



UNIVERSIDAD DEL PAPALOAPAN

Campus Tuxtepec

Licenciatura en Ciencias Químicas

“Evaluación *in silico* -Docking y QSAR- de derivados del ácido valproico sobre la HDAC8 con potencial uso antineoplásico”

TESIS

Presenta:

Heidy Martínez Pacheco

Para obtener el título de:

LICENCIADA EN CIENCIAS QUÍMICAS

Dr. Guillermo Ramírez Galicia
Director de Tesis

Dr. José Correa Basurto
Codirector de Tesis

San Juan Bautista Tuxtepec, Oaxaca, Noviembre de 2012.

terra uberrima, mens aperta
Bo Lo-tama, chí jí jú

El presente trabajo se realizó en el Laboratorio de Simulación de la División de Estudios de Posgrado e Investigación de la Universidad del Papaloapan Campus Tuxtepec y en el laboratorio de Modelado Molecular y Bioinformática de la Sección de Estudios de Posgrado e Investigación de la Escuela Superior de Medicina del Instituto Politécnico Nacional, bajo la dirección del Dr. Guillermo Ramírez Galicia y el Dr. José Correa Basurto.

Este trabajo forma parte del proyecto 132353 del Consejo Nacional de Ciencia y Tecnología (CONACyT).

Productos del trabajo presentado en esta tesis:

- ✓ Presentación oral en el 13^o Foro Estatal de investigación e innovación. Oaxaca de Juárez, Oaxaca, 1 y 2 de diciembre de 2011.
- ✓ Presentación oral en el Seminario Estudiantil de la Universidad del Papaloapan. San Juan Bautista Tuxtepec, Oaxaca, 3 de febrero de 2012.
- ✓ Presentación oral en el 2^o Taller de Dinámica Molecular: Algoritmos, Análisis y Aplicaciones en Programas Paralelos. Guanajuato, Guanajuato, del 20 al 24 de agosto de 2012.
- ✓ 1 publicación en las Memorias del 13^o Foro estatal de Investigación e Innovación Oaxaca. pp. 130-132, 2011. ISBN: 978-607-7849-19-3.

DEDICATORIA

A DIOS y a la VIRGEN de GUADALUPE. Porque aún sabiendo que son entes espirituales me han dado la oportunidad de respirar cada día de mi vida y la fuerza necesaria para vencer todos los obstáculos que se me han presentado en el camino.

A mis padres “las personas más importantes de mi vida”: Sr. Marcos Martínez Lavida y Sra. Lucia Pacheco Quintero, han sido y siguen siendo un ejemplo y un orgullo para mí, les agradezco la vida, la educación, los principios, la confianza y los valores que me han inculcado durante todos estos años; y sin temor a equivocarme puedo decir lo siguiente: *sin ustedes no soy nada*.

A mis paninos: Dra. Delia E. Páramo Calderón y Dr. Alejandro Aparicio Saguilán. Por todas las muestras de cariño, los consejos y la confianza que han depositado en mí. *Han sido mi mayor apoyo*.

A mi sensei: Dr. Guillermo Ramírez Galicia. Por guiarme en un mundo nuevo y desconocido que jamás pensé conocer “el mundo de la química computacional”, donde *una duda no debe de hacerse más pequeña sino más grande*.

A mí querida burris mayor: Dra. María C. Salas Solís. Por los momentos de complicidad, por su apoyo incondicional y frente a cualquier problema, por luchar tanto por mí y *por quererme así como yo la quiero a usted*.

A mi hermanita adoptiva: Milagros del C. Antele López. Es todo un privilegio para mí que seas mi amiga y mi hermanita adoptiva, *eres la energía, la luz y la paz*, gracias por compartir conmigo los mejores momentos de tu vida.

A mis amigos, a los que tanta lata di mi mal genio: Jair, Conchita, Anita, Fernando, Perla, Rubén, Brenda, Mariacel, la química Lety y la química Luz.

Gracias por formar parte de esta aventura CON CARIÑO
Heidy Martínez Pacheco

AGRADECIMIENTOS

La elaboración de esta tesis no habría sido posible sin la colaboración y el apoyo de muchas personas, a las que quiero expresar mis agradecimientos.

Sin duda, gran parte del mérito que este trabajo pueda tener corresponde en especial a mi director de tesis, el Dr. Guillermo Ramírez Galicia y a mi codirector, el Dr. José Correa Basurto, gracias por guiarme en el mundo de la investigación, por sus consejos, sugerencias, discusiones y críticas.

Al laboratorio de Simulación de la División de Estudios de Posgrado e Investigación de la Universidad del Papaloapan Campus Tuxtepec.

Al laboratorio de Modelado Molecular y Bioinformática de la Sección de Estudios de Posgrado e Investigación de la Escuela Superior de Medicina del IPN.

Asimismo, ha sido de gran importancia el apoyo de los estudiantes, que me acogieron durante mi estancia en la Escuela Superior de Medicina del IPN.

No menos importante ha sido la ayuda y los ánimos prestados por cada uno de mis profesores que he tenido durante mi vida de estudiante. Especialmente quiero agradecerle a la Dra. Martha E. Poisot Vázquez y a la Dra. Judith Amador Hernández, por esos consejos tan oportunos que lograron mi permanencia en la universidad.

A mis sinodales: Dr. Enrique Villalobos Amador, Dr. Ramón Garduño Juárez y Dra. Hermenegilda Moreno Díaz, por cada una de sus aportaciones a este trabajo.

Al proyecto 132353 de ciencia básica de CONACyT.

A la jefatura de la licenciatura en Ciencias Químicas y a la Universidad del Papaloapan Campus Tuxtepec.

Gracias

CONTENIDO GENERAL

CONTENIDO GENERAL	VI
ÍNDICE	VII
ÍNDICE DE FIGURAS	X
ÍNDICE DE TABLAS	XI
ABREVIATURAS	XII
RESUMEN	XIII

ÍNDICE

1.	INTRODUCCIÓN	1
2.	ANTECEDENTES	3
2.1.	Epidemiología del cáncer	3
2.2.	Génesis del cáncer	3
2.3.	Epigenética	4
2.4.	Histonas	5
2.5.	Acetilación y deacetilación	6
2.6.	Histonas deacetilasas (HDAC's)	7
2.6.1.	Mecanismo catalítico de las HDAC's pertenecientes a la clase I	8
2.6.2.	Inhibidores de las HDAC's	9
2.7.	Ácido valproico	11
2.8.	Procedimientos <i>in silico</i>	12
2.8.1.	Mecánica molecular	14
2.8.1.1.	Acoplamiento Molecular (docking)	15
2.8.2.	Mecánica cuántica	17
2.8.2.1.	Aproximación de Born-Oppenheimer	19
2.8.3.	Ventajas y desventajas de los métodos computacionales	23
2.8.4.	Optimización geométrica	24
2.8.5.	Relación cuantitativa estructura-actividad (QSAR)	25
2.8.5.1.	Grupos funcionales activos de un fármaco	28
3.	PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN	30
3.1.	Planteamiento del problema	30
3.2.	Justificación	30
4.	HIPÓTESIS	32
5.	OBJETIVOS	33
5.1.	Objetivo general	33
5.2.	Objetivos particulares	33
6.	MATERIAL Y METODOLOGÍA	34

6.1.	Instalaciones y equipo de cómputo	34
6.2.	Metodología	34
6.3.	Docking	35
6.3.1.	Minimización de la energía de los derivados del AVP	36
6.3.2.	Estructura tridimensional	36
6.3.3.	Docking ciego	36
6.3.4.	Probabilidad	37
6.3.5.	Análisis de los sitios catalíticos de los derivados del AVP y los análogos del hidroxamato	40
6.4.	QSAR	40
6.4.1.	Elección de un conjunto de compuestos con actividad biológica conocida	41
6.4.2.	Identificación de los descriptores moleculares de los análogos del hidroxamato	41
6.4.3.	Selección de los descriptores de los análogos del hidroxamato	42
6.4.4.	Revisión de colinealidad de los análogos del hidroxamato	42
6.4.5.	Identificación de los descriptores moleculares de los derivados del AVP	42
6.4.6.	Regresión Lineal Múltiple (RLM)	42
6.4.7.	Expresión matemáticamente de la actividad biológica de los derivados del AVP	43
6.4.8.	Redes Neuronales Artificiales (ANN)	43
6.4.9.	Predicción de la actividad biológica teórica de los derivados del AVP	44
6.4.10.	Modelo matemático de la actividad biológica teórica	44
7.	RESULTADOS Y DISCUSIÓN	45
7.1.	Docking	45
7.1.1.	Minimización de la energía	45
7.1.2.	Estructura tridimensional de la HDAC8	46
7.1.3.	Docking ciego	48

7.2.	QSAR	55
7.2.1.	Conjunto de compuestos con actividad biológica conocida	55
7.2.2.	Identificación de los descriptores moleculares de los análogos del hidroxamato	56
7.2.3.	Selección de los descriptores de los análogos del hidroxamato	56
7.2.4.	Regresión lineal por pasos en SPSS de los análogos del hidroxamato	56
7.2.5.	Identificación de los descriptores moleculares de los derivados del AVP	58
7.2.6.	Regresión Lineal Múltiple (RLM)	59
7.2.7.	Expresión matemática de la actividad biológica de los derivados del AVP	60
7.2.8.	Regresión Lineal Múltiple en relación al descriptor Belm4	62
7.2.9.	Regresión Lineal Múltiple en relación al descriptor IC ₅₀	64
7.2.10.	Redes Neuronales Artificiales (ANN)	65
7.2.11.	Regresión Lineal Múltiple en relación al descriptor Belm4 (57DAVP)	66
7.2.12.	Regresión Lineal Múltiple en relación al descriptor IC ₅₀ (57DAVP)	67
7.2.13.	Predicción de la actividad biológica teórica de los derivados del AVP	68
8.	RESUMEN DE RESULTADOS	71
9.	PERSPECTIVAS	71
10.	CONCLUSIÓN	72
11.	REFERENCIAS	73

ÍNDICE DE FIGURAS

Fig. 1. Nucleosoma.	5
Fig. 2. Remodelación de la cromatina.	6
Fig. 3. Sitio activo de la HDAC8 código PDB: 1W22, inhibidor: hidroxamato.	8
Fig. 4. Mecanismo catalítico de las HDAC pertenecientes a la clase I.	9
Fig. 5. Sitio catalítico de las HDAC clase I.	9
Fig. 6. Estructuras químicas de SAHA y CBHA.	10
Fig. 7. Estructuras químicas del ácido hidroxámico y uno de sus derivados.	11
Fig. 8. Esquema general de herramientas informáticas.	13
Fig. 9. Modelo de dos átomos unidos mediante un resorte.	14
Fig. 10. Complejo estable unión ligando-receptor de la molécula DAVP76.	16
Fig. 11. Modelo de acoplamiento molecular.	17
Fig. 12. Sistema de coordenadas X-Y-Z de n-electrones y N-núcleos.	21
Fig. 13. Estructura base de los derivados del AVP.	31
Fig. 14. Estructura base de los DAVP.	35
Fig. 15. Estructura de las Redes Neuronales Artificiales (ANN).	44
Fig. 16. Derivados del AVP A) DAVP1, B) DAVP5, C) DAVP45 y D) DAVP82.	45
Fig. 17. Derivados del AVP A) HID2, B) HID3, C) HID4 y D) HID5.	46
Fig. 18. Información general de la enzima HDAC8 encontrada en PDB: 3F07.	46
Fig. 19. Enzima HDAC8.	47
Fig. 20. Enzima HDAC8.libre de grupos ajenos.	48
Fig. 21. Docking de la molécula DAVP5 sobre la HDAC8.	49
Fig. 22. Interacciones de los mejores compuestos la superficie de la HDAC8.	51
Fig. 23. Estructuras moleculares de la A) HID26, B) DAVP266, y C) DAVP281.	52
Fig. 24. Modelo M158.	59
Fig. 25. Modelo M158Belm4.	63
Fig. 26. Modelo M158IC ₅₀ .	64
Fig. 27. Modelo M87Belm4.	66
Fig. 28. Modelo M87IC ₅₀ .	67
Fig. 29. Estructura base de los derivados del AVP (<i>orto, meta y para</i>).	68

ÍNDICE DE TABLAS

Tabla 1. Ventajas y desventajas de los diferentes Métodos computacionales.	24
Tabla 2. Grupos electroattractores y electrodonadores.	35
Tabla 3. Energías y K_d de los mejores análogos del hidroxamato.	50
Tabla 4. Energías y K_d de los mejores derivados del AVP.	50
Tabla 5. Sustituyentes químicos de los derivados del AVP más prometedores.	52
Tabla 6. Aminoácidos del sitio catalítico de los mejores análogos del hidroxamato.	54
Tabla 7. Aminoácidos del sitio catalítico de los mejores derivados del AVP.	54
Tabla 8. Actividad biológica de los 30 análogos del hidroxamato.	55
Tabla 9. Modelos de la 2 ^a RLM del conjunto de entrenamiento.	57
Tabla 10. Coeficientes estadísticos de la 2 ^a RLM del conjunto de entrenamiento.	57
Tabla 11. Coeficientes estadísticos de la 2 ^a RLM de los derivados del AVP.	58
Tabla 12. Modelos de la 3 ^a RLM del conjunto de entrenamiento.	61
Tabla 13. Coeficientes estadísticos de la 3 ^a RLM del conjunto de entrenamiento.	61
Tabla 14. Coeficientes estadísticos de la 3 ^a RLM del conjunto de predicción.	62
Tabla 15. Coeficientes estadísticos de la 4 ^a RLM del conjunto de predicción.	62
Tabla 16. Modelo de la 4 ^a RLM del conjunto de predicción.	62
Tabla 17. Grupos funcionales de los 57 DAVP con su respectiva IC_{50} teórica.	69

ABREVIATURAS

AVP	Ácido valproico
HDAC8	Histona deacetilasa 8
QSAR	Relación cuantitativa estructura actividad biológica
Docking	Acoplamiento molecular
HDAC	Histona deacetilasa
OMS	Organización Mundial de la Salud
p53	Proteína humana localizada en el cromosoma 17p
ADN	Ácido desoxiribonucleico
ARN	Ácido ribonucleico
HAT	Histona acetiltransferasa
Ac	Acetilación
FT	Factor de Transcripción
ARNm	Ácido ribonucleico mensajero
SIRT	Sirtuinas (clase de enzimas)
NAD ⁺	Nicotinamida Adenina Dinucleótido
PDB	Protein Data Bank
TSA	Tricostatina A
TPX	Tetrapéptido trapoxina
SAHA	Ácido hidroxámico de suberoilánilina
CBHA	Ácido bis-(<i>m</i> -carboxicinamil) hidroxámico
Folding	Plegamiento de proteínas
DFT	Teoría de los Funcionales de la Densidad
HEP	Hipersuperficie de Energías Potenciales
ADME	Propiedades farmacocinéticas
CoMFA	Análisis comparativo en el campo molecular.
COMSIA	Análisis comparativo por índice de semejanza molecular.
ΔG	Energía libre de Gibbs
K_d	Constante de disociación
K_A	Constante de Afinidad
AM1	Austin Model 1
CLOA-OM	Combinación Lineal de Orbitales Atómicos y Moleculares
R	Coeficiente de correlación lineal
SD	Desviación estándar
ANN	Redes neuronales artificiales
AIC	Criterio de información de Akaike
FIT	Función de aptitud de Kubinyi

RESUMEN

El tema central de este estudio es la evaluación *in silico* de 500 derivados mono-, di- y tri-sustituídos del ácido valproico (AVP) sobre la histona deacetilasa 8 (HDAC8) con potencial uso antineoplásico.

A partir de métodos computacionales, se establecieron estudios de acoplamiento molecular (Docking) entre la HDAC8-38 análogos del hidroxamato y la HDAC8-500 derivados del AVP, así como la relación cuantitativa estructura-actividad biológica (QSAR) de los derivados del AVP con sitio catalítico de inhibición similar a los observados en los análogos del hidroxamato. Estos modelos permitieron establecer a los mejores derivados del AVP y su actividad biológica teórica, por lo que al unir ambos métodos computacionales, Docking y QSAR, con estudios experimentales, estos arrojan resultados cercanos a los valores reales, favoreciendo así el análisis de datos.

1. INTRODUCCIÓN

En los últimos años el cáncer ha sido considerado como uno de los principales problemas de salud pública, ya que a pesar de los avances en investigación científica y médica, actualmente fallecen más de 7 millones de personas en el mundo [1]. Médicamente es considerado como la pérdida de la capacidad de la célula para controlar su crecimiento y desarrollo, es decir, pierde su proliferación normal [2]. Para contrarrestar esta excesiva proliferación y poder erradicar esta patología, se han desarrollado una serie de inhibidores de las enzimas denominadas histonas deacetilasas (HDAC's), las cuales están asociadas con la eliminación de los grupos acetilo de los residuos de lisina de las histonas, responsables de la proliferación excesiva de las células. Muchos de estos fármacos han fracasado debido a los efectos secundarios que provocan en el organismo. Actualmente se ha descubierto que el ácido 2-propilpentánico o ácido valproico (AVP), un fármaco con actividad antiepiléptica, también actúa como inhibidor de las HDAC's [3]. Esta actividad inhibitoria se puede separar de la actividad antiepiléptica mediante modificaciones apropiadas en la estructura química del AVP [4], utilizando este fármaco en pacientes con cáncer.

Para poder entender la acción y el sitio de unión que tienen los derivados del AVP sobre la HDAC8, se puede recurrir a la utilización de diversos programas computacionales tales como AutoDock 4.0 y Gaussian 03, los cuales permiten calcular algunas propiedades moleculares como la geometría, la energía mínima, la energía libre de Gibbs (ΔG), la constante de disociación (K_d), entre otras. Para ello se propusieron y construyeron en GaussView 500 derivados del AVP, posteriormente se realizó la optimización de las estructuras utilizando el método semiempírico AM1. Una vez concluidos los cálculos de optimización se recurrió al uso del programa AutoDock para poder realizar el docking ciego de todas las estructuras químicas, con el fin de reconocer el sitio de inhibición de la HDAC8, se realizó un análisis QSAR (entre derivados del AVP y análogos de hidroxamato), encontrando el mejor modelo matemático que describe la actividad teórica de los derivados del AVP en relación a los análogos de hidroxamato.

La estructura de la tesis está dividida en diversos capítulos. En el capítulo 2 se detallan los antecedentes del cáncer, su epidemiología y la génesis; además con este capítulo sabremos que son las histonas deacetilasas, su mecanismo de inhibición y los inhibidores que se han estado utilizando en los últimos años, por otro lado tendremos

una idea general acerca de los métodos teóricos-computacionales, los cuales han avanzado de manera exponencial tanto en la capacidad de cálculo como en la resolución de problemas de tipo médico-químico-biológico.

En el capítulo 3 se planea el problema y la justificación por la cual se eligió realizar la Evaluación *in silico* –Docking y QSAR- de derivados del ácido valproico sobre la HDAC8 con potencial uso antineoplásico, mientras que en los capítulos 4 y 5 se muestran la hipótesis y los objetivos general y específicos que se plantearon desde un inicio, los cuales se ven reflejados en los resultados obtenidos en cada una de las metodologías realizadas (ver capítulos 6, 7 y 8). Pero como en todo trabajo de tesis al final surgen nuevas ideas que nos podrían ayudar a corroborar nuestros resultados, por lo que en el capítulo 9 se presentan las perspectivas que surgen al finalizar esta tesis.

2. ANTECEDENTES

2.1. Epidemiología del cáncer

De acuerdo a la Organización Mundial de la Salud (OMS), el cáncer es la primera causa de mortalidad a nivel mundial y se le atribuyen 7.6 millones de defunciones ocurridas en 2008, aproximadamente el 13% de la mortalidad global. Desafortunadamente se prevé que las muertes por cáncer en todo el mundo seguirán aumentando hasta rebasar los 11 millones en 2030 [1]. De acuerdo a las estadísticas más de 70 mil mexicanos mueren al año a causa de diversos tipos de cáncer, los más importantes son la leucemia, cáncer de mama, de próstata, de cuello uterino, de tráquea, de bronquios y de pulmón [5].

2.2. Génesis del cáncer

A pesar de que existen múltiples factores asociados a la génesis del cáncer, algunos de ellos no han sido identificados totalmente, aunque el estilo de vida está relacionado con determinados tipos de neoplasias. El comportamiento anormal de las células cancerosas es el resultado de la acumulación de numerosos defectos, los cuales se pueden agrupar en dos tipos: alteraciones genéticas y alteraciones de los mecanismos epigenéticos.

Hay información suficiente que sugiere que las causas para que una neoplasia se desarrolle son las mutaciones (deleciones, sustituciones, inversiones, translocaciones, etc.) en genes específicos, entre ellos se encuentran los oncogenes que son los genes dañados, cuya presencia en cierta forma y/o la sobre actividad pueden estimular el desarrollo del cáncer, los oncogenes están relacionados con los genes normales llamados protooncogenes, los cuales codifican los componentes del mecanismo de control de crecimiento normal de la célula. Algunos de estos componentes son factores de crecimiento, receptores, enzimas señaladoras y factores de transcripción [6].

Por otro lado los genes supresores de tumor producen la síntesis de productos que frenan el crecimiento y la multiplicación celular [7]. Un gen supresor de tumor en particular codifica la proteína conocida como "p53" (proteína humana localizada en el cromosoma 17p). La proteína p53 actúa como un "freno" que detiene el crecimiento y la división de las células. En el caso que no se pueda reparar el daño, iniciará el suicidio

celular (apoptosis), en células que han sufrido daño en su ADN, previniendo así el crecimiento descontrolado de las células genéticamente dañadas.

Existe otro grupo de genes, los genes de reparación del ADN, los cuales mantienen la estructura del material genético y participan en la reparación del mismo [8].

2.3. Epigenética

En los últimos años se ha está llevando a cabo un esfuerzo importante para el desarrollo de nuevas terapias frente al cáncer basadas en la modificación de las histonas. El conocimiento de la bioquímica de las células cancerígenas ha permitido el desarrollo de fármacos con capacidad de interferir en el crecimiento de las células transformadas frente a otras no tumorales. Así, en los últimos años se ha puesto de manifiesto la importancia de la epigenética en la búsqueda de nuevos blancos terapéuticos, útiles en un tratamiento contra el cáncer [9].

En la actualidad la epigenética se utiliza para hacer referencia a los cambios en la expresión de genes, sin que se produzcan modificaciones en la estructura de su secuencia de ADN. Entre los mecanismos epigenéticos más conocidos y generales podemos destacar los siguientes [10]:

a) Metilación del ADN. Consiste en la adquisición de grupos metilo, de forma que la metilación se asocia con la expresión del gen.

b) Remodelación de la cromatina. La cromatina está constituida por el ADN y un grupo de 4 proteínas. La cadena de ADN se enrolla en un octágono de esas cuatro proteínas constituyendo subunidades que se llaman nucleosomas. Dado que las histonas están esencialmente compuestas de aminoácidos cargados positivamente, estas se pueden unir al ADN que tiene carga negativa en forma estable. Cada una de las ocho histonas tiene una cola amino-terminal que se extiende fuera del nucleosoma, y que puede ser modificada epigenéticamente. Cuando la cromatina se condensa, se acorta, y los nucleosomas se juntan comprimiendo el ADN. Esta es una situación estructural del ADN en la que no se puede realizar la transcripción del mensaje genético. Por el contrario, cuando se va a transcribir el mensaje de un gen, la cromatina de esa zona se relaja separando los nucleosomas, lo que permite la entrada de los factores de transcripción, activando funcionalmente al gen y transmitiendo el mensaje.

c) Por la acción de los micro-ARN. Recientemente se ha identificado un gran número de ARN pequeños, que generalmente provienen de zonas no codificantes, llamados micro-ARN. Aunque para muchos de ellos su función aún no es completamente conocida, se sabe que algunos tipos de micro-ARN tienen una importante contribución en la regulación de la expresión génica durante el desarrollo prenatal.

En general, los procesos epigenéticos se realizan a través de diferentes familias de enzimas como las ADN metiltransferasas, acetiltransferasas, deacetilasas, demetilasas y fosforilasas, entre otras.

2.4. Histonas

El genoma humano se localiza dentro del núcleo celular en la cromatina, que es un complejo macromolecular dinámico formado por nucleosomas. Un único nucleosoma (Fig.1) se compone de un fragmento de ADN (146 pares de bases) enrollado alrededor de un octámero de histona formado por cuatro núcleos de histona: un tetrámero H3-H4 y dos dímeros H2A-H2B (Fig. 1B).

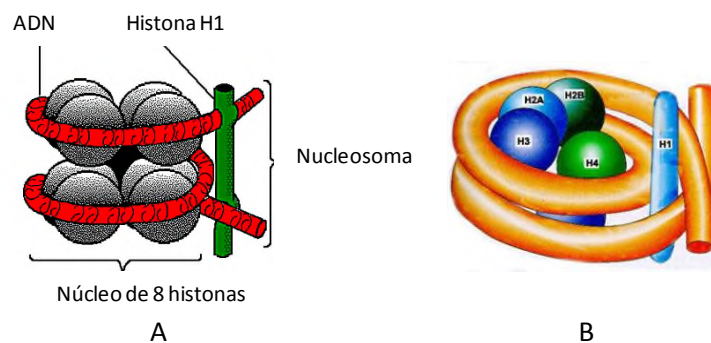


Fig. 1. Nucleosoma. A) Esquema general del nucleosoma y B) ubicación del tetrámero H3-H4 y de los dímeros H2A-H2B.

Las histonas son pequeñas proteínas básicas ricas en los aminoácidos lisina y arginina. Las cuatro histonas nucleosómicas contienen dos dominios: el dominio C-terminal localizado dentro del nucleosoma y el dominio N-terminal con residuos de lisina extendidos fuera del mismo [11]. En estas proteínas se enrolla el material genético y por tanto las interacciones interhistonas facilitan o limitan la accesibilidad de los promotores nucleares a las cadenas de ADN, lo que induce o inhibe la transcripción del gen en cuestión (Fig. 2B). Los mecanismos que provocan cambios conformacionales y por tanto de interacción entre las histonas son la acetilación, metilación, fosforilación y la

ubiquitinación. La acetilación-deacetilación es el principal moderador de las interacciones interhistonas [12].

2.5. Acetilación y deacetilación

El extremo N-terminal de las histonas pasa a través y alrededor de la doble hélice del ADN, y su modificación, mediante acetilación-deacetilación es crucial en la modulación de la expresión genética, al afectar la interacción del ADN con los complejos proteicos no nucleosomales reguladores de la transcripción [13].

La acetilación de residuos de lisina en las secuencias N-terminales está mediada por las enzimas denominadas histona acetiltransferasas (HAT). Los grupos acetilo son eliminados de las ξ -N-acetil-lisinas por la actividad de las histonas deacetilasas. Las actividades de las HAT y las HDAC se asocian a los genes diana a través de los complejos constituidos por factores de transcripción específicos para ciertas secuencias y sus respectivos cofactores. El balance entre las actividades opuestas de las HAT y las HDAC regula el estado de acetilación de las histonas [14].

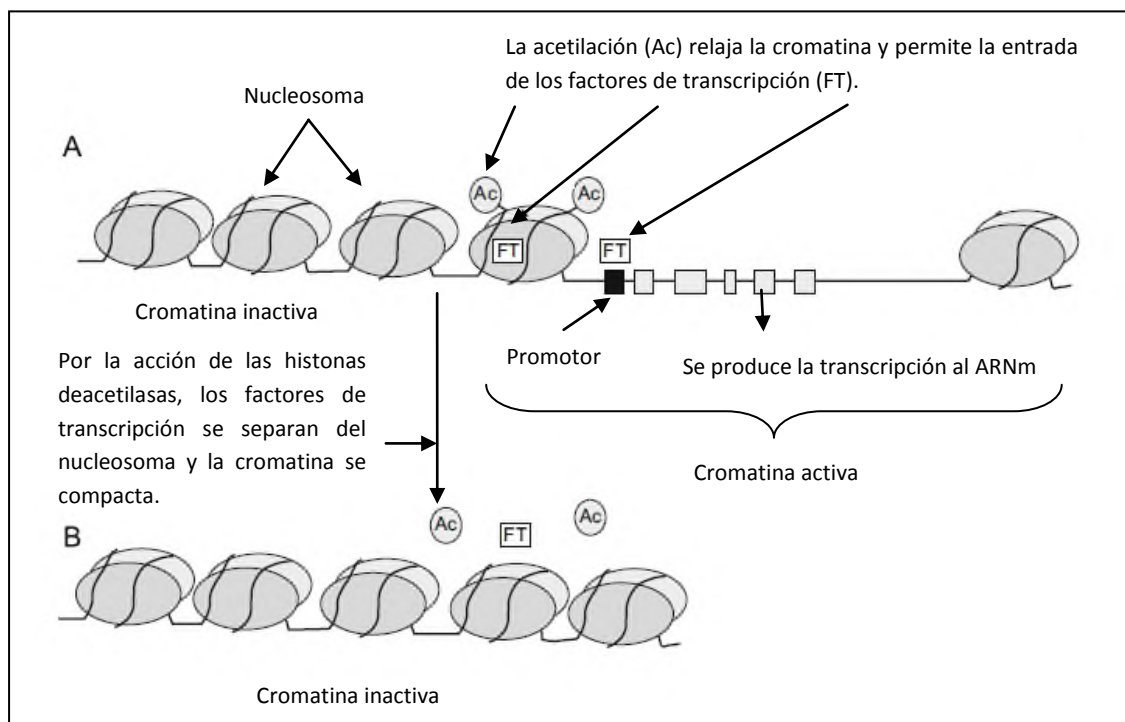


Fig. 2. Remodelación de la cromatina: A) Cuando las histonas de los nucleosomas se acetilan (Ac), la cromatina se relaja y deja pasar a los factores de transcripción (FT), es decir, se activa la función del gen y se transcribe el mensaje al ARNm. B) Cuando actúan las histonas deacetilasas, se pierde la acetilación, la cromatina se compacta y se impide la transcripción.

En la Fig. 2A se observa que la acetilación de las histonas relaja la estructura de la cromatina separando los nucleosomas, permitiendo la transcripción del gen, esto debido a que se neutralizan sus cargas positivas. En consecuencia, la pérdida de la acetilación tendrá el efecto contrario, concentrando la cromatina y silenciando la función del gen (Fig. 2B).

En general, los altos niveles de acetilación (hiper-acetilación) se asocian a un incremento de la actividad transcripcional, mientras que bajos niveles de acetilación (hipo-acetilación) se asocian a la represión de la expresión genética. Esto es debido a que al disminuir el espacio entre el nucleosoma y el ADN enrollado a su alrededor se dificulta el acceso de los factores de transcripción, lo que conduce a una represión transcripcional [15].

2.6. Histonas deacetilasas (HDAC's)

Las HDAC's son un tipo de enzimas implicadas en la eliminación de los grupos acetilo de los residuos de lisina en las histonas [16-17], las cuales causan enfermedades en el humano tales como, leucemia mieloide aguda, tumores sólidos, carcinoma de mama, entre otras [18].

Actualmente se han encontrado 11 isoformas de las HDAC's [19], las cuales se dividen en cuatro clases principales:

- I. La clase I está conformada por las HDAC1-HDAC3 y HDAC8, las cuales se encuentran principalmente en el núcleo celular, mientras que la HDAC3 también se encuentra en el citoplasma y está asociada a la membrana, las cuales son homólogas al gen RPD3 de levadura (potasio reducido dependiente 3) del que se sabe que es necesario para la represión normal de numerosos genes de levadura [20-26].
- II. La clase II incluye las HDAC4-HDAC7, HDAC9-HDAC10, constituidas por dos sitios catalíticos en su estructura, distribuidas entre el citoplasma y el núcleo, similar a la HDA1 [27-34].
- III. En la clase III (o sirtuin SIRT) se agrupan todas aquellas enzimas que utilicen NAD^+ [35-37] como cofactor en lugar de Zn^{2+} .
- IV. Finalmente, en la clase IV se encuentra la HDAC11 que comparte homología con la clase II pero carece de uno de los sitios catalíticos.

descomposición de dicho estado de transición es aportado por el sistema His-Asp (Fig. 4C) [42].

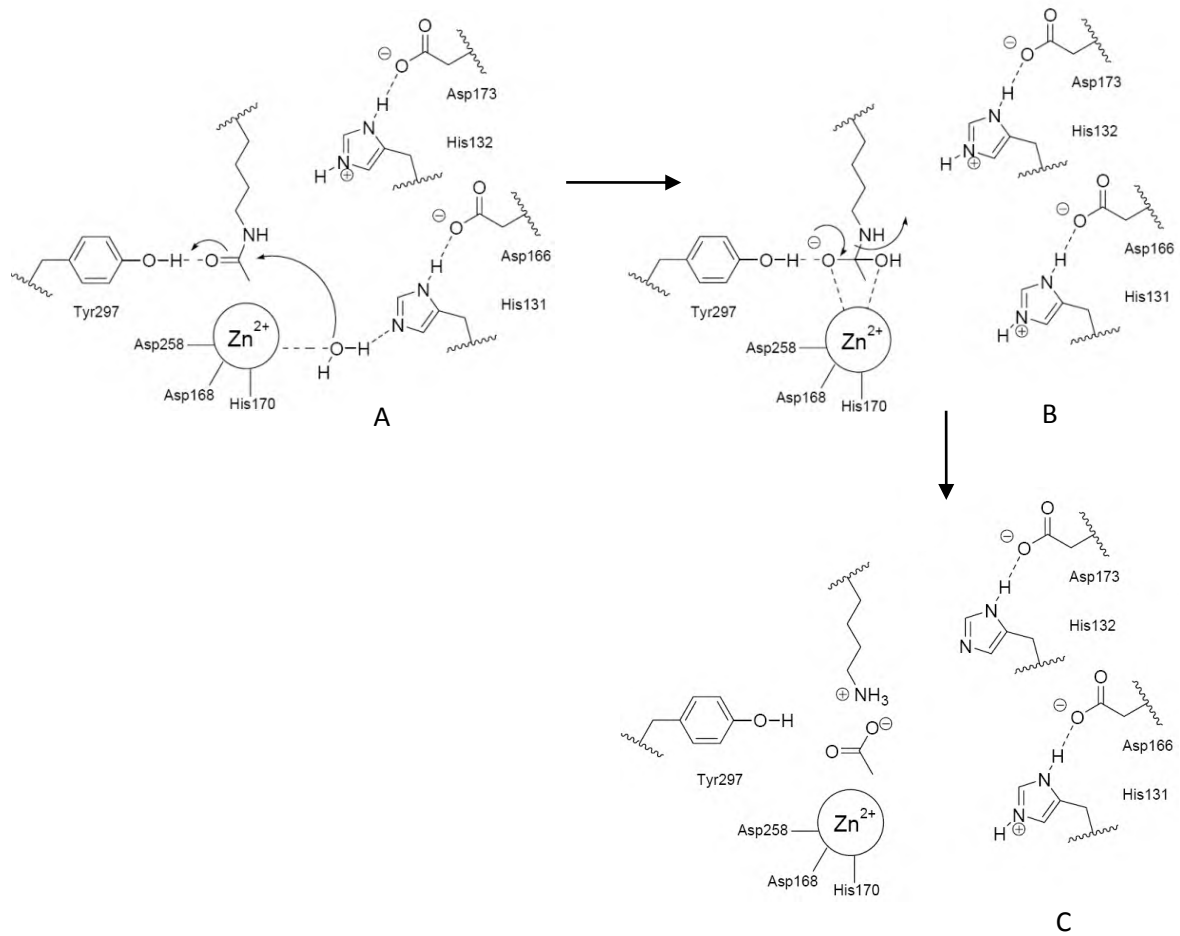


Fig. 4. Mecanismo catalítico de las HDAC pertenecientes a la clase I [42].

2.6.2. Inhibidores de las HDAC's

El diseño de inhibidores de las histonas deacetilasas de la clase I, está basado en la existencia de un átomo de Zn^{2+} catalítico coordinado con los residuos de Asp, His y Tyr. Todo ello constituye una cavidad polar dentro del sitio catalítico enzimático de naturaleza polar [43] (Fig. 5).

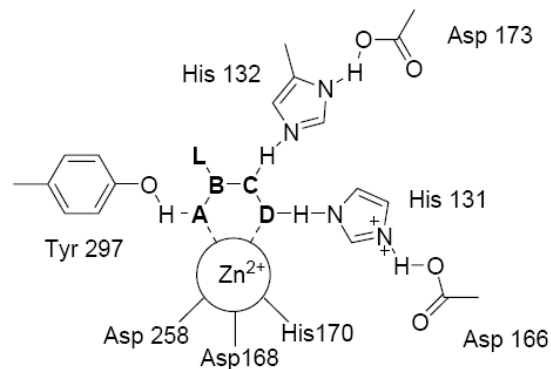


Fig. 5. Sitio catalítico de las HDAC clase I. Incluye los fragmentos A, B, C, D y L.

Por tanto, un fragmento químico que le confiere la capacidad para situarse dentro del sitio activo enzimático y coordinarse con el átomo de Zn^{2+} debería poseer las siguientes características (Fig. 5):

1. El fragmento A debería ser una base blanda con capacidad para donar electrones y coordinarse con el átomo de Zn^{2+} , un aceptor de enlaces de hidrógeno que permita establecer un puente de hidrógeno con el grupo hidroxilo de la Tyr, o un dador de enlaces de hidrógeno, capaz de establecer dicho enlace con el oxígeno de la Tyr.
2. El fragmento B debe ser un grupo trivalente que permita unir el fragmento quelante del Zn^{2+} con el fragmento espaciador.
3. El fragmento C debe ser un grupo trivalente donador de enlaces de hidrógeno con capacidad para establecer este tipo de enlaces con el resto de la His132.
4. El fragmento D debe ser un grupo con capacidad para coordinarse al átomo de Zn^{2+} y también establecer un puente de hidrógeno con el resto de la His.
5. El fragmento L permite la unión del grupo quelante del átomo de Zn^{2+} con la cabeza lipófila.

Se han descrito varias clases de inhibidores de las HDAC's, que comúnmente se suelen clasificar en dos tipos: los naturales y los sintéticos.

Entre los inhibidores naturales se encuentran la tricostatina A (TSA) [44], la apicidina [45], el tetrapéptido trapoxina (TPX) [46], entre otros; la TSA ha sido identificada como el fármaco más potente y específico inhibidor de la HDAC. Entre los inhibidores sintéticos se encuentran el fenil butirato de sodio [47], el ácido hidroxámico de suberoilánilida (SAHA), el ácido bis-(*m*-carboxicinamil) hidroxámico (CBHA) [48], la piroxamida, la oxamflatina [49] y el ácido valproico [50-51].

La Fig. 6 muestra las estructuras químicas de SAHA (1) y CBHA (2), las cuales son inhibidores activos frente a la HDAC1 (IC_{50} = 0.01 μ M para ambos compuestos) y HDAC3 a concentraciones micromolares (IC_{50} = 0.02 μ M y 0.07 μ M, respectivamente) [52].

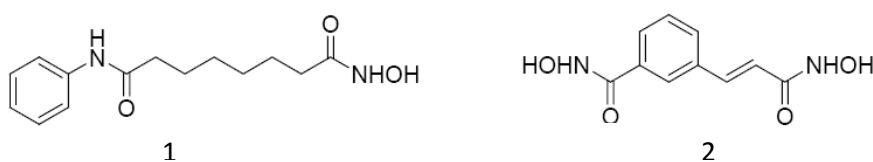


Fig. 6. Estructuras químicas de SAHA (1) y CBHA (2).

El ácido hidroxámico (**3**) y sus derivados (Fig. 7) son otra clase de inhibidores de las HDAC's [53]. El compuesto más potente frente a las HDAC's humanas recombinantes es el compuesto 4 con una IC₅₀ de 0.06 μM.

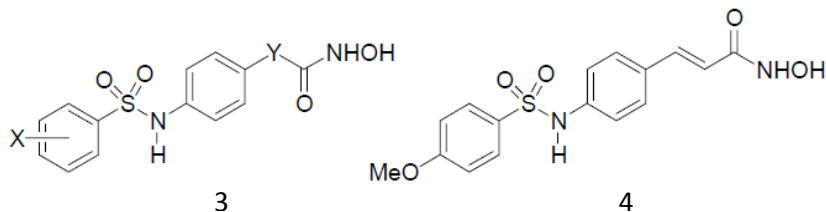


Fig. 7. Estructuras químicas del ácido hidroxámico (3) y uno de sus derivados (4).

En 2005 Eyal S. y colaboradores determinaron la inhibición de las HDAC's usando isómeros estructurales del AVP, los más potentes fueron el 4-ene-valproico y el 2-ene-valproico, con una IC₅₀ de 1.5 μM y 2.8 μM respectivamente, mientras que para el AVP fue de 1.5 μM [54].

La mayoría de estos compuestos, poseen un anillo aromático (fragmento lipofílico) y un grupo funcional con la habilidad de coordinarse con el ión Zn²⁺ ubicado en el sitio catalítico [55].

2.7. Ácido valproico

El ácido valproico, ha sido utilizado como fármaco en psiquiatría y neurología ya que presenta una acción estabilizadora y anti-epiléptica [56-59], bloqueando los disparos sostenidos y repetitivos de alta frecuencia de las neuronas a concentraciones terapéuticas (25–30 mg/kg por día), aunque el mecanismo de este fenómeno no se conoce aún [60].

Se sabe que el AVP es un fuerte inhibidor de las histonas deacetilasas, cuyo efecto va a impedir que se separen los grupos acetilo de los nucleosomas, produciendo una hiperexpresión génica (Fig. 2A). De hecho, este mecanismo se ha considerado como la base de los efectos teratogénicos del AVP [61]. Recientemente en pruebas preclínicas se ha observado su actividad antiproliferativa en células neoplásicas [62-66], lo que ha llevado a este fármaco a ser empleado en pacientes con cáncer. Su acción antineoplásica se debe a la inhibición de la enzima HDAC [67-68] dependiente del ion Zn²⁺, consiguiendo recuperar la expresión de genes supresores de tumores.

El AVP es farmacéuticamente aceptable, ya que es capaz de sensibilizar las células cancerosas humanas. Es por esta razón que se ha estado explotando este blanco farmacológico desde el punto de vista experimental, aunque presenta efectos secundarios tales como sedación y cambios en la masa corporal [69]. Más del 70% de los pacientes adultos que reciben tratamiento con AVP han aumentado de peso entre 5 a 14 kg [70].

2.8. Procedimientos *in silico*

La computación es un área muy extensa y prácticamente se encuentra en constante interacción con otros campos de la ciencia; en este caso, el interés se centra en el papel que juega la computación en la Química. A principios de los años 70's apareció un nuevo campo de conocimiento dirigido al diseño de moléculas asistido por computadora: la química computacional, la cual implica el uso de modelos matemáticos para predecir las propiedades químicas, biológicas y físicas de los compuestos, permitiendo la investigación de átomos, moléculas y macromoléculas [71].

La química computacional incluye aspectos como:

- a) El modelado molecular.
- b) Los métodos computacionales.
- c) El diseño molecular asistido por ordenador.
- d) Las bases de datos químicos.
- e) El diseño en síntesis orgánica.

En la química computacional podemos encontrar diversos métodos y niveles de cálculo, los cuales pueden dividirse en dos grandes categorías [72-73]:

- a) La mecánica molecular:
 - I. Docking (acoplamiento molecular).
 - II. Dinámica molecular.
 - III. Folding (plegamiento de proteínas).

Estos métodos aplican las leyes de la física clásica al núcleo molecular sin considerar explícitamente a los electrones.

- b) La mecánica cuántica:
 - I. Métodos semiempíricos.
 - II. Métodos de Hartree-Fock.

III. Métodos posHartree-Fock.

IV. Métodos de Teoría de los Funcionales de la Densidad (DFT).

Los cuales se basan en resolver la ecuación de Schrödinger para describir una molécula con un tratamiento directo de la estructura electrónica.

Dichos métodos teóricos-computacionales han avanzado de manera exponencial tanto en la capacidad de cálculo como en la resolución de problemas de tipo médico-químico-biológico. La metodología utilizada se lleva al cabo a partir de algoritmos basados en las propiedades fisicoquímicas de moléculas pequeñas y grandes para predecir, simular y estudiar sus interacciones con otras moléculas, que usan estrategias muy avanzadas en la búsqueda de soluciones para diversas situaciones, basándose en parámetros obtenidos de manera experimental, por ejemplo, los cálculos de mecánica cuántica para la optimización de geometrías de moléculas pequeñas, el uso de bases de datos con parámetros de mecánica molecular para el plegamiento de proteínas, los cálculos termodinámicos para la predicción de interacciones ligando-proteína, entre otros. Estas ramas se engloban en un campo llamado modelado molecular, el cual nos da un modelo en 2D o 3D de la estructura de una molécula y sus propiedades. Los métodos computacionales calculan la geometría molecular y las propiedades electrónicas para realizar el modelo. En la Fig. 8 se indica una metodología para la obtención de nuevos fármacos a partir de técnicas basadas en la química computacional.

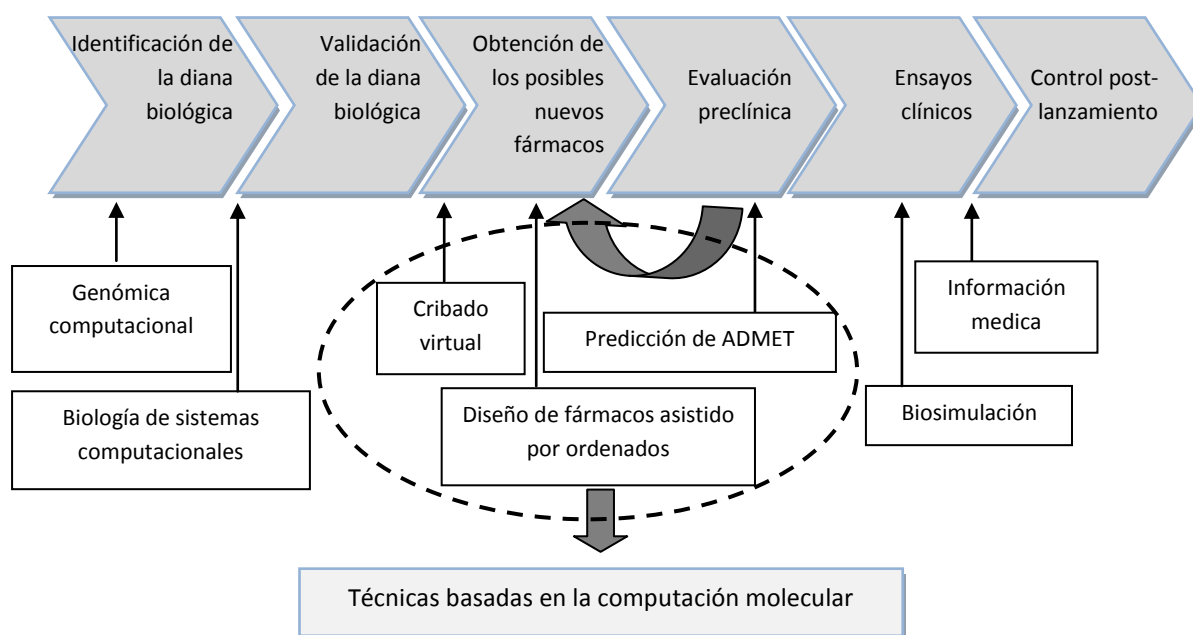


Fig. 8. Esquema general de herramientas informáticas en el diseño de nuevos fármacos.

2.8.1. Mecánica molecular

La mecánica molecular describe la energía de una molécula en términos de un conjunto de funciones de energía potencial. Las funciones de energía potencial y los parámetros usados para su evaluación se conocen como "campo de fuerza".

Los métodos de mecánica molecular más difundidos e incluidos en el programa Gaussian 03 son AMBER [74], DREIDING [75] y UFF [76], englobados en los métodos teóricos, y se suelen utilizar en cálculos tipo capas (ONIOM).

Los métodos de la mecánica molecular se basan en los siguientes principios [77]:

- ✓ Se considera a la molécula como una colección de partículas
- ✓ El núcleo y los electrones están juntos y se tratan como una única partícula, ya que los efectos electrónicos se encuentran implícitos en los campos de fuerza.
- ✓ Estas partículas se tratan como si fueran esferas.
- ✓ Los enlaces entre partículas se comportan como osciladores armónicos (Fig. 9).



Fig. 9. Modelo de dos átomos unidos mediante un resorte.

- ✓ Se utilizan funciones de potencial individuales para describir las diferentes interacciones: tensión de enlaces, flexión de ángulos de enlace como ángulos torsionales, así como interacciones no enlazantes.

$$U = \sum u_s + \sum u_b + \sum u_t + \sum u_{vdW} + \dots$$

donde:

U es la energía potencial total de la molécula.

u_s es la suma de la energía de vibración longitudinal de los enlaces entre los átomos.

u_b es la energía de deformación del ángulo entre tres átomos.

u_t es la energía de torsión, que se relaciona con la rotación alrededor del enlace.

y u_{vdW} es la energía proveniente de las interacciones de van der Waals.

- ✓ Las funciones de energía potencial se relacionan con parámetros derivados empíricamente (constantes de fuerza y valores de equilibrio) que describen las interacciones entre los átomos.
- ✓ La suma de todas las interacciones determina la distribución espacial (conformación) de un conjunto de átomos.
- ✓ Las energías obtenidas en la mecánica molecular no son cantidades absolutas, sino que solo se pueden utilizar para comparar energías estéricas relativas entre dos o más conformaciones de la misma molécula.

Por lo tanto la energía total de una molécula E se puede describir, como suma de las diferentes interacciones.

Es decir,

$$E = E. \text{ de Tensión} + E. \text{ de Flexión} + E. \text{ de Torsión} + E. \text{ de Interacciones no Enlazantes.}$$

2.8.1.1. Acoplamiento Molecular (docking)

El uso del campo del acoplamiento molecular se ha visto aumentado en los últimos años, entre algunas de las principales aportaciones se encuentran las de David S. Goodsell [78], quien fue el fundador del programa llamado AutoDock, cuyo objetivo es el estudio de la compatibilidad de moléculas pequeñas llamadas ligandos con macromoléculas como candidatos a la formación de fármacos.

La aplicación de métodos computacionales para el estudio de la formación de complejos intermoleculares, ha sido objeto de intensa investigación; de ahí que hoy en día existen diversos tipos de programas para el estudio de docking; siendo la implementación de uno u otro algoritmo computacional, así como la fórmula para la optimización de la función de energía, los que determinan un mejor funcionamiento de un programa en comparación con otros.

El docking consiste en la predicción de la orientación y posición preferida de una molécula con respecto a una segunda con el fin de formar un complejo estable, con esta información, y teniendo en cuenta las teorías termodinámicas, es posible conocer el estado de equilibrio de esta interacción [79]. Por lo que el docking juega un papel importante en el diseño racional de nuevos fármacos [80].

Dos aproximaciones son las más populares dentro de la comunidad del docking. La primera usa la técnica de “matching” que describe a la proteína y al ligando como superficies complementarias. La segunda aproximación simula el proceso de docking como el par ligando-proteína y se calculan las energías de unión, las energías libres, o alguna cantidad numérica medible como función, para determinar cuál es el complejo más estable (Fig. 10).

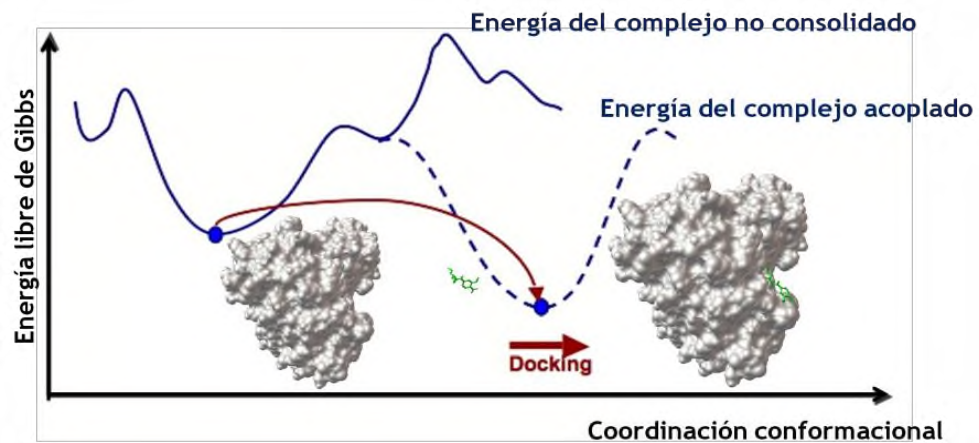


Fig. 10. Complejo estable unión ligando-receptor de la molécula DAVP76 con la HDAC8.

En la simulación del docking, la proteína y el ligando están separadas a una distancia física, de ahí el ligando encuentra su posición en el sitio afín de la proteína, esto después de un cierto número de movimientos en el espacio conformacional, estos movimientos se incorporan como transformaciones de cuerpo rígido tales como traslaciones y rotaciones, también con cambios internos en la estructura del ligando incluyendo torsiones. Cada uno de estos movimientos en el espacio conformacional del ligando induce un costo en la energía total del sistema.

Al calcular el cambio de energía total en el sistema ($\Delta G_{\text{unión}}$) derivada de cada una de las interacciones en condiciones estándar, es posible conocer la constante de disociación, la cual corresponde a la constante de equilibrio de la disociación del complejo proteína-ligando (Fig. 11). Cuanto menor sea la constante de disociación menor es la concentración de ligando necesaria para saturar la proteína y, en consecuencia, mayor es la afinidad (K_A) de la proteína hacia el ligando, suponiendo que esta mayor afinidad corresponde a la afinidad del sitio activo.

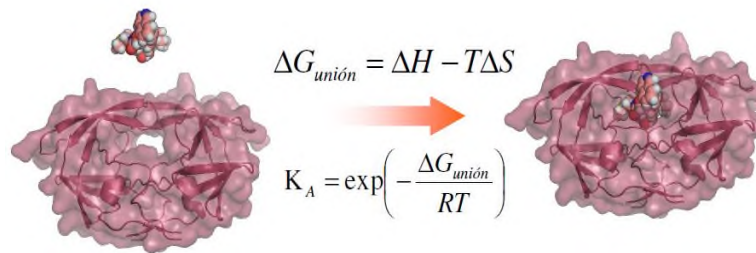


Fig. 11. Modelo de acoplamiento molecular.

Podemos definir la afinidad de una enzima hacia un ligando como un parámetro semicuantitativo relacionado con la inversa de la constante de disociación; al conocer estos dos parámetros podemos llevar los resultados a la segunda ley de la termodinámica:

$$\Delta G_{unión} = -RT \ln K_A$$

El programa que se utiliza para realizar el docking molecular, tiene como característica principal el uso de un algoritmo genético lamarkiano, el cual genera los resultados de las probables conformaciones del ligando en el espacio de búsqueda de la proteína y, al término cada proceso, se enlistan a partir de su energía de unión:

$$\Delta G_T = \Delta G_H + \Delta G_{vdW} + \Delta G_D + \dots$$

donde ΔG_T es el cambio en la energía total del sistema, ΔG_H es la energía de interacción atribuida a puentes de hidrógeno; ΔG_{vdW} es la energía de interacción atribuible a las interacciones de van der Waals, y así sucesivamente con todos los tipo de interacción débiles como las iónicas, las aromáticas, las electrostáticas, etc.

La ventaja de este método es que es mucho más manejable para incorporar la flexibilidad en el ligando, otra ventaja es que el proceso es físicamente mucho más cercano a lo que sucede en la realidad. Una desventaja de esta técnica es que necesita una gran cantidad de tiempo para evaluar las conformaciones óptimas para la unión de la molécula estable y se tiene que explorar una gran cantidad de posibilidades.

2.8.2. Mecánica cuántica

El punto de vista de la mecánica cuántica ha revolucionado la manera de interpretar fenómenos espectroscópicos, eléctricos, pero en especial los fenómenos químicos; el gran desarrollo experimentado por la química cuántica ha permitido, al

menos para moléculas aisladas, hacer predicciones de propiedades estructurales, termodinámicas, espectroscópicas y cinéticas.

Dichas propiedades moleculares se pueden determinar resolviendo la ecuación de Schrödinger, donde los distintos métodos de cálculo de la estructura electrónica se caracterizan por sus distintos niveles de aproximación a la solución exacta de la misma.

Básicamente se pueden clasificar en dos tipos: los métodos semiempíricos y los métodos *ab initio* ("a primeros principios").

El más antiguo de los métodos semiempíricos es el propuesto por Hückel [81] el cual sólo toma en cuenta los electrones π de moléculas aromáticas, los métodos semiempíricos más difundidos son los conocidos con las siglas AM1, MNDO, MINDO/3 [82], CNDO [83], INDO [84], PM3 y PM3MM [85], los cuales presentan diferentes parametrizaciones utilizadas para aproximar las interacciones entre átomos.

La elección de un método de mecánica cuántica depende de la disponibilidad de los recursos computacionales, de la calidad de las soluciones requeridas y de la parametrización del método.

Teniendo en cuenta lo anterior, en el presente trabajo se realizaron cálculos basados en métodos semiempíricos Austin Model 1 (AM1). Este método usualmente no da resultados tan precisos en las geometrías de las moléculas (longitudes y ángulos), en las energías de los orbitales moleculares, energías de estados de transición y energías de formación como los resultados *ab initio* o DFT, debido a que el método AM1 utiliza un hamiltoniano más simple que el hamiltoniano molecular completo.

A pesar de estas desventajas, este tipo de método es ampliamente usado ya que el tiempo de cálculo y el consumo de recursos computacionales es mucho menor en comparación con los otros métodos computacionales "completos". Con él se obtienen resultados de geometría molecular, energía de las conformaciones, calor de formación, energías de estados de transición, cargas atómicas, densidad electrónica, momento dipolar, potencial electrostático molecular, orbitales moleculares, espectros IR y UV-Visible de forma rápida, los cuales están ajustados a formulas empíricas y parámetros atómicos experimentales [86].

Es importante mencionar que los métodos *ab initio* como los semiempíricos se basan en los siguientes principios:

- ✓ Los núcleos y los electrones se distinguen unos de otros.

- ✓ Las interacciones electrón-electrón y electrón-núcleo están dirigidas por el movimiento y las cargas de los electrones.
- ✓ Los métodos de la mecánica cuántica se resuelven mediante aproximaciones de soluciones aproximadas de la ecuación de onda de Schrödinger (solo es posible para un sistema de un núcleo y un electrón).
- ✓ El operador hamiltoniano, \hat{H} , contiene información que describe la energía cinética y potencial de los electrones y los núcleos de un sistema.
- ✓ La función electrónica de onda, ψ , describe el estado de los electrones en términos de su movimiento y su posición.
- ✓ La energía, E , es la energía asociada un estado electrónico en particular.
- ✓ Las distintas soluciones describen un conjunto de posibles estados para un electrón. La magnitud observable de estos estados se describe con una función de probabilidad, la cual es el cuadrado de la función de onda.
- ✓ La función de probabilidad normalizada indica la probabilidad de encontrar un electrón en ese estado.

2.8.2.1. Aproximación de Born-Oppenheimer

Los métodos computacionales se basan en el cálculo de la hipersuperficie de Energías Potenciales (HEP) de las moléculas, de tal manera que este procedimiento puede ahorrar considerablemente tiempo de cálculo. De las HEP se puede derivar información estructural y química de la molécula, por esta razón la principal aplicación de la mecánica cuántica es la descripción y predicción del estado de un átomo o molécula, la cual se establece por medio de una función de onda que depende de las coordenadas de las partículas y del tiempo. Esta función, contiene toda la información que puede ser determinada para el sistema. Para poder obtener esta información, es necesario tener un operador hermitiano lineal. El operador más importante de la mecánica cuántica es el hamiltoniano \hat{H} [87], que está asociado a la energía total del sistema el cual es un operador de la energía cinética \hat{K} y de la energía potencial \hat{U} ,

$$\hat{H} \Psi(x, t) = -\frac{\hbar^2}{2m} \frac{d^2 \Psi(x, t)}{dx^2} + V(x) \Psi(x, t)$$

juntando estos dos elementos se obtiene la ecuación de Schrödinger dependiente del tiempo:

$$-\frac{\hbar}{i} \frac{\partial \Psi(x, t)}{\partial t} = -\frac{\hbar^2}{2m} \frac{d^2 \Psi(x, t)}{dx^2} + V(x) \Psi(x, t)$$

Resolviendo la ecuación mediante la técnica de separación de variables y suponiendo que la función de energía potencial es independiente del tiempo, se obtiene la ecuación de Schrödinger independiente del tiempo para las funciones propias y los valores propios de la energía;

$$\hat{H} \Psi(x, t) = E \Psi(x, t)$$

$$-\frac{\hbar}{i} \frac{\partial \Psi(x, t)}{\partial t} = E \Psi(x, t)$$

para esto es necesario que la función de onda sea el producto de dos funciones, una que dependa exclusivamente del tiempo y la segunda que dependa de las coordenadas de las partículas, así que

$$\Psi(x, t) = f(t) \psi(x)$$

donde la función $\psi(x)$ satisface la ecuación de Schrödinger independiente del tiempo.

$$\hat{H} \psi(x) = E \psi(x)$$

$$\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} + V(x) \psi(x) = E \psi(x)$$

Cuando se ha de resolver la ecuación de Schrödinger para una molécula, incluso la más sencilla, es necesaria la aplicación de la aproximación de Born-Oppenheimer [88]. Esta aproximación se basa en el hecho de que los núcleos son mucho más pesados que los electrones, lo que conlleva en la separación de los movimientos nucleares y electrónicos; esto es desacoplar el movimiento nuclear del movimiento electrónico [89], permitiendo la determinación de las soluciones de la ecuación de Schrödinger electrónica molecular para cada una de las de posiciones relativas de los diferentes núcleos que componen la molécula.

Las moléculas son sistemas microscópicos que están formadas por n-electrones y N-núcleos:

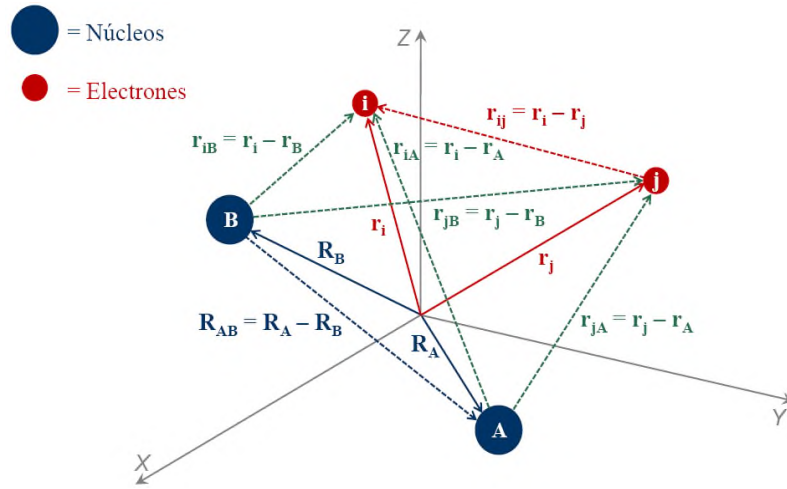


Fig. 12. sistema de coordenadas X-Y-Z de n-electrones y N-núcleos.

Si se asume que los núcleos y los electrones son partículas puntuales con masa y carga se puede escribir el hamiltoniano molecular no relativista de la siguiente manera:

$$\hat{H} = -\frac{\hbar^2}{2} \sum_{A=1}^N \frac{1}{M_A} \nabla_A^2 - \frac{\hbar^2}{2m_e} \sum_{i=1}^n \nabla_i^2 + \sum_{i=1}^n \sum_{i>j}^n \frac{Ze^2}{r_{ij}} - \sum_{i=1}^n \sum_{A=1}^N \frac{Z_A e^2}{r_{iA}} + \sum_{A=1}^N \sum_{B>A}^N \frac{Z_A Z_B e^2}{r_{AB}}$$

Energía potencial electrónica i, j
Energía potencial nuclear

↓
↓

↑
↑
↑

Energía cinética del A-ésimo núcleo
Energía cinética del i -ésimo electrón
Energía potencial del i -ésimo electrón i -y l A-ésimo núcleo A

donde los subíndices i, j representan a los electrones; los subíndices A y B a los núcleos; $\hbar = \frac{h}{2\pi}$; m_e es la masa del electrón; M_A es la masa del núcleo A-ésimo; ∇^2 es el operador laplaciano; e es la carga del electrón; Z_A y Z_B son los números atómicos de los núcleos, r_{AB} es la distancia entre los núcleos A-B y r_{ij} la distancia entre los electrones i - j .

Así que el hamiltoniano anterior se puede separar en un operador electrónico y en un operador de energía cinética nuclear. La ecuación de Schrödinger dependiente de las coordenadas de electrones y núcleos adquiere la siguiente forma:

$$\hat{H} \psi(q_i, q_A) = E\psi(q_i, q_A)$$

Dado que la velocidad de los núcleos es tan pequeña en comparación con la velocidad de los electrones, se desprecia el movimiento de los núcleos, considerando que estos se encuentran en un estado estacionario con respecto a los electrones, por tanto la función de onda electrónica es relativamente insensible a los cambios de la posición y al momento nuclear. De esta manera el hamiltoniano se simplifica omitiendo el término de energía cinética de los núcleos, obteniendo la ecuación de valores propios siguiente,

$$(H_{el} + V_{NN})\psi_{elec} = E_{el}\psi_{elec}$$

donde el hamiltoniano electrónico H_{el} tiene la siguiente forma [90]:

$$H_{el} = -\frac{\hbar^2}{2m_e} \sum_{i=1}^n \nabla_i^2 + \sum_{i=1}^n \sum_{i>j}^n \frac{Ze^2}{r_{ij}} - \sum_{i=1}^n \sum_{A=1}^N \frac{Z_A e^2}{r_{iA}}$$

este hamiltoniano describe el movimiento de los electrones para la posición de los núcleos, es decir, depende de la posición y no del momento.

$$H_{el}\psi_{el} = E_{el}\psi_{el}$$

$$H_{el} = E_{el}\{\mathbf{R}_A\}$$

$$V_{NN} = \sum_{A=1}^N \sum_{B>A}^N \frac{Z_A Z_B e^2}{r_{AB}}$$

V_{NN} es la repulsión entre núcleos.

En la literatura de la química cuántica [89] la suma definida por

$$E_{total} = E_{el}\{\mathbf{R}_A\} + V_{NN}$$

se considera como la función de energía potencial para el movimiento nuclear, y en general da lugar a una hipersuperficie de energías potenciales, mientras que la energía electrónica está en función de las distancias de enlace, ángulos de enlace y diedros de rotación molecular, la cual depende de la configuración nuclear dada por r y por tal razón depende paramétricamente de estas distancia.

Una vez encontrada la función de onda así como la ecuación de Schrödinger electrónica, podemos realizar el cálculo de la energía esperada del sistema. Partiendo de

la aproximación CLOA-OM (Combinación Lineal de Orbitales Atómicos y Moleculares), se pretende encontrar la mejor combinación de orbitales atómicos para la formación de los orbitales moleculares.

$$\psi = a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_Nx_N = \sum_{i=1}^N a_ix_i$$

donde a son los coeficientes de combinación lineal y x los orbitales atómicos.

$$\hat{H} \sum_{i=1}^N a_ix_i = E \sum_{i=1}^N a_ix_i$$

Se prueban distintas funciones variacionales y luego se calculan los coeficientes (vectores propios que minimizan la E de la molécula). Cuando se aplica normalmente el método de variaciones se debe cumplir que la energía calculada E a partir de la ecuación de Schrödinger, proveniente de la función de onda será más grande o igual que la energía que tiene en estado fundamental E_0 , es decir [91]:

$$\int \psi^* \hat{H} \psi d\tau \geq E_0$$

$$E \geq E_0$$

2.8.3. Ventajas y desventajas de los métodos computacionales

Como se comentó anteriormente los métodos computacionales se basan en el cálculo de las HEP de las moléculas, para esto existen algunos tipos básicos de cálculos como [92]:

- a) Cálculo de la Energía en un Punto (1SCF): es el valor de la HEP para un conjunto de coordenadas atómicas fijas.
- b) Optimización geométrica: consiste en la modificación sistemática de las coordenadas atómicas de un modelo dando como resultado una geometría donde las fuerzas netas en la estructura son iguales a cero.
- c) Cálculo de propiedades: es la predicción de determinadas propiedades físicas y químicas, tales como la carga, momento dipolar.

Para cada aplicación, cada método posee ventajas y desventajas que debemos conocer, dependiendo la elección del método de una serie de factores como son:

- ✓ La naturaleza de la molécula.
- ✓ El tipo de información que necesitemos.
- ✓ La disponibilidad para poder aplicar parámetros experimentales determinados, ya que algunos de los métodos así lo requieren.
- ✓ Recursos de computación.

En la Tabla 1 se muestra una visión general de las características de cada método y sus aplicaciones.

Tabla 1. Ventajas y desventajas de los diferentes métodos computacionales.

Tipo	Características	Ventajas	Desventajas	Aplicaciones
Mecánica molecular Gaussian y AutoDock	Usa la física clásica	Es rápido	Solo aplicable a un limitado número de moléculas	Grandes sistemas como enzimas
	Depende de campos de fuerza con parámetros obtenidos empíricamente	Utiliza pocos recursos de ordenador	No calcula propiedades electrónicas. Requiere de datos experimentales	Sistemas en los que no hay rompimiento ni creación de enlaces
Mecánica cuántica Semiempírico Gaussian	Usa la física cuántica	Requiere menos recurso de ordenador que el método <i>ab initio</i>	Requiere de datos experimentales	Sistemas de tamaño medio (cientos de átomos)
	Utiliza parámetros empíricos derivados experimentalmente	Es capaz de calcular estados de transición y estado excitados	Menos riguroso que el método <i>ab initio</i> en el procedimiento de cálculo	Sistemas en los que se producen transiciones electrónicas
	Usa aproximaciones extensivas			
Mecánica cuántica <i>Ab initio</i> Gaussian	Usa la física cuántica	Útil para un amplio campo de sistemas	Necesita de muchos recursos de ordenador	Sistemas Pequeños
	Riguroso matemáticamente y no utiliza parámetros experimentales	No utiliza datos experimentales		Sistemas en los que existen transiciones electrónicas
		Capaz de calcular estados de transición y estados excitados		Sistemas que requieran de una precisión rigurosa

2.8.4. Optimización geométrica

La energía y la geometría son dos de las propiedades moleculares que se obtienen directamente de un cálculo. La optimización geométrica se utiliza para localizar la

conformación más estable de un modelo, de tal manera que se encuentra un punto estacionario en la HEP.

Los puntos críticos de una HEP son:

- ✓ Mínimo global: es la energía más baja y nos indica la conformación más estable.
- ✓ Mínimo local: constituyen regiones donde un cambio en la geometría en cualquier dirección nos da una geometría de mayor energía.
- ✓ Punto de silla: Es el punto entre dos energías extremas. El punto silla se define como un punto en la HEP, en el cual hay un incremento de energía en todas las direcciones excepto una (la pendiente “primera derivada” de la superficie es cero).

En el caso del mínimo de energía global y local se logra a través de la minimización de la energía. Mientras que la determinación del punto silla nos lleva a un estado de transición.

La capacidad de una optimización geométrica para converger a un mínimo depende de la geometría de partida, de la función de energía de potencial usada, y de las condiciones impuestas para conseguir un gradiente mínimo aceptable entre pasos.

La optimización geométrica es un proceso iterativo y comienza con el cálculo de la energía en un punto único, el cual se realiza sobre la geometría de partida, después se cambian las coordenadas para el conjunto de átomos y se recalcula un nuevo punto de energía, con el fin de determinar la energía de la nueva conformación, en este caso, la primera o segunda derivada de la energía (según el método) con respecto a las coordenadas atómicas determina cuanto y en qué dirección se debe cambiar el siguiente incremento de geometría, se determina de nuevo la energía y sus derivadas y el proceso continua hasta que se alcanza la mínima energía.

2.8.5. Relación cuantitativa estructura-actividad (QSAR)

Los métodos de las relaciones cuantitativas entre la estructura química y la actividad biológica (más conocida por sus siglas en inglés QSAR: *Quantitative Structure-Activity Relationships*) fueron introducidos por Hansch y Fujita en 1960, los que se han agrupado históricamente a todas las técnicas que han intentado establecer modelos empíricos de comportamiento sobre familias de compuestos biológicamente activos [93]. El establecimiento de relaciones cuantitativas entre la estructura química (a través de sus propiedades electrónicas, estéricas o de lipofilia) y la actividad biológica permite

determinar los parámetros más influyentes en la actividad biológica así como predicción de potencia para compuestos aun desconocidos sobre la base de los ya evaluados. Es, precisamente, en este aspecto predictivo donde radica el potencial de las diversas técnicas disponibles para el establecimiento de relaciones cuantitativas estructura-actividad en su aplicación al diseño de fármacos [94].

Los métodos QSAR se basan en dos principios:

- a) la posibilidad de cuantificar diversas propiedades fisicoquímicas de un compuesto, y
- b) la posibilidad de describir matemáticamente la relación existente entre la actividad biológica que se desea optimizar y las propiedades moleculares calculadas a partir de la estructura.

Se trata así de transformar la estructura del compuesto en un grupo de datos numéricos (los descriptores) y después establecer una relación cuantitativa entre estos descriptores y la actividad biológica. En otras palabras la actividad biológica es una función de los parámetros (o descriptores) fisicoquímicos y estructurales como:

- Parámetros fisicoquímicos $\left\{ \begin{array}{l} \checkmark \text{ Estéricos} \\ \checkmark \text{ Electrónicos} \\ \checkmark \text{ Hidrofóbicas} \end{array} \right.$
- Parámetros Estructurales: Teoría de grafos.
- Parámetros Teóricos $\left\{ \begin{array}{l} \text{Orbitales Moleculares} \\ \checkmark \text{ HOMO} \\ \checkmark \text{ LUMO} \\ \checkmark \text{ Cargas atómicas} \end{array} \right.$
- Parámetros geométricos: a partir de cálculos de superficies.

QSAR es una relación matemática entre la actividad biológica de una molécula de interés, como su geometría y sus propiedades químicas [95]. Se fundamenta en el planteamiento de una ecuación matemática que permita relacionar la actividad biológica con parámetros moleculares, es decir,

Actividad biológica = f (parámetros fisicoquímicos y/o estructurales)

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n$$

donde:

Y es la variable dependiente (actividad biológica).

$X_1 \dots X_n$ son las variables Independientes o descriptores (parámetro estérico, parámetro electrónico, parámetro lipofílico, parámetro topológico, etc.).

a_0, \dots, a_n son los parámetros ajustables en la regresión lineal.

La finalidad de QSAR es construir un modelo multilineal que permita predecir si un compuesto nuevo:

- ✓ Tendrá una actividad.
- ✓ Será selectivo.
- ✓ Proveerá nueva información.
- ✓ Tiene propiedades farmacocinéticas (ADME).
- ✓ Permite mejorar el modelo matemático.

Los estudios de correlaciones entre propiedades físicas y químicas o biológicas, requieren disponer de datos fisicoquímicos o biológicos obtenidos de forma experimental, así como definir los descriptores químicos y elegir el modelo adecuado que describa de forma cuantitativa estas variables. Los datos biológicos y fisicoquímicos pueden analizarse estadísticamente de dos formas:

- ✓ En forma conjunta o separada con métodos clasificatorios de reconocimiento de tendencias.
- ✓ Usando técnicas de regresión para establecer correlaciones cuantitativas.

Unos de los métodos más populares, por ser uno de los más antiguos (1962) es el denominado de las relaciones de energía libre, método extra termodinámico o, método de Hansch. En él se establecen relaciones cuantitativas entre la potencia en una serie de fármacos estructuralmente relacionados y algunas propiedades fisicoquímicas de los sustituyentes presentes en dicha serie, como la solubilidad, el grado de ionización o determinados parámetros estéricos. Una gran parte de la aceptación de este método se basa en la facilidad de determinación de estas propiedades, en la existencia de datos tabulados para muchas de ellas y, sobre todo, por las buenas correlaciones obtenidas con el comportamiento farmacocinético de los fármacos objeto de estudio.

El método de Hansch se basa, en parte, en el análisis de relaciones lineales de energía libre de Hammett, desarrollado durante la década de los cuarenta y utilizado extensamente en química orgánica para el estudio de mecanismos de reacción. Pero en los años más recientes se tiende a emplear los métodos QSAR 3D, en los que las

propiedades tridimensionales de una molécula se consideran como conjunto, más que considerar sustituyentes o subunidades aisladas.

Los 2 métodos más usados de QSAR-3D son:

- a) CoMFA: Comparative Molecular Field Analysis, este análisis se enfoca en mapas 3D de las interacciones electrostáticas y estéricas de los ligandos en estudio que luego se correlacionan con la actividad biológica [96].
- b) CoMSIA: Comparative Molecular Similarity Index Analysis, este análisis calcula índices de similitud para un conjunto de moléculas previamente alineadas en relación a las propiedades fisicoquímicas (como lo son las interacciones de distintos tipos, puentes H, etc.). Esos índices luego se correlacionan con la actividad [97].

En general los estudios de QSAR son empíricos que pueden ser usados para evaluar nuevas propiedades químicas, siendo esta una técnica retrospectiva que depende, clásicamente, de la actividad farmacológica de compuestos con núcleos similares, sirve para optimizar un líder y no para descubrir uno nuevo. Actualmente se han realizado estudios QSAR mezclando diferentes tipos y fuentes de compuestos para llegar a modelos más generales de la actividad biológica [98, 99].

2.8.5.1. Grupos funcionales activos de un fármaco

Los grupos activos, son aquellos que modifican la reactividad de un fármaco influyendo en su actividad biológica, se clasifican en tres grupos fundamentales [100]:

- a) Elementos: halógenos (flúor, cloro, bromo, yodo), arsénico, fósforo, antimonio, azufre, etc.
- b) Grupos apolares: fenilo, bencilo, metilo, metileno, grupos alquílicos, etc.
- c) Grupos polares: hidroxilo, carbonilo, amino, imina, amida, etc.

Los grupos apolares aumentan la liposolubilidad, mientras que los polares incrementan la hidrosolubilidad, pudiendo modificar el paso de los fármacos a través de las membranas biológicas.

Efecto de ciertos grupos estructurales

- a) Grupos alquílicos: si un grupo alquílico reemplaza a un átomo de hidrogeno reactivo (HCN-RCN, ArOH-ArOR, RNH₂-R₂NH, etc.), la reactividad biológica del compuesto resultante es menor.

- b) Núcleos aromáticos y heterocíclicos: los hidrocarburos cíclicos, especialmente los bencénicos, incrementan la actividad y la toxicidad.
- c) Aminas: las aminas primarias son mas toxicas y activas que las secundarias, siéndolo a su vez éstas con respecto a las aminas terciarias.
- d) Alcoholes y fenoles: los alcoholes son menos depresores que sus correspondientes hidrocarburos. Los alcoholes monovalentes muestran mayor intensidad de acción que los dioles, y estos, a su vez, que los trioles, etc. Los alcoholes terciarios son más activos que los secundarios y estos, a su vez, más que los primarios.
- e) Aldehídos y cetonas: los aldehídos tienen mayor capacidad de reacción que las cetonas, poseyendo, en consecuencia, mayor actividad biológica.
- f) Ácidos: la introducción de grupos ácidos (carboxilo, sulfónico, fosfórico, etc.) en una molécula orgánica disminuye considerablemente su actividad biológica y su toxicidad.
- g) Halógenos: los halógenos ejercen una marcada influencia en la conducta biológica de los compuestos orgánicos, generalmente incrementan la utilidad terapéutica y las propiedades tóxicas de los compuestos activos, aunque no en la misma proporción. Los derivados fluorados parecen ser menos activos fisiológicamente que los derivados que contienen otros halógenos, debido quizá a la mayor estabilidad del enlace C-F.

3. PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN

3.1. Planteamiento del problema

El cáncer es una de las principales enfermedades que provocan la muerte de miles de personas al año a nivel mundial, debido a algunos factores como los hereditarios y los agentes carcinógenos como las sustancias químicas, virus, bacterias, radiación e incluso el propio estilo de vida a las que los seres humanos nos hemos expuesto día con día. Todos estos factores ocasionan mutaciones en el material genético que se acumulan y finalmente llevan a alterar el comportamiento celular [101].

En los últimos años se han desarrollado potentes fármacos contra el cáncer; en algunas ocasiones este puede ser curado con el uso de esta modalidad terapéutica y en otras, se consigue un efecto paliativo importante de incremento de la supervivencia [102]. A pesar de ello, la mayoría de las neoplasias muestran resistencia inicial a estos fármacos constituyendo un gran obstáculo para la curación de los pacientes. Se han identificado varios genes, proteínas y vías metabólicas implicados en este fenómeno, pero hasta el momento el gen más estudiado es el *mdr 1* y su producto, la glicoproteína P, lo cual ha inducido a la búsqueda de nuevos blancos farmacológicos susceptibles de ser alterados. Uno de estos blancos es la familia de proteínas HDAC, en donde existe evidencia de que al ser inhibida, la replicación disminuye y aumenta la diferenciación.

3.2. Justificación

El AVP es un fuerte inhibidor de las HDAC's, cuyo efecto impide que se separen los grupos acetilo de los nucleosomas, disminuyendo el crecimiento celular de las células cancerígenas. Este fármaco parece potenciar el efecto de la quimioterapia, de ahí que actualmente se encuentre en estudios clínicos. Sin embargo este fármaco es hepatotóxico (nocivo para las células del hígado), lo que ha motivado a desarrollar nuevos derivados.

Es por esta razón que el presente trabajo propone la evaluación *in silico* de una serie de 500 derivados mono- di- y tri-sustituídos del AVP con una gama amplia de grupos electroattractores y electrodonadores, a través de una sustitución aromática en el grupo carboxilo, lo que posiblemente aumentará la estabilidad de los derivados del AVP al interactuar con la HDAC8, y por consiguiente superar la capacidad inhibitoria del VPA sobre dicha proteína, debido a que la estructura de las HDAC's permite el diseño de nuevos inhibidores por modificación de pseudotipos, en los que se establecen tres zonas

estructurales susceptibles de mejora desde el punto de vista de la química farmacéutica:
a) un grupo voluminoso, con afinidad hacia la zona situada al comienzo del canal hidrofóbico, b) un espaciador y c) un grupo quelante del Zn^{2+} situado en el sitio catalítico.

Es importante mencionar que estos compuestos tienen una gran probabilidad de éxito, ya que al compararlo con los fármacos usados actualmente para el mismo fin [103], conservan el anillo aromático que al ingresar al sitio catalítico de la HDAC8 interactuará por uniones tipo π - π o π -catión, por puentes de hidrógeno, o formará un complejo con el ión Zn^{2+} .

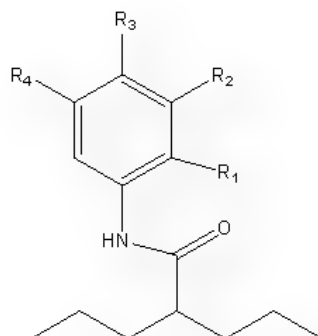


Fig. 13. Estructura base de los derivados del AVP con sustituciones R_1 (*orto*), R_2 , R_4 (*meta*) y R_3 (*para*).

4. HIPÓTESIS

El cribado virtual permitirá seleccionar los compuestos más prometedores dentro de los derivados propuestos del ácido valproico (AVP) los cuales, en principios deberán de poseer un mejor efecto antineoplásico a través de la inhibición de la histona deacetilasa 8 (HDAC8), ya que contienen grupos electroattractores y electrodonadores, de esta manera aumentan la estabilidad del AVP al interactuar con la enzima HDAC8, mientras que el estudio de la relación cuantitativa de la estructura química y la actividad biológica permitirá establecer cuáles son las propiedades fisicoquímicas que rigen la actividad enzimática de aquellos compuestos sustituidos con los grupos funcionales óptimos resultado del cribado virtual.

5. OBJETIVOS

5.1. Objetivo general

Identificar el sitio de reconocimiento de los derivados del ácido valproico con posible actividad antineoplásica sobre la HDAC8 por docking ciego y establecer parámetros fisicoquímicas que rigen la actividad enzimática por medio del método QSAR.

5.2. Objetivos particulares

- ✓ Proponer y realizar la optimización de la energía de un conjunto de 500 derivados del ácido valproico.
- ✓ Identificar el (los) sitio (s) de reconocimiento de los derivados del AVP sobre la HDAC8 mediante un docking ciego, en función de parámetros fisicoquímicos como la energía del sistema, la constante de disociación y por ende la constante de afinidad.
- ✓ Encontrar el mejor modelo matemático que describa la relación cuantitativa entre el efecto del inhibidor y las características químico-estructurales, entre los compuestos que presentan actividad biológica experimental (30 moléculas análogos al hidroxamato) con los derivados del AVP que se encuentren en el mismo sitio catalítico de inhibición, estableciendo así la actividad biológica teórica.

6. MATERIAL Y METODOLOGÍA

6.1. Instalaciones y equipo de cómputo

El cribado virtual se realizó en las instalaciones del laboratorio de modelado molecular y bioinformática de la Sección de Estudios de Posgrado e Investigación (SEPI) de la Escuela Superior de Medicina del Instituto Politécnico Nacional, bajo la dirección del Dr. José Correa Basurto. Para la realización de los cálculos de docking ciego se utilizaron los siguientes programas: PyMOL [104], Molekel [105], AutoDockTools 1.5.4 y AutoDock 4.0 [106], en 6 procesadores de 4 núcleos con plataforma LINUX, distribuciones Fedora 8, así como los programas: GaussView 4.1.2 [107] y Gaussian 03 [108] en 2 procesadores de 4 núcleos con plataforma Windows XP.

Mientras que el análisis QSAR se realizó en el Laboratorio de Simulación de la División de Estudios de Posgrado e Investigación de la Universidad del Papaloapan Campus Tuxtepec bajo la dirección del Dr. Guillermo Ramírez Galicia. Para realizar los cálculos de mecánica cuántica se utilizaron los siguientes programas: GaussView 4.1.2, Gaussian 03, además de los programas estadísticos, MATLAB 7.2.0.232 (R2006a) [109], DRAGON 2.1 [110], Origin 7 [111] y Babel 16 [112], en 2 procesadores de 4 núcleos, con plataforma Windows XP y un Cluster de 2 nodos más 1 nodo maestro, cada uno con 16 procesadores, en plataforma Linux.

6.2. Metodología

La evaluación *in silico* –Docking y QSAR- de derivados del ácido valproico sobre la HDAC8, se llevó a cabo a partir de proponer 500 diseños moleculares, con sustituyentes electroattractores y electrodonadores (Tabla 1H del anexo), con el fin de incrementar la afinidad y, de manera indirecta, su capacidad inhibitoria sobre la HDAC8.

Los compuestos propuestos se diseñaron de tal manera que contengan los requerimientos establecidos en la literatura para este tipo de inhibidores: A) un grupo quelante del ion Zn^{2+} , B) un grupo espaciador y C) una cola lipófila (Fig. 14). Específicamente, se investigó la influencia en la actividad biológica tras los cambios en el grupo espaciador, debido a la introducción de grupos electroattractores y electrodonadores (Tabla 2), que funcionarán como grupo quelante del ion Zn^{2+} .

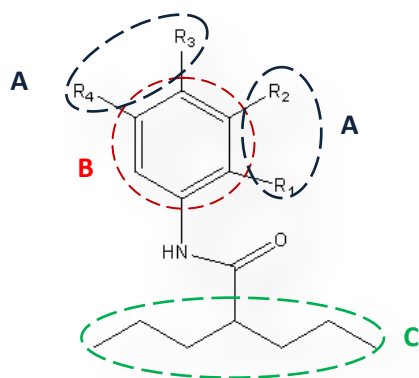


Fig. 14. Estructura base de los DAVP: A) grupo quelante del ion Zn^{2+} , B) grupo espaciador y C) cola lipófila.

Tabla 2. Grupos electroattractores y electrodonadores.

Electroattractores	Electrodonadores
-NO ₂	-OCH ₃
-CN	-OH
-COO ⁻	-N(CH ₃) ₂
-CHO	-NH ₂
-COOCH ₃	-CH ₃
-SO ₃ H	-CH ₂ CH ₃
-COCH ₃	-CH ₂ CH ₂ CH ₃
-CON(CH ₃) ₂	-CH(CH ₃) ₂
-COCH=CH ₂	-CH ₂ CH(CH ₃) ₂
-COCl	-C(CH ₃) ₃
-COOCOCH ₃	
-F, -Cl, -I, Br	

6.3. Docking

En la metodología del análisis de docking ciego, el algoritmo toma en cuenta una serie de descriptores de acoplamiento geométrico, donde la superficie de la macromolécula se aproxima como una serie de segmentos de superficie accesible al disolvente, incluyendo la capacidad de donar y aceptar puentes de hidrógeno y potenciales en los ángulos de torsión. Por otra parte, el ligando se mueve a través de estos espacios accesibles y escoge un lugar de forma aleatoria. En ese punto se realiza una serie de movimientos conformacionales del ligando, lo que induce un costo energético total del sistema. Una vez finalizado este proceso se escoge un nuevo lugar geométrico y se repite el proceso. Durante esta búsqueda de sitios accesibles, todas las

interacciones obtienen un puntaje, en especial a las interacciones débiles que estabilizan al complejo, disminuyendo la energía total [113].

6.3.1. Minimización de la energía de los derivados del AVP

Los 500 derivados del AVP se crearon en 3D con el programa GaussView 4.1.2, a los archivos obtenidos se les efectuó una optimización geométrica con el método semiempírico AM1 incluido el programa Gaussian 03. Posteriormente con el programa Molekel se realizó la transformación de coordenadas Gaussian 03 *.out a coordenadas *.PDB, que son necesarios para el docking ciego.

6.3.2. Estructura tridimensional

Se realizó la búsqueda de la estructura tridimensional de la HDAC8 en la base de datos Protein Data Bank (PDB) obtenida experimentalmente a través de cristalización o difracción de rayos X, encontrándose la estructura con clave PDB: 3f07 [114]. En la preparación del archivo de entrada: se eliminaron las moléculas de agua y de los ligandos ajenos a la HDAC8 con el programa PyMOL.

6.3.3. Docking ciego

El procedimiento general consistió en preparar archivos de entrada para el programa AutoDock a partir de las estructuras optimizadas de los compuestos y la estructura de la macromolécula (HDAC8).

La estructura de la HDAC8 se visualizó con el programa AutoDockTools versión 4.1.2. Fue necesario adicionar los átomos de hidrógenos polares, es decir, aquellos unidos a átomos electronegativos que pueden participar en interacciones por puente de hidrógeno; se agregaron las cargas parciales de Kollman [106] con la finalidad de evaluar la contribución electrostática, así como los parámetros de solvatación. Al igual que la HDAC8 los derivados del AVP se visualizaron con el mismo programa, obteniendo los archivos *.pdbqt; posteriormente se delimitó a la HDAC8 en el espacio de búsqueda, el cual está representado por una caja de 116 x 96 x 112 Å que incluye completamente a la HDAC8, esta metodología genera el archivo *.gpf. A continuación se utilizó el programa autogrid 4 [106] para establecer las interacciones. Los mapas de afinidad atómica se generaron para cada archivo *.pdbqt; con el fin de determinar la afinidad de los

inhibidores (derivados del AVP) por la enzima. El siguiente paso fue crear el archivo *.dpf, para lograr la ejecución del programa autodock 4 [106] para esto, se utilizaron los siguientes parámetros: 100 conformaciones, con 10, 000,000 de interacciones.

En AutoDock el GRID es la energía de interacción debida a factores estéricos, enlaces de hidrógeno y electrostáticos.

$$E_{xyz} = \sum E_{lj} + \sum E_{el} + \sum E_{hb}$$

donde

E_{xyz} Es la energía de interacción de enlace en las coordenadas x, y, z.

E_{lj} Es la energía de van der Waals.

E_{el} Es la energía electrostática.

E_{hb} Es la energía de enlaces hidrógeno.

6.3.4. Probabilidad

En esta parte se toma en cuenta la probabilidad de un microestado en el colectivo canónico. En el colectivo canónico, el estado macroscópico queda fijo al conocer N (número de moles), V (volumen) y T (temperatura), por lo que la energía interna se obtendrá como función de estas tres variables [U(N, V, T)]. Sin embargo, la energía de los microestados es función únicamente de N y V [E(N, V)]. Pensemos en un conjunto de N partículas independientes de masa m en una caja monodimensional de longitud a, la energía será:

$$E = \sum_{i=1}^N \frac{h^2}{8ma^2} n_{xj}^2$$

De acuerdo con la siguiente igualdad las probabilidades de ocupación de los microestados deben de ser función de la temperatura:

$$U(N, V, T) = \sum_j p_j E_j(N, V)$$

Por otra parte parece lógico suponer que la relación entre el número de sistemas del colectivo que se encuentren en el microestado j y en el microestado i debe depender de las energías de ambos E_i, E_j . Podríamos escribir por tanto la siguiente relación:

$$\frac{p_j}{p_i} = f(E_i, E_j)$$

¿Cómo deben de combinarse las energías en esta función? La energía de un microestado depende de la elección arbitraria del cero de energías potenciales, sin embargo la relación entre las probabilidades de dos microestados no puede depender de esta elección. En consecuencia, la relación de probabilidades debe ser función de la diferencia de energías:

$$\frac{p_j}{p_i} = f(E_i, E_j) = f(E_i - E_j)$$

Si consideramos ahora tres microestados distintos (i, j, k), podremos escribir lo siguiente:

$$\begin{aligned} \text{-----}E_i & & \frac{p_k}{p_j} &= f(E_j - E_k) \\ \text{-----}E_j & & & \\ \text{-----}E_k & & \frac{p_j}{p_i} &= f(E_i - E_j) \\ & & \frac{p_k}{p_i} &= f(E_i - E_k) \end{aligned}$$

Ahora bien, también se debe cumplir la relación matemática:

$$\frac{p_k}{p_i} = \frac{p_j}{p_i} \frac{p_k}{p_j}$$

y de acuerdo con las relaciones anteriores queda:

$$f(E_i - E_k) = f(E_i - E_j)f(E_j - E_k)$$

Si nos fijamos en esta ecuación, el término que aparece dentro de la primera función es la suma de lo que aparece dentro de las otras dos:

$$(E_i - E_k) = (E_i - E_j) + (E_j - E_k)$$

La única función matemática que cumple esta relación (el producto de dos funciones es igual a la función de la suma) es la exponencial e^x . $e^y = e^{x+y}$. Por lo tanto la relación entre las probabilidades de dos microestados viene dada por:

$$\frac{p_j}{p_i} = e^{\beta(E_i - E_j)}$$

lo que implica que las probabilidades vienen dadas por:

$$p_i = C e^{-\beta E_i}$$

Para determinar completamente las probabilidades necesitamos conocer C, la cual se puede determinar por la condición de normalización. Efectivamente la suma de las probabilidades de todos los microestados debe ser la unidad:

$$1 = \sum_j p_i = C \sum_j e^{-\beta E_i}$$

por lo tanto C es:

$$C = \frac{1}{\sum_j e^{-\beta E_i}}$$

Quedando entonces la probabilidad de ocupación de un determinado microestado como:

$$p_i = \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_i}}$$

Al denominador de esta expresión se le conoce como función de partición canónica (Q) que es función de β y de las energías de los microestados, donde β es $\frac{1}{KT}$.

La relación entre las probabilidades de ocupación de dos microestados (k y j) de diferente energía es:

$$\frac{p_k}{p_l} = \frac{e^{-\beta E_k}/Q}{e^{-\beta E_l}/Q} = \frac{e^{-\beta E_k}}{e^{-\beta E_l}} = e^{-\left(\frac{E_k}{KT} - \frac{E_l}{KT}\right)} = e^{-\frac{\Delta E}{KT}}$$

Sabiendo lo anterior se procedió a calcular la probabilidad de ocupación p_i de los resultados arrojados por el docking ciego mediante la siguiente fórmula:

$$p_i = \frac{e^{-\frac{\Delta E_i}{KT}}}{\sum_j e^{-\frac{\Delta E_j}{KT}}}$$

o

$$p_i = \frac{e^{-\frac{\Delta E_i}{RT}}}{\sum_j e^{-\frac{\Delta E_j}{RT}}}$$

donde la T es la temperatura a 298 K, ΔE_i es la diferencia de energía libre de Gibbs por cada molécula de los derivados del AVP, K es la constante de Boltzmann $K_B = \frac{R}{N_A} = 1.3806504 \times 10^{-23} \frac{J}{K}$, R es la constante de los gases ideales 1.98 cal/mol. K y N_A es el número de avogadro $6.02214179 \times 10^{23} mol^{-1}$. Para este caso en particular la probabilidad propuesta en el presente trabajo de tesis es $p_1 = 0.7$ y $p_2 = 0.3$, esto con el

fin de descartar sitios catalíticos con baja probabilidad de éxito, por lo que el intervalo de la energía libre de Gibbs de cada uno de los derivados del AVP debe estar entre 0 y 0.49993963 kcal/mol ($\Delta E = RT \ln \frac{p_1}{p_2}$).

Todo esto se realizó con la finalidad de establecer una relación entre la probabilidad y la energía libre de Gibbs y así poder observar solo aquellos sitios catalíticos con menor energía (mayor estabilidad en el complejo ligando-enzima).

Se realizó la minimización de la energía, el docking ciego y la probabilidad de ocupación p_i de los 38 análogos del hidroxamato (Tabla 2H del anexo), con el objetivo de establecer la relación cuantitativa estructura-actividad biológica con los derivados del AVP, los cuales poseen características similares con los análogos del hidroxamato en cuanto a su estructura química y actividad inhibitoria sobre la HDAC8.

6.3.5. Análisis de los sitios catalíticos de los derivados del AVP y los análogos del hidroxamato

Se analizaron solo aquellos sitios catalíticos con menor energía libre de Gibbs en función de la probabilidad de ocupación p_i de los resultados arrojados por el docking ciego.

6.4. QSAR

La metodología de investigación en los estudios cuantitativos de la relación estructura actividad biológica depende de las características de los compuestos que se desean estudiar, comúnmente para obtener un modelo predictivo se requieren tres conjuntos de datos: el conjunto de *entrenamiento*, *validación* (en ambos es necesario conocer el valor experimental de la propiedad que se desea estudiar) y *predicción*. El punto inicial es la elección de un conjunto de datos de entrenamiento seleccionado con base a la diversidad estructural de los compuestos, el cual es utilizado para generar el modelo. Sobre él se realizan todas las pruebas necesarias para elegir los descriptores que mejor ajusten al valor experimental de la propiedad, utilizando parámetros estadísticos como: coeficiente de correlación lineal (R), desviación estándar (SD), F-estadístico total (F), el criterio de información de Akaike (AIC), la función de aptitud de Kubinyi (FIT), coeficiente de correlación por validación cruzada (q^2_{LOO} y q^2_{LGO}), entre otras. Luego se

emplean dichos descriptores para calcular el valor de la propiedad de las moléculas que conforman el conjunto de validación (grupo de compuestos que se reserva para la comprobación del modelo). Si el valor calculado se aproxima en buena medida al experimental para todas las moléculas del conjunto de validación o por lo menos, la mayoría, significa que el modelo tiene poder predictivo, y, en principio, se puede usar para predecir valores de la propiedad de un conjunto nuevo (conjunto de predicción) [115-117].

Cabe mencionar que en el presente trabajo solo se toman en cuenta dos conjuntos de datos: el conjunto de entrenamiento (análogos del hidroxamato) y el conjunto de predicción (derivados del AVP), debido a que las estructuras químicas en ambos conjuntos son diferentes.

Para este estudio se siguió la siguiente metodología:

6.4.1. Elección de un conjunto de compuestos con actividad biológica conocida

Se eligió un conjunto de compuestos con actividad conocida (conjunto de entrenamiento) de los análogos del hidroxamato, debido a que estos poseen actividad biológica experimental y gran similitud estructural con los derivados del AVP, lo que es indispensable para estudio de la relación cuantitativa estructura-actividad biológica.

6.4.2. Identificación de los descriptores moleculares de los análogos del hidroxamato

Primero se realizó el cambio de coordenadas *.docked.pdbq obtenidos por docking ciego a coordenadas *.gjf con el programa Babel 16, con el objetivo de adicionar los átomos de hidrógeno con el programa GaussView 4.1.2. A continuación se efectuó un cálculo único con el método semiempírico AM1 usando el programa Gaussian 03, conservando de esta manera las características geométricas obtenidas del docking ciego, posteriormente con el programa Babel 16 se transformaron las coordenadas *.out de los archivos de salida de los cálculos únicos a coordenadas *.mol, este cambio de coordenadas se realizó para que el programa DRAGON pudiera determinar los descriptores moleculares que mejor definan la estructura de los compuestos elegidos. Además de los descriptores obtenidos por DRAGON se adicionaron otros descriptores,

como la actividad biológica experimental (IC_{50}), los descriptores teóricos obtenidos del cálculo único: energía mínima, energía del orbital molecular más bajo desocupado (LUMO), energía del orbital molecular más alto ocupado (HOMO), dureza, blandura, dipolo X, Y, Z y total, así como los descriptores obtenidos por el docking ciego: energía libre de Gibbs, intermolecular, interna y la constante de disociación.

6.4.3. Selección de los descriptores de los análogos del hidroxamato

Con el programa MATLAB se seleccionaron los descriptores con un coeficiente de correlación de colinealidad de 0.8 en relación a la actividad biológica.

6.4.4. Revisión de colinealidad de los análogos del hidroxamato

Con ayuda del programa SPSS se realizó una regresión lineal múltiple, con los descriptores seleccionados anteriormente, estos fueron establecidos como variables independientes, mientras que el descriptor energía libre de Gibbs fue establecido como la variable dependiente, estableciendo así el conjunto de entrenamiento con las que se puedan obtener una serie de modelos matemáticos que muestren el comportamiento de los derivados de AVP en función de ciertos descriptores de los análogos del hidroxamato, además se realizó una regresión lineal en SPSS estableciendo como variables independientes los descriptores obtenidos en DRAGON, y como variable dependiente la actividad biológica, con el único fin de encontrar aquellos descriptores que tengan una relación directa con el IC_{50} .

6.4.5. Identificación de los descriptores moleculares de los derivados del AVP

Se realizó el mismo procedimiento que en la identificación de los descriptores moleculares de los análogos del hidroxamato, con la única diferencia de que en este caso el descriptor de la actividad biológica no se contempla, debido a que estos valores son los que se desean encontrar de manera teórica. Estableciendo así el conjunto de predicción.

6.4.6. Regresión Lineal Múltiple (RLM)

Se realizaron las RLM de los modelos obtenidos en la revisión de colinealidad de los análogos del hidroxamato, junto con los modelos del conjunto de predicción, los

cuales fueron obtenidos de los datos numéricos de los descriptores seleccionados en la regresión lineal con programa SPSS de los análogos del hidroxamato. Las gráficas de las regresiones lineales múltiples se trazaron con el programa Origin 7.

La selección del mejor modelo para la RLM se llevó a cabo de acuerdo al análisis de los siguientes parámetros estadísticos: R, SD, F, AIC, FIT, q^2_{LOO} , q^2_{LGO} , entre otras.

6.4.7. Expresión matemáticamente de la actividad biológica de los derivados del AVP

En el programa SPSS se realizó una regresión lineal con los descriptores que resultaron exitosos del mejor modelo de la RLM, donde el descriptor Belm4 se consideró como la variable dependiente, mientras que los demás descriptores fueron considerados como variables independientes, encontrando modelos matemáticos que relacionan la actividad biológica-descriptores. Se trazaron estos modelos en Origin 7 (modelos del conjunto de entrenamiento y predicción), y se seleccionó el mejor modelo de la RLM de acuerdo a los parámetros estadísticos establecidos anteriormente.

6.4.8. Redes Neuronales Artificiales (ANN)

De acuerdo a la literatura las Redes Neuronales Artificiales, ANN (*Artificial Neuron Networks*) son mecanismos de procesamiento de la información inspirados en las redes de neuronas biológicas. Su funcionamiento, explicado en numerosos textos entre ellos [118-120], se basa en operaciones sencillas realizadas en paralelo por un gran número de células elementales, las neuronas. En las ANN, la unidad análoga a la neurona biológica es el elemento procesador, PE (*process element*). Un elemento procesador tiene varias entradas y las combina (Fig. 15) normalmente con una suma básica. La suma de las entradas es modificada por una función de transferencia y el valor de la salida de esta función de transferencia se pasa directamente a la salida del elemento procesador, por lo que los modelos de ANN son dirigidos a partir de los datos, es decir, son capaces de encontrar relaciones (patrones) de forma inductiva por medio de los algoritmos de aprendizaje basado en los datos.

Las características que hacen más interesantes a las ANN para el modelado de sistemas son:

- ✓ Son modelos no lineales, puesto que la función de activación es no lineal.

- ✓ Son modelos paramétricos y los parámetros corresponden a los pesos de las conexiones entre neuronas.
- ✓ Son modelos adaptivos, puesto que la aparición de nuevos datos permite el reaprendizaje de los parámetros adaptando los valores anteriores a los datos actuales.
- ✓ Son tolerantes a fallos, ya que su comportamiento está distribuido entre todos los parámetros.

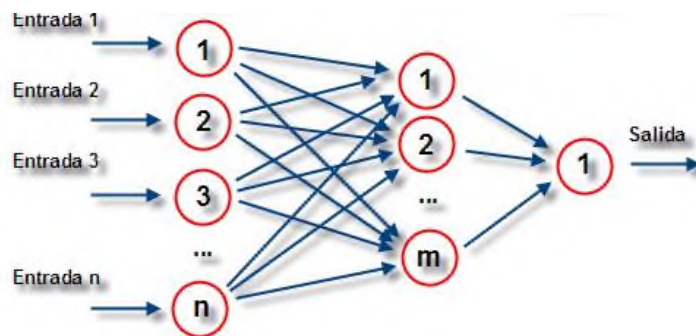


Fig. 15. Estructura de las Redes Neuronales Artificiales (ANN).

El motivo que, probablemente, ha dado más impulso a la utilización de las redes neuronales en el campo de la identificación y control de procesos, es su demostrada capacidad de aproximar funciones con un grado arbitrario de precisión. Es por esta razón que se procedió a la aplicación de las ANN al mejor modelo obtenido de la RLM, utilizando 3 células elementales, es decir, 3 neuronas por cada entrada.

6.4.9. Predicción de la actividad biológica teórica de los derivados del AVP

Utilizando la ecuación matemática donde se relaciona la actividad biológica de los análogos del hidroxamato con los descriptores moleculares, se determinó la actividad teórica de los derivados del AVP.

6.4.10. Modelo matemático de la actividad biológica teórica

Se representó gráficamente en Origin 7 el modelo matemático de la actividad biológica teórica de los derivados del AVP en relación a la actividad biológica experimental de los análogos del hidroxamato.

7. RESULTADOS Y DISCUSIÓN

7.1. Docking

7.1.1. Minimización de la energía

Se optimizaron geoméricamente las 500 moléculas de los derivados del AVP y los 38 análogos del hidroxamato con el método semiempírico AM1, obteniendo dos de las propiedades moleculares más importantes: la energía mínima y la geometría molecular más estable de cada una de las estructuras. Estos resultados se obtuvieron un tiempo de cálculo de aproximadamente 1 minuto, siendo esta una enorme ventaja de los métodos semiempíricos.

En la Fig. 16 se muestra la conformación más estable de 4 estructuras de los derivados del AVP con su respectiva mínima energía. De acuerdo a estas propiedades moleculares, estas estructuras se encuentran en un punto estacionario en la HEP, correspondiente al mínimo global, donde la energía es la más baja y la conformación es la más estable.

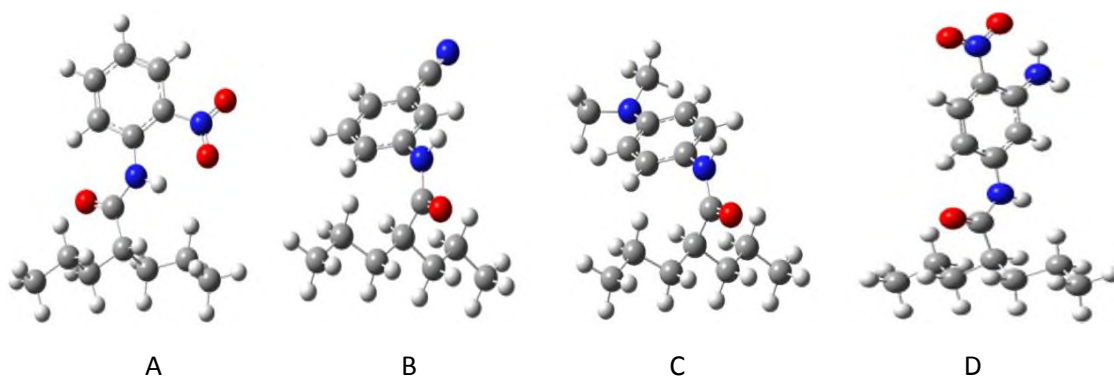
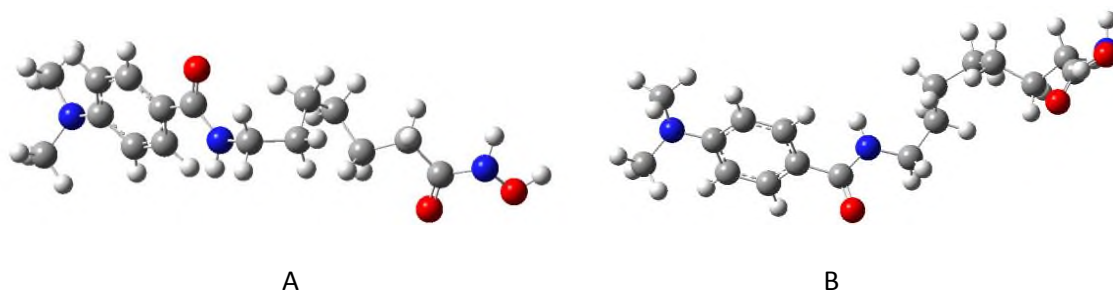


Fig. 16. Derivados del AVP. A) DAVP001 ($E=-0.0787191$ hartres, $R_1=NO_2$), B) DAVP005 ($E=-0.0283518$ hartres, $R_2=CN$), C) DAVP045 ($E=-0.0656739$ hartres, $R_3=N(CH_3)_2$), D) DAVP082 ($E=-0.0852064$ hartres, $R_2=NH_2$, $R_3=NO_2$).

Mientras que en la Fig. 17 se muestran 4 estructuras de los análogos del hidroxamato, en su mínimo global.



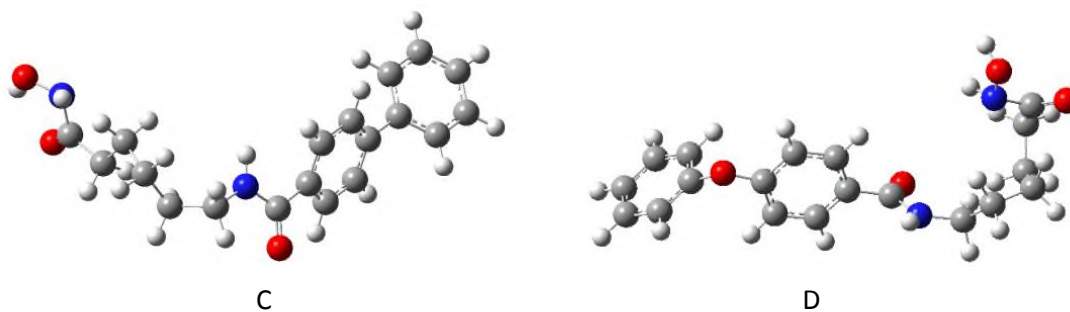


Fig. 17. Derivados del AVP. A) HID02 ($E=-0.1307932$ hartres), B) HID03 ($E=-0.1419026$ hartres), C) HID40 ($E=-0.0925125$ hartres), D) HID05 ($E=-0.1310265$ hartres).

El mínimo global se determina matemáticamente con el cambio consecutivo de las coordenadas del conjunto de átomos de cada estructura y finaliza hasta que alcanza la energía mínima.

Analizando ambos grupos de compuestos la energía de los derivados del AVP son menores, lo que indica que las estructuras están en su conformación más estable, aunque esta energía depende en sí al tipo de estructura que se esté evaluando, afortunadamente ambos tipos de compuestos son muy similares en su estructura química.

7.1.2. Estructura tridimensional de la HDAC8

Al realizar la búsqueda de la HDAC8 en Protein Data Bank, se encontró su estructura 3D determinada por rayos X con las siguientes características: en un complejo tricostatina A y 3-(1-metil-4-fenilacetil-1H-2-pirrolil)-N-hidroxi-2-propenamida (APHA) con resolución de 3.30 Å y código PDB: 3F07 [114], dependiente del Zn^{2+} el cual tiene la función de regular la acetilación de residuos de lisina en las histonas y otras proteínas en células eucariotas.

	3F07 Crystal Structure Analysis of Human HDAC8 complexed with APHA in a new monoclinic crystal form
	Authors: Dowling, D.P., Gantt, S.L., Gattis, S.G., Fierke, C.A., Christianson, D.W.
	Release: 2008-12-30 Classification: Hydrolase
	Experiment: X-RAY DIFFRACTION with resolution of 3.30 Å
	Compound: 1 Polymer [Display Full Polymer Details Display for All Results] 3 Ligands [Display Full Ligand Details Display for All Results]
	Citation: Structural studies of human histone deacetylase 8 and its site-specific variants complexed with substrate and inhibitors. (2008) <i>Biochemistry</i> 47 : 13554-13563 [Display Full Abstract Display for All Results]

Fig. 18. Información general de la enzima HDAC8 encontrada en PDB: 3F07.

En la preparación del archivo de entrada de la HDAC8 con el programa PyMOL se observaron las siguientes situaciones: la presencia de 6 iones K^{2+} , 1 grupo acetilo y 2

ligandos, todos ajenos a la HDAC8 (círculos rojos en la Fig. 19). Además se observó que la HDAC8 obtenida de PDB (código: 3F07) estaba compuesta de tres partes estructurales casi idénticas, dando a entender que la HDAC8 encontrada es un trímero.

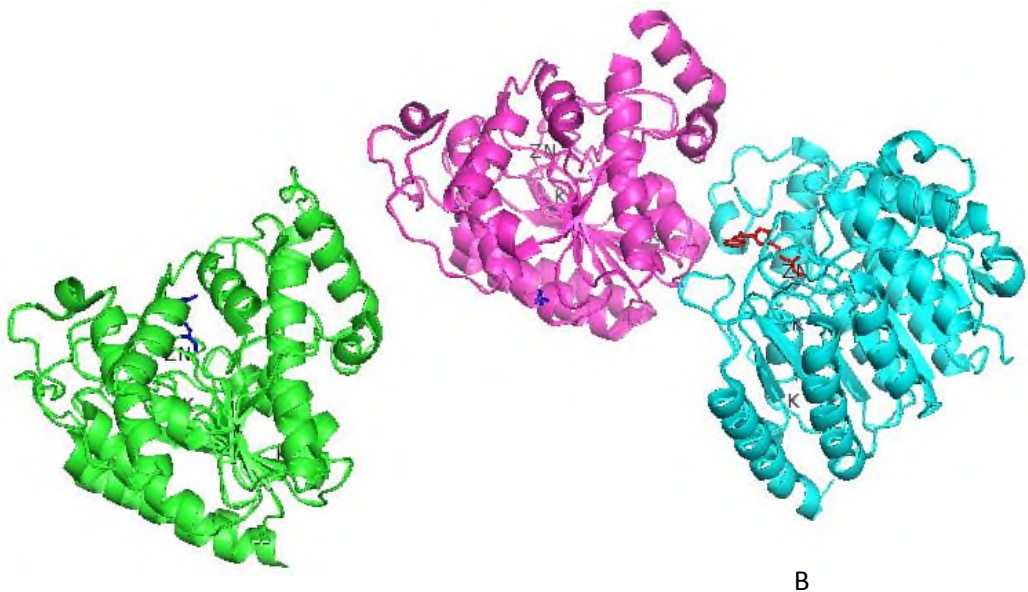
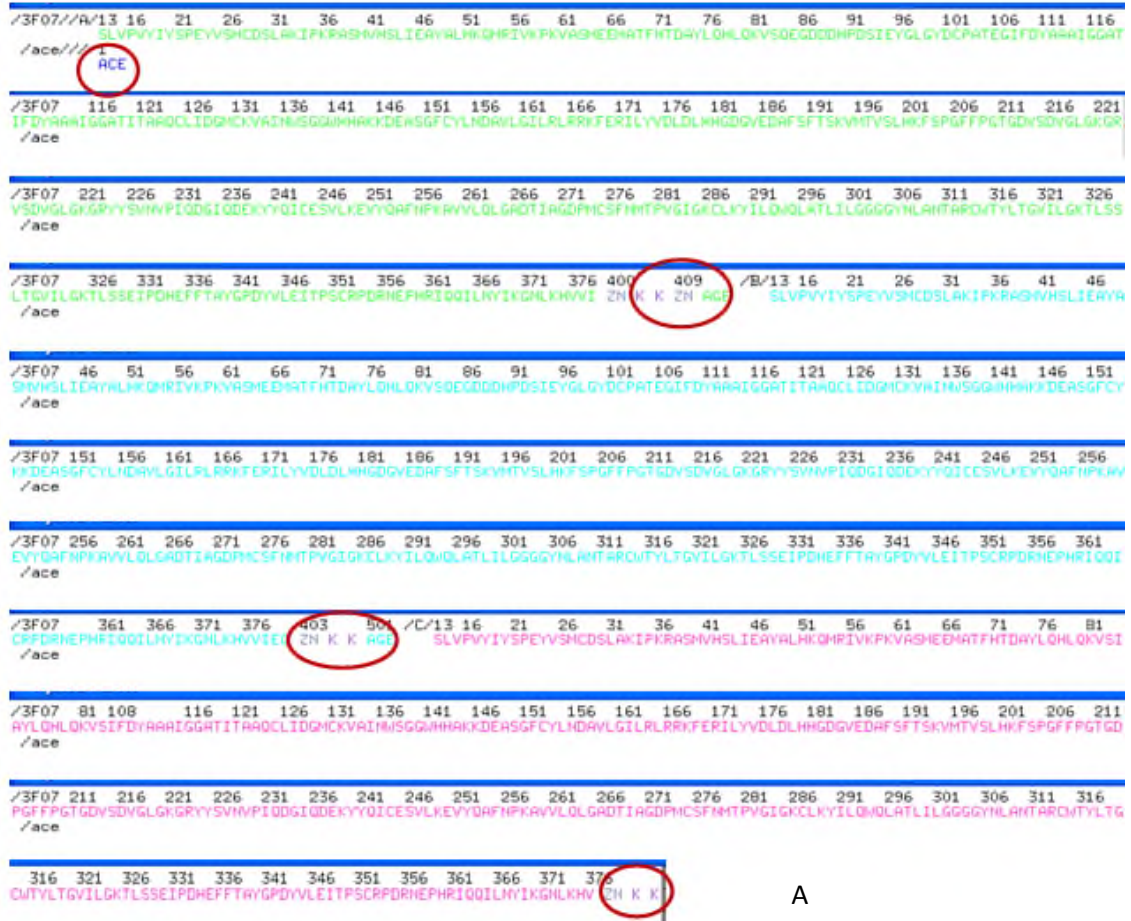


Fig. 19. Enzima HDAC8. Visualización de la enzima HDAC8 con el programa PYMOL. A) Secuencia de aminoácidos de la HDAC8, B) Estructura cuaternaria de la HDAC8.

Realizada la eliminación de estos grupos ajenos (6 iones K^{2+} , 1 grupo acetilo y 2 ligandos) a la HDAC8, así como la eliminación de 2 monómeros, se obtuvo la estructura adecuada para el presente estudio:

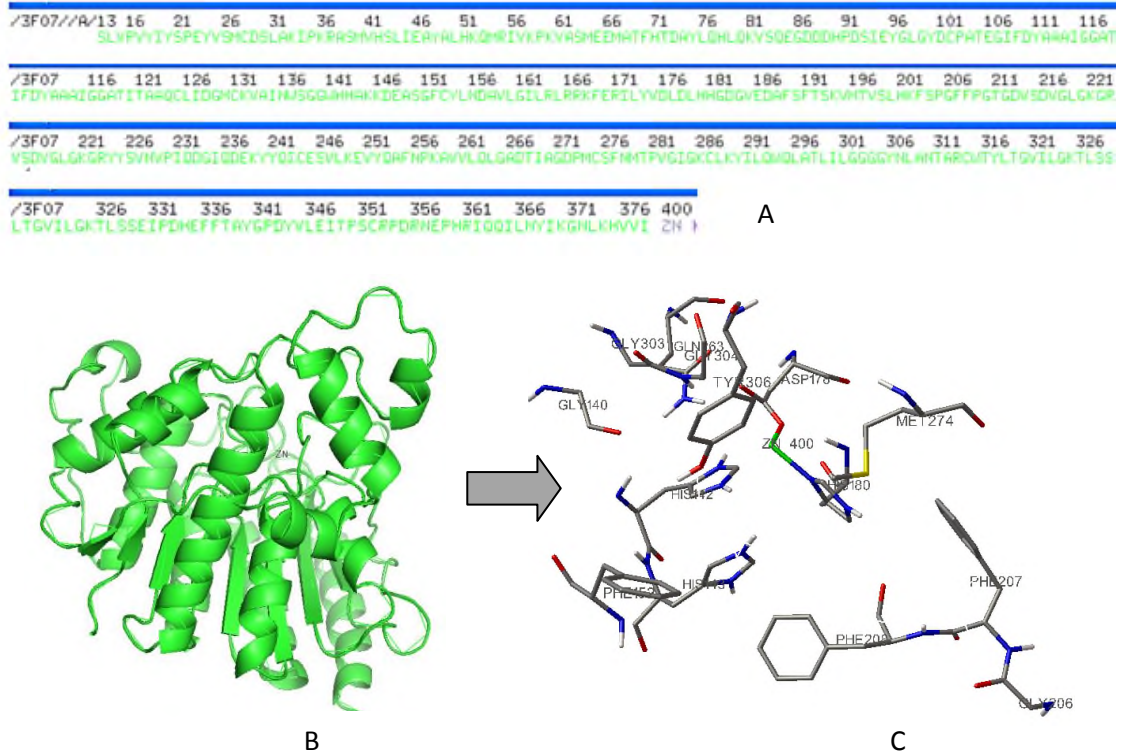


Fig. 20. Enzima HDAC8.libre de grupos ajenos. A) Secuencia de aminoácidos de la HDAC8, B) Estructura terciaria de la HDAC8 con la presencia del ion Zn^{2+} , C) Sitio catalítico de la HDAC8.

En la Fig. 20A la HDAC8 se encuentra libre de grupos y ligandos ajenos, manteniendo solo una interacción con el ion Zn^{2+} , ya que este permite formar complejos con los grupos electroattractores y electrodonadores de los derivados del AVP, así como de los análogos del hidroxamato. La permanencia del ión Zn^{2+} en la proteína tiene como finalidad regular la acetilación de residuos de lisina en las histonas, el ion Zn^{2+} forma parte del sitio catalítico, este sitio catalítico como se ve en la Fig. 20B está altamente conservado por residuos de aminoácidos aromáticos (Tyr306, His142, His143, Phe152 y Phe207), lo que concuerda con la literatura.

7.1.3. Docking ciego

Mediante la metodología explicada en la sección anterior se obtuvieron los resultados de los 500 derivados del AVP y los 38 análogos del hidroxamato por docking

ciego. Estos resultados fueron analizados exhaustivamente mediante la aplicación de la probabilidad de ocupación p_i , obteniendo solo aquellos sitios catalíticos con menor energía libre de Gibbs. De los 500 derivados del AVP solo 128 moléculas presentaron mayor estabilidad ligando-enzima, mientras que de los 38 análogos del hidroxamato solo 30 moléculas presentaron dicha estabilidad. Al ir analizando estos sitios catalíticos con el programa AutoDock se logró observar cómo el ligando se movía a través de los espacios accesibles de la HDAC8, escogiendo un lugar de forma aleatoria, con una serie de movimientos adicionales, lo que indujo al final un costo energético total del sistema (ver Fig. 21). Cabe mencionar que el espacio escogido por los ligandos en la HDAC8 fue el mismo para todos, en otras palabras todos los ligandos tanto los derivados del AVP como los análogos del hidroxamato se encontraban en el mismo sitio catalítico de la HDAC8, satisfactoriamente este sitio catalítico es el que se encuentra reportado en la literatura para este tipo de enzimas.

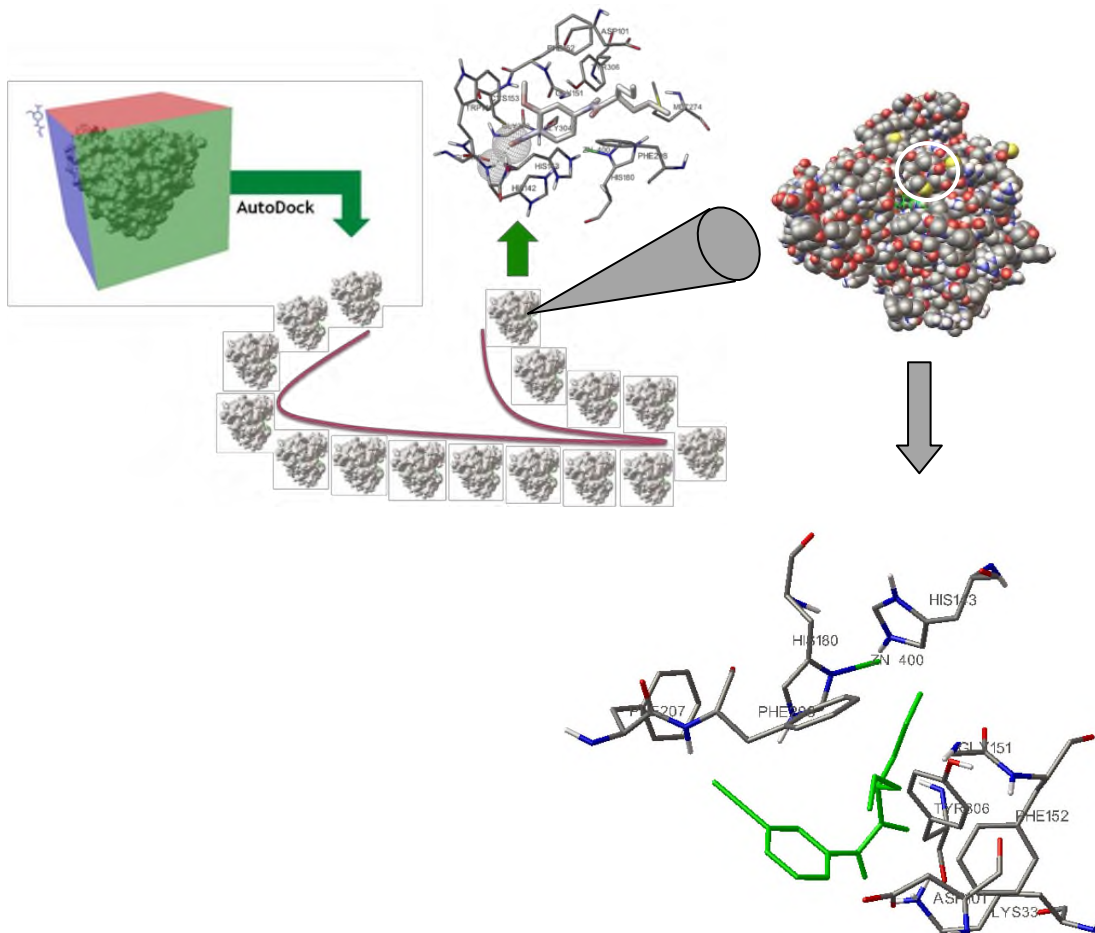


Fig. 21. Docking de la molécula DAVP005 sobre la HDAC8 (aminoácidos encontrados en el sitio catalítico: Phe208, His180, Zn²⁺, His143, Gly151, Asp101, Phe152, Tyr306, Lys33 y Phe207 con R₂=CN).

Mediante la simulación de docking ciego, se obtuvieron las energías de acoplamiento, intermolecular, interna y torsional así como las constantes de disociación de los 128 derivados del AVP y de los 30 análogos del hidroxamato contra la HDAC8. En la Tabla 3 y la Tabla 4 se resumen los mejores compuestos de acuerdo a los parámetros fisicoquímicos descritos anteriormente. (En la Tabla 3H y 4H del anexo se muestran los resultados totales).

Tabla 3. Energías y constantes de disociación de los mejores análogos del hidroxamato.

Molécula	ΔG	K_d	E. Intermolecular	E. Interna	E. Torsional
HID0 6	-7.47	3.34	-10.16	-0.22	2.68
HID 15	-7.19	5.33	-10.18	-0.76	2.98
HID 23	-8.1	1.15	-10.49	-1.2	2.39
HID 25	-7.54	2.97	-9.03	-0.52	1.49
HID 26	-8.01	1.35	-10.1	-0.59	2.09
HID 29	-7.39	3.8	-10.97	-0.81	3.58
HID 34	-7.65	2.49	-10.33	-0.93	2.68
HID 35	-7.37	3.93	-9.46	-0.36	2.09
HID 37	-7.91	1.59	-10.89	-0.62	2.98

[Los datos en rojo indican los mejores compuestos] [Las energías están dadas en Kcal/mol y la K_d en μM]

Tabla 4. Energías y constantes de disociación de los mejores derivados del AVP.

Molécula	ΔG	K_d	E. Intermolecular	E. Interna	E. Torsional
DAVP 266	-7.02	7.2	-9.1	-0.49	2.09
DAVP 269	-7.52	3.08	-9.61	-0.38	2.09
DAVP 271	-6.97	7.84	-8.76	-0.57	1.79
DAVP 277	-6.77	10.86	-8.86	-0.6	2.09
DAVP 281	-7.04	6.94	-9.13	-0.63	2.09
DAVP 282	-6.85	9.49	-8.94	-0.28	2.09
DAVP 283	-6.68	12.65	-9.37	-0.53	2.68
DAVP 389	-6.18	29.28	-7.48	-1.09	2.39
DAVP 405	-6.01	39.37	-8.4	-0.88	2.39
DAVP 458	-6.51	16.83	-8.6	-0.63	2.09

[Los datos en rojo indican los mejores compuestos] [Las energías están dadas en Kcal/mol y la K_d en μM]

Como se mencionó anteriormente la constante de disociación está relacionada con la afinidad de los ligandos, ya que cuanto menor sea K_d menor es la concentración de ligando necesaria para saturar la proteína y, en consecuencia, mayor es la afinidad de la proteína hacia el ligando, de acuerdo a esto los compuestos que se encuentran marcados

de color rojo en las tablas 3 Y 4 son los mejores valores, hablando solo con respecto a la energía libre de Gibbs y la constante de disociación.

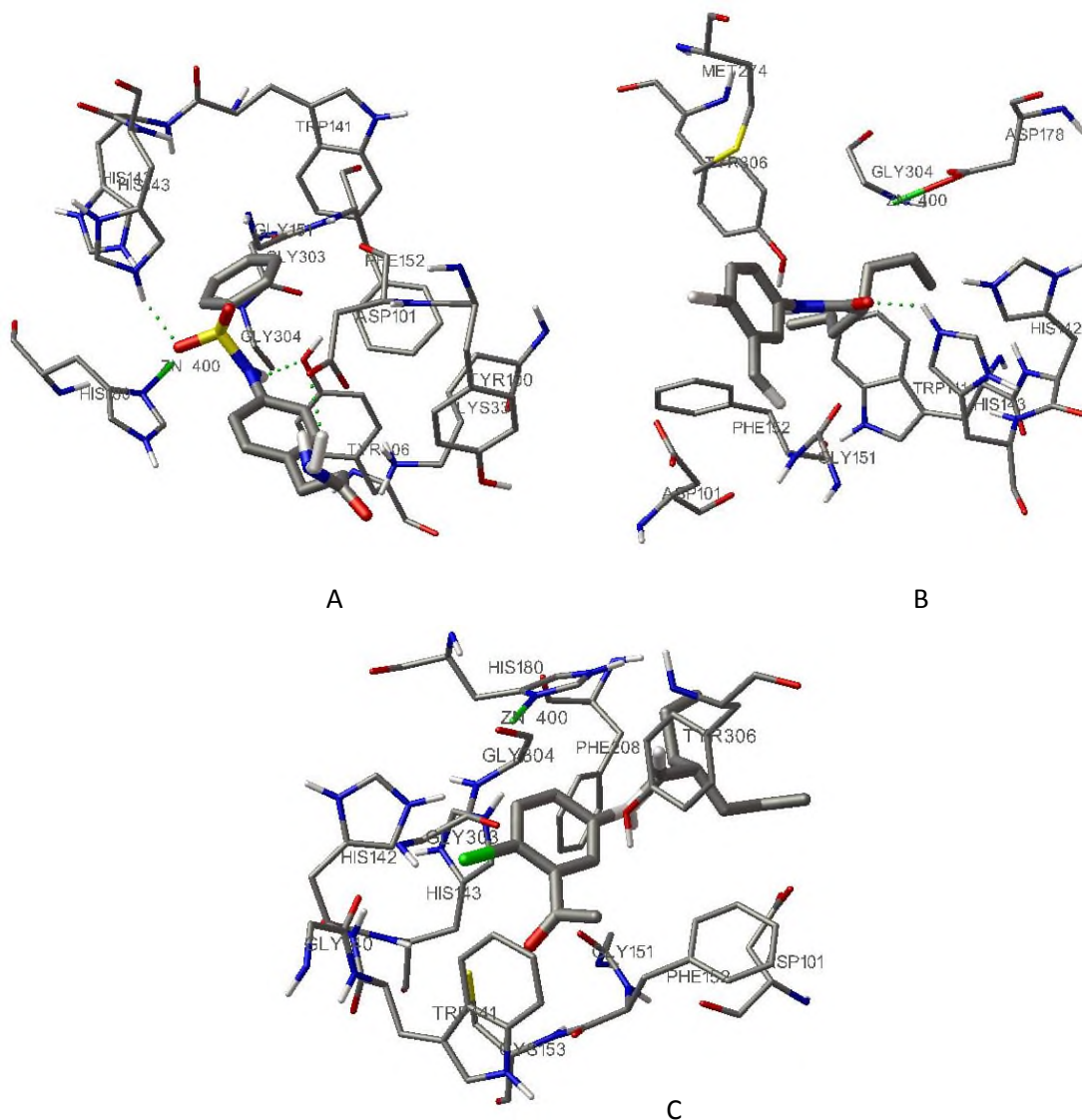


Fig.22. Interacciones del compuesto A) HID26, B) DAVP266 ($R_2=CHO$ y $R_3=I$), y C) DAVP281 ($R_2=CON(CH_3)_2$ y $R_3=Br$) con la superficie de la HDAC8.

En la Fig. 22 se observa la interacción de los derivados con las mejores afinidades de acuerdo al docking ciego, y puede apreciarse que independientemente de la sustitución de grupos electroattractores y electrodonadores de los 128 derivados del AVP, así como de los 30 análogos del hidroxamato, la orientación en el espacio del sitio catalítico es muy semejante, esto es debido a que las cadenas alifáticas estabilizan a los compuestos en la entrada del sitio catalítico de inhibición, y la diferencia en la afinidad del ligando hacia el sitio radica en el tipo de interacciones débiles que se pueden establecer en el complejo ligando-enzima, como fuerzas de Van der Waals, interacciones

aromáticas como π - π o π -catión, puentes de hidrogeno-oxígeno, provocado por los grupos químicos que presentan los compuestos con los residuos de aminoácidos del sitio catalítico (principalmente His180, Tyr306, Trp141, His142), y que el grupo funcional carboxilo interactúa electrostáticamente con el ión Zn^{2+} en aquellos derivados que otorgan una carga parcial negativa al compuesto.

Los resultados arrojados en el análisis exhaustivo del docking ciego para los análogos del hidroxamato y de los derivados del AVP indican que los mejores compuestos fueron la HID26, el DAVP266, y el DAVP281 (ver Fig. 23), ya que presentaron las constantes de disociación más bajas, que indican una mayor afinidad con el sitio catalítico de inhibición, además de que en sus estructuras químicas contienen grupos aceptores (interactúa el O del ligando con el H de la His143) y donadores (interactúa el H del ligando con el-O del Aps101 y de la Tyr306) de enlaces de hidrógeno, los cuales son favorecidas debido al tipo de sustituyente químico que presentan estas moléculas (ver Tabla 5) así como por su posición en el espacio.

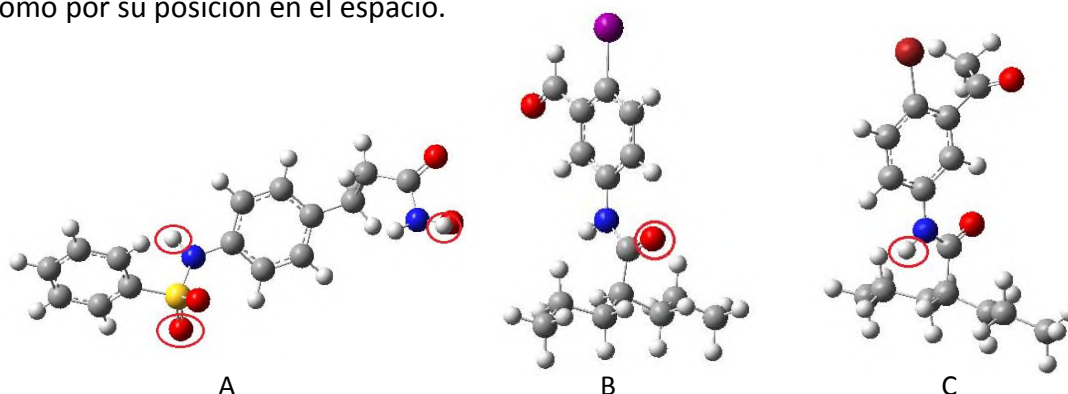


Fig.23. Estructuras moleculares de la A) HID26, B) DAVP266, C) DAVP281. (Los círculos rojos indican el O y el H de los ligandos que formaron los puentes H-O con los aminoácidos de sitio catalítico).

Tabla 5. Sustituyentes químicos de los derivados del AVP más prometedores.

Molécula	R ₁ (<i>orto</i>)	R ₂ (<i>meta</i>)	R ₃ (<i>para</i>)
DAVP266	-H	-CHO	-I
DAVP269	-H	-COCH ₃	-I
DAVP271	-H	-COCl	-I
DAVP277	-H	-CHO	-Br
DAVP281	-H	-CON(CH ₃) ₂	-Br
DAVP282	-H	-COCl	-Br
DAVP283	-H	-COOCOCH ₃	-Br
DAVP289	-H	-COOCH ₃	-NH ₂
DAVP389	-OCH ₃	-H	-CHO
DAVP405	-CH ₃	-H	-COOCH ₃
DAVP458	-NH ₂	-H	-F

En la Tabla 5 se muestran los mejores compuestos sustituidos con grupos electroattractores y electrodonadores en las posiciones *orto*, *meta* y *para* con respecto al grupo arilo en la estructura de los derivados del AVP. Los resultados de docking ciego obtenidos coinciden, en cierto sentido, con los mejores derivados de AVP obtenidos por Bermúdez y colaboradores [121] mediante un docking dirigido hacia el sitio catalítico reportado para la clase I de las HDACs, en este trabajo los mejores compuestos poseen en su estructura química grupos carbonilos y grupos aminos en posición *orto*, *meta* y *para*, teniendo gran similitud con nuestros mejores derivados del AVP. Al realizar el análisis de los grupos activos de los mejores derivados del AVP, se encontró lo siguiente:

- a) Halógenos: Bromo y Iodo. Generalmente podrán incrementar la utilidad terapéutica y las propiedades tóxicas de los compuestos activos.
- b) Grupos no polares: metilos y metilenos. Aumentaran la liposolubilidad.
- c) Grupos polares: carbonilo y amino. Incrementaran la hidrosolubilidad.

De acuerdo a la literatura el núcleo aromático puede incrementar la actividad biológica y por ende la toxicidad de los compuestos, los grupos carbonilo disminuyen considerablemente la actividad biológica y su toxicidad, mientras que los compuestos con grupos aldehídos tienen una mayor capacidad de reacción y mayor actividad biológica. Debido a lo anterior la reactividad de los derivados del AVP puede modificarse influyendo en la actividad biológica.

En las tablas 6 y 7 se muestran los aminoácidos que conforman el sitio catalítico para los mejores derivados del AVP y los mejores análogos del hidroxamato (en la Tabla 5H y 6H del anexo se muestran los resultados totales). En estas tablas se puede observar que todos los ligandos presentan casi el mismo sitio catalítico, ya que no todos los ligandos interactúan al 100 % con los mismos aminoácidos (el ión Zn^{2+} es el único que interactúa con todos los ligandos), por lo que se decidió seleccionar solo los aminoácidos más representativos de todos los ligandos, con el fin de restablecer un solo sitio catalítico, este sitio está conformado por los siguientes aminoácidos: Gly304, Aps178, Tyr306, His142, Zn^{2+} , Phe152, His143, His180, Met274, Phe208, Gly151, Trp141 y Asp101, de los cuales 7 son aminoácidos aromáticos.

Tabla 6. Aminoácidos que conforman el sitio catalítico de los mejores análogos del hidroxamato.

Molécula	Aminoácidos													
HID06	Gly304		Tyr306	His142	Zn ²⁺	phe152	His143	His180		Phe208	Phe207	Gly151	Trp141	
HID15	Gly304	Asp178	Tyr306	His142	Zn ²⁺		His143	His180	Met274	Phe208		Gly151	Trp141	
HID23	Gly304	Asp178	Tyr306	His142	Zn ²⁺	Phe152	His143	His180	Met274	Phe208		Gly151	Trp141	Asp101
HID25			Tyr306	His142	Zn ²⁺		His143	His180		Phe208		Gly151	Trp141	Asp101
HID26	Gly304		Tyr306	His142	Zn ²⁺	Phe152	His143	His180				Gly151	Trp141	Asp101
HID29	Gly304		Tyr306	His142	Zn ²⁺		His143	His180	Met274	Phe208		Gly151	Trp141	Asp101
HID34	Gly304		Tyr306	His142	Zn ²⁺	Phe152	His143	His180	Met274	Phe208	Phe207	Gly151	Trp141	
HID35		Asp178		His142	Zn ²⁺	Phe152	His143	His180				Gly151	Trp141	Asp101
HID37	Gly304	Asp178	Tyr306	His142	Zn ²⁺	Phe152	His143	His180		Phe208		Gly151	Trp141	Asp101

Tabla 7. Aminoácidos que conforman el sitio catalítico de los mejores derivados del AVP.

Molécula	Aminoácidos												
DAVP 266				Zn ²⁺	His143	His142	Met274	Gly151	Asp101	Phe152	Tyr306		
DAVP 269	Phe208	His180		Zn ²⁺	His143	His142	Met274	Gly151	Asp101	Phe152	Tyr306		Phe207
DAVP 271	Phe208	His180		Zn ²⁺	His143	His142		Gly151	Asp101	Phe152	Tyr306	Lys33	
DAVP 277	Phe208	His180		Zn ²⁺	His143	His142		Gly151	Asp101	Phe152	Tyr306	Lys33	
DAVP 281	Phe208	His180		Zn ²⁺	His143	His142		Gly151	Asp101	Phe152	Tyr306		
DAVP 282	Phe208	His180		Zn ²⁺	His143	His142	Met274	Gly151	Asp101	Phe152	Tyr306		
DAVP 283	Phe208	His180		Zn ²⁺	His143	His142		Gly151	Asp101	Phe152	Tyr306	Lys33	
DAVP 289	Phe208	His180		Zn ²⁺	His143		Met274	Gly151		Phe152			Phe207
DAVP 389	Phe208	His180		Zn ²⁺	His143		Met274	Gly151	Asp101	Phe152			
DAVP 405	Phe208	His180		Zn ²⁺	His143	His142	Met274	Gly151	Asp101	Phe152	Tyr306		Phe207
DAVP 458	Phe208	His180		Zn ²⁺	His143	His142		Gly151		Phe152	Tyr306		

7.2. QSAR

7.2.1. Conjunto de compuestos con actividad biológica conocida

De acuerdo a los resultados obtenidos por el docking ciego, sobre los 38 análogos del hidroxamato se observó que solo 30 de estas moléculas se encontraban dentro del mismo sitio catalítico (sitio de inhibición reportado para la clase I de las HDAC). Estas moléculas fueron establecidas como el conjunto de entrenamiento para la realización del análisis QSAR. En la Tabla 8 se muestran los 30 análogos del hidroxamato con sus respectivas actividades biológica.

Tabla 8. Actividad biológica de los 30 análogos del hidroxamato.

Molécula	pIC₅₀	IC₅₀
HID02	8	0.903
HID03	7.54	0.877
HID 04	7.28	0.862
HID 05	7.36	0.867
HID 06	6.95	0.842
HID 07	6.99	0.844
HID 08	7.15	0.854
HID 09	6.3	0.799
HID 11	6.82	0.834
HID 13	7.02	0.846
HID 14	7.35	0.866
HID 15	8.3	0.919
HID 16	8.7	0.939
HID 17	8.02	0.904
HID 18	8.35	0.922
HID 19	7.6	0.881
HID 21	7.13	0.853
HID 22	6	0.778
HID 23	6	0.778
HID 25	6.05	0.782
HID 26	7	0.845
HID 27	8.3	0.919
HID 28	8.07	0.907
HID 29	8.4	0.924
HID 33	7.05	0.848
HID 34	7.52	0.876
HID 35	5.82	0.765
HID 37	6.87	0.837
HID 38	7.46	0.873
HID 39	7.82	0.893

[pIC₅₀ es logaritmo negativo de IC₅₀]

Si comparamos la Tabla 2H del anexo con la Tabla 8, nos podemos dar cuenta que la mayoría de las moléculas que fueron removidas presentan grupos muy voluminosos de peso molecular alto, razón que justifica la formación de un complejo en un sitio catalítico diferente. Por lo tanto no es factible hacer la comparación estructural entre estas moléculas y los derivados del AVP. Además de que no formaron interacciones con el sitio catalítico, ni con el ión Zn^{2+} , requisito propuesto anteriormente.

7.2.2. Identificación de los descriptores moleculares de los análogos del hidroxamato

En el programa DRAGON se lograron determinar 1179 descriptores moleculares que mejor definieron la estructura de los 30 compuestos elegidos de los análogos del hidroxamato, más los descriptores obtenidos por los cálculos cuánticos como la energía mínima, la energía LUMO, HOMO, la dureza, la blandura, dipolo X, Y, Z y total, así como los descriptores obtenidos por el docking ciego: la energía libre de Gibbs, intermolecular, interna y la constante de disociación.

7.2.3. Selección de los descriptores de los análogos del hidroxamato

Los 1179 descriptores fueron analizados en su totalidad, encontrando la existencia de variables constantes, por lo que estos fueron eliminados. Después de este proceso 997 los descriptores fueron seleccionados, y mediante una subrutina escrita en MATLAB se seleccionaron finalmente 165 descriptores con un coeficiente de correlación de colinealidad menor a 0.8 entre ellos.

7.2.4. Regresión lineal por pasos en SPSS de los análogos del hidroxamato

Al realizar la regresión lineal por etapas dentro del programa SPSS, se establecieron como variables independientes los descriptores obtenidos en DRAGON, los descriptores de energía mínima, energía HOMO, LUMO, N dureza, S blandura, dipolo X, Y, Z y total, energía libre de Gibbs, intermolecular, interna y la constante de disociación y como variable dependiente la actividad biológica; obteniendo los siguientes descriptores:

BELm4	MATS5m	Mor05e	Mor04u	R4e+	H0e
--------------	--------	--------	--------	------	-----

y al hacer el cambio de variable dependiente de IC₅₀ a energía libre de Gibbs en SPSS, se encontraron los siguientes descriptores:

K_d	E. Intermolecular	BEIm4	Mor15u	L3u	Mor17e
----------------------	--------------------------	--------------	--------	-----	--------

Por lo que en el conjunto de entrenamiento se propusieron 6 modelos matemáticos en función de la energía libre de Gibbs, correspondientes a los 30 análogos del hidroxamato (ver Tabla 9 y 10).

Tabla 9. Modelos de la 2ª regresión lineal múltiple del conjunto de entrenamiento.

Modelo	Constante	K_d	E. Intermolecular	BEIm4	Mor15u	L3u	Mor17e
1	-7.275	0.022					
2	-3.217	0.015		0.412			
3	-3.119	0.008		0.659	1.636		
4	-2.89	0.008		0.681	1.736	-0.293	
5	-3.008	0.008		0.658	1.828	-0.235	-0.318
6	-2.922	0.009		0.667	1.723	-0.137	-0.433
							-0.296

[En rojo se muestra el mejor modelo con sus respectivos descriptores]

Tabla 10. Coeficientes estadísticos de la 2ª RLM del conjunto de entrenamiento.

Modelo	R	RSS	AIC	FIT
1	0.844	3.993	0.138	2.247
2	0.905	2.513	0.090	3.594
3	0.971	0.784	0.029	10.997
4	0.982	0.498	0.019	14.690
5	0.989	0.314	0.012	19.508
6	0.993	0.207	0.009	24.630

[En rojo se muestra el mejor modelo que consta de los siguientes descriptores: K_d, E. Intermolecular y BEIm4]

Los coeficientes de correlación de la Tabla 10 demostraron una relación lineal de 0.971 entre el descriptor BEIm4 y la energía libre de Gibbs, ya que estos son valores muy cercanos a 1. Por otra parte el error cuadrático (RSS) permitió evaluar la bondad del ajuste de los valores de la variable independiente predichos por el modelo con respecto a los valores observados, si observamos la Tabla 10 el error cuadrático para las primeras dos variables es muy alto, mientras que las demás son menores a 1, por lo que el ajuste de los valores se encuentran en un intervalo más amplio, y si tomamos en cuenta el criterio de información de Akaike (AIC) [122] el cual nos ofrece una medida relativa de la pérdida de información de un modelo, nos podemos dar cuenta que en el modelo 3 hay una menor pérdida de información (AIC= 0.029); mientras que la función de aptitud de Kubinyi (FIT) [123] solo muestra la comparación de los modelos con las 6 variables.

De acuerdo a los parámetros anteriores el mejor modelo es el modelo 3 (marcado de color rojo en las tablas 9 y 10). La combinación de estos dos parámetros justifica la adición de una nueva variable en el modelo.

7.2.5. Identificación de los descriptores moleculares de los derivados del AVP

Se encontraron 1183 descriptores moleculares de la base de datos del programa DRAGON que mejor definieron a las 128 estructuras de los derivados del AVP, más los descriptores cuánticos: la energía mínima, la energía LUMO, HOMO, la dureza, la blandura, dipolo X, Y, Z y total, así como los descriptores obtenidos del docking ciego: la energía libre de Gibbs, intermolecular, interna y la constante de disociación. En toda esta gama de descriptores solo se buscaron aquellos descriptores que conformaron las 6 variables de la regresión lineal por pasos en SPSS con la variable dependiente energía libre de Gibbs, estableciendo el conjunto de predicción.

Realizando la estadística correspondiente para los derivados del AVP se obtuvieron los siguientes datos:

Tabla 11. Coeficientes estadísticos de la 2ª RLM de los derivados del AVP.

Modelo	R	AIC	FIT
1	0.835	6.603	2.258
2	0.860	3.947	2.690
3	0.897	1.611	3.719
4	0.892	1.962	3.344
5	0.881	1.364	2.754
6	0.876	1.626	2.427

[En rojo se muestra el mejor modelo que consta de los siguientes descriptores: K_d , E. Intermolecular y Belm4]

De acuerdo a la Tabla 11 que corresponde a la estadística realizada para el conjunto de predicción (derivados del AVP) el mejor modelo fue el modelo 3, pero a diferencia del conjunto de entrenamiento los parámetros estadísticos R, AIC, FIT, tuvieron los mejores valores para este modelo, por ejemplo el coeficiente de correlación lineal es de 0.897, mientras que AIC es de 1.611 el menor valor de pérdida de información. El mejor modelo para esta regresión lineal corresponde a los siguientes descriptores: K_d , E. Intermolecular y Belm4, donde K_d y E. intermolecular, que son descriptores obtenidos del docking ciego y Belm4 es un descriptor de la base de datos del programa DRAGON.

7.2.6. Regresión Lineal Múltiple (RLM)

La RLM es el método tradicional para generar modelos QSAR, se basa en el método de los mínimos cuadrados, esto es, encontrar la combinación lineal de variables que minimice la suma de los cuadrados de los residuales (diferencia entre los valores observados de la variable dependiente y los valores predichos) [124].

El modelo inicial M158 (ver Fig. 24) fue construido con 3 variables: K_d , E. Intermolecular y Belm4 en relación a la energía libre de Gibbs; el cual fue construido con los 30 análogos del hidroxamato (conjunto de entrenamiento) y los 128 derivados de AVP (conjunto predicción), obteniendo las ecuaciones 1 y 2. Es claro que los compuestos de los derivados del AVP con respecto a la energía libre de Gibbs mayores a -6.0 presentan deficiencias en la predicción (círculo rojo en la Fig.24).

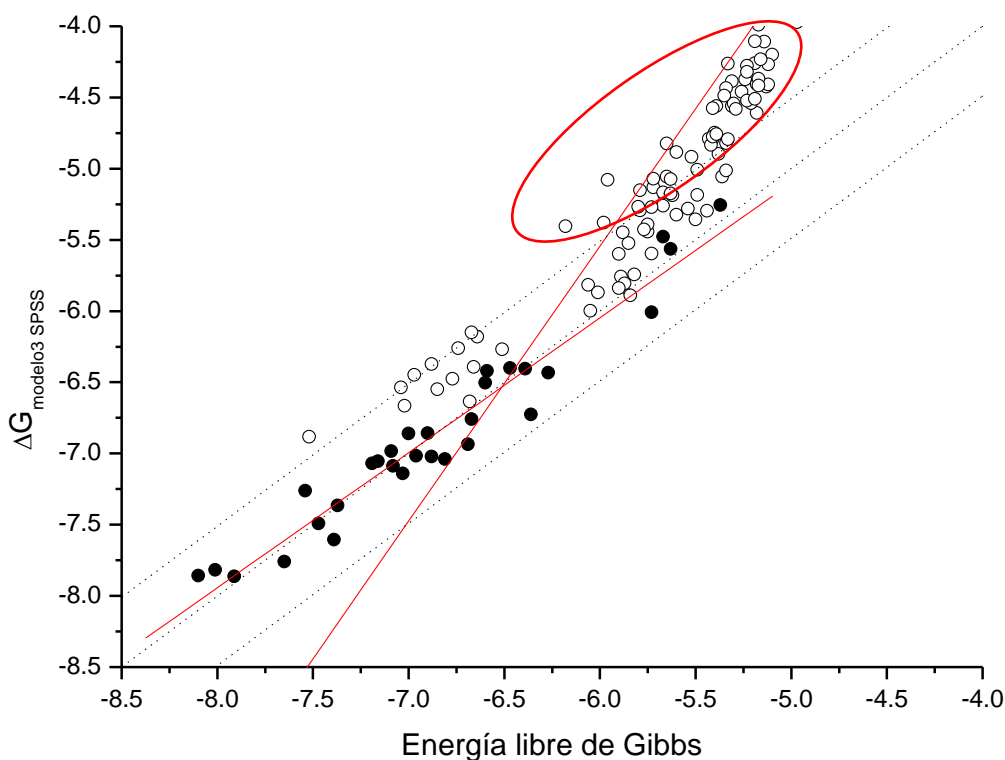


Fig. 24. Modelo M158. Modelo establecido en relación a la energía libre de Gibbs de los 30 análogos del hidroxamato (círculos negros) y los 128 derivados del AVP (círculos blancos), del modelo 3 de la Tabla 9 (El círculo rojo muestra aquellas moléculas que presentan deficiencias en el conjunto de predicción).

Conjunto de entrenamiento (30 análogos del hidroxamato)

$$\Delta G_{\text{modelo3 SPSS}} = -0.362 (\pm 0.302) + 0.948 (\pm 0.044) \Delta G$$

$$R = 0.971, SD = 0.163, p < 0.0001, N = 30, SEE = 0.784, q^2_{\text{LOO}} = 0.995, q^2_{\text{LGO}} = 0.976,$$

$$\text{FIT} = 10.997, \text{AIC} = 0.029, \text{F} = 145.263$$

(ecuación 1)

Conjunto de predicción (128 derivados del AVP)

$$\Delta G_{\text{modelo3 SPSS}} = 6.043 (\pm 0.466) + 1.931 (\pm 0.085) \Delta G$$

$$R = 0.897, SD = 0.582, p < 0.0001, N = 128, FIT = 3.7190, AIC = 1.611 \quad (\text{ecuación 2})$$

Se estableció una probabilidad del 99.7 %, con el objetivo de que todos los datos estuvieran dentro de tres desviaciones estándares de la media (σ), aunque esto no se cumplió ya que en la Fig. 24 se observa que la gran mayoría de las moléculas de los derivados del AVP se salen de los límites establecidos a 3 σ (círculo rojo).

Para confirmar que el modelo de regresión lineal múltiple M158 es el modelo óptimo, se determinó el F-estadístico total (ecuación 1) en SPSS, lo cual permitió contrastar la hipótesis nula de que R es cero, es decir, contrastar la hipótesis de que la pendiente de la recta de regresión vale cero. En el modelo M158 el coeficiente de regresión es de 0.971, dando a entender que los análogos del hidroxamato están linealmente relacionados. Este modelo también fue sometido a un análisis de validación cruzada con el objetivo de determinar la robustez y capacidad predictiva del modelo, para esto se realizó una validación cruzada (CV) por Leave-One-Out (q^2 LOO “dejar un dato afuera”) y otra por Leave-Group-Out (LGO “dejar un grupo de datos afuera”). Para la ecuación 1, los valores fueron $q^2_{\text{LOO}} = 0.9948$ y $q^2_{\text{LGO}} = 0.9759$. Aunque en general se considera aceptable un valor de q^2 mayor a 0.5, se ha demostrado que un valor adecuado de q^2 es condición necesaria pero no suficiente para afirmar la capacidad predictiva del modelo [125]. El q^2 tiende a ser sobreoptimista, en especial en aquellos casos en los que el conjunto de entrenamiento no es estructuralmente diverso, por ejemplo, cuando está formado por derivados más o menos similares de un mismo compuesto líder, lo cual no es del todo inusual en el campo de la Química Farmacéutica [126]. En este caso los algoritmos de LOO-CV y de LGO-CV fueron escritos en MATLAB.

7.2.7. Expresión matemáticamente de la actividad biológica de los derivados del AVP

En la primera regresión lineal en relación a la actividad biológica se obtuvieron los siguientes descriptores:

BELm4	MATS5m	Mor05e	Mor04u	R4e+	H0e
--------------	--------	--------	--------	------	-----

mientras que en la segunda regresión lineal en relación a la energía libre de Gibbs se obtuvieron estos otros:

K_d	E. Intermolecular	BELm4	Mor15u	L3u	Mor17e
-------	-------------------	-------	--------	-----	--------

El mejor modelo matemático de acuerdo a los parámetros estadísticos mencionados anteriormente es el modelo 3, por esta razón los siguientes descriptores están relacionados matemáticamente con la actividad biológica.

Energía libre de Gibbs	BELm4	K_d	E. Intermolecular
------------------------	-------	-------	-------------------

Donde BELm4 es un descriptor 2D-BCUT de DRAGON que representa el más bajo valor propio n. 4 de la matriz ponderado por masas atómicas, desarrollado por Burden F. R. en 1989 [122]. 2D-BCUT son descriptores moleculares obtenidos a partir de los valores propios positivos y negativos de la matriz en relación a los pesos atómicos. Las tres variantes de este tipo de descriptor son: carga de un átomo, un átomo de polarizabilidad y las propiedades de enlaces de hidrógeno aceptor y donador. El número de valores propios menores o más alto se calcula en específico con los parámetros de configuración.

Después se realizó otra regresión lineal en SPSS colocando a Belm4 como variable dependiente, ya que este descriptor fue el único que coincidió en las dos regresiones anteriores, y como variables independientes: energía libre de Gibbs, K_d y energía intermolecular. En la tablas 12 y 13 (los datos en rojo son los mejores valores) se muestran los datos obtenidos de SPSS y de la estadística realizada para dicha regresión lineal para los 30 análogos del hidroxamato:

Tabla 12. Modelos de la 3^a RLM del conjunto de entrenamiento.

Modelo	Constante	Energía libre de Gibbs	E. Intermolecular
1	2.145	0.100	
2	0.852	0.352	-0.317

Tabla 13. Coeficientes estadísticos de la 3^a RLM del conjunto de entrenamiento.

Modelo	R	RSS	AIC	FIT
1	0.388	0.782	0.027	0.160
2	0.870	0.223	0.008	2.472

[En la Tabla 12 y 13 el mejor modelo (en rojo) consta de los siguientes descriptores: Energía libre de Gibbs y E. Intermolecular]

Mientras que en la Tabla 14 se muestran los datos obtenidos de la estadística realizada para los 128 derivados del AVP:

Tabla 14. Coeficientes estadísticos de la 3ª RLM del conjunto de predicción.

Modelo	R	RSS	AIC	FIT
1	0.067	5.117	0.041	0.004
2	0.519	1.760	0.014	0.350

[En rojo se muestra el mejor modelo que consta de los siguientes descriptores: Energía libre de Gibbs y E. Intermolecular]

En esta regresión lineal solo permanecieron dos descriptores, la energía libre de Gibbs y la energía intermolecular y al igual que el conjunto de entrenamiento el mejor modelo es el 2 (datos en color rojo).

Se realizó una cuarta regresión lineal en SPSS colocando a IC_{50} como la variable dependiente y como variables independientes, K_d , energía intermolecular, Belm4 y la energía libre de Gibbs, con el objetivo de encontrar la ecuación matemática que prediga la actividad biológica teórica de cada derivado del AVP.

Tabla 15. Modelo de la 4ª RLM del conjunto de predicción.

Modelo	Constante	BELm4
1	0.541	0.220

Tabla 16. Coeficientes estadísticos de la 4ª RLM del conjunto de predicción.

Modelo	R	RSS	AIC	FIT
1	0.832	0.020	0.0007	2.031

[El modelo 1 consta únicamente del descriptor Belm4]

Con los datos obtenidos en esta última regresión lineal se puede observar que el único descriptor que mantiene una relación matemática con la actividad biológica es el descriptor de DRAGON Belm4 (ver Tabla 15).

La Tabla 16 muestra que el coeficiente de regresión lineal es 0.832, lo que indica que la relación es muy estrecha entre ambas variables Belm4 y IC_{50} . A partir de los datos de la última regresión lineal se encontró la siguiente ecuación matemática con la que fue posible encontrar la actividad biológica teórica de los derivados de AVP.

$$IC_{50} = 0.541 + 0.22 \text{ Belm4}$$

7.2.8. Regresión lineal múltiple en relación al descriptor Belm4

La Fig. 25 representa el segundo modelo descrito en las tablas 13 y 14, correspondiente a la variable dependiente Belm4, tanto para el modelo de

entrenamiento y predicción con 3σ . Este modelo se seleccionó de acuerdo a los parámetros estadísticos mencionados anteriormente (R, AIC y FIT). Al igual que el primer modelo M158 se determinó el F-estadístico total (ver ecuación 3), la CVLOO y la CVLGO.

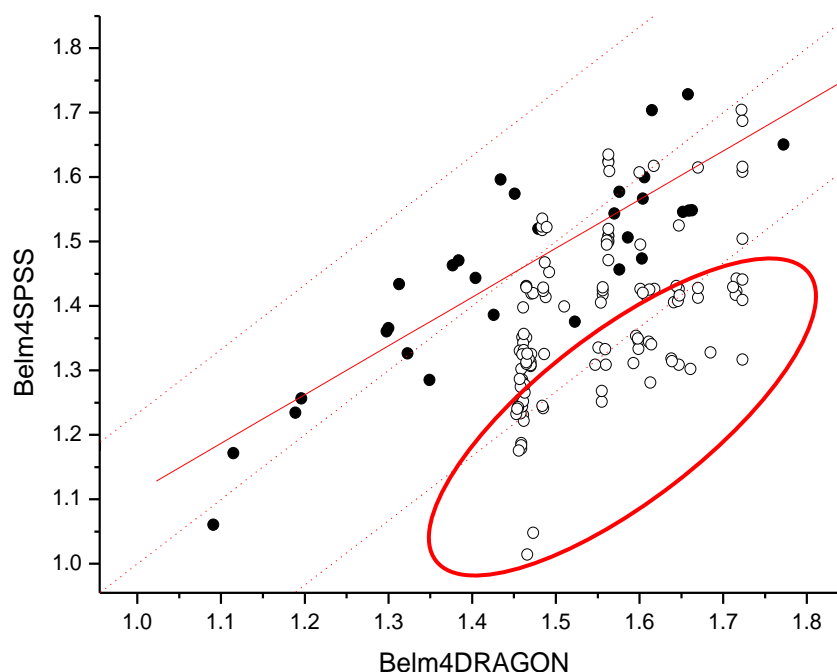


Fig. 25. Modelo M158Belm4. Modelo establecido a partir del descriptor Belm4 de los 30 análogos del hidroxamato (círculos negros) y los 128 derivados del AVP (círculos blancos), del modelo 2 de la Tabla 12 (El círculo rojo muestra aquellas moléculas que presentan deficiencias en el conjunto de predicción).

Conjunto de entrenamiento (30 análogos del hidroxamato)

$$\text{Belm4}_{\text{modelo2 SPSS}} = 0.354 (\pm 0.119) + 0.757 (\pm 0.081) \text{Belm4}_{\text{DRAGON}}$$

$$R = 0.870, SD = 0.078, p < 0.0001, N = 30, SEE = 0.223, q^2_{\text{LOO}} = 0.609, q^2_{\text{LGO}} = 0.681,$$

$$\text{FIT} = 2.472, \text{AIC} = 0.008, F = 42.148$$

(ecuación 3)

Conjunto de predicción (128 derivados del AVP)

$$\text{Belm4}_{\text{modelo2 SPSS}} = 0.254 (\pm 0.165) + 0.727 (\pm 0.106) \text{Belm4}_{\text{DRAGON}}$$

$$R = 0.519, SD = 0.106, p < 0.0001, N = 128, \text{FIT} = 0.350, \text{AIC} = 0.014$$

(ecuación 4)

Este modelo fue construido con 2 descriptores, donde los parámetros estadísticos son los mejores, por ejemplo el coeficiente de correlación, demostró el valor más alto de los dos modelos (Tabla 13) con $R = 0.870$ y $q^2 > 0.5$. Este modelo como se observa en la Fig. 25 muestra que una gran cantidad de los compuestos de DAVP se encuentran por debajo al límite inferior.

7.2.9. Regresión lineal múltiple en relación al descriptor IC_{50}

La Fig. 26 y las ecuaciones 5 y 6 representan a la actividad biológica teórica de los derivados del AVP en relación a la actividad biológica experimental de los análogos del hidroxamato.

La RLM del modelo $M158IC_{50}$ fue construido con 1 solo descriptor, obtenido a partir del único modelo descrito en la Tabla 16; al igual que en los casos anteriores los valores de q^2 son mayores a 0.5 (ver ecuación 5).

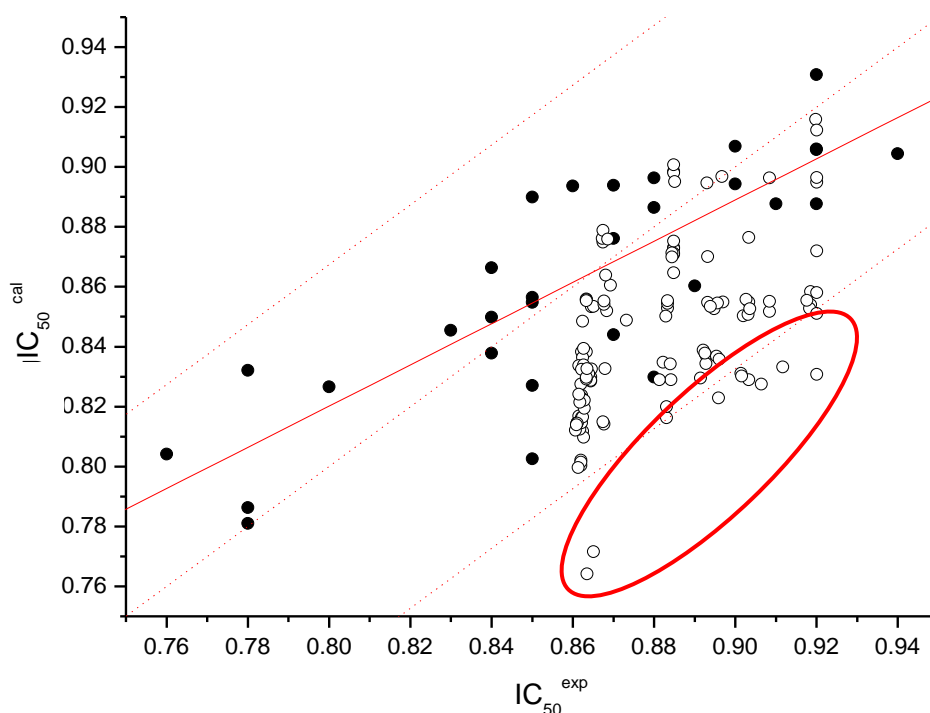


Fig. 26. Modelo $M158IC_{50}$. Modelo en función al IC_{50} de los 30 análogos del hidroxamato (círculos negros) y los 128 derivados del AVP (círculos blancos), del modelo 1 de la Tabla 15 (El círculo rojo muestra aquellas moléculas que presentan deficiencias en el conjunto de predicción).

Conjunto de entrenamiento (30 análogos del hidroxamato)

$$IC_{50}^{cal} = 0.270 (\pm 0.076) + 0.688 (\pm 0.088) IC_{50}^{exp}$$

$$R = 0.826, SD = 0.022, p < 0.0001, N = 30, SEE = 0.020, q^2_{LOO} = 1, q^2_{LGO} = 0.522,$$

$$FIT = 2.031, AIC = 0.0007, F = 62.802$$

(ecuación 5)

Conjunto de predicción (128 derivados del AVP)

$$IC_{50}^{cal} = 0.204 (\pm 0.094) + 1.931 (\pm 0.106) IC_{50}^{expHID}$$

$$R = 0.519, SD = 0.023, p < 0.0001, N = 128,$$

(ecuación 6)

El modelo M158 IC₅₀ también muestra como algunos derivados del AVP se salen del límite inferior de la gráfica, para ser exactos 11 son las moléculas de DAVP que se salen de este límite establecido a 3σ .

Con todo lo anterior es posible decir que el parámetro más importante de todo la estadística es q^2 ya que este no aumenta indefinidamente a medida que se adicionan otros parámetros al modelo, por el contrario alcanza un máximo, por lo que q^2 es un parámetro que debe de considerarse en la determinación del número óptimo de variables que se deben de incluir en el modelo.

7.2.10. Redes Neuronales Artificiales (ANN)

Como se comentó anteriormente el motivo que, probablemente, ha dado más impulso a la utilización de las redes neuronales en el campo de la identificación y control de procesos es su demostrada capacidad de aproximar funciones con un grado arbitrario de precisión. Es por esta razón que se procedió a la aplicación de las ANN con varias combinaciones de neuronas (2, 2.5, 3.5 y 4 neuronas) al modelo M158Belm4 de la RLM, lo que permite facilitar la predicción de la categoría a la que pertenecen las moléculas o el valor de alguna propiedad fisicoquímica. Realizado lo anterior se encontró que el modelo M158Belm4 con 3 neuronas presentaba una mejor descripción en relación a las demás, sin embargo esta regresión lineal no mejoró con respecto a la obtenida a la RLM, esto a pesar de que la red neuronal se entrenó varias veces (desventaja principal de esta metodología), con lo que se esperaba generar un modelo lineal, indicando que muy probablemente hubo una sobreparametrización, es decir, que el modelo “memorizó”. Con respecto a lo anterior se esperaba que la aplicación de las redes neuronales al modelo M158Belm4 cumpliera con el objetivo de incorporar aquellos derivados del AVP que se salen del límite inferior de la RLM y así poder mejorar la regresión lineal para este modelo.

Por otro lado, es de suma importancia saber que ANN no se trata de un método de selección de variables, sino de generación de modelos, por lo que se debe tener cierto conocimiento teórico o intuitivo sobre cuáles del conjunto total de variables son las que contribuyen más a la propiedad biológica estudiada. Por lo que se decidió eliminar a todas aquellas moléculas de los derivados del AVP que se encontraban fuera del límite inferior de la RLM, siendo 71 moléculas las removidas, las cuales no mantienen algún tipo

de relación con las propiedades de la estructura y la actividad biológica de los análogos del hidroxamato.

7.2.11. Regresión lineal múltiple en relación al descriptor Belm4 (57DAVP)

Al realizar la RLM en función del descriptor Belm4 de los 30 análogos del hidroxamato y los 57 derivados del AVP, se obtuvo la gráfica de la Fig. 27 junto con las ecuaciones 7 y 8:

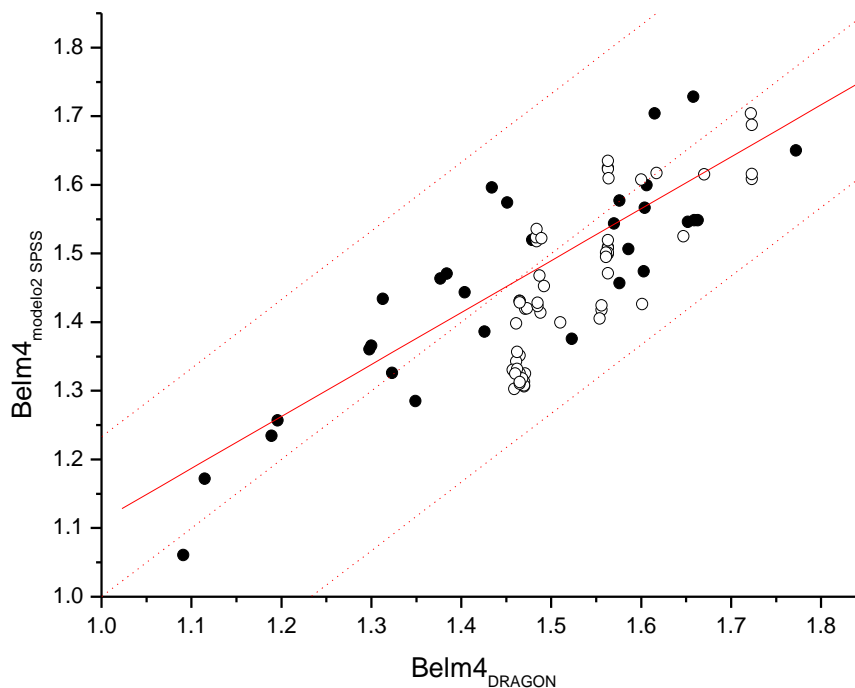


Fig. 27. Modelo M87Belm4. Modelo establecido en función de descriptor Belm4 de los 30 análogos del hidroxamato (círculos negros) y los 75 derivados del AVP (círculos blancos), del modelo 2 de la Tabla 12.

Conjunto de entrenamiento (30 análogos del hidroxamato)

$$\text{Belm4}_{\text{modelo2 SPSS}} = 0.354 (\pm 0.119) + 0.757 (\pm 0.081) \text{Belm4}_{\text{DRAGON}}$$

$$R = 0.870, SD = 0.078, p < 0.0001, N = 30, SEE = 0.223, q^2_{\text{LOO}} = 0.609, q^2_{\text{LGO}} = 0.681,$$

$$\text{FIT} = 2.472, \text{AIC} = 0.008, F = 42.148$$

(ecuación 7)

Conjunto de predicción (57 derivados del AVP)

$$\text{Belm4}_{\text{modelo2 SPSS}} = -0.348 (\pm 0.180) + 1.180 (\pm 0.118) \text{Belm4}_{\text{DRAGON}}$$

$$R = 0.804, SD = 0.067, p < 0.0001, N = 57$$

(ecuación 8)

Los descriptores BCUT se han aplicado con éxito en diversos estudios de QSAR. Por ejemplo, Stanton [127] encontró que los descriptores BCUT pueden proporcionar una

información única sobre las estructuras moleculares y contribuir de manera significativa a las ecuaciones resultantes. En el 2000, Pirard y Pickett [128] clasificaron correctamente los inhibidores de la quinasa utilizando el análisis discriminante PLS junto con los descriptores BCUT, y después [129] obtuvieron modelos altamente predictivos con descriptores BCUT y el algoritmo genético en el análisis QSAR. Además, Ford et al. [130] trabajaron con parámetros BCUT como variables de entrada para reconocer los compuestos que actúan sobre blancos biológicos pertenecientes a las proteínas quinasas. Los resultados ilustran que los descriptores BCUT tienen utilidad en compuestos exigentes que interactúan con las familias de genes particulares [131].

7.2.12. Regresión lineal múltiple en relación al descriptor IC_{50} (57DAVP)

En la RLM en función del descriptor IC_{50} de los 30 análogos del hidroxamato y los 57 derivados del AVP, se obtuvo la gráfica de la Fig. 28 junto con las ecuaciones 9 y 10:

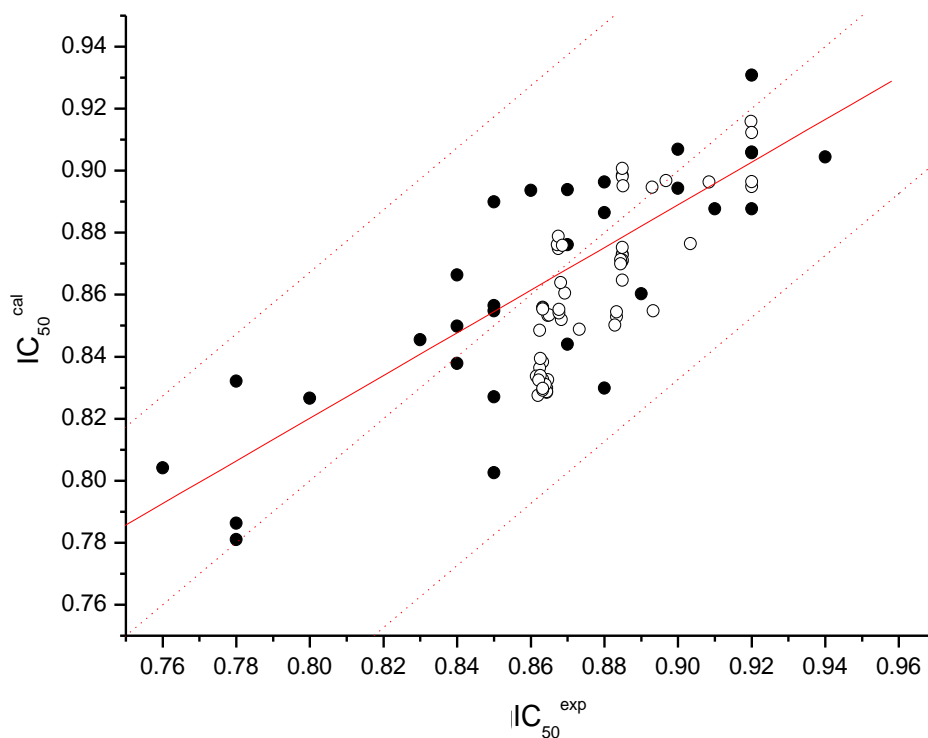


Fig. 28. Modelo M87 IC_{50} . Modelo establecido en relación al IC_{50} de los 30 análogos del hidroxamato (círculos negros) y los 57 derivados del AVP (círculos blancos), del modelo 1 de la Tabla 15.

Conjunto de entrenamiento (30 análogos del hidroxamato)

$$IC_{50}^{cal} = 0.270 (\pm 0.076) + 0.688 (\pm 0.088) IC_{50}^{exp}$$

$$R = 0.826, SD = 0.022, p < 0.0001, N = 30, SEE = 0.020, q^2_{LOO} = 1, q^2_{LGO} = 0.522,$$

$$FIT = 2.031, AIC = 0.0007, F = 62.802$$

(ecuación 9)

Conjunto de predicción (57 derivados del AVP)

$$IC_{50}^{cal} = -0.174 (\pm 0.103) + 1.180 (\pm 0.118) IC_{50}^{expHID}$$

$$R = 0.804, SD = 0.015, p < 0.0001, N = 57, \quad (\text{ecuación 10})$$

$$IC_{50}^{calc} = 0.541 (\pm 0.041) + 0.220 (\pm 0.028) \text{Belm4} \quad (\text{ecuación 11})$$

En los modelos M87Belm4 y M87IC₅₀ (Fig. 27 y 28 respectivamente), el coeficiente de correlación lineal (marcado de color rojo en la ecuación 8 y 10) correspondiente a los 57 derivados del AVP se mejora drásticamente en comparación a los modelos M158Belm4 y M158IC₅₀ donde se consideran los 128 derivados del AVP, ya que R cambia de 0.520 a 0.804. Con el resultado anterior se da por entendido que el mejor modelo se establece a partir de 57 derivados del AVP y 30 análogos del hidroxamato.

En términos generales la ecuación 11 es el modelo matemático que mejor predice la actividad biológica de este tipo de compuestos, mucho mejor que donde se aplica la metodología de las ANN.

7.2.13. Predicción de la actividad biológica teórica de los 57 derivados del AVP

Utilizando la ecuación matemática donde se relaciona la actividad biológica de los 30 análogos del hidroxamato con el descriptor Bem4, se determinó la actividad teórica de los 57 derivados del AVP (ver Tabla 17). Los 57 compuestos seleccionados de los derivados del AVP han demostrado que la predicción de la actividad inhibitoria fue arreglada en un solo conjunto de datos mediante un análisis de RLM.

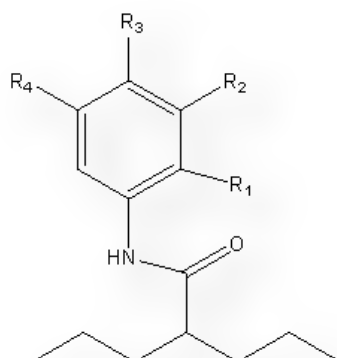


Fig. 29. Estructura base de los derivados del AVP con sustituciones R₁ (*orto*), R₂, R₄ (*meta*) y R₃ (*para*).

Tabla 17. Grupos funcionales de los 57 DAVP con su respectiva IC₅₀ teórica.

Molécula	R ₁ (<i>orto</i>)	R ₂ (<i>meta</i>)	R ₃ (<i>para</i>)	R ₄ (<i>meta</i>)	IC ₅₀
DAVP012	-H	-H	-CHO	-H	0.832
DAVP015	-H	-H	-COOCH ₃	-H	0.852
DAVP030	-H	-H	-COCl	-H	0.833
DAVP032	-H	-COOCOCH ₃	-H	-H	0.873
DAVP033	-H	-H	-COOCOCH ₃	-H	0.873
DAVP038	-H	-OCH ₃	-H	-H	0.838
DAVP042	-H	-H	-OH	-H	0.833
DAVP056	-H	-NH ₂	-H	-H	0.836
DAVP057	-H	-H	-NH ₂	-H	0.833
DAVP116	-H	-OH	-CHO	-H	0.853
DAVP118	-H	-Cl	-CHO	-H	0.830
DAVP119	-H	-I	-CHO	-H	0.828
DAVP195	-H	-N(CH ₃) ₂	-COOCOCH ₃	-H	0.895
DAVP199	-H	-NH ₂	-COOCOCH ₃	-H	0.898
DAVP201	-H	-CH ₂ CH ₃	-COOCOCH ₃	-H	0.897
DAVP202	-H	-CH ₂ CH ₂ CH ₃	-COOCOCH ₃	-H	0.916
DAVP205	-H	-C(CH ₃) ₃	-COOCOCH ₃	-H	0.895
DAVP220	-H	-CN	-OCH ₃	-H	0.829
DAVP222	-H	-CHO	-OCH ₃	-H	0.853
DAVP223	-H	-COOCH ₃	-OCH ₃	-H	0.875
DAVP225	-H	-COCH ₃	-OCH ₃	-H	0.853
DAVP229	-H	-F	-OCH ₃	-H	0.831
DAVP231	-H	-CN	-OH	-H	0.827
DAVP233	-H	-CHO	-OH	-H	0.856
DAVP234	-H	-COOCH ₃	-OH	-H	0.876
DAVP236	-H	-COCH ₃	-OH	-H	0.854
DAVP238	-H	-COCl	-OH	-H	0.848
DAVP239	-H	-COOCOCH ₃	-OH	-H	0.898
DAVP240	-H	-F	-OH	-H	0.834
DAVP244	-H	-CHO	-N(CH ₃) ₂	-H	0.855
DAVP272	-H	-COOCOCH ₃	-I	-H	0.871
DAVP278	-H	-COOCH ₃	-Br	-H	0.871
DAVP283	-H	-COOCOCH ₃	-Br	-H	0.865
DAVP286	-H	-CN	-NH ₂	-H	0.834
DAVP287	-H	-COO	-NH ₂	-H	0.856
DAVP289	-H	-COOCH ₃	-NH ₂	-H	0.879
DAVP294	-H	-COOCOCH ₃	-NH ₂	-H	0.901
DAVP295	-H	-F	-NH ₂	-H	0.832

DAVP305	-H	-COOCOCH ₃	-CH ₃	-H	0.875
DAVP372	-OH	-H	-CN	-H	0.839
DAVP398	-OCH ₃	-H	-COOCH ₃	-H	0.860
DAVP399	-OH	-H	-COOCH ₃	-H	0.864
DAVP401	-Cl	-H	-COOCH ₃	-H	0.854
DAVP403	-Br	-H	-COOCH ₃	-H	0.855
DAVP404	-NH ₂	-H	-COOCH ₃	-H	0.876
DAVP405	-CH ₃	-H	-COOCH ₃	-H	0.849
DAVP406	-CH ₂ CH ₂ CH ₃	-H	-COOCH ₃	-H	0.896
DAVP417	-OH	-H	-COCH ₃	-H	0.850
DAVP435	-OH	-H	-COCl	-H	0.855
DAVP437	-Cl	-H	-COCl	-H	0.829
DAVP439	-Br	-H	-COCl	-H	0.830
DAVP443	-OCH ₃	-H	-COOCOCH ₃	-H	0.895
DAVP447	-I	-H	-COOCOCH ₃	-H	0.871
DAVP448	-Br	-H	-COOCOCH ₃	-H	0.870
DAVP451	-CH ₂ CH ₂ CH ₃	-H	-COOCOCH ₃	-H	0.912
DAVP497	-H	-OH	-CON(CH ₃) ₂	-OH	0.896
DAVP500	-H	-NH ₂	-CON(CH ₃) ₂	-NH ₂	0.876

El dato en rojo (DAVP42) es uno de los mejores compuestos encontrado por Bermúdez y colaboradores [121], mientras que los datos de azul son nuestros mejores resultados en función de la actividad biológica teórica.

En la Tabla 17 se muestran los sustituyentes electroattractores y electrodonadores en posición *orto*, *meta* y *para* con respecto al grupo arilo de la estructura de los 57 derivados del AVP, así como la actividad biológica teórica de cada uno de estos compuestos, las cuales se encuentran por arriba de $IC_{50} = 0.827 \mu M$, para reducir al 50 % la actividad de la HDAC8 (impedir que los grupos acetilo se separen de los nucleosomas), esto bajo condiciones específicas de reacción.

Como se comentó anteriormente Bermúdez y colaboradores, mostraron por un docking dirigido que los mejores derivados del AVP poseen en su estructura química, grupos carbonilos y grupos aminos, lo que concuerda con nuestros derivados propuestos por docking ciego y QSAR, pero en particular, el mejor compuesto para ellos fue el derivado del AVP con un grupo hidroxilo en posición *meta*, encontrándose este compuesto dentro de los 57 derivados del AVP, correspondiente al DAVP42. De esta manera se corrobora nuestros resultados.

8. RESUMEN DE RESULTADOS

- ✓ El sitio catalítico está conformado por los siguientes aminoácidos: Gly304, Aps178, Tyr306, His142, Zn²⁺, Phe152, His143, His180, Met274, Phe208, Gly151, Trp141 y Asp101.
- ✓ Los mejores compuestos obtenidos por el docking ciego fueron HID26, DAVP266, y DAVP281, ya que presentaron las constantes de disociación más bajas lo que indica una mayor afinidad con el sitio catalítico de inhibición.
- ✓ La ecuación matemática con la es posible encontrar la actividad biológica teórica de los derivados de AVP es:

$$IC_{50} = 0.541 + 0.22 \text{ Belm4}$$

- ✓ El mejor modelo matemático está en función del descriptor de DRAGON Belm4 (descriptor 2D-BCUT que toma en cuenta la carga de la molécula, la polarizabilidad y las propiedades de los enlaces de hidrógeno: grupo funcional aceptor y donador de hidrógenos) con respecto a 30 análogos del hidroxamato y 57 derivados del AVP, donde los mejores compuestos de los 57 derivados del AVP en relación a IC₅₀ teórica son: DAVP220, DAVP231 y DAVP437.

9 PERSPECTIVAS

- ✓ Realizar la dinámica molecular de los 57 derivados del AVP, con el objetivo de encontrar los mejores derivados del AVP.
- ✓ Sintetizar y caracterizar los mejores derivados del AVP, obtenidos en la dinámica molecular.
- ✓ Confirmar los resultados de la predicción de aquellos derivados que presentaron mejores actividades biológicas de manera experimental.

10. CONCLUSIÓN

En este estudio se identificó el sitio catalítico de inhibición (altamente conservado por residuos de aminoácidos aromáticos como la Tyr306, His142, Phe152, His143, His180, Phe208 y el Trp141) de 128 derivados del AVP con posible actividad antineoplásica sobre la HDAC8 por docking ciego, el cual coincide con el sitio catalítico reportado para la clase I de las HDAC's.

Por otra parte el análisis QSAR permitió encontrar el mejor modelo de RLM que describe la relación cuantitativa entre el efecto del inhibidor y las características químico-estructurales, entre los 30 análogos del hidroxamato (presentan actividad biológica experimental) con los 57 derivados del AVP, así como la ecuación matemática que predice la actividad biológica teórica de los derivados del AVP, en función del descriptor de DRAGON Belm4 que es un descriptor 2D-BCUT.

Como se ha mencionado, varios modelos matemáticos han tomado a los descriptores BCUT y han tenido éxito en diversos estudios de QSAR. En particular, el descriptor Belm4 ofrece la relación entre carga de la molécula, polarizabilidad y las propiedades de los enlaces de hidrógeno de los compuestos, con respecto al sitio catalítico. Se ha encontrado que los descriptores BCUT pueden proporcionar una información única sobre las estructuras moleculares y contribuir de manera significativa a las ecuaciones resultantes, actuando sobre blancos biológicos pertenecientes a proteínas. Aquí que nuestros modelos matemáticos son altamente predictivos e indican que los descriptores BCUT son útiles en estudios QSAR y deben ser ampliamente aplicada en el estudios de relación cuantitativa estructura actividad.

11. REFERENCIAS

1. <http://www.who.int/mediacentre/factsheets/fs297/es/> (Página consultada en julio de 2012).
2. Lopez J., Percharde M., Coley H. M., Webb A., Crook T. (2009). The context and potential of epigenetics in oncology. *Br. J. Cancer*. 100: 571-577.
3. Kennichi C. Dowdell, Lesley Pesnicak, Victoria Hoffmann, Kenneth Steadman, Alan T. Remaley, Jeffrey I. Cohen, Stephen E. Straus and V. Koneti Rao (2009). Valproic Acid (VPA), a Histone Deacetylase Inhibitor, Diminishes Lymphoproliferation in the Fas Deficient MRL/lpr^{-/-} Murine Model of Autoimmune Lymphoproliferative Syndrome. *Exp. Hematol*. 37: 487–494.
4. <http://www.espatentes.com/A61/2252519.html> (pagina consultada en Julio del 2012).
5. <http://www.inegi.org.mx/inegi/contenidos/espanol/prensa/contenidos/estadisticas/2011/cancer11.asp?s=inegi&c=2781&ep=51> (pagina consultada en Julio de 2012).
6. Ana Villar Garea (2005). Epigenetic transcriptional repression of tumor suppressor genes and its reversion by drugs. *UNIVERSITAT DE VALENCIA, Servei de Publicacions*.
7. Spandidos D. A. (2007). Oncogenes and tumor suppressor genes as paradigms in oncogenesis. *J. BUON*. 12 Suppl. 1: 9-12.
8. Vogelstein B., Kinzler K. W. (2004). Cancer genes and the pathways they control. *Nat. Med*. 10: 789-799.
9. Yoo C. B., Jones P. A. (2006). Epigenetic therapy of cancer: past, present and future. *Nat. Rev. Drug Discov*. 5: 37-50.
10. M. L. Martínez-Frías (2010). Estructura y función del ADN y de los genes II. Tipos de alteraciones de la función del gen por procesos epigenéticos. *Medicina de Familia Semergen*. 36: 332–335.
11. Monneret, Claude (2005). Histona deacetylase inhibitors. *Eur. J. Med. Chem*. 40: 1-13.
12. Biel M. Wascholowski V., Giannis A. (2005). Epigenetics, an epicenter of gene regulation: histones and histone-modifying enzymes. *Angew Chem Int Ed Engl*. 44: 3186-3216.
13. Marks P. A., Rifkind R. A., Richon. V. M., Breslow R., Miller T., Kelly W. K. (2001). Histone deacetylase and cancer: causes and therapies. *Nat. Rev. Cancer*. 1: 194-202.

14. Marks P. A., Rifkind R. A., Richon. V. M. (2000). Histone deacetylase inhibitors: inducers of differentiation or apoptosis of transformed cells. *J. Natl. Cancer Inst.* 1210-1216.
15. Ruijter A. J. M., Gennip A. H., Caron H. N. (2003). Histone deacetylases (HDAC's): characterization of the classical HDAC family. *Biochem. J.* 370: 737-749.
16. Chunaram Choudhary, Chanchal Kumar, Florian Gnad, Michael L. Nielsen, Michael Rehman, Tobias C. Walther, Jesper V. Olsen and Matthias Mann (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science.* 325: 834–840.
17. Yang X. J., Seto E. (2007). HATs and HDAC's: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene.* 26: 5310–5318.
18. <http://www.genome.jp/kegg/> (Página consultada en Julio de 2012).
19. Lina Pan, Jun Lu and Baiqu Huang (2007). HDAC Inhibitors: A Potential New Category of Anti-Tumor Agents. *Review, Cellular & Molecular Immunology.* 4:337-343.
20. Taunton J., Hassig C. A. and Schreiber S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* 272: 408-411.
21. Yang W. M., Yao Y. L., Sun J.M., Davie J.R. and Seto E. (1997). Isolation and characterization of cADNs corresponding to an additional member of the human histone deacetylase gene family. *J. Biol. Chem.* 272: 28001-28007.
22. Emiliani S., Fischle W., Van Lint C., Al-Abed Y. and Verdin E. (1998). Characterization of a human RPD3 ortholog, HDAC3. *Proc. Natl. Acad. Sci.* 95: 2795-2800.
23. Hu E., Chen Z., Fredrickson T., Zhu Y., Kirkpatrick R., Zhang G.F., Johanson K., Sung C. M., Liu R. and Winkler J. (2000). Cloning and characterization of a novel human class I histone deacetylase that functions as a transcription repressor. *J. Biol. Chem.* 275: 15254-15264.
24. Van den Wyngaert I., de Vries W., Kremer A., Neefs J., Verhasselt P., Luyten W. H., and Kass S. U. (2000). Cloning and characterization of human histone deacetylase 8. *FEBS Lett.* 478: 77-83.

25. Buggy J. J., Sideris M. L., Mak P., Lorimer D. D., McIntosh B. and Clark J.M. (2000). Cloning and characterization of a novel human histone deacetylase, HDAC8. *Biochem. J.* 350: 199-205.
26. Gao L., Cueto M. A., Asselbergs F. and Atadja P. (2002). Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family. *J. Biol. Chem.* 277: 25748-25755.
27. Grozinger C. M., Hassig C. A. and Schreiber S. L. (1999). Three proteins define a class of human histone deacetylases related to yeast Hda1p. *Proc. Natl. Acad. Sci.* 96: 4868-4873.
28. Fischle W., Emiliani S., Hendzel M. J., Nagase T., Nomura N., Voelter W. and Verdin E. (1999). A new family of human histone deacetylases related to *Saccharomyces cerevisiae* HDA1p. *J. Biol. Chem.* 274: 11713-11720.
29. Kao H. Y., Downes M., Ordentlich P. and Evans R. M. (2000). Isolation of a novel histone deacetylase reveals that class I and class II deacetylases promote SMRT-mediated repression. *Genes Dev.* 14: 55-66.
30. Zhou X., Marks P. A., Rifkind R. A. and Richon V. M. (2001). Cloning and characterization of a histone deacetylase, HDAC9. *Proc. Natl. Acad. Sci.* 98: 10572-10577.
31. Fischer D. D., Cai R., Bhatia U., Asselbergs F. A., Song C., Terry, R., Trogani N., Widmer R., Atadja P. and Cohen D. (2002). Isolation and characterization of a novel class II histone deacetylase, HDAC10. *J. Biol. Chem.* 277: 6656-6666.
32. Guardiola A. R. and Yao T. P. (2002). Molecular cloning and characterization of a novel histone deacetylase HDAC10. *J. Biol. Chem.* 277: 3350-356.
33. Kao H. Y., Lee C. H., Komarov A., Han C. C. and Evans R. M. (2002). Isolation and characterization of mammalian HDAC10, a novel histone deacetylase. *J. Biol. Chem.* 277: 187-193.
34. Tong J. J., Liu J., Bertos N. R. and Yang X. J. (2002). Identification of HDAC10, a novel class II human histone deacetylase containing a leucine-rich domain. *Nucleic Acids Res.* 30: 1114-1123.
35. Frye R. A. (1999). Characterization of five human cADNs with homology to the yeast SIR2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have protein ADPribosyltransferase activity. *Biochem. Biophys. Res. Commun.* 260: 273-279.

36. Imai S., Armstrong C. M., Kaeberlein M. and Guarente L. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nat.* 403: 795-800.
37. Landry J., Sutton A., Tafrov S. T., Heller R. C., Stebbins J., Pillus L. and Sternglanz R. (2000). The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases. *Proc. Natl. Acad. Sci.* 97: 5807-5811.
38. Bottomley M. J., Lo Surdo P., Di Giovine P., Cirillo A., Scarpelli R., Federica Ferrigno F., Jones P., Neddermann P., De Francesco R., Steinkuhler C., Gallinari P., Carfi A. (2008). Structural and functional analysis of the human HDAC4 catalytic domain reveals a regulatory structural Zinc-binding domain. *J. Biol. Chem.* 283: 26694–26704.
39. Ruijter A. J. M., Gennip A. H., Caron H. N. (2003). Histone deacetylases (HDAC's): characterization of the classical HDAC family. *Biochem. J.* 370: 737-749.
40. Lu Q., Yang Y. T., Chen C. S., Davis M., Byrd J. C., Etherton M. R., Umar A. (2004). Zn²⁺ chelating motif-tethered short-chain fatty acids as a novel class of histone deacetylase inhibitors. *J. Med. Chem.* 47: 467-474.
41. Yarnell A. (2006). Blocking genome gatekeepers. Scientists pin hopes on histone deacetylase inhibitors for cancer therapy biological studies. *Chem. Eng. News.* pp. 43-44.
42. Grozinger C. M., Shreiber S. L. (2002). Deacetylase enzymes: biological functions and the use of small molecules inhibitors. *Chem. Biol.* 9: 3-16.
43. Van Ommeslaeghe K., Loverix S., Geerlings P., Tourwe D. (2005). DFT-based ranking of zinc-binding groups in histone deacetylase inhibitors. *Bioorg. And Med. Chem.* 13: 6070-6082.
44. M. Yoshida, M. Kijima, M. Akita and T. Beppu (1990). T. Potent and Specific Inhibition of Mammalian Histone Deacetylase both In Vivo and In Vitro by Trichostatin A. *J. Biol. Chem.* 265: 17174–17179.
45. J. W. Han, S. H. Ahn, S. Y. Wang and G. U. Seo (2000). Apicidin, a histone deacetylase inhibitor, inhibits proliferation of tumor cells via induction of p21WAF1/Cip1 and gelsolin. *Cancer Res.* 60: 6068–6074.

46. M. Kijima, M. Yoshida, K. Suguta, S. Horinouchi and T. Beppu (1993). T. Trapoxin, an Antitumor Cyclic Tetrapeptide, Is an Irreversible Inhibitor of Mammalian Histone Deacetylase. *J. Biol. Chem.* 268: 22429–22435.
47. Gore S. D, Baylin S., Sugar E., Carraway H., Miller C. B., Carducci M., Grever M., Galm O., Dauses T., Karp J. E., Rudek M. A., Zhao M., Smith B. D., Manning J., Jiemjit A., Dover G., Mays A., Zweibel J., Murgo A., Weng L. J., Herman J. (2006). Combined DNA methyltransferase and histone deacetylation inhibition in the treatment of myeloid neoplasms. *Cancer Res.* 66: 6361-6369.
48. L. M. Butler, D. B. Agus, H. I. Scher, B. Higgins, A. Rose, C. Cordon-Cardo, H. T. Thaler, R. A. Rifkind, P. Marks and V. M. Richon (2000). Suberoylanilide hydroxamic acid, an inhibitor of histone deacetylase, suppresses the growth of prostate cancer cells in vitro and in vivo. *Cancer Res.* 60: 5165–5170.
49. Y. B. Kim, K. H. Lee, K. Sugita, M. Yoshida and S. Horinouchi (1999). Oxamflatin is a novel antitumor compound that inhibits mammalian histone deacetylase. *Oncogene.* 18: 2461–2470.
50. Johannessen C. U., Johannessen S. I. (2003). Valproate: past, present, and future. *CNS Drug Rev.* 9: 199-216.
51. M. Göttlicher, S. Minucci, P. Zhu, O. H. Kramer, A. Schimpf, S. Giavara, J. P. Sleeman, F. Lo Coco, C. Nervi, P. G. Pelicci and T. Hainzel (2001). Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *EMBO J.* 6969–6978.
52. DiGiuseppe J. A., Weng L. J., Yu K. H., Fu S., Kastan M. B., Samid D., Gore S. D. (1999). Phenylbutyrate-induced G1 arrest and apoptosis in myeloid leukemia cells: structure-function analysis. *Leukemia.* 13: 1243-1253.
53. Lavoie R., Bouchain G., Frechette S., Woo S. H., Khanlil E. A., Leit S. Fournel M., Yan P. T., Trachy-Bourget M. C., Beaukieu C., Li Z., Besterman J., Delorme D. (2001). Design and synthesis of a novel class of histone deacetylase inhibitors. *Bioorg. Med. Chem. Lett.* 11: 2847-2850.
54. Eyal S., Yagen B., Shimshoni J., Bialer M. (2005). Histone deacetylases inhibition and tumor cells cytotoxicity by CNS-active VPA constitutional isomers and derivatives. *Biochem. Pharmacol.* 69: 1501-1508.

55. M. S. Finnin, J. R. Dongian, A. Cohen, V. M. Richon, R. A. Rifkina, P. A. Marks, R. Breslowand, N. P. Pavletich (1999). Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nat.* 401: 188–193.
56. Hahnen E., Hauke J., Tränkle C., Eyüpoglu I. Y., Wirth B., Blümcke I. (2008). Histone deacetylase inhibitors: possible implications for neurodegenerative disorders. *Expert. Opin. Investig. Drugs.* 17: 169–184.
57. <http://news.bbc.co.uk/2/hi/health/6606315.stm> (pagina consultada en Julio del 2012).
58. Chapman A., Keane P. E., Meldrum B. S. (1982). Mechanism of anticonvulsant action of valproate. *Prog. Neurobiol.* 19: 315-359.
59. Davis R., Peters D. H., McTavish D. (1994). Valproic acid. A reappraisal of its pharmacological properties and clinical efficacy in epilepsy. *Drugs.* 47: 332-372.
60. Katzung, Bertram G. (2007). Basic & Clinical Pharmacology. *McGraw-Hill.* pp. 569-570.
61. Gurvich N., Berman M. G., Wittner B. S., Gentleman R. C., Klein P. S., Green J. B. (2005). Association of valproate-induced teratogenesis with histone de acetylase inhibition in vivo. *FASEBJ.* 19: 1166–1168.
62. Cinatl J. Jr., Cinatl J., Scholz M. (1996). Antitumor activity of sodium valproate in cultures of human neuroblastoma cells. *Anticancer Drugs.* 7: 766-773.
63. Cinatl J. Jr., Cinatl J., Driever P. H. (1997). Sodium valproate inhibits in vivo growth of human neuroblastoma cells. *Anticancer Drugs.* 8: 958-963.
64. Knupfer M. M., Hernaiz-Driever P., Poppenborg H. (1998). Valproic acid inhibits proliferation and changes expression of CD44 and CD56 of malignant glioma cells in vitro. *Anticancer Res.* 18: 3585-3589.
65. Knupfer M. M., Pulzer F., Schindler I. (2001). Different effects of valproic acid on proliferation and migration of malignant glioma cells in vitro. *Anticancer Res.* 21: 347-351.
66. Samid D., Shack S., Sherman L. T. (1992). Phenylacetate: a novel nontoxic inducer of tumor cell differentiation. *Cancer Res.* 52: 1988-1992.
67. Krämer O.H., Zhu P., Ostendorff H.P., Golebiewski M., Tiefenbach J., Peters M. A., Brill B., Groner B., Bach I., Heinzl T., Göttlicher M. (2003). The histone deacetylase

- inhibitor valproic acid selectively induces proteasomal degradation of HDAC2. *EMBO J.* 22: 3411-3420.
68. Paul A. Marks, Richard A. Rifkind, Victoria M. Richon and Ronald Breslow (2001). Inhibitors of histone deacetylase are potentially effective anticancer agents. *Cancer Res.* 7: 759-760.
69. Macritchie K. A., Geddes J. R., Scott J., Haslam D. R., Goodwin G. M. (2001). Valproic acid, valproate and divalproex in the maintenance treatment of bipolar disorder. *Cochrane Database Syst Rev.* 3: CD003196.
70. Jallon P., Picard F. (2001). Bodyweight Gain and Anticonvulsants: A Comparative Review. *Drug Saf.* 24: 969–978.
71. Cuevas G., Cortés F. (2003). Introducción a la química computacional. *Fondo de cultura económica.* México, DF. pp. 15-20.
72. Jensen F. (1999). Introduction to Computational Chemistry. *John Wiley & sons.* Inglaterra. pp. 6-29.
73. Grand G. H., Richards W. G. (1995). Computational chemistry. *Oxford science publications. Universidad de Oxford.* pp. 1-3, 32-45.
74. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman (1995). A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* 117: 5179.
75. S. L. Mayo, B. D. Olafson and W. A. Goddard III. (1990). DREIDING: A Generic Force Field for Molecular Simulations. *J. Phys. Chem.* 94: 8897.
76. A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard III and W. M. Skiff (1992). UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* 114: 10024.
77. Mari Inés Nicolás Vázquez, Erick Marín Chiñas, Francisco Miguel Castro Martínez, René Miranda Ruvalcaba (2006). Algunos aspectos básicos de la química computacional. *Fondo de cultura económica.* México, DF. pp. 21-23.
78. <http://mgl.scripps.edu/people/goodsell/> (pagina consultada en julio del 2012).
79. Lengauer T., Rarey M. (1996). Los métodos computacionales para el atraque biomolecular. *Curr. Opin. Estructura. Biol.* 6: 402-406.

80. Cocina D.B., Decornez H., Furr J.R., Bajorath J. (2004). Docking y la puntuación en la detección virtual para el descubrimiento de fármacos: métodos y aplicaciones. *Nat. Reviews*. 8: 402
81. R. Hoffmann (1963). An Extended Hückel Theory. I. Hydrocarbons. *J. Chem. Phys.* 39: 1397.
82. M. Dewar and W. Thiel (1998). Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.* 99: 4499.
83. G. Segal and J. Pople (1966). Application of the pople-santry-segal CNDO method to the cyclopropylcarbinyl and cyclobutyl cation and to bicyclobutane. *J. Chem. Phys.* 44: 3289.
84. J. A. Pople, D. Beveridge and P. Dobosh (1967). Approximate Self-consistent Molecular Orbital Theory V. Intermediate Neglect of Differential Overlap. *J. Chem. Phys.* 47: 2026.
85. J. J. P. Stewart (1989). Optimization of parameters for semiempirical methods I. Method. *J. Comp. Chem.* 10: 209.
86. Dres. D. Murgida y E. Marceca (2005). Tópicos de fisicoquímica en sistemas biológicos: Métodos de estructura electrónica. Guía de Trabajos Prácticos. pp. 5
87. Ira N. Levine (2001). Química Cuántica. Quinta edición. *Pearson Educación, S. A.* Madrid. pp. 11-13, 45-48.
88. Kenneth Krane (1991). Física moderna. *McGraw-Hill*. México, DF. pp.21-23, 141-145.
89. Hernán Estrada Bustos (2005). Física computacional. *Universidad Nacional de Colombia*. Bogotá, Colombia. pp.46-48.
90. Cuevas G., Cortés F. (2003). Introducción a la química computacional. *Fondo de cultura económica*. México, DF. pp. 47.
91. María Font (2001). Mecánica cuántica: Química Computacional Aplicada. Departamento de Química Orgánica y Farmacéutica, Sección de Modelización Molecular. *Universidad de Navarra*. pp. 35.
92. <http://es.scribd.com/doc/57968541/Principios-Quimica-Computacional-3-Edicion> (pagina consultada en Julio de 2012).
93. Hansch C.; Fujita T. (1964). A new substituent constant, π , derived from partition coefficients. *J. Amer. Chem. Soc.* 86, 1616.

94. Antonio Delgado Cirilo, Cristina Minguillón Llombart, Jesús Joglar Tamargo (2003). Introducción a la química terapéutica. *Díaz de Santos*. Madrid España. pp. 143,144.
95. Hansch, C.; Leo, A. (1995). In Exploring QSAR: Fundamental and application in chemistry and biology. *American Chemical Society*.
96. Richard D. Cramer III, David E. Patterson and Jeffrey D. Bunce (1988). Comparative Molecular Field Analysis (COMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* 110: 5959.
97. Klebe G., Abraham U., Mietzner T. (1994). Comparative Molecular Similarity Indices Analysis of 1-(Naphthylalkyl)-1H-imidazole Analogs with Antiepileptic Activity. *J. Med. Chem.* 37: 4130.
98. Ramírez-Galicia G., Garduño-Juárez R., Correa-Basurto J., Deeb O. (2012) Exploring QSAR for inhibitory effect a set of heterocyclic thrombin inhibitors by multilinear regression refined by artificial neural network and molecular docking simulations. *J. Enzyme Inhib. Med. Chem.* 27: 174-186.
99. Guillermo Ramírez-Galicia, Heidy Martínez-Pacheco, Ramón Garduño-Juárez y Omar Deeb (2012). Exploring QSAR of antiamebic agents of isolated natural products by MLR, ANN, and RTO. *Med. Chem. Res.* 21: 2501–2516.
100. Alfonso Velasco Martín (2001). Compendio de farmacología general. *Díaz de Santos*. Madrid España. pp. 155-157.
101. Daniel W. Nebert y Ross A. McKinnon (2001). 33. Toxicología; determinantes genéticos de la respuesta tóxica. *Enciclopedia de salud y seguridad en el trabajo*. pp. 33.21-33.24.
102. Dr. Gustavo Aldereguía Lima (2007). Temas para la asistencia clínica al adulto mayor. *Revista Electrónica de las Ciencias Médicas en Cienfuegos*. 5:1
103. Dhanshi C. Juvale, Vishal V. Kulkarni, Hemantkumar S. Deokar, Nilesh K. Wagh, Subhash B. Padhye and Vithal M. Kulkarni (2006). 3D-QSAR of histone deacetylase inhibitors: hydroxamate analogues. *Org. Biomol. Chem.* 4: 2858-2868.
104. PyMOL (TM) Educational Product - Copyright (C) 2010 Schrodinger, LLC.
105. P. F. Flukiger (1992). Development of the molecular graphics package MOLEKEL and its application to selected problems in organic and organometallic chemistry. Theses No 2561, *Department de chimie physique, University de Geneve, Geneve*.

106. Morris G. M. D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.* 19: 1639.
107. GaussView 3.0, Gaussian, Inc. Carnegie Office Park-Building 6, Pittsburgh PA 15106 USA. Copyright (c) 2000-2003. Semichem, Inc.
108. Frisch, M. J. Trucks, G. W. H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Aprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. ennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Peterson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, J. A. Pople (2003). Gaussian 03, Revision E.01, Gaussian, Inc, Pittsburgh, PA.
109. Caballero J., Fernández M. (2008). Artificial neural networks from MATLAB in medicinal chemistry. Bayesian-regularized genetic neural networks (BRGNN). Application to the prediction of antagonist activity against human platelet thrombin receptor (PAR-1). *Curr Top Med Chem.* 8: 1580–1605.
110. Todeschini R., Consonni V., Pavan M. (2002) Dragon Software version 2.1-2002 Pisani 13, Milano. Dragon Software and references therein.
111. Origin 7 SR2 v7.0383 (B383) Copyright (C) 1991-2002 OriginLab Corporation, Northampton, MA 01060 USA.
112. Babel version 1.3 Copyright (C) 1992-1996 by Patt Walters and Matt Stahl. babel @ mercury.aichem.arizona.edu.
113. Bermudez-Lugo J. A., Correa-Basurto J. (2009). Diseño de Inhibidores de Acetilcolinesterasa Asistido por Métodos Computacionales: Modelado por Homología, Docking y Dinámica Molecular. *Tesis de licenciatura. ESM-IPN.*
114. <http://www.rcsb.org/pdb/explore/explore.do?structureId=3F07> (pagina consultada en Julio del 2011).

115. Kunal Roy, Indrani Mitra, Supratik Kar, Probir Kumar Ojha, Rudra Narayan Das and Humayun Kabir (2011). Comparative Studies on Some Metrics for External Validation of QSPR Models. *J. Chem. Inf. Model.* 52: 396–408.
116. Guillermo Ramirez-Galicia, Ramon Garduño-Juarez, Bahram Hemmateenejad, Omar Deeb and Samuel Estrada-Soto (2007). QSAR Study on the Relaxant Agents from Some Mexican Medicinal Plants and Synthetic Related Organic Compounds. *Chem. Biol. Drug Des.* 70: 143–153.
117. Javier García, Pablo R. Duchowicz, Eduardo A. Castro (2005). Consideración de la flexibilidad conformacional molecular en los estudios QSAR. *División Química Teórica, CONICET-Universidad Nacional de La Plata.* pp. 5-10.
118. Xavier Basoqain Olabe. Redes neuronales artificiales y sus aplicaciones (1998). Departamento de ingeniería de sistemas y automática. *Escuela superior de ingeniería de Bilbao.* EHU. pp. 2-6.
119. Pedro Isasi Viñuela, Inés M. Galván León (2003). Redes de Neuronas Artificiales. Un enfoque Práctico. *Pearson Prentice Hall.* pp. 2-14.
120. Bonifacio Martín del Brío, Alfredo Sanz Molina (2001). Redes Neuronales y Sistemas Borrosos. *ALfaomega Ra-Ma.* pp. 2-10.
121. Bermudez-Lugo J. A., Correa-Basurto J. (2011). Síntesis y evaluación *in silico* e *in vitro* de derivados del ácido valproico como antineoplásicos. *Tesis de maestría. ESM-IPN.*
122. Akaike H. (1974). A new look at the statistical model identification. *IEEE Trans Automat. Contr.* AC-19: 716-723.
123. Kubinyi H. (1994). Variable selection in QSAR studies: 2. A highly efficient combination of systematic search and evolution. *Quant. Struct.-Act. Relatsh.* 13: 393-401.
124. Hill, T., Lewicki, P. (2006). Statistics Methods and Applications. *Statsoft, Tulsa, OK.*
125. Golbraikh, A., Tropsha, A. (2002). Beware of q^2 !. *J. Mol. Graph. Model.* 20: 269-276.
126. Hawkins, D. M. (2004). The problem of overfitting. *J. Chem. Inf. Comput. Sci.* pp. 1-12.
127. Stanton D. T. (1999). Evaluation and use of BCUT descriptors in QSAR and QSPR studies. *J Chem Inf Comput Sci.* 39:11–20.

128. Pirard B., Pickett S. (2000). Classification of kinase inhibitors using BCUT descriptors. *J Chem Inf Comput Sci.* 40:1431–1440.
129. Gao H. (2001). Application of BCUTmetrics and genetic algorithm in binary QSAR analysis. *J Chem Inf Comput Sci.* 41:402–407.
130. Ford M., Pitt W., Whitley D. (2004). Selecting compounds for focused screening using linear discriminant analysis and artificial neural networks. *J Mol Graph Model.* 22:467–472.
131. Ming Hao, Yan Li, Yonghua Wang and Shuwei Zhang (2011). A classification study of human β 3-adrenergic receptor agonists using BCUT descriptors. *Mol Divers.* 15:877–887