

**ÍNDICE DEL ESTADO ELECTROTOPOLÓGICO EN LA PREDICCIÓN DEL  
RENDIMIENTO CUÁNTICO DE PRODUCCIÓN DE OXÍGENO SINGULETE  
PARA DERIVADOS DE 1H-FENALEN-1-ONA**

**CÉSAR AUGUSTO GUARÍN DURÁN**

**Trabajo de grado presentado para optar al título de Químico**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE CIENCIAS  
ESCUELA DE QUÍMICA  
BUCARAMANGA  
2008**

**ÍNDICE DEL ESTADO ELECTROTOPOLÓGICO EN LA PREDICCIÓN DEL  
RENDIMIENTO CUÁNTICO DE PRODUCCIÓN DE OXÍGENO SINGULETE  
PARA DERIVADOS DE 1H-FENALEN-1-ONA**

**CÉSAR AUGUSTO GUARÍN DURÁN**

**Trabajo de grado presentado para optar al título de Químico**

**Directora del proyecto:**

**Dra. Martha Cecilia Daza Espinoza**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE CIENCIAS  
ESCUELA DE QUÍMICA  
BUCARAMANGA  
2008**

## AGRADECIMIENTOS

*A la doctora Martha Cecilia Daza Espinosa, mi directora de proyecto, por sus orientaciones, correcciones, comprensión y paciencia.*

*Al profesor Guillermo Restrepo Rubio, PhD, por su valiosa y generosa colaboración.*

*A los profesores Enrique Mejía Ospino y José Carlos Gutiérrez, por ser los evaluadores de este trabajo de investigación.*

*A mis padres Tano Guarín y Cleotilde Durán por todo el apoyo brindado durante mis estudios.*

*A mis demás familiares, amigos y conocidos, por ser compañeros en este peregrinar por el mundo.*

## DEDICATORIA

*Al Señor, Dios todopoderoso y eterno, por ser el motor de mi búsqueda de la verdad.*

*“buscad el reino y todo lo demás se os dará por añadidura...conoced la verdad y la verdad os hará libres”*

*EVANGELIOS*

# TABLA DE CONTENIDO

	Pág.
1. INTRODUCCIÓN	1
2. ESTADO DEL ARTE	4
2.1 DESCRIPTORES MOLECULARES	6
2.1.1 Grafos e índices topológicos	7
2.2 ÍNDICE DEL ESTADO ELECTROTOPOLÓGICO	10
2.2.1 Estrategia para el uso del E-State a nivel-átomo	12
2.2.2 Hiper-espacio E-State	12
2.3 MODELOS ESTADÍSTICOS	13
2.3.1 Modelo multivariado de regresión lineal	14
2.3.2 Análisis de componentes principales	19
2.4 OXÍGENO SINGULETE Y SENSIBILIZADORES	19
3. METODOLOGÍA	21
3.1 CONJUNTO DE ESTUDIO Y DESCRIPTORES MOLECULARES	21
3.2 CÁLCULOS ESTADÍSTICOS Y FUNCIÓN DE SIMILITUD	23
4. RESULTADOS Y DISCUSIÓN	25

4.1	MODELOS CON DESCRIPTORES CUASI-ORTOGONALES	28
4.2	MODELOS CON DESCRIPTORES ORTOGONALIZADOS	36
4.3	MODELO FINAL	41
4.3.1	Conjunto de prueba	42
4.3.2	Hiper-espacio E-State a partir del modelo N193	44
5	CONCLUSIONES	49
6	RECOMENDACIONES	51
	BIBLIOGRAFÍA	52
	ANEXOS	58

# LISTA DE FIGURAS

	Pág.
Figura 1. Estructura molecular de m-metoxi-Fenol y su representación como un grafo molecular.	8
Figura 2. Espacio 3D para los índices E-State de glicoles (gli), alcanos (a) y alcoholes (o).	13
Figura 3. Estructura de la 1H-Fenalen-1-ona junto a su representación como grafo químico con sus sitios de sustitución. A la derecha se muestra la conversión de los sustituyentes en grafos.	25
Figura 4. Representación como grafos químicos de los diez derivados y de la 1H-Fenalen-1-ona.	26
Figura 5. A. Validación cruzada para el modelo M22	35
B. Validación Cruzada para el modelo M22 sin 4fn2 ni 9fn1.	36
Figura 6. A. Validación cruzada para el modelo N193.	38
B. Validación cruzada para el modelo N193 sin la molécula FN.	38
C. Validación cruzada para el modelo N16.	39
D Validación cruzada para el modelo N191.	39

Figura 7.	Ploteo de los residuales en su orden de aparición.	40
Figura 8.	Descriptores que representan las características estructurales relacionadas con la propiedad $\Phi_{\Delta}$ .	42
Figura 9.	Resultado de la validación cruzada para el modelo construido con 9 moléculas, no se incluyó la 4fn1 ni la 9fn2.	43
Figura 10.	Gráfica por trazas 3D y 1D para el hiperspacio 4D E-State compuesto por cuatro descriptores del estado electrotopológico y nombrado como $V(es1,es2,es3,es7)$ .	45
Figura 11.	Distribución de las moléculas según distancia euclidiana.	46
Figura 12.	Conjuntos del espacio químico y del espacio de los descriptores moleculares.	47

## LISTA DE TABLAS

	Pág.
Tabla 1. Características deseables de los índices topológicos.	9
Tabla 2. Modelos QSPR para puntos de ebullición de alcanos.	15
Tabla 3. Rendimiento cuántico de producción de oxígeno singulete de las moléculas 1H-Fenalen-1-ona y sus diez derivados.	22
Tabla 4. Índices del estado electrotopológico a nivel-átomo para la estructura común FN.	27
Tabla 5. Valores de los índices del estado electrotopológico para cada átomo etiquetado de las estructuras de los diez derivados y la 1H-Fenalen1-ona.	29
Tabla 6. Matriz de correlación entre los 14 descriptores electrotopológicos.	30
Tabla 7. Porcentaje de varianza explicada por los componentes principales CP.	31
Tabla 8. Residuales para el modelo M22 con los descriptores es1 es2 es8 y es11.	31
Tabla 9. Tendencia de los valores del rendimiento cuántico de producción de oxígeno singulete.	32

Tabla 10.	Residuales para modelos de regresión con cuatro descriptores electrotopológicos.	32
Tabla 11.	Parámetros estadísticos para modelos con cuatro descriptores E-State.	34
Tabla 12.	Residuales validación cruzada modelo M22, sin 9fn1 ni 4fn2.	35
Tabla 13.	Base 14 dimensional con todos los valores de los descriptores E-State despues del proceso de ortogonalización.	37
Tabla 14.	Bases de diferentes dimensiones y la cantidad de modelos posibles con cuatro descriptores E-State.	37
Tabla 15.	Correlaciones de los descriptores E-State de la base 5 con $\exp(\Phi_{\Delta})$ .	39
Tabla 16.	Parámetros para los cinco mejores modelos con descriptores ortogonles E-State.	41
Tabla 17.	Datos obtenidos para el modelo N193'.	44
Tabla 18.	Agrupación uno para las Fenalen-1-onas.	46
Tabla 19.	Agrupación dos para las Fenalen-1-onas.	47

# LISTA DE ANEXOS

	Pág.
Anexo 1. Nomenclatura de los modelos construidos con tres, cuatro y cinco descriptores cuasi-ortogonales.	51
Anexo 2. Nomenclatura de los modelos construidos con cuatro descriptores ortogonales provenientes de cinco bases diferentes.	52
Anexo 3. Nomenclatura de los modelos construidos con cuatro descriptores ortogonales provenientes de cinco bases diferentes.	53

**TÍTULO:** ÍNDICE DEL ESTADO ELECTROTOPOLÓGICO EN LA PREDICCIÓN DEL RENDIMIENTO CUÁNTICO DE PRODUCCIÓN DE OXÍGENO SINGULETE PARA DERIVADOS DE 1H-FENALEN-1-ONA\*

**Autor:** César Augusto Guarín Durán\*\*

*Palabras Claves:* Índice del estado electrotopológico, nivel-átomo, 1H-Fenalen-1-ona, rendimiento cuántico de producción de oxígeno singulete, hiperespacio E-State.

### Resumen

En este trabajo se estudió un conjunto de once estructuras moleculares correspondientes al sensibilizador 1H-Fenalen-1-ona y diez de sus derivados, por medio del índice del estado electrotopológico (E-State) como descriptor de aspectos electrónicos y topológicos de las moléculas en cuestión. El objetivo de esta investigación fue la construcción de modelos predictivos para la propiedad rendimiento cuántico de producción de oxígeno singulete ( $\Phi_{\Delta}$ ) tomando como base descriptores electrotopológicos, los cuales, debido a su contribución al modelo, permitieron determinar qué características moleculares disminuyen  $\Phi_{\Delta}$ . Se eligió la descripción estructural a nivel-átomo para ponderar la variación electrotopológica causada por los sustituyentes en cada átomo de la estructura común (la molécula de 1H-Fenalen-1-ona). Mediante la aplicación de 14 E-State se desarrolló varios modelos estructura-propiedad para  $\Phi_{\Delta}$ , haciendo uso de técnicas pertenecientes a la estadística multivariada, particularmente análisis de regresión lineal.

De estos modelos se escogió el que presentó los mejores parámetros de, coeficiente de correlación lineal  $R$ , desviación estandar  $S$  y prueba de Fisher  $F$ , con valores 0,997; 235,926 y 0,067 respectivamente, y con coeficiente de validación cruzado  $Q$ , de 0,995, el cual indica alto poder predictivo. El análisis de las moléculas se realizó con base en los cuatro E-State presentes en la mejor ecuación de regresión, por medio de los cuales, se generó un subespacio 4D-dimensional del hiperespacio E-State y se interpretó a las estructuras moleculares como vectores sobre los que se aplicó la función de similitud de distancia euclidiana para obtener el aporte de cada E-State y las agrupaciones que evidenciaron la similitud con la tendencia de  $\Phi_{\Delta}$ .

---

\*Trabajo de grado para optar al título de químico.

\*\*Facultad de ciencias. Escuela de química. Directora: Martha Cecilia Daza Espinosa, Dra. Grupo de bioquímica teórica.

**TITLE:** ELECTROTOPOLOGICAL STATE INDEX IN THE PREDICTION OF THE QUANTUM YIELD OF SINGLET OXYGEN FOR DERIVATIVES OF THE 1H-PHENALEN-1-ONE\*

**Author:** César Augusto Guarín Durán\*\*

**Keywords:** *Electrotopological state index, atom-level, 1H-Phenalen-1-one, quantum yields of singlet oxygen, E-State hyperspace.*

### **Abstract**

In this work the set of eleven molecular structures was studied. These structures correspond to the sensitizer 1H-Phenalen-1-one and ten of its derivatives, through the electrotopological state index (E-State) as a descriptor of the electronics and topological features of the molecules in study. The objective of this investigation was the construction of predictive models for the quantum yields of singlet oxygen property ( $\Phi_{\Delta}$ ), based in electrotopological descriptors, which due to its contribution to the model, permit determine which molecular characteristics decrease  $\Phi_{\Delta}$ . The structural description atom-level was chosen to weigh the electrotopological variation caused by the substituents over each atom of the common structure (the 1H-phenalen-1-one molecule). By the application of 14 E-State, it was developed several structure-property models for  $\Phi_{\Delta}$ , using techniques of the multivariate statistics, particularly the lineal regression analysis.

The model that showed the best parameters of lineal correlation coefficient R, standard deviation S and Fisher test F, with values of 0,997; 235,926 and 0,067 respectively, and with a crossed validation correlation Q, of 0,995, that indicates a high predictive power. The analysis of the molecules was achieve based in the four E-State presents in the better regression equation, by which was generated a subspace 4D-dimensional of the E-State hyperspace, and was interpreted to molecular structures as vectors, over which the similarity function of Euclidian distance was applied to obtain the contribution of each E-State and the sets that evidenced the similarity with the  $\Phi_{\Delta}$  tendency.

---

\*Undergraduate Degree Work.

\*\*Faculty of Science. Chemistry Department. Director: Martha Cecilia Daza Espinosa, Dr. Grupo de bioquímica teórica.

# 1. INTRODUCCIÓN

La definición de estructura molecular es importante en química. Si se considera la estructura molecular como la representación de las moléculas, puede afirmarse que es la responsable de las propiedades observadas [1-8]. La idea generalizada que se tiene de estructura molecular es la de un conjunto de átomos y de enlaces covalentes en una o varias disposiciones espaciales. Este es el modelo geométrico incompleto, aunque muy útil, derivado del mundo macroscópico de las “esferas y barras” [1]. Desde hace tres siglos muchos investigadores han trabajado en la idea de estructura molecular y han desarrollado diversos modelos para explicar algún aspecto de esta. Por ejemplo los modelos de Boyle, Dalton, Kekulé, Lewis, Pauling, Schrödinger o Bader por citar algunos [3,4,9]. Una de las definiciones, establece que la estructura molecular *“es una entidad de orden topológico y geométrico que corresponde a la realidad de una molécula, y que permite distinguirla de otras agrupaciones y moléculas compuestas por los mismos átomos [1]”*. También Kier y Hall, enuncian que la estructura molecular *“es una expresión general de códigos, mediante la cual se describe una molécula en términos cuantitativos. Este modelo hace referencia a la “forma” de una molécula, la cual produce una serie de funciones llamadas propiedades, que son las respuestas a estímulos físicos o químicos aplicados a un sistema que la contiene. El modelo puede ser incompleto, pero adecuado para un propósito particular, haciendo posible el entendimiento y la predicción de alguna propiedad” [4]*.

Los estudios pioneros de la relación estructura-propiedad de Crum-Brown y Fraser, citados por Selassie [6], y los de la relación estructura-actividad de Hansch y Fujita [7] permitieron encontrar relaciones cuantitativas entre

propiedades experimentales y la constitución de las moléculas, y entre la actividad biológica y algunas propiedades fisicoquímicas. Sin embargo, los modelos topológicos estructura-propiedad-actividad constituyen el mejor intento para demostrar la hipótesis de que la estructura es la responsable de las propiedades determinadas experimentalmente [5,8,10-13]. Actualmente los estudios QSAR/QSPR; de sus siglas en inglés *Quantitative structure-activity relationship* y *Quantitative structure-property relationship*; son usados para la elaboración de modelos estadísticos para cuantificar la relación entre la estructura de un conjunto de moléculas y sus propiedades determinadas experimental o teóricamente. Este tipo de investigaciones permite calcular la propiedad en estudio para otra molécula con estructura semejante y proponer la síntesis de mejores candidatos con actividad deseada [12-13]. En QSPR las características estructurales de una molécula se cuantifican mediante el uso de descriptores moleculares, los cuales tienen como objetivo generar un valor numérico -a partir del grafo químico- para relacionarlo con las magnitudes de alguna propiedad, es decir buscar correlaciones entre dos conjuntos numéricos vía una expresión matemática.

En este trabajo se eligieron los índices del estado electrotopológico E-State, introducidos por Kier y Hall [4], por su capacidad para cuantificar de una manera unificada, aspectos electrónicos y topológicos para los átomos constituyentes de cada molécula. Además, este descriptor permite realizar una interpretación estructural directa, es de fácil cómputo, posee simplicidad como expresión matemática, no está basado en propiedades fisicoquímicas y cambia gradualmente con los cambios graduales de la estructura [8,11]. Utilizamos la metodología QSPR [8,13] y el E-State como descriptor, para evaluar su potencial para la descripción del rendimiento cuántico de producción de oxígeno singulete de la 1H-Fenalen-1-ona y diez de sus derivados. Esta tesis se desarrolló en el grupo de Bioquímica Teórica de la Universidad Industrial de Santander y contó con la colaboración del profesor,

Guillermo Restrepo Rubio de la Universidad de Pamplona. Nuestros resultados nos permiten afirmar que el descriptor E-State, implementado en el paquete computacional MOLCONNZ versión 4.10, es un buen descriptor de la estructura molecular de estos sensibilizadores.

## 2. ESTADO DEL ARTE

La hipótesis QSAR/QSPR sostiene que la estructura de una molécula es la responsable de sus propiedades, este punto de vista fue introducido por Crum-Browm y Fraser quienes son considerados los pioneros de los análisis QSAR/QSPR, según reportan Selassie [6], y Mihalić y Trinajstić [8]. Cuando se investiga un conjunto de estructuras moleculares y alguna propiedad de las sustancias que las moléculas representan, se puede buscar una función matemática que permita cuantificar la relación entre las estructuras y la propiedad. QSAR/QSPR es una hipótesis matemática que tiene su fundamento en el hecho de que la estructura molecular es la responsable de las propiedades químicas, físicas y biológicas [12-13]. Esta hipótesis parece contradecir el concepto de que las propiedades son producidas por conjuntos de moléculas y no por entes individuales, sin embargo muchos estudios QSAR/QSPR han dado resultados satisfactorios que avalan su potencial predictivo [5,14].

La obtención de un buen modelo QSAR/QSPR comienza, en primer lugar con la elección del conjunto de estudio denominado “conjunto de prueba”, que es un conjunto de moléculas para las que se conoce la propiedad que se pretende modelar [13]. En el presente trabajo se eligió el conjunto de sensibilizadores derivados de la molécula 1H-Fenalen-1-ona debido a que hay suficientes datos experimentales para la propiedad “rendimiento cuántico de producción de oxígeno singulete ( $\Phi_{\Delta}$ )”. Los estudios realizados en diversas áreas de la química sobre los sensibilizadores, han permitido un mejor entendimiento de los mecanismos de formación del primer estado triplete excitado y de su potencial para la activación del oxígeno molecular

triplete. También ha sido posible la cuantificación del  $\Phi_{\Delta}$  para diversas sustancias en diferentes solventes y el desarrollo de nuevas aplicaciones para el oxígeno reactivo generado por los sensibilizadores [15-16]. Una de ellas es la búsqueda y desarrollo de sensibilizadores eficientes en terapia fotodinámica, donde las características del sensibilizador son importantes. En este caso es vital que el sensibilizador tenga la capacidad de activarse por longitudes de onda en la región visible del espectro electromagnético. Además, el oxígeno singulete tiene un potencial enorme para la purificación de aguas con una alta carga de materia orgánica y para controlar el desarrollo de patógenos nocivos para la salud humana [17-19].

En segundo lugar la realización de un buen modelo QSAR/QSPR implica la elección de los descriptores, estos pueden ser medidas experimentales de alguna propiedad o entes matemáticos computados a partir del grafo químico, cuya representación se deriva de dibujar la estructura molecular con la regla de intercambio de los átomos por vértices y los enlaces por aristas. De esta última clase de descriptores existen los que codifican información de fragmentos moleculares y los que codifican la molécula como un todo [5,13,20-21]. Actualmente hay descriptores extensamente empleados, como los índices de conectividad de Randić, de conteo de no adyacencia de Hosoya, de suma de distancias de Balaban y de Schultz; los mecanocuánticos, como la energía del HOMO (highest occupied molecular orbital) y del LUMO (lowest occupied molecular orbital), la energía total y la energía de repulsión electrónica, entre otros [22].

Hace más de una década fue propuesto otro descriptor, el índice del estado electrotopológico (E-State), el cual codifica aspectos topológicos y electrónicos para cada átomo o grupo funcional de una molécula [4]. El poder predictivo de los modelos construidos con el E-State, radica en que este descriptor codifica información estructural asociada con el tamaño, la forma, el grado de ramificación, los atributos estéricos y el efecto de los vecinos

sobre cada átomo. Además el descriptor codifica aspectos electrónicos para cada átomo relacionados con su polaridad, carga eléctrica y niveles de energía; ambos atributos, topológicos y electrónicos, aportan al entendimiento de diversas interacciones moleculares [22-26].

En tercer lugar, para obtener un buen modelo QSAR/QSPR se recurre a la aplicación de métodos estadísticos multivariados, que correlacionan el conjunto de datos de los descriptores moleculares, con el conjunto de las propiedades medidas experimentalmente [8,13]. Un método ampliamente usado es el de análisis de regresión multivariable, que tiene como objetivo representar una variable dependiente mediante un conjunto de variables independientes. La calidad del modelo así obtenido usualmente se caracteriza con los parámetros: coeficiente de correlación (R), desviación estándar (S) y prueba de Fisher (F) [4,12-13,27]. Otra técnica estadística usada es el análisis de componentes principales, que tiene como objetivos generar nuevas variables que puedan expresar la información contenida en el conjunto original de datos, reducir la dimensionalidad del problema como paso previo para futuros análisis y eliminar algunas variables originales si ellas aportan poca información [28-29].

## **2.1 DESCRIPTORES MOLECULARES**

Como no existe un método directo ni generalizado que encuentre todas las posibles relaciones entre todos los átomos constituyentes de las moléculas de un conjunto en estudio, ni que cuantifique la totalidad de las características estructurales de las moléculas [13], se recurre al uso de los descriptores moleculares, entendidos como una técnica indirecta que codifica información relevante de la estructura molecular. Los descriptores moleculares permiten la manipulación y el análisis de la información de la estructura química [13,20] por medio de magnitudes numéricas que caracterizan a una sustancia. Algunos descriptores son experimentales, por

ejemplo el índice de refracción, la hidrofobicidad o el coeficiente de partición agua/octanol [7], o pueden provenir de un cálculo grafo teórico, como el índice de Wiener, el índice de conectividad de Randić o el índice del estado electrotopológico [4].

Se han creado tantos descriptores, que se ha planteado una clasificación jerárquica de los mismos en las categorías topoquímicas, topoestructurales, geométricas y mecanocuánticas [5]. Aunque no se ha llegado a crear un descriptor global [11], existen algunos que codifican características importantes de la estructura molecular, como longitud de la cadena, cantidad de ciclos, ramificaciones, etc [4,21], siendo el índice del estado electrotopológico, E-State, creado por Kier y Hall en 1991 [4], útil para codificar aspectos topológicos y electrónicos de cada fragmento de una molécula cualesquiera.

### **2.1.1 Grafos e índices topológicos**

Los índices topológicos se obtienen mediante la aplicación de algoritmos que actúan sobre la representación de la molécula como un grafo químico [5,12,30-31]. Este último concepto proviene de la teoría de grafos, la cual hace parte de las matemáticas discretas y está relacionada con la topología y la combinatoria. Los grafos son conjuntos de vértices y aristas representables en una, dos o más dimensiones [30]. El grafo químico es el conjunto de las conexiones entre átomos, electrones, moléculas, fragmentos moleculares o intermediarios, particularmente el grafo molecular está basado en la equivalencia de las fórmulas químicas moleculares con los grafos matemáticos, donde se usa la regla sencilla de transformación de átomos por vértices y enlaces por aristas (figura 1) [4,13,30,32].

El grafo se puede ponderar en sus vértices o aristas, dependiendo del interés que se le quiera dar al cálculo de algún descriptor para su posterior estudio QSAR/QSPR [22,33].

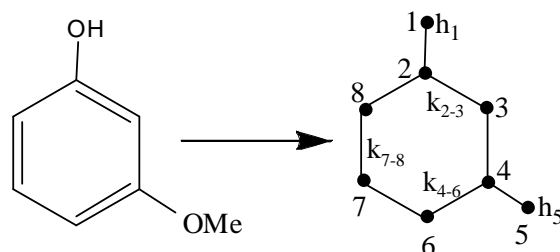


Figura 1. "Estructura molecular" de m-Metoxi-Fenol y su representación como un grafo molecular, donde  $k_{i-j}$  es la ponderación de la arista que se halla entre los vértices etiquetados como  $i$  y  $j$ , y sirve para representar un doble enlace.  $h_i$  es la ponderación del vértice  $i$ , significando que se trata de un heteroátomo, en este caso del oxígeno.

Como expresión matemática, el grafo es una colección de dos conjuntos, el de vértices y el de aristas, cuya forma es:

$$\mathbf{G} = (\mathbf{V}(\mathbf{S}(h)), \mathbf{A}(\mathbf{B}(k))) \quad (\text{Ecuación 1})$$

Donde  $\mathbf{G}$  es el grafo molecular,  $\mathbf{V}$  el conjunto de vértices,  $\mathbf{A}$  el conjunto de aristas,  $\mathbf{S}(h)$  y  $\mathbf{B}(k)$  son las ponderaciones. La aplicabilidad directa de los grafos moleculares en la química, radica en el hecho de que vértices y aristas representan diferentes objetos, y esto permite en primer lugar, etiquetar los vértices para representar heteroátomos y distinguirlos por ejemplo de un átomo de carbono. Para los vértices, la ponderación podrá considerar el número atómico, la masa o una relación con respecto a los átomos de carbono [33]. En segundo lugar se pueden ponderar las aristas para representar por ejemplo el orden de enlace, así un enlace sencillo tendrá un valor de 1, un enlace aromático de 1,5 y uno doble de 2.

Los índices topológicos computan diversos aspectos de la estructura, tales como conectividad, distancia, cantidad de átomos de un mismo elemento, grupos funcionales, cantidad de enlaces, etc. Los valores obtenidos por estos

descriptores se denominan invariantes. Cada índice recibe un nombre específico por el que lo reconocen los químicos teóricos y las revistas especializadas; este nombre puede ser el de su creador, el de los aspectos que codifica, o el de la propiedad que representa.

Tabla 1. Características deseables de los índices topológicos [20].

<b>Característica</b>
Interpretación estructural directa.
Buena correlación por lo menos con una propiedad.
Discriminación de isómeros.
Localización definida.
Generalizable a altas analogías.
Independencia lineal.
Simplicidad.
No basado en propiedades fisicoquímicas.
No estar relacionado trivialmente con otros descriptores.
Eficiencia en su elaboración.
Basado en conceptos estructurales.
Mostrar una correcta dependencia con el tamaño.
Cambios graduales con cambios graduales de la estructura.

La cuantificación de la estructura molecular por medio de un índice topológico está acompañada de una pérdida de información, sin embargo las características estructurales que codifica el índice pueden ser de interés en un estudio de relaciones estructura-propiedad. Algunas investigaciones han

determinado los criterios sobre los requerimientos que debe poseer un índice, la tabla 1 muestra las características deseables reportadas por Randić [20].

Los índices topológicos poseen las ventajas sobre los índices geométricos y mecanocuánticos de ser computados directamente desde el grafo molecular, tener un único valor para un compuesto químico en particular y ser calculados con poco costo computacional.

## 2.2 ÍNDICE DEL ESTADO ELECTROTOPOLÓGICO

Este índice se simboliza por la letra  $S$ . Su cálculo se realiza a partir del grafo molecular y mediante la ecuación:

$$S_i = I_i + \sum_j \Delta I_{ij} \quad (\text{Ecuación 2})$$

Donde,  $S_i$  simboliza el índice del estado electrotopológico para el  $i$ -ésimo átomo o fragmento. El valor de  $S_i$  se obtiene por la suma entre el valor intrínseco  $I_i$  para el átomo  $i$ -ésimo, y la perturbación que este experimenta por el resto de los componentes  $j$ -ésimos de la estructura molecular  $\Delta I_{ij}$ . Ambas variables,  $I_i$  y  $\Delta I_{ij}$ , se calculan por conteo de electrones  $\sigma$ ,  $\pi$ ,  $n$ , átomos de hidrógeno ( $h$ ) enlazados y " $r$ ", que es la distancia topológica entre los vértices del grafo desplazada en una unidad. (Ecuaciones 3, 4, 5 y 6) [4].

$$\delta = \sigma - h \quad (\text{Ecuación 3}) \quad \delta^v = Z^v - h = \sigma + \pi + n - h \quad (\text{Ecuación 4})$$

$$I_i = \frac{\left(\frac{2}{N}\right)^2 \delta^v + 1}{\delta} \quad (\text{Ecuación 5}) \quad \Delta I_{ij} = \frac{I_i - I_j}{r_{ij}^m} \quad (\text{Ecuación 6})$$

Las variables  $I_i$  y  $\Delta I_{ij}$  dependen de delta ( $\delta$ ) y de delta de valencia ( $\delta^v$ ), el primero codifica la cantidad de átomos adyacentes diferentes al hidrógeno, los electrones sigma ( $\sigma$ ) en un átomo excluyendo los enlazados al hidrógeno, la cantidad de enlaces diferentes a los establecidos con hidrógenos y el entorno topológico inmediato de un átomo en una molécula. El segundo está basado en el número total de electrones de valencia substrayendo los electrones enlazados a los hidrógenos, y en el conteo de electrones sigma excluyendo los enlazados al hidrógeno, y también en el conteo de electrones pi ( $\pi$ ) y pares libres ( $n$ ). Estos deltas proveen información acerca de los átomos en la molécula y son útiles para la descripción de la estructura molecular. Como el E-State es un descriptor de los átomos en una molécula, existen tantos índices como átomos hay en la molécula. Los símbolos para estos se derivan del símbolo atómico y del tipo de enlaces que posee el átomo en cuestión, s para sencillo, d para doble, t para triple y a para aromático. Por ejemplo SsssCH, SdsCH, StCH y SaaCH, son los símbolos del E-State para un carbono metínico, con enlaces sencillos, doble, triples y aromáticos, respectivamente. También existen extensiones de los índices E-State, como los denominados átomo-tipo, que son la suma de todos los descriptores E-State del mismo tipo dentro de la molécula. Otros casos son los, HE-State el estado electrotopológico para cada hidrógeno y los grupo-tipo, definido como el E-State para ciertas funciones químicas como cetona, aldehído o metileno [4]. Por último, cabe mencionar que el E-State ha dado buenos resultados en estudios QSAR/QSPR de aminas aromáticas, metanos halogenados, compuestos nitro-aromáticos, derivados de xantonas, inhibición de la enzima monoamina oxidasa por hidrazidas, umbral de olor de alquilpirazinas, inhibición de la actividad de la adenosina A por xantinas, benzimidazoles como inhibidores del virus influenza, afinidad de enlazamiento al receptor dopamina, actividad opioide, toxicidad de bencenos substituidos, puntos de ebullición de policloroalcanos, alcoholes y alcanos, entre otros [4].

### **2.2.1 Estrategia para el uso del E-State a nivel-átomo**

Las moléculas que presentan una estructura común, donde hay sustituyentes que varían en posiciones de sustitución, producen cambios graduales en el carácter electrotopológico de los átomos en la molécula [4]. En estas estructuras los átomos comunes pueden ser numerados de igual forma, de manera que los E-State para estas posiciones pueden ser usados como variables en QSAR/QSPR. Por ejemplo, en un conjunto de derivados del fenol se conservan siempre los mismos siete átomos, sobre los cuales se pueden computar los valores para los índices del estado electrotopológico que serán diferentes dependiendo del tipo de sustituyente en el anillo aromático. Estos valores obtenidos en cada átomo, por ejemplo en un carbono aromático, codificarán su accesibilidad electrónica. Estos resultados son útiles para un análisis QSAR/QSPR ya que permiten plantear la posibilidad de determinar qué codificación estructural varía “proporcionalmente” con los datos de alguna medida experimental [4].

### **2.2.2 Hiper-espacio E-State**

Según Kier y Hall, los valores netos, la suma o los promedios del E-State, pueden servir como ejes de un espacio  $n$ -dimensional, generando una “base vectorial” que cubre todas las posibles estructuras moleculares en términos de todos los posibles átomos que las constituyen [4]. Cada dimensión es un parámetro calculado para un átomo particular o grupo. Por ejemplo los grupos OH,  $-\text{CH}_2-$  y  $\text{CH}_3$  de alcanos, alcoholes y glicoles pueden servir de ejes en un espacio de tres dimensiones, en donde una combinación de los índices calculados para estos tres grupos puede representar al conjunto de moléculas. Moviéndose a través de los ejes se describe el cambio que experimentan los E-State en un átomo particular, cambio generado por efecto de la proximidad y distinta electronegatividad de otro átomo.

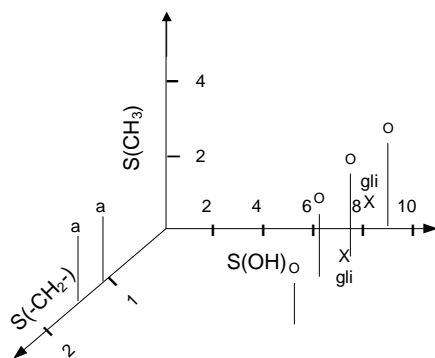


Figura 2. Espacio 3D para los índices E-State de glicoles (gli), alcanos (a) y alcoholes (o).

El despliegue de las moléculas en un espacio  $n$ -dimensional es útil en estudios de similitud molecular. De esta manera, en el ejemplo citado arriba, cada molécula del conjunto se ubicará en un punto compuesto por tres coordenadas [4], la figura 2 muestra esta situación para el conjunto de ocho moléculas localizadas en el espacio definido por el promedio de los valores del E-State átomo-tipo para el índice del estado electrotopológico de OH, -CH<sub>2</sub>- y CH<sub>3</sub>. En este gráfico se visualizan dos agrupaciones, uno para los alcanos, y el otro para los alcoholes y glicoles. Un camino más riguroso para cuantificar la semejanza molecular, es el empleo de una función de similitud [34], siendo la distancia euclidiana una de las más generalizadas.

### 2.3 MODELOS ESTADÍSTICOS

En el estudio de las relaciones de QSAR/QSPR, se busca construir una función capaz de determinar óptimamente el valor numérico de la propiedad en estudio. Esta relación conocida como el “modelo”, se representa matemáticamente por  $Y = f(x)$  donde  $Y$  es la propiedad o actividad,  $f$  es la función construida mediante métodos estadísticos y  $x$  son los aspectos relevantes de la estructura molecular que determinan la propiedad [13].

La utilidad de un modelo QSAR/QSPR radica en que ahorra tiempo en la obtención de estructuras candidatas que tengan actividad química deseada,

disminuye el costo de reactivos debido a que reduce la síntesis de sustancias derivadas de un patrón que presente cierta actividad conocida, sugiere estructuras moleculares con actividad más potente y también permite estimar el valor de una propiedad para moléculas similares de las que no se tenga ese dato ya sea por tratarse de sustancias tóxicas o muy inestables.

En la obtención del modelo QSAR/QSPR es importante la elección de los descriptores; si bien es imprescindible disponer para un conjunto dado de moléculas del valor de sus propiedades, también es de suma importancia elegir aquellos índices que sirvan para la elaboración de un modelo con buenos parámetros estadísticos de coeficiente de correlación  $R$ , desviación estándar  $S$  y parámetro de Fisher  $F$ , que son los que cuantifican la calidad del modelo.

En la construcción del modelo se emplean relaciones lineales y métodos estadísticos, entre los más empleados están el análisis de componentes principales, las redes neuronales artificiales, los mínimos cuadrados parciales y el análisis de regresión multivariable [13,28].

### **2.3.1 Modelo multivariado de regresión lineal**

Esta técnica estadística se usa en la hipótesis QSAR/QSPR para encontrar, mediante correlación entre la propiedad (actividad) y los descriptores moleculares, varias funciones de la forma mostrada en la ecuación 7. De estas ecuaciones se escoge aquella que muestre el mejor coeficiente de correlación  $R$ , siempre mayor a 0,80, la menor desviación estándar  $S$  y el mayor parámetro de Fisher  $F$ . Por ejemplo, en un estudio QSPR para modelar puntos de ebullición de alcanos [8], se obtuvieron las funciones mostradas en la tabla 2 para los descriptores moleculares Wiener  $W$ , Hosoya  $Z$ , Randić  $\chi$ , Balaban  $J$ , Schultz  $MTI$  y Harary  $H$ , de los cuales los modelos contruidos con el índice Hosoya  $Z$  y el índice Randić  $\chi$  tienen los mejores parámetros estadísticos  $R$ ,  $S$  y  $F$ .

Una mirada rápida a las bases matemáticas detrás de la elaboración de un modelo de regresión lineal comienza con el entendimiento del objetivo del análisis, que consiste en establecer una correlación entre una variable dependiente y otras variables independientes.

Tabla 2. Modelos QSPR para puntos de ebullición de alcanos [11].

Función	R	S	F
$pe = 77,93(W) - 3,35 \cdot 10^{-5} (W^3) - 164,24$	0,9969	5,09	1906
$pe = 122,66 (\ln Z) - 11,534(\ln Z)^2 - 164,48$	<b>0,9975</b>	<b>4,52</b>	<b>3638</b>
$pe = 81,14 \chi + 6,60(\chi^2) - 2,168(\chi^3) - 165,71$	<b>0,9977</b>	<b>4,52</b>	<b>2593</b>
$pe = 139,4 (J) - 14,61 (J)^2 - 186,68$	0,8824	29,89	65
$pe = 48,92 (MTI) - 7,67 (MTI) - 163,93$	0,9964	5,48	1644
$pe = 80,54 H - 165,8$	0,9864	10,46	664

Los valores en negrilla son los mejores valores, desde el punto de vista estadístico, para coeficiente de correlación R, desviación estándar S y parámetro de Fisher F

El modelo tiene la forma:

$$Y_{CALC} = b + B_1 X_1 + B_2 X_2 + \dots + B_n X_n \text{ (Ecuación 7)}$$

Donde  $Y_{CALC}$  es la variable dependiente calculada,  $b$  y  $B_1, \dots, B_n$  son la constante y los coeficientes a estimar, y  $X_1, X_2, \dots, X_n$  son el conjunto de variables independientes. La obtención de los coeficientes  $B_i$  y de la constante  $b$ , se realiza vía la minimización del cuadrado de los errores – denominados residuales en la literatura QSAR/QSPR– (ecuación 8). Para hacer esto se define el residual  $E_i$ , (ecuación 9), como la diferencia entre el valor observado  $Y_{OBS}$  (variable dependiente) y el valor calculado  $Y_{CALC}$

(variable dependiente calculada), se reemplaza  $Y_{CALC}$  y se define el cuadrado de los residuales (ecuaciones 7,10-11).

$$\frac{\partial E_i^2}{\partial B_i} = \frac{\partial E_i^2}{\partial b} = 0 \quad (\text{Ecuación 8})$$

$$E_i = Y_{OBS} - Y_{CALC} \quad (\text{Ecuación 9})$$

$$E_i = Y_{OBS} - B_i X_i - b \quad (\text{Ecuación 10})$$

$$\sum_{i=1}^n E_i^2 = SS \quad (\text{Ecuación 11})$$

Como la definición de SS, es la sumatoria del cuadrado de los residuales, se reemplazan los valores de las funciones Y, y se desarrolla el binomio (ecuaciones 12-14).

$$SS = \sum_{i=1}^n (Y_{OBS} - Y_{CALC})^2 \quad (\text{Ecuación 12})$$

$$SS = \sum_{i=1}^n (Y_{OBS} - B_i X_i - b)^2 \quad (\text{Ecuación 13})$$

$$SS = \sum_{i=1}^n (Y_{OBS}^2 - Y_{OBS} B_i X_i - Y_{OBS} b - Y_{OBS} B_i X_i + B_i^2 X_i^2 + B_i X_i b - b Y_{OBS} + b B_i X_i + b^2) \quad (\text{Ecuación 14})$$

Luego se deriva la ecuación 14 y se minimiza con respecto a  $b$  y  $B_i$  (ecuaciones 15-18), se reescriben, y por solución simultanea del sistema de ecuaciones se consiguen los valores de  $b$  y  $B_i$  (ecuaciones 19-20).

$$\frac{\partial SS}{\partial b} = \sum_{i=1}^n -2(Y_{OBS} - b - B_i X_i) \quad (\text{Ecuación 15})$$

$$\frac{\partial SS}{\partial B_i} = \sum_{i=1}^n -2X_i(Y_{OBS} - b - B_i X_i) \quad (\text{Ecuación 16})$$

$$\sum_{i=1}^n (Y_{OBS} - b - B_i X_i) = 0 \quad (\text{Ecuación 17})$$

$$\sum_{i=1}^n X_i (Y_{OBS} - b - B_i X_i) = 0 \quad (\text{Ecuación 18})$$

$$b \sum_{i=1}^n X_i + B_i \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_{OBS} \quad (\text{Ecuación 19})$$

$$b + B_i \sum_{i=1}^n X_i = \sum_{i=1}^n Y_{OBS} \quad (\text{Ecuación 20})$$

La “calidad” de un modelo de regresión lineal se mide por los parámetros estadísticos coeficiente de correlación  $R$ , desviación estándar  $S$  y parámetro de Fisher  $F$ . El coeficiente  $R$  (a veces se evalúa el  $R^2$ ) toma valores desde 0 a 1, e indica la proporción de la variación en la variable dependiente que es explicada por la ecuación de regresión, así un valor cercano a cero indica que la ecuación de regresión no explica la variable dependiente. El  $R^2$  se define con la ecuación 21, el  $R$  es la raíz cuadrada de esta ecuación.

La desviación estándar es una medida absoluta de la calidad del ajuste. Idealmente debería ser cero, pero en situaciones experimentales no es así. Se calcula por medio de la raíz cuadrada de la ecuación 22.

$$R^2 = 1 - \frac{SS}{TSS} = 1 - \frac{\sum_{i=1}^n (Y_{OBS} - Y_{CALC})^2}{\sum_{i=1}^n (Y_{OBS} - \bar{Y})^2} \quad (\text{Ecuación 21})$$

El parámetro de Fisher F es usado como una medida del nivel de significancia estadística del modelo de regresión. Los valores calculados para F son comparados con valores reportados en tablas estadísticas, permitiendo analizar diferentes niveles de confianza. Si el valor calculado es más grande que el tabulado, se dice que la ecuación es significativa a un nivel de confianza particular. Altos valores de F corresponden a altos niveles de significancia. Por medio de la ecuación 23 se calcula el parámetro F.

$$S^2 = \frac{SS}{n - p - 1} = \frac{\sum_{i=1}^n (Y_{OBS} - Y_{CALC})^2}{n - p - 1} \quad (\text{Ecuación 22})$$

$$F = \frac{ESS}{p} = \frac{\sum_{i=1}^n (Y_{CALC} - \bar{Y})^2}{p} \cdot \frac{n - p - 1}{\sum_{i=1}^n (Y_{OBS} - Y_{CALC})^2} \quad (\text{Ecuación 23})$$

Donde  $n$  es la cantidad de casos,  $p$  es la cantidad de variables independientes,  $ESS$  es la sumatoria de cuadrados entre la resta de la variable dependiente calculada y su media  $\bar{Y}$  y  $SS$  es la sumatoria de cuadrados de los errores o residuales [12].

### 2.3.2 Análisis de componentes principales

Los objetivos principales de este análisis son en primer lugar encontrar una manera simplificada de representar un conjunto de datos en otro conjunto de menor dimensión, en segundo lugar examinar la interdependencia de las variables desde la independencia total hasta la colinealidad, evitando así redundancias. Como último objetivo está el desarrollo de modelos con base en los componentes principales generados [28-29].

### 2.4 OXÍGENO SINGULETE Y SENSIBILIZADORES

La química del oxígeno singulete  $O_2(^1\Delta_g)$  tiene importancia biológica, atmosférica, terapéutica y de síntesis orgánica. Existen muchas investigaciones tanto teóricas como experimentales en cada uno de los campos arriba mencionados [35-40], pero especialmente en los concernientes al impacto del  $O_2(^1\Delta_g)$  en la química y la medicina, tanto que se han hecho investigaciones importantes que han conducido al desarrollo de terapias fotodinámicas donde las moléculas que interaccionan con el oxígeno molecular, para producir oxígeno singulete  $O_2(^1\Delta_g)$  juegan un papel importante por su capacidad de activación por efecto de la luz [16,41]. De la misma forma en las plantas existen moléculas que usan el oxígeno singulete para rechazar naturalmente invasores fúngicos o bacterianos, o para defenderse de agresiones mecánicas. Estas moléculas, denominadas fitoalexinas, son fotosensibilizadores de oxígeno singulete que exhiben una gran variedad de estructuras; desde las terpenoides como isoprenos y sesquiterpenos, hasta las fenólicas con varias unidades aromáticas y grupos substituyentes. Este último grupo de sensibilizadores son un conjunto importante de fitoalexinas que ha sido estudiado recientemente [15,17].

La producción de oxígeno singulete se cuantifica por medio del rendimiento cuántico de producción de oxígeno singulete ( $\Phi_\Delta$ ) que es definido como la

medida de la cantidad de moléculas de  $O_2(^1\Delta_g)$  producidas por absorción de cuantos de luz.

Dicho valor se calcula por medio de la ecuación reportada por Derosa y Crutchley [36] (Ecuación 24). Donde  $\Phi_T$  es el rendimiento cuántico de la formación de oxígeno triplete,  $\Phi_{en}$  la eficiencia en la transferencia de energía,  $k_{en}$  la constante de velocidad de transferencia de energía,  $k_q$  la constante de velocidad de atrapamiento (quenching) del fotosensibilizador por oxígeno, y  $k_{nr}$  y  $k_r$  son las constantes de velocidad para la relajación del sensibilizador. El estándar para el  $\Phi_\Delta$  se referencia como 1 y corresponde a la cantidad de moléculas de  $O_2(^1\Delta_g)$  producidas por la 1H-Fenalen-1-ona que según estudios realizados en varios solventes dicho valor no cambia significativamente [16].

$$\Phi_\Delta = \Phi_T \Phi_{en} = \Phi_T \left( \frac{k_{en} [O_2]}{k_r + k_{nr} + k_q [O_2]} \right) \quad (\text{Ecuación 24})$$

### 3. METODOLOGÍA

La metodología desarrollada en este trabajo se ejecutó en dos partes. La primera correspondió al cálculo de los descriptores del índice del estado electrotopológico para las once moléculas reportadas por Flors [15]. La segunda parte consistió en el tratamiento estadístico de los datos, la consecuente validación de los mejores modelos obtenidos y el uso de la función de similitud de distancia euclidiana para “comparar” las once moléculas.

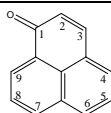
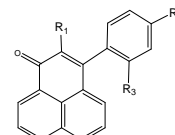
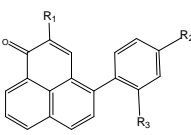
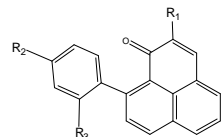
#### 3.1 CONJUNTO DE ESTUDIO Y DESCRIPTORES MOLECULARES

Las moléculas de las que se tomaron los valores de la propiedad experimental cumplen los criterios exigidos en la hipótesis QSPR [4,8,13], de presentar una estructura molecular común, de poseer diferentes sustituyentes que en parte permutan en todas las posiciones de sustitución, y de poseer una propiedad ( $\Phi_{\Delta}$ ) para la cual sus valores numéricos reportados no son discretos. El conjunto de once moléculas, correspondientes al sensibilizador 1H-Fenalen-1-ona y diez de sus derivados, se eligió por tener la estructura molecular común FN (1H-Fenalen-1-ona) y por ser una colección de moléculas en las que el valor de la propiedad rendimiento cuántico de producción de oxígeno singulete, varía por el intercambio de los sustituyentes OH, MeO, Fenil, NO<sub>2</sub> y H en las posiciones tres, cuatro, y nueve, cambio que en principio puede ser representado por uso de algún descriptor molecular.

Las moléculas fueron: 3-(4-metoxi)-Fenilfenalen-1-ona, 2-metoxi-4-Fenilfenalen-1-ona, 2-Hidroxi-4-(4-metoxi)-Fenilfenalen-1-ona, 9-(2-Nitro)-Fenilfenalen-1-ona, 9-Fenilfenalen-1-ona, 2-Metoxi-9-Fenilfenalen-1-ona, 9-

(4-Metoxi)-Fenilfenalen-1-ona, 2-Hidroxi-9-Fenilfenalen-1-ona, 2-Hidroxi-9-(4-Hidroxi)-Fenilfenalen-1-ona, 2-Hidroxi-9-(4-Metoxi)-Fenilfenalen-1-ona y 1H-Fenalen-1-ona [15-16]. Sus representaciones estructurales se muestran en la tabla 3.

Tabla 3. Rendimiento cuántico de producción de oxígeno singlete de las moléculas 1H-Fenalen-1-ona y sus diez derivados. Se muestra la numeración de la IUPAC [15].

MOLÉCULA	REPRESENTACIÓN ESTRUCTURAL	SUSTITUYENTES			$\Phi_{\Delta}$
		R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	
1H-Fenalen-1-ona		---	---	---	<b>1</b>
3-Fenilfenalenona		H	MeO	H	<b>0,94</b>
4-Fenilfenalenonas		H	H	H	<b>0,72</b>
		OH	MeO	H	<b>0,12</b>
9-Fenilfenalenonas		H	MeO	NO <sub>2</sub>	<b>0,51</b>
		H	H	H	<b>0,08</b>
		MeO	H	H	<b>0,08</b>
		H	MeO	H	<b>0,01</b>
		OH	H	H	<b>0,03</b>
		OH	OH	H	<b>0,01</b>
		OH	MeO	H	<b>0,002</b>

Se escogió el conjunto de índices del estado electrotopológico como descriptores para las estructuras de las moléculas mencionadas arriba, debido a su capacidad de cuantificar átomos y grupos funcionales dentro de una molécula. Este descriptor permitió ponderar aspectos electrónicos y topológicos para cada uno de los átomos dentro de las moléculas derivadas de la 1H-Fenalen-1-ona.

La representación estructural de las moléculas en 2D, los denominados grafos moleculares, se desarrollaron con el software GAUSSVIEW [42], En las entradas construidas se etiquetó y se ponderó los vértices y aristas, para después calcular por uso del paquete computacional MOLCONNZ versión 4.10\_WIN2K EDUSOFT 2003 [43], el valor numérico de cada E-State calculado con la ecuación 2.

### **3.2 CÁLCULOS ESTADÍSTICOS Y FUNCIÓN DE SIMILITUD**

Para el tratamiento estadístico se usó el paquete computacional STATISTICA versión 6.0 STATSOFT 2001 [44], con este programa se construyó la matriz de correlación para los datos, se realizó el análisis componentes principales, se desarrollaron los modelos de correlación estructura-propiedad entre el conjunto de descriptores de los derivados de la 1H-Fenalen-1-ona y el rendimiento cuántico de producción de oxígeno singulete, se validó cada uno de los mejores modelos obtenidos y se ortogonalizó el conjunto de los descriptores E-State.

La validación cruzada para cada uno de los modelos desarrollados, se usó para verificar la calidad estadística de los mismos, el coeficiente de correlación cruzado Q se calculó realizando varias iteraciones con modelos construidos con (n-1) individuos. Se construyeron once nuevos modelos por cada modelo obtenido, por medio de la metodología de dejar una molécula afuera (LOO) [5,13].

Con cada nuevo modelo obtenido se realizó la predicción de  $\Phi_{\Delta}$  para la molécula retirada. Al final de todos los cálculos se obtuvo un conjunto de valores predichos de la propiedad para las once moléculas. Este conjunto de valores se correlacionó con los datos experimentales para el conjunto de moléculas y el resultando fue el coeficiente Q. Este procedimiento se desarrolló para los mejores modelos desde el punto de vista estadístico, obteniéndose varios coeficientes Q. Todos los modelos se construyeron por el método de regresión lineal múltiple.

Los descriptores E-State presentes en el mejor modelo QSPR, se usaron como ejes de un espacio 4-dimensional, y como componentes de las estructuras vistas como vectores. Se utilizó la métrica euclidiana como función de similitud.

## 4. RESULTADOS Y DISCUSIÓN

En esta investigación se usaron los resultados experimentales para el valor de  $\Phi_{\Delta}$  (tabla 3), reportados por Flors y Nonelli [15] para un conjunto de derivados de 1H-Fenalen-1-ona, y el valor del  $\Phi_{\Delta}$  reportado en el documento de Schmidt et al para este sensibilizador [16]. Estos valores están dentro del rango de 0,002 a 1, y poseen la característica de ser adimensionales. En el primero de estos artículos no se reporta el error experimental para cada valor del  $\Phi_{\Delta}$ , sin embargo en otros documentos los autores indican una incertidumbre del 10% sobre el valor de la propiedad para los derivados [19], y para la 1H-Fenalen-1-ona una incertidumbre de  $\pm 0,05$  [16,41]. Las representaciones estructurales de cada una de estas moléculas se dibujaron como grafos químicos mediante la conversión de átomos por vértices y enlaces por aristas [12].

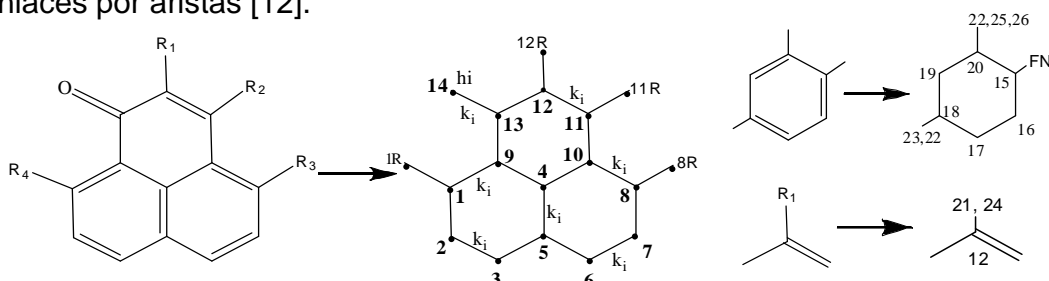


Figura 3. Estructura de la 1H-Fenalen-1-ona junto a su representación como grafo químico con sus sitios de sustitución. A la derecha se muestra la conversión de los sustituyentes en grafos. Cada grafo está etiquetado.

Los vértices etiquetados fueron aquellos del grafo común, el cual denominamos la estructura común FN (1H-Fenalen-1-ona). La figura 3 muestra la conversión de la “estructura molecular” de la 1H-Fenalen-1-ona a su respectivo grafo molecular, algunos de los vértices y las aristas llevan la ponderación  $h_i$ ,  $k_i$ , es decir se les asigna un valor para distinguir los vértices (heteroátomos) y las aristas (tipos de enlace). La asignación de los  $h_i$ ,  $k_i$  en todos los grafos se incluyó en las entradas para el cómputo del E-State.

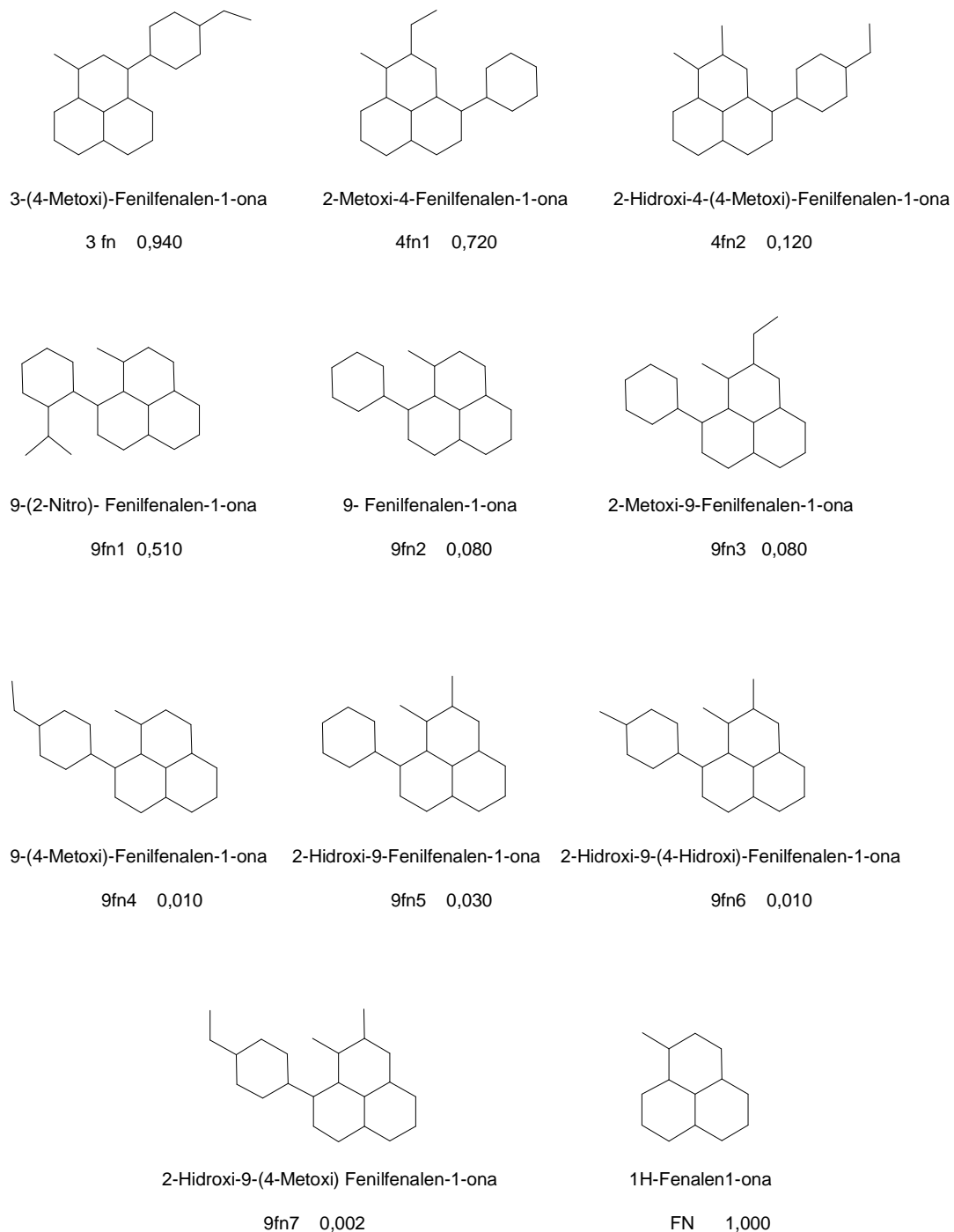









Figura 4. Representación como grafos químicos de los diez derivados y de la 1H-Fenalen-1-ona. Los grafos son de hidrógenos suprimidos. Se muestra la etiqueta usada en este trabajo para cada nombre IUPAC de las moléculas, seguido de la magnitud del  $\Phi_{\Delta}$ .

La figura 4 muestra la representación de los grafos moleculares para todas las moléculas de este trabajo, se incluye el nombre sistemático, la abreviatura usada en este documento y el valor numérico de la propiedad para cada una. Los tipos de átomos presentes en cada “estructura” y los descriptores a calcular se muestran en la tabla 4 donde se especifica cada clase de índice, el grupo funcional presente y la codificación como esi para los E-State. Para los átomos de la estructura común FN la *i* va de 1 a 14 y representa el número de la etiqueta del átomo de la figura 3.

Tabla 4. Índices del estado electrotopológico a nivel-átomo para la estructura común FN.

CÓDIGO	E-STATE	FUNCIÓN QUÍMICA	DESCRIPCIÓN
es1 es8	SaasC		Carbono aromático substituido
es2 es3 es6 es7	SaaCH		Carbonos metínicos aromáticos
es4	SaaaC		Carbono aromático sin accesibilidad
es5 es9 es10	SaaaC		Carbonos aromáticos accesibles
es11 es12	SdssC SdsCH		Carbono olefínico con o sin substitución
es13	SdssC		Carbono carbonílico
es14	SdO		Oxígeno cetónico

Una vez construidas las matrices de entrada para cada estructura, se realizó el cálculo de los descriptores moleculares según el algoritmo del E-State. Sobre la 1H-Fenalen-1-ona y su conjunto de derivados se evaluaron los 14 E-State a nivel-átomo para los vértices ponderados de la estructura común FN (figura 1), cuyo resultado se muestra en la tabla 5; para estos datos se construyó la matriz de correlación por el método de Pearson [28-29] (tabla 6), mediante la cual se determinaron los descriptores menos correlacionados que se emplearon más adelante en los modelos de correlación QSAR/QSPR.

Mediante la aplicación del análisis de componentes principales PCA se analizó la varianza total acumulada por los componentes principales, de esta forma se determinó que cuatro componentes son suficientes para la elaboración del modelo de regresión lineal (columna 4 de la tabla 7). Este resultado está de acuerdo con la literatura, la cual estima que para conjuntos pequeños de datos se debe usar 1 descriptor por cada tres moléculas, en nuestro caso 3,7 es decir cuatro descriptores [4-5].

Alcanzadas las etapas de la metodología QSAR/QSPR, correspondientes a la elección del conjunto de valores de la propiedad de interés y de los descriptores para codificar cada estructura molecular, el siguiente paso que se realizó fue la construcción de los modelos estructura-propiedad mediante la búsqueda de correlación entre los descriptores del índice del estado electrotopológico y el rendimiento cuántico de producción de oxígeno singulete. El fundamento para elaborar estos modelos descansa en la estrategia de uso del E-State a nivel-átomo [4].

Los primeros tres modelos que se desarrollaron por regresión lineal múltiple fueron en los que se utilizaron tres y cuatro de los componentes principales obtenidos por PCA. De estos se escogió el de mejores parámetros R, S y F con valores de 0,887, 0,236 y 5,5 respectivamente. De la posterior validación cruzada se obtuvo como resultado un Q de 0,35 el cual indica bajo poder predictivo, por lo que el modelo se descartó [5].

#### **4.1 MODELOS CON DESCRIPTORES CUASI-ORTOGONALES**

De los resultados de la tabla 5 se observa que los descriptores menos correlacionados son los es1, es2, es8, es11, es12, es13 y es14 [13,14 28]. En el caso del descriptor es13 que no cumple estos criterios, este se eligió debido a que tres de sus cuatro correlaciones de 0,90 son con el es4, es9 y es10, los cuales no se usaron en la elaboración de los modelos.

La elección de sólo estos siete descriptores se justifica ya que las posibles correlaciones con la propiedad serían redundantes si se usara los descriptores interdependientes estadísticamente [33]. La cantidad de modelos que se pueden construir por uso de tres, cuatro y cinco de los 14 índices electrotopológicos es de 3367, según la fórmula combinatoria [13],

$$\binom{N}{l} = \frac{N!}{(N-l)!l!} \text{ (Ecuación 25)}$$

Tabla 5. Valores de los índices del estado electrotopológico para cada átomo etiquetado de las estructuras de los diez derivados y la 1H-Fenalen1-ona.

	$\Phi_{\Delta}$	$\exp(\Phi_{\Delta})$	es1*	es2	es3	es4	es5	es6
<b>FN</b>	1,000	2,7183	1,8866	1,9291	2,0365	1,0856	1,1368	2,0365
<b>3fn</b>	0,940	2,5600	1,8983	1,9361	2,0482	1,0447	1,0995	2,0482
<b>4fn1</b>	0,720	2,0544	1,8700	1,9168	2,0377	1,0089	1,0759	2,073
<b>4fn2</b>	0,120	1,1275	1,7591	1,8383	1,9758	0,8922	0,9925	2,0053
<b>9fn1</b>	0,510	1,6653	0,6033	1,7925	1,9036	0,8571	0,9561	1,9424
<b>9fn2</b>	0,080	1,0833	1,0138	2,0494	2,0944	1,0707	1,1203	2,0591
<b>9fn3</b>	0,080	1,0833	0,9520	2,0211	2,0730	1,0089	1,0759	2,0377
<b>9fn4</b>	0,010	1,0101	0,9694	2,0281	2,0777	1,0447	1,0995	2,0482
<b>9fn5</b>	0,030	1,0305	0,8613	1,9517	2,022	0,9182	1,0132	1,9867
<b>9fn6</b>	0,010	1,0101	0,7543	1,8793	1,9663	0,8573	0,9649	1,9508
<b>9fn7</b>	0,002	1,0020	0,8169	1,9303	2,0053	0,8922	0,9925	1,9758
	es7	es8	es9	es10	es11	es12	es13	es14
<b>FN</b>	2,0155	2,0365	0,8200	1,1368	1,8866	1,6360	0,1058	11,5852
<b>3fn</b>	2,0363	2,0777	0,7827	1,1091	0,9694	1,7350	0,0649	12,4778
<b>4fn1</b>	2,1075	1,1135	0,7126	1,0516	1,8611	0,3920	-0,0502	12,4888
<b>4fn2</b>	2,0167	0,9784	0,5489	0,8753	1,5667	-0,2167	-0,3241	12,1995
<b>9fn1</b>	1,9367	1,9424	0,5369	0,9561	1,7925	1,5232	-0,1227	12,4906
<b>9fn2</b>	2,0315	2,0591	0,8257	1,1203	1,9092	1,6702	0,0909	12,3793
<b>9fn3</b>	2,0032	2,0199	0,7349	1,0294	1,8258	0,3920	-0,0503	12,8351
<b>9fn4</b>	2,0225	2,0482	0,7924	1,0995	1,8983	1,6569	0,0649	12,4778
<b>9fn5</b>	1,9337	1,9199	0,5919	0,8865	1,5481	-0,1959	-0,3073	12,4789
<b>9fn6</b>	1,9040	1,8840	0,5128	0,8382	1,5122	-0,2442	-0,3682	12,4964
<b>9fn7</b>	1,9247	1,9090	0,5586	0,8657	1,5372	-0,2167	-0,3332	12,5773

\*esi es el índice electrotopológico para el carbono etiquetado como i, según la figura 3.

pero con el uso de los E-State menos correlacionados esta cantidad disminuye a un total de 91 modelos posibles (anexo I). Con base en esto se construyeron, por permutación de los siete descriptores mencionados arriba,

21 modelos de correlación lineal de cinco descriptores, 35 modelos de cuatro y 35 modelos con tres descriptores del índice del estado electrotopológico.

Tabla 6. Matriz de correlación entre los 14 descriptores electrotopológicos.

	es1	es2	es3	es4	es5	es6	es7	es8	es9	es10	es11	es12	es13	es14
es1	1,00	-0,07	0,26	0,51	0,52	0,63	0,75	-0,46	0,44	0,45	-0,22	0,19	0,38	-0,54
es2	-0,07	1,00	0,94	0,71	0,72	0,66	0,33	0,45	0,73	0,55	0,25	0,27	0,51	0,19
es3	0,26	0,94	1,00	0,85	0,86	0,85	0,58	0,27	0,84	0,67	0,17	0,31	0,62	0,04
es4	0,51	0,71	0,85	1,00	1,00	0,90	0,73	0,29	0,99	0,94	0,23	0,67	0,90	-0,32
es5	0,52	0,72	0,86	1,00	1,00	0,90	0,73	0,27	0,99	0,92	0,22	0,64	0,88	-0,34
es6	0,63	0,66	0,85	0,90	0,90	1,00	0,92	-0,09	0,88	0,81	0,20	0,43	0,75	-0,13
es7	0,75	0,33	0,58	0,73	0,73	0,92	1,00	-0,38	0,71	0,73	0,20	0,41	0,68	-0,18
es8	-0,46	0,45	0,27	0,29	0,27	-0,09	-0,38	1,00	0,35	0,32	-0,06	0,49	0,37	0,08
es9	0,44	0,73	0,84	0,99	0,99	0,88	0,71	0,35	1,00	0,96	0,24	0,73	0,93	-0,27
es10	0,45	0,55	0,67	0,94	0,92	0,81	0,73	0,32	0,96	1,00	0,26	0,86	0,99	-0,30
es11	-0,22	0,25	0,17	0,23	0,22	0,20	0,20	-0,06	0,24	0,26	1,00	0,16	0,28	-0,17
es12	0,19	0,27	0,31	0,67	0,64	0,43	0,41	0,49	0,73	0,86	0,16	1,00	0,90	-0,31
es13	0,38	0,51	0,62	0,90	0,88	0,75	0,68	0,37	0,93	0,99	0,28	0,90	1,00	-0,27
es14	-0,54	0,19	0,04	-0,32	-0,34	-0,13	-0,18	0,08	-0,27	-0,30	-0,17	-0,31	-0,27	1,00

Para elaborar los modelos se usó el método de regresión lineal múltiple donde la variable dependiente fue el rendimiento cuántico de producción de oxígeno singlete y las variables independientes fueron los descriptores es1, es2, es8, es11, es12, es13 y es14. Los resultados conseguidos para el parámetro estadístico de coeficiente de correlación de los mejores de estos modelos, están en el rango de 0,850 a 0,960. Si bien estos valores son significativos, en todos los modelos se presenta el problema de sobreestimación de los valores [27], ya que los residuales tienen mayor magnitud que algunos valores de la propiedad experimental. La tabla 8 muestra los valores de uno de los modelos obtenidos (M22, anexo I) para el que el valor del residual para las cuatro estructuras moleculares 4fn2, 9fn4, 9fn6 y 9fn7

tiene un mayor valor que  $\Phi_{\Delta}$ . Para solucionar este inconveniente se transformó la magnitud de la propiedad, por medio de un operador matemático [5,23], que se escogió dependiendo de los resultados de la tendencia de la propiedad experimental para las once moléculas y de la correlación lineal con los valores originales reportados por Flors y Nonelli [15]. En la tabla 9 se observa que el mejor ajuste se da para un operador exponencial.

Tabla 7. Porcentaje de varianza explicada por los componentes principales CP.

CP	Valor propio	% Varianza	% Varianza acumulada
1	8,24	58,86	58,86
2	2,67	19,05	77,91
3	1,35	9,62	87,54
4	1,13	8,05	95,58
5	0,61	4,36	99,94
6	0,01	0,04	99,98
7	~0,00	0,01	99,99
8	~0,00	~0,00	100,00

Tabla 8. Residuales para el modelo M22 con los descriptores es1 es2 es8 y es11.

Molécula	Observado	Calculado	Residual
FN	1,0000	1,1346	-0,1346
3fn	0,9400	0,8417	0,0983
4fn1	0,7200	0,4321	0,2879
4fn2	0,1200	0,3679	-0,2479
9fn1	0,5100	0,3708	0,1392
9fn2	0,0800	0,0782	0,0018
9fn3	0,0800	0,0520	0,0280
9fn4	0,0100	0,0924	-0,0824
9fn5	0,0300	0,0105	0,0195
9fn6	0,0100	0,0968	-0,0868
9fn7	0,0020	0,0250	-0,0230

Los resultados que eliminaron la sobre-estimación de los valores, fueron, en todos los casos, los obtenidos por uso del operador Euler (**exp**), lo que confirma los resultados mostrados en la tabla 9. Un resultado adicional fue

que la mayoría de estos modelos mejoró la “calidad” de los parámetros estadísticos R, S y F, y la sobre-estimación de los valores de la propiedad desapareció. Los máximos valores porcentuales mostrados para el residual con respecto al exponencial de la propiedad fueron del 31%.

Tabla 9. Tendencia de los valores del rendimiento cuántico de producción de oxígeno singulete.

Tendencia de $\Phi_{\Delta}$	R <sup>2</sup>	$\Phi_{\Delta}$ vs operador( $\Phi_{\Delta}$ )	R <sup>2</sup>
Exponencial	0,958	Exponencial	0,995
Polinómica	0,957	Raíz 2	0,981
Logarítmica	0,908	Logarítmica	0,857
Lineal	0,808	Lineal	-----

Tabla 10. Residuales para modelos de regresión con cuatro descriptores electrotopológicos. Los modelos se etiquetaron como M seguido de un número que representa el orden en que se elaboraron.

Molécula	exp( $\Phi_{\Delta}$ )	Residual				
		M1	M2	M22	M23	M24
FN	2,7183	-0,0361	-0,1136	-0,1565	0,0849	0,0575
3fen	2,5600	0,0232	0,0812	0,1132	-0,1142	-0,0654
4fen1	2,0544	0,3721	0,4191	0,4240	0,4573	0,3728
4fen2	1,1275	-0,3554	-0,3818	-0,3757	-0,4218	-0,3612
9fen1	1,6653	0,0331	0,1111	0,1554	0,1279	0,0241
9fen2	1,0833	0,0951	-0,0690	-0,0320	-0,0878	-0,1028
9fen3	1,0833	0,0263	0,0482	0,0203	0,0995	0,0330
9fen4	1,0101	-0,1848	-0,1595	-0,1209	-0,1779	-0,1918
9fen5	1,0305	0,1387	0,0833	0,0517	0,0689	0,1441
9fen6	1,0101	-0,0149	-0,0566	-0,0866	-0,0611	-0,0098
9fen7	1,0020	0,0932	0,0376	0,0071	0,0243	0,0993

En la tabla 10 se observan los residuales para cinco modelos de regresión lineal, de estos, dos corresponden a los construidos con cinco descriptores electrotopológicos, **M1**: es1, es2, es8, es11, es12; **M2**: es1, es2, es8, es11, es13; y 3 modelos con cuatro descriptores, **M22**: es1, es2, es8, es11; **M23**: es1, es2, es8, es12; y **M24**: es1, es2, es8, es13.

Ningún dato es mayor en magnitud, que el valor correspondiente al  $\Phi_{\Delta}$  transformado por el operador matemático, por el contrario la mayoría de los

residuales está en el segundo orden de cifra decimal. Para la búsqueda del mejor modelo de regresión lineal se eligió el criterio de la menor cantidad de descriptores E-State.

De la comparación de los parámetros R, S y F entre los 70 modelos construidos con el mínimo de descriptores, es decir con tres y cuatro de los E-State menos correlacionados, se eligió el de mayor coeficiente de correlación R, menor desviación estandar S y mayor parámetro de Fisher F. En la tabla 11 se observan algunos de estos resultados, donde se incluye un modelo con cinco descriptores E-State, el modelo M1, para mostrar que estos modelos generan altos coeficientes de correlación.

La ecuaciones de regresión para los modelos M1, M22 y M57 fueron:

$$\begin{aligned} \exp(\Phi_{\Delta}) = & 0,6477*(es1)\pm(0,1458) - 8,9435*(es2)\pm(1,1238) + 1,3542*(es8)\pm(0,1854) \\ & - 1,1201*(es11)\pm(0,2544) + 7,0437*(es12)\pm(1,3602) + 17,2026\pm(2,4220) \end{aligned}$$

*(Ecuación 26)*

$$\begin{aligned} \exp(\Phi_{\Delta}) = & 1,4220*(es1)\pm(0,1955) - 4,6629*(es2)\pm(1,1289) + 1,3709*(es8)\pm(0,1854) \\ & + 0,5108*(es11)\pm(0,3309) + 5,4323\pm(2,1093) \end{aligned}$$

*(Ecuación 27)*

$$\begin{aligned} \exp(\Phi_{\Delta}) = & 1,3153*(es1)\pm(0,2000) - 3,9001*(es2)\pm(1,3024) + 1,2137*(es8)\pm(0,2988) \\ & + 5,2233\pm(2,3034) \end{aligned}$$

*(Ecuación 28)*

Para validar los mejores modelos obtenidos (de la tabla 11 por cada modelo M1,M22,...,M38 se generan once nuevos modelos, para un total de 143 modelos de validación) se usó la metodología LOO (leave-one-out, dejar uno afuera) [12-14]. Con las ecuaciones de regresión obtenidas en cada uno de

los modelos se calculó el valor de la propiedad experimental para cada molécula que se retiró del conjunto inicial, para después realizar la regresión lineal para cada conjunto de datos; obteniéndose los mejores coeficientes de validación cruzada Q en un rango de 0,60 a 0,66. La figura 5A muestra la dispersión de la correlación entre los valores experimentales y los calculados por la ecuación de regresión (ecuación 27), para uno de los 35 modelos de cuatro descriptores electrotopológicos. Las estructuras representadas por los E-State es1, es2, es8 y es11, que más se dispersan en la correlación lineal para la validación cruzada, son las correspondientes a las moléculas 4fn1, 4fn2, 9fn1 y 9fn3. En general, las estructuras que se dispersarán en la mayoría de los modelos de validación cruzada son las de las moléculas denominadas como 4fn1, 4fn2, y 9fn1.

Tabla 11. Parámetros estadísticos para modelos con cuatro descriptores E-State.

<b>Modelo</b>	<b>R</b>	<b>F</b>	<b>S</b>
M23	0,958	16,894	0,244
M22	0,952	14,557	0,261
M1	0,960**	11,875**	0,261**
M27	0,943	12,116	0,283
M57	0,932*	15,540*	0,286*
M30	0,934	10,166	0,306
M26	0,922	8,505	0,331
M47	0,913	7,509	0,348
M29	0,896	6,108	0,379
M53	0,889	5,656	0,391
M49	0,885	5,402	0,398
M51	0,864	4,430	0,429
M38	0,856	4,118	0,441

\*Parámetros para modelo de con 3 índices E-State.

\*\* Parámetros para el modelo con 5 índices E-State.

Al construir nuevos modelos donde se excluyan dos de estas “estructuras”, los resultados para los modelos de correlación con los mismos descriptores mejoran mucho en los parámetros estadísticos R, S y F. Además cuando se desarrolló los modelos de validación cruzada, el coeficiente Q, se elevó desde 0,70 a 0,98, que por ser muy cercanos a la unidad indican que los

modelos de correlación QSAR/QSPR poseen una excelente capacidad predictiva del  $\Phi_{\Delta}$  [5,13,33], como se puede ver en la figura 5B y la tabla 11 donde el coeficiente Q toma el valor de 0,859, claro está si pasamos por alto el hecho de haber retirado las dos moléculas 4fn2 y 9fn1.

Estos resultados muestran que los modelos con descriptores cuasi-ortogonales son predictivos para nueve de las once moléculas del conjunto de derivados de la 1H-Fenalen-1-ona, y que el principal aporte en las correlación lo realizan los descriptores es1, es8, es11 y es12.

Tabla 12. Residuales validación cruzada modelo M22, sin 9fn1 ni 4fn2.

	Observado	Predicho	Residual
<b>FN</b>	2,7183	2,4180	0,3003
<b>3fn</b>	2,5600	2,4640	0,0959
<b>4fn1</b>	2,0544	2,0975	-0,0431
<b>9fn2</b>	1,0833	1,0654	0,0179
<b>9fn3</b>	1,0833	1,0251	0,0582
<b>9fn5</b>	1,0101	1,0283	-0,0182
<b>9fn4</b>	1,0305	1,0837	-0,0533
<b>9fn6</b>	1,0101	0,9509	0,0591
<b>9fn7</b>	1,0020	1,0352	-0,0332
		Q	<b>0,859</b>

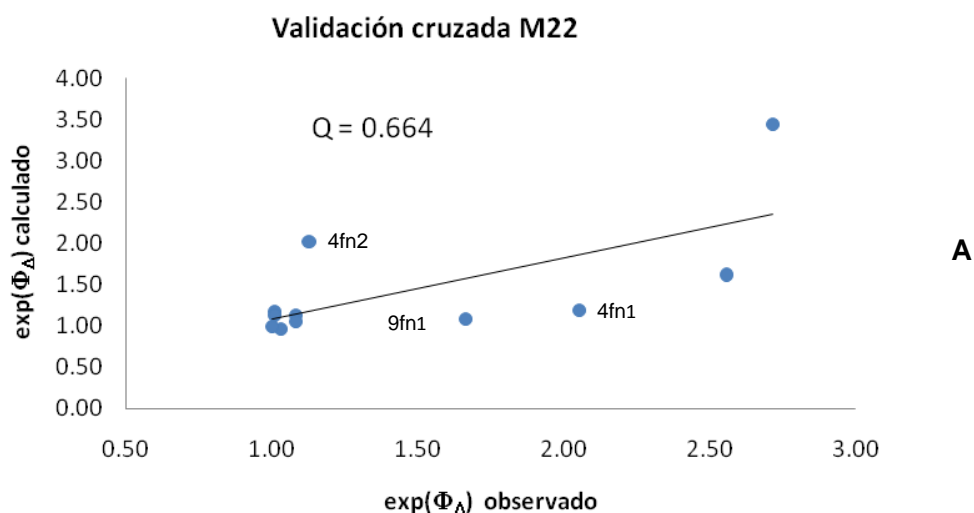


Figura 5A. Validación cruzada para el modelo M22.

Los resultados obtenidos hasta esta parte del trabajo, indican que aquellos descriptores empleados en los modelos con mejores parámetros estadísticos representan al exponencial de la propiedad experimental  $\exp(\Phi_{\Delta})$ . Estos índices codifican el estado intrínseco de los carbonos aromáticos de la estructura común FN y la perturbación inmediata que experimentan por todos los átomos restantes, en mayor medida por el sustituyente fenil [4], siendo el carbono aromático etiquetado como 1, equivalente a la posición de sustitución 9 de las 9-Fenilfenalen-1-onas quien está más correlacionado con el rendimiento cuántico de producción de oxígeno singulete.

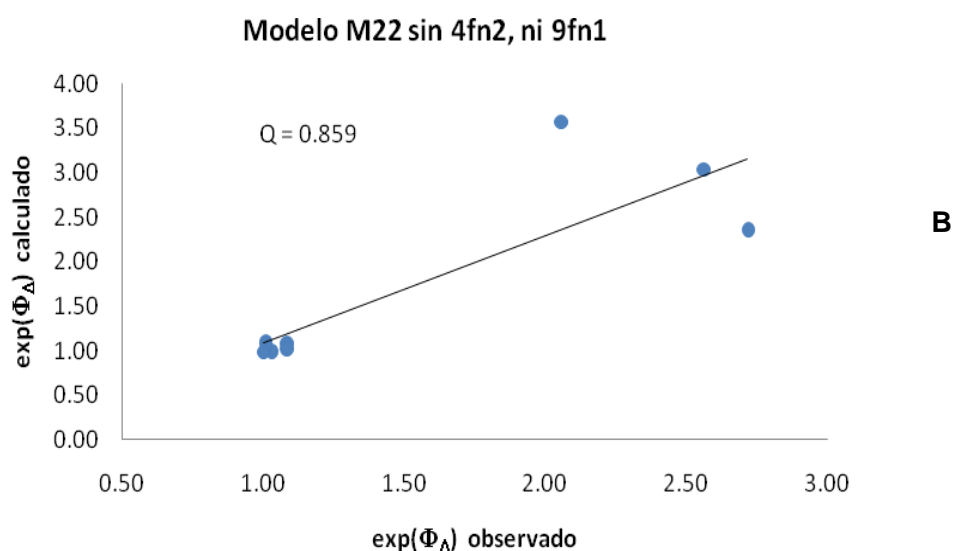


Figura 5B. Validación Cruzada para el modelo M22 sin 4fn2 ni 9fn1.

#### 4.2 MODELOS CON DESCRIPTORES ORTOGONALIZADOS

Se construyó la base 14-dimensional con los E-State de la estructura común FN (tabla 13) [4], la cual presentó valores muy pequeños para los últimos descriptores ortogonalizados, debido a esto se decidió reducir la dimensión de la base, sumado a la generación de otras bases con orden diferente de ortogonalización [11,13,20-21], con el objetivo de incluir aquellos descriptores

que no se han utilizado. Se construyeron cinco bases ortonormales, de diferentes dimensiones, con algunos de los 14 E-State (tabla 14) para poder utilizarlos en la elaboración de una segunda clase de modelos QSAR/QSPR.

Tabla 13. Base 14 dimensional con todos los valores de los descriptores E-State despues del proceso de ortogonalización.

ES1	DESCRIPTORES ORTOGONALIZADOS*								ES10	ES12-
	ES2	ES3	ES4	ES5	ES6	ES7	ES8	ES9	ES11	ES14
1,8866	0,0021	-0,0057	0,0231	0,0019	0,0008	-0,0004	0,0145	0,0001	0,0000	0,0000
1,8983	0,0092	0,0007	0,0109	-0,0030	-0,0129	0,0000	0,0020	-0,0001	0,0000	0,0000
1,8700	-0,0104	0,0045	0,0147	0,0022	0,0035	-0,0009	0,0278	-0,0004	0,0000	0,0000
1,7591	-0,0900	0,0005	-0,0477	-0,0011	0,0085	0,0013	-0,0426	0,0004	0,0000	0,0000
0,6033	-0,1477	0,0006	0,0665	-0,0001	-0,0003	0,0002	-0,0025	0,0003	0,0000	0,0000
1,0138	0,1134	0,0003	0,0239	0,0003	0,0017	0,0017	-0,0682	-0,0006	0,0001	0,0000
0,9520	0,0845	0,0006	-0,0069	-0,0004	0,0031	-0,0012	0,0773	0,0000	0,0000	0,0000
0,9694	0,0917	-0,0002	0,0173	-0,0018	0,0064	-0,0002	-0,0117	0,0008	-0,0001	0,0000
0,8613	0,0142	0,0004	-0,0313	0,0026	-0,0083	0,0009	-0,0346	0,0000	-0,0001	0,0000
0,7543	-0,0593	-0,0017	-0,0331	-0,0012	0,0015	-0,0007	0,0174	-0,0016	0,0000	0,0000
0,8169	-0,0077	0,0000	-0,0374	0,0007	-0,0041	-0,0008	0,0205	0,0012	0,0001	0,0000

\* Los descriptores fueron ortogonalizados siguiendo el orden según la etiqueta establecida para cada uno de los índices del estado electrotopológico.

Tabla 14. Bases de diferentes dimensiones y la cantidad de modelos posibles (C) con cuatro descriptores E-State.

BASE	DESCRIPTORES ORTOGONALIZADOS								C
1	ES1	ES2	ES3	ES4	ES5	ES6	ES7	ES8	70
2	ES1	ES3	ES5	ES7	ES9	ES10	ES11	ES14	70
3	ES1	ES8	ES2	ES11	ES6	ES12	ES13		35
4	ES1	ES5	ES9	ES10	ES13	ES12			15
5	ES1	ES2	ES6	ES7	ES3				5

TOTAL 195

\*El simbolo "ES" pertenece al descriptor ortogonalizado proveniente del correspondiente descriptor "es".

De igual forma que en la parte anterior, las posibilidades de construcción de los modelos de correlación entre el  $\Phi_{\Delta}$  y los descriptores E-State ortogonalizados, presentes en las bases de ocho, siete, seis y cinco, es extensa (ecuación 25), dando un total de 512 modelos posibles [13]. Para solucionar este problema se eligió los modelos con cuatro descriptores junto al uso de la estrategia de inclusión de a pasos [12-13,27], comenzando con

la elección del descriptor de mayor correlación con la propiedad, (el ES1, según la tabla 15), seguido de un proceso iterativo entre los demás descriptores y el residual, incluyendo el de mejor correlación en el modelo [11-13], todo esto disminuyó el costo computacional y permitió obtener 195 modelos (anexo II), de los cuales cinco mostraron los mejores parámetros estadísticos.

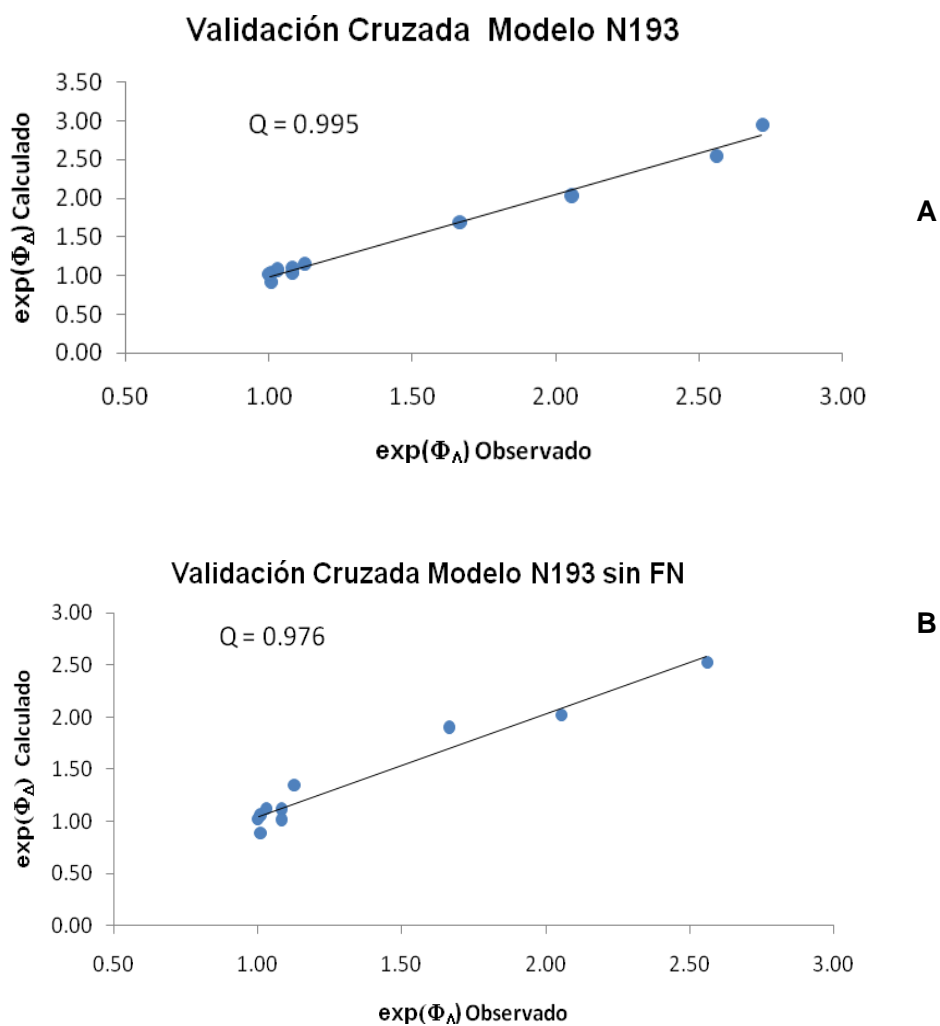
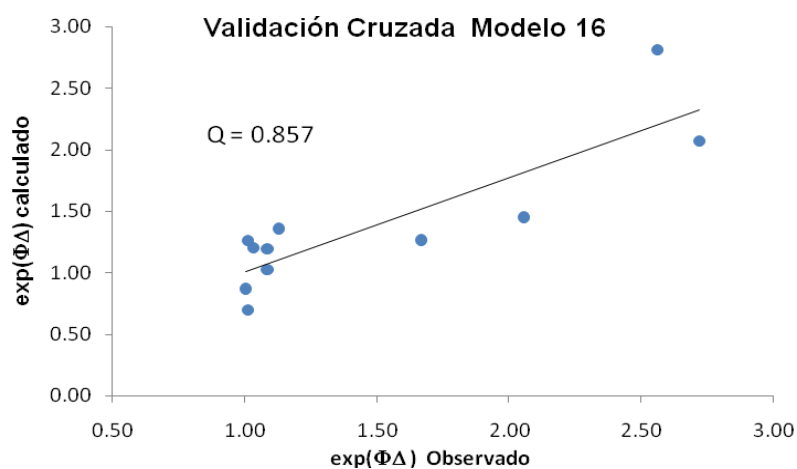
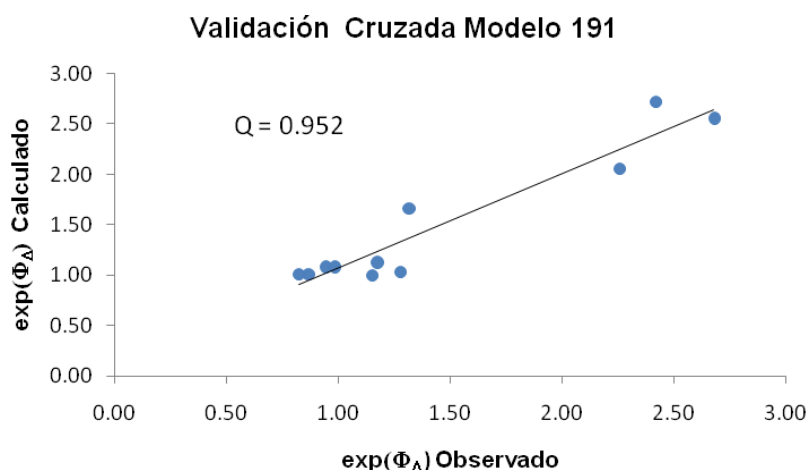


Figura 6. A. Validación cruzada para el modelo N193. B. Validación cruzada para el modelo N193 sin la molécula FN.



C



D

Figura 6. C. Validación cruzada para el modelo N16. D Validación cruzada para el modelo N191.

Tabla 15. Correlaciones de los descriptores E-State de la base 1 con  $\exp(\Phi_{\Delta})$ .

Descriptor	ES1	ES2	ES3	ES4	ES5	ES6	ES7	ES8
Correlación con $\exp(\Phi_{\Delta})$	0,73	-0,16	-0,22	0,48	0,06	-0,33	-0,20	-0,19

En la tabla 16 se muestran en orden descendente los mejores modelos con sus respectivos parámetros, donde resaltan los coeficientes de validación cruzada Q, para los modelos etiquetados como N193 y N191. Los otros modelos aunque posean un buen ajuste lineal, no sirven para la predicción

del  $\Phi_{\Delta}$  debido a que sus coeficientes de validación se alejan del límite de 0,80 propuesto por diversos estudios QSAR/QSPR [5,12].

Las gráficas de validación cruzada obtenidas para los tres mejores modelos, mostradas en las figuras 6A, 6C y 6D, revelan un agrupamiento para las moléculas sustituidas por el fenil en la posición 9, también muestran diferentes grados de dispersión para las once “estructuras” y valores predichos por debajo (deficiencia) o por encima (suficiencia) de la línea de tendencia.

Por otra parte destaca que en el modelo N193 el grado de dispersión es mínimo, mostrando valores que no se alejan demasiado del ajuste lineal. La gráfica de los residuales en orden de aparición del modelo N193 (figura 7) muestran una tendencia de bloque, y una distribución en forma aleatoria, indicando que la hipótesis estadística sobre el ajuste lineal de los datos por este modelo es correcta [27].

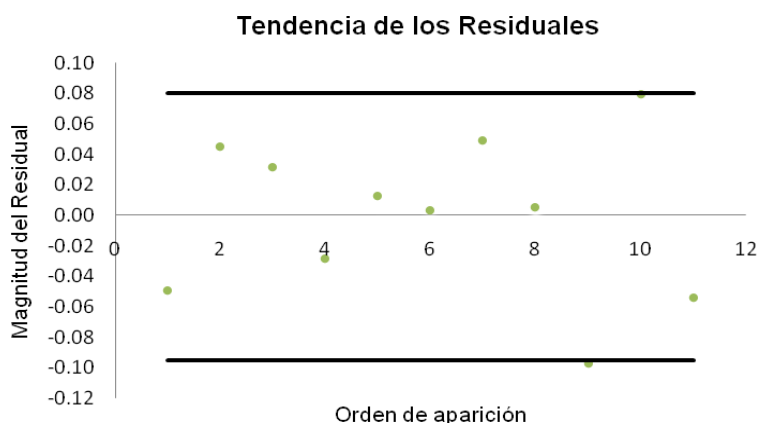


Figura 7. Ploteo de los residuales en su orden de aparición.

Para comprobar el poder predictivo de este último modelo se realizó una prueba más, en la que se eliminó la molécula patrón 1H-Fenalen-1-ona, se construyó el correspondiente modelo con las diez restantes, se calculó el

coeficiente de validación cruzada Q, cuyo valor fue de 0,976, y se concluyó que este modelo mantiene un buen ajuste lineal [5].

La figura 6B muestra estos resultados, que al compararlos gráficamente con los del modelo de la figura 6A construido con once moléculas, permite observar claras diferencias que indican una disminución en la precisión del modelo. Por ejemplo hay mayor dispersión de los datos y no se conserva la agrupación para todas las 9-Fenilfenalenonas, lo que se debe a que la calibración del modelo, es decir las características estructurales codificadas por los descriptores ortogonales ES1, ES2, ES7 y ES3 que son utilizadas en la regresión lineal multivariada, son óptimas para un modelo con las once estructuras.

Tabla 16. Parámetros para los cinco mejores modelos con descriptores ortogonales E-State.

MODELO	R	S	F	Q
N193*	0,997	0,067	235,93	0,995
N191	0,987	0,145	47,70	0,952
N141	0,952	0,261	14,55	0,692
N144	0,933	0,305	10,22	0,632
N16	0,925	0,355	5,95	0,857

\*Estos modelos se etiquetaron con la letra N seguido del número que indica el orden de construcción, ver anexo II.

### 4.3 MODELO FINAL

Obtenidos diferentes modelos con las dos clases de descriptores, los cuasi-ortogonales y los ortogonalizados, se escogió el mejor tanto en sus parámetros de ajuste como en su coeficiente de validación cruzado Q, correspondiendo a el nombrado como N193. La ecuación de regresión para este modelo es:

$$\exp(\Phi_{\Delta}) = 0,9340*(ES1) \pm (0,0414) - 1,3731*(ES2) \pm (0,2719) - 125,0474*(ES7) \pm (6,3247) - 53,5728*(ES3) \pm (12,2268) + 0,3493 \pm (0,0544)$$

*(Ecuación 29)*

Debido a que el aporte de cada pareja descriptor/coeficiente de regresión es aditivo-sustractivo, de la ecuación 29 y de los porcentajes entre los valores de cada pareja, se concluye que la mayor contribución se da por parte del ES1, solo en seis casos el aporte del ES7 esta en el rango del orden porcentual del 21-81% con respecto al anterior. La obtención de los descriptores ES1, ES2, ES3 y ES7 indican claramente que los átomos de estas posiciones y su perturbación por parte del entorno topológico y electrónico del resto de la estructura para cada molécula, estan relacionados con el efecto de disminución de la propiedad  $\Phi_{\Delta}$  observada para los derivados de la 1H-Fenalen-1-ona. Es pertinente aclarar que “no se puede extraer desde la hipótesis QSAR/QSPR información sobre los mecanismos de reacción, frecuencias de colisión, procesos de difusión, ni procesos mecanocuánticos como relajación de estados excitados o tiempo de vida media [18-19,35-41]”, lo que si se puede afirmar es que en esta región molecular codificada por estos cuatro descriptores E-State, los átomos de carbono por un proceso dependiente de la disponibilidad electrónica, afectan el proceso por el que todas las moléculas generan  $O_2(^1\Delta_g)$ . La figura 8 muestra la ubicación de los descriptores E-State con y sin ortogonalizar y sus respectivos rangos.

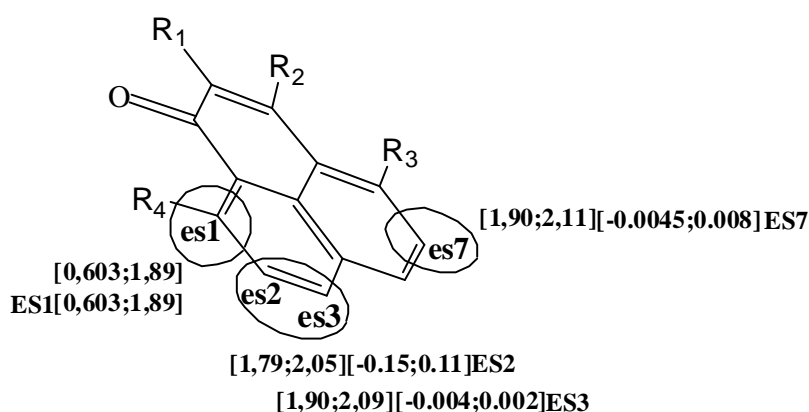


Figura 8. Descriptores que representan las características estructurales relacionadas con la propiedad  $\Phi_{\Delta}$ .

### 4.3.1 Conjunto de prueba

Una vez elegido este modelo como el más representativo de los realizados por uso del E-State, se procedió a retirar dos moléculas del conjunto para hacer una validación del modelo, tipo “conjunto de prueba” [13]. Si bien hay que considerar que el modelo es válido con la totalidad de las moléculas, retirar dos de estas con características similares al resto de la población, y luego probar si el modelo es capaz de predecir  $\exp(\Phi_{\Delta})$  para estas, constituyó una prueba útil de la capacidad predictiva para moléculas con características semejantes a las usadas en la calibración de la ecuación de regresión [4-5,13]. La ecuación correspondiente al modelo N193' fue:

$$\exp(\Phi_{\Delta}) = 0,9220*(ES1)\pm(0,0556) - 1,3710*(ES2)\pm(0,3712) - 125,0530*(ES7)\pm(7,5544) - 56,2730*(ES3)\pm(15,8591) + 0,3590\pm(0,0705)$$

(Ecuación 30)

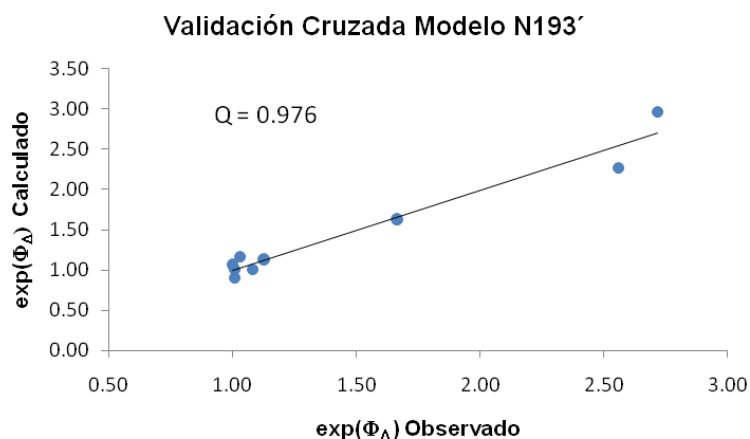


Figura 9. Resultado de la validación cruzada para el modelo construido con 9 moléculas, no se incluyó la 4fn1 ni la 9fn2.

La figura 9 muestra el ajuste entre valores calculados y observados para el  $\exp(\Phi_{\Delta})$ , la tabla 17 muestra los resultados para los dos modelos desarrollados por uso de los descriptores del modelo N193, pero con once y

nueve moléculas. Los porcentajes de error mostrados allí son del orden de 10%, semejantes a los obtenidos para los datos experimentales. Cuando se realiza la transformación de los datos por aplicación del inverso del operador Euler, el modelo se muestra predictivo para los datos de  $\Phi_{\Delta}$  superiores de la tabla 1.

Tabla 17. Datos obtenidos para el modelo N193'.

Molécula	EXP(FD)			Residual		% error	
	Observado	Calculado	Predicho	Calculado	Predicho	Calculado	Predicho
FN	2,7183	2,7671	2,9491	-0,0488	-0,2308	-1,79	-8,49
3fn	2,5600	2,5148	2,5496	0,0451	0,0104	1,76	0,40
4fn1	2,0544		<b>2,0348*</b>		<b>0,0196</b>		<b>0,95</b>
4fn2	1,1275	1,1557	1,1574	-0,0282	-0,0299	-2,50	-2,65
9fn1	1,6653	1,6523	1,6925	0,0129	-0,0272	0,78	-1,63
9fn2	1,0833		<b>1,1053*</b>		<b>-0,0220</b>		<b>-2,03</b>
9fn3	1,0833	1,0336	1,0321	0,0497	0,0512	4,58	4,72
9fn4	1,0101	1,0046	1,0384	0,0054	-0,0283	0,54	-2,80
9fn5	1,0305	1,1277	1,0795	-0,0973	-0,0490	-9,44	-4,76
9fn6	1,0101	0,9304	0,9219	0,0796	0,0882	7,88	8,73
9fn7	1,0020	1,0561	1,0147	-0,0541	-0,0127	-5,4	-1,27

\*Los números en negrilla son valores predichos por la ecuación 30 en forma interpolativa.

#### 4.3.2 Hiper-espacio E-State

Se trabajó con un sub-espacio E-State 4D,  $V(es1,es2,es3,es7)$ , en el que se desplegó las once estructuras [4]. Además se comparó por medio de la métrica euclidiana [4,34], la semejanza entre las estructuras vistas como vectores cuyos componentes son los índices del estado electrotopológico utilizados en el modelo N193. La distribución de las moléculas en las cuatro coordenadas de  $V$  muestra que el descriptor **es1** -cuyo valor del estado intrínseco como carbono aromático es de 1,666- para las moléculas FN, 3fn, 4fn1 y 4fn2 es perturbado aditivamente por el entorno electrotopológico, mientras que para las moléculas 9fn1, 9fn2, 9fn3, 9fn4, 9fn5, 9fn6 y 9fn7 el efecto neto es sustractivo [23]. Esto indica que este carbono tiene más disponibilidad electrónica en el primer grupo.

Un análisis de los descriptores *es1* para el grupo 2, revela que el carbono que es menos accesible topológicamente y de menor disponibilidad electrónica es el que pertenece a la molécula 9fn1 [4,23]. Los descriptores *es2* y *es3* para las once moléculas muestran un efecto aditivo neto del entorno electrotopológico para todas la moléculas, presentando el de menor magnitud para la 9fn1. El *es7* muestra efecto neto aditivo, además de dos grupos bien diferenciados. La figura 10 muestra estos resultados gráficamente.

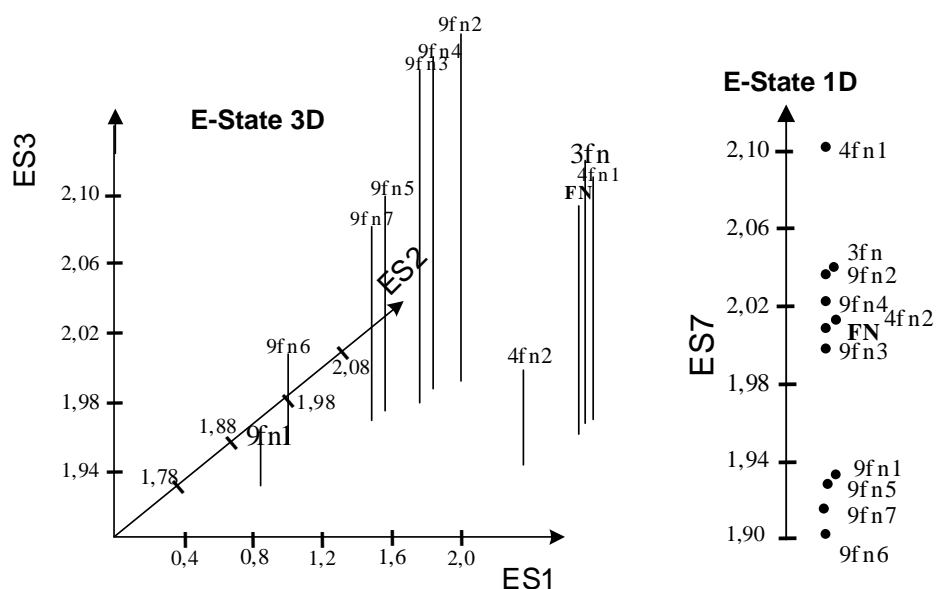


Figura 10. Gráfica por trazas 3D y 1D para el hiperspacio 4D E-State compuesto por cuatro descriptores del estado electrotopológico y nombrado como  $V(es1, es2, es3, es7)$ .

La comparación por medio de la métrica euclidiana mostró una distribución en dos grupos para la 1H-Fenalen-1-ona y sus diez derivados; el grupo uno, de cuatro moléculas correspondió a FN, 3fn, 4fn1 y 4fn2, el grupo dos con siete moléculas, compuesto por todas aquellas con posición de sustitución 9 para el fenil. La figura 11 muestra una marcada separación de estos dos grupos. Las estructuras de esta figura y los valores de las tablas 18 y 19, muestran rasgos interesantes, por ejemplo los derivados de la 1H-Fenalen-

1ona sustituidos en la posición 9, que tienen dos sustituyentes, 9fn6 y 9fn7, aparecen separadas de las que poseen uno, 9fn1-9fn5; la 9fn1 con el sustituyente nitro se ve aparte del resto, esto sugiere un comportamiento diferente a las demás por causa de este grupo. Otro resultado importante es que la distancia es altamente análoga para la mayoría de los casos, a la tendencia que muestra la propiedad experimental y aun más análoga al exp de la propiedad  $\Phi_{\Delta}$ .

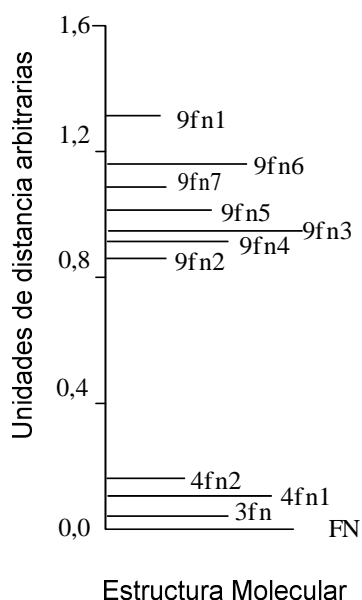


Figura 11. Distribución de las moléculas según distancia euclidiana.

Tabla 18. Agrupación uno para las Fenalen-1-onas.

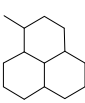
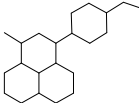
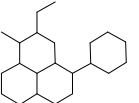
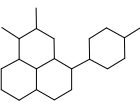
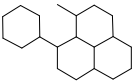
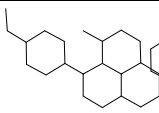
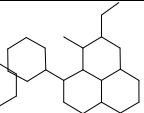
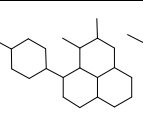
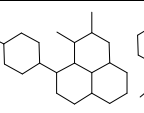
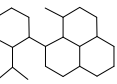
Etiqueta	FN	3fn	4fn1	4fn2
Grafo				
Distancia	0,0	0,0275	0,0943	0,1679

Tabla 19. Agrupación dos para las Fenalen-1-onas

Etiqueta	9fn2	9fn4	9fn3	9fn5	9fn7	9fn6	9fn1
Grafo							
Distancia	0,8831	0,9235	0,9399	1,0289	1,0740	1,1410	1,2998

Por último se afirma que el conjunto de la propiedad rendimiento cuántico de producción de oxígeno singulete, cuyas componentes son las once estructuras moleculares FN, 3fn, 4fn1, 4fn2, 9fn1, 9fn2, 9fn3, 9fn4, 9fn5, 9fn6 y 9fn7, es representado por el conjunto de los índices del estado electrotopológico, es1, es2, es3, es7, mediante la ecuación haciendo a ambos conjuntos numéricos superponibles; así en esta investigación se desarrolló la hipótesis de que la estructura es la responsable de las propiedades.

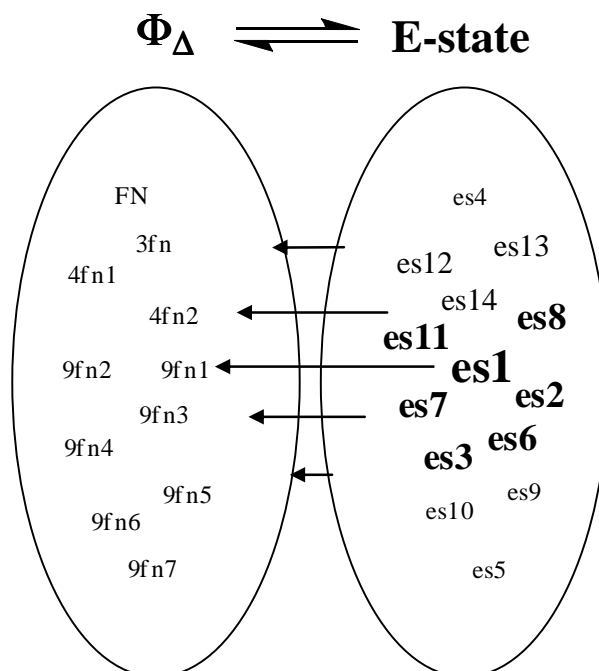


Figura 12. Conjuntos del espacio químico y del espacio de los descriptores moleculares.

La figura 12 muestra el resultado obtenido, cinco de los 14 descriptores E-State son capaces de representar la propiedad rendimiento cuántico de producción de oxígeno singulete, la flecha de mayor magnitud indica que el descriptor *es1* es el que posee mayor correlación con la propiedad [3-4].

## 5. CONCLUSIONES

Se obtuvo un modelo con altos parámetros estadísticos de coeficiente de correlación R, parámetro de Fisher F, desviación estandar S y coeficiente de validación cruzada Q, 0,997; 235,926; 0,067 y 0,995 respectivamente; indicando que la ecuaciones construidas por regresión lineal son útiles en la predicción del rendimiento cuántico de producción de oxígeno singulete para el conjunto de derivados de la 1H-Fenalen-1-ona.

El modelo encontrado incluye los descriptores ortogonalizados en el orden, es1, es2, es7 y es3, correspondientes a los carbonos aromáticos de la estructura común de la 1H-Fenalen-1-ona en la inmediata vecindad de la posición 9 de las cetonas poliaromáticas; esto nos permite concluir que la caída en magnitud de la propiedad esta relacionada con la accesibilidad electrónica de estos fragmentos moleculares. Observamos que para todos los modelos, el descriptor es1 fue el de mayor correlación con la propiedad. La magnitud del es1 esta agrupada en los rangos, [0,6 a 1,1] y [1,7 a 1,89]. Esto indica que para valores superiores a 1,6666, (el estado intrínseco de un carbono aromático)<sup>2</sup> el efecto del entorno es aditivo, lo que indica mejor accesibilidad topológica y mayor disponibilidad electrónica. Para valores inferiores todo lo contrario.

A pesar de que las medidas experimentales de esta propiedad claramente dependen de la concentración de oxígeno, del solvente, de los procesos de decaimiento de los estados excitados; y bajo las consideraciones de que las

medidas fueron realizadas por el mismo grupo de investigación, en el mismo solvente y de que todas las moléculas son derivadas de la 1H-Fenalen-1-ona; se encontró una correlación entre la caída del rendimiento cuántico de producción de oxígeno singulete y la variabilidad estructural de los sustituyentes sobre la estructura común. La clave de esta relación está en la cuantificación de las estructuras moleculares por medio de los índices del estado electrotopológico y la correlación desarrollada en los modelos QSPR construidos por regresión lineal multivariada.

La distancia topológica en un espacio 4D E-State hallada para las once estructuras, revela dos grupos definidos para el conjunto de moléculas, estando de acuerdo en líneas generales con la tendencia de la propiedad  $\Phi_{\Delta}$ .

Existen otros modelos con buenos parámetros estadísticos que permiten ver la contribución de los E-State en los modelos de correlación. Los descriptores es8, es11 y es12 empleados en estos modelos indican una dependencia marcada de la propiedad para con las posiciones de sustitución cuyo sustituyente es el grupo fenil.

Para los valores pequeños de la propiedad el modelo deja de ser predictivo. Este comportamiento es observado en varios trabajos QSAR/QSPR [4,9] donde se usa la técnica matemática de regresión lineal multivariable, y no es una estimación de que el E-State no sirve para modelar la propiedad en estudio.

## 6. RECOMENDACIONES

Para ampliar la capacidad predictiva del modelo, se sugiere usar mayor cantidad de moléculas donde se presente más variabilidad en las posiciones de sustitución y en el tipo de sustituyente.

Para una mayor generalidad del modelo QSPR, se sugiere usar descriptores que codifiquen el efecto del solvente, y a las moléculas como un todo. Se sugiere el empleo de descriptores mecano-cuánticos.

## BIBLIOGRAFÍA

- [1] DEL RE, G. Ontological Status of Molecular Structure. HYLE-International Journal for Philosophy of Chemistry. (1998), 4, 81.
- [2] HAWKINS, D.; BASAK, S.; SHI, X. M. QSAR with Few Compounds and Many Features. Journal of Chemical Information and Computer Science. (2001), 663, 41.
- [3] RANDIĆ, M. Chemical Structure: What Is "She"? Journal of Chemical Education. (1992), 69, 713.
- [4] KIER, L. B.; HALL, L. H. Molecular Structure Description: The Electrotopological State. Academic Press, San Diego, California, (1999).
- [5] BENIGNI, R. Quantitative Structure-Activity Relationship Models of Mutagens and Carcinogens. CRC PRESS, Boca Ratón, USA, cap 7 (2003), 208.
- [6] SELASSIE, C. D. Burger's Medicinal Chemistry and Drug Discovery. History Of Quantitative Structure-Activity Relationships. Jonh Wiley & Sons Inc, cap 1 (2003), 3.
- [7] HANSCH, C.; FUJITA, T.  $\rho$ - $\sigma$ - $\pi$  analysis. A Method for the Correlation of Biological Activity and Chemical Structure. Journal American Chemical Society. (1964), 86, 1616.
- [8] MIHALIĆ, Z.; TRINAJSTIĆ, N.; A Graph-Theoretical Approach to Structure-Property Relationships. Journal of Chemical Education. (1992), 69, 701.

- [9] BADER, R. Atoms in Molecules: A Quantum Theory. Clarendon Press. Oxford. International Series of Monographs on Chemistry. Cap 3 (2003), 53.
- [10] ZHOU, Z.; DAI, Q.; GU, T. A QSAR Model of PAHs Carcinogenesis Based on Thermodynamic Stabilities of Biactive Sites. Journal of Chemical Information and Computer Science. (2003), 43, 615.
- [11] RANDIĆ, M. Resolution of Ambiguities in Structure-Property Studies by Use of Orthogonal Descriptors. Journal of Chemical Information and Computer Science. (1991), 31, 311.
- [12] LEACH, A. R.; GILLET, V. J. An Introduction of Chemoinformatics. Springer, cap 3 (2007), 53.
- [13] FERNADEZ, F. M.; CASTRO, E. A.; DUCHOWICZ, P. R. Los descriptores ortogonales en la teoría QSAR-QSPR. Afinidad. (2004), 61, 476.
- [14] NÚÑEZ, M. B.; MAGUNA, F. P.; OKULIK, N.; CASTRO, E. A. Predicción de la actividad inhibitoria MAO A de derivados de xantonas. Universidad nacional del nordeste. Comunicaciones científicas y tecnológicas (2004). Consultado en: [www.unne.edu.ar/Web/cyt/com2004/8-Exactas/E-065.pdf](http://www.unne.edu.ar/Web/cyt/com2004/8-Exactas/E-065.pdf)
- [15] FLORS, C.; NONELLI, S. Light and Singlet Oxygen in Plant Defense Against Pathogens: Phototoxic Phenalenone Phytoalexins. Accounts of Chemical Research. (2006), 39, 293.
- [16] SCHMIDT, R.; TANIELIAN, C.; DUNSBACH, R.; WOLFF, C. Phenalenone, a Universal Reference Compound for the Determination of Quantum Yields of Singlet Oxygen ( $^1\Delta_g$ ) Sensitization. Journal of Photochemistry and Photobiology A. (1994), 79, 11.
- [17] LAZZARO, A.; COROMINAS, M.; MARTÍ, C.; FLORS, C.; IZQUIERDO, L. R.; GRILLO, T. A.; LUIS J. G.; NONELL S. Light- and Singlet Oxygen-

mediated Antifungal Activity of Phenylphenalenone Phytoalexins. *Photochemical & Photobiological Sciences*. (2004), 3, 706.

[18] WILKINSON, F.; HELMAN, W. P.; ROSS, A. B. Quantum Yields for the Photosensitized Formation of the Lowest Electronically Excited Singlet State of Molecular Oxygen in Solution. *Journal of Physical and Chemical Reference Data*. (1993), 22, 113.

[19] FLORS, C.; OGILBY, P. R.; LUIS, J. G.; GRILLO, T. A.; IZQUIERDO L. R.; GENTILI, P.; BUSSOTTI, L.; NONELL, S. Phototoxic Phytoalexins. Processes that Compete with the Photosensitized Production of Singlet Oxygen by 9-Phenylphenalenones. *Photochemistry and Photobiology*. (2006), 82, 95.

[20] RANDIĆ, M. Generalized Molecular Descriptors. *Journal of Mathematical Chemistry*. (1991), 7, 155.

[21] ESTRADA, E. Generalization of Topological Indices. *Chemical Physics Letters*. (2001), 336, 248.

[22] GARCIA, V.; MOLINS, J.; FORNS, A. Determinación de los índices topológicos de mayor influencia en la estructura molecular de los hidrocarburos acíclicos saturados. *Afinidad*. (2002), 59, 504.

[23] HALL, L.; MONHNEY, B. J. The Electrotological State: Structure Information at the Atomic Level for Molecular Graphs. *Journal of Chemical Information and Computer Science*. (1991), 31, 76.

[24] CONTRERA, J.; MATTHEWS, E.; BENZ, D. Predicting the Carcinogenic Potential of Pharmaceuticals in Rodents Using Molecular Structural Similarity and E-State Indices. *Regulatory Toxicology and Pharmacology*. (2003), 38, 243.

- [25] HUUSKONEN, J. QSAR Modeling with the Electrotopological State Indices: Predicting the Toxicity of Organic Chemicals. *Chemosphere*. (2003), 50, 949.
- [26] CHAKRABORTY, S.; SENGUPTA, C. ROY, K. Exploring QSAR with E-State Index: Selectivity Requirements for COX-2 Versus COX-1 Binding of Terphenyl Methyl Sulfones and Sulfonamides. *Bioorganic & Medicinal Chemistry Letters*. (2004), 14, 4665.
- [27] DRAPER, N. R.; SMITH, H. *Applied Regression Analysis*. John Wiley & Sons, New York, cap 5, 6 y 7 (1981), 218.
- [28] JOHNSON, D. E. *Métodos multivariados aplicados al análisis de datos*. International Thomson editores, México, (1988), 1.
- [29] PLA, L. E. *Análisis Multivariado: método de componentes principales*. Monografías Serie matemáticas. Secretaria general de la organización de los estados americanos, Washington, cap 5 (1986), 93.
- [30] TORANZOS, A. *Introducción a la teoría de grafos*. Secretaria general de la organización de los estados americanos, Washington, Cap 1 (1976), 7.
- [31] SEYBOLD, P. G.; MAY, M.; BAGAL, U. A. Molecular Structure-Property Relationships. *Journal of Chemical Education*. (1987), 64, 575.
- [32] GUTMAN, I.; MILUN, M.; TRINAJSTIĆ, N. Graph Theory and Molecular Orbitals. 19. Nonparametric Resonance Energies of Arbitrary Conjugated Systems. *Journal of the American Chemical Society*. (1977), 99, 1692.
- [33] IVANCIUC, O.; TARAVIRAS, S. L. CABROL-BASS, D. Quasi-ortogonal Basis Sets of Molecular Graph Descriptors as a Chemical Diversity. *Journal of Chemical Information and Computer Science*. (2000), 40, 126.

- [34] RESTREPO, G.; MESA, H.; LLANOS, E. J.; VILLAVECES, J. L. Topological Study of the Periodic System. *Journal of Chemical Information and Computer Science*. (2004), 44, 68.
- [35] SCHWEITZER, C.; SCHMIDT, R. Physical Mechanisms of Generation and Deactivation of Singlet Oxygen. *Chemical Reviews*. (2003), 103, 1685.
- [36] DEROSA, C.; CRUTCHLEY, J. Photosensitized Singlet Oxygen and its Applications. *Coordination Chemistry Reviews*. (2002), 233, 351.
- [37] ANDO, Y.; NIWA, K.; YAMADA, N.; ENOMOTO, T.; IRIE, T.; KUBOTA, H.; OHMIYA, Y.; AKIYAMA, H. Firefly Bioluminescence Quantum Yield and Colour Change by pH-sensitive Green Emission. *Nature Photonics letter 2* (2007), 44.
- [38] KOLAROVA, H.; BAJGAR, R.; NEVRELOVA P.; MOSINGER, P. Comparison of Sensitizers by Detecting Reactive Oxygen Species after Photodynamic Reaction in Vitro. *Toxicology in Vitro*. (2007), 21, 1287.
- [39] ARNBJERG, J.; PATERSON, M. J.; NIELSEN, C. B.; JORGENSEN, M.; CHRISTIANSEN, O.; OGILBY, P. R. One- and Two-Photon Photosensitized Singlet Oxygen Production: Characterization of Aromatic Ketones as Sensitizer Standards. *Journal of Physical Chemistry*. (2007), 111, 5576.
- [40] EID, M.; FANCHAOUY, M.; GURNY R.; DELIE F.; VARGAS, A. In Vivo Photodynamic Activity of Photosensitizer-loaded Nanoparticles: Formulation Properties, Administration Parameters and Biological Issues Involved in PDT Outcome. *European Journal of Pharmaceutics and Biopharmaceutics*. (2007).
- [41] MARTI, C.; JÜRGENS, O.; CUENCA, O.; CASALS, O. M.; NONELL, S. Aromatic Ketones as Standards for Singlet Molecular Oxygen  $O_2(^1\Delta_g)$  photosensitization. Time Resolved Photoacoustic and Near-IR Emission Studies. *Photochemistry and Photobiology*. (1996), 97, 11.

[42] GAUSSVIEW 3.0, Gaussian, Inc, Pittsburgh PA, **2003**.

[43] MOLCONNZ, Versión 4.10\_WIN2K, EDUSOFT **2003**. Kier y Hall

[44] STATISTICA, versión 6.0, STATSOFT 2001.

# ANEXOS

## ANEXO I

Nomenclatura de los modelos construidos con tres, cuatro y cinco descriptores cuasi-ortogonales.

Modelos con 5 E-State		Modelos con 4 E-State		Modelos con 3 E-State	
<b>M1</b>	es1 es2 es8 es11 es12	<b>M22</b>	es1 es2 es8 es11	<b>M57</b>	es1 es2 es8
<b>M2</b>	es1 es2 es8 es11 es13	<b>M23</b>	es1 es2 es8 es12	<b>M58</b>	es1 es2 es11
<b>M3</b>	es1 es2 es8 es11 es14	<b>M24</b>	es1 es2 es8 es13	<b>M59</b>	es1 es2 es12
<b>M4</b>	es1 es2 es8 es12 es13	<b>M25</b>	es1 es2 es8 es14	<b>M60</b>	es1 es2 es13
<b>M5</b>	es1 es2 es8 es12 es14	<b>M26</b>	es1 es2 es11 es12	<b>M61</b>	es1 es2 es14
<b>M6</b>	es1 es2 es8 es13 es14	<b>M27</b>	es1 es2 es11 es13	<b>M62</b>	es1 es8 es11
<b>M7</b>	es1 es2 es11 es12 es13	<b>M28</b>	es1 es2 es11 es14	<b>M63</b>	es1 es8 es12
<b>M8</b>	es1 es2 es11 es12 es14	<b>M29</b>	es1 es2 es12 es13	<b>M64</b>	es1 es8 es13
<b>M9</b>	es1 es2 es11 es13 es14	<b>M30</b>	es1 es2 es12 es14	<b>M65</b>	es1 es8 es14
<b>M10</b>	es1 es2 es12 es13 es14	<b>M31</b>	es1 es2 es13 es14	<b>M66</b>	es1 es11 es12
<b>M11</b>	es1 es8 es11 es12 es13	<b>M32</b>	es1 es8 es11 es12	<b>M67</b>	es1 es11 es13
<b>M12</b>	es1 es8 es11 es12 es14	<b>M33</b>	es1 es8 es11 es13	<b>M68</b>	es1 es11 es14
<b>M13</b>	es1 es8 es11 es13 es14	<b>M34</b>	es1 es8 es11 es14	<b>M69</b>	es1 es12 es13
<b>M14</b>	es1 es8 es12 es13 es14	<b>M35</b>	es1 es8 es12 es13	<b>M70</b>	es1 es12 es14
<b>M15</b>	es1 es11 es12 es13 es14	<b>M36</b>	es1 es8 es12 es14	<b>M71</b>	es1 es13 es14
<b>M16</b>	es2 es8 es11 es12 es13	<b>M37</b>	es1 es8 es13 es14	<b>M72</b>	es2 es8 es11
<b>M17</b>	es2 es8 es11 es12 es14	<b>M38</b>	es1 es11 es12 es13	<b>M73</b>	es2 es8 es12
<b>M18</b>	es2 es8 es11 es13 es14	<b>M39</b>	es1 es11 es12 es14	<b>M74</b>	es2 es8 es13
<b>M19</b>	es2 es8 es12 es13 es14	<b>M40</b>	es1 es11 es13 es14	<b>M75</b>	es2 es8 es14
<b>M20</b>	es2 es11 es12 es13 es14	<b>M41</b>	es1 es12 es13 es14	<b>M76</b>	es2 es11 es12
<b>M21</b>	es8 es11 es12 es13 es14	<b>M42</b>	es2 es8 es11 es12	<b>M77</b>	es2 es11 es13
		<b>M43</b>	es2 es8 es11 es13	<b>M78</b>	es2 es11 es14
		<b>M44</b>	es2 es8 es11 es14	<b>M79</b>	es2 es12 es13
		<b>M45</b>	es2 es8 es12 es13	<b>M80</b>	es2 es12 es14
		<b>M46</b>	es2 es8 es12 es14	<b>M81</b>	es2 es13 es14
		<b>M47</b>	es2 es8 es13 es14	<b>M82</b>	es8 es11 es12
		<b>M48</b>	es2 es11 es12 es13	<b>M83</b>	es8 es11 es13
		<b>M49</b>	es2 es11 es12 es14	<b>M84</b>	es8 es11 es14
		<b>M50</b>	es2 es11 es13 es14	<b>M85</b>	es8 es12 es13
		<b>M51</b>	es2 es12 es13 es14	<b>M86</b>	es8 es12 es14
		<b>M52</b>	es8 es11 es12 es13	<b>M87</b>	es8 es13 es14
		<b>M53</b>	es8 es11 es12 es14	<b>M88</b>	es11 es12 es13
		<b>M54</b>	es8 es11 es13 es14	<b>M89</b>	es11 es12 es14
		<b>M55</b>	es8 es12 es13 es14	<b>M90</b>	es11 es13 es14
		<b>M56</b>	es11 es12 es13 es14	<b>M91</b>	es12 es13 es14

## ANEXO II

Nomenclatura de los modelos construidos con cuatro descriptores ortogonales provenientes de cinco bases diferentes.

	BASE1		BASE1		BASE 2
<b>N1</b>	ES1 ES2 ES3 ES4	<b>N36</b>	ES2 ES3 ES4 ES5	<b>N71</b>	ES1 ES3 ES5 ES7
<b>N2</b>	ES1 ES2 ES3 ES5	<b>N37</b>	ES2 ES3 ES4 ES6	<b>N72</b>	ES1 ES3 ES5 ES9
<b>N3</b>	ES1 ES2 ES3 ES6	<b>N38</b>	ES2 ES3 ES4 ES7	<b>N73</b>	ES1 ES3 ES5 ES10
<b>N4</b>	ES1 ES2 ES3 ES7	<b>N39</b>	ES2 ES3 ES4 ES8	<b>N74</b>	ES1 ES3 ES5 ES11
<b>N5</b>	ES1 ES2 ES3 ES8	<b>N40</b>	ES2 ES3 ES5 ES6	<b>N75</b>	ES1 ES3 ES5 ES14
<b>N6</b>	ES1 ES2 ES4 ES5	<b>N41</b>	ES2 ES3 ES5 ES7	<b>N76</b>	ES1 ES3 ES7 ES9
<b>N7</b>	ES1 ES2 ES4 ES6	<b>N42</b>	ES2 ES3 ES5 ES8	<b>N77</b>	ES1 ES3 ES7 ES10
<b>N8</b>	ES1 ES2 ES4 ES7	<b>N43</b>	ES2 ES3 ES6 ES7	<b>N78</b>	ES1 ES3 ES7 ES11
<b>N9</b>	ES1 ES2 ES4 ES8	<b>N44</b>	ES2 ES3 ES6 ES8	<b>N79</b>	ES1 ES3 ES7 ES14
<b>N10</b>	ES1 ES2 ES5 ES6	<b>N45</b>	ES2 ES3 ES7 ES8	<b>N80</b>	ES1 ES3 ES9 ES10
<b>N11</b>	ES1 ES2 ES5 ES7	<b>N46</b>	ES2 ES4 ES5 ES6	<b>N81</b>	ES1 ES3 ES9 ES11
<b>N12</b>	ES1 ES2 ES5 ES8	<b>N47</b>	ES2 ES4 ES5 ES7	<b>N82</b>	ES1 ES3 ES9 ES14
<b>N13</b>	ES1 ES2 ES6 ES7	<b>N48</b>	ES2 ES4 ES5 ES8	<b>N83</b>	ES1 ES3 ES10 ES11
<b>N14</b>	ES1 ES2 ES6 ES8	<b>N49</b>	ES2 ES4 ES6 ES7	<b>N84</b>	ES1 ES3 ES10 ES14
<b>N15</b>	ES1 ES2 ES7 ES8	<b>N50</b>	ES2 ES4 ES6 ES8	<b>N85</b>	ES1 ES3 ES11 ES14
<b>N16</b>	ES1 ES3 ES4 ES5	<b>N51</b>	ES2 ES4 ES7 ES8	<b>N86</b>	ES1 ES5 ES7 ES9
<b>N17</b>	ES1 ES3 ES4 ES6	<b>N52</b>	ES2 ES5 ES6 ES7	<b>N87</b>	ES1 ES5 ES7 ES10
<b>N18</b>	ES1 ES3 ES4 ES7	<b>N53</b>	ES2 ES5 ES6 ES8	<b>N88</b>	ES1 ES5 ES7 ES11
<b>N19</b>	ES1 ES3 ES4 ES8	<b>N54</b>	ES2 ES5 ES7 ES8	<b>N89</b>	ES1 ES5 ES7 ES14
<b>N20</b>	ES1 ES3 ES5 ES6	<b>N55</b>	ES2 ES6 ES7 ES8	<b>N90</b>	ES1 ES5 ES9 ES10
<b>N21</b>	ES1 ES3 ES5 ES7	<b>N56</b>	ES3 ES4 ES5 ES6	<b>N91</b>	ES1 ES5 ES9 ES11
<b>N22</b>	ES1 ES3 ES5 ES8	<b>N57</b>	ES3 ES4 ES5 ES7	<b>N92</b>	ES1 ES5 ES9 ES14
<b>N23</b>	ES1 ES3 ES6 ES7	<b>N58</b>	ES3 ES4 ES5 ES8	<b>N93</b>	ES1 ES5 ES10 ES11
<b>N24</b>	ES1 ES3 ES6 ES8	<b>N59</b>	ES3 ES4 ES6 ES7	<b>N94</b>	ES1 ES5 ES10 ES14
<b>N25</b>	ES1 ES3 ES7 ES8	<b>N60</b>	ES3 ES4 ES6 ES8	<b>N95</b>	ES1 ES5 ES11 ES14
<b>N26</b>	ES1 ES4 ES5 ES6	<b>N61</b>	ES3 ES4 ES7 ES8	<b>N96</b>	ES1 ES7 ES9 ES10
<b>N27</b>	ES1 ES4 ES5 ES7	<b>N62</b>	ES3 ES5 ES6 ES7	<b>N97</b>	ES1 ES7 ES9 ES11
<b>N28</b>	ES1 ES4 ES5 ES8	<b>N63</b>	ES3 ES5 ES6 ES8	<b>N98</b>	ES1 ES7 ES9 ES14
<b>N29</b>	ES1 ES4 ES6 ES7	<b>N64</b>	ES3 ES5 ES7 ES8	<b>N99</b>	ES1 ES7 ES10 ES11
<b>N30</b>	ES1 ES4 ES6 ES8	<b>N65</b>	ES3 ES6 ES7 ES8	<b>N100</b>	ES1 ES7 ES10 ES14
<b>N31</b>	ES1 ES4 ES7 ES8	<b>N66</b>	ES4 ES5 ES6 ES7	<b>N101</b>	ES1 ES7 ES11 ES14
<b>N32</b>	ES1 ES5 ES6 ES7	<b>N67</b>	ES4 ES5 ES6 ES8	<b>N102</b>	ES1 ES9 ES10 ES11
<b>N33</b>	ES1 ES5 ES6 ES8	<b>N68</b>	ES4 ES5 ES7 ES8	<b>N103</b>	ES1 ES9 ES10 ES14
<b>N34</b>	ES1 ES5 ES7 ES8	<b>N69</b>	ES4 ES6 ES7 ES8	<b>N104</b>	ES1 ES9 ES11 ES14
<b>N35</b>	ES1 ES6 ES7 ES8	<b>N70</b>	ES5 ES6 ES7 ES8	<b>N105</b>	ES1 ES10 ES11 ES14

### ANEXO III

Nomenclatura de los modelos construidos con cuatro descriptores ortogonales provenientes de cinco bases diferentes.

	BASE 2		BASE 3		BASE 4
<b>N106</b>	ES3 ES5 ES7 ES9	<b>N141</b>	ES1 ES8 ES2 ES11	<b>N176</b>	ES1 ES5 ES9 ES10
<b>N107</b>	ES3 ES5 ES7 ES10	<b>N142</b>	ES1 ES8 ES2 ES6	<b>N177</b>	ES1 ES5 ES9 ES13
<b>N108</b>	ES3 ES5 ES7 ES11	<b>N143</b>	ES1 ES8 ES2 ES13	<b>N178</b>	ES1 ES5 ES9 ES12
<b>N109</b>	ES3 ES5 ES7 ES14	<b>N144</b>	ES1 ES8 ES2 ES12	<b>N179</b>	ES1 ES5 ES10 ES13
<b>N110</b>	ES3 ES5 ES9 ES10	<b>N145</b>	ES1 ES8 ES11 ES6	<b>N180</b>	ES1 ES5 ES10 ES12
<b>N111</b>	ES3 ES5 ES9 ES11	<b>N146</b>	ES1 ES8 ES11 ES13	<b>N181</b>	ES1 ES5 ES13 ES12
<b>N112</b>	ES3 ES5 ES9 ES14	<b>N147</b>	ES1 ES8 ES11 ES12	<b>N182</b>	ES1 ES9 ES10 ES13
<b>N113</b>	ES3 ES5 ES10 ES11	<b>N148</b>	ES1 ES8 ES6 ES13	<b>N183</b>	ES1 ES9 ES10 ES12
<b>N114</b>	ES3 ES5 ES10 ES14	<b>N149</b>	ES1 ES8 ES6 ES12	<b>N184</b>	ES1 ES9 ES13 ES12
<b>N115</b>	ES3 ES5 ES11 ES14	<b>N150</b>	ES1 ES8 ES13 ES12	<b>N185</b>	ES1 ES10 ES13 ES12
<b>N116</b>	ES3 ES7 ES9 ES10	<b>N151</b>	ES1 ES2 ES11 ES6	<b>N186</b>	ES5 ES9 ES10 ES13
<b>N117</b>	ES3 ES7 ES9 ES11	<b>N152</b>	ES1 ES2 ES11 ES13	<b>N187</b>	ES5 ES9 ES10 ES12
<b>N118</b>	ES3 ES7 ES9 ES14	<b>N153</b>	ES1 ES2 ES11 ES12	<b>N188</b>	ES5 ES9 ES13 ES12
<b>N119</b>	ES3 ES7 ES10 ES11	<b>N154</b>	ES1 ES2 ES6 ES13	<b>N189</b>	ES5 ES10 ES13 ES12
<b>N120</b>	ES3 ES7 ES10 ES14	<b>N155</b>	ES1 ES2 ES6 ES12	<b>N190</b>	ES9 ES10 ES13 ES12
<b>N121</b>	ES3 ES7 ES11 ES14	<b>N156</b>	ES1 ES2 ES13 ES12		
<b>N122</b>	ES3 ES9 ES10 ES11	<b>N157</b>	ES1 ES11 ES6 ES13		BASE 5
<b>N123</b>	ES3 ES9 ES10 ES14	<b>N158</b>	ES1 ES11 ES6 ES12		
<b>N124</b>	ES3 ES9 ES11 ES14	<b>N159</b>	ES1 ES11 ES13 ES12	<b>N191</b>	ES1 ES2 ES6 ES7
<b>N125</b>	ES3 ES10 ES11 ES14	<b>N160</b>	ES1 ES6 ES13 ES12	<b>N192</b>	ES1 ES2 ES6 ES3
<b>N126</b>	ES5 ES7 ES9 ES10	<b>N161</b>	ES8 ES2 ES11 ES6	<b>N193</b>	ES1 ES2 ES7 ES3
<b>N127</b>	ES5 ES7 ES9 ES11	<b>N162</b>	ES8 ES2 ES11 ES13	<b>N194</b>	ES1 ES6 ES7 ES3
<b>N128</b>	ES5 ES7 ES9 ES14	<b>N163</b>	ES8 ES2 ES11 ES12	<b>N195</b>	ES2 ES6 ES7 ES3
<b>N129</b>	ES5 ES7 ES10 ES11	<b>N164</b>	ES8 ES2 ES6 ES13		
<b>N130</b>	ES5 ES7 ES10 ES14	<b>N165</b>	ES8 ES2 ES6 ES12		
<b>N131</b>	ES5 ES7 ES11 ES14	<b>N166</b>	ES8 ES2 ES13 ES12		
<b>N132</b>	ES5 ES9 ES10 ES11	<b>N167</b>	ES8 ES11 ES6 ES13		
<b>N133</b>	ES5 ES9 ES10 ES14	<b>N168</b>	ES8 ES11 ES6 ES12		
<b>N134</b>	ES5 ES9 ES11 ES14	<b>N169</b>	ES8 ES11 ES13 ES12		
<b>N135</b>	ES5 ES10 ES11 ES14	<b>N170</b>	ES8 ES6 ES13 ES12		
<b>N136</b>	ES7 ES9 ES10 ES11	<b>N171</b>	ES2 ES11 ES6 ES13		
<b>N137</b>	ES7 ES9 ES10 ES14	<b>N172</b>	ES2 ES11 ES6 ES12		
<b>N138</b>	ES7 ES9 ES11 ES14	<b>N173</b>	ES2 ES11 ES13 ES12		
<b>N139</b>	ES7 ES10 ES11 ES14	<b>N174</b>	ES2 ES6 ES13 ES12		
<b>N140</b>	ES9 ES10 ES11 ES14	<b>N175</b>	ES11 ES6 ES13 ES12		