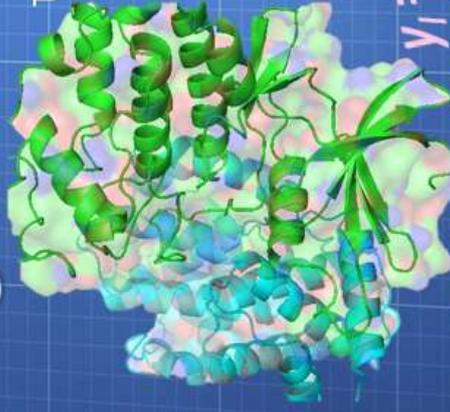
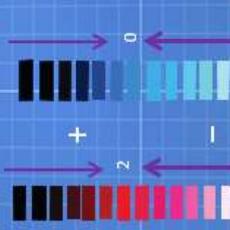


Análisis de semejanza cuántica molecular de dos series de compuestos: catecolaminas e indenopirazoles

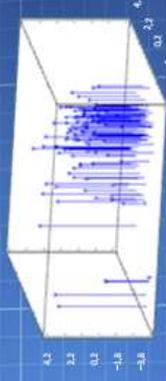
Juan Enrique Torres Castellar



Tutor: Ricardo Vivas



$$y_l = \langle \Omega \rangle \approx \beta^T f(z)$$



$$\hat{H} |\psi(t)\rangle = i\hbar \frac{\partial}{\partial x} |\psi(t)\rangle$$
$$\alpha \pm \beta$$

Requisito parcial para optar por el título de doctor en Ciencias biomédicas
2017

ANÁLISIS DE SEMEJANZA CUÁNTICA MOLECULAR
APLICADA A DOS SERIES DE COMPUESTOS:
MONOAMINAS E INDENOPIRAZOLES



JUAN ENRIQUE TORRES CASTELLAR

2017

**ANÁLISIS DE SEMEJANZA CUÁNTICA MOLECULAR
APLICADA A DOS SERIES DE COMPUESTOS:
MONOAMINAS E INDENOPIRAZOLES**



Juan Enrique Torres Castellar

**Tesis como requisito parcial para obtener el Título de
Doctor en Ciencias Biomédicas**

Tutor: RICARDO VIVAS REYES

Cotutores:

ROSA BALDIRIS AVILA

DARIO MENDEZ CUADRO

2017

NOTA DE APROBACIÓN DEL JURADO

PRESIDENTE DEL JURADO

JURADO

JURADO

JURADO

CARTAGENA DE INDIAS, 2017

La Universidad de Cartagena ni el jurado examinador, se hacen responsables de los conceptos emitidos en el presente trabajo.

CARTAGENA DE INDIAS, 2017

CONTENIDO

	RESUMEN.....	I
	SUMMARY.....	II
	CAPÍTULO I.....	1
1	INTRODUCCIÓN.....	1
1.1	DESCRIPCIÓN DEL PROBLEMA.....	1
1.2	REFERENCIAS DEL CAPÍTULO I.....	8
	CAPÍTULO II.....	11
2	BASES BIOQUÍMICAS DEL CÁNCER.....	11
2.1	Mecanismos reguladores del ciclo celular.....	11
2.1.1	La cinasa dependiente de ciclina 2.....	14
2.2	LA SEÑALIZACIÓN β -ADRENÉRGICA Y EL CÁNCER.....	18
2.3	REFERENCIAS.....	20
	CAPÍTULO III.....	22
3	BIOFÍSICA DE LA INHIBICIÓN DE ENZIMAS.....	22
3.1	EL ESTADO DE TRANSICIÓN.....	22
3.1.1	Formulación cuantitativa del postulado de Hammond.....	25
3.1.2	Aplicación del ET a la catálisis enzimática.....	26
3.2	DINÁMICA CONFORMACIONAL DE UNA PROTEÍNA EN LA CATÁLISIS.....	31
3.3	REFERENCIAS.....	35
	CAPÍTULO IV.....	37
4	SEMEJANZA CUÁNTICA.....	37
4.1	La mecánica cuántica.....	37
4.1.2	Postulados de la mecánica cuántica.....	38
4.1.3	La ecuación de Schrödinger.....	39
4.1.4	Funcionales de densidad DF.....	41
4.1.5	Los principios de Hohenberg–Kohn.....	43
4.1.6	Valores de expectación.....	47
4.2	Medición de la Semejanza Cuántica Molecular.....	48
4.2.1	Aproximación de la Cáscara Atómica (ASA).....	50
4.2.2	Matriz de Semejanza.....	52
4.3	REFERENCIAS.....	53
	CAPÍTULO V.....	55
5	QSAR.....	55
5.1	El modelo de Hammett.....	55
5.2	Algoritmos de regresión.....	58
5.3	Validación estadística.....	62
5.4	Ecuación Fundamental del QQSAR.....	64
5.5	REFERENCIAS.....	65
	CAPÍTULO VI.....	67

6	QSAR CUÁNTICO DE INDENOPIRAZOLES ANTAGONISTAS DE LA CDK2.....	68
6.1	METODOLOGÍA.....	
6.1.1	Acoplamiento molecular y Optimización.....	68
6.2	Desarrollo y validación de un QQSAR.....	70
6.3	RESULTADOS.....	71
6.3.1	Interacciones Cinasa-Indenopirazol.....	71
6.3.2	Análisis por PLS del QQSAR.....	74
6.3.3	CONCLUSIONES CAPÍTULO VI.....	78
6.4	REFERENCIAS.....	81
	CAPÍTULO VII.....	86
7	ESCALA DE CONVERGENCIA DE LAS CATECOLAMINAS.....	86
7.1	Fisiología de las catecolaminas.....	86
7.2	Metodología.....	87
7.2.1	Alineamiento de las estructuras de las aminas.....	87
7.2.2	Diseño de moléculas, optimización de la geometría y cálculo de los descriptores moleculares.....	89
7.2.3	Índices de Semejanza.....	70
7.2.4	Índices Alfa y beta propuestos.....	71
7.3	RESULTADOS.....	94
7.4	CONCLUSIONES.....	97
7.5	REFERENCIAS.....	106
	ANEXO.....	116

RESUMEN

Las mediciones de semejanza cuántica son un vehículo natural y no sesgado para obtener parámetros moleculares que se utilizan como descriptores en distintos problemas químicos. Estos descriptores tienen un fundamento geométrico que viene de consideraciones que derivan en últimas los postulados de la mecánica cuántica. Sus valores surgen objetivamente de la naturaleza del conjunto de moléculas estudiadas. En esta tesis, se logró calcular mediciones de semejanza cuántica dentro de un esquema rápido, aproximado pero suficientemente preciso, basado en el ajuste de la función de densidad a un conjunto de bases atómicas de simetría esférica, mediante la llamada aproximación de la Capa Atómica. Esto permitió obtener varios descriptores para dos conjuntos de compuestos que han atraído mucho interés en el campo del desarrollo de agentes contra el cáncer.

De un lado, fueron calculados índices de semejanza de compuestos deducidos a partir de consideraciones en el marco de la semejanza cuántica y se correlacionaron con sus valores de actividad biológica. La evaluación del modelo de regresión construido dio un coeficiente de predicción $q^2 = 0.89$ para siete componentes, lo cual implica que el modelo QSAR fue exitoso. La diana de esta serie de ligandos derivados del indenopirazol es la enzima CDK2, la cual es un regulador crucial del ciclo celular. De otro lado, recientemente también las catecolaminas han sido involucradas en la progresión de los tumores. Por lo tanto, se llevó a cabo un análisis de semejanza cuántica de las catecolaminas, usando cuatro descriptores de semejanza. A partir de estos descriptores, fueron definidos dos nuevos parámetros cuantitativos de convergencia llamados índices alfa y beta, los cuales podrían ser útiles en el estudio del estado de transición de la ruta sintética mediante la interpretación del postulado de Hammond introducido por Cioslowski.

Palabras clave: Cinasas, Catecholaminas, QSAR, semejanza cuántica.

ABSTRACT

Quantum similarity measures are a natural and unbiased vehicle to obtain molecular parameters acting as descriptors that can be connected to all fields of chemistry. These descriptors are based on purely geometrical grounds, as they are a consequence of considerations which stem from the application of quantum-mechanical postulates. Their values arise from the nature of the studied system and are not influenced by *a priori* design implemented by the researcher. In this thesis, the calculation of several quantum similarity measures were achieved in a fast, approximate although highly accurate framework, based on density function fitting to a spherically symmetric atomic basis set, by means of the so-called atomic shell approximation. This allowed to obtain several descriptors for two sets of compounds which have attracted the interest in the area of development of anticancer agents.

For one, they were calculated similarity indices of indenopyrazole derivatives deduced from considerations attached to the molecular quantum similarity framework, and they were correlated to their biological activities. Assessment built regression model yielded a squared coefficient of prediction $q^2 = 0.89$ for seven components, which means that a successful QSAR model has been developed. This series of indenopyrazol ligands targets the CDK2 enzyme, which is a key regulator of the cell cycle. On the other side, catecholamines have been involved in the tumor progression as well. Therefore, it was performed a quantum molecular similarity analysis of catecholamines, using four similarity descriptors: the molecular quantum similarity index of overlap, the similarity index of Coulomb, and the corresponding electronic Euclidean distances overlap and Coulomb, respectively. From these descriptors, it were defined a couple of new quantitative convergence parameters named alpha and beta indices, which are proposed to be useful in the studies of the transition state of the synthesis reaction paths by means of the Hammond postulate interpretation introduced by Cioslowski.

Key words: Cyclin Dependent Kinase 2, Catecholamines, QSAR, quantum similarity.

CAPÍTULO I

INTRODUCCIÓN

1.1 Descripción del problema

Desde un punto de vista reduccionista, la química es la ciencia que deriva de la física de los electrones de valencia. Adicionalmente, la química teórica es una rama de la química que desarrolla herramientas matemáticas basadas en las leyes fundamentales de la física para modelar la reactividad y otras propiedades de las moléculas o agregados de átomos en los que el enlace que los une tiene suficiente estabilidad para ser consideradas como una especie discreta. En tanto que la química computacional se enfoca en la obtención de resultados relevantes para problemas químicos, no directamente en el desarrollo de métodos teóricos [1]. La interacción entre la química teórica y la química computacional es fuerte y evidente. El desarrollo de nuevos métodos, posibilita el estudio de nuevos problemas [2]. En ese sentido, los métodos de medición de la semejanza cuántica permiten abordar sistemas con un número de partículas suficientemente grande para tener interés en las ciencias biomédicas. Por ejemplo es posible cuantificar la semejanza entre un inhibidor enzimático y el estado de transición del sustrato nativo de la enzima.

La presente tesis doctoral es una investigación enmarcada en la química computacional que tiene como objetivo desarrollar computacionalmente índices de semejanza y modelos estadísticos que sean útiles en el desarrollo de ligandos que actúen sobre dianas terapéuticas involucradas en el cáncer. Se

han incluido secciones donde se introducen conceptos propios de las heterogéneas disciplinas en las que se apoyó la investigación. Este documento consta de los siguientes capítulos:

Capítulo 1, en el cual se presenta el planteamiento del problema, los objetivos de la investigación, la justificación, los alcances y limitaciones de la misma, además de una exposición general del marco teórico referida a la biofísica de la inhibición de enzimas y el postulado de Hammond-Leffler así como los enfoques y métodos computacionales empleados.

Capítulo 2, en el cual se presenta una descripción de varios de los circuitos bioquímicos más importantes involucrados en la patogenia del cáncer, especialmente los que implican a las CDK2. También se resalta el papel en la evolución del cáncer recientemente documentado en la literatura para las catecolaminas.

Capítulo 3, que está centrado en los aspectos termodinámicos de la inhibición enzimática. Incluye la comparación las barreras energéticas en dependencia de las entalpías, profundiza en la formulación cuantitativa del postulado de Hammond y muestra como la ocurrencia de la catálisis enzimática se explica a través del enlace de la enzima al estado de transición más fuertemente que cuando interacciona con el estado basal de los reactantes.

Capítulo 4, en el cual se presenta una exposición de la teoría de la semejanza cuántica que parte de una descripción de los postulados de la mecánica cuántica y los funcionales de densidad, hasta los fundamentos del enfoque conocido como Aproximación de la Cáscara Atómica (ASA) que permitió realizar los cálculos de los índices de semejanza a un costo computacional razonable.

Capítulo 5, en el cual se presentan las bases teóricas de la metodología QSAR incluyendo su conexión con la teoría de la semejanza cuántica y

también las bases matemáticas de la Proyección de las Variables Latentes o PLS que fue el método de regresión implementado en la validación del modelo.

Capítulo 6, en el cual se presenta un estudio llevado a cabo por el autor de esta tesis concerniente a los derivados indenopirazólicos que según la literatura muestran actividad inhibitoria frente a las enzimas cinasas dependientes de ciclina CDK2 y CDK4.

Capítulo 7, en el cual se presenta un estudio concerniente a las catecolaminas las cuales son un grupo de compuestos involucrados en el desarrollo del cáncer.

En el desarrollo de esta tesis se abordó el estudio de dos series de compuestos de interés en el cáncer. En una primera de línea, se aplicó la semejanza cuántica para generar descriptores que a continuación permitieron elaborar un modelo de Relación Cuantitativa Estructura Actividad (QSAR). Un descriptor es un rasgo molecular que se emplea en el modelo estadístico para predecir una propiedad fisicoquímica o biológica. La hipótesis de trabajo fue la suposición de que las variables latentes de un grupo de descriptores obtenidos a partir de la semejanza cuántica podrían predecir de manera robusta la afinidad de los inhibidores indenopirazol por la CDK2. A partir de la química básica es claro que tiene que haber una conexión entre la estructura de una molécula y su actividad biológica. No obstante, la conexión es muy compleja y es muy improbable que una propiedad como el IC_{50} pueda ser alguna vez predicha completamente *a priori* [3]. La concentración inhibitoria media (IC_{50}) corresponde en este caso a una medida de la efectividad de un compuesto en la inhibición de la función de transferencia de fosfatos de la enzima CDK2.

Un componente esencial de los cálculos es calibrar nuevos métodos y utilizar los resultados de los cálculos para predecir o racionalizar la información que arrojan los experimentos. En ambos tipos de investigación se comparan dos conjuntos de datos y el interés está en caracterizar qué tan bien uno de los conjuntos de datos puede representar o predecir el otro. Afortunadamente, uno

o ambos conjuntos de datos usualmente contiene “ruido”, y puede dificultar decidir si una correlación pobre es debido a lo ruidoso de los datos o a una falta fundamental de conexión. La estadística es una herramienta para cuantificar tales relaciones [4]. La calibración externa del QSAR mostró que el modelo podía predecir la actividad de antagonistas de la enzima CDK2 con un q^2 más alto que el obtenido mediante un estudio CoMFA. Con esta ecuación sería posible guiar de manera económica la investigación básica de compuestos con actividad anticancerígena dirigidos a la diana en mención. Toda vez que la evaluación experimental de compuestos sintéticos es enormemente costosa [5]. Los datos para construir el modelo QSAR de los indenopirazoles se basó en los valores de actividad biológica *in vitro* que fueron tomados de varios artículos que reportaron las IC_{50} para la inhibición de la CDK2 y la CDK4 correspondientes a derivados del indenopirazol [6–10] y correspondían a moléculas ensayadas bajo las mismas condiciones experimentales.

En general, las cinasas constituyen una de las familias más importantes de dianas, dando cuenta de 20–30% de los programas de descubrimiento de fármacos de las farmacéuticas, superadas sólo por los receptores acoplados a proteína G [11]. Las enzimas CDK1, CDK2, CDK4 y CDK6 y sus ciclinas asociadas A, B, D, E pueden ser consideradas *bona fide* reguladoras del ciclo celular. La Inhibición de las actividades esenciales de la CDK2 puede producir apoptosis en células cancerosas, pero sólo arresto reversible del ciclo celular en células normales. De aquí que los inhibidores de la CDK2 tienen potencial como agentes contra el cáncer con una buena ventana terapéutica [12, 13]. En consecuencia, el trabajo teórico presentado en esta propuesta de investigación, es un enfoque alternativo para estudiar la interacción CDK2-indenopirazoles aplicando descriptores de semejanza cuántica. En la teoría de Eyring se plantea que las reacciones químicas transcurren a través de un *estado de transición*, en el cual los enlaces en el sustrato que se van a romper y los enlaces en el producto que se van a formar se hallan parcialmente rotos y formados, respectivamente [14]. En el estado de transición, el sistema se halla en su energía más alta a lo largo de la coordenada de reacción [15].

También en la literatura se presenta un número creciente de publicaciones acerca del impacto de las catecolaminas en la progresión de los tumores. Debido al interés científico creciente sobre estos compuestos, se emplearon medidas de semejanza cuántica molecular y descriptores basados en la Teoría del Funcional de Densidad para proponer una metodología que definiera cuantitativamente grupos individuales de estas monoaminas. Todas las transformaciones químicas pasan a través de una estructura cuya energía la hace inestable llamada el estado de transición y su configuración representa una transición entre las estructuras químicas de los sustratos y los productos [16].

En la segunda línea de la presente tesis se utilizó también la teoría de la semejanza cuántica, esta vez con el objetivo de proponer una convergencia de escalas cuantitativa que permitiera definir las catecolaminas como grupo de compuestos. Las catecolaminas son objeto de gran interés por su papel en varios procesos fisiológicos. La literatura científica presenta un número creciente de publicaciones acerca del impacto de las catecolaminas en la progresión de los tumores [17] que estaría mediado por el efecto de esta familia de hormonas en la señalización β -adrenérgica [18]. En consecuencia, se emplearon descriptores basados en la Teoría del Funcional de Densidad para proponer una metodología que definiera cuantitativamente grupos individuales de monoaminas. Se respaldaron las interpretaciones presentadas con resultados experimentales de la ruta de la reacción de síntesis. El nuevo índice de semejanza propuesto que caracteriza las catecolaminas, aparte de sus aplicaciones en el diseño racional de fármacos puede ser útil en correlacionar reactividades de grupos funcionales en sistemas orgánicos, en la comparación de moléculas semejanza o en la interpretación computacional de espectros IR.

En sus esfuerzos para entender y sistematizar los mecanismos de las reacciones químicas, frecuentemente los investigadores recurren a varias correlaciones que involucran estructuras y energías de reactantes, productos y estados de transición. El estado de transición es un arreglo inestable de átomos en los cuales los enlaces químicos están en el proceso de formación y ruptura. Sirven como ejemplos las relaciones de Hammett y Brønsted en las

cuales se utilizan cantidades bien definidas como son las energías de Gibbs y las constantes de disociación [19]. En contraste, las correlaciones ampliamente utilizadas con base en el postulado de Hammond si bien son atractivas por su carácter intuitivo, son de naturaleza cualitativa. Esto deriva parcialmente de los conceptos vagamente definidos de estados de transición tempranos y tardíos. Es obvio que, a fin de derivar una formulación más cuantitativa del postulado de Hammond, primero se deben definir rigurosamente los conceptos de exotermicidad y el progreso a lo largo de la coordenada de reacción para determinar si el estado de transición en cuestión es de carácter temprano o tardío. Se profundizará sobre este tópico en el capítulo 3.

La justificación de una investigación sobre diseño racional de ligandos con posible actividad anticancerígena es que con el envejecimiento de la población y de la creciente exposición a los factores de riesgo, la morbilidad y mortalidad por cáncer en Colombia ha aumentado en los últimos años con el consiguiente impacto social y económico para los servicios de salud. De manera similar, en el mundo el cáncer se ubica junto con las enfermedades cardiovasculares entre las causas líderes de muerte [20]. Por lo anterior se consideró importante estudiar dos series de compuestos relacionados con el cáncer. De un lado, los derivados del indenopirazol los cuales tienen potencial como agentes anticancerosos con una ventana terapéutica amplia por ser antagonistas de una enzima comprometida en el control del ciclo celular. De otro lado, las catecolaminas, que recientemente han atraído el interés de los investigadores al encontrarse evidencia que las asocian también con algunos tipos de tumor. Así mismo, el estrés emocional que es un rasgo característico de la modernidad y a la vez un factor asociado con el cáncer, es mediado también por hormonas de la familia de las catecolaminas [21].

Los químicos, farmacólogos, toxicólogos y farmacéuticos habitualmente discuten y clasifican la respuesta biológica de los alimentos, fármacos, cosméticos y xenobióticos empleando medidas intuitivas de semejanza molecular. Por ejemplo, se piensa a menudo que el aroma de las fragancias obedece a la forma molecular. En ese mismo sentido, los químicos orgánicos constantemente emplean el concepto de semejanza molecular para aludir a las

nociones de grupos funcionales y sintones [22]. Sin embargo no se había definido mediante un índice matemático explícito a las catecolaminas como un grupo separado. También son escasos los esfuerzos en la literatura por utilizar los índices de semejanza en la construcción de modelos que predigan la actividad de los ligandos contra una diana biológica relevante en el desarrollo del cáncer. Estas razones motivaron los estudios presentados en este trabajo, los cuales pudieron implementarse gracias a que los métodos computacionales requeridos son factibles de llevar a cabo mediante avances recientes.

Las propiedades moleculares que tienen definiciones teóricas rigurosas, como la energía y el momento multipolar, son fácilmente accesibles mediante enfoques teóricos. En cambio, propiedades de naturaleza más intuitiva, como la reactividad, aromaticidad o semejanza, han sido más difíciles de definir cuantitativamente. Es natural suponer que una definición cuantitativa adecuada para la semejanza entre las moléculas esté formulada en el contexto formal de la química cuántica, dado que solo los potenciales cuánticos son consistentes con las leyes físicas que subyacen a problemas químicos como las propiedades, estructuras y reacciones de las moléculas. La autoría de la primera medida cuantitativa de la semejanza cuántica recae en Carbó quien desarrolló una expresión numérica basada en la superposición de las densidades de carga electrónica de dos moléculas [23]. Sus contribuciones a los métodos de relación estructura-actividad con descriptores cuánticos constituyen parte medular del contenido del presente documento.

En esa línea, Cioslowski propuso un índice de semejanza relacionado con el número de electrones en el fragmento molecular común de ambas moléculas. Este índice denominado NOEL depende de la topología de las moléculas comparadas y puede constituir una herramienta para sistematizar las propiedades de los sistemas químicos de una manera rigurosa. Además de la topología, el otro enfoque para definir medidas cuantitativas de semejanza molecular es la mecánica cuántica. La mecánica cuántica ofrece la mejor herramienta para modelar el comportamiento de la materia, especialmente a escala ultramicroscópica, mediante la resolución de la ecuación de Schrödinger [24]. Pero aún en su aproximación más cruda, el empleo de la mecánica

cuántica se limita a sistemas con centenares de átomos en diferentes configuraciones, lo cual es insuficiente para muestrear el espacio de fases entero de un sistema biológico. Para llevarla al terreno de aplicaciones biomédicas, además de las simplificaciones adicionales acostumbradas, se han ensayado enfoques como el de la semejanza cuántica calculada a partir de Aproximación de la Cáscara Atómica, ASA [25].

1.2 REFERENCIAS

1. Cramer, CJ. Essentials of computational chemistry. Theories and Models. 2a Edición. Ed. Wiley. 2004. Pag 4.
2. Weymuth, T. & Reiher, M. Inverse quantum chemistry: Concepts and strategies for rational compound design. (2014). *Int. J. Quantum Chem.* **114** (13), 823–837.
3. Jensen, F. Introduction to Computational Chemistry. 2a Ed. Wiley. 2007. Pag 559.
4. Dehmer, M., Varmuza, K. & Bonchev, D. Statistical Modelling of Molecular Descriptors in QSAR/QSPR. Primera Edición. 2012 Wiley-VCH Verlag GmbH & Co. KGaA. 2012. Pag 7.
5. Young, DC. Computational drug design. A guide for Computational and medicinal chemists. Ed. Wiley. 2009. Pag 2.
6. Nugiel, DA., Etkorn, AM., Vidwans, A., Benfield, PA., Boisclair, M., Burton, CR., Cox, S., Czerniak, PM., Doleniak, D. & Seitz, SP. (2001). Indenopyrazoles as Novel Cyclin Dependent Kinase (CDK) Inhibitors. *J. Med. Chem.* **44** (9): 1334–1336.
7. Nugiel, DA., Vidwans, A., Etkorn, A-M., Rossi, KA., Benfield, PA., Burton, CR., Cox, S., Doleniak, D. & Seitz, SP. (2002). Synthesis and Evaluation of

Indenopyrazoles as Cyclin-Dependent Kinase Inhibitors. 2. Probing the Indeno Ring Substituent Pattern. *J. Med. Chem.* **45** (24): 5224–5232.

8. Yue, EW., Higley, CA., DiMeo, SV., Carini, DJ., Nugiel, DA., Benware, C., Benfield, PA., Burton, CR., Cox, S., Grafstrom, RH., Sharp, DM., Sisk, LM., Boylan, JF., Muckelbauer, JK., Smallwood, AM., Chen, H., Chang, Ch-H., Seitz, SP. & Trainor, GL. (2002). Synthesis and Evaluation of Indenopyrazoles as Cyclin-Dependent Kinase Inhibitors. 3. Structure Activity Relationships at C3(1,2). *J. Med. Chem.* **45** (24): 5233–5248.

9. Yue, EW., DiMeo, SV., Higley, CA., Markwalder, JA., Burton, CR., Benfield, PA., Grafstrom, RH., Cox, S., Muckelbauer, JK., Smallwood, AM., Chen, H., Chang, CH-H., Trainor, GL. & Seitz, SP. (2004). Synthesis and evaluation of indenopyrazoles as cyclin-dependent kinase inhibitors. Part 4: Heterocycles at C3. *Bioorg. Med. Chem. Lett.* **14** (2): 343–346.

10. Nugiel, DA., Vidwans, A. & Dzierba, CD. (2004). Parallel synthesis of acylsemicarbazide libraries: preparation of potent cyclin dependent kinase (cdk) inhibitors. *Bioorg. Med. Chem. Lett.* **14** (22): 5489–5491.

11. Cohen, P. (2002). Protein kinases — the major drug targets of the twenty-first century?. *Nat. Rev. Drug Discov.* **1**: 309-315.

12. Nguyen, MD., Boudreau, M., Kriz, J., Couillard-Després, S., Kaplan, DR. & Julien JP. (2003). Cell Cycle Regulators in the Neuronal Death Pathway of Amyotrophic Lateral Sclerosis Caused by Mutant Superoxide Dismutase. *J Neurosci.* **23** (6): 2131-40.

13. Shapiro, G I J. (2006). Cyclin-Dependent Kinase Pathways As Targets for Cancer Treatment. *Clin Oncol.* **24**: 1770–1783.

14. Eyring, H. (1935). The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **3** (2): 107-115.

15. Parang, K., Till, JH., Ablooglu, AJ., Kohanski, RA., Hubbard, SR. & Cole, PA. (2001). Mechanism-based design of a protein kinase inhibitor. *Nat Struct Biol.* **8** (1): 37-41.

16. Thomas, MP. & McInnes, C. (2006). Structure-based discovery and optimization of potential cancer therapeutics targeting the cell cycle. *IDrugs*. **9**: 273–278.
17. Moreno-Smith, M., Lutgendorf, SK. & Sood, AK. (2010). Impact of stress on cancer metastasis. *Future Oncol*. **6** (12): 1863–1881.
18. Cole, S.W. & Sood, A.K. (2012). Molecular Pathways: Beta-Adrenergic Signaling in Cancer. *Clin Cancer Res*. **18** (5): 1201-1206.
19. Lente, G. Deterministic kinetics. In chemistry and systems biology. The dynamics of complex reaction networks. Springer. Pag. 108.
20. Siegel, RL., Miller, KD. & Jemal, A. (2017). Cancer statistics. *CA Cancer J Clin*. **67** (1): 7–30.
21. Ranabir, S. & Reetu, K. (2011). Stress and hormones. *Indian J Endocrinol Metab*. **15** (1): 18–22
22. Ganguly, P. & Desiraju, GR. (2008). Van der Waals and Polar Intermolecular Contact Distances: Quantifying Supramolecular Synthons. *Chem Asian. J*. **3** (5): 868-880.
23. Carbó-Dorca, R. (2016). Aromaticity, quantum multimolecular polyhedra, and quantum QSPR fundamental equation. *J Comput Chem*. **37** (1): 78 - 82.
24. Cioslowski, J. & Fleischmann, ED. (1991). Assessing molecular similarity from results of ab initio electronic structure calculations. *J. Am. Chem. Soc*. **113** (1): 64-67.
25. Davis, B., Laing, B. & Hefferhn, R. en: Proceedings of the 1997 International Arctic Seminar (Pedagogical Institute, Murmansk, Russia, 1997), pp. 31-36.

CAPÍTULO II

CDKs Y RUTA DE SEÑALIZACIÓN β - ADRENÉRGICA EN EL CÁNCER

El cáncer es un proceso microevolutivo que se traduce en un conjunto de enfermedades caracterizadas por la proliferación continua de células genéticamente anormales con capacidad de invasión y destrucción de otros tejidos. La homeostasis celular está regulada en primer término por la presencia de factores externos como son los factores de crecimiento, interacciones con otras células y nutrientes [1]. Una vez unido el factor de crecimiento a su receptor, la actividad cinasa de éste inicia un sistema de transducción de señales que permiten la llegada de la señal proliferativa hasta el núcleo, donde por inducción de la expresión de diversos mediadores, incluidas las ciclinas, se favorece la proliferación celular [2]. Sin embargo, cuando el estímulo proliferativo llega al núcleo, la división celular queda regulada por los puntos de control del ciclo celular que impiden la progresión de células dañadas y mantienen la homeostasis celular [3].

2.1 MECANISMOS REGULADORES DEL CICLO CELULAR

Dos rutas supresoras de tumores que bloquean la progresión a través del ciclo celular, la ruta del retinoblastoma (RB) y la del p53 están inactivadas en la mayoría, si no en todas, las células cancerosas [4,5]. El RB reprime la activación transcripcional de genes controlados por el E2Fs, una familia de

factores de transcripción que regula la expresión de varios genes involucrados en la progresión del ciclo celular [6]. En las células animales ambas rutas están integradas con las cascadas de fosforilaciones de proteínas citoplasmáticas efectuadas por heterodímeros cuyos constituyentes son una subunidad cinasa y una subunidad ciclina (Fig. 2.1). Las proteínas cinasas pertenecen a la familia de las transferasas que catalizan la transferencia de un grupo fosforilo desde un donante hacia un aceptor. En el caso del complejo CDK/ciclina algunos de los aceptores son la proteína lámina nuclear, la condensina y el APC.

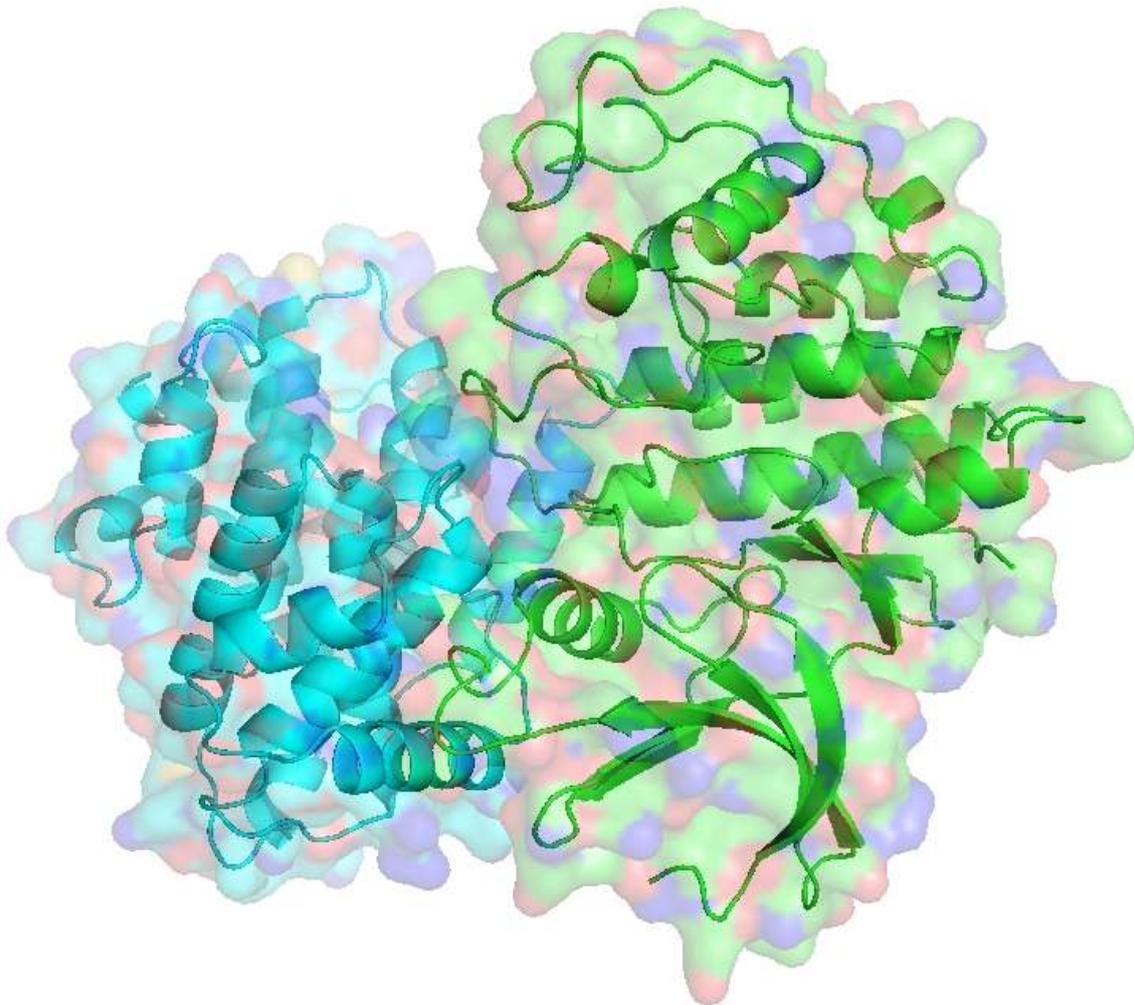


Figura 2.1. Conformación activa del complejo ciclina (estructura con superficie azul)-cinasa dependiente de ciclina (estructura con superficie verde).

La subunidad catalítica aislada de la CDK2 presenta una actividad catalítica relativamente despreciable. La actividad cinasa máxima de la enzima no se alcanza hasta que la subunidad catalítica se enlaza con una ciclina alostérica y el dominio catalítico se ha fosforilado en la Thr-160. Tanto el enlace de la ciclina como la fosforilación de la Thr-160 estabilizan los cambios conformacional de gran escala en el dominio catalítico los cuales incrementan la afinidad por el sustrato (reducen la K_M) y también elevan la velocidad de la reacción catalítica (aumentan k_{cat}) [7]. La concentración de las ciclinas varía en forma cíclica durante el transcurso del ciclo celular debido a la proteólisis dependiente de ubiquitinación y posterior traslado al proteasoma [8]. De otro lado, la actividad de las holoenzimas CDK/ciclina es regulada negativamente por la interacción con proteínas de la familia CKI (*CDK inhibitory protein*) como también mediante la fosforilación de la Tyr-15 y la Thr-14 de la subunidad catalítica. Las CKI se clasifican en las subfamilias CIP/KIP (p21, p27 y p57) e INK4 (p15, p16, p18, p19) [9].

Debido a su participación en procesos regulatorios de la homeostasis celular y la desregulación de las cinasas dependientes de ciclina (CDKs) en varios desordenes, los inhibidores de las CDK tendrían un amplio espectro de aplicaciones que iría desde infecciones protozoarias (como malaria, leishmania, tripanosomiasis [10-12], infecciones virales (por HCMV, HSV, HIV, HPV) [13], desordenes reproductivos [14], enfermedades cardiovasculares (como aterosclerosis, restenosis, hipertrofia cardiaca [15]), glomerulonefritis [16] y cánceres [17, 18] hasta enfermedades del sistema nervioso (como enfermedades de Alzheimer, convulsión, enfermedad amiotrófica y abuso de drogas [19]). Particularmente en la investigación de agentes antineoplásicos una diana farmacológica relevante es la enzima CDK2 (cinasa dependiente de ciclina 2) que fue utilizada para desarrollar uno de los estudios de esta tesis.

2.1.2 Estructura de la cinasa dependiente de ciclina 2.

La CDK2 es una cadena polipeptídica monómerica de 298 residuos que presenta un pliegue bilobular conservado; la bisagra une el lóbulo N-terminal más pequeño al lóbulo C-terminal más grande jugando un importante papel en el enlace específico del ATP, el cofactor común de las cinasas. Como se

muestra en la **Figura 2.2**, estructuralmente el lóbulo amino-terminal de la CDK2 contiene primariamente láminas β y la hélice PSTAIRE. La proteína sustrato se enlaza en la entrada del surco interactuando principalmente con la superficie del lóbulo carboxi-terminal. Los residuos cercanos catalizan la transferencia del fosfato del ATP al oxígeno del hidroxilo en la proteína sustrato [20].

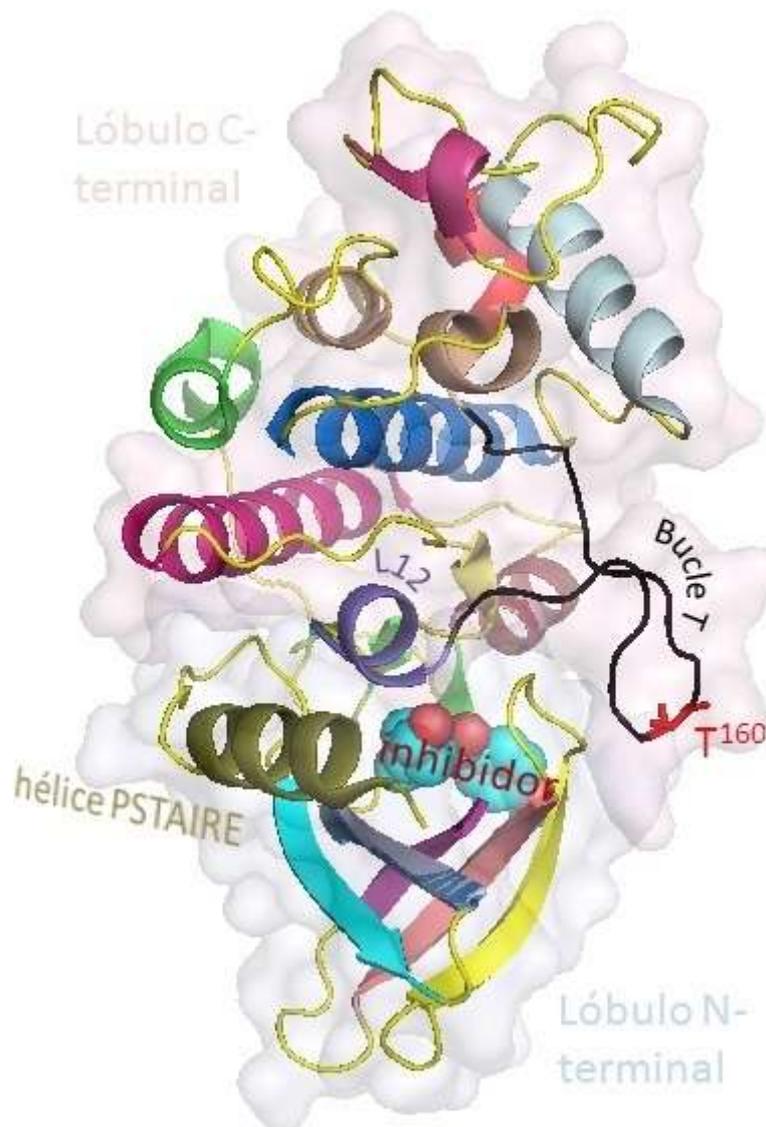


Figura 2.2. Estructura 3D de la CDK2. El lóbulo carboxi-terminal se ha representado como una superficie de color rosado pálido mientras que el lóbulo amino-terminal está como una superficie azul claro. Los bucles se muestran en amarillo excepto el bucle T que aparece en negro. El residuo de Thr160 es

representado en barras rojas. Se señalan la hélice PSTAIRE en oliva y la hélice L12 en morado.

Actualmente se dispone de múltiples estructuras cristalinas de distintos estados funcionales de la CDK2. En el estado inactivo de la CDK2 varias cadenas laterales de aminoácidos están colocadas de manera incorrecta de modo que el ATP no está idealmente orientado para la reacción cinasa. Una vez que la CDK2 se enlaza con la ciclina, la hélice L12 que ocluía el sitio activo se transforma en una hoja β [21]. Una vez que la Thr 160 es fosforilada, su grupo fosfato se inserta en un bolsillo catiónico y actúa como nodo central para una red de puentes de hidrógeno que estabilizan la interacción con la ciclina (**Figura 2.3**).

La hélice PSTAIRE es equivalente a la hélice C en la proteína cinasa A. Esta hélice interacciona directamente con la ciclina mediante la formación de un puente de hidrógeno entre el residuo Glu51 y el residuo Lys33 de la Ciclina E. La hélice PSTAIRE es virtualmente paralela a la hélice $\alpha 5$ de la Ciclina E y casi perpendicular a la hélice $\alpha 3$ del extremo C-terminal [22]. La hélice PSTAIRE se mueve hacia adentro tras el enlace con la ciclina resultando en la reorientación de residuos clave del sitio activo. El nombre de esta hélice viene de su secuencia aminoacídica la cual está conservada en las CDKs. La hélice L12 es una pequeña hélice α (residuos 147-151) adyacente al bucle T que evita el movimiento de la hélice PSTAIRE en la CDK2 libre [23]. Su estructura cambia a hebra β después de la unión con la ciclina permitiendo a la hélice PSTAIRE rotar casi 90 grados alrededor de su eje helicoidal, reconfigurando en consecuencia el sitio activo.

El bucle T (**Figura 2.3**), llamado a veces bucle de activación, es un elemento flexible adyacente al sitio activo de la cinasa. Su nombre obedece al residuo Thr cuya fosforilación es requerida para la máxima actividad y suministra selectividad para el enlace con el sustrato mediante el bloqueo del enlace de sustratos proteínicos en la entrada del surco del sitio activo [24]. La orientación precisa del bucle T se requiere para el reconocimiento selectivo de un motivo secuencial presente en la Ciclina A y en la Ciclina E. La fosfo-Thr160 actúa

como centro organizador y hace contactos con tres residuos Argininas (Arg 50, 126 y 150) también en la molécula de CDK2. La Arg50 y la Arg150 forman puentes de hidrógeno con los grupos carbonilos en la Leu187 y Glu188 en la Ciclina E [25].

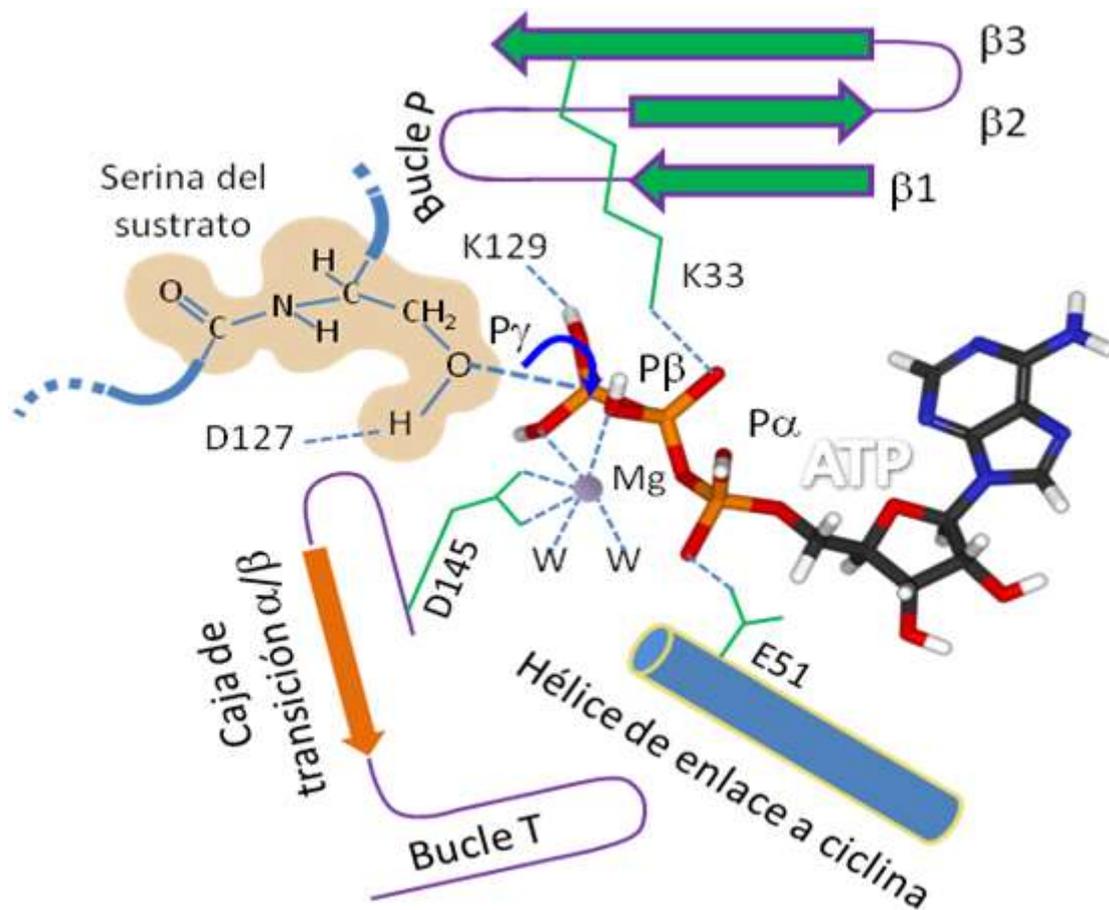


Figura 2.3. Complejo terciario CDK2-ATP-serina nucleofílica de la proteína diana. Se han destacado varios motivos de estructura secundaria de la CDK2 próximos a su bolsillo de ATP: los bucles P y T, la hélice de enlace a ciclina, la caja de transición y el átomo de magnesio.

Se ha reconocido de antaño que la capacidad de una enzima de formar complejo con su sustrato es crítica para la tasa de catálisis. Esta cualidad ha recibido denominaciones como propincuidad y aproximación, entre otras [26]. El bolsillo de enlace al ATP de las cinasas consiste en un número de elementos estructurales conservados que interactúan con varios grupos funcionales del ATP (**Figura 2.3**). Por ejemplo, el enlace de los fosfatos involucra residuos

catalíticos conservados, incluyendo el aspartato del bucle de activación (bucle A), asparaginas del bucle catalítico, lisina de la tercera hoja del lóbulo N-terminal y el fragmento rico en glicina de enlace al fosfato (también llamado bucle P). No obstante, es el sitio de enlace a la adenina el que ha sido centro de atención para descubrimiento de inhibidores de cinasa. El surco hidrofóbico formado por residuos conservados de la hoja principal del lóbulo N-terminal y los residuos conservados de una hoja pequeña del lóbulo C-terminal muestra una preferencia global por una estructura anular, mientras que los grupos amino y carbonilo del esqueleto la bisagra dicta el reconocimiento específico de la adenina y sus competidores [27, 28].

El complejo CDK2/Ciclina E coordina la entrada en la fase S a través de la fosforilación adicional del bolsillo de sus proteínas sustrato, inactivando sus funciones y promueve la iniciación de la síntesis de ADN, replicación y progresión a través de la fase S junto con la CDK2/Ciclina A. Más aún, varios estudios muestran que la CDK2/Ciclina A coordina la activación del CDK1/Ciclina B en el centrosoma y en el núcleo durante la fase G2. En una etapa posterior, la Ciclina A se asocia con la CDK1 y es finalmente reemplazada por la Ciclina B para coordinar la progresión a través de la mitosis [29]. Aparte de las CDKs, otra ruta bioquímica importante en el cáncer es la señalización β -adrenérgica.

2.2. LA SEÑALIZACIÓN β -ADRENÉRGICA Y EL CÁNCER

La evidencia reciente apoya una hipótesis de larga data según la cual el estrés crónico puede influenciar el crecimiento y la progresión de los tumores. Se ha mostrado que los neurotransmisores simpáticos, tales como las catecolaminas y neuropéptidos, pueden afectar tanto el crecimiento celular en el cáncer y la vascularización tumoral. Dependiendo del neurotransmisor y tipo de tumor, estos efectos pueden ser tanto estimulatorios como inhibitorios [30]. La noradrenalina (NA) y la adrenalina (A) son potentes estimulantes de la vascularización, actuando tanto al inducir la liberación de factores angiogénicos de las células tumorales y directamente en las funciones de las células endoteliales (EC). Como resultado, la activación del sistema adrenérgico

incrementa el crecimiento de varios tipos de tumores y se ha mostrado que media el aumento inducido por estrés de la progresión tumoral (**Figura 2.4**).

La dopamina (DA), de otro lado, interfiere con la señalización del VEGF en las células endoteliales, bloquea sus funciones angiogénicas e inhibe el crecimiento del tumor [31]. Otro neurotransmisor simpático que se libera con la NA, el neuropéptido Y (NPY), estimula directamente la angiogénesis. No obstante, las acciones proangiogénicas del NPY pueden ser alteradas por su efecto directo sobre la proliferación y supervivencia de la célula tumoral. En consecuencia, el NPY puede tanto estimular como inhibir el crecimiento del tumor, dependiendo del tipo de tumor. De aquí que los neurotransmisores simpáticos sean poderosos moduladores del crecimiento tumoral y que sus dianas puedan llegar a jugar un papel en la terapia del cáncer [32].

El interés en la eficacia de agentes β -bloqueadores como posibles fármacos en el tratamiento del cáncer ha estado ganando momento en años recientes. Zhang *et al.* reportaron que los receptores β -adrenérgicos pueden jugar un papel en la invasión del cáncer pancreático concretamente hallaron que las líneas celulares del cáncer pancreático MIA PaCa-2 y BxPC-3 expresan ARNm y proteína para los receptores β -adrenérgicos β_1 y β_2 [33]. Observaron que los β -bloqueadores pudieran ser útiles en la supresión de la invasión y proliferación del cáncer pancreático. De otro lado, el grupo de Yang ha mostrado que la NA puede promover la progresión tumoral al estimular la secreción de VEGF, MMP-2, MMP-9, interleucina (IL)-6 y IL-8. De esta manera, la NA aumenta las propiedades invasiva y pro-angiogénicas de otros tumores incluyendo el carcinoma nasofaríngeo, mieloma múltiple y el melanoma cutáneo [34].

La respuesta al estrés del Sistema Nervioso Simpático para pelear o huir, libera epinefrina y norepinefrina en el microambiente del tumor desde las fibras nerviosas simpáticas. Ambas catecolaminas se enlazan a los receptores β -adrenérgicos, resultando en la activación mediada por $G_{\alpha s}$ de la adenilil ciclasa y la subsiguiente conversión del ATP en AMPc. El flujo transitorio de AMPc intracelular activa dos sistemas bioquímicos efectores. En el primero, el AMPc activa la PKA para fosforilar múltiples proteínas diana, incluyendo factores de

transcripción de las familias CREB/ATF y GATA [35], como también la cinasa del receptor β -adrenérgico (BARK). El BARK recluta la β -arrestina inhibiendo la señalización del receptor β -adrenérgico y activa la cinasa Src, resultando en la activación de factores de transcripción como el STAT3 y cinasas corriente abajo tales como la cinasa de adhesión focal (FAK).

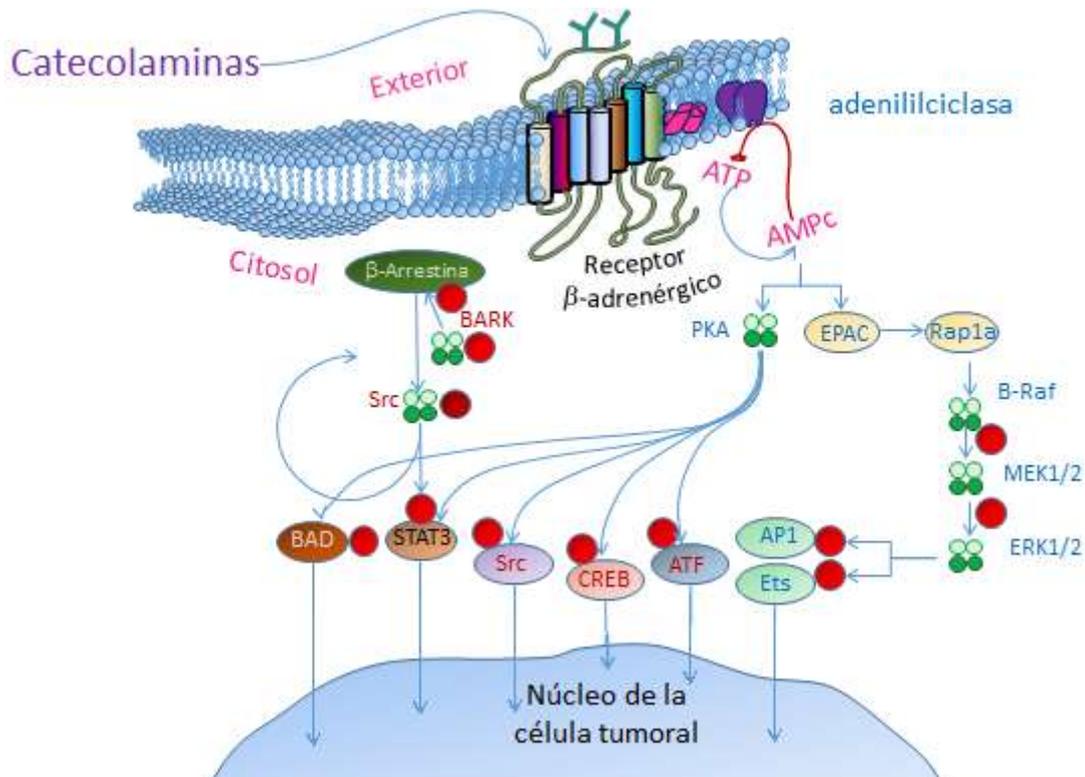


Figura 2.4. Señalización β -adrenérgica en el cáncer.

La activación del FAK modula el tráfico y la movilidad celular a través de su acción sobre la dinámica del citoesqueleto, como también la resistencia celular a la apoptosis. La activación dependiente de PKA del BAD también puede hacer que las células cancerosas sean resistentes a apoptosis inducida por quimioterapia. En la segunda ruta efectora, la AMPc activa la proteína de intercambio activada por adenilil ciclasa (EPAC) que a su turno lleva a la activación mediada por Rap1A de la ruta de señalización de la B-Raf/proteína cinasa activada por mitógenos y los efectos corriente abajo incluyen la transcripción génica mediada por familia de factores de transcripción AP-1 y Ets [36].

2.3 REFERENCIAS

1. Pruitt, K. Progress in molecular biology and translational science. Vol 144. Molecular and cellular changes in the cancer cell. Elsevier. 2016. Pag 410.
2. Moschovi, M., Critselis, E., Cen, O., Adamaki, M., G Lambrou, G.I., Chrousos,G.P. & Vlahopoulos, S. (2015). Drugs acting on homeostasis: challenging cancer cell adaptation. *Expert Rev Anticancer Ther.* **15** (12): 1405-1417.
3. Felton, M J. Survival of the Fittest in Drug Design. American Chemical Society (2000). In: *Modern Drug Discovery.* **3** (9): 49-50.
4. Levine, AJ. (1997). p53, the cellular gatekeeper for growth and division. *Cell.* **88** (3): 323-31.
5. Sellers, WR. & Kaelin, WG Jr. (1997). Role of the retinoblastoma protein in the pathogenesis of human cancer. *J Clin Oncol.* **15** (11): 3301-12.
6. Trimarchi, JM. & Lees, JA. Sibling rivalry in the E2F family. (2002). *Nat Rev Mol Cell Biol.* **3** (1): 11-20.
7. Brown, N.R., Noble, M.E., Lawrie, A.M., Morris, M.C., Tunnah, P., Divita, G., Johnson, L.N., and Endicott, J.A. (1999b). Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity. *J. Biol. Chem.* **274**: 8746–8756.
8. Gutierrez, GJ., Vöggtlin, A., Castro, A., Ferby, I., Salvagiotto, G., Ronai, Z., Lorca, T. & Nebreda, AR. (2006). Meiotic regulation of the CDK activator RINGO/Speedy by ubiquitin-proteasome-mediated processing and degradation. *Nat Cell Biol.* **8** (10): 1084-1094.

9. Pavletich, NP. (1999). Mechanisms of cyclin-dependent kinase regulation: structures of cdk2s, their cyclin activators, and cip and INK4 inhibitors. *J. Mol. Biol.* **287** (5): 821-828.
10. Gutierrez, GJ., Vöggtlin, A., Castro, A., Ferby, I., Salvagiotto, G., Ronai, Z., Lorca, T. & Nebreda, AR. (2006). Meiotic regulation of the CDK activator RINGO/Speedy by ubiquitin-proteasome-mediated processing and degradation. *Nat Cell Biol.* **8** (10): 1084-1094.
11. Osorio, EJ., Montoya, GL. & Arango, GJ. (2006). Productos Naturales Alcaloidales con Actividad Antiprotozoaria. *Vitae.* **13** (1): 61-84. ISSN 0121-4004
12. Tanowitz, HB., Machado, FS., Jelicks, LA., Shirani, J., Campos de Carvalho, AC., Spray, DC., Factor, SM., Kirchhoff, LV. & Weiss, LM. (2009). Perspectives on Trypanosoma cruzi-induced heart disease (Chagas disease). *Prog Cardiovasc Dis.* **51** (6): 524-539.
13. Laman, H., Coverley, D., Krude, T., Laskey, R. & Jones, N. (2001). Viral Cyclin–Cyclin-Dependent Kinase 6 Complexes Initiate Nuclear DNA Replication. *Mol. Cell. Biol.* **21** (2): 624–635
14. Tempfer, CB., Simoni, M., Destenaves, B. & Fauser BCJM. (2009). Functional genetic polymorphisms and female reproductive disorders: Part II—endometriosis. *Human Reproduction Update.* **15** (1): 97–118.
15. Suzuki, J., Isobe, M., Morishita, R., Aoki, M., Horie, S., Okubo, Y., Kaneda, Y., Sawa, Y., Matsuda, H., Ogihara, T. & Sekiguchi, M. (1997). Prevention of graft coronary arteriosclerosis by antisense cdk2 kinase oligonucleotide. *Nat Med.* **3** (8): 834-835.
16. Zoja, C., Casiraghi, F., Conti, S., Corna, D., Rottoli, D., Cavinato, RA., Remuzzi, G. & Benigni, A. (2007). Cyclin-dependent kinase inhibition limits

glomerulonephritis and extends lifespan of mice with systemic lupus. *Arthritis Rheum.* **56** (5): 1629-1637.

17. Li, HL., Ma, Y., Ma, Y., Li, Y., Chen, XB., Dong, WL. & Wang, RL. (2017). The design of novel inhibitors for treating cancer by targeting CDC25B through disruption of CDC25B-CDK2/Cyclin A interaction using computational approaches. *Oncotarget.* **8** (20): 33225-33240.

18. Otto T. & Sicinski P. (2017). Cell cycle proteins as promising targets in cancer therapy. *Nature Reviews Cancer.* **17**: 93–115.

19. Nguyen, MD., Boudreau, M., Kriz, J., Couillard-Després, S., Kaplan, DR. & Julien, JP. (2003). Cell Cycle Regulators in the Neuronal Death Pathway of Amyotrophic Lateral Sclerosis Caused by Mutant Superoxide Dismutase. *J Neurosci.* **23** (6): 2131-40.

20. Davies TG, Bentley J, Arris CE, Boyle FT, Curtin NJ, Endicott JA, Gibson AE, Golding BT, Griffin RJ, Hardcastle IR, Jewsbury P, Johnson LN, Mesguiche V, Newell DR, Noble ME, Tucker JA, Wang L, Whitfield HJ. (2002). Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. *Nature Struct. Biol.* **9** (10): 745-749.

21. Morgan, DO. (2007). The Cell cycle. Principles of control. *Yale J Biol Med.* **80** (3): 141–142.

22. Morgan, DO. (1996). The dynamics of cyclin dependent kinase structure. *Curr. Opin. Cell Biol.* **8** (6): 767-772.

23. Russo, A.A., Jeffrey, P.D. & Pavletich, N.P (1996). Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat Struct Biol.* **3**: 696 – 700.

24. Moshinsky, DJ., Bellamacina, CR., Boisvert, DC., Huang, P., Hui, T., Jancarik, J., Kim, S.-H. & Rice, AG. (2003). SU9516: biochemical analysis of

cdk inhibition and crystal structure in complex with cdk2. *Biochem Biophys Res Commun.* **310**: 1026–1031.

25. Davies TG, Tunnah P, Meijer L, Marko D, Eisenbrand G, Endicott JA, Noble ME. (2001). Inhibitor binding to active and inactive CDK2: the crystal structure of CDK2-cyclin A/indirubin-5-sulphonate. *Structure.* **9** (5): 389-397.

26. Bruice, TC. (1970). Proximity Effects and Enzyme Catalysis. *The Enzymes.* **2**: 217-279.

27. Radzio-Andzelm, E., Lew, J. & Taylor, S. (1995). Bound to activate: conformational consequences of cyclin binding to CDK2. *Structure.* **3**: 1135–1141.

28. May, C., Morris, Gondeau, C., Tainer, J.A. & Divita, G. (2002). Kinetic Mechanism of Activation of the Cdk2/Cyclin A Complex key role of the C-lobe of the Cdk. *J. Biol. Chem.* **277**: 23847-23853.

29. Stevenson, L.M., Deal, M.S., Hagopian, J.C., and Lew, J. (2002). Activation mechanism of CDK2: role of cyclin binding versus phosphorylation. *Biochemistry.* **41**: 8528–8534.

30. Yang, EV. (2010). Role for catecholamines in tumor progression. Possible use for β -blockers in the treatment of cancer. *Cancer Biol Ther.* **10** (1): 30-32.

31. Chakroborty, D., Sarkar, C., Basu, B., Dasgupta, P.S. & Basu, S. (2009). Catecholamines regulate tumor angiogenesis. *Cancer Res.* **69** (9): 3727–3730.

32. Tilan, J & Kitlinska, J. (2010). Review Article. Sympathetic Neurotransmitters and Tumor Angiogenesis—Link between Stress and Cancer Progression. *J Oncol.* **2010**: 1-6.

33. Zhang, D., Ma, Q-Y., Hu, H-T. & Zhang, M. (2010). β 2-adrenergic antagonists suppress pancreatic cancer cell invasion by inhibiting CREB, NF κ B and AP-1. *Cancer Biol Ther* . **10**: 19-29

34. Yang, EV., Donovan, EL., Benson, DM. & Glaser, R. (2008). VEGF is differentially regulated in multiple myeloma-derived cell lines by norepinephrine. *Brain Behav Immun* . **22**: 318-23.

CAPÍTULO 3

BIOFÍSICA DE LA INHIBICIÓN DE ENZIMAS

3.1. El estado de transición. La teoría del estado de transición (ET) o teoría del complejo activado fue desarrollada simultáneamente de un lado por Henry Eyring en la Universidad de Princeton [1] y de otro por Meredith Gwynne Evans y Michael Polanyi en la Universidad de Manchester [2]. El principio de Evans-Polanyi se refiere a una relación lineal entre la energía de activación y la entalpía de una reacción que algunas veces se observa en una serie de reacciones estrechamente relacionadas:

$$E_a = \alpha \Delta H^0 + E_0 \quad (\text{Ec. 3.1})$$

donde

E_a es la energía de activación de una reacción de referencia

ΔH^0 es la entalpía de la reacción

α es un número que caracteriza la posición del estado de transición a lo largo de la coordenada de la reacción ($0 \leq \alpha \leq 1$) para una serie dada de reacciones

La constante aditiva E_0 es la barrera intrínseca, es decir, la energía de activación de la reacción de esa familia para la cual $\Delta H^0 = 0$.

La ecuación de Evans-Polanyi implica que las barreras de las reacciones exotérmicas son menores que las barreras de las endotérmicas [3]. En esa línea, el postulado de Hammond-Leffler, es una hipótesis ampliamente utilizada

en la química física orgánica el cual dicta que si las especies intermediarias que se forman a lo largo de una reacción presentan energías similares, sus estructuras serán similares. El objetivo de Hammond era el de establecer la interpretación mecanística, postulando que los cambios en la estructura del estado de transición son afectados por la manera en la cual los sustituyentes afectan las energías de los intermediarios en rutas alternativas desde los reactantes hasta los productos [4].

Se denomina exoérgica a aquella reacción química que libera energía, en contraposición a la reacción endoérgica donde la energía se absorbe. Generalmente los cambios energéticos en una reacción se asocian con desprendimiento o absorción de calor y se denominan exotérmicas y endotérmicas respectivamente. En una reacción elemental con una diferencia de energía significativa entre los productos y los reactantes, el ET cambia gradualmente desde estructuras semejantes al reactante en reacciones altamente exergónicas, a un intermediario en carácter para reacciones ergoneutrales y a estructuras semejantes al producto en reacciones endergónicas, como se ilustra en la **Figura 3.1**. Esto se expresa cuantitativamente en una relación de energía de Gibbs [5]:

$$\Delta\Delta G^\ddagger = \alpha \Delta G_p + (1 - \alpha) \Delta G_R \quad (\text{Ec. 3.2})$$

La cual relaciona cambios en la energía del ET provocados por los sustituyentes, con los cambios la energía de Gibbs provocados por los mismos sustituyentes en reactantes (R) y productos (P).

Si $\alpha \approx 0$ entonces $\Delta\Delta^\ddagger G \approx \Delta G_R$ y el ET responde a cambios en los sustituyentes en la misma manera que los reactantes, una característica de los pasos exoenergéticos elementales. El ET es más cercano a los reactantes y en consecuencia, los arreglos estructurales para alcanzar los complejos activados son minimizados y la barrera de activación es pequeña. En la **Figura 3.2** se observa la correlación entre la localización de la barrera energética, indicada por una flecha y la estabilidad relativa de los reactantes y productos.

En una reacción exotérmica, el ET se parecerá más a los reactantes, mientras que en una endotérmica a los productos [6].

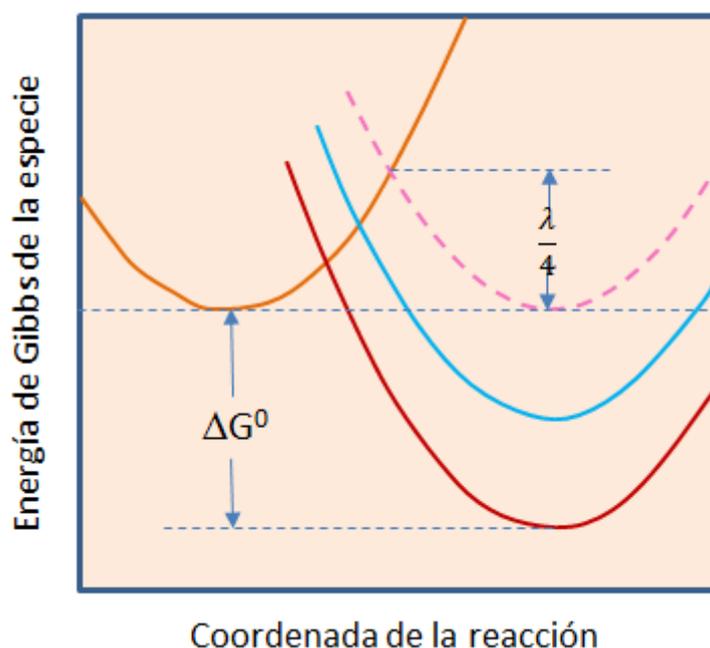


Figura 3.1. Cambio en la localización del estado de transición con la energía de Gibbs de la reacción.

De manera inversa, si $\alpha \approx 1$ entonces $\Delta\Delta^\ddagger G \approx \Delta G_P$ y el ET responde a cambios en los sustituyentes en una manera similar a los cambios en los productos, una característica de una reacción elemental altamente endoenergética. El ET es más cercano a los productos y sus rasgos estructurales son parecidos a los de los materiales finales. Se necesitan grandes rearrreglos estructurales para alcanzar el ET y la barrera de activación es alta [7].

3.2. Formulación cuantitativa del postulado de Hammond. Cioslowski formuló de manera cuantitativa el postulado de Hammond. Para lograrlo, él cuantificó los conceptos de la exotermicidad de la reacción (simbolizada por γ) y la proximidad estructural entre el estado de transición y los reactantes (simbolizada por β). Adicionalmente, Cioslowski introdujo un parámetro al que denominó isosincronicidad (simbolizada por α) el cual describe que tan directo es el camino de una reacción. Él definió α y β utilizando distancias estructurales

calculadas con la ayuda del índice de semejanza cuántica llamado NOEL (número de electrones solapantes, por sus siglas en inglés) que él mismo había desarrollado en un trabajo previo [8]. A continuación, él encontró una correlación positiva, consistente con el postulado de Hammond, entre los valores calculados de β y γ .

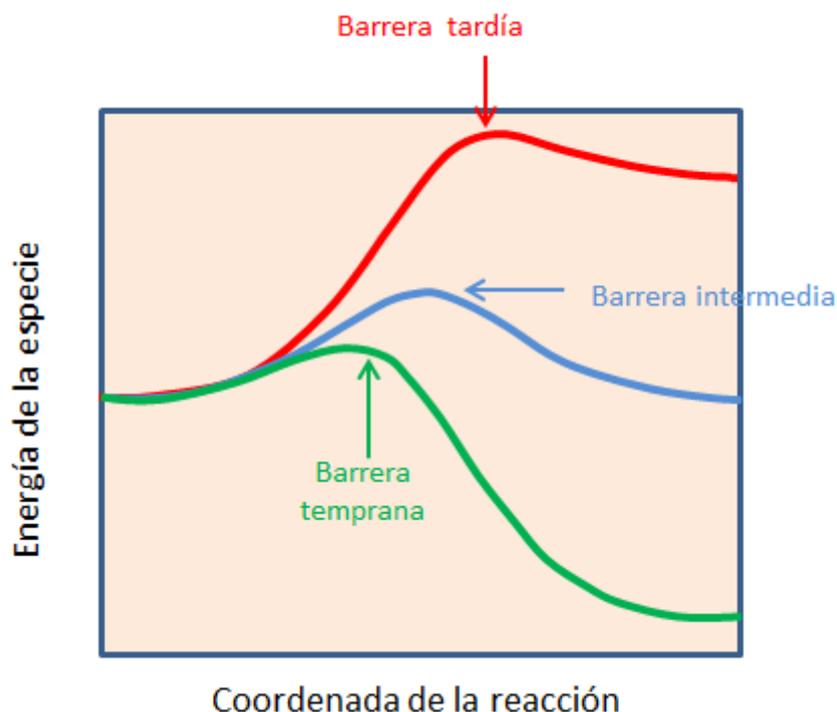


Figura 3.2. Perfil de energía potencial a lo largo de la coordenada de reacción. Se muestran los dos casos asimétricos que corresponden a una reacción exoérgica y endoérgica.

La exotermicidad (γ) se cuantifica a partir de las energías de los reactantes (A), productos (B) y del estado de transición (TS), las cuales respectivamente se simbolizan por E_A , E_B y E_{ET} . El siguiente cociente considera la diferencia y la suma de las energías de activación, $E_{ET} - E_A$ y $E_{ET} - E_B$, para las reacciones directa e inversa, respectivamente:

$$\gamma = \frac{E_B - E_A}{2E_{ET} - E_A - E_B} \quad (Ec. 3.3)$$

Los valores de γ están en el rango entre -1 y $+1$. Las reacciones exotérmicas están caracterizadas por valores negativos de γ , mientras que las reacciones endotérmicas tienen valores positivos de γ [9]. Las definiciones de β y γ serán dadas en el capítulo VII.

3.3 Aplicación del ET a la catálisis enzimática. A fin de entender la catálisis es necesario apreciar la importancia del ET y los intermediarios en las reacciones químicas. La velocidad de una reacción química depende de la frecuencia con la que las moléculas reactantes colisionen con la orientación y la energía correctas para que la reacción sea favorecida. Se puede representar el progreso de una reacción mediante un diagrama de energía. La **Figura 3.3** es un ejemplo que muestra la conversión de un sustrato (reactante) en un producto en un solo paso. El ET se presenta en el pico de la barrera de activación. Esta es la energía que se debe exceder para que la reacción proceda. Cuanto menor sea la barrera, el ET es más estable y mayor la velocidad de la reacción [10].

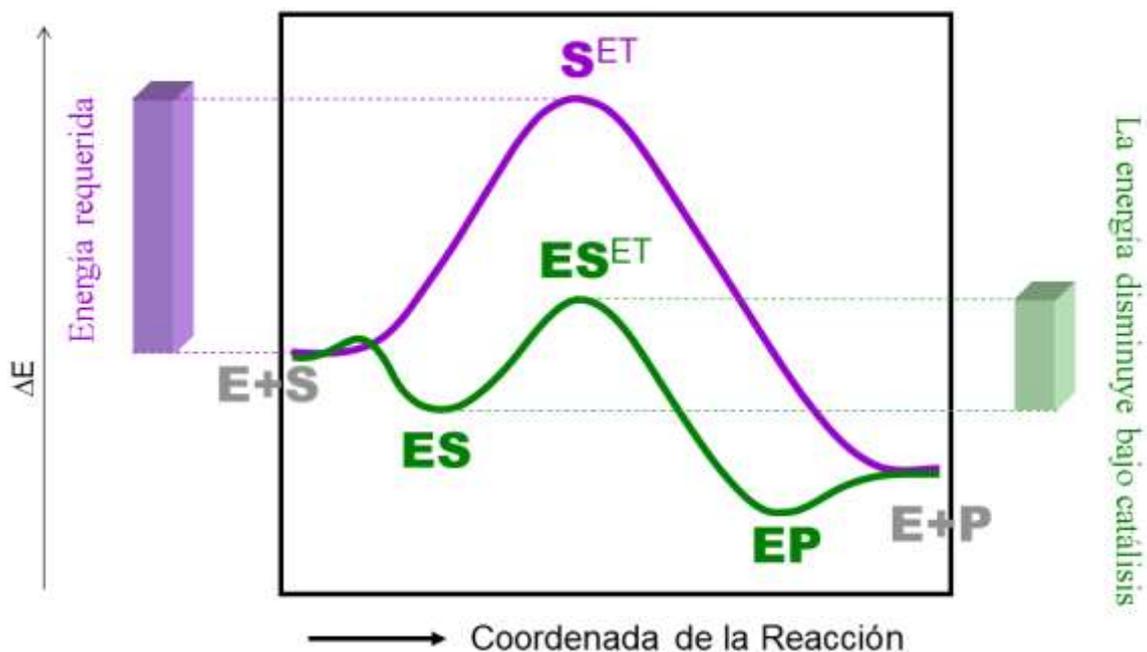


Figura 3.3. Diagrama energético de una reacción de biocatálisis. El eje x, denominado la coordenada de la reacción, mide el progreso de la reacción, comenzando con el sustrato en la izquierda y procediendo hacia el producto en

la derecha. Este eje no es tiempo, sino más bien el progreso de la ruptura de los enlaces y la formación de los enlaces de una molécula particular. El eje y muestra la energía de Gibbs de las especies.

La estructura del ET es crucial para entender la catálisis enzimática, puesto el aumento en la velocidad de reacción que logran las enzimas es debido a la disminución de la energía de activación cuando ellas ajustan la forma de su sitio activo de tal manera que la estructura inestable del ET encaje aún mejor que el sustrato. Linus Pauling sugirió que las moléculas inhibidoras de la actividad enzimática son análogos estructurales del ET del sustrato de una reacción particular. Siendo los inhibidores mimetizadores al ET del sustrato, frecuentemente ellos pueden enlazarse a la enzima mucho más fuertemente que el sustrato. De otro lado, puesto que la velocidad de reacción es proporcional a la fracción de reactante acomplexado con la enzima en el ET, se propuso también que la enzima incrementaba la concentración de estas especies reactivas [11]. Esta propuesta fue posteriormente formalizada en una hipótesis de Wolfenden (1972) y sus colaboradores de acuerdo con la cual el aumento de velocidad es proporcional a la afinidad de la enzima con la estructura del ET relativa al complejo de Michaelis.

Se han explorado las estructuras del ET de dianas enzimáticas para el cáncer, la autoinmunidad, malaria y antibióticos mediante la aplicación sistemática de los efectos isotópicos cinéticos y la química computacional [12]. Hoy la combinación de acceso experimental and computacional a la información del ET permite el diseño de análogos del ET como poderosos inhibidores enzimáticos y la exploración de los rasgos de la proteína vinculados a la estructura del ET. Los mapas de potencial electrostático del ET sirven como planos para guiar la síntesis de inhibidores análogos del ET de las enzimas escogidas. Se considera que las sustancias que mimetizan idealmente los rasgos geométricos y electrostáticos de un ET (u otros intermediarios de alta energía) como excelentes inhibidores enzimáticos. Ellos se enlazan hasta 100 veces más fuerte que el sustrato. Así, el propósito de diseñar análogos del ET es crear estructuras químicas estables con geometría de van der Waals y

superficies de potencial electrostático tan cercanas como sea posible a las del ET [13].

El estado de transición asociativo se define algunas veces como una formación de enlace por encima del 50% entre el oxígeno nucleofílico y el fósforo, la cual ocurre con al menos 50% del enlace residual del grupo saliente aún presente. El estado de transición disociativo puede definirse como una formación de enlace por debajo del 50% entre el nucleófilo y el fósforo que tiene lugar cuando se ha roto en al menos 50% del enlace del grupo saliente y el fósforo [14]. La descripción actual de la transferencia del grupo fosforilo es un mecanismo en el cual el donador y el aceptor del fosforilo interactúan débilmente con el grupo fosforilo en un estado de transición en el cual el enlazamiento total del fósforo está disminuido con relación al estado basal (**Fig. 3.4**) [15].

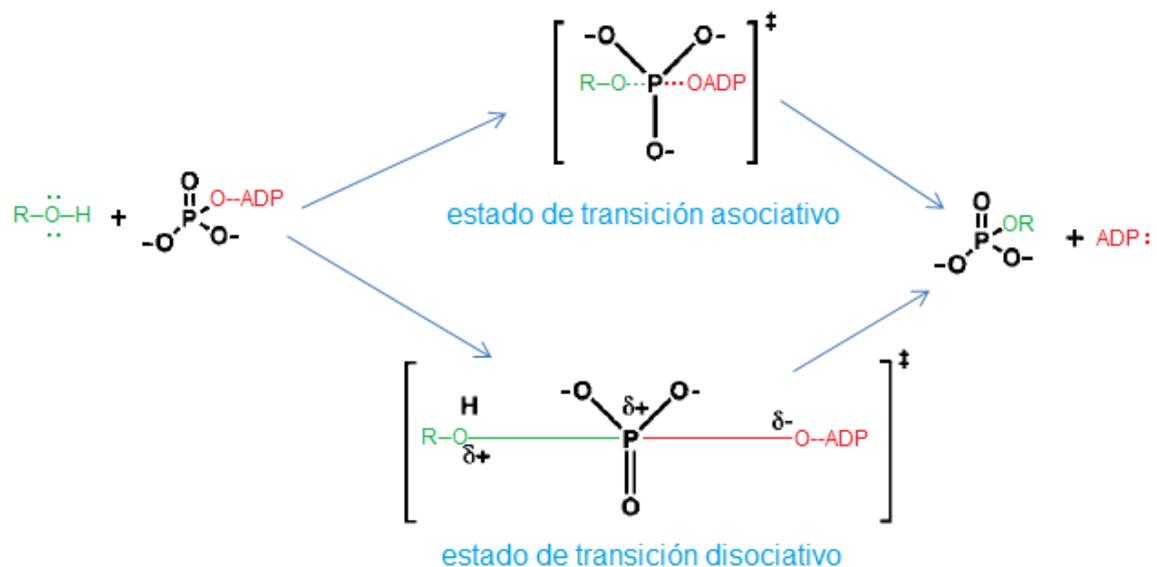


Figura 3.4. Probables Mecanismos de transferencia del grupo fosforilo en la CDK2. ROH es el nucleófilo (serina o treonina) el cual ataca el grupo fosforilo del ATP, y ADP es el grupo saliente. En la parte inferior del esquema se observa que en el estado de transición el enlace es débil entre el fósforo y el grupo saliente R-X y entre el fósforo y el grupo aceptor. Debido al enlazamiento disminuido al fósforo, este estado de transición ha sido descrito como *disociativo*.

Con el enlazamiento a la proteína, ésta puede cambiar bien en su conformación o en el estado de protonación de los residuos aminoacídicos localizados en el sitio de enlace. Estos cambios a nivel de la proteína contribuyen al cambio de entalpía resultante. No obstante, el enfoque preferido en el diseño de fármacos es buscar cambios de entalpía favorables que son específicos del fármaco, tales como la formación de puentes de hidrógeno y contactos de van der Waals. La desolvatación de grupos polares tras el enlazamiento representa una contribución entálpica desfavorable que se ha de compensar mediante interacciones favorables. El cambio en la entropía obedece a dos contribuciones. Antes del enlace, el ligando se halla solvatado y el sitio de unión no ocupado contiene agua. Tras el enlace, el agua es liberada resultando en una contribución entrópica favorable (**Figura 3.5**). Sin embargo, el enlace pudiera también resultar en la pérdida de las conformaciones disponibles y disminuir así el número de grados de libertad redundando en un cambio entrópico desfavorable [16]. Esta contribución desfavorable se puede minimizar restringiendo la flexibilidad del fármaco.

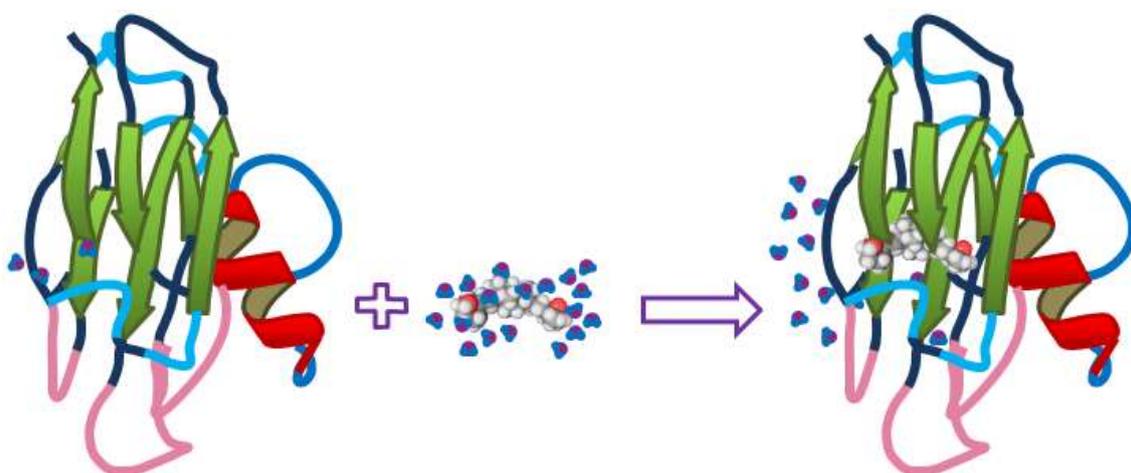


Figura 3.5. Reacción de formación del complejo diana-ligando. En la figura se considera el desplazamiento de moléculas de solvente tras la formación del enlace del fármaco con una proteína diana.

La comprensión de los mecanismos bioquímicos requiere el conocimiento del campo de fuerza que controla la evolución de la geometría molecular tanto en

el estado electrónicamente excitado como en el estado electrónico fundamental. Este campo de fuerzas se expresa en términos de superficies de energía potencial del estado excitado y del estado fundamental. La forma en que la energía de un sistema molecular varía al modificar su estructura, para un estado electrónico determinado, viene especificada por su superficie de energía potencial la cual constituye una representación matemática que relaciona la estructura molecular con la energía resultante. Para conocer la hipersuperficie de energía potencial, se ha de hacer un barrido de todas las variables internas de esta molécula. Por ejemplo, las distancias entre los átomos y el ángulo y las torsiones entre ellos, puede ser una elección posible [17].

Los estudios de descubrimiento de fármacos, inicialmente suelen enfocarse en la optimización de la afinidad de enlace. Al medir la dependencia de la afinidad del enlace con la temperatura, se identifican individualmente las contribuciones entálpicas y entrópicas. Los fármacos que son seleccionados con base en la optimización entálpica usualmente son más selectivos y tienen una afinidad del enlace más alta que los obtenidos utilizando optimización entrópica. La afinidad más alta es debida a la naturaleza menos específica de las interacciones hidrofóbicas que influyen fuertemente los cambios entrópicos [18, 19].

3.4. Dinámica conformacional de una proteína la catálisis. La dinámica de las proteínas juega un papel en la afinidad y la cinética del enlace de los fármacos [20]. Se puede concebir al fármaco como si se enlazara al estado de energía más bajo de la proteína, el cual debe sufrir re-arreglos estructurales a fin de acomodarse al fármaco. Debido a la dinámica de la proteína, un fármaco se encuentra con múltiples conformaciones de una proteína previamente al enlace (**Figura 3.6**). En algunos casos, el enlace del fármaco puede inducir un cambio conformacional sustancial debido a las interacciones hidrofóbicas entre la proteína y el fármaco. Si bien la energía global de una conformación de la proteína pudiera estar en un mínimo para el complejo no enlazado, pudiera no ser el más favorable para el enlace del fármaco, por cuanto los cambios conformacionales asociados con el enlace del fármaco puede llevar a un complejo más estable con la conformación no enlazada que tiene una energía mayor.

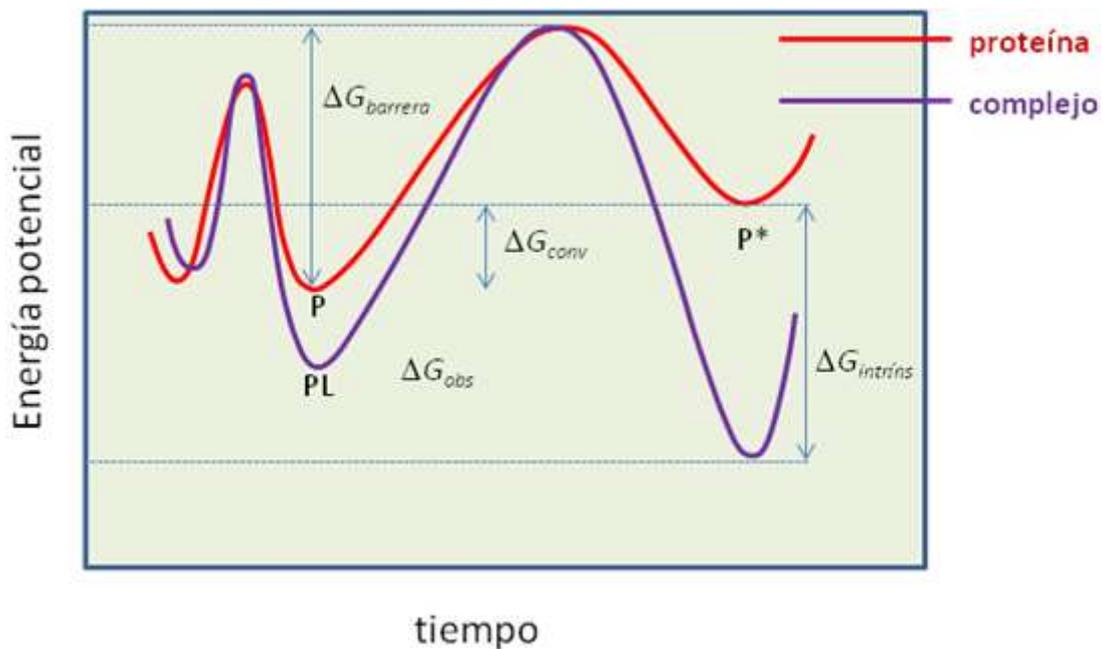


Figura 3.6. Dinámica conformacional de una proteína. Se han representado dos conformaciones posibles como P y P^* , cada una con una energía total diferente. El enlace de un fármaco puede tener diferentes interacciones con estas conformaciones produciendo un cambio observado de energía de Gibbs para el enlace del ligando.

Muchos aspectos de la química pueden reducirse a cuestiones relacionadas con las superficies de energía potencial (PES). Una PES presenta la energía de una molécula en función de su geometría. La energía se grafica sobre el eje vertical y las coordenadas geométricas (p. ej. Longitudes de enlace, ángulos de valencia, etc.) Se grafican sobre los ejes horizontales. Una PES real tiene muchas dimensiones, pero los rasgos clave pueden representarse mediante una PES 3-dimensional. Las estructuras moleculares de equilibrio corresponden a las posiciones de los mínimos en los valles sobre una PES. La energética de las reacciones se puede calcular a partir de las energías o altitudes de los mínimos para los reactantes y productos. Un camino de reacción conecta los reactantes y productos a través de un paso entre montañas. Una estructura de transición es el punto más alto sobre la trayectoria de energía más baja [21].

Los puentes de hidrógeno Individuales y las fuerzas de van der Waals son reversibles y se rompen fácilmente afectando la estructura de la proteína. Como resultado la estructura de la enzima libre (es decir sin ligando) es dinámica y de hecho representa un colectivo de sub-estados conformacionales, o microestados, que son interconvertibles [22]. Las transiciones entre estos microestados reflejan excursiones electrónicas, traslacionales, rotacionales y principalmente vibracionales a lo largo de la superficie de energía potencial del colectivo de microestados (**Figura 3.7**).

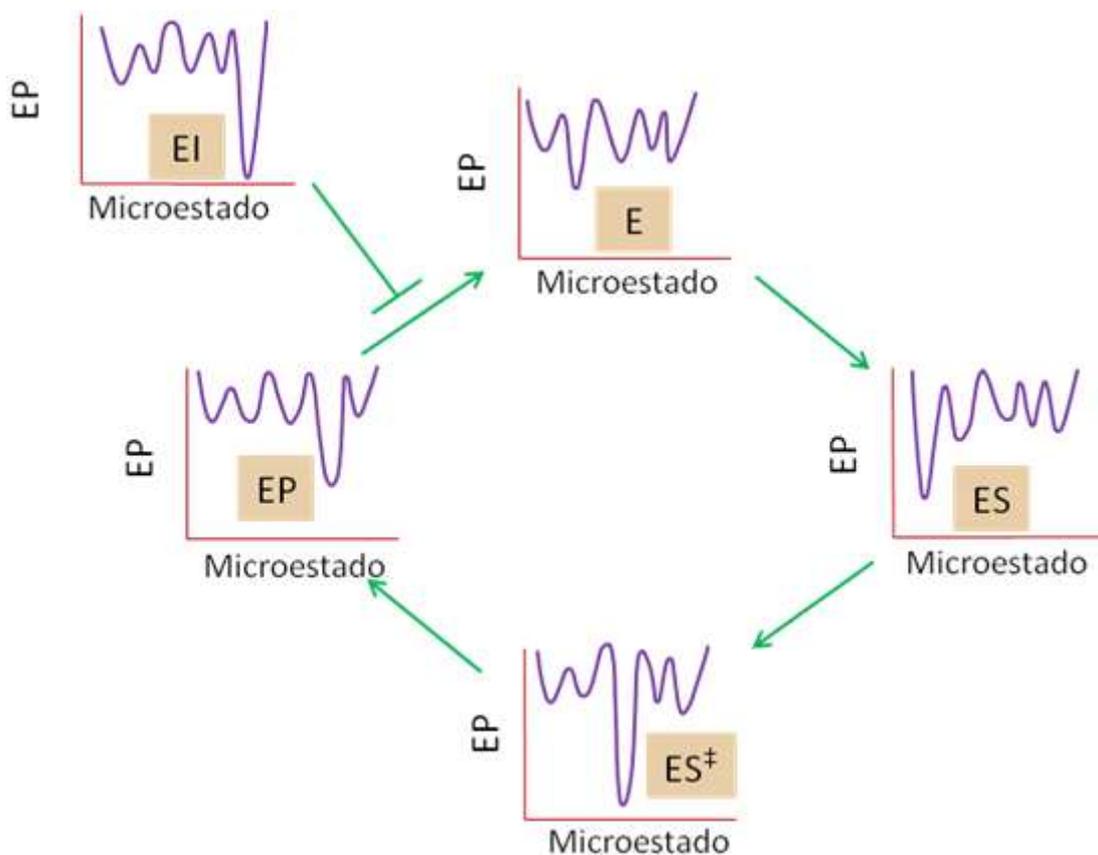


Figura 3.7. Cambios en distribución de los microestados conformacionales de la proteína durante la catálisis enzimática. Por cada paso del ciclo de reacción, la distribución de los microestados conformacionales es representada como un diagrama de energía potencial.

Los ligandos (sean sustratos, estados de transición, productos, o inhibidores) se unen preferencialmente a un microestado específico, o a un subconjunto de los microestados disponibles, el cual representa la mejor complementariedad

entre el bolsillo de enlace de ese microestado(s) y la estructura del ligando [23]. Así, el enlace con el ligando es un evento que estabiliza un microestado particular (o subconjunto de microestados) y con ello implica un desplazamiento en la distribución de estados, con relación a la enzima libre, hacia una población mayor en un pozo de potencial más profundo y estrecho (es decir, un mínimo de energía potencial).

La profundidad del pozo de potencial para el microestado preferido representativo del complejo enzima–ligando refleja el grado de estabilización de ese estado, el cual directamente se relaciona con la afinidad del ligando por ese estado. Cuanto más profundo es este pozo de potencial, mayor es la barrera de energía para la interconversión entre este microestado y los otros microestados potenciales del sistema [24]. Así, como se ilustra en la Figura 3.6, un ciclo catalítico mínimo de la enzima refleja una serie de cambios en la distribución de microestados a medida que la enzima se une al sustrato (ES), lo convierte en la estructura del estado de transición (ES^\ddagger), y convierte este en estructura del estado producto (EP).

Similarmente, las moléculas de Inhibidor se enlazan al microestado particular, o subconjunto de microestados, que mejor complementa su estructura. El microestado de mayor afinidad con el inhibidor puede presentarse en cualquier parte de la ruta de la reacción de la enzima; en la **Figura 7** se ilustra un ejemplo donde el inhibidor se enlaza preferencialmente a un microestado que está más poblado después del paso de liberación del producto en la ruta de la reacción. Si el pozo de potencial resultante del microestado(s) del complejo enzima–inhibidor es suficientemente profundo, el inhibidor atrapa la enzima en este microestado, evitando así interconversiones posteriores entre los microestados que se requieren para la catálisis [25].

3.5. REFERENCIAS

1. Eyring H. (1935). The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **3** (2): 107-115.
2. Evans, M. G. & Polanyi, M. (1936). Further considerations on the thermodynamics of chemical equilibria and reaction rates. *J. Chem. Soc. Faraday Trans.* **32**: 1340-1360.
3. IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: <http://goldbook.iupac.org> (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. ISBN 0-9678550-9-8. <https://doi.org/10.1351/goldbook>.
4. Hammond, G. S. (1955). A Correlation of Reaction Rates. *J. Am. Chem. Soc.* **77**: 334-338.
5. Gibbs, JW. (1873). A Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces. *Trans. Conn. Acad.* **2**: 382-404.
6. Lienhard, G.E. (1973). Enzymatic Catalysis and Transition-State Theory. *Science*. **180** (4082): 149-154.
7. Pechukas, P. (1981). Transition State Theory. *Ann. Rev. Phys. Chem.* **32**: 159-177.
8. Cioslowski, J. & Fleischmann, E.D. (1991). Assessing molecular similarity from results of ab initio electronic structure calculations. *J. Am. Chem. Soc.* **113** (1): 64-67.
9. J. Cioslowski. (1991). Quantifying the Hammond postulate: intramolecular proton transfer in substituted hydrogen catecholate anions. *J Am Chem Soc.* **113** (18): 6756-6760.

10. Horton, HR., Moran LA, Scrimgeour KG, Perry MD, Rawn JD. Principles of Biochemistry. Pearson Prentice Hall. 4a Ed. Página 161.
11. Wolfenden, R. (1972). Analog approaches to the structure of the transition state in enzyme reactions. **5** (1): 10-18.
12. Drug Discovery. Edited by Hany A. El-Shemy. InTech. 2013. Capítulo 12. Transition State Analogues of Enzymatic Reaction as Potential Drugs. Karolina Gluza and Pawel Kafarski. Página 325.
13. Schramm, V.L. (2007). Enzymatic Transition State Theory and Transition State Analogue Design. *J. Biol. Chem.* **282**: 28297-28300.
14. Matte, A., Tari, LW. & Delbaere, LTJ. (1998). Minireview. How do kinases transfer phosphoryl groups? *Structure*. **6** (4): 413-419.
15. Smith, HJ. & Simons, C. (2004). Enzymes and Their Inhibitors. Drug Development. CRC Press.
16. Freire, E. (2004) Isothermal titration calorimetry: controlling binding forces in lead optimization. *Drug Discovery Today: Technologies*. **1**: 295–9.
17. Landau, LD & Lifshitz, EM. (1996). Statistical Physics. 3rd Edition Part 1. Butterworth-Heinemann, Oxford.
18. Martin, IC. (2010). Quantitative Drug Design: A Critical Introduction, 2^a Edition. Ed. CRC Press.
19. Warshel, A & Levitt, M (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**: 227–49.
20. Landau LD, Lifshitz EM. (1996). Statistical Physics. 3rd Edition Part 1. Butterworth-Heinemann, Oxford.
21. Martin IC.(2010). Quantitative Drug Design: A Critical Introduction, 2^a Edition. Ed. CRC Press.
22. McGowan, LC. & Hamelberg, D. (2013). Conformational Plasticity of an Enzyme during Catalysis: Intricate Coupling between Cyclophilin A Dynamics and Substrate Turnover. *Biophys. J.* **104** (1): 216–226.

23. Herschlag, D. & Natarajan, A. (2013). Fundamental Challenges in Mechanistic Enzymology: Progress toward Understanding the Rate Enhancements of Enzymes. *Biochemistry*. **52** (12): 2050-2067.
24. Stromgaard, K., Krogsgaard-Larsen, P. & Madsen, U. Textbook of Drug Design and Discovery, Fourth Edition. CRC Press. 2016. Pág. 176.
25. Keedy, D.A., Kenner, L.R., Warkentin, M., Woldeyes, R.A., Hopkins, J.B., Thompson, M.C., Brewster, A.S., Van Benschoten, A.H., Baxter, E.L., Uervirojnangkoorn, U., McPhillips, S.E., Song, J., Alonso-Mori, R., Holton, J.M., Weis, W.I., Brunger, A.T., Soltis, S.M., Lemke, H., Gonzalez, A., Sauter, N.K., Cohen, A.E., van den Bedem, H. Thorne, R.E. & Fraser, JS. (2015). Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *eLife*. **4**: e07574.

CAPÍTULO IV

SEMEJANZA CUÁNTICA

4.1. LA MECÁNICA CUÁNTICA

La mecánica cuántica es uno de los pilares fundamentales de la física moderna. Ella suministra una manera factible de asociar una función descriptiva a cada molécula. Dos términos de la mayor relevancia en esta teoría son el de observable y el de vector de estado [1]. Un observable es una variable dinámica que puede ser medida. En el marco de la mecánica cuántica, los observables se encuentran a partir de operadores lineales autoadjuntos. Un vector de estado es aquel que determina de manera única el estado del sistema. Todo estado cuántico está representado por un vector normalizado que pertenece a un espacio de Hilbert [2]. P.A.M. Dirac introdujo una notación concisa conocida como la notación bra-ket la cual permite escribir el vector de estado como un *ket*, $|\Psi\rangle$. De acuerdo con la interpretación probabilística de la función de onda, $|\psi(\mathbf{r})|^2 d\mathbf{r}$ representa la probabilidad de encontrar la partícula, en el elemento de volumen $d\mathbf{r}$. Como consecuencia, la probabilidad de encontrar la partícula en todo el espacio será la unidad.

Además, el cuadrado del módulo de la función de onda es un número real que está vinculado a un significado físico: se le puede interpretar como la densidad de probabilidad de que las partículas se hallen en una determinada región del espacio [3]. La teoría de la semejanza cuántica permite encontrar descriptores de semejanza calculados a partir de la función de densidad. A su turno estos descriptores de semejanza han encontrado una importante aplicación en la

predicción teórica de las propiedades de los compuestos químicos. Otro término crucial, tanto en la teoría del enlace químico como en la semejanza cuántica, es la integral de solapamiento la cual da una medida de la superposición de dos orbitales atómicos en sendos átomos A y B [4]. Por ejemplo, para la más sencilla de las moléculas, el H_2^+ , los orbitales 1s en los átomos A y B son $\psi_A(\mathbf{r})$ y $\psi_B(\mathbf{r})$ respectivamente, la integral de solapamiento se define como:

$$S \equiv \langle \psi_A | \psi_B \rangle = \int \psi_A^*(\mathbf{r}) \psi_B(\mathbf{r}) d^3\mathbf{r} \quad (\text{ec. 4.1})$$

donde \mathbf{r} es la posición del electrón en la molécula

Si un sistema está en un estado descrito por una función de onda ψ , el valor de expectación del observable correspondiente al operador \hat{A} es:

$$\langle A \rangle = \frac{\langle \psi | \hat{A} | \psi \rangle}{\langle \psi | \psi \rangle} \quad (\text{ec. 4.2})$$

4.1.1 Postulados de la mecánica cuántica. El desarrollo axiomático de la mecánica cuántica parte de los siguientes postulados:

Postulado 1. *Asociado con cada observable físico hay un operador \hat{A} correspondiente a partir del cual se pueden deducir los resultados de la medición del observable.*

Se asume que cada operador es lineal y que satisface una ecuación de la forma $\hat{A} \psi_n = a_n \psi_n$, en la cual los eigenvalores a_n son números reales y las eigenfunciones ψ_n forman un conjunto ortogonal completo en el espacio de la función de estado. Se garantiza que los eigenvalores, los cuales pueden tomar valores discretos o existen en un rango continuo de valores, sean reales (y por ende mensurables) si el operador correspondiente es Hermitiano. Se nota

también que, en general, las eigenfunciones mismas son complejas y por ello no directamente mensurables.

Postulado 2. *En un sistema particular el único posible resultado de una medición de un observable físico asociado con el operador \hat{A} es un eigenvalor del operador \hat{A} .*

Postulado 3. *Para cada sistema existe siempre un vector de estado ψ que contiene toda la información que se conoce del sistema.*

El vector de estado ψ puede ser empleado para hallar la probabilidad relativa de obtener el eigenvalor a_n asociado con el operador \hat{A} para un sistema particular en un tiempo dado.

Postulado 4. *La evolución temporal del vector de estado ψ está determinado por una ecuación con derivadas parciales conocida como la ecuación de Schrödinger la cual se amplía a continuación.*

4.1.2 La ecuación de Schrödinger.

La energía, uno de los observables en la mecánica cuántica, y su operador asociado se relacionan mediante la ecuación de Schrödinger la cual gobierna la variación que sufre el estado físico de un sistema a lo largo del tiempo [5]:

$$\hat{H} |\psi(t)\rangle = i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle \quad (\text{ec. 4.3})$$

Donde $|\Psi(t)\rangle$ es el vector de estado y \hat{H} es el operador Hamiltoniano. A fin de obtener una solución físicamente relevante de la ecuación de Schrödinger, la función de onda debe ser continua, univaluada, normalizable y antisimétrica con respecto al intercambio de electrones.

El Hamiltoniano contiene las operaciones asociadas con las energías cinética y potencial, por lo tanto se expresa mediante la suma:

$$\hat{H} = \frac{-\hbar^2}{2\mu} \langle \psi | \sum_{i=1}^N \nabla_i^2 | \psi \rangle + V(\mathbf{r}, t) \quad (\text{ec. 4.4})$$

Donde μ es la masa reducida de la partícula, V la energía potencial y ∇ es el operador Laplaciano que efectúa las derivadas de segundo orden sobre las coordenadas del electrón i -ésimo [6].

La solución de onda estacionaria de la ecuación de Schrödinger dependiente del tiempo es:

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r}) e^{-\frac{iEt}{\hbar}} \quad (\text{ec. 4.5})$$

La cual lleva a la forma independiente del tiempo de la ecuación de Schrödinger.

$$\hat{H}\psi(\mathbf{r}) = E\psi(\mathbf{r}) \quad (\text{ec. 4.6})$$

En los casos donde \hat{H} sea independiente del tiempo, el vector de estado estacionario puede descomponerse en una combinación lineal de los eigenestados del Hamiltoniano que resuelven la ecuación de Schrödinger:

$$|\psi(t)\rangle = \sum_{n=1}^N e^{\frac{i}{\hbar}E_n t} |\phi_n(t)\rangle \quad (\text{ec. 4.7})$$

Donde $|\phi_n\rangle$ son los valores del eigenestado del Hamiltoniano con energía E_n . Así, la ecuación de Schrödinger independiente del tiempo toma la forma:

$$\hat{H} |\phi_n(t)\rangle = E_n |\phi_n(t)\rangle \quad (\text{ec. 4.8})$$

La ecuación de Schrödinger así formulada no está embebida en un marco espacio-temporal 4D y se aplica con bajo la condición de contorno de energías cinéticas $\langle \psi | \hat{T} | \psi \rangle$ muy inferiores a la energía en reposo de los electrones [7].

Los sistemas atómicos consisten en electrones y núcleos, por ende la **Ec. 4** toma la forma:

$$\hat{H} = - \sum_i^{elec} \frac{\hbar^2}{2m_e} \nabla_i^2 + \sum_{i<j}^{elec} \frac{q_e^2}{r_{ij}} - \sum_i^{elec} \sum_A^{núcleos} Z_A \frac{q_e^2}{r_{iA}} - \sum_A^{núcleos} \frac{\hbar^2}{2m_A} \nabla_A^2 + \sum_{A<B}^{núcleos} Z_A Z_B \frac{q_e^2}{r_{AB}}$$

(ec. 4.9)

donde m_e y q_e son la masa y la carga de un electrón, m_A y Z_A son la masa y el número atómico del núcleo A. En los softwares que utilizan los químicos, el Hamiltoniano arriba mostrado casi nunca es utilizado. En su lugar se emplea la aproximación de Born–Oppenheimer la cual simplifica la solución de la **Ec. 6** asumiendo que el acoplamiento de los electrones y núcleos es despreciable. La aproximación de Born–Oppenheimer se expresa así:

$$- \left(\sum_i^{elec} \frac{\hbar^2}{2m_e} \nabla_i^2 + \sum_{i<j}^{elec} \frac{q_e^2}{r_{ij}} - \sum_i^{elec} \sum_A^{núcleos} Z_A \frac{q_e^2}{r_{iA}} \right) \psi(\mathbf{r}, \mathbf{R}) + \sum_{A<B}^{núcleos} Z_A Z_B \frac{q_e^2}{r_{AB}}$$

$$= E_e(\mathbf{R}) \psi(\mathbf{r}, \mathbf{R}) \quad (ec. 4.10)$$

Para que la ecuación de Schrödinger de un sistema de n cuerpos sea soluble, es necesario recurrir a aproximaciones. Existe un amplio rango de métodos concernientes a la determinación de la estructura electrónica, desde el muy aproximado enfoque de enlace-ajustado [8] al método esencialmente exacto de la interacción de configuración total [9]. Uno de estos enfoques aproximados es la Teoría del Funcional Densidad (DFT, por sus siglas en inglés) el cual es el método mecano cuántico subyacente en la medición de la semejanza cuántica.

4.2 FUNCIONALES DE DENSIDAD DF

La semejanza molecular es un concepto basado en la información molecular y toda información acerca de una molécula está contenida en la densidad electrónica molecular. La rama de la química que trata sobre la manera en que las moléculas representan, almacenan, procesan y exhiben información se llama informática molecular. Evidentemente, la informática molecular descansa fuertemente sobre la densidad electrónica. El papel central de la densidad electrónica en la informática molecular subraya la importancia del teorema de Hohenberg–Kohn: densidad electrónica del estado basal no degenerado de una molécula determina la energía molecular y a través del Hamiltoniano, todas las otras propiedades moleculares [10].

En el marco de la teoría DFT, se utilizan funcionales cuyo argumento son las coordenadas espaciales y cuya imagen es la densidad electrónica. Y también aplicaciones que convierten la densidad electrónica en la energía total del estado basal para lo cual se emplea un procedimiento variacional que minimiza la energía con respecto al funcional de densidad (DF por su sigla en inglés) [11]. La base teórica para la DFT fue dada por Hohenberg, Kohn y Sham quienes demostraron que la densidad de cualquier sistema determina todas sus propiedades en su estado basal y que a partir del funcional de la densidad es posible escribir una ecuación para orbitales los electrónicos, de los cuales se obtiene la densidad.

En esta formulación, la densidad electrónica se expresa como una combinación lineal de funciones base matemáticamente similares a los orbitales HF. Después se forma un determinante a partir de estas funciones, llamados los orbitales s Kohn-Sham. Lo que se utiliza para calcular la energía es la densidad electrónica de este determinante de orbitales. Este procedimiento es necesario porque los sistemas fermiónicos solo pueden tener densidades electrónicas que se derivan de unan función de onda antisimétrica [12].

La Teoría del Funcional de Densidad tiene como objetivo encontrar la densidad electrónica del estado basal en lugar de la función de onda. Para calcular E a partir de la ecuación de Schrödinger se debería resolver:

$$E = \text{Min}_\psi \langle \psi | \hat{H} | \psi \rangle \quad (\text{ec. 4.11})$$

En DFT, se sigue un enfoque diferente [13]. En lugar de focalizar ψ se hace uso de la densidad de una partícula $\rho(\mathbf{r})$, la cual está relacionada cuantitativamente con ψ mediante la ecuación:

$$\rho(\mathbf{r}) = \int |\psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 d\mathbf{r}_2, \dots, d\mathbf{r}_N \quad (\text{ec. 4.12})$$

La densidad es una cantidad bastante más sencilla que ψ ya que depende solo de las tres coordenadas espaciales en lugar de $3N$ variables, reduciendo enormemente la complejidad de la tarea. Los principios de Hohenberg–Kohn prueban que la densidad electrónica es la cantidad que determina las interacciones electrónicas y forma la base de una expresión exacta del estado electrónico basal.

4.2.1 Los principios de Hohenberg–Kohn.

El lema básico de Hohenberg y Kohn [14] enuncia que la densidad electrónica del estado basal de un sistema de electrones interactuantes en un potencial externo arbitrario determina unívocamente este potencial. La demostración está dada mediante el principio variacional. Si se considera un Hamiltoniano \hat{H}_1 de un potencial externo V_1 como

$$\hat{H}_1 = \hat{T} + \hat{U} + \hat{V}_1 \quad (\text{ec. 4.13})$$

donde \hat{T} es el operador de energía cinética y \hat{U} es el operador de interacción electrón–electrón. La solución de la ecuación de Schrödinger

$$\hat{H}_1 \psi = E\psi \quad (\text{ec. 4.14})$$

es la función de onda del estado basal ψ_1 , la cual corresponde a la densidad electrónica ρ_1 . Por lo tanto, la energía del estado es:

$$E_1 = \langle \psi_1 | \hat{H}_1 | \psi_1 \rangle = \int V_1(\mathbf{r})\rho(\mathbf{r}) + \langle \psi_1 | \hat{T} + \hat{U} | \psi_1 \rangle \quad (ec. 4.15)$$

Considerando otro potencial V_2 , el cual no puede ser obtenido como $V_1 +$ constante, con una función de onda del estado basal ψ_2 , que genera la misma densidad electrónica, la energía del estado es:

$$E_2 = \int V_2(\mathbf{r})\rho(\mathbf{r}) + \langle \psi_2 | \hat{T} + \hat{U} | \psi_2 \rangle \quad (ec. 4.16)$$

De acuerdo con el principio variacional:

$$E_1 < \langle \psi_2 | \hat{H}_1 | \psi_2 \rangle = \int V_1(\mathbf{r})\rho(\mathbf{r}) + \langle \psi_2 | \hat{T} + \hat{U} | \psi_2 \rangle = E_2 + \int [V_1(\mathbf{r}) - V_2(\mathbf{r})]\rho(\mathbf{r}) \quad (ec. 4.17)$$

y

$$E_2 < \langle \psi_1 | \hat{H}_2 | \psi_1 \rangle = \int V_2(\mathbf{r})\rho(\mathbf{r}) + \langle \psi_1 | \hat{T} + \hat{U} | \psi_1 \rangle = E_1 + \int [V_2(\mathbf{r}) - V_1(\mathbf{r})]\rho(\mathbf{r}) \quad (ec. 4.18)$$

Sumando las inecuaciones 4.16 y 4.17, se llega a una contradicción:

$$E_1 + E_2 < E_1 + E_2 \quad (ec. 4.19)$$

Esta es la demostración indirecta de que dos potenciales externos diferentes no pueden generar la misma densidad electrónica. El segundo teorema de Hohenberg–Kohn establece un vínculo entre la energía total y la densidad electrónica, afirmando que existe un funcional de energía universal, el cual es válido para cada potencial externo y su mínimo global corresponde al estado basal del sistema y a la densidad electrónica del estado basal [15]. Para

demostrar este teorema, se escribe el funcional de la energía total como un funcional universal:

$$E[\rho] = F_{HK}[\rho] + \int V(\mathbf{r})\rho(\mathbf{r}) + E_{ZZ} \quad (ec. 4.20)$$

donde F_{HK} aplica a cada sistema electrónico. It determines the entire electronic energy except the energy due to the external potential $V(\mathbf{r})$. E_{ZZ} es la interacción entre los núcleos. La energía del estado basal está dada por:

$$E = \langle \psi | \hat{H} | \psi \rangle = E[\rho] \quad (ec. 4.21)$$

De acuerdo con el principio variacional, al cambiar la función de onda a una ψ' diferente, que a su vez corresponde a una densidad electrónica diferente ρ_0 , la energía resultante:

$$E < E' = \langle \psi' | \hat{H} | \psi' \rangle \quad (ec. 4.22)$$

es mayor que E , así que ρ no puede corresponder al estado basal exacto.

La función de onda del estado basal se puede encontrar a partir del principio variacional:

$$E = \min_{\psi'} \langle \tilde{\psi} | \hat{T} + \hat{U} + V | \tilde{\psi} \rangle \quad (ec. 4.23)$$

donde $\tilde{\psi}$ es una función de onda de ensayo. El principio variacional puede ser reformulado en términos de las densidades de ensayo, $\tilde{\rho}$:

$$E = \min_{\tilde{\rho}} E[\tilde{\rho}] \quad (ec. 4.24)$$

donde la minimización se efectúa con respect a las distribuciones de densidad que preservan n , el número total de electrones. Conceptualmente, la **ec. 4.24**

es una gran simplificación si se compara con la **ec. 4.11**. El funcional de energía de Kohn-Sham viene entonces dado por:

$$E^{KS} = -\frac{1}{2} \sum_i \int d\mathbf{r} \psi_i(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) + \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[\rho(\mathbf{r})] + \int d\mathbf{r} V(\mathbf{r})\rho(\mathbf{r}) \quad (ec. 4.25)$$

Donde E_{xc} es el funcional de energía de intercambio-correlación, $V(\mathbf{r})$ es el potencial externo creado por los iones y $\rho(\mathbf{r})$ la densidad electrónica. Esta última corresponde a:

$$\rho(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2 \quad (ec. 4.26)$$

En la práctica funcional de energía de la **ec. 4.25** se usa así [16]:

$$E_v[\rho] = \int V(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + F[\rho] \quad (ec. 4.27)$$

El primer término en el lado derecho de la **ec. 4.27** representa la energía de interacción de los electrones con el potencial externo V . El segundo término es la energía de Coulomb de una distribución de densidad ρ . La cantidad $F[\rho]$ es un funcional universal de la densidad, que significa que está especificado unívocamente por la densidad ρ de los electrones interactuantes y no depende del potencial externo particular V que actúa sobre los electrones.

Para minimizar E_v con respecto a ψ en lugar de hacerlo con respecto ρ se logra resolviendo la ecuación Kohn-Sham:

$$\left(-\frac{\nabla^2}{2} + V(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{xc}[\mathbf{r}] \right) \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}) \quad (ec. 4.28)$$

donde ε_i son los eigenvalores de la matriz de multiplicadores de Lagrange y se conocen como los eigenvalores de Kohn-Sham.

En muchos casos, los resultados a partir de cálculos QM son los únicos datos disponibles para la simulación de una reacción enzimática. No obstante, estos datos deben ser empleados con cautela [17]. Adicionalmente, los datos QM, incluyendo los DFs, típicamente son obtenidos en la fase gas y pueden ocurrir cambios significativos en las geometrías, vibraciones y energética conformacional al pasar de la fase gas a la fase condensada.

El algoritmo de cálculo de los DFs debe involucrar los siguientes pasos [18]:

- 1) Resolver la *ec. 4.28*
- 2) Cálculo de la densidad electrónica mediante electrones no interactuantes ficticios.
- 3) Cálculo de la energía del estado basal en conformidad con:

$$E_v[\rho] = \int V(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + 2 \sum_i^N \left\langle \psi_i \left| -\frac{\nabla^2}{2} \right| \psi_i \right\rangle + E_{xc}[\rho] \quad (ec. 4.29)$$

Se puede considerar a los DFs tanto como aplicaciones como operadores. Por lo tanto, a partir de dos DFs, correspondientes a dos moléculas, se han extraído coeficientes numéricos que expresan el grado de semejanza entre los sistemas comparados [19]. Esto se puede efectuar por medio de una técnica del valor de expectación estadístico, análogo al procedimiento que se utiliza para obtenerlo dentro de un estado de un sistema único.

4.3 VALORES DE EXPECTACIÓN

En sus aspectos teóricos y computacionales, la semejanza cuántica está basada en el esquema de los valores de expectación de la *ec. 4.2*. Si se conoce el DF de un estado, se pueden extraer de él todos los valores de las

propiedades observables del sistema, ω , como valores de expectación, $\langle \omega \rangle$, del operador Hermitiano asociado, Ω [20]. La manera estadística habitual de calcular los valores de expectación de los operadores, es expresándolos mediante una integral definida dentro del espacio al cual pertenecen tanto el operador $\Omega(\mathbf{r})$ como la función de densidad del estado del sistema, ρ , sobre la cual actúa el operador. También equivalentemente esta integral se expresa mediante un producto escalar:

$$\langle \omega \rangle \equiv \int \Omega(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} = \langle \Omega | \rho \rangle \quad (\text{ec. 4.30})$$

donde \mathbf{r} representa una matriz de coordenadas de las partículas [21].

4.4 MEDICIÓN DE LA SEMEJANZA CUÁNTICA MOLECULAR

Las mediciones de la semejanza cuántica se basan en el principio de similitud: “cuanto más similares son dos moléculas, más similares son sus propiedades” [22]. Se puede cuantificar la similitud entre las moléculas por medio del valor de expectación o comparando sus DFs correspondientes. La Medida de Semejanza Cuántica Molecular (MQSM por sus siglas en inglés) se puede ponderar mediante varios índices, entre ellos el operador $\Omega(\mathbf{r}_1, \mathbf{r}_2)$. Este operador es definido a través de un producto interno o equivalentemente a partir de los funcionales de densidad, los cuales se multiplican e integran sobre las coordenadas electrónicas en un dominio conveniente [23]:

$$Z_{AB}(\Omega) = \langle \rho_A | \Omega | \rho_B \rangle = \iint \rho_A(\mathbf{r}_1) \Omega(\mathbf{r}_1, \mathbf{r}_2) \rho_B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (\text{ec. 4.31})$$

Donde A y B son las dos moléculas comparadas, $\{\mathbf{r}_1, \mathbf{r}_2\}$ las coordenadas electrónicas asociadas con las correspondientes funciones de onda, $\{\rho_1, \rho_2\}$ las correspondientes densidades de probabilidad de los electrones y $\Omega(\mathbf{r}_1, \mathbf{r}_2)$ un operador que depende de las coordenadas de los electrones. La integral

anterior resulta siempre en un número real positivo debido a la presencia del operador definido positivo y el DF.

Otros índices son el operador de superposición, el Operador coulombico, la autosemejanza y el índice de Carbó. El operador más usado en el cálculo de la SMC es la función delta de Dirac, $W = \delta(r_1, r_2)$ el cual tiene la forma [24]:

$$\begin{aligned} Z_{AB} &= \iint \rho_A(\mathbf{r}_1) \delta(\mathbf{r}_1 - \mathbf{r}_2) \rho_B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \\ &= \int \rho_A(\mathbf{r}) \rho_B(\mathbf{r}) d\mathbf{r} \quad (\text{ec. 4.32}) \end{aligned}$$

Esta cantidad es una medida de la superposición entre dos densidades electrónicas.

Además del operador de superposición, el otro operador ampliamente usado es el operador Coulombico [25]:

$$Z_{AB} = \iint \rho_A(\mathbf{r}_1) |\mathbf{r}_1 - \mathbf{r}_2|^{-1} \rho_B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (\text{ec. 4.33})$$

La autosemejanza resulta de la comparación de dos densidades electrónicas idénticas mediante los elementos diagonales de la matriz de semejanza molecular:

$$Z_{AA} = \iint \rho_A(\mathbf{r}_1) \Omega(\mathbf{r}_1, \mathbf{r}_2) \rho_A(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (\text{ec. 4.34})$$

Notablemente, las medidas de autosemejanza han sido usadas recientemente en la identificación de sitios activos en farmacología.

Un índice de semejanza es una medida normalizada de la similitud molecular. En el programa SIMILARITY se implementó el índice de Carbó [26], C_{AB} , el cual se define como:

$$C_{AB} = \frac{Z_{AB}}{(Z_{AA}Z_{BB})^{\frac{1}{2}}} \quad (\text{ec. 4.35})$$

Para la definición de semejanza molecular de la **ec. 4.8**, este índice está comprendido entre cero y uno, $C_{AB} \hat{=} (0,1)$. Al igual que las medidas de similitud molecular, los índices de similitud para un conjunto de n moléculas pueden ser representados en una matriz cuadrada y simétrica de orden n , $C = [C_{AB}]$. En este caso, la matriz C tendría unos en la diagonal, y los elementos no diagonales estarían entre uno y cero.

4.4.1 Aproximación de la Capa Atómica (ASA).

Un cálculo de la MQSM mediante la **ec. 25** puede presentar serias dificultades en el caso del DF de una molécula grande o cuando se aplican niveles altos de cálculo computacional. Una manera simplificada de construir los DFs para la implementación práctica de la QSM es la llamada Aproximación ASA (Atomic Shell Approximation) [27]. La técnica ASA permite obtener los funcionales de densidad atómica de manera bastante precisa, ajustando un conjunto de funciones de densidad esférica a las densidades atómicas *ab initio*. La función de densidad en la forma ASA puede ser escrita en términos del conjunto de contribuciones de las funciones atómicas $\{\sigma_a\}$:

$$\rho_A^{PASA}(\mathbf{r}) = \sum_{a \in A} \sigma_a(\mathbf{r}) \quad \text{ec. 4.35}$$

donde la suma abarca todos los átomos presentes en la molécula A.

Al mismo tiempo, el conjunto atómico de funciones $\{\sigma_a\}$, se puede construir empleando un conjunto de capas atómicas $\{s_i\}$, por medio de la suma:

$$\sigma_a(\mathbf{r}) = \sum_{i \in A} s_i(\mathbf{r}) \quad \text{ec. 4.36}$$

Y la suma se efectúa sobre todas las capas atómicas del átomo a -ésimo. El conjunto de funciones esféricas $\{s_i\}$, se define como:

$$s_i(\mathbf{r}) = \sum_{k \in i} c_k \varphi_k(\mathbf{r}) \quad \text{ec. 4.37}$$

donde la suma se lleva a cabo sobre el conjunto de funciones positivas $\{\varphi_k\}$, que pertenecen a la capa i .

Análogamente, en el modelo PASA (promolecular atomic Shell approximation) la densidad *ab initio*, se expresa como una combinación lineal de funciones 1S [28]:

$$\rho_A^{PASA}(\mathbf{r}) = \frac{1}{P_A} \sum_{a \in A} P_a \rho_a^{PASA}(\mathbf{r} - \mathbf{R}_a) \quad \text{ec. 4.38}$$

donde P_a es el número atómico de cada átomo presente en la molécula A y P_A es el número total de electrones. De esta manera, se considera la densidad molecular como una adición de densidades atómicas discretas $\rho_a^{PASA}(\mathbf{r} - \mathbf{R}_a)$, colapsadas en las coordenadas atómicas \mathbf{R}_a , las cuales a su vez se expresan como:

$$\rho_a^{PASA}(\mathbf{r} - \mathbf{R}_a) = \sum_{i \in A} w_i |S_i(\mathbf{r} - \mathbf{R}_a)|^2 \quad \text{ec. 4.39}$$

De esta manera, se calcula la **ec. 25** utilizando solo funciones gaussianas 1S, reduciendo el costo computacional y por lo tanto ampliando el campo de aplicación de la MQSM.

4.4.2 Matriz de Semejanza.

Una bella característica del índice de semejanza es que compara las propiedades 3D de parejas de moléculas arrojando una única cifra [29]. En el caso de una serie de n moléculas M se puede calcular la MQSM comparando todas las posibles parejas en el conjunto a través de la Matriz de Semejanza (SM) $n \times n$ la cual se define como [30]:

$$\mathbf{Z} = \{Z_{ij}(\Omega) \mid \forall i, j \in M\} \quad \text{ec. 4.40}$$

Se puede considerar esta matriz como un hipervector formado por un conjunto de vectores columna:

$$\mathbf{Z} = \{ \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n \} \quad \text{ec. 4.41}$$

Donde cada vector columna está formado por todas las MQSM relacionadas con la molécula α [31]. La matriz de semejanza para n moléculas puede representarse así:

$$\mathbf{Z} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & \dots & i & \dots & n \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ j \\ \vdots \\ n \end{matrix} & \left[\begin{array}{ccccccc} z_{11} & z_{12} & z_{13} & \dots & z_{1i} & \dots & z_{1n} \\ & z_{22} & z_{23} & \dots & z_{2i} & \dots & z_{2n} \\ & & z_{33} & \ddots & z_{3i} & \dots & z_{3n} \\ & & & \ddots & & & \vdots \\ & & & & z_{ii} & \ddots & z_{jn} \\ & & & & & \ddots & \vdots \\ \vdots & \vdots & & & & & z_{nn} \end{array} \right] & \text{ec. 4.42} \end{matrix}$$

$$\begin{matrix} \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ \mathbf{z}_1 & \mathbf{z}_2 & \mathbf{z}_3 & & \mathbf{z}_i & & \mathbf{z}_n \end{matrix}$$

Cada una de las columnas o filas de la SM se puede considerar la colección de todas las QSM entre la i -ésima molécula y cada otra del conjunto, incluyéndola a ella misma. Consecuentemente, cada vector \mathbf{z}_i se puede interpretar como una representación discreta N-dimensional de la i -ésima molécula y puede ser relacionada con una proyección de la correspondiente función de densidad en un subespacio generado por la función de densidad del conjunto de moléculas.

Tales colecciones de columnas de la SM se pueden considerar como un conjunto de descriptores moleculares y ser utilizadas como insumo del QSAR cuántico.

4.5. REFERENCIAS

1. Levi, A. F. J. (2003). Applied Quantum Mechanics. Cambridge University Press. Pag 245
2. Ohya, M., Volovich, I. (2011). Mathematical Foundations of Quantum Information and Computation and Its Applications to Nano- and Bio-systems. Springer. Pag 70.
3. Doucet, J-P., Weber, J. (1996). Computed-aided molecular design. Theory and applications. Academic Press. Pag 308.
4. Saspe, A-M. (1998). Molecular Orbital Calculations for Biological Systems. Oxford University Press. Pag 17.
5. Field MJ. (2007). A practical introduction to the simulation of molecular systems. 2a Edición. Cambridge University Press. Pag. 52.
6. Cramer CJ. (2004). Essentials of computational chemistry. Theories and Models. 2a Edición. Ed. Wiley. Pag 107.
7. Wei Y. (2015). Some Solutions to the Fractional and Relativistic Schrödinger Equations. *International Journal of Theoretical and Mathematical Physics*. **5** (5): 87-111.
8. Martin R.M. (2004). Electronic Structure. Basic Theory and Practical Methods. 1a Edición. Cambridge University Press. Pag 120.
9. Hohenberg, P., Kohn, W. (1964). Inhomogeneous electro gas. *Phys. Rev.* **136**: B864 – 871.
10. Wesolowski, T.A. Hohenberg-Kohn-Sham Density Functional Theory. Capítulo Molecular Materials with Specific Interactions – Modeling and Design. Volumen 4 de la serie Challenges and Advances in Computational Chemistry and Physics. 2007. Pág. 153-201

11. Mc. Weeny, R. *Methods of Molecular Quantum Mechanics*; Academic Press: London, 1978.
12. Kohn, W. (1998). Nobel lecture: Electronic structure of matter—wave functions and density functionals. *Rev. Mod. Phys.* **71**: 1253-1266.
13. Jones, R.O. & Gunnarsson, O. (1989). The density functional formalism, its applications and prospects. *Rev. Mod. Phys.* **61**: 689.
14. Sahni, V. *Quantal Density Functional Theory*. Springer. 2004, Pág 99.
15. Perdew, J. P., Ruzsinszky, A., Tao, J., Staroverov, V. N., Scuseria, G. E. & Csonka, G. I. (2005). Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. *J. Chem. Phys.* **123**: 062201.
16. Becke, A.D. (2014). Perspective: Fifty years of density-functional theory in chemical physics. *J. Chem. Phys.* **140**: 18A301.
17. Comba, P. *Modeling of Molecular Properties*. Wiley. 2011. Pág. 17.
18. Sholl, D., Steckel, JA. *Density functional theory. A practical introduction*. Wiley. Pag 44. 2009.
19. Mc. Weeny, R. *Methods of Molecular Quantum Mechanics*; Academic Press: London, 1978.
20. Carbó-Dorca, Besalú, R., Gironés, X. (2000). Extended density functions. *Advances in Quantum Chemistry*. **38**: 1-63.
21. Carbó-Dorca, R.; Amat, L.; Besalu, E.; Gironés, X.; Robert, D. *Fundamentals of Molecular Similarity*; Kluwer/Academic/Plenum: New York, 2001. Capítulo 12, 187–320.
22. Gallegos S. (2004). *Molecular Quantum similarity in QSAR: Applications in Computer-aided Molecular Design*. Tesis. Pag 52.

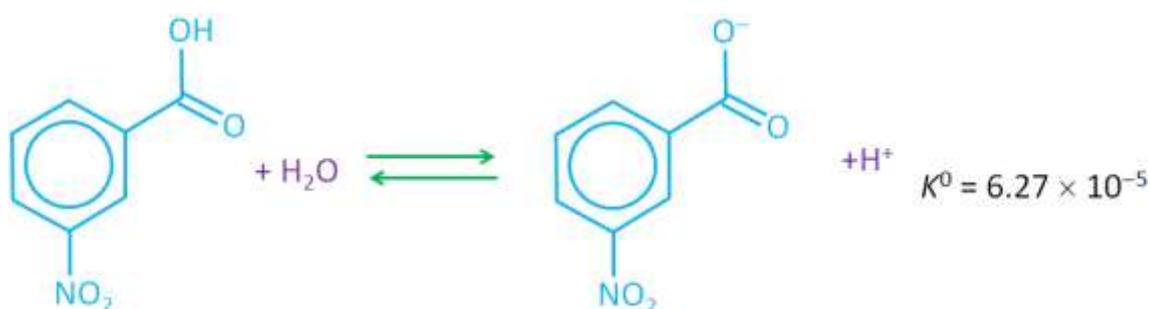
23. Carbó-Dorca, R., Leyda, L., Arnau, N. (1980). How similar is a molecule to another? An electron density measure of similarity between two molecular structures. *Int. J. Quantum Chem.* **17** (6): 1185 – 1189.
24. Ponec, R., Amat, L., Carbó-Dorca, R. (1999). Molecular basis of quantitative structure–properties relationships (QSPR): a quantum similarity approach. *J Comput-Aided Mol Des.* **13**: 259–270.
25. Carbó, R., Arnau, M. & Leyda, L. (1980). *Int. J. Quant. Chem.* **17**: 1185-1189.
26. Besalú, E., Gironés, X., Amat, LI. & Carbó –Dorca, R. (2002). Molecular Quantum Similarity and the Fundamentals of QSAR. *Acc. Chem. Res.* **35**: 289-295
27. Constans, P. & Carbó, R. (1995). Atomic shell approximation: electron density fitting algorithm restricting coefficients to positive values. *J Chem Inf Comput Sci.***35**:1046–1053.
28. Amat, L., y Carbó-Dorca, R. (1997). Quantum similarity measures under Atomic Shell Approximation: first order density fitting using elementary jacobi rotations. *J. Comput. Chem.* **18** (16): 2023-2039.
29. Richards, WG. (1995). Molecular Similarity and Dissimilarity. *Molecular Engineering.* 5 (1): 229–233.
30. Carbó –Dorca, R., Robert, D., Amat, LI., Gironés, X. & Besalú, E. *Molecular Quantum similarity in QSAR and drug design.* Springer. 2000. Pag 14.
31. Bultinck, P., De Winter, H., Langenaeker, W., Tollenaere, JP. *Computational Medicinal Chemistry for Drug Discovery.* Marcel Dekker. 2004. Pag 368.

CAPÍTULO 5

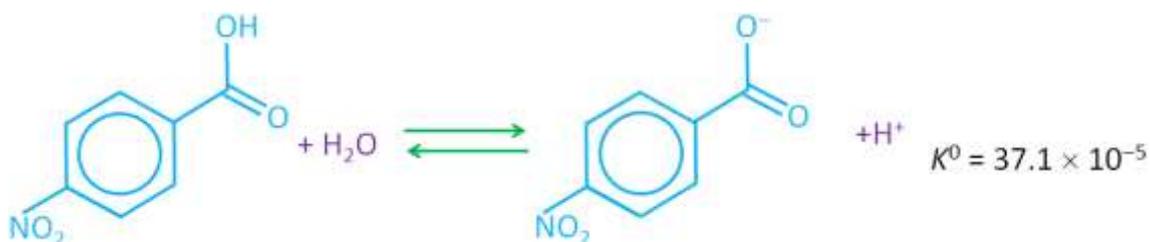
MODELOS QSAR

5.1 EL MODELO DE HAMMETT

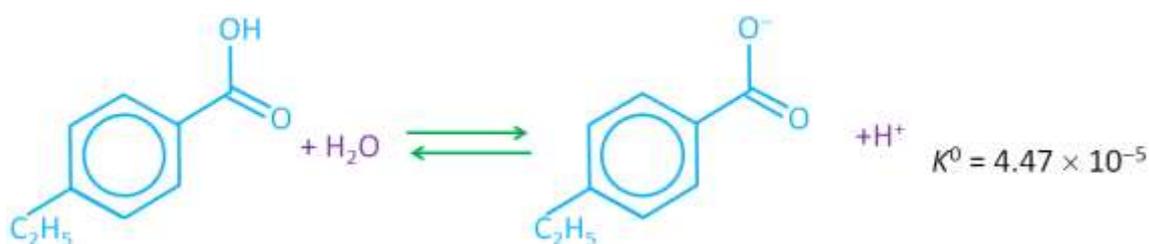
Hammett desarrolló una ecuación que describe una relación lineal de la energía de Gibbs de ciertos compuestos orgánicos con las velocidades de reacción y las constantes de equilibrio de algunas reacciones [1]. En 1937, Hammett publicó un artículo sobre los “Efectos de la estructura en las reacciones de derivados del benceno”. La reacción de disociación del ácido benzoico es:



Hammett se dio cuenta de que la adición de un sustituyente en el anillo aromático tiene un efecto cuantitativo en la constante de disociación K . La reacción de disociación del ácido 4-nitrobenzoico y su constante son:



El incremento de la constante indica un incremento en la estabilización de la carga negativa generada por la disociación. Si el sustituyente es un grupo etilo en posición *para*, se observa el efecto opuesto con respecto a la reacción no sustituida



Indicando que el grupo etilo desestabiliza la carga negativa generada en la disociación. Hammett encontró, por ejemplo, que el grupo nitro tiene un efecto estabilizante sobre otras reacciones de disociación las del ácido fenilacético. En consecuencia propuso una relación cuantitativa que daba cuenta de estos hallazgos [2]

$$\log K = \log K^0 + \rho \sigma \quad ec.5.1$$

La misma clase de expresión también es válida para las constantes de velocidad, k

$$\log k = \log k^0 + \rho \sigma \quad ec.5.2$$

K^0 y k^0 denotan las constantes correspondientes para el compuesto "padre" o no sustituido. La *constante del sustituyente* σ es una medida del efecto electrónico de reemplazar H por un sustituyente dado (en la posición *para* o *meta*) y es, en principio, independiente de la naturaleza de la reacción. La *constante de reacción* ρ depende de la naturaleza de la reacción, incluyendo condiciones como el solvente y la temperatura; ρ es una medida de la

susceptibilidad de la reacción a los efectos electrónicos de los sustituyentes. La reacción de referencia es la ionización del ácido benzoico no sustituido en agua a 25 °C, con $\rho = 1$.

Así es claro que solo unos pocos valores de σ y ρ pueden resumir un extenso compendio de mediciones de equilibrios y velocidades y puede ayudar a predecir los coeficientes de velocidad y las constantes de equilibrio para reacciones que aún no han sido estudiadas. La *constante del sustituyente* σ es un número positivo o negativo. Por ejemplo, $\sigma_p = 0.71$ para $-\text{NO}_2$ indica que este grupo tiene un efecto sustractor de electrones mayor que los átomos de H, mientras que $\sigma_p = -0.66$ para el $-\text{NH}_2$ indica que este grupo tiene un efecto donador de electrones más fuerte que los átomos de H. Las reacciones con $\rho > 0$ son facilitadas por sustituyentes sustractores de electrones.

Estas correlaciones empíricas demandan una explicación. De la ec. (5.2) se puede escribir:

$$\Delta G^{\ddagger 0} = \Delta G^{\ddagger 0}_0 - 2.3 RT \sigma \rho \quad \text{ec.5.3}$$

Donde $\Delta G^{\ddagger 0}$ corresponde a la energía de Gibbs de activación de la reacción de referencia [3]. Para otras reacciones homologas, con el mismo conjunto de sustituyentes, se tiene

$$\Delta G^{\ddagger 0'} = \Delta G^{\ddagger 0'}_0 - 2.3 RT \sigma' \rho \quad \text{ec.5.4}$$

Los modelos de relación cuantitativa estructura-actividad (quantitative structure-activity relationship QSAR) fueron introducidos por Hansch y colaboradores sobre la base de las implicaciones de la relación lineal en la ecuación de Hammett [4]. El enfoque del QSAR se basa en la premisa de que la diferencia en las propiedades fisicoquímicas dan cuenta de las diferencias en la actividad biológica de los compuestos [5]. En los modelos de regresión QSAR se establece una relación entre un conjunto de variables predictoras (X)

y la variable de respuesta (Y). Las propiedades moleculares en una ecuación QSAR son llamados descriptores.

A diferencia del modelo de Hamlet donde los conjuntos de datos correspondían a vectores de dimensión $N \times 1$, a menudo se tienen varios conjuntos de variables \mathbf{x} ($\mathbf{x1}$, $\mathbf{x2}$, $\mathbf{x3}$, ... , \mathbf{xM}), cada uno de los cuales puede potencialmente describir algo de la variación en el conjunto de datos \mathbf{y} [6]. Las variables \mathbf{x} se pueden arreglar en una matriz $N \times M$:

$$\mathbf{X} = \begin{bmatrix} x1_1 & x2_1 & x3_1 & \cdots & xM_1 \\ x1_2 & x2_2 & x3_2 & \cdots & xM_2 \\ x1_3 & x2_3 & x3_3 & \cdots & xM_3 \\ \vdots & \vdots & \vdots & \ddots & \cdots \\ x1_N & x2_N & x3_N & \cdots & xM_N \end{bmatrix} \quad ec. 5.5$$

Si los x descriptores derivan de múltiples fuentes, tendrán diferentes unidades, medias y varianzas. Por lo tanto antes del análisis de correlación, cada vector \mathbf{x} puede ser *centrado* para que su media sea cero.

5.1 ALGORITMOS DE REGRESIÓN

Puesto que se tendrán cientos de puntos de muestreo alrededor de las moléculas, no se podrá emplear el tradicional análisis de regresión para establecer las relaciones entre la actividad biológica y los valores de los campos a través de la red de puntos. En su lugar, será implementada la técnica estadística de los mínimos cuadrados parciales (*Partial Least Squares* PLS, algunas veces llamado también *Proyección a estructuras Latentes*) para efectuar la selección de las variables latentes mediante ponderación de la matriz \mathbf{X} con el vector \mathbf{y} previo a la diagonalización. Por consiguiente, se diagonalizará la matriz $\mathbf{X}^t\mathbf{y}\mathbf{y}^t\mathbf{X}$ en lugar de la $\mathbf{X}^t\mathbf{X}$ que se usa en el análisis de componentes principales. Ello asegura que los eigen vectores con los eigen valores más grandes serán sesgados hacia la descripción de la variación en \mathbf{y} . De este modo será soslayada la necesidad de efectuar una búsqueda combinatoria para los vectores latentes que se utilizarán en la regresión. Para

casos óptimos, un gráfico de Q^2 contra el número de componentes PLS rápidamente alcanzará un máximo y proporcionará un modelo compacto con buena capacidad predictiva. El resultado será una función que podrá ser presentada en contornos 3D [7].

En una regresión de PLS se combinan los rasgos de la regresión de cuadrados múltiples o mínimos cuadrados *Ordinarios* OLS, con el de los componentes principales. En esencia, los PLS calculan las cargas de las propiedades moleculares sobre los componentes, denominados *variables latentes*, para explicar una propiedad dependiente tal como la actividad biológica, en lugar de buscar maximizar la varianza explicada en las propiedades físicas como en el PCA. Cada variable latente sucesiva será ortogonal a todas las variables latentes previas. En otro contraste con el PCA, cada variable latente sucesiva de los PLS aportará una explicación adicional de la varianza en la propiedad dependiente, lo que resultará en un incremento en el R^2 del ajuste [8].

En la regresión de componentes principales, la información en las variables independientes es resumida en un número pequeño de componentes. No obstante, no hay una razón *a priori* de por qué los componentes principales asociados con los valores singulares más grandes deberían ser los más útiles para la regresión. El componente principal PC 1 cubre la varianza más grande, pero aún pudiera tener solo poder predictivo limitado. Puesto que rutinariamente se escogen los componentes principales comenzando con el número 1 y subiendo, hay una posibilidad de incluir variables que de hecho no contribuyan al modelo de regresión. Al comprimir información en X sin importar lo que se va a predecir, nunca se estará seguro de que la parte esencial de los datos se preserva [9].

El PLS constituye una alternativa. Igual que en el PCR, en la PLS se definen variables latentes ortogonales para comprimir la información y descartar lo que sea irrelevante. No obstante, el PLS explícitamente ayuda a construir las variables latentes de una manera tal que a medida que se captura la mayoría de la varianza en X y Y mientras que se maximiza la covarianza entre estas matrices. El algoritmo es ligeramente más complicado que el PCR; en efecto,

existen varios algoritmos casi equivalentes para ejecutar el PLS. Las diferencias radican en pequeñas variaciones en el criterio que es optimizado, diferentes implementaciones para obtener mejoras en la velocidad en situaciones específicas o en opciones diferentes para escalar los resultados intermedios [10].

Igual que en el PCR, el PLS es configura para centrar la media de los datos de manera que no hay necesidad de estimar un vector intercepto; este se obtiene posteriormente. La notación es un tanto más complicada que con el PCR, dado que ambas matrices \mathbf{X} y \mathbf{Y} tienen puntajes y cargas. Más aún, en muchos algoritmos se emplean matrices de peso adicionales. Otra diferencia con el PCR es que los componentes de PLS son extraídos secuencialmente mientras que los PCs en el PCR se puede obtener en un paso SVD. En cada iteración en el algoritmo PLS, la variación asociada con el componente estimado se remueve de los datos en un proceso llamado deflación y el remanente (que se indica con \mathbf{E} para la matriz \mathbf{X} deflatada y \mathbf{F} para la \mathbf{Y} deflatada) se utiliza para estimar el próximo componente. Esto continua hasta que el usuario decide que es suficiente o hasta que todos los componentes han sido estimados [11].

El primer componente se obtiene de un SVD de la matriz de producto cruz $\mathbf{S} = \mathbf{X}^T \mathbf{Y}$, el cual incluye información tanto sobre la variación en \mathbf{X} y \mathbf{Y} como de la correlación entre ambas. El primer vector singular left, w , puede ser visto como la dirección de la máxima varianza en la matriz de producto cruz y usualmente se indica con la descripción algo vaga de "pesos". Las proyecciones de la matriz \mathbf{X} sobre este vector se llaman "puntajes":

$$t = \mathbf{X} w = \mathbf{E} w \quad (\text{ec. 5.6})$$

Eventualmente, estos puntajes t se reúnen en una matriz \mathbf{T} que cumple el mismo papel que la matriz de puntajes en el PCR; es un estimado de baja dimensión de la información en \mathbf{X} . Por lo tanto, una regresión de \mathbf{Y} sobre \mathbf{T} es fácil de hacer y el vector coeficiente para \mathbf{T} se puede convertir en un vector coeficiente para las variables originales [12].

El siguiente paso en el algoritmo es obtener las cargas para \mathbf{X} y \mathbf{Y} mediante la regresión contra el *mismo* vector de puntaje t :

$$p = \frac{E^T t}{t^T t} \quad (\text{ec. 5.7}) \quad y \quad q = \frac{F^T t}{t^T t} \quad (\text{ec. 5.8})$$

Se divide por la suma de todos los elementos cuadrados en t . esto lleva a cargas "normalizadas". No es esencial hacer el escalamiento de esta manera. En efecto, hay numerosas posibilidades para escalar las cargas, pesos o puntajes. Se puede optar por tener los puntajes o las cargas ortogonales. Desafortunadamente, esto puede dificultar la comparación de los puntajes y las cargas en PLS de diferente implementación [13].

Finalmente, las matrices de datos son deflatadas: se sustrae la información relacionada con esta variable latente, en la forma de los productos externos tp^T y tq^T de las matrices de datos (actuales):

$$E_{n+1} = E_n - tp^T \quad y \quad F_{n+1} = F_n - tq^T \quad (\text{ec. 5.9})$$

La estimación del siguiente componente puede iniciar a partir del SVD de la matriz del producto cruz $E_{n+1}^T F_{n+1}$. Después de cada iteración, los vectores w , t , p y q son guardados como columnas en matrices \mathbf{W} , \mathbf{T} , \mathbf{P} y \mathbf{Q} , respectivamente. El algoritmo puede ser resumido de la siguiente forma: los vectores w constituyen la dirección de la mayor parte de la variación en la matriz del producto cruz $\mathbf{X}^T \mathbf{Y}$. Los puntajes t son las coordenadas de los objetos en este eje. Las cargas para \mathbf{X} y \mathbf{Y} se obtienen mediante regresión de ambas matrices contra los puntajes y se remueven los productos de los puntajes y cargas para \mathbf{X} y \mathbf{Y} de las matrices de datos \mathbf{E} y \mathbf{F} .

Una complicación es que las columnas de la matriz \mathbf{W} no se pueden comparar directamente: derivan de las matrices deflatadas sucesivamente \mathbf{E} y \mathbf{F} [14]. Una manera alternativa de representar los pesos, de forma tal que todas las columnas se relacionan con la matriz original \mathbf{X} , está dada por:

$$R = W(P^T W)^{-1} \quad (ec. 5.10)$$

La Matriz R tiene algunas propiedades interesantes, una de ellas es que es una inversa generalizada para P^T . También se preserva que $T = XR$.

En lugar de la regresión de Y en X , Se utilizan los puntajes T para calcular los coeficientes de regresión A y después se convierten estos al reino de las variables originales:

$$Y = \tilde{X}B + \varepsilon = T(P^T B) + \varepsilon = TA + \varepsilon \quad (ec. 5.11)$$

$$A = (T^T T)^{-1} T^T Y \quad (ec. 5.12)$$

$$B = RA \quad (ec. 5.13)$$

A diferencia del PCR, el cálculo de T incluye la información sobre Y [15].

5.3 VALIDACIÓN ESTADÍSTICA

El significado estadístico del R^2 para un ajuste PLS se evaluará utilizando el número de variables latentes incluidas como los p grados de libertad atendiendo los cuadrados de las medias MS según la ecuación [16]:

$$F_{p, n-p-1} = \frac{MS_{regr}}{MS_{error}} \quad ec. 5.14$$

Donde los cuadrados de las medias de regresión y error serán encontrados por:

$$MS_{regr} = \frac{\sum(\hat{Y} - \bar{Y})^2}{p} \quad ec. 5.15$$

y

$$MS_{error} = \frac{\sum(Y - \hat{Y})^2}{n - p - 1} = s^2 \quad ec. 5.16$$

Si bien el significado de cada variable latente sucesiva podría ser evaluado con pruebas F parciales, el número de variables latentes a incluir será escogido a partir de los resultados de la validación cruzada.

El ajuste para los dos conjuntos se puede evaluar mediante el coeficiente de correlación, r^2 , el cual se define como [17]:

$$r^2 = 1 - \frac{\sum_{i=1}^N (y_c - y_e)^2}{\sum_{i=1}^N y_c^2 - \frac{1}{N} (\sum_{i=1}^N y_c)^2} \quad ec. 5.17$$

donde Y son los valores de la propiedad investigada, como sigue: Y_e son los valores medidos experimentalmente; Y_m es la media de la actividad de los N compuestos.

Debido a la manera como es formulado, el coeficiente de correlación está restringido al intervalo $[-1, 1]$, donde un valor $r = 1$ indica que al graficar los datos todos los puntos yacen exactamente en una recta con pendiente positiva ($a > 0$), $r = -1$ significa todos los puntos yacen exactamente en una recta con pendiente negativa ($a < 0$), mientras que $r = 0$ implica que los dos conjuntos de datos no se correlacionan.

De otro lado, es posible evaluar la calidad del ajuste producido por un modelo de correlación dado por medio de la suma de los cuadrados de predicción (*Predicted Residual Sum of Squares, PRESS*) [18]:

$$PRESS = \sum_{i=1}^N (y_i^{experimental} - y_i^{predicho})^2 \quad ec. 5.18$$

Y también se puede medir ese ajuste mediante el coeficiente de correlación q^2 definido mediante la *PRESS* normalizada:

$$q^2 = 1 - \frac{PRESS}{N\sigma_y^2} \quad ec. 5.19$$

donde (σ_y^2) es la varianza de los datos \mathbf{y} :

$$\sigma_y^2 = \frac{1}{N} \sum_{i=1}^N (y_i^{experimental} - \bar{y})^2 \quad ec. 5.20$$

5.2 ECUACIÓN FUNDAMENTAL DEL QSAR CUÁNTICO

Teniendo en cuenta la naturaleza desconocida del operador del QQSAR (Quantum QSAR), $W(\mathbf{r})$, se le descompone como el producto del operador que lleva al valor de expectación $W_\omega(\mathbf{r})$ que se va a determinar por un operador conocido $\Omega(\mathbf{r}, \mathbf{r}_0)$ [19]. Así, el valor de expectación de la **ec. 4.24** se expresa como:

$$\langle \omega \rangle = \langle W_\omega | \Omega | \rho \rangle \quad ec. 5.21$$

Esto equivale a transformar la **ec. 4.24** en la integral más general:

$$\langle \omega \rangle = \int W_\omega(\mathbf{r}) \Omega(\mathbf{r}, \mathbf{r}_0) \rho(\mathbf{r}_0) d\mathbf{r} d\mathbf{r}_0 \quad (ec. 5.22)$$

Se puede escoger el operador positivo $\Omega(\mathbf{r}, \mathbf{r}_0)$ como la función delta de Dirac: $\delta(\mathbf{r}, \mathbf{r}_0)$ y al hacer esto la **ec. 4.24** reemerge [90]. Si ese es el caso, se le llama una QSM de solapamiento, la cual da una medida del volumen encerrado en la superposición de ambas funciones de densidad electrónica [91]. Otra posibilidad es la utilización del operador de Coulomb, $|\mathbf{r}_1 - \mathbf{r}_2|^{-1}$, definiendo

una QSM de Coulomb la cual expresa la energía de Coulomb repulsiva entre las dos densidades de carga [20]. Así se obtienen la **ec. 4.26** y **ec. 4.27**.

La **ec. 4.24** para una molécula A puede escribirse como:

$$\langle \omega_A \rangle = \int W(\mathbf{r}) \rho_A(\mathbf{r}) d\mathbf{r} \quad (\text{ec. 5.23})$$

Cuandoquiera que el operador y la función de densidad en la **ec. 4.24** pertenezcan ambos al mismo semiespacio vectorial y expresen aproximadamente el operador desconocido, $W_\omega(\mathbf{r})$, como una combinación lineal del conjunto etiquetado de moléculas que actúa como base:

$$W_\omega(\mathbf{r}) \approx \sum_l W_l \rho(\mathbf{r}) \quad (\text{ec. 5.24})$$

Se obtiene una nueva relación al sustituir la **ec. 5.24** en la **ec. 5.23** mientras se toma en cuenta la descomposición de la **ec. 5.21**:

$$\begin{aligned} \langle \omega_A \rangle &= \sum_l W_l \iint \rho_l(\mathbf{r}) \Omega(\mathbf{r}, \mathbf{r}_0) \rho_A(\mathbf{r}_0) d\mathbf{r} d\mathbf{r}_0 \\ &= \sum_l W_l \mathbf{z}_{lA}(\Omega) \quad (\text{ec. 5.25}) \end{aligned}$$

Si la **ec. 43** rige para todas las moléculas y la SM, \mathbf{Z} , obtenida al utilizar el operador $\Omega(\mathbf{r}, \mathbf{r}_0)$ es simétrica, se puede escribir una expresión equivalente en matricial [21]:

$$\mathbf{Z}\mathbf{w} = \pi \quad (\text{ec. 5.26})$$

donde π es un vector columna que contiene los valores de expectación o la propiedad molecular de interés y \mathbf{w} es otro vector columna que contiene el conjunto de coeficientes desconocido $\{w_i\}$ del operador en la **ec. 4.25**.

5.4. REFERENCIAS

1. LP, Hammett. (1935). Some relations between reaction rates and equilibrium constants. *Chem Rev.* **17**: 125–136.
2. Hansch, C., Muir, RM., Fujita, T., Maloney, PP., Geiger, E. & Streich, M. The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. *J Am Chem Soc.* **85**: 2817–2824, 1963.
3. Fujita, T., Iwasa, J. & Hansch C. (1964). A new substituent constant, π , derived from partition coefficients. *J Am Chem Soc.* **86**: 5175–5180.
4. Leszczynski, J. & Shukla, MK. Practical Aspects of Computational Chemistry. Methods, Concepts and Applications. Springer. 2009. Pag 195.
5. Famini, GR. & Wilson, LY. (2002). Linear Free Energy Relationships Using Quantum Mechanical Descriptors. *Rev. Comp. Chem.* **18**: 211–255.
6. Patrick, GL. An introduction to Medicinal Chemistry. 4^a Ed. Oxford University Press. 2009. Pag 691.
7. Jensen, F. Introduction to Computational Chemistry. 2a Ed. Wiley. 2007. Pag 547.
8. Höskuldsson, A. Prediction Methods in Science and Technology. Copenhagen: Thor, 1996.
9. Wold, S., Sjostrom, M. & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems.* **58**: 109–130.
10. Carbó-Dorca, R. & Besalú, E. (2000). Quantum theory of QSAR. *Contributions to science.* **1** (4): 399-422.

11. Fradera, X., Amat, L., Besalú, E. & Carbó-Dorca, R. (1997). Application of molecular quantum similarity to QSAR. *Quant Struct-Act Relatsh.* **16**: 25–32
12. Ponec, R., Amat, L. & Carbó-Dorca, R. (1999). Molecular basis of quantitative structure–properties relationships (QSPR): a quantum similarity approach. *J Comput-Aided Mol Des.* **13**: 259–270.
13. Gironés, X., Amat, L., Robert, D. & Carbó-Dorca, R. (2000). Use of electron–electron repulsion energy as a molecular descriptor in QSAR and QSPR studies. *J Comput-Aided Mol Des.* **14**: 477–485.
14. Carbó-Dorca, R. & Mezey, PG. *Advances in molecular similarity.* (1998). Volumen 2. Jai Press Inc. Pag 44.

CAPÍTULO VI

QQSAR DE INDENOPIRAZOLES

ANTAGONISTAS DE LA CDK2

Este capítulo es concerniente a un trabajo de diseño racional de ligandos recientemente publicado por el autor de esta tesis [1] que comprende los pasos mostrados en la **Figura 6.1**:

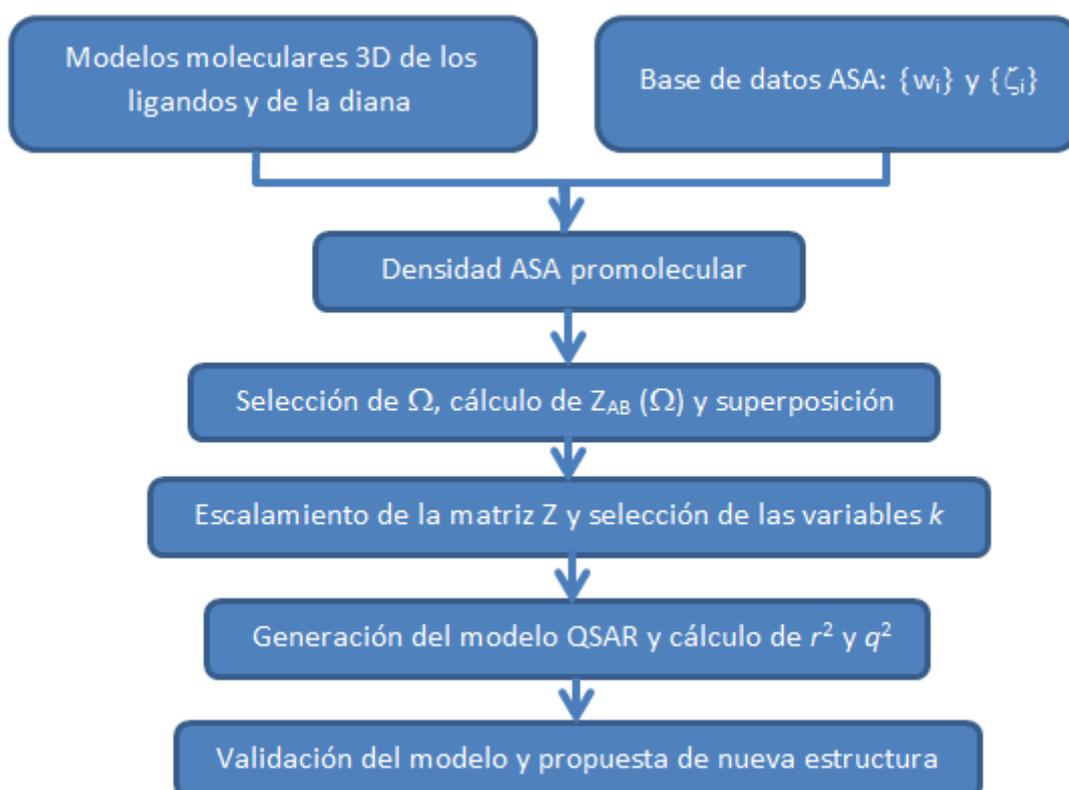


Figura 1. Procedimiento para la generación de un modelo QSAR.

6.1 ACOPLAMIENTO MOLECULAR Y OPTIMIZACIÓN

El modelo de la diana utilizada en este estudio fue la estructura cristalográfica (Código en el PDB: 1GZ8), resuelta a 1.3 Å, con un valor R de 0.153 y publicada por Gibson *et al.* la cual corresponde a la CDK2 en complejo con el inhibidor 1-[(2-amino-6,9-dihidro-1H-purin-6-il)oxi]-3-metil-2-butanol [2] (**Fig. 6.2**). De los 447 modelos de la CDK2 depositados hasta la fecha en el Protein Databank este fue seleccionado debido a que las estructuras de alta resolución son cruciales, en la medida que se suministra el nivel de definición química de la estructura y de las interacciones no covalentes (puentes de hidrogeno, contactos, conformaciones de ángulo torsional) las cuales son fundamentales para perseguir diseño molecular, reconocimiento y procesos catalíticos. Basados en esta estructura, se adoptará un sistema compuesto por 5 residuos de aminoácidos (región bisagra) y el inhibidor.



Figura 6.2. Estructura del ligando co-cristalizado (MBP) de la entrada 1GZ8 del Protein Databank.

Los residuos de aminoácidos incluidos fueron F⁸⁰, E⁸¹, F⁸², L⁸³ e H⁸⁴. Con la única excepción de la tríada catalítica K³³, E⁵¹ y D¹⁴⁵. Se reconstruyeron los 7 residuos faltantes en el modelo 1GZ8 (los residuos 37-43).

Se adicionaron los átomos de hidrógeno faltantes al sistema modelo tomado de la estructura cristalográfica. Las posiciones de estos hidrógenos se optimizarán con el método semi-empírico PM6. La estructura convergida totalmente se utilizó como modelo inicial para realizar los cálculos restantes. Los ligandos indenopirazol fueron acoplados con esta estructura convergida de la 1GZ8 utilizando el programa FlexiDock disponible en Sybyl-X [94]. Los IC₅₀ de estos ligandos fueron tomados de la literatura luego de constata que fueron ensayados en iguales condiciones [3-7].

El programa **Flexible Docking** basado en Algoritmos Genéticos trabaja en el espacio torsional, manteniendo las longitudes de enlace y los ángulos constantes [8]. Por tanto las interacciones de Van der Waals grandes sólo pueden relajarse mediante rotación(es) de enlace, la optimización no altera los centros quirales ni la estereoquímica. En FlexiDock los átomos del esqueleto de la proteína se mantienen fijos en el espacio, pero el ligando es móvil (se pueden aplicar rotación/traslación). Para acelerar los cálculos, FlexiDock considera como rotables únicamente los enlaces sencillos no anulares y los enlaces amida. FlexiDock utiliza Algoritmos Genéticos para determinar la geometría óptima del ligando. Los Algoritmos Genéticos son optimizadores globales relativamente robustos, con requerimientos de desempeño que se escalan bien con el aumento en el tamaño del sistema.

Se realizó luego el cálculo de optimización de los ligandos indenopirazoles con el fin de utilizarlos para el estudio de semejanza cuántica empleando el método ab iniito DFT/B3LYP para describir los átomos [9]. También se especificó la carga y la multiplicidad de estos. Esta decisión automáticamente definió el Hamiltoniano electrónico, donde se utilizará para ello el conjunto junto con la base de cálculo 6-31G seguido por una minimización total (microiteración) de los átomos utilizando un optimizador. Para este fin se utilizó la opción Opt=QuadMacro del paquete GAUSSIAN09 [10].

6.2 DESARROLLO DE SEMEJANZA CUÁNTICA Y ANÁLISIS POR PLS

Para la ejecución del cálculo de semejanza cuántica y la obtención de la matriz de semejanza se utilizó el programa SIMILARITY que forma parte de una estrategia computacional para cálculos de QSAR en 3 dimensiones (3D-QSAR) y que actualmente se encuentra en desarrollo en el grupo de Química Cuántica y Teórica de la Universidad de Cartagena. Este programa está compuesto por una rutina de iniciación y otra de alineamiento [11].

La rutina de inicialización lee las coordenadas de las moléculas que serán comparadas. En principio, las coordenadas pueden ser obtenidas a partir de datos de difracción de Rayos-X o del archivo de salida de un algoritmo de optimización de estructura ya sea por métodos *ab-initio*, semiempíricos o de campos de fuerza clásicos. Además de las coordenadas atómicas, el programa requiere que se especifiquen las cargas atómicas. En general, la evaluación del producto interno involucrado en la semejanza molecular resulta computacionalmente prohibitivo en un estudio QSAR si se efectúa mediante cálculos de estructura electrónica. Para simplificar la evaluación de la densidad electrónica se utilizó la aproximación ASA.

Tanto la semejanza molecular entre moléculas diferentes como los índices de semejanza correspondientes, dependen crucialmente de la orientación relativa de estas moléculas en el espacio. El problema de encontrar la orientación relativa óptima entre dos moléculas en el espacio, es conocido en la literatura como el problema del alineamiento. Este problema fue abordado mediante el algoritmo de alineamiento que trae incorporado el paquete SYBYL para los análisis CoMFA. El análisis estadístico por PLS se llevó a cabo en el programa STATGRAPHICS Centurion XVI versión 16.02.4, el cual es bastante completo y proporciona información soportada de la mayoría de los datos que se necesitan en este estudio como los gráficos de factores de componentes, gráficos de modelos y de predicción.

6.3 RESULTADOS

6.3.1. Interacciones Cinasa-Indenopirazol

Una vez analizados los resultados del acoplamiento molecular se observó que las principales interacciones entre la proteína CDK2 y todos los ligandos se presentan en el sitio catalítico por medio de puentes de hidrógenos entre los aminoácidos de naturaleza polar (Glu 81 e His 84) y los oxígenos presentes en este grupo. Además, se presentan uniones entre el ligando y el sitio activo de la proteína por medio de los oxígenos del grupo carboxilo presentes en la cadena lateral del residuo Lys 33 y los nitrógenos presentes en los ligandos indenopirazoles. Por otro lado, las interacciones estabilizantes de naturaleza no polar se dan entre los residuos Phe 80, Phe 82, Leu 83 y el anillo aromático adyacente en la base indenopirazolica de los ligandos. Todas las interacciones detalladas anteriormente son esenciales para la unión entre la enzima y los ligandos y se presentan en todas las moléculas estudiadas.

Haciendo un enfoque más detallado en los diferentes núcleos básicos estructurales de los derivados de indenopirazoles, puede decirse que los de mayor puntuación total de acoplamiento y mayor afinidad por la diana CDK2 fueron los de la primera base estructural (1-93), esto se debe a que la mayoría de los ligandos de este grupo poseen interacciones fuertes de enlaces de hidrogeno que se forman entre los grupos NH y tioles con los residuos aminoacídicos del sitio activo de la CDK2, mientras que con una puntuación relativamente media y significativa encontramos los ligandos de la segunda base estructural (94-110), en estos últimos se evidenciaron pocas interacciones de hidrógeno e interacciones de van der Waals con menor valor de afinidad que los de la primera base, debido a que existen más interacciones lipofílicas entre los anillos tanto piperidinius como pirrolidinius y los residuos de fenilalanina de la región bisagra. Por último encontramos la tercera base estructural (111-117) como la de más baja puntuación promedio del total de acoplamiento, estos últimos ligandos derivados de indenopirazoles poseen radicales bastante simples y con densidades electrónicas bajas, y por ende se crean pocas interacciones con los residuos de los aminoácidos del sitio activo

como las de Van der Waals y de torsión, que no contribuyen tanto a la puntuación total como las hidrofílicas y lipofílicas.

Los resultados del acoplamiento llamados “poses” comprenden tanto la posición respecto a las coordenadas cartesianas, la orientación y la conformación de la estructura. La **Figura 6.3** muestra las pose del ligando que presentó la mayor puntuación de acoplamiento, el 062, con los residuos de aminoácidos del sitio activo:

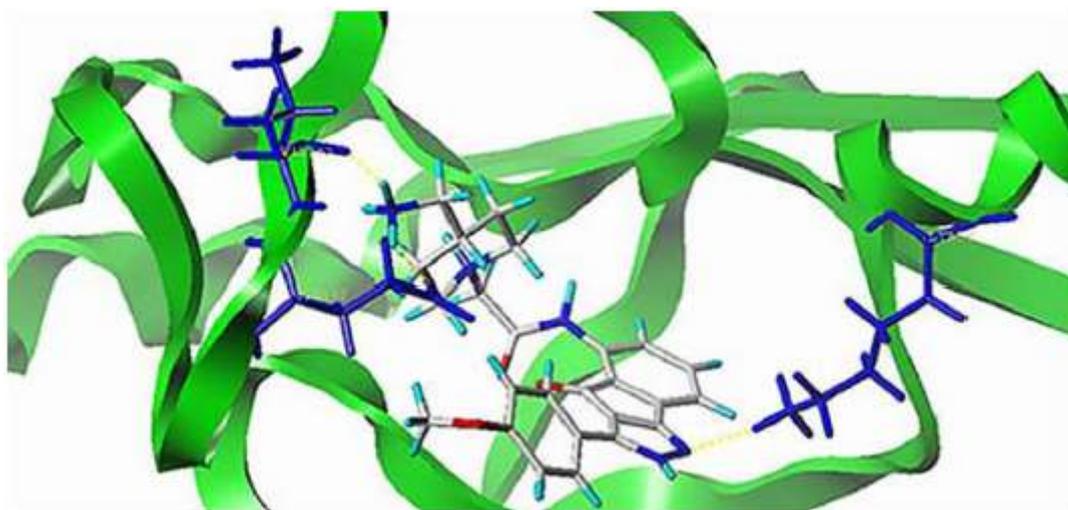


Figura 6.3. Pose ligando indenopirazol 062 con el sitio activo de la proteína.

Se realizó una regresión de la actividad biológica y el Total Score, con el fin de encontrar evaluar si la afinidad de los ligandos derivados de indenopirazoles se correlaciona con la inhibición por la diana CDK2.

De acuerdo a los resultados obtenidos de la regresión realizada se puede decir que existe una buena correlación entre la actividad biológica medida en concentración efectiva (IC_{50}) y la puntuación total del motor de acoplamiento (Total Score) de los ligandos derivados de indenopirazoles, con un coeficiente de determinación obtenido (R^2) de 0,89, y como se aprecia en la **Figura 6.4** estos valores aumentan linealmente de manera proporcional. El anterior

análisis fue hecho con dos variables mientras que el análisis por PLS se desarrolló para construir modelos que relaciona múltiples variables independientes X, en este caso los descriptores de semejanza cuántica molecular, con múltiples variables dependientes Y (actividad biológica de los ligandos indenopirazoles). El objetivo principal es predecir las variables de respuesta para explicar mejor el universo de datos obtenidos que una regresión lineal, con el menor número de descriptores posibles.

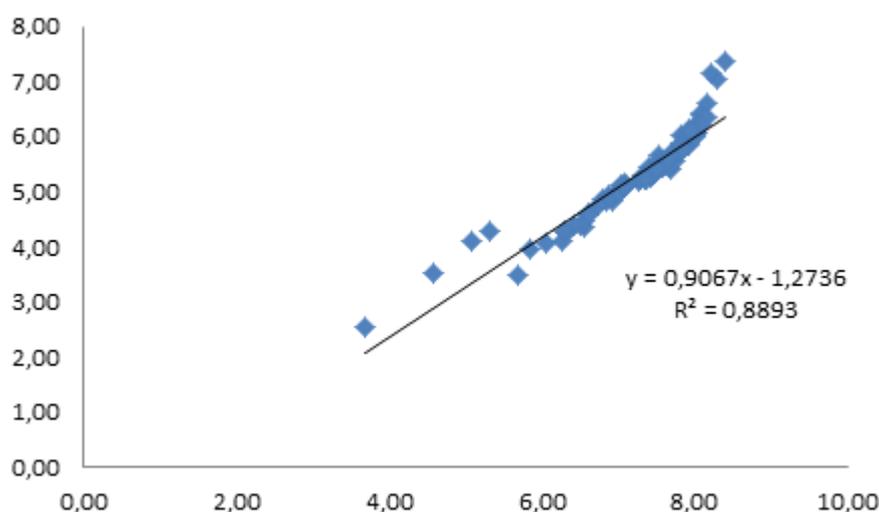


Figura 6.4. Correlación entre el pIC₅₀ vs Total Score de los ligandos indenopirazoles.

6.3.2. Evaluación del QQSAR

Todos los ligandos de la tabla 6.1 fueron optimizados como se describió anteriormente para ser utilizados en el cálculo de los índices de semejanza cuántica, el cual arroja a su vez una matriz del índice de semejanza. En dicha matriz cada columna es un descriptor y cada descriptor representa la semejanza entre dos ligandos. Los valores de semejanza para los 117 compuestos se obtuvieron utilizando el operador de superposición, en combinación con el índice de Carbó.

Tabla 6.1. Análisis de varianza para el pIC₅₀ experimental					
<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	8,787	10	0,878	$3,69 \times 10^{10}$	0,0
Residuo	$4,76 \times 10^{-11}$	2	$2,38 \times 10^{-11}$		
Total (corr.)	8,787	12			

En los resultados del análisis de regresión PLS se asume que las variables independientes están representadas por las columnas de la matriz de Semejanza de los ligandos indenopirazoles, y las variables dependientes están representadas por la actividad biológica (pIC₅₀ real) mostrada en la Tabla 6.1.

En las Tablas 6.3 y 6.4 se indican los valores de los coeficientes de la regresión lineal múltiple obtenidos con la técnica PLS. El análisis estadístico fue realizado con los 117 ligandos derivados de indenopirazoles, donde se usaron 10 descriptores, utilizando las columnas como componentes mediante validación cruzada conocida como “sacando 2 a la vez”, (k-ésima con k=2), el cual consiste en ajustar n veces, cada vez dejando fuera 2 de las observaciones y volviendo a ajustar el modelo usando las otras $n - 1$ observaciones. Luego la observación omitida se predice con el modelo del que fue excluida. Esta validación es utilizada para acortar el proceso en grandes conjuntos de datos, como es el caso de este estudio.

En la tabla 6.2 encontramos valores relativamente altos para R² con 10 componentes alcanzando un valor de 100,0% en el séptimo componente, por lo que queda notoriamente demostrada la correlación entre las variables propuestas con pocos descriptores utilizados. A partir del grupo de prueba de validación cruzada se calculó el PRESS, (*Prediction Error Sum of Squares* – Suma de Cuadrados del Error de Predicción). Esta estadística es comparable al cuadrado medio de los residuos de la tabla 6.3, excepto que la primera se calcula a partir de las predicciones para las observaciones cuando estas no se

emplean para ajustar el modelo. A menor valor del cuadrado medio PRESS menor error de predicción y por ende mejor es el modelo con el número de componentes a extraer. Podemos observar que en el séptimo componente PRESS alcanza su valor mínimo (0,63) por lo que se sugiere que un modelo con 7 componentes sería el de menor error de predicción.

Tabla 6.2. Modelo para el pIC₅₀ Experimental				
	<i>% Variación</i>		<i>Cuadrado Medio</i>	<i>Predicción</i>
<i>Componente</i>	<i>en Y</i>	<i>R-Cuadrada</i>	<i>PRESS</i>	<i>R-Cuadrada</i>
1	88,33	88,3339	0,914139	0,847782
2	9,475	97,8093	0,851819	0,858159
3	1,98	99,7849	0,994633	0,834378
4	$1,8 \times 10^{-1}$	99,962	0,893983	0,851138
5	$3,3 \times 10^{-2}$	99,995	0,911568	0,84821
6	$4,9 \times 10^{-3}$	99,9998	0,914641	0,847698
7	$1,67 \times 10^{-4}$	100,0	0,638122	0,893743
8	$3,8 \times 10^{-6}$	100,0	0,913973	0,847809
9	$5,89 \times 10^{-7}$	100,0	0,914067	0,847794
10	$1,26 \times 10^{-8}$	100,0	0,914062	0,847795

La última columna muestra el q^2 o *R Cuadrada de Predicción* promedio a través de todas las variables dependientes (actividad biológica). El promedio alcanza un pico en 7 componentes, con un valor de $R^2 = 89,37\%$, sugiriendo que un modelo con siete componentes sería una excelente elección, aunque el valor para 10 componentes explica perfectamente la correlación y predicción del modelo ($q^2 = 0.848$). Este resultado es notable si se compara con el $q^2 = 0.765$

reportado por Singh *et al.* [12] llevado a cabo con el mismo conjunto de ligandos indenopirazólicos.

Tabla 6.3. Variables independientes y dependientes del análisis estadístico por PLS.					
	<i>% Variación</i>	<i>% Acumulado</i>	<i>% Variación</i>	<i>% Acumulado</i>	<i>Predicción Promedio</i>
<i>Componente</i>	<i>en X</i>	<i>de X</i>	<i>en Y</i>	<i>de Y</i>	<i>R-Cuadrada</i>
1	13,0381	13,0381	88,3339	88,3339	0,847782
2	13,1011	26,1392	9,4754	97,8093	0,858159
3	8,53256	34,6718	1,97559	99,7849	0,834378
4	9,2057	43,8775	0,177094	99,962	0,851138
5	7,84852	51,726	0,0330028	99,995	0,84821
6	6,49628	58,2223	0,00485691	99,9998	0,847698
7	7,93268	66,155	0,000166736	100,0	0,893743
8	6,86759	73,0226	0,00000383992	100,0	0,847809
9	5,94447	78,967	5,89338E-7	100,0	0,847794
10	8,58347	87,5505	1,26086E-8	100,0	0,847795

La tabla 6.3 presenta datos similares para los porcentajes de la variación total en las variables X y Y explicada conforme aumenta el número de componentes.

Para explicar el número de componentes escogido o extraído se graficó el porcentaje de variación de las variables independientes, dependientes y PRESS vs el número de componentes indicado, como se ilustra en la **Figura 6.5.**

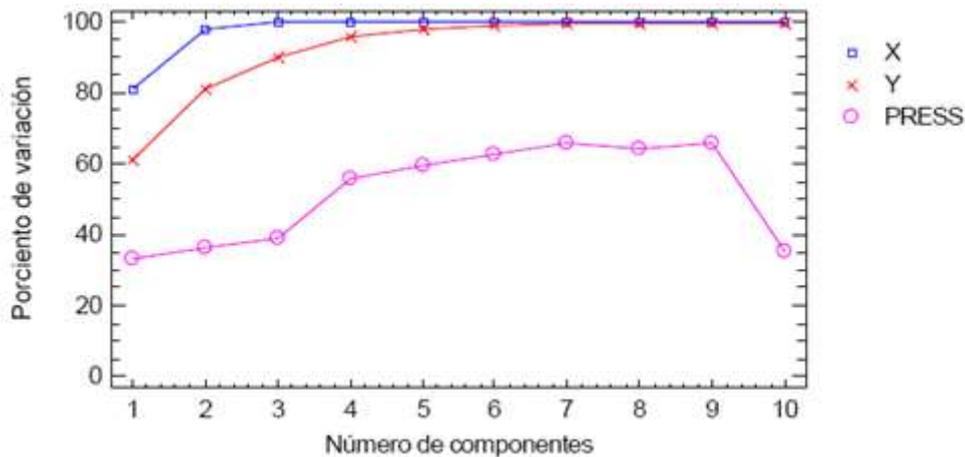


Figura 6.5. Gráfico de comparación de modelos para elegir el mejor componente extraído.

Este gráfico es útil para visualizar cuántos componentes tienen que extraerse. Cabe resaltar que el porcentaje de variación para PRESS aumenta a lo largo de 7 componentes.

También se realizó un gráfico de componentes en 2D (**Figura 6.6**), en el cual se seleccionan dos factores, par por cada eje, y se grafican los puntos que representan las filas en las correspondientes columnas. En situaciones donde los factores son interpretables, este gráfico muestra el valor de cada una de las muestras para estos factores.

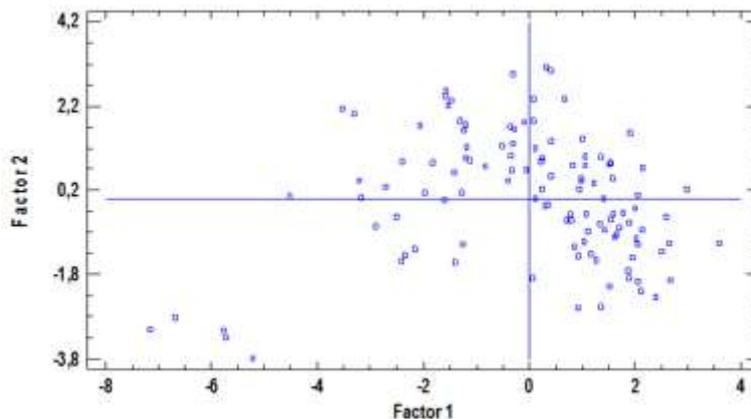


Figura 6.6. Gráfico de componentes para valores de factor de PLS en 2D.

Análogamente se hizo *Gráfico de Componentes 3D*, el cual es análogo al gráfico 2D excepto que se eligen tres componentes (**Figura 6.7**).

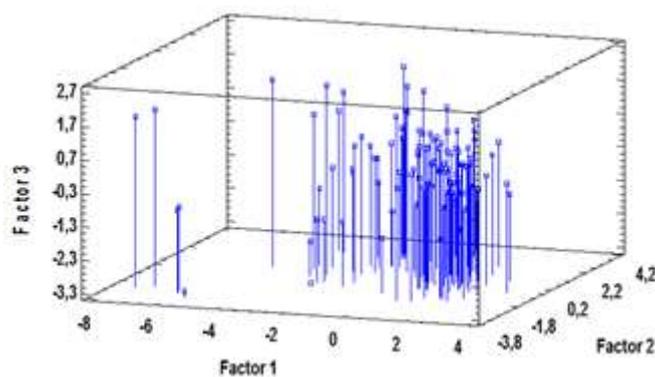


Figura 6.7. Gráfico de componentes para valores de factor de PLS en 3D.

Para analizar los resultados de la Tabla 6.4, cabe recordar que cada componente escogido en el análisis por PLS es una combinación lineal de las columnas de la matriz de semejanza, y esta tiene la particularidad de generar nuevos descriptores no correlacionados entre sí y aumentar a su vez la correlación con las actividades biológicas. La diferencia entre el análisis tradicional de Componentes Principales, y el análisis de PLS usados en este estudio, radica en implantar la correlación entre las variables independientes a la hora de generar los componentes.

Finalmente, en la **Figura 6.8** se muestran los valores de la actividad biológica predicha vs la actividad biológica medida experimentalmente, donde la actividad biológica predicha se calculó con los valores de los coeficientes obtenidos por PLS para los 10 componentes. Como es el caso de este estudio, si existe un buen modelo de ajuste, los puntos se alinean a lo largo de la línea diagonal, por lo que a pesar de que existe un mayor R^2 de predicción para 7 componentes ($R^2= 89,37$), la predicción del modelo con 10 componentes es bastante precisa ($R^2= 84,78$).

6.4 CONCLUSIONES

Del presente estudio se puede concluir que existe una alta afinidad entre los ligandos indenopirazoles utilizados y la enzima diana CDK2. Esto se pudo comprobar mediante el análisis estructural y electrónico en el sitio de interacción (sitio activo) y el análisis estadístico de las correlaciones entre la actividad biológica de los ligandos con los descriptores de acoplamiento y semejanza cuántica molecular.

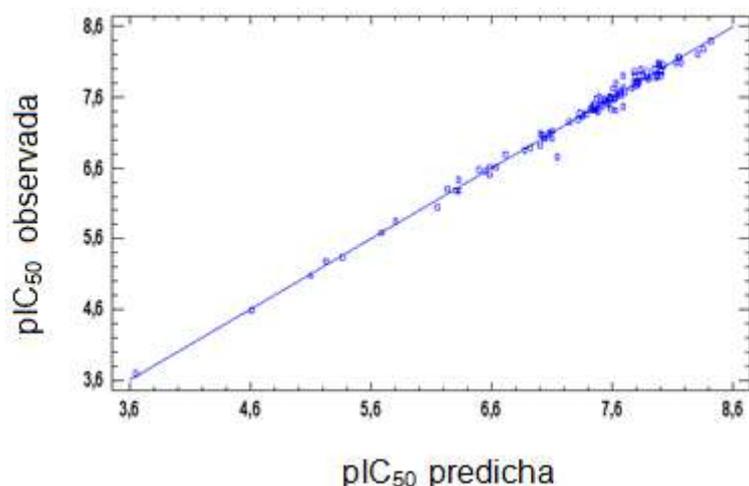


Figura 6.8. Gráfico de actividad biológica observada vs la predicha.

Para el análisis estructural se encontró que las que las principales interacciones entre la proteína CDK2 y todos los ligandos se presentan en el sitio catalítico por medio de puentes de hidrógenos entre los aminoácidos de naturaleza polar (Glu 81 e His 84) y los oxígenos presentes en este grupo. Además, se presentan uniones entre el ligando y el sitio activo de la proteína por medio de los oxígenos del grupo carboxilo presentes en la cadena lateral del residuo Lys 33 y los nitrógenos presentes en los ligandos indenopirazoles. Por otro lado, las interacciones estabilizantes de naturaleza apolar se dan entre los residuos Phe 80, Phe 82, Leu 83 y el anillo aromático adyacente en la base indenopirazolica de los ligandos. Todas las interacciones detalladas anteriormente son esenciales para la unión entre la enzima diana y los ligandos derivados de indenopirazoles y se presentan en todas las moléculas estudiadas.

En un enfoque más detallado en los diferentes núcleos básicos estructurales de los derivados de indenopirazoles, puede decirse que los de mayor puntuación total de acoplamiento y mayor afinidad por la diana CDK2 fueron los de la primera base estructural (1-93), esto se debe a que la mayoría de los ligandos de este grupo poseen fuertes interacciones de enlaces de hidrogeno que se forman entre los grupos NH y tioles con los residuos aminoacídicos del sitio activo de la CDK2, mientras que con una puntuación relativamente media y significativa encontramos los ligandos de la segunda base estructural (94-110),

en estos últimos se evidenciaron pocas interacciones de hidrógeno e interacciones de van der Waals con menor valor de afinidad que los de la primera base, debido a que existen más interacciones lipofílicas entre los anillos tanto piperidínicos como pirrolidínicos y los residuos de fenilalanina de la región bisagra.

Por último encontramos la tercera base estructural (111-117) como la de más baja puntuación promedio del total de acoplamiento, estos últimos ligandos derivados de indenopirazoles poseen radicales bastante simples y con densidades electrónicas bajas, y por ende se crean pocas interacciones con los residuos de los aminoácidos del sitio activo como las de Van der Waals y de torsión, que no contribuyen tanto a la puntuación total de acoplamiento como las de naturaleza polar (puentes de hidrógeno), y apolar como las presentadas por los anillos con naturaleza bencénica, lo que permite a su vez una mayor estabilización del complejo CDK2-indenopirazol.

Luego en el cálculo de semejanza cuántica se obtuvo la matriz del índice de semejanza, la cual indica que las bases estructurales indenopirazólicas son fundamentales para establecer una similitud entre dos ligandos, y estos descriptores fueron utilizados para realizar el análisis estadístico donde se buscó predecir la relación entre los parecidos estructurales de cada ligando y su actividad biológica en la diana, mediante la técnica de PLS.

Para el análisis estadístico se evidenció una correlación significativa entre la actividad inhibitoria de los ligandos indenopirazoles y la puntuación total de acoplamiento del complejo, con un coeficiente de determinación de $R^2= 0,89$, así como una muy buena correlación de predicción entre semejanza cuántica molecular y la actividad biológica de los ligandos, con un coeficiente cuadrado de predicción de $q^2= 0.894$ con el menor número de componentes posibles, lo que supone que se ha desarrollado un modelo exitoso QSAR que permite explicar con claridad la interacción de los inhibidores estudiados con los aminoácidos de la región bisagra en el sitio activo de la CDK2, siendo este más robusto que los modelos recurrentes de división simple.

6.5 REFERENCIAS CAPÍTULO VI

1. Torres, J.E., Toro, J.P., Vergara, J., Baldiris, R., Vivas, R. (2017). Computational study of the interaction between indene pyrazole and cyclin dependent kinase 2. *Quim. Nova.* **40** (4): 413-417.
- 2 Gibson, A. E., Arris, C. E., Bentley, J., Boyle, F. T., Curtin, N. J., Davies, T. G., Endicott, J. A., Golding, B. T., Grant, S., Griffin, R. J., Jewsbury, P., Johnson, L. N., Mesguiche, V., Newell, D. R., Noble, M. E., Tucker, J. A. & Whitfield, H. J. (2002). Probing the ATP Ribose-Binding Domain of Cyclin-Dependent Kinases 1 and 2 with O6-Substituted Guanine Derivatives. *J. Med. Chem.* **45** (16): 3381-3393
3. Nugiel, DA.; Etkorn, AM.; Vidwans, A.; Benfield, PA.; Boisclair, M.; Burton, CR.; Cox, S.; Czerniak, PM.; Doleniak, D.; Seitz, SP. (2001). Indenopyrazoles as Novel Cyclin Dependent Kinase (CDK) Inhibitors. *J. Med. Chem.* **44** (9): 1334–1336.
4. Nugiel, DA., Vidwans, A., Etkorn, A-M., Rossi, KA., Benfield, PA., Burton, CR., Cox, S., Doleniak, D. & Seitz, SP. (2002). Synthesis and Evaluation of Indenopyrazoles as Cyclin-Dependent Kinase Inhibitors. 2. Probing the Indeno Ring Substituent Pattern.. *J. Med. Chem.* **45** (24): 5224–5232.
5. Yue, EW., Higley, CA., DiMeo, SV., Carini, DJ., Nugiel, DA., Benware, C., Benfield, PA., Burton, CR., Cox, S., Grafstrom, RH., Sharp, DM., Sisk, LM., Boylan, JF., Muckelbauer, JK., Smallwood, AM., Chen, H., Chang, Ch-H., Seitz, SP. & Trainor, GL. (2002). Synthesis and Evaluation of Indenopyrazoles as Cyclin-Dependent Kinase Inhibitors. 3. Structure Activity Relationships at C3(1,2). *J. Med. Chem.* **45** (24): 5233–5248.
6. Yue, EW., DiMeo, SV., Higley, CA., Markwalder, JA., Burton, CR., Benfield, PA., Grafstrom, RH., Cox, S., Muckelbauer, JK., Smallwood, AM., Chen, H., Chang, CH-H., Trainor, GL., Seitz, SP. (2004). Synthesis and evaluation of

indenopyrazoles as cyclin-dependent kinase inhibitors. Part 4: Heterocycles at C3. *Bioorg. Med. Chem. Lett.* **14** (2): 343–346.

7. Nugiel, DA.; Vidwans, A.; Dzierba, CD. (2004). Parallel synthesis of acylsemicarbazide libraries: preparation of potent cyclin dependent kinase (cdk) inhibitors. *Bioorg. Med. Chem. Lett.* **14** (22) 5489–5491.

8. Tripos International; SYBYL X; Molecular Modeling from Sequence Through Lead Optimization; South Hanley Rd., St. Louis, Missouri, USA, 1997.

9. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R.; Montgomery, Jr., J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X.; Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W.; Ayala, P. Y., Morokuma, K.; Voth, G. A.; Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G.; Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T.; Al-Laham, M. A., Peng, C. Y.; Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C. & Pople, J. A. Gaussian 09; Gaussian Inc, Wallingford CT, 2004.

10. Rincón, L. (2007). SIMILARITY: Un programa para el calculo de la similitud molecular cuántica. *CIENCIA.* **15** (3): 406 - 420.

11. Singh SK, Dessalew N, Bharatam PV. (2006). 3D-QSAR CoMFA study on indenopyrazole derivatives as cyclin dependent kinase 4 (CDK4) and cyclin dependent kinase 2 (CDK2) inhibitors. *Eur. J. Med. Chem.* **41**: 1310–1319.

CAPÍTULO VII

ESCALA DE CONVERGENCIA DE LAS CATECOLAMINAS

En este estudio, se formuló una definición cuantitativa propuesta para las catecolaminas con respecto a las monoaminas biogénicas asociadas [1]. Esta cuantificación permite definir las escalas alfa y beta, las cuales expresan matemáticamente el grupo de correlación de las catecolaminas con respecto a su fragmento común (la molécula de la feniletilamina) y con respecto a su precursor, la tirosina. Para este fin se efectuó un análisis de semejanza cuántica molecular empleando cuatro descriptores: índices de semejanza cuántica molecular de solapamiento, Coulomb, interacción de solapamiento, interacción de Coulomb, distancias euclídeas de solapamiento y distancias Euclídeas de Coulomb. El método analítico fue complementado mediante los índices alfa y beta. Se utilizó el enfoque de la superposición topo-geometrical como método de alineamiento. También se respaldaron las interpretaciones con resultados experimentales de la ruta de la reacción de síntesis evaluando el estado de transición de la ruta de la reacción [2,3] de acuerdo con los criterios introducidos por Cioslowski [4]. A partir de la escala calculada con los índices alfa y beta se propuso una cuantificación de la semejanza del fragmento común.

7.1 FISIOLÓGÍA DE LAS CATECOLAMINAS

Las catecolaminas son un grupo de compuestos que incluyen a la adrenalina, noradrenalina y dopamina, cuya biosíntesis parte de aminoácido tirosina [5,6]. Estas moléculas presentan un grupo catecol y un grupo amino. Algunas catecolaminas son secretadas en las glándulas suprarrenales, ejerciendo una función hormonal y otras como neurotransmisores en las neuronas. Las catecolaminas son importantes en la respuesta al estrés y están asociadas la obesidad [7]. Las catecolaminas presentan una estructura distintiva de un anillo bencénico con dos grupos hidroxilo, una cadena de acilo y un grupo amino terminal. Las catecolaminas y sustancias relacionadas estudiadas en el presente trabajo son mostradas en la Tabla 1. Las moléculas 1–4 son las catecolaminas en tanto que 5 y 6 son las monoaminas biogénicas relacionadas tirosina y feniletilamina. La tirosina es la catecolamina precursora y feniletilamina es la molécula que es el fragmento molecular común.

Las catecolaminas también son consideradas aminas biogénicas, las cuales constituyen el grupo principal de neurotransmisores en el sistema nervioso. El rasgo distintivo de estas sustancias es la presencia de un grupo amino ($-NH_2$), la reactividad de esta función química es un campo importante de interés científico [8–11]. La Monoamino oxidasa es una enzima que cataliza la desaminación oxidativa de las monoaminas. La Monoamino oxidasa opera con el FAD (dinucleotido de flavina y adenina) como cofactor. Ambas MAOs son también cruciales en la inactivación de los neurotransmisores monoaminérgicos. Existen dos formas de la enzima, la MAO-A y la MAO-B las cuales presentan distinta afinidad por los sustratos y distinta sensibilidad frente a los inhibidores. La serotonina, noradrenalina y adrenalina son degradadas en su mayoría por la MAO-A. La fenetilamina es degradada por la MAO-B. Ambas degradan la dopamina.

7.2 METODOLOGÍA

7.2.1 Alineamiento de las estructuras de las aminas. El índice de semejanza de Carbó y el de Hodgkin-Richards dependen de la orientación relativa de las moléculas, por esta razón, se pueden alcanzar diferentes conclusiones sobre el grado de semejanza dependiendo del esquema de alineamiento. Por ende, es crucial escoger cuidadosamente un método de alineamiento basado en el objetivo del estudio. Considerando esta dependencia se han propuesto muchos métodos de alineamiento, tales como el incluido en los algoritmos de CoMFA y CoMSIA, los cuales son tridimensionales y permiten obtener mapas estéricos, electrostáticos e hidrológicos.[12] Para encontrar los patrones de alineamiento que produzcan los mejores resultados, en este estudio se utilizó el algoritmo de superposición topo-geométrico (TGSA) introducido por Gironés *et al* [13]. Se puede emplear el TGSA para sistemas con alguna (sub)estructura común y se pueden alinear grupos funcionales similares entre las moléculas que son comparadas en una manera que es consistente con su realidad bioquímica.



Figura 7.1. Estructura de los compuestos analizados.

Tabla 7.1. Estructura de los compuestos analizados.					
Compuesto	R ¹	R ²	R ³	R ⁴	R ⁵
1	H	H	OH	OH	OH
2	H	H	H	OH	OH
3	CH ₃	H	OH	OH	OH
4	H	H	H	H	OH
5	H	COOH	H	H	OH
6	H	H	H	H	H

Esta base de datos molecular tiene la particularidad de que el fragmento estructural común a todas las estructuras es una molécula que es parte del conjunto de moléculas en la base de datos y no una porción de una de ellas. El fragmento común sobre el cual se hizo la superimposición se muestra en la Figura 2.

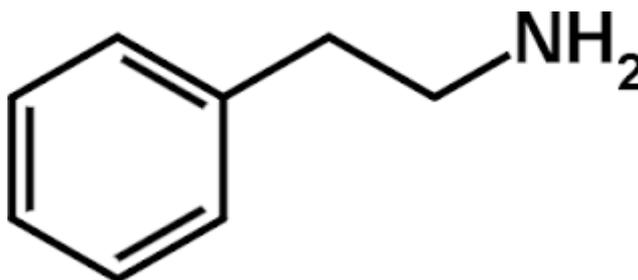


Figura 7.2. Estructura química de la feniletilamina empleada para el alineamiento.

7.2.2 Diseño de moléculas, optimización de la geometría y cálculo de los descriptores moleculares. El algoritmo TGSA de Gironés *et al.*[13] comienza con la identificación de la mayor subestructura común entre las moléculas de la base de datos. Los algoritmos para hacer esta búsqueda constituyen un importante campo de interés científico.[14] Todas las estructuras fueron optimizadas con el programa Gaussian 98.[15] El cálculo de los descriptores moleculares, se realizó mediante el enfoque de la teoría del funcional de densidad, empleando el funcional híbrido B3LYP[16] y la base de cálculo 6-31G* [17]. Las moléculas de la base de datos tienen dos grupos hidroxilo, un sustituyente de acilo y un grupo amino terminal como parte de sus características estructurales, pero hay algunos casos en la Tabla 1 como la tiramina donde los compuestos fueron clasificados como catecolaminas (4), la cual tiene solo un grupo -OH, no los dos requeridos para la caracterización química de las catecolaminas, por esta razón en este trabajo se propusieron dos índices en la definición de dos escalas, a partir de los cuales se buscaba una conceptualización más rigurosa de las catecolaminas como grupo. También en este estudio se plantearon algunas consideraciones de la ruta de

reacción de las catecolaminas con respecto a las monoaminas biogénicas asociadas con estas moléculas.

7.2.3 Índices de Semejanza

El índice de Carbó.[18,19] fue expresado en la ec. 32 e incluye la distancia euclídias, D_{AB} , entre las densidades electrónicas $\rho_A(\mathbf{r})$ y $\rho_B(\mathbf{r})$ de dos moléculas A y B. El índice de Carbó puede ser representado geométricamente usando el concepto del coseno del ángulo subtendido entre las funciones de densidad, considerándolas como vectores; este índice se define matemáticamente as:

$$R_{AB} = \frac{\int \rho_A(\mathbf{r})\rho_B(\mathbf{r})d\mathbf{r}}{[(\int \rho_A^2(\mathbf{r})d\mathbf{r})(\int \rho_B^2(\mathbf{r})d\mathbf{r})]^{1/2}} = \frac{Z_{AB}}{\sqrt{Z_{AA}Z_{BB}}} \quad (ec. 7.1)$$

Esta ecuación involucra la integral de solapamiento Z_{AB} , entre las densidades electrónicas de las moléculas A y B. Z_{AA} y Z_{BB} con las medidas de auto-semejanza de las moléculas A y B [20,21]. En el caso particular de la molécula 4, la cual difiere de la molécula 6 (que fue la molécula empleada como fragmento molecular común) por la presencia de un grupo –OH adicional, esta medida se calcula tomando la parte correspondiente de cuadrado de la norma de la función de densidad que pertenece al fragmento de interés (la molécula 6). Esta auto-semejanza ocupa los elementos de la diagonal de las matrices de semejanza.

El uso de este tipo de descriptores en fragmentos molecular [22,12] se basa parcialmente en el teorema holográfico de la densidad electrónica, el cual asegura que la información contenida en la densidad electrónica total está presente también en la densidad local de un fragmento particular [23]. El índice de Carbó tiene el mismo rango de valores que la desigualdad de la integral de Schwartz [24]:

$$\left[\int \rho_A(\mathbf{r})\rho_B(\mathbf{r})d\mathbf{r} \right]^2 \leq \int \rho_A^2(\mathbf{r})d\mathbf{r} \int \rho_B^2(\mathbf{r})d\mathbf{r} \quad (\text{ec. 7.2})$$

7.2.4. Índices Alfa y beta propuestos

Los estados de transición de las reacciones químicas tienen vidas medias del orden de 10^{-13} s, el tiempo de una sola vibración del enlace. Los estados de transición pueden ser estudiados computacionalmente porque corresponden a los puntos de silla sobre la superficie de energía potencial. Así, el estado de transición es la configuración crítica de un sistema que se ubica en el punto más alto en la superficie de energía potencial, sus características influyen decisivamente en el comportamiento dinámico de la reacción.

Los índices Alfa y beta fueron postulados por Cioslowski [11] a través del postulado de Hammond, el cual se refiere al estado de transición en una reacción química en una configuración particular a lo largo de la coordenada de reacción. El estado de transición (TS) se define como el estado que corresponde a la energía máxima a lo largo de la coordenada de reacción. En este punto, se asume que la colisión de las especies reactantes siempre conduce a la formación de productos. De acuerdo con el postulado de Hammond el estado de transición tendrá rasgos estructurales y energía intermedios entre los de los compuestos de partida y los productos pero estructura tendrá mayor similitud a la especie que tenga una mayor entalpía. En términos estructurales cuanto más exoérgica es la reacción, más similar será la configuración en la barrera a la del reactante [25].

Para cuantificar el postulado de Hammond, Cioslowski [11] mostró que la cuantificación de este postulado puede ser obtenido mediante la inclusión de dos nuevos parámetros, el parámetro de la isosincronicidad (α) entre los estados de transición (TSs) y los reactantes (β). Estos parámetros α y β se definen mediante las distancias entre A y B ($d_{A,B}$), las cuales satisfacen las propiedades de las distancias euclídeas. En particular, α se obtiene mediante:

$$\alpha = \frac{[d_{A.TS} + d_{B.TS}]}{d_{A,B}} \quad (ec.7.3)$$

El parámetro α nunca es menor que 1 y toma el valor de 1 cuando la distancia de A o B es muy pequeña en el TS o en el caso de reacciones, donde el TS pierde su similitud con los reactantes en exactamente el mismo grado que adquiere su similitud con los productos.

El índice β fue definido por Cioslowski:

$$\beta = \frac{[d_{A.TS} + d_{B.TS}]}{d_{A,B}} \quad (ec.7.4)$$

El índice β toma valores que van de -1 a 1 . Si los reactantes (A) y TS están cerca en las superficies de energía potencial a los productos (B) y al TS, el valor de β es negativo, de otra manera es positivo. Un valor negativo de β corresponde a la reacción con un TS (temprano) y un valor positivo a una reacción con un TS (tardío). Teniendo en cuenta algunas ideas previas,[11] este trabajo intenta aplicar los índices alpha and beta para caracterizar de manera cuantitativa a las catecolaminas como grupo. El nuevo índice se definió como:

$$\alpha = \frac{Z_{AC}^x + Z_{CB}^x}{Z_{AB}^x} \quad (ec.7.5)$$

donde Z_{AB}^x cuando $x = 0$ representa el índice de solapamiento cuántico molecular, $x = C$ representa el índice cuántico molecular de Coulomb y Z_{AB} es el índice de Carbó [Eq. (2)]. El índice α toma un rango definido mediante el intervalo $(0, 2)$. Sus valores están determinados por el límite

$$\lim_{Z_{AB}^x \rightarrow 0} \alpha$$

donde un valor de cero significa que hay disimilitud entre A y C como también C y B, cuando toma el valor de 2 es porque $A = B = C$. Se puede interpretar α como un índice de correlación grupal, el cual permite justificar que las catecolaminas sean un grupo separado de neurotransmisores con respecto a las monoaminas biogénicas. Se toma la feniletilamina (Fig. 1) como molécula C, para ambos índices de semejanza.

Dado que Cioslowski calculó los índices α , β y γ , para transferencias intramoleculares de proton entre aniones catecolato, en la Universidad de Cartagena se desarrolló su propuesta aplicándola en el estudio de las aminas biogénicas.

Como segundo índice de caracterización de las catecolaminas se utilizó el índice beta el cual permitirá evaluar las distancias euclidianas, suministrando otra manera de describir y vincular las catecolaminas con respecto a las monoaminas biogénicas. Matemáticamente este índice se define así:

$$\beta = \frac{D_{AC}^x - D_{CB}^x}{D_{AB}^x} \quad (ec.7.6)$$

donde D_{AB}^x , cuando $x = 0$ es la distancia euclídea de solapamiento, para $x = C$ es la distancia Euclídea de Coulomb. Este índice tiene un rango de convergencia (1,1), donde un valor negativo significa que la molécula A es más similar a la molécula C con respecto a la molécula B, cuando toma el valor de 1 significa que la molécula B es más similar a la molécula C, con respecto a la molécula A, teniendo en consideración que a medida que la distancia euclídea disminuye la semejanza cuántica aumenta.

Posteriormente, en el mismo Grupo de Química Cuántica y Teórica de la Universidad de Cartagena, se propuso una nueva interpretación de los índices alfa y beta para definir las medidas cuantitativas de solapamiento y Coulomb de los inhibidores multidiana que enlazan tanto la enzima colinesterasa como la monoamina oxidasa. La base de datos fue la serie de inhibidores duales

reportadas por Bolea *et al.* [13] quienes sintetizaron moléculas con una fracción benzilpiperidina del donepezil y otra fracción indolil propargilamino del inhibidor de la MAO N-[(5-benziloxi-1-metil-1H-indol-2-il)metil]-N-metilprop-2-in-1-amina. En este sentido, el estudio utilizó los índices alfa y beta para cuantificar la naturaleza híbrida de los compuestos estudiados. Los índices propuestos pueden ser útiles en el desarrollo de fármacos para el tratamiento de la enfermedad de Alzheimer [14]

7.3 RESULTADOS

En este estudio, se efectuó un análisis QMS para cuantificar la semejanza estructural y electrónica de las catecolaminas con respecto a las monoaminas biogénicas. Posteriormente estos datos se emplearon para caracterizar las catecolaminas como grupo aislado de neurotransmisores, teniendo en cuenta que la catecolamina precursora es la L-tirosina se asociaron los resultados teóricos con información experimental tomada de la literatura, permitiendo determinar la cuantificación de cada monoamina estudiada con respecto al reactante y el producto en la síntesis de catecolaminas.

Las Tablas 2–6 muestran los índices de semejanza cuántica molecular de las catecolaminas (Tablas 2–5) y la tirosina precursora (Tabla 6), con respecto a las monoaminas biogénicas.

Tabla 7.2. Índice de semejanza para la NA vs. Base de datos.						
$Z_{1-B}^{[a]}$	$I_{1B} (O)^{[b]}$	$I_{1B} (C)^{[c]}$	$Z_{1B} (O)^{[d]}$	$Z_{1B} (C)^{[e]}$	$D_{1B} (O)^{[f]}$	$D_{1B} (C)^{[g]}$
Z_{1-1}	12.158	909.718	1.000	1.000	0.000	0.000
Z_{1-2}	7.807	856.399	0.676	0.954	2.741	9.076
Z_{1-3}	7.239	813.400	0.707	0.929	2.509	11.302
Z_{1-4}	8.566	859.975	0.723	0.944	2.425	10.045
Z_{1-5}	7.086	772.468	0.763	0.961	2.329	8.643
Z_{1-6}	6.390	836.531	0.586	0.929	3.027	11.281

[a] semejanza norepinefrina versus conjunto molecular

[b] interacción de solapamiento

[c] interacción de Coulomb

[d] índice de semejanza de solapamiento

- [e] índice de semejanza de Coulomb
 [f] distancia euclídea de solapamiento
 [g] distancia euclídea de Coulomb

Tabla 7.3. Índice de semejanza para la DA vs. Base de datos.						
$Z_{2-B}^{[a]}$	$l_{2B} (O)^{[b]}$	$l_{2B} (C)^{[c]}$	$Z_{2B} (O)^{[d]}$	$Z_{2B} (C)^{[e]}$	$D_{2B} (O)^{[f]}$	$D_{2B} (C)^{[g]}$
Z ₂₋₂	10.970	885.452	1.000	1.000	0.000	0.000
Z ₂₋₃	6.129	813.322	0.630	0.940	2.708	10.179
Z ₂₋₄	5.710	758.485	0.623	0.939	2.695	10.227
Z ₂₋₅	7.454	880.873	0.693	0.949	2.568	9.807
Z ₂₋₆	8.566	859.975	0.827	0.968	1.902	7.490

Tabla 7.4. Índice de semejanza para la A vs. Base de datos.						
$Z_{3-B}^{[a]}$	$l_{3B} (O)^{[b]}$	$l_{3B} (C)^{[c]}$	$Z_{3B} (O)^{[d]}$	$Z_{3B} (C)^{[e]}$	$D_{3B} (O)^{[f]}$	$D_{3B} (C)^{[g]}$
Z ₃₋₃	8.619	844.812	1.000	1.000	0.000	0.000
Z ₃₋₄	7.614	779.813	0.923	0.989	2.708	4.615
Z ₃₋₅	7.918	871.026	0.831	0.961	1.820	8.673
Z ₃₋₆	5.967	824.411	0.650	0.950	2.509	9.305

Tabla 7.5. Índice de semejanza TA vs. Base de datos.						
$Z_{4-B}^{[a]}$	$l_{4B} (O)^{[b]}$	$l_{4B} (C)^{[c]}$	$Z_{4B} (O)^{[d]}$	$Z_{4B} (C)^{[e]}$	$D_{4B} (O)^{[f]}$	$D_{4B} (C)^{[g]}$
Z ₄₋₄	7.893	736.115	1.000	1.000	0.000	0.000
Z ₄₋₅	8.145	822.697	0.893	0.972	1.461	7.949
Z ₄₋₆	7.614	779.813	0.923	0.989	1.133	4.615

Tabla 7.6. Índice de semejanza para la Y vs. Base de datos.						
$Z_{5-B}^{[a]}$	$l_{5B} (O)^{[b]}$	$l_{5B} (C)^{[c]}$	$Z_{5B} (O)^{[d]}$	$Z_{5B} (C)^{[e]}$	$D_{5B} (O)^{[f]}$	$D_{5B} (C)^{[g]}$
Z ₅₋₅	10.531	972.463	1.000	1.000	0.000	0.000
Z ₅₋₆	7.918	871.026	0.831	0.961	1.821	8.673

En la Tabla 7.7, se muestran los valores de los índices alfa y beta que permitirán caracterizar las catecolaminas como grupo y relacionar las monoaminas biogénicas, mediante la definición de la escala de la convergencia alfa y beta. Como ejemplo de la aplicación de la relación de grupo de estos índices se determinó la feniletilamina con respecto a la norepinefrina.

Tabla 7.7. Coeficientes α y β propuestos para la FEA con respecto a la NE			
Alfa (solapamiento)	Beta (solapamiento)	Alfa (Coulomb)	Beta (Coulomb)
$\alpha^{1o[a]} = 1.857$	$\beta^{1o[a]} = -0.518$	$\alpha^{1c} = 1.967$	$\beta^{1c} = -0.302$
$\alpha^{2o} = 2.392$	$\beta^{2o} = -0.032$	$\alpha^{2c} = 2.032$	$\beta^{2c} = -0.121$
$\alpha^{3o} = 1.782$	$\beta^{3o} = -0.378$	$\alpha^{3c} = 1.989$	$\beta^{3c} = -0.073$
$\alpha^{4o} = 1.964$	$\beta^{4o} = -0.471$	$\alpha^{4c} = 2.006$	$\beta^{4c} = -0.879$

El mayor valor de coeficiente alfa (Tabla 7) a medida que la tiramina se asocia con la molécula, la cual es el fragmento molecular común y el precursor de la tirosina, teniendo en cuenta la escala de convergencia propuesta en este estudio, un valor cercano a 2 del coeficiente alfa, esto es una indicación de que la tiramina es la que tiene la mayor semejanza cuántica de solapamiento y de coulomb con el fragmento común (Fig. 1) y su tirosina precursor. También en esta Tabla 7 se pueden observar los valores beta, estos valores permiten asociar el fragmento molecular común, el cual en esta base de datos es la feniletilamina con el camino de reacción de las catecolaminas como también el coeficiente beta, esto tiene una escala de convergencia para las catecolaminas, dentro del rango de la escala definida de convergencia para las catecolaminas.

Las distancias Euclídeas (Tablas 2–6) indican que la dopamina tiene una mayor semejanza estructural con el fragmento común (**Fig. 7.2**) que la L-tirosina, esta evidencia puede observarse en la ruta de síntesis de la dopamina (**Fig. 7.1**).

La Figura 7.1 muestra que la dopamina tiene dos grupos –OH en las posiciones 4 y 5 y adicionalmente tiene el fragmento común feniletilamina (Tabla 1) a diferencia de la L-tirosina la cual solo tiene un grupos –OH y un grupo carboxilo, lo cual justifica el signo negativo en el índice beta, lo cual resulta en una mayor correlación en el nivel grupal de las catecolaminas a partir de cada otra, en la medida que el índice alfa se relaciona con la correlación estérica a nivel de grupo a diferencia del índice beta el cual se relaciona con la correlación electrónica a nivel de grupo. Tomando en cuenta que estos índices

involucran el precursor y el producto de la reacción de cada catecolamina se obtuvieron las escalas alfa y beta, respectivamente (**Fig. 7.3**). Los índices alfa y beta calculados son consistentes con el comportamiento estérico y electrónico de la ruta de reacción experimental de las catecolaminas.

La Figura 3 presenta la tendencia en la relación de una monoamina piogénica en particular, la feniletilamina con respecto a las catecolaminas de la Tabla 1. Se observa que esta relación está cuantificada mediante la escala de convergencia que se propone en este trabajo. Los valores beta indican el grado de relación entre las catecolaminas y las monoaminas biogénicas, teniendo en cuenta que en este conjunto el fragmento molecular común consiste en una molécula en particular, la monoamina biogénica feniletilamina y no un fragmento molecular como es usual en tales relaciones.

Tabla 7.8. Índices de reactividad global para los inhibidores AChE/iMAO.			
Compuesto	μ^a (eV)	η^b (eV)	ω^c (eV)
1	-3.9031	3.2956	2.3113
2	-2.6169	4.8877	0.7006
3	-2.2839	4.3457	0.6001
4	-2.4981	4.9049	0.6362
5	-2.4937	4.9084	0.6334
6	-2.5021	4.9079	0.6378
7	-2.5085	4.9136	0.6432
8	-2.4945	4.9062	0.6341
9	-2.4916	4.9043	0.6329

^aPotencial químico

^bDureza

^cElectrofilicidad

La importancia de formular herramientas que permitan la cuantificación de relaciones biológicas y generen escalas de grupo es importante en la síntesis de catecolaminas y para determinar qué tipo de similitudes son importantes para consideraciones experimentales. La **Figura 7.3** muestra las tendencias de estas relaciones ejemplificada con una amina en particular, la feniletilamina con relación a la NA. Los valores de β indican el grado de relación entre las catecolaminas y las animas biogénicas,

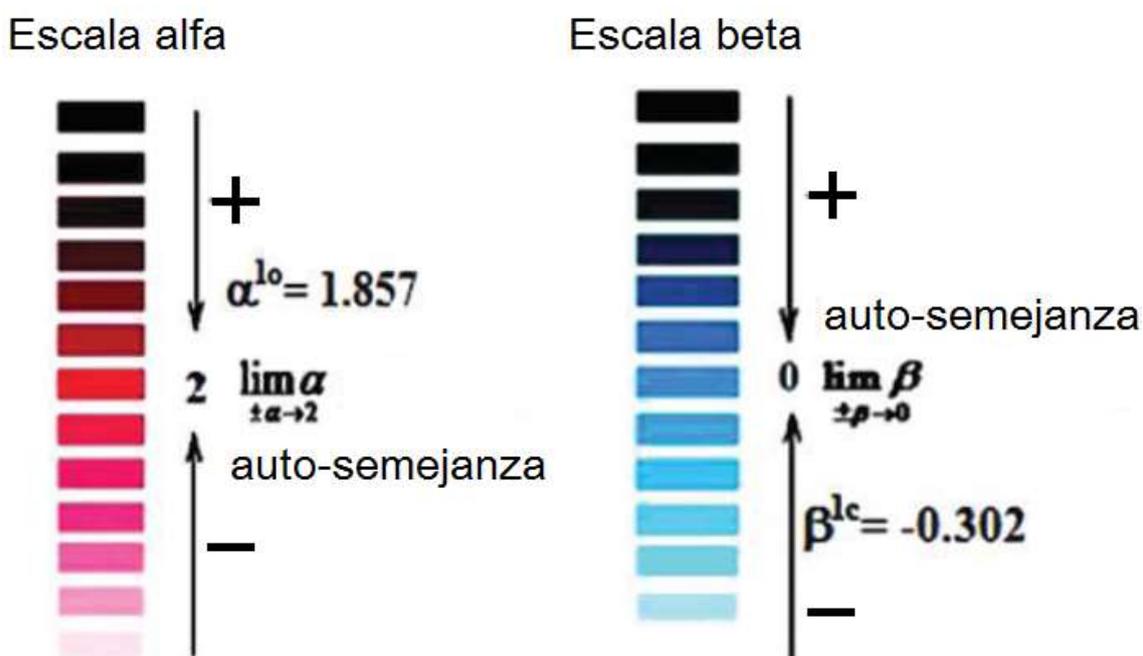


Figura 7.3 Escala de convergencia de la feniletilamina con respecto a la NA.

7.4 CONCLUSIONES

En este estudio se encontraron los índices α y β de convergencia los cuales permitieron relacionar las catecolaminas aisladas como grupo, para efectuar una caracterización más rigurosa de estas, para ayudar a diferenciar las monoaminas biogénicas relacionadas. Para lograr esta cuantificación se efectuaron cálculos de QMS, observándose que el descriptor de Coulomb produce mejores resultados que la semejanza de solapamiento para describir un nivel de convergencia de una monoamina biogénica particular con las

características estructural y electrónicas de la ruta de reacción se la síntesis de las catecolaminas.

Por lo tanto, las escalas de convergencia son ampliamente usadas en la síntesis experimental para suministrar una herramienta para cuantificar los conceptos biológicos para correlacionar grupos de relevancia biológica. También se propuso una metodología para definir grupos aislados a través de relaciones cuantitativas de transitividad. Con base en los descriptores de semejanza cuántica y apoyando las interpretaciones con resultados experimentales, se formuló una herramienta para agrupar las aminas biogénicas que suministra una clara correlación, definiendo los grupos de compuestos aislados de una manera diferente a la que se había presentado previamente en la literatura (series químicas y funciones), teniendo en cuenta que la búsqueda de descriptores de caracterización grupal es un área de creciente interés en años recientes.

Los resultados de la función de Fukui suministran información sobre la interacción posible de estas moléculas y ofrece explicaciones acerca de la inhibición de la AChE y la MAO B que pueden ser considerados en términos de interacciones de retrodonación sobre el anillo central, el cual participa en la inhibición dual de la colinesterasa. La inhibición de la MAO puede ser estudiada mediante los efectos estéricos y electrónicos. Por ejemplo en el caso de la molécula 5 el anillo central de esta molécula es favorable a los ataques nucleofílicos y electrofílicos sugiriendo que el proceso de retrodonación sobre el anillo central es una interacción adicional que estabiliza la molécula.

7.5. REFERENCIAS

1. Morales-Bayuelo, A., Torres, J. & Vivas-Reyes, R. (2012), Quantum molecular similarity analysis and quantitative definition of catecholamines with respect to biogenic monoamines associated: Scale alpha and beta of quantitative convergence. *Intern. J. Quantum Chemistry*. **112** (14): 2637–2642.

2. Hammond, G.S. (1995). A Correlation of Reaction Rates. *J Am Soc.* **77**: 334-338.
3. G. A. Arteca, P. G. Mezey. (1988). Shape characterization of some molecular model surfaces. *J Comput Chem.* **9**: 554-563.
4. Cioslowski, J. (1991). Quantifying the Hammond postulate: intramolecular proton transfer in substituted hydrogen catecholate anions. *J Am Chem Soc.* **113** (18): 6756-6760.
5. Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., McNamara, J. O. & White, L. E. Neuroscience, 4th ed.; Sinauer Associates: Sunderland, Mass, 2008; pp.137–138.
6. Joh, T. H. & Hwang, O. (1987). Dopamine beta-hydroxylase: biochemistry and molecular biology. *N Y, A. Acad Sci.* **493**: 342-350.
7. Zouhal, H., Lemoine-Morel, S., Mathieu, ME., Casazza, GA. & Jabbour, G. (2013). Catecholamines and obesity: effects of exercise and training. *Sports Med.* **43** (7): 591-600.
- 8 Dahlstrom, A. & Fuxe, K. (1966). Monoamines and the pituitary gland. *Acta Endocrinol* (Copenh). **51**: 301-314.
9. Reis, D. J. (1971). Brain monoamines in aggression and sleep. *Clin Neurosurg.* **18**: 471-502.
10. Rosenberg, P. H., Klinge, E. & Anaesth, Br J. (1974). Some effects of enflurane anaesthesia on biogenic monoamines in the brain and plasma of rats. **46** (10): 708-713.
11. Carlson, H. E. & Robbins, J. (1974). Effects of hormones and nucleotides on ciliary beating in frog esophagus and guinea pig trachea. *Life Sci.* **14**: 2413-2426.

12. Morales-Bayuelo, A., Ayazo, H. & Vivas-Reyes, R. (2010). Three-dimensional quantitative structure–activity relationship CoMSIA/CoMFA and LeapFrog studies on novel series of bicyclo [4.1.0] heptanes derivatives as melanin-concentrating hormone receptor R1 antagonists. *Eur J Med Chem.* **45**: 4509-4522.

13. (a) X. Gironés, D. Robert, R. Carbó -Dorca, *J Comput Chem* 2001, 22, 255; (b) P. Bultinck, T. Kuppens, X. Gironés, R. Carbó-Dorca, *J Chem Inf Comput Sci* 2003, 43, 1143; (c) X. Gironés, R. Carbó-Dorca., *J Comput Chem.* 2004, 25, 153.

14. Morales-Bayuelo, A., Baldiris, R. & Ricardo Vivas-Reyes, R. (2013). Scale Alpha and Beta of Quantitative Convergence and Chemical Reactivity Analysis in Dual Cholinesterase/Monoamine Oxidase Inhibitors for the Alzheimer Disease Treatment Using Density Functional Theory (DFT). *Journal of Theoretical Chemistry.* **2013**: 1- 13.

Anexo 1

Tabla 1: Descripción del Total Score de los ligandos con sus respectivos sustituyentes e IC₅₀.

Número	Sustituyentes			Actual pIC ₅₀	Total Score
	R ₁	X, Y	R ₂		
001	(CH ₃) ₂	CH	-Ph-4-OMe	5,85	3,96
002	(CH ₃) ₃	C	-Ph-4-OMe	5,68	3,50
003	(Me) ₂ N-	CH ₂	-Ph-4-OMe	7,36	5,31
004	Morfolina-4-yl	CH ₂	-Ph-4-OMe	7,68	5,42
005	Piperazina-1-yl	CH ₂	-Ph-4-OMe	7,48	5,37
006	Etil NH-	CH ₂	-Ph-4-OMe	6,89	4,92
007	N-metil piperazina	CH ₂	-Ph-4-OMe	7,92	5,91
008	4-Aminometilpiperidina	CH ₂	-Ph-4-OMe	7,92	5,86
009	4-Aminopiperidina	CH ₂	-Ph-4-OMe	8,10	6,44
010	4-Hidroximetilpiperidina	CH ₂	-Ph-4-OMe	7,96	6,01
011	4-Amidopiperazina	CH ₂	-Ph-4-OMe	8,10	6,38
012	4-Amidinopiperazina	CH ₂	-Ph-4-OMe	8,16	6,62
013	H	CH ₂	-Ph-4-OMe	6,57	4,34
014	Bencil	NH	-Ph-4-OMe	7,02	5,13
015	Fenil	NH	-Ph-4-OMe	7,50	5,38
016	n-butil	NH	-Ph-4-OMe	7,08	5,15
017	(Me) ₂ N	NH	-Ph-4-OMe	8,30	7,03
018	4-metilpiperazina	NH	-Ph-4-OMe	7,92	6,14
019	Morfolina-4-yl	NH	-Ph-4-OMe	7,75	5,55
020	Piperidina-1-yl	NH	-Ph-4-OMe	7,66	5,43
021	Pirrolidina-1-yl	NH	-Ph-4-OMe	7,82	6,01
022	H	CH ₂	-Ph	6,62	4,61
023	H	CH ₂	-PH-4-Me	6,80	4,86
024	H	CH ₂	-PH-4-Et	6,55	4,48
025	H	CH ₂	-PH-4-n-Pr	6,31	4,25
026	H	CH ₂	-PH-4-OH	6,28	4,12
027	-Ph-4-NH ₂	CH ₂	-PH-4-OMe	7,42	5,27
028	H	CH ₂	-PH-4-NMe ₂	6,50	4,40
029	H	CH ₂	-PH-4-piperidino	6,05	4,07
030	H	CH ₂	-PH-4-morfolino	6,43	4,39
031	H	CH ₂	-PH-4-SMe	6,92	4,82

032	Morfolino	CH ₂	-PH-4-NMe ₂	7,44	5,31
033	4-(OH)piperidin-1-yl	CH ₂	-PH-4-NMe ₂	7,48	5,33
034	4-(Aminometil)piperidin-1-yl	CH ₂	-PH-4-NMe ₂	7,82	5,87
035	N-metilpiperazin-1-yl	CH ₂	-PH-4-NMe ₂	7,54	5,66
036	Morfolino	CH ₂	-PH-4-morfolino	7,43	5,30
037	4-(OH)piperidina-1-yl	CH ₂	-PH-4-morfolino	7,07	5,12
038	4-(Aminometil)piperidin-1-yl	CH ₂	-PH-4-morfolino	7,59	5,57
039	H	NH	3-tienil	8,05	6,23
040	N-metilpiperazin-1-yl	CH ₂	-PH-4-morfolino	7,38	5,22
041	4-(Aminometil)piperidin-1-yl	CH ₂	Et	6,62	4,58
042	4-(Aminometil)piperidin-1-yl	CH ₂	Ciclopropil	7,26	5,19
043	4-(Aminometil)piperidin-1-yl	CH ₂	Ciclohexano	7,02	5,05
044	H	NH	Ciclopropil	6,85	4,84
045	H	CH ₂	4-Piridil	7,12	5,11
046	H	CH ₂	2-Tienil	6,75	4,76
047	H	NH	2-Tienil	7,96	5,99
048	H	NH	2-Tienil-1,3-OMe	7,77	5,75
049	H	NH	2-Tienil-1,5-Me	7,89	5,82
050	H	NH	2-Furanil	7,07	5,05
051	H	NH	2-Tienil-1,5-CO ₂ Et	6,89	4,90
052	H	NH	2-Tienil-1,5-Cl	7,89	5,88
053	H	NH	3-Pirrolil,1-Me	7,59	5,56
054	Dimetilamino	NH	2-Tienil	7,44	5,32
055	Dimetilamino	NH	5-(OMe) tien-2-yl	7,50	5,46
056	Dimetilamino	NH	5-(Me)tien-2-yl	7,60	5,51
057	Dimetilamino	NH	5(CO ₂ EtMe)tien-2-yl	7,55	5,48
058	Dimetilamino	NH	3-Tienil	7,62	5,57
059	Dimetilamino	NH	5-(Cl) tien-3-yl	8,16	6,36
060	Dimetilamino	NH	2,5-(di-Me)tien-3-yl	7,59	5,43
061	Dimetilamino	NH	Furan-2-yl	7,59	5,47
062	Dimetilamino	NH	2,4-(di-Me)tiazol-5-yl	8,40	7,38
063	Morfolina-4-yl	NH	5-(Me) tien-2-yl	8,00	6,05
064	Morfolina-4-yl	NH	5-(CO ₂ EtMe) tien-2-yl	7,51	5,43

065	Morfolina-4-yl	NH	5-(Cl) tien-3-yl	8,00	6,18
066	4-(Metil)piperazin-1-yl	NH	5-(CO ₂ EtMe) tien-2-yl	7,44	5,32
067	4-(Aminometil) piperidin-1-yl	CH ₂	Isopropil	7,32	5,29
068	4-(Metil)piperazin-1-yl	NH	2,5-(di-Me)tien-3-yl	7,92	6,00
069	4-(Metil)piperazin-1-yl	NH	2,4-(di-Me)tiazol 5-yl	8,10	6,19
070	(Me) ₂ CHCONH-	NH	-Ph-4-OMe	7,72	5,67
071	4-(OH)Ph(CH ₂) ₂ CONH-	NH	-Ph-4-OMe	7,89	5,86
072	4-(Ome)PhCONH-	NH	-Ph-4-OMe	7,82	5,75
073	3-(NO ₂)PhCONH-	NH	-Ph-4-OMe	7,62	5,57
074	3,4,5-(tri-OMe)PhCONH-	NH	-Ph-4-OMe	7,70	5,63
075	3-(Me)PhCONH-	NH	-Ph-4-OMe	7,89	5,84
076	3,4-(di-OMe)PhCONH-	NH	-Ph-4-OMe	7,39	5,35
077	(4-OH,3-NH ₂)PhCONH-	NH	-Ph-4-OMe	7,62	5,52
078	2,5-(di-Cl)PhCONH-	NH	-Ph-4-OMe	7,44	5,45
079	3,4-(di-OH)PhCONH-	NH	-Ph-4-OMe	8,00	6,04
080	3,5-(di-NH ₂)PhCONH-	NH	-Ph-4-OMe	7,41	5,46
081	MeOCONH-	NH	-Ph-4-OMe	7,80	5,76
082	2-(OH)PhCONH-	NH	-Ph-4-OMe	8,05	6,13
083	Naftalen-2-yl CONH-	NH	-Ph-4-OMe	7,46	5,45
084	BnCONH-	NH	-Ph-4-OMe	7,48	5,44
085	PhCONH-	NH	-Ph-4-OMe	7,92	5,94
086	4-pirridiloCONH-	NH	-Ph-4-OMe	7,92	5,88
087	3-pirridiloCONH-	NH	-Ph-4-OMe	7,80	5,79
088	MeCONH-	NH	-Ph-4-OMe	7,43	5,36
089	4-(OH)PhCONH-	NH	-Ph-4-OMe	8,05	6,07
090	H ₂ NCONH-	NH	-Ph-4-OMe	7,01	4,98
091	3-(NH ₂)PhCONH-	NH	-Ph-4-OMe	7,89	5,90
092	2,4-(di-OH)PhCONH-	NH	-Ph-4-OMe	7,80	5,75
093	4-NH ₂ PhCONH-	NH	-Ph-4-OMe	8,00	6,00
094	2(Dimetilamino)etilamino			7,68	5,69
095	2(Pirrolidin-1-yl)etilamino			7,60	5,47
096	2(Piperidina-1-yl)etilamino			7,51	5,52
097	2-(Morfolin-4-yl)etilamino			7,29	5,23
098	Piperidin-1-yl			7,60	5,55
099	3(Dimetilamino)Piperidin-1-yl			7,62	5,57
100	4(Dimetilamino)Piperidin-1-yl			7,64	5,54
101	Piperazin-1-yl			7,42	5,38

102	4-(Etil)piperazin-1-yl			7,72	5,65
103	3-(Amino)pirrolidin-1-yl			7,41	5,39
104	3-(Metilamino)pirrolidin-1-yl			7,96	6,10
105	3(Dimetilamino)pirrolidin-1-yl			7,36	5,23
106	Azepan-1-yl			7,64	5,59
107	4-(Metil)piperazin-1-yl			7,50	5,41
108	[1,4]Diazepan-1-yl			8,22	7,14
109	4-(Metil)-[1,4]diazepan-1			7,82	5,86
110	4-(Etil)-[1,4]diazepan-1-yl			7,85	5,78
111	H	H	C	4,59	3,51
112	Acetamida	H	C	6,29	4,27
113	H	NH ₂	C	3,68	2,56
114	NH ₂	H	C	5,08	4,09
115	H	H	N	5,33	4,27
116	OH	H	C	5,28	5,18
117	Formamido	H	C	7,10	5,07