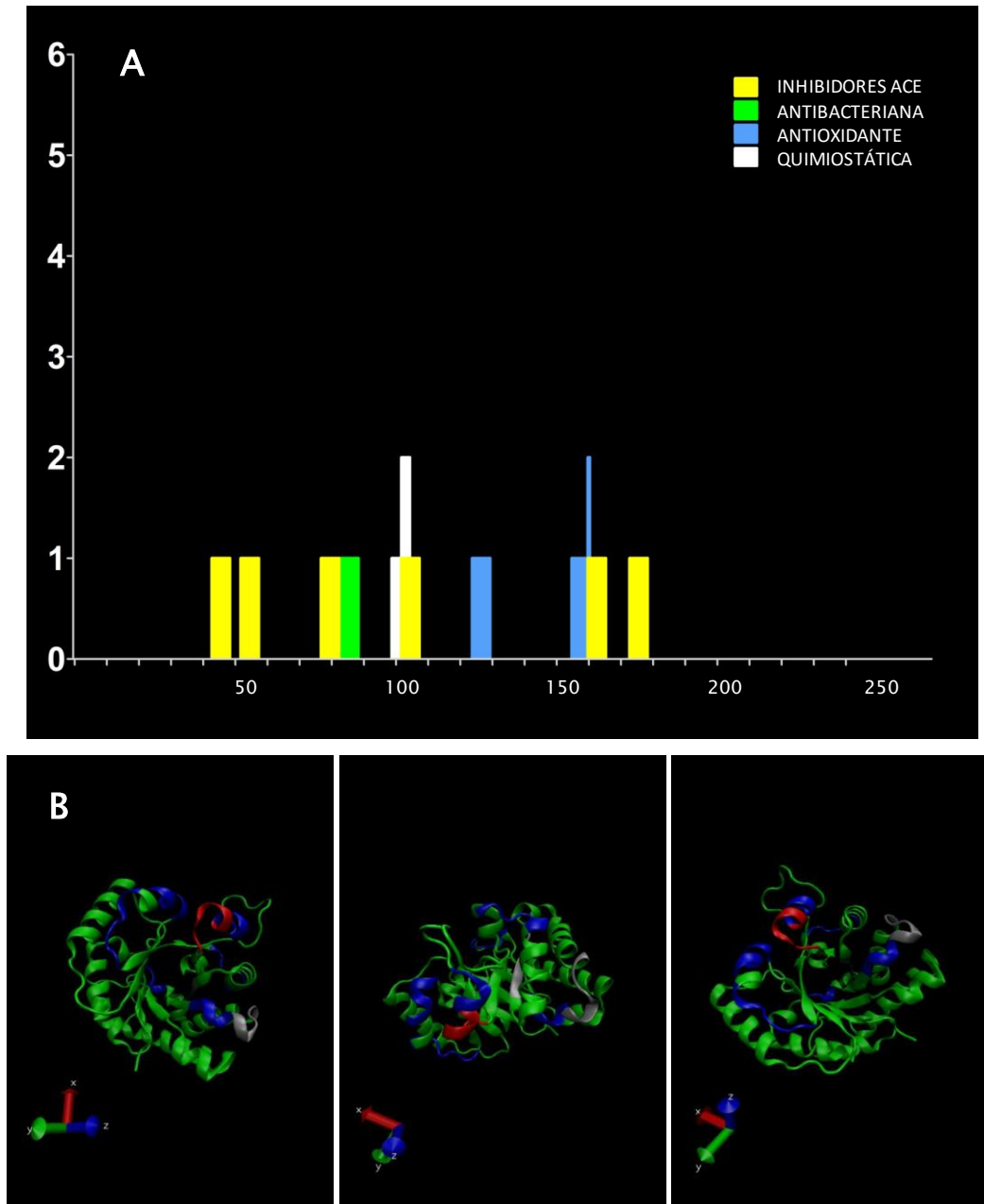


FIGURA I.15 (A) Hots spots en la secuencia de la estructura de referencia de la superfamilia 3.20.20.70 por actividad. **(B)** Dominio 2vxnA00, representante superfamilia 3.20.20.70 (verde). Hot-spots según actividad detallada en BIOPEP: Inhibidores ACE (azul); antibacteriana (rojo); antioxidante (gris); quimio-stática (ocre).

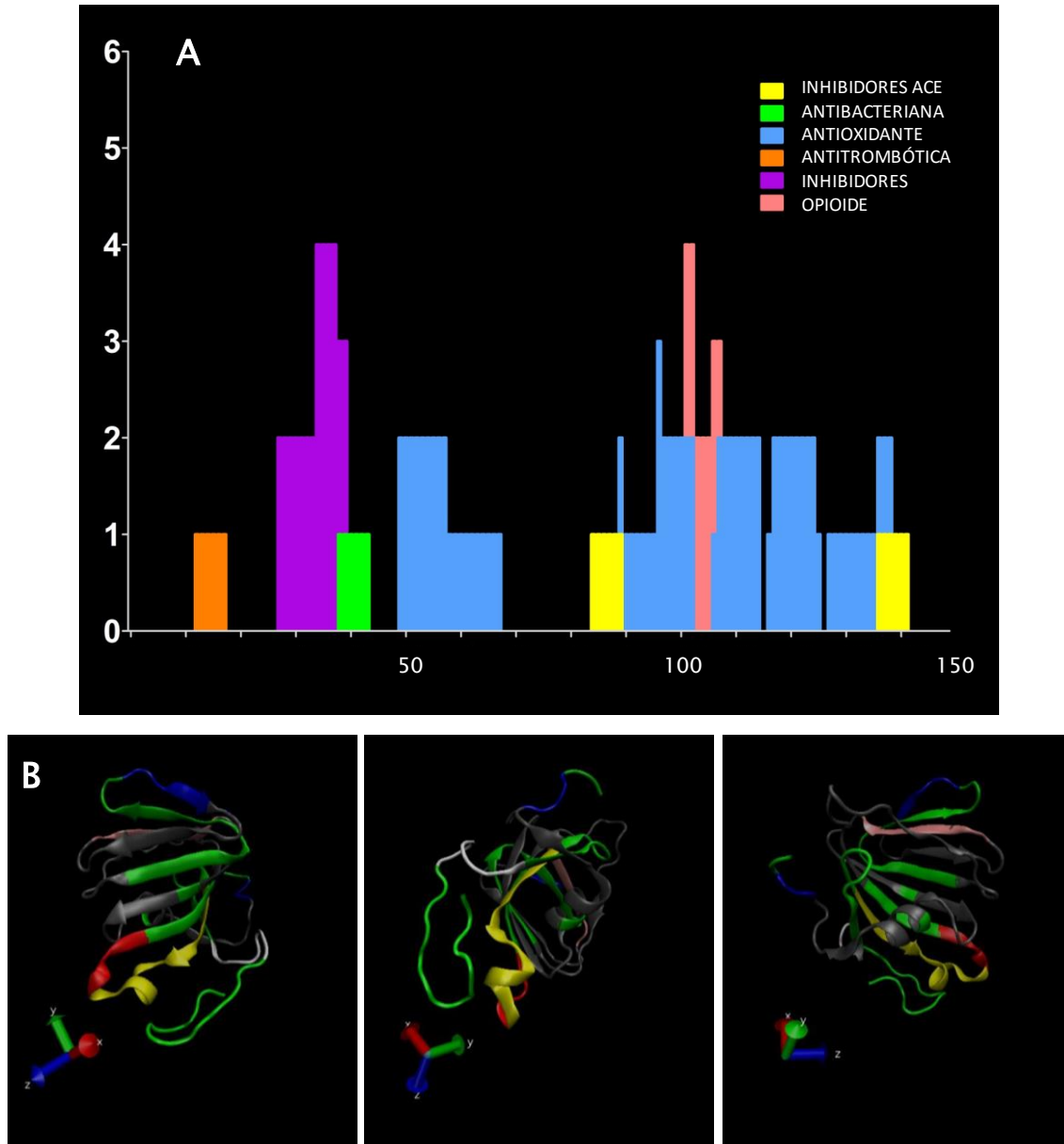


FIGURA I.16 (A) Hots spots en la secuencia de la estructura de referencia de la superfamilia 2.60.120.10 por actividad. **(B)** Dominio 1juA02, representante superfamilia 2.60.120.10 (verde). Hot-spots según actividad detallada en BIOPEP: Inhibidores ACE (azul); antibacteriano (rojo); antioxidante (gris); antitrombótica (blanco); Inhibidores (amarillo); opioide (rosa).

I.5 CONCLUSIONES CAPÍTULO 1

Dentro del espacio secuencial muestreado se encontraron que 89.909 secuencias que tienen al menos un péptido con actividad biológica ya descrita encriptado en su secuencia. Fue posible realizar la asignación estructural al 60% de las secuencias que contienen al menos un péptido mediante la búsqueda de similitud estructural empleando la base de datos de CATH.

Encontramos que sólo 4 plegamientos que contienen PBs dan cuenta de casi 90% de las secuencias a las que se pudo designar un plegamiento. Creemos que este resultado se encuentra enmascarado por la abundancia relativa de determinadas secuencias en la base de datos secuencial. Por ejemplo, uno de los cuatro plegamientos encontrados se corresponde con la rubisco, una de las proteínas más abundantes de la Tierra ya que se encuentra en la gran mayoría de las plantas.

Uno de los resultados más promisorios para desarrollar herramientas bioinformáticas para identificar posibles proteínas conteniendo PBs, resulta del análisis del número de PBs en una determinada superfamilia estructural. Como explicamos anteriormente, las proteínas con más de 5 PBs se encuentran en superfamilias que tienen gran diversidad funcional, secuencial y estructural. Si bien resta estudiar el porqué de esta correlación, es un dato que podría utilizarse para realizar predicciones y en nuestra opinión indica una de las primeras propiedades globales que identifican a proteínas con gran cantidad de PBs. En cuanto a la predicción, las proteínas con estas características, serían las primeras candidatas a analizar para la caracterización de PBs en proteínas de interés alimentario. Ya que la gran mayoría de los métodos de predicción de PBs se basan en el conocimiento previo de PBs ya caracterizados y depositados en bases de datos, las propiedades globales de proteínas encontradas en este trabajo

doctoral permitirían explorar la ocurrencia de nuevos PBs en las proteínas candidatas.

De la misma forma, el hallazgo de hot-spots, regiones con gran tendencia a la co-localización de distintos PBs en una determinada superfamilia, podría servir para el desarrollo de herramientas bioinformáticas de predicción de nuevos PBs en regiones equivalentes a la de los hot-spots en proteínas de la misma superfamilia. La ocurrencia de hot-spots para PBs podría indicar que el péptido crítico en ese hot-spots conservaría parte de la conformación en la proteína origen, aunque trabajos adicionales son necesarios para explorar esta hipótesis.

CAPÍTULO 2

INHIBIDORES PEPTÍDICOS DE HMGR DE AMARANTO

2.1 INTRODUCCIÓN

2.1.1 LA VÍA DEL MEVALONATO

La mayoría de las células de mamíferos pueden producir colesterol. La biosíntesis del colesterol es un proceso complejo que involucra a más de 30 enzimas, los detalles de la vía biosintética fueron estudiados y obtenidos en muchas instituciones, principalmente en los años 50 y 60. La vía, llamada vía del Mevalonato (VM), es un blanco natural en la búsqueda de fármacos para reducir las concentraciones plasmáticas de colesterol y reducir el riesgo de ECV.

La VM provee a la célula de moléculas esenciales que participan en múltiples procesos celulares además de la biosíntesis del colesterol, entre las cuales se destacan las ubiquinonas, hemo-A e isopenteniladenina, e intermediarios como farnesilpirofosfato y geranilpirofosfato (Buhaescu y Izzedine, 2007; Goldstein y Brown, 1990). Es una vía compleja en la cual participan componentes citoplasmáticos y microsomales. Comienza en el citosol con la síntesis de acetoacetil-CoA a partir de dos moléculas de acetil-CoA, que sirve como esqueleto carbonado de todos los esteroides producidos en la vía, en una reacción catalizada por la enzima acetoacetil-CoA tiolasa (Tiolasa). Seguidamente, una molécula de acetoacetil-CoA vuelve a combinarse con otra de acetil-CoA para formar 3-hidroxi-3-metilglutaril-CoA (HMG-CoA) por acción de la HMG-CoA sintasa (HMGCS). Luego se produce la conversión de HMG-CoA a mevalonato (ácido mevalónico) a partir de la HMG-CoA reductasa (HMGR). La mevalonato quinasa (MK) cataliza la fosforilación de mevalonato para dar fosfomevalonato. En la siguiente reacción se produce una nueva fosforilación, en la cual la fosfomevalonato quinasa (PMK) convierte el fosfomevalonato en bifosfomevalonato (MP), que posteriormente es decarboxilado por la MPP

decarboxilasa para formar isopentenilpirofosfato (IPP). Este último compuesto constituye la unidad estructural cuya polimerización dará origen a los distintos productos poliisoprenoides de la vía. Una isomerasa cataliza la interconversión de IPP en su isómero dimetilalil pirofosfato (DMAPP), siendo ambos isoprenoides precursores de distintos productos celulares (Buhaescu y Izzedine, 2007). En la siguiente reacción se produce la condensación entre IPP y DMAPP para generar una molécula de geranil pirofosfato (GPP), el cual reacciona con otra molécula de IPP en una segunda reacción de condensación para formar farnesil pirofosfato (FPP). Ambas reacciones de condensación del tipo "cabeza-cola" son catalizadas por la enzima FPP sintasa. Este punto representa la primera ramificación de la vía. Una de las ramas continúa hacia la síntesis de colesterol en una serie de aproximadamente 20 reacciones (FIGURA II.1).

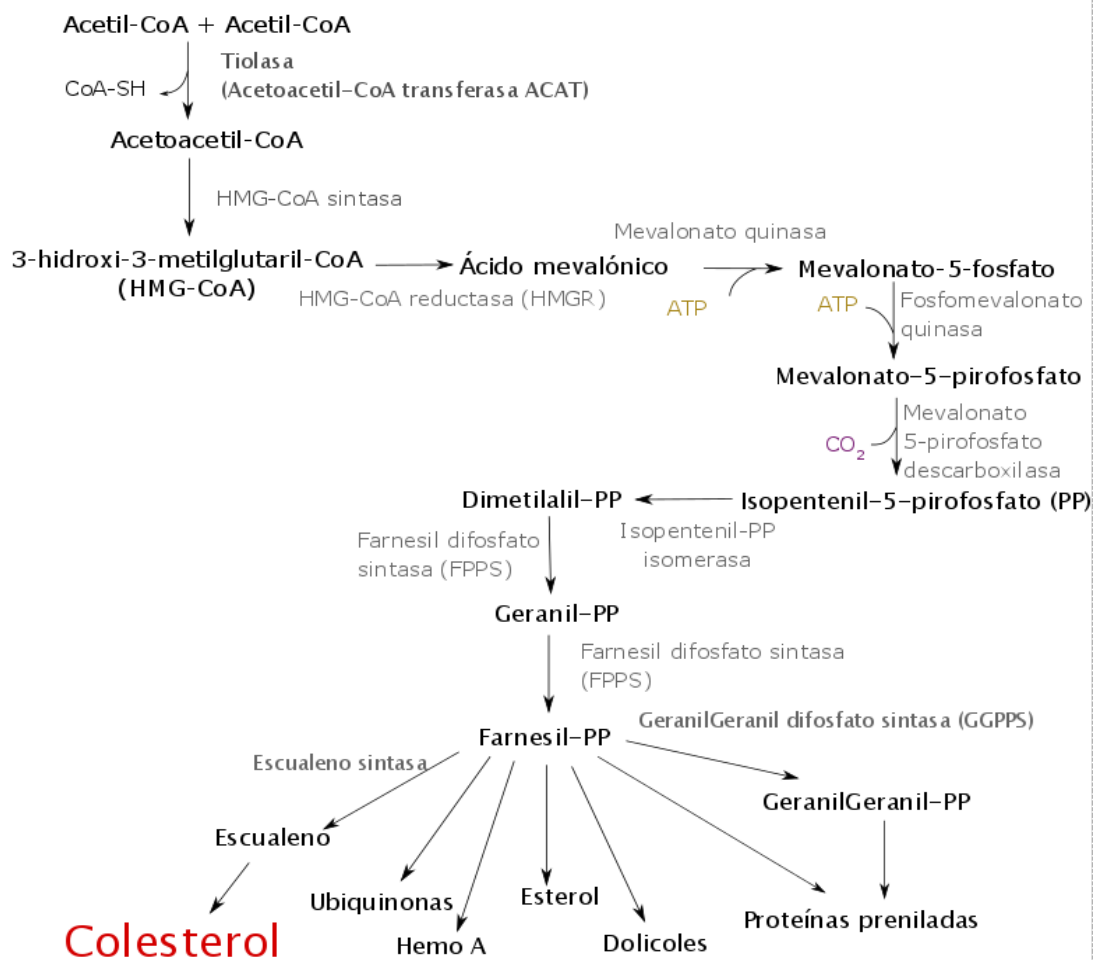


FIGURA II.1 Esquema de la vía del mevalonato en humanos.

2.1.2 HMGR

La enzima limitante de la VM, la 3-hidroxi-3-metilglutaril-coenzima-A reductasa (HMGR) (EC 1.1.1.34). Se encuentra presente en Eukarya, Archaea y Bacteria, los tres dominios de los seres vivos. La comparación de secuencias entre HMGR revela dos clases distintas de enzimas: HMGR de eucariotes (clase I) y HMGR de procariotes (clase II) (Bochar y col., 1999). En Archaea el análisis filogenético reveló la presencia de las dos clases de HMGR lo que sugiere una divergencia evolutiva entre los dos tipos de enzimas (Friesen y Rodwell, 2004).

Las HMGR de diferentes organismos son multímeros de un número específico de especies de monómeros idénticos. Las enzimas de clase I, a excepción de las presentes en Archaea, contienen dominios de membrana N-terminal que están implicados en la degradación regulada por esterol de HMGR, mientras que las HMGR clase II carecen de este dominio y son solubles. Los dominios de membrana de las HMGR de clase I son diversos, presentando dos (plantas), siete (levaduras) u ocho (mamíferos) hélices transmembrana. El dominio catalítico es conservado entre las HMGR de una misma clase, pero difieren entre clases (con sólo 14–20% de identidad), lo que sugiere divergencia evolutiva entre ambas clases (Bochar y col., 1999).

En humanos la enzima se localiza en el retículo endoplasmático (RE) y está formada por una única cadena polipeptídica de 888 aminoácidos. El extremo amino terminal de 339 residuos corresponde al dominio transmembrana; mientras que la región catalítica se ubica en el extremo C-terminal (residuos 460–888) orientado hacia el citosol. Ambos dominios se conectan por un segmento de 119 residuos llamado dominio de unión (**FIGURA II.2, A**).

La estructura de la porción catalítica de la HMGR humana (residuos 460–888) ha sido descrita por Istvan y Deisenhofer (2000) (PDB 1DQ8, 1DQ9, 1DQA). La enzima forma tetrámeros cuyas dimensiones aproximadas son 110 Å x 80 Å x 70 Å con simetría D₂. En esta estructura dos monómeros se pliegan uno sobre otro formando dímeros, que se disponen de a dos para formar la estructura del tetrámero (**FIGURA II.2, B**).

Los monómeros están formados por un dominio N, un gran dominio L y un pequeño dominio S (**FIGURA II.2, C**). El dominio N el más pequeño de los tres dominios y presenta una estructura α -helicoidal y en la proteína completa conecta la porción catalítica de HMGR al dominio de la membrana. El dominio L (d1dq8b4) contiene el sitio de unión del sustrato, HMG-CoA. Su plegamiento (residuos 528–590 y 694–872) es exclusivo de las HMGR y presentan una

arquitectura semejante a un prisma. El dominio S (d1dq8b1) (residuos 592–682) se inserta en el dominio L. Contiene el sitio de unión para NADPH y adopta un plegamiento de sándwich α/β que se asemeja a la ferredoxina y no al plegamiento de Rossmann usualmente observado en las enzimas dependientes de NADP (Istvan, 2001a). Los dominios S y L están conectados por una cadena β casi continua (L β 3 y S β 1) y un bucle (residuos 682–694). Este bucle, llamado 'cis-loop' contiene un *cis*-péptido entre los residuos C688 y T689 y es esencial en la formación del sitio activo de la HMGR (Istvan y col., 2000).

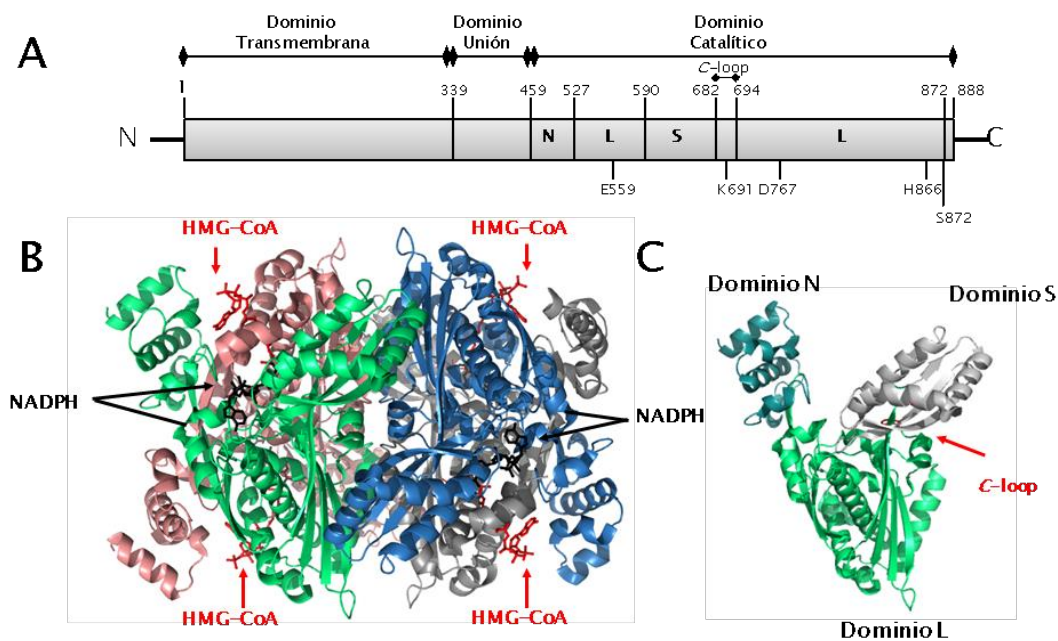
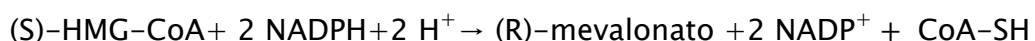


FIGURA II.2 HMGR humana. **(A)** Representación esquemática de la secuencia de HMGR **(B)** Diagrama de cintas del tetrámero de la HMGR humana (PDB 1DQA). Cada monómero se presenta en rosa, verde, azul y gris. El tetrámero está formado por dos dímeros conformados uno por el dímero verde y rosa y el otro por el azul y gris. Cada monómero presenta una molécula de HMG-CoA (en rojo) y una de NADPH (en negro). **(C)** Diagrama de cintas de uno de los monómeros. En azul se indica el dominio N, en verde el dominio S y en gris el dominio L.

La reacción catalizada por la HMGR es:

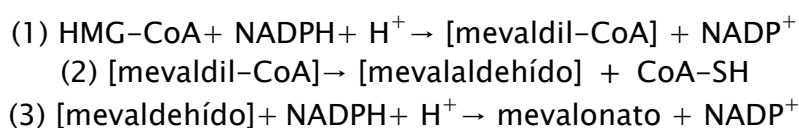


donde (S)-HMG-CoA y (R)-mevalonato corresponden a los estereoisómeros del sustrato y producto respectivamente. El sitio activo está situado en la interfase de los dos monómeros, de manera que cada dímero posee dos sitios activos donde participan residuos de ambos monómeros (diferenciados con las letras α y β).

La porción HMG de una molécula de HMG-CoA se une predominantemente a uno de los monómeros y se localiza cercana al sitio de unión del NADPH ubicado en el otro monómero. La porción CoA del sustrato adopta una conformación extendida, donde su grupo ADP se ubica en un bolsillo con carga positiva cerca de la superficie de la enzima, con la parte de ácido pantoténico extendiéndose profundamente en el interior de la proteína. La unión del CoA se completa con la participación del residuo Y479 del monómero vecino. El anillo de esta tirosina se ubica en forma paralela sobre la base adenina del CoA estableciendo interacciones del tipo van der Waals, y forma un puente de hidrógeno mediante su grupo OH con el 3'-fosfato de la ribosa, cerrando de esta forma la cavidad. El *cis*-péptido se curva sobre la parte superior de HMG, es esencial en la formación del sitio de unión a HMG permitiendo la correcta orientación de los residuos S684, D690, K691, y K692 (Istvan, 2000).

El sitio activo se completa con el sitio de unión de NADPH el cual se ubica predominantemente en el monómero vecino. La unión es estabilizada por el motivo altamente conservado, DAMGMN, y por interacciones del tipo dipolo-dipolo con la hélice S α 3. La unión de NADPH provoca un cambio conformacional en el extremo C-terminal de la enzima que da como resultado el ordenamiento de la hélice L α 11 y el cierre completo del sitio activo. En las estructuras de la enzima en las cuales falta el NADPH, 26 residuos C-terminales están desordenados incluyendo el residuo H866.

La determinación de la estructura cuaternaria de HMGR humana permitió junto con estudios cinéticos previos dilucidar el mecanismo de reacción (Frimpong y Rodwell, 1994; Istvan, 2000). La reacción ocurre en tres etapas:



La primera reacción de reducción da como producto intermediario mevaldil-CoA, que es estabilizado por la cadena lateral del residuo K691. La unión de NADPH provoca un cambio conformacional que da como resultado el ordenamiento de los residuos C-terminales. Esto cierra el sitio activo y orienta el residuo H866 dentro de la distancia de puente de hidrógeno al grupo tiol de CoA, permitiendo que el residuo H866 done un protón al tioanión en el segundo paso de la reacción. La proximidad de E559 a D767 podría potencialmente elevar el pKa de la cadena lateral de ácido glutámico de manera que pueda ser

protonado. En consecuencia, se ha propuesto que el E559 es el donador de protones para la formación del mevaldehído. D767 es crítico en la catálisis, porque su cadena lateral se sitúa cerca de E559 y también forma interacciones iónicas con K691, estabilizando la cadena lateral de lisina en el sitio activo (Istvan, 2000) (FIGURA II.3).

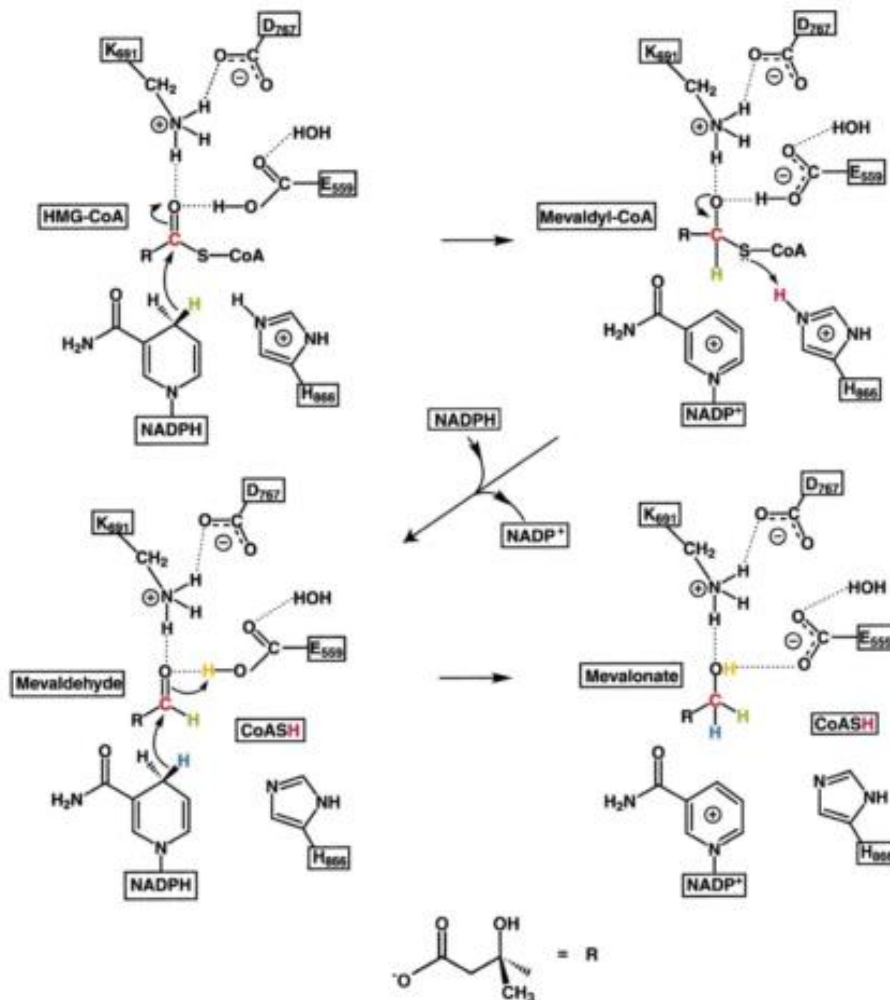


FIGURA II.3 Mecanismo de reacción de HMGR humana (adaptado de Istvan 2000)

Los estudios de dinámica molecular QM/MM sobre el sitio activo de HMGR humana con distintos estados de protonación de los residuos Glu559, Asp767 e His752 (Oliveira y col., 2016) concuerda con la hipótesis de que Glu559 actúa como un ácido general durante la primera etapa de reducción, tal como lo sugirieron tanto la estructura cristalográfica como el estudio computacional sobre la HMGR de procariontes (Haines y col., 2012). El mecanismo observado en simulaciones de dinámica molecular genera un intermediario, el mevaldil-CoA estable, que permite el intercambio del cofactor y no impide que ocurra la reacción inversa, hecho que concuerda con observaciones experimentales.

2.1.3 Regulación HMGR

La HMGR presenta uno de los mecanismos de regulación más complejos conocidos hasta el momento (Ačimovič y Rozman, 2013). Está regulada a nivel de transcripción, traducción, modificaciones postraduccionales y degradación (Goldstein y Brown, 1990, 2015; Jo y DeBose-Boyd, 2010).

El modo principal por el cual esta enzima es regulada es por medio del control a largo plazo mediante la modulación de la velocidad de síntesis y degradación por retroalimentación. Cuando los niveles de LDL-colesterol o mevalonato caen, la cantidad de HMGR presente en las células puede ascender hasta 200 veces, debido a un incremento en la síntesis de la enzima, combinado con una disminución de su degradación. Cuando los niveles de LDL-colesterol o mevalonato aumentan, estos efectos se revierten. El colesterol puede controlar la expresión de más de 20 genes involucrados en su metabolismo tales como los que codifican a la HMGR y al receptor de LDL (Horton y col., 2003). Todos estos genes contienen una secuencia de ADN por delante del sitio de iniciación de la transcripción llamada elemento regulador de esteroides (SRE). Para que estos genes se transcriban, un factor de transcripción específico, la proteína de unión al elemento regulador de esteroides (SREBP), debe unirse al SRE. La SREBP es sintetizada como una proteína integral de membrana que se fija en la membrana del retículo endoplasmático formando un complejo con la proteína activadora de la ruptura del SREBP (SCAP) – cuando la concentración de colesterol se eleva por encima de un umbral de 5% del total de lípidos– (Motamed y col., 2011). El SCAP funciona como un sensor de esteroides. Cuando el colesterol en la membrana del retículo endoplasmático disminuye, la SCAP cambia su conformación e inmediatamente después acompaña a su SREBP unida hacia el aparato de Golgi por medio de vesículas membranosas. En el aparato de Golgi, la SREBP es clivada por dos proteasas unidas a membrana. La proteasa del sitio 1 (S1P), cliva a la SREBP en la región luminal que conecta sus dos hélices transmembrana, pero solamente cuando se halla asociada a la SCAP. Esta ruptura expone un enlace peptídico de la SREBP para que sea clivado por la proteasa de sitio 2 (S2P). Esta libera un dominio que migra al núcleo donde activa la transcripción de su gen diana. Por consiguiente, el nivel de colesterol en la célula aumenta hasta que la SCAP no induce más la translocación de la SRBEP al Golgi, un ejemplo clásico de inhibición por retroalimentación.

Por otro lado, la actividad de HMGR es atenuada a través de la fosforilación de un residuo serina localizado en el dominio catalítico (S872) (Sato y col., 1993).

La HMGR es fosforilada en un sistema de cascada bicíclico por medio de una enzima quinasa dependiente de AMP (AMPK).

2.1.4 Inhibidores de la HMGR: Estatinas

Las estatinas son un grupo de fármacos utilizados para disminuir el colesterol. Todas las estatinas comparten un componente estructural que imita a la porción HMG de HMG-CoA unido covalentemente a un grupo hidrofóbico rígido. Esta similitud les permite actuar como inhibidores competitivos de la HMGR con respecto a la unión de HMG-CoA, pero no con respecto a la unión de NADPH (Istvan, 2003).

Las estatinas se clasifican en dos grupos estructurales (**FIGURA II.4**). Las estatinas naturales conocidas como las de tipo I, presentan en común el anillo hexahidronaftaleno con un grupo éster α -metilbutírico y una β -hidroxi- δ -lactona enlazada por un puente etilénico; pero difieren las unas de las otras por la posición de los metilos sobre el anillo y la cadena lateral. El anillo de la β -hidroxi- δ -lactona de estos compuestos puede sufrir una apertura hidrolítica y convertirse en los correspondientes ácidos 3',5'-dihidroxiheptanoicos, que son las formas activas y las responsables del efecto inhibitor de la HMGR. Las estatinas del tipo II, surgieron como producto de las diversas investigaciones llevadas a cabo con el fin de obtener medicamentos hipocolesterolémicos con una acción farmacológica potenciada con respecto a las estatinas naturales (Goldstein y Brown, 2015; Istvan, 2001b). Estructuralmente conservan la porción afín con la HMGR responsable de su acción inhibitoria, pero presentan un tamaño molecular mayor que las de tipo I. Tienen en común una cadena lateral fluorofenil, un grupo metiletil y un ácido heptenoico. Las diferencias estructurales entre las distintas estatinas de este grupo dan lugar a diferentes características de unión que se suman a las de unión de la porción que imita la porción HMG del sustrato.

La resolución de la estructura de HMGR co-cristalizada con diferentes estatinas, junto con estudios de cinética enzimática, han permitido dilucidar el mecanismo de acción de estos compuestos a nivel molecular (Istvan, 2003). El grupo similar a HMG (común a las moléculas de estatina) ocupa el sitio de unión de HMG de la enzima, mientras que los grupos hidrofóbicos se posicionan en una ranura poco profunda formada por el reordenamiento de los residuos C-terminales de la enzima, uniéndose fuertemente a la misma. De esta forma a pesar de ser moléculas más voluminosas que el sustrato, logran ubicarse en el sitio activo y

2.2 OBJETIVOS ESPECÍFICOS

De los objetivos específicos del presente trabajo de Tesis, los correspondientes a este capítulo son:

- Obtener, caracterizar y purificar los péptidos que provienen de un aislado proteico de amaranto y sus hidrolizados.
- Determinar la capacidad de inhibición de la fracción proteica de amaranto de la enzima 3-hidroxi-3-metil-CoA reductasa (HMGR).
- Determinar la capacidad de los productos de hidrólisis de un aislado proteico de amaranto de inhibir a la HMGR.
- Estimar mediante estudios computacionales la presencia de péptidos bioactivos en la fracción globulinas de amaranto de inhibidores de HMGR.
- Caracterizar posibles mecanismos de acción de los péptidos con diferentes actividades utilizando técnicas *in silico*.

2.3 MATERIALES Y MÉTODOS

2.3.1 Obtención de la harina de amaranto

Se utilizaron semillas de *Amaranthus hypochondriacus* cosechadas en el Instituto de Agrobiotecnología de Rosario (INDEAR), Argentina (2013). Las semillas enteras se molieron en un molino ciclónico UDY (UDY Corporation), de 1 mm de malla y posteriormente se desgrasaron mediante extracción con n-hexano, en una relación 10 % p/v, durante 5 h con agitación constante a temperatura ambiente. Una vez transcurrido ese tiempo se dejó en contacto 20 h más en reposo. Para separar el solvente se realizó una filtración a temperatura ambiente empleando una bomba de vacío y luego se dejó secar en iguales condiciones durante otras 20 h para asegurar la evaporación completa del hexano. La harina desgrasada obtenida se almacenó en recipientes herméticos a 4°C hasta su uso.

2.3.2 Preparación de los aislados proteicos

Los aislados proteicos de amaranto fueron obtenidos de acuerdo a lo propuesto por Martínez y Añón (1996). La harina desgrasada se dispersó en agua, en una relación 10% p/v. La extracción de las proteínas se realizó ajustando el pH a 9,0 con NaOH 2 N y agitando la suspensión durante 1 h a temperatura ambiente. El siguiente paso consistió en la centrifugación a 9000 g a 10°C durante 20 min (centrífuga Beckman). El sobrenadante obtenido en la centrifugación se ajustó a pH 5,0 con HCl 2 N con el fin de precipitar las proteínas presentes al alcanzar su punto isoeléctrico. La dispersión obtenida se centrifugó a 9000g y a 4°C durante 20 min a efectos de separar las proteínas. Se descartó el sobrenadante;

el precipitado fue resuspendido en agua, en una relación 1:3 y se neutralizó con NaOH 2 N hasta pH 7. Luego se congeló a -80°C y se liofilizó. El sólido obtenido fue molido y almacenado a 4°C en recipientes herméticos hasta su uso (FIGURA II.5).

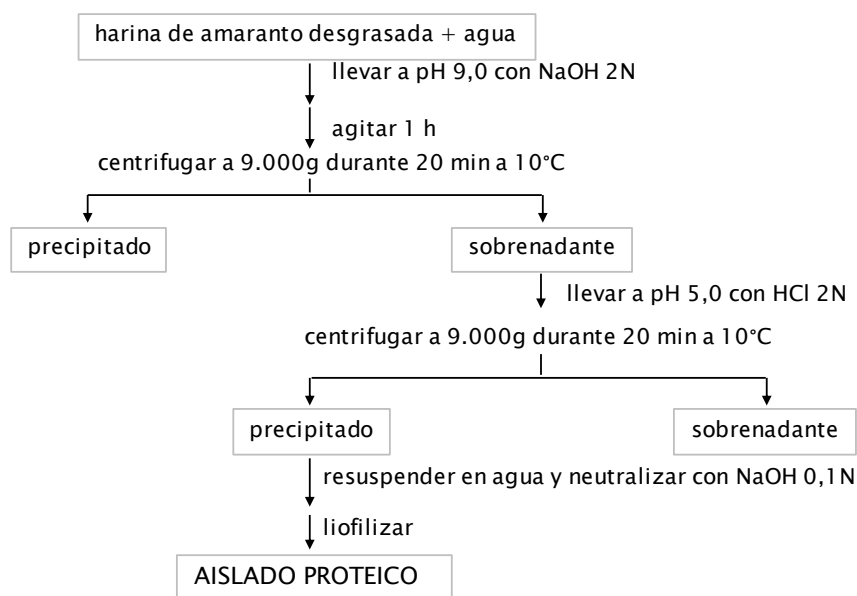


FIGURA II.5 Esquema de obtención de aislados de amaranto.

2.3.3 Preparación de los hidrolizados proteicos con alcalasa

El aislado nativo se dispersó en agua utilizando una relación sólido:líquido de 1:10 p/v y se ajustó el pH a 9 con NaOH 2 N. Las dispersiones obtenidas se agitaron durante 1h a 37°C en un agitador orbital con termostato (OrbitEnvironShaker de Lab-Line). Se agregó la enzima alcalasa, se utilizó una proteasa bacteriana proveniente de *Bacillus licheniformis*, marca Sigma-Aldrich ($\geq 2.4 \text{ U/g}$). La mezcla de reacción se incubó durante 6 h a 37°C controlando y corrigiendo el pH durante todo el proceso. Se adicionaron $0,16 \mu\text{l}$ de la enzima comercial por mg de aislado. Esta relación ha sido previamente utilizada y optimizada en nuestro laboratorio (Vecchi, 2007), de modo de obtener hidrolizados extensivos (GH $>10\%$) que generen la liberación de péptidos bioactivos (Benítez y col., 2008).

La reacción de hidrólisis se detuvo por calentamiento de la mezcla de reacción a 90°C durante 10 min en baño de agua. Posteriormente la muestra se congeló a -80°C y se liofilizó. Los hidrolizados obtenidos se molieron y se almacenaron a 4°C en recipientes herméticamente cerrados hasta su utilización. Se tomaron

alícuotas a distintos tiempos de proteólisis con el objetivo de estudiar la cinética de reacción y realizar un seguimiento de la misma.

2.3.4 Digestión gastrointestinal simulada

El proceso de digestión es un proceso complejo en el cual intervienen factores intrínsecos propios del individuo. Los protocolos de simulación utilizados, son diversos, tanto desde el punto de vista cualitativo como cuantitativo, considerando la composición química de los fluidos y enzimas empleadas. Estas diferencias en el protocolo en muchos casos dificultan la comparación de los resultados obtenidos por diferentes autores. Recientemente se ha propuesto un protocolo estandarizado de digestión *in vitro* (Minekus y col., 2014) aceptado por la comunidad científica, sin embargo, al inicio del presente trabajo no se había implementado este modelo razón por la cual se empleó otro protocolo.

El aislado nativo de *Amaranthus hypochondriacus* se sometió a una digestión gastrointestinal simulada empleada por Gawlik-Dziki y col., (2009) con modificaciones (Minekus y col., 2014) por duplicado.

Se preparó una suspensión de 2 g de aislado en 40 ml de fluido similar saliva con α -amilasa (final 70 Unidades/ml), se dejó 5 min en agitación a 37°C. Luego se adicionaron 160 ml de fluido similar jugo gástrico (termostatizado a 37°C), se ajustó el pH a 2 con HCl y se dejó en agitación durante 60 min. Finalmente se adicionaron 160 ml de fluido similar jugo duodenal, se ajustó el pH a 7 y se incubó durante 60 min a 37°C con agitación. La digestión se detuvo por calentamiento de la mezcla de reacción a 80°C (medida en el seno de la dispersión) durante 10 min en baño de agua.

Se tomaron alícuotas al inicio y final de cada etapa del proceso de digestión con el objetivo de estudiar la cinética de proteólisis y realizar un seguimiento de la misma. Todas las alícuotas fueron tratadas a 80°C durante 10 min en baño de agua.

Reactivos

-*Fluido similar saliva*: 2,38 g/l Na_2HPO_4 ; 0,19 g/l KH_2PO_4 ; 0,3 g/l NaCl pH=6,8; α -amilasa (Sigma-Aldrich) 60 mg (70 Unidades/ml).

-*Fluido similar jugo gástrico*: se utilizó pepsina comercial 1:15000 marca MP Biomedicals. Se prepararon 2,5 g/l pepsina en 0,03 M de NaCl en HCl 0,1 N. La solución se incubó a 37°C durante 30 min con agitación antes de utilizarla.

-*Fluido símil jugo duodenal*: se utilizó pancreatina comercial 4 X USP marca MP Biomedicals 3 g/l pancreatina en NaHCO_3 0,1 M pH 8,1. Se incubó 30 min a 37°C con agitación antes de utilizarla.

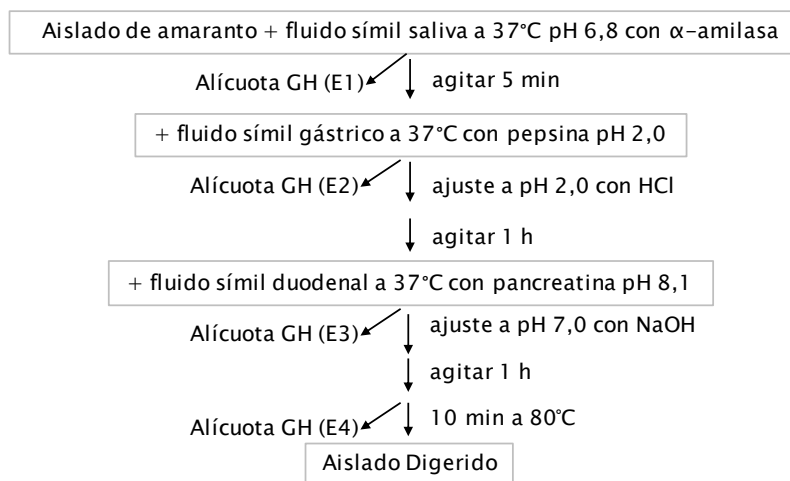


FIGURA II.6 Esquema del proceso de digestión gastrointestinal simulado. Todas las alícuotas fueron tratadas térmicamente para detener la acción de las enzimas.

2.3.5 Caracterización de las muestras

2.3.5.1 Determinación de proteínas

➤ *Método de Kjeldahl*

El contenido de proteínas de los aislados y sus respectivos hidrolizados se determinó mediante el método de Kjeldahl (AOAC, 1984). Para convertir el nitrógeno en proteínas se utilizó como factor de conversión 5,85 g proteína/g de nitrógeno, que es el más apropiado para las proteínas de amaranto (Paredes-López, 1994)

Reactivos

Catalizador: Na_2SO_4 anhidro + $\text{CuSO}_4 \cdot 5 \text{H}_2\text{O}$ (relación 10:1).

H_2SO_4 : 98% p/p ácido sulfúrico concentrado, $\delta = 1,84$ g/ml.

H_3BO_3 : 4% p/v ácido bórico

Indicador Mortimer: 0,016% p/v rojo de metilo, 0,083% p/v verde de bromocresol en etanol.

H_2SO_4 valorado: ácido sulfúrico de valorado (aproximadamente 0,2N).

La digestión de cada muestra se realizó colocándola en un balón de Kjeldahl y agregando catalizador y 15 ml de H₂SO₄ concentrado. Una vez terminada la digestión, se alcalinizó la mezcla con el objetivo de liberar el nitrógeno presente como amoníaco y destilarlo por arrastre con vapor de agua. El amoníaco destilado se recogió sobre ácido bórico 4% p/v utilizando Mortimer como indicador. La valoración se realizó con ácido sulfúrico de concentración conocida. El porcentaje de proteínas de las distintas muestras se calculó mediante la siguiente expresión:

$$\% \text{ Proteínas} = (V_{\text{muestra}} - V_{\text{blanco}}) \times N_{\text{ácido}} \times 0,014 \times 5,85 \times 100 / \text{g muestra}$$

donde,

V_{muestra}: volumen (ml) de ácido sulfúrico gastado en la valoración de la muestra.

V_{blanco}: volumen (ml) de ácido sulfúrico gastado en la valoración del blanco.

N_{ácido}: concentración normal del ácido sulfúrico.

0,014: masa en g del miliequivalente de nitrógeno.

5,85: factor de conversión de nitrógeno a proteína de amaranto.

g muestra: masa de la muestra en gramos.

➤ *Método de MicroKjeldahl*

El contenido de proteínas de las dispersiones proteicas preparadas para los estudios de solubilidad y actividad biológica, se determinó empleando el método de MicroKjeldahl, el cual, a diferencia del Kjeldahl, permite utilizar una menor cantidad de muestra para su cuantificación.

La digestión de cada una de las muestras se realizó en un tubo digestor junto con 1 g de catalizador (Na₂SO₄ anhidro + CuSO₄.5H₂O relación 10:1) y 3 ml de H₂SO₄ concentrado. Los tubos se colocaron en un digestor (2040 Digestor Foss Tecator) el tiempo necesario hasta aparición de color verde. Una vez finalizado el proceso, el contenido de los tubos se trasvasó a matraces de 50 ml y se completó el volumen con agua destilada. Por último, el amonio liberado se cuantificó por el método colorimétrico descrito por Nkonge y Ballance (1982). Se incluyeron dos tubos con controles junto con las muestras, uno de ellos con sulfato de amonio y el otro con urea. El sulfato de amonio utilizado como control fue secado en estufa durante 12 h a 105°C y cuidadosamente pesado una vez frío. Ambos controles se realizaron a fin de conocer exactamente el contenido de nitrógeno que aportan y evaluar a partir de los datos obtenidos la exactitud

del método. La lectura de la absorbancia a 660 nm permitió calcular la concentración de nitrógeno en las muestras y convertir el nitrógeno a proteínas para lo cual se utilizó el factor de conversión antes mencionado, 5,85 g proteína/g nitrógeno (Paredes-López, 1994).

Método colorimétrico de Nkonge y Balance

Reactivos

- Solución A: Na₂HPO₄ 0,2 M, NaOH 0,2 M
- Solución B: NaOH 2,5 M
- Solución salicilato-nitroprusiato (N-S): 20% p/v salicilato de sodio y 0,03% p/v nitroprusiato de sodio.
- Lavandina que asegure 5% de cloro activo.

Curva de calibración

Se preparó empleando albúmina de suero bovino (Sigma-Aldrich), en concentraciones de 0; 1,5; 3 y 6 g de nitrógeno cada 100 ml de solución.

Cálculos

El porcentaje de proteínas se obtuvo usando la curva de calibración para calcular los g de N, teniendo en cuenta las diluciones previamente realizadas, y luego aplicando el factor de conversión (5,85 mg de proteína/mg de N) para obtener los mg de proteína. Finalmente, estos últimos se refieren a la masa pesada en forma porcentual.

➤ *Determinación de proteína soluble*

La proteína soluble, preparada a diferentes concentraciones, previa a la determinación de la actividad biológica, se determinó dispersando las muestras en el buffer de estudio y cuantificando el contenido de proteínas en los sobrenadantes. Se prepararon dispersiones de los sólidos en agua en una concentración de 10 mg/ml, se agitaron a 500 rpm durante 1 h a 25°C en un termomixer Eppendorff. Luego se centrifugaron a 10000g a 20°C, durante 15 min en una microcentrífuga Hermle Z 233 MK-2. El contenido de proteína del sobrenadante se determinó utilizando el método de MicroKjeldahl.

La solubilidad (S) de cada muestra se calculó utilizando la siguiente expresión:

$$S (\%) = \frac{(s \times 100)}{mg \text{ Prot Total}}$$

donde,

s: contenido de proteína en el sobrenadante determinado por MicroKjeldahl.

mg proteína total: contenido total de proteína en la muestra determinado por Kjeldahl.

2.3.5.2 Determinación de cenizas

La determinación del contenido de minerales del aislado, albúminas e hidrolizados se realizó mediante calcinación. Las muestras se colocaron en cápsulas de porcelana taradas, se carbonizaron por acción de la llama y finalmente se calcinaron en una mufla a 550°C hasta cenizas blancas. En algunos casos, en los que no se obtuvieron cenizas blancas, fue necesario disolver los residuos en agua destilada, secar la dispersión con mechero sobre tela metálica y volver a colocar en la mufla a 550°C. La cantidad de cenizas de la muestra se calculó por diferencia de peso entre la cápsula con las cenizas y la cápsula vacía, y se refirió a la masa inicial de muestra para calcular el porcentaje de cenizas.

2.3.5.3 Determinación del contenido de humedad

La determinación de humedad se realizó utilizando el método indirecto (AOAC, 1990). Para ello las muestras se colocaron en recipientes limpios y tarados en estufa a $103 \pm 2^\circ\text{C}$ y presión atmosférica, hasta alcanzar peso constante. La humedad porcentual se calculó determinando la cantidad total de agua perdida por parte de la muestra. Ésta se calculó por diferencia de peso antes y después de haberla colocado en la estufa, y calculando el porcentaje de la misma respecto a su peso inicial.

2.3.5.4 Determinación del contenido de hidratos de carbono

El contenido de hidratos de carbono se determinó por el método colorimétrico de antrona (Loewus, 1951). Con el fin de mejorar la homogeneidad de las muestras estas fueron sometidas a hidrólisis ácida con HCl a ebullición durante 2 h a reflujo antes del ensayo colorimétrico. Se utilizó una relación de 1 ml HCl concentrado ($\delta = 1,19 \text{ g/ml}$) cada 11 ml de solución.

Para la realización de la curva de calibración se utilizaron diluciones adecuadas de glucosa patrón marca Sigma-Aldrich. Las lecturas de absorbancia se realizaron en un espectrofotómetro a una longitud de onda de 620 nm.

2.3.5.5 Determinación del grado de hidrólisis (GH%) por OPA

El grado de hidrólisis se determinó dosando los grupos amino libres mediante el método del ortoftaldehído (OPA) descrito por Nielsen y col. (2001) y Dinnella y col. (2002) con algunas modificaciones como describió Cian y col. (2012).

La curva de calibración se construyó a partir de una solución de leucina (Sigma-Aldrich) 1 mg/ml. La reacción consiste en adicionar 750 µl del reactivo de OPA sobre 100µl de la muestra, agitar durante 5 s, dejar reposar 2 min y finalmente medir la absorbancia a 340 nm, para lo cual se utilizó un lector de placas.

Reactivos

- Buffer del ensayo: Borato de sodio 0,15 M, 0,15% p/v SDS, pH 9,5.
- Reactivo OPA: buffer del ensayo.
- OPA 5,97 mM.
- DTT 5,7 mM.

Cálculos

Para calcular el grado de hidrólisis se utilizó la siguiente expresión:

$$GH \% = \frac{(NH_2 t - NH_2 t=0)}{NH_2 Totales} \times 100$$

donde,

$NH_2 t$: grupos amino libres a tiempo de hidrólisis t.

$NH_2 t=0$: grupos amino libres a tiempo cero de hidrólisis.

NH_2 totales: grupos amino totales presentes en la proteína.

NH_2 totales: $NH_2^\infty - NH_2 t=0$

NH_2^∞ : $(1 / PM \text{ aa promedio}) \times (1 + f_{Lys}) \times [\text{proteína proteínas en mg/ml}]$

$NH_2 t=0$: $(1 / PM \text{ aa promedio}) \times (1 / L + f_{Lys}) \times [\text{proteína proteínas en mg/ml}]$, y se mide experimentalmente con OPA utilizando una muestra sin hidrolizar.

PM aa promedio: promedio de los pesos moleculares de los aminoácidos de las proteínas de la muestra. Para amaranto se considera 130 g/mol.

f_{Lys} : proporción de moles de Lys (L) por cada mol de aa para proteínas de amaranto, 1/20.

L: peso promedio de los polipéptidos (300 g/mol).

2.3.5.6 Electroforesis

Se realizaron electroforesis desnaturalizantes tricina-SDS-PAGE y tricina en condiciones reductoras Tricina-SDS-PAGE + β -ME (Schägger, 2006).

Muestras

Las muestras para las electroforesis se prepararon de diferente modo según provinieran de muestras líquidas o de muestras liofilizadas. En el primer caso se mezclaron 20 μ l de las mismas (concentración aproximada 10 mg/ml), con 10 μ l del buffer de muestra adecuado, según el tipo de electroforesis a realizar. Cuando se trabajó con sólidos, las muestras se dispersaron directamente en el buffer de muestra correspondiente, en la relación 1 mg de sólido cada 500 μ l buffer aproximadamente.

Para asegurar la acción del β -mercaptoetanol presente en el buffer de muestra en condiciones reductoras, las dispersiones fueron calentadas a 100°C en baño maría durante 1 min.

Antes de sembrar las muestras, en los geles de poliacrilamida, las mismas fueron centrifugadas a 10000g a 15°C, durante 15 min. Posteriormente en cada calle se colocó un volumen de muestra, disuelta en buffer de muestra, que permitiera sembrar entre 20 y 50 μ g de proteína por calle.

Patrones

Se utilizaron diferentes patrones de masa molecular. Para las electroforesis de Tricina-SDS-PAGE se emplearon patrones de muy bajo peso molecular (VLMW) marca BIO-RAD. Las proteínas incluidas en esta mezcla son: triosa-fosfato isomerasa (26,6 kDa), mioglobina (16,95 kDa), α -lactoalbúmina (14,4 kDa), aprotinina (6,51 kDa), cadena β de la insulina (3,5 kDa) y bacitracina (1,4 kDa).

Reactivos

Reactivos Tricina SDS-PAGE y Tricina SDS-PAGE+ β -ME

- Buffer de electrodo anódico: Tris-HCl 0,1 M, pH 8,9
- Buffer de electrodo catódico: Tris 0,1 M, Tricina 0,1 M, SDS 0,1% p/v, pH 8,25
- Buffer de gel: Tris-HCl 1M, SDS 0,1% p/v, pH 8,45
- N',N',N',N' tetrametiletilendiamina (TEMED): se agregó en la preparación de los distintos geles a una concentración final de 0,01% v/v

- Solución AB3: 48 g acrilamida, 1,5 g bisacrilamida/100 ml de solución
- Solución AB6: 46,5 g acrilamida, 3 g bisacrilamida/100 ml de solución
- Persulfato de amonio al 10% p/v. Se coloca durante la preparación de los geles en concentración final 0,075% v/v en gel apilador, de 0,05% v/v en gel espaciador y 0,03% v/v en gel separador.

- Buffer de muestra: Tris-HCl 50 mM, pH 6.8, glicerol 12% p/v, SDS 4%, Servablue-G 0,01% p/v, pH 6,8, con o sin β -mercaptoetanol (2-ME) 5% v/v, para obtener condiciones reductoras o no reductoras, respectivamente.

- Solución fijadora: metanol al 50% v/v, ácido acético al 10% v/v.

- Solución colorante: Coomassie Brilliant Blue R-250 al 0,192% p/v disuelto en agua/metanol/ácido acético (10:10:4).

- Solución decolorante: 10% v/v ácido acético.

➤ *Reactivos tinción con plata*

- Solución fijadora: 60 ml de 50% v/v acetona, 1,5 ml de 50% p/v TCA y 25 μ l de 37% v/v formaldehído.

- Solución de pretratamiento A: 50% v/v acetona.

- Solución de pretratamiento B: 100 μ l de 10% $\text{Na}_2\text{S}_2\text{O}_3$ en 60 ml de agua bidestilada.

- Solución de impregnación: 0,8 ml de 20% p/v AgNO_3 , 0,6 ml de 37% v/v formaldehído y 60 ml de agua bidestilada.

- Solución de revelado: 1,2 g de Na_2CO_3 , 25 μ l de 37% v/v formaldehído, 25 μ l de 10% p/v $\text{Na}_2\text{S}_2\text{O}_3$ y 60 ml de agua bidestilada.

- Solución de detención (Stop): 1% v/v ácido acético glacial.

Geles

Los geles se prepararon de modo de generar tres zonas diferentes:

- Gel Apilador: 0,0825 ml solución AB3/ml solución gel.

- Espaciador: 0,2 ml solución AB3/ml solución gel.

- Separador: 0,33 ml solución AB6/ml solución gel.

Desarrollo de los geles y condiciones de corrida

Las electroforesis se realizaron en miniplacas (10 cm x 7 cm x 1 mm) en un equipo BIO-RAD, modelo mini-protean II.

Las corridas se llevaron a cabo a voltaje constante. Se empleó voltaje bajo, aproximadamente 30V, hasta que las proteínas atravesaron el gel apilador.

Luego la diferencia de potencial se aumentó a 100V hasta finalizar la corrida. Se requirieron entre 4 y 5 h para completarlas.

Fijación, coloración y decoloración

Los geles se colocaron en la solución fijadora una vez finalizada la corrida durante al menos 45 min, luego se le adicionó la solución colorante con Coomassie Brilliant Blue R-250 durante un mínimo de 5 h y finalmente se decoloraron a temperatura ambiente durante 2 h, renovando el decolorante cada 40 min.

En los casos en los que realizó la tinción con plata se colocaron los geles durante 5 min en la solución fijadora. Posteriormente se lavaron y enjuagaron varias veces con agua bidestilada. Después se pusieron en las soluciones para realizar los pretratamientos A y B durante 5 y 1 min, respectivamente. Una vez enjuagados los geles se colocaron en la solución de impregnación durante 8 min. Después de enjuagarlos nuevamente se agregó la solución reveladora durante el tiempo necesario para revelar la imagen (aproximadamente 60 s) y se detuvo la reacción manteniendo los geles en contacto con ácido acético glacial 1% (v/v) durante 60 s.

Obtención de imágenes y análisis de los geles

Para procesar los geles, las imágenes se digitalizaron utilizando un scanner HP ScanJet G2710.

2.3.6 Análisis por cromatografía líquida rápida para proteínas (FPLC)

La cromatografía fue utilizada en este trabajo con el fin de caracterizar, separar o purificar las muestras proteicas de amaranto.

2.3.6.1 Cromatografía FPLC de exclusión molecular

Durante la ejecución de este trabajo se utilizó la cromatografía FPLC de exclusión molecular, que permite separar especies peptídicas por el tamaño que presenten. Se utilizó una columna Superdex Peptide 10/300 GL pre-empacada marca GE Healthcare.

Muestras

Las muestras liofilizadas fueron solubilizadas en buffer de corrida, buffer Na₂HPO₄ 50 mM/NaCl 0,15 mM, pH 8, para obtener una concentración no superior a los 20 mg de proteína soluble/ml. Se agitaron durante 1 h a 25°C y 500 rpm en un agitador térmico y luego se centrifugaron a 14000g a 20°C durante 20 min utilizando una microcentrífuga. El sobrenadante obtenido se filtró a través de filtros de nylon de 0,22 µm de poro y luego se inyectó en el equipo de FPLC, utilizando un loop de 200 µl.

Desarrollo

Para desarrollar las corridas cromatográficas se utilizó el equipo de ÄKTA Purifier (GE Healthcare). Todos los buffers de corrida empleados fueron filtrados y desgasificados a través de filtros de nylon de 0,22 µm de poro antes de su uso. Los ensayos se realizaron a temperatura ambiente. El solvente de corrida fue el buffer utilizado para solubilizar a la muestra. Inicialmente se equilibró la columna suministrando 50 ml del buffer correspondiente (equivalente a dos volúmenes de la misma) a una velocidad de flujo de 0,5 ml/min. Luego se procedió a inyectar la muestra y se eluyó a un flujo de 0,5 ml/min. Se colectaron las fracciones del eluato durante la totalidad de la corrida con un colector automático mientras se determinaron los valores de absorbancia a 214 nm de modo continuo a la salida de la columna. Los perfiles obtenidos se procesaron con el programa OriginPro 8.

Calibración

La columna empleada presenta un rango de exclusión de 100–7000 Da para proteínas globulares; se calibró con azul dextrano (volumen muerto, V₀), aprotinina (6512 Da), vitamina B12 (1855 kDa) y ácido hipúrico (179,14 Da). Luego se obtuvo una curva de calibración relacionando el logaritmo de la masa molecular (en Da) con el K_{AV}, siendo K_{AV}:

$$K_{AV} = (V_e - V_0) / (V_T - V_0)$$

donde,

V_e: volumen de elución en mililitros del pico de interés.

V_T: corresponde al volumen total del lecho (24 ml).

V₀: volumen muerto.

2.3.6.1 Cromatografía FPLC de interacción hidrofóbica

En este trabajo se utilizó la cromatografía FPLC de interacción hidrofóbica a efectos de separar las muestras proteicas de amaranto. En esta metodología de separación las especies se separan por partición entre la fase estacionaria, de características apolares, y la fase móvil de carácter más polar que el relleno de la columna. Se utilizó una columna Alkyl Superose HR 5/5 marca GE Healthcare.

Reactivos

–Buffer A: Na₂HPO₄ 50 mM pH 7, NH₄SO₄ 2M.

–Buffer B: Na₂HPO₄ 50 mM pH 7.

Muestras

Las muestras fueron dispersadas en buffer A a una concentración de 25 mg de proteína soluble/ml. Se agitaron durante 1 h a 25°C y 500 rpm en un agitador térmico y luego se centrifugaron a 14000g a 20°C durante 20 min utilizando una microcentrífuga. El sobrenadante obtenido se filtró a través de filtros de nylon de 0,22 µm de poro y luego se inyectó en el equipo de FPLC, utilizando un loop de 200 µl.

Desarrollo

Para desarrollar las cromatografías se utilizó el equipo de ÄKTA Purifier (GE Healthcare). Todos los buffers de corrida empleados fueron filtrados y desgasificados a través de filtros de nylon de 0,22 µm de poro antes de su uso. Los ensayos se realizaron a temperatura ambiente. Como solvente de corrida se empleó el mismo buffer utilizado para solubilizar a la muestra. Inicialmente se equilibró la columna con 5 volúmenes de columna del buffer A (5ml) a una velocidad de flujo de 0,5 ml/min. Se inyectaron 200 µl de la muestra. La proteína no retenida se eluyó con 5 volúmenes de columna de buffer A. La proteína retenida se eluyó empleando un gradiente de 0 a 100% de buffer B en 15 volúmenes de columna. Finalmente se volvió a equilibrar la columna con 5 volúmenes de columna de buffer A. Se utilizó una velocidad de flujo de 0,5 ml/min durante toda la corrida. Las fracciones del eluato se colectaron durante la totalidad de la corrida con un colector automático mientras se determinaron los valores de absorbancia a 214 nm de modo continuo a la salida de la columna. Los perfiles obtenidos se procesaron con el programa OriginPro 8.

2.3.7 Ensayo de Inhibición de HMGR *in vitro*

Se determinó la inhibición *in vitro* de la enzima 3-hidroxi-3-metil-coenzima A reductasa (HMGR) empleando el kit comercial marca SIGMA-ALDRICH ([CS1090](#)) de acuerdo a las indicaciones de uso, utilizando como buffer del ensayo buffer fosfato pH 7,4; 400 mM KCl.

Reactivos

-*Buffer de ensayo*: 1,4 g/l Na₂HPO₄; 0,24 g/l NaH₂PO₄) pH 7,4; KCl 400 mM.

-*NADPH*: 25 mg NADPH (95 % de pureza) en 1500 ul de buffer de ensayo.

-*HMG-CoA*: 5 mg HMG-CoA /ml de buffer de ensayo.

-*HMGR*: 0,5-0,7 mg HMGR/ml de buffer (50 mM tris-HCl pH=7,5; 5 mM DTT, 1:200 v/v protease inhibitor cocktail (Sigma-Aldrich); 50% p/v glicerol).

-*Pravastatina*: 100 µM en buffer de ensayo.

Muestras

Las muestras se solubilizaron en el buffer del ensayo durante 1h con agitación constante a 500 rpm y 25°C empleando un agitador térmico Thermomixer Comfort marca Eppendorf, luego se centrifugaron durante 20 min a 14000g y temperatura ambiente utilizando una microcentrífuga (Hermle Z 233 MK-2).

Desarrollo

El dominio catalítico de la enzima HMGR cataliza, tal como se discutió previamente, la reducción del β-hidroxi-β-metilglutaril-coenzima A (HMG-CoA) dependiente de NADPH para dar mevalonato y coenzima A. La reacción fue monitoreada determinando la velocidad de consumo de NADPH mediante la medición de absorbancia a 340 nm. En presencia de sustancias que inhiban la actividad de la HMGR, las tasas de velocidad de consumo de NADPH serán menores.

Las cinéticas de reacción se realizaron en microplacas de 96 pocillos; para la medición de absorbancia a 340 nm se utilizó un lector de placas BioTek Synergy HT. En cada pocillo de la microplaca se colocó en el orden indicado buffer de reacción (para alcanzar un volumen final de reacción fijo de 200 µl), 20 µl de muestra, 10 µl solución de NADPH (concentración final 400 µM) y 12 µl HMG-CoA (concentración final 0,3 mg/ml). Las reacciones se iniciaron (tiempo cero) por adición de 12 µl de HMGR (concentración final 6 µg/ml). Se incubó a 37°C, agitando energicamente la placa durante 10 s y se determinó la absorbancia a 340 nm durante 10 min. Durante el ensayo se incluyó un control

positivo de inhibición, utilizando pravastatina, y un control de actividad máxima de HMGR, y un blanco de enzima (sin enzima). Todas las medidas se realizaron al menos por duplicado. En la TABLA II.1 indican las cantidades y el orden de agregado de los reactivos en las mezclas de reacción de inhibición HMGR *in vitro*.

TABLA II.1 Detalle de mezclas de reacción de Inhibición HMGR *in vitro*.

Orden	Reactivo	Concentración Final	Blanco Enzima	Blanco Sustrato	Inhibición	Actividad Máxima	Muestra
1	Buffer		178 µl	188 µl	181 µl	182 µl	156 µl
2	Pravastatina	500 nM	-	-	1 µl	-	
3	Muestra		-	-	-	-	20 µl
4	NADPH	400 µM	10 µl	10 µl	10 µl	10 µl	10 µl
5	HMG-CoA	0,3 mg/ml	12 µl	-	12 µl	12 µl	12 µl
6	HMGR	6 µg/ml	-	2 µl	2 µl	2 µl	2 µl

Cálculos

La actividad de HMGR se expresó como Unidades/mg de proteína. Una unidad convierte 1,0 µmol de NADPH a NADP⁺ por min a 37°C por mg de proteína (Unidades/mgP).

$$\text{Unidades/mg P} = \frac{\frac{\Delta \text{Abs}}{\text{min}} \text{muestra} - \frac{\Delta \text{Abs}}{\text{min}} \text{blanco enzima}}{12,44 \times V \times 0,6 \times 0,55}$$

donde:

$\frac{\Delta \text{Abs}}{\text{min}} \text{muestra}$: pendiente de la curva de absorbancia vs tiempo correspondiente a cada muestra.

$\frac{\Delta \text{Abs}}{\text{min}} \text{blanco enzima}$: pendiente de la curva de absorbancia vs tiempo correspondiente al ensayo blanco enzima.

0,2: volumen total de reacción en ml.

12,44: coeficiente de extinción molar del NADPH a 340 nm x2 ($\epsilon=6,22\text{mM}^{-1}\text{cm}^{-1}$)

V: volumen de enzima empleado en la reacción en ml (2 µl).

0,6: concentración de la enzima en mg P/ml (0,5–0,7 mg/ml, de acuerdo al lote de la enzima).

0,55: camino óptico en cm para las cubetas de la placa.

Las pendientes fueron obtenidas mediante el ajuste lineal de la curva de Abs en función del tiempo en el rango de 3 a 10 min.

El porcentaje de inhibición % *INH* fue calculado mediante la siguiente expresión:

$$\% INH = \left(1 - \frac{\frac{\text{Unidades}}{\text{mg P}} \text{ Muestra}}{\frac{\text{Unidades}}{\text{mg P}} \text{ Actividad Máxima}} \right) \times 100$$

La actividad máxima se determinó en iguales condiciones que las utilizadas para la muestra. La evaluación de cada condición (realizada por triplicado) incluyó los controles de blanco sustrato, blanco enzima y la medida de actividad máxima por duplicado.

2.3.8 Análisis estadístico de los resultados

Los resultados obtenidos fueron evaluados estadísticamente empleando el programa GraphPad Prism 6. Se efectuó el análisis de varianza (ANOVA), y para la comparación de medias se utilizaron el Test de la menor diferencia significativa (LSD) descrita por (Fisher, 1954) y el Test HSD (Honestly-Significant-Difference) de Tukey (Tukey, 1949), ambos con un nivel de significación de $\alpha = 0,05$ (95% de confianza).

2.4 RESULTADOS Y DISCUSIÓN

2.4.1 Obtención y caracterización del aislado proteico de amaranto y sus hidrolizados

A partir de la harina desgrasada de amaranto se prepararon los aislados proteicos. El rendimiento de la obtención de los mismos se calculó después de realizar varias preparaciones y resultó igual a $6,6 \pm 0,3$ g de aislado proteico cada 100 g de harina desgrasada.

El contenido de proteínas, hidratos de carbono, cenizas y humedad del aislado proteico (Ais) se determinó por las metodologías detalladas en Materiales y Métodos del presente capítulo. La composición resultante se detalla en la **TABLA II.2**. Los valores obtenidos coinciden, en general, con los valores previamente determinados para otras especies y variedades de amaranto (Condés y col., 2009; Vecchi, 2007). En particular el contenido de proteínas del aislado fue $84,2 \pm 0,5$ g/100g de aislado, siendo del orden de los valores obtenidos en preparaciones realizadas en nuestro laboratorio (Condés y Col., 2009; $75,4 \pm 0,5$ g/100 g aislado; Orsini y col., 2011; $72,4 \pm ,4$ g/100 g aislado; Ventureira y col., 2012: $83,1 \pm 0,5$ g/100 g aislado). Los aislados de amaranto normalmente presentan un contenido de proteínas inferior al que presentan los aislados proteicos de soja, debido principalmente al hecho que estos aislados son obtenidos a partir de harina integral por lo que contienen una mayor cantidad residual de hidratos de carbono, representados principalmente por polisacáridos presentes en la fibra, y sales. Los aislados proteicos de amaranto están principalmente constituidos por albúminas y globulinas, plegadas en una conformación cercana a la nativa (Martínez y Añón 1996).

TABLA II.2 Composición del aislado proteico

Composición centesimal (g/100g de muestra)				
Proteínas	Hidratos de Carbono	Cenizas	Humedad	Fibras
84,2 ± 0,5	3,3 ± 0,1	2,7 ± 0,2	5,9 ± 0,1	3,9

El contenido de fibras se determinó por diferencia

A partir del aislado se prepararon los hidrolizados, por hidrólisis exógena con alcalasa, de ahora en más denominados Hid, y por simulación del proceso de digestión gastrointestinal, Dig, como se detalla en la Sección 2.3.3 y 2.3.4 de Materiales y Métodos del presente capítulo.

La alcalasa es obtenida por fermentación sumergida de una cepa de *Bacillus licheniformis* y su principal componente enzimático, es la serinproteasa subtilisina de Carlsberg, (EC 3.4.21.62). Esta enzima actúa a pHs comprendidos entre 6 y 10, y si bien es una proteasa relativamente inespecífica, estudios sobre su cinética de hidrólisis señalan que la mínima secuencia de reconocimiento para su acción es un residuo de glicina flanqueado por dos aminoácidos hidrofóbicos voluminosos (Bratovanova y Petkov, 1987). La alcalasa libera, por lo tanto, en su mayoría péptidos con aminoácidos hidrofóbicos en su extremo C-terminal. La acción de la alcalasa se siguió midiendo el grado de hidrólisis (GH) a diferentes tiempos de reacción, mediante la determinación, como se indicó previamente, de la cantidad de grupos amino libres con el reactivo o-ftalaldehído (OPA). Para eso se tomaron alícuotas representativas de las dispersiones del aislado de amaranto tratadas a pH 9 y 37°C a distintos tiempos de reacción (0, 1, 2, 3, 5, y 6 h). Como control se incluyó una alícuota tomada inmediatamente después del agregado de la enzima (0 min). Los resultados se muestran en la **FIGURA II.7 A**. El proceso de simulación de digestión gastrointestinal fue monitoreado del mismo modo. Se tomaron alícuotas representativas después de dispersar la muestra en el fluido símil saliva (E1), inmediatamente después del agregado del fluido símil gástrico (E2); luego del agregado del fluido símil duodenal, antes de llevar a pH 7 e incubar 1h (E3) y al finalizar el tratamiento térmico para detener el proceso de digestión (E4). Todas las alícuotas fueron calentadas durante 10 min a 80°C en baño de agua. Los resultados obtenidos se muestran en la **FIGURA II.7 B**.

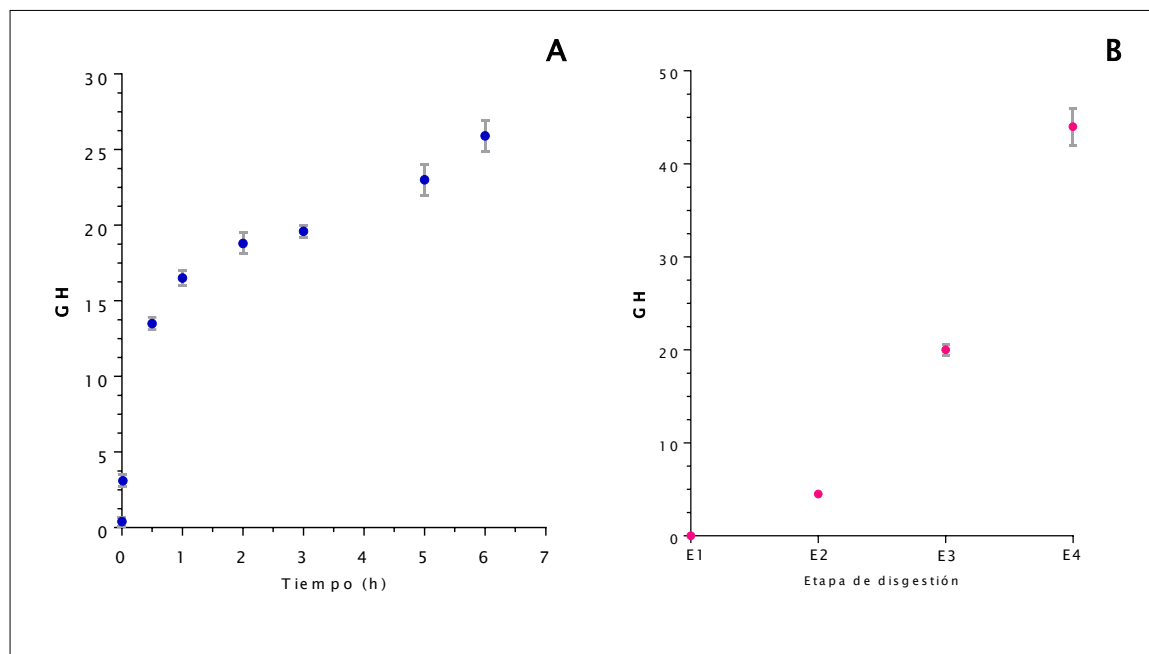


FIGURA II.7 Cinética de hidrólisis del aislado proteico de amaranto **(A)** por alcalasa y **(B)** sometido a digestión gastrointestinal simulada. Las barras de error corresponden a la desviación estándar.

En los dos procesos de hidrólisis se observa, como es esperado, un incremento del grado de hidrólisis con el tiempo de reacción alcanzando un GH % final de $25 \pm 3\%$ y $44 \pm 2\%$ para Hid y Dig, respectivamente. Resulta además evidente que la acción de la alcalasa logró un menor número de rupturas de las uniones peptídicas presentes, hecho que fue superado por la acción de la pepsina y las enzimas pancreáticas.

Durante la simulación del proceso de digestión gastrointestinal se observó un incremento del GH de $0 \pm 0,2\%$ en E1; a $4,5 \pm 0,4\%$ en E2; a $20 \pm 0,6\%$ en E3, a $44 \pm 1\%$ en E4. El fluido símil saliva no presenta actividad proteolítica, y su agregado responde al pedido de unificación de protocolos de digestión gastrointestinal realizado por Minekus y col. (2014). El incremento del GH observado inmediatamente después del agregado del fluido símil gástrico es producto de la acción del medio ácido y de la acción proteolítica de la pepsina en un corto tiempo comprendido entre la adición del reactivo y el tratamiento térmico para detener la reacción enzimática. Luego del período de incubación con pepsina, el GH se incrementó a $20 \pm 0,6\%$ y alcanzó, al finalizar el período de incubación con pancreatina -mezcla de enzimas proteolíticas-, un valor final de $44 \pm 1\%$. El grado de hidrólisis final y el contenido de proteínas (determinado por microkjeldahl) de las tres muestras obtenidas (Ais, Hid y Dig) se detallan en la **TABLA II.3**.

TABLA II. 3 Contenido de proteínas y grado de hidrólisis porcentual (GH) del aislado de amaranto (Ais), Aislado hidrolizado con alcalasa durante 6h (Hid) y aislado sometido a un proceso de digestión gastrointestinal simulada (dig).

	Ais	Hid	Dig
Proteínas	84,2 ± 0,5	70 ± 1	50,0 ± 0,3
GH	0%	25 ± 3%	44 ± 2%

Tanto la composición en proteínas como el grado de hidrólisis porcentual obtenidos coinciden con los valores informados previamente en nuestro laboratorio. El contenido de proteínas de los hidrolizados, en particular del obtenido en el proceso de digestión gastrointestinal simulado, es inferior al correspondiente al aislado proteico debido a la incorporación de sales durante su obtención.

La caracterización de los perfiles polipeptídicos presentes en las diferentes muestras se llevó a cabo mediante electroforesis y cromatografía de filtración en gel (FPLC). En la **FIGURA II.8** se muestran los perfiles electroforéticos en geles de tricina (SDS-PAGE) en condiciones reductoras y no reductoras correspondientes a las tres muestras.

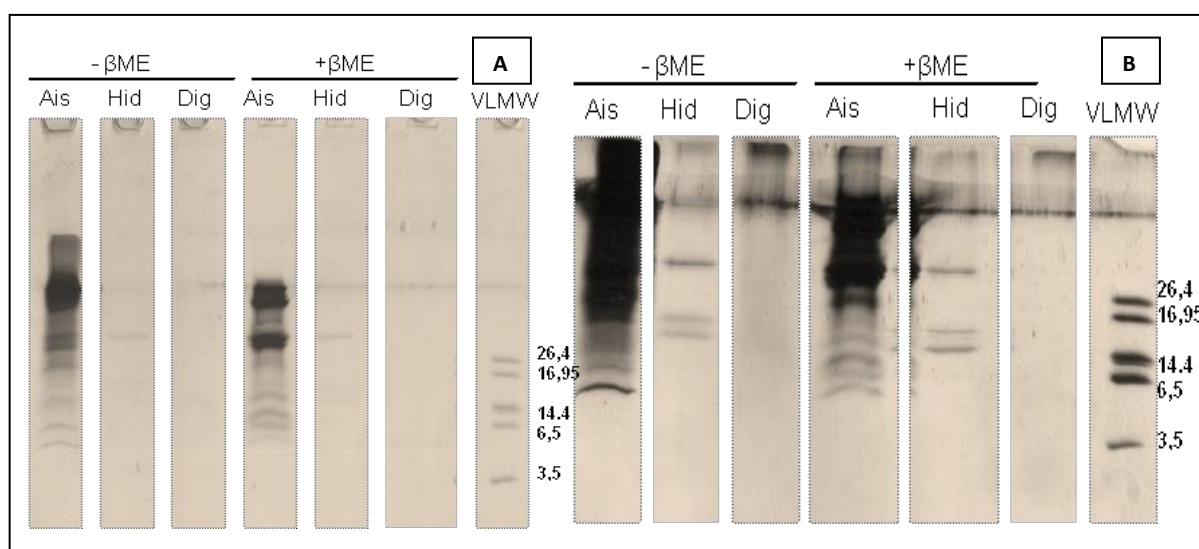


FIGURA II.8 Electroforesis Tricina-SDS-PAGE en condiciones reductoras (+βME) y no reductoras (-βME) del aislado de amaranto (Ais), su hidrolizado (Hid) y digerido (Dig) con tinción convencional (A) y tinción de plata (B). VLMW: patrones de muy bajo peso molecular.

Los perfiles obtenidos revelan diferencias significativas en la composición polipeptídica de las muestras. En el caso del aislado de amaranto, el cual está constituido principalmente por las globulinas de almacenamiento del grano (11S, globulina-P y 7S), se obtuvo el perfil electroforético típico en concordancia

con lo informado por otros autores (Martínez y Añón, 1996; Abugoch y col., 2010). Se puede observar la presencia de bandas de 67 kDa, constituyentes de la globulina 7S (Quiroga y col., 2010), bandas de masa molecular de 56–54 kDa correspondientes a las subunidades AB de las globulinas 11S y al polipéptido P-54 de la globulina-P; bandas de aproximadamente 30 kDa y 20 kDa correspondientes a los polipéptidos A y B de la globulina 11S, y bandas de masa molecular menor a 20 kDa, que pueden corresponder a la fracción albúminas. También se observaron especies polipeptídicas de alta masa molecular que no pudieron entrar al gel y que se corresponden con la existencia de agregados proteicos solubles bajo las condiciones de ensayo.

En condiciones reductoras decrecen los agregados antes indicados y se disocian las subunidades AB de las globulinas 11S y su equivalente de globulina-P, incrementándose el contenido de polipéptidos ácidos y básicos.

Cuando el aislado es sometido al proceso de hidrólisis, tanto exógena como gastrointestinal simulada, el perfil electroforético se modifica radicalmente. En el caso de los hidrolizados obtenidos por acción de alcalasa, tanto en condiciones reductoras como no reductoras, se pueden observar cantidades reducidas de polipéptidos de masas moleculares aproximadas de 55, 17 y 14 kDa. En tanto que en el caso de los hidrolizados obtenidos por digestión gastrointestinal simulada no fue posible detectar ninguna especie polipeptídica en ninguna de las condiciones de ensayo, lo que sugiere la presencia de péptidos de masa molecular inferior a 3,5 kDa. Estos resultados son concordantes con los grados de hidrólisis determinados.

Por último, cabe señalar que en el caso de Hid y Dig desaparecen también del perfil los agregados solubles presentes en el aislado.

A efectos de visualizar los cambios sufridos por las proteínas de amaranto como consecuencia de la hidrólisis, se analizaron también las variaciones en la masa molecular de las mismas mediante cromatografía de exclusión molecular, empleando columnas de gel filtración y un sistema FLPC. Teniendo en cuenta que las muestras se encuentran enriquecidas en polipéptidos de baja masa molecular, producto de los procesos de degradación durante los procesos de hidrólisis empleados, se seleccionó para este ensayo la columna Superdex Peptide, cuyo rango de exclusión es de 7000 a 100 Da. Los perfiles de elución obtenidos se muestran en la **FIGURA II.9**.

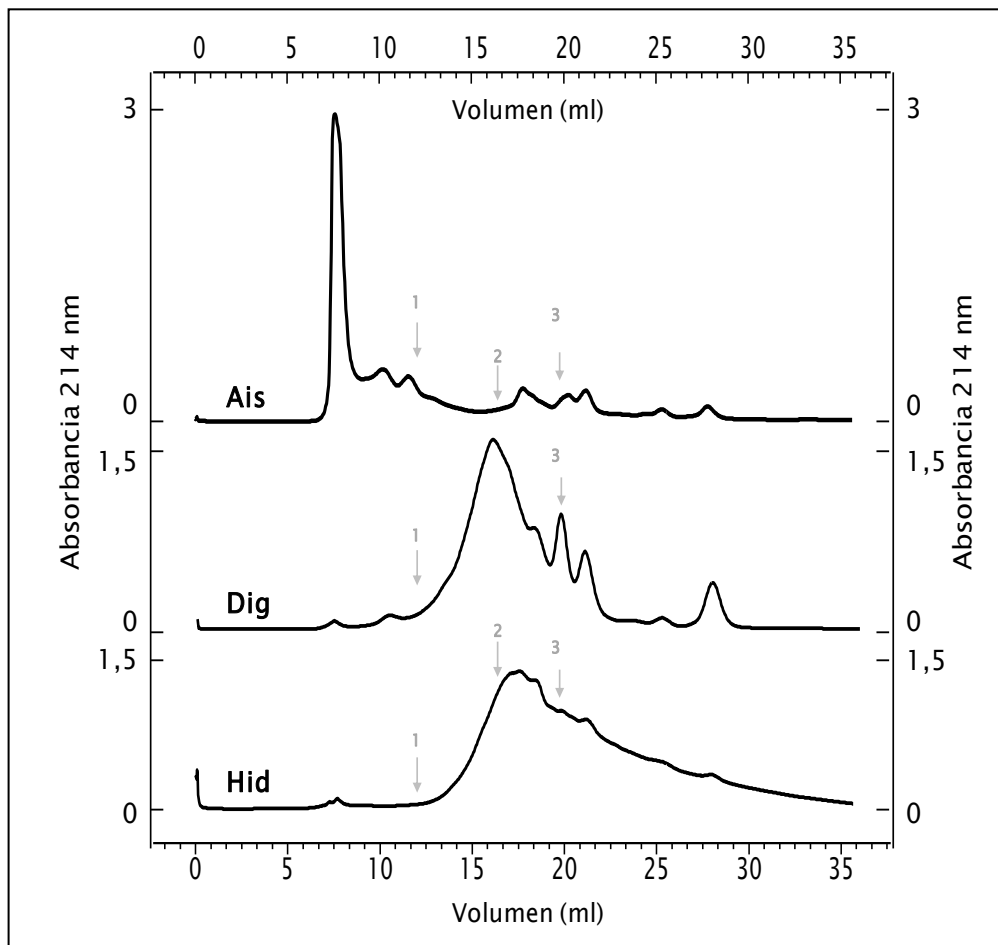


FIGURA II.9 Cromatograma del aislado de amaranto (Ais) e hidrolizados obtenidos por digestión gastrointestinal simulada (Dig) y acción de alcalasa (Hid). Las muestras fueron eluidas con buffer Na_2HPO_4 50 mM/ NaCl 0,15 mM, pH 8. La columna se calibró con azul dextrano (volumen muerto, V_0), aprotinina (1), vitamina B12 (2) y ácido hipúrico (3).

El aislado de amaranto está constituido principalmente por especies polipeptídicas cuyas masas moleculares son superiores a 7000 Da (Scilingo y col., 2002; Martínez y Añón, 1996), las cuales eluyen en el volumen muerto. Las especies cuyos pesos moleculares son menores a 7000 Da, constituidas principalmente por polipéptidos de menos de 70 residuos aminoacídicos, que pueden corresponder a especies polipeptídicas que se encuentran libres, como los inhibidores de tripsina (Tamir y col., 1996) o a productos de hidrólisis provenientes de la ruptura de polipéptidos mayores.

Los perfiles cromatográficos correspondientes al Dig y el Hid mostraron un incremento de los picos correspondientes a las especies polipeptídicas de menor tamaño, inferiores a 6.500 Da. Al aumentar el grado de hidrólisis se observó un incremento en la población de especies de menor peso molecular. Resulta evidente que la simulación del proceso de digestión gastrointestinal conduce a la liberación de péptidos de largo de cadena inferior a 10 aminoácidos.

2.4.2 Determinación de la Inhibición de HMGR

La potencial actividad hipocolesterolemizante de las proteínas y péptidos de amaranto se evaluó mediante el ensayo de inhibición *in vitro* de la enzima 3-hidroxi-3-metilglutaril-CoA reductasa (HMGR). En este ensayo se enfrenta el sustrato, hidroximetil-CoA (HMG-CoA) con la enzima HMGR que cataliza su reducción dependiente de NADPH a coenzima A y mevalonato. La reacción es monitoreada, como se indicó previamente, por el descenso de la absorbancia a 340 nm debido al consumo de NADPH. De este modo, el método permite estudiar la habilidad de una muestra para modular la actividad de la HMGR observándose los cambios en la velocidad de conversión de NADPH a NADP⁺ mediante la lectura espectrofotométrica.

Se analizó el efecto de la fracción proteica soluble de las proteínas o péptidos de amaranto presentes en cada muestra en estudio (Ais, Hid, Dig) sobre la actividad enzimática de la HMGR empleando un kit comercial marca Sigma-Aldrich. Antes de iniciar los ensayos de inhibición se estudió la solubilidad de las muestras en el buffer de ensayo del kit, cuantificando proteína soluble por el método de Lowry (Lowry y col., 1951). Los valores de absorbancia obtenidos para la curva de calibración, preparada con el buffer provisto en el kit, fueron superiores a los obtenidos rutinariamente en nuestro laboratorio. Del mismo modo, la absorbancia desarrollada en las muestras fue superior a la esperada de acuerdo a la cantidad de muestra solubilizada y su contenido de proteína. Estos resultados nos llevaron a pensar en la presencia de alguna/s sustancia/s que interfiera/n con el método de cuantificación de proteínas, como el caso del doecilsulfto de sodio (SDS). Para comprobar esta hipótesis se repitieron los ensayos de solubilidad de proteínas y/o polipéptidos modificando el método de cuantificación. Se optó por el método del ácido bicinonínico (Smith y col., 1985) por tratarse de un método con mayor tolerancia a sustancias interferentes. Sin embargo, la respuesta obtenida mostró la misma tendencia que la observada en los ensayos previos, confirmando así la presencia de sustancias interferentes. Sin embargo, no pudiendo descartar la presencia de componentes peptídicos, además de proteínas, se optó por cuantificar la proteína soluble por el método de microkjeldahl, determinándose además la posible presencia de proteína en el buffer incluido en el kit. Este análisis reveló la presencia de proteína en el buffer del kit, probablemente albúmina, utilizada frecuentemente como estabilizante para otras proteínas solubles (por ejemplo,

enzimas lábiles), lo que impide el uso del mismo como tal en nuestros estudios. En forma paralela se contactó a los representantes de Sigma–Aldrich a efectos de solicitar información sobre la composición del buffer del kit, la cual no fue suministrada por ser de su propiedad. Los reclamos efectuados por nosotros dieron lugar a la incorporación de una leyenda, en la información del kit, advirtiendo que el buffer de ensayo no es compatible con los ensayos de determinación de proteínas.

Por lo tanto, para poder llevar adelante los estudios de determinación de la actividad de la HMGR se decidió remplazar el buffer del kit. Se utilizó en su reemplazo un buffer fosfato (1,4 g/l Na_2HPO_4 , 0,24 g/l NaH_2PO_4) de pH 7,4 con la adición de KCl 400 mM (Bischoff y Rodwell, 1996; Frimpong y col., 1993; Polo y García de Bravo, 2006). La solubilidad del Ais, Dig e Hid en el nuevo buffer de ensayo se determinó de acuerdo al protocolo detallado en Pág. 88 del presente trabajo y los resultados obtenidos se presentan en la **TABLA II.4**.

TABLA II. 4 Solubilidad % (g de proteína en el sobrenadante/100 g proteína total) en buffer fosfato pH 7,4; KCl 400 mM del Ais, Hid y Dig.

	Ais	Hid	Dig
Solubilidad %	49±1	95±2	98±2

Los valores encontrados resultaron similares a los valores de solubilidad del aislado de amaranto y sus hidrolizados en agua (Sabbione, 2015). A baja fuerza iónica y pH cercano a 6–7 las globulinas de amaranto, muestran una solubilidad reducida. La misma se ve francamente incrementada por efecto de la hidrólisis la cual provoca una disminución sensible del tamaño molecular de los péptidos presentes en ambos hidrolizados.

Seguidamente se evaluó la actividad de la HMGR en función de la concentración de proteína ($\mu\text{g}/\text{ml}$) de Ais, Dig e Hig. La respuesta observada se expresó como el porcentaje de inhibición (% INH) haciendo uso de los cálculos detallados en la Sección 2.3.7 de Materiales y Métodos. En la **FIGURA II.10** se muestra un ejemplo de las cinéticas de reacción obtenidas para los ensayos de actividad máxima y los controles.

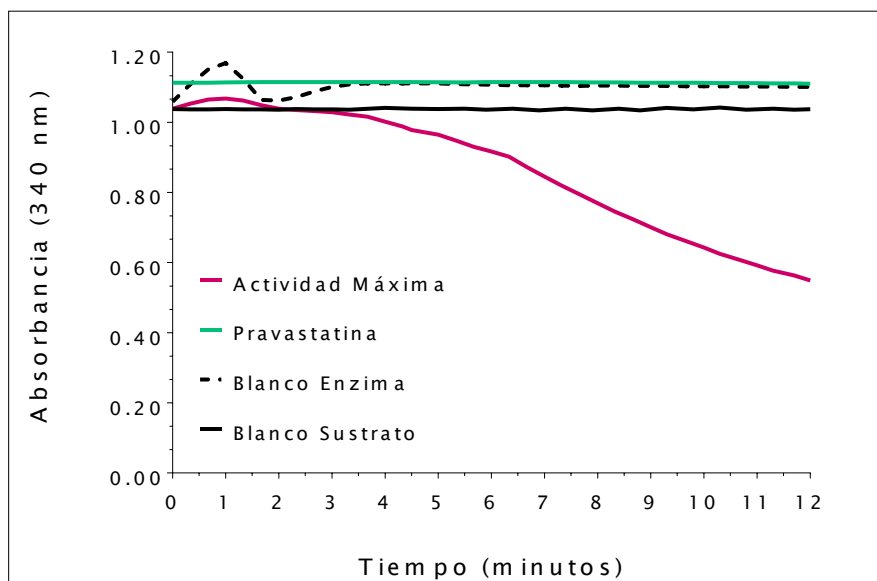


FIGURA II.10 Actividad de la HMGR.

Para cada condición ensayada se obtuvieron las curvas de absorbancia a 340 nm en función del tiempo, por triplicado. A partir de estos gráficos, se calcularon los porcentajes de inhibición de la enzima. Los resultados obtenidos se muestran en la **TABLA II.5** y **FIGURA II.11**.

Los resultados obtenidos indican que los polipéptidos y/o péptidos presentes tanto en el Ais como en el Hid y el Dig son capaces de inhibir la acción de la HMGR. Los valores de inhibición más altos fueron alcanzados por Ais seguido por Hid. El proceso de digestión gastrointestinal simulado, Dig, redujo ligeramente el grado de inhibición enzimática alcanzado, lo que indicaría que algunos péptidos bioactivos han sido hidrolizados durante este proceso transformándose en productos no activos.

En ninguna de las muestras y/o condiciones ensayadas, se logró alcanzar una inhibición del 100 % de la actividad enzimática. Este hecho podría deberse a la presencia de sustancias que estarían actuando como inhibidores no competitivos y/o a la existencia simultánea de inhibidores y activadores de la enzima. La HMGR es una enzima multisustrato, con un mecanismo de acción del tipo *pin-pong* (Panda y col., 2011) y sus inhibidores pueden mostrar distintos tipos de inhibición dependiendo del sustrato que se considere (HMGR-CoA y NADPH).

TABLA II.5 Porcentajes de inhibición (%INH) de HMGR del Ais, Hid y Dig a diferentes concentraciones de proteína expresada en $\mu\text{g/ml}$.

Ais			Hid			Dig			
Proteína ($\mu\text{g/ml}$)									
10	52	104	7	33	66	10	30	62	90
% INH									
58 ± 7^a	63 ± 10^a	53 ± 12^a	55^{ab}	47 ± 1^{ab}	43 ± 2^{ab}	45 ± 4^{ab}	24 ± 4^{abc}	12 ± 4^c	5 ± 10^c

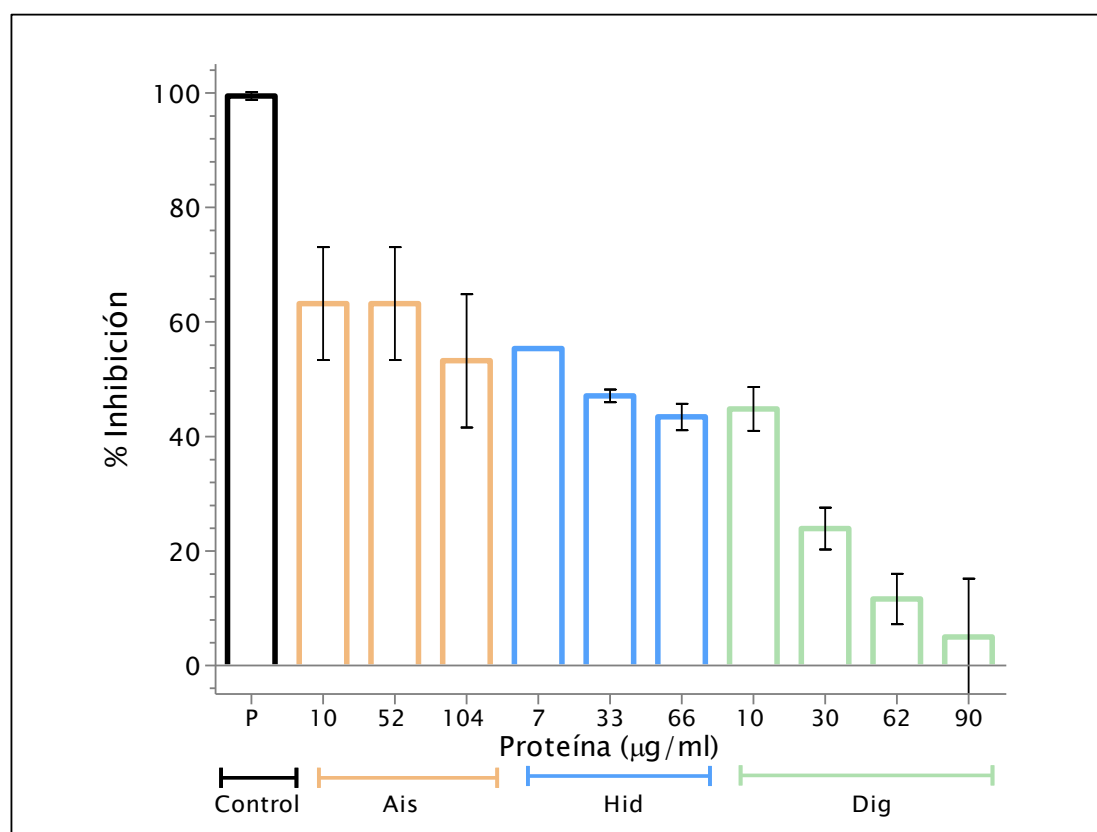


FIGURA II.11 Porcentaje de inhibición de HMGR del Aislado (Ais), aislado hidrolizado (Hid) y aislado digerido (Dig). P corresponde al control de inhibición con Pravastatina. Diferentes letras sobre las barras indican diferencia significativa ($\alpha = 0,05$, Tukey).

Bajo las condiciones de ensayo empleadas las muestras de Ais e Hid no mostraron un comportamiento dosis–respuesta. En el caso del Dig, si bien se obtuvo una respuesta del porcentaje de inhibición con la concentración de muestra, este es contrario al esperado; dado que la inhibición se incrementa al reducirse la concentración de proteína de la muestra.

En el caso de Ais e Hid la falta de un comportamiento dosis–respuesta podría ser atribuido, tal como se indicó previamente, a la existencia de especies polipeptídicas, resistentes a la acción de la alcalasa, capaces de inhibir o activar

la HMGR o a la aparición de péptidos productos de la hidrólisis con estas características. Otra causa probable del efecto encontrado es que el rango de concentración de proteína seleccionado para realizar las determinaciones de la actividad enzimática no sea el adecuado y que el mismo deba ser, probablemente, reducido. Por otro lado, el hecho de haber modificado la composición del buffer de reacción podría incidir en la conformación y/o comportamiento del sitio catalítico de la enzima.

Resultados informados por Soares y col. (2015) a partir de hidrolizados de proteínas de amaranto obtenidos por acción de una mezcla de enzimas (tripsina, α -quimioripsina y peptidasas de la mucosa intestinal porcina) mostraron la existencia de tri y tetrapéptidos, capaces de inhibir la HMGR. El mayor grado de inhibición, 55%, fue alcanzado por el hidrolizado, seguido por el péptido VGVI el cual fue capaz de inhibir en un 50% la acción de la enzima. Para todos los otros péptidos identificados el porcentaje de inhibición alcanzado fue inferior variando entre 39 y 11%. Todos estos porcentajes de inhibición fueron obtenidos para una concentración de proteínas de 4 $\mu\text{g/ml}$. Soares y col., (2015) no mostraron resultados de inhibición en función de la concentración de los hidrolizados o péptidos identificados por lo que no se ha podido verificar si los mismos exhiben un comportamiento dosis-respuesta.

Los porcentajes de inhibición alcanzados en nuestro estudio son superiores al obtenido con el hidrolizado incluido en el trabajo de Soares y col., para una concentración proteica 2,5 veces más alta.

Cabe por último señalar que prácticamente no existen otros antecedentes en bibliografía relacionados con la potencial acción inhibitoria de proteínas y/o péptidos de amaranto de la HMGR.

2.4.3 Fraccionamiento del aislado sometido a digestión gastrointestinal simulada

Teniendo en cuenta el incremento del % INH observado al reducirse la concentración de la muestra se decidió fraccionar con el fin de identificar a las sustancias responsables de la bioactividad. Para ello se decidió emplear la cromatografía como paso de purificación de la muestra sometida a digestión gastrointestinal simulada. Teniendo en cuenta los antecedentes descritos en bibliografía de inhibidores peptídicos de la HMGR (ver próxima sección) se decidió emplear una cromatografía (FPLC) de interacción hidrofóbica que separa

las especies proteicas de acuerdo a su carácter no polar, sin poder descartar un efecto del tamaño de las proteínas o péptidos. Las condiciones de corrida se describieron en Materiales y Métodos (2.3.6.1). El Dig se solubilizó en buffer Na_2HPO_4 50 mM pH 7 NH_4SO_4 2M y se eluyó con un gradiente lineal de 0 a 100% de buffer Na_2HPO_4 50 mM. Se realizaron varias corridas sucesivas obteniéndose cromatogramas superpuestos al que se muestra en la **FIGURA II.12**.

El eluido de la columna fue colectado en 7 fracciones de acuerdo a los picos observados en los cromatogramas. Las fracciones I y II corresponden a componentes de mayor carácter hidrofílico. Las fracciones III, IV, V, VI y VII presentan hidrofobicidad creciente en el mismo sentido de su numeración.

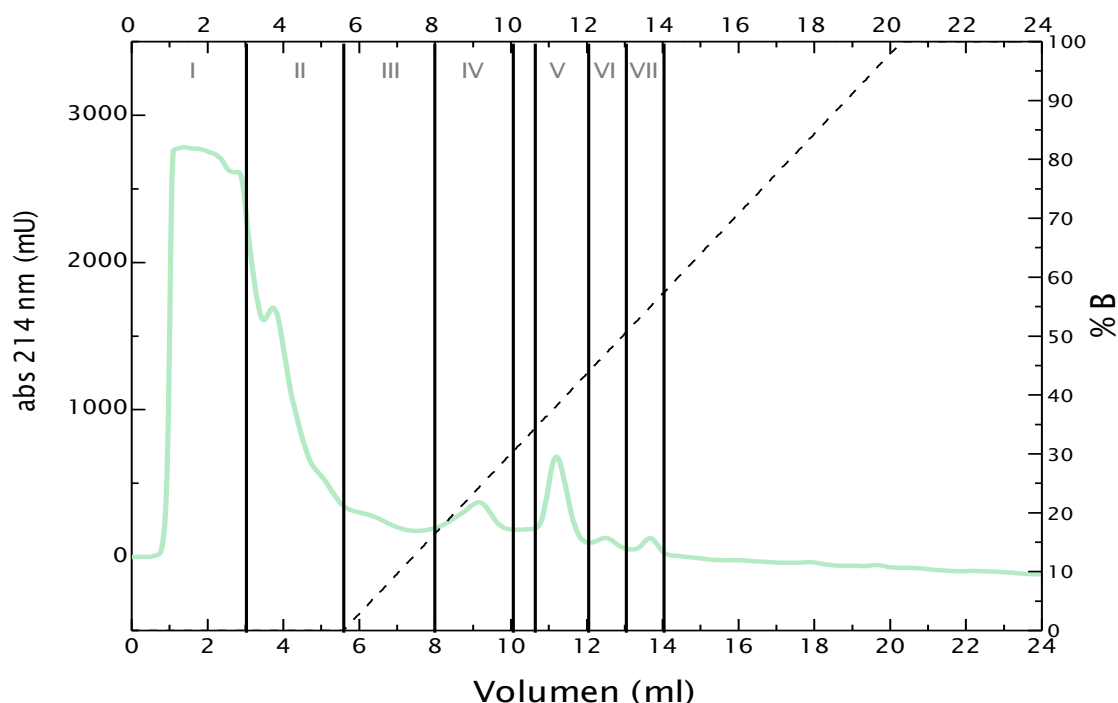


FIGURA II.12 Cromatografía de interacción hidrofóbica FPLC (columna Alkyl Superose, GE Healtcare) del aislado sometido a digestión gastrointestinal simulada (Dig) en buffer Na_2HPO_4 50 mM pH 7). En línea punteada se indica el gradiente de NH_4SO_4 de 2 a 0M.

A cada fracción colectada se le analizó su potencial acción inhibitoria de la HMGR. Como las fracciones eluidas contienen cantidades variables de NH_4SO_4 primero se estudió la actividad de la HMGR reemplazando el volumen de muestra empleado en los ensayos por soluciones de NH_4SO_4 1, 1,5 y 2 M en buffer Na_2HPO_4 50 mM pH 7. Dado que la concentración final de NH_4SO_4 en las mezclas motivo de estudio es 0,1–0,2 M, las mismas se encuentran comprendidas en el rango de no precipitación por agregado de sal (Wingfield, 2001), conservándose la actividad de la HMGR en estas condiciones.

Seguidamente se estudió la potencial actividad inhibitoria de todas las fracciones colectadas, no encontrándose inhibición de la actividad enzimática.

La ausencia de actividad podría deberse a una baja concentración de péptidos bioactivos en las fracciones cromatográficas obtenidas, a pesar de haberse colectado las fracciones correspondientes a más de una corrida; concentración inferior a la presente en el digerido sin fraccionar.

2.4.4 Bibliotecas de inhibidores de HMGR

Con el objetivo de encontrar inhibidores de la HMGR en la fracción proteica de amaranto se realizó una búsqueda bibliográfica exhaustiva de inhibidores de esta enzima, de secuencias conocidas. Como resultado de esta búsqueda se encontraron antecedentes sólo para soja y más recientemente en amaranto.

En hidrolizados de aislados proteicos de soja fueron identificados tres péptidos con actividad inhibitoria de la HMGR: LPYP, LPYPR e SLYPR provenientes de la globulina 7S de soja (glicinina). Pak y col. (2008; 2012; 2005; 2007). Estos autores demostraron que el ácido glutámico, aminoácido E, era un residuo reconocido por el bolsillo de unión de la enzima a HMG-CoA. Sin embargo, no encontraron, mediante estudios con dicroísmo circular, patrones relacionados con una conformación a nivel de estructura secundaria de los péptidos IAVE, IVAE, YAVE y YVAE (Pak y col., 2006). Sí lograron hallar una conformación de giro en el extremo N-terminal de otra serie de péptidos tales como FFYVAE, FPYVAE y FGYVAE. Todos estos péptidos fueron capaces de inhibir a la HMGR de manera competitiva (Park y col., 2008). Este tipo de estructura se encontraría favorecida por la presencia de residuos de prolina. Finalmente, estos autores, haciendo uso de diferentes estrategias como: la superposición de las estructuras de los tetrapéptidos con las correspondientes a las estatinas en los complejos HGMR-estatina (PDB 1HW8 (compactina), 1HW9 (simvastatina), 1HWI (fluvastatina), 1HWJ (cerivastatina) y 1HWL (rosuvastatina)); el análisis de componentes principales (PCA) y el estudio mediante dinámica molecular del comportamiento conformacional del modelo cíclico del hexapéptido de glicina, lograron demostrar la importancia de la conformación tipo giro β en la estructura de los PB inhibidores de la enzima en estudio y arribaron al diseño de un nuevo inhibidor muy potente, el SFGYVAE (K_i estimada $12 \pm 0,4$ nM) (Pak y col., 2012). En base a estos conocimientos estos autores construyeron una biblioteca sintética de péptidos (TABLA II.6).

Tabla II.6 Biblioteca de inhibidores peptídicos derivada de los trabajos de Pak y col. (2008a; 2008b, 2012) Los residuos se representan en colores de acuerdo a sus propiedades fisicoquímicas: Residuos ácidos (verde): D, E; básicos (celeste): R, H y K; hidrofóbicos (rojo): I,V,F,Y,W, P.

Péptido	IC50 (μM) <i>Homo sapiens</i>	IC50 (μM) <i>Rattus norvegicus</i>
AFGYVAE	0,49	-
DFGYVAE	0,16	-
efgyvae	0,24	-
FFGYVAE	0,32	-
FFYVAE	2,5	-
FGYVAE	0,4	-
FPYVAE	1,4	-
FVAE	-	43,8
GFGYVAE	0,27	-
GFPDGG	-	1,5
GFPEGG	-	1,7
GFPTGG	-	16,9
GLPDGG	-	22,3
GLPEGG	-	27,2
GLPTGG	-	19,4
IAVE	52	75,2
IAVP	-	97
IAVPGEVA	201	152
IAVPTGVA	152	-
IFGYVAE	0,35	-
IVAE	75	44,1
LFGYVAE	0,37	-
LPYP	-	484
PFGYVAE	0,43	-
SFGYVAE	0,033	-
TFGYVAE	0,26	-
VFGYVAE	0,45	-
YAVE	44	52,6
YVAE	41	41,8

Cabe señalar, que todas las secuencias identificadas contienen una proporción importante de aminoácidos hidrofóbicos, característica bastante común en péptidos que presentan diferentes actividades biológicas, por ejemplo, actividad antihipertensiva debida a la inhibición de la enzima convertidora de angiotensina.

Soares y col. (2015), como se mencionó oportunamente, estudiaron la capacidad de inhibición de HMGR de hidrolizados de proteínas de amaranto e identificaron tri y tetrapéptidos activos. Las secuencias identificadas se detallan en la **TABLA II.6**. Estos autores indican que dado que los péptidos bioactivos identificados no tienen una relación estructural con las estatinas los mismos

pueden representar una clase nueva de inhibidores de la HMGR y podrían interactuar directamente con la enzima bloqueando así el camino del mevalonato.

Por su parte, Li y col., (2015), empleando una biblioteca de *Phago Display*, identificaron 12 dodecapéptidos capaces de inhibir la HMGR. Para evaluar la afinidad de unión de los péptidos con la enzima realizaron un análisis de docking empleando la estructura PDB 1HWK y considerando las particularidades del sitio activo de HMGR, en el que participan residuos de la subunidad A y subunidad B de HMGR, e incluyendo todos los residuos críticos informados previamente por Istvan (2001b). Dado que el tamaño de los dodecapéptidos era demasiado largo para encajar en el sitio activo de la enzima, para facilitar el proceso de docking estos autores dividieron cada secuencia en dos hexapéptidos. Los resultados obtenidos mostraron que el carácter del entorno del sitio de unión es principalmente hidrófilico y que las interacciones hidrofóbicas entre el péptido y HMGR pueden ser insignificantes. Basándose en estos resultados diseñaron cuatro hexapéptidos y seis tetrapéptidos los cuales sintetizaron (TABLA II.7). La evaluación de estos péptidos reveló grandes discrepancias entre las puntuaciones scoring de los análisis de docking y las capacidades de inhibición de la HMGR determinada experimentalmente.

Los inhibidores peptídicos descritos hasta el momento actúan como inhibidores competitivos al HMG-CoA, al igual que las estatinas.

Tabla II.7 Porcentajes de inhibición (% INH) de HMGR de péptidos. Los residuos se representan en colores de acuerdo a sus propiedades fisicoquímicas: residuos hidrofóbicos (rojo); cargados positivamente (celeste); cargados negativamente (verde).

Secuencia	% IHN	concentración de proteínas/péptido	Autores
TPMASD	~60	100 µM	Lin y col., 2015
HFKW	~60	100 µM	
PMAS	~70	100 µM	
GGV ^{bc}	39	17 µM	Soares y col., 2015
IVG ^{bc}	37	14 µM	
LVG ^{abc}	19	14 µM	
VGVI ^{ab}	11	10 µM	
VGVL ^{cd}	50	10 µM	
Hidrolizado de amaranto ^{cd}	55	4 µg/ml	

2.4.5 Identificación de péptidos inhibidores de HMGR en secuencias de amaranto

El aislado proteico de amaranto está constituido principalmente por proteínas correspondientes a la fracción globulina, particularmente globulina 11S y en menor proporción 7S. Con el objetivo de identificar por métodos secuenciales y estructurales secuencias aminoácidicas encriptadas en las proteínas de amaranto que coincidan con las correspondientes a los inhibidores antes descritos (**TABLA II.6**) se realizó, en principio, una búsqueda de las secuencias de las proteínas de amaranto.

Hasta el presente se han identificado 825 secuencias perteneciente a la división Angiospermae, orden Centrospermae, clase Dicotyledonae, familia Amaranthaceae, distribuidas en 49 especies del género *Amaranthus*, (<http://www.ncbi.nlm.nih.gov/protein>, junio 2016).

Estas secuencias pueden ser agrupadas en 177 clusters de 90% de identidad (<http://weizhonglab.ucsd.edu/cd-hit/>). De las 49 especies correspondientes al género *Amaranthus*, de acuerdo a la información existente hasta el presente, solo 4 son comestibles (*Amaranthus caudatus*, *cruentus*, *hypochondriacus* y *mantegazzianus*).

Particularmente en la especie *Amaranthus hypochondriacus* encontramos 148 secuencias agrupables en 105 clusters con 90% de identidad de las cuales 69 fueron identificadas en cloroplastos correspondientes a semillas germinadas, estadio posterior al de nuestro interés. Por esta razón nos quedamos con 36 secuencias representativas de cada grupo (**TABLA II.8**).

Del contenido de esta Tabla nos resultan de particular interés para realizar las búsquedas secuenciales la globulina 11S (CAA57633.1) y la lectina o aglutinina (AAD33922.1) En la **FIGURA II.13** se presentan las secuencias correspondientes a estas dos proteínas en amaranto.

Dado que en la búsqueda secuencial realizada no se encontró ocurrencia de las secuencias de inhibidores de HMGR conocidos (**TABLA II.7**), encriptadas en las proteínas seleccionadas (globulina 11S y lectina de amaranto), nos planteamos diseñar otra estrategia para identificar secuencias peptídicas potencialmente activas.

Tabla II.8 Proteínas secuenciadas en la especie *Amaranthus hypocondriacus*.

ACCESSION	Descripción
IQFD	Cadena A, Estructure en solución de inhibidor de Alfa-Amilasa (Aai)
AAB58165.1	Betaina aldehído deshidrogenasa
AAD33922.1	Aglutinina
AAD34639.1	Precursor proteico antimicrobiano
AAD37438.1	Precursor de la subunidad pequeña de la ribulosa 1,5 bisfosfato carboxilasa
AAM43922.1	Poliamina oxidasa
AAV37459.1	Prosistemina
ABG89856.1	Cistatina
ABS72164.1	Acetolactato sintasa
ABS72165.1	Protoporfirinógeno oxidasa mitocondrial
ADH59419.1	Cobre/zinc superóxido dismutase citosólica
ADM35971.1	RING zinc finger protein
ADO15315.1	C4 Fotosintética fosfenolpiruvato
ADO15316.1	Precursor de la subunidad alfa de la enzima málica NAD dependiente
ADZ48644.1	4,5-DOPA extradiol dioxigenasa
AEH84413.1	Piruvato ortofosfato di quiinasa
AFS17278.1	Sucrosa sintasa II
AFS17281.1	Invertasa neutra/alcalina, parcial
AFS17283.1	ADP glucosa pirofosforilasa, parcial
AHI59253.1	Proteína de tipo 4 similar a la superfamilia transmembrana 9 putativa
AHI59309.1	Box DEAE de la RNA helicasa ATP-dependiente, parcial
AHI59366.1	Substilisina putativa, parcial
AHI59422.1	Glutaredoxina putativa-como proteína, parcial
AHV78224.1	4,5-DOPA dioxigenasa proteína similar al extradiol
AHV78225.1	Glucosil ltransfera I
AHV78226.1	cyclo-DOPA 5-O-glucosiltransferasI
AHV78227.1	CYP76, parcial
AKO71445.1	Ah24
BAJ09328.1	Amidón Sintasa
BAR73034.1	Enzima desramificante de almidón
CAA52749.1	Proteína de unión a la clorofila a/b, parcial
CAA52750.1	Proteína de unión a la clorofila a/b
CAA57633.1	Globulina 11S. Proteína de almacenamiento de semillas
CAB61327.1	Inhibidor de tripsina, parcial
P83167.1	Proteína no-específica de transferencia de lípidos 1

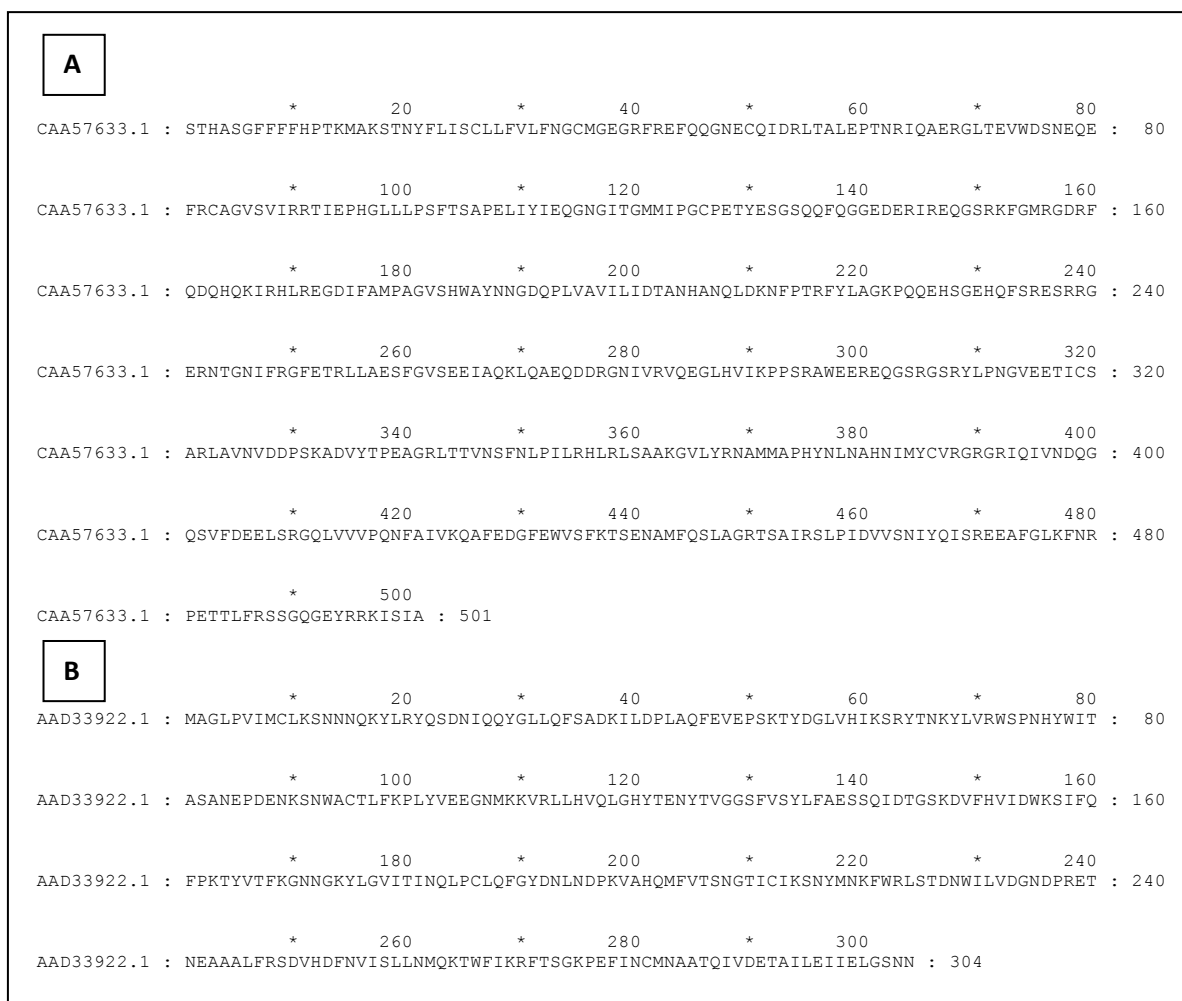


Figura II.13: Secuencias de las proteínas globulina 11S (CAA57633.1) **(A)** y aglutinina (AAD33922.1) **(B)** de *Amaranthus hypochondriacus*

Teniendo en cuenta lo discutido previamente se buscará completar el presente trabajo, en ensayos posteriores, mediante una aproximación integral que tenga en consideración las características de los inhibidores peptídicos ya identificados.

2.4 CONCLUSIONES CAPÍTULO 2

La actividad hipocolesterolémica de péptidos bioactivos puede ser debida, entre otras alternativas, a la inhibición de la HMGR, enzima reguladora de la vía del mevalonato. La evaluación de la actividad de esta enzima, de acuerdo a la información contenida en fuentes bibliográficas, puede llevarse a cabo de diferentes formas: (i) midiendo la actividad de la HMGR directamente en preparados de hígado de ratón; (ii) empleando un homogenado de *Escherichia coli* transformada con el plásmido codificante del dominio catalítico de la HMGR de hamster (Frimpong y col., 1993; Pak y col., 2005) y/o humano (Lin y col., 2015); (iii) utilizando el dominio catalítico purificado correspondiente a la enzima humana (Sigma–Aldrich) (Soares y col., 2015).

En el presente trabajo de tesis, se optó por la última opción haciendo uso de un kit comercial provisto por Sigma–Aldrich. Esta metodología presentó una serie de inconvenientes los cuales tuvieron que ser superados, particularmente aquellos asociados a la existencia de proteínas no declaradas en el buffer a utilizar en la reacción. Una vez logrado esto, los resultados obtenidos mostraron que las muestras de Ais e Hid eran capaces de inhibir la HMGR, al igual que la muestra de Dig, aunque en una menor proporción. Estos resultados llevan a pensar que en el Ais se encuentran especies proteicas o polipéptidos que logran reducir la actividad enzimática, actividad que no es incrementada por efecto de la hidrólisis exógena ni del proceso de digestión gastrointestinal simulada aunque el tamaño molecular de especies presentes se reduce drásticamente. Por lo tanto, las reacciones de hidrólisis deben estar generando péptidos con actividad inhibitoria de la enzima, para compensar la pérdida de actividad que podría estar asociada a proteínas y/o polipéptidos de mayor tamaño. El hecho que la actividad inhibitoria de la enzima no sea función de la concentración de

muestra, falta de un comportamiento dosis–respuesta, nos induce a pensar en la existencia simultánea en las muestras de activadores e inhibidores. Los primeros estarían en baja cantidad en la muestra de Ais y en mayor contenido en Dig, siendo a su vez su afinidad a la enzima superior a la de los inhibidores y su sitio de unión diferente. Estas hipótesis deben ser comprobadas en un futuro próximo.

Es necesario también tener en consideración que el ensayo utilizado en este trabajo hace uso solo del dominio catalítico de la enzima y que este pudo haber sufrido cambios conformacionales cuando se reemplazó el buffer del Kit por otro carente de proteína cuya función más probable es estabilizar la enzima, por ende, la respuesta de esta pudo ser diferentes frente a potenciales inhibidores o activadores de su función o presencia de diferente cantidad de proteína. Esto también deberá ser comprobado mediante el empleo de otras metodologías de la actividad de la HMGR.

Cabe por último señalar que los resultados de inhibición obtenidos puedan ser fruto de una acción inespecífica, sin embargo, llama la atención que sean del mismo orden de los obtenidos por Soares y col. (2015) con la misma matriz proteica. Por otra parte, resulta llamativo que estos autores no hayan informado una relación entre la dosis de los péptidos identificados y la respuesta de la inhibición.

Los péptidos de amaranto, como indicó Soares y col. (2015) podrían eventualmente constituir una clase de inhibidores de la HMGR diferentes del de las estatinas.

El uso de herramientas bioinformáticas tampoco arrojó mayor claridad a la problemática que enfrentamos dado que la búsqueda secuencial realizada mostró la inexistencia de secuencia de péptidos identificados como inhibidores de la HMGR en la estructura primaria de globulina 11S y lectina, únicas secuencias conocidas entre las proteínas de almacenamiento de amaranto. Nuevamente aquí la falta de información es un factor limitante para la búsqueda debiéndose emplear otra estrategia para lograr identificar secuencias peptídicas potencialmente activas.

Finalmente cabe destacar que la HMGR es una enzima regulada a diferentes niveles: de transcripción, traducción, modificaciones postraduccionales y degradación y además es estructuralmente compleja. Debemos recordar que la misma está constituida por una sola cadena polipeptídica con tres dominios: uno transmembrana en el extremo N–terminal, uno catalítico en el extremo

C-termina y un dominio intermedio de unión. La porción catalítica en humanos de la HMGR humana forma tetrámeros constituidos también por tres dominios, L, S y N donde se ubican el sitio de unión al sustrato, el sitio de unión para NADPH y la porción de unión de la porción catalítica al dominio transmembrana, respectivamente. Además, se encuentra presente un bucle que contiene un *cis*-péptido esencial en la formación del sitio activo.

Resulta evidente que los potenciales péptidos inhibidores de la acción de la HMGR podrían en principio actuar sobre distintos sitios de la molécula, ya sea el sitio activo o el sitio de unión de NADPH, etc. En el caso de mezclas como las muestras de Ais, Hid y Dig diferentes péptidos podrían actuar a través de diferentes mecanismos. Hasta el presente, como se indicó previamente, solo se han descrito inhibidores de tipo competitivo con el sustrato.

CAPÍTULO 3

**APLICACIÓN DE TÉCNICAS
BIONFORMÁTICAS AL ESTUDIO DE
OTRAS ACTIVIDADES BIOLÓGICAS DE LA
FRACCIÓN PROTEÍCA DE AMARANTO**

3.1 INTRODUCCIÓN

Las proteínas de almacenamiento de semillas de amaranto, tal como se indicara previamente (INTRODUCCIÓN, Sección III.3), contienen secuencias peptídicas encriptadas que exhiben diferentes actividades biológicas. Nuestro grupo de trabajo se encuentra estudiando las siguientes actividades: antihipertensiva, inmunomoduladora, hipocolesterolémica, antitrombótica y antioxidante. Para ello se emplean técnicas *in vitro* e *in vivo* haciendo uso de diferentes sistemas animal modelo. En el caso del estudio de la actividad antihipertensiva, hemos además realizado estudios *in silico* los cuales nos han permitido identificar dos tetrapéptidos, ALEP y VIKP, encriptados en la globulina 11S de amaranto que actúan como inhibidores competitivos de la enzima Convertidora de Angiotensina, ACE, tal como demostráramos haciendo uso de docking ligando-proteína (Vecchi y Añón, 2009). Este fue el primer trabajo en el que se utilizó el docking de péptidos para encontrar inhibidores de ACE a partir de una fuente proteica alimentaria. Como hemos visto previamente (Introducción General, Sección I.4) la aproximación *in silico* es una herramienta muy útil y poderosa para complementar, ayudar en el diseño de estrategias y/o corroborar los ensayos *in vitro* e *in vivo*. Por esta razón se decidió en principio consolidar este tipo de estudios en el seno del grupo de trabajo: formar recursos humanos especializados en el marco de este trabajo de Tesis Doctoral y por otro lado complementar estudios en curso haciendo uso de estas herramientas. En particular se utilizó herramientas bioinformáticas en el análisis de las actividades antitrombótica y antioxidante; el estudio experimental correspondiente fue llevado a cabo por las Lic. Ana Clara Sabbione y Lic. María Cecilia Orsini Delgado en sus respectivas Tesis Doctorales.

A continuación, se hará una breve introducción de las dos actividades biológicas seleccionadas y con posterioridad la descripción de los estudios llevados a cabo haciendo uso de herramientas bioinformáticas.

3.1.1 Actividad Antitrombótica

La trombosis es la obstrucción local del flujo de sangre en algún vaso sanguíneo arterial o venoso debido a la formación de un trombo, coágulo que se forma en el lugar y tiempo equivocado. Actualmente la trombosis se considera una enfermedad multicausal resultante del efecto de factores genéticos y adquiridos que interactúan entre sí (Lensing y col., 1999).

Durante el funcionamiento adecuado de los mecanismos de hemostasis la formación de un trombo/coágulo es desencadenada por una lesión en la pared de un vaso sanguíneo. En una primera etapa se produce la adhesión plaquetaria en la zona lesionada formándose un agregado plaquetario. En una segunda etapa se forma una red de fibrina que atrapa los elementos sanguíneos en su interior para dar lugar a la formación del trombo (FIGURA III.1).

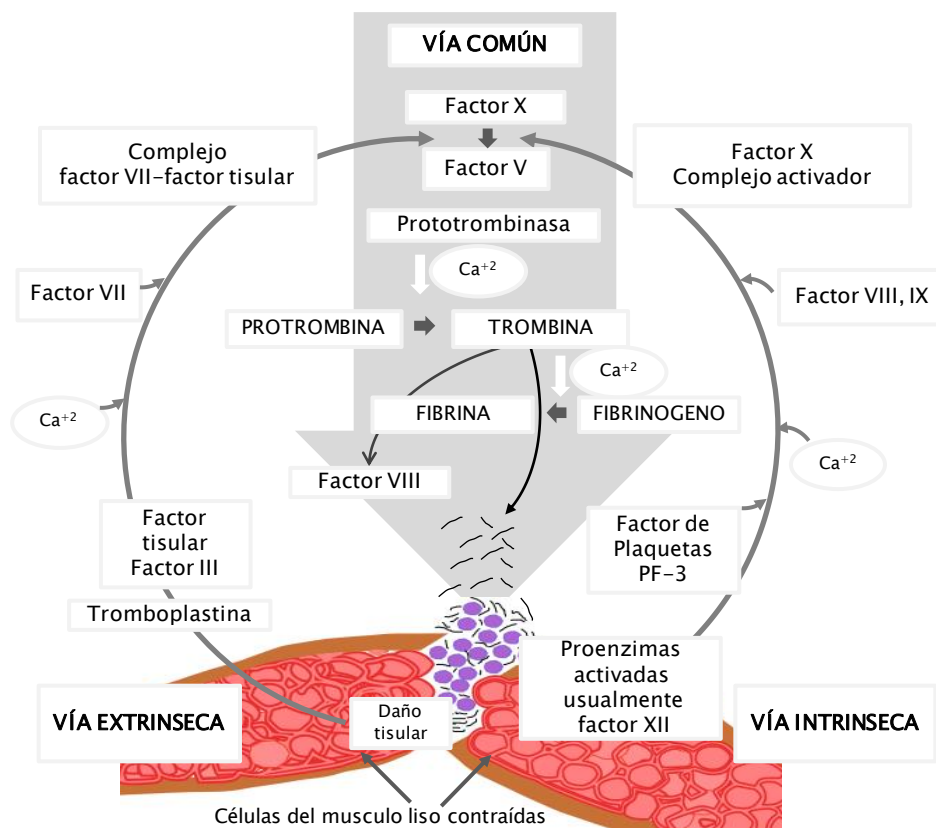


FIGURA III.1 Esquema condensado representando la cascada de la coagulación.

Una falla en la regulación de los mecanismos de coagulación origina la formación de trombos patológicos debido a la precipitación y adhesión de los elementos celulares que componen la sangre. Dependiendo del tipo de vaso sanguíneo ocluido y la duración de este fenómeno se pueden producir diferentes daños (Cruz-Gonzalez y col., 2008).

Los tratamientos antitrombóticos existentes se basan en impedir la formación o crecimiento del trombo y facilitar la disolución de coágulos ya existentes. Los agentes antitrombóticos se clasifican en antiplaquetarios, anticoagulantes y fibrinolíticos, según su mecanismo de acción. Los dos primeros actúan impidiendo la formación de coágulos en la sangre; mientras que los agentes fibrinolíticos se emplean para disolver coágulos ya existentes.

La prevención de enfermedades arteriales trombóticas es prioritaria en una gran cantidad de países desarrollados. Estudios epidemiológicos han demostrado que una dieta inadecuada constituye un importante factor de riesgo para los eventos trombóticos. Por otro lado, dietas ricas en frutas y verduras han mostrado que tienen el potencial de prevenir estas enfermedades (Bazzano y col., 2002; Gillman y col., 1995). En los últimos años se han informado efectos antitrombóticos de extractos alcohólicos de distintas variedades de tomate, ajos, cebollas y proteínas de soja (Corzo-Martínez y col., 2007; Gibbs y col., 2004; Yamada y col., 2004)

3.1.2 Actividad Antioxidante

En los organismos aerobios el oxígeno representa la fuerza motriz para el mantenimiento de su metabolismo y su viabilidad celular, sin embargo, representa en sí un peligro potencial debido a su capacidad de formación de intermediarios parcialmente reducidos y dotados de una alta reactividad, conocidas como especies de oxígeno reactivas (ROS). Estas especies pueden ser radicalarias, es decir especies moleculares activadas con un electrón en un nivel energético mayor y, en cuyo caso, poseen una reactividad alta e indiscriminada. En sistemas biológicos las ROS pueden tener origen tanto endógeno como exógeno. Por ejemplo, además de ser intermediarios en la cadena respiratoria, el organismo también utiliza a los radicales libres como mecanismo de defensa para la destrucción de bacterias y patógenos invasores. El estrés oxidativo es causado por un desbalance entre la producción de ROS y la capacidad de detoxificación del organismo.

Frente a esta situación el organismo humano responde con sistemas de defensa antioxidante, pero en determinadas ocasiones puede ser insuficiente, desencadenando diferentes procesos fisiológicos y fisiopatológicos. El estrés oxidativo se asocia al proceso de envejecimiento celular y diversas enfermedades como enfermedades cardiovasculares, cáncer y enfermedades degenerativas del sistema nervioso como Párkinson y de Alzheimer (Halliwell, 2001).

Los mecanismos de defensa contra los radicales libres pueden ser enzimáticos o no enzimáticos. En el primer grupo se encuentran las enzimas como la superóxido dismutasa, la catalasa y la glutatión peroxidada entre otras. Todas ellas son dependientes de oligoelementos (selenio, cobre, manganeso y zinc). Como sistemas no enzimáticos las células cuentan con sustancias “scavenger”, que son sustancias capaces de captar radicales libres. En este grupo se incluyen los carotenoides, las vitaminas C y E, entre otros. Los mecanismos se complementan ya que actúan sobre distintos oxidantes o en diferentes compartimentos celulares.

Las ROS pueden producir diversos efectos sobre moléculas biológicas pudiendo causar daño celular. Por ejemplo las ROS pueden actuar oxidando lípidos de las membranas biológicas, sobre proteínas provocando su desnaturalización y consecuente pérdida de actividad en el caso de enzimas; y sobre los ácidos nucleicos modificando bases y produciendo mutagénesis y carcinogénesis (Pihlanto, 2006)

Los antioxidantes son un grupo de sustancias estructuralmente complejas que se agrupan de acuerdo a sus mecanismos de acción. Existen evidencias de que el consumo de antioxidantes en la dieta puede tener efectos positivos adicionales en la disminución del estrés oxidativo que se suman a las defensas propias del organismo (Fang y col., 2002). Entre los antioxidantes más conocidos y estudiados de la dieta se encuentran las vitaminas, como la vitamina C y E, los carotenoides, los polifenoles y los flavonoides. Los ácidos fenólicos y flavonoides son antioxidantes muy eficientes debido a la presencia en su estructura de anillos aromáticos con grupos hidroxilos que son capaces de estabilizar el electrón desapareado de los radicales libres. Además, los flavonoides pueden prevenir la auto-oxidación de lípidos insaturados debido a su capacidad de formar complejos con el ión cúprico, que actúa como catalizador en este proceso.

Recientemente han sido aislados péptidos con capacidad antioxidante proveniente de proteínas alimentarias de diversas fuentes como las caseínas de la leche, proteínas de trigo, huevo y soja, entre otras (Chen y col., 1998; Gibbs y col., 2004; Hernández-Ledesma y col., 2005; Peña-Ramos y col., 2004). La actividad se ha asociado con la presencia de ciertos aminoácidos en la secuencia. La histidina parece ser el aminoácido más relevante en esta actividad debido a la presencia del grupo imidazol en su estructura, y la hidrofobicidad de los péptidos también parece ser un factor importante en esta actividad ya que incrementa su accesibilidad hacia sitios hidrofóbicos donde, por ejemplo, ocurre la peroxidación lipídica (Chen y col. 1998). También se ha demostrado que varios aminoácidos por si mismos pueden actuar como antioxidantes, entre ellos tirosina, triptófano, metionina y lisina (Chen y col., 1996).

3.2 APLICACIÓN DE TÉCNICAS BIOINFORMÁTICAS AL ESTUDIO DE LA ACTIVIDAD ANTITROMBÓTICA

3.2.1 Obtención de la fracción activa

Esta fracción fue obtenida por la Dra. Sabbione durante el desarrollo de su Tesis Doctoral. Para ello se preparó un aislado proteico a partir de harina desgrasada obtenida de semillas de *Amaranthus hypochondriacus* (UNRC, Córdoba, Argentina) mediante extracción a pH 9 y precipitación isoeléctrica a pH 5 como fue descrito anteriormente. El aislado se sometió a un proceso de digestión gastrointestinal simulada; inicialmente fue tratado con pepsina (0,1g pepsina/g proteína en 0,1N HCl y 0,03M NaCl, pH 2) a 37°C durante 1 h, y luego con pancreatina (0,1g pancreatina/g proteína en 0,1M NaHCO₃, pH 6) en iguales condiciones. Las muestras se caracterizaron mediante la determinación de la concentración proteica (Kjeldahl), del grado de hidrólisis (OPA) y de la composición polipeptídica por electroforesis Tricina-SDS-PAGE con y sin agente reductor. Empleando el método de las microplacas se determinó la actividad antitrombótica *in vitro* del aislado y del digerido. Este método descrito por Yang y col. (2007), modificado por Zhang y col. (2008), se basa en la detección de la formación del coágulo, una vez que la trombina convierte el fibrinógeno en fibrina, mediante la medida de la absorbancia a 405 nm (Sabbione, 2015). Los resultados obtenidos muestran que en las proteínas de amaranto se encuentran péptidos con actividad antitrombótica que se liberan una vez que la proteína es

sometida a la digestión gastrointestinal simulada (Digerido: $68,4 \pm 0,8$ g proteína/g hidrolizado y GH% $51,1 \pm 3,8$). Los resultados alcanzados por la Dra. Sabbione muestran que el aislado proteico ($82,4 \pm 1,6$ g proteína/g aislado) no fue capaz de inhibir la coagulación sanguínea a las concentraciones estudiadas, mientras que el digerido presentó una curva de inhibición del tipo dosis-respuesta que permitió calcular un IC_{50} de $0,23 \pm 0,02$ mg/ml. Mediante FPLC de exclusión molecular se logró separar, a partir del digerido, una fracción activa (FA) con comportamiento dosis-dependiente cuyo IC_{50} fue $0,07 \pm 0,01$ mg/ml. En experimentos posteriores la Dra. Sabbione estudió, haciendo uso de un ensayo *in vitro*, la capacidad de absorción de los péptidos a través del epitelio intestinal. Para ello se utilizaron monocapas de células Caco 2-TC7 en filtros semipermeables llamados insertos (FIGURA III.2). La fracción peptídica activa fue incubada con células de epitelio intestinal y posteriormente se analizaron mediante HPLC en fase reversa distintas alícuotas correspondientes a las cámaras apical y basolateral. Esta metodología permite conocer los péptidos presentes en dichos compartimentos y con posterioridad las secuencias de los péptidos que atravesaron la monocapa celular (Sabbione, 2015).

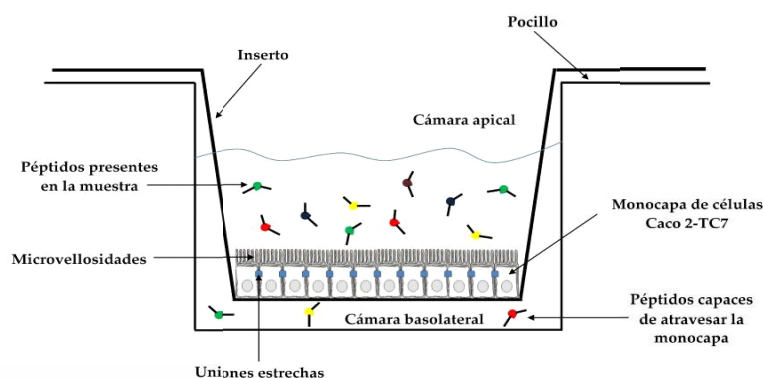


FIGURA III.2 Diagrama del montaje utilizado para las mediciones de la actividad transportadora de la monocapa de células Caco 2-TC7 (Sabbione, 2015).

3.2.2 Identificación de péptidos presentes en la fracción activa y en la cámara basolateral

Las muestras FA y CB fueron enviadas al INRA (*Unité BIA-Plateforme Biopolymères Biologie Structurale*, Nantes, Francia) para ser analizadas por espectrometría de masas MALDI-TOF, con el objetivo de estudiar la complejidad de las mismas y determinar los pesos moleculares de las especies presentes en cada una de ellas (FIGURA III.3).

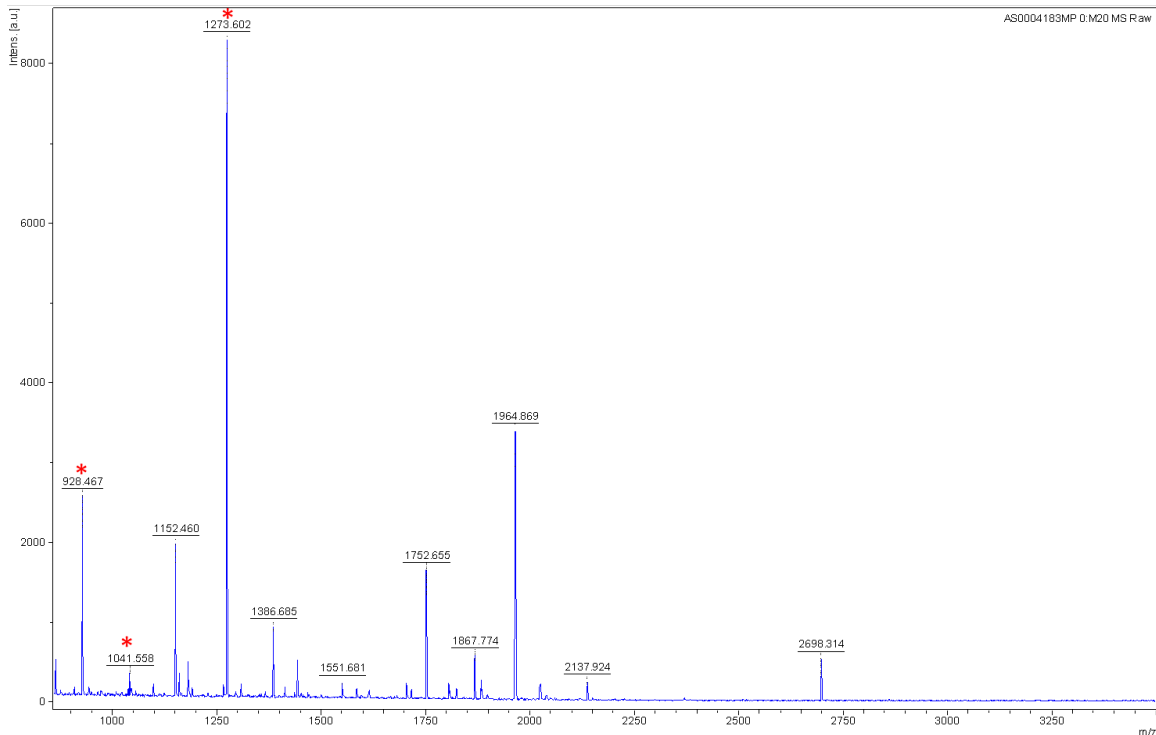


FIGURA III.3 Espectro de masas MALDI-TOF de la fracción activa.

En esta figura se puede apreciar que una gran cantidad de masas presentaron pesos moleculares inferiores a los 4000 Da, razón por la cual se procedió a analizar los péptidos presentes por nano-cromatografía líquida acoplada a detectores de masa en tándem (LC-MS/MS) y así poder obtener los datos de las secuencias. Al comparar los resultados del MALDI-TOF con los del LC-MS/MS se encontró un gran número de picos que no pudieron ser identificados. De hecho, de todas las especies halladas, solamente tres, correspondientes a los picos marcados con el asterisco rojo en la FIGURA III.3, pudieron identificarse. Otras masas significativas (ejemplo 1964,86 o 1752,65) fueron detectadas por LC-MS/MS pero no lograron identificarse haciendo uso de la base de datos utilizada en el análisis de datos. La imposibilidad de identificación de estos picos se debe a que pocas proteínas de amaranto se encuentran secuenciadas. Cabe remarcar que algunas proteínas de importancia dentro de la taxonomía *Amaranthus*, como por ejemplo la globulina 7S, una de las proteínas de reserva mayoritarias en las semillas, aún no ha sido secuenciada; es de esperarse que péptidos presentes en la FA y por lo tanto presentes en la cámara basolateral del inserto provengan de esta proteína.

Teniendo en cuenta estos datos se realizó una secuenciación *de novo* de los picos correspondientes a las masas moleculares que no se identificaron. Sin

embargo, debido a la complejidad de la muestra este análisis no permitió alcanzar resultados confiables.

La FA resultó compleja en cuanto a su composición polipeptídica. Se identificaron en ella 57 secuencias de 10 proteínas del género *Amaranthus*. Dado que la fracción activa en estudio, FA, proviene de un aislado proteico de amaranto y siendo que el aislado proteico se compone principalmente de las proteínas de reserva, se decidió seleccionar para los posteriores análisis las secuencias encontradas dentro de estas proteínas. La mayor cantidad de péptidos detectados en la fracción activa pertenecen a la globulina 11S (CAA57633.1) y a la aglutinina (AAD33922.1) (Tabla III.1).

TABLA III.1 Número de secuencias y familias en las cuales se agruparon las secuencias identificadas por LC-MS/MS acoplado a MALDI-TOF en la fracción activa (FA) y el contenido de la cámara basolateral (CB). Se utilizó el programa MASCOT 2.2 contra la base de datos Uniprot restringido al género *Amaranthus*.

<i>Proteína a la que pertenece</i>	<i>FA</i>		<i>CB</i>	
	Secuencias identificadas	familias	secuencias identificadas	familias
Globulina 11S (<i>A. hypochondriacus</i>)	32	9	2	2
Aglutinina (<i>A. caudatus</i>)	11	6	2	2
Alérgeno del polen (<i>A. retroflexus</i>)	8	5	–	–
Cistaina (<i>A. hypochondriacus</i>)	2	2	–	–
Enzima málica dependiente de NAD (<i>A. hypochondriacus</i>)	2	2	1	1
Betaína (aldehído deshidrogenasa) (<i>A. hypochondriacus</i>)	1	1	–	–
NADH deshidrogenasa (<i>A. blitoides</i>)	1	1	–	–
ATP sintetasa	–	–	2	2
Inhibidor de tripsina	–	–	1	1
Almidón sintasa	–	–	1	1

Los péptidos pertenecientes a una misma proteína, fueron agrupados en familias, que corresponden a secuencias pertenecientes a una misma región de la secuencia de la proteína madre. De esta forma se identificaron 9 familias de péptidos provenientes de la globulina 11S y 6 familias correspondientes a la lectina (aglutinina) de amaranto para la FA y, 2 y 2 para CB respectivamente. Las secuencias identificadas corresponden al 24 % de la secuencia de la 11S y al 21 % de la secuencia de la aglutinina (FIGURAS III.4 Y III.5).

```

      *      20      *      40      *      60      *      80
: MEGRFREFQGNQCIDRLTALEPTNR IQAERGLTEVWDSNEQE FRCAGVSVIRRTIEPHGLLLPSFTSAPEL IYIEQGN :
      *      100     *      120     *      140     *      160
: GITGMMIPGCPETYESGSGQQFQGGEDERIREQGSRKFGMRGDRFDQHQKIRHLREGDIFAMPAGVSHWAYNNGDQPLVA :
      *      180     *      200     *      220     *      240
: VIL IDTANHANQLDR NFPTRFY LAGKPOQEHSGEHQFS RESRRGERNTGNI FRGFETRLLAESFGVSEEIAQK LQAEQDD :
      *      260     *      280     *      300     *      320
: RGNIVRVQEGLHVIKPPSRAWEEREQGSRGSRYLP NGVEETICSAR LAVNVDDPSKADVYTP EAGRLTTVNSFNLPILRH :
      *      340     *      360     *      380     *      400
: LRLSAAKGVLYRNAMMAPHYNLNAHNIMYCVRGRGR IQIVNDQGSVFDEELS RGQLVVVPQNFAIVK QAFEDGF EWVSF :
      *      420     *      440     *      460
: KTSENAMFQSLAGRTSAIRSLPIDVVSNIYQISREEAFGLKFNRPETTLFRSSGQGEYRRKISIA :

```

Figura III.4. Secuencia de la globulina 11S de *Amaranthus hypochondriacus* (CAA57633.1). En azul se encuentran marcadas las familias de péptidos identificadas en la fracción activa; subrayadas en negro se encuentran marcadas las familias de péptidos identificadas en la cámara basolateral. En negro se marca el sitio de corte que da origen a los polipéptidos ácido y básico.

```

      *      20      *      40      *      60      *      80
: XAGLPVIMCLKSNHDKQYLRYQSDNIQQYGLLQ FSADKILDPL LAQFEVEPSKTYDGLVHIKSRYTNKYLVWRWSPNH WIT :
      *      100     *      120     *      140     *      160
: ASANEPDENKSNWACTL FKPLYVEEGNMKKVRLLLHVQLGHYQNYTVGGSFVSY LFAESSQIDTGSKDVHFVIDWKSIFQ :
      *      180     *      200     *      220     *      240
: FPKGYVTFKGNNGKYLGVITINQLPCLQFGY DNLNDPKVAHQMFVTSNGTICIKSNYMNKFWRLSTDDW ILVDGNDPRET :
      *      260     *      280     *      300
: NEAAALFRSDVHDFNVI SLLNMQKTWFIKRFTSGKPGFINCMNAATQNVDETA ILEIIELGQNN : 304

```

Figura III.5 Secuencia de la lectina de *Amaranthus caudatus* (AAD33922.1). En gris se encuentran marcadas las familias de péptidos identificadas en la fracción activa. Subrayadas en negro se encuentran marcadas las familias de péptidos identificadas en la cámara basolateral.

Con estos resultados se trabajó entonces en la identificación de las secuencias candidatas a ser las responsables de la bioactividad para su posterior síntesis. Se analizó la composición aminoacídica de los péptidos identificados comparándola con péptidos antitrombóticos identificados y disponibles en bases de datos, y secuencias de los sustratos fisiológicos de la enzima trombina. Las secuencias identificadas fueron entonces comparadas con secuencias de péptidos con actividad antitrombótica derivados de proteínas alimentarias tomando como referencia la base de datos BIOPEP (Minkiewicz y col., 2010), en la que, en mayo de 2014, se encontraron registrados 56 péptidos antitrombóticos. Ninguno de ellos presentó coincidencias con las secuencias encontradas en la fracción activa. Por esta razón se decidió emplear otros criterios que nos permitieran seleccionar aquellas secuencias que se sintetizarán para realizar posteriores estudios.

Teniendo en cuenta que el ensayo empleado para la determinación de la actividad antitrombótica se basa en la inhibición de la enzima trombina se optó

por consensuar criterios para la selección de péptidos en base a las interacciones entre péptidos y la enzima trombina.

La trombina (EC 3.4.21.5) es una glicoproteína de 32 kDa, formada por dos cadenas polipeptídicas, A y B, de 259 y 36 aminoácidos respectivamente que se encuentran unidas por un puente disulfuro. Participa en el final de la cascada de la coagulación, conocida también como vía común (Sección 3.1.1, Pág. 122). Esta proteasa escinde las moléculas de fibrinógeno en fibrinopéptidos, que espontáneamente comienzan a agruparse formando asociaciones altamente ordenadas que dan inicio a la formación del coágulo (Fuss y col., 2001). Esta enzima pertenece a la familia de las serinproteasas, como la tripsina y quimiotripsina. Inicia el ataque nucleofílico en el enlace peptídico mediante un residuo de serina que posee en su centro activo e interacciona preferentemente con sustratos que poseen aminoácidos cargados positivamente como Arg/Lys en la posición P1 del sustrato de acuerdo a la nomenclatura de Schechter y Berger (1967) (**FIGURA III.6**).

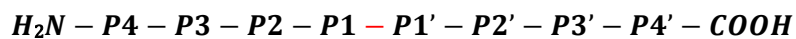


FIGURA III.6 Nomenclatura de Schechter y Berger para residuos aminoacídicos de acuerdo a su localización con respecto al enlace escindido por proteasas. El enlace escindido se muestra en rojo.

Sin embargo, al igual que otras proteasas la preferencia de sustrato se extiende más allá de la posición P1 y está dictada por interacciones con otros dominios estructurales que no están en contacto con el bolsillo de especificidad primaria. Las inspecciones de las secuencias de consenso reconocidas por la trombina revelan que interacciona preferentemente con sustratos que contienen aminoácidos cargados positivamente, como arginina, en la posición P1, prolina en P2 y un residuo hidrofóbico, el cual suele ser aromático, en la posición P4 (Bajusz y col., 1990; Huntington, 2005; Hirsh y col., 2005; Mehta y col., 2013). La trombina presenta dos exositos electropositivos físicamente alejados del sitio activo y localizados casi en lugares opuestos en la superficie de la trombina. Estos exositos desempeñan un papel crucial en el reconocimiento de sustratos, efectores e inhibidores específicos que participan en el complejo mecanismo de regulación de la cascada de coagulación y la actividad de la trombina; sus propiedades se han caracterizado por extensos estudios cristalográficos, mutagénicos, biofísicos e enzimológicos (**Figura III.7**) (Marino y col., 2010; Bock y col., 2007).

Mientras el sitio activo tiene preferencia por aminoácidos cargados positivamente, ambos exositos tienen preferencia por sustratos cargados negativamente. El exosito I exhibe además afinidad por sustratos hidrofóbicos (Huntington, 2005).

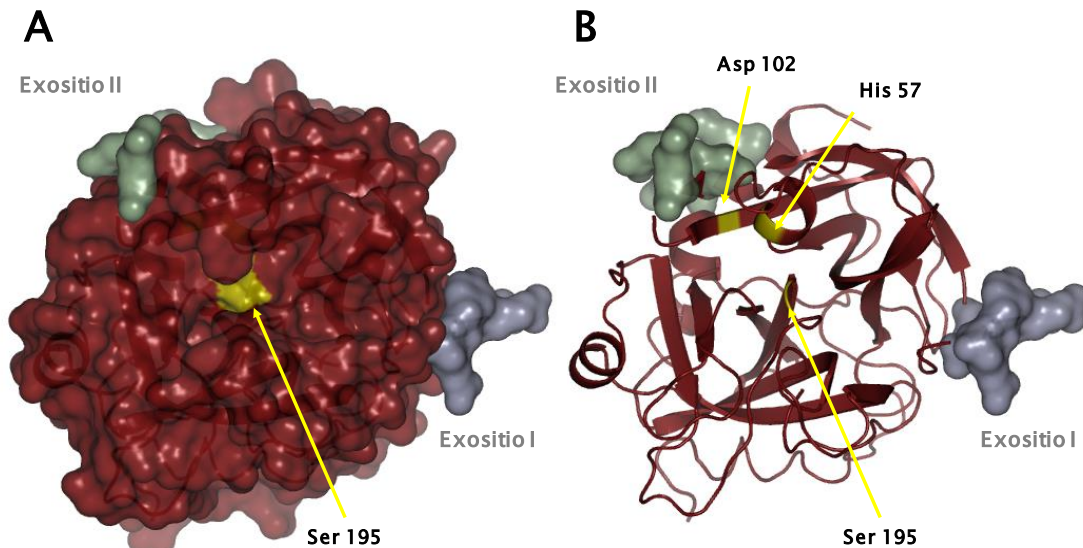


Figura III.7 Estructura cristalina de la trombina (PDB 1PPB) representada con y sin superficie (A y B respectivamente). En amarillo se indican los residuos serina (Ser195), histidina (His57) y aspartato (Asp102) que forman parte de la triada catalítica del sitio activo. Destacado en lavanda se indica el exosito I y en verde el exosito II.

Para poder seleccionar algunos péptidos entre todos los encontrados en la fracción activa se establecieron criterios que nos permitieran tener en consideración la afinidad y el mecanismo de reconocimiento entre los péptidos y la trombina. Los criterios utilizados se basaron en el trabajo de Huntington (2005), para definirlos se establecieron algunos parámetros que involucran las características y cantidades relativas de los distintos tipos de aminoácidos: la *fracción neta de carga negativa (Fa)* y la *fracción neta hidrofóbica (Fh)*. La primera se calculó sustrayendo los residuos básicos de los ácidos y dividiendo por el número total de residuos y la segunda se calculó haciendo la relación entre los residuos hidrofóbicos y el número total de residuos. Una vez establecidas estas consignas se definieron los criterios que se indican a continuación:

- *Probable interacción del péptido con el sitio activo de la trombina:* para este sitio se determinó la cantidad de aminoácidos cargados positivamente y la distribución de los mismos en las secuencias encontradas.
- *Probable interacción de los péptidos con el exosito I:* cuando la fracción neta de carga negativa se divide por la fracción neta de carga hidrofóbica

y el valor resulta < 2 existe una alta probabilidad de interacción de la secuencia con el exosítio I. Los contactos hidrofóbicos proveen la mayor parte de las energías de enlace para las interacciones del péptido con el exosítio I, mientras que las interacciones electrostáticas estarían principalmente involucradas en orientar las superficies hidrofóbicas complementarias (Karshikov, y col., 1992; Myles, y col., 2001).

- *Probable interacción de los péptidos con el exosítio II:* cuando la fracción neta de carga negativa se divide por la fracción neta de carga hidrofóbica y el valor resulta > 2 , existe una alta probabilidad de interacción de la secuencia en cuestión con el exosítio II. Las interacciones electrostáticas son las que predominan cuando los péptidos establecen interacciones con este exosítio.

Basándonos en los criterios de selección anteriormente mencionados y aplicándolos a las secuencias de las familias de péptidos hallados, se seleccionaron las secuencias que se remarcan en la **TABLA III.2**.

A la lista de péptidos candidatos a ser sintetizados se sumaron los hallados en la cámara basolateral que fueron también encontrados en la fracción activa para sintetizarlos y realizar futuros estudios. La **TABLA III.2** muestra además de las secuencias mencionadas, la familia a la que pertenecen, el criterio por el cual han sido seleccionadas, indicando la probabilidad de interaccionar con el sitio activo, el exosítio I o el II o su presencia en la cámara basolateral después de los ensayos de simulación de absorción y transporte en los insertos.

TABLA III.2. Detalle del número de residuos totales (T), hidrofóbicos (H), ácidos (A), y básicos (B); la *fracción neta de carga negativa (Fa)* y la *fracción neta hidrofóbica (Fh)* de las secuencias identificadas en FA y CB de globulina 11S y aglutinina. Resaltados en gris se indican cuales fueron seleccionados para su síntesis. Residuos ácidos: D, E; básicos: R, H y K; hidrofóbicos: I, V, F, Y, W, P.

Familia	Secuencia	Número de Residuos				Fa	Fh	Fa/Fh	Criterio
		T	H	A	B				
<i>Globulina 11S</i>									
1	LLFVLFNGCM	10	2	0	0	0,00	0,20	0,00	
2	IQA EAGL TEV-----	10	2	2	0	0,20	0,20	1,00	Exosítio 1
	-----TEVWDSNE--	8	2	3	0	0,38	0,25	1,50	
	-----TEVWDSNEQ-	9	2	3	0	0,33	0,22	1,50	
	-----TEVWDSNEQE	10	2	4	0	0,40	0,20	2,00	Exosítio 2
3	--IEQGNGIT-	8	2	1	0	0,13	0,25	0,50	
	IYIEQGNGITG	11	4	1	0	0,09	0,36	0,25	

4	IDTANHANQLD-	11	1	2	1	0,09	0,09	1,00	Exositio 1
	IDTANHANQLDK	12	1	2	2	0,00	0,08	0,00	Sitio Activo
	----NHANQLDK	8	0	1	2	-0,13	0,00	-	
5	-LAGKPQQE-----	8	1	1	1	0,00	0,13	0,00	
	-LAGKPQQEH-----	9	1	1	2	-0,11	0,11	-1,00	
	YLAGKPQQEH-----	10	2	1	2	-0,10	0,20	-0,50	
	-----PQQEHSGEHQFS	12	2	2	2	0,00	0,17	0,00	
	-LAGKPQQEHSGEH---	13	1	2	3	-0,08	0,08	-1,00	
	-LAGKPQQEHSGEHQ--	14	1	2	3	-0,07	0,07	-1,00	Sitio Activo
	YLAGKPQQEHSGEH---	14	2	2	3	-0,07	0,14	-0,50	
	YLAGKPQQEHSGEHQ--	15	2	2	3	-0,07	0,13	-0,50	
-LAGKPQQEHSGEHQFS	16	3	2	3	-0,06	0,19	-0,33		
6	LQAEQDDR	8	0	3	1	0,25	0,00	-	*
7	LAVNVDDPS-----	9	3	2	0	0,22	0,33	0,67	
	-AVNVDDPSKA-----	11	3	2	1	0,09	0,27	0,33	
	-AVNVDDPS-----	8	3	2	0	0,25	0,38	0,67	
	-AVNVDDPSK-----	9	3	2	1	0,11	0,33	0,33	*
	-----DVYTPEAG	8	4	2	0	0,25	0,50	0,50	
	-----DVYTPE--	6	3	2	0	0,33	0,50	0,67	CB
8	IQIVNDQGG-----	9	3	1	0	0,11	0,33	0,33	
	IQIVNDQGGQS-----	10	3	1	0	0,10	0,30	0,33	
	IQIVNDQGGQSV-----	11	4	1	0	0,09	0,36	0,25	
	IQIVNDQGGQSVFDEELS	17	5	4	0	0,24	0,29	0,80	
	--IVNDQGGQSV-----	9	3	1	0	0,11	0,33	0,33	
	----NDQGGQSVFDEELS	13	2	4	0	0,31	0,15	2,00	Exositio 2
9	-AFEDGFE	7	2	3	0	0,43	0,29	1,50	Exositio 1
	QAFEDGFE	8	2	3	0	0,38	0,25	1,50	
Familia	Secuencia	Número de Residuos				Fa	Fh	Fa/Fh	Criterio
		T	H	B	A				
<i>Aglutinina</i>									
1	-SADKILDPL	9	1	2	1	0,11	0,11	1,00	
	FSADKILDPL	10	2	2	1	0,10	0,20	0,50	
2	WITASANEPDEN-	12	3	2	0	0,17	0,25	0,67	
	-ITASANEPDENK	12	2	2	1	0,08	0,17	0,50	
3	LFAESSQIDTGSK	13	2	1	1	0,00	0,15	0,00	
	-FAESSQIDTGS-	11	2	1	0	0,09	0,18	0,50	
	-FAESSQIDTGSK	12	2	1	1	0,00	0,17	0,00	
	--AESSQIDTGSK	11	1	1	1	0,00	0,09	0,00	
	--AESSQIDTGSK	11	1	1	1	0,00	0,09	0,00	CB
4	DNLNDPK	7	1	0	1	-0,14	0,14	-1,00	Sitio Activo
5	ILVDGNDPR	9	3	1	1	0,00	0,33	0,00	
	-LVDGNDPR	8	2	1	1	0,00	0,25	0,00	Sitio Activo
6	AILEIHELGSNN	12	3	2	0	0,17	0,25	0,67	

En la **FIGURA III.8** se presentan las estructuras de la globulina 11S y de la aglutinina. Sobre las mismas se encuentran marcados los péptidos seleccionados para ser sintetizados (**TABLA III.2**). Se puede apreciar en ambas estructuras que la mayoría de los péptidos seleccionados se encuentran ubicados en la zona superficial de las moléculas proteicas, quedando más expuestos a posibles ataques hidrolíticos. En este sentido cabe mencionar que el péptido DVYTPE hallado en la cámara basolateral de los insertos, se encuentra ubicado en la cara IE del homotrímero de la globulina 11S (**FIGURA III.8**). En la molécula nativa hexamérica esta cara se encuentra unida por interacciones no covalentes a la del otro trímero. Por esta razón en el caso de que la molécula se encuentre nativa, con su estructura cuaternaria intacta, este péptido no quedaría expuesto al medio (Tandang-Silvas y col., 2012). Sin embargo, la globulina 11S se separa en sus trímeros por efecto del pH, lo que asegura la accesibilidad de las proteasas a esa región de la molécula. La simulación *in silico* de la proteólisis de la globulina 11S y de la aglutinina, haciendo uso de la herramienta de corte de péptidos *PeptideCutter* de ExPASy (Gasteiger y col., 2005), permitió comprobar que el ataque de diversas proteasas, tales como pepsina, quimotripsina y tripsina, permitiría la liberación de los péptidos identificados.

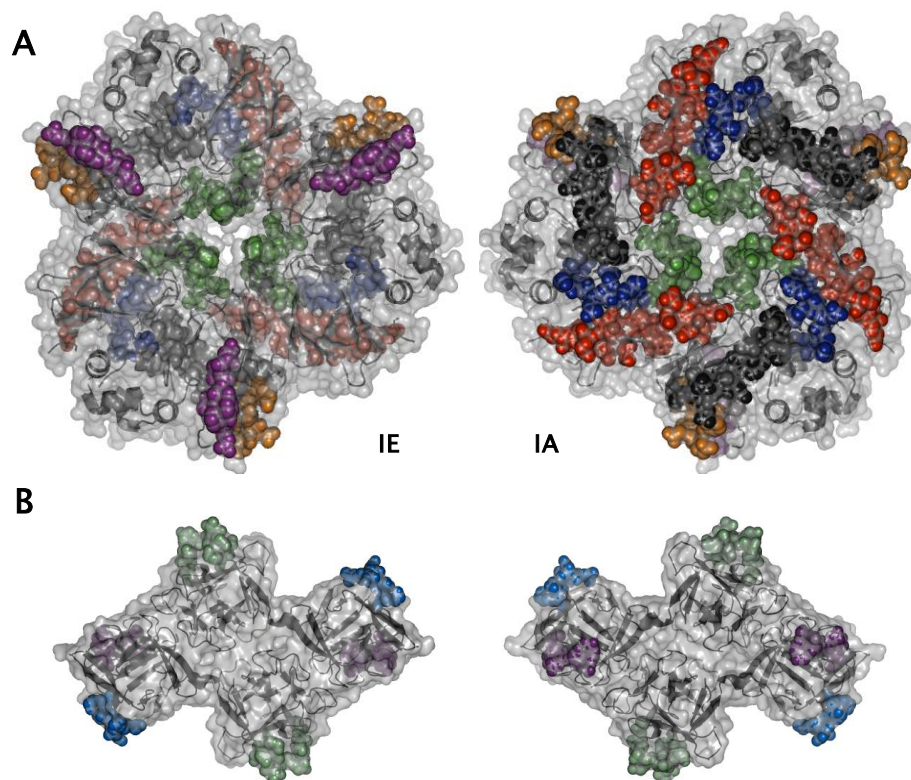


FIGURA III.8: Ubicación de los péptidos seleccionados para su síntesis marcados en la estructura de la globulina 11S (**A**) y aglutinina de amaranto (**B**). Los colores coinciden con los utilizados para presentar los péptidos en la **TABLA III.2**.

La aproximación realizada haciendo uso de herramientas bioinformáticas ha permitido la identificación de péptidos con potencial actividad antitrombótica encriptados en la secuencia de proteínas de amaranto. Cabe señalar que estos péptidos no mostraron actividad antitrombótica haciendo uso de las técnicas *in vitro* indicadas previamente, al menos en las concentraciones en que han sido evaluados. Por consiguiente, es necesario buscar otra metodología para detectar su bioactividad y además replantear, en un futuro próximo, otro tipo de estrategia bioinformática para identificar nuevos péptidos antitrombóticos.

3.3 APLICACIÓN TÉCNICAS COMPUTACIONALES AL ESTUDIO DE ACTIVIDADES ANTIOXIDANTE

3.3.1 Obtención de péptidos antioxidantes de *amaranthus mantegasianus*

El objetivo de esta parte del trabajo, realizada por la Dra. Orsini Delgado durante el desarrollo de su Tesis Doctoral, fue separar e identificar péptidos antioxidantes originados en el proceso simulado de digestión gastrointestinal. Mediante RP-HPLC preparativa se separaron dos muestras: aislado proteico de amaranto mantegazzianus (I) y su respectivo hidrolizado obtenido por digestión gastrointestinal simulada (Id). Posteriormente se determinó la actividad antioxidante de cada fracción mediante el método de ORAC (Orsini Delgado y col., 2015). Las fracciones del Id que presentaron mayor actividad (7, 16, 20, 22, 25 y 30) fueron analizadas en la *Plateforme Biopolymères Biologie Structurale* perteneciente al *Unité BIA* en el INRA (Nantes, Francia) por espectrometría de masas con el objetivo de estudiar la complejidad de las mismas y determinar los pesos moleculares de las especies presentes en cada una de ellas. En base a esta información se emplearon herramientas bioinformáticas para establecer posibles mecanismos de acción. Los estudios desarrollados se detallan a continuación.

Todas las fracciones analizadas por MALDI del aislado sometido a digestión gastrointestinal (Id) mostraron la presencia de polipéptidos inferiores a 4 kDa. Posteriormente, las muestras fueron analizadas por cromatografía líquida de nanoflujo-espectrometría de masas en tándem (LC-MS/MS). La identificación de péptidos se realizó comparando espectros de masa la base de datos UniProt

restringida a la taxonomía de *Amaranthus*, utilizando el servidor MASCOT 2.2 (Matrix Science). Las secuencias identificadas en las fracciones se detallan en la TABLA III.3.

TABLA III.3 Péptidos identificados en las fracciones de Id analizadas por LC-MS/MS. Se indica la fracción, el peso molecular (Mr experimental y calculado), la secuencia y la proteína de origen.

<i>Fracción</i>	<i>Mr experimental</i>	<i>Mr calculada</i>	<i>Secuencia</i>	<i>Proteína</i>
7	821,4282	821,4283	VVSNIYQ	Globulina 11S
	882,3593	882,3607	ESFGVSEE	Globulina 11S
	907,4392	907,4399	ISREEAFG	Globulina 11S
	978,3938	978,3931	TEVWDSNE	Globulina 11S
	990,4408	990,4407	AYNNGDQPL	Globulina 11S
	1019,4299	1019,4308	WEEREQGS	Globulina 11S
	1106,4524	1106,4516	TEVWDSNEQ	Globulina 11S
	1163,5819	1163,5822	IYIEQGNGITG	Globulina 11S
	1294,6229	1294,6227	IYIEQGNGITGM	Globulina 11S
	812,464	812,4644	DKILDPL	Aglutinina
	814,3817	814,3821	DNLNDPK	Aglutinina
	993,5351	993,5356	STHNGLRPL	Almidón Sintasa
	1017,4958	1017,4978	EEQKGSIL	Almidón Sintasa
	1228,6113	1228,6088	VQEWNPITDK	Almidón Sintasa
	830,4248	830,4246	RCNKDEI	Cistaína
	888,43	888,4301	ENREVDK	Proteína de shock termico-cloroplasto
1172,5564	1172,5561	DITPDDKQEL	Fosfoenol Piruvato-C4 Fotosintético	
916,4652	916,4688	LDVLVCNAA	Protoclorofilida reductasa	
16	875,3775	875,3773	WEEREQ	Globulina 11S
	1018,4723	1018,4753	IEQGNGITGM	Globulina 11S
	1129,489	1129,4901	GDRFQDQHQ	Globulina 11S
	1431,6485	1431,6491	AGKPQQEHSGEHQ	Globulina 11S
	1159,4986	1159,4993	ITASANEPDEN	Aglutinina
	1160,4937	1160,4945	ASANEPDENKS	Aglutinina
	1122,473	1122,473	NNPNFSGPY	Almidón Sintasa
	844,4303	844,429	NNEDLLK	Enzima Málica
	830,3911	830,3882	QQAGGENK	Proteína de shock termico-cloroplasto
	1061,4615	1061,4625	SNDREEIAE	Sacarosa Sintasa
20	855,4093	855,4086	NNGDQPLV	Globulina 11S
	893,3864	893,3879	NDQQQSVF	Globulina 11S
	973,4464	973,4465	LQAEQDDR	Globulina 11S
	1544,7336	1544,7332	LAGKPQQEHSGEHQ	Globulina 11S
	890,474	890,4749	EILYPEK	Almidón Sintasa
	1422,5799	1422,58	NNPNFSGPYGEN	Almidón Sintasa
22	802,3102	802,3102	NAMMAPH	Globulina 11S
	1246,567	1246,569	AWEEREQGSR	Globulina 11S
	1601,7475	1601,7481	GMRGDRFQDQHQK	Globulina 11S
	1665,7485	1665,7495	AGKPQQEHSGEHQFS	Globulina 11S
	1169,5835	1169,5829	YLAGKPQQEH	Globulina 11S
	833,4124	833,4131	IDTGSKDV	Aglutinina
	916,4978	916,4978	LQSDRVVT	Almidón Sintasa
	939,4279	939,4298	FTTSAERE	Actina
25	869,4608	869,4607	LAGKPQQE	Globulina 11S
	1030,4675	1030,4679	LQAEQDDR	Globulina 11S
	1386,5901	1386,5953	VSHWAYNNGDQP	Globulina 11S
	1121,5196	1121,52	AESSQIDTGSK	Aglutinina
	1046,4659	1046,4669	FKHTEDSPS	Cistaína
30	863,4133	863,4137	FNRPETT	Globulina 11S
	866,5338	866,5338	VIKPPSRA	Globulina 11S
	958,4736	958,472	IVNDQGGQSV	Globulina 11S
	1003,5909	1003,5927	HVIKPPSRA	Globulina 11S
	1014,4976	1014,4982	AVNVDDPSKA	Globulina 11S
	991,5074	991,5087	KFNRPETT	Globulina 11S
	1034,4673	1034,4669	GYDNLNDPK	Aglutinina

Se encontró que las fracciones contenían diversos péptidos de diferentes proteínas de amaranto, cuya masa molecular osciló entre 800 y 1700 Da (7–15 aminoácidos). Es importante señalar, como ya se mencionó anteriormente, que sólo una parte de las proteínas de semilla de amaranto tienen secuencia informada en la base de datos UniProt, debido al conocimiento incompleto del genoma de *Amaranthus*. Por lo tanto, sólo los péptidos pertenecientes a esas proteínas referenciadas pueden ser identificados mediante este procedimiento y no podemos descartar la presencia en cada fracción de otros péptidos que no se han identificado contra la base de datos empleada.

3.3.2 Relación entre la estructura de péptidos y actividad antioxidante

Diferentes estudios realizados por distintos autores han analizado la relación entre la estructura de péptidos y su actividad antioxidante. Chen y col., (1995) obtuvieron seis péptidos con actividad antioxidante, medida por su capacidad de peroxidación del ácido linoleico, a partir de digeridos de β -conglucina de soja. Estos péptidos contienen aminoácidos hidrofóbicos, tales como Val y Leu en posición N-terminal y Pro (P), His (H) o Tyr (Y) en otras zonas de sus secuencias. Posteriormente, utilizando péptidos sintéticos derivados de LLP_{HH}, estos autores pudieron demostrar que el segmento HH era importante para la actividad antioxidante. La supresión de residuo His (H) de la zona C-terminal reducía significativamente el poder antioxidante. Por otra parte la supresión de Leu (L) de la posición N-terminal no tenía efecto sobre la actividad en estudio, mientras que His (H) y Pro (P) jugaban un rol importante (Chen y col., 1996). A efectos de dilucidar cuales secuencias peptídicas derivadas de proteínas de suero de leche podrían ser más activas, Hernández-Ledesma y col., (2005) evaluaron la actividad ORAC de cada aminoácido. Estos autores mostraron que Trp (W), Tyr (Y) y Met (M) mostraron las mayores actividades antioxidantes, seguidos de Cys (C), His (H) y Phe (F); el resto de los aminoácidos no presentó actividad. La mayor actividad antioxidante mostrada por Trp (W) y Tyr (Y) puede ser explicada por su capacidad para donar hidrógeno. Por otra parte, Met (M) podría ser oxidada a sulfóxido de Met (M) y Cys (C) puede donar el hidrógeno del sulfuro. Con posterioridad a la evaluación por ORAC de los péptidos seleccionados, estos autores concluyeron que la conformación de los péptidos puede conducir tanto a efectos sinérgicos como antagónicos en comparación a la actividad antioxidante de los aminoácidos. En otras palabras, la actividad

antioxidante, evaluada por ORAC, de péptidos derivados de proteínas de β -lactoglobulina mostraron alta actividad en la medida que contuvieran en su secuencias residuos de Trp (W), Tyr (Y) y Met (M) (Hernández-Ledesma y col., 2007). Adicionalmente, estos autores postularon que la presencia de Tyr (Y) y Met (M) en posición C-terminal, así como la existencia de Trp (W) y Tyr (Y) en la posición N-terminal aumentaba la actividad secuestrante de los péptidos. La presencia de Tyr (Y) en la posición N-terminal también ha sido descrita como un factor determinante en la actividad secuestrante de radicales para péptidos derivados de ovoalbúmina (Dávalos y col., 2004). Otro aminoácido que aumentaría la capacidad secuestrante del péptidos sería el residuo de Leu (Alemán y col., 2011). Algunos péptidos carentes de aminoácidos donantes de protones, por ejemplo GGAN, presentaron también actividad secuestrante de radicales (Li y col., 2007). Debido a sus características estructurales, Gly (G) – que confiere flexibilidad– y Pro (P) –que interrumpe la estructura secundaria e impone restricciones conformacionales– pueden incrementar la disponibilidad de los aminoácidos que pueden actuar como antioxidantes (Alemán y col., 2011).

La relación entre estructura y actividad antioxidante de péptidos ha sido también analizada usando la técnica de modelado QSAR (Li y Li 2013). Estos autores obtuvieron a partir de información bibliográfica secuencias de péptidos antioxidantes y los organizaron en tres bases de datos: capacidad antioxidante equivalente a trolox (TEAC), a ORAC y capacidad para secuestrar radicales superóxido (SOR). El modelado QSAR mostró las relaciones entre las propiedades fisicoquímicas de las regiones C-terminal y N-terminal y la potencia antioxidante, revistiendo una mayor importancia la zona C-terminal. Las posiciones más relevantes fueron: $C2 > C1$ para TEAC, $C3 > C4 > C1 > N1 > N2$ para ORAC, y $C4 > C1 > N1$ para SOR. En los tres casos, los aminoácidos hidrofóbicos voluminosos localizados en la región C-terminal se relacionaron con la actividad antioxidante. En el caso de ORAC, los aminoácidos polares/cargados en la posición C1 como también los residuos aminoacídicos con bajas propiedades electrónicas en las posiciones N1 y N2 contribuyen a la capacidad antioxidante de los péptidos (Li y Li, 2013).

3.3.3 Selección de péptidos potencialmente antioxidantes para su síntesis y evaluación de su actividad

La comparación de las fracciones analizadas, a partir de las obtenidas en nuestro laboratorio, mostró la presencia de péptidos (especialmente los de la globulina 11S) con secuencias superpuestas. Para la globulina 11S, estas secuencias se clasificaron en 9 grupos (TABLA III.4). Esta proteína, también llamada amarantina, en su forma nativa adopta una conformación hexamérica heterogéneas (330 y 400 kDa). La mayoría de los grupos de péptidos detectados (1 a 6) se localizaron en la subunidad ácida, mientras que los grupos 7, 8 y 9 eran parte de la subunidad básica de la globulina 11S. Teniendo en cuenta la información detallada previamente que relaciona la estructura de péptidos con su actividad antioxidante se seleccionaron 10 péptidos –representativos de diferentes grupos para ser sintetizados, los mismos se encuentran resaltados con negrita en la TABLA III.4.

TABLA III.4 Grupos de péptidos obtenidos por digestión gastrointestinal de globulina 11S de amaranto. Se indica el peso molecular (Mr) y la localización de estos péptidos dentro de la secuencia de la globulina 11S, así como la actividad antioxidante (ORAC) de los péptidos sintetizados seleccionados.

Grupo	Fracción	Secuencia	Mr	N° aa	Localización	IC50	TE	ID
1	F7	TEVWDSNE	1019,4308	8	35–42			
	F7	TEVWDSNEQ	1090,4679	9	35–43	0,020 (0,017–0,023)	1,82	1
2	F7	IYIEQGNGITG	1163,5822	11	74–85			
	F7	IYIEQGNGITGM	1294,6227	12	74–86	0,017 (0,015–0,020)	2,52	2
	F16	IEQGNGITGM	1018,4753	10	76–86			
3	F16	GDRFQDQHQ	1129,4901	9	121–130	0,60 (0,49–0,72)	0,06	3
	F22	GMRGDRFQDQHQK	1601,7481	13	118–130			
4	F16	AGKPQQEHSGEHQFS	1665,7485	15	184–198			
	F22	YLAGKPQQEHSGEH	1579,7379	14	183–196			
	F20	LAGKPQQEHSGEHQ	1544,7332	14	182–195	0,14 (0,12–0,17)	0,36	4
	F16	AGKPQQEHSGEHQ	1431,6491	13	184–196			
	F22	YLAGKPQQEH	1169,5829	10	182–191	0,016 (0,013–0,018)	2,34	5
5	F25	LQAEQDDRG	1030,4679	9	234–241			
	F20	LQAEQDDR	973,4465	8	234–242	na	-	6
6	F30	VIKPPSRA	866,5338	8	251–259			
	F30	HVIKPPSRA	1003,5927	9	250–259	0,57 (0,50–0,65)	0,06	7
7	F16	WEEREQ	875,3773	6	261–266			
	F7	WEEREQGS	1019,4308	8	261–268			
	F22	AWEEREQGSR	1246,5690	10	260–269	0,0067 (0,0063–0,0073)	6,07	8
8	F30	AVNVDDPSKA	1014,4982	10	288–297	na	-	9
9	F22	KFNRPETT	991,5074	8	441–448	na	-	10
	F30	FNRPETT	863,4137	7	442–448			

Los péptidos seleccionados para la síntesis y la evaluación de la actividad antioxidante se indican en negrita. na: no hay actividad para concentraciones de hasta 1 mg/ml. TE: actividad expresada como equivalentes de Trolox: $\mu\text{mol Trolox} / \mu\text{mol péptido}$. IC50: se obtuvo a partir del ajuste de las curvas dosis - respuesta; Los valores entre paréntesis representan el intervalo de confianza del 95%.

La actividad antioxidante de los péptidos sintetizados fue evaluada por el método de ORAC por la Dra. Orsini Delgado. Los resultados obtenidos (**TABLA III.4**) mostraron que los péptidos 6, 9 y 10, correspondientes a los grupos 5, 8 y 9 respectivamente, no mostraron actividad antioxidante al menos hasta una concentración de 1 mg/ml. Por otra parte, los siete péptidos restantes mostraron una actividad antioxidante creciente en función de la concentración de cada uno de ellos, por lo que fue posible calcular los respectivos valores de IC_{50} . Como se muestra en la **TABLA III.4**, el péptido más activo fue el péptido 8 (grupo 7, fracción 22) seguido por el péptido 5 (grupo 4, fracción 22), 2 (grupo 2, fracción 7) y el 1 (grupo 1, fracción 7).

En la **FIGURA III.9** se muestra la secuencia completa de la globulina 11S y se han resaltado las secuencias peptídicas seleccionadas para su síntesis. Además, dichas secuencias han sido localizados en la estructura homotrimérica de la globulina 11S con excepción del péptido AWEEREQGSR por estar ubicado en la región móvil localizada en la cara interna del trímero (cara IE). La cara IA corresponde al lado expuesto al medio de la estructura hexamérica.

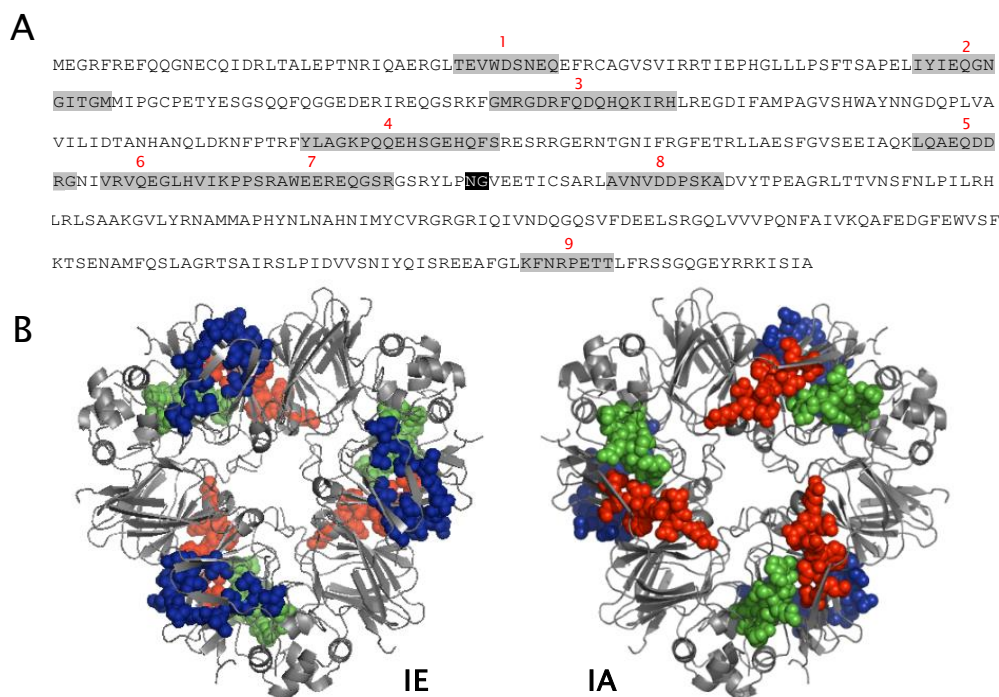


FIGURA III.9 (A) Secuencia completa de globulina 11S de *A. hypochondriacus*. Los rectángulos grises indican la posición de los péptidos detectados por MALDI-TOF. **(B)** Péptidos sintetizados marcados en la estructura del homotr trímero de globulina 11S (PDB 3QAC): TEVWDSNEQ (ROJO), YIEQNGITGM (AZUL), LAGKPQQEHSGEHQY YLAGKPQQEH (VERDE). Con las siglas IE se indica la cara que participa en las interacciones hidrofóbicas entre dos trímeros para formar la estructura hexamérica de la proteína nativa. La cara IA queda expuesta al medio. AWEEREQGSR no se muestra en la estructura por encontrarse en una región móvil en la cara IE.

Dado que los péptidos que presentaron mayor potencia antioxidante contienen entre 9 y 12 residuos aminoacídicos, ellos pueden adoptar alguna conformación tridimensional que podría afectar su actividad. Por esta razón se generaron las estructuras peptídicas de los mismos haciendo uso del programa PEPFOLD, el cual es un recurso en línea para la predicción *de novo* de la estructura de péptidos a partir de la secuencia de aminoácidos (Maupetit y col., 2009). La conformación seleccionada representa la conformación de energía más baja de las estructuras generadas por PEP-FOLD. Con el programa PyMOL (Molecular Graphics System, Version 1.7.4 Schrödinger, LLC) se generaron las imágenes y se realizó el análisis de cada estructura (FIGURA III.10).

El péptido 8 resultó mucho más potente que la fracción completa digerida. Su secuencia muestra un aminoácido aromático voluminoso en posición N2, otro aminoácido voluminoso en posición C4 y un aminoácido cargado positivamente en posición C1, características relevantes según QSAR para la actividad ORAC. Este péptido es el que exhibe la estructura más plegada, con cinco giros, y más compacta. El mismo gira, con los anillos de Trp y Arg expuestos al medio (FIGURA III.10, E).

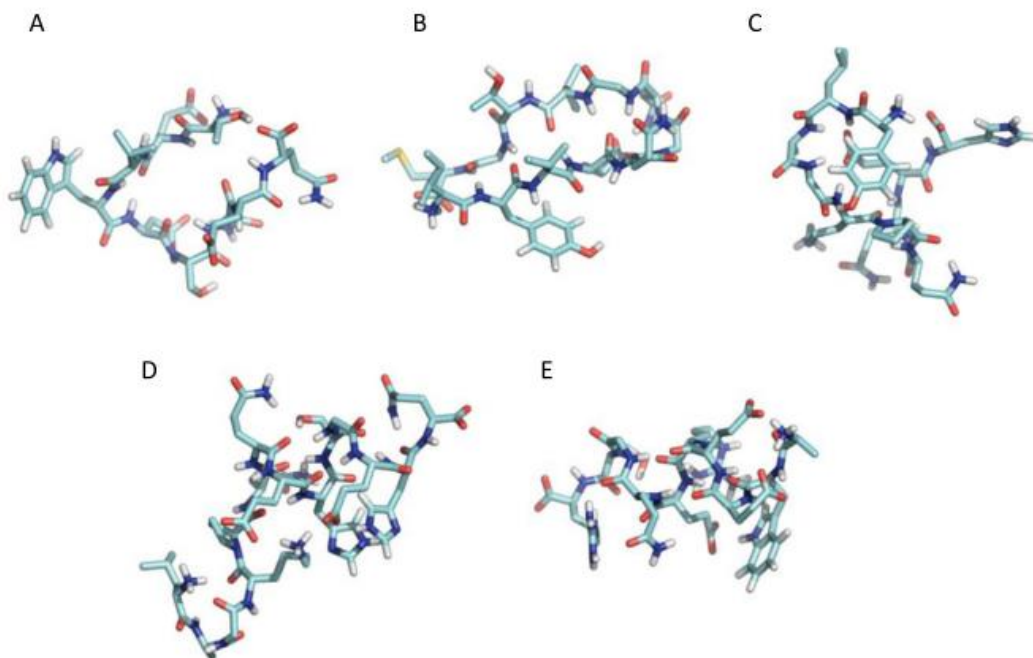


FIGURA III.10 Estructuras obtenidas por PEPFOLD de péptidos de amaranto con capacidad antioxidante: (A) TEVWDSNEQ; (B) IYIEQGNIGTGM; (C) LAGKPQQEHSGEHQ; (D) YLAGKPQQEH; (E) AWEEREQGSR

El péptido 5 muestra un aminoácido voluminoso en la posición N1, un aminoácido positivamente cargado en posición C1 y residuos aminoacídicos voluminosos en C3, C4 y N2. Este péptido a pesar de poseer el mismo número de aminoácidos que el péptido 8, diez residuos aminoacídicos, presenta sólo 3 giros, con los residuos de Tyr e His expuestos al medio (**FIGURA III.10, D**).

El péptido 4 contiene una secuencia adicional, a la correspondiente al péptido 5, en el extremo C-terminal y ha perdido el residuo de Tyr en el extremo N-terminal, esto ha sido suficiente para reducir su actividad antioxidante 10 veces respecto a la exhibida por el péptido 5. Si se comparan las estructuras de los péptidos 4 y 5 se puede observar en el péptido 4 una estructura más plegada con dos anillos de His localizados uno muy cercano al otro (**FIGURA III.10, C**).

El péptido 2, presenta aminoácidos voluminosos en posición N2 (Tyr), N1, N3 y C4 (Ile) y Met en posición C1. El anillo aromático del residuo Tyr y la Met se encuentran muy expuestos al medio (**FIGURA III.10, B**).

Finalmente, el péptido 1 contiene Trp en el medio de la molécula, un aminoácido cargado en posición C2 y diversos aminoácidos voluminosos. En este péptido el anillo aromático del Trp está localizado hacia afuera de la estructura (**FIGURA III.8, A**). Tanto el péptido 1 como el 2 mostraron solamente un giro con las regiones N y C terminal en la misma dirección (**FIGURA III.10, A y B**).

Cabe remarcar que los cuatro péptidos más activos (8, 5, 2 y 1) están localizados en la región menos estructurada de la subunidad ácida de la globulina 11S. Los péptidos 1 y 5 fueron localizados en una región de conformación extendida + giro (turns), mientras que el péptido 2 se encontró en una región de conformación extendida + región helicoidal. El péptido 8, el más activo de todos los péptidos sintetizados, se encuentra localizado en una región helicoidal móvil. Adicionalmente, los péptidos 2 y 8 se encuentra ubicados en el lado hidrofóbico, IE, de la estructura de la glicina (globulina 11S), mientras que los péptidos 1 y 5 se localizan en el lado expuesto, IA, sugiriendo un carácter más hidrofílico de sus secuencias (**FIGURA III.9**).

Los resultados alcanzados indican que los péptidos de amaranto con capacidad antioxidante identificados exhiben algunas características estructurales compatibles con características previamente descritas como relevantes para esta actividad biológica haciendo uso del método QSAR. Todos ellos, como hemos visto presentaron al menos un aminoácido voluminoso de naturaleza aromática. Además, los péptidos más activos se encuentra localizados en pequeñas regiones estructuradas, internas o expuestas de la globulina 11S.

3.4 CUPINAS Y LECTINAS

Como vimos en el CAPÍTULO 1, las proteínas que presentan mayor cantidad de péptidos bioactivos pertenecen a superfamilias con mayor diversidad, estructural y secuencial (**FIGURA I.11**, Pág. 61). Por otra parte, existen evidencias suficientes de la presencia de péptidos encriptados en las proteínas de amaranto (sección II.3, Introducción General). En particular, en este capítulo hemos comprobado la presencia de secuencias activas en la globulina 11S y aglutinina de amaranto, entonces no preguntamos si la familia estructural a la cual pertenecen se encuentra entre las que presentan mayor diversidad.

La globulina 11 S (PDB 3QAC) forma parte de las globulinas de reserva de amaranto las cuales son incluidas dentro de una superfamilia más amplia denominada “superfamilia de cupinas”. Dentro de esta clasificación se agrupan distintas proteínas que cubren funciones biológicas muy diversas (Khuri y col., 2001; Dunwell y col., 2001, 2004). Estas proteínas comparten dos motivos de secuencias cortas de consenso conservadas y un dominio estructural denominado *cupin* (Pfam id PF00190) de donde proviene el término cupinas (Dunwell, 1998).

La lectina de amaranto pertenece a la fracción albúmina que no es de reserva. Su estructura nativa es homodimérica constituida por dos subunidades de 33 kDa, cada una de las cuales comprende dos dominios homólogos de amarantina (Pfam id PF07468) en tándem. Su función aún no ha sido asignada, pero se ha sugerido que está implicada en la señalización del mecanismo de defensa de plantas (Lannoo y Van Damme, 2014; Manning y col., 2017).

Es interesante resaltar que tanto la globulina 11S como la aglutinina pertenecen a superfamilias con gran diversidad estructural (superfamilias 2.60.120.10. y 2.80.10.50 respectivamente). En la **FIGURA III.11** se muestran los gráficos de diversidad estructural en función de la diversidad secuencial para ambas

superfamilias. Del mismo modo que se observó en la **FIGURA I.11** del Capítulo 1 (Pág. 62) se muestra el número de clusters al 35% de identidad vs el número de clusters estructurales para cada una de las superfamilias.

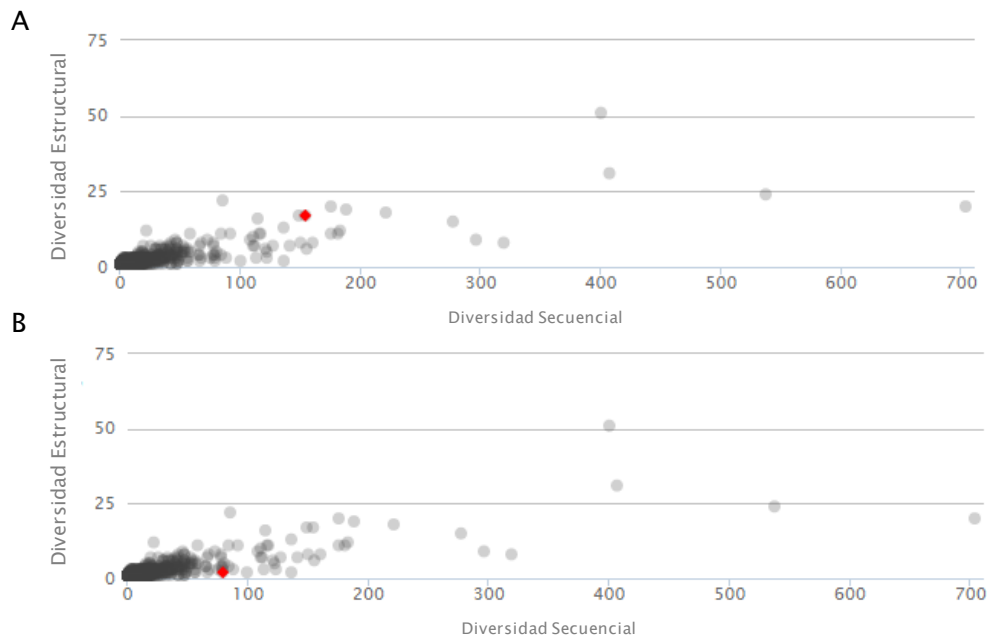


FIGURA III.11: Número de familias de homólogas al nivel 35% de identidad (S35) (diversidad secuencial) vs número de clusters estructurales (diversidad estructural) para todas las superfamilias contenidas en CATH (puntos grises). Las superfamilias **(A)** 2.60.120.10 (11S) y **(B)** 2.80.10.50 (aglutinina) se encuentran coloreadas en rojo.

Estas observaciones nos abren la posibilidad de usar la diversidad secuencial y estructural para seleccionar posibles candidatos proteicos que puedan contener PBs. Como vimos en el capítulo 1, una vez detectadas las superfamilias con PBs, la existencia de posibles hot-spots ayudaría en la predicción de la ocurrencia de PBs en regiones específicas de las estructuras proteicas. Otra ventaja adicional que supondría este método de selección de candidatos, es que sería aplicable al estudio de proteínas candidatas que aún no han sido secuenciadas pero que sí se conozcan proteínas homólogas cercanas secuenciadas. Por ej. en nuestro caso particular la 7S, una proteína mayoritaria en los aislados de amaranto, que no se conoce su secuencia. En cambio, se conoce la secuencia de una proteína homóloga cercana, una vicilina perteneciente a la especie *Sesamum indicum* (Quiroga y col., 2010). Estas globulinas pertenecen a la superfamilia de las cupinas, y a la superfamilia 2.60.120.10, que como se mencionó anteriormente muestra una diversidad secuencial y estructural alta comparada con el resto de las superfamilias en CATH, lo que la convierte también en una candidata para evaluar la ocurrencia de PBs en su secuencia. Al tratarse de una homóloga cercana a la 7S de amaranto, se podría extrapolar la predicción realizada.

Evidentemente, como comentamos en el Capítulo 1 estos resultados preliminares son promisorios para el desarrollo de un método bioinformático para predecir qué proteínas serían las mejores candidatas para evaluar la existencia de PBs. Claramente, este método deberá ser desarrollado y correctamente evaluado para medir su capacidad predictiva.

3.5 CONCLUSIONES CAPÍTULO 3

La aplicación de técnicas bioinformáticas resulta una herramienta fundamental en el estudio de la ocurrencia y actividad de PBs. La aproximación experimental empleada en laboratorios como el nuestro, en general, permiten obtener fracciones con actividad biológica comprobada, purificadas, con respeto al material de partida, pero lejos están de resultar lo suficientemente puras como para poder identificar en ellas una única secuencia. Dichas fracciones, complejas desde el punto de vista de su composición, hacen necesarias la aplicación de sucesivas etapas de purificación (usualmente técnicas cromatografías) incrementando los costos y tiempo de experimentación.

El uso de espectrometría de masas aplicada a fracciones que presentan esta composición tan compleja requiere necesariamente del uso de bases de datos secuenciales, hecho que introduce una limitación adicional al proceso de identificar la sustancia responsable de la actividad biológica evaluada.

En este capítulo se aplicaron aproximaciones bioinformáticas para la selección de secuencias identificadas en fracciones activas basadas en las particularidades de la actividad antioxidante y antitrombótica.

En el estudio de la actividad antitrombótica se empleó una aproximación basada en las características de los sustratos de la enzima trombina para seleccionar, de una posible lista de 57 secuencias activas, 10 para ser sintetizadas. En la posterior evaluación *in vitro* de la actividad antitrombótica ninguna de las secuencias seleccionadas resultó activa en las concentraciones evaluadas. Restaría plantear a futuro una estrategia distinta para identificar los péptidos responsables de la actividad. Aún, teniendo en cuenta que varios componentes de la fracción activa no han podido ser secuenciados por espectrometría de masas, el/los péptidos activos podrían encontrarse en la lista de secuencias

descartadas. A pesar de estos resultados negativos, hemos aprendido que las características fisicoquímicas de los aminoácidos y por ende de los péptidos, no fue suficiente para seleccionar los mejores péptidos candidatos. Posiblemente, su disposición espacial (conformaciones) enmascare ciertas propiedades o simplemente la interacción con el agua y/o interacciones intramoleculares cambien ciertas propiedades evaluadas en este trabajo. En este sentido, pensamos aplicar métodos de docking molecular proteína–proteína. Esta metodología abarca un conjunto de técnicas que simulan computacionalmente la unión de un ligando a una proteína. Estos métodos permitirían utilizar todas las secuencias identificadas en la fracción activa para reconocer ligandos particulares de una determinada enzima. Así, estos estudios podrían permitir seleccionar una nueva lista de candidatos a sintetizar. Sin embargo los protocolos de docking proteína–proteína no se encuentran optimizados para secuencias peptídicas mayores a 3 o 4 residuos (como es en nuestro caso), y aunque en general se entienda que los costos de las aproximaciones *in silico* son menores a los costos experimentales, para sistemas proteína–proteína los tiempos de cálculo para obtener resultados confiables suelen ser muy altos debido a la gran variedad de conformaciones que adquieren polipéptidos de más de 5 aminoácidos de longitud. A pesar de estos inconvenientes, creemos que una combinación de datos y metodologías (análisis fisicoquímico, docking, estudio conformacional de péptidos, etc) reduciría el mínimo los errores en las predicciones de PBs.

En cuanto a la actividad antioxidante se aplicó una estrategia basada en conocimientos generados mediante modelos QSAR. Esto fue posible porque la actividad antioxidante de componentes peptídicos ha sido estudiada más en profundidad que la actividad antitrombótica. Se logró seleccionar de una lista compleja de posibles secuencias activas 10 para su posterior síntesis y evaluación de la actividad *in vitro*. El criterio de selección empleado permitió la identificación de 4 secuencias nuevas, que portan nuevas características estructurales para relacionar estructura–actividad. Las estructuras obtenidas por Pep–Fold constituyen solo un modelo para visualizar una de las probables estructuras que podrían adoptar los péptidos.

CONCLUSIONES GENERALES

En el presente trabajo de tesis se han analizado y discutido las diferentes formas en las que han sido abordados los estudios relacionados con péptidos bioactivos derivados de proteínas alimentarias. En particular se han descrito las bases de datos de referencia en esta temática – BIOPEP– la cual presenta limitaciones en su construcción, las que dificultan el empleo de la información biológica almacenada en las secuencias peptídicas reportadas, desde un enfoque bioinformático.

Se ha caracterizado el espacio secuencial de las proteínas que contienen encriptados PB que poseen un largo de cadena superior a 5 aminoácidos, a fin de generar conocimiento acerca de la relación evolutiva y estructural de estas proteínas. Es aceptado por la comunidad científica que cualquier proteína puede ser fuente de péptidos bioactivos ya que si tomamos una secuencia de 2 o más aminoácidos es altamente probable que la misma se encuentre en alguna zona del espacio secuencial total. Sin embargo, cuando estudiamos la distribución de PB observamos que la cantidad de péptidos encriptados que contiene una proteína no se correlaciona con su longitud. Este hecho nos sugiere que la ocurrencia de PB no se debe solo a la probabilidad de existencia al azar de una determinada secuencia aminoacídica. La misma tendencia fue encontrada cuando se analizaron los resultados obtenidos mediante la asignación estructural llevada a cabo. Las secuencias que presentan mayor número de PBs se encuentran entre las superfamilias estructurales de mayor diversidad estructural y secuencial. Esto constituye un conocimiento novedoso, que trasladado a la investigación aplicada implicaría que al momento que se decide

investigar una nueva fuente de proteínas como fuente de péptidos bioactivos el conocimiento de su clasificación estructural puede contribuir a la selección o descarte de esta fuente. Esta nueva alternativa podría ser aplicada aún cuando se desconozca la secuencia de las proteínas de interés, en la medida que se sepa la secuencia de proteínas homólogas cercanas. Por otra parte, este conocimiento puede ser la base del desarrollo de un nuevo método bioinformático para predecir que proteínas serían las mejores candidatas para evaluar la existencia de PBs. Nos proponemos en un futuro llevar adelante esta tarea y una vez desarrollado el método evaluar correctamente su capacidad predictiva.

En el presente trabajo de tesis también se ha descrito la aplicación de herramientas bioinformáticas y ensayos de bioactividad experimentales para la identificación de péptidos bioactivos en amaranto. Se estudió la actividad hipocolesterolémica mediante ensayos de inhibición de la HMGR humana, enzima central en la biosíntesis de colesterol. Los resultados obtenidos mostraron, en principio, la presencia de inhibidores de esta enzima cuya actividad resulta independiente de la concentración de proteínas y/o péptidos presentes en aislados, en hidrolizados y en digeridos de amaranto. Como era esperable la liberación de péptidos de las secuencias de las proteínas de almacenamiento del grano –ya sea por acción de enzimas endógenas o simulación del proceso gastrointestinal– produjo un mayor grado de inhibición. Cabe señalar que este es uno de los escasísimos estudios relacionados con una posible actividad hipocolesterolémica asociada a péptidos de origen vegetal; se han descrito solamente tres trabajos en soja y uno en amaranto.

Las dificultades experimentales encontradas en la determinación de la HMGR, así como la limitación que representa el desconocimiento de la secuencia aminoacídica de la totalidad de las proteínas de almacenamiento del grano de amaranto, nos impidió arribar a resultados más precisos sobre la identificación de péptidos con potencial actividad hipocolesterolémica inducida por la inhibición de la HMGR, así como profundizar en el mecanismo de acción de los mismos. Sin embargo, resulta promisorio el haber detectado que tanto los Ais como los Hid y Dig presentan una acción inhibitoria de la HMGR, siendo requeridos futuros estudios para corroborar estos resultados y determinar su posible sitio de acción.

El poder abordar el estudio de potenciales péptidos de origen alimentario con actividad fisiológica a través de la combinación de herramientas bioinformáticas y ensayos experimentales *in vitro* es fundamental, particularmente en nuestra

región donde resulta muy difícil lograr una excelente purificación e identificación experimental de las secuencias aminoacídicas de los probables péptidos bioactivos. Esta aproximación reduce sustancialmente el número de péptidos a sintetizar y contribuye además en la determinación del probable mecanismo de acción.

Como se indicó precedentemente, la bioinformática es hoy una herramienta potente que permite la aplicación de métodos computacionales para manejar, curar e interpretar información sobre sistemas biológicos. La aplicación de este tipo de herramientas al estudio de la actividad antioxidante y antitrombótica realizada ha enriquecido el trabajo de las Dras. Sabbione y Orsini Delgado, y ha permitido particularmente identificar cuatro nuevos péptidos antioxidantes encriptados en la secuencia de las proteínas de almacenamiento de amaranto. Uno de estos péptidos, sintetizado haciendo uso de la información derivada del abordaje bioinformático, presentó *in vitro* una alta bioactividad restando aún la confirmación de la misma mediante ensayos *in vivo* en un sistema modelo animal.

Una de las perspectivas futuras para continuar con lo iniciado en este trabajo de Tesis es la implementación de técnicas de docking proteína-proteína como herramienta para seleccionar péptidos a sintetizar de una posible lista de péptidos antitrombóticos, inhibidores de la enzima trombina, e hipocolesterolémicos, inhibidores de la enzima HMGR.

Finalmente, durante este trabajo de tesis he adquirido conocimientos básicos y capacidad para utilizar diferentes herramientas bioinformáticas, el lenguaje de programación C, shell scripts, etc.; conocimientos que me permitirán iniciar una nueva sub-línea de investigación en el grupo de proteínas vegetales dirigido por la Dra. M. Cristina Añón.

BIBLIOGRFÍA

- Abugoch, L. E. (2006). Relación estructura–funcionalidad de glutelinas y aislados proteicos de amaranto (*Amaranthus hypochondriacus*). Universidad Nacional de la Plata.
- Abugoch, L. E., Martínez, E. N., & Añón, M. C. (2010). Influence of pH on Structure and Function of Amaranth (*Amaranthus hypochondriacus*) Protein Isolates. *Cereal Chemistry*, 87(5), 448–453. <http://doi.org/10.1094/CCHEM-09-09-0125>
- Ačimovič, J., & Rozman, D. (2013). Steroidal triterpenes of cholesterol synthesis. *Molecules*, 18(4), 4002–4017. <http://doi.org/10.3390/molecules18044002>
- Agyei, D., & Danquah, M. K. (2012). Rethinking food–derived bioactive peptides for antimicrobial and immunomodulatory activities. *Trends in Food Science & Technology*, 23(2), 62–69. <http://doi.org/10.1016/j.tifs.2011.08.010>
- Akhavan, T., Luhovyy, B. L., Brown, P. H., Cho, C. E., & Anderson, G. H. (2010). Effect of premeal consumption of whey protein and its hydrolysate on food intake and postmeal glycemia and insulin responses in young adults. *American Journal of Clinical Nutrition*, 91(4), 966–975. <http://doi.org/10.3945/ajcn.2009.28406>
- Alemán, A., Giménez, B., Pérez–Santín, E., Gómez–Guillán, M. C., & Montero, P. (2011). Contribution of Leu and Hyp residues to antioxidant and ACE–inhibitory activities of peptide sequences isolated from squid gelatin hydrolysate. *Food Chemistry*, 125(2), 334–341. <http://doi.org/10.1016/j.foodchem.2010.08.058>
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI–BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25(17), 3389–3402. <http://doi.org/gka562>
- Alvarez–Jubete, L., Arendt, E. K., & Gallagher, E. (2010). Nutritive value of pseudocereals and their increasing use as functional gluten–free ingredients. *Trends in Food Science and Technology*. 21(2), 106–113. <http://doi.org/10.1016/j.tifs.2009.10.014>
- Amin, I., Norazaidah, Y., & Hainida, K. I. E. (2006). Antioxidant activity and phenolic content of raw and blanched *Amaranthus* species. *Food Chemistry*, 94(1), 47–52. <http://doi.org/10.1016/j.foodchem.2004.10.048>
- AOAC. (1984). Association of Official Analytical Chemists Inc. Official Methods of Analysis (Ed. 14th). Ed Williams S., Arlington, VA. Washington, DC.
- AOAC. (1990). Association of Official Analytical Chemists Inc. Official Methods of Analysis. (Ed.15th). Ed Williams S., Arlington VA. Washington, DC.
- Atkinson, A. B., & Robertson, J. I. S. (1979) Captopril in the treatment of clinical hypertension and cardiac failure. *The Lancet*, 314(8147):836–839. [http://doi.org/10.1016/S0140-6736\(79\)92186-X](http://doi.org/10.1016/S0140-6736(79)92186-X)
- Bale, J. R., & Kauffman, C. S. (1992). Editors’ overview. *Food Reviews International*, 8(1), 1–4. <http://doi.org/10.1080/87559129209540926>
- Barba de la Rosa, A. P., Gueguen, J., Paredes–Lopez, O., & Viroben, G. (1992). Fractionation procedures, electrophoretic characterization, and amino acid composition of amaranth seed proteins. *Journal of Agricultural and Food Chemistry*, 40(6), 931–936. <http://doi.org/10.1021/jf00018a002>
- Barba de la Rosa, A. P., Barba Montoya, A., Martínez–Cuevas, P., Hernández–Ledesma, B., León–Galván, M. F., De León–Rodríguez, A., & González, C. (2010). Tryptic amaranth glutelin digests induce endothelial nitric oxide production through inhibition of ACE: Antihypertensive role of amaranth peptides. *Nitric Oxide – Biology and Chemistry*, 23(2), 106–111. <http://doi.org/10.1016/j.niox.2010.04.006>
- Barrio, D. A., & Añón, M. C. (2010). Potential antitumor properties of a protein isolate obtained from the seeds of *Amaranthus mantegazzianus*. *European Journal of Nutrition*, 49(2), 73–82.

<http://doi.org/10.1007/s00394-009-0051-9>

Bazzano, L. A., He, J., Ogden, L. G., Loria, C. M., Vupputuri, S., Myers, L., & Whelton, P. K. (2002). Fruit and vegetable intake and risk of cardiovascular disease in US adults: the first National Health and Nutrition Examination Survey Epidemiologic Follow-up Study. *The American Journal of Clinical Nutrition*, 76(1), 93-9.

Benítez, R., Ibarz, A., & Pagan, J. (2008). Hidrolizados de proteína: procesos y aplicaciones. *Acta Bioquím. Clín. Latinoam.*, 42, 227-236.

Bischoff, K. M., & Rodwell, V. W. (1996). 3-Hydroxy-3-methylglutaryl-coenzyme A reductase from *Haloferax volcanii*: purification, characterization, and expression in *Escherichia coli*. *Journal of Bacteriology*, 178(1), 19-23.

Bochar, D. A., Stauffacher, C. V., & Rodwell, V. W. (1999). Sequence comparisons reveal two classes of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Molecular Genetics and Metabolism*, 66(2), 122-127. <http://doi.org/10.1006/mgme.1998.2786>

Bock, P. E., Panizzi, P., & Verhamme, I. M. A. (2007). Exosites in the substrate specificity of blood coagulation reactions. *Journal of Thrombosis and Haemostasis*, 5(suppl. 1), 81-94. <http://doi.org/10.1111/j.1538-7836.2007.02496.x>

Bouglé, D., & Bouhallab, S. (2017). Dietary Bioactive Peptides: Human Studies. *Critical Reviews in Food Science and Nutrition*, 57(2), 335-343. <http://doi.org/10.1080/10408398.2013.873766>

Brantl, V., Teschemacher, H., Henschen, A., & Lottspeich, F. (1979). Novel opioid peptides derived from casein (β -casomorphins). I. Isolation from bovine casein peptone. *Hoppe-Seyler's Zeitschrift Fur Physiologische Chemie*, 360(9), 1211-1216. <http://doi.org/10.1515/bchm2.1979.360.2.1211>

Bratovanova, E. K., & Petkov, D. D. (1987). Glycine flanked by hydrophobic bulky amino acid residues as minimal sequence for effective subtilisin catalysis. *The Biochemical Journal*, 248(3), 957-60.

Bressani, R., & Garcia-Vela, A. L. (1990). Protein Fractions in Amaranth Grain and Their Chemical Characterization. *J. Agric. Food Chem*, 38(5), 1205-1209. <http://doi.org/10.1021/jf00095a010>

Buhaescu, I., & Izzedine, H. (2007). Mevalonate pathway: A review of clinical and therapeutical implications. *Clinical Biochemistry*, 40(9-10), 575-584. <http://doi.org/10.1016/j.clinbiochem.2007.03.016>

Castellani, O. F., Martínez, E. N., & Añón, M. C. (1998). Structural Modifications of an Amaranth Globulin Induced by pH and NaCl. *Journal of Agricultural and Food Chemistry*, 46(12), 4846-4853. <http://doi.org/10.1021/jf9802427>

Castellani, O. F., Martínez, E. N., & Añón, M. C. (2000). Amaranth globulin structure modifications induced by enzymatic proteolysis. *Journal of Agricultural and Food Chemistry*, 48(11), 5624-5629. <http://doi.org/10.1021/jf000624o>

Chanson-Rolle, A., Aubin, F., Braesco, V., Hamasaki, T., & Kitakaze, M. (2015). Influence of the Lactotriptides Isoleucine-Proline-Proline and Valine-Proline-Proline on Systolic Blood Pressure in Japanese Subjects: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *PloS One*, 10(11), e0142235. <http://doi.org/10.1371/journal.pone.0142235>

Chelliah, V., Chen, L., Blundell, T. L., & Lovell, S. C. (2004). Distinguishing structural and functional restraints in evolution in order to identify interaction sites. *J Mol Biol*, 342(5), 1487-1504. <http://doi.org/10.1016/j.jmb.2004.08.022>S0022-2836(04)00983-0

Chen, H., Muramoto, K., & Yamauchi, F. (1995). Structural analysis of a antioxidative peptides from soybean beta-conglycinin. *Journal Agricultural Food Chemistry*, 44, 2619-2623.

- Chen, H. M., Muramoto, K., Yamauchi, F., & Nokihara, K. (1996). Antioxidant Activity of Designed Peptides Based on the Antioxidative Peptide Isolated from Digests of a Soybean Protein. *Journal of Agricultural and Food Chemistry*, 44(9), 2619–2623. <http://doi.org/10.1021/jf950833m>
- Chen, H., Muramoto, K., Yamauchi, F., Fujimoto, K., & Nokihara, K. (1998). Antioxidative Properties of Histidine-Containing Peptides Designed from Peptide Fragments Found in the Digests of a Soybean Protein. *Journal of Agricultural and Food Chemistry*, 46(1), 49–53. <http://doi.org/Doi 10.1021/jf970649w>
- Choi, J., Sabikhi, L., Hassan, A., & Anand, S. (2012). Bioactive peptides in dairy products. *International Journal of Dairy Technology.*, 65(1), 1–12. <http://doi.org/10.1111/j.1471-0307.2011.00725.x>
- Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J*, 5(4), 823–826.
- Cian, R. E., Martínez-Augustin, O., & Drago, S. R. (2012). Bioactive properties of peptides obtained by enzymatic hydrolysis from protein byproducts of *Porphyra columbina*. *Food Research International*, 49(1), 364–372. <http://doi.org/10.1016/j.foodres.2012.07.003>
- Cicero, A. F. G., Colletti, A., Rosticci, M., Cagnati, M., Urso, R., Giovannini, M., ... D'Addato, S. (2016). Effect of Lactotriptides (Isoleucine-Proline-Proline/Valine-Proline-Proline) on Blood Pressure and Arterial Stiffness Changes in Subjects with Suboptimal Blood Pressure Control and Metabolic Syndrome: A Double-Blind, Randomized, Crossover Clinical Trial. *Metabolic Syndrome and Related Disorders*, 14(3), 161–166. <http://doi.org/10.1089/met.2015.0093>
- Claverie, J. (2000). From Bioinformatics to Computational Biology From Bioinformatics to Computational Biology. *Genome Research*, 1277–1279. <http://doi.org/10.1101/gr.155500>
- Condés, M. C., Scilingo, A. A., & Añón, M. C. (2009). Characterization of amaranth proteins modified by trypsin proteolysis. Structural and functional changes. *LWT – Food Science and Technology*, 42(5), 963–970. <http://doi.org/10.1016/j.lwt.2008.12.008>
- Conforti F., Loizzo, M.R, Statti, G.A & Menichini, F. (2005). Comparative Radical Scavenging Methanolic Extract and Fractions and Antidiabetic Activities of from *Aehiuea ligústica* All. *Biological and Pharmaceutical Bulletin* 28(9) 1791–1794
- Corzo-Martínez, M., Corzo, N., & Villamiel, M. (2007). Biological properties of onions and garlic. *Trends in Food Science and Technology*, 18(12), 609–625. <http://doi.org/10.1016/j.tifs.2007.07.011>
- Cruz-Gonzalez, I., Sanchez-Ledesma, M., Baron, S. J., Healy, J. L., Watanabe, H., Osakabe, M., ... Jang, I.-K. (2008). Efficacy and safety of argatroban with or without glycoprotein IIb/IIIa inhibitor in patients with heparin induced thrombocytopenia undergoing percutaneous coronary intervention for acute coronary syndrome. *Journal of Thrombosis and Thrombolysis*, 25(2), 214–218. <http://doi.org/10.1007/s11239-007-0071-3>
- Dávalos, A., Miguel, M., Bartolomé, B., & López-Fandiño, R. (2004). Antioxidant Activity of Peptides Derived from Egg White Proteins by Enzymatic Hydrolysis. *Journal of Food Protection*, 67(9), 1939–1944. <http://doi.org/10.4315/0362-028X-67.9.1939>
- De Gennaro Colonna, V., Rossoni, G., Rigamonti, A. E., Bonomo, S., Manfredi, B., Berti, F., & Muller, E. E. (2002). Enalapril and quinapril improve endothelial vasodilator function and aortic eNOS gene expression in l-NAME-treated rats. *European Journal of Pharmacology*, 450(1), 61–66. [http://doi.org/10.1016/S0014-2999\(02\)02046-0](http://doi.org/10.1016/S0014-2999(02)02046-0)
- Dia, V. P., Torres, S., De Lumen, B. O., Erdman, J. W., & De Mejia, E. G. (2009). Presence of lunasin in plasma of men after soy protein consumption. *Journal of Agricultural and Food Chemistry*, 57(4), 1260–1266. <http://doi.org/10.1021/jf803303k>
- Dias, R., & de Azevedo, W. F. (2008). Molecular docking algorithms. *Current Drug Targets*, 9(12), 1040–7.

- Dinnella, C., Gargaro, M. T., Rossano, R., & Monteleone, E. (2002). Spectrophotometric assay using o-phthaldialdehyde for the determination of transglutaminase activity on casein. *Food Chemistry*, 78(3), 363–368. [http://doi.org/10.1016/S0308-8146\(02\)00109-7](http://doi.org/10.1016/S0308-8146(02)00109-7)
- Doyon, M., & Labrecque, J. (2008). Functional foods: a conceptual definition. *British Food Journal*, 110(11), 1133–1149. <http://doi.org/10.1108/00070700810918036>
- Dudek, A., Arodz, T., & Gálvez, J. (2006). Computational methods in developing quantitative structure–activity relationships (QSAR): a review. *Combinatorial Chemistry and High Throughput Screening*, 9, 213–228. <http://doi.org/10.2174/138620706776055539>
- Erdmann, K., Cheung, B. W. Y., & Schröder, H. (2008). The possible roles of food–derived bioactive peptides in reducing the risk of cardiovascular disease. *The Journal of Nutritional Biochemistry*, 19(10), 643–654. <http://doi.org/10.1016/j.jnutbio.2007.11.010>
- Fang, Y.–Z., Yang, S., & Wu, G. (2002). Free radicals, antioxidants, and nutrition. *Nutrition*, 18(10), 872–879. [http://doi.org/10.1016/S0899-9007\(02\)00916-4](http://doi.org/10.1016/S0899-9007(02)00916-4)
- Fisher, E. A., Feig, J. E., Hewing, B., Hazen, S. L., & Smith, J. D. (2012). High–Density Lipoprotein Function, Dysfunction, and Reverse Cholesterol Transport. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 32(12), 2813–2820. <http://doi.org/10.1161/ATVBAHA.112.300133>
- Fischer, M., Coleman, R. G., Fraser, J. S., & Shoichet, B. K. (2014). Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat Chem*, 6(7), 575–583.
- Fisher, R. A. (1954). *Statistical Methods for Research Workers*. (O. and Boyd, Ed.).
- Fitzgerald, C., Gallagher, E., Tasdemir, D., & Hayes, M. (2011). Heart Health Peptides from Macroalgae and Their Potential Use in Functional Foods. *Journal of Agricultural and Food Chemistry*, 59(13), 6829–6836. <http://doi.org/10.1021/jf201114d>
- Floris, R., Recio, I., Berkhout, B., & Visser, S. (2003). Antibacterial and Antiviral Effects of Milk Proteins and Derivatives Thereof. *Current Pharmaceutical Design*, 9(16), 1257–1275. <http://doi.org/10.2174/1381612033454810>
- Franzosa, E. A., & Xia, Y. (2006). Structural Determinants of Protein Evolution Are Context–Sensitive at the Residue Level. *Molecular Biology and Evolution*, (Dickerson 1971). <http://doi.org/10.1093/molbev/msp146>
- Friesen, J. A., & Rodwell, V. W. (2004). The 3–hydroxy–3–methylglutaryl coenzyme–A (HMG–CoA) reductases. *Genome Biology*, 5(11), 248. <http://doi.org/10.1186/gb-2004-5-11-248>
- Frimpong, K., Darny, B. G., & Rodwell, V. W. (1993). Syrian–Hamster–3–Hydroxy–3–methylglutaryl–coenzyme–A Reductase Expressed in *Escherichia coli* Production of Homogeneous Protein. *Protein Expression and Purification*, 4, 337–344.
- Frimpong, K., & Rodwell, V. W. (1994). The Active Site of Hamster 3–Hydroxy–3–methylglutaryl–CoA. *The Journal Of Biological Chemistry*, 2(14), 1217–1221.
- Fritz, M., Vecchi, B., Rinaldi, G., & Añón, M. C. (2011). Amaranth seed protein hydrolysates have in vivo and in vitro antihypertensive activity. *Food Chemistry*, 126(3), 878–884. <http://doi.org/10.1016/j.foodchem.2010.11.065>
- Furnham, N., Sillitoe, I., Holliday, G. L., Cuff, A. L., Rahman, S. a., Laskowski, R. a., ... Thornton, J. M. (2012). FunTree: A resource for exploring the functional evolution of structurally defined enzyme superfamilies. *Nucleic Acids Research*, 40(D1), 776–782. <http://doi.org/10.1093/nar/gkr852>
- Fuss, C., Palmaz, J. C., & Sprague, E. A. (2001). Fibrinogen : Structure , Function , and Surface Interactions, 12(6), 677–682. [http://doi.org/10.1016/S1051-0443\(07\)61437-7](http://doi.org/10.1016/S1051-0443(07)61437-7)

- García Fillería, S. F., & Tironi, V. A. (2015). Application of amaranth protein isolate and hydrolysate on a reduced salt fish restructured product: antioxidant properties, textural and microbiological effects. *International Journal of Food Science & Technology*, 50(6), 1452–1460. <http://doi.org/10.1111/ijfs.12777>
- García-Mora, P., Martín-Martínez, M., Angeles Bonache, M., González-Múniz, R., Peñas, E., Frias, J., & Martínez-Villaluenga, C. (2017). Identification, functional gastrointestinal stability and molecular docking studies of lentil peptides with dual antioxidant and angiotensin I converting enzyme inhibitory activities. *Food Chemistry*, 221, 464–472. <http://doi.org/10.1016/j.foodchem.2016.10.087>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExpASY Server. In *The Proteomics Protocols Handbook* (pp. 571–607). Totowa, NJ: Humana Press. <http://doi.org/10.1385/1-59259-890-0:571>
- Gawlik-Dziki, U., Dziki, D., Baraniak, B., & Lin, R. (2009). The effect of simulated digestion in vitro on bioactivity of wheat bread with Tartary buckwheat flavones addition. *LWT – Food Science and Technology*, 42(1), 137–143. <http://doi.org/10.1016/j.lwt.2008.06.009>
- Geerts, B. F., Van Dongen, M. G. J., Flaming, B., Moerland, M. M., Kam, M. L. de, Cohen, A. F., ... Burggraaf, J. (2011). Hydrolyzed casein decreases postprandial glucose concentrations in T2DM patients irrespective of leucine content. *Journal of Dietary Supplements*, 8(3), 280–292. <http://doi.org/10.3109/19390211.2011.593617>
- Gibbs, B. F., Zougman, A., Masse, R., & Mulligan, C. (2004). Production and characterization of bioactive peptides from soy hydrolysate and soy-fermented food. *Food Research International*, 37(2), 123–131. <http://doi.org/10.1016/j.foodres.2003.09.010>
- Gillman, M. W., Cupples, L. A., Gagnon, D., Posner, B. M., Ellison, R. C., Castelli, W. P., & Wolf, P. A. (1995). Protective Effect of Fruits and Vegetables on Development of Stroke in Men. *Journal of the American Medical Association*, 273, 1113–1117.
- Goedeke, L., & Fernández-Hernando, C. (2012). Regulation of cholesterol homeostasis. *Cellular and Molecular Life Sciences*, 69(6), 915–930. <http://doi.org/10.1007/s00018-011-0857-5>
- Goldstein, J. L., & Brown, M. S. (1990). Regulation of the mevalonate pathway. *Nature*, 343(6257), 425–430. <http://doi.org/10.1038/343425a0>
- Goldstein, J. L., & Brown, M. S. (2015). A century of cholesterol and coronaries: From plaques to genes to statins. *Cell*, 161(1), 161–172. <http://doi.org/10.1016/j.cell.2015.01.036>
- Gorinstein, S., Arnao de Nue, I., & Arruda, P. (1991). Alcohol-soluble and total proteins from amaranth seeds and their comparison with other cereals. *Journal of Agricultural and Food Chemistry*, 39(5), 848–850. <http://doi.org/10.1021/jf00005a005>
- Gorinstein, S., Zemser, M., Friedman, M., Rodrigues, W. A., Martins, P. S., Vello, N. A., ... Paredes-López, O. (1996). Physicochemical characterization of the structural stability of some plant globulins. *Food Chemistry*, 56(2), 131–138. [http://doi.org/10.1016/0308-8146\(95\)00144-1](http://doi.org/10.1016/0308-8146(95)00144-1)
- Gorinstein, S., Zemser, M., Fliess, A., Shnitman, I., Paredes-Lopez, O., Yamamoto, K., ... Taniguchi, H. (1998). Computational analysis of the amino acid residue sequences of amaranth and some other proteins. *Biosci Biotechnol Biochem*, 62(10), 1845–51. <http://doi.org/10.1271/bbb.62.1845>
- Greene, L. H., Lewis, T. E., Addou, S., Cuff, A., Dallman, T., Dibley, M., ... Orengo, C. A. (2007). The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic Acids Res*, 35(Database issue), D291–7. <http://doi.org/gkl959> 10.1093/nar/gkl959

- Haines, B. E., Steussy, C. N., Stauffacher, C. V., & Wiest, O. (2012). Molecular Modeling of the Reaction Pathway and Hydride Transfer Reactions of HMG-CoA Reductase. *Biochemistry*, 51(40), 7983–7995. <http://doi.org/10.1021/bi3008593>
- Halliwell, B. (2001). Role of Free Radicals in the Neurodegenerative Diseases. *Drugs & Aging*, 18(9), 685–716. <http://doi.org/10.2165/00002512-200118090-00004>
- He, R., Aluko, R. E., & Ju, X.-R. (2014). Evaluating Molecular Mechanism of Hypotensive Peptides Interactions with Renin and Angiotensin Converting Enzyme. *PLoS ONE*, 9(3), e91051. <http://doi.org/10.1371/journal.pone.0091051>
- Hernández-Ledesma, B., Dávalos, A., Bartolomé, B., & Amigo, L. (2005). Preparation of antioxidant enzymatic hydrolysates from α -lactalbumin and β -lactoglobulin. Identification of active peptides by HPLC-MS/MS. *Journal of Agricultural and Food Chemistry*, 53(3), 588–593. <http://doi.org/10.1021/jf048626m>
- Hernández-Ledesma, B., Quirós, A., Amigo, L., & Recio, I. (2007). Identification of bioactive peptides after digestion of human milk and infant formula with pepsin and pancreatin. *International Dairy Journal*, 17(1), 42–49. <http://doi.org/10.1016/j.idairyj.2005.12.012>
- Higgins, D. G., Bleasby, A. J., & Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *Computer Applications in the Biosciences: CABIOS*, 8(2), 189–91.
- Higgins, D., & Taylor, W. (Eds.). (2001). *Bioinformatics: Sequence, Structure and Databanks. A practical Approach*. Oxford, University Press. <http://doi.org/10.1093/bib/2.2.202>
- Hirota, T., Ohki, K., Kawagishi, R., Kajimoto, Y., Mizuno, S., Nakamura, Y., & Kitakaze, M. (2007). Casein hydrolysate containing the antihypertensive tripeptides Val-Pro-Pro and Ile-Pro-Pro improves vascular endothelial function independent of blood pressure-lowering effects: contribution of the inhibitory action of angiotensin-converting enzyme. *Hypertension Research: Official Journal of the Japanese Society of Hypertension*, 30(6), 489–496. <http://doi.org/10.1291/hyres.30.489>
- Holton, T. A., Vijayakumar, V., & Khaldi, N. (2013). Bioinformatics: Current perspectives and future directions for food and nutritional research facilitated by a Food-Wiki database. *Trends in Food Science and Technology*, 34(1), 5–17. <http://doi.org/10.1016/j.tifs.2013.08.009>
- Horton, J. D., Shah, N. a, Warrington, J. a, Anderson, N. N., Park, S. W., Brown, M. S., & Goldstein, J. L. (2003). Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. *Proceedings of the National Academy of Sciences of the United States of America*, 100(21), 12027–12032. <http://doi.org/10.1073/pnas.1534923100>
- Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics*, 26(5), 680–682. <http://doi.org/10.1093/bioinformatics/btq003>
- Huerta-Ocampo, J. Á., Barba de La Rosa, P. A., Angel Huerta-Ocampo, J., & Paulina Barba De La Rosa, A. (2011). Amaranth: A Pseudo-Cereal with Nutraceutical Properties. *Current Nutrition & Food Science*, 7(1), 1–9. <http://doi.org/10.2174/157340111794941076>
- Ikonen, E. (2006). Mechanisms for Cellular Cholesterol Transport: Defects and Human Disease. *Physiological Reviews*, 86(4), 1237–1261. <http://doi.org/10.1152/physrev.00022.2005>
- Illergård, K., Ardell, D. H., & Elofsson, A. (2009). Structure is three to ten times more conserved than sequence – A study of structural response in protein cores. *Proteins: Structure, Function and Bioinformatics*, 77(3), 499–508. <http://doi.org/10.1002/prot.22458>
- Istvan, E. S., Palnitkar, M., Buchanan, S. K., & Deisenhofer, J. (2000). Crystal structure of the catalytic portion of human HMG-CoA reductase: insights into regulation of activity and catalysis. *The EMBO Journal*, 19(5), 819–830. <http://doi.org/10.1093/emboj/19.5.819>

- Istvan, E. S. (2000). The structure of the catalytic portion of human HMG-CoA reductase. *Biochimica et Biophysica Acta (BBA) – Molecular and Cell Biology of Lipids*, 1529(1-3), 9-18. [http://doi.org/10.1016/S1388-1981\(00\)00134-7](http://doi.org/10.1016/S1388-1981(00)00134-7)
- Istvan, E. S. (2001a). Bacterial and mammalian HMG-CoA reductases: Related enzymes with distinct architectures. *Current Opinion in Structural Biology*, 11(6), 746-751. [http://doi.org/10.1016/S0959-440X\(01\)00276-7](http://doi.org/10.1016/S0959-440X(01)00276-7)
- Istvan, E. S. (2001b). Structural Mechanism for Statin Inhibition of HMG-CoA Reductase. *Science*, 292(5519), 1160-1164. <http://doi.org/10.1126/science.1059344>
- Istvan, E. S. (2003). Statin inhibition of HMG-CoA reductase: A 3-dimensional view. In *Atherosclerosis Supplements* (Vol. 4, pp. 3-8). [http://doi.org/10.1016/S1567-5688\(03\)00003-5](http://doi.org/10.1016/S1567-5688(03)00003-5)
- Iwaniak, A., Minkiewicz, P., Darewicz, M., Protasiewicz, M., & Mogut, D. (2015). Chemometrics and cheminformatics in the analysis of biologically active peptides from food sources. *Journal of Functional Foods*, 16, 334-351. <http://doi.org/10.1016/j.jff.2015.04.038>
- Jayaprakasam, B., Zhang, Y., & Nair, M. G. (2004). Tumor Cell Proliferation and Cyclooxygenase Enzyme Inhibitory Compounds in *Amaranthus tricolor*. *Journal of Agricultural and Food Chemistry*, 52(23), 6939-6943. <http://doi.org/10.1021/jf048836z>
- Jimsheena, V. K., & Gowda, L. R. (2010). Arachin derived peptides as selective angiotensin I-converting enzyme (ACE) inhibitors: Structure-activity relationship. *Peptides*, 31(6), 1165-1176. <http://doi.org/10.1016/j.peptides.2010.02.022>
- Jo, Y., & DeBose-Boyd, R. A. (2010). Control of cholesterol synthesis through regulated ER-associated degradation of HMG CoA reductase. *Critical Reviews in Biochemistry and Molecular Biology*, 45(3), 185-98. <http://doi.org/10.3109/10409238.2010.485605>
- Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*, 8(3), 275-282.
- Kigel, J. (1994) Development and ecophysiology of Amaranths. *Amaranth biology, chemistry and technology*. Ed. Paredes-López. Boca Raton Florida. CRC Press. p:39-73.
- Kimura, M., & Ota T. (1974). On Some Principles Governing Molecular Evolution. *Proceedings of the National Academy of Sciences of the United States of America* 71 (7): 2848-52. <http://doi.org/10.1073/pnas.71.7.2848>.
- Konishi, Y., Horikawa, K., Oku, Y., Azumaya, J., & Nakatani, N. (1991). Extraction of two albumin fractions from amaranth grains : comparison of some physicochemical properties and the putative localization in the grains growing c4 plant that produces high yields of amaranth grains . A large amount of proteins method. *Agric. Biol. Chem.*, 55(II), 2745-2750.
- Korhonen, H., & Pihlanto, A. (2006). Bioactive peptides: Production and functionality. *International Dairy Journal*, 16(9), 945-960. <http://doi.org/10.1016/j.idairyj.2005.10.012>
- Kwak, J. H., Kim, M., Lee, E., Lee, S. H., Ahn, C. W., & Lee, J. H. (2013). Effects of black soy peptide supplementation on blood pressure and oxidative stress: A randomized controlled trial. *Hypertension Research Clinical & Experimental*, 36(12), 1060-1066. <http://doi.org/http://dx.doi.org/10.1038/hr.2013.79>
- Lado, M. B., Burini, J., Rinaldi, G., Añón, M. C., & Tironi, V. A. (2015). Effects of the Dietary Addition of Amaranth (*Amaranthus mantegazzianus*) Protein Isolate on Antioxidant Status, Lipid Profiles and Blood Pressure of Rats. *Plant Foods for Human Nutrition*, 70(4), 371-379. <http://doi.org/10.1007/s11130-015-0516-3>

- Lamothe, L. M., Srichuwong, S., Reuhs, B. L., & Hamaker, B. R. (2015). Quinoa (*Chenopodium quinoa* W.) and amaranth (*Amaranthus caudatus* L.) provide dietary fibres high in pectic substances and xyloglucans. *Food Chemistry*, 167, 490–496. <http://doi.org/10.1016/j.foodchem.2014.07.022>
- Lannoo, N., & Van Damme, E. J. (2014). Lectin domains at the frontiers of plant defense. *Front Plant Sci*, 5, 397. <http://doi.org/10.3389/fpls.2014.00397>
- Lensing, A. W., Prandoni, P., Prins, M. H., & Büller, H. (1999). Deep-vein thrombosis. *The Lancet*, 353(9151), 479–485. [http://doi.org/10.1016/S0140-6736\(98\)04298-6](http://doi.org/10.1016/S0140-6736(98)04298-6)
- Li, B., Chen, F., Wang, X., Ji, B., & Wu, Y. (2007). Isolation and identification of antioxidative peptides from porcine collagen hydrolysate by consecutive chromatography and electrospray ionization–mass spectrometry. *Food Chemistry*, 102(4), 1135–1143. <http://doi.org/10.1016/j.foodchem.2006.07.002>
- Li, H., & Aluko, R. E. (2010). Identification and Inhibitory Properties of Multifunctional Peptides from Pea Protein Hydrolysate. *Journal of Agricultural and Food Chemistry*, 58(21), 11471–11476. <http://doi.org/10.1021/jf102538g>
- Li, Y.-W., & Li, B. (2013). Characterization of structure–antioxidant activity relationship of peptides in free radical systems using QSAR models: Key sequence positions and their amino acid properties. *Journal of Theoretical Biology*, 318, 29–43. <http://doi.org/10.1016/j.jtbi.2012.10.029>
- Li-Chan, E. C. Y. (2015). Bioactive peptides and protein hydrolysates: Research trends and challenges for application as nutraceuticals and functional food ingredients. *Current Opinion in Food Science*, 1(1), 28–37. <http://doi.org/10.1016/j.cofs.2014.09.005>
- Lin, B.-F., Chiang, B.-L., & Lin, J.-Y. (2005). *Amaranthus spinosus* water extract directly stimulates proliferation of B lymphocytes in vitro. *International Immunopharmacology*, 5(4), 711–722. <http://doi.org/10.1016/j.intimp.2004.12.001>
- Lin, S. H., Chang, D. K., Chou, M. J., Huang, K. J., & Shiuan, D. (2015). Peptide inhibitors of human HMG–CoA reductase as potential hypocholesterolemia agents. *Biochemical and Biophysical Research Communications*, 456(1), 104–109. <http://doi.org/10.1016/j.bbrc.2014.11.042>
- Loewus, F. (1951). Improvement in anthrone method for determination of carbohydrates. *Anal. Chem.*, 24, 219.
- London, N., Raveh, B., & Schueler–Furman, O. (2013). Peptide docking and structure–based characterization of peptide binding: from knowledge to know–how. *Current Opinion in Structural Biology*, 23(6), 894–902. <http://doi.org/10.1016/j.sbi.2013.07.006>
- Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. *The Journal of Biological Chemistry*, 193(1), 265–75.
- Lozoya–Gloria E. (1994) Biotechnology for an ancient crop: amaranth. *Amaranth: Biology, Chemistry and Technology*, Ed. Paredes–López. Boca Raton FL. CRC Press. p:1–7.
- Lu, S., Deng, P., Liu, X., Luo, J., Han, R., Gu, X., ... Pongor, S. (1999). Solution structure of the major α -amylase inhibitor of the crop plant amaranth. *Journal of Biological Chemistry*, 274(29), 20473–20478. <http://doi.org/10.1074/jbc.274.29.20473>
- Luna–Suárez, S., Medina–Godoy, S., Cruz–Hernández, A., & Paredes–López, O. (2010). Modification of the amaranth 11S globulin storage protein to produce an inhibitory peptide of the angiotensin I converting enzyme, and its expression in *Escherichia coli*. *Journal of Biotechnology*, 148(4), 240–247. <http://doi.org/10.1016/j.jbiotec.2010.06.009>
- Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). Review: What is bioinformatics? An introduction and overview. *Yearbook of Medical Informatics*, 83–100.

- Madureira, A. R., Pereira, C. I., Gomes, A. M. P., Pintado, M. E., & Malcata, F. X. (2007). Bovine whey proteins—Overview on their main biological properties. *Food Research International* (1984), 1197–1211.
- Maguid, S., Fernández-Alberti, S., Parisi, G., & Echave, J. (2006). Evolutionary conservation of protein backbone flexibility. *Journal of Molecular Evolution*, 63(4), 448–57. <http://doi.org/10.1007/s00239-005-0209-x>
- Majumder, K., & Wu, J. (2010). A new approach for identification of novel antihypertensive peptides from egg proteins by QSAR and bioinformatics. *Food Research International*, 43(5), 1371–1378. <http://doi.org/10.1016/j.foodres.2010.04.027>
- Majumder, K., & Wu, J. (2014). Molecular targets of antihypertensive peptides: Understanding the mechanisms of action based on the pathophysiology of hypertension. *International Journal of Molecular Sciences*, 16(1), 256–283. <http://doi.org/10.3390/ijms16010256>
- Maldonado-Cervantes, E., Jeong, H. J., León-Galván, F., Barrera-Pacheco, A., De León-Rodríguez, A., González De Mejía, E., ... Barba De La Rosa, A. P. (2010). Amaranth lunasin-like peptide internalizes into the cell nucleus and inhibits chemical carcinogen-induced transformation of NIH-3T3 cells. *Peptides*, 31(9), 1635–1642. <http://doi.org/10.1016/j.peptides.2010.06.014>
- Manning, J. C., Romero, A., Habermann, F., García-Caballero, G., Kaltner, H., Roth, J., & Gabius, H.-J. (2017). Lectins: a primer for histochemists and cell biologists. *Histochemistry & Cell Biology*, 147(2), 199–222. <http://doi.org/10.1007/s00418-016-1524-6>
- Marcone, M. F., Kakuda, Y., & Yada, R. Y. (1998). Salt soluble seed globulins of dicotyledonous and monocotyledonous plants II. Structural characterization. *Food Chemistry*, 63(2), 265–274. [http://doi.org/10.1016/S0308-8146\(97\)00159-3](http://doi.org/10.1016/S0308-8146(97)00159-3)
- Marcone, M. F., & Kakuda, Y. (1999). A comparative study of the functional properties of amaranth and soybean globulin isolates. *Nahrung – Food*, 43(6), 368–373. [http://doi.org/10.1002/\(SICI\)1521-3803\(19991201\)43:6<368::AID-FOOD368>3.0.CO;2-R](http://doi.org/10.1002/(SICI)1521-3803(19991201)43:6<368::AID-FOOD368>3.0.CO;2-R)
- Marcone, M. F. (2000). First report of the characterization of the threatened plant species *Amaranthus pumilus* (seabeach amaranth). *Journal of Agricultural and Food Chemistry*, 48(2), 378–382. <http://doi.org/10.1021/jf991133a>
- Marino, F., Pelc, L. A., Vogt, A., Gandhi, P. S., & Di Cera, E. (2010). Engineering thrombin for selective specificity toward protein C and PAR1. *Journal of Biological Chemistry*, 285(25), 19145–19152. <http://doi.org/10.1074/jbc.M110.119875>
- Martin, A. C. R., & Porter, C. T. (n.d.). ProFit.
- Martínez, E. N., & Añón, M. C. (1996a). Composition and structural characterization of amaranth proteins isolates. An electrophoretic and calorimetric study. *J. Agric. Food Chem.*, 44(9), 2523–2530. <http://doi.org/10.1021/jf960169p>
- Martínez, E. N., & Añón, M. C. (1996b). Composition and structural characterization of amaranth proteins isolates. An electrophoretic and calorimetric study. *J. Agric. Food Chem.*, 44(9), 2523–2530. <http://doi.org/10.1021/jf960169p>
- Martínez, E. N., Castellani, O. F., & Añón, M. C. (1997). Common Molecular Features among Amaranth Storage Proteins. *Scanning*, 3832–3839. <http://doi.org/10.1021/JF9700384>
- Martínez-Maqueda, D., Miralles, B., Recio, I., & Hernández-Ledesma, B. (2012). Antihypertensive peptides from food proteins: a review. *Food & Function*, 3(4), 350–61. <http://doi.org/10.1039/c2fo10192k>

- Martinez-Villaluenga, C., Rupasinghe, S. G., Schuler, M. A., & Gonzalez De Mejia, E. (2010). Peptides from purified soybean β -conglycinin inhibit fatty acid synthase by interaction with the thioesterase catalytic domain. *FEBS Journal*, 277(6), 1481–1493. <http://doi.org/10.1111/j.1742-4658.2010.07577.x>
- Martirosyan, D. M., & Singh, J. (2015). A new definition of functional food by FFC: what makes a new definition unique? *Functional Foods in Health and Disease*, 5(6), 209–223.
- Maupetit, J., Derreumaux, P., & Tuffery, P. (2009). PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Research*, 37 (Web Server), W498–W503. <http://doi.org/10.1093/nar/gkp323>
- McLachlan, A. D. (1982). Rapid comparison of protein structures. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*. International Union of Crystallography.
- Mengoni, A. (2015). *Lectina de Amaranto: un potencial compuesto antiproliferativo*. Tesis de Grado. Universidad Nacional de la Plata.
- Minikus, M., Alminger, M., Alvito, P., Ballance, S., Bohn, T., Bourlieu, C., ... Brodtkorb, A. (2014). A standardised static in vitro digestion method suitable for food – an international consensus. *Food & Function*, 5(6), 1113. <http://doi.org/10.1039/c3fo60702j>
- Minkiewicz, P., Dziuba, J., Iwaniak, A., Dziuba, M., & Darewicz, M. (2008). BIOPEP Database and Other Programs for Processing Bioactive Peptide Sequences. *Journal of AOAC International*, 91(4), 965.
- Möller, N. P., Scholz-Ahrens, K. E., Roos, N., & Schrezenmeir, J. (2008). Bioactive peptides and proteins from foods: indication for health effects. *European Journal of Nutrition*, 47(4), 171–182. <http://doi.org/10.1007/s00394-008-0710-2>
- Montoya-Rodríguez, A., Gómez-Favela, M. A., Reyes-Moreno, C., Milán-Carrillo, J., & González de Mejía, E. (2015). Identification of Bioactive Peptide Sequences from Amaranth (*Amaranthus hypochondriacus*) Seed Proteins and Their Potential Role in the Prevention of Chronic Diseases. *Comprehensive Reviews in Food Science and Food Safety*, 14(2), 139–158. <http://doi.org/10.1111/1541-4337.12125>
- Montoya-Rodríguez, A., Milán-Carrillo, J., Díaz, V. P., Reyes-Moreno, C., & González de Mejía, E. (2014). Pepsin-pancreatin protein hydrolysates from extruded amaranth inhibit markers of atherosclerosis in LPS-induced THP-1 macrophages-like human cells by reducing expression of proteins in LOX-1 signaling pathway. *Proteome Science*, 12(1), 30. <http://doi.org/10.1186/1477-5956-12-30>
- Moronta, J., Smaldini, P. L., Docena, G. H., & Añón, M. C. (2016a). Peptides of amaranth were targeted as containing sequences with potential anti-inflammatory properties. *Journal of Functional Foods*, 21, 463–473. <http://doi.org/10.1016/j.jff.2015.12.022>
- Moronta, J., Smaldini, P. L., Fossati, C. A., Añón, M. C., & Docena, G. H. (2016b). The anti-inflammatory SSEDIKE peptide from Amaranth seeds modulates IgE-mediated food allergy. *Journal of Functional Foods*, 25, 579–587. <http://doi.org/10.1016/j.jff.2016.06.031>
- Motamed, M., Zhang, Y., Wang, M. L., Seemann, J., Kwon, H. J., Goldstein, J. L., & Brown, M. S. (2011). Identification of luminal loop 1 of scap protein as the sterol sensor that maintains cholesterol homeostasis. *Journal of Biological Chemistry*, 286(20), 18002–18012. <http://doi.org/10.1074/jbc.M111.238311>
- Mulero Cánovas, J., Zafrilla Rentero, P., Martínez-Cachá Martínez, A., Leal Hernández, M., & Abellán Alemán, J. (2011). Péptidos bioactivos. *Clínica E Investigación En Arteriosclerosis*, 23(5), 219–227. <http://doi.org/10.1016/j.arteri.2011.04.004>
- Nagpal, R., Behare, P., Rana, R., Kumar, A., Kumar, M., Arora, S., ... Yadav, H. (2011). Bioactive peptides derived from milk proteins and their health beneficial potentials: an update. *Food & Function*, 2(1), 18–27. <http://doi.org/10.1039/c0fo00016g>

- Nakano, Y., Fujitani, K., Kurihara, J., Ragan, J., Usui-Aoki, K., Shimoda, L., ... Yamamoto, D. (2001). Mutations in the novel membrane protein spinster interfere with programmed cell death and cause neural degeneration in *Drosophila melanogaster*. *Molecular and Cellular Biology*, 21(11), 3775–3788. <http://doi.org/10.1128/MCB.21.11.3775-3788.2001>
- Ndiaye, F., Vuong, T., Duarte, J., Aluko, R. E., & Matar, C. (2012). Anti-oxidant, anti-inflammatory and immunomodulating properties of an enzymatic protein hydrolysate from yellow field pea seeds. *European Journal of Nutrition*, 51(1), 29–37. <http://doi.org/10.1007/s00394-011-0186-3>
- Nielsen, P. M., Petersen, D., & Dambmann, C. (2001). Improved Method for Determining Food Protein Degree of Hydrolysis. *Journal of Food Science*, 66(5), 642–646. <http://doi.org/10.1111/j.1365-2621.2001.tb04614.x>
- Nkonge, C., & Ballance, M. (1982). Sensitive colorimetric procedure for nitrogen determination in micro-Kjeldahl digests. *J. Agric. Food Chem*, 30, 416–420.
- Nongonierma, A. B., & Fitzgerald, R. J. (2014). An *in silico* model to predict the potential of dietary proteins as sources of dipeptidyl peptidase IV (DPP-IV) inhibitory peptides. *Food Chemistry*, 165, 489–498. <http://doi.org/10.1016/j.foodchem.2014.05.090>
- Nongonierma, A. B., & FitzGerald, R. J. (2015). The scientific evidence for the role of milk protein-derived bioactive peptides in humans: A Review. *Journal of Functional Foods*, 17, 640–656. <http://doi.org/10.1016/j.jff.2015.06.021>
- Nongonierma, A. B., Mooney, C., Shields, D. C., & Fitzgerald, R. J. (2014). In silico approaches to predict the potential of milk protein-derived peptides as dipeptidyl peptidase IV (DPP-IV) inhibitors. *Peptides*, 57, 43–51. <http://doi.org/10.1016/j.peptides.2014.04.018>
- Norris, R., Casey, F., FitzGerald, R. J., Shields, D., & Mooney, C. (2012). Predictive modelling of angiotensin converting enzyme inhibitory dipeptides. *Food Chemistry*, 133(4), 1349–1354. <http://doi.org/10.1016/j.foodchem.2012.02.023>
- Oliveira, E. F., Cerqueira, N. M. F. S. A., Ramos, M. J., & Fernandes, P. A. (2016). QM/MM study of the mechanism of reduction of 3-hydroxy-3-methylglutaryl coenzyme A catalyzed by human HMG-CoA reductase. *Catal. Sci. Technol.*, 6(19), 7172–7185. <http://doi.org/10.1039/C6CY00356G>
- Orengo, C. A., Jones, D. T., & Thornton, J. M. (1994). Protein superfamilies and domain superfolds. *Nature*, 372(6507), 631–634. <http://doi.org/10.1038/372631a0>
- Orengo, C. A., Pearl, F. M. G., & Thornton, J. M. (2003). The CATH domain structure database., *Methods of biochemical analysis*, 44, 249–271, <http://doi.org/10.1002/0471721204.ch13>
- Orsini Delgado, M. C., Tironi, V. A., & Añón, M. C. (2011). Antioxidant activity of amaranth protein or their hydrolysates under simulated gastrointestinal digestion. *LWT – Food Science and Technology*, 44(8), 1752–1760. <http://doi.org/10.1016/j.lwt.2011.04.002>
- Orsini Delgado, M. C., Galleano, M., Añón, M. C., & Tironi, V. A. (2015). Amaranth Peptides from Simulated Gastrointestinal Digestion: Antioxidant Activity Against Reactive Species. *Plant Foods for Human Nutrition*, 70(1), 27–34. <http://doi.org/10.1007/s11130-014-0457-2>
- Orsini Delgado, M. C., Nardo, A., Pavlovic, M., Rogniaux, H., Añón, M. C., & Tironi, V. A. (2016). Identification and characterization of antioxidant peptides obtained by gastrointestinal digestion of amaranth proteins. *Food Chemistry*, 197, 1160–1167. <http://doi.org/10.1016/j.foodchem.2015.11.092>
- Osborne, T. B. (1924). *The vegetable proteins*. *Journal of the Society of Chemical Industry* (2nd editio, Vol. 43). London: John Wiley & Sons, Ltd. <http://doi.org/10.1002/jctb.5000431704>

- Pak, V. V., Koo, M., Lee, N., Kim, M. S., & Kwon, D. Y. (2005). Structure—Activity Relationships of the Peptide Ile–Ala–Val–Pro and Its Derivatives Revealed Using the Semi–Empirical AM1 Method. *Chemistry of Natural Compounds*, 41(4), 454–460. <http://doi.org/10.1007/s10600-005-0176-x>
- Pak, V. V., Koo, M., Yun, L., & Kwon, D. Y. (2007). Recognized sequence and conformation in design of linear peptides as a competitive inhibitor for HMG–CoA reductase. *Journal of Molecular Recognition*, 20(3), 197–203. <http://doi.org/10.1002/jmr.824>
- Pak, V. V., Koo, M., Kim, M. J., Yang, H. J., Yun, L., & Kwon, D. Y. (2008). Modeling an active conformation for linear peptides and design of a competitive inhibitor for HMG–CoA reductase. *Journal of Molecular Recognition*, 21(4), 224–232. <http://doi.org/10.1002/jmr.889>
- Pak, V. V., Koo, M., Kwon, D. Y., & Yun, L. (2012). Design of a highly potent inhibitory peptide acting as a competitive inhibitor of HMG–CoA reductase. *Amino Acids*, 43(5), 2015–2025. <http://doi.org/10.1007/s00726-012-1276-0>
- Pan, D., Cao, J., Guo, H., & Zhao, B. (2012). Studies on purification and the molecular mechanism of a novel ACE inhibitory peptide from whey protein hydrolysate. *Food Chemistry*, 130(1), 121–126. <http://doi.org/10.1016/j.foodchem.2011.07.011>
- Panda, T., Basak, T., Saraswathi, G., & Théodore, T. (2011). Kinetic Mechanisms of Cholesterol Synthesis: A Review. *Industrial & Engineering Chemistry Research*, 50(23), 12847–12864. <http://doi.org/10.1021/ie200073a>
- Paredes–López, O. (1994). *Amaranth: Biology, Chemistry and Technology*. CRC Press.
- Peña–Ramos, E. A., Xiong, Y. L., & Arteaga, G. E. (2004). Fractionation and characterization for antioxidant activity of hydrolysed whey protein. *Journal of the Science of Food and Agriculture*, 84(14), 1908–1918. <http://doi.org/10.1002/jsfa.1886>
- Phelan, M., Aherne, A., FitzGerald, R. J., & O'Brien, N. M. (2009, November). Casein–derived bioactive peptides: Biological effects, industrial uses, safety aspects and regulatory status. *International Dairy Journal*. <http://doi.org/10.1016/j.idairyj.2009.06.001>
- Pihlanto, A. (2006). Antioxidative peptides derived from milk proteins. *International Dairy Journal*, 16(11), 1306–1314. <http://doi.org/10.1016/j.idairyj.2006.06.005>
- Pimenta, D. C., & Lebrun, I. (2007). Cryptides: Buried secrets in proteins. *Peptides*, 28(12), 2403–2410. <http://doi.org/10.1016/j.peptides.2007.10.005>
- Polo, M. P., & de Bravo, M. G. (2006). Effect of geraniol on fatty–acid and mevalonate metabolism in the human hepatoma cell line Hep G2. *Biochemistry and Cell Biology*, 84(1), 102–111. <http://doi.org/10.1139/o05-160>
- Prupp, A. H., Isaksson, T., Stepaniak, L., Sorhaug, T., & Ardo, Y. (2005). Quantitative structure activity relationship modelling of peptides and proteins as a tool in food science. *Trends in Food Science & Technology*, 16(11), 484–494. <http://doi.org/10.1016/j.tifs.2005.07.003>
- Puchalska, P., Marina Alegre, M. L., & Garcia Lopez, M. C. (2015). Isolation and characterization of peptides with antihypertensive activity in foodstuffs. *Critical Review Food Science Nutrition*, 55(4), 521–551. <http://doi.org/10.1080/10408398.2012.664829>
- Quiroga, A. V., Martínez, E. N., & Añón, M. C. (2007). Amaranth Globulin Polypeptide Heterogeneity. *The Protein Journal*, 26(5), 327–333. <http://doi.org/10.1007/s10930-007-9075-2>
- Quiroga, A., Martínez, E. N., Rogniaux, H., Geairon, A., & Añón, M. C. (2010). Amaranth (*Amaranthus hypochondriacus*) vicilin subunit structure. *Journal of Agricultural and Food Chemistry*. <http://doi.org/10.1021/jf103296n>

- Quiroga, A. V., Aphalo, P., Ventureira, J. L., Martínez, E. N., & Añón, M. C. (2012). Physicochemical, functional and angiotensin converting enzyme inhibitory properties of amaranth (*Amaranthus hypochondriacus*) 7S globulin. *Journal of the Science of Food and Agriculture*, 92(2), 397-403. <http://doi.org/10.1002/jsfa.4590>
- Quiroga, A. V., Barrio, D. A., & Añón, M. C. (2015). Amaranth lectin presents potential antitumor properties. *LWT – Food Science and Technology*, 60(1), 478-485. <http://doi.org/10.1016/j.lwt.2014.07.035>
- Quiroga, A., Aphalo, P. y Añón, M.C. (2017). Antihypertensive properties of Amaranth (*Amaranthus Hypochondriacus*) protein-derived peptides. Alternative blood pressure regulating mechanisms different from ACE inhibition. *Enviado marzo 2017*.
- Rawlings, N. D., Waller, M., Barrett, A. J., & Bateman, A. (2014). MEROPS : the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 42(D1), D503-D509. <http://doi.org/10.1093/nar/gkt953>
- Repo-Carrasco-Valencia, R., Hellström, J. K., Pihlava, J.-M., & Mattila, P. H. (2010). Flavonoids and other phenolic compounds in Andean indigenous grains: Quinoa (*Chenopodium quinoa*), kañiwa (*Chenopodium pallidicaule*) and kiwicha (*Amaranthus caudatus*). *Food Chemistry*, 120(1), 128-133. <http://doi.org/10.1016/j.foodchem.2009.09.087>
- Rodas, B., & Bressani, R. (2009). [The oil, fatty acid and squalene content of varieties of raw and processed amaranth grain]. *Archivos Latinoamericanos de Nutricion*, 59(1), 82-7.
- Ros, E. (2003). Inhibición de la absorción intestinal del colesterol: nueva diana terapéutica en la reducción de la colesterolemia. *Clínica E Investigación En Arteriosclerosis*, 15(6), 261-275. [http://doi.org/10.1016/S0214-9168\(03\)78938-4](http://doi.org/10.1016/S0214-9168(03)78938-4)
- Rosenfeld, R., Vajda, S., & DeLisi, C. (1995). Flexible Docking and Design. *Annual Review of Biophysics and Biomolecular Structure*, 24(1), 677-700. <http://doi.org/10.1146/annurev.bb.24.060195.003333>
- Rummery, N. M., Grayson, T. H., & Hill, C. E. (2005). Angiotensin-converting enzyme inhibition restores endothelial but not medial connexin expression in hypertensive rats. *Journal of Hypertension*, 23(2), 317-28. <http://doi.org/10.1097/00004872-200502000-00014>
- Sabbione, A. C. (2015). Actividad antitrombótica de proteínas de amaranto. Tesis Doctoral. Universidad Nacional de la Plata.
- Sabbione, A. C., Scilingo, A., & Añón, M. C. (2015). Potential antithrombotic activity detected in amaranth proteins and its hydrolysates. *LWT – Food Science and Technology*, 60(1), 171-177. <http://doi.org/10.1016/j.lwt.2014.07.015>
- Sabbione, A. C., Ibañez, S. M., Martínez, E. N., Añón, M. C., & Scilingo, A. A. (2016 a). Antithrombotic and Antioxidant Activity of Amaranth Hydrolysate Obtained by Activation of an Endogenous Protease. *Plant Foods for Human Nutrition*, 71(2), 174-182. <http://doi.org/10.1007/s11130-016-0540-y>
- Sabbione, A. C., Nardo, A. E., Añón, M. C., & Scilingo, A. (2016b). Amaranth peptides with antithrombotic activity released by simulated gastrointestinal digestion. *Journal of Functional Foods*, 20, 204-214. <http://doi.org/10.1016/j.jff.2015.10.015>
- Sabbione, A. C., Rinaldi, G., Añón, M. C., & Scilingo, A. A. (2016c). Antithrombotic Effects of *Amaranthus hypochondriacus* Proteins in Rats. *Plant Foods for Human Nutrition*, 71(1), 19-27. <http://doi.org/10.1007/s11130-015-0517-2>
- Salnikow, J., Zemser, M., Gorinstein, S., Friedman, M., Scheler, C., & Paredes-Lopez, O. (1998). Structure-Function Relationships of Folded and Unfolded Amaranth Proteins. *Protein Journal*, 17(6).
- Sato, R., Goldstein, J. L., & Brown, M. S. (1993). Replacement of serine-871 of hamster 3-hydroxy-3-methylglutaryl-CoA reductase prevents phosphorylation by AMP-activated kinase and blocks inhibition of

- sterol synthesis induced by ATP depletion. *Proceedings of the National Academy of Sciences of the United States of America*, 90(20), 9261–5. <http://doi.org/10.1073/pnas.90.20.9261>
- Schägger, H. (2006). Tricine-SDS-PAGE. *Nature Protocols*, 1(1), 16–22. <http://doi.org/10.1038/nprot.2006.4>
- Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F., & Schomburg, D. (2002). BRENDA: A resource for enzyme data and metabolic information. *Trends in Biochemical Sciences*, 103, 439–4. [http://doi.org/10.1016/S0968-0004\(01\)02027-8](http://doi.org/10.1016/S0968-0004(01)02027-8)
- Schrezenmeir, J., Korhonen, H., Williams, C., Gill, H. S., & Shah, N. (2000). Foreword. *British Journal of Nutrition*, 84(SUPPL. 1).
- Scilingo, A. A., Molina Ortiz, S. E., Martínez, E. N., & Añón, M. C. (2002). Amaranth protein isolates modified by hydrolytic and thermal treatments. Relationship between structure and solubility. *Food Research International*, 35(9), 855–862. [http://doi.org/10.1016/S0963-9969\(02\)00089-3](http://doi.org/10.1016/S0963-9969(02)00089-3)
- Segura-Nieto, M., Barba de la Rosa, A., & Paredes-López, O. (1994). Biochemistry of amaranth proteins. In O. Paredes-López (Ed.), *Amaranth. Biology, chemistry and technology* (pp. 129–185). Boca Raton, FL: CRC Press.
- Segura-Nieto, M., Vazquez-Sanchez, N., Rubio-Velazquez, H., Olguin-Martinez, L. E., Rodriguez-Nester, C. E., & Herrera-Estrella, L. (1992). Characterization of amaranth (*Amaranthus hypochondriacus* L.) seed proteins. *Journal of Agricultural and Food Chemistry*, 40(9), 1553–1558. <http://doi.org/10.1021/jf00021a016>
- Shahidi, F., & Zhong, Y. (2008). Bioactive peptides. *J AOAC Int*, 91(4), 914–931.
- Shtatland, T., Guettler, D., Kossodo, M., Pivovarov, M., & Weissleder, R. (2007). PepBank—a database of peptides based on sequence text mining and public peptide data sources. *BMC Bioinformatics*, 8, 280. <http://doi.org/10.1186/1471-2105-8-280>
- Shtatland, T., Guettler, D., Kossodo, M., Pivovarov, M., & Weissleder, R. (2007). PepBank—a database of peptides based on sequence text mining and public peptide data sources. *BMC Bioinformatics*, 8, 280. <http://doi.org/10.1186/1471-2105-8-280>
- Silva-Sánchez, C., Barba de la Rosa, A. P., León-Galván, M. F., de Lumen, B. O., de León-Rodríguez, A., & de Mejía, E. G. (2008). Bioactive Peptides in Amaranth (*Amaranthus hypochondriacus*) Seed. *Journal of Agricultural and Food Chemistry*, 56(4), 1233–1240. <http://doi.org/10.1021/jf072911z>
- Sipola, M., Finckenberg, P., Korpela, R., Vapaatalo, H., & Nurminen, M.-L. (2002). Effect of long-term intake of milk products on blood pressure in hypertensive rats. *Journal of Dairy Research*, 69(1), 103–111. <http://doi.org/10.1017/S002202990100526X>
- Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., ... Klenk, D. C. (1985). Measurement of protein using bicinchoninic acid. *Analytical Biochemistry*, 150(1), 76–85. [http://doi.org/10.1016/0003-2697\(85\)90442-7](http://doi.org/10.1016/0003-2697(85)90442-7)
- Soares, R., Mendonça, S., de Castro, L. Í., Menezes, A., & Arêas, J. G. (2015). Major Peptides from Amaranth (*Amaranthus cruentus*) Protein Inhibit HMG-CoA Reductase Activity. *International Journal of Molecular Sciences*, 16(2), 4150–4160. <http://doi.org/10.3390/ijms16024150>
- Spanova, M., & Daum, G. (2011, November). Squalene – biochemistry, molecular biology, process biotechnology, and applications. *European Journal of Lipid Science and Technology*. <http://doi.org/10.1002/ejlt.201100203>
- Steinbrecher, U. P., Zhang, H. F., & Loughheed, M. (1990). Role of oxidatively modified LDL in atherosclerosis. *Free Radical Biology & Medicine*, 9(2), 155–68.
- Tabernero, L., Bochar, D. a, Rodwell, V. W., & Stauffacher, C. V. (1999). Substrate-induced closure of the flap domain in the ternary complex structures provides insights into the mechanism of catalysis by 3-hydroxy-3-

- methylglutaryl-CoA reductase. *Proceedings of the National Academy of Sciences of the United States of America*, 96(13), 7167–71. <http://doi.org/10.1073/pnas.96.13.7167>
- Tamir, S., Bell, J., Finlay, T. H., Sakal, E., Smirnoff, P., Gaur, S., & Birk, Y. (1996). Isolation, characterization, and properties of a trypsin–chymotrypsin inhibitor from amaranth seeds. *Journal of Protein Chemistry*, 15(2), 219–229. <http://doi.org/10.1007/BF01887402>
- Tandang–Silvas, M., Carrazco–Pena, L., Barba de la Rosa, A. P., Osuna–Castro, J., Utsumi, S., Mikami, B., & Maruyama, N. (n.d.). Structure of amaranth 11S proglubulin, a major seed storage protein from *Amaranthus hypochondriacus* L. To Be Published. <http://doi.org/10.2210/PDB3QAC/PDB>
- Tandang–Silvas, M. R., Cabanos, C. S., Carrazco Peña, L. D., De La Rosa, A. P. B., Osuna–Castro, J. A., Utsumi, S., ... Maruyama, N. (2012). Crystal structure of a major seed storage protein, 11S proglubulin, from *Amaranthus hypochondriacus*: Insight into its physico–chemical properties. *Food Chemistry*, 135(2), 819–826. <http://doi.org/10.1016/j.foodchem.2012.04.135>
- Tiengo, A., Faria, M., & Netto, F. M. (2009). Characterization and ACE–inhibitory activity of Amaranth proteins. *Journal of Food Science*, 74(5), H121–H126. <http://doi.org/10.1111/j.1750-3841.2009.01145.x>
- Tironi, V. A., & Añón, M. C. (2014). Antioxidant activity of amaranth protein hydrolysate against thermal oxidation of vegetable oils. *JAOCS, Journal of the American Oil Chemists' Society*, 91(9), 1583–1594. <http://doi.org/10.1007/s11746-014-2509-z>
- Todeschini, R., & Consonni, V. (2008). Frontmatter. In *Handbook of Molecular Descriptors* (pp. i–xxi). Wiley-VCH Verlag GmbH. <http://doi.org/10.1002/9783527613106.fmatter>
- Tovar–Pérez, E. G., Guerrero–Legarreta, I., Farrés–González, A., & Soriano–Santos, J. (2009). Angiotensin I–converting enzyme–inhibitory peptide fractions from albumin and globulin as obtained of amaranth grain. *Food Chemistry*, 116(2), 437–444. <http://doi.org/10.1016/j.foodchem.2009.02.062>
- Transue, T. R., Smith, A. K., Mo, H., Goldstein, I. J., & Saper, M. A. (1997). Structure of benzyl T–antigen disaccharide bound to *Amaranthus caudatus* agglutinin. *Nature Structural Biology*, 4(10), 779–83. <http://doi.org/10.1038/nsb1097-779>
- Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics*, 29(6–7), 476–488. <http://doi.org/10.1002/minf.201000061>
- Tukey, J. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5(2), 99–114.
- Tyagi, A., Tuknait, A., Anand, P., Gupta, S., Sharma, M., Mathur, D., ... Raghava, G. P. S. (2015). CancerPPD: A database of anticancer peptides and proteins. *Nucleic Acids Research*, 43(D1), D837–D843. <http://doi.org/10.1093/nar/gku892>
- Udenigwe, C. C., Lu, Y. L., Han, C. H., Hou, W. C., & Aluko, R. E. (2009). Flaxseed protein–derived peptide fractions: Antioxidant properties and inhibition of lipopolysaccharide–induced nitric oxide production in murine macrophages. *Food Chemistry*, 116(1), 277–284. <http://doi.org/10.1016/j.foodchem.2009.02.046>
- Udenigwe, C. C., Li, H., & Aluko, R. E. (2012). Quantitative structure–activity relationship modeling of renin–inhibiting dipeptides. *Amino Acids*, 42(4), 1379–1386. <http://doi.org/10.1007/s00726-011-0833-2>
- Udenigwe, C. C. (2014). Bioinformatics approaches, prospects and challenges of food bioactive peptide research. *Trends in Food Science & Technology*, 36(2), 137–143. <http://doi.org/10.1016/j.tifs.2014.02.004>
- Valdes–Rodriguez, S., Segura–Nieto, M., Chagolla–Lopez, A., Verver Vargas–Cortina, A., Martinez–Callardo, N., & Blanco–Labra, A. (1993). Purification, Characterization, and Complete Amino Acid Sequence of a Trypsin Inhibitor from Amaranth (*Amaranthus hypochondriacus*) Seeds'. *Plant Physiol*, 103(4), 1407–1412. <http://doi.org/10.1104/pp.103.4.1407>

- Van der Wulp, M. Y. M., Verkade, H. J., & Groen, A. K. (2013). Regulation of cholesterol homeostasis. *Molecular and Cellular Endocrinology*, 368(1-2), 1-16. <http://doi.org/10.1016/j.mce.2012.06.007>
- Vecchi, B. (2007). Actividad antihipertensiva de péptidos derivados de proteínas de amaranto. Tesis de grado. Universidad Nacional de la Plata.
- Vecchi, B., & Añón, M. C. (2009). ACE inhibitory tetrapeptides from *Amaranthus hypochondriacus* 11S globulin. *Phytochemistry*, 70(7), 864-870. <http://doi.org/10.1016/j.phytochem.2009.04.006>
- Velarde-Salcedo, A. J., Barrera-Pacheco, A., Lara-González, S., Montero-Morán, G. M., Díaz-Gois, A., González De Mejía, E., & Barba de la Rosa, A. P. (2013). In vitro inhibition of dipeptidyl peptidase IV by peptides derived from the hydrolysis of amaranth (*Amaranthus hypochondriacus* L.) proteins. *Food Chemistry*. <http://doi.org/10.1016/j.foodchem.2012.08.032>
- Venskutonis, P. R., & Kraujalis, P. (2013). Nutritional Components of Amaranth Seeds and Vegetables: A Review on Composition, Properties, and Uses. *Comprehensive Reviews in Food Science and Food Safety*, 12(4), 381-412. <http://doi.org/10.1111/1541-4337.12021>
- Ventureira, J. L., Martínez, E. N., & Añón, M. C. (2012). Effect of acid treatment on structural and foaming properties of soy amaranth protein mixtures. *Food Hydrocolloids*, 29(2), 272-279. <http://doi.org/10.1016/j.foodhyd.2012.03.013>
- Wang, G., Li, X., & Wang, Z. (2016). APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research*, 44 (November 2015), D1087-D1093. <http://doi.org/10.1093/nar/gkv1278>
- Wang, J. H., Liu, Y. Le, Ning, J. H., Yu, J., Li, X. H., & Wang, F. X. (2013). Is the structural diversity of tripeptides sufficient for developing functional food additives with satisfactory multiple bioactivities? *Journal of Molecular Structure*, 1040, 164-170. <http://doi.org/10.1016/j.molstruc.2013.03.004>
- Wang, W., & González De Mejía, E. (2005). A new frontier in soy bioactive peptides that may prevent age-related chronic diseases. *Comprehensive Reviews in Food Science and Food Safety*, 4, 63-78.
- Wang, X., Chen, H., Fu, X., Li, S., & Wei, J. (2017). A novel antioxidant and ACE inhibitory peptide from rice bran protein: Biochemical characterization and molecular docking study. *LWT - Food Science and Technology*, 75, 93-99. <http://doi.org/10.1016/j.lwt.2016.08.047>
- Wingfield, P. (2001). Protein precipitation using ammonium sulfate. *Current Protocols in Protein Science*, Appendix 3, Appendix 3F. <http://doi.org/10.1002/0471140864.psa03fs13>
- Witztum, J. L., & Steinberg, D. (1991). Role of oxidized low density lipoprotein in atherogenesis. *The Journal of Clinical Investigation*, 88(6), 1785-92. <http://doi.org/10.1172/JCI115499>
- Worth, C. L., Gong, S., & Blundell, T. L. (2009). Structural and functional constraints in the evolution of protein families. *Nat Rev Mol Cell Biol*, 10(10), 709-720. <http://doi.org/nrm276210.1038/nrm2762>
- Yamada, K., Naemura, A., Sawashita, N., Noguchi, Y., & Yamamoto, J. (2004). An onion variety has natural antithrombotic effect as assessed by thrombosis/thrombolysis models in rodents. *Thrombosis Research*, 114(3), 213-220. <http://doi.org/10.1016/j.thromres.2004.06.007>
- Yamaguchi, N., Kawaguchi, K., & Yamamoto, N. (2009). Study of the mechanism of antihypertensive peptides VPP and IPP in spontaneously hypertensive rats by DNA microarray analysis. *European Journal of Pharmacology*, 620(1-3), 71-77. <http://doi.org/10.1016/j.ejphar.2009.08.005>
- Yang, W. G., Wang, Z., & Xu, S. Y. (2007). A new method for determination of antithrombotic activity of egg white protein hydrolysate by microplate reader. *Chinese Chemical Letters*, 18(4), 449-451. <http://doi.org/10.1016/j.ccllet.2007.02.014>

Yawadio Nsimba, R., Kikuzaki, H., & Konishi, Y. (2008). Antioxidant activity of various extracts and fractions of *Chenopodium quinoa* and *Amaranthus* spp. seeds. *Food Chemistry*, 106(2), 760-766
<http://doi.org/10.1016/j.foodchem.2007.06.004>

Yu, G., Wang, F., Zhang, B., & Fan, J. (2016). In vitro inhibition of platelet aggregation by peptides derived from oat (*Avena sativa* L.), highland barley (*Hordeum vulgare* Linn. var. *nudum* Hook. f.), and buckwheat (*Fagopyrum esculentum* Moench) proteins. *Food Chemistry*, 194, 577-586.
<http://doi.org/10.1016/j.foodchem.2015.08.058>

Zhang, S. B., Wang, Z., & Xu, S. Y. (2008). Antioxidant and antithrombotic activities of rapeseed peptides. *JAACS, Journal of the American Oil Chemists' Society*, 85(6), 521-527. <http://doi.org/10.1007/s11746-008-1217-y>

Zhao, C., & Dahlman-Wright, K. (2010). Liver X receptor in cholesterol metabolism. *The Journal of Endocrinology*, 204(3), 233-40. <http://doi.org/10.1677/JOE-09-0271>

Zioudrou, C., Streaty, R. A., & Klee, W. A. (1979). Peptides Derived from Food Proteins The exophins. *The Journal of Biological Chemistry*, 254(7), 2446-2449.

ANEXO

TABLA A.1 Lista completa de péptidos bioactivos extraída de BIOPEP (mayo de 2011). Se detalla el Id del péptido y su secuencia. La actividad (ACT) se indica con el siguiente código numérico y corresponde a la clasificación provista en la base de datos: 1-Antiamnésico; 2-Inhibidor de ACE; 3-Toxina celíaca; 4-Antibacteriano; 5-Antitrombótica; 6-Immunomoduladora; 7-Opioide; 8-Estimulante; 9-Immunoestimulante; 10-Neuropéptido; 11-Regulatoria; 12-Anticancerígena; 13-Antioxidante; 14-Contractil; 15-Ligando de permeasa bacteriana; 16-Antiinflamatoria; 17-Natriurético; 18-inhibitoria; 19-Quimiotáctico; 20-Acelerante; 21-Hemolítica; 22-Anorexígeno; 23-Agonista Opioide; 24-Antagonista Opioide; 25-Reactivo; 26-Unión; 27-Hipotensivo; 28-Estimulante de otras actividades; 29-Antibiótico; 30-Embriotóxico; 31-Antifúngica; 32-Antiviral; 33-Vasoconstrictor; 34-Agonista del Receptor Neuropeptídico GPR14; 35-Activación de la proteólisis mediada por ubiquitina; 36-Péptido promotor de fertilización; 37-Inhibidor de Proteína Kinasa C (CPK); 38-Mineralización de tejidos y fluidos corporales (mineralizante); 39-Péptidos activos en membrana; 40-Unión a Heparina.

Id	ACT	SECUENCIA
2605	1	IHPFAQTQ
2606	1	HLPLPL
2607	1	LYPFVEPI
2608	1	IYPFPGPI
2609	1	IYVPEPS
2610	1	IYPFPGPIH
2611	1	IYPFPGPI
2612	1	VYPFPGPIA
2613	1	VYPFPGPI
2614	1	VYPFPGPIH
2615	1	VYPFPGPI
2616	1	IFPFVEPI
2802	1	QPPPAPGPGT
2803	1	LPPPVH
2804	1	PPPVHL
2805	1	RPPPAQAPAP
2806	1	HLPPP
2807	1	MPPPLPARVDFSLAG ALN
2808	1	MTPLP
2809	1	PPPLP
2810	1	MPPPLP
2811	1	VHLP
2812	1	LPPP
2813	1	GVQVETISPGDGR
2814	1	KPPI
2815	1	MTPLPARVDFSLA GALN
2816	1	GVQVETISPGDGRTF PK
2817	1	HLPPPVLPPP
2818	1	KPPV
2819	1	PRPQHPQHP
2820	1	LPPV
2821	1	TPPV
2822	1	PPPPGGQPRPPQG RYDWWPYGNLFGG
2823	1	HTFISP MPPPLPTRVDFSLAG
2824	1	ALN
2825	1	LPPPLPLLA
2826	1	APTMPPLP
2827	1	MPPPLPA
2828	1	PPAYPPPPV
2829	1	APPTPPPLP
2830	1	PPP
3166	1	VPL
3455	1	LLSPWNINA
3456	1	LSPFWNINA
3457	1	SPFWNINA
3458	1	GPGG
3459	1	PGP
3460	1	PG
3461	1	GP
3774	1	MPPPL
3775	1	MPPLP
3776	1	MPLP
2617	2	VMP
2618	2	GRVMP
2619	2	TVYTKGRVMP
2620	2	PVPQP
2621	2	PAVVLP
2622	2	NILP
2623	2	PQNILP
2624	2	LPQNILP
2625	2	YGFLP
2626	2	EPIPYGFLP
2627	2	IYP
2628	2	LIYP
2629	2	PLIYP
2630	2	QPLIYP
2631	2	PQPLIYP
2632	2	QPQPLIYP
2633	2	FQPQPLIYP
2636	2	PSFQPQPLIYP
2637	2	PSFQP
2638	2	YPSFQPQPLIYP
2639	2	IYPSFQPQPLIYP
2640	2	KIYPSFQPQPLIYP
2641	2	DKIYPSFQPQPLIYP
2642	2	ILP
2643	2	PVRAVP
2644	2	TDQHQDKIYP
2645	2	HQIYP
2646	2	PHQIYP
2649	2	NPPHQIYP
2651	2	VLPYPV
2652	2	VLPYP
2653	2	VLP
2654	2	PQEVLP
2655	2	IPQEVLP
2656	2	VLPIQ
2657	2	VLPIP
2658	2	PQTLALP
2660	2	QPIP
2661	2	PQPIP
2662	2	VPQPIP
2663	2	QVPQPIP
2664	2	PLP
2665	2	LPLP
2666	2	HLPLP
2667	2	LHLPLP

2668	2	LHLP
2669	2	NLHLPLP
2671	2	NLHLP
2672	2	ENLHLPLP
2673	2	LENLHLPLP
2975	2	ALKAWSVAR
3255	2	YLLF
3257	2	RL
3258	2	IR
3259	2	HIRL
3260	2	HIR
3261	2	ALPMHIR
3331	2	LVYP
3332	2	SLVYP
3333	2	TQSLVYP
3334	2	QTQSLVYP
3335	2	AQTQSLVYP
3336	2	FAQTQSLVYP
3337	2	HPFAQTQSLVYP
3338	2	IHPFAQTQSLVYP
3339	2	KIHFAQTQSLVYP
3341	2	FQP
3342	2	GPA
3343	2	ALPHA
3344	2	IVGRPRHQG
3345	2	DMIPAQK
3346	2	IKPLNY
3347	2	DYGLYP
3348	2	IWHHT
3349	2	IKP
3357	2	GYALPHA
3358	2	YALPHA
3359	2	KVLAGM
3360	2	VKAGF
3361	2	LKL
3362	2	LKVGVKQY
3363	2	GPAGAPGAA
3364	2	HGLF
3365	2	FQKVVAG
3366	2	FQKVVAK
3367	2	GKKVLQ
3368	2	FQKVVA
3369	2	YQQPVLGPVR
3370	2	AVP
3371	2	AVPYP
3372	2	PYP
3373	2	PQR
3374	2	PFPE

3375	2	AAP
3376	2	DFVAP
3377	2	FGK
3378	2	GRP
3379	2	AKK
3380	2	RY
3381	2	LY
3382	2	GWAP
3383	2	IY
3384	2	VF
3385	2	MF
3386	2	KW
3387	2	RFH
3388	2	MY
3389	2	LW
3391	2	LPP
3393	2	FAP
3394	2	FPFEVFGK
3395	2	AMKPWIQPK
3397	2	LQSW
3398	2	MKPWIQPK
3399	2	AMKPW
3400	2	WKVIP
3401	2	YPVQPFTE
3402	2	LVVYPWTQR
3403	2	YPWTQR
3404	2	VRP
3405	2	YANPAVVRP
3406	2	TAP
3407	2	MYY
3408	2	PFDFPQIP
3409	2	LLNPPHQIYP
3410	2	ALPP
3411	2	PPHQIYP
3412	2	LALPP
3413	2	LNPPHQIYP
3415	2	PANIKWGD
3416	2	PSKIKWGD
3417	2	PTHIKWGD
3418	2	HQAAGW
3419	2	IRPVQ
3420	2	GVYPHK
3421	2	LVL
3422	2	SAYPGQITSN
3423	2	QVSLNSGY
3425	2	NRVYVHPF
3426	2	QSLVYP
		ENGLPVHLDQSIFRR
3465	2	P

3466	2	ENGLPVHLDQSIFRR
3474	2	GDAP
3480	2	AVPYPQR
3482	2	SFQPQPLIYP
3483	2	GKKIATYQER
3484	2	GHKIATFQER
3485	2	PANLPWGSSNV
3486	2	VW
3487	2	VWY
3488	2	YW
3489	2	RF
3490	2	FWN
3491	2	IYPRY
3492	2	VY
3493	2	YGGY
3494	2	HY
3495	2	YIPIQYVLSR
3496	2	FFVAPFPFEVFGK
3497	2	FFVAP
3498	2	KVLPVPQ
3499	2	SKVLPVPE
3500	2	YKVPQL
3501	2	AYFYPE
3502	2	FP
3503	2	TPVVVPPFLQP
3504	2	VYPFPG
3505	2	VYP
3506	2	GKP
3507	2	IPA
3508	2	GLDIQK
3509	2	VAGTWY
3510	2	YQQPVL
3511	2	EMPFK
3512	2	DAYPSGAW
3513	2	LAYFYP
3514	2	LAHKAL
3515	2	GGY
3516	2	IRAQQ
3517	2	PPPVHL
3518	2	VAA
3519	2	AYFYP
3521	2	VAP
3522	2	IPP
3523	2	LYPVK
3524	2	VPP
3525	2	VHIPP
3526	2	VHLPP
3527	2	VHLPPP

3528	2	LVR
3529	2	AVNPIR
3530	2	TTMPLW
3531	2	PLW
3532	2	GY
3534	2	YPRY
3535	2	YPR
3536	2	PRY
3537	2	PR
3538	2	VSP
3539	2	LAA
3540	2	IYPR
3541	2	LSP
3542	2	LQP
3543	2	LRP
3544	2	LNP
3545	2	VHLAP
3546	2	VAY
3547	2	IRA
3550	2	YL
3551	2	LF
3553	2	YG
3554	2	YGLF
3556	2	FY
3557	2	FNQ
3558	2	LAY
3563	2	AY
3564	2	LSPA
3565	2	SQPQAFP
3566	2	EPKAIP
3567	2	SEPKAIP
3568	2	QSEPKAIP
3569	2	IQSEPKAIP
3570	2	GIQSEPKAIP
3571	2	SGIQSEPKAIP
3572	2	FHSGIQSEPKAIP
3573	2	AFP
3575	2	QPQAFP
3577	2	QSQPQAFP
3578	2	IQSQPQAFP
3579	2	SIQSQPQAFP
3585	2	PQAFP
3595	2	KAIP
3597	2	AIP
3627	2	FRAEHPL
3666	2	YP
3712	2	FYQQ
3713	2	LLP

3714	2	LQQ
3715	2	YSQQQQ
3742	2	PKAIP
3763	2	LNPA
3770	2	AHSY
3969	2	RADHPF
3970	2	WLAHK
3971	2	VGINYWLAHK
3973	2	YGL
3974	2	YGLVAGTW
3975	2	VPK
3976	2	LAMA
3977	2	LDAQSAPLR
3978	2	CMENSA
3979	2	VLDTDYK
3980	2	VAGTW
7070	2	LDIQK
7414	2	LRW
7415	2	VLIVP
7478	2	HHL
7479	2	HLL
7480	2	FALPQY
7481	2	VTSTAV
7483	2	RPKHPIKHQ
7484	2	AMPKPW
7485	2	TKVIP
7486	2	YFPFGPIP
7487	2	KVLPVP
7488	2	VFK
7489	2	ALPMH
7490	2	GTW
7491	2	GVW
7492	2	VYFPFGPI
7493	2	DKIHPF
7494	2	LNVPGEIVE
7495	2	NIPPLTQTPV
7496	2	FPEVFGK
7497	2	AMKPWIQPK
7498	2	TVY
7499	2	FVAP
7500	2	SLVLPVPE
7501	2	YFPFGPIP
7502	2	EMFPFK
7503	2	TQVY
7504	2	SKVYP
7505	2	SKVYFPFGPI
7506	2	GPL
7507	2	PGL

7508	2	LGP
7509	2	GLP
7510	2	PLG
7511	2	LPG
7512	2	GP
7513	2	PL
7514	2	NWGPLV
7515	2	VKKVLGNP
7516	2	RMLGQTPWK
7541	2	IVY
7542	2	DLP
7543	2	AW
7544	2	IW
7545	2	GPV
7546	2	IKW
7547	2	IRP
7548	2	IWH
7549	2	LKP
7550	2	LYP
7551	2	YQY
7552	2	VFPS
7553	2	CWLPHY
7554	2	GEP
7555	2	PGTAVFK
7556	2	YPLDL
7557	2	YPLDLF
7558	2	VK
7559	2	PSY
7560	2	FFL
7561	2	IYLL
7562	2	IA
7563	2	YLAGNQ
		LVYFPFGPIPNSLPQN
7564	2	IPP
7565	2	LHLPLPL
7566	2	VRGPFPIIV
7567	2	KVREGTTY
7568	2	KVREGT
7569	2	LKA
7570	2	LAP
7571	2	MNPPK
7572	2	ITTNP
7573	2	RMLGNTPTK
7574	2	VLAQYK
7575	2	FQKPKR
7576	2	IVGRRRHQG
7577	2	FKGRYYP
7578	2	WSVPQPK
7579	2	GW

7580	2	RW
7581	2	IP
7582	2	RP
7583	2	AF
7584	2	AP
7585	2	LA
7586	2	KR
7587	2	VP
7588	2	RA
7589	2	YA
7590	2	AA
7591	2	GF
7592	2	FR
7593	2	IF
7594	2	VG
7595	2	IG
7596	2	GI
7597	2	GM
7598	2	GA
7599	2	GL
7600	2	AG
7601	2	GH
7602	2	HL
7603	2	GR
7604	2	KG
7605	2	FG
7606	2	DA
7607	2	GS
7608	2	GV
7609	2	MG
7610	2	GQ
7611	2	GK
7612	2	GT
7613	2	WG
7614	2	HG
7615	2	GE
7616	2	GG
7617	2	QG
7618	2	SG
7619	2	LG
7620	2	GD
7621	2	TG
7622	2	EG
7623	2	EA
7624	2	NG
7625	2	PG
7626	2	IAK
7627	2	KDERF

7628	2	VR
7629	2	EKDERF
7630	2	YIPIQY
7631	2	LPYPY
7632	2	MKG
7633	2	PAP
7634	2	YVP
7635	2	VAV
7636	2	SHP
7637	2	GHF
7638	2	LTF
7639	2	IFG
7640	2	MNP
7641	2	TNP
7642	2	NPP
7643	2	TTN
7644	2	ITT
7645	2	PPK
7646	2	GYK
7647	2	YGG
7648	2	FCF
7649	2	LRY
7650	2	MKY
7651	2	YKY
7652	2	KFY
7653	2	KYY
7654	2	NKL
7655	2	YNK
7656	2	IYK
7657	2	LIY
7658	2	FALPQYLK
7659	2	FPQYLQY
7660	2	AMKPW
7661	2	LFRQ
7662	2	YQEPVL
7663	2	YQEPVLQPVR
7664	2	YAEERYPIL
7665	2	YPPFGPI
7666	2	LLYQQPV
7667	2	IAIP
7668	2	AIPP
7669	2	IAIPP
7670	2	PNSHP
7671	2	QAFT
7672	2	SSIQSQQAFT
7673	2	HSSIQSQQAFT
7674	2	VHSSIQSQQAFT
7675	2	KVHSSIQSQQAFT

7676	2	HSGIQSEPKAIP
7677	2	KFHSGIQSEPKAIP
7678	2	LKPMN
7679	2	YRPY
7680	2	QK
7681	2	DG
7682	2	NY
7683	2	NF
7684	2	SY
7685	2	SF
7686	2	FFGRCVSP
7687	2	ERKIKVYL
7688	2	FGRCVSP
7689	2	NIFYCP
7690	2	AKYSY
7691	2	KY
7692	2	KF
7693	2	KL
7694	2	YKYY
7695	2	KFYG
7696	2	AIYK
7697	2	YK
7698	2	NK
7699	2	KKIATYQER
7700	2	YLYEIAR
7701	2	YLYEIARR
7702	2	YLYEIA
7741	2	RR
7742	2	AR
7743	2	KA
7744	2	GAAELPCSADWW
7745	2	RMLGQTPTK
7746	2	LVE
7747	2	FGASTRGA
7748	2	MIFPGGPQL
7749	2	VIY
7750	2	VVYPW
7751	2	CF
7752	2	EY
7798	2	VVRP
7799	2	AVVRP
7800	2	FFVAPFPEVFGK
7801	2	VRYL
7802	2	NMAINPSK
7803	2	IPY
7804	2	ALNEINQFY
7805	2	ALNEINQFYQK
7806	2	LQKW

7807	2	LLF
7808	2	LVRT
7809	2	GVHGV
7810	2	KP
7811	2	VWIG
7812	2	EVMAGNYLPG
7813	2	MIFPGAGGPEL
7814	2	MDFLI
7815	2	MFDL
7816	2	MDLA
7817	2	GEP
7818	2	VIEKYP
7819	2	IAP
7820	2	GPP
7821	2	RIY
7822	2	IAPG
7823	2	FAL
7824	2	IAE
7825	2	PSGQYY
7826	2	EI
7827	2	IE
7828	2	EV
7829	2	VE
7830	2	TE
7831	2	LQ
7832	2	LN
7833	2	PT
7834	2	TQ
7835	2	AH
7836	2	PP
7837	2	PQ
7838	2	EW
7839	2	ME
7840	2	EK
7841	2	KE
7842	2	HP
7843	2	PH
7844	2	HK
7859	2	IEP
8096	2	AGSS
8154	2	TIIPLPV
8155	2	LTLTDVE
8156	2	YPQRDMPIQ
8157	2	PGPIP
8158	2	PKHKEMPFPPKYPVE PFT
8159	2	GPFPIIV
8160	2	RGPFPIIV
8169	2	GPVIRGPFPII

8178	2	PYVRYL
8179	2	LKKISQ
8180	2	IQY
8181	2	FSDKIAK
8182	2	ALEP
8183	2	VIKP
8184	2	IQP
8185	2	TF
8193	2	AI
8199	2	KAPVA
8200	2	PTPVP
8201	2	MYPGIA
8202	2	VIPEL
8227	2	IEY
8228	2	IEW
8229	2	IKY
8254	2	AGSP
8286	2	GPVIRGPFPII
8290	2	GPLGLLGLGFLGLS
8291	2	PGPLGLTGP
8292	2	QLGFLGPR
8296	2	GVHHA
8297	2	MNVKHWPWMK
8298	2	VTVNPYKWLP
8299	2	RMLGQTP
8300	2	EKERERQ
8301	2	KRQKYDI
8307	2	KVLILA
8308	2	VVPP
8309	2	VRSP
8328	2	VYNEGLPAP
2576	3	YQLLQELCCQHL
2577	3	QEQVPLVQQF
2578	3	PSQQQP
2797	3	YPQPQ
2798	3	YPQPQP
2800	3	QQPYPQ
3619	3	VRVPVPLQPNPS QQQPQEQVPLPQQ QQF
3638	3	QQQQPSSQVSFQQP LQQYPLGQGSFRPS QQNPQA
3640	3	LGQQQPFPQPYP QPQPFPSSQPY
3819	3	SQQQQPPFSQQQQP PFSQQQQPPFSQQQ PPF
3986	3	LGQQQPFPQPYP

3987	3	LGQQQPFPQPYP QPQP
6986	3	QQQP
6987	3	PSQQ
6988	3	QQPY
6989	3	QPYP
6990	3	PQNPSQQQPQEQVP
6991	3	QQFLGQQQPFPQP QQKQQQQPSSQV S
6992	3	LQPQNPSQQQPQEQ
6993	3	QPQPFPSSQPYLQL
6994	3	QGSFRPSQQNPQAQ
6995	3	QPFPQPYPQPQP
6996	3	QPFPSSQPYLQLQP
6997	3	QPFRPQPYPQPQP
6998	3	PFRPQPYPQPQP
6999	3	VRFPVPLQPN
7000	3	QLQPQNPSQQQP
7001	3	PSQQQPQEQVPL
7002	3	QEQVPLVQQQF
7003	3	VQQQPLGQQQP
7004	3	VQQQPLGQQQP
7005	3	FPPQPYPQPQP
7006	3	YPQPFPSSQQP
7007	3	FPSQQPYLQLQP
7008	3	YLQLQFPQPQL
7009	3	FPQPPLPYSQPQ
7010	3	PYSQPFPFRPQQ
7011	3	PFRPQPYPQPQP
7012	3	PYPQPQPYSQP
7013	3	PQYSQPQPISQ
7014	3	QQPISQQILQQI
7015	3	QILQQILQQLI
7016	3	LQQQLIPCMDVV
7017	3	PCMDVVLQQHNI
7018	3	LQQHNIAHGRSQ
7019	3	AHGRSQVLQOST
7020	3	VLQQSTYQLLQE
7021	3	YQLLQELCCQHL
7022	3	LCCQHLWQIPEQ
7023	3	WQIPEQSQCQAI
7024	3	SQCQAIHNVVHA
7025	3	HNVVHAIILHQQ
7026	3	IILHQQKQQQQ
7027	3	QKQQQQPSSQVS
7028	3	PSSQVSFQQPLQ
7029	3	FQQPLQQYPLGQ
7030	3	QYPLGQGSFRPS
7031	3	

7032	3	GSFRPSQQNPQA
7033	3	QQNPQAQGSVQP
7034	3	QGSVQPQQLPQF
7035	3	QQLPQFEEIRNL
7036	3	EEIRNLALQTLQ
7037	3	ALQTLPAMCNVY
7038	3	AMCNVYIPPYCT
7039	3	IPPYCTIAPFGI
7040	3	IAPFGIFGTN
7044	3	PFPQPQPYPQPQPF
7059	3	LQLPFPQPQLPY
7060	3	PQPQLPYQPQLPY
7061	3	QPQQSFPQQQ
7062	3	VQGGQHQPQQPA QL
7063	3	QQPFQQQQPLPQ
7064	3	QPQPFQSQEQSQQ PFPQPQPF
7065	3	QQQQPPFSQQQQSP FSQQQQ
7066	3	QFPQPQLPYQPQL PY
7067	3	PYPQPQLPY
7068	3	FPQQPQQPYPQQP
7069	3	FSQPQQQFQPQPQ
7308	3	QYPQPQQPI
7309	3	IFPQPQQTF
7310	3	PFPQPQQPQ
7311	3	PFPQPQQQF
7312	3	PFPQPQQQL
7313	3	PFPQPQQTF
7314	3	PFPQPQQQQ
7315	3	PFPQPQQQT
7316	3	PFPQQQQPL
7317	3	PFPQTQQPQ
7318	3	PHQQQQQQF
7319	3	PLPQPQQPQ
7320	3	PQPQQPQQQ
7321	3	PQPQQQLPQ
7322	3	PQQQFPQPQ
7323	3	PQQQFPQTQ
7324	3	PQQQIPQPQ
7325	3	PQQQLPQPQ
7326	3	PQTQQPQQQ
7327	3	PQQQQPFPQ
7328	3	QFPQPQQPQ
7329	3	QFPQTQQPQ
7330	3	QIPQTQQPQ
7331	3	QLPFPQQPF
7332	3	QLPFPQQPQ
7333	3	QQPQLPFPQ
7334	3	QQPQQPFPL
7335	3	QQPQQPLPQ
7336	3	QQPQQQFPQ
7337	3	QQQQFPQPQ
7338	3	QQQQQFPQ
7339	3	QQTFFPQPQ
7340	3	QQTYPQRQPQ
7341	3	QRPQQPFPQ
7342	3	QTQQPQQPF
7343	3	QWPQQQQPF
7344	3	PIQQQPHQF
7345	3	PQQHQQFPQ
7346	3	QFPQQQIPV
7347	3	QQPIQQQPH
7348	3	PFPQPQLPF
7349	3	PFPQPQQPI
7350	3	PQPQLPFPQ
7351	3	PQPQQPIPV
7352	3	PYPQQPYPS
7353	3	QQPYPQQPY
7354	3	PFPQPQEPI
7355	3	PFPQPQQPN
7356	3	PFPQQPFPQ
7357	3	PFPQQPIPQ
7358	3	PFPQQPQPY
7359	3	PIPQQPQPY
7360	3	PLPRPQQPF
7361	3	PLPQPQQPF
7362	3	PQHQPFPQPT
7363	3	PQPFPQQPI
7364	3	PQPFPQQPQ
7365	3	PQPQEPIPQ
7366	3	PQPQQPNPQ
7367	3	PQPTQQFPQ
7368	3	PQPYPQQPF
7369	3	PQPYPQQPQ
7370	3	PQPYPQQPY
7371	3	PQQLPQQQF
7372	3	PQQQFPQQM
7373	3	PRPQQPFPW
7374	3	PWPQQPFPQ
7375	3	PYPQQPQPY
7376	3	QEPIPQQPQ
7377	3	QFPQPQQPF
7378	3	QLPQQQFPQ
7379	3	QQPFPQQQE
7380	3	QQPFPQQPF
7381	3	QQPFPQQPI
7382	3	QQPFPQQPV
7383	3	QQPFPQQPY
7384	3	QQPFPQQTI
7385	3	QQPLPQQPQ
7386	3	QQPNPQQPQ
7387	3	QQPQHQPFPQ
7388	3	QQPQLPHGH
7389	3	QQPQQPFPR
7390	3	QQPQQPLPR
7391	3	QQQQLPQQH
7392	3	QQTIPQQPQ
7393	3	PFPQQPYPQ
7394	3	PQPQQPFPW
7395	3	QQPIPQQPQ
7396	3	IQPQQPFPQ
7397	3	PFPQPQQPT
7398	3	PFPQQPQQI
7399	3	PQPQQPIPQ
7400	3	PQPQQPTPI
7401	3	PQPQQQLPL
7402	3	PYPQQPFPQ
7403	3	PYPQQPQQQ
7404	3	QFPQQPQQI
7405	3	QHQQFPQQQ
7406	3	QQHPQQPQ
7407	3	QQPFPQRPQ
7408	3	QQPQRPEQQ
7409	3	QQPYPQQPF
7410	3	QQQFPQQPQ
7411	3	QQTQQPFPL
7412	3	QRPEQQFPQ
7413	3	PQPQQPYPQ
7416	3	QQFPQPQQPFPQQP
7417	3	QPFPQPQQPFPQSQ
7418	3	QLQFPQPQLPYPQ
7419	3	PQQPFPQPQQPFRQ
7420	3	PQQPFPQPQQPFPQ
7421	3	QYQPYPEQQEPFVQ
7422	3	QYQPYPEQQQPFVQ
7423	3	PQFRPQQPYPQPQ PQ
7424	3	PQQPQQSFPQQQRP F
7425	3	PQQPQQSFPQQPQR
7426	3	QQPFPQPQQPFPQ
7427	3	QQPFPQPQQPFPQ

7428	3	QQPFVQQQPFVQ Q
7429	3	QQPPFSQQQPVL PQ
7430	3	QQPPFSQQQLVLP Q
7431	3	QQPPFSEEEEPVLP Q
7432	3	QQPPFSEQQLVLP Q
7433	3	QLQPPFPEQLPY
7434	3	QLQPPFPPELQPY
7435	3	QLQPPFPQQLPY
7436	3	QLQPPFPQQLPY
7437	3	VQCGQIIPQPPAQ L
7438	3	QQQQPPFSQQQSP FSQQQQ
7439	3	QQPPFSQQQQPLP Q
7440	3	QPQPPQSQSEQSQ PFQPQPF
7441	3	QQXSQPQXPQQQ XPQQPQQ
7442	3	PLVQQQFLGQQQP PPQ
7443	3	YYPTSPQSQGQ
7444	3	QGQQGYPTSPQ
7445	3	SSQVSFQPSQLN
7446	3	FPQTQQPQLFPQS Q
7447	3	GQQGYPTSVQQSG Q
7448	3	GSVQPQQQLPQFE IR
7449	3	FPQHNYQQQPQTF PQPF
7450	3	QQPIQPQQFPQQF F
7451	3	QSGQYQQPQMOT T
7452	3	ESQQSQQDEPQPF
7453	3	FPQPEDQQSQQSE
7454	3	LQQVQQGPQQPQ PQPQPF
7455	3	QVQWPQQPFPQP QQPF
7456	3	PLLQPPFSQQEQP QF
7457	3	WQQQPPFSQEQPQ F
7458	3	QSNLPQPAQQPQ PQVPPQP
7459	3	WFQPSQLNPQAQQ DQPQ
7460	3	LQQPQQPQFQPQQ QF

7461	3	VQQQIPVVQPSIL
7462	3	VQQQFPQPPQPP PPQQPYPQPPFSQ QPY
7463	3	VRVPVPLQPPQNP SQQPPQEQVPLVQQ QQF
7464	3	LGQGSFRPSQQN
7465	3	AGQQQPPFPQPP PQPQPF
7466	3	LGQQQAFPPQPP PQPQPF
7467	3	PQPQPPFSQQPY
7468	3	LGQQQFPFPQPP PQPQPPFSQQPY
7469	3	VVKVGGSSSLGW
7470	3	VIQVGGSNVLDI
7471	3	VIPVGGSSALGA
7472	3	VIRVGGASEVEV
7473	3	RIRVQSMNMG
7474	3	VLKVGSTSDVEV
7475	3	VEKIGGASSRGE
7476	3	VLKMAGNSFQEN
7477	3	LQLQPPFPQPLP PQPLPYPQPLPYP PQP
2579	4	NGKRVVCVR
2580	4	NGKRVVCVRG
2581	4	CNGKRVVCVRG
2582	4	CNGKRVVCVR
2583	4	CNGKRVVC
2584	4	CNGKRVCCG
2585	4	YCNGKRVVCVRG
2586	4	YCNGKRVVCVR
2587	4	YCNGKRVVC
2588	4	YCNGKRVCCG
2589	4	DAACAAHCLFR
2590	4	DAACAAHCLFRG
2591	4	NDAACAAHCLFR GRSGG
2592	4	NDAACAAHCLFR GRSGGG
2593	4	VTCDILSVEAKGV KLG
2594	4	VTCDILSVEAKGV KLG
2595	4	KWCFRVCYRGICY RRCR
2596	4	KWCFRVCYRGICY RRCR
2597	4	RWCFRVCYRGICY RRCR

2598	4	RWCFRVCYRGICY RRCR
2702	4	SDEKASPKHHRFSL SRYAKLANRLANPKL LETFLSSKWIGDRGN RSV
2703	4	CIGNGRCNENVGP PYCCSGFCLLRQPN QGYGVCRNA
2757	4	ACYCRIGACVSGERL TGACGLNGRIYRLC CR
2758	4	CSCRTSSCRFGERLS GACRLNGRIYRLCC
2759	4	VTCYCRSTRCGFRER LSGACGYRGRIYRLC CR
2760	4	VTCYCRTRCGFRE RLSGACGYRGRIYRL CCR
2761	4	VCYCRSRGCKGRER MNGTCRKGHLLYTL CCR
2762	4	VFCTCRGFLCGSER ASGCTINGVRHTLC CRR
2763	4	VSCTCRRFSCGFRER ASGCTVNGVRHTL CCRR
2764	4	GRCVCRKQLLSYR ERRIGDCKIRGVRFPF CCPR
2765	4	GICACRRFCPNSER FSGYCRVNGARYVR CCSRR
2766	4	VVCACRRALCLPLER RAAGFCRIRGRIHPL CCRR
2767	4	VVCACRRALCLPRER RAGFCRIRGRIHPLC CRR
2792	4	RRWCFRVCYKGFY RKCRG
2793	4	RRWCFRVCYKGFY RKCR
2794	4	RRWCFRVCYKGFY RKCR
2795	4	RRWCFRVCYKGFY RKCRG
2948	4	GIGAVLKVLTGLPA LISWIKRKRQQ
2955	4	GWASKIGQTLGKIAK VGLKELIQPK
2956	4	GVLSNVIGYLLKLT GALNAVLKQ

2957	4	GMASKAGAIAGKIA KVALKALG
2958	4	GMASKAGAIAGKIA KVALKAL
2959	4	GWASKIGQTLGKIAK VGLQGLMQPK
2960	4	GFGSFLGKALKAAALK IGANALGGSPQQ
2961	4	GIGKFLHSAKFKGKA FVGEIMNS
2962	4	GIGKFLHSAKFKGKA FVGEIMKS
2963	4	GLLQTIKEKLESLESL AKGIVSGIQA
2964	4	GLLQTIKEKLESLESL AKGIVSGIQAG
2965	4	GFVDFLKKVAGTIAN VVT
2966	4	GFVDFLKKVAGTIAN VVTG
2967	4	GLFGVLAKVASHVVP AIAEHFQA
2968	4	GLFGVLAKVASHVVP AIAEHFQAG
2969	4	FVGLAKVAHVVP AEHFG
2970	4	FVGLAKVAHVVP AEHF
2971	4	GLFGVLAKVAHVVP PAIAEHFG
2972	4	GLFGVLAKVAHVVP PAIAEHF
3035	4	RPKHPIKHQGLPQEV LNENLLRF
3053	4	ILPWKWPWWPWRR VCSCRLVFCRRTEL VGNCLIGGVSFTYCC
3070	4	TRV
3071	4	DCYCRIPACIAGERR YGTCIYQGRWLWAF C
3072	4	CYCRIPACIAGERRY GTCIYQGRWLWAFCC
3073	4	ACYCRIPACIAGERR YGTCIYQGRWLWAF C
3086	4	RRCICTTRTCRFPYR RLGTICIFQNRVYTF C
3137	4	GIFSKLGRKKIKNLLIS GLKNVGEVGMDDVV
3138	4	RTGIDIAGCKIKGEC
3185	4	ALWKTMLKLGTMAL LHAGKAALGAAADT ISQGTQ

3187	4	KKQLCSFEIYVWPWE D
3188	4	VNYFFDVEVGRTI
3189	4	EDRMSLVDSRCQEA
3190	4	SSSKEENRIIPGGI
3204	4	WNPFKLEKVGQQRV RDAVISAGPAVATV AQATALAKG
3205	4	WNPFKLEKVGQQRV RDAVISAGPAVATV AQATALAK
3206	4	KWKVFKKIEKMGRNI RNGIVKAGPAIAVLG EAKALG
3207	4	KWKVFKKIEKMGRNI RNGIVKAGPAIAVLG EAKAL
3208	4	AIAVLGEAKALMGR NIRNGIVKAGPAIAV LGEAKALG
3209	4	KWKVFKKIEKMGRNI RNGIVKAGPKWKVF KKIEKG
3210	4	KWKVFKKIEKMGRNI RNGIVKAGPKWKVF KKIEK
3211	4	KWKLFFKIEKVGQNI RDGIKAGPAVAVVG QATQIAKG
3212	4	KWKLFFKIEKVGQNI RDGIKAGPAVAVVG QATQIAK
3221	4	GLLSVLGSVAKHVLP HVVPVIAEHLG
3222	4	GLLSVLGSVAKHVLP HVVPVIAEHL
3224	4	GIMDTLKNLAKTAG KGALQSLLNKASCKL SGQC
3225	4	FLPLLAGLAANFLPKI FCKITRKC
3229	4	FKCRRWQWRMCKL GAPSITCVRRAF
3240	4	GIGAILSAGKSIKIG LANGLAEHEG
3241	4	GIGSAILSAGKSALKG GLAGHFANG
3242	4	GIGSAILSAGKSALKG GLAGHFAN
3243	4	GIGASILSAGKSALKG LAKGLAGHFAN
3244	4	GIGASILSAGKSALKG LAKGLAGHFANG
3245	4	GIGAILSAGKSALK GLAKGLAEHF

3246	4	GIGAILSAGKSALK GLAKGLAEHFG
3247	4	GIGALSAGKALKGLA KGLAEHFAN
3248	4	GIGALSAGKALKGLA KGLAEHFANG
3249	4	GIFALLSAAKVGLK GLAKGLAEHFAN
3254	4	RLCRIVVIRVCR
3432	4	KTKLTHEEKNRNLFL KKISQRYQKFAFPQY LKTVYQHQQ
3433	4	QKIAEKFSGTRRG
3435	4	VYQHQAAMKPIWQP KTKVIPYVRY
3436	4	EWSKCYQWQRRMR KKLGAPSITCIRRTSA
3437	4	APRKNVRW
3438	4	KWKSFIKLEKVLKP GGLSNIVTSL
3439	4	KWKLFFKIGAVLK VLTGGLPALIS
3440	4	KWKSFIKLEKVLKK GPILANLVISIV
3441	4	KWKKFIKELQKVLAP GGLSNIVTSL
3442	4	KWKEFIKLTAVKK VLTGGLPALIS
3467	4	VFIDILDKVENAIHN AAQVGIGFAKPFEL INPK
3609	4	GLLRRLRDLKKGIGE KFKKIGY
3610	4	LKLLKLLKLLKLLK
3648	4	KLKLLLLLKLK
3652	4	KKWHKRKK
3667	4	KKFHLRK
3668	4	KKFDLRK
3669	4	KKWHLRKR
3670	4	KGWHKRKK
3671	4	QYRHRCCAOWGPR KYCKRWC
3673	4	RHITSLEVIKAGPHSP TAQLITATLKNRGI SLD
3674	4	SIQDLNVSMKLFKQ AKWKILVKNLNDGREL SLD
3675	4	SIQDLNVSMKLFKQ AKWKIIAKLNDGREL SLD
3676	4	SIQDLNVSMKLFKQ AKWKIILKNDGREL SLD

3677	4	SIQDLNVSMKLFQRKQ AKWKLVKLNDRGREL SLD
3678	4	SIQDLNVSMKLFQRKQ AKWKIVKLNDRGRE LSLD
3679	4	ANIKLSVKWKAQKR FLKMSIIVKLNDRGREL SLD
3680	4	ANIKLSVKWKAQKR FLKMSINVDLSDGRE LSLD
3681	4	SIQDLNVSMKLFQRKQ AKWKIIVKLNDRGREL SLD
3683	4	KFIVTLRVIKAGPHSP TAQIIVELKNGRKLSD D
3684	4	KKFHIRKR
3685	4	ANIKISGKWAQKRF LKMSGNFDLSIEGMS ISAD
3694	4	ANIKLSVEMKLFKRH LKWKIIVKLNDRGREL SLD
3696	4	KKWHRKK
3718	4	PNGLHTRPA
3728	4	SIQDLNVSLKLFQRKQ AKWKIIVKLNDRGREL SLD
3740	4	SIQDLKVSMLFRKQ AKWKIIVKLNDRGREL SLD
3743	4	FKCRRWQWRMCKL GAPSITCVRRFAFA
3754	4	KLKLLKLLKLG
3755	4	FKCRRWQWRMCKL GAPSITCVRRFAFAL
3756	4	QWGRRCGGWGPGR RYCVRWC
3757	4	QYGRRCNNWGPGR RYCKWRC
3758	4	QWGRRCGGWGPGR RYCRRWC
3759	4	GLLRRLRDFLKKIGE KFKKIGYG
3799	4	GILSKLGKALKKAAK HAAKA
3800	4	SIQFLKVSNLDRKQ AKWKIIVKLNDRGREL SLD
3801	4	SIQDLKVSNLFRKQ AKWKIIVKLNDRGREL SLD

3802	4	SIQDLNLSMKLFQRKQ AKWKIIVKLNDRGREL SLD
3806	4	SIQKLVSMKLFQRKQ AKWKIIVKLNDRGREL SLD
3821	4	GRFRRLGRKFKLKF KYGP
3822	4	GLLRRLRKKIGEIFK YG
3823	4	GILSKLGKALKKAAK HAAKAG
3824	4	KLKLLKLLKLG
3878	4	LRDLVCYCRSRGCK GRERMNGTCKRKGHL LYTLCCR
3880	4	FLPLIGRVLSGIL
3881	4	FLPLIGRVLSGILG
3882	4	LLPIVGNLLKSL
3883	4	LLPIVGNLLKSLG
3884	4	LLPILGNLLNGLL
3885	4	LLPILGNLLNGLL
3886	4	LLPIVGNLLNLL
3887	4	LLPIVGNLLNLLG
3888	4	VLPIIGNLLNLL
3889	4	VLPIIGNLLNLLG
3890	4	FLPLIGKVLGIL
3892	4	FFPVGIRILNGIL
3893	4	FFPVGIRILNGILG
3894	4	LSPVIGNLLKSL
3895	4	LSPVIGNLLKSLG
3896	4	LLPVIGNLLKSL
3897	4	LLPVIGNLLKSSLG
3898	4	FVQWFSKFLGRIL
3899	4	FVQWFSKFLGRILG
3900	4	FIGASALKVLAGVLP SVISWVKLGRILQ
3903	4	ELCEKASKTWSGNC GNTGHCDNQCKSW EGAAHGACHVRNG KHMCFYFNC
3904	4	EVCEKASKTWSGNC GNTTGHC
3905	4	DGVKLCDVPSGTWS GHCGSSSKCSQCK DREHFAYGGACHYQ FPSVKCFCKRQC
3906	4	RNCESLSHRFKGPCT RDSNC
3907	4	RVCMGKSAGFKGLC MRDQNAQVCLQE GWGGNCDGVMR QCKCIRQCW

3908	4	KICRRRSAGFKGPPC MSNKNCAQVCQEQE GWGGNCDGPFRR CKCIRQC
3909	4	RICRRRSAGFKGPCV SNKNCAQVCMQEQE WGGNCDGPLRRC KCMRRC
3943	4	GFGCPNNYQCHRH CKSIPGRCCGYCGG WHRLRCTCYRGC
3945	4	GCASRCKAKCAGR CJGWASASFRGRCY CKCFRC
3946	4	SCASRCKGHCRARR CGYYVSVLYRGRCY CKCLRC
3947	4	SCASRCKSRARR CRYVSVRYGGFCY CRC
3948	4	GCASRCKAKCAGR CKGWASASFRRCY CKCFRC
3949	4	VVTCGSLCAHCTF RKCGYFMAVLYHGR CYCRLLC
3950	4	HSHACTSYWCGKFC GTASCTHYLCRVLH PGKMCACVHCSR
3961	4	LKKISQ
3967	4	FLSFPTTKTYFPFDL SHGSAQVKGHGAK
3991	4	APRKNVRWCTISQPE W
3992	4	CIRA
3993	4	FKCRRWQWRMCKL GAPSITCVRRFAFALE CIR
3994	4	APRKNVRWCTI
3995	4	TKCFQWQRNMRKV RGPPVSCIKR
3997	4	YCNSSTRAFDCLG QCGRCDFHKLQCV H
3999	4	YCNSSTRAFDCLG QCGRCDFHKLQCV H
4000	4	YCNSSTRAADCLG QCGRCDFHKLQCV H
4009	4	EQLTK
4010	4	GYGGVSLPEWVCTT F
4011	4	ALCSEK
4012	4	CKDDQNP

4013	4	ISCDKF
4014	4	VAGTWY
4015	4	GLFDVIKKVASVIGG L
4016	4	GLFDVIKKVASVIGG LG
5291	4	FIGASALKVLAGVLP SVISWVKLGRILQG
5292	4	LCNERPSQTSWGN GNTAHCDKQCQDW EKASHGACHKRENH WKCFYFNC
5293	4	NLCERASLTWTGNC GNTGHCDTQCRNW ESAKHGACHRGNW KCFCYFNC
5297	4	RVCMGKSAGFKGLC MRDQNC AQVCLQE GWGGGNC DGVMR QCKCIRQCW
5330	4	GFGCPNNYACHQH CKSIRGYCGGYCAG WFRLRCTCYRCG
5337	4	HPHVCTSYYCSKFC GTAGCTRYGCRNLH RGKLCFCLHCSR
5338	4	LTCEIDRSLCLLHCR L KGYLRAYCSQQKVC RCVQ
5339	4	DQMSDGVNYGKGS SLSKGGAKCLGIVG GLATIPSGPLGWLAG AAGVINSCKM
5340	4	AISYGNVYCNKEK CWWNKAENKQAITG IVIGGWASSLAGMG H
5341	4	VNYGNVSCSKTKC SVNWGQAFQERYTA GINSFVSGVASGAGS IGRRP
5343	4	KYYGNVSCNKNKC TVDWSKAIGIIGNNA AANLTTGGAAGWN KG
5344	4	ARSYGNVYCNKK CWWNRGEATQSIIGG MISGWASGLAGM
5345	4	ILPWKWPWWPWRR VPLPNVPQGRPPFP TFPGQGPFPNPKIKWP QGY
5346	4	FVPYNNPPRPGQSKPF PSFPGHGPFNPKIQW PYPLPNPGH

5348	4	GYVSGAVIEIPDEILD SARFISLYSDGLRQK RQLNLSGPGSEHAG TIRLDGQRNIFDNGR TRVDGTGSYQLDYA RGMKPIHGAGLGAE VNHNWIRGRGGQSL DLYGGATRQFNFGN RPNEWGAHGGIRYN F
5349	4	SLQPGAPNVNNDQ PWQVSPHISRDDSG NTRTDINVRHGEN NDFEAGWSKVVRGP NKAKPTWHIGGTHR W
5351	4	VTCDLLSFKGQVND SACAANCLSLGKAG GHCEKVGICRKTFS KDLWDKRFG
5352	4	QKLCERPSGTXSGVC GNNNACKNQ CIR
5353	4	DGIWTLQIFLTKNL ATLWQSGDFQFLDH ECHYRIKPTFRRLKW KYKGFWCPSWTSIT GRATKSSRSGAVEHS VRNFVQAKSSGLIT QRQAEQFISQYN
5354	4	GRKSDCFRKSFGCAF LKCPSLTLISGKCSRF YLCKRIR
5355	4	SAFTVWSGPGCNR AERYSKCGCSAIHQK GGYDFSQYGTAAAL YNQAGCSGVAHTRF GSSARACNPFGWKSI FIQC
5356	4	GKREKCLRRNGFCA FLKCPYTLISGTCR FQVCC
5357	4	GRKSDCFRKNFCA FLKCPYTLISGLCSX FHLC
5358	4	QAEESNLQSLVSQYF QTVADYDKDLVEKA KSELQTOAKAYFE KTQEELTPFFKAGT DLLNFLSSFIDPKKQP AT
5359	4	GLLSRLRDFLSDRGR RLGKIERIGQKIKDL SEFFQS
5360	4	GWLKKIGKIERVGG HTRDATIQTIAVAQ QAANVAATARG

5361	4	SWLSKTAKKLENSAK KRISERIAIAIQGPR
5362	4	WNPFKELERAGQRV RDAVISAAPAVATV GQAAAIARG
5363	4	RWKIFKKIEKVGQNI RDGIVKAGPAVAVV GQAATIG
5364	4	RWKLKFKIEKVRNV RDGLIKAGPAIAVIG QAASLG
5365	4	GWLRKLGKKIERIGQ HTRDASIQLGIAQQ AANVAATARG
5366	4	KWKIFKKIEKVRNIR NGIIRKAGPAVAVLGE AKALG
5367	4	GWLKKLGKRIERIGQ HTRDATIQGLGIAQ QAANVAATARG
5368	4	GWLRKIGKIERVGG HTRDATIQVGLGIAQ QAANVAATARG
5369	4	GWIRDFGKRIERVGG HTRDATIQTIAVAQ QAANVAATLKG
5370	4	MASSFMFAVVVLFIS LAANVESYLAFCRGR YSPCLDDGPNVNL SCCSFYNCHKCLARL ENCPKGLHYNAYLK VCDWPSKAGCTSVN KECHLWKTGRK
5371	4	GLWSKIEVKGKAAK AAKAAGKAALGAV SEAVG
5372	4	ALWKNMLKIGIGKLA GQAALGAVKTLVGA EG
5373	4	ALWKDILKNVKGAA GKAVLNTVTDMVN QG
5374	4	GLWNKIEAASKAA GKAALGFVNEMVG
5375	4	GIMSIVKDVAKTAAK EAAKALSTLSCKLA KTC
5376	4	GILDTLKQFAKGVGK DLVKGAAQGVSTV SCKLAKTC
5377	4	FLGALFKVASKVLPS VFCAITKCC
5378	4	FLPLLAGLAANFLPTI ICKISYKC
5379	4	FLPAIFRMAAKVVPTI ICSITKCC

5380	4	VIPFVASVAEMQHV YCAASRKC
5381	4	GILDTLKNLAISAAK GAAQGLVNKASCKL SGQC
5382	4	GILDTLKNLAKTAGK GALQGLVKMASCKL SGQC
5383	4	GILLDKLKNFAKTAG KGVLSLLNTASCKL SGQC
5384	4	GILDSLKNLAKNAG QILLNKASCKLSGQC
5385	4	GILSLVKGVAKLAGK GLAKEGGKFGLELIA CKIAKQC
5386	4	GIFSLVKGAAKLAGK GLAKEGGKFGLELIA CKIAKQC
5387	4	GIFSKLAGKLLKNLLI SGLKNVGKEVGMDV VRTGIDIAGCKIKGE C
5388	4	GWMSKIASGIGTFLS GIGQQG
5389	4	AMWKDVLKKGITVA LHAGKAALGAVADT ISQG
5390	4	ALWKDILKNAGKAA LNEINQLVNQG
5391	4	FLGGLIKIVPAMICAV TKKC
5392	4	GLLNTFKDWAISIAK GAGKGVLTLSCKL DKSCG
5393	4	SLFSLIKAGAKFLGK NLLKQGAQYAACKV SKECG
5394	4	GILDSFKQFAKGVGK DLIKGAAQGVLSTM SCKLAKTC
5395	4	SLFSLIKAGAKFLGK NLLKQGACYAACKA SKQC
5396	4	GIMSIVKDVAKNAAK EAAKGALSTLSCKLA KTC
5397	4	GIMDTLKNLAKTAG KGALQSLVKMASCK LSGQC
5398	4	FLPVLAGIAAKVPPA LFCKITKKC
5399	4	GLLDSLKGFAATAG KGVLSLLSTASCKL AKTC

5400	4	GIFDKLKNFAKQVA QSLLNKASCKLSGQ C
5401	4	GIFSKLAGKKIKNLLI SGLKNVGKEVGMDV VRTGIDIAGCKIKGE C
5402	4	GIFSKLGRKKIKNLLIS GLKNVGKEVGMDVV RTGIDIAGCKIKGEC
5403	4	GRPNPVNNKPTPHP RL
5404	4	RSGRGECLRR HEGQPWETQECMRR CRRRG
5405	4	FLPLLAGLAANFFPKI FCKITRKC
5406	4	GWLKKGKIERVVGQ HTRDATIQGLGIAQ QAANVAATARG
5407	4	GWLKKGKIERVVGQ HTRDATIQVIGVAQ QAANVAATARG
5408	4	GIFSSRCKTSPKTFK GICTRDSNCDTSCR YEGYPAGDCKGIRRR CMCSKPC
5409	4	GIFSSRCKTVSKTF RGICTRNANC
5410	4	MFFSSKCKTVSKTF RGPCVRNA
5411	4	CRRWQWRMCKLGA PSITCV
5412	4	ALWFTMLKKGTLTMA LHAGKAALGAAANT ISQGTQ
5413	4	ALWKNMLKGIGKLA GKAALGAVKLVGA ES
5414	4	ALWMTLLKVLKAA AKALNAVLVGANA
5415	4	GLWSKIKTAGKSVAK AAAKAAVKAVTNAV
5416	4	INLKALAALAKKILG
5417	4	KWKLFFKIGAVLKVL EKCLRWNEMRKG G
5418	4	SKCRQWSKIRRTNP
5419	4	TKCFQWQWNMRKV RG
5420	4	GRPNPVNTKPTPYPR L
5421	4	GKPRPYSRPTSHPR PIRV
5422	4	

5423	4	DLRFLYPRGKLPVPT PPPFNPKPIYIDMGN RY
5424	4	DLRFLYPRGKLPVPT LPPPFNPKPIYIDMGN RY
5425	4	KSCCPTTAARNQYN ICRLPGTPRPVCAAL SGCKIISGTGCPPGY RH
5426	4	KSCCPSTTARNIYNT CRLTGTSRPTCASLS GCKIISGSTCBSGWB H
5427	4	KSCCPBTGRBIYBT CRFGGSRZVCARIS GCKIISASTCPSYPBK
5428	4	KSCCPNTTGRNIYNT CRFGGSRVPCASLS GCKIISASTCPSYDCK
5429	4	KSCCPNTTGRNIYN ACRLTGAPRPTCAK LSGCKIISGSTCPSDY PK
5430	4	MFFSSKCKTVXKTF RGPCVRNAN
5431	4	GIFSNMYXRTPAYGF RGPXGGYXXN
5432	4	GIFSSRCKTSPKTFK GYCTRDSNCDTSCR YEGYPAGD
5433	4	XTCESPSHKFKGPCA TNRNCES
5434	4	TKCFQWQRNMRKV R
5435	4	CFQWQRNMRKVRG PPVSCI
5436	4	FKCRRWQWRMCKL G
5437	4	IWLTALKFLGKHAAK HLAKQQLSKL
5438	4	KIKWFKTMKSIKAFIA KEQMKKHLGGE
5439	4	YPGPQAKEDSEGPS QGPASREK
5440	4	LPVNSPMNKGDTTEV MKCIVEISDTLSKPS PMPVSKECFETLRGD ERILSILRHQNLKEL QDLALQGAKERTHQ Q
5441	4	LRKAQEKFGKNKSQ SFQL
5442	4	PEWSKCYQWQRRM RKLGAAPSITCVRRTS A

5443	4	SKCYQWQWRMRKL GA
5444	4	EKCLRWQWEMRKV GG
5445	4	SKCRQWQWKIRRTN P
5446	4	GLLDGLLGTLLGLG
5447	4	GLLDGLLGGGLGLG
5448	4	GLLDVMTLLGNLGLG
5449	4	GLLDVVGNTLLGGLG LG
5450	4	GLFDAIGNLLGGLGL G
5451	4	GLLDVVGNTLHSLG LG
5452	4	GLLGMVGSLLGGLG LG
5453	4	GLLGFVGSLLGGLGI G
5454	4	GLFDIHKIAESIG
5456	4	GLLDIVKKVVGAFGS LG
5457	4	LDIVKKVVGAFGSLG
5458	4	GLFDIVKKVVGALGS LG
5459	4	GLFDIVKKVVGAGIS LG
5461	4	FDIVKKVVGTIAGLG
5462	4	FDIVKKIAGHIAGSIG
5465	4	FDIVKKIAGHIVSSIG
5466	4	QGVNRHVTCRINRG FCVPIRCPGRTRQIG TCFGRPKCCRSW
5467	4	QGVNRHVTCRIYGG FCVPIRCPGRTRQIG TCFGRPVKCCRRW
5469	4	VYQHQAAMKPIQIP KTKVIPYVRYL
5470	4	QGVNFVTCRINRG FCVPIRCPGHRRQIG TCLGPRKCCCR
5471	4	GLFDIVKKVVGAFGS LG
5472	4	GLFDIAKKVIGVIGSL G
5473	4	GLFDIVKKIAGHIAGS IG
5474	4	GLFDIVKKIAGHIA
5475	4	GLMSSIGKALGGLIV DVLKPKTPAS
5476	4	IIGGR
5477	4	HPQYNQR

5478	4	DFASCHTNGGICLP NRCPGHMIQIGICFR PRVKCCRSW
5479	4	VRNHVTCRINRGFC VPIRCPGRTRQIGTC FGPRIKCCRSW
5480	4	VRNFVTCRINRGFCV PIRCPGHRRQIGTCL GPQIKCCCR
5481	4	QGVRSYLSWCWGNRG ICLLNRCPRMRQIG TCLAPRVKCCCR
5482	4	GPLSCRRNGGVCPI RCPGPMRQIGTCFG RPVKCCRSW
5483	4	GPLSCGRNGGVCPI RCPVPMRQIGTCFG RPVKCCRSW
5484	4	SGISGPLSCGRNGGV CIPRCPVPMRQIGTC FGRPVKCCRSW
5485	4	SIGSAFKKALPVAKKI GKAALPIAKAALP
5486	4	SLGGVISGAKKVAKV AIPIGKAVLPVAKLV G
5487	4	SIGTAVKKAVPIAKK VGKVAIPIAKAVLSV VGQLVG
5488	4	VFQFLGKIIHHVGNF VHGFSHFV
5489	4	VFQFLGRIIHHVGNF VHGFSHFV
5490	4	GWFGKAFRSVSNFY KKHKTYIHAGLSAAT LLGDMTDEEFQEFM QDIEQAREEELLSRQ
5491	4	GNNRPIYIPQPRPPH RL
5492	4	GNNRPVYIPQPRPPH PRL
5493	4	DDMTMKPTPPPQYP LNLQGGGGGQSGD GFGFAVQGHQKVV TSDNGRHEIHLNGG YQHLGGPYGNSEP SWKVGSTYTRFPNF
5494	4	DSHEKRHHGYRRKF HEKHHSHREFPFY DYGSNYLYDN
5495	4	RKFHEKHHSHREFPF YGDYGSNYLYDN
5496	4	HRHQGPIFDTRPSPF NPNQPRPGPIY

5497	4	MPCSCKKYCDPWEV IDGSCGLFNSKYICC REK
5498	4	GRKSDCFRKSFGCAF LKCPSTLISGKCSR YLCKRIWG
5499	4	ATCDLLSGTGINHSA CAAHLLRGNRGGY CNGKAVCVCRN
5500	4	DCLSGRYKGPCAVW DNETCRRVCKEEGR SSGHCSPLKWCCE GC
5501	4	AVDFSSCARMDVPG LSKVAQGLCISSCKF QNCGTGHCEKRRG RPTCVCDRCGRGGG EWPSVPMKGRSSR GRRHS
5502	4	ATCDLLSGTGINHSA CAAHLLRGNRGGY CNGKGVVCVCRN
5503	4	ATCDLLSGTGINHSA CAAHLLRGNRGGY CNRKGVVCVCRN
5504	4	YRGGYTGPIRPPPIG RPPLRLVVCACYRLS VSDARNCCIKFGSCC HLVK
5505	4	YRGGYTGPIRPPPIG RPPFRVCNACYRLS VSDARNCCIKFGSCC HLVKG
5506	4	QVRNPQSCRWNM GVCIPFLCRVGMROI GTCFGRPVCCRR
5507	4	QVVRNPQSCRWNM GVCIPISCPGNMRI GTCFGRPVCCRRW
5508	4	QGVNFVTCRINRG FCVPIRCPGHRRQIG TCLAPQIKCCCR
5509	4	SIGSALKKALPVAKKI GKIALPIAKAALP
5510	4	VFHLLGKIIHHVGNF VYGFHFV
5511	4	AFKLLGRIIHHVGNF VYGFHFVFG
5512	4	GLLRKGGGKIGEKLK KIGQKIKNFFQKLV QPEQ
5514	4	RQRVEELSKFSKKA AARRRK
5515	4	GXFGKAFXSVSNAK KHKTA

5516	4	GXFGPAFHSVSNFAK KHKTA
5517	4	AFPPPNVPGPRFPPP NFPGPRFPPNFPGP RFPPNFPGPRFPPN FPGPPFPPFPGPWF PPPPFRPPFGPPRF P
5518	4	VDKGSYLPRPTPPRPI YNRN
5519	4	RRRPRPPYLPRRPPP FFPRPRLPRIPPFP FPPRFP
5521	4	RSVCRQIKICRRRGG CYYKCTNRPY
5522	4	RGGRLCYCRRRFCV CVGR
5523	4	AGRKQGGKVRRAK AKTRSSRAGLQFPVG RVHLLRKGNY
5524	4	TRSSRAGLQFPVGRV HLLRK
5525	4	DTHFPICFCCGCCH RSKCGMCKCT
5526	4	ICIFCCGCCHRSKCG MCKCT
5527	4	QVYKGGYTRPIRPP PFVRPLGGPIGPYN GCPVSCRGISFSQAR SCCSRLGRCCHVGK GYSG
5528	4	QVYKGGYTRPVPRP PPFVRPLGGPIGPYN GCPVSCRGISFSQAR SCCSRLGRCCHVGK GYSG
5529	4	QVYKGGYTRPIRPP FVRPVGGPIGPYNG CPVSCRGISFSQARS CCSRLGRCCHVGK YSG
5530	4	DPTCSVLGDFKCNP GRCCSKINYCGACY QWRFGTPAR
5531	4	CPNNLCCSQWGC G
5532	4	FALLGDFFRKSKEKIG KEFKRIVQRIKDFLRN LVPRTES
5533	4	LLGDFFRKSKEKIGKE FKRIVQRIKDFLRNLV PRTES
5534	4	LCNERPQTSWGN GNTAAHCDKQCQD WEKASHGACCHKRE NHWKCFYFNC

5535	4	GLFGRLRDSLQRGG QKILEKAERIWCICKD IFRG
5536	4	GLRKRLRKFNRKIKE KLKKGQKIQGFVPK LAPRTDY
5537	4	SLQPGAPSPMPGSQ LPTSVSGNVEKQGR NTIATIDAQHKTDY DVRGTWTKVVDGP GRSKPNFRIGGSYRW
5538	4	YGPGDGHGGGGHGG GHGGGHNGQGGG HGHGPGGGGGGGH GGGGGGGGGGGG SGGGGSPGHGAGG GYPGGGGGHHGG YQTHGY
5539	4	NLCERASLTWTGNC GNTGHCDTQCRNW ESAKHGACHKRG WKCFYFNC
5542	4	DGVKLCDVPSGTWS GHCSSSKCSQCC KDREHFAYGGACHY QFPSVKCFCKRQC
5543	4	RGLRRLGRKIAHGK KYGPTVLRIRIAG
5544	4	GILDTIKSIASKVWNS KTVQDLKRKGINWV ANKLGVSPQAA
5545	4	GRFKRFRKFKKLFK KLSPVILLHLG
5546	4	GGLRSLGRKILRAWK KYGPIIPIRIG
5547	4	AGLFRRLRDSIRRGQ QKILEKARRIGERIKD IFRG
5548	4	SLQGGAPNFPQPSQ QNGGWQVSPDLGR DDKGNTRGQIEIQN KGDHDFNAGWGK VIRGPNKAKPTWHV GGTYRR
5549	4	QERGSIVIQGTKEGK SRPSLDIDYKQRVYD KNGMTGDAYGGLNI RPGQPSRQHAGFEF GKEYKNGFIKQSEV QRGPGGRLSPYFGIN GGFRF
5550	4	RFRPPIRRPPIRPPFYP PFRPPIRPPFPIRPPF RPPLGPPF

5551	4	RFRPPIRRPPIRPPFNP PFRPPVPPFRPPFRP PFRPPIGFP
5552	4	RRIRPRPRLPRRPR PLPFRPGPRPIRPLP FPRGPRPIRPLPFP RPGPRPIRPL
5553	4	RRLRPRRRLPRRPR PRRPRSLPLRPQPR RIPRILLPWRPPRPI RPQPPIRPL
6985	4	ATCDLLSGIGVQHS CALHCVFRGNRGGY CTGKICVCRN
7770	4	GSKKPVIIYCNRR GKCQRM
7775	4	QLKTADLPAGRDET TSFVLV
7776	4	QGRVEVLYRGSWGT VC
7777	4	SGSLSTFFRLEFRSFT QALGK
7778	4	TFRLEFRSFTQALG KGGGNLRIIRKGIHI IKKY
7779	4	TFRLEFRN
8161	4	LKKISQRYQKQFALPQ Y
8162	4	QELLNPTHQYQVPT QPLAPVHNPIV
8163	4	GLPQE
8164	4	RPKHPIK
8192	4	TSKYR
8194	4	MAIPPKNQDKTEIP TINTIASGEPSTPTT EAVESTVATLEDSPE VIESPPEINTVQVTST AV
8195	4	YQRRPAIANNPYVP RTYYANPAVVRPHA QIPQRQYLPNSHPPT VRRPNLHPSF
8196	4	IPAVFK
8197	4	AASDISLLDAQSAPL R
8198	4	VLVLDTDYK
8255	4	KTVYQHQAAMKPM QPKTKVIPYVRYL
8256	4	LKTVYQHQAAMKPM WIQPKTKVIPYVRYL
8257	4	ALPQYLKTVYQHQA AMKPMWIQPKTKVIPY VRYL

8258	4	QKFALPQYLKTVYQ HQKAMKPWQPKTK VIPYVRYL
8259	4	LKKISQRYQKFALPQ YLKTVYQHQAAMKP WQPKTKVIPYVRYL
8260	4	LKKISQYYQKFAWP QYL
8261	4	VDQHQAAMKPWTQ PKTNAIPYVRYL
8262	4	PYVRYL
8263	4	YYQKQKPA
8264	4	VQVTSTAV
8265	4	STVATL
8266	4	FSDKIAK
8267	4	IQY
8268	4	YVL
8269	4	EIPT
8270	4	VESTVATL
8271	4	PAAVRSPAQILQ
8272	4	DNIADAVACAQRV RDPQGIRAWVAWRN R
8273	4	PEWSKCYWQRRM RKLGAISITCVRRTS A
3039	5	GPRPP
3040	5	GPRG
3041	5	GPRGP
3042	5	GPRGPA
3043	5	GPRP
3044	5	GPRGPP
3047	5	GPR
3121	5	DGEA
3122	5	KDGEA
3123	5	PAGKNGETGPQG
3124	5	GPAGKDGEAGAQG
3125	5	RGDSPASSKP
3126	5	RGDS
3266	5	KKQRFHRNRQ
3268	5	TAQVTSTEV
3269	5	QVTSTEV
3270	5	KDQDK
3271	5	VTGRGDSPASSKPI
3272	5	VTRGDSPA
3273	5	GRGDS
3274	5	VTRGD
3275	5	RGDSPA
3276	5	KPGEFGPK
3277	5	KREE

3278	5	GHRP
3279	5	ACRGGWCG
3280	5	ACDCRGDCFCG
3281	5	CRRETAWAC
3282	5	NGRAHA
3283	5	GP
3284	5	PGP
3285	5	PG
3287	5	RGDF
3288	5	LGGAKQAGDV
3289	5	NIMEILRGDF
3290	5	KRDS
3291	5	NQDK
3292	5	MAIPPKKNQDK
3293	5	MAIPPK
3294	5	MAIPPKK
3352	5	WGC
3353	5	SSGE
3354	5	DEE
3462	5	GPGG
3468	5	GRGDSPK
3549	5	SFLLRNPNDKYEFP
3990	5	GAQCTAGPCCWPCE GTICRRARGDDLDD YCNGISADCPNPYY
4017	5	ECESGPCCRNCKFLK EGTICKRARGDDMD DYCNGKTCDCPRNP HKGPAT
4018	5	EAGEDCDGSPANP CCDAATCKLIPGAQ CGEGLCCDQCSFIEE GTVCRIARGDDLDD YCNGRSAGCPNPF H
4019	5	GKECDCSSPENPCC DAATCKLRPGAQCG EGLCCEQCKFSRAG KICRIPRGDMPDDRC TGQSADCPRYH
4020	5	VSPVCGNKILEQGE DCDCGSPANCQDQ CCNAATCKLTPGSQ CNHGECCDQCKFK KARTVCRIARGDWN DDYCTGKSSDCPWN H
7756	5	GELTPESGPDLFVHF LDGNPSYSLYADAVP R
8165	5	KNQDK
8166	5	HHLGGAKQAGDV

8167	5	RGDX
8168	5	PPK
2575	6	TKPR
2840	6	GYPMYPLPR
2882	6	YG
2934	6	LVCYPQ
2935	6	LVVYPW
2936	6	FRPRIMTP
2937	6	FLGFPT
2938	6	VDPP
2939	6	VVYPD
3027	6	TPRK
3055	6	RELEELNVPGEIVESL SSSEESITR
3056	6	GKVLKRR
3057	6	YGKVLKRR
3059	6	YQQPVLGVRGPFPII V
3060	6	SIKVAV
3061	6	GFL
3062	6	RETIESLSSEESIPEY K
3063	6	QPTIPFFDPQIPK
3064	6	VEPIPY
3066	6	EAE
3083	6	DSDPR
3095	6	RKEVY
3105	6	ASHLGLAR
3119	6	PGIPN
3213	6	YIPIQYVLSR
3223	6	QRRQRKSRRTI
3228	6	AGIYGTKESPQTHYY
3252	6	YQQPVLGPVR
3469	6	VGGIPY
3470	6	AEPIPY
3562	6	RKNVY
3584	6	GRKSSK
3608	6	RTAKV
3616	6	KRLDG
3618	6	KTPKI
3620	6	GRKP
3621	6	RKPK
3622	6	KTPK
3623	6	TKPI
3624	6	TRKP
3625	6	RKP
3626	6	KRP
3628	6	RKPKV
3653	6	DKPTL

3654	6	DETRL
3655	6	IKPR
3656	6	VTDFYF
3657	6	VTKFYFQ
3658	6	ITDFQI
3659	6	ITDFQILE
3660	6	EIPKTITGSETNL
3662	6	ITGSE
3691	6	VTKFYF
3692	6	EIPKT
3693	6	TKPIE
3699	6	NRKQDK
3700	6	RKQDKR
3701	6	ENRQDKRF
3702	6	VTKFY
3703	6	VTDFY
3734	6	RKSSK
3735	6	RKQDK
3739	6	TKRP
3741	6	YGG
3783	6	HCQRPR
3793	6	QRPR
3864	6	PKNYP
8170	6	KNTMEHVSSEESIIS QETYKQEKMAINP SK
8171	6	RPKHPIKHQGLPQEV LNENLLRF
8172	6	TTMPLW
8173	6	RELEELNVPGEIVESL SSSEESITRINK
8174	6	LLYQEPVLGPVRGPF PIIV
8175	6	FKCRRWQWRMCKL GAPSITCVRRAF
8176	6	VESYVPLFP
8177	6	YKPR
2570	7	VVYPWTQRF
2571	7	VVYPWTQ
2841	7	FGGFTGARKSA
2842	7	FGGFTGARKSARKL ADE
2843	7	FGGFTGA
2844	7	YGGFTGARKSA
2845	7	FGGFTGARKSY
2846	7	FGGFTGARKYA
2847	7	YPFT
2848	7	YGGFLRARKSARKLA NQ

2849	7	YGGFLRRIRFARKLA NQ
2850	7	FGGFTGRIPKLWDN Q
2851	7	YGGFLRRIRPKLKLA NQ
2852	7	YGGFTGARKSARKL ANQ
2853	7	YGGFLGARKSARKL ANQ
2854	7	YGGFLRRIRPKRCLA NQ
2855	7	FGGFTRRIRPKLKW NQ
2856	7	YGGFLRRIKSARKLA NQ
2858	7	YGGFLRRIRPKLKW NQ
2859	7	YGGFLRRKSARKLA NHQ
2860	7	FGGFLRRIRPKLKW NQ
2861	7	YVPPF
2862	7	LVVYPWTQR
2868	7	YPPF
2869	7	YPPFGPIP
2870	7	YPF
2871	7	KKGE
2872	7	NAGA
2874	7	YIFHLMN
2875	7	YIFHLM
2876	7	YAFGYPS
2878	7	YGGFLR
2880	7	YGGFLRKYP
2881	7	YGGFLRKYPK
2973	7	LVVYPWTQRF
2974	7	LVVYPWTQ
3078	7	YPWTQRFF
3079	7	YPWTQRF
3080	7	YPWTQR
3081	7	YPWTQ
3082	7	YPWT
3087	7	YPISL
3088	7	YGGWL
3089	7	YGGW
3090	7	GYPT
3091	7	GYYP
3093	7	YPLG
3094	7	PLG
3127	7	TTMPLW
3186	7	GGFLTRH

3198	7	GICACRRRFCPNSER FSGYCRVNGARYVR CCSRR
3214	7	YPSYGLN
3218	7	YPYY
3250	7	YPSF
3265	7	YLGYLE
3472	7	GLF
3479	7	YGGFMRR
3574	7	YPPGFF
3587	7	WWPK
3588	7	WWPKHG
3592	7	YPPGFE
3599	7	YPPGFW
3600	7	YPPGFCQ
3602	7	YPPGFD
3606	7	YPPGFV
3629	7	YPPGFS
3630	7	YPPGFL
3631	7	YPPGFT
3632	7	YPPGFI
3633	7	YPPGFY
3634	7	YPPGFH
3635	7	YPPGFM
3636	7	YPPGFQ
3637	7	YPPGFP
3639	7	YPPGFA
3641	7	YPPGFK
3642	7	YPPGFG
3643	7	YPPGFR
3650	7	YPPGFN
3686	7	TSKYR
3695	7	YPPGFGG
3697	7	YPPGFNG
3698	7	YPPGFAG
3704	7	YPPGFQG
3705	7	YPPGFKG
3706	7	YPPGFMG
3707	7	YPPGFRG
3716	7	YPPGFSG
3717	7	YPPGFDG
3719	7	WWPRG
3720	7	WWPKG
3721	7	WWPKHHG
3722	7	WWPKHKG
3723	7	WWPKHNG
3724	7	WWPKHGG
3725	7	YPPGFEG
3726	7	YPPGFPG

3727	7	YPFGFCQG
3729	7	YPFGFFG
3730	7	YPFGFVG
3731	7	YPFGFLG
3732	7	YPFGFTG
3733	7	YPFGFIG
3736	7	YPFGFYG
3737	7	YPFGFHG
3738	7	YPFGFWG
3784	7	YYPT
3785	7	YYP
3786	7	GY Y
3787	7	YPVSL
3788	7	YPLSL
3789	7	YPASL
3790	7	YPTSL
3791	7	YPFSL
3794	7	RYYP
3795	7	WWPR
3796	7	WWPKHH
3797	7	WWPKHK
3798	7	WWPKHN
3804	7	YPWSL
3815	7	WYYP
3817	7	SYYP
3825	7	GYPTS
7753	7	YPFVV
7754	7	YPFVVN
7755	7	YPFVNA
7846	7	YPFW
7847	7	YPFWG
7848	7	YFFF
7849	7	YFFFG
7850	7	YGGFMTSEKSQTPLV TLFKNAIKNAYKKG E
7851	7	YGGFMRF
7852	7	YGGFMRGL
7853	7	YGGFM
7854	7	YGGFMRRV
7855	7	YGGFMRRVG
7856	7	YGGFLRRQQFKVVT FGGFTGARKSARKL
7857	7	ANQ
7858	7	TEPGLEEVGEIEQKQ LQ
2676	8	GGVYACRMGPITWV CSPLGG
2677	8	GGCRIGPITWVCGG

2678	8	GGTYSCHFGPLTWV CKPQGG
2679	8	GGDYHCRMGPLTW VCKPLGG
2680	8	VGNYMCHFGPITWV CRPQGG
2682	8	DLEGCRLGWVGHCN VWGGDEYTKRTSHS VPPSHKSKLL
2683	8	GGTYSCHFGPATWV CKPQGG
2684	8	GGTYSCHFGALTWV CKPQGG
2685	8	GGTYSCHFGPLTFVC KPQGG
2686	8	GNADGPTLRQWLEG RRPKN
2687	8	GGCADGPTLREWISF CGG
2689	8	GGLYLRCRFGPVTWD CGYKGG
2694	8	GGCTLREWLHGGFC GG
2695	8	TIKGPTLRQWLKSRE HTS
2696	8	IEGPTLRQWLAARA
2697	8	GRVRDQIMLSLGG
2698	8	LAIEGPTLRQWLHGN GRDT
2699	8	SFLLRN
3096	8	CSRARKQAASIKVAV SADR
3350	8	VPL
3351	8	EEE
3355	8	SSS
3356	8	LLL
3555	8	FLGFPTTKTYFPHF
8320	8	VL
8321	8	LV
8322	8	IV
8323	8	IL
8324	8	LI
8325	8	II
8326	8	LL
8329	8	EE
8330	8	SE
3054	9	RKD VY
3065	9	LLY
3098	9	GVM
3099	9	LG Y
3120	9	KEEAE
3613	9	RKLDG

3614	9	RTPKV
3615	9	RNVRV
3708	9	RNKPFK FVL
3709	9	SSRDRNKPFK FML
3710	9	RNKPFK FML
3711	9	SSRDRNKPFK FVL
3984	9	GMNRSPIL
3985	9	GMNRRPIL
2572	10	CYFQNCPRG
2573	10	CYFQNCPRGG
2574	10	YKPR
2601	10	APSGFLGMRG
2602	10	ASFSPWG
2603	10	ASFSPWGG
2604	10	APSGFLGM R
2700	10	AGCKNFFWKTFTSC SANSNPAMAPRERK AGCKNFFWKTFTSC
2701	10	AGCKNFFWKTFTSC
2790	10	KRQH PGKR
2791	10	KRQH PG
2835	10	SYIQNSPLGG
2836	10	CYIQNCPLGG
2837	10	CYIQNCPLG
2885	10	KQKLQ
2886	10	EIEQKQLQ
2887	10	EVQKQLQ
2888	10	AEPGADDAEEVEQK QLQ
2889	10	FGGFTGARKSARKL ANQ
2892	10	GYNRSFLRF
2893	10	FMR F
2894	10	NRNFLRF
2895	10	DRNFLRF
2896	10	KNEFIRF
2897	10	TNRNFLRF
2899	10	NYDKNFLRF
2900	10	ADKNFLRF
2903	10	ANRSPSLRLRFG
2904	10	GGRSPSLRLRFG
2905	10	DRTPALRLRFG
2906	10	APALRLRFG
2907	10	NYDKNFLRFG
2909	10	DRNFLRFG
2911	10	SDRNFLRF
2914	10	NRNFLRFG
2915	10	GYNRSFLRFG
2916	10	TNRNFLRFG
2917	10	SDRNFLRFG

2918	10	KNEFIRFG
2919	10	FMRFG
2920	10	ADKNFLRFG
2921	10	GGRSPSLRLRF
2923	10	APALRLRF
2924	10	DRTPALRLRF
2925	10	ANRSPSLRLRF
2930	10	YGGFLGARKSARKL ANQ
2931	10	TEPGLEEVGEIEQKQ LQ
2932	10	CPTATATALAPRGG SRV
2940	10	FGGFTGRIRPKLKW NQ
2941	10	FGGFTGARKSKLKW DNQ
2942	10	YGGFTGARKSARKL ANQ
2943	10	FGGFTGARKSARKL DNQ
2944	10	FGGFTGARKSALKW DNQ
2945	10	FGGFTGARRPKLWD NQ
2946	10	FGGFTGARRPKLKW DNQ
2947	10	FGGFTGARKSARKW DNQ
2978	10	IARRHPYFL
3092	10	KRNKNNIA
3100	10	DAEFRHDSGYQVHH QKLVFFAEDVGSNK
3101	10	GSNKGAIIGLM
3102	10	VHHQKLVFFAEDVG SNK
3103	10	DAEFRHDSGYEVHH QKLVFFAEDVGSNK
3104	10	DAEFRHDSGYQVHH QK
3749	10	KPS
3760	10	KPT
4001	10	FFGLM
4002	10	RPKPQQFFGLM
4003	10	FFGLMG
4004	10	RPKPQQFFGLMG
7703	10	HVVNFTPRL
7704	10	HVVNFTPRLG
7705	10	SPPFAPRL
7706	10	SPPFAPRLG
7707	10	HLSPFSPRL
7708	10	HLSPFSPRLG

7709	10	DWNKDLHIW
7710	10	DWNKDLHIWG
7711	10	GWNNLHEGW
7712	10	GWNNLHEGWG
7713	10	AWQSLQSGW
7714	10	AWQSLQSGWG
7715	10	NWQGFHGGW
7716	10	NWQGFHGGWG
7717	10	SKWDNFRGSW
7718	10	SKWDNFRGSWG
7719	10	EPAWSNLGIW
7720	10	EPAWSNLGIWG
7721	10	GIEALKYHNMDLGT ARGY
7722	10	GIEALKYHNMDLGT ARGYG
7723	10	NKLASVYALTPSLRV G
7724	10	NKLASVYALTPSLRV
7725	10	RIGKMVSFPRI
7726	10	RIGKMVSFRIG
7727	10	PGANSJGMWFGPRL G
7728	10	PGANSJGMWFGPRL G
7729	10	TYRKPPFNGSIFG
7730	10	TYRKPPFNGSIF
7731	10	RYLPT
7732	10	SENFTPWAYIILNGEA PIIREVHYSPL
7733	10	SENFTPWAYIILNGEA PIIREVHYSPLG
7734	10	ARYLPT
7735	10	APSLRLRF
7736	10	APSLRLRFG
7737	10	IAYKPEG
7738	10	IAYKPE
7739	10	SPTISITAPIDVLRKT WEQERARKQMVKN REFLNSLN
7740	10	AGALGESGASLSIVN SLDVLNRLLLEIARK KAKEGANRNRQILLS L
8310	10	YL
2566	11	NYKKPKL
2567	11	NYKKPKLAAAAPALLA LLVAPLLAVAA
2568	11	AAVALLPAVLLALLA PAAANYKKPKL
2569	11	NYKKPKLAAAAAVA LLPAVLLALLAP

2732	11	GWFGGADWHA
2733	11	DRFGDKDIAF
2734	11	ARRCYNDPKC
2735	11	IWGCSGKLIC
2736	11	GFW
2737	11	GFL
2738	11	LGY
2739	11	GLY
2740	11	GLF
2741	11	KLDAPTNKWG
2742	11	EPFYVNWPD
2743	11	EPFYVNDPD
2744	11	EPFYENVPD
2745	11	EPFYQNVDP
2746	11	EPFYVNVDP
2747	11	EPFYMNVDP
2748	11	EPFYENWPD
2749	11	DY
2750	11	TSLYR
2751	11	EPFYENDPF
2752	11	DPSYVNVQ
2753	11	GP
2754	11	PG
2755	11	GPGG
2756	11	PGP
2831	11	SVSEIQLMHNLGKHL NSMERVEWLKRLQ DVHNF
3075	11	SYSMEHFRWGKPVG KKRRPVKVYPNGAE DESAEAFPLEF
3106	11	YMEHFRW
3107	11	FRWGKPVGKKRRPV KVYPNGAEDESAEAF PLE
3108	11	MEHFRWGK
3109	11	MEHFRWG
3110	11	KPVGKKRRPVKVYP
3111	11	YSYM
3112	11	SYSMEHFRWGKPVG KKRRPVKVYP
3113	11	SYSMEHFRWGKPVG KKR
3114	11	SYSMEHFRWGKPVG KK
3115	11	SYSMEHFRWGKPVG
3116	11	SYSMEHFRWGKPV
3117	11	SYSMEHFRWG
3118	11	YMEHFRWG
2674	12	RPKPQQFFGLM

2675	12	RPKPQQFFGLMG
3427	12	YIGSR
3428	12	EILDVPST
3429	12	EILDV
3430	12	QTSMTDFY
3651	12	SIKVAV
3663	12	DLYYMLDL
3664	12	RNIAEIKDI
3910	12	GLFDVIKKVASVIGG L
3911	12	GLFDVIKKVASVIGG LG
3912	12	GVIDAAKKVVNVLK NLF
3913	12	GVIDAAKKVVNVLK NLFG
5455	12	GLFDIIKKIAESF
5460	12	GLFDIVKKVGTIAG L
5463	12	GLFDIVKKIAGHIASS I
5464	12	GLFDIVKKIAGHIVSS I
7045	12	GLFDIVKKVVGAFGS L
7046	12	GLFDIAKKVIGVIGSL
7047	12	GLFDIVKKIAGHIAGS I
7048	12	VYINKLTPPCGTMYY ACEAV
7049	12	TAQTPIHVYSEIGKL
7050	12	HGVSGHGQHGVBHG FKCRRWQWRMKKL GAPSITCVRRAF
7051	12	RRWQWR
7052	12	KWKLFKKIKFLHSAK KF
7053	12	GIGFLHSACKFGKA FVGEIMNS
7054	12	KWKLFKKIPKFLHLA KKF
7055	12	KWKLFKKIKFLHLA KKF
7056	12	KWKLFKKILKFLHLA KKF
7057	12	KLAKLAKKLAKLAK
8137	12	VECYGPNRPQF
8186	12	AYRNRYRRQYRY
8252	12	EQRPR
8276	12	PMDYMT
8277	12	LPTSEAAKY
8311	12	FFVAPFPEVFGK

8312	12	ENLLRFFVAPFPEVFG NENLLRFFVAPFPEVF G
8313	12	LNENLLRFFVAPFPEV FG
8314	12	NLHLPLPLL
8315	12	NLHLPLPLL
8316	12	ENLHLPLPLL
8317	12	VENLHLPLPLL
8318	12	VVV
8319	12	VWV
3296	13	HHP
3297	13	YHH
3298	13	HHPL
3299	13	LHPH
3300	13	PHH
3301	13	HLH
3302	13	LHH
3303	13	HPLH
3304	13	LLPHH
3305	13	LH
3306	13	HPHL
3307	13	PYY
3308	13	HHL
3309	13	LPYY
3310	13	LPHY
3311	13	HPH
3312	13	LLHH
3313	13	PLHH
3314	13	LLPH
3315	13	HLHP
3316	13	HLPH
3317	13	HL
3318	13	LPHH
3319	13	HH
3320	13	HHPLL
3772	13	LLPHH
3792	13	VNPHDHQN
3807	13	LLPHHADADY
3808	13	VIPAGYP
3809	13	LQSGDALRVPSGTT YY
3810	13	MQFHT
3811	13	PHCKRM
3826	13	LVNPHDHQN
7757	13	PSKYEPFV
7793	13	AHK
7794	13	VHH
7795	13	VHHANEN
7796	13	VKEAMAPK

7797	13	ARHPHPLSFM
7860	13	LLGPLTNHA
7861	13	DLGLGLPGAH
7862	13	IKK
7863	13	FKK
7864	13	FIKK
7865	13	LPHSGY
7866	13	AY
7867	13	GY
7868	13	ADF
7869	13	SDF
7870	13	FGHPY
7871	13	TTY
7872	13	LY
7873	13	IY
7874	13	KHNRGDEF
7875	13	AVPYPQR
7876	13	KVLPVPQK
7877	13	VLPVPQK
7878	13	YQEP
7879	13	YQEPVLP
7880	13	DSGVT
7881	13	IEAEGE
7882	13	DAQEKLE
7883	13	EELDNALN
7884	13	VPSIDDQEELM LNSGDALRVPSGTTY Y
7885	13	Y
7886	13	AH
7887	13	YFYPEL
7888	13	EL
7889	13	HGPLGPL
7890	13	YLYP
7891	13	WYSLAMAASDI
7892	13	MHIRL
7893	13	YVEEL
7894	13	QVVPYPQ
7895	13	HQIYPV
7896	13	PYPQ
7897	13	WSVPQPK
7898	13	WY
7899	13	IYPF
7900	13	WYS
7901	13	WYSL
7902	13	WYSLA
7903	13	WYSLAM
7904	13	SALAM
7905	13	WYSLAMA
7906	13	HDH

7907	13	HEH
7908	13	AHH
7909	13	IHH
7910	13	FHH
7911	13	WHH
7912	13	HFH
7913	13	HWH
7914	13	HYH
7915	13	HHF
7916	13	HHW
7917	13	HHY
7918	13	GHH
7919	13	NHH
7920	13	QHH
7921	13	MHH
7922	13	SHH
7923	13	THH
7924	13	DYY
7925	13	YDY
7926	13	YYD
7927	13	EYY
7928	13	YEY
7929	13	YYE
7930	13	HYY
7931	13	KYY
7932	13	RYY
7933	13	YHY
7934	13	YKY
7935	13	YRY
7936	13	YYH
7937	13	YYK
7938	13	YYR
7939	13	YYA
7940	13	YYI
7941	13	YYL
7942	13	YYV
7943	13	YYF
7944	13	YYW
7945	13	YYY
7946	13	YYG
7947	13	YYN
7948	13	YYQ
7949	13	YYM
7950	13	YYS
7951	13	YYT
7952	13	YYC
7953	13	AYY
7954	13	IYY

7955	13	LYY
7956	13	VYY
7957	13	YAY
7958	13	YIY
7959	13	YLY
7960	13	YVY
7961	13	FYY
7962	13	WYY
7963	13	YFY
7964	13	YWY
7965	13	NYY
7966	13	QYY
7967	13	YGY
7968	13	YNY
7969	13	YQY
7970	13	MYY
7971	13	YMY
7972	13	YSY
7973	13	YTY
7974	13	SY Y
7975	13	TY Y
7976	13	YCY
7977	13	CYY
7978	13	DHH
7979	13	EHH
7980	13	HHH
7981	13	HKH
7982	13	HRH
7983	13	HAH
7984	13	HIH
7985	13	HVH
7986	13	KHH
7987	13	RHH
7988	13	LHA
7989	13	LHD
7990	13	LHE
7991	13	LHF
7992	13	LHG
7993	13	LHI
7994	13	LHK
7995	13	LHL
7996	13	LHM
7997	13	LHN
7998	13	LHQ
7999	13	LHR
8000	13	LHS
8001	13	LHT
8002	13	LHV

8003	13	LHW
8004	13	LHY
8005	13	LWA
8006	13	LWD
8007	13	LWE
8008	13	LWF
8009	13	LWG
8010	13	LWH
8011	13	LWI
8012	13	LWK
8013	13	LWM
8014	13	LWN
8015	13	LWQ
8016	13	LWR
8017	13	LWS
8018	13	LWT
8019	13	LWV
8020	13	LWW
8021	13	LWY
8022	13	PHA
8023	13	PHD
8024	13	PHE
8025	13	PHF
8026	13	PHG
8027	13	PHI
8028	13	PHK
8029	13	PHL
8030	13	PHM
8031	13	PHN
8032	13	PHQ
8033	13	PHR
8034	13	PHS
8035	13	PHT
8036	13	PHV
8037	13	PHW
8038	13	PHY
8039	13	PWA
8040	13	PWD
8041	13	PWE
8042	13	PWG
8043	13	PWH
8044	13	PWI
8045	13	PWL
8046	13	PWN
8047	13	PWQ
8048	13	PWR
8049	13	PWS
8050	13	PWT

8051	13	PWV
8052	13	PWW
8053	13	PWY
8054	13	RHA
8055	13	RHD
8056	13	RHE
8057	13	RHF
8058	13	RHG
8059	13	RHI
8060	13	RHK
8061	13	RHL
8062	13	RHM
8063	13	RHN
8064	13	RHQ
8065	13	RHR
8066	13	RHS
8067	13	RHT
8068	13	RHV
8069	13	RHW
8070	13	RHY
8071	13	RWA
8072	13	RWE
8073	13	RWG
8074	13	RWH
8075	13	RWI
8076	13	RWL
8077	13	RWN
8078	13	RWR
8079	13	RWT
8080	13	RWW
8081	13	RWQ
8082	13	RWY
8083	13	PSHDAHPE
8084	13	SHDAHPE
8085	13	VDHDHPE
8086	13	PKAVHE
8087	13	PAGY
8088	13	PHHADS
8089	13	VDYP
8090	13	MY
8091	13	NIQFNEQKAL
8092	13	EEKNRLTKKTKLT
8093	13	QGAR
8094	13	LQGM
8095	13	FRDEHKK
8097	13	NAPVSIPQ
8098	13	SALLRSIPA
8099	13	FPLEMMPF

8100	13	LKQELEDLLEKQE
8101	13	LNLPTAVYMT
8102	13	RPDFDLEPPY
8103	13	VKL
8104	13	VVKL
8105	13	VKV
8106	13	PKAV
8107	13	IKL
8108	13	LGLNGDDVN
		ESTVPERTHPACPDF
8109	13	N
8110	13	VKAGFAWTANQQLS
8111	13	LMSYMWSTSM
8112	13	LELHKLRSSHWFSSR
8113	13	LARL
8114	13	GGE
8115	13	GAH
8116	13	GAWA
8117	13	PHYL
8118	13	GALAAH
8119	13	NGLEGLK
8120	13	NADFGNLGLEGLA
8121	13	HFGDPFH
8122	13	YGGFL
8123	13	YGGFM
8124	13	DRVYIHPFHL
8125	13	DRVYIHPF
8126	13	PSKYEPFV
		LEQQVDDLEGSLEQE
8127	13	KK
8128	13	LQPGQGQQ
8129	13	VPSGK
8130	13	EAK
8131	13	FVAGK
8132	13	KAI
8133	13	KVI
8134	13	KD
8135	13	DHHQ
8136	13	DHHQVYSPGEQ
8138	13	VECYGPNRPQF
8139	13	PEL
8140	13	YPEL
8141	13	FYPEL
8142	13	AQIPQQ
8143	13	LQPGQGQQG
8144	13	FDSGPAGVL
8145	13	NGPLQAGQPGER
8146	13	AIRQGDVF
8147	13	VLEANPRSF

8148	13	YFPVGGDRPESF
8149	13	IQY
8150	13	YVL
8151	13	FALPQYLK
8152	13	PYVRYL
8153	13	TIIPLPV
8187	13	WPL
8188	13	VPW
8189	13	VFPW
8190	13	PW
8191	13	HNPI
8203	13	QYDQGV
8204	13	YEDCTDCGN
8205	13	AADNANELFPPN
8206	13	AMRLTYNRPDIYAT
8207	13	AMRLTYNRPCLYAT
8208	13	AMRLTYNRPGLYAT
8209	13	AMRLTYNRPGLYAT
8210	13	AMRLTYNRPCLGAT
8211	13	AMRLTYNRPGLGAT
8212	13	AMRLTYNRPCLGAT
8213	13	IRW
8214	13	RW
8215	13	IR
8216	13	LKP
8217	13	LK
8218	13	KP
8219	13	TY
8220	13	TFE
8221	13	TSTA
8222	13	NYKQM
8223	13	HVCYMF
8224	13	VY
8225	13	SWN
8226	13	MWC
		IANNEADAISLDGGQ
8230	13	VFEAG
		PIAAEVYEHTEGSTT
8231	13	SY
		IAAEVYEHTEGSTTS
8232	13	Y
8233	13	AGWNIPIGT
8234	13	AGWNIP
8235	13	AGWNI
8236	13	WNIP
8237	13	AIEWEGIESGSVEQA
8238	13	IEWEGIESGSVEQA
8239	13	LSKAQSDFG
8240	13	LGFEYY

8241	13	LGFEY
8242	13	VIPMGL
8243	13	LVEKGDVAFI
8244	13	GWNIP
8245	13	GWNI
8253	13	LEEEEEELEGCE
8274	13	LPTSEAAKY
8275	13	PMDYMT
8278	13	VPYPQ
8279	13	VPNSYP
8280	13	NPYVPR
8281	13	YGYTGA
8282	13	ISELGW
8283	13	NHAV
8284	13	HVRETALV
8285	13	IPIQYVL
8287	13	VLPVPQKKVLPVPQK
8288	13	YFYPQL
8289	13	GPLGLLGLPLGLS
8305	13	QLGNLGV
8306	13	SVMPVVA
2838	14	GYPMPYPLPR
3215	14	YIPIQYVLSR
3805	14	YPLPR
3812	14	ASHLGLAR
3813	14	YVLSR
3814	14	YPMYPLPR
3816	14	MYPLPR
3818	14	PLPR
3820	14	WWGKKYRASKLGLA R
3827	14	PMYPLPR
3751	15	KK
3752	15	KKK
3753	15	KKKA
3321	16	KIPYIL
3322	16	RRPYIL
3323	16	ELYENKPRRPYIL
3324	16	HDMNKVLDL
3327	16	DSDPR
3328	16	DTEAR
8327	16	KGHYAERVG
2933	17	RSSCFGGRIDRIGAC
2704	18	MGQKL
2705	18	MVVGW
2706	18	MNWKR
2707	18	MFKAL
2708	18	MTITC
2709	18	MSMRL

2710	18	MTIAI
2711	18	MFSKL
2712	18	MKLNLM
2713	18	MIIQS
2714	18	MSMVV
2715	18	MINQT
2716	18	MFLLF
2717	18	MRMLT
2718	18	MVIL
2719	18	MVLTV
2720	18	MIITI
2721	18	MVQLR
2722	18	MSRHV
2723	18	MSLRV
2724	18	MVLKV
2725	18	MILVV
2726	18	MSLKV
2727	18	MYMLT
2728	18	MILKT
2729	18	MLIRV
2772	18	LLEYSI
2773	18	LEYSI
2774	18	LEYS
2775	18	LLEY
2776	18	LLEYS
2777	18	LLEYSL
2796	18	MAIPPKNQDK
2801	18	ASTTTNYT
2834	18	GGYR
2890	18	GQ
2979	18	KKAVRRQEAVDAL
2980	18	KKAMRRQEAVDAL
2981	18	KKAIRRQEAVDAL
2982	18	KKALKRQEAVDAL
2983	18	KKALHRQEAVDAL
2984	18	KKAGRRQEAVDAL
2985	18	KKALRKQEAVDAL
2986	18	KKAFRRQEAVDAL
2987	18	KKALRREEAVDAL
2988	18	KKALRRQEAYDAL
2989	18	KKALRRDEAVDAL
2990	18	KKALRRNEAVDAL
2991	18	KKALRRQEGVDAL
2992	18	KKALRRQECVDAL
2993	18	KKALRRQEVVDAL
2994	18	KKALRRQEIVDAL
2995	18	KKALRRQEAFDAL
2996	18	KKALRRQEALDAL

2997	18	KKALRHQEAVDAL
2998	18	KKALRRAEAVDAL
2999	18	AKALRRQEAVDAL
3000	18	YKALRRQEAVDAL
3001	18	KAALRRQEAVDAL
3002	18	KYALRRQEAVDAL
3003	18	KKYLRRQEAVDAL
3004	18	KKAARRQEAVDAL
3005	18	KKALARQEAVDAL
3006	18	KKALYRQEAVDAL
3007	18	KKALRRQEAVYAL
3008	18	KKALRYQEAVDAL
3009	18	KKKLRRQEAVDAL
3010	18	KKALRRYEAVDAL
3011	18	KKALRRQAAVDAL
3012	18	KKALRRQYAVDAL
3013	18	KKALRRQEIDAL
3014	18	KKALRRQEAGDAL
3015	18	KKALRRQEAADAL
3016	18	KKALRRQEAVAAL
3017	18	KKALRRQEYVDAL
3018	18	KKALRRQEAVDYL
3019	18	KKALRAQEAVDAL
3020	18	KKALRRQEAVDAY
3021	18	KKALRRQELVDAL
3022	18	KKKLRRQEAFDAL
3023	18	KKALRRQEAVDAL
3024	18	KKKLRRQEAFDAY
3025	18	KKALRRQEAVDAA
3026	18	KKAYRRQEAVDAL
3036	18	YNVEHAVDYVERA
3037	18	DRIEYNVEHAVDY
3038	18	IEYNVEHAVDYVE
3049	18	LREIQILLRF
3050	18	YCQRTLREIQILLRF
3051	18	ECESGPCCRNCKFLK EGTICKRARGDDMD DYCNGKTCDCPRNP HKGPA
3076	18	DFEEIPEEYLQ
3077	18	GDFEEIPEEYLQ
3129	18	WRMW
3130	18	WRMWYV
3131	18	WRMWY
3132	18	WRMWYVP
3133	18	RMWYVPA
3134	18	TWRMWYVPA
3135	18	WRMWYVPA
3136	18	VTWRMWYVPA

3167	18	IPI
3168	18	PPPP
3169	18	GP
3170	18	PP
3171	18	MP
3172	18	VA
3173	18	MA
3174	18	KA
3175	18	LA
3176	18	FA
3177	18	AP
3178	18	FP
3179	18	PA
3180	18	LP
3181	18	VP
3182	18	LL
3183	18	VV
3184	18	HA
3199	18	PGVY
3200	18	PGAY
3201	18	GAW
3202	18	TPGVY
3203	18	PHPHLSF
3219	18	TIPPEYR
3220	18	EKLGERRDDTIPPEYRE LLEKKTGVACE
3750	18	PPPA
3761	18	PPPK
3766	18	PPRP
3767	18	PPEP
3768	18	PPSP
3769	18	PPLP
3771	18	PPAP
3773	18	EPPPPPEPPI
3777	18	PPPE
3778	18	PPPS
3779	18	PPPF
3781	18	PPPG
3782	18	PPGP
8247	18	IR
8249	18	KF
8250	18	EF
8293	18	KNPQLR
8294	18	EITPEKNPQLR
8295	18	RKQEEDEDEEQRE
8302	18	VAWWMY
8304	18	IPA
3464	19	PGP
3744	19	VGAPG

3745	19	GVAPG
3746	19	PGAIPG
3747	19	LGTIPG
3762	19	VGVPG
3803	19	VGVAPG
3548	20	IKVAV
2832	21	GFFALIPKIISSPLFKT LLSAVGSALSSSGEQ EG
2833	21	GFFALIPKIISSPLFKT LLSAVGSALSSSGEQ E
2949	21	GIGAVLKVLTTGLPA LISWIKRKRQQ
2950	21	GIGAVLKVLTTGLPA LISWIKRKRQQG
2951	21	LKLSIVAWAKKVL
2952	21	LKLSIVAWAKKVLG
2953	21	INLKALAALAKKIL
2954	21	INLKALAALAKKILG
3192	21	FLPAIAGILSQLFG
3193	21	FLPLAIGLLGKLF
3194	21	FLPAIAGILSQLF
3195	21	FLPLAIGLLGKLF
3196	21	FLPLILRKIVTAL
3197	21	FLPLILRKIVTALG
3230	21	INVLGILGLLGKALSH LG
3231	21	INVLGILGLLGKALSH L
3232	21	INIKDILAKLVKVLGH V
3233	21	INIKDILAKLVKVLGH VG
3234	21	IKIMDILAKLGKVLAH V
3235	21	IKIMDILAKLGKVLAH VG
3236	21	SKITDILAKLGKVLA HVG
3237	21	SKITDILAKLGKVLA HV
3238	21	IKITTMLAKLGKVLA HVG
3239	21	IKITTMLAKLGKVLA HV
3862	21	FLPLILGKLVKGLL
3865	21	FLPLILGKLVKGLLG
3866	21	FLPILGKIAGFLF
3867	21	FLPILGKIAGFLFG
3868	21	FLPLLFGALSAILPKIF

3869	21	FLPLLFGALSAILPKIF G
3871	21	FLPILGKILGGLL
3872	21	FLPILGKILGGLLG
3873	21	FLPILGKILGQLF
3875	21	FLPILGKILGQLFG
3876	21	FLPLIAGLLGKLF
3877	21	FLPLIAGLLGKLF
3962	21	LKKISQRYQKFALPQ Y
3963	21	LKKISQRYQKFALPQ YLKT
3964	21	VYQHQAAMKPIQP KTKVIPYVRYL
3965	21	VYQHQAAMKPIQP KTKVIPYVRY
3966	21	VYQHQAAMKPIQP KTKVIPY
7760	21	GWGSFFKAAHVGK HVGKAALTHYL
7761	21	GWGSFFKAAHVGK HVGKAALTHYL
7764	21	RWRSFFKAAHRGK HVGKRARTHYL
7765	21	RWRSFFKAAHRGK HVGKRARTHYL
7768	21	SWSSFFKAAHSGK HVGKSASTHYL
7769	21	SWSSFFKAAHSGK HVGKSASTHYL
3046	22	PGP
3163	22	APGPR
2863	23	YPFVEP
2864	23	YPFV
2865	23	YPFVEPIP
2866	23	YGFPL
2867	23	YPFVE
2877	23	YGGFM
2879	23	YGGFL
3251	23	YPFPGPIPNL
3256	23	YLLF
3262	23	YPFPGPI
3263	23	YPFPG
3473	23	RYLGYL
3552	23	YGLF
2839	24	GYPMYPLPR
3128	24	RYLGYL
3216	24	YIPIQYVLSR
3217	24	SRYPY
2768	25	KNNQKSEPLIGRKT
3580	25	VQGEESNDK

3581	25	FRHSVV
3582	25	YGGF
3583	25	DFGLEKI
3617	25	HIGPGRAF
3030	26	AMKPWIQPKTKVIPY VRYL
3031	26	KPWIQPKTKVIPYVR YL
3032	26	LKKISQRYQKFALPQ Y
3033	26	VYQHQAAMKPWIQ KTKVIPYVRY
3034	26	VYQHQAAMKPWIQ KTKVIPYVRYL
3191	26	GGHTRH
3475	26	TRWLVIYFSRPLYVAT
3476	26	IVRHLFLHVYPRVLM
3477	26	PRHVFYRWFLSNPRI
3665	26	VCDWWGWGIC
7772	26	GKVLKKRR
7792	26	DWLKAFYDKVAEKL KEAF
8303	26	VAWWMY
2730	27	PHPFHFFVYK
2731	27	PHPFHFFVY
3226	27	KRPAGFSPFR
3227	27	EGGWPRPGPEIPP
3340	27	RPPGFSPFR
8246	27	IR
8248	27	KF
8251	27	EF
2769	28	YRVRVTPKEKTGPM KEM
2770	28	WQPPRARI
2771	28	YEKPGSPREVPRP RPGV
2789	28	DAENLIDSFQEIV
3074	28	LCNTNTCATQRLAN FLVHSSNNGAILSST NVGSNTTG
3139	28	EAQGCRGWVWGNC KEWLGDEYAKNTGT PAEKGKSRNPP
3140	28	DVEVCRGGWVGHC NAWLRDEYNRQPKK PVQQQVVYSTR
3141	28	SLEPCRGWVGHEN EWQRDEYAINPKWP NAPIEDPNPL
3142	28	QVDTCTRGWVGHC NAWMGDEYAKTPG LPPMPQSDYLKPR

3143	28	GLGACRRGWVGHC NDWLNDEYAKKPGY AMPDGYPHNGT
3144	28	EGEVCLPGWVGHC YWLMDYANIPRNP TPRSNELKPP
3145	28	DFDVCRRGWVGHC KDWSDEYASNPSYP VPHSYLNP
3146	28	GKEVCRGGWVGHC QEWPMDEYTRNPSH PVPHNSRHKTP
3148	28	DVHECRPGWVGHC KDWSDEYASNRRS PEPHRYPIPP
3149	28	NIQGCIRGWVQCK DWLRDEYAREHTNQ ETPNLLNPP
3150	28	DREGCRGGWVQGC KAWFN
3151	28	REGCRGGWVQCK AWFN
3152	28	EGCRGGWVQCKA WFN
3153	28	DVEACGGWVGHC NYWLRDEYASKPIKQ VPPGNHNQPS
3154	28	DREGCRGGWVQGC KAWFNDEYAKPPK PFRNSYSLGPA
3155	28	LLQMCSPGWVGHC NDWPRDEYANNPPN PVVDRQALTPP
3156	28	DFEDCQGGWVGHC NDWLGDEYARHPRY GATQTLVNRH
3157	28	DLEVCRGGWVGHC KDWIWEYARNPRY PDPQRKEVKS
3159	28	AKVVCRRGWVGHC SAWLTDEYESNPNT RIPNTFDMKTP
3160	28	DLAGCRGGWVGHC SEWLRDEYTSNPRYP VAPSYRLQPP
3161	28	YSDVCRGGWVGHC DWLGDEYSSQPSYA LPHSTSLNPR
3644	28	PQPFPSQQPYLQ
3645	28	RPQQPYQPQPQ
3646	28	CPQLQPQNPSQQQP QEQQ
3647	28	GSFRPSQQNPQAQG
3661	28	LQGCTLRWRAGM C

3672	28	CMGLSLRPWMLCAK
3687	28	STRSESRHPFPWLL
3688	28	VRRQIVEYKHRLTLP
3689	28	FEWNYVEFSWASV
3690	28	VRQWNLTFLVLDTH P
3915	29	GLFDVIKKVASVIGG L
3916	29	GLFDVIKKVASVIGG LG
3917	29	GVIDAAKKVVNVLK NLF
3918	29	GVIDAAKKVVNVLK NLFG
3919	29	GLFDIIKKIAESI
3920	29	GLFDIIKKIAESIG
3921	29	GLFDIIKKIAESF
3922	29	GLFDIIKKIAESFG
3923	29	GLLDIVKKVVGAFGS L
3924	29	GLLDIVKKVVGAFGS LG
3925	29	GLFDIVKKVVGALGS L
3926	29	GLFDIVKKVVGALGS LG
3927	29	GLFDIVKKVVGAIKS L
3928	29	GLFDIVKKVVGAIKS LG
3929	29	GLFDIVKKVVGTIAG L
3930	29	GLFDIVKKVVGTIAG LG
3933	29	GLFDIAKKVIGVIGSL
3934	29	GLFDIAKKVIGVIGSL G
3935	29	GLFDIVKKIAGHIAG GSI
3936	29	GLFDIVKKIAGHIAG GSIG
3939	29	GLFDIVKKIAGHIVSS I
3940	29	GLFDIVKKIAGHIVSS IG
3941	29	GLMSSIGKALGLIV DLVKPTPAS
5318	29	GLFDIVKKVVGAFGS L
5319	29	GLFDIVKKVVGAFGS LG
3164	30	RGD
3165	30	YIGSR

3968	31	FLSFPTTKTYFPDFDL SHGSAQVKGHGAK
7758	31	GWGSFFKKAHVGVK HVGKAALTHYL
7759	31	GWGSFFKKAHVGVK HVGKAALTHYLG
7762	31	RWRSFFKKAHRGK HVGKRARTHYL
7763	31	RWRSFFKKAHRGK HVGKRARTHLYG
7766	31	SWSSFFKKAHSGK HVGKSASTHYL
7767	32	EDLIWK
3982	32	ADRDQYELL
3983	32	EDLIWK
3067	33	CTCFPTYKDKECVYY CHLDIIW
3068	33	CSCSSWLDKECVYF CHLDIIW
3069	33	CSCSSLMDKECVYF CHLDIIW
3463	33	DRVYIHPF
3988	33	ETPDCFWKYCV
3989	34	ETPDCFWKYCV
4005	35	RA
4006	35	LA
4007	35	WA
4008	36	FLGYDCNLEKCNHH GVCNNKNCH
7771	37	HKPS
7773	38	DDPNDSSDESNGND D
7774	38	DSKSDSSKSESDDS
7780	39	NPVENYIDEVLNEVL VVPNINSSNC
7781	39	GRKKRRQRRR
7782	39	GLFEAEGFIENGWEG MIDGWYG
7783	39	GLFEAIEGFIENGWEG MIWDYG
7784	39	KFTIVFPHNQKGNW KNVPSNYHYCP
7785	39	GIGAVLKVLTTGLPA LISWIKRKRQQ
7786	39	MVKSIGSWILVLFV AMWSDVGLCKKRPK P
7787	39	WEAKLAKALAKALA KHLAKALKAKALCE A
7788	39	GALFLGFLGAAGST MGAWSQPKKRKKV
7789	39	GLFEALLELLESLWEL LLEA

7790	39	GWTLSAGYLLGKI NLKALAALAKKIL
7791	39	LIRLWSHLIHWFQNR RLKWKKK
3559	40	APRANVAWCT
3560	40	APRKSSRWCT
3561	40	APRKAARWCT
3576	40	APAANVRWCT
3586	40	WQARMKKAGA
3589	40	FACRRWQWRMAAL GA
3590	40	FKCAAWQWAMKKL GA
3591	40	WQWRMKKLGA
3593	40	WQWRGKKLGA
3594	40	WQWRCKKLGA
3596	40	WQWRSKKLGA
3598	40	WQCRMKKCGA
3601	40	WQWRMKALGA
3603	40	PSITCVRRAF
3604	40	APRKNVRWCT
3605	40	APRKNVAWCT
3607	40	WQWRAKKLGA
3649	40	APAKNVAWCT
3764	40	FACAAWQWRMKKL GA
3765	40	FKCRRWQWRMKKL GA

