



Universidad del Azuay
Facultad de Ciencia y Tecnología

Escuela de Ingeniería en Alimentos

**Actividad Antihipertensiva de Péptidos obtenidos por hidrólisis
y/o fermentación de alimentos – Modelización QSAR**

**Trabajo de graduación previo a la obtención del título de Ingeniero en
Alimentos**

Autor:

Bolívar Andrés Pérez González

Director:

Piercosimo Tripaldi Cappelletti

Cuenca-Ecuador

2014

DEDICATORIA

“El presente trabajo está dedicado a, Bertha González, mi madre y a Pastora González, mi abuela”

AGRADECIMIENTO

Quisiera agradecer al Dr. Piercosimo Tripaldi por las enseñanzas, apoyo y afecto que ha sabido entregarme durante el desarrollo de este trabajo y el tiempo que le conozco.

INDICE DE CONTENIDOS

DEDICATORIA	ii
AGRADECIMIENTO	iii
INDICE DE CONTENIDOS	iv
INDICE DE TABLAS.....	viii
INDICE DE FIGURAS.....	ix
INDICE DE ANEXOS.....	x
RESUMEN.....	xi
ABSTRACT	xii
INTRODUCCION	xiii
Antecedentes.....	xiii
Hipótesis.....	xv
Hipótesis Nula	xv
Hipótesis Alternativa	xv
Objetivos	xv
Objetivo General.....	xv
Objetivos Específicos:	xv

CAPÍTULO I: HIPERTENSIÓN

Introducción.....	1
1.1. Hipertensión.....	1
1.1.1. Angiotensina II	3
1.1.1.1. Acciones de la Angiotensina II	3

1.1.1.1.1 Sistema Nervioso Central	3
1.1.1.1.2. Corazón	4
1.1.1.1.3 Sistema Vascular	5
1.1.1.1.4 Riñón	5
1.1.2 Sistema Renina-Angiotensina	6

CAPÍTULO II: PÉPTIDOS BIOACTIVOS

Introducción	7
2.1 Péptidos Bioactivos.....	7
2.2. Obtención de Péptidos Bioactivos.....	7
2.2.1. Proteolisis	8
2.3. Separación y Purificación de Péptidos Bioactivos	12
2.4. Actividad de Péptidos Bioactivos	12
2.4.1. Actividad Antihipertensiva de Péptidos Bioactivos	13
2.4.2. Cuantificación de la Actividad Antihipertensiva de Péptidos Bioactivos.....	14
2.5. Mecanismo antihipertensivo de los péptidos bioactivos	14
2.5.1 Inhibición del sistema renina-angiotensina.....	14

CAPITULO III: QSAR

Introducción	16
3.1 QSAR (Quantitative Structure-Activity Relationship)	16
3.1.1. Utilización de modelos matemáticos en Química Teórica	16
3.1.2. Fundamentos QSPR-QSAR.....	17
3.2. Descriptores Moleculares.....	18
3.2.1. Descriptores 0D	19
3.2.2. Descriptores 1D	19
3.2.3. Descriptores 2D o Topológicos	19

3.2.3.1. Descriptores de Autocorrelación	19
3.2.3.2. Descriptores BCUT	20
3.2.3.3. Índices de carga Topológica de Gálvez	20
3.2.3.4. Cuenta de Trayectos Moleculares.....	20
3.2.3.5. Índices de Conectividad	21
3.2.4. Descriptores 3D	21
3.2.4.1. Perfiles moleculares de Randić.....	21
3.2.4.2. Descriptores RDF (Funciones de Distribución Radial).....	21
3.2.4.3. Descriptores WHIM.....	22
3.2.4.4. Descriptores 3D MoRSE	22
3.2.4.5. Descriptores GETAWAY	22
3.2.4.6. Descriptores Geométricos.....	22
3.2.4.7. Descriptores de Carga	23

CAPITULO IV: ANÁLISIS MULTIVARIANTE

Introducción.....	24
4.1. Análisis Multivariante	24
4.1.1. Selección de Variables.....	24
4.1.1.1 Algoritmos Genéticos.....	24
4.1.2. Modelos de Regresión	26
4.1.2.1. Análisis de Regresión Multivariante	26
4.1.2.2. Métodos de los Mínimos Cuadrados Ordinarios (OLS)	27
4.1.3. Modelos de Clasificación	28
4.1.3.1. K-means	28
4.1.3.2. KNN	28

CAPÍTULO V: METODOLOGÍA

Introducción	30
5.1. Metodología	30
5.1.1. Base de Datos	30
5.1.2. Digitalización y construcción de las moléculas	31
5.1.3. Cálculo de Descriptores Moleculares	36
5.1.4. Cálculo de los Modelos de Regresión y Clasificación	37

CAPÍTULO VI: RESULTADOS Y DISCUSIÓN

Introducción	38
RESULTADOS Y DISCUSIÓN	38
CONCLUSIONES	50
RECOMENDACIONES.....	52
BIBLIOGRAFÍA	53
ANEXOS	56

INDICE DE TABLAS

Tabla 1: Sistemas y órganos afectados por la Hipertensión Arterial y sus condiciones patológicas	2
Tabla 2: Proteasas Comerciales	9
Tabla 3: Microorganismos responsables de la fermentación de alimentos, generadores de péptidos bioactivos	10
Tabla 4: Di y Tripéptidos Inhibidores de la ECA Obtenidos por Hidrolisis y/o Fermentación.....	11
Tabla 5: Alimentos Funcionales comerciales con Péptidos Bioactivos.....	12
Tabla 6: Opciones para el desarrollo de OLS	39
Tabla 7: Parámetros del Modelo generados por OLS	39
Tabla 8: Opciones para el desarrollo de KNN	40
Tabla 9: Modelos calculados en función de los descriptores moleculares seleccionados con su valor de NER%.....	42
Tabla 10: Centroides de los descriptores para cada clase	43
Tabla 11: Matriz de confusión del modelo con 8 descriptores y 3 vecinos	45
Tabla 12: Clasificación de di y tripéptidos del Training Set	46
Tabla 13: Selectividad y especificidad de las clases del modelo de clasificación con 8 descriptores	47
Tabla 14: Matriz de confusión de validación del modelo	48
Tabla 15: Clasificación de di y tripéptidos del Test Set	48
Tabla 16: Selectividad y especificidad de la validación del modelo	49

INDICE DE FIGURAS

Figura 1: Vía de síntesis de la Angiotensina II	15
Figura 2: Funcionamiento Algoritmos Genéticos.....	26
Figura 3: Características de los aminoácidos de la base de datos	32
Figura 4: Opciones de la herramienta MM+	33
Figura 5: Componentes de la herramienta MM+	34
Figura 6: Optimización mecánica molecular.....	35
Figura 7: Molécula de Alanina-Alanina optimizada utilizando el software Hyperchem 8.0	36
Figura 8: Relación entre número de descriptores considerados en la clasificación	41
Figura 9: Comportamiento de los centroides de cada clase del modelo.....	44

INDICE DE ANEXOS

ANEXO 1 (Base de datos de dipéptidos y tripéptidos con actividad inhibidora de la enzima convertidora de la Angiotensina)	31
ANEXO 2 (Descriptores Moleculares Calculados y Seleccionados del Training Set) ...	41
ANEXO 3: Descriptores Moleculares Calculados y Seleccionados del Test Set.	47

Revisado
09/01/14

"Actividad Antihipertensiva de Péptidos obtenidos por hidrólisis y/o fermentación de alimentos – Modelización QSAR"

RESUMEN

En la actualidad se ha demostrado que los péptidos poseen actividad antihipertensiva y su consumo regular disminuye la hipertensión. En este trabajo se empleó la metodología QSAR y obtener un modelo matemático que permita predecir esta actividad en péptidos. Se empleó el método de clasificación KNN, para desarrollar un modelo matemático que prediga la capacidad de inhibición de la ECA, a partir de una base de datos de 262 moléculas y 956 descriptores moleculares, el modelo obtenido está compuesto por 8 descriptores moleculares. Este modelo puede clasificar a los péptidos en una de tres clases, estas clases son: Clase 1 baja capacidad de inhibición, Clase 2 mediana capacidad de inhibición y Clase 3 alta capacidad de inhibición, con una precisión del 94%. Este modelo servirá como herramienta para futuros trabajos de investigación en el desarrollo de alimentos funcionales con capacidad antihipertensiva.

Palabras Claves: Actividad antihipertensiva, péptidos bioactivos, QSAR, KNN, ECA.



Dr. Piercosimo Tripaldi

Director



Ing. Fausto Parra

Presidente de la Junta Académica



Bolívar Andrés Pérez González

Autor

Received
13/01/14

ABSTRACT


"Antihypertensive activity of peptides obtained by hydrolysis and / or fermentation of food-QSAR Modelling"


At present it has been proven that peptides have antihypertensive activity and that its regular use decreases hypertension. In this work we used QSAR (quantitative structure-activity relationship) methodology and obtain a mathematical model to predict this activity in peptides. The KNN classification method was used to develop a mathematical model that predicts the capacity of ACE inhibition from a database of 262 molecules and 956 molecular descriptors. The model obtained consists of eight molecular descriptors. This model can classify peptides into one of three classes: Class 1 low inhibition capacity. Class 2 medium inhibition capacity, and Class 3 high Inhibition capacity, with an accuracy of 94%. This model will serve as a tool for future research on the development of functional foods with antihypertensive capacity.

Keywords: Antihypertensive Activity, Bioactive Peptides, QSAR, KNN, ECA


Dr. Piercosimo Tripaldi
Director


Ing. Fausto Parra
President of the Academic Board


Bolívar Andrés Pérez González
Author


Lourdes Crespo
UNIVERSIDAD DEL
AZUAY
DPTO. IDIOMAS


Translated by,
Lic. Lourdes Crespo

Pérez González Bolívar Andrés

Trabajo de Graduación

Tripaldi Capelleti Piercosimo

Enero del 2014

ACTIVIDAD ANTIHIPERTENSIVA DE PÉPTIDOS OBTENIDOS POR HIDROLISIS Y/O FERMENTACIÓN DE ALIMENTOS MODELIZACIÓN QSAR

INTRODUCCION

Antecedentes

A partir de los años 90, se ha comenzado a trabajar con el estudio de péptidos bioactivos con una actividad nutraceútica específica. Para la obtención de estos péptidos se ha trabajado con técnicas de hidrólisis, tanto enzimática como química, y fermentación, siendo los sustratos utilizados inicialmente los productos lácteos. El uso de péptidos bioactivos pueden ayudar a reducir el número de muertes prematuras causadas por enfermedades crónicas, que son de 58 millones anuales. Las proteínas funcionales y los péptidos bioactivos constituyen una categoría importante dentro de los alimentos nutraceúticos que están valorados en \$75 billones anuales (*Yoshimori, et al. 2010*).

Los péptidos bioactivos en general poseen entre 2 y 20 aminoácidos, los cuales actúan como compuestos moduladores y regulatorios en los organismos animales, estas actividades dependen de la composición de los péptidos y de la secuencia en la que estos se encuentran enlazados entre sí. Entre las actividades relevantes que poseen los péptidos bioactivos se encuentran: actividad antioxidante, actividad antitrombótica, actividad hipocolesterolemicos, actividad antiobesidad, actividad antitumoral, actividad inmunomoduladora, actividad antihipertensiva y actividad antiinflamatoria (*Yoshimori, et al 2010*)

En la actividad antihipertensiva se encuentra la inhibición de la enzima ECA, que provoca la transformación de la angiotensina I en angiotensina II, de esta forma al

conservar la angiotensina I se logra prevenir la degradación de la bradiquina que tiene un efecto vasodilatador. Mientras que la angiotensina II aumenta la vasoconstricción que afecta a la síntesis y liberación de la aldosterona (*Chibuike & Rotimi, 2012*).

El desarrollo de modelos matemáticos que permitan estudiar los comportamientos de moléculas bioactivas, es una herramienta que nos permite predecir la actividad de otras de semejantes. Para esto se han desarrollado los estudios QSAR que están basados en una serie de aplicaciones matemáticas y estadísticas que tienen como objetivo encontrar un modelo matemático cuantitativo que relacione la estructura molecular y una actividad definida (Frank & Todeschini, 2006). QSAR es un tipo de estrategia que es aplicada sobre todo para resolver problemas en química, farmacología, ciencias ambientales y toxicología y está apoyada principalmente de técnicas estadísticas multivariantes y quimiometría (*Frank & Todeschini, 2006*).

En base a todo lo detallado se pretende realizar el estudio QSAR, para desarrollar un modelo matemático que relacione parámetros estructurales de los péptidos (descriptores moleculares) con la actividad antihipertensiva que inhiban la enzima ECA. Con la obtención de este modelo se puede predecir la actividad antihipertensiva de otros péptidos que se obtengan de procesos de hidrólisis enzimática y/o fermentación, sin ser necesarias las pruebas in-vitro e in-vivo.

Según los datos de la Organización Mundial de la Salud (OMS), 1 de cada 3 personas adultas presentan problemas de hipertensión, de estos 1 de cada 3 adultos desconocen que la poseen (OPS/OMS, 2013). En nuestro país según los datos entregados por la OMS de cada 100000 ecuatorianos 1373 tienen problemas de hipertensión (*Lizarzaburo, 2013*)

La hipertensión es una de las principales causas de infarto al miocardio, accidentes cerebrovasculares e insuficiencia renal. Para reducir la hipertensión arterial desde el punto de vista nutricional, es necesario reducir el consumo de sal, evitar el uso nocivo de alcohol, evitar el consumo de tabaco y mantener una dieta equilibrada con presencia de productos de origen natural que ayudan a prevenir, entre estos productos alimenticios están los obtenidos por hidrólisis y/o fermentación, por lo cual es necesario el desarrollo de pruebas farmacológicas sobre ellos que puedan brindarnos información de su actividad antihipertensiva.

Mediante el desarrollo de un modelo matemático que relacione la actividad antihipertensiva con los péptidos presentes en los alimentos, se puede identificar cuáles son los alimentos que tendrían propiedades nutraceuticas antihipertensivas que podrán ser utilizados para combatir la enfermedad antes de que sea necesario el uso de medicamentos.

Hipótesis

Existe una relación estadísticamente significativa entre la estructura molecular de los diferentes di y tripéptidos bioactivos y su respectiva respuesta biológica (inhibición de la ECA)

Hipótesis Nula

H_0 = No existe una relación estadística significativa entre la estructura molecular y la inhibición de la ECA de los diferentes di y tripéptidos analizados.

Hipótesis Alternativa

H_1 = existe una relación estadística significativa entre la estructura molecular y la inhibición de la ECA de los diferentes di y tripéptidos analizados.

Objetivos

Objetivo General

Predecir la actividad Antihipertensiva de péptidos derivados de hidrólisis y/o fermentación de alimentos mediante un modelo matemático

Objetivos Específicos:

Obtener el modelo matemático que relacione la actividad Antihipertensiva con la estructura de péptidos derivados de hidrólisis y/o fermentación

Validar el modelo, sobre nuevos tipos de péptidos

CAPÍTULO I

HIPERTENSIÓN

Introducción

En este capítulo se abordará las características generales de la hipertensión arterial (HTA), qué es la Angiotensina II, efectos de la Angiotensina II en el organismo y su metabolismo en forma general.

1.1. Hipertensión

La hipertensión arterial (HTA) es un síndrome multifactorial con un factor genético de fácil diagnóstico, que es difícil de controlar a nivel mundial.

Los problemas existentes entre diagnóstico y la terapia de la misma se relacionan directamente con el control de esta. La falta de control de la HTA ha provocado el incremento de las enfermedades cardiovasculares y cerebrovasculares, causantes de muerte de un gran número de personas a nivel mundial. Otro de los sistemas que se ven afectados por la HTA es el sistema renal, en el cual se presenta la insuficiencia renal. Dejando claro que la HTA es uno de los problemas más importantes de la medicina contemporánea, por lo cual el control adecuado de la misma permitiría la disminución de la morbilidad y la mortalidad de la población. *(Sellen, 2007)*

Los estudios médicos realizados hasta el momento han definido que la HTA, se produce por cuatro razones teóricas esencialmente: genética, neurogénica, humoral y autorregulación.

La Hipertensión Arterial se manifiesta en diferentes órganos y/o sistemas en el organismo con diferentes condiciones patológicas, como se presenta en la tabla 1.

SISTEMA ÓRGANO	CONDICIÓN PATOLÓGICA
Endocrinopatías	Diabetes
	Feocromocitoma
	Síndrome de Cushing
	Glucocorticoides exógenos
	Mineralocorticismos
	Enfermedad de Addison
	Hipertiroidismo
	Pseudohipoparatiroidismo
Alteraciones del sistema nervioso	Insuficiencia autonómica
	Atrofia cerebral
	Enfermedad cerebro vascular
	HTA neurogénica
	Insomnio familiar fatal
Sistema Respiratorio	Síndrome de apnea del sueño
Riñón	HTA renovascular
	Insuficiencia renal crónica
Hipertensión esencial	HTA sal sensible
	HTA con hipertrofia ventricular izquierda
Corazón	Insuficiencia cardíaca congestiva

Tabla 1: Sistemas y órganos afectados por la Hipertensión Arterial y sus condiciones patológicas

Fuente: (Sellen, 2007).

La teoría genética sugiere que una alteración del ADN provoca la alteración de distintas moléculas, por lo cual su función original se ve modificada. Entre los genes que se han identificado como responsables se encuentran el gen de renina, los genes codificadores de la kinina, la kaliceína y las prostaglandinas renales, genes que codifican factores que regulan la homeostasis del calcio y el sodio, la bomba de sodio-potasio, proteína C y el fosfoinositol. (Sellen, 2007)

Entre las actividades biológicas que se desarrollan en el organismo, que se ven afectadas por las condiciones genéticas se encuentra la inhibición de la producción de la Angiotensina II (A-II), la cual se regula la presión arterial. La A-II está relacionada directamente con la cardiopatía isquémica, la insuficiencia cardiaca congestiva y la insuficiencia renal. *(Sellen, 2007)*

1.1.1. Angiotensina II

La angiotensina II (A-II) es uno de los principales factores vasoactivos implicados en el desarrollo y las complicaciones de diversas patologías cardiovasculares y renales como la hipertensión arterial, la insuficiencia cardíaca, la insuficiencia renal, la nefropatía diabética, etc. Por ello, la inhibición de las acciones de la A-II ha sido y continúa siendo de los objetivos primordiales de la terapéutica cardiovascular. Los inhibidores de la enzima de conversión de la angiotensina (IECA) han demostrado una gran eficacia en el tratamiento de las alteraciones cardiovasculares anteriormente mencionadas. *(Lahera, et al 2000)*

1.1.1.1. Acciones de la Angiotensina II

1.1.1.1.1 Sistema Nervioso Central

Se ha demostrado que en todo el sistema nervioso central se encuentran distribuidos receptores AT_1 . La A-II presenta acciones relacionadas con la regulación del equilibrio hidroelectrolítico y la presión arterial. Cuando la A-II llega al cerebro por vía intracerebroventricular provoca el aumento en la presión arterial, permite la liberación de vasopresina y provoca sensación de sed. Asimismo, la A-II, provoca la activación, en diferentes puntos, del sistema nervioso simpático. *(Lahera, et al, 2000)*

En el sistema nervioso periférico, amplifica y facilita la actividad del sistema nervioso simpático, contribuyendo de esta forma a la acción vasoconstrictora de la A-II. *(Lahera, et al, 2000)*

1.1.1.1.2. Corazón

Los efectos directos de la A-II en el corazón son inotrópicos y cronotrópicos positivos debidos a la estimulación simpática en los receptores AT₁ del mismo. En estudios recientes se ha propuesto que la A-II participa en la regulación de la comunicación entre cardiomiocitos y en la propagación del impulso cardíaco. Entre los efectos indirectos de la A-II en el corazón, se destaca la participación de la A-II en la regulación del flujo coronario por el efecto vasoconstrictor que produce en el territorio arterial general. Estos efectos en el corazón se dan cuando las concentraciones de A-II no tienen efecto vasoconstrictor. (*Lahera, et al, 2000*)

La A-II puede llegar a ejercer una función trófica sobre los cardiomiocitos, debido a la síntesis de proteínas contráctiles y a la expansión intersticial, provocada por la síntesis anormal de fibras de colágeno, lo que conduce a una hipertrofia de los cardiomiocitos y a la aparición de fibrosis intersticial, que genera la hipertrofia ventricular izquierda, fenómeno asociado a la hipertensión arterial. La hipertrofia ventricular izquierda es el principal factor mecánico que afecta al miocardio. (*Lahera, et al, 2000*)

Sellen en el 2007, describe los efectos producidos por la A-II en el corazón son los siguientes:

- Activación de la cascada de la caspasa: muerte celular programada o apoptosis.
- Induce necrosis.
- Fibrosis.
- Hipertrofia miocárdica.
- Propicia la acumulación de colágeno intersticial en el corazón.
- Intervienen en la lesión por isquemia-reperfusión.
- En la etapa pos IMA provoca el remodelado ventricular del músculo no infartado que se hipertrofia o fibrosa debido a la vasoconstricción, proliferación y crecimiento celular que provoca la A-II.
- Regula la hemodinámica intrarrenal, la filtración glomerular y la reabsorción tubular de solutos y agua.
- Libera aldosterona en la corteza suprarrenal e incrementa la reabsorción de sodio en la nefrona distal, aumentando la resistencia periférica total y la vasoconstricción, lo que contribuye a la patogenia de la nefropatía diabética y a perpetuar la HTA.

1.1.1.1.3 Sistema Vascular

La A-II favorece el incremento del tono vascular, mediante su acción facilitadora de la transmisión adrenérgica y la estimulación de factores endoteliales constrictores de la endotelina-1 (ET-1) y el tromboxano A₂ (TXA-2). La ET-1 permite mediar algunos de los efectos hipertróficos cardíacos que la A-II provoca, mientras que la producción de TXA-2 de acción vasoconstrictora y proagregante plaquetaria permite mediar parcialmente los efectos sistémicos de la hipertensión. Otra de las acciones destacables de la A-II, es ser uno de los principales estímulos en la producción de radicales libres de oxígeno en la pared vascular mediante la estimulación de varias actividades enzimáticas como la NADH/NADPH oxidasa, estos radicales libres permiten la oxidación del NO, que es el principal factor relajante regulador del tono vascular, permitiendo así que se mantenga un tono vascular elevado. Estos mecanismos y acciones se dan cuando se activan los receptores AT₁. (*Lahera, et al, 2000*)

La presencia elevada de A-II y el estrés hemodinámico son los principales causantes de las alteraciones estructurales de la pared vascular. La A-II al estimular los receptores AT₁, activan señales intracelulares que modifican la expresión de genes reguladores del crecimiento, de esta forma se estimula el crecimiento de la pared vascular y estimulan la producción y acción de factores mitogénicos que estimulan la proliferación de las células musculares lisas y la síntesis de proteínas. La consecuencia de estas acciones es el engrosamiento en el espacio subintimal, así como la acumulación de colágeno en la adventicia.

1.1.1.1.4 Riñón

En el riñón la A-II ejerce una acción vasoconstrictora selectiva sobre la arteriola eferente, que es la responsable de producir la disminución del flujo sanguíneo renal (FSR), mantiene la tasa de filtración glomerular (TFG). Estos efectos provocan que aumente la fracción de filtración, la disminución de la presión hidrostática y aumenta la presión oncótica de los capilares peritubulares, lo que facilita la reabsorción de agua y electrolitos. Cuando la A-II no produce el efecto de autorregulación del FSR provoca que la TFG provoque cambios en la presión de perfusión renal, esto supone que las

concentraciones de sodio y potasio disminuyan teniendo como consecuencia la secreción de la renina y la posterior generación de A-II. La A-II ejerce un efecto antinatriurético al estimular la síntesis y liberación de la aldosterona. (*Lahera, et al, 2000*)

Por estas razones la A-II posee una participación decisiva en los mecanismos de daño renal y en la progresión de insuficiencia renal por sus acciones sobre las arteriolas aferente y eferente. La constricción de la arteriola eferente favorece a la hipertensión glomerular que lleva al desarrollo de glomerulosclerosis, proteinuria y al deterioro progresivo de la función renal. (*Lahera, et al, 2000*) (*Matsui & Tanaka, 2010*)

1.1.2 Sistema Renina-Angiotensina

Dentro de los sistemas hormonales y enzimáticos que generan aumento o disminución de la presión arterial se encuentra el Sistema Renina-Angiotensina, siendo este uno de los más importantes debido a que su efecto no se produce solamente en el sistema circulatorio sanguíneo, sino también en otros órganos como son el cerebro, los pulmones, la arteria aorta y el riñón. (*Matsui & Tanaka, 2010*)

En este sistema, el angiotensinógeno producido por el hígado es transformado en A-I que es un decapeptido, la A-I por acción de la enzima convertidora de la angiotensina (ECA) se transforma en la A-II por escisión del dipéptido Histidina-Leucina. (*Matsui & Tanaka, 2010*)

CAPÍTULO II PÉPTIDOS BIOACTIVOS

Introducción

En este capítulo se detalla información de los péptidos, su definición general, información de péptidos bioactivos, la obtención de péptidos bioactivos, los métodos de laboratorio como se los extrae, identifica y purifican, y sobre todo la actividad antihipertensiva de los péptidos bioactivos por medio de la inhibición de la Enzima Convertidora de la Angiotensina (ECA)

2.1 Péptidos Bioactivos

Los péptidos bioactivos, son péptidos que a más de ser fuentes alimenticias de aminoácidos esenciales, tienen características nutraceuticas, capaces de ejercer efectos biológicos específicos. La generación de péptidos bioactivos se da de forma espontánea en el organismo cuando se digieren las proteínas proceso que se realiza *in vivo*. En la experimentación *in vitro* se ha logrado obtener péptidos bioactivos a partir de proteína, por vía enzimática, mediante la utilización de bacterias. Los sustratos alimenticios más utilizados, para la obtención de péptidos bioactivos son la leche, la sardina, la soya, el huevo, la gelatina y el maíz. En estudios científicos realizados se ha logrado demostrar que los péptidos bioactivos tienen acción tanto a nivel local (tracto gastrointestinal), como a nivel sistémico, debido a que por su tamaño pueden atravesar el epitelio intestinal y llegar a los tejidos periféricos por medio de la circulación sanguínea (Martínez, 2006).

2.2. Obtención de Péptidos Bioactivos

Para la obtención de péptidos bioactivos se ha desarrollado estudios sobre diferentes alimentos, tanto de procedencia animal como vegetal. La mayoría de los estudios realizados sobre alimentos de origen animal se han centrado en las proteínas de la leche, tanto en las caseínas como en el lacto suero, en huevos, en carne y en

alimentos de origen marino, en los que se incluyen al salmón, las ostras, el calamar, el erizo de mar, los camarones, centollas y al hipocampo. En los estudios realizados sobre alimentos de origen vegetal se incluyen la soya, lenteja, garbanzo, arveja, frejol, avena, trigo, semillas de cáñamo, canola y linaza. (*Chibuike & Rotimi, 2012*)

Los criterios que se utilizan para justificar la obtención de péptidos bioactivos, son dos de gran importancia. Uno la utilización de las abundantes proteínas subutilizadas producidas por la industria alimentaria y dos la utilización de proteínas que contengan aminoácidos con un interés farmacológico particular. (*Chibuike & Rotimi, 2012*)

2.2.1. Proteolisis

La proteólisis es una hidrolisis enzimática, provocada por la acción de enzimas proteasas, las cuales provocan la ruptura de los enlaces químicos en puntos específicos en la cadena de aminoácidos de acuerdo a la especificidad de la enzima. (*Benítez, et al, 2008*) Al producirse la proteólisis se potencian algunas características funcionales de los alimentos, lo que genera una ventaja nutricional de estos alimentos sobre los alimentos similares que poseen las proteínas originales, por este motivo se han desarrollado hidrolizados proteicos que se utilizan industrialmente para generar alimentos de consumo especial.

La hidrolisis enzimática de las proteínas depende tanto de la enzima que realiza el proceso como de las condiciones de hidrolisis, siendo necesario conocer las condiciones óptimas de trabajo de la enzima. De esta forma se puede mantener y controlar el proceso, los parámetros que deben controlarse en un proceso de hidrolisis enzimática son el pH, la temperatura y el tiempo. El pH es un indicador del grado de hidrólisis de la proteína, puesto que conforme avanza el tiempo el pH del sustrato, proveedor de proteína, desciende debido a la ruptura de los enlaces peptídicos de la proteína. La temperatura de trabajo, tiene que ser la temperatura ideal de trabajo de la enzima debido a que las enzimas proteasas son sensibles a los cambios de temperatura. Cuando se conocen las condiciones de trabajo de la enzima, tanto el pH como la temperatura, se pueden determinar los tiempos de hidrólisis para esto el proceso debe ya haber sido estandarizado. Cuando se conocen estos parámetros, y se desea llegar a un grado de hidrolisis conocido, es necesario detener el proceso, para

ello se puede optar por 3 métodos, la disminución del pH, el aumento de la temperatura y una combinación de los dos métodos anteriores. (Benítez, et al, 2008)

Debido a la importancia de la hidrólisis de proteínas, se han desarrollado concentrados de enzimas para uso industrial, en la tabla 3 se muestra el tipo de proteasa, el origen, el nombre de la enzima, la fuente de obtención, el rango de temperatura y el rango de pH de actividad de la enzima.

Tipo de proteasa	Origen	Nombre	Fuente	Temperatura (°C)	pH
Serinproteasa	Animal	Tripsina	Porcino, Bovino	30 - 60	7 a 9
		Quimotripsina		45 -55	8 a 9
		Elastasa			6 a 8
	Bacteriano	Subsilitisin Carlsberg alcalasa	Bacillus licheniformis	50 - 60	6 a 10
		Subst. BPN, Subsilitisin Novo	Bacillus amyloliquefaciens	40 -55	6 a 10
Cisteinproteasa	Vegetal	Papaína	Papaya	40 - 75	5 a 8
		Bromelaina	Piña	20 - 65	5 a 8
		Ficina	Latex de Ficus		5 a 8
Aspartato proteasas	Animal	Pepsina	Porcino, Bovino		1 a 4
		Quimosina	Becerro		4 a 6
	Fúngico	Aspergilopeptidasa A	Aspergillus saitoi	35 - 50	2 a 5
		Newlasa	Rhizopus sp.	40 - 50	3 a 6
Metaloproteasas	Animal	Carboxipeptidasa A	Páncreas		7 a 8
	Bacteriano	Neutrasa®	Bacillus amyloliquefaciens	40 - 55	6 a 7.5
		Termolisina	B. Termoproteolyticus		7 a 9
Preparaciones enzimáticas					
Mezcla de papaina, quimopapaina y lisozima		Papaína cruda	Fruto de la papaya		5 a 9
Mezcla de tripsina, quimiotripsina, elastasa y carboxipeptidasa		Pancreatina	Páncreas (bovino y porcino)	30 - 80	7 a 9
Mezcla de serin-, aspartato- y metalo- proteasas		Veron P. Sumicina LP. Bocina A	Aspergillus oryzae	40 - 55	4 a 8
Mezcla de endo- y exo- proteasas, actividad en pH alcalino y neutro		Pronasa	Streptomyces griseus		7 a 9

Tabla 2: Proteasas Comerciales

Fuente: (Benítez, 2008)

El efecto proteolítico se puede generar a partir de la fermentación de alimentos, por medio de la acción de microorganismos. Los microorganismos entre los subproductos de su metabolismo generan enzimas y compuestos responsables de sabor y aroma. Los cuales provocarán la transformación del alimento sobre el cual actúan, por todas las transformaciones que se producen por este mecanismo, se ha considerado a la fermentación como un proceso general para la obtención de péptidos bioactivos a

pesar que dentro de este se realice hidrólisis enzimática también, en la tabla 4, se muestran los microorganismos utilizados en la fermentación de alimentos que se ha demostrado, ser los responsables de la generación de péptidos bioactivos.

Microorganismos responsables de la fermentación de alimentos			
Microorganismo	Sustrato Alimenticio	Péptidos Bioactivos	Referencia
<i>Lactobacillus helveticus</i>	Leche y Productos Lacteos	VPP, IPP	Nakamura Y, et al 1996
<i>Saccharomyces cerevisiae</i>	Leche, Productos Lacteos, cereales, harinas, carne	VPP, IPP	Nakamura Y, et al 1996
<i>Lactobacillus helveticus LBK-16H</i>	Leche y Productos Lacteos	VPP, IPP	Sipola MP, et al 2002
<i>Enterococcus faecalis</i>	Leche y Productos Lacteos	VPP, IPP, LHLPLP	Quirós A, et al 2007

Tabla 3: Microorganismos responsables de la fermentación de alimentos, generadores de péptidos bioactivos

Fuente: Varias fuentes (Nakamura, et al, 1995) , (Nakamura, et al, 1996), (Sipola, et al, 2002), (Quirós, et al, 2007))

Los péptidos bioactivos que se pueden generar a partir de la hidrólisis y la fermentación de alimentos, ofrece un número considerable de estos, la información de los péptidos que se utilizarán en este estudio, corresponden a di y tripéptidos bioactivos con capacidad de inhibición de la ECA, en la tabla 5, se han listado algunos de los péptidos generados por hidrólisis y/o fermentación, que alimento es el proveedor de los mismos, y las enzimas que intervienen en este proceso en el caso de la hidrólisis enzimática.

Di y Tripéptidos Inhibidores de la ECA Obtenidos por Hidrolisis y/o Fermentación			
Mecanismo de Obtención	Péptido	Fuente	Enzimas utilizadas
Hidrólisis Enzimática	LKP	Músculo de Pollo	Aldolasa
	IPA	β -lactoglobulina	Proteínasa K
	GKP	β -microglobulina	Proteínasa K
	LW	Salmón	Termolisina
	IW	Salmón	Papaina
	GPL	Adebajo de Alaska (pez)	Alcalasa, pronasa y colagenasa
	GPM	Adebajo de Alaska (pez)	Alcalasa, pronasa y colagenasa
	KY	Sardina	Proteasa alcalina
	AKK	Sardina	Proteasa alcalina
	VY	Sardina	Proteasa alcalina
	LY	Sardina	Proteasa alcalina
	GRP	Sardina	Proteasa alcalina
	IY	Bonito	Termolisina
	IKP	Bonito	Termolisina
	FQP	Bonito	Termolisina
FP	β -caseína, albumina	Proteínasa K	
Fermentados	RF	lías de sake	N/A
	VW	lías de sake	N/A
	YW	lías de sake	N/A
	VY	sake	N/A
	VPP	beta-caseína	N/A
	IPP	beta y k caseína	N/A
	YP	alfa, beta y k caseína	N/A

Tabla 4: Di y Tripéptidos Inhibidores de la ECA Obtenidos por Hidrolisis y/o Fermentación

Fuente: (Yamamoto, 2010)

Industrialmente, esta información ha permitido el desarrollo de alimentos funcionales a nivel mundial. En la tabla 6 se muestran algunos alimentos con una actividad nutraceútica específica, la marca del producto, el tipo de alimento, los péptidos bioactivos identificados presentes en el alimento, y el productor.

Alimentos Funcionales comerciales con Péptidos Bioactivos				
Actividad	Marca	Tipo de Alimento	Péptidos Bioactivos	Productor
Inhibidores de la ECA	Calpis/Ameal S	Leche agria	VPP, IPP	Calpis Co., Japón
Inhibidores de la ECA	Peptide soup	Sopa	LKPNM	Japan supplement, Japón
Inhibidores de la ECA	Marine peptide	Suplemento Alimenticio	VY	SenmiEkisu, Japón
Inhibidores de la ECA	Sesame tea	Te	LVY	Suntory, Japón
Inhibidores de la ECA	Evolus	Leche fermentada	VPP, IPP	Valio Oy, Finlandia
Inhibidores de la ECA	BioZate	Proteína de suero hidrolizada	Fragmentos β -lactoglobulina	Davisco, USA
Anticancerígeno, antimicrobiano,	BioPURE-GMP	Aislado de proteína de Suero de leche	Glucomacropéptido	Davisco, USA
Disminución de los síntomas del estrés	ProDiet F200/Lactium	Leche saborizada	YLGYLEQLLR	Ingredia, Francia
Mejora Atlética	PeptoPro	Bebida saborizada	Di y tripéptidos derivados de la	DSM Food, Holanda
Relajante	Vivinal Alpha	Ingrediente	α -lactoalbumina	Borculo Domo Ingredients, Holanda

Tabla 5: Alimentos Funcionales comerciales con Péptidos Bioactivos

Fuente: (Matsui & Tanaka, 2010)

2.3. Separación y Purificación de Péptidos Bioactivos

Para la purificación y separación de péptidos se utilizan diferentes métodos que son la cromatografía de filtración en gel, cromatografía de intercambio iónico, y la cromatografía líquida de alta resolución de fase inversa (RP-HPLC)

2.4. Actividad de Péptidos Bioactivos

Los péptidos bioactivos obtenidos por hidrólisis de proteínas, han demostrado que tienen potentes actividades biológicas entre las cuales destacan las actividades antihipertensiva, antioxidante, inmunomoduladora, anticancerígenas, antimicrobiana y antilipídica. (Meisel , 2004); (Wang & Gonzalez de Mejia, 2005); (Korhonen & Pihlanto, 2006)

La actividad biológica de los péptidos se debe a sus propiedades estructurales y a las características físico-químicas de los aminoácidos que los conforman. Estas actividades son consideradas como inferiores a la que ofrecen peptidomiméticos sintéticos y los fármacos desarrollados, pero las grandes ventajas que ofrecen los

péptidos bioactivos que son ingeridos por medio de la alimentación, incluyen la seguridad de consumir un producto de origen natural, el bajo costo comparado con el uso de fármacos y los beneficios nutricionales que se obtienen debido a que están compuestos por aminoácidos esenciales. El tamaño de estos péptidos varían de di-, tri-, y oligopéptidos a polipéptidos de alto peso molecular.

2.4.1. Actividad Antihipertensiva de Péptidos Bioactivos

La actividad antihipertensiva de los péptidos bioactivos se centra en la inhibición de la enzima convertidora de la Angiotensina (ECA), esta enzima es parte del Sistema Renina-Angiotensina. Este efecto de inhibición se da cuando los péptidos bioactivos actúan sobre los centros activos de la ECA, estos centros activos están conformados por los residuos de funcionales de los aminoácidos Tyr, Arg, Glu y Lys. Los péptidos bioactivos que actúen sobre estos centros activos deben ser pequeños o que tengan una afinidad alta con los centros activos de los residuos funcionales de la ECA. A partir de estos datos se han diseñado e identificado péptidos a partir de proteínas naturales. La evidencia farmacológica en la producción de medicamentos que inhiban la ECA, muestra que en el caso del captopril o el enalapril, responden a la misma teoría, en el caso de estos dos medicamentos, la estructura basal corresponde a la unión de aminoácidos, Ala-Pro y Phe-Ala-Pro respectivamente. (*Matsui & Tanaka, 2010*)

Se han reportado más de 400 péptidos que tienen acción inhibitoria de la ECA, entre las características de estos péptidos se destaca, que los residuos de los aminoácidos son hidrófobos, y que el extremo C-terminal son aromáticos como son la Tyr, Phe, Trp y Pro. Estos péptidos han mostrado tener una alta capacidad de inhibir la ECA con IC_{50} menor a 100 $\mu\text{mol/Litro}$. El IC_{50} se define como la concentración necesaria de un compuesto para reducir, in vitro, la actividad de la ECA, en este caso, en un 50% de la población. (*Matsui & Tanaka, 2010*)

En estudios en humanos se observó que después de 4 semanas en las cuales se los voluntarios consumieron 4g/día de sardina hidrolizada, la presión arterial disminuyó en (Presión Arterial Sistólica/Presión Arterial Diastólica: 9.3/5.2 mm Hg). (*Kawasaki, et al, 2000*) En este estudio se pudo obtener información muy útil, pues al consumir por medio de la alimentación antihipertensivos naturales no se presentó fenómeno de

rebote (cambios bruscos de la presión arterial) ni efectos secundarios, como presentan los fármacos inhibidores de la ECA. Evidenciando de esta forma que el consumo de productos alimenticios FOSHU (termino japonés adoptado por la FAO para alimentos de uso específico en la salud), son una herramienta para combatir enfermedades, en este caso disminuir la presión arterial e inhibir la ECA.

2.4.2. Cuantificación de la Actividad Antihipertensiva de Péptidos Bioactivos

Para cuantificar la Actividad Antihipertensiva de los Péptidos Bioactivos se utiliza la técnica desarrollada por Cushman y Cheung en 1971 en la cual, se hace reaccionar la ECA con hipuril histidil leucina dando como resultado ácido hipúrico y histidil leucina. La cantidad de ácido hipúrico liberado está relacionado directamente con la actividad de la ECA. El producto obtenido de esta reacción se trata con acetato de etilo para extraer el ácido hipúrico. La capa de acetato de etilo se separa del resto de la mezcla con mucho cuidado, se evapora para obtener el ácido hipúrico, finalmente se disuelve el ácido hipúrico en agua y se hace la medición de absorbancia a 228nm (*Cushman & Cheung, 1971*). Para cuantificar la actividad de los péptidos bioactivos se procede de la misma forma, aumentando inicialmente el péptido a la ECA y al hipuril histidil leucina.

2.5. Mecanismo antihipertensivo de los péptidos bioactivos

2.5.1 Inhibición del sistema renina-angiotensina

La enzima convertidora de la angiotensina (ECA), interviene el sistema renina-angiotensina provocando la hidrólisis de la Angiotensina I, decapeptido, obteniendo de esta forma Angiotensina II y el dipéptido Hys-Leu. Como ya se había descrito anteriormente la Angiotensina II eleva la potencia vasoconstrictora. La inhibición de la ECA provoca el descenso de la presión arterial, esto se debe a que los péptidos pueden relacionarse con los centro activos de la ECA. (*Hernández Ledesma, 2002*)

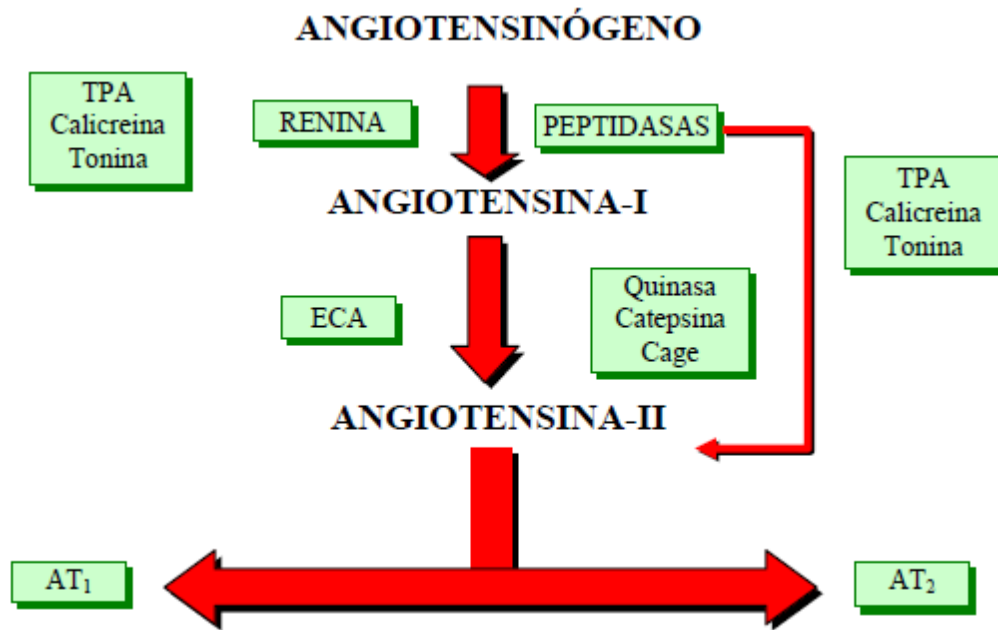


Figura 1: Vía de síntesis de la Angiotensina II

Fuente: Sellen 2007. Vía de síntesis de la Angiotensina II independientes de la ECA

El péptido con secuencia C-terminal Phe-Ala-Pro, es el más favorable para unirse al centro activo de la ECA, mientras que el grupo guanidino y el grupo ϵ -amino de la Arg y Lys, respectivamente, contribuyen a potenciar la acción inhibitoria de los péptidos. Los péptidos con presencia de aminoácidos dicarboxílicos o Pro como penúltimo aminoácido disminuyen o anula la actividad inhibitoria de la ECA. Cuando consideramos el extremo N-terminal la presencia de Val o de Ile incrementan la actividad inhibitoria del péptido (*Ondetti, et al, 1971*)

CAPITULO III

QSAR

Introducción

En el presente capítulo se dará a conocer información acerca de la técnica QSAR, como se utiliza y en que se fundamenta, al hablar de QSAR, es necesario también indicar el empleo de Descriptores Moleculares, se hará una explicación de los Descriptores Moleculares, su clasificación y una breve explicación de cada componente de la clase.

3.1 QSAR (Quantitative Structure-Activity Relationship)

3.1.1. Utilización de modelos matemáticos en Química Teórica

La utilización de modelos matemáticos en química teórica, responde a la necesidad de ajustar datos obtenidos de mediciones experimentales u observaciones, a un modelo que permita cuantificar, predecir o evaluar propiedades de compuestos químicos, con fines fisicoquímicos, biológicos o farmacológicos. Cuando el desarrollo de un modelo responde al empirismo permitiría realizar predicciones previas a la ejecución de nuevos experimentos, así mismo cuando el modelo responde a una teoría o ley, la evaluación del ajuste de datos, a un modelo, permitiría obtener una perspectiva de los procesos que se realizarán posteriormente a las observaciones efectuadas al momento del desarrollo del modelo. Las características predictivas de un modelo desarrollado no necesariamente tienen como finalidad la cuantificación de propiedades, sino también la cualificación de estas. De esta forma el desarrollo de un modelo matemático en química teórica facilitará el desarrollo y formulación de nuevas hipótesis de investigación en base a los datos experimentales y las observaciones obtenidas en procesos previos. (*Duchowicz, 2005*).

Al estudiar las propiedades de las moléculas químicas, se debe considerar que estas no dependen exclusivamente de la reactividad química de las moléculas sino también

de comportamientos teóricos derivados de la Mecánica Cuántica, los que permiten describir de forma más precisa estos fenómenos. Se debe considerar que, en la actualidad, los cálculos mecanocuánticos se puede realizar con una buena aproximación en niveles teóricos adecuados, esto se refiere a que estos cálculos son factibles en moléculas de pocos átomos que no presenten interacciones intermoleculares, esto se justifica debido a que la calidad de los resultados finales obtenidos no es totalmente satisfactoria. (*Duchowicz, 2005*).

Para el desarrollo de estos modelos se utiliza la teoría QSPR-QSAR

3.1.2. Fundamentos QSPR-QSAR

Los estudios de las estructuras de las moléculas químicas ha permitido el desarrollo de las teorías de estudio de las Relaciones Cuantitativas Estructura-Propiedad (QSPR en sus siglas en inglés) y el estudio de las Relaciones Cuantitativas Estructura-Actividad (QSAR en sus siglas en inglés), los cuales tienen sus aplicación más importantes en Fisicoquímica, Química Medicinal, Quimiometría, la Teoría de Grafos Químicos, Química Matemática, Química Computacional y Modelado Matemático.

Al observar en un conjunto de estructuras moleculares sus propiedades, medidas experimentalmente, cabe hipotetizar si existe algún tipo de relación entre ellas, considerando que esta hipótesis sea válida nace la pregunta, ¿Se podrá cuantificar la relación existente entre la estructura de la molécula y su propiedad? Esta hipótesis se encaja en la teoría QSRP-QSAR, que es una hipótesis matemática fundamentada en que la estructura de una molécula está relacionada directamente con la propiedad experimental medida de esta. (*Duchowicz, 2005*).

Los investigadores al correr de los años y el desarrollo científico y tecnológico, han logrado establecer relaciones estructura-actividad (SAR en sus siglas en inglés). En los estudios SAR se utiliza la información de proporcionada por la estructura molecular de los compuestos químicos, para predecir la actividad de estos de acuerdo al carácter del estudio que se esté realizando (*Todeschini, Introduzione alla Chemiometria, 2003*). Las suposiciones en las cuales se fundamenta los estudios SAR, según Todeschini, son:

- Un compuesto similar corresponde a una actividad similar

- Una modificación es la estructura molecular coincide con cambios en la actividad del compuesto

Basándose en estas suposiciones los estudios SAR responden a la siguiente relación matemática:

ACTIVIDAD = f (propiedades fisicoquímicas y/o propiedades estructurales)

En los estudios QSPR-QSAR, se combinan métodos de Mecánica Estadística con la Química Computacional, que son los que permiten desarrollar, validar y establecer modelos matemáticos que cuantifiquen las relaciones SAR.

En la teoría QSPR-QSAR, se define a la estructura molecular como la disposición tridimensional de los átomos en una molécula, la información requerida para este propósito se obtiene de los denominados descriptores moleculares, que son variables numéricas que reflejan distintos aspectos de la estructura molecular. Estos descriptores se han desarrollado utilizando la Teoría de Grafos, las aproximaciones de la Química Cuántica, la Teoría de la Información y Aproximaciones Geométricas. (*Todeschini, 2003*).

En la actualidad se han desarrollado miles de descriptores moleculares, de los cuales es necesario realizar una selección adecuada acorde a los objetivos del estudio, esta selección se realiza al utilizar la Química Computacional para el desarrollo del modelo matemático buscado. (*Todeschini, 2003*).

3.2. Descriptores Moleculares

Los descriptores moleculares han sido desarrollados por diferentes grupos de investigación tanto en forma aislada o en grupos interdisciplinarios para lo cual se han utilizado varias teorías para su desarrollo y validación. Al existir miles de descriptores moleculares la explicación de cada uno de ellos sería muy extensa, por tal motivo se puede describirlos en conjuntos de acuerdo a las características comunes existentes entre ellos. La clasificación de los descriptores responde a la representación de la estructura de las moléculas, en este caso se los agrupa de acuerdo a su dimensión, los descriptores de dimensión cero se catalogan como 0D, los de dimensión 1 como 1D, los de dimensión 2 como 2D y los de dimensión 3 como 3D. Vale acotar que dentro de esta clasificación existen subclasificaciones. (*Mercader, 2009*).

3.2.1. Descriptores 0D

Estos descriptores describen la constitución de la molécula y son independientes de la conectividad o conformación molecular, entre estos descriptores se encuentran el número de átomos de la molécula, algunos tipos de enlaces químicos, peso molecular, peso atómico, suma de propiedades atómicas (Volúmenes de Van der Waals). Vale acotar que en este grupo de descriptores no se puede diferenciar los isómeros de las moléculas.

3.2.2. Descriptores 1D

En este grupo se considera las fracciones de la molécula como un subconjunto de átomos, entre estos descriptores se consideran al número de grupos funcionales químicos, el número total de átomos de carbono primario, hidrógenos enlazados a un heteroátomo, hidrógenos enlazados a un carbono alfa, Flúor enlazado a un Carbono primario entre otros.

3.2.3. Descriptores 2D o Topológicos

Estos descriptores se han obtenido de la teoría de grafos, y son independientes de la conformación de la molécula. En este grupo de descriptores ya se presentan subclasificaciones de acuerdo a la teoría que se utilizó para su desarrollo. Estos descriptores son los de mayor estudio en la actualidad para interpretación de conceptos cuánticos.

3.2.3.1. Descriptores de Autocorrelación

En estudios realizados por Broto, Merleau y Geary, introdujeron diferentes variables estructurales que corresponden a autocorrelaciones bidimensionales entre pares de átomos, utilizadas para reflejar la contribución de una propiedad atómica con la propiedad de la molécula, las variables que se utilizan en estos descriptores son la electronegatividad, masa atómica, polarizabilidad atómica y el volumen de Van der Waals.

3.2.3.2. Descriptores BCUT

Estos descriptores se obtienen de la matriz de Burden, esta matriz es un grafo molecular al cual se le han quitado los hidrógenos, definido de la siguiente manera:

- Los elementos diagonales son los números atómicos de los elementos químicos
- El elemento fuera de la diagonal (B_{ij}), representa átomos enlazados i y j iguales a $\pi^* \times 10^{-1}$, donde π^* es el orden del enlace convencional (1, 2, 3, 1.5 correspondientes a simple, doble, triple y aromático respectivamente).
- Los elementos fuera de la diagonal corresponden a enlaces terminales se incrementan en 0.01
- El resto de elementos de la matriz se fija en 0.001 contribuciones de todos los átomos, reflejando de esta forma la topología de la molécula.

Los descriptores BCUT, son una extensión de los autovalores de Burden, consideran tres clases de matrices cuyos elementos diagonales corresponden a valores relacionados con la carga atómica, polarizabilidad y habilidad de unirse a átomos de hidrógeno (*Mercader, 2009*)

3.2.3.3. Índices de carga Topológica de Gálvez

Estos índices describen la transferencia de carga entre pares de átomos, y en su conjunto la transferencia de carga global de la molécula. Estos descriptores correlacionan la distribución de cargas en las moléculas, y por lo tanto también con los momentos dipolares de hidrocarburos aromáticos e insaturados. Su desarrollo se basa en la utilización de la matriz de adyacencia y el recíproco cuadrado de la matriz de distancia. (*Mercader, 2009*)

3.2.3.4. Cuenta de Trayectos Moleculares

Son descriptores obtenidos contando trayectos y caminos moleculares tanto de ida como de ida y vuelta a diferentes longitudes dentro de la molécula, estos trayectos son de orden 1 a orden 10, y dependerán del tipo de enlace existente entre cada una de los átomos que se encuentren conformando la molécula. (*Mercader, 2009*)

3.2.3.5. Índices de Conectividad

Son descriptores calculados a partir de grado del vértice del átomo a partir de un grafo libre de hidrógenos, siendo la base para el desarrollo de estos descriptores moleculares, las matrices que se generan en la teoría de los grafos, base del desarrollo de la mayoría de descriptores moleculares. (Mercader, 2009)

3.2.4. Descriptores 3D

En el desarrollo de estos descriptores se considera la conformación de la estructura molecular, de esta forma se puede describir propiedades estereoquímicas de las moléculas. La matriz molecular M, está conformada por las coordenadas cartesianas de cada átomo en la geometría molecular optimizada, esta es la base para el cálculo de estos descriptores. El método de optimización geométrica utilizado es el Campo de Fuerza de la Mecánica Molecular MM+, el cual resulta de la combinación de métodos derivados de la Mecánica Cuántica

3.2.4.1. Perfiles moleculares de Randić

Estos descriptores son derivaciones de la distribución de la distancia de la matriz G, que es la matriz de la distancia Geométrica, están definidos como el promedio de las sumas por columnas de los elementos elevados a una potencia n, normalizados por el factor n!. (Mercader, 2009)

3.2.4.2. Descriptores RDF (Funciones de Distribución Radial)

Estos descriptores se basan en la distribución de las distancias en la geometría molecular, esta distribución puede interpretarse como la probabilidad de encontrar un átomo en un volumen esférico de cierto radio centrado en un punto determinado, parte de la información que estos descriptores ofrecen son las longitudes de enlace, distancias interatómicas, tipos de anillos, sistemas planos y no planos y el tipo de átomos presentes. (Mercader, 2009)

3.2.4.3. Descriptores WHIM

Los descriptores WHIM se basan en índices estadísticos calculados durante la proyección de los átomos a lo largo de los ejes principales, la información obtenida tiene relación con el tamaño molecular, la forma, simetría y distribución atómica con respecto a una referencia invariante. Los descriptores WHIM pueden ser direccionales y globales, que se diferencian entre sí la información obtenida proviene de los ejes de la estructura o de toda la molécula. (*Mercader, 2009*)

3.2.4.4. Descriptores 3D MoRSE

Estos descriptores son la Representación Molecular 3D de las Estructuras basada en la Difracción Electrónica, la información de estos descriptores se ha obtenido de los estudios de difracción electrónica que permite reflejar la distribución tridimensional de distintas propiedades atómicas en la molécula. Estos descriptores han demostrado describir correctamente la ramificación molecular (*Mercader, 2009*)

3.2.4.5. Descriptores GETAWAY

Los descriptores GETAWAY consisten en un ensamblado de: Pesos de Átomos, Geometría y Topología. Fueron diseñados con el propósito de tratar de conectar la estructura tridimensional de la molécula, provista por la matriz de influencia y la información química utilizando diferentes esquemas de pesos atómicos (*Mercader, 2009*).

3.2.4.6. Descriptores Geométricos

Estos descriptores se obtienen del cálculo de la geometría molecular, donde se consideran: las longitudes de enlace (determinadas por los tipos de átomos y la multiplicidad), ángulos de enlace, ángulos diedros, etc., como ejemplos de estos descriptores se encuentran los índices gravitacionales y el índice de Wiener tridimensional (*Mercader, 2009*).

3.2.4.7. Descriptores de Carga

Estos descriptores describen la distribución de cargas en la molécula, estos descriptores son útiles solo cuando la molécula ha sido optimizada mediante la utilización de algún método mecano cuántico. Entre los descriptores de este grupo se encuentra: la suma de las cargas atómicas, la carga atómica positiva máxima, la carga atómica negativa máxima, la carga positiva total, la carga cuadrática total y las energías HOMO y LUMO. (*Mercader, 2009*).

CAPITULO IV

ANALISIS MULTIVARIANTE

Introducción

En este capítulo, se hará una breve descripción del Análisis Multivariante específicamente de las Técnicas Multivariantes que se utilizarán para el desarrollo de este trabajo. Para ello se tratará primero la Selección de Variables y como técnica específica los Algoritmos Genéticos, se hará una descripción tanto de los modelos de Regresión como de los modelos de Clasificación, para ello se considerará el empleo del Análisis de Regresión Multivariante, el método de los Mínimos Cuadrados Ordinales y los métodos K-means y KNN para clasificación.

4.1. Análisis Multivariante

El Análisis Multivariante, como su nombre lo indica, relaciona muchas variables independientes con una respuesta o propiedad, en este análisis lo que se pretende es desarrollar modelos matemáticos que relacionen estas variables independientes con la respuesta o propiedad medida.

4.1.1. Selección de Variables

En los estudios QSAR, se generan un gran número de variables independientes (descriptores moleculares), considerando que existen más de 4000 descriptores moleculares, es necesario realizar una selección previa de variables, para ello se utilizan diversos métodos de selección de variables, uno de estos métodos son los Algoritmos Genéticos.

4.1.1.1 Algoritmos Genéticos

El método de Algoritmos Genéticos se basa en las reglas de la evolución biológica de los organismos vivos, este método consiste en crear una población de individuos

(regresiones lineales), los que luego se reproducirán apareándose o mutando, lo que provoca una evolución a través de generaciones sucesivas hasta alcanzar una solución final óptima. La característica más importante de los Algoritmos Genéticos es que busca muchas soluciones posibles al mismo tiempo. Cada una de estas soluciones posibles explora diferentes regiones del espacio de los descriptores.

Cada individuo de la población está representado por un vector de números enteros, cuyos elementos son la numeración de los descriptores llamados genes.

El primer paso es crear una población de individuos (IND), cada individuo posee el mismo número de genes de forma aleatoria a partir de todos los descriptores utilizados (D) cumpliendo las siguientes reglas:

- Dos individuos no pueden tener los mismos genes
- Todos los genes del individuo deben ser diferentes entre sí.

El siguiente paso es la obtención de hijos, estos hijos se producen por entrecruzamiento de sus padres y por mutación. La forma en la que se obtienen los hijos es por reproducciones sexual y asexual de esta forma la nueva generación tendrá características de ambos o uno de los padres. En la reproducción sexual los individuos que serán los padres se eligen probabilísticamente en base a su posicionamiento en la escala de ajuste, en este entrecruzamiento los padres contribuyen aleatoriamente con la mitad de sus genes. La reproducción asexual se obtiene cuando los padres elegidos al azar mutan en forma aleatoria por consecuencia uno de sus genes es remplazado por otro, dentro de esta reproducción, existe la probabilidad que el individuo no se altere y transmita los mismo genes a su hijo (*Mathworks Inc, 2004*) Para una mejor comprensión del funcionamiento de los algoritmos genéticos la figura 2, muestra su funcionamiento.

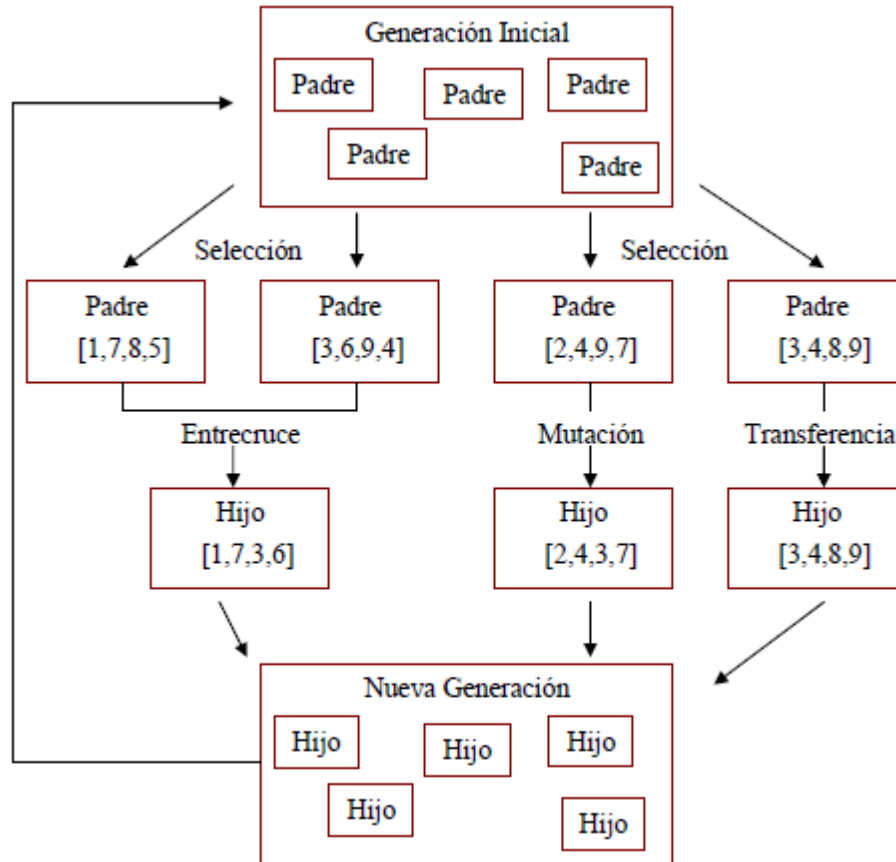


Figura 2: Funcionamiento Algoritmos Genéticos

Fuente: (Duchowicz, 2005)

Los procesos de selección, entrecruzamiento y mutación, se repite las veces que sean necesarias hasta que los IND padres de la población son reemplazados por IND hijos, la escala de ajuste de cada miembro de la nueva generación es evaluada en forma repetida, hasta que el 90% de las generaciones tengan el mismo ajuste.

4.1.2. Modelos de Regresión

Los modelos de regresión son aquellos que generan un modelo matemático que permite cuantificar la propiedad buscada.

4.1.2.1. Análisis de Regresión Multivariante

El Análisis de Regresión Multivariante (MRA), resulta de la combinación del método de Mínimos Cuadrados y otros métodos derivados de la Estadística Matemática. Cuando se realiza una MRA lineal, dentro de los estudios QSPR-QSAR, su éxito resulta si se obtiene el cumplimiento de la relación propuesta (relación lineal), entre la variable dependiente (propiedad o respuesta a estudiarse) con las variables independientes (descriptores moleculares).

Este Análisis en su desarrollo no considera: la precisión de los experimentos utilizados para medir la propiedad a estudiarse, los errores sistemáticos resultantes de la medición de la propiedad, ni la precisión teórica con la que se calculan los descriptores moleculares. El MRA solo cuantifica el comportamiento de los errores aleatorios o residuos de la regresión. (*Duchowicz, 2005*)

El método MRA se basa en las siguientes admisiones matemáticas:

- Para cada variable independiente, las variables dependientes están distribuidas normalmente, cumpliendo la ley límite de distribución de Gauss. La distribución de Gauss supone que la probabilidad de cometer errores negativos es igual a la de cometer errores positivos, por lo tanto, los errores son completamente aleatorios, y que la probabilidad de cometer errores muy grandes es mínima y tiende a cero
- Las variables dependientes para distintos valores de las variables independientes son estadísticamente independientes, lo que indica que la probabilidad de cometer un error no está influenciada por la historia previa, y su generación no influirá en la historia posterior
- La precisión de las variables independientes es mucho mayor que la precisión de las variables dependientes. Esto supone que los datos de partida son precisos.

4.1.2.2. Métodos de los Mínimos Cuadrados Ordinarios (OLS)

Es un método utilizado en regresión lineal, el cual estima los parámetros desconocidos de un modelo de regresión lineal, el cual minimiza la suma de cuadrados de las distancias verticales entre las respuestas observadas en el grupo de datos y las respuestas predichas por aproximación lineal.

El estimador calculado es consistente cuando los regresores son exógenos y no hay perfecta multicolinealidad, correlación entre las variables, y es óptimo en la clase de

estimadores lineales cuando los errores son homoscedásticos, cuando la varianza del error de la variable endógena se mantiene a lo largo de las observaciones, y no hay correlación serial.

Con estas condiciones, el método OLS proporciona la mínima varianza media-insesgada estimada cuando los errores tienen varianzas finitas. Bajo la suposición adicional de que los errores se distribuyen normalmente, el estimador OLS es el de máxima verosimilitud. (*Todeschini, 2003*)

4.1.3. Modelos de Clasificación

En los modelos de Clasificación, lo que se obtiene son modelos matemáticos que agrupan las respuestas en clases, definidas por atributos específicos que permitan identificar características distintivas para cada una de ellas.

4.1.3.1. K-means

Es un algoritmo creado en 1967 y de amplia difusión, debido a su aplicación simple. Consiste en una aproximación por etapas sucesivas un cierto número de clusters, para ello utiliza los centroides de los puntos que se representan. Para realizar K-means se siguen los siguientes pasos:

- Situa K puntos en el espacio donde se encuentran los objetos a clasificar. Estos K puntos representan los centroides iniciales de los grupos.
- Cada objeto es asignado al grupo que tiene el centroide más cercano.
- Cuando todos los objetos han sido asignados a una clase, se recalcula las posiciones de los K centroides.
- Este paso se repite las veces que sean necesarias hasta que los centroides de las clases se mantengan estables, para así producir la clasificación de los objetos en grupos que permitan dar una métrica entre ellos. (*Sancho Caparrini, 2009*)

4.1.3.2. KNN

El método KNN, es un método de clasificación no paramétrico, el cual se base en el concepto de analogía, se fundamenta en la elección de una distancia y una selección

oportuna de los K objetos más cercanos a la molécula que se está estudiando. (*Gemperline, 2007*), posterior a esto enumera cuántos de los K objetos se ubican en cada clase, asignando a la molécula a la clase más numerosa, extendiéndose a cada uno de los objetos en estudio. (*Todeschini, 2003*)

Todeschini en el 2003 explica las etapas del algoritmo KNN, las cuales son:

- Escalado de los datos
- Selección de la distancia a utilizar
- Elección del número de K útil para la clasificación
- Cálculo de la matriz de distancias
- Para cada objeto se consideran los K objetos más cercanos
- El objeto se asigna a la clase más representativa en función de los K objetos más cercanos.

Cuando se aplica el método KNN es necesario probar diversos valores de K, estos valores de K oscilan entre 1 y 10, hasta lograr obtener el valor óptimo, el cual se puede identificar al calcular el error de los datos utilizados para el desarrollo del modelo. (*Todeschini, 2003*)

CAPÍTULO V

METODOLOGÍA

Introducción

En el presente capítulo se describirá la metodología utilizada para el desarrollo de los modelos, empezando por la obtención de los datos que generará la base de datos que se utilizarán, la digitalización y construcción de las moléculas de la base de datos, el cálculo de los descriptores moleculares para cada una de las moléculas de la base de datos y los métodos Multivariante que se utilizarán para el desarrollo de los modelos matemáticos.

5.1. Metodología

5.1.1. Base de Datos

El primer paso para el desarrollo de un estudio QSAR, es la obtención de las moléculas de dipéptidos y tripéptidos con su respectiva actividad inhibidora de la ECA, para ello se realizó una revisión bibliográfica, estudiando diferentes fuentes como son artículos científicos y bases de datos digitales. La base de datos utilizada fue Peptide Database of Central Food Technological Research Institute de la India (Central Food Technological Research Institute), la cual consta de 135 artículos, publicados en diferentes revistas, de donde se obtiene 865 péptidos bioactivos con actividad inhibidora de la ECA, de las cuales 262 corresponden a los dipéptidos y tripéptidos que serán utilizados para el desarrollo del modelo.

La actividad de inhibición de la ECA (IC_{50} , la cantidad del péptido necesaria para inhibir el 50% de la ECA) se encuentra medida en μM . Los datos correspondientes al IC_{50} , fueron tratados matemáticamente aplicando el logaritmo al inverso de la propiedad

$$respuesta = \log\left(\frac{1}{IC_{50}}\right)$$

Esta base de datos generada se ha dividido en dos grupos, Training Set y Test Set. El Training Set, consta de 184 moléculas de di y tripéptidos elegidos al azar utilizando el programa Matlab R2008a, con el utilitario “ga toolbox”. Los valores tratados de las actividades de inhibición de la ECA de estas moléculas serán relacionados con los descriptores moleculares que se generen para el desarrollo del o los modelos. Mientras que los datos correspondientes a las 78 moléculas restantes, conformarán el Test Set, el Test Set será utilizado para validar el o los modelos obtenidos del Training Set.

Para calcular un modelo de Clasificación se someterán las respuestas a un proceso de clasificación previa para asignar una clase hipotética. El método utilizado para generar esta clasificación es el algoritmo K-means, y la clasificación propuesta es de tres niveles.

- Clase 1: Baja actividad de inhibición de la ECA
- Clase 2: Media actividad de inhibición de la ECA
- Clase 3: Alta actividad de inhibición de la ECA

Los centroides, el promedio de los baricentros, presentaron un comportamiento escalado en función de la actividad para cada clase. Siendo los valores promedio de cada clase los siguientes:

- Baja actividad de inhibición de la ECA: -1.0867
- Media actividad de inhibición de la ECA: 0.0948
- Alta actividad de inhibición de la ECA: 0.2329

La base de datos en la cual se encuentran los valores de la actividad inhibidora de la ECA, la respuesta tratada con el logaritmo del inverso de la actividad inhibidora, la asignación de Training y Test Set, y las clases asignadas por el método K-means se encuentra detallada en el ANEXO 1 (Base de datos de dipéptidos y tripéptidos con actividad inhibidora de la enzima convertidora de la Angiotensina)

5.1.2. Digitalización y construcción de las moléculas

Con estos datos el siguiente paso es construir y optimizar las moléculas, para ello se utilizó el programa Hyperchem 8.0 (*Hypercube Inc, 2008*). Las moléculas se construyen utilizando la base datos de aminoácidos del programa, siguiendo la secuencia de

aminoácidos, se ha considerado que por su pequeño tamaño la estructura de los aminoácidos responde a una hélice alfa, un ángulo omega de 180° , y a un isómero levorrotatorio. Luego de construida la molécula ésta es optimizada. Para la optimización de las moléculas, primero se debe definir el método a utilizar para el cálculo del campo de fuerza mecánico molecular como se muestra en la figura 3.

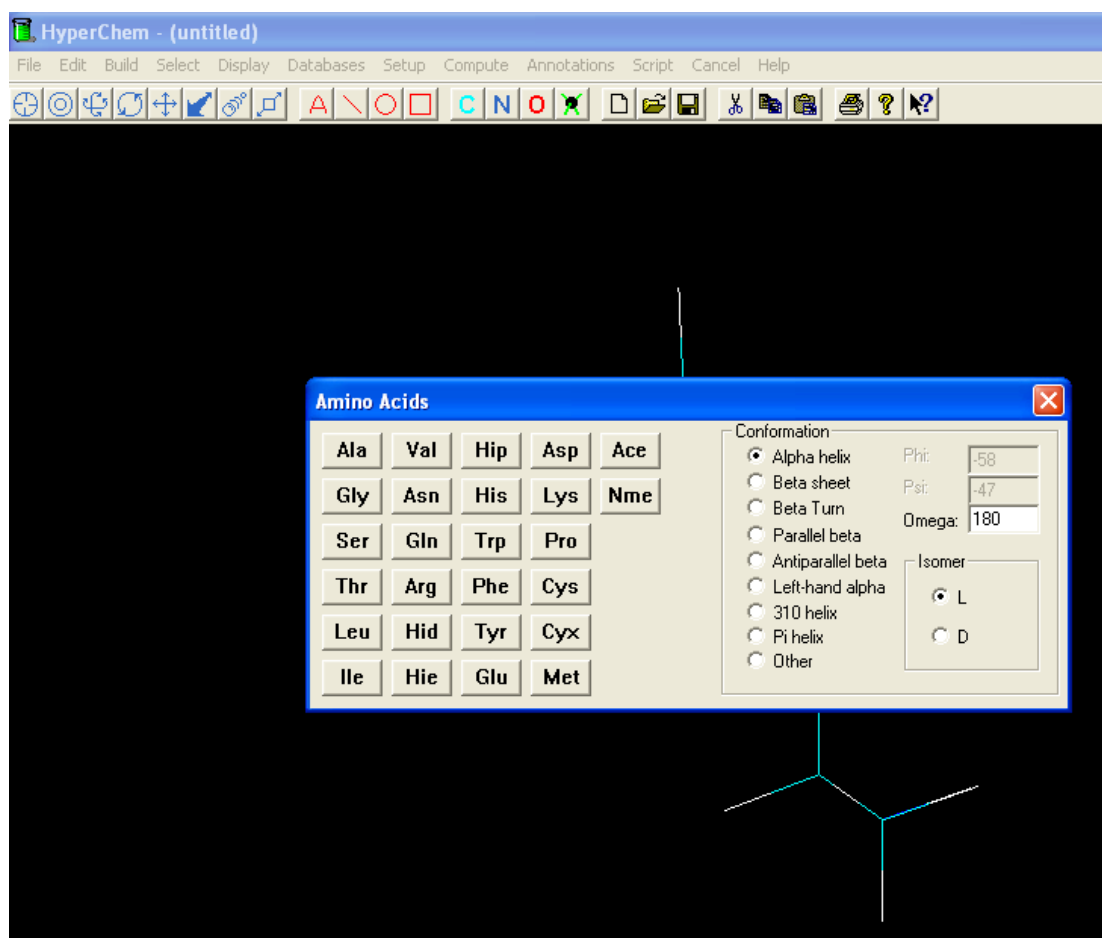


Figura 3: Características de los aminoácidos de la base de datos

Fuente: Software Hyperchem 8.0

En este caso el cálculo del campo de fuerza mecánico molecular, se realizó utilizando la herramienta MM+. Se han considerado todos los parámetros que esta herramienta puede utilizar para el cálculo, estos parámetros son:

- Enlace
- Ángulo
- Torsión
- Electroestática (enlaces dipolares)

- Enlaces con Hidrógenos

Las opciones y componentes de la herramienta MM+, se muestran en la figura 4 y 5

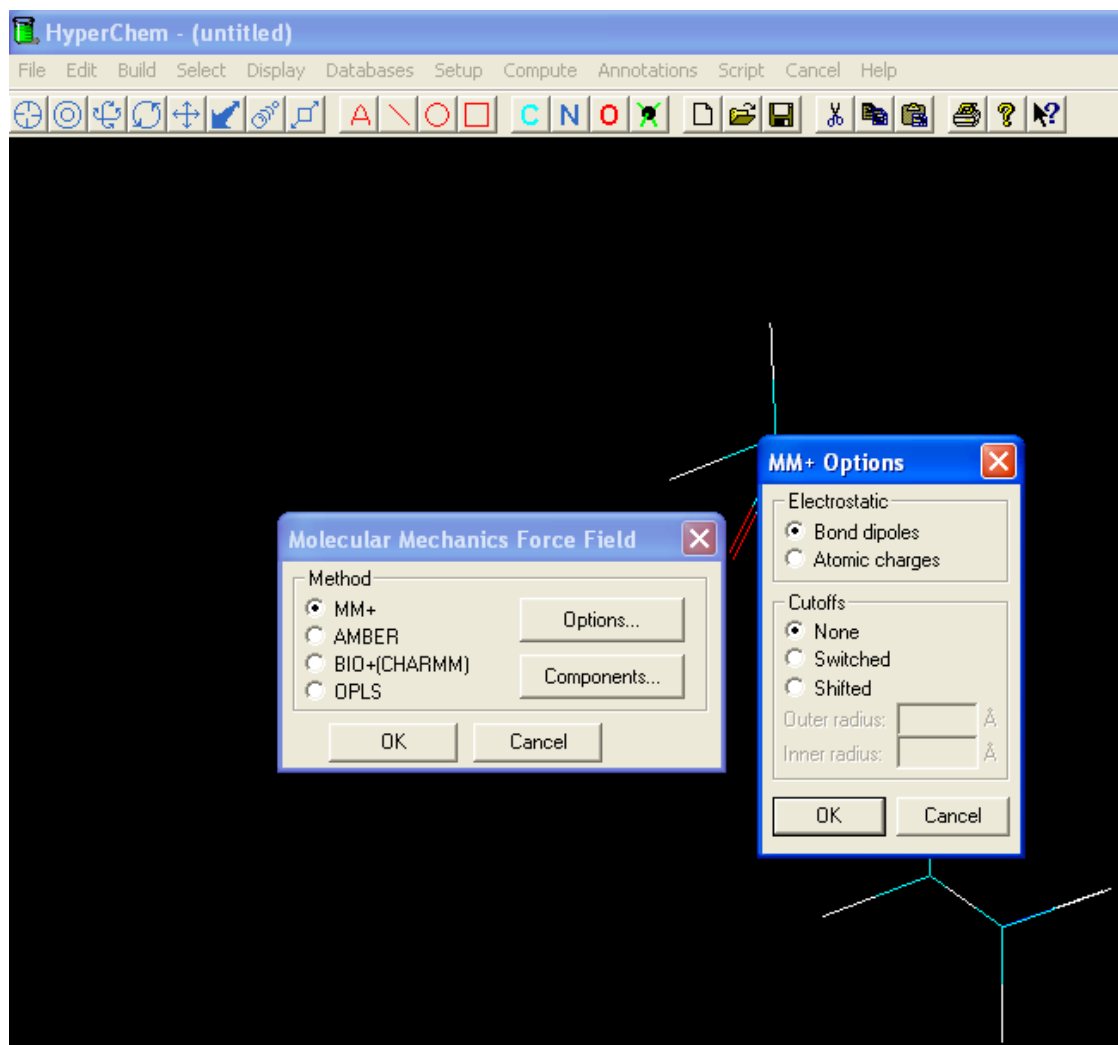


Figura 4: Opciones de la herramienta MM+

Fuente: Software Hyperchem 8.0

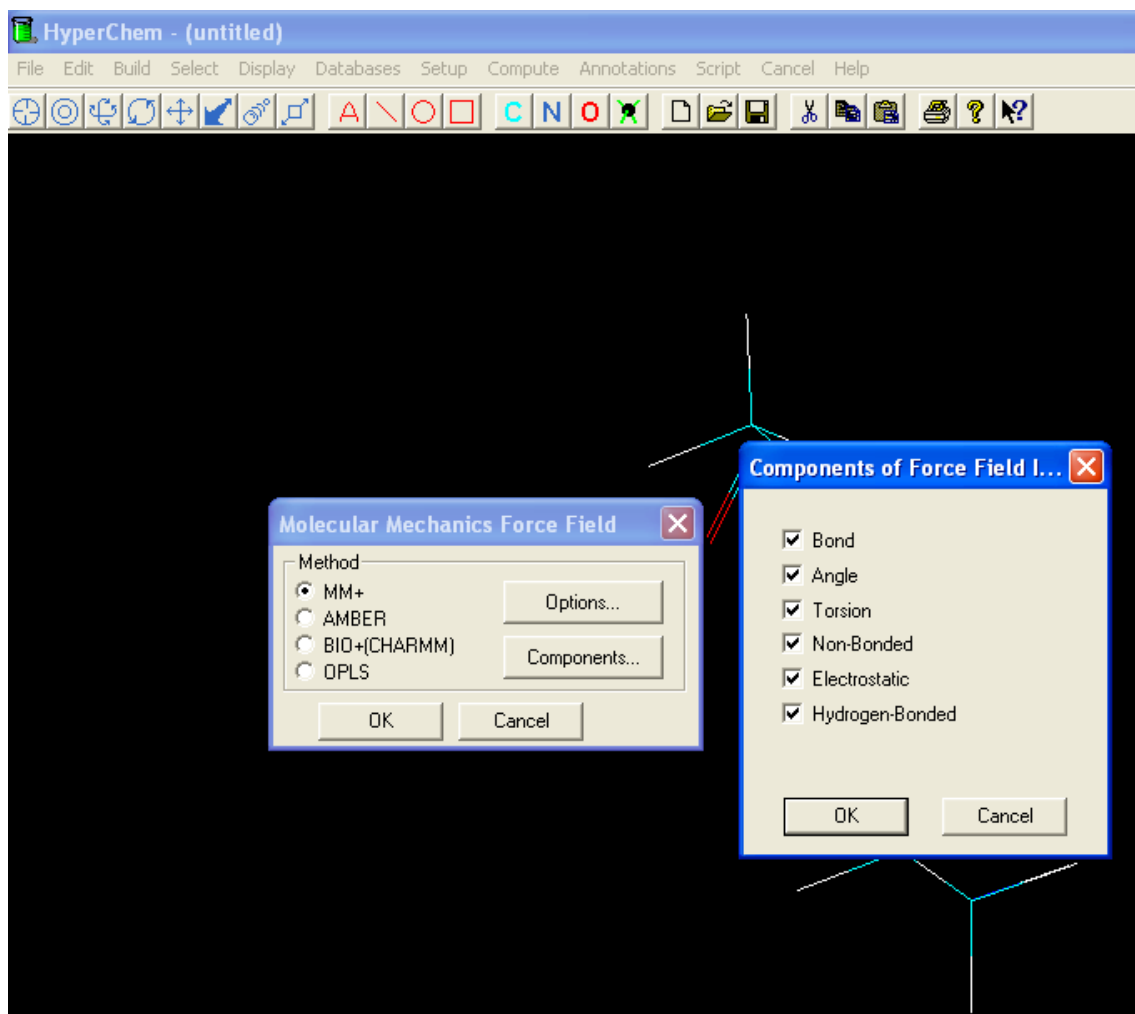


Figura 5: Componentes de la herramienta MM+

Fuente: Software Hyperchem 8.0

El siguiente paso en la optimización de la molécula es definir el algoritmo que se utilizará. En este caso se utilizó el algoritmo Polak-Ribiere, que es una optimización numérica que utiliza el método de gradiente conjugado no lineal [Navon], las opciones para la aplicación se establecen, mediante el ajuste del gradiente de la raíz cuadrada media (RMS), el cual determinará el final de los cálculos cuando la pendiente de RMS es menor al valor introducido para los cálculos, el valor introducido es de 0.01 Kcal como se muestra en la figura 6.

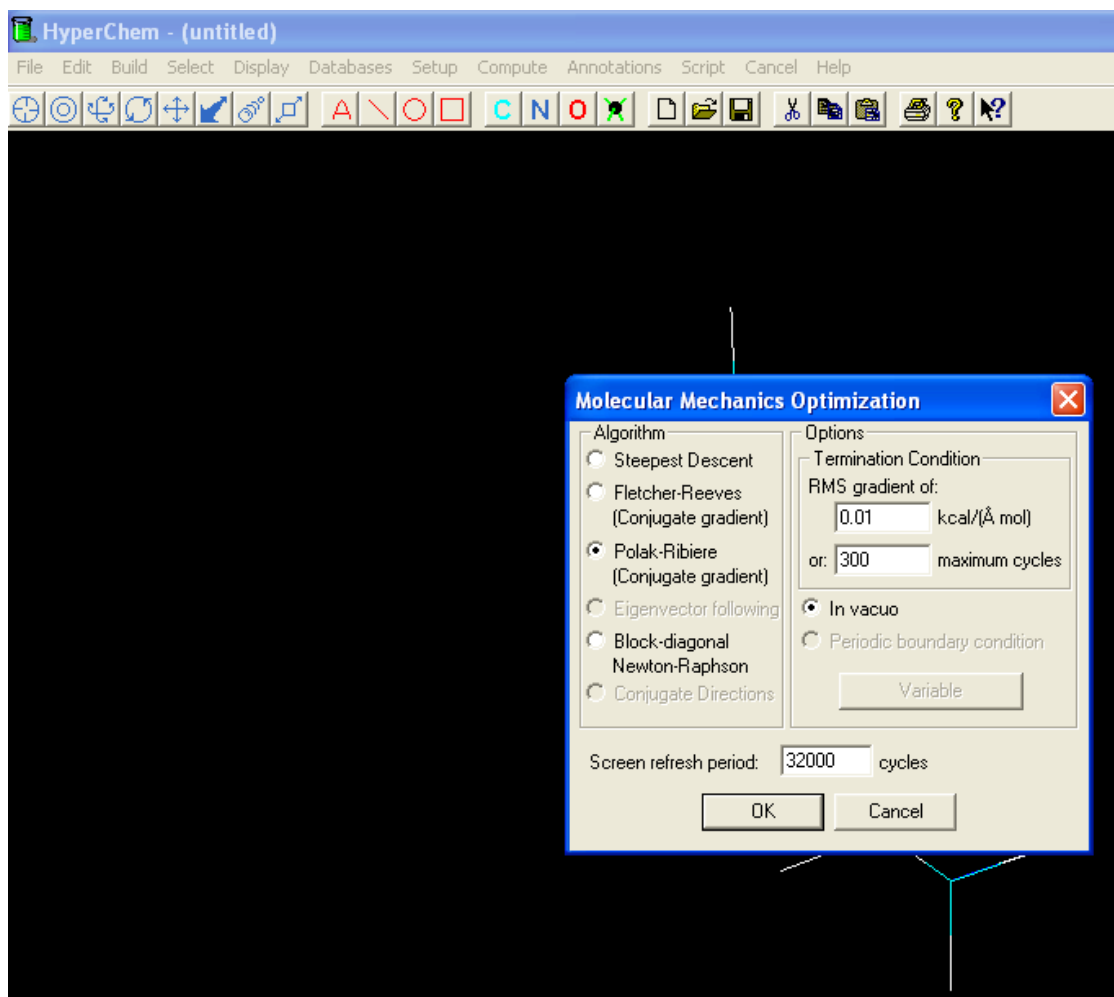


Figura 6: Optimización mecánica molecular

Fuente: Software Hyperchem 8.0

Establecidas estas condiciones se procede a calcular la optimización de la molécula. La molécula generada se guarda con un nombre específico en el formato del programa Hyperchem. Estas moléculas serán utilizadas para el cálculo de los descriptores moleculares. El ejemplo de una molécula optimizada se muestra en la figura 7

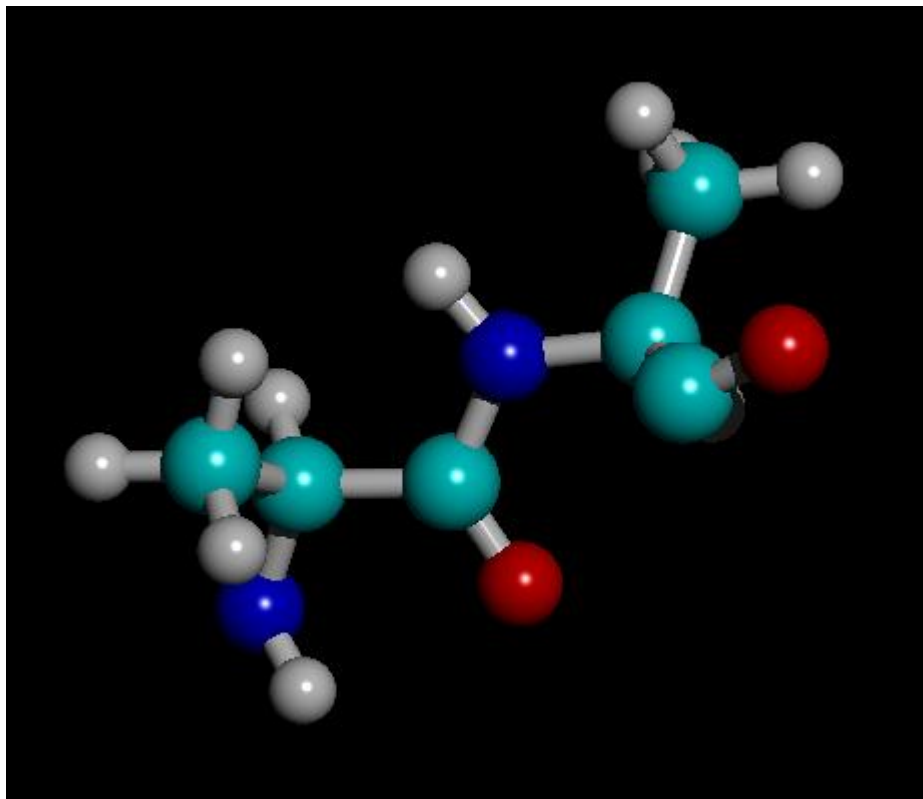


Figura 7: Molécula de Alanina-Alanina optimizada utilizando el software Hyperchem 8.0

Fuente: Software Hyperchem 8.0

5.1.3. Cálculo de Descriptores Moleculares

Para el cálculo de descriptores moleculares se utiliza el programa DRAGON 6.0 (Talete sr, 2010). El programa DRAGON, calculó 956 descriptores moleculares para las moléculas de di y tripéptidos digitalizadas.

Los descriptores que suministra DRAGON pueden describir la estructura de moléculas que contengan alguno de los siguientes tipos de átomos:

C, H, O, N, S, P, F, Cl, Br, I, B, Si, Ni, Fe, Co, Al, Cu, Zn, Sn, Cd

Por lo cual los di y tripéptidos, que contienen aminoácidos azufrados no fueron utilizados en el cálculo de descriptores moleculares. Los tipos de descriptores moleculares que se calcularon fueron:

- Descriptores constitucionales
- Descriptores topológicos
- Número de trayectos moleculares
- Descriptores BCUT

- Índices de carga topológica de Gálvez
- Descriptores de autocorrelación 2D
- Descriptores de carga
- Índices de aromaticidad
- Perfiles moleculares de Randic
- Descriptores geométricos
- Descriptores RDF
- Descriptores 3D-MoRSE
- Descriptores WHIM
- Descriptores GETAWAY
- Número de grupos funcionales
- Fragmentos centrados en el átomo

Debido a la cantidad de descriptores generadas, es necesario la selección de las variables a utilizar, para esto se utilizará el método de Algoritmos Genéticos.

5.1.4. Cálculo de los Modelos de Regresión y Clasificación

Para el cálculo de los modelos, se utilizará el programa Matlab R2008a, con el utilitario "ga toolbox", desarrollado por el Milano Chemometrics and QSAR de la Universidad Milano Bicocca (Italia) en este utilitario se encuentran desarrollados algoritmos y métodos para desarrollo de modelos de Regresión y Clasificación.

Para la obtención de un modelo de Regresión se utilizará el Método de los Mínimos Cuadrados Ordinarios (OLS), mientras que para la obtención de un modelo de clasificación se utilizará en método KNN.

CAPÍTULO VI

RESULTADOS Y DISCUSION

Introducción

En este capítulo se presentarán los resultados obtenidos de la aplicación del trabajo, donde se mostrarán los parámetros que se utilizaron, los descriptores moleculares que permitieron obtener el modelo matemático buscado, las características del modelo desarrollado y los parámetros de validación del modelo.

RESULTADOS Y DISCUSIÓN

En primera instancia se trató de obtener un modelo de regresión, aplicando el método OLS, el método OLS resulta ser un método muy robusto para el tratamiento de datos de diferentes fuentes, de los cuales no se tiene la certeza de que las condiciones de trabajo puedan fidelizar los datos. Se aplicó el método OLS, a los datos del Training Set, para obtener el modelo de regresión, usando el utilitario “ga toolbox” del programa Matlab R2008a, las opciones con los que se realizó el método OLS, se muestran en la tabla 6, y los parámetros obtenidos para el desarrollo del modelo se muestran en la tabla 7.

Opciones	
Método	OLS
Escalado	Ninguno
Grupos de Cross Validación	5
Tipo de Cross Validación	Veneciana
Número de Cromosomas	30
Probabilidad de Mutación	0.01
Probabilidad de Cruzamiento	0.5
Repeticiones	100
Evaluaciones	100
Tipo de distancia	Euclideana

Tabla 6: Opciones para el desarrollo de OLS

Fuente: Matlab R2008a

Parámetros Modelo OLS	
R^2 CV	-229.9546
Numero de Variables	6

Tabla 7: Parámetros del Modelo generados por OLS

Fuente: Matlab R2008a

Después de aplicar el método OLS se puede observar que se han escogido 6 variables, descriptores moleculares, pero su R^2 CV es igual a -229.9546, lo que significa que no existe un modelo posible de regresión para este tipo de moléculas y actividad. Se aplicó el método KNN, a los datos del Training Set, para obtener el modelo de clasificación, usando el utilitario “ga toolbox” del programa Matlab R2008a, para el desarrollo del modelo es necesario una selección de variables mediante la aplicación

de algoritmos genéticos, que el mismo utilitario posee. Para la ejecución del método es necesario establecer las condiciones de trabajo. Las condiciones de trabajo se muestran en la tabla 8

Opciones	
Método	KNN
Escalado	Ninguno
Grupos de Cross Validación	5
Tipo de Cross Validación	Veneciana
Número de Cromosomas	30
Probabilidad de Mutación	0.01
Probabilidad de Cruzamiento	0.5
Repeticiones	100
Evaluaciones	100

Tabla 8: Opciones para el desarrollo de KNN

Fuente: MatLab R2008a

Terminado el cálculo del modelo se procede a evaluar los parámetros de validación de la clasificación, si el No Error Rate (NER%) de un modelo es satisfactorio, este es aceptado. El NER%, de cada modelo en función del número de descriptores utilizados (variables) para su desarrollo, se muestran en la figura 8.

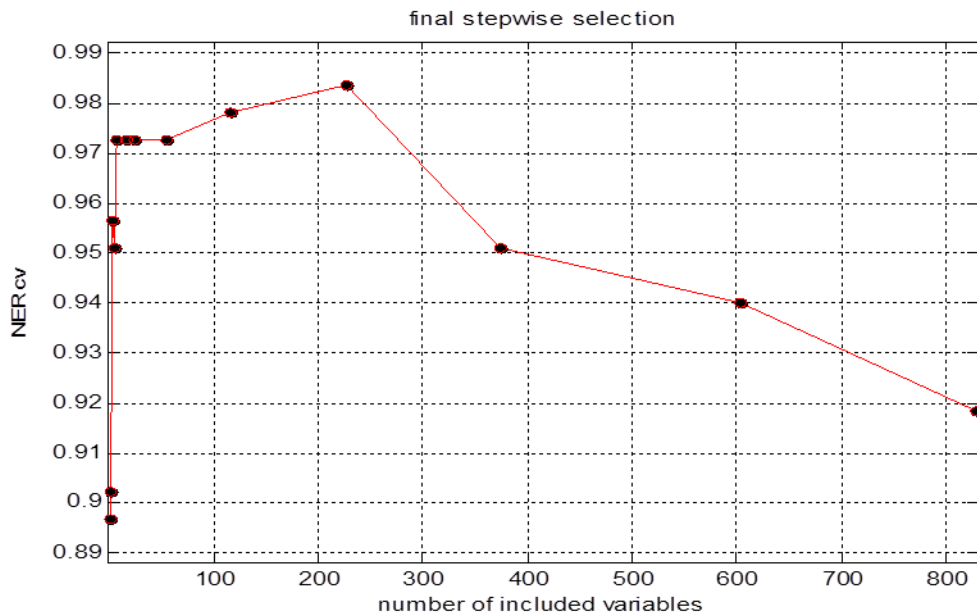


Figura 8: Relación entre número de descriptores considerados en la clasificación KNN y la calidad del modelo en predicción (NER cv)

Fuente: Matlab R2008a

El modelo con el NER más satisfactorio, considerando que para el desarrollo de un modelo es necesario que este tenga el menor de número de variables, es de 0.9728 que corresponde al modelo de 8 variables este valor quiere decir que la capacidad predictiva del modelo es del 97%. Las cuales fueron seleccionadas de los 956 descriptores moleculares, desarrollados para este estudio, por el método de Algoritmos Genéticos. Los valores del NER% para los 8 primeros modelos de clasificación calculados se muestran en la tabla 9. Los datos correspondientes a los descriptores moleculares calculados para las moléculas del Training Set se encuentran en el ANEXO 2 (Descriptores Moleculares Calculados y Seleccionados del Training Set)

Numero de Variables	NER	Descriptor
1	0.90217	Eig15_EA(bo)
2	0.89673	Eig15_AEA(dm)
3	0.92663	ATS7s
4	0.95652	RDF085u
5	0.95108	P_VSA_s_6
6	0.95833	Eig15_EA(dm)
7	0.96557	Mor02s
8	0.97282	R3s+

Tabla 9: Modelos calculados en función de los descriptores moleculares seleccionados con su valor de NER%

Fuente: MatLab R2008a

Después de obtener el modelo, se procede a calcular los centroides de cada uno de los descriptores moleculares para cada clase, en la tabla 10 se muestran los valores de los centroides para cada una de las clases, los valores de los centroides se graficaron para poder observar su comportamiento en cada clase y definir como cada descriptor aporta en cada clase. En la figura 9 se puede observar el comportamiento de los centroides.

Centroides de Clases del Modelo			
	KnnCen1	KnnCen2	KnnCen3
Eig15_EA(bo)	1.53308	-0.76811	0.383062
Eig15_AEA(dm)	0.73636	-0.846907	0.68357
ATS7s	-1.86761	-0.111247	0.750769
RDF085u	-1.23156	-0.483885	0.989326
P_VSA_s_6	-0.77527	-0.284323	0.632038
Eig15_EA(dm)	0.63478	-0.672481	0.494603
Mor02s	-0.5872	-0.384575	0.659346
R3s+	1.66651	0.194642	-0.79365

Tabla 10: Centroides de los descriptores para cada clase

Todeschini en el 2009 define a los descriptores de la siguiente forma:

Los descriptores Eig15_EA(bo), se el quinceavo autovalor de matriz M de la teoría de Grafos, estos autovalores corresponden a las raíces del polinomio característico de la matriz M, se calculan a partir de la matriz de adyacencia expandida y pesado por el orden de enlace.

Los descriptores Eig15_EA(dm), se el quinceavo autovalor de matriz M de la teoría de Grafos, estos autovalores corresponden a las raíces del polinomio característico de la matriz M, se calculan a partir de la matriz de adyacencia expandida y pesado por el momento del dipolo.

Los descriptores Eig15_AEA(dm), se el quinceavo autovalor de matriz de adyacencia M de la teoría de Grafos, estos autovalores corresponden a las raíces del polinomio característico de la matriz M, pesado por el momento del dipolo.

El descriptor ATS7s, es un descriptor de autocorrelación del potencial de ionización de la molécula, que se recalcula después de 7 enlaces.

El descriptor RDF085u, es un descriptor de función de distribución radial, se obtienen de la interpretación de la difracción de rayos X.

El descriptor P_VSA_s_6 está definido como el valor del área de la superficie de van der Waals de una propiedad P. Este descriptor corresponde a la partición de la área de la superficie de van der Waals de acuerdo a los valores atómicos de una propiedad P.

El descriptor Mor02s, describe la distribución de los átomos, pesados por el Potencial de Ionización, que produce una figura de difracción electrónica.

El descriptor R3s+, es un descriptor GETAWAY (Geometry, Topology, and Atom-Weights Assembly), es un descriptor topológico que se derivan de la Matriz de

Influencia Molecular (MIM), esta matriz es la representación de moléculas denotadas por H.

En la figura 9, se puede observar que el comportamiento de cada descriptor en las clases en su mayoría es diferente entre sí, por lo cual se puede caracterizar las clases.

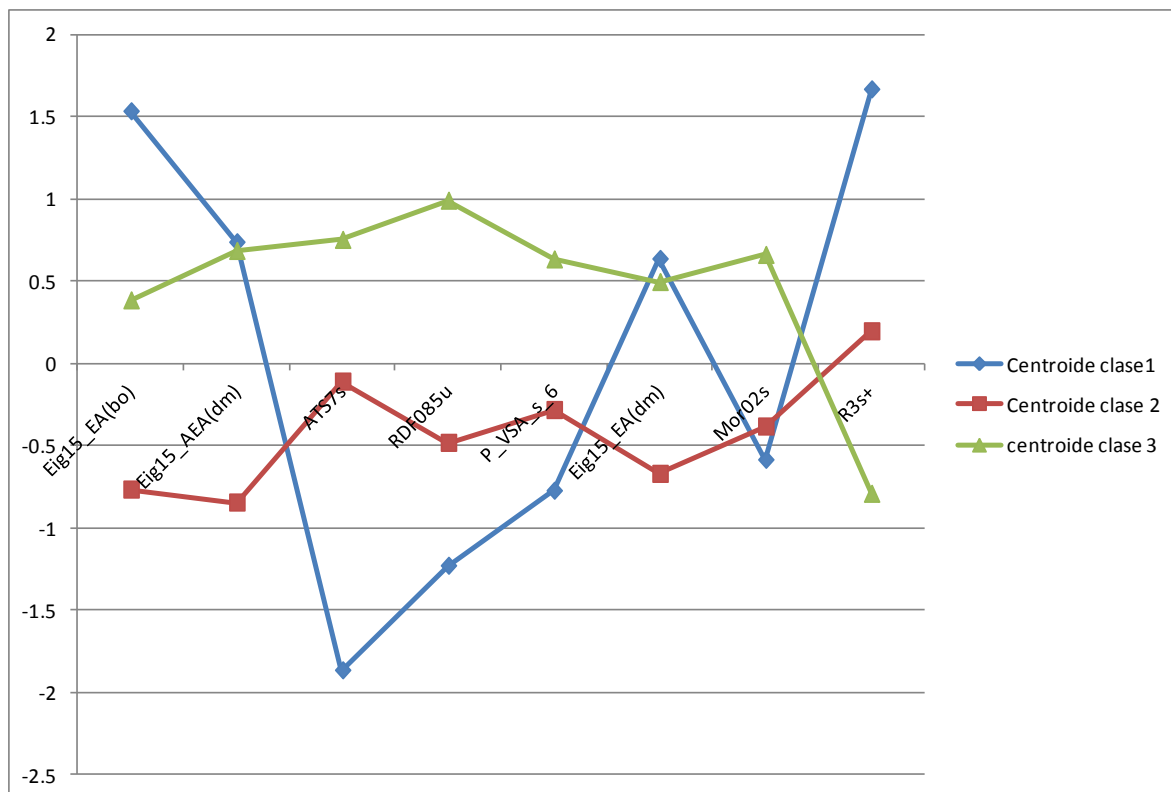


Figura 9: Comportamiento de los centroides de cada clase del modelo

Posterior a esto se evaluó la predicción del modelo para asignar a los péptidos en las clases generadas, para ello se utilizó el mejor modelo calculado, y utilizando el programa MatLab R2088a se realizó la asignación a las clases. Como resultado el modelo calculado pudo asignar a los péptidos del Training Set de la siguiente forma: 29 péptidos en la clase 1, 78 en la clase 2 y 77 en la clase 3, los péptidos asignadas por el método K-means a las clases verdaderas, a la clase 1 corresponden 28 péptidos, a la clase 2 corresponde 79 péptidos y a la clase 3 corresponden 77 péptidos. Con estos datos se construye la matriz de confusión con la cual se calculará la especificidad y la selectividad de las clases, la matriz de confusión se muestra en la tabla 11. La clasificación de los di y tripéptidos del Training Set se muestran en el tabla 12.

NER	0.9728	Clases asignadas		
		1	2	3
Clases Reales	1	28	0	0
	2	1	76	2
	3	0	2	75

Tabla 11: Matriz de confusión del modelo con 8 descriptores y 3 vecinos

Fuente: MatLab R2008a

Clase 1	Clase 2		Clase 3	
AA	AF	IY	AQL	LQQ
AG	AGP	KF	DLP	LRP
AGS	AIP	KL	DYG	LTF
AP	AKK	KP	FAL	LVL
DG	AR	KW	FDK	LVR
GA	ASP	LAP	FEP	LVY
GD	AVF	LGI	FFL	LWA
GF	AVL	LSA	FGK	LYP
GG	AVP	LSP	FNE	PLW
GK	AW	LW	FNF	PRY
GP	AY	LY	FNQ	PYP
GQ	DF	NF	GHF	RFH
GT	DFG	NP	GLY	RIY
GV	DGL	NPP	GWV	RPK
HG	DL	PAP	HHL	RPP
KG	DY	PLP	HHT	RVY
LG	EAP	PR	HIR	RW
RG	EP	QK	HLL	TNP
SG	ER	RF	HQG	TTN
TG	EY	RY	IAQ	VAY
TP	FE	SF	IAY	VFK
VG	FG	SY	IEP	VIY
	FGG	TAP	IEY	VKK
	FL	TF	IFL	VKP
	FP	TK	IKY	VQV
	FY	VAA	ILP	VRK
	GGY	VAV	IQP	VRP
	GKV	VL	IRA	VTR
	GPA	VP	ITT	VVF
	GQP	VPA	IVQ	VVY
	GRP	VPP	IVY	VYP
	GSH	VQ	IWH	YGL

Tabla 12: Clasificación de di y tripéptidos del Training Set

Con los valores de la matriz de confusión se procede a calcular la selectividad y especificidad de cada clase, estos valores se presentan en la tabla 13.

	Clases		
	1	2	3
Sensibilidad	1.00000	0.96203	0.97403
Especificidad	0.96552	0.97436	0.97403

Tabla 13: Selectividad y especificidad de las clases del modelo de clasificación con 8 descriptores

La sensibilidad se define como la capacidad de representar los péptidos en cada clase, mientras que la especificidad es la capacidad de aislar los objetos de una clase de las otras clases, es decir el grado de pureza de la clase [Todeschini, 2003]. Por lo tanto la sensibilidad de la clase 1 es 100%, de la clase 2 es 96.20% y de la clase 3 es 97.40% y la especificidad de la clase 1 es 96.55%, de la clase 2 es 97.44% y de la clase 3 es 97.40%.

Para la validación del modelo, se utilizan los datos correspondientes a los descriptores del Test Set, y se sigue el procedimiento de evaluación del modelo, los descriptores calculados para el Test Set se muestran en el ANEXO 3: Descriptores Moleculares Calculados y Seleccionados del Test Set.

Como resultado el modelo asignó 6 péptidos a la clase 1, 38 a la clase 2 y 34 a la clase 3, los péptidos asignadas por el método K-means a las clases verdaderas, a la clase 1 corresponden 6 péptidos, a la clase 2 corresponden 37 péptidos y a la clase 3 corresponden 35 péptidos. El NER calculado para el Test Set es igual a 0.9359. Con estos datos se construye la matriz de confusión con la cual se calculará la especificidad y la selectividad de las clases, la matriz de confusión se muestra en la tabla 14. La clasificación de los di y tripéptidos del Test Set se muestran en el tabla 15.

NER	0.9359	Clases Asignadas		
		1	2	3
Clases Reales	1	5	1	0
	2	1	35	1
	3	0	2	33

Tabla 14: Matriz de confusión de validación del modelo

Clase 1	Clase 2		Clase 3	
EG	GH	LF	IG	LIY
GE	GS	LPP	AFL	LQP
GI	IA	NK	AQK	LRY
GL	KA	NY	FWN	LVQ
GR	PG	PPK	FYN	NKL
VAP	QG	RL	GTW	PSY
	AAP	RP	IEW	RPR
	ALP	SHP	IKP	RR
	AVV	VF	IKW	TKY
	FAP	VGP	IRP	TVY
	FQ	VK	ITF	VLY
	GHG	VLP	KFY	VSW
	GKP	VY	KYY	VWY
	GPL	WA	LAY	YEY
	GPP	YK	LEE	YVA
	GPV	YV	LEK	YVP
	IAP	YW	LEP	
	IP	FQP		
	KY	KLP		
	LAA			

Tabla 15: Clasificación de di y tripéptidos del Test Set

Con los valores de la matriz de confusión se procede a calcular la selectividad y especificidad de cada clase, estos valores se presentan en la tabla 16.

	Clases		
	1	2	3
Selectividad	0.83333	0.94595	0.94286
Especificidad	0.83333	0.92105	0.97059

Tabla 16: Selectividad y especificidad de la validación del modelo

La sensibilidad de la clase 1 es 83.33%, de la clase 2 es 94.59% y de la clase 3 es 94.29% y la especificidad de la clase 1 es 83.33%, de la clase 2 es 92.10% y de la clase 3 es 97.06%.

CONCLUSIONES

El estudio QSAR aplicado de la actividad inhibidora de la ECA de di y tripéptidos bioactivos, ha permitido desarrollar un modelo de clasificación con el cual se puede predecir el nivel de inhibición de la ECA por parte nuevos di y tripéptidos bioactivos que no se encuentren en la base de datos utilizada para la elaboración del modelo.

Los valores obtenidos de NER en el modelo desarrollado 0.9728 y en la validación del modelo 0.9359, nos muestran la calidad del modelo, indicándonos que este modelo es confiable para ser aplicado en la clasificación de di y tripéptidos bioactivos.

La Clase 1, Baja capacidad de inhibición de la ECA, se caracteriza por la presencia de di y tripéptidos que contengan el aminoácido glicina, la glicina representa el 44.44% de los aminoácidos presentes en esta clase, sin que necesariamente los di y tripéptidos de esta clase contengan glicina. También se caracteriza por la ausencia de los aminoácidos: Isoleucina, Triptófano, Tirosina, Asparragina.

La Clase 2, Media capacidad de inhibición de la ECA, se caracteriza por la presencia de di y tripéptidos que contengan el aminoácido Prolina, la Prolina representa el 14.71% de los aminoácidos presentes en esta clase, existe también la presencia representativa de otros aminoácidos que son Alanina en 10.79%, Glicina en 10.29% y Tirosina en 9.31%, analizando los datos se puede observar que el 29% de los di y tripéptidos asignados a esta clase presentan una interacción entre los cuatro aminoácidos presentes en mayor porcentaje en esta clase. En esta clase, están presentes todos los aminoácidos que componen los di y tripéptidos estudiados.

La Clase 3, Alta capacidad de inhibición de la ECA, se caracteriza por la presencia de di y tripéptidos que contengan el aminoácido Leucina, la Leucina representa el 13.91% de los aminoácidos presentes en esta clase, existe también la presencia representativa de otros aminoácidos que son Prolina 10.43% y Tirosina 10.87% analizando los datos se puede observar que el 32.5% de los di y tripéptidos asignados a esta clase presentan una interacción entre los tres aminoácidos presentes en mayor porcentaje en esta clase. En esta clase, están presentes todos los aminoácidos que componen los di y tripéptidos estudiados.

Los descriptores moleculares que definen las clases son: Eig15_EA(bo), R3s+, ATS7s, RDF085u y P_VSA_s_6.

En la clase 1 se puede observar que los valores de los descriptores Eig15_EA(bo) y R3s+ son altos, mientras que los valores de los descriptores ATS7s, RDF085u y P_VSA_s_6, son bajos.

La clase 2 se mantiene más lineal en su comportamiento, los valores de los descriptores Eig15_EA(bo) es bajo, los valores de descriptores ATS7s, RDF085u P_VSA_s_6 y R3+ son valores medios.

La clase 3 el valor del descriptor Eig15_EA(bo) es medio, los valores de los descriptores ATS7s, RDF085u y P_VSA_s_6 son altos y el valor del descriptor R3+ es bajo.

RECOMENDACIONES

Para el desarrollo modelos matemáticos utilizando la teoría QSAR, es necesario contar con una base de datos amplia, por cual se recomienda en un estudio posterior incrementar la base de datos para el trabajo, tanto en número de di y tripéptidos como también en péptidos de mayor tamaño, para la construcción de nuevos modelos.

La información obtenida de la actividad inhibidora de la enzima ECA, fue recopilada de varias fuentes bibliográfica, por lo cual se recomienda que se trabaje en la fase experimental in vitro, de la cuantificación de la actividad inhibidor a de la ECA de péptidos bioactivos. De la misma forma trabajar en la generación de péptidos bioactivos de sustratos alimenticios y la identificación de los mismos.

BIBLIOGRAFÍA

- AHLQUIST. (1948). A study of Adrenotropic Receptors. *Amer J Physiol.* , 586-600.
- BENÍTEZ, R., IBARZ, A., & PAGAN, J. (2008). Hidrolizados de proteínas: Procesos y aplicaciones. *Acta Bioquímica Clínica Latinoamericana*, 227-236.
- Central Food Technological Research Institute. (s.f.). *Peptide Database of Central Food Technological Reserch Institute*. Recuperado el 16 de Julio de 2013, de <http://www.cftri.com/pepdb/>
- CHIBUIKE, C., & ROTIMI, E. (2012). Food Protein-Derived Bioactive Peptides: Production, Processing, and Potential Health Benefits. *Journal of Food Science*, R11-R24.
- CUSHMAN, D. W., & CHEUNG, H. S. (1971). Spectrophotometric assay and properties of the angiotensin-converting enzyme of rabbit lung. *Biochem. Pharmacol*, 1637-1648.
- DUCHOWICZ, P. R. (2005). *Aplicaciones Fisicoquímicas y Biológicas de la Teoría QSPR y QSAR*. Mar del Plata: UNIVERSIDAD NACIONAL DE LA PLATA.
- FRANK, I., & TODESCHINI, R. (2006). *The Data Analysis Handbook*. . Napoli-Italia .
- GEMPERLINE, P. (2007). *Practical guide to Chemometrics*. New York: Taylor & Francis.
- HERNÁNDEZ LEDESMA, B. (2002). *Caracterización y Bioactividad de Péptidos obtenidos a partir de Proteínas Lácteas mediante Hidrólisis Enzimática y Procesos Fermentativos*. Madrid: UNIVERSIDAD COMPLUTENSE DE MADRID.
- HORTON, H., MORAN, L., SCRIMGEOUR, K., & PERRY, M. (2008). *Principios de Bioquímica (Cuarta ed.)*. Mexico. Mexico: Pearson Educación.
- Hypercube Inc. (2008). Hyperchem (Software for molecular modeling) Version 8.0.
- KAWASAKI , T., SEKI, E., OSAJIMA, K., YOSHIDA , M., & ASAD, M. (2000). Antihypertensive effect of Valyl - Tyrosine, a short chain peptide derived from sardine muscle hydrolyzate, on mild hypertensive subjects. *J Hum Hypertens*, 519 –523.

- KORHONEN , H., & PIHLANTO, A. (2006). Bioactive peptides: production and functionality. *Int Dairy J*, 945–960.
- LAHERA, S., DE LA HERAS, N., CEDIEL, E., NAVARRO, J., & CACHOFEIRO, V. (2000). Angiotensina II e hipertensión arterial: consecuencias del antagonismo de sus receptores. *Hipertensión y Riesgo Vascular*.
- LIZARZABURO, G. (4 de Febrero de 2013). *En Ecuador, de cada 100000 personas, 1373 tienen problemas de hipertensión*. Recuperado el 14 de Junio de 2013, de <http://www.andes.info.ec/es/sociedad/ecuador-cada-100000-personas-1373-tienen-problemas-hipertension.html>
- MARTÍNEZ, O. (2006). Proteínas y péptidos en nutrición enteral. *Nutrición Hospitalaria*, 1-14.
- Mathworks Inc. (2004). Genetic Algorithms and Direct Search Toolbox. En T. Matgroup, *User´s Guide for Use With MATLAB*. Massachusetts.
- MATSUI, T., & TANAKA, M. (2010). Antihypertensive Peptides and Their Underlying Mechanisms. En Y. Mine, E. Li-Chan, & B. Jiang, *Bioactive Protein and Peptides as Funtional Foods and Nutraceutical* (págs. 43-54). Iowa-USA: Willey-Blackwell.
- MEISEL , H. (2004). Multifunctional peptides encrypted in milk proteins. *BioFactors*, 55-61.
- MERCADER, A. G. (2009). *Desarrollo y Aplicación de la Teoría QSAR/QSPR*. Mar del Plata: Universidad de la Plata.
- NAKAMURA, Y., MASUDA, O., & TAKANO, T. (1996). Decrease of tissue angiotensin I-converting enzyme activity upon feeding sour milk in spontaneously hypertensive rats. *Bioscience Biotech Biochem*, 488-489.
- NAKAMURA, Y., YAMAMOTO, N., SAKAI, K., TAKANO, T., OKUBO, A., & YAMAZAKI, S. (1995). Purification and characterization of angiotensin-converting enzyme inhibitors from sour milk. *Journal Dairy Science*, 777-783.
- ONDETTI, M. A., WILLIAMS, N. J., SABO, E. F., PLUSCEC, J., WEAVER, E. R., & KOCY, O. (1971). Angiotensin-converting enzyme inhibitors from the venom of Bothrops jararaca. Isolation, elucidation of structure and synthesis. *Biochemistry*, 4033-4039.
- OPS/OMS. (23 de Abril de 2013). *Comer con menos sal previene la hipertensión y reduce en un 25% el riesgo de ataques cardíacos y accidentes*

- Cerebrovasculares*. Recuperado el 14 de Junio de 2013, de http://new.paho.org/hq/index.php?option=com_content&view=artic
- PÉREZ-DÍAZ, I., HIRIART, M., OLIVARES-REYES, J. A., & ROBLES-DÍAZ, G. (2006). Receptores para la angiotensina II diferentes a los clásicos receptores membranales AT1 y AT2: características y su papel en el funcionamiento celular. *Revista de Educación Bioquímica*, 55-60.
- QUIRÓS, A., RAMOS, M., MURGUERZA, B., DELGADO, M., MIGUEL, M., ALEIXANDRE, A., y otros. (2007). Identification of novel antihypertensive peptides in milk fermented with *Enterococcus faecalis*. *International Dairy*, 33-41.
- SANCHO CAPARRINI, F. (2009). Aplicación K-means. *Magazine Theme for PivotX*.
- SELLEN, J. (2007). Hipertensión arterial: diagnóstico, tratamiento y contro. *Hospital General Calixto García*.
- SIPOLA, M., FINCKENBERG, P., VAPAATALO, H., & NURMINEN, M. (2002). Effect of long-term intake of milk products on blood pressure in hypertensive rats. *Journal Dairy Res*, 103-111.
- Taleta sr. (2010). DRAGON (Software for Molecular Descriptor Calculation) .
- TODESCHINI, R. (2003). *Introduzione alla Chemiometria*. Napoli: EdiSES.
- TODESCHINI, R., & CONSONNI, V. (2009). *Handbook of Molecular Descriptors*. Weinheim: Wiley-VCH.
- WANG , W., & GONZALEZ DE MEJIA, E. (2005). A new frontier in soy bioactive peptides that may prevent age-related chronic diseases. *Compr Rev Food Sci Food Safety*, 63-78.
- YAMAMOTO, N. (2010). Functional Food Products with Antihypertensive Effects. En M. Yoshinori , E. Li-Chan, & B. Jiang, *Bioactive Proteins and Peptides as Functional Foods and Nutraceuticals* (págs. 169-178). Iowa: Wiley-Blackwell.
- YANG, Y., TAO, G., LIU, P., & LIU, J. (2007). Peptide with angiotensin I-converting enzyme inhibitory activity from hydrolyzed corn gluten meal. *Journal of Agriculture and Food Chemistry*, 7891-7895.
- YOSHIMORI, M., EUNICE, L., & JIANG. (2010). *Bioactive Protein and Peptides as Funtional Foods and Nutraceutical*. Iowa-USA: Willey-Blackwell.

ANEXOS

ANEXO 1. BASE DE DATOS DE DIPÉPTIDOS Y TRIPÉPTIDOS CON ACTIVIDAD INHIBIDORA DE LA ENZIMA CONVERTIDORA DE LA ANGIOTENSINA (ECA)

Péptido	IC50 μM	(log(1/I50))	SET	Clase Kmeans
AA	51.4	0.1708	Training	1
AAP	30	0.3983	Test	2
AF	1.01	0.0063	Training	2
AFL	64.15	0.0771	Test	3
AG	0.03	-1.4727	Training	1
AGP	560	-0.8384	Training	2
AGS	527.9	-0.8135	Training	1
AIP	350	-0.6398	Training	2
AKK	3.13	1.3534	Training	2
ALP	240	-0.4804	Test	2
AP	129.04	-0.2182	Training	1
AQK	1800	-1.3318	Test	3
AQL	1.33	0.1231	Training	3
AR	95.5	-0.0910	Training	2
ASP	242.3	-0.4844	Training	2
AVF	2123	-1.4016	Training	2
AVL	10.20	1.0085	Training	2
AVP	340	-0.6276	Training	2
AVV	66.6	0.0613	Test	2
AW	15.32	0.6823	Training	2
AY	19	0.5913	Training	2
DF	361.54	-0.6535	Training	2
DFG	44.7	0.2298	Training	2
DG	12.3	0.7751	Training	1
DGL	32.69	1.5144	Training	2
DL	0.04	-1.3754	Training	2
DLP	4.8	1.1727	Training	3
DY	100	-0.1105	Training	2
DYG	2700	-1.5032	Training	3
EAP	0.20	-0.7040	Training	2
EG	0.01	-1.9300	Test	1
EP	1200	-1.1605	Training	2
ER	667	-0.9123	Training	2
EY	2.68	1.4190	Training	2
FAL	2.84	0.4539	Training	3
FAP	18.67	1.2712	Test	2
FDK	0.21	-0.6845	Training	3

Péptido	IC50 μM	(log(1/IC50))	SET	Clase Kmeans
FE	1.45	1.6785	Training	2
FEP	6.09	0.7847	Training	3
FFL	2.03	0.3079	Training	3
FG	0.02	-1.6381	Training	2
FGG	82.5	-0.0292	Training	2
FGK	160	-0.3091	Training	3
FL	4.66	0.6679	Training	2
FNE	0.24	-0.6261	Training	3
FNF	10.43	1.0182	Training	3
FNQ	335	-0.6213	Training	3
FP	315	-0.5953	Training	2
FQ	1.48	0.1717	Test	2
FQP	12	0.7855	Test	3
FWN	18.3	0.6072	Test	3
FY	5.10	1.1471	Training	2
FYN	4.07	0.6096	Test	3
GA	0.04	-1.3754	Training	1
GD	0.01	-2.0175	Training	1
GE	0.01	-1.9105	Test	1
GF	454.43	-0.7502	Training	1
GG	8670	-1.9961	Training	1
GGY	1.3	1.7247	Training	2
GH	0.03	-1.5602	Test	1
GHF	1100	-1.1237	Training	3
GHG	122	-0.1945	Test	2
GI	0.06	-1.1905	Test	1
GK	0.02	-1.7937	Training	1
GKP	352	-0.6422	Test	2
GKV	18.26	1.2615	Training	2
GL	0.03	-1.4727	Test	1
GLY	8.15	0.9112	Training	3
GP	360	-0.6517	Training	1
GPA	405	-0.7015	Training	2
GPL	27.33	1.4366	Test	2
GPP	4.16	0.6193	Test	2
GPV	15.26	1.1836	Test	2
GQ	5496.87	-1.8036	Training	1
GQP	21.84	1.3393	Training	2
GR	1699.07	-1.3074	Test	1

Péptido	IC50 μM	(log(1/IC50))	SET	Clase Kmeans
GRP	20	0.5696	Training	2
GS	0.02	-1.6478	Test	1
GSH	2.32	0.3663	Training	2
GT	0.02	-1.8229	Training	1
GTW	464.4	-0.7593	Test	3
GV	0.02	-1.7256	Training	1
GVW	240	-0.4804	Training	3
GVY	400	-0.6963	Training	2
GW	2.49	0.3955	Training	2
GY	0.31	-0.5094	Training	2
HG	0.01	-1.8619	Training	1
HHL	1.42	0.1523	Training	3
HHT	800	-0.9892	Training	3
HIR	0.09	-1.0640	Training	3
HLL	1.33	0.1231	Training	3
HQG	745	-0.9591	Training	3
HY	26.1	0.4571	Training	2
IA	0.52	-0.2856	Test	1
IAP	2.97	0.4733	Test	2
IAQ	2.17	0.3371	Training	3
IAY	5.82	0.7652	Training	3
IEP	9.79	0.8715	Training	3
IEW	104	-0.1270	Test	3
IEY	182	-0.3635	Training	3
IF	468.73	-0.7633	Training	2
IFL	44.8	0.2288	Training	3
IG	0.07	-1.1613	Test	1
IKP	4.71	1.1807	Test	3
IKW	9.77	0.8724	Test	3
IKY	17.1	0.6358	Training	3
IL	1.39	0.1425	Training	2
ILP	151.18	-0.2851	Training	3
IP	0.61	-0.2175	Test	2
IPA	141	-0.2556	Training	2
IPP	5	1.1554	Training	2
IQP	3.8	1.2714	Training	3
IR	830	-1.0047	Training	2
IRA	9.64	0.8780	Training	3
IRP	1.8	1.5872	Test	3

Péptido	IC50 μM	(log(1/I50))	SET	Clase Kmeans
ITF	1.55	0.1912	Test	3
ITT	0.12	-0.9180	Training	3
IVQ	0.81	-0.0910	Training	3
IVY	29.23	1.4658	Training	3
IW	15.26	1.1836	Training	2
IWH	20.42	1.3101	Training	3
IY	3.4	1.3184	Training	2
IYK	177	-0.3517	Training	3
IYP	61	0.0984	Training	3
KA	31.5	0.3777	Test	1
KF	28.3	0.4230	Training	2
KFY	45	0.2270	Test	3
KG	0.03	-1.5797	Training	1
KL	50.2	0.1808	Training	2
KLP	500	-0.7906	Test	3
KP	26.10	0.4571	Training	2
KW	10.8	0.8300	Training	2
KY	9.32	0.9696	Test	2
KYY	79	-0.0108	Test	3
LAA	13	0.7517	Test	2
LAP	20.42	1.3101	Training	2
LAY	3.9	1.2604	Test	3
LEE	0.78	-0.1105	Test	3
LEK	800	-0.9892	Test	3
LEP	22.28	0.5240	Test	3
LF	0.23	-0.6359	Test	2
LG	0.01	-1.9981	Training	1
LGI	2.60	0.4150	Training	2
LIY	83.77	1.9231	Test	3
LKL	0.42	-0.3732	Training	3
LKP	43.75	1.6409	Training	3
LKY	5.44	1.1198	Training	3
LLF	79.8	-0.0151	Training	3
LLP	57	0.1271	Training	3
LNP	50.27	0.1802	Training	3
LNY	81	-0.0214	Training	3
LPF	1.90	0.2787	Training	3
LPP	9.6	0.8798	Test	2
LQP	1.9	1.5643	Test	3

Péptido	IC50 μM	(log(1/I50))	SET	Clase Kmeans
LQQ	100	-0.1105	Training	3
LRP	0.27	2.3888	Training	3
LRY	14.27	1.1544	Test	3
LSA	7.81	0.9670	Training	2
LSP	1.7	1.6113	Training	2
LTF	25.55	1.4074	Training	3
LVL	5.96	0.7750	Training	3
LVQ	5.21	0.7166	Test	3
LVR	5.21	0.7166	Training	3
LVY	1.8	1.5872	Training	3
LW	20.42	0.5609	Training	2
LWA	5.82	0.7652	Training	3
LY	2.32	0.3663	Training	2
LYP	6.6	1.0381	Training	3
NF	46.3	0.2149	Training	2
NK	810	-0.9944	Test	2
NKL	88	-0.0564	Test	3
NP	2300	-1.4354	Training	2
NPP	0.28	-0.5580	Training	2
NY	32.6	0.3632	Test	2
PAP	87	-0.0516	Training	2
PG	0.01	-2.2802	Test	1
PLP	430	-0.7268	Training	2
PLW	2.08	0.3177	Training	3
PPK	0.08	-1.0835	Test	2
PR	4.1	1.2393	Training	2
PRY	2.5	1.4483	Training	3
PSY	4.66	0.6679	Test	3
PYP	220	-0.4436	Training	3
QG	0.01	-2.0564	Test	1
QK	885	-1.0318	Training	2
RF	93	-0.0798	Training	2
RFH	330	-0.6150	Training	3
RG	0.07	-1.1613	Training	1
RIY	28	0.4275	Training	3
RL	0.03	-1.4629	Test	2
RP	21	0.5490	Test	2
RPK	0.05	-1.3462	Training	3
RPP	1.27	0.1036	Training	3

Péptido	IC50 μM	(log(1/I50))	SET	Clase Kmeans
RPR	382	-0.6768	Test	3
RR	267.1	-0.5256	Test	3
RVY	205.6	-0.4150	Training	3
RW	18.86	0.5944	Training	3
RY	30.75	0.3879	Training	2
SF	130.2	-0.2220	Training	2
SG	0.01	-1.9883	Training	1
SHP	280	-0.5455	Test	2
SY	66.3	0.0632	Training	2
TAP	3.5	1.3062	Training	2
TF	18	0.6142	Training	2
TG	0.01	-2.0564	Training	1
TK	0.05	-1.2878	Training	2
TKY	30.57	1.4853	Test	3
TNP	0.38	-0.4218	Training	3
TP	290	-0.5604	Training	1
TTN	0.12	-0.9180	Training	3
TVY	15	0.6912	Test	3
VAA	13	0.7517	Training	2
VAP	2	1.5426	Test	2
VAV	260	-0.5142	Training	2
VAY	16	0.6639	Training	3
VF	43.7	0.2394	Test	2
VFK	0.08	-1.0932	Training	3
VG	0.08	-1.1224	Training	1
VGP	2.84	0.4539	Test	2
VIY	6.04	1.0756	Training	3
VK	5.70	0.7555	Test	2
VKK	1045	-1.1021	Training	3
VKP	3.7	1.2827	Training	3
VL	3.9	1.2604	Training	2
VLP	0.95	-0.0229	Test	2
VLV	2.43	0.3858	Test	3
VP	570	-0.8459	Training	2
VPA	4.4	1.2095	Training	2
VPP	9	0.9071	Training	2
VQ	1300	-1.1943	Training	2
VQV	8.34	0.9209	Training	3
VRK	700	-0.9327	Training	3

Péptido	IC50 μM	(log(1/I50))	SET	Clase Kmeans
VRP	2.2	1.5024	Training	3
VSP	10	0.8625	Training	2
VSW	3.18	0.5025	Test	3
VTR	0.58	-0.2369	Training	3
VVF	2.13	0.3274	Training	3
VVY	22	0.5294	Training	3
VW	2.81	1.3989	Training	2
VWY	9.4	0.8887	Test	3
VY	21.44	0.5403	Test	2
VYP	288	-0.5574	Training	3
WA	277.3	-0.5414	Test	2
WG	0.01	-1.8327	Training	2
WL	2.49	0.3955	Training	2
YE	0.13	-0.8889	Training	2
YEY	17.85	1.2517	Test	3
YG	1306.78	-1.1965	Training	2
YGG	0.01	-2.0564	Training	2
YGL	0.20	-0.7040	Training	3
YH	5.1	1.1471	Training	2
YK	610	-0.8746	Test	2
YKY	43.5	0.2413	Training	3
YL	82	-0.0266	Training	2
YN	51	0.1741	Training	2
YNK	125	-0.2048	Training	3
YP	890	-1.0342	Training	2
YPR	16.5	0.6509	Training	3
YV	0.14	-0.8499	Test	2
YVA	48.93	1.6896	Test	3
YVP	200	-0.4034	Test	3
YW	10.5	0.8419	Test	2

ANEXO 2: DESCRIPTORES MOLECULARES CALCULADOS Y SELECCIONADOS DEL TRAINING SET

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
AA	1.533	0.736	-2.754	-1.386	-1.427	0.635	-0.982	2.938
AF	1.533	0.736	-0.946	-1.261	0.273	0.635	-0.583	0.845
AG	1.533	0.736	-1.520	-1.386	-0.219	0.635	-1.101	2.512
AGP	1.533	0.736	-1.721	-1.386	-0.219	0.635	-0.520	2.244
AGS	1.533	0.736	-4.649	-1.386	-1.427	0.635	-1.068	3.114
AIP	1.533	0.736	-1.338	-1.377	-0.219	0.635	-0.389	2.502
AKK	1.533	0.736	-0.828	-1.203	-0.219	0.635	0.010	2.089
AP	1.533	0.736	-5.108	-1.386	-1.427	0.635	-0.929	3.799
AQL	1.533	0.736	-1.671	-1.254	-1.427	0.635	-1.012	1.082
AR	1.533	0.736	-1.070	-0.866	-0.582	0.635	-0.115	1.508
ASP	1.533	0.736	-1.509	-0.945	-1.427	0.635	-0.866	0.999
AVF	1.533	0.736	-4.524	-1.386	-1.427	0.635	-1.392	2.152
AVL	1.533	0.736	-0.886	-1.081	0.021	0.635	0.082	2.048
AVP	1.533	0.736	-0.781	-0.821	0.262	0.635	0.818	0.784
AW	1.533	0.736	-2.620	-1.386	-0.331	0.635	-1.380	2.124
AY	1.533	0.736	-2.332	-1.386	-1.427	0.635	-1.118	1.312
DF	1.533	0.736	-1.603	-1.384	-1.427	0.635	-0.919	1.052
DFG	1.533	0.736	-1.449	-1.131	-1.427	0.635	-1.124	1.085
DG	1.533	0.736	-1.819	-1.215	-1.427	0.635	-0.902	1.060
DGL	1.533	0.736	-0.897	-1.178	-0.582	0.635	0.028	1.316
DL	1.533	0.736	-1.065	-1.129	-0.582	0.635	0.139	1.340
DLP	1.533	0.736	-2.008	-1.176	-1.427	0.635	-0.949	1.119
DY	1.533	0.736	-2.358	-1.386	-1.809	0.635	-0.736	0.150
DYG	1.533	0.736	-1.507	-1.314	0.021	0.635	-0.186	1.610
EAP	1.533	0.736	-0.774	-1.003	0.262	0.635	0.992	0.954

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
EP	1.533	0.736	-1.638	-1.311	-0.331	0.635	-0.404	2.729
ER	1.533	0.736	-0.935	-0.972	-0.331	0.635	-0.975	0.952
EY	1.533	0.736	-1.985	-1.386	-1.427	0.635	-0.862	1.246
FAL	-1.109	-1.361	-0.438	-1.079	-0.823	-2.600	-1.723	0.017
FDK	-1.778	-1.413	-0.795	-1.126	-0.823	-3.122	-1.463	0.060
FE	-0.570	-0.750	0.437	-0.632	-0.823	-0.339	-1.319	-0.196
FEP	-1.123	-1.372	-0.033	-0.086	0.262	-0.676	0.874	0.571
FFL	-0.570	-0.809	0.371	0.125	-0.823	-0.339	-0.942	-0.321
FG	-1.825	-2.029	-0.504	-0.819	0.262	-3.124	0.951	0.459
FGG	-0.790	-1.276	0.093	-1.108	0.273	-0.686	-1.667	-0.010
FGK	-0.120	-0.440	0.501	0.131	-0.823	0.162	-0.904	-0.055
FL	-0.569	-0.809	0.337	0.280	-0.823	-0.666	-0.535	-0.247
FNE	-0.749	-0.809	0.266	-1.199	-0.823	-0.668	-1.755	-0.141
FNF	-0.863	-0.809	0.187	-0.928	-0.823	-0.685	-1.615	-0.204
FNQ	-0.599	-0.707	-0.347	-0.406	-1.427	0.150	-0.210	0.577
FP	-1.653	-1.308	-0.468	-0.693	-0.331	-0.850	-0.230	1.042
FY	-1.825	-1.940	-0.425	-0.909	-0.219	-3.124	-1.015	0.802
GA	-0.522	-0.053	-0.098	-0.819	0.385	-0.339	-1.590	-0.347
GD	-1.339	-1.749	-1.191	-1.382	-0.219	-3.070	-0.790	0.722
GF	-0.621	-0.098	-0.124	-0.230	0.877	0.113	0.110	0.866
GG	-0.227	-0.050	0.082	-0.545	-0.823	0.331	-1.908	-0.666
GGY	-0.621	-0.706	0.001	-0.744	-0.219	0.082	0.445	0.708
GHF	-0.621	-0.809	-0.476	-0.784	-0.823	-0.677	-0.560	-0.472
GK	-0.621	-0.710	-0.029	-0.290	0.021	0.082	0.480	0.602
GKV	-0.621	-0.678	-0.200	-0.620	0.273	-0.219	-0.436	0.880
GLY	-0.519	-0.625	0.296	-0.713	0.021	-0.290	-0.521	0.182

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
GP	-0.532	-0.809	0.253	-0.119	0.021	-0.290	-0.743	0.250
GPA	-1.811	-1.503	-0.739	-1.118	-0.823	-3.124	-1.432	0.808
GQ	-0.754	-0.220	0.446	-0.979	0.625	-0.339	-0.032	0.289
GQP	-0.190	-0.165	0.391	-0.242	0.866	0.017	0.052	0.105
GRP	-0.280	-0.168	0.422	-0.075	0.273	0.635	-0.931	0.459
GSH	-1.833	-1.797	-0.745	-1.100	-0.331	-2.600	-0.322	1.154
GT	-0.369	-0.095	-0.071	-0.267	-0.331	0.351	-0.222	0.590
GV	-0.621	-1.098	-0.298	-0.637	-1.427	0.082	-0.436	0.467
GVW	-1.825	-1.956	-0.573	-0.254	-1.427	-3.124	-0.667	0.352
GVY	-1.287	-1.984	-1.098	-1.117	-1.427	-2.602	-1.203	0.187
GW	-0.545	-0.951	0.059	-0.414	-0.823	-0.364	-0.989	-0.760
GY	-0.285	-0.030	0.265	-0.357	-0.823	0.315	-1.395	-1.143
HG	-0.713	-1.226	-0.144	0.399	0.262	-0.279	0.988	-0.012
HHL	-0.299	-0.542	-0.020	-0.137	-1.427	0.635	0.043	0.154
HHT	-0.621	-0.809	-0.118	-0.710	-0.331	0.150	-0.331	0.455
HIR	-0.586	-0.809	-0.015	0.161	-0.582	0.150	0.613	0.606
HLL	-1.136	-1.925	-0.232	-0.343	-0.582	-2.600	0.648	0.538
HQG	-1.136	-1.933	-0.765	-0.939	-0.582	-2.600	-0.133	0.389
HY	-0.226	-0.070	0.217	0.154	-0.582	0.635	1.167	0.260
IAQ	-0.513	-0.809	-0.182	-0.713	-0.823	-0.339	-1.687	-0.636
IAY	-0.621	-0.809	-0.396	0.133	-1.427	0.082	-0.399	0.522
IEP	-0.474	-0.809	-0.329	-0.087	-0.823	-0.273	-1.061	-0.891
IEY	-0.779	-0.815	0.086	-0.600	0.273	-0.685	-0.617	-0.300
IF	-0.435	-0.556	0.270	-0.611	0.273	0.073	-1.444	-0.642
IFL	-0.273	-0.543	-0.101	0.197	-1.427	0.635	0.061	0.238
IKY	-1.932	-1.964	-0.872	-1.270	0.021	-3.159	-0.926	0.776

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
IL	-0.391	0.322	0.338	-1.088	0.625	0.091	-1.810	-0.454
ILP	-0.918	-0.775	-0.065	-0.346	1.117	0.082	-0.062	0.986
IPA	-0.148	-0.098	0.409	0.029	-1.205	0.331	-0.740	-1.020
IPP	-0.111	0.412	0.354	0.572	-0.361	0.379	0.552	-0.636
IQP	-0.790	-1.288	-0.352	-0.547	-0.120	-0.364	0.965	-0.779
IR	-1.125	-1.316	-0.239	-0.094	0.866	-2.600	0.802	0.626
IRA	-0.375	-0.191	0.132	0.171	0.262	0.150	1.259	0.450
ITT	-0.206	-0.107	0.268	0.172	1.357	0.635	1.440	0.432
IVQ	-1.274	-1.345	-0.293	-0.790	-0.331	-0.850	-0.319	1.093
IVY	-0.166	0.556	0.692	-0.361	0.273	0.135	-0.185	-0.237
IW	-0.919	-1.095	-0.113	-0.857	0.764	-0.676	-0.175	1.121
IWH	-0.757	-1.087	0.080	-0.591	0.273	-0.686	-1.199	-0.437
IY	-0.869	-1.107	-0.141	-0.436	-0.331	-0.693	-0.020	0.843
IYK	-1.825	-2.069	-0.167	-0.402	0.513	-3.124	0.452	0.958
IYP	-1.790	-1.432	-0.311	-0.626	-0.823	-3.122	-0.926	-0.178
KF	-0.869	-0.992	-0.081	-0.586	-0.823	-0.685	-1.156	-0.529
KG	-0.869	-1.281	-0.487	-0.537	-0.823	-0.677	-1.226	-0.517
KL	-1.825	-2.069	-0.368	-0.499	-0.582	-3.124	0.068	0.440
KP	-0.207	-0.763	0.540	0.343	-0.823	0.135	-0.606	-0.795
KW	1.533	0.736	-1.232	-1.386	-1.427	0.635	-1.307	0.309
LAP	-0.737	-0.809	0.345	-0.686	0.273	-0.339	-1.463	-0.591
LG	-0.621	-0.138	0.187	-0.661	0.877	0.113	1.323	0.684
LGI	-1.784	-1.772	-0.844	-1.307	-0.331	-1.728	-0.235	0.876
LKL	-0.433	-0.064	0.011	0.145	-0.331	0.351	0.730	0.810
LKP	-0.533	-0.778	0.080	0.119	0.513	0.635	0.687	0.207
LKY	-0.621	-0.809	-0.103	-0.359	-0.331	0.150	-0.074	0.440

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
LLF	-0.918	-0.809	-0.031	-0.500	1.117	0.082	0.693	0.776
LLP	-0.753	-0.885	-0.550	-1.149	-0.331	0.091	-0.640	0.317
LNP	-0.873	-0.831	-0.314	-0.537	-0.331	0.082	-0.130	0.583
LNy	0.425	0.688	0.314	0.184	-0.331	0.635	0.638	0.555
LPF	0.058	-0.189	0.610	1.144	-0.823	0.635	0.760	-0.445
LQQ	-0.212	0.231	0.603	0.504	0.385	0.113	-0.619	-0.529
LRP	-0.346	1.024	0.573	-0.093	1.481	0.136	1.087	-0.327
LSA	1.085	1.557	1.002	1.395	1.229	0.635	1.388	-1.110
LSP	0.400	1.313	0.888	-0.035	0.385	0.635	0.552	-0.844
LTF	1.295	1.370	0.877	1.563	-0.823	0.635	0.685	-0.996
LVL	-0.199	0.556	0.221	0.382	0.021	0.635	0.473	-1.047
LVR	1.318	1.614	0.966	0.971	0.625	0.635	1.030	-0.822
LVY	0.413	1.120	1.002	0.456	2.074	0.635	1.408	-0.969
LW	0.400	0.908	0.859	0.223	0.625	0.635	0.654	-0.920
LWA	2.116	1.895	1.012	2.109	0.625	0.635	0.955	-1.231
LY	1.336	1.673	0.943	1.371	1.721	0.635	0.663	-1.065
LYP	0.081	0.991	0.540	0.521	-0.823	0.635	0.188	0.661
NF	-0.079	0.154	0.548	0.334	0.273	0.635	-0.419	0.346
NP	1.115	1.364	0.749	1.318	-0.823	0.635	1.019	-0.810
NPP	1.104	1.353	0.682	0.874	0.273	0.351	-0.228	-0.797
PAP	0.097	0.500	0.658	2.253	-0.823	0.635	0.409	-0.889
PLP	-0.464	-0.225	0.370	0.144	0.625	0.071	0.336	-0.748
PLW	-0.213	0.386	0.435	0.278	0.273	0.635	-0.160	-0.857
PR	1.680	1.861	1.088	1.426	0.385	0.635	1.936	-0.607
PRY	0.514	1.480	1.042	1.247	1.481	0.635	2.094	-0.977
PYP	-0.143	0.372	0.642	0.803	0.021	0.635	-0.060	-0.861

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
QK	1.682	1.580	0.993	2.191	0.021	0.635	1.013	-0.679
RF	-0.204	-0.143	0.557	0.880	-0.823	0.331	-0.656	-0.906
RFH	-0.243	0.597	0.757	0.323	0.625	0.135	0.678	-0.830
RG	-0.121	-0.135	0.573	1.703	0.866	0.118	1.564	-0.893
RIY	0.097	0.622	0.717	1.406	0.866	0.635	0.467	-0.938
RPK	-0.289	0.007	0.720	0.778	0.625	0.118	0.478	-0.922
RPP	0.097	0.906	0.772	0.436	0.273	0.635	-0.091	-0.961
RVY	0.505	0.967	0.774	1.210	0.273	0.635	-0.188	-0.881
RW	1.583	1.701	1.104	1.643	1.117	0.635	1.480	-0.871
RY	-0.046	0.229	0.727	1.534	0.021	0.635	0.263	-0.771
SF	1.740	1.701	1.163	1.627	2.212	0.635	1.315	-0.826
SG	0.097	0.209	0.370	-0.257	0.273	0.635	-0.202	-0.799
SY	0.097	0.560	0.916	0.865	1.229	0.635	2.293	-1.063
TAP	-0.087	0.216	0.743	0.307	0.385	0.135	0.836	-0.738
TF	0.514	0.820	0.832	1.604	0.273	0.635	0.009	-0.902
TG	0.097	0.200	0.688	2.752	0.021	0.635	0.671	-0.902
TK	0.053	0.212	0.592	0.921	0.021	0.635	-0.104	-0.838
TNP	0.097	0.228	0.714	1.981	-0.823	0.635	0.391	-0.834
TP	-0.048	-0.143	0.505	1.298	-0.823	0.331	-0.622	-0.859
TTN	-0.191	0.212	0.698	-0.219	0.625	0.113	-0.064	-0.803
VAA	0.106	0.461	0.443	0.887	-0.823	0.635	-0.984	-0.777
VAV	0.097	0.783	0.939	1.534	2.074	0.118	1.927	-0.938
VAY	0.097	0.212	0.669	1.890	0.866	0.635	0.468	-0.922
VFK	0.097	0.208	0.736	0.248	0.273	0.635	-0.524	-0.808
VG	0.041	-0.740	0.472	0.483	-0.823	0.162	-0.375	-0.904
VIY	0.097	0.109	0.708	0.938	0.866	0.635	1.137	-1.120

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
VKK	0.379	0.492	0.665	1.017	-0.823	0.635	1.267	-0.934
VKP	0.502	0.644	0.728	1.337	0.273	0.635	-0.269	-0.877
VL	0.051	0.381	0.750	1.465	1.470	0.162	0.880	-0.644
VP	0.089	1.033	0.626	0.548	0.986	0.635	0.494	-1.020
VPA	0.516	1.192	0.653	0.931	-0.109	0.635	-0.107	-0.979
VPP	1.908	1.659	1.058	1.709	0.866	0.635	2.388	-0.914
VQ	1.536	1.466	1.087	1.055	1.961	0.635	1.189	-0.973
VQV	0.492	0.563	0.737	1.340	1.710	0.635	1.367	-1.065
VRK	0.097	0.413	0.525	0.198	0.866	0.371	-0.080	-0.898
VRP	0.578	1.423	0.825	0.856	2.554	0.635	1.657	-1.161
VSP	-0.151	-0.191	0.247	0.501	1.950	0.017	2.288	0.107
VTR	0.097	-0.039	0.342	0.509	0.262	0.635	1.459	0.133
VVF	0.514	1.157	1.045	1.660	2.212	0.635	1.186	-0.548
VVY	-0.367	-0.114	0.854	-0.065	1.721	-0.339	0.072	-0.593
VW	-0.364	-0.186	0.859	-0.218	2.816	-0.685	0.572	-0.503
VYP	0.103	0.698	0.877	2.784	0.021	0.635	1.699	-0.988
WG	0.097	0.347	0.839	1.346	0.866	0.635	1.039	-0.861
WL	-0.143	-0.114	0.626	0.652	0.021	0.331	0.099	-0.746
YE	0.097	0.392	0.836	1.035	0.273	0.635	0.422	-0.648
YG	0.097	0.559	0.901	1.642	1.710	0.635	1.552	-0.889
YGG	0.097	0.021	0.701	0.623	0.866	0.331	0.546	-0.820
YGL	0.097	1.019	0.786	0.570	0.273	0.635	0.224	-0.345
YH	0.097	-0.180	0.871	0.636	1.961	0.118	1.056	-0.963
YKY	0.097	0.382	0.760	0.482	0.273	0.635	-0.488	-0.580
YL	2.237	1.855	1.097	1.484	0.273	0.635	0.892	-0.820
YN	0.097	0.641	0.760	1.092	0.273	0.635	-0.015	-0.793

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
YNK	1.700	2.282	1.214	0.914	2.576	0.635	2.499	-1.106
YP	-0.084	0.250	0.219	0.064	0.273	0.635	-0.037	-1.035
YPR	1.704	1.705	1.127	1.856	2.212	0.635	1.583	-1.118

ANEXO 3: DESCRIPTORES MOLECULARES CALCULADOS Y SELECCIONADOS DEL TEST SET

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
AAP	1.533	0.736	-3.253	-1.386	-1.427	0.635	-0.794	3.547
AFL	1.533	0.736	-2.620	-1.386	-1.427	0.635	-1.251	1.160
ALP	-2.775	-2.355	-1.034	-1.003	-1.427	-3.124	-0.629	1.154
AQK	1.533	0.736	-1.234	-1.224	-1.427	0.635	-0.336	1.622
AVV	1.533	0.736	-3.588	-1.386	-0.331	0.635	-0.773	3.013
EG	1.533	0.736	-2.015	-1.386	-0.331	0.635	-0.757	3.330
FAP	-1.828	-1.777	-0.703	-0.633	-1.427	-2.600	-0.446	1.050
FQ	-0.944	-0.809	-0.177	-0.685	-0.219	-0.676	-0.642	1.058
FQP	-0.398	0.693	0.484	-0.242	0.385	0.113	0.363	-0.310
FWN	-0.944	-0.809	-0.111	-0.705	0.385	-0.677	-0.601	-0.257
FYN	-0.918	-0.769	-0.010	-0.753	0.877	0.082	-0.496	1.050
GE	-0.713	-0.191	-0.150	-0.157	1.470	-0.279	1.248	0.485
GH	-2.775	-2.355	-1.253	-1.381	-1.427	-3.124	-0.998	0.909
GHG	-0.621	-0.809	-0.333	-0.423	-1.427	0.082	-0.573	0.452
GI	-0.757	-1.060	-0.877	-1.181	-1.427	-0.339	-1.499	0.389
GKP	-0.278	-0.094	0.059	-0.128	-0.331	0.635	-0.330	0.794
GL	-1.159	-1.016	-0.084	-1.143	-0.823	-0.933	0.211	0.690
GPL	-0.756	-0.857	-0.242	-0.122	-0.823	-0.364	-0.672	0.195
GPP	-0.705	-1.220	-0.379	-1.162	-0.823	-0.344	-2.079	0.583
GPV	-0.873	-1.180	-0.484	-1.089	-0.823	-0.680	-1.678	0.287
GR	-0.665	-0.492	0.108	-0.683	0.273	-0.290	0.626	0.514
GS	-1.090	-1.087	-0.595	-0.505	-1.427	0.145	-0.316	0.815
GTW	-0.520	-1.086	-0.091	-0.537	-0.823	-0.339	-1.680	-0.660
IA	-0.553	-0.766	0.135	0.039	0.513	0.635	0.685	0.590
IAP	-1.030	-1.273	-0.385	-0.661	-0.823	-2.600	-1.116	-0.296

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
IEW	-0.183	-0.038	0.168	-0.878	-0.823	0.315	-1.879	-0.865
IG	-0.621	-0.809	-0.203	0.073	-0.331	0.150	-0.325	0.502
IKP	-0.944	-0.809	-0.239	-0.307	0.021	-0.676	-0.086	0.990
IKW	-1.339	-1.545	-0.262	-0.741	0.866	-3.070	0.375	0.988
IP	-0.757	-1.090	-0.281	-1.178	-1.205	-0.364	-1.548	-0.834
IRP	-0.705	-1.009	-0.064	-0.230	0.262	-0.279	1.179	0.385
ITF	-0.749	-1.219	-0.513	-0.135	0.262	-0.304	1.125	0.225
KA	-0.757	-1.089	-0.123	-0.708	-0.823	-0.668	-1.577	-0.556
KFY	-0.869	-1.114	-0.342	-0.767	-1.427	-0.676	-0.423	0.522
KLP	1.533	0.736	-0.627	-0.765	-1.427	0.635	-0.750	0.381
KY	-0.750	-0.954	0.010	-0.692	-0.823	-0.674	-1.169	-0.720
KYY	-0.339	-0.755	0.223	-0.606	-0.823	0.091	-1.549	-0.963
LAA	-1.825	-1.860	-0.260	-0.690	0.021	-3.124	0.433	0.696
LAY	-0.357	-0.648	-0.056	-0.230	-1.427	0.635	-0.038	0.217
LEE	-0.868	-0.809	-0.156	-0.917	-0.331	0.082	-0.281	0.504
LEK	-0.601	-0.763	-0.611	-0.990	-1.427	0.150	-0.953	0.675
LEP	-1.099	-1.121	-0.749	-0.921	-1.427	0.095	-0.702	0.677
LF	-0.249	-0.553	-0.046	-0.329	-1.427	0.635	-0.682	0.338
LIY	-0.621	-0.809	-0.224	-0.798	0.273	-0.265	-0.040	-0.564
LPP	-0.286	-0.131	0.790	0.649	1.470	0.113	1.314	-0.388
LQP	-0.354	-0.220	0.668	0.614	0.625	0.071	1.573	-0.249
LRY	-0.201	-0.185	0.279	0.282	-0.823	0.635	-0.766	-0.920
LVQ	0.413	1.577	1.016	0.007	1.833	0.635	1.311	-0.928
NK	0.081	0.687	0.574	0.520	0.273	0.635	-0.908	0.524
NKL	0.081	0.687	0.492	0.404	-0.823	0.635	-0.764	0.301
NY	1.145	1.253	0.893	1.795	0.866	0.635	1.162	-0.965

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
PG	-0.325	0.730	0.538	-0.228	0.625	0.074	1.148	-0.343
PPK	-0.243	0.645	0.787	-0.024	0.385	0.135	0.730	-0.795
PSY	0.097	0.693	0.776	1.422	-0.823	0.635	0.810	-1.078
QG	1.682	1.580	0.993	2.191	0.021	0.635	1.013	-0.679
RL	0.097	0.926	0.802	0.599	0.273	0.635	-0.213	-0.912
RP	-0.298	-0.206	0.695	-0.062	1.368	-0.666	-1.612	-0.818
RPR	2.063	1.857	0.986	2.087	-0.823	0.635	2.035	-1.233
RR	0.983	1.103	0.959	2.083	1.117	0.635	2.084	-1.080
SHP	0.097	1.172	0.975	0.494	1.593	0.118	1.811	-0.754
TKY	0.844	1.180	0.901	1.571	1.117	0.635	0.690	-0.720
TVY	0.514	1.089	0.913	0.826	1.721	0.635	1.118	-0.967
VAP	-0.087	0.212	0.711	0.361	0.625	0.135	0.786	-0.799
VF	0.861	1.472	0.960	1.823	1.961	0.635	1.273	-0.824
VGP	0.010	0.004	0.668	0.303	0.625	0.118	0.243	-0.936
VK	0.097	0.382	0.721	1.246	0.273	0.635	-0.221	-0.898
VLP	1.134	1.582	0.789	1.549	-1.205	0.635	0.562	-1.049
VLY	1.550	1.701	0.903	1.793	1.579	0.635	1.368	-1.084
VSW	0.518	1.464	0.993	0.653	1.961	0.635	0.935	-0.940
VWY	0.097	0.382	0.905	0.998	1.368	0.635	-0.403	-0.527
VY	-0.214	-0.155	0.414	0.285	0.273	0.635	-0.202	-0.636
WA	0.097	0.800	0.867	1.090	0.273	0.635	-0.270	-0.593
YEY	-0.275	-0.220	0.701	0.665	0.625	0.113	0.973	-0.679
YK	0.097	-0.140	0.665	0.389	-0.823	0.635	-0.512	-0.603
YV	0.975	1.495	1.057	1.232	2.565	0.635	2.393	-1.180
YVA	1.667	1.506	0.810	1.022	1.961	0.635	1.706	-1.233
YVP	-0.023	0.010	0.543	-0.210	0.273	0.635	-0.592	-0.789

Valor del descriptor calculado								
Péptidos	Eig15_EA(bo)	Eig15_AEA(dm)	ATS7s	RDF085u	P_VSA_s_6	Eig15_EA(dm)	Mor02s	R3s+
YW	0.097	0.608	0.717	-0.160	0.273	0.635	-1.099	-0.918