UNIVERSIDAD AUTONOMA DE NUEVO LEON
Facultad de Ciencias Químicas
División de Estudios Superiores

Nuevos Enfoques en Estudios de Relación Estructura-Respuesta para el Diseño de Moléculas (Productos)

TESIS

Presentada como Requisito Parcial para Obtener el Grado Académico de Maestro en Ciencias, con Especialidad en Química Orgánica

Por:
JACINTO GUADALUPE RODRIGUEZ GOMEZ

Agosto 1991

UNIVERSIDAD AUTONOMA DE NUEVO LEON

Facultad de Ciencias Químicas

División de Estudios Superiores

Nuevos Enfoques en Estudios de Relación Estructura-Respuesta para el Diseño de Moléculas (Productos)

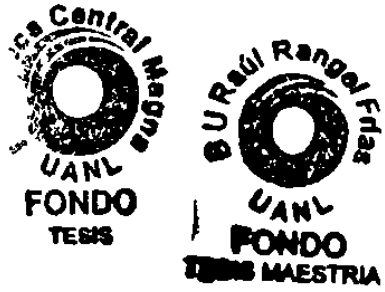
TESIS

Presentada como Requisito Parcial para Obtener el Grado Académico de Maestro en Ciencias, con Especialidad en Química Orgánica

Por:

JACINTO GUADALUPE RODRIGUEZ GOMEZ

Agosto 1991


Central Magna
UANL
FONDO
TESIS
B U Raúl Rangel Frías
UANL
FONDO
MAESTRIA

UNIVERSIDAD AUTONOMA DE NUEVO LEON

Facultad de Ciencias Químicas

Division de Estudios Superiores

M.C Blanca Najera de Quilantan
Coordinador de la Escuela de Graduados en Ciencias
PRESENTE

Mediante este conducto hacemos de su conocimiento que la TESIS elaborada por el Sr. JACINTO GUADALUPE RODRIGUEZ GOMEZ, titulada:

"NUEVOS ENFOQUES EN ESTUDIOS DE RELACION ESTRUCTURA-RESPUESTA PARA EL DISEÑO DE MOLECULAS (PRODUCTOS)"

ha sido aceptada como requisito parcial para obtener el grado académico de:

MAESTRO EN CIENCIAS

ESPECIALIDAD EN QUIMICA ORGANICA

en virtud de haber cumplido integramente con el reglamento de tesis vigente.

ATENTAMENTE

EL COMITE DICTAMINADOR

Dr. Gabriel Gójon Zorrilla
PRESIDENTE

Dr. Mario H. Gutiérrez V.
SECRETARIO

Dr. Porfirio Caballero Mata
VOCAL

Vo. Bo. EL COORDINADOR ACADEMICO

M.C. CORALIA MARTINEZ HINOJOSA

RESUMEN

Con el inicio de la apertura del País al mercado de libre comercio, aparecieron nuevos competidores extranjeros y nuevos requerimientos internacionales para exportar, así como la pronta posibilidad de patentar productos en México. Estos cambios fuerzan a todos aquellos dedicados a la Ciencia y Tecnología, a enfocarse hacia la generación de conocimiento y formación de recursos humanos en aspectos relacionados con las estrategias y métodos para realizar IyD en tecnologías de producto, de tal manera que también se tenga la capacidad de aplicar, mejorar e innovar en esta área de relevancia contemporánea.

Dentro de las implicaciones particulares para el desarrollo de tecnologías de producto, sobre todo en la etapa de diseño del producto consistente en una molécula (*e.g.* fármaco) es común aplicar ciertas metodologías con el fin de eficientar el proceso de IyD, *i.e.*, sintetizar y evaluar un mínimo de estructuras químicas (sustancias) a partir de las cuales presumiblemente, pueda llegarse (predecir) a la estructura que posteriormente se convertirá en el producto comercial, de una manera rápida y económica.

El trabajo presente se enmarca en el desarrollo de metodologías aplicables al diseño de moléculas (productos).

El Capítulo I versa sobre una nueva aplicación de los diseños fatoriales (y su modalidad de "compuesto centrado") en estudios de relación estructura-respuesta (*i.e.* propiedad física, biológica, química, etcétera). Las ventajas que presenta son: La introducción relativamente fácil , de los términos de interacción ($b_{ij}X_iX_j$) y curvatura ($b_{ii}X_i^2$) en el modelo, así como, la factibilidad del análisis multirrespuesta y el requerimiento de un mínimo de estructuras para determinar la superficie de respuesta o el grado de contribución de los rasgos estructurales en la respuesta bajo estudio. La principal dificultad para su aplicación detectada en el desarrollo del presente trabajo, parece estar relacionada con la selección de una escala que transforme las características estructurales discontinuas en continuas, para determinar la superficie de respuesta y lograr la capacidad de predicción, dentro de la región experimental seleccionada.

Cuando el número de variables (*e.g.* posiciones en un anillo aromático) y sus valores (*e.g.* tipo de sustutiyente) implican un gran número (combinaciones) de estructuras por sintetizar y evaluar, en la búsqueda de una molécula con la respuesta óptima, el método Monte Carlo parece ser una herramienta útil para determinar (predecir) a partir de unas cuantas estructuras (combinaciones) sintetizadas y evaluadas, la fracción de moléculas (del total de combinaciones) con cierto valor o intervalo de la propiedad bajo estudio (*e.g.* $LD_{50} \geq$

5.0). En el Capítulo II se presenta la aplicación y limitaciones de este enfoque generado y contrastado durante el presente trabajo, en algunos análisis retrospectivos. Al igual que los diseños factoriales, parece una estrategia muy eficiente. Su eficacia dependerá del tamaño de la muestra (N) o número de moléculas pre-evaluadas y de la suerte del experimentador.

Por otra parte, dentro de las múltiples formas de representar algún rasgo o característica estructural de una molécula (estructura), para estudios de correlación (y predicción), estan los índices topológicos. En el presente estudio también se analiza otro nuevo enfoque aquí generado y relacionado con ciertas tablas de conectividad usadas para obtener el índice (topológico) de Wiener. La principal ventaja detectada es su capacidad de representar UNIVOCAMENTE cada molécula y la posibilidad de extender el enfoque a otras tablas de conectividad (representaciones moleculares), aprovechando su amplio uso en sistemas de cómputo. En el capítulo III se ilustra su aplicación en la correlación de algunas propiedades físicas, para cierto número de alcanos y alcoholes alifáticos.

En resumen, la idea fué (ver Introducción General) el generar o re-enfocar metodologías para el diseño de moléculas (de cualquier tipo), que permitan el desarrollo de productos con mayor eficiencia y sean eficaces en sí mismas para tal fin.

Los nuevos enfoques aquí generados y contrastados como el método Monte Carlo y las Tablas de Conectividad resultaron muy prometedores. La aplicación de los diseños factoriales, sin embargo, involucra ciertos problemas aún por resolver (selección de escalas y adecuación de las características estructurales a los niveles alto y bajo exigidos por el diseño), aunque las ventajas potenciales que representa justifican continuar su investigación al respecto.

MAESTRIA EN QUIMICA ORGANICA

TESIS:

"NUEVOS ENFOQUES EN ESTUDIOS DE RELACION ESTRUCTURA-RESPUESTA PARA EL DISEÑO DE MOLECULAS (PRODUCTOS)"

RECONOCIMIENTO

Los estudios de Maestría del Autor fueron financiados en su mayor parte por el Centro de Investigacion en Quimica Aplicada (CIQA).

La realizacion del presente trabajo fue financiada principalmente por el Consejo Nacional de Ciencia y Tecnologia (CONACyT; Proyecto D111-904038,1990) y llevado a cabo en las instalaciones del CIQA.

Desde el punto de vista tecnico se conto con la atinada asesoria del Dr. Gabriel Gojon Zorrilla, de la Fac. de Ciencias Quimicas, Division de Estudios Superiores, UANL.

Mi reconocimiento y eterno agradecimiento a dichas Instituciones y al Dr. Gojon Zorrilla por tan apreciable contribucion.

JACINTO GUADALUPE RODRIGUEZ GOMEZ

AUTOR

Julio 1991.

a mi hijo:

CRISTIAN

INDICE

INTRODUCCION GENERAL

MARCO DE REFERENCIA.

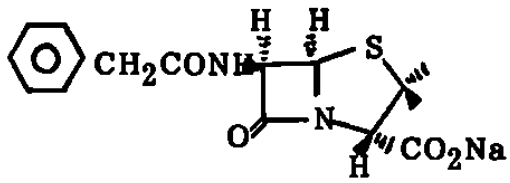
Raison d'Etre.

Con el inicio de apertura de la economía del país, a partir de 1985 se iniciaron profundos cambios y nuevos retos (consecuencias) a enfrentar por la industria nacional. La aparición de nuevos competidores extranjeros; la necesidad de llenar los requerimientos internacionales para exportar y la pronta posibilidad de patentar productos en México, fuerzan a todos aquellos dedicados a la C y T a enfocarse hacia la generación de conocimiento y formación de recursos humanos en aspectos relacionados con las estrategias y métodos para realizar I y D en tecnología de productos (ya no únicamente de proceso-operación), de tal manera que también se tenga la capacidad de aplicar, mejorar e innovar en esta área[1].

A su vez, la tecnología de producto tiene sus propias implicaciones. Aunque generalmente se argumenta que el objetivo en el desarrollo de productos es incrementar las ventas o el margen de utilidad de la empresa, desde otro punto de vista, se debe a la necesidad de hacer frente a la competencia, a las regulaciones y a la satisfacción de nuevas demandas-oportunidades detectadas en el mercado. Y este desarrollo no tiene que ver únicamente con el producto, sino también con aspectos del desarrollo en tecnologías de proceso, operación y de equipo. El porciento de contribución de cada una de estas tecnologías depende del tipo de producto (p. ej. fármaco, plastificante, colorante)[2].

También a su vez, el desarrollo de un producto consistente en una formulación (mezcla de diferentes moléculas), tiene algunas implicaciones diferentes al desarrollo de un producto consistente en una sola molécula con ciertas propiedades.

Supóngase el caso de un fármaco. Hasta 1974 se presumía que 1/15 000 moléculas estudiadas llegan al mercado con un costo entre 11-40 millones de dólares americanos y en un lapso de 8 a 10 años[3]. Aquí se estan considerando todos los aspectos involucrados desde su diseño hasta la comercialización. Ello implica las innumerables evaluaciones técnico-económicas donde muchos proyectos claudican, aunado a los aspectos de comercialización, inversión y las capacidades técnicas y económicas de la empresa; el entendimiento entre los recursos humanos involucrados, la estrategia planteada para conformar el paquete tecnológico, etc[4]. Dentro de este intrincado proceso global, esta otro más particular y relativamente sutil: el encontrar la molécula (producto) adecuada. Considérese la estructura de la penicilina G sódica. Si se desea proporcionarle ciertas propiedades mediante la incorporación de algunos sustituyentes en 3 de las 5 posibles regiones del anillo bencénico, correspondería sintetizar (y evaluar) 36 537 análogos, si se plantearan 20 posibles sustituyentes[3a]: un mundo de posibilidades!



Penicilina G sódica

Cómo evitar la obtención de tantos análogos y aumentar la probabilidad de encontrar el mejor de ellos? Cómo disminuir la mortandad de productos hasta la comercialización? En forma global Cómo aumentar la eficiencia del proceso de desarrollo de un producto y la eficacia del mismo, sobre todo la de aquel consistente en el diseño de una sola molécula?

El Paradigma de Hansch.

Enfocándose particularmente al caso del diseño de moléculas (productos), resulta atractivo el poder predecir (teóricamente o en base a ciertas estructuras pre-evaluadas) los parámetros de desempeño o propiedades deseadas en otras moléculas aún no sintetizadas. Esto abre la posibilidad de reducir grandemente el costo y tiempo de desarrollo de un producto, desde la etapa de generación de la estructura global de la molécula (Lead generation) hasta su optimización-"afinación" de sus propiedades (Lead optimization).

Para que lo anterior sea válido, los métodos actuales -y posiblemente también los futuros- tendrán que considerar ciertas suposiciones engloba-das en el "QSAR paradigm" o Paradigma de Hansch (Pomona College; Claremont, California) pionero (1963) en la cuantificación de las relaciones estructura-actividad (QSAR) de moléculas bioactivas. Tales suposiciones son:

1. La actividad biológica [y muy posiblemente cualquier otra propiedad o respuesta] es función de la estructura molecular.

2. La estructura involucra ciertas propiedades globales (p. ej. lipofilicidad) y ciertas locales (p. ej. distribución de la lipofilicidad).

3. Estas propiedades pueden cuantificarse mediante parámetros extratermodinámicos (π, σ, E_s, etc.), o cualquier otro descriptor de la estructura.

4. **Existe un modelo estadístico** que relaciona la respuesta con la estructura. Con los cambios en las propiedades globales y locales, aunque no sea un modelo simple o fácil de descubrir.

Sobre estas suposiciones descansan todos los métodos actuales para la correlación (y predicción), de las propiedades de una molécula, los cuales han aumentado la eficiencia del proceso de desarrollo de nuevos productos (moléculas). Sin embargo, aún no esta dicha la última palabra y en el proceso del descubrimiento y diseño de nuevas estructuras, empleando la intuición-empirismo; modelando su "modo de acción" o usando correlaciones empíricas de la estructura-respuesta, falta incrementar la eficiencia la sistematización y la eficacia del método mismo que se utilice.

En resumen.

Dentro de las implicaciones globales y particulares mencionadas arriba para el desarrollo de tecnologías de producto, resalta como una necesidad (sobre todo en el país), el desarrollar la capacidad para diseñar sustancias con ciertas características predeterminadas. De tal manera que también se pueda seleccionar, aplicar, asimilar, mejorar e inventar nuevas

tecnologías de producto.

Sobre esta percepción somera y parcial de la realidad, descansa la idea de generar y re-enfocar metodologías para el diseño de moléculas (de cualquier tipo), que permitan el desarrollo de productos con mayor eficiencia y sean eficaces en sí mismos para tal fin. Como consecuencias de esta idea, cada capítulo que comprende la presente TESIS, trata una metodología; un nuevo enfoque para el "quehacer diario" en esta área. Una nueva oportunidad para el éxito: para quien las use y para quien las mejore.

REFERENCIAS.

1. a) E. Rubio del Cueto, "El reto de la industria mexicana ante las tendencias del comercio internacional", REVISTA INDUSTRIA, 1989, 1(7), 39. b) "Programa Nacional de Ciencia y Modernización Tecnológica 1990-1994". SPP-CONACyT, Cap. I.

2. "Desarrollo Tecnológico: Una posibilidad al alcance de su Empresa". FONEI.

3. a) G. Redl, et. al., "Quantitative Drug Design" Chem. Soc. Rev., 1974, 273. b) Sobre agroquímicos ver J.J. Menn, "Contemporary Frontiers in Chemical Pesticide Research", J. Agric. Food Chem., 1980, 28, 2.

4. R. Polacek, "New Product Development: From research to commercialization", ACS short course.

5. S.H. Unger, "Consequences of the Hansch paradigm", in "Drug Design", Vol. IX, Cap. 2, Academic Press (1980).

CAPITULO I

"APLICACION DE LOS DISEÑOS FACTORIALES EN EL DISEÑO DE MOLECULAS"

JACINTO G. RODRIGUEZ GOMEZ

Julio 1991

INTRODUCCION

En la mayoría de los estudios de "cuantificación" de la relación estructura-actividad (QSAR) se proponen modelos de la forma:

$$ACTIVIDAD = b_o + \sum_{i}^{t} b_i X_i$$

donde X_i son las características estructurales y los coeficientes b_i sus "contribuciones" a la actividad, obtenidos mediante regresión lineal[1]. Algunas veces se involucra el término de curvatura ($b_{ii}X_i^2$) y rara vez el de interacción ($b_{ij}X_iX_j$), aunque se haya reconocido la necesidad de introducirlo[1,2].

Para llevar a cabo estos estudios se me ocurre que los diseños factoriales[3], comunmente usados en la IyD de procesos químicos, podrían ser útiles en estos casos, puesto que a partir de ellos pueden obtenerse modelos matemáticos que involucren términos de interacción y curvatura, además de la posibilidad del análisis multirrespuesta (otra debilidad encontrada en la estrategia arriba mencionada[2]) y aumentar la capacidad de predicción, dado que puede determinarse la superficie de dicha(s) respuesta(s).

Aunque los diseños factoriales se emplean para estudios de causación, resulta interesante este enfoque a estudios de

correlación. No existe alguna violación al respecto, puesto que realmente el arreglo de los ensayos o experimentos (en este caso moléculas) dadas por el diseño factorial, serían con el fin de determinar la superficie de respuesta de una manera eficiente, independientemente de que los factores o variables (X_i) sean causa de la respuesta o simplemente exista una correlación. Opcionalmente, tal vez puedan ser condiciones necesarias para la causación, pero no una condición suficiente. Lo importante, sobre todo, es la capacidad de predicción que resulte y esto se puede alcanzar teniendo un mejor modelo que describa adecuadamente la superficie de respuesta. Un mejor modelo obtenido de manera eficiente, empleando los diseños factoriales.

Por "mejor modelo obtenido de manera eficiente" debe entenderse aquél obtenido con un mínimo de experimentación. Es decir, con unas cuantas estructuras sintetizadas y evaluadas, obtener una superficie de respuesta que permita predecir la propiedad bajo estudio, para otras estructuras aun no sintetizadas. He aquí el gran potencial de este nuevo enfoque.

Los diseños factoriales normalmente permiten obtener modelos que involucran el término de interacción (referido de aquí en adelante como "interactivo", por simplificación):

$$\text{RESPUESTA} = b_o + b_i X_i \ldots + b_{ij} X_i X_j \qquad \text{("interactivo")}$$

útiles para determinar las "contribuciones" significativas de los rasgos estructurales bajo estudio (*e.g.*, sustituyentes), en un ambiente de multirrespuesta inclusive y análogamente a la metodología de Free-Wilson[4], pero incluyendo la interacción y mediante un simple cálculo, si se desea a mano. Esto resulta sumamente útil en la etapa de optimización (Lead optimization)[1].

En cierto momento, el diseño factorial anterior podría extenderse con algunos cuantos experimentos (estructuras) extra, a uno más particular, *e.g.*, diseño "compuesto centrado", para obtener un modelo tipo:

$$\text{RESPUESTA} = b_o + b_i X_i \ldots + b_{ij} X_i X_j \ldots + b_{ii} X_i^2 \qquad \text{(cuadrático)}$$

mediante el cual puede aproximarse la superficie de respuesta y con ello, la posibilidad de predecir el valor de respuesta para estructuras aun no sintetizadas o evaluadas. Aún más, es posible que con este enfoque, se pueda reducir el número de modelos opcionales a prácticamente UNO (interactivo o cuadrático), restando unicamente el encontrar la correlación con, por ejemplo, algún parámetro extratermodinámico (π, σ, E_s, etc.), tal como típicamente se procede en el área de los estudios de relación estructura-propiedad (SPR o SAR) y en su cuantificación (QSPR o QSAR)[1,5].

APLICACIONES: EL NUEVO ENFOQUE.

A continuación se ilustrará entonces,el potencial de este nuevo enfoque (método) y sus posibles limitaciones hasta ahora visualizados en el presente trabajo, con algunos casos de estudio retrospectivos llevados a cabo por el autor, usando datos de estructura-propiedad descrita en la literatura.

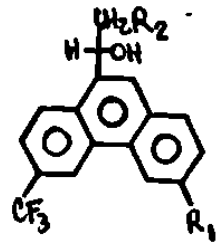
CASO 1

En la tabla I-1 se muestra la actividad antimalaria (expresada como el *log* $1/c$, donde c es la concentración en mol/Kg del animal referente al ED_{50} en ratas) para cuatro fenantrenaminoalquilcarbinoles reportados por Craig y Hansch[5] en un estudio de correlación. Se ordenaron de acuerdo a un diseño factorial 2^2 añadiendo la columna correspondiente al término de interacción. Su análisis, de acuerdo al procedimiento ampliamente descrito en la literatura[3] o por regresión múltiple, arroja el modelo "interactivo" siguiente:

$$log\ 1/c = 3.4825 + 0.0225R_1 + 0.0075R_2 - 0.2225R_1R_2$$

De acuerdo a los valores de los parámetros (b_i's) obtenidos

Tabla I-1. Actividad antimalaria para cuatro fenantrenaminoalquilcarbinoles ordenados de acuerdo a un diseño factorial 2^2



No.[a]	variables en "unidades" originales.		variables en "unidades" codificadas.			RESPUESTA[a]
	R_1	R_2	R_1	R_2	R_1R_2	log 1/c
	H	$NBut_2$	-1	-1	+1	3.23
	I	$NBut_2$	+1	-1	-1	3.72
	H	$NHept_2$	-1	+1	-1	3.69
	I	$NHept_2$	+1	+1	+1	3.29

[a]Orden de compuestos y actividades según la Tabla II en referencia 5. c(mol/kg del-animal) correspondientes al ED_{50} en ratas.

se puede estimar el efecto o contribución a la respuesta (*log* 1/c) por el hecho de variar el sustituyente (R_1) hidrógeno por yodo y por el intercambio del butil por pentil en el grupo amina (R_2):

FACTOR	EFECTO
R_1	+0.045
R_2	+0.015
R_1R_2	-0.445

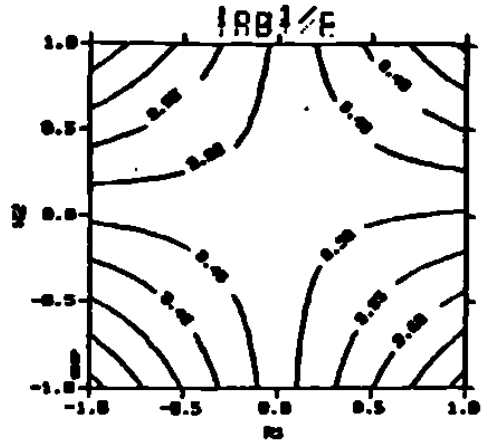
La contribución o efecto del R_1 es de 3 veces mayor que la de R_2 y éste sustituyente no debe mantenerse entonces constante, cuando se intenta optimizar la respuesta, porque el último término indica una gran interacción NEGATIVA entre R_1 y R_2. Dependiendo del valor (tipo de sustituyente) de alguno de ellos, será el efecto del otro y del valor global (más alto o más bajo) de la respuesta.

Hasta aquí se demuestra como introducir el término de interacción en el modelo, de una forma relativamente simple y para este caso, se nota que es la más significativa (ver figura I-1).

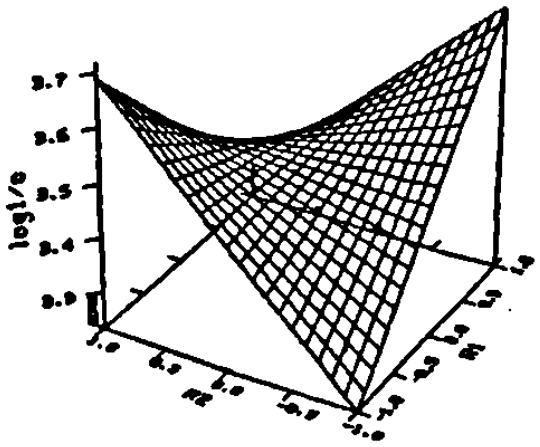
En segunda instancia, nótese que los cambios de valor en los sustituyentes (R's) son, hasta ahora, discontinuos: I por H; Heptil por Butil. Esto es bueno cuando se buscan efectos "gruesos" de los sustituyentes (screening), pero no para

predecir, puesto que se desconoce la superficie real de respuesta. ¿Porqué real? Pues porque a la superficie obtenida (figura I-1) no se le estimó la curvatura[3]. Si ésta no fuera significativa, el modelo "interactivo" permitiría predecir la respuesta para otro sustituyente comprendido DENTRO de la superficie o intervalo de valores de R_1 y R_2. Pero, ¿Cuál intervalo de valores? ¿Cómo estimar la curvatura?

Puesto que los valores de los R's son discontinuos es necesario buscar algún otro parámetro que los transforme en continuos. Supóngase que el efecto de los sustituyentes en la respuesta se debe a sus contribuciones electrónicas. Podría usarse entonces, los valores σ de Hamett como una escala continua. Análogamente, supóngase que se debe a las contribuciones de los sustituyentes en la lipofilia de las moléculas. Podría usarse entonces, los valores π de Hansch. También podrían ser ambos efectos o simplemente otros, solos o mezclados con éstos. He aquí el dilema cuando se trabaja en QSAR. Hay múltiples combinaciones posibles, como también puede haber modelos. Sin embargo, en este enfoque solo habrá uno de dos modelos opcionales: El factorial o el cuadrático. Por consiguiente, el número de modelos a probar se reduce, pero no el número de combinaciones entre los parámetros por estudiar, con el fin de determinar una escala continua que permita introducir la curvatura (si es necesario) y predecir el valor de la respuesta para otros sustituyentes (valores de las R's) aún no sintetizados o evaluados.



(a)



(b)

Figura I-1. Superficie de respuesta (a) en curvas de nivel y (b) en 3-dimensiones para $Y=3.48 + 0.02R_1 + 0.07R_2 - 0.222R_1R_2$.

CASO 2

Existe una importante consideración en este enfoque presumiblemente relacionada con el hecho de que los sustituyentes sean variables discontinuas, en las primeras etapas del cálculo.

Supóngase el caso de las tetraciclinas estudiadas por Free y Wilson[4] (ver tabla I-2). En el análisis de acuerdo a la metodología establecida[3] e igual al Caso 1, se obtuvo el modelo "interactivo" siguiente:

$$\text{BIOACTIVIDAD} = 167.5 + 130R_2 - 22.5R_1R_2$$

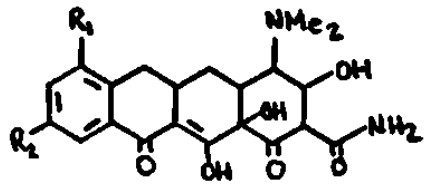
Como puede notarse, la interacción no es tan "significativa". R_2 posee un gran efecto y R_1 NO tiene efecto o contribución alguna en la respuesta. Al intercambiar en el diseño Cl en lugar del Br, para el nivel bajo de R_1 (tabla I-3) se obtiene:

$$\text{BIOACTIVIDAD} = 220.25 - 52.75R_1 + 176.75R_2 - 72.25R_1R_2$$

donde, R_2 sigue teniendo mayor contribución y R_1 ahora sí contribuye en poco menos de 1/3 respecto a R_2.

Como primera regla puede argumentarse que: "Si un factor resulta no-significativo (b_i=0), probar al menos con otro

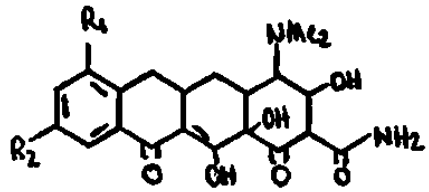
Tabla I-2 . Actividad biológica de algunos derivados de la tetraciclina, ordenados de acuerdo a un diseño factorial 2^2.



Variables en "unidades" originales.		variables en unidades codificadas.			RESPUESTA[a]
R_1	R_2	R_1	R_2	R_1R_2	"Bioactividad"
Br	NO_2	-1	-1	+1	15
NO_2	NO_2	+1	-1	-1	60
Br	NH_2	-1	+1	-1	320
NO_2	NH_2	+1	+1	+1	275

[a] Potencia inhibitoria contra S. Aureus, in vitro.

Tabla I-3. Actividad biológica de algunos derivados de la tetraciclina, ordenados de acuerdo a un diseño factorial 2^2.



variables en "unidades" originales.		variables en unidades codificadas.			RESPUESTA[a]
R_1	R_2	R_1	R_2	R_1R_2	"Bioactividad"
Cl	NO_2	-1	-1	+1	21
NO_2	NO_2	+1	-1	-1	60
Cl	NH_2	-1	+1	-1	525
NO_2	NH_2	+1	+1	+1	275

[a] Potencia inhibitoria contra S. aureus, in vitro.

sustituyente".

CASO 3

El hecho de que las variables puedan considerarse discontinuas o cualitativas al inicio, es lo que permite el "screening" de variables y el incluir no solo sustituyentes, sino algún otro rasgo estructural, tal como centros asimétricos. Por ejemplo, en la tabla I-4 se presenta un diseño factorial 2^3 para algunas estructuras con actividad "*anti-tremor*" (en Inglés) donde se incluyen como factores, sustituyentes y centros asimétricos. Analizando el diseño se obtiene el modelo siguiente:

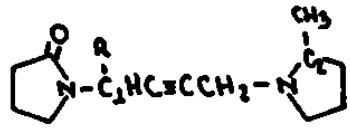
$$\text{ACTIVIDAD} = 147 + 0.25R - 0.65C_1 + 0.22C_2 - 0.27RC_1 + 0.01RC_2 - 0.06C_1C_2 - 0.06RC_1C_2$$

El cual indica que el centro quiral C_1 posee el mayor efecto (negativo). De este análisis puede recomendarse lo siguiente:

1. Dejar C_2 con la configuración S (+1).
2. Variar R y C_1, puesto que incluso su interacción (RC_1) es la siguiente en importancia.
3. Opcionalmente, mantener la configuración de C_1 en R (-1) y experimentar exclusivamente con el sustituyente R.

A tales conclusiones llegó Lehmann[7] en su estudio de correlación, con dichas sustancias.

Tabla I-4. Actividad anti-tremoral de algunas oxotremorin-derivados, ordenados de acuerdo a un diseño factorial 2^3.



FACTOR	VALOR −1	VALOR +1
R	Pr	Me
C_1 [a]	R	S
C_2 [a]	R	S

R	C_1	C_2	RC_1	RC_2	C_1C_2	RC_1C_2	actividad anti-tremoral [b].
-1	-1	-1	+1	+1	+1	-1	1.38
-1	+1	-1	-1	+1	-1	+1	0.62
-1	-1	+1	+1	-1	-1	+1	1.82
-1	+1	+1	-1	-1	+1	-1	1.05
+1	-1	-1	-1	-1	+1	+1	2.30
+1	+1	-1	+1	-1	-1	-1	0.70
+1	-1	+1	-1	+1	-1	-1	3.00
+1	+1	+1	+1	+1	+1	+1	0.82

[a] Se refiere al cambio de configuración R o S. [b] logaritmo de la potencia = $100/ED_{50}$ en μmol/kg.

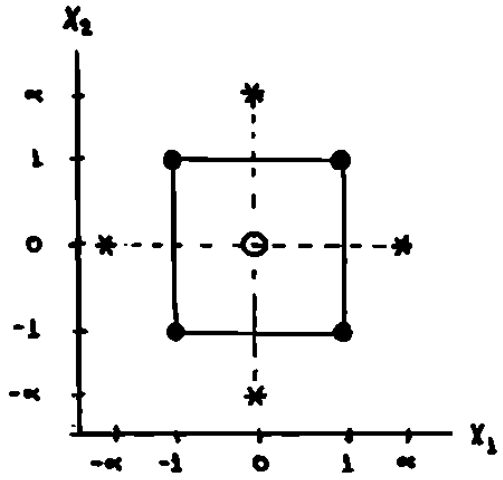
CASO 4

En la investigación y desarrollo de procesos químicos, cuando se emplean diseños factoriales, la estrategia consiste en llevar a cabo algunos experimentos para determinar el modelo "interactivo" y la importancia de la curvatura (mediante puntos centrales, figura I-2). Si la curvatura resulta significativa, se realizan algunos ensayos extra (puntos estrella) para estimar un modelo cuadrático, el cual permite una mejor descripción de la superficie de respuesta.

La Figura I-2 ilustra el diseño, llamado particularmente "compuesto centrado".

Para el enfoque presente, esta estrategia de extensión no es tan simple, por la necesidad de transformar los factores a una escala continua, donde los valores de la escala para cada factor (sustituyente) no necesariamente coinciden con el arreglo del diseño "compuesto centrado" de la Figura I-2.

Para ejemplificar estos argumentos fue difícil encontrar en la literatura los datos adecuados. Sin embargo, parece que empleando algunos derivados del ácido benzóico y sus puntos de fusión (única propiedad disponible) para las estructuras requeridas puede ilustrarse el procedimiento.



	NIVEL	
	X_1	X_2
Puntos Factoriales	-1	-1
	+1	-1
	-1	+1
	+1	+1
Punto Central	0	0
Puntos Estrella	0	$-\alpha$
	0	$+\alpha$
	$-\alpha$	0
	$+\alpha$	0

Figura I-2. **Diseño factorial (2^2) y su extensión a un "compuesto centrado" para dos variables. El nivel recomendado para α (distancia a partir del punto central) puede estar en el intervalo desde 1 ("centrado en las caras") hasta $2^{n/4}$ (n=número de variables). Tomado de referencia 3.**

Supóngase que cualquiera de las combinaciones mostradas en la tabla I-5 pudieron ser escogidas para iniciar el estudio. En este caso, se usurán las tres opciones con el fin de demostrar que la selección de una escala continua es independiente de las combinaciones. Empleando el procedimiento general[3] se obtuvieron los modelos "interactivos" siguientes.

4a $pf = 171.75 - 10.75X_1 + 40.25X_2 - 20.25X_1X_2$

4b $pf = 187.5 - 4.0X_1 + 7.5X_2 - 10X_1X_2$

4c $pf = 171.7 + 10.75X_1 - 40.25X_2 - 20.25X_1X_2$

Ahora supóngase que se está de acuerdo en ellos y se pretende encontrar una escala continua para correlacionar (predecir) el punto de fusión com algún rasgo estructural. La única manera de realizarlo, en este enfoque, es probando con otras estructuras que tengan sus valores de la escala continua dentro del intervalo de los sustituyentes usados para obtener el modelo.

Para encontrar entonces una escala continua se usaron la *refraccion molar* (MR), el parámetro electrónico de Hammett (σ) y el estérico (E_s) de Taft[6]. Los valores para los sustituyentes usados (tabla I-5) y para los que se pretende utilizar como medio de comprobación del modelo y de la susceptibilidad del parámetro para correlacionar la

Tabla I-5. Algunos derivados del ácido benzóico y sus puntos de fusión (pf), ordenados de acuerdo a un diseño factorial 2^2.

CASOS	Variables en "unidades" originales[a]		Variables en unidades codificadas.			RESPUESTA[b]
	Rm	Rp	X_1	X_2	X_1X_2	pf, ° C
4a	H	H	-1	-1	+1	122
	NO_2	H	+1	-1	-1	141
	H	Cl	-1	+1	-1	243
	NO_2	Cl	+1	+1	+1	181
4b	Me	OH	-1	-1	+1	174
	NO_2	OH	+1	-1	-1	186
	Me	Cl	-1	+1	-1	209
	NO_2	Cl	+1	+1	+1	181
4c	NO_2	Cl	-1	-1	+1	181
	H	Cl	+1	-1	-1	243
	NO_2	H	-1	+1	-1	141
	H	H	+1	+1	+1	122

[a] Sustituyentes en posición meta (Rm) y para (Rp). [b] valores tomados de referencia 8.

respuesta, se muestra en la tabla I-6. Dichos valores se codifican usando la fórmula establecida:

$$\text{Valor codificado} = \frac{(\text{valor no-cod.}) - \frac{(\text{nivel alto} + \text{nivel bajo})}{2}}{\frac{(\text{nivel alto} - \text{nivel bajo})}{2}}$$

donde nivel alto y bajo son los valores de la escala para los sustituyentes +1 y -1 en el diseño (tabla I-6). Por ejemplo, para el *m*Cl, en la escala de MR (4a) resulta:

$$\frac{6.09 - \frac{7.96 + 1.09}{2}}{\frac{7.96 - 1.09}{2}} = 0.58 \quad \text{en codificado}$$

De esta manera se introdujeron en cada uno de los modelos (4a, b y c) y se obtuvo la respuesta (pf) predicha para las estructuras con los sustituyentes mostrados en la tabla I-7 en donde se añade además el punto de fusión reportado[8]. La diferencia (desviación o residuo) del punto de fusión reportado con respecto al predicho con los modelos, es una medida de su capacidad de predicción y de la capacidad de correlación de la escala (MR, σ, E_s). Ambas cosas no necesariamente van juntas. Si las desviaciones no son pequeñas relativamente, puede deberse a lo siguiente:

1. El modelo no es adecuado.
2. La escala no es adecuada.
3. Ambas cosas.

Tabla I-6. Valores de algunos sustituyentes "aromáticos" en las escalas de - MR, σ y Es, según referencia 6.

SUSTITUYENTES	VALOR EN LA ESCALA		
	MR	σ	Es
mH	1.03	0.0	1.24
pH	1.03	0.0	1.24
mCl	6.03	0.37	0.27
pCl	6.03	0.23	0.27
pHO	2.85	- 0.37	0.69
mNO_2	7.36	0.71	- 1.28
mMe	5.65	- 0.07	0.0

Analizando los datos de la tabla I-7 resulta difícil seleccionar algunas de las escalas. Considerando que se trata de puntos de fusión, las desviaciones son relativamente grandes. El hecho de que E_s no haya podido calcularse usando el modelo 4b, complica la selección. Hasta aquí y para este caso (uno de los peores, espero) no puede seleccionarse la escala, presumiblemente, debido a que el modelo "interactivo" no es adecuado.

Partiendo de esta suposición se extendió el modelo "interactivo" a uno cuadrático escogiendo los puntos que aproximadamente se ubiquen de acuerdo a un diseño compuesto centrado "en las caras" segun la Figura I-2. Los diseños completos para las tres escalas se muestran en las Tablas I-8, 9 y 10. Para este caso, resulta mejor determinar los modelos en función de las unidades originales (ahora MR, σ o E_s) en lugar de las codificadas. Usando regresión múltiple se obtuvieron los siguiente modelos:

$$4d \quad pf = 68.23 - 8.26(mMR) + 70.28(pMR) + 1.37(mMR)^2 - 6.11(pMR)^2 - 2.12(mMR)(pMR)$$

$n=9$; $R^2=0.96$ ($\alpha=0.026$); $s=8.75$

$$4e \quad pf = 143.08 + 362.41(m\sigma) + 175.2(p\sigma) - 562.26(m\sigma)^2 + 923.49(p\sigma)^2 - 64.17(m\sigma)(p\sigma)$$

$n=9$; $R^2=0.898$ ($\alpha=0.100$); $s=14.02$

Tabla I-7. Puntos de fusión reportados y la desviación de los productos (Modelos - 4a, b y c) usando las escalas de MR, σ y Es, para algunos derivados - del ácido benzóico.

Derivado.	pf. reportado.	MR	Desviación[a] Modelo 4a σ	Es	Modelo 4b Mr	σ	Es	Modelo 4c MR	σ	Es
mCl, pH	156	18.90	24.12	26.68	-	32.41	-	18.99	24.12	26.68
mCl, pHO	170	12.18	-	10.06	- 6.64	6.64	-	12.18	-	10.06
mCl, pCl	208	13.98	2.76	11.13	5.16	14.82	-	13.58	-2.76	-11.13

[a] Desviación = (pf reportado) - (predicho por el modelo); El signo - indica que el dato no pudo obtenerse, por que el valor en la escala (para alguno de los- sustituyentes) cae fuera del intervalo bajo estudio, delimitado por los- grupos correspondientes a los niveles -1, +1.

Tabla I-8. Algunos derivados del Ac. benzoico ordenados de acuerdo a un diseño - "compuesto centrado" y de acuerdo a la escala de la MR para su correlación con el punto de fusión.

ENSAYO	"Unidades originales" mR	pR	mMR	pMR	Unidades en codificado. X_1	X_2	RESPUESTA pf, °C
1	H	H	1.03	1.03	-1	-1	122
2	NO_2	H	7.36	1.03	+1	-1	141
3	H	Cl	1.03	6.03	-1	+1	243
4	NO_2	Cl	7.36	6.03	+1	+1	181
5	Me	OH	5.65	2.85	0.46	-0.27	174
6	Me	H	5.65	1.03	0.46	-1	112
7	Me	Cl	5.65	6.03	0.46	+1	209
8	H	OH	1.03	2.85	-1	-0.27	215
9	NO_2	OH	7.36	2.85	+1	-0.27	186

Tabla I-9. Algunos derivados del Ac. benzoico, ordenados de acuerdo a un diseño "compuesto centrado" y de acuerdo al parámetro $\bar{\sigma}$ de Hammett, — por su correlación con el punto de fusión.

ENSAYO	"Unidades originales" mR	pR	$m\bar{\sigma}$	$p\bar{\sigma}$	Unidades en codificado. X_1	X_2	RESPUESTA pf, °C
1	Me	OH	-0.07	-0.37	-1	-1	174
2	NO_2	OH	0.71	-0.37	+1	-1	186
3	Me	Cl	-0.07	0.23	-1	+1	209
4	NO_2	Cl	0.71	0.23	+1	+1	181
5	H	H	0.0	0.0	-0.82	0.23	122
6	H	OH	0.0	-0.37	-0.82	-1	215
7	H	Cl	0.0	0.23	-0.82	+1	243
8	Me	H	-0.07	0.0	-1	0.23	112
9	NO_2	H	0.71	0.0	+1	0.23	141

Tabla I-10. Algunos derivados del Ac benzoico, ordenados de acuerdo a un diseño- "compuesto centrado" y de acuerdo al parámetro esterico (Es) de Taft, para su correlación con el punto de fusión.

ENSAYO	mR	"Unidades originales" pR	mEs	pEs	unidades en codificado. X_1	X_2	RESPUESTA pf, °C
1	NO_2	Cl	-1.28	0.27	-1	-1	181
2	H	Cl	1.24	0.27	+1	-1	243
3	NO_2	H	-1.28	1.24	-1	+1	141
4	H	H	1.24	1.24	+1	+1	122
5	Me	OH	0.0	0.69	0.02	-0.13	174
6	Me	Cl	0.0	0.27	0.02	-1	209
7	Me	H	0.0	1.24	0.02	+1	112
8	NO_2	OH	-1.28	0.69	-1	-0.13	186
9	H	OH	1.24	0.69	+1	-0.13	215

4f $pf = 198.43 + 34.34(mE_s) + 27.93(pE_s) + 10.41(mE_s)^2 - 77.5(pE_s)^2 - 33.27(mE_s)(pE_s)$

$n=9$; $R^2=0.988$ ($\alpha=0.0045$); $s=4.86$.

De los cuales el 4f parece mas adecuado según el coeficiente de correlación (R^2) y la desviación estándar (s) de las desviaciones de los residuos. Por consiguiente, el parámetro estérico E_s parece explicar (correlacionar) mejor el punto de fusión. Algunas veces resulta adecuado transformar la escala de la respuesta o de algún factor. Por ejemplo, transformando el punto de fusión (respuesta) a su *log* o 1/pf se obtuvieron los modelos siguientes:

4g $log\ pf = 2.27 + 0.08(mE_s) + 0.14(pE_s) + 0.03(mE_s)^2 - 0.25(pE_s)^2 - 0.08(mE_s)(pE_s)$

$n=9$; $R^2=0.983$ ($\alpha=0.007$); $s=6.44$

4h $1/pf = 5.73 - 0.93(mE_s) - 3.03(pE_s) - 0.41(mE_s)^2 + 4.25(pE_s)^2 + 1.04(mE_s)(pE_s)$

$n=9$; $R^2=0.973$ ($\alpha=0.014$); $s=10.64$.

Para los cuales el modelo 4f resulta también superior, aunque las desviaciones predichas para otros derivados no incluídos en el diseño y mostrados en la tabla I-11 parecen indicar lo contrario. Ahora, comparando los primeros tres derivados de

dicha tabla con la tabla I-7, parece que el modelo 4f y por consiguiente la extensión del modelo "interactivo" al cuadrático resultó contraproducente: el "interactivo" parece mejor que el cuadrático.

Introduciendo los derivados de la tabla I-11 en el diseño, como experimento extra, se obtiene para el "interactivo":

4i $pf = 231.94 + 31.47(mE_s) - 74.91(pE_s) - 30.04(mE_s)(pE_s)$

$n=9$; $R^2=0.86$ ($\alpha=0.00005$); $s=13.77$

y para el cuadrático:

4j $pf = 219.01 + 32.04(mE_s) - 40.16(pE_s) + 4.92(mE_s)^2 - 22.69(pE_s)^2 - 30.04(mE_s)(pE_s)$

$n=15$; $R^2=0.87$ ($\alpha=0.0007$); $s=12.97$.

de donde se concluye que, prácticamente, ambos modelos resultan igualmente útiles y el parámetro E_s correlaciona mejor la respuesta. Sin embargo, las desviaciones aún son demasiado grandes, relativamente, para la predicción de una propiedad como el punto de fusión. Deberá continuarse la búsqueda de algún otro rasgo estructural de la molécula (*e.g.* relacionada con las interacciones intermoleculares) o combinaciones de ellos que pueda correlacionar la respuesta, ajustándose exclusivamente a los modelos "interactivo" o cuadrático, según el presente enfoque.

Tabla I-11. Puntos de fusión reportados y desviación de los predichos por los modelos 4f, g y h, usando la escala de Es y transformando la respuesta, pf, en log pf y 1/pf.

Derivado	p.f.reportado °C	Desviación[a] modelo 4f	modelo 4g.	modelo 4h.
mCl, pH	156	43.223	40.787	39.399
mCl, pHO	170	-14.629	-11.802	- 9.208
mCl, pCl	208	0.078	4.108	8.888
mBr, pH	155	41.600	39.175	37.819
mBr, pHO	177	- 4.770	- 2.349	- 0.091
mBr, pCl	215	12.588	16.774	21.669

[a]Desviación=(pf reportado)-(pf predicho por el modelo); logicamente las desviaciones para los modelos 4g y 4h, corresponden a los transformados al pf-normal, para comparación.

CONCLUSIONES

De acuerdo al procedimiento tradicional en estudios de relación estructura-respuesta (*e.g.* biológica, física, química) se intenta, por regresión múltiple encontrar un modelo y algún rasgo estructural (cuantificable) que mejor correlacione la respuesta. En el presente enfoque la búsqueda del modelo se redujo exclusivamente a dos: el "interactivo" o el cuadrático, bajo la suposición de que son suficientes como para hacer una estimación aproximada de la superficie de respuesta y se presume que ésta es una gran ventaja, no una limitación. De acuerdo a los análisis retrospectivos o casos presentados anteriormente, puede notarse que tal presunción parece positiva.

La segunda ventaja resultante del enfoque es la menor cantidad de ensayos (estructuras) necesarias para estimar la superficie de respuesta, o sea, una gran eficiencia aumentada aún más por la factibilidad del análisis multirrespuesta, al igual que en el estudio de procesos químicos empleando estos diseños. Ello es mas importante en la etapa de "screening", usando los sustituyentes u otra característica estructural, como variables discontinuas o cualitativas (*e.g.* Caso 3). La tercera ventaja radica en la introducción relativamente fácil del término de interacción.

Lo anterior permite visualizar la utilidad potencial del enfoque demostrada en algunos de los casos discutidos. Sin embargo, podría pensarse que dicha utilidad no quedo tajantemente demostrada, según los resultados obtenidos para el Caso 4. Pero debe notarse que dicho caso no está concluído y al parecer el obtener una buena conclusión depende del encontrar los rasgos estructurales o sus combinaciones que mejor se ajusten al modelo, más que al modelo en sí. Por consiguiente, este problema (típico del área) se dejó parcialmente resuelto, bajo la consideración de que principalmente se está ilustrando la metodología, más que resolver un problema concreto por ahora. Pero nótese las implicaciones reales del enfoque. Se estan ajustando "los datos" al modelo (!) y esto que desde otros puntos de vista es prohibido, en esta área resulta común y más complejo: ajustar datos y modelos al mismo tiempo. En el presente enfoque se reduce la tarea al ajuste de los datos prácticamente, además de ser menor la cantidad necesaria de ellos.

Hasta ahora, el enfoque presente puede bosquejarse a manera de procedimiento como sigue:

1. Seleccionar algunos parámetros o descriptores (e.g. MR o E_s) de la molécula, los cuales posiblemente se correlacionan con la respuesta y permitan la transformación de los

sustituyentes (por ejemplo) a una escala continua.

2. Seleccionar los sustituyentes que más se ajusten a todas las escalas escogidas, de acuerdo a un diseño "compuesto centrado", por si resulta necesario extender el modelo "interactivo" a uno cuadrático. Además, tratar de que otros sustituyentes caigan dentro de la región experimental, con el fin de predecir posteriormente la respuesta, para otras sustancias análogas sin necesidad de sintetizarlas.

3. Seleccionar las estructuras (moléculas) que se ajusten a un diseño factorial y llevarlo a cabo, así como probar otras estructuras para evaluar la validéz del modelo obtenido y de la escala seleccionada. Estas otras estructuras pueden ser algunas de las necesarias para extender el modelo "interactivo" al cuadrático. De esa manera se optimiza el uso de los datos de las estructuras.

4. Determinar si las variables seleccionadas son relevantes, como para seguirlas considerando en el estudio. Recuerde que si algún parámetro del modelo (b_i) resulta igual a cero, deberá probarse (al menos) con otro sustituyente.

5. Juzgar la validéz del modelo "interactivo" y su posibilidad de extensión a uno cuadrático. Si resulta necesario, llevarlo a cabo. El algoritmo general para deducir el diseño "compuesto centrado en las caras" (ver Figura I-2),

para p variables o factores es como sigue:

Puntos factoriales = 2^p

Puntos estrella (α =1) = 2^p

Punto central = el centro de todos los factores

Las combinaciones para los factoriales pueden obtenerse por lógica o en referencia 3. Los puntos estrella también por lógica, según la Figura I-2 y para tres factores en referencia 3.

6. Seleccionar el mejor modelo y la escala que mejor correlacione a la respuesta.

7. Si los resultados no son satisfactorios, iniciar el procedimiento con otras escalas (parámetros o descriptores) o combinaciones de ellas (relaciones, sumas, diferencias, etc), incluyendo la posibilidad de su transformación (*e.g. log y* o 1/y).

El último paso 7 resume, prácticamente, el típico problema de encontrar el rasgo estructural cuantificable para correlacionar la respuesta. La condición aparecer de este enfoque, es el ajustar las estructuras o más bién, los valores de las características estructurales (por ejemplo, sustituyentes) a un diseño factorial o compuesto centrado (paso 1). Si hablamos de sustituyentes ellos deben tener de preferencia y para los que conforman los puntos factoriales

(niveles +1 y -1) los extremos de dichas escalas. De esta manera, una gran cantidad de sustituyentes caen en el intervalo de la escala o región experimental y tendrán la posibilidad de conformar los puntos extra para obtener el modelo cuadrático o para evaluar el modelo y las escalas finalmente, para predecir la respuesta de otras estructuras sin necesidad de sintetizarlas o evaluarlas.

Para el Caso 4 por ejemplo, los sustituyentes en posición meta quedaron distribuidos según la Figura 1-3, para las escalas MR, σ y Es. Nótese como "los valores originales": HO, NH_2, Cl y CN, en la escala de MR, pueden conformar otras estructuras y los sustituyentes OCH_3, Br, Et y I no pueden utilizarse, por caer fuera de la región experimental. Aunado a esto, queda el problema de si las estructuras o moléculas requeridas para el estudio y exigidas por los diseños estén disponibles o puedan sintetizarse y evaluarse.

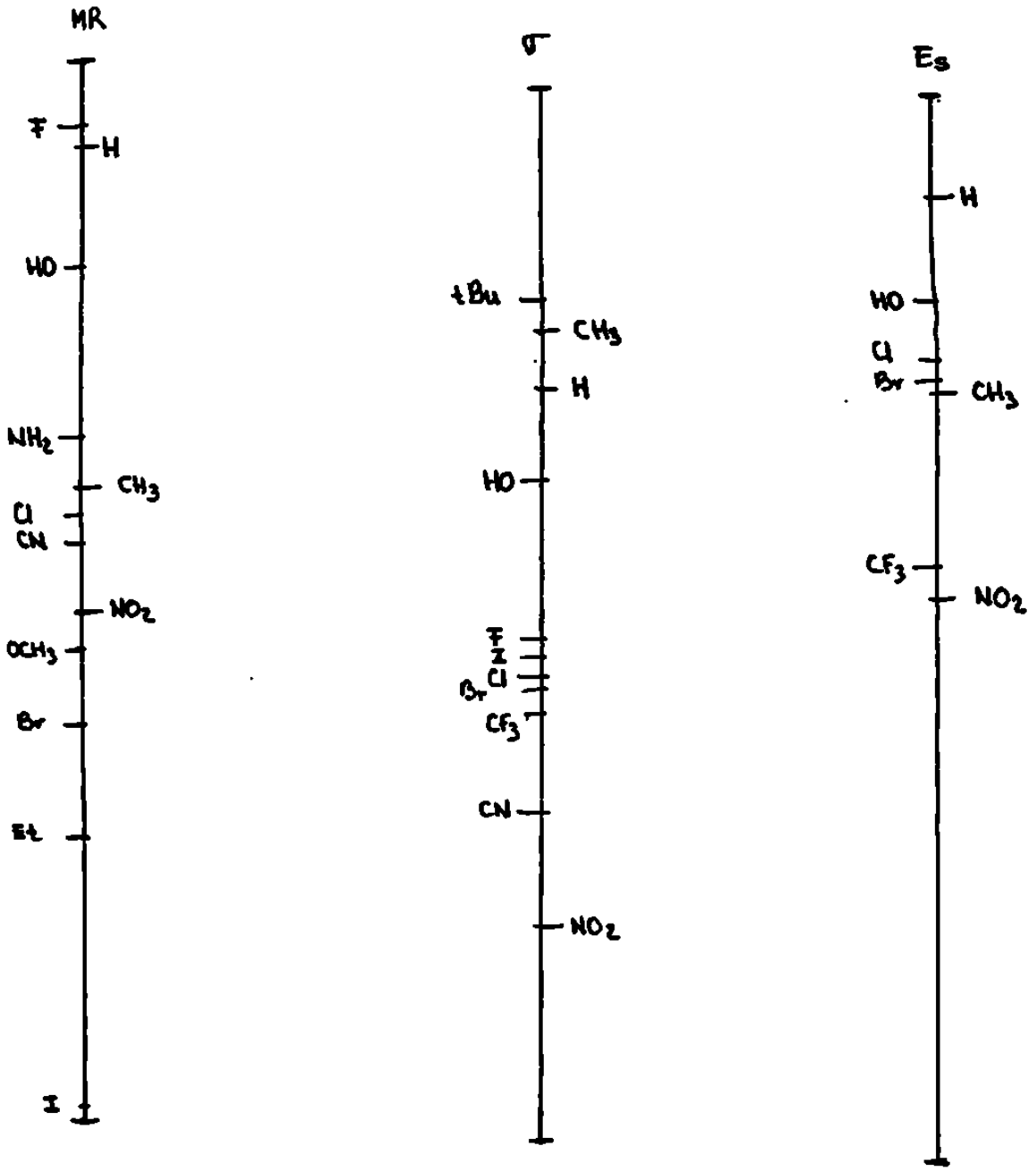


Figura I-3. Algunos sustituyentes "meta-aromáticos" distribuidos según su valor en las escalas MR, σ y Es, utilizados en el Caso 4 (ver Figura I-8,9 y 10).[6]

REFERENCIAS

1. Redl, G. "Quantitative Drug Design". *Chem. Soc. Rev.* 1974, *273*
2. Weiner, M.L.; Weiner, P.H. "A study of structure-activity relationships of a serie of diphenylaminopropanols by Factor Analysis" *J. Med. Chem.* 1973, *16*, 655.
3. Murphy, T.D. "Design and Analysis of Industrial experiments". *Chem. Eng.* 1977, (*Jun 6*), 168.
4. Free, S.M.; Wilson, J.W. "A mathematical contribution to structure-activity studies". *J. Med. Chem.* 1964, 7, 395
5. Craig, P.N.; Hansch, C.H. "Structure-Activity correlations of antimalarials compounds. 2. Phenantreneaminoalkylcarbinol antimalarials". *J. Med. Chem.* 1980, *16*, 661.
6. a) Hansch, C.H.; *et. al.* "Aromatic susbtituent constants for structure-activity correlations" *J. Med. Chem.* 1973, *16*, 661. b) Craig, P.N. "Interdependence between physical parameters and selection of substituent groups for correlations studies". *J. Med. Chem.* 1977, *14*, 680.
7. Lehmann, P.A. "Quantifying stereoselectivity or How to choose a pair of shoes when you have two left". *Trends in Pharmacological Sciences.* 1982, *3*, 103.
8. Weast, R.C.(Ed.) "Handbook of Chemistry and Physics". 55th Ed. CRC Press, USA(1974), pags.C174-194.

CAPITULO II

"APLICACION DEL METODO MONTE CARLO EN EL DISEÑO DE MOLECULAS"

JACINTO G RODRIGUEZ GOMEZ

Julio 1991

INTRODUCCION

Durante la IyD de nuevas moléculas con cierta propiedad (respuesta) física, biológica, etcétera, es común usar la intuición (experiencia) para determinar (sesgar) las estructuras a evaluar primeramente. Los aspectos relacionados con la dificultad técnica para su síntesis y los económicos, juegan un papel importante en esta selección. Sin embargo, en este intento por eficientar o acelerar el desarrollo, puede perderse la oportunudad de hacer grandes descubrimientos.

Por otra parte, supóngase que se desea examinar alguna propiedad para cierta familia de estructuras análogas, típicamente variando algunos sustituyentes. A medida que aumenta el número de sustituyentes y posiciones deseables, aumenta también la cantidad de moléculas por sintetizar y por evaluar (recuérdese el caso de la penicilina G).

En ambas situaciones que pueden ser una misma, puede opcionalmente "antojarse" dejarlo a la casualidad y llevar a cabo experimentos aleatorios, o sea, obtener y evaluar algunas cuantas moléculas seleccionadas al azar, del total de posibilidades.

Charles Hendrix[1] enfatizó su recomendación de usar experimentos aleatorios (Método Monte Carlo) en la IyD de procesos químicos. Hendrix argumenta que la posibilidad de

obtener la mejor respuesta está gobernada por la ecuación siguiente:

Supóngase F = fracción del ESPACIO (área, volumen) de gran INTERES (e.g. rendimiento >60%).

Entonces, 1-F = fracción del espacio INDESEADO.

$(1-F)^N$ = probabilidad de que todos los N ensayos, CAIGAN en dicha región INDESEADA y,

$1-(1-F)^N$ = probabilidad de que AL MENOS UNO de los ensayos CAIGA en el ESPACIO de gran INTERES.

Por ejemplo, si la región de interés es un 10% del total (F=0.1) se tiene que:

N	$(1-F)^N$	$1-(1-F)^N$
10	0.35	0.65
15	0.21	0.80
20	0.12	0.88

Por lo tanto, parece poco atractivo correr más de 20 ensayos o experimentos, para este caso y para cualquier otro de ¡10-20 variables! presume Hendrix, puesto que la probabilidad

de que al menos uno de ellos caiga en la región de interés es razonablemente alta (88%).

¿Podría aplicarse este razonamiento al desarrollo de productos (moléculas), sobre todo cuando se trata de situaciones como las mencionadas al inicio? Veámoslo a continuación.

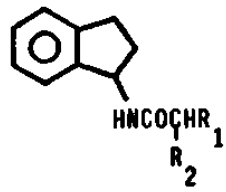
APLICACIONES: EL NUEVO ENFOQUE.

A partir del famoso artículo de Free y Wilson[2] se tomaron las actividades (LD_{50}) para algunos analgésicos del tipo indanamina, los cuales se muestran en la tabla II-1.

Supóngase el caso ficticio donde se pretende estudiar la actividad (LD_{50}) para estos análogos cuando justamente R_1= H, Me, Et y R_2= NMe_2, NEt_2 y Morfolina. Añádase también, que se desea sintetizar exclusivamente 5 de las 9 estructuras posibles e interesa una actividad $LD_{50} \geq 5.0$.

De acuerdo a la tabla II-1, tres de las nueve estructuras corresponderían a la región de interés, luego F=3/9=0.33. De acuerdo a Hendrix o más bien, al método Monte Carlo:

Tabla II-1. LD_{50} para algunos analgésicos del tipo indanamina.



No	No[a]	R_1	R_2	LD_{50}
1	19	H	NMe	2.75
2	20	H	NEt_2	1.90
3	21	H		5.00
4	22	CH_3	NMe_2	1.77
5	23	CH_3	NEt_2	1.55
6	24	CH_3		5.20
7	25	C_2H_5	NMe_2	1.75
8	26	C_2H_5	NEt_2	1.58
9	27	C_2H_5		5.00

[a]Orden de compuestos según la Tabla de Free y Wilson[2].

N	$(1-F)^N$	$1-(1-F)^N$
3	0.30	0.70
5	0.13	0.87
9	0.02	0.98

Por consiguiente, hay una probabilidad del 87% de que, al sintetizar únicamente CINCO moléculas al azar, se obtenga AL MENOS UNA con el valor de $LD_{50} \geq 5.0$.

¿Verdadero o falso? Para comprobarlo se empleó el procedimiento sugerido por Hendrix: los valores (sustituyentes) de R_1 y R_2 se anotaron en pequeños papeles (seis) y se pusieron en dos recipientes adecuados, según R_1 y R_2. De esta manera los sustituyentes fueron seleccionados a manera de rifa, hasta obtener las CINCO estructuras deseadas. La figura II-1 muestra los resultados obtenidos, los cuales se explican como sigue:

1. Se determinaron (rifa) 5 bloques de CINCO (N) estructuras cada uno, los cuales representan las combinaciones de R_1 y R_2. Para cada estructura se tomó el valor LD_{50} correspondiente de la tabla II-1 (como si hubieran sido sintetizadas y evaluadas), encontrándose que la frecuencia de haber obtenido un valor $LD_{50} \geq 5.0$ fué de 4:1 (figura

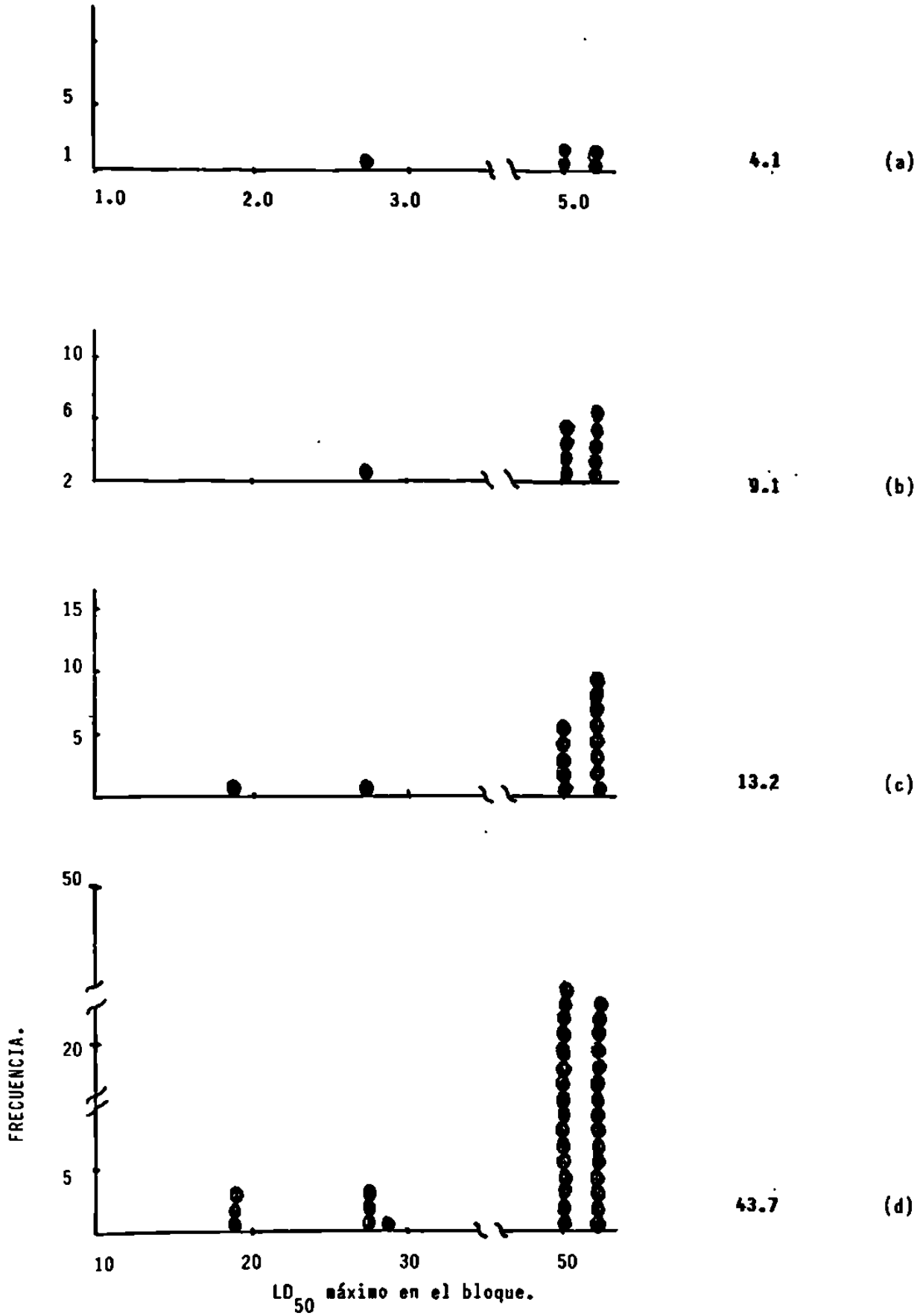


Figura II-1. Frecuencia de los valores máximos obtenidos en cada bloque de cinco ensayos.

II-1,a), tomando como base el máximo valor obtenido de las cinco moléculas y para los cinco bloques. Solo un bloque presentó un compuesto con un LD_{50} máximo de 2.75. La frecuencia 4:1 equivale a 80%, lo cual esta muy cercana al 87% calculado.

2. De la misma forma y considerando 10 bloques, también de CINCO estructuras (combinaciones de R_1 y R_2), se logró una frecuencia del 90%, como predice aproximadamente el cálculo previo (figura II-1,b).

3. Tomando 15 bloques se obtuvieron resultados excelentes de 13:2 equivalentes al 87% predicho (figura II-1,c).

4. Finalmente, para 50 bloques de CINCO estructuras se obtuvo el también excelente resultado de 43:7, o sea, 86% (figura II-1,d).

Por consiguiente, dentro de la incertidumbre existente en este tipo de problemas, la formula $1-(1-F)^N$ predice adecuadamente la probabilidad mencionada, lo cual fundamenta su aplicabilidad en este tipo de estudios. Sin embargo, en la realidad es imposible pre-determinar o establecer inicialmente un valor (nivel) de F. Así que, lo contrario es mas interesante: Predecir el valor de F a partir de unas cuantas moléculas pre-evaluadas. Nótese en el ejemplo anterior que, la capacidad de predicción, depende del número

de experimentos (moléculas) o valor de N. Por consiguiente N debe determinarse según el caso particular bajo estudio.

Pero, ¿Es posible determinar F? Veámoslo usando el mismo ejemplo.

CASO 1

En la tabla II-2 se muestra el primer bloque de CINCO (N) combinaciones obtenidas para el estudio anterior. Tomando sus valores de LD_{50} de la tabla II-1, ordenándolos de menor a mayor y haciendo los cálculos de acuerdo a Hendrix, como se muestra y deduce en la tabla II-3, se concluye lo siguiente:

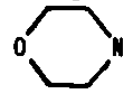
1. Aproximadamente el 50% de las nueve estructuras o moléculas (!) posibles deben tener un $LD_{50} \geq 2.75$. Según la tabla II-1, cuatro poseen tal aseveración (no puede haber 4.5 estructuras).

2. Igualmente, ≈34% deben tener un $LD_{50} \geq 5.0$, o sea, aproximadamente 3, como es el caso.

3. Finalmente, el 17% de las estructuras, deberían tener un $LD_{50} \geq 5.20$, o sea UNA, lo cual es cierto para el caso real.

En este momento puede resumirse lo siguiente:

Tabla II-2. Relación de las 5 estructuras obtenidas al azar (rifa) en el primer bloque.

No.[a]	No.[b]	R_1	R_2	LD_{50}
1	9	Et		5.0
2	7	Et	NMe_2	1.75
3	2	H	NEt_2	1.90
4	6	Me		5.20
5	1	H	NMe_2	2.75

[a] Orden de aparición en la rifa

[b] Orden según la Tabla II-1.

Tabla II-3. Estimación de la región de interés (número de moléculas) empleando cinco estructuras (de las 9 posibles) obtenidas al azar.

No[a]	LD_{50}	ORDEN PRIORIDAD.	RANGO DE[b] PROBABILIDAD
2	1.75	1	16.6
3	1.90	2	33.3
5	2.75	3	50.0
1	5.00	4	66.6
4	5.20	5	83.33

[a] Orden acuerdo a la Tabla II-2. [b] Determinando acuerdo a C. Hendrix[1]: $\frac{\text{orden} \times 100}{N+1}$ = rango de probabilidad.

1. A partir de unas cuantas moléculas obtenidas al azar, del total de posibilidades, se predijo cuantas hay (en dicho total) con cierto valor de respuesta.

2. Si como en este caso, la(s) molécula(s) de interés o con la mejor respuesta predicha ya se obtuvo(ieron) en los primeros experimentos al azar, pues no se sigue buscando. En caso contrario, se puede entonces visualizar la magnitud o esfuerzo de continuar la búsqueda, según la probabilidad de encontrarla(s) de acuerdo al cálculo.

El primer punto quedó demostrado y así mismo se demuestra el objetivo pretendido en este estudio. El segundo punto implica que dependiendo de los valores de la respuesta obtenidos, o más bien, de que tan cerca se encuentran del mejor valor (la molécula con la mejor respuesta) será mayor la exactitud de la predicción. En el caso anterior resulto que la mejor molécula apareció casualmente en el primer bloque de 5, lo cual puede suceder para otros casos. Pero, supóngase que el primer bloque fuese el que comprende la molécula con un valor máximo de LD_{50}=2.75 (figura II-1,a) tal como resultó en el estudio inicial, ¿cómo se ven ahora las predicciones?

Aplicando de nuevo el procedimiento para este bloque (tabla II-4) se deduce claramente que hay un 17% de estructuras, de las 9 posibilidades, que tienen un $LD_{50} \geq 2.75$ o sea UNA, lo cual es fálso! como se vió en el ajemplo anterior y en tabla

TABLA II-4 Estimación de la región de interés (número de moleculas) empleando cinco estructuras (de las 9 posibles) obtenidas al azar. En este caso el bloque con la peor máxima respuesta de los primeros 5 correspondientes a la Figura II-1,a

No[a]	No[b]	R_1	R_2	LD_{50}	ORDEN DE PRIORIDAD	RANGO DE PROBABILIDAD
1	5	Me	NEt_2	1.55	1	16.66
3	7	Et	NMe_2	1.75	2	33.33
5	4	Me	NMe_2	1.77	3	50.00
4	2	H	NEt_2	1.90	4	66.6
2	1	H	NMe_2	2.75	5	83.33

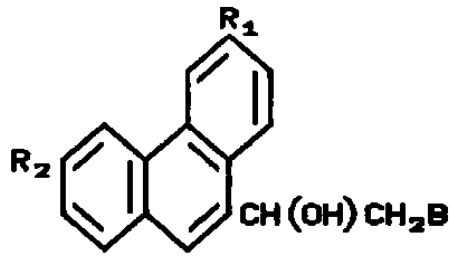
[a] Orden de aparición en la rifa.

[b] Orden según la Tabla II-1

II-1. En este caso, lógicamente ya se obtuvo la mejor de 2.75, por consiguiente, al detener el estudio, se perdió la oportunidad de detectar las que tienen valores de $LD_{50} \geq 5.0$, o sea TRES aún mejores! Pero como se mencionó al inicio "se dejó a la casualidad" y éste es uno de los peores caso; los de mala suerte y es el riesgo que se corre, al adoptarse esta estrategia. Sin embargo, el primer ejemplo mostró que es el menos favorecido (figura II-1). De 4:1... de 20% ... ó 13% de probabilidad, como predice el cálculo y fue demostrado. ¿Vále la pena correr el riesgo? Para este caso de 9 combinaciones, posiblemente no valga la pena, pero ¿qué tal para otros casos con mayor número de posibilidades?

CASO 2

Craig y Hansch[3] reportan un estudio con algunos fenantrenaminoalquilcarbinoles en su búsqueda de alguna correlación con la actividad antimalaria:



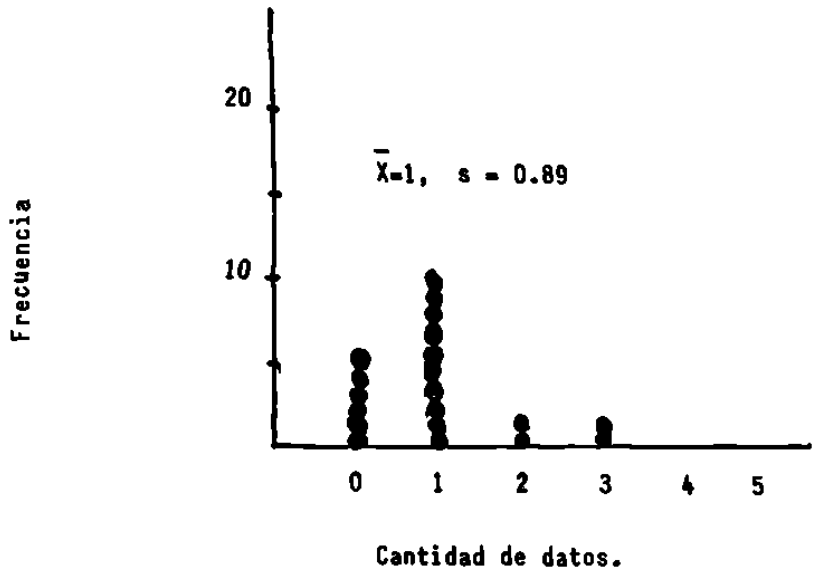
Si R_1= H, F, Cl, Br, I, CF_3; R_2= H, F, Cl, Br, CF_3, OMe, y B= $NBut_2$, $NHept_2$ y 2-piperidina; el total de moléculas por sintetizar y evaluar serían 108, de los cuales se reportan

exclusivamente 29 (tabla II-5). Dado este número limitado de ensayos, no será posible obtener todas las respuestas para las N moléculas (combinaciones) correspondientes al bloque seleccionado. Sin embargo, puede servir como ilustración, a falta de un ejemplo más adecuado.

En analogía al CASO 1, se corrieron inicialmente 20 bloques de CINCO combinaciones (N). Dado el número limitado de combinaciones (moléculas) reportadas, para 6 de los 20 bloques, no hubo respuesta correspondiente para las cinco moléculas implicadas en cada uno de ellos. En promedio hubo un dato reportado por bloque y una respuesta promedio de *log r/c*=3.31 (figura II-2).

Extendiendo el estudio a 20 bloques de 10 moléculas (N) cada uno, se obtuvo un promedio de 2.2 datos reportados por bloque y una respuesta de *log r/c*=3.60, no muy alejada de la anterior, pero respuestas cercanas a 4.0 hubo con mas frecuencia (figura II-3). Extendiéndose mejor a 20 bloques de 15 (N) combinaciones cada uno, se obtuvo un promedio de 4 datos reportados por bloque, ningúno tuvo cero reportados y un promedio en la respuesta de *log r/c*=3.72, lo cual nuevamente no está muy alejada del anterior, pero puede observarse (figura II-4) que hay una mayor frecuencia de los datos alrededor de 4.0. Tomándose pues un tamaño de 15 ensayos por bloque, se seleccionaron (de los 20 bloques) los valores (en el orden obtenidos al azar) no repetidos de cada

a)



b)

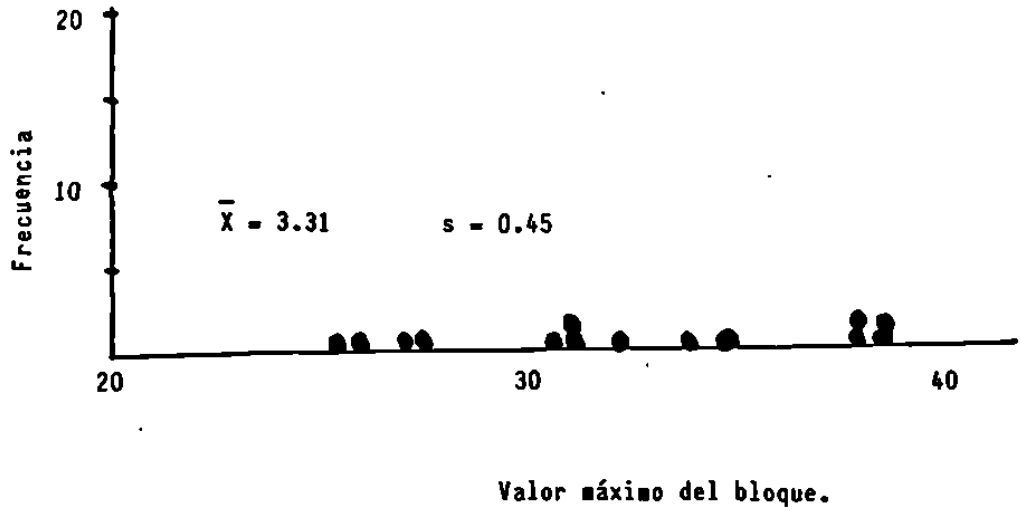
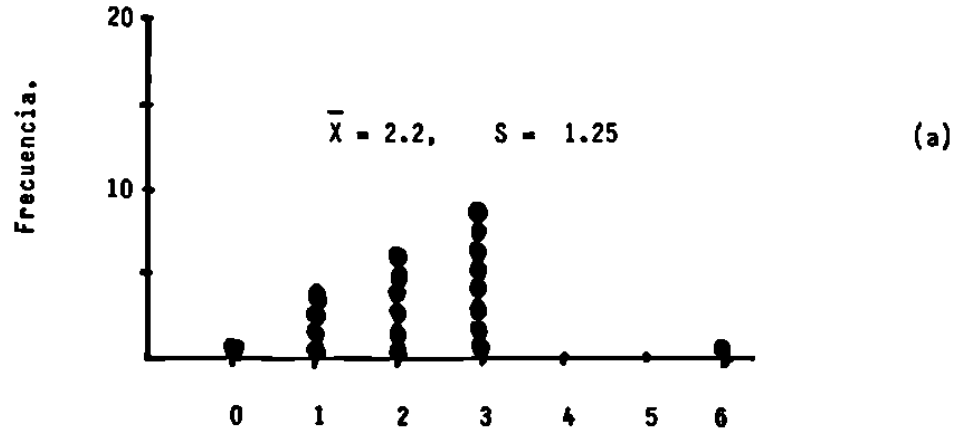


Figura II-2. Frecuencias a) del número de datos reportados para cada bloque y b) de la máxima respuesta obtenida en el bloque; para el caso de 20 bloques con 5 ensayos.



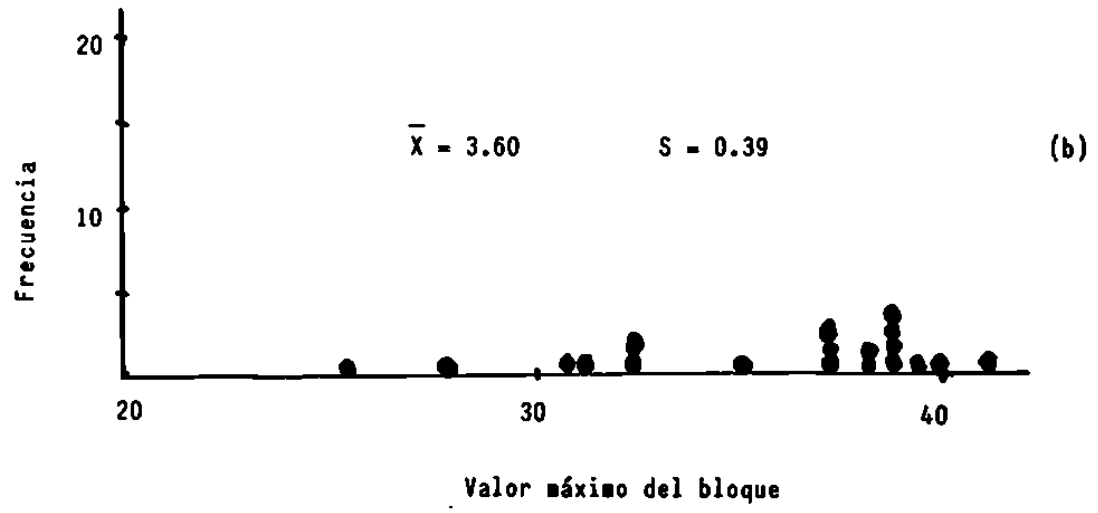


Figura II-3. Idém para 20 bloques con 10 ensayos.

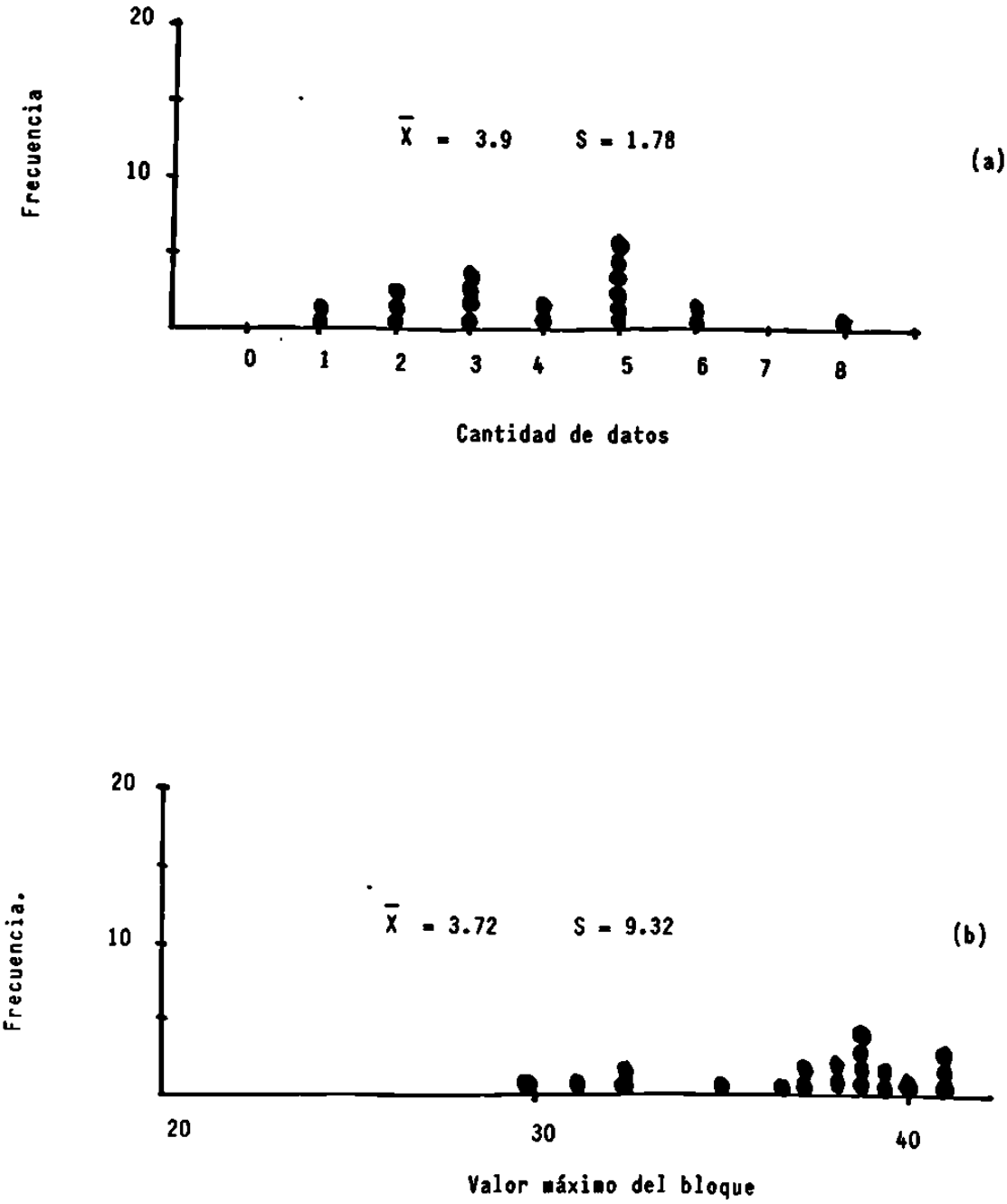


Figura II-4. Idém para 20 bloques con 15 ensayos.

Tabla II-5. Algunos fenantrenaminoalquilcarbinoles con actividad antima - laria (c=moles/kg., correspondientes al ED_{50} en ratas).

R_1						R_2						B			log 1/c
H	F	Cl	Br	I	CF_3	H	F	Cl	Br	CF_3	OMe	$NBut_2$	$NHept_2$	2-pip	
					1					1		1			4.10
			1							1				1	3.98
					1					1				1	3.93
					1			1				1			3.88
		1								1		1			3.87
		1								1				1	3.81
				1						1		1			3.72
		1								1			1		3.72
1										1			1		3.69
1									1				1		3.55
		1							1					1	3.50
					1			1						1	3.40
1										1					3.37
1										1		1			3.23
1									1					1	3.23
			1		1								1		3.12
	1							1					1		3.12
	1							1					1		3.08
			1							1			1		2.98
	1							1					1		2.92
					1					1			1		2.88
	1											1			2.78
			1							1		1			2.74
		1						1					1		2.59
	1												1		2.59
		1							1				1		
1			1		1										2.58
	1										1	1			2.54
1			1					1				1			2.49
	1				1							1			2.43

[a] Tomados de la Tabla II en referencia 3.

bloque hasta completar uno de 15 (tabla II-6). Aplicando el razonamiento usado por Hendrix como en el caso anterior, se encontró (tabla II-6) que 6.75 moléculas de las 108 totales deben tener un valor *log r/c* ≥ 4.1, de los cuales una combinación ya fue encontrada en este bloque. Ahora, toca decidir el continuar buscando (sintetizando y evaluando) las 5.75 ó las 5 ó 6 moléculas deseadas en las 93 restantes. El "grado de alejamiento" de la realidad o qué tan errónea es la predicción, se encuentra en que tan cerca del mejor valor (*log r/c*) se encuentra el 4.1 máximo obtenido, como se dedujo en el Caso 1 anterior.

Aplicando ahora la formula $1-(1-F)^N$, supóngase que son 6 moléculas las que tienen un *log r/c* ≥ 4.1. Esto quiere decir que hubo una probabilidad de $1-(1-6/108)^{15}= 0.57$ de que al menos una molécula cayera en la región de gran interés (*log r/c* ≥ 4.1).

Desde otro punto de vista e igualmente válido, supóngase que las 29 moléculas de la tabla II-5 fueron seleccionadas al azar (rifa) de las 108 posibilidades, o sea, se prefirió un bloque de 29 en lugar de uno de 15. Aplicando el análisis anterior, resulta que (29)(100)/30=96.67, *i.e.*, 3.34 moléculas con un valor de *log r/c* ≥ 4.1 de las cuales UNA ya fué obtenida. Toca decidir si continuar buscando las otras 2 en las 79 restantes.

Tabla II-6. Estimación de la región de interés (número de moléculas) empleando - 15 estructuras (de las 108 posibles) obtenidas al azar.

R	R	B	log 1/c[a]	prioridad	rango[b] de probabilidad.
F	H	$NBut_2$	2.43	1	6.25
F	H	$NHept_2$	2.58	2	12.50
H	Cl	$NHept_2$	2.59	3	18.75
H	Cl	$NBut_2$	2.74	4	25.00
CF	Cl	$NHept_2$	2.88	5	31.25
Cl	OMe	$NHept_2$	2.98	6	37.50
H	F	$NHept_2$	3.08	7	43.75
H	CF_3	$NBut_2$	3.23	8	50.00
Cl	Cl	2 pip	3.50	9	56.25
H	CF_3	$NHept_2$	3.69	10	62.50
I	CF_3	$NBut_2$	3.72	11	68.75
Cl	CF_3	$NBut_2$	3.87	12	75.00
CF	Cl	$NBut_2$	3.88	13	81.25
CF	CF_3	2 pip	3.93	14	87.5
CF	CF_3	$NBut_2$	4.10	15	93.75

[a] actividad de acuerdo a la Tabla II-5. [b] Determinado acuerdo casos anteriores - y Hendrix[1].

Aplicando ahora la formula $1-(1-F)^N$, supóngase que son 3 las moléculas que tienen un *log* $I/C \geq 4.1$. Esto quiere decir que hubo una probabilidad de $1-(1-3/108)^{29}=0.56$ de que al menos UNA molécula cayera en la región de gran interés: !Una probabilidad igual que para el bloque de 15! lo cual significa que no era necesaria la selección del bloque de 29.

CONCLUSIONES

En el Caso 2 arriba mencionado, nótese que hay una diferencia de ≈3 moléculas en las predicciones del número total de combinaciones con un valor de $log\ r/c \geq 4.1$, entre los bloques de 15 y 29 opciones. Nuevamente surge el problema mencionado en el Caso 1: La capacidad de predicción de la formula (orden)(100)/N+1 dada por Hendrix[1].

En la tabla II-7 puede observarse que a medida que aumenta N (el tamaño del bloque o el número de combinaciones obtenidas al azar) el "rango de probabilidad" aumenta y por consiguiente el número de moléculas predichas con cierta respuesta disminuye. Luego, entre más alejado este el máximo valor obtenido del bloque respecto a la región de interés, mayor será el error de la predicción y más probable el perder la oportunidad de encontrar otras moléculas deseadas. En contraposición, $1-(1-F)^N$ nos dice que a mayor N, mayor probabilidad de que al menos una molécula (combinación) tenga la mejor respuesta o que caiga en la región de gran interés. Asi que nuevamente, la primera clave está en la selección del valor de N: Un compromiso entre lo deseado y lo viable.

La segunda clave es aceptar el riesgo y dejarlo a la casualidad... a la suerte. Se justifica si un análisis más

Tabla II-7. Cambio del valor del "rango de probabilidad" conforme al aumento de N o tamaño del bloque, según la formúla:

PRIORIDAD X 100/N+1 dada por Hendrix[1].

ORDEN PRIORIDAD.	N=3	N=4	N=5	N=6	N=7	N=8	N=9
1	25.00	20.00	16.67	14.28	12.50	11.12	10.00
2	50.00	40.00	33.34	28.57	25.00	22.23	20.00
3	75.00	60.00	50.00	42.86	37.50	33.34	30.00
4		80.00	66.67	56.14	50.00	44.45	40.00
5			83.34	71.43	62.50	55.56	50.00
6				85.71	75.00	66.67	60.00
7					87.50	77.78	70.00
8						88.89	80.00
9							90.00

"sistemático" como el presente, se utiliza para tomar desiciones de continuar o detener la búsqueda (síntesis y evaluación). Lo importante es que no se aplica la intuición y se evita el sesgar la búsqueda a cierta región experimental con la posibilidad de disminuir la probabilidad de hacer grandes descubrimientos. Sin embargo, es posible que a partir del primer bloque realizado, mediante un análisis por inspección o aplicando la metodología de Free-Wilson[2], se visualicen las restantes combinaciones con mejor respuesta y justifiquen tal vez un segundo bloque ya no tan al azar.

En conclusión, es posible predecir o estimar el número de moléculas que poseen igual o mejor valor de la propiedad bajo estudio, a partir de unas cuantas pre-evaluadas del total de combinaciones posibles. Hay una gran eficiencia en el Método Monte Carlo. La eficacia tiene que ver con N y la suerte del experimentador!

REFERENCIAS

1. Hendrix, Ch. C. "Through the response surface with test tube end pipe wrench". *CHEMTECH*, 1980, (*Aug.*), 488; Biles, W. E. and Swain, J. J. "Optimization and Industrial Experimentation". Jhon Wiley & Sons, New York (1980), pags. 189-192.
2. Free, S. M. and Wilson, J. W. "A mathematical contribution to structure-activity studies". *J. Med. Chem.* 1964, 7, 395.
3. Craig, P. M. and Hansh, C. H. "Structure-Activity Correlations of antimalarial compounds.
2. Phenanthreneaminoalkylcarbinol antimalarials". *J. Med. Chem.* 1980, *16*, 661.

CAPITULO III

"APLICACION DE LAS TABLAS DE CONECTIVIDAD EN ESTUDIOS DE RELACION ESTRUCTURA-PROPIEDAD"

JACINTO G. RODRIGUEZ GOMEZ

Julio 1991

INTRODUCCION.

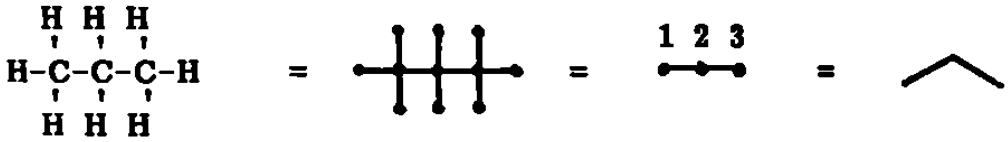
Es fácil suponer que las propiedades físicas y químicas de una sustancia dependen de las características (geométricas y electrónicas) de su estructura molecular; pero resulta difícil suponer que existan formas relativamente simples para describir dichas características y además que se puedan utilizar en estudios de relación estructura propiedad (QSPR).

Tales formas relativamente simples pueden ser algunas propiedades físicas (p. ej. peso molecular), algunos descriptores topológicos (índice de Wiener, Randic, etcétera) o los *ad hoc* (como el número de carbonos y metilos terminales), para el caso de -por ejemplo- compuesto alifáticos. Esta simpleza, comparada con el uso de la química cuántica, mecánica molecular o de los parámetros extratermodinámicos (π, σ, E_s, etc.), fomenta la búsqueda de sus aplicaciones y del desarrollo de otros descriptores nuevos[1] que sean simples, unívocos y que consideren aspectos estereoquímicos. Además, abren la posibilidad de incorporar a la enseñanza (dada su simpleza) los temas de estudio de relación estructura-respuesta (donde, por respuesta, se entiende cualquier propiedad física, química, biológica, etc.), a nivel de licenciatura.

Con el advenimiento de los ordenadores (o computadoras) y su aplicación en química se demandaron maneras de procesar (representar unívocamente) las estructuras químicas en estos sistemas que a la vez, sean inteligibles por el químico. Las tablas (matrices) de conectividad fueron desarrolladas para ese fin. Representan generalmente el cómo

los átomos estan interconectados en la molécula y son, por consiguiente, equivalentes a la estructura molecular. Las analogías con la teoría de grafos (graph theory) de dichas tablas de conectividad facilita el conceptualizar y mejorar el procesamiento de la estructura molecular, como un dato, lo cual ha permitido el uso de los ordenadores en el diseño de rutas sintéticas y simulación de interacciones químicas. También han ayudado al desarrollo de descriptores topológicos, los cuales sirven (por ejemplo) para los estudios de relación estructura-respuesta.

De acuerdo a la teoría de grafos en su aplicación de la química[1], la estructura molecular puede finalmente representarse tal como la usamos en síntesis orgánica. Por ejemplo para el propano:



La matriz de adyacencia (adjacency matriz) representaría cuales átomos (excluyendo los hidrógenos) están interconectados, mediante un 1. Para el propano:

$$A = \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix}$$

La matriz de distancias (distance matrix) representa las distancias o número de enlaces (path's) entre dos átomos, tomando uno como referencia. Para el caso del propano y tomando el átomo 1 como referencia:

$$D = \begin{vmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{vmatrix}$$

Puesto que ambas matrices representan el cómo estan interconectados los átomos en la estructura del propano, se considerarán aquí como matrices o tablas de conectividad, en general.

Un índice topológico es una cantidad numérica derivada matemáticamente de forma directa e inambigua del grafo estructural de una molécula o de su tabla o matriz de conectividad. Al parecer existen 39 índices topológicos reportados en la literatura[1]. Su importancia radica en que reflejan la forma y el tamaño molecular, por lo que pueden usarse para correlacionar propiedades que principalmente estén en función de esas características estructurales.

El número de Wiener[1,2].

El primer uso de los índices topológicos en química se debe a los trabajos pioneros de H. Wiener (1947) en su investigación de la relación entre la estructura y las propiedades de los hidrocarburos saturados.

Wiener partió de la suposición de que las propiedades de estos compuestos (p. ej. el punto de ebullición) varian de acuerdo a su tamaño y ramificación e introdujo un índice basado en la suma de todas las distancias (path's) más cortas en la molécula. Por ejemplo, el índice de Wiener (w) para el propano, puede calcularse formando una tabla de las distancias más cortas (vía enlaces) entre todos los carbonos y sumando los elementos:

	1	2	3
1	0		
2	1	0	
3	2	1	0

$$w = 3 + 1 + 0 = 4$$

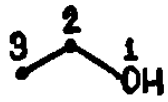
De igual manera se puede obtener w, tomando para cada enlace C-C, el producto del número de átomos de carbono en cada uno de sus lados y sumar este valor para todos los enlaces de la molécula. Para el propano:

$$w = (1 \times 2) + (2 \times 1) = 4$$

$$\quad a \qquad\quad b$$

Al parecer, entre más bajo es el valor de w, más "compacta" es la molécula. Para extender el índice de Wiener a heteroátomos (como N,O,S, etc.), se definió el "atomic site index" (S_i), el cual representa la suma de todas las distancias más cortas (en términos de enlaces) desde el heteroátomo i hasta los demás átomos de carbono. Por ejemplo, para el

etanol:



	1	2	3
1	0		
2	1	0	
3	2	1	0

$S_i = 3$

$w = 3 + 1 + 0 = 4$

Observando las matrices obtenidas según la teoría de grafos, el índice de Wiener es igual a la mitad de la suma de todos los elementos d_{ij} de la matriz de distancias, D:

$$w = \tfrac{1}{2} \sum_i \sum_j d_{ij}$$

y,

$$S_i = \sum_j d_{ij}$$

para el heteroátomo i.

Ahora bien, observando las "tablas de conectividad" del propano y etanol se nota que son exactamente iguales, incluyendo el número de Wiener. Esto equivale a la aseveración de que los índices topológicos tienen más que ver con la forma y tamaño de la molécula, que con su naturaleza y por consiguiente la idea de llegar a "índices de conectividad" unívocos para cada estructura (sustancia) donde se incluya su naturaleza, resulta sumamente atractivo.

APLICACIONES: EL NUEVO ENFOQUE.

En base a la similitud de las tablas de conectividad del propano y etanol mostrados anteriormente, puede generalizarse dicha similitud de la siguiente manera:

Para el caso de cualquier par posicionalmente isomérico, el número de Wiener es idéntico, cuando (según la figura III-1a y b) G es un heteroátomo o cualquier grupo representado como átomo 1. Cuando G=CH_3 se elimina la isomería y la tabla corresponde a la de alguno de los isomeros (Fig. III-1c), luego w no es unívoco. Tampoco lo es S_1 como puede compararse inclusive con la estructura mostrada en la figura III-1d, para una misma familia.

Sin embargo, analizando la tabla empleada en cada uno de los isómeros posicionales (figuras III-1a y b), puede notarse que los números de la fila externa horizontal (FEH) cuya suma dan los valores de w son diferentes. Por otra parte, dentro de la tabla, los números que la comprenden son exactamente los mismos, pero en la última fila horizontal (Fl H) el ORDEN es diferente.

Ahora, viendo la matriz en sentido vertical, se notan más discrepancias pero, considerando la columna uno (CI) correspondiente al átomo 1 (G, en este caso), la diferencia se encuentra principalmente en el último par de números (PN's).

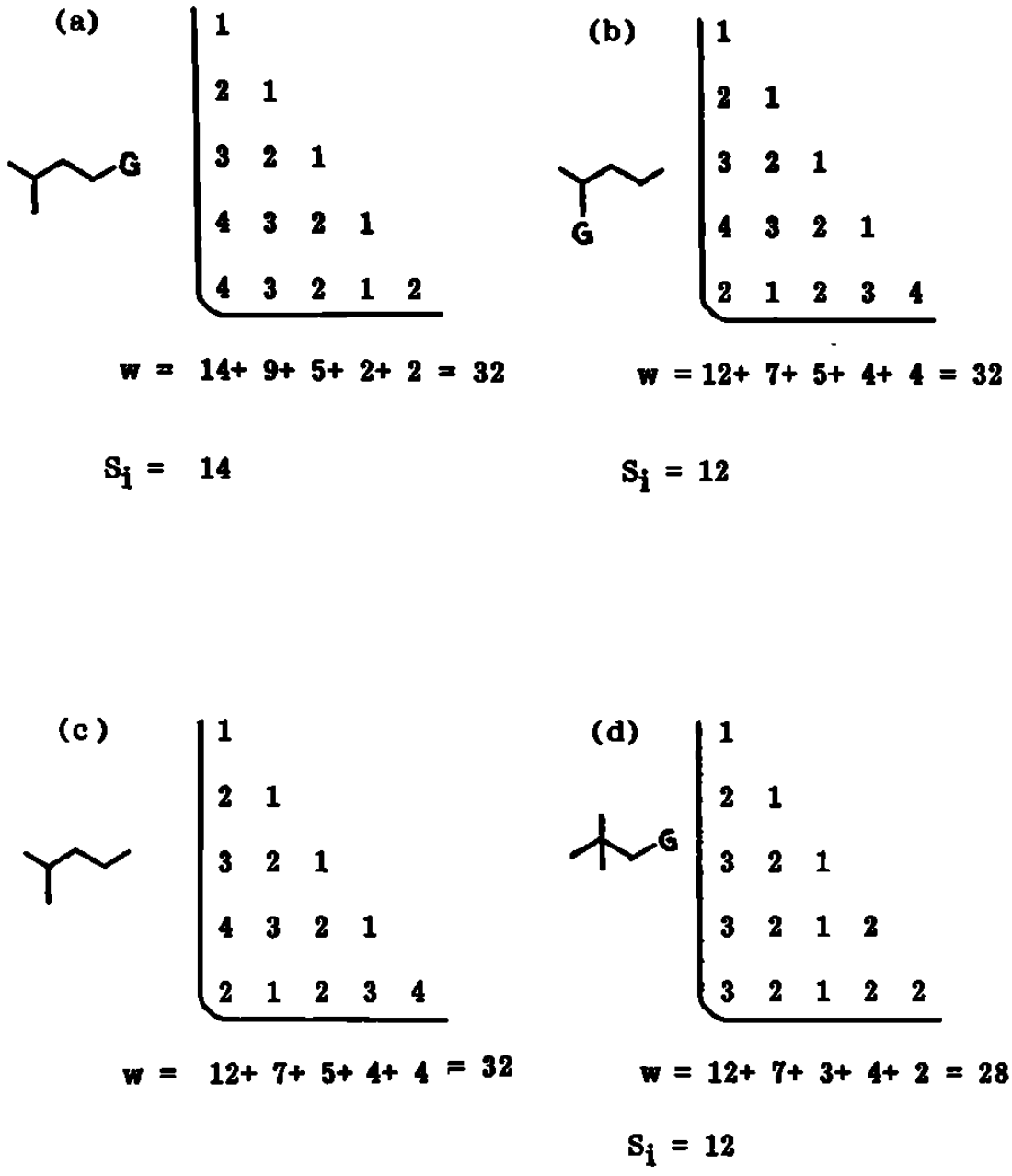


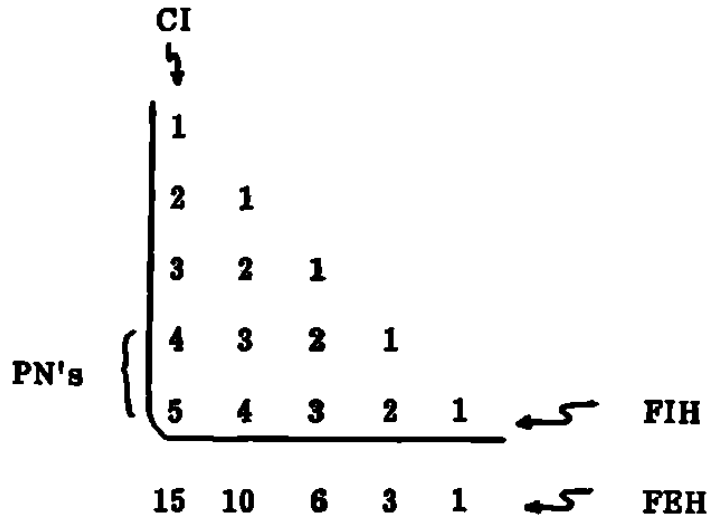
Figura III-1. Comparación de w, S_i y tablas de conectividad según Wiener, para algunas estructuras alifáticas.

En resumen y a manera de HIPOTESIS: ¿Será posible usar las FEH, FIH, CI o los PN's como descriptores moleculares unívocos? ¿Es posible que, por ejemplo, usando los FEH se obtenga una ecuación (correlación) que al sustituir $X_1 = 14$, $X_2 = 9$, $X_3 = 5$, $X_4 = 2$ se obtenga Y = Teb = 131.2°C para el 3-metilbutanol (G = OH)? ¿Idém para las FIH, CI y PN's?.

Si tal cosa es posible, considérese la siguiente enorme ventaja (ver figura III-1 y III-2): para una misma forma o esqueleto estructural, la matriz (y por ende FEH, FIH, CI y PN's) es idéntica, independientemente del grupo funcional (G), lo cual facilita el cálculo. ¿Dónde esta lo unívoco? Primeramente, los descriptores arriba mencionados permitirán diferenciar entre isómeros posicionales dentro de una familia (alcanos, alcoholes, etc.). La diferenciación entre familias se encontraría en los coeficientes (b_i) obtenidos al relacionar la respuesta (propiedad) con dichos descriptores, mediante un modelo lineal:

$$\text{PROPIEDAD} = y = b_o + b_i X_i \ldots + b_{ii} X_i^2 \ldots + b_{ij} X_i X_j$$

de tal manera que, cada familia, tendría un modelo lineal diferente. De esta forma, descriptores y modelos brindan la posibilidad de hacer estudios de relación estructura-propiedad en un sentido unívoco para cada molécula.



G	Teb,°C
OH	137.80
CH_3	68.74
hal.	.
SH	.
etc.	.

Figura III-2. Posibles "descriptores topológicos" obtenidos de las tablas de conectividad, según Wiener, y su diferenciación en el modelo (b_i's) según la propiedad y familia bajo estudio.

Caso 1.

En la figura 3 se presentan las tablas de conectividad, usadas por Wiener, para estructuras alifáticas desde C_1 hasta C_5 con un grupo funcional, G. A partir de ellas pueden obtenerse los "descriptores topológicos" aquí planteados y los cuales son generales, independientemente del significado de G.

En la tabla III-1 se muestra concretamente estos descriptores, los cuales representan los valores que deben tener las X_i para encontrar un modelo lineal (correlación) mediante regresión múltiple. Como primer ejemplo, supóngase que se intenta correlacionar los puntos de ebullición (Teb) de los alcoholes alifáticos (C_5 - C_1) con dichos descriptores.

Empleando los PN's puede observarse en la tabla III-1 que únicamente permitirán diferenciar estructuras con igual número de carbonos. Para el caso de C_5 (n = 8) se obtuvo como mejor correlación:

$$Teb = 87.41 - 9.74\ X_1 + 14.75\ X_2 + 3.82\ X_i^2 - 2.29\ X_2^2$$

con un coeficiente de determinación múltiple $R^2 = 0.9656$ y una medida de la significancia de la regresión $F_{4,3} = 21.06$ (98.44%). La desviación de los residuos de los datos (Teb) predichos por este modelo con respecto a los reportados fue s = 2.1991 (ver tabla III-2). Esto último, como una medida de su capacidad de predicción.

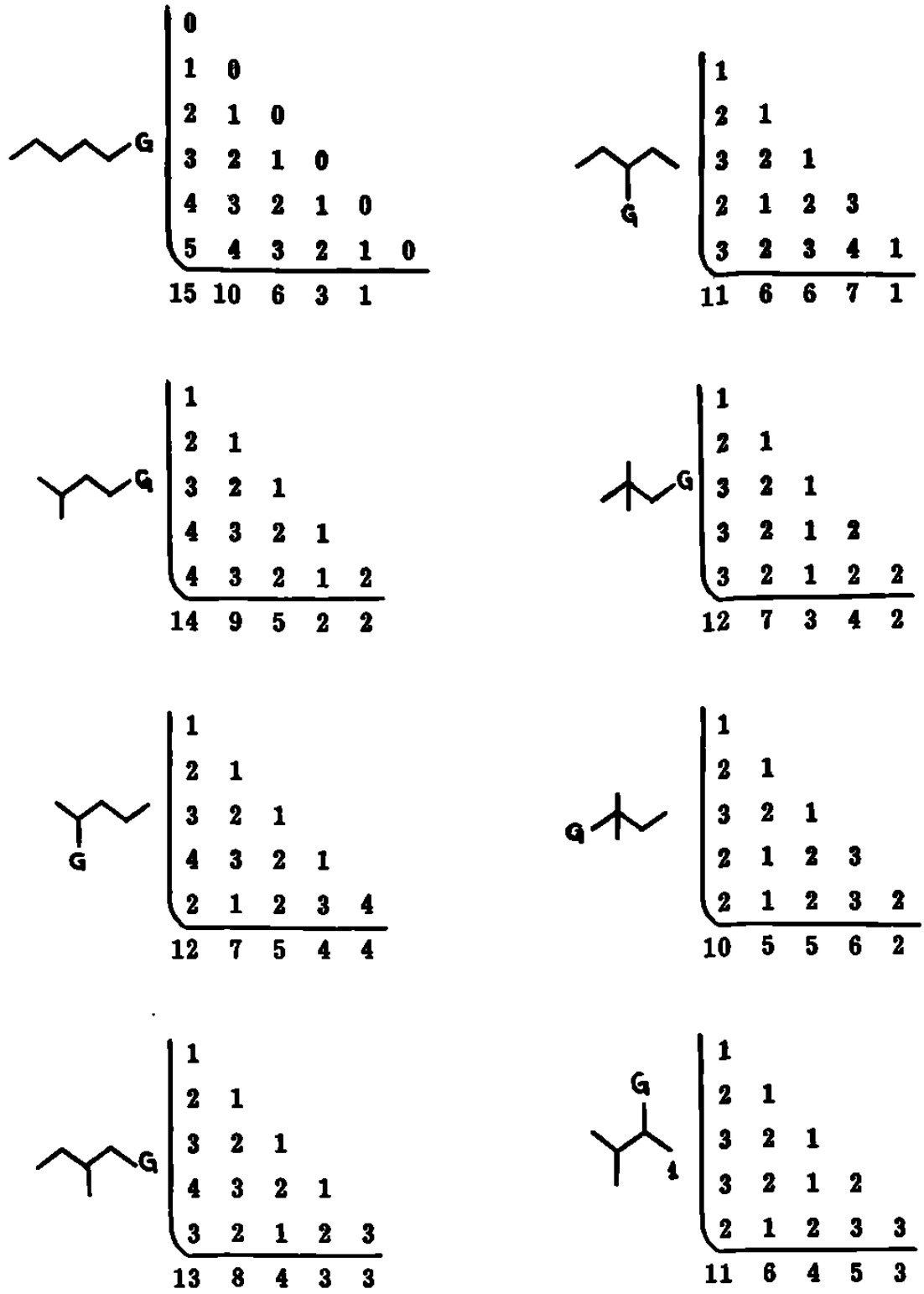
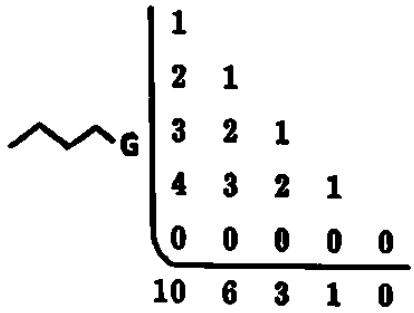
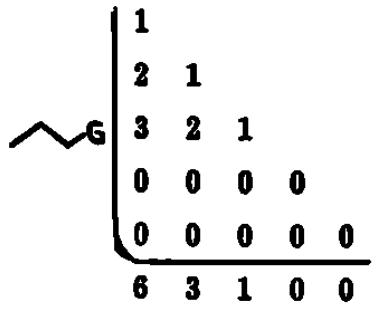
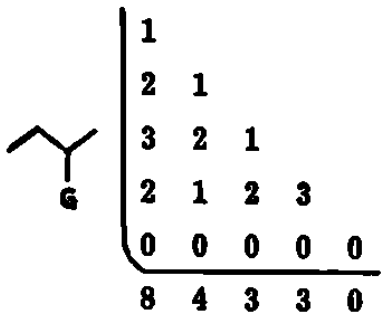
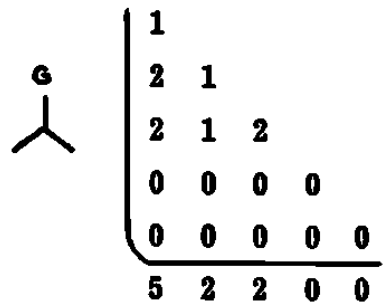


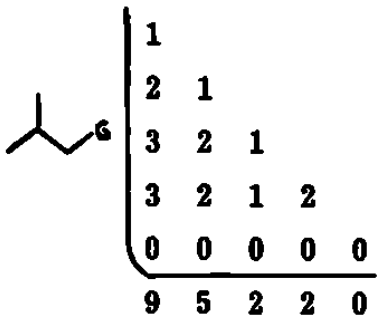
Figura III-3. FEH, FIH, CI y PN's para estructuras alifáticas, desde C_5 hasta C_1, con un grupo funcional G.

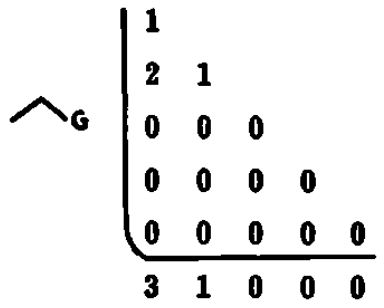


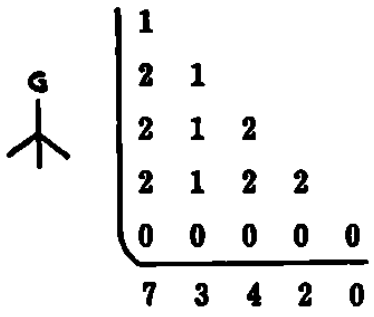












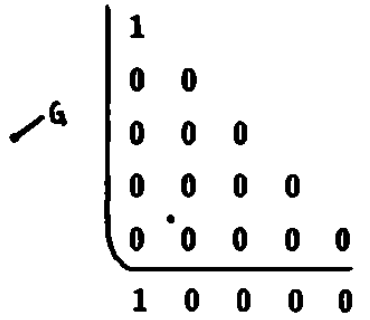


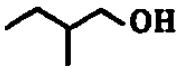
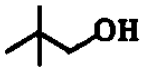
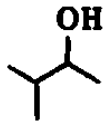
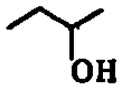
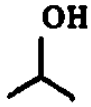
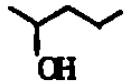
Figura III-3. ... Continuación

Tabla III-1. Valores de las X_i a introducir en el modelo, para correlacionar los "descriptores topológicos" planteados con, por ejemplo, las Teb de alcoholes alifáticos hasta C_5.

No. de carbones	PN's		CI					FIH					FEH					Teb[a] (°C)
	X_1	X_2	X_1	X_2	X_3	X_4	X_5	X_1	X_2	X_3	X_4	X_5	X_1	X_2	X_3	X_4	X_5	
5	4	5	1	2	3	4	5	5	4	3	2	1	15	10	6	3	3	137.8
5	4	4	1	2	3	4	4	4	3	2	1	2	14	9	5	2	2	131.2
5	4	2	1	2	3	4	2	2	1	2	3	4	12	7	5	4	4	119.0
5	4	3	1	2	3	4	3	3	2	1	2	3	13	8	4	3	3	128.7
5	2	3	1	2	3	2	3	3	2	3	4	1	11	6	6	7	1	115.3
5	3	3	1	2	3	3	3	3	2	1	2	2	12	7	3	4	2	113.1
5	2	2	1	2	3	2	2	2	1	2	3	2	10	5	5	6	2	102.0
5	3	2	1	2	3	3	2	2	1	2	3	3	11	6	4	5	3	111.5
4	3	4	1	2	3	4	0	4	3	2	1	0	10	6	3	1	0	117.7
4	3	2	1	2	3	2	0	2	1	2	3	0	8	4	3	3	0	99.6
4	3	3	1	2	3	3	0	3	2	1	2	0	9	5	2	2	0	107.9
4	2	2	1	2	2	2	0	2	1	2	2	0	7	3	4	2	0	82.4
3	2	3	1	2	3	0	0	3	2	1	0	0	6	3	1	0	0	97.2
3	2	2	1	2	2	0	0	2	1	2	0	0	5	2	2	0	0	87.3
2	1	2	1	2	0	0	0	2	1	0	0	0	3	1	0	0	0	78.3
1	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	64.7

[a] tomados de referencia 2.

Tabla III-2. Capacidad de correlación de las Teb para alcoholes alifáticos desde C_1 hasta C_5 empleando PN's, FIH, FEH, y CI.

ALCOHOL[a]	PN's[c]	DESVIACIONES[b]		
		CI	FEH	FIH
	2.31	3.52	3.13	3.86
	-3.13	-2.76	-2.46	-5.18
	3.13	1.39	0.42	0.12
	-	4.32	-0.62	7.89
	-	-5.98	-1.25	-5.62
	-	-5.83	0.30	1.65
	-	-2.28	-3.07	1.06

[a] se presentan los tres más desviados (subrayado) para cada descriptor Los demás son añadidos dada su presencia en la tabla. [b] Desviacón = (Teb reportada)-(Teb predicha). [c] PN's para C_5 exclusivamente.

Para el caso de CI se desechó X_1 puesto que su valor es constante (Tabla III-1) en el intervalo C_1 - C_5. Empleando el resto de los factores se obtuvo:

$$Teb = 64.7 + 4.83\ X_2 + 6.89\ X_3 - 5.79\ X_4 + 6.36\ X_5 + \\ + 2.96\ X_4^2 + 0.95\ \ X_5^2 - 1.87\ X_4X_5$$

con $R^2 = 0.9735$, $F_{7,8} = 42.05$ (99.99%) y los residuos o desviaciones $s = 3.35$. Se introdujo la interacción y curvatura de X_4 y X_5 puesto que estos corresponden a los PN's, dando el mejor modelo presentado. Algunos ejemplos se muestran en la tabla III-2. Respecto a FEH:

$$Teb = 59.42 + 10.54\ X_1 + 5.05\ X_2 + 0.67\ X_3 + 3.45\ X_4 + 2.59\ X_5 + 9.77\ X_1X_2 - 4.67\ X_1^2 - 5.59\ X_2^2$$

con $R^2 = 0.9936$, $F_{8,7} = 136.84$ (99.99%) y las desviaciones con $s = 1.64$. Este no necesariamente es el mejor modelo, puesto que pueden existir todas las interacciones posibles entre los cinco factores. Dejando exclusivamente la parte lineal de los parámetros se obtiene:

$$Teb = 67.67 + 0.74\ X_1 + 7.15\ X_2 - 1.73\ X_3 + 0.45\ X_4 - 0.92\ X_5$$

con $R^2 = 0.9660$, $F_{5,10} = 56.8$ (99.99%) y las desviaciones con $s = 3.79$, lo cual muestra como mejor ecuación la anterior y tal vez puedan obtenerse otras mejores, introduciendo las interacciones y curvatura para los otros factores. En la tabla III-2 se muestra el caso de la ecuación con interacciones y curvatura anteriormente mostrada.

De igual manera puede emplearse FIH. En este caso un modelo obtenido es:

$$Teb = 63.77\ X_1 - 44.31\ X_2 - 0.57\ X_3 + 3.69\ X_4 + 6.56\ X_5 - 0.72\ X_1X_2$$

con $R^2 = 0.9989$, $F_{6,10} = 1657$ (100%) y las desviaciones con $s = 3.51$

Para este caso no resultó adecuado la introducción del término b_o. Excluyendo el término de interacción (X_1X_2) la desviación de los residuos es ligeramente mayor, $s = 3.61$ y $R^2 = 0.9691$. Someramente, parece que la FEH's resultaron más adecuadas para este caso.

CASO 2.

Tal como sucede con cualquier descriptor molecular, algunos permiten correlacionar bien algunas propiedades (respuestas) mientras que otros no resultan tan adecuados.

Tomando ahora el caso de los alcanos alifáticos desde C_2 hasta C_5 e intentando correlacionar sus calores de combustión (ΔH_c°) con los "descriptores topológicos" aquí planteados, se obtiene la relación de las X_i 's y respuestas a considerar en la búsqueda del modelo, según se muestra en la Tabla III-3. Como ya se mencionó, la isomería se reduce, pero se mantienen los mismos descriptores (tabla de conectividad) según la forma molecular.

Nuevamente los PN's sólo permiten diferenciar entre estructuras con el mismo número de carbonos. Para este caso se desecho su estudio por ser entonces, poco interesante.

Tabla III-3. Valores de las X_i a introducir en el modelo, para correlacionar los "descriptores topológicos" planteados con, por ejemplo, los ΔH_c° de alcanos desde C_2 hasta C_6.

No. de carbones	PN's		CI					FIH					FEH					ΔH_c°, 25 °C (Kcal/mol)
	X_1	X_2	X_1	X_2	X_3	X_4	X_5	X_1	X_2	X_3	X_4	X_5	X_1	X_2	X_3	X_4	X_5	
6	4	5	1	2	3	4	5	5	4	3	2	1	15	10	6	3	3	1002.57
6	4	2	1	2	3	4	2	2	1	2	3	4	12	7	5	4	4	1000.87
6	4	3	1	2	3	4	3	3	2	1	2	3	13	8	4	3	3	1001.51
6	2	2	1	2	3	2	2	2	1	2	3	2	10	5	5	6	2	998.17
5	3	2	1	2	3	3	2	2	1	2	3	3	11	6	4	5	3	1000.04
3	3	4	1	2	3	4	0	4	3	2	1	0	10	6	3	1	0	838.80
5	3	2	1	2	3	2	0	2	1	2	3	0	8	4	3	3	0	843.24
5	2	2	1	2	2	2	0	2	1	2	2	0	7	3	4	2	0	840.49
4	2	3	1	2	3	0	0	3	2	1	0	0	6	3	1	0	0	687.98
4	2	2	1	2	2	0	0	2	1	2	0	0	5	2	2	0	0	686.34
3	1	2	1	2	0	0	0	2	1	0	0	0	3	1	0	0	0	530.60
2	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	372.82

[a] tomados del Perry y Chilton, Manual del Ingeniero Químico, 2° Ed., McGraw-Hill, México(1973).

Empleando las columnas internas de las tablas de conectividad correspondientes al átomo UNO (CI's), se obtiene como una de las mejores ecuaciones (también quitando la constante X_1):

$$\Delta H_c^\circ = 372.82 + 88.79\ X_2 + 51.01\ X_3 + 127.74\ X_4 + 82.39\ X_5 - 23.23\ X_4^2 - 12.85\ X_5^2 + 3.33\ X_4X_5$$

con $R^2 = 0.9925$, $F_{7,4} = 75.94$ (99.95%) y las desviaciones con $s = 18.11$. En la Tabla III-4 se presentan las desviaciones (respecto al valor reportado) de los valores producidos por este modelo, para algunas estructuras.

De igual manera, empleando las FIH's se obtuvo (como uno de los mejores modelos):

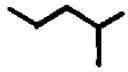
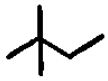
$$\Delta H_c^\circ = 372.8\ X_1 - 136.97\ X_2 + 68.22\ X_3 + 66.68\ X_4 + 42.95\ X_5 - 35.11\ X_1X_2$$

con $R^2 = 0.999$, $F\ 6,6 = 1069.18$ (99.99%). Las desviaciones con $s = 26.86$, lo cual implica menor capacidad de predicción respecto a los CI's como puede notarse en la Tabla III-4.

Para los FEH's se obtuvo, también como uno de los posibles modelos:

$$\Delta H_c^\circ = 218.33 + 155.47\ X_1 - 154.85\ X_2 - 0.3\ X_3 + 0.03\ X_4 + 0.6\ X_5$$

Tabla III-4. Capacidad de correlación de los ΔH_c° para alcanos desde C_2 hasta C_5 empleando CI, FIH y FEH.

ALCANO[a]	DESVIACIONES[b]		
	CI	FIH	FEH
	-22.77	-31.65	1.33
	55.46	32.28	-0.46
	33.90	11.50	0.95
	18.14	-45.82	-0.19
	-12.68	37.30	0.15
	5.96	37.38	-0.52
	2.53	7.74	1.77
	-3.90	-23.24	-4.28
	-15.47	-14.01	1.66

[a,b] Idém consideraciones como en tabla III-2.

con R^2 = 0.999, $F_{5,6}$ = 20024.58 (100%). Las desviaciones con s = 1.62. Sorpresivamente, este simplísmo modelo, produce alta correlación y predicción, por demás excelente, según se observa en la Tabla III-4.

Las desviaciones podrían ser despreciables, comparadas con la mayor cifra para la respuesta, de varios cientos de Kcal/mol.

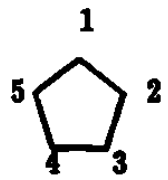
CONCLUSIONES.

En primera instancia, no puede decirse cual "descriptor topológico" aquí planteado resulta más adecuado. Ello dependerá del caso particular que se trate, por consiguiente, parece razonable sugerirlos todos e iniciar la búsqueda del modelo más simple y con mejor correlación y capacidad de predicción. Podría pensarse que las columnas internas (CI's) deberían ser mejores, por corresponder al átomo UNO (G) a partir del cual se está "cuantificando" la forma de la estructura. Los casos anteriores parecen estar más acordes con las filas externas horizontales (FEH's) según se observa más tajantemente para los alcanos (Caso 2) y un poco menos para los alcoholes (Caso 1). Dado que los FEH's representan la suma de todas las distancias más cortas para todos los átomos que comprenden la estructura, podrían entonces ser más representativas de la forma estructural y tamaño de la molécula. ¿Serán pues los FEH's los "descriptores" más útiles para para QSPR? ¿Fue acaso mera coincidencia, para los ejemplos anteriores? ¿Se favoreció para el caso de los alcanos, dada la menor isomería y por ende, menos puntos "discordantes" por correlacionar?

El hecho de una mera coincidencia es el punto en segunda instancia. ¿Cómo se explican estas correlaciones? ¿Tienen algún sentido "físico" los "descriptores" planteados?

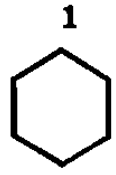
Visualizando estos "descriptores" (p. ej. FEH's) como una medida de las distancias "interatómicas" vía enlaces, representan entonces la forma y tamaño molecular, al igual que lo consideraba Wiener, pero insertados en un modelo matemático como los aquí empleados ¿Cuál es el sentido? ¿Pueden considerarse realmente como variables o factores continuos o discontinuos? Para el caso de los FEH's en los alcanos (Caso 2) ¿Puede suponerse que X_1 puede tomar valores desde 1 hasta 15 como puede observarse en la Tabla III-3? ¿De manera similar X_2, X_3, X_4 y X_5 pueden tomar el intervalo de valores allí mostrados?.

De resultar cierto debería ser posible que al emplear la FEH para el ciclopentano o ciclohexano:



1
2 1
2 2 1
1 2 2 1
0 0 0 0 0

FEH 6 5 3 1 0



1
2 1
3 2 1
2 3 2 1
1 2 3 2 1

FEH 9 8 6 3 1

se obtuvieron los ΔH_c° = 793.39 y 944.79 respectivamente. El modelo obtenido en el Caso 2 para FEH's predice los valores de 376.0358 y 377.6549 Kcal/mol lo cual puede deberse a lo inadecuado del modelo; a la forma en que se estan obteniendo los valores (tablas de conectividad) para el caso de estructuras cíclicas o resulta que estos "descriptores" representan factores discontinuos o categóricos.

Es necesario enfatizar que la idea es obtener modelos simples, o sea, evitar demasiados términos X_i^2 y X_iX_j, a menos que se justifique lo contrario. Sobre esta premisa se trabajó en el presente estudio, con el fin general de ejemplificar el uso de los "descriptores topológicos" planteados y de evitar el encuentro de una correlación causada por el exceso de parámetros en el modelo.

¿Existen otras opciones o formas de usar las diferencias en las tablas de conectividad, para estudios de correlación estructura-propiedad? ¿Podrá hacerse la analogía con otras tablas o matrices de conectividad?

Parece atractivo ahondar más en este asunto, sobre todo porque estas Tablas son la más general representación de la estructura molecular y su manipulación, en nuestra actual era de las computadoras. También abren la posibilidad de usar las tablas de conectividad que representa la molécula en tres dimensiones (descriptores topográficos) en analogía de las de Wiener[3], con la susceptibilidad de introducir rasgos estereoquímicos en un futuro y por ende, en estudios de relación estructura-respuesta, en el significado amplio de la palabra respuesta.

REFERENCIAS.

1.- P.J. Hansen y P.C. Jurs, "Chemical Applications of Graph Theory", J. Chem. Educ., 1988, 65(7), 574-580; Ibid, 65(8), 661-664.

2.- P.G. Seybold, et. al., "Molecular Structure-Property Relationships", J. Chem. Educ., 1987, 64(7), 575-581.

3.- B. Bogdariou, et. al, "On the three-dimensional Wiener number", J. Math. Chem., 1989, 3, 299-309.

GLOSARIO DE TERMINOS

Análisis multirespuesta. Cuando en un, *e.g.*, diseño factorial se miden varias respuestas para cada ensayo o combinación (N), pueden determinarse los efectos de los factores bajo estudio para cada respuesta, lo cual es sumamente eficiente. En este sentido permite llevarse a cabo un análisis multirespuesta.

Característica estructural. Dícese de algún aspecto o rasgo de la estructura química (*e.g.* grupo funcional, No. de carbonos, índice topológico, parámetro extratermodinámico, configuración de algun centro quiral) que permita distinguirla de otras estructuras y que muy probablemente tenga cierta correlación con la "respuesta" bajo estudio.

Contribución (de un factor). Significa la magnitud y signo de los parámetros b_i's del modelo para el factor (característica estructural) bajo estudio. Es análogo al efecto de las variables en estudios de causación y "contribución" es más utilizado en estudios de correlación.

Descriptor topológico. Cantidad numérica derivada matematicamente de forma directa e inambigüa del grafo estructural de una molécula o de su tabla o matriz de conectividad. El grafo es la representación de la estructura con líneas y puntos como el ejemplo del propano en la página

70. su índice topológico (el de Wiener) se dedujo en ambas formas como se muestra en la página 72. Puesto que un grafo es un concepto topológico (y matematico) o sea, no-geométrico, ángulos y configuraciones moleculares no han podido traducirse en algun descriptor topológico.

Diseño "compuesto centrado". Diseño experimental caracterizado por la yuxtaposición de un diseño factorial y un "star design", de tal forma que con un mínimo de experimentación es posible obtener una buena aproximación de la superficie de respuesta usando un modelo completo de segundo orden (*i.e.* cuadrático). La figura I-2 muestra un diseño "compuesto centrado" para dos factores a dos niveles.

Diseño factorial. Diseño experimental caracterizado por la variación simultánea y sistemática de todos los factores o variables (p). El número de ensayos o corridas (N) se determina mediante todas las posibles combinaciones de los valores (alto, bajo y medio; m) de las variables, es decir $N=m^p$. Los puntos factoriales de la figura I-2 muestran un diseño factorial 2^2.

ED_{50}. Dosis específica para obtener cierta respuesta biológica o actividad en el 50% de los "entes", *e.g.* ratas, que comprende la muestra bajo estudio.

Factor continuo o discontinuo. Un factor o variable continuo

(o cuantitativo) es aquél que puede tomar cualquier valor dentro de cierto intervalo (dominio). Por ejemplo, la presión, volumen, peso, tiempo y concentración. Un factor o variable discontinua (discreta o cualitativa) puede tomar un número limitado de valores como: tipo de solvente o de catalizador, número de extracciones, etcétera. En el mismo sentido se habla de respuesta continua o discontinua.

Indice de conectividad. sinónimo de descriptor topológico.

Indice topológico. sinónimo de descriptor topológico.

Interacción (de variables). En un sistema multivariable es posible que el efecto de un factor X_i dependa del nivel o valor de otro factor X_j. Estadísticamente, la estimación de este efecto interactivo se logra introduciendo en el modelo matematico el término de interacción $b_{ij}X_iX_j$ usando un diseño factorial. Esta interacción "estadística" difiere de la llamada interacción química.

Lead generation. Se refiere a la búsqueda (screening) de una estructura que posea cierta respuesta por ahora independientemente de su estructura, aunque se emplean metodologías para encontrar un "común denominador" entre ellas.

Lead optimization. Se refiere a la búsqueda de la estructura

química que proporcionaría la respuesta óptima dentro de una serie de compuestos, los cuales generalmente poseen la misma estructura básica y solo se varían los sustituyenes o la configuración, por ejemplo.

LD_{50}. Dosis letal o concentración necesaria de la sustancia "activa" para matar al 50% de los animales (*e.g.*, ratas) que comprende la muestra bajo estudio.

Matriz de conectividad. Representación matricial de los átomos que constituyen la molécula y su interacción, *e.g.*, interconección *vía* enlaces.

Parámetro estructural. sinónimo de característica estructural.

Parámetro extratermodinámico. A las constantes σ, π, Es, etcétera, se les llama así por relacionarse linealmente con la energía libre (cantidad termodinámica) y dicha relación matematica fué deducida empíricamente sin utilizar las leyes de la termodinámica, *e.g.*, $-\Delta\Delta G=2.3\ RT\sigma_x$.

QSPR o QSAR. ver relación.

Rasgo estructural. sinónimo de característica estructural.

Relación. (a) Estructura-Propiedad. Cuando se correlaciona

una propiedad física como el punto de fusión, con alguna(s) característica(s) estructural (en inglés, SAR); (b) Estructura-Actividad. Para el caso de alguna propiedad biológica (actividad) como la Dosis Letal de un fármaco (en inglés, SAR). Cuando se obtiene alguna relación matematica de la propiedad o actividad, en inglés se conoce como QSPR o QSAR (Q de quantitative) respectivamente.

Respuesta. Es toda aquella propiedad física (punto de fusion), química (reactividad), biológica (dosis letal), etcétera, que se intenta asociar (correlacionar) con alguna(s) característica(s) de la estructura química. El término ha sido tomado del área de diseño de experimentos al de diseño de moleculas, por el autor.

Screening. sinónimo de "lead generation".

Star design. Diseño experimental, en analogía al diseño factorial, caracterizado por un punto central a partir del cual otras combinaciones de los factores (puntos estrellas) se generan al moverse en una distancia positiva y una negativa en cada escala del factor (p) a un tiempo, obteniéndose $2p+1$ combinaciones, las cuales representan el número de experimentos (N) por ejecuter. Este diseño no permite estimar el término de interacción en el modelo. La figura I-2 muestra los puntos estrella y el punto central para un "star design", el cual combinado con un factorial,

produce un "compuesto centrado" para dos factores a dos niveles.

Superficie de respuesta. Es la gráfica que describe el comportamiento de cierta respuesta con respecto a los factores y dentro del intervalo (niveles) bajo estudio. La figura I-1 muestra dos tipos de representacion más comunes.

Variable continua o discontinua. ver Factor...

BIBLIOTECA DIVISION
ESTUDIOS SUPERIORES

IMPRENTA "PEREZ SANTOS"
OBREGON SUR 568
TEL. 4-32-78
SALTILLO, COAHUILA