



Acta Universitaria

ISSN: 0188-6266

actauniversitaria@ugto.mx

Universidad de Guanajuato

México

Villanueva, Manuel; Robles, Juvencio; Martínez-Richa, Antonio; Villaseñor, Tayde O.; Reyna Hernández, María

Obtención de Relaciones Cuantitativas Estructura-Propiedad en Cristales Líquidos

Acta Universitaria, vol. 19, núm. 2, septiembre, 2009, pp. 82-92

Universidad de Guanajuato

Guanajuato, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=41612893013>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Obtención de Relaciones Cuantitativas Estructura-Propiedad en Cristales Líquidos

Manuel Villanueva*, Juvencio Robles**, Antonio Martínez-Richa*, Tayde O. Villaseñor*** y María Reyna Hernández***.

RESUMEN

Una serie de 69 cristales líquidos termotrópicos se sometió a una colección de Métodos Quimiométricos incluyendo Análisis de Componentes Principales; Algoritmos Genéticos y Relaciones Cuantitativas Estructura Propiedad para la obtención de modelos matemáticos que permitan predecir las Temperaturas de Transición Nemática, (T_N). En este trabajo se aplica por primera vez el Análisis de Componentes Principales al estudio de cristales líquidos para la identificación de subconjuntos de moléculas para los cuales se obtuvieron modelos locales. Los resultados obtenidos muestran que la capacidad de predicción de los modelos locales es superior a la del modelo global, lo cual resultó consistente con la hipótesis planteada. La metodología utilizada se extenderá al estudio de cristales líquidos con actividad farmacológica.

ABSTRACT

Sixty nine thermotropic liquid crystal molecules were analyzed by different chemometric methods, including Principal Component Analysis (PCA), Genetic algorithms and Quantitative Structure-Property Relationships (QSPR). Mathematical models to predict nematic transition temperatures (T_N) were derived. This is the first time that Principal Component Analysis was used to predict liquid crystal properties, leading to the derivation of local (specific) models for different molecule sets. Results indicate that local models have higher prediction capabilities than global models, which is consistent to our initial assumption. This methodology will be used in the study of liquid crystals used in pharmaceutical applications.

Recibido: 11 de Junio de 2009
Aceptado: 29 de Julio de 2009

INTRODUCCIÓN

Cristales líquidos

Los cristales líquidos fueron descubiertos hace más de cien años por Friedrich Reinitzer, un botánico austriaco (Reinitzer, 1888) quien encontró que algunos compuestos orgánicos, al pasar de la fase sólida a la fase líquida, exhibían dos cambios o transiciones de fase intermedias sucesivas. En 1889, O. Lehmann (Lehmann, 1889) descubrió que el líquido de apariencia turbia, intermedio entre el cristal y el líquido transparente poseía propiedades ópticas parecidas a las de un cristal sólido pero con la viscosidad más parecida a un líquido y acuñó el nombre de *cristal líquido*, (CL). Las diversas fases cristal líquido son fases intermedias (mesofases) entre las fases sólida y líquida, por lo que el término cristal líquido (Friedel, 1922), denota un estado de agregación cuyas propiedades son intermedias entre las del sólido cristalino y el líquido amorfico. Las transiciones a estos estados intermedios pueden ser producidas por procesos puramente térmicos (*mesomorfismo termotrópico*) o por la influencia de disolventes (*mesomorfismo liotrópico*) y desde este punto de vista los CL's se clasifican como Termotrópicos y Liotrópicos y de manera muy general las fases cristal líquido se clasifican en dos tipos principales: fases *nemáticas* y *esmécticas* (Chandrasekhar, 1992) existiendo también otras fases de gran interés tecnológico como las fases columnares o fases mas complejas como las colestéricas. Existen muchos factores que determinan la presencia de

Palabras clave:

Relaciones cuantitativas estructura propiedad; QSPR; Cristales líquidos; Temperatura de transición nemática; Componentes principales.

Keywords:

Quantitative structure-property relations; QSPR; Liquid crystals; Nematic transition temperature; Principal components.

* Departamento de Química, División de Ciencias Naturales y Exactas del Campus Guanajuato. Universidad de Guanajuato. Noria Alta s/n, Guanajuato, Gto. C.P. 36050, México, Teléfono / Fax: (+52) 473 732 0006 Ext. 8120. Correos electrónicos: villagm@quijote.ugto.mx y richa@quijote.ugto.mx.

** Departamento de Farmacia, División de Ciencias Naturales y Exactas del Campus Guanajuato. Universidad de Guanajuato. Noria Alta s/n, Guanajuato, Gto. C.P. 36050, México. Correo electrónico: roblesj@quijote.ugto.mx.

*** Alumnos de la Licenciatura en Química, División de Ciencias Naturales y Exactas del Campus Guanajuato. Universidad de Guanajuato. Noria Alta s/n, Guanajuato, Gto. C.P. 36050, México. Correos electrónicos: mechanical_teen@hotmail.com y reyna386@gmail.com.

una fase cristal líquido. Generalmente se reconoce que la aparición de las fases cristal líquido se deben principalmente a la forma geométrica de las moléculas, pero otros factores tales como repulsiones intermoleculares, interacciones de tipo atractivas y enlaces de hidrógeno también juegan un papel importante, por lo que la exploración de sus relaciones cuantitativas estructura-propiedad es un excelente tema de estudio. La aproximación molecular para la estimación de temperaturas de transición nemática no siempre es fácil ya que son altamente dependientes de la conformación de las moléculas. Mientras que la conformación de más baja energía se puede calcular para una molécula, ésta puede ser solamente aproximada para una molécula en la mesofase.

En las últimas tres décadas ha habido un gran avance en la síntesis y caracterización de cristales líquidos (Demus, D., 1998). Esto ha permitido tener un avance considerable en el entendimiento de las fases cristal líquido y en la obtención de los diagramas de fases que caracterizan a tales sistemas. La ocurrencia de diversas mesofases presentadas por un mismo tipo de material constituye un comportamiento que debe de ser explicado de manera teórica. Por otro lado, las propiedades de los cristales líquidos han tenido un gran impacto en muchas áreas de la ciencia así como en aplicaciones tecnológicas y, recientemente, varios investigadores han declarado que el campo está experimentando un renacimiento. Los avances de las últimas décadas han permitido a los dispositivos de cristal líquido penetrar en casi todos los campos de la óptica y están trascendiendo en el campo de la medicina de diferentes formas. De particular interés, observamos que en años recientes algunos CL's son usados como trasportadores de fármacos (Woltman, 2007) y como fármacos anticancerígenos (ScienceDaily, 2009), lo cual hace aún más atractivo el estudio de estos materiales.

Relaciones Cuantitativas Estructura-Propiedad

El diseño *in silico* de moléculas con propiedades específicas tanto químicas, fisicoquímicas, bioquímicas y estructurales, mediante el Modelado Molecular Asistido por Computadora, se ha convertido en una rama de la Química de un gran interés actual. La posibilidad de diseñar compuestos con propiedades bien definidas disminuyendo sustancialmente el alto costo de la síntesis experimental ha motivado un gran esfuerzo en el desarrollo de metodologías que permitan un diseño efectivo. Una forma de lograr esto es mediante las técnicas que permiten obtener *relaciones cuantitativas estructura-propiedad* (QSPR, por sus siglas en inglés), esto se logra en una forma racionalizada y sistematizada (Hansch, 1995). Las técnicas QSPR suponen

que existe una relación entre las propiedades de una molécula y su estructura y tratan de establecer una relación matemática simple que permita reproducir (y posteriormente predecir) una propiedad dada para un conjunto de compuestos bajo la suposición de que la propiedad observada es el resultado de la contribución de diversos factores, los cuales se comportan de manera independiente. Las relaciones QSPR se fundamentan matemáticamente mediante la aceptación de la existencia de una dependencia continua entre la propiedad experimentalmente medida y algunos factores estructurales intrínsecos de la molécula. Por simplicidad, se supone que la propiedad experimental depende de los factores estructurales linealmente en una cierta región del dominio.

Cada contribución a la propiedad se representa por un descriptor y los análisis tipo QSPR involucran la definición y obtención de descriptores físico-químicos que dependen de la estructura de la molécula, así como el tratamiento matemático-estadístico aplicado a los descriptores para que la propiedad de un conjunto de compuestos se ajuste a un modelo multilineal (Katritzky, 2001). La estrategia básica es encontrar la relación cuantitativa estructura-propiedad óptima que pueda ser usada para la predicción de propiedades de moléculas, incluyendo aquellas que todavía no han sido sintetizadas.

Los descriptores son parámetros que a través de valores numéricos representan una información estructural, electrónica (mecano cuántica), estérica, y fisicoquímica de las moléculas en estudio. La mayoría de los descriptores se obtienen por métodos computacionales y su número se ha incrementado en los últimos años. Desde este punto de vista, se debe seguir un proceso sistemático que permita la elección del conjunto de descriptores más significativos que puedan describir la propiedad en estudio. Uno de dichos procesos consiste en elegir aquellos descriptores que se piense están más relacionados con la manifestación de la propiedad de la sustancia. Esto requiere un conocimiento teórico acerca del fenómeno en estudio. Luego, se implementa algún otro proceso que nos permita una selección adicional de propiedades; de tal suerte que en el modelo o modelos finales se incluya el menor número de descriptores, siendo éstos los que tengan mayor sentido físico para describir al problema en cuestión. (Todeschini, 2000).

Componentes Principales

En el tratamiento computacional del diseño molecular mediante computadora (y específicamente en las metodologías QSPR), muy frecuentemente tenemos que

manejar un conjunto de objetos (moléculas) que se caracterizan mediante variables independientes (un conjunto de descriptores). En estos casos es necesaria la utilización de herramientas estadísticas de naturaleza multivariante y la exactitud de un modelo predictivo es una de las principales preocupaciones de la mayoría de los químicos computacionales interesados en la obtención de modelos QSPR. En la búsqueda de estos modelos se supone que el modelo que se obtenga a partir de moléculas análogas exhibirá un carácter predictivo mucho mejor que el modelo obtenido a partir de un gran conjunto de moléculas con diversidad estructural. Sin embargo, el gran número de moléculas y de descriptores dificultan el reconocimiento de semejanzas o diferencias entre las moléculas del conjunto, impidiéndonos reconocer grupos de moléculas semejantes. Entonces, ¿cómo podríamos visualizar grupos de objetos semejantes cuando están caracterizados por más de tres variables? La respuesta es: reduciendo la dimensionalidad, es decir, si logramos describir con precisión los valores de p variables por un pequeño subconjunto $r < p$ de ellas se habrá reducido la dimensionalidad del problema a costa de una pequeña perdida de información.

El análisis de componentes principales (PCA, por sus siglas en inglés) es una técnica estadística muy útil que tiene aplicaciones en campos como reconocimiento de rostros y compresión de imágenes y, además, es una técnica común para identificar patrones en datos de dimensión alta, y expresar los datos de tal manera que se resalten sus similitudes y diferencias. Dado que los patrones en datos de dimensión alta puede ser difícil hallarlos, donde no nos podemos dar el lujo de graficarlos porque no es posible directamente, PCA se vuelve una herramienta poderosa para el análisis de datos. PCA utiliza una aproximación al problema de los errores más complicado y completo que el que utiliza un análisis ANOVA (Brereton 1990; Jackson, 1991). Dadas n observaciones de p variables, se analiza si es posible representar adecuadamente esta información con un número menor de variables construidas como combinaciones lineales de las originales. Por ejemplo, con variables con alta dependencia es frecuente que un pequeño número de nuevas variables (menos del 20 por 100 de las originales) expliquen la mayor parte (más del 80 por 100 de la variabilidad original). La técnica de componentes principales tiene una doble utilidad:

1. Permite representar en un espacio de dimensión pequeña observaciones de un espacio general p -dimensional.

2. Permite transformar las variables originales, en general correlacionadas, en nuevas variables no correlacionadas, facilitando la interpretación de los datos.

Las p variables originales (x) son reemplazadas por un número menor de m nuevas variables (t), llamadas componentes principales, de forma que la pérdida de información sea mínima. Ello se consigue expresando cada nueva variable t_i como un vector combinación lineal de las p variables originales,

$$\begin{aligned} t_1 &= c_{11}x_1 + c_{12}x_2 + \dots + c_{1p}x_p \\ t_2 &= c_{21}x_1 + c_{22}x_2 + \dots + c_{2p}x_p \\ &\vdots \\ t_m &= c_{m1}x_1 + c_{m2}x_2 + \dots + c_{mp}x_p \end{aligned} \quad (1)$$

imponiendo la condición de ortogonalidad, es decir, las componentes principales no estarán relacionadas entre sí. Cada nueva variable t_i acumula, de forma decreciente (desde 1 hasta m), el máximo de información original. En teoría se pueden obtener tantas componentes principales como variables originales, si bien con un número reducido $m < p$ se retiene una cantidad suficiente de la información (varianza) de las variables originales.

La elección de los componentes se realiza de tal forma que el primero recoja la mayor proporción posible de la variabilidad original; el segundo componente debe recoger la máxima variabilidad posible no recogida por el primero, y así sucesivamente. Del total de componentes se elegirán aquéllos que recojan el porcentaje de variabilidad que se considere suficiente. A éstos se les denominará componentes principales. En muchos casos, solamente algunas de las variables utilizadas originalmente tienen un significado estadístico en el modelo obtenido. Las otras variables, que en realidad proporcionan poca información, pueden ser ignoradas. De esta manera, se puede considerar que el análisis de componentes principales es una forma de *reducción de dimensionalidad*.

Muchos algoritmos computacionales han sido desarrollados para PCA. Algunos métodos obtienen todos los componentes en un solo paso, mientras que otros encuentran primero los componentes más significativos. La matemática involucrada en PCA se basa en los principios del álgebra matricial. Para los estudios usando este método, es necesario entender la importancia del escalamiento, utilizar el concepto de error de manera acertada, el papel del ruido estadístico en

los datos y tener criterios para fijar el número de componentes a utilizar. Mediante PCA, es posible visualizar la distribución que adoptan los objetos (moléculas) en el espacio de descriptores y su agrupamiento en conglomerados (clusters). También es posible determinar cómo contribuyen las variables originales a esta distribución (Pla, 1986).

PLANTEAMIENTO DEL PROBLEMA

En este trabajo intentamos obtener relaciones cuantitativas estructura propiedad para cristales líquidos termotrópicos mediante el estudio de una serie de 69 moléculas que presentan la transición de fase nemática. La propiedad en estudio es la temperatura de transición nemática determinada experimentalmente. La originalidad de este trabajo reside en que por primera vez se ha aplicado el PCA al estudio de cristales líquidos termotrópicos como una herramienta que ha permitido la obtención de subconjuntos de CL's para la obtención de modelos QSPR locales. El otro aspecto importante es que los modelos locales nos permitirán descubrir qué tipo de propiedades moleculares tienen mayor influencia en el comportamiento cristal líquido de las moléculas estudiadas, lo cual será de suma importancia en la racionalización de las transiciones de fase que estos materiales experimentan. La aproximación que se propone se extenderá al estudio de cristales líquidos con actividad farmacológica.

HIPÓTESIS

Los modelos QSPR locales obtenidos para cristales líquidos con características semejantes, tendrán un mayor potencial predictivo de la temperatura de transición nemática que un modelo global obtenido de una serie de compuestos con diversidad estructural amplia.

OBJETIVO

El objetivo general de este proyecto es aplicar la metodología del Análisis de Componentes Principales a moléculas de cristales líquidos para la investigación de la relación entre las propiedades moleculares y la manifestación de la fase cristal líquido. La estrategia que se propone nos permitirá obtener un modelo QSPR global que permita predecir las temperaturas de transición nemática de series de moléculas de cristales líquidos termotrópicos considerando todas las moléculas. Enseguida se utilizará PCA con la finalidad de reconocer similaridades estructurales e identificar tendencias que nos permitan seleccionar subconjuntos

moleculares para obtener modelos QSPR locales y evaluar sus capacidades predictivas con respecto al modelo global y, en función de los descriptores finales en los modelos locales, racionalizar los factores que influyen en las transiciones de fase.

METODOLOGÍA

“Any alternative viewpoint with a different emphasis leads to an inequivalent description.

There is only one reality but there are many viewpoints. It would be very narrowminded to use only one: we have to learn to be able to imagine several.”

Hans Primas

In: Chemistry, Quantum Mechanics and Reductionism
(Primas, 1981)

“A molecular descriptor can be thought of as a mythological animal having several different meanings which depend on one's point of view.”

In: Handbook of Molecular Descriptors,
(Todeschini, 2000)

Conjunto de datos y generación de descriptores

Se estudiará un conjunto de 69 cristales líquidos termotrópicos (iminas aromáticas sustituidas en posición *para*, ver Figura 1), cuyas estructuras y temperaturas nemáticas experimentales asociadas se presentan en la tabla 1. Estos datos han sido tomados de resultados experimentales publicados (Thiemann, 1997, Liquid Crystals Database 1998). La estructura de la parte superior es la estructura básica de los compuestos estudiados, R1 y R2 son los grupos laterales colocados en la posición *para* en los anillos aromáticos, por lo que se espera que los efectos estéricos no sean importantes. Las estructuras de las moléculas fueron construidas utilizando el programa SPARTAN (Spartan, 2006). Con el mismo programa se llevó a cabo un análisis conformacional a nivel de Mecánica Molecular con el campo de fuerza SYBYL (Mayo, 1990, Boyd, 1982) y se seleccionó la conformación de mínima energía de cada molécula para llevar a cabo la optimización de geometría sin restricciones de simetría mediante un cálculo

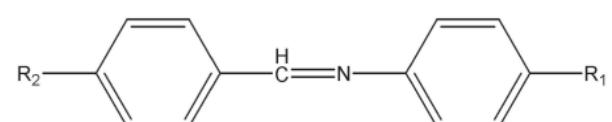


Figura 1. Estructura base de la serie de Cristales Líquidos estudiados.

mecano-cuántico a nivel de teoría y precisión numérica semiempírico PM3 y precisión numérica con base mínima (Stewart, 1989). Una vez que se obtuvieron las geometrías de mínima energía se llevó a cabo un análisis de frecuencias para asegurarnos de haber alcanzado un mínimo en la superficie de energía potencial. Las geometrías optimizadas se transfirieron al programa Dragon Systems, paquete desarrollado por Milano Chemometrics y QSAR Group (Talete, 2002) para la obtención de los descriptores diferentes a los obtenidos mediante SPARTAN. El software MobyDigs (Talete, 2002) fue utilizado para la obtención de los diferentes modelos.

Para describir las moléculas seleccionamos aquellos descriptores que pueden tener un papel determinante en la T_N y, por consiguiente, en las transiciones de fase. Los descriptores moleculares de todas las moléculas fueron generados mediante Dragon (Talete, 2002) y SPARTAN (Spartan, 2006). Los descriptores fueron seleccionados de tal forma que contienen información relevante relacionada con la forma geométrica, la estructura electrónica y la distribución de carga, se presentan en tabla 2.

Selección del subconjunto de descriptores

La selección de los subconjuntos de descriptores para los modelos global y locales se llevó a cabo mediante la aproximación de algoritmos genéticos (GA), basándose en la evolución de una población de modelos a fin de encontrar el más óptimo de ellos. El término algoritmo genético significa un vector binario llamado cromosoma (un modelo), en el cual tiene en cada posición un gen que corresponde a un descriptor (1 si esta incluido en el modelo y 0 no esta incluido en el mismo). Cada cromosoma corresponde a un modelo con una serie

Tabla 1.
Temperaturas de Transición de fase nemática, T_N , °C, de los cristales líquidos en estudio.

Molécula	R1	R2	T_N (°C)	Molécula	R1	R2	T_N (°C)
cl1	C_3H_7O	C_4H_9	55.7	cl36	C_4H_9O	$OCOOC_3H_7$	106.0
cl2	C_4H_9O	C_4H_9	75.0	cl37	C_4H_9O	$OCOOC_4H_9$	103.0
cl3	$C_5H_{11}O$	C_4H_9	69.2	cl38	C_4H_9O	$OCOOC_5H_{11}$	100.0
cl4	$C_6H_{13}O$	C_4H_9	77.8	cl39	C_4H_9O	$OCOOC_6H_{13}$	96.0
cl5	C_4H_9O	OC_6H_9	122.0	cl40	$C_4H_{11}O$	$OCOOC_3H_9$	105.0
cl6	C_4H_9O	OC_6H_{13}	119.0	cl41	$C_4H_{11}O$	$OCOOC_2H_5$	104.0
cl7	C_4H_9O	$OOCCH_3$	108.0	cl42	$C_4H_{11}O$	$OCOOC_4H_9$	96.0
cl8	C_4H_9O	$OOCC_2H_5$	113.0	cl43	$C_4H_{11}O$	$OCOOC_4H_9$	95.0
cl9	C_5H_7O	$OOCC_5H_{11}$	109.0	cl44	$C_5H_{11}O$	$OCOOC_5H_{11}$	93.0
cl10	C_4H_9O	$OOCC_6H_3$	100.0	cl45	$C_4H_{11}O$	$OCOOC_6H_3$	93.0
cl11	C_4H_9O	$OOCCH_3$	114.2	cl46	$C_6H_{13}O$	$OCOOC_3H_9$	113.0
cl12	C_4H_9O	$OCOCC_2H_5$	120.0	cl47	$C_6H_{13}O$	$OCOOC_2H_5$	108.0
cl13	C_4H_9O	$OOCC_4H_7$	119.0	cl48	$C_6H_{13}O$	$OCOOC_4H_7$	100.0
cl14	C_4H_9O	$OOCC_4H_9$	113.0	cl49	$C_6H_{13}O$	$OCOOC_4H_9$	99.0
cl15	C_4H_9O	$OCOCC_4H_1$	115.5	cl50	$C_6H_{13}O$	$OCOOC_4H_1$	96.0
cl16	C_4H_9O	$OCOCC_6H_3$	110.0	cl51	$C_6H_{13}O$	$OCOOC_6H_3$	96.0
cl17	$C_5H_{11}O$	$OOCCH_3$	105.0	cl52	C_3H_7COO	C_4H_9	56.5
cl18	$C_4H_{11}O$	$OCOCC_2H_5$	109.0	cl53	C_4H_9COO	CH_3	66.5
cl19	$C_5H_{11}O$	$OOCC_4H_9$	110.5	cl54	C_2H_5COO	C_2H_5	58.0
cl20	$C_3H_{11}O$	$OCOCC_6H_1$	111.0	cl55	C_4H_9COO	C_4H_9	71.2
cl21	$C_5H_{11}O$	$OCOCC_6H_{13}$	104.0	cl56	C_3H_7COO	C_2H_5	63.5
cl22	$C_6H_{13}O$	$OOCCH_3$	109.0	cl57	C_3H_7COO	C_4H_9	74.0
cl23	C_4H_9O	$COCH_3$	101.7	cl58	C_4H_9COO	C_2H_5	56.0
cl24	C_4H_9O	$COCH_3$	111.0	cl59	C_4H_9COO	C_4H_9	68.0
cl25	C_4H_9O	COC_4H_6	146.0	cl60	$C_4H_{11}COO$	C_2H_5	66.5
cl26	$C_5H_{11}O$	$COCH_3$	109.0	cl61	$C_5H_{11}COO$	C_4H_9	76.0
cl27	$C_6H_{13}O$	$COCH_3$	114.5	cl62	$C_6H_{13}COO$	C_2H_5	64.5
cl28	C_4H_9O	$OCOOC_3H_9$	108.0	cl63	$C_4H_{11}COO$	C_4H_9	74.5
cl29	C_4H_9O	$OCOOC_2H_5$	106.0	cl64	CH_3COO	$COCH_3$	105.0
cl30	C_4H_9O	$OCOOC_4H_7$	95.0	cl65	C_4H_9COO	$COCH_3$	113.0
cl31	C_4H_9O	$OCOOC_4H_9$	93.0	cl66	C_3H_7COO	$COCH_3$	115.5
cl32	C_3H_7O	$OCOOC_5H_{11}$	90.0	cl67	C_4H_9COO	$COCH_3$	105.5
cl33	C_4H_9O	$OCOOC_4H_{13}$	88.0	cl68	$C_4H_{11}COO$	$COCH_3$	110.5
cl34	C_4H_9O	$OCOOC_3H_9$	114.0	cl69	$C_6H_{13}COO$	$COCH_3$	107.0
cl35	C_4H_9O	$OCOOC_2H_5$	114.0				

Tabla 2.
Breve descripción de los diversos tipos de descriptores utilizados.

Tipo de descriptores	Descripción
Constitucionales	Molecular Wt.(amu), nSK (número de átomos diferentes a Hidrógeno)
Geométricos	RGyr (radio de giro ponderado por la masa), G1 (índice gravitacional), L/Bw (Relación longitud-anchura), Ovalidad, CPK Vol (Volumen Molecular, \AA^3), CPK Área (Área superficial molecular, \AA^2), SPH (Esfericidad), ASP (índice de asimetría) MOVINERCI1, MOVINERCI2, MOVINERCI3, Momentos de inercia principales.
Descriptores de Carga	qmax (carga máxima positiva), qmax (carga máxima negativa), Dipole (magnitud del momento dipolar eléctrico molecular)
Fisicoquímicos	G° (Energía libre de Gibbs, kJ/mol) S° (Entropía, $\text{kJ/mol}^\circ\text{C}$) H° (Entalpía, kJ/mol)
Estructura Electrónica	E LUMO (energía del orbital molecular más bajo desocupado, kJ/mol) E HOMO (energía del orbital molecular más alto ocupado, kJ/mol)
Otros	CFD (HBA), Descriptores de funciones químicas, aceptores de enlaces de hidrógeno.

de variables. Una vez que se define el parámetro a optimizar, que en este caso fue Q^2 (que es el cuadrado del coeficiente de validación cruzada y es una medida de la predicción interna), comienza la evolución del algoritmo genético basado en tres pasos y una condición de finalización:

- 1) Creación de una población: se comienza con la construcción aleatoria de los modelos de una población.
- 2) Entrecruzamiento: se genera el entrecruzamiento aleatorio de los genes (descriptores) entre los mejores modelos generados (cromosomas) en el primer paso, a fin de encontrar un mejor modelo. Si esto es así, éste se incluye en la población, este procedimiento se genera por iteración.
- 3) Mutación: por cada modelo (cromosoma) presente en la población, se genera un cambio de algunos de los genes (variables) del modelo, de esta manera aumenta el tamaño de la población, a fin de generar nuevas aproximaciones, que si son altos sus valores del parámetro a optimizar se incluyen en la población. También este procedimiento es generado por iteración.
- 4) Condición de finalización: el segundo y el tercer paso son repetidos hasta que se encuentra la condición de finalización, eso está definido por el número máximo de iteraciones o algún proceso arbitrario (Clark, 2000).

Obtención de modelos y su validación

Los modelos QSPR se obtienen mediante un análisis de Correlación y Regresión Lineal Múltiple utilizando Mínimos Cuadrados Ordinarios (OLS). La metodología calcula un modelo lineal paramétrico para una sola respuesta y proporciona los coeficientes de los descriptores en el modelo. Debido a que los descriptores caracterizan la información estructural molecular sus magnitudes son extensamente diferentes. Para impedir que los descriptores con los rangos más grandes pesen más que aquéllos con los rangos más pequeños, los descriptores originales son todos transformados a un rango más consistente mediante un proceso de normalización previo a la obtención de los modelos. En este tipo de modelos se supone que la variable dependiente es una función multilineal de los descriptores. La técnica de validación constituye una herramienta fundamental para la evaluación de la capacidad reproductiva y predictiva de los modelos obtenidos de una serie de datos por regresión multivariante. Se utilizan los siguientes parámetros estadísticos para establecer

la condición necesaria de la validez de un modelo de regresión: 1) R^2 que es el cuadrado del coeficiente de correlación, es una medida relativa de la calidad de ajuste entre la variable dependiente (T_N) y las variables independientes (descriptores) del modelo construido a partir de un conjunto de entrenamiento. Sus valores van desde 0 hasta 1, cuanto más cercano a 1 mayor será la calidad del modelo; 2) Q^2 que es el cuadrado del coeficiente de validación cruzada, es una medida de la predicción interna. Dicho en otras palabras mide el poder predictivo del modelo. Sus valores van desde 0 hasta 1, cuanto más cercano a 1 mayor será la calidad predictiva del modelo. En este trabajo R^2 y Q^2 se expresan como un porcentaje, $0 \leq R^2 \leq 100$. s es la desviación estándar, que es una medida de la calidad absoluta del modelo, cuanto más pequeño este valor menor será la dispersión de la respuesta dada por el modelo y 4) F de Fisher (razón de variancia), es una medida de la significación estadística con respecto a la variabilidad del modelo (Todeschini, 2004).

Para la obtención de los modelos, los conjuntos de datos se dividieron en una relación de 3:1 en los subconjuntos de entrenamiento y de prueba respectivamente (marcados con un asterisco en las correspondientes tablas). Los subconjuntos de entrenamiento fueron utilizados para la generación de los modelos. Los conjuntos de prueba sirvieron como muestras externas para evaluar su capacidad predictiva. Nuestros modelos fueron obtenidos mediante el programa Mo-byDigs (Todeschini, 2004) utilizando la aproximación de GA para la selección de las variables más relevantes en los modelos. El método GA proporciona un conjunto de modelos aceptables con diferentes subconjuntos de variables, esto permite su evaluación para poder seleccionar los de mejor calidad. Se partió de 10 poblaciones de 50 modelos cada una. Esto generó diversos modelos finales con un número de variables de 5 a 7, siendo siete el valor máximo de variables definido para los modelos. Se obtuvieron algunos modelos con más de 8 variables, sin embargo, los modelos con siete variables fueron los mejores.

RESULTADOS Y DISCUSIÓN

Resultados modelo Global

Para encontrar el modelo global el conjunto de datos ($n=69$) se dividió en dos grupos: conjunto de calibración ($n=52$) y conjunto de prueba ($n=17$). El número total de descriptores fue de 22 y el conjunto de descriptores más conveniente para el modelo global se obtuvo mediante GA utilizando el procedimiento de va-

lidación cruzada (cross-validated) y como función objetivo Q^2 leave-one-out, (Cramer, 1988) para evaluar la calidad del modelo durante el proceso de evolución. Se formaron diez poblaciones con un tamaño de 50 individuos (modelos) y con un número máximo de descriptores igual a siete. Los modelos finales seleccionados fueron 10 con un número total de variables retenidas igual a 10 y distribuidas en los modelos. Se seleccionó el modelo 7 debido a que en él aparecen variables que no se encuentran correlacionadas de manera significativa ($r^2 < 0.1$, en promedio, en la matriz de correlación).

Modelo Global obtenido:

$$\begin{aligned}
 T_N &= 342.31453^*S^{\circ} + 1.09718^*H^{\circ} + 2 \\
 &0.3636^*Wt + 19.41403^*Vol + 1.750 \\
 &79^*Dipole \\
 &+ 0.42123^*LUMO + 0.60322^* \\
 &HOMO - 126.76048
 \end{aligned} \tag{2}$$

Parámetros estadísticos:

$$\begin{aligned}
 R^2 &= 80.95, Q^2 = 72.09, F_{7,44} = 26.72 \\
 (p &= 0.05), s = 8.581
 \end{aligned}$$

Como se puede observar el método GA, seleccionó una combinación de diversos descriptores: Termodinámicos, de distribución de Carga, Geométricos y de Estructura electrónica. La contribución más relevante corresponde a la entropía, lo cual sugiere que en la generación de la fase nemática el factor entrópico juega un papel importante, sugiriendo estudiar la forma en que se pueda determinar la conformación molecular en la mesofase que en principio debe ser diferente de la conformación de mínima energía. Se ha establecido que en las transiciones de fase observadas en cristales líquidos, ocurren cambios importantes en el orden estructural de las fases involucradas (Collings, 1997), lo cual es congruente con

este resultado. Las contribuciones restantes contribuyen de manera positiva.

Los parámetros de la regresión poseen una buena calidad estadística del modelo. Los valores de las T_N 's de la predicción calculadas utilizando el modelo dado por la ecuación (2) se presentan en la tabla 3. Se puede observar que el modelo predice las T_N 's para las moléculas en estudio con un error absoluto promedio de 0,56 %.

Componentes Principales

Con la finalidad de probar nuestra hipótesis se procedió a aplicar el análisis de Componentes principales al conjunto total de las moléculas en estudio. Del total de los componentes retenidos (PC_1 , PC_2 , PC_3 y PC_4), PC_1 y PC_2 conservan una varianza acumulada del 71%, lo cual garantiza que los

Tabla 3.

Temperaturas de Transición Nemática, $^{\circ}\text{C}$, experimentales y calculadas mediante el modelo Global, para los cristales líquidos estudiados.

Comp.	Y Exp.	Y-Pred	Err. Pred.	% Err. Abs.	Comp.	Y Exp.	Y-Pred	Err. Pred.	% Err. Abs.
cl1	55.70	71.26	15.56	27.94	cl36	106.00	103.09	-2.91	-2.75
cl2*	75.00	67.50	-7.50	-10.00	cl37	103.00	102.78	-0.22	-0.21
cl3*	69.20	66.15	-3.05	-4.41	cl38	100.00	102.83	2.83	2.83
cl4	77.80	58.12	-19.68	-25.30	cl39	96.00	87.29	-8.71	-9.07
cl5	122.00	115.90	-6.10	-5.00	cl40	105.00	105.31	0.31	0.30
cl6*	119.00	117.96	-1.04	-0.87	cl41	104.00	101.74	-2.26	-2.17
cl7	108.00	102.81	-5.19	-4.81	cl42*	96.00	103.91	7.91	8.24
cl8*	113.00	108.90	-4.10	-3.63	cl43	95.00	102.85	7.85	8.26
cl9*	109.00	111.64	2.64	2.42	cl44*	93.00	90.73	-2.27	-2.44
cl10	100.00	112.44	12.44	12.44	cl45	93.00	88.21	-4.79	-5.15
cl11	114.20	103.67	-10.53	-9.22	cl46*	113.00	104.31	-8.69	-7.69
cl12	120.00	108.71	-11.29	-9.41	cl47	108.00	96.42	-11.58	-10.72
cl13	119.00	111.25	-7.75	-6.51	cl48	100.00	102.75	2.75	2.75
cl14	113.00	111.67	-1.33	-1.18	cl49	99.00	102.17	3.17	3.20
cl15	115.50	110.94	-4.56	-3.95	cl50	96.00	82.70	-13.30	-13.85
cl16*	110.00	108.11	-1.89	-1.72	cl51	96.00	101.69	5.69	5.93
cl17	105.00	105.09	0.09	0.09	cl52	56.00	70.66	14.66	26.18
cl18	109.00	109.92	0.92	0.84	cl53	66.50	69.43	2.93	4.41
cl19	110.50	108.44	-2.06	-1.86	cl54*	58.00	68.62	10.62	18.31
cl20	111.00	97.28	-13.72	-12.36	cl55	71.00	66.45	-4.55	-6.41
cl21	104.00	115.88	11.88	11.42	cl56	63.00	72.18	9.18	14.57
cl22	109.00	103.03	-5.97	-5.48	cl57	74.00	70.04	-3.96	-5.35
cl23	101.70	119.28	17.58	17.29	cl58*	56.00	70.95	14.95	26.70
cl24	111.00	120.87	9.87	8.89	cl59	68.00	68.25	0.25	0.37
cl25	146.00	118.19	-27.81	-19.05	cl60	66.50	70.85	4.35	6.54
cl26	109.00	116.06	7.06	6.48	cl61*	76.00	68.63	-7.37	-9.70
cl27*	114.50	114.68	0.18	0.16	cl62*	64.50	69.78	5.28	8.19
cl28*	108.00	106.13	-1.87	-1.73	cl63	74.50	64.35	-10.15	-13.62
cl29	106.00	102.79	-3.21	-3.03	cl64	105.00	100.79	-4.21	-4.01
cl30	95.00	104.94	9.94	10.46	cl65*	113.00	107.21	-5.79	-5.12
cl31	93.00	104.15	11.15	11.99	cl66*	115.50	111.62	-3.88	-3.36
cl32	90.00	103.63	13.63	15.14	cl67	105.50	112.87	7.37	6.99
cl33	88.00	103.13	15.13	17.19	cl68	110.50	110.07	-0.43	-0.39
cl34	114.00	104.66	-9.34	-8.19	cl69	107.00	109.88	2.88	2.69
cl35	114.00	101.58	-12.42	-10.89					

* Compuestos utilizados como conjunto de validación

primeros Componentes Principales contemplan la mayor parte de la información y ya son suficientes para representar las muestras de un modo satisfactorio. En la figura 2 se presenta cada muestra en el sistema de ejes PC_1 y PC_2 . En este sistema uno de los ejes (que se denomina primer componente principal, o PC_1 , del inglés Principal Component), se orienta en la dirección de máxima variación (dispersión). El segundo eje (PC_2) también se orienta en la dirección de la máxima dispersión. Pero, como debe ser perpendicular al primer eje, sigue la dirección de la máxima dispersión residual que no ha explicado el PC_1 y recoge información que no ha podido recoger el PC_1 . Mediante un análisis de las contribuciones de las variables a los componentes principales, se identifican tendencias y se observa que en el espacio bidimensional de los primeros componentes principales PC_1 y PC_2 se proyectan dos tipos de agrupamientos debido a la relevancia del componente PC_2 . El primer subconjunto se forma con las moléculas ubicadas en los cuadrantes I y II del diagrama y el otro subconjunto con las ubicadas en los cuadrantes III y IV.

Modelo Local Subconjunto 1

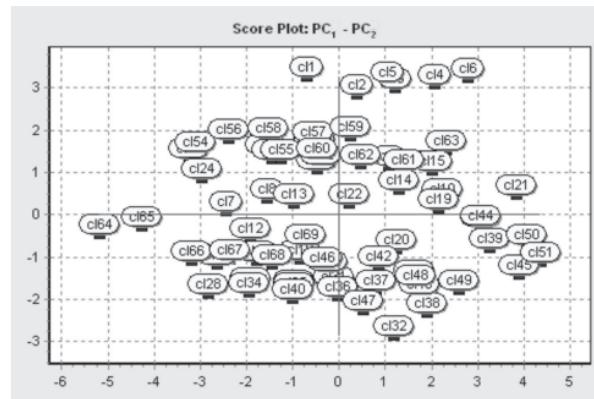


Figura 2. Proyección de las moléculas en el espacio bidimensional de los componentes principales PC_1 (eje de ordenadas) y PC_2 (eje de abscisas). Las etiquetas identifican a las moléculas.

Para encontrar el modelo local, el conjunto de datos ($n=33$) se dividió en dos grupos: conjunto de calibración ($n=24$) y conjunto de prueba ($n=8$). Se excluyó una molécula por análisis de marginales (outliers). El número total de descriptores fue de 22 y el conjunto de descriptores más conveniente para el modelo se obtuvo mediante GA utilizando como función objetivo Q^2 leave-one-out para evaluar la calidad del modelo durante el proceso de evolución. Se formaron diez poblaciones;

población con un tamaño de 50 individuos (modelos) y con un número máximo de descriptores igual a siete en cada modelo. Los modelos finales seleccionados fueron 10 con un número total de variables retenidas igual a 9 y distribuidas en los modelos. Se seleccionó el modelo 1 aunque en él aparecen variables altamente correlacionadas como el volumen molecular y el área superficial ($r^2=0.99$). Esto se fundamentó mediante un análisis de la frecuencia de las variables en los mejores modelos de la población final ya que el Área y el Volumen tienen una frecuencia relativa del 100%, lo cual sugiere que, aunque altamente correlacionadas son de gran importancia para modelar la T_n .

Modelo Local 1:

$$\begin{aligned}
 T_n = & 1105.6597 * \text{OVALIDAD} + 44.9304 * \text{CFD(HBA)} + \\
 & 7.8385 * \text{Vol-8.8019 * Area} \\
 & -3.21194 * \text{Dipole} + 0.0032 * \text{Inercia} - 36.39221 * \\
 & \text{qnmax- 1214.0349.}
 \end{aligned} \tag{3}$$

Parámetros estadísticos: $R^2=97.00$, $Q^2=94.37$, $F_{7,16}=73.98$ ($p=0.05$), $s= 5.207$

Como se puede observar, en el modelo aparecen cuatro descriptores relacionados con la geometría molecular (Vol, Area, Inercia, Ovalidad); dos con la distribución de carga (qnmax y dipole) y uno relacionado con las interacciones intermoleculares mediante enlaces de hidrógeno. En este modelo la contribución más relevante corresponde a la ovalidad. Esta propiedad es un indicador de que tan cercana está la geometría de la molécula a un cilindro o a una esfera. Si $O>1$ la molécula tiende a adoptar una forma cilíndrica y si $O<1$ una forma esférica. Para los cristales líquidos del subconjunto 1, los valores de ovalidad ($1.53 < O < 1.73$) sugieren que las moléculas tienden a adoptar una forma cilíndrica. Además, se ha observado experimentalmente (Demus, 1998) que para valores de $L/Bw>3$ se detecta la fase nemática y que para moléculas muy elongadas se presenta la fase esmética. Los valores de la relación de aspecto en las moléculas de este subconjunto están en el rango $8.93 < L/Bw < 50.1$, por lo que los cristales líquidos estudiados es probable que presenten también la fase esmética.

Lo anterior sugiere que la metodología que hemos empleado resulta adecuada para la obtención de modelos del tipo *relación cuantitativa estructura-propiedad* (QSPR), que ayuden a predecir, entender y racionalizar las propiedades moleculares que favorecen la aparición de mesofases.

Los parámetros de la regresión indican una excelente calidad estadística del modelo. Los valores de las T_N 's calculadas utilizando el modelo dado por la ecuación (3) se presentan en la tabla 4.

Tabla 4.
Temperaturas de Transición Nemática, °C, experimentales y calculadas mediante el modelo del subconjunto 1.

Comp.	Y Exp.	Y-Pred	Err. Pred.	% Err. Abs.
cl1	55.70	68.79	13.09	23.50
cl2	75.00	71.79	-3.21	-4.28
cl3*	69.20	67.09	-2.11	-3.05
cl4	77.80	66.74	-11.06	-14.22
cl5	122.00	120.08	-1.92	-1.57
cl6	119.00	121.64	2.64	2.22
cl7*	108.00	97.24	-10.76	-9.96
cl8	113.00	105.71	-7.29	-6.45
cl9*	109.00	108.32	-0.68	-0.62
cl10*	100.00	108.58	8.58	8.58
cl13	119.00	122.11	3.11	2.61
cl14	113.00	116.64	3.64	3.22
cl15	115.50	112.44	-3.06	-2.65
cl19	110.50	118.03	7.53	6.81
cl21	104.00	105.37	1.37	1.32
cl22	109.00	109.92	0.92	0.84
cl23	101.70	102.44	0.74	0.73
cl24	111.00	108.48	-2.52	-2.27
cl26	109.00	110.95	1.95	1.79
cl27	114.50	110.47	-4.03	-3.52
cl52	56.00	67.45	11.45	20.45
cl53*	66.50	68.54	2.04	3.07
cl54	58.00	59.96	1.96	3.38
cl55*	71.00	57.86	-13.14	-18.51
cl56	63.00	54.34	-8.66	-13.75
cl57	74.00	65.23	-8.77	-11.85
cl58	56.00	62.25	6.25	11.16
cl59	68.00	67.02	-0.98	-1.44
cl60	66.50	65.15	-1.35	-2.03
cl61*	76.00	65.43	-10.57	-13.91
cl62	64.50	63.96	-0.54	-0.84
cl63*	74.50	65.53	-8.97	-12.04

*Compuestos utilizados como conjunto de validación. No se utilizó el Cl25 por resultar un marginal (outlier)

Modelo Local Subconjunto 2

La metodología para encontrar el modelo local del subconjunto dos fue igual que las utilizadas anteriormente. De los modelos finales se seleccionó el modelo 1. Para este modelo, dos compuestos (Cl16 y Cl32) se clasificaron como marginales (outliers) mediante el criterio de la desviación entre las T_N 's observadas y calculadas ($obs-cal > 3s$).

Modelo Local 2:

$$T_N = -120.4858 * \text{ASP} - 12.8230 * \text{qnmax} + 6.9231 * \text{Vol} - 8.2999 * \text{Area}$$

$$+ 2.68196 * \text{Wt} + 0.4053 * \text{LUMO} + 0.40396 * \text{L/Bw} \\ + 127.06844 \quad (4)$$

$$\text{Parámetros estadísticos: } R^2=81.99, Q^2=63.56, F_{7,18}= 11.78 \text{ (p=0.05), s=3.092}$$

Como se puede observar en el modelo local 2, aparecen cuatro descriptores relacionados con la geometría molecular (ASP, Vol, Área, L/Bw); uno con la distribución de carga (qnmax) y uno relacionado con la estructura electrónica. En este modelo la contribución más relevante corresponde a la ASP.

La forma molecular está relacionada con varios procesos fisicoquímicos, tales como fenómenos de transporte, contribuciones entrópicas y a la capacidad de interacción entre ligando y receptor en el ámbito del diseño de fármacos. El grado de desviación de la forma esférica es llamado anisometría, ASP es un descriptor geométrico tridimensional que se obtiene a partir de los eigenvalores de la matriz de inercia y mide la anisometría de las moléculas. ASP está relacionado con el tamaño y la forma de las moléculas, sus valores están en el rango: $0 \leq \text{ASP} \leq 1$, cuando $\text{ASP}=0$ corresponde a una forma esférica y si $\text{ASP}=1$ indica moléculas lineales, valores entre cero y uno corresponden a moléculas tipo cilíndricas (Todeschini, 2000). Para las moléculas del subconjunto dos: $0.722 \leq \text{ASP} \leq 0.943$ con un valor promedio de 0.874, por lo que estas moléculas tienen una forma geométrica ovoide (prolata) la cual favorece las interacciones intermoleculares que ayudan a la aparición de la fase nemática.

Los parámetros de la regresión indican una excelente calidad estadística del modelo. Los valores de las T_N 's calculadas utilizando el modelo dado por la ecuación (4) se presentan en la tabla 5.

CONCLUSIONES

Este trabajo aplica PCA a un conjunto de 69 cristales líquidos termotrópicos cuyas T_N 's se conocen experimentalmente, para separar el conjunto de moléculas en subconjuntos, con la finalidad de confirmar la hipótesis de que los modelos QSPR locales exhibirán una mejor capacidad de predecir las T_N 's que el modelo derivado del conjunto global. El análisis PCA nos permitió obtener dos subconjuntos de moléculas para validar la hipótesis y encontramos que los resultados obtenidos mediante modelos locales son mejores a los obtenidos globalmente.

Tabla 5.

Temperaturas de Transición Nemática, °C, experimentales y calculadas mediante el modelo del subconjunto 2.

Comp.	Y Exp.	Y-Pred	Err. Pred.	% Err. Abs
cl11	114.20	111.62	-2.58	-2.26
cl12	120.00	121.30	1.30	1.08
cl17	105.00	107.68	2.68	2.55
cl18	109.00	108.10	-0.90	-0.83
cl20	111.00	109.22	-1.78	-1.60
cl28*	108.00	113.72	5.72	5.30
cl29	106.00	105.75	-0.25	-0.24
cl30*	95.00	103.94	8.94	9.41
cl31	93.00	102.70	9.70	10.43
cl32*	90.00	111.76	21.76	24.18
cl34	114.00	109.27	-4.73	-4.15
cl35*	114.00	104.75	-9.25	-8.11
cl36	106.00	103.19	-2.81	-2.65
cl37	103.00	98.63	-4.37	-4.24
cl38	100.00	110.46	10.46	10.46
cl39	96.00	90.30	-5.70	-5.94
cl40	105.00	110.74	5.74	5.47
cl41	104.00	103.19	-0.81	-0.78
cl42	96.00	95.93	-0.07	-0.07
cl43	95.00	97.56	2.56	2.69
cl44	93.00	98.64	5.64	6.06
cl45	93.00	93.27	0.27	0.29
cl46	113.00	109.83	-3.17	-2.81
cl47	108.00	99.22	-8.78	-8.13
cl48	100.00	100.76	0.76	0.76
cl49*	99.00	98.89	-0.11	-0.11
cl50*	96.00	94.96	-1.04	-1.08
cl51	96.00	98.13	2.13	2.22
cl64	105.00	111.18	6.18	5.89
cl65	113.00	109.27	-3.73	-3.30
cl66	115.50	110.50	-5.00	-4.33
cl67	105.50	108.32	2.82	2.67
cl68*	110.50	106.97	-3.53	-3.19
cl69	107.00	106.50	-0.50	-0.47

* Compuestos utilizados como conjunto de validación. No se utilizaron los compuestos cl16 y cl33 por resultar marginales (outliers)

La reducción del número de variables de entrada se logró aplicando la metodología de Algoritmos Genéticos. Los resultados sugieren que un número pequeño de descriptores (7 de 22 en este trabajo) relacionados con la geometría molecular (Vol. Área, Inercia, Ovalidad, ASP), la distribución de carga (qnmax y dipolo), la termodinámica del proceso de transición de fase (Entropía, S y Entalpía, H) y la estructura electrónica (HOMO, LUMO) son los que codifican mejor las transiciones de fase. El análisis de los descriptores nos permite afirmar que estas moléculas pueden también experimentar transiciones de fase esmécticas.

Concluimos que la metodología PCA-GA-QSPR que hemos empleado resulta adecuada para la obtención de modelos del tipo relación cuantitativa estructura-propiedad (QSPR), que ayuden entender y raciona-

lizar las propiedades moleculares que favorecen la aparición de mesofases. Este procedimiento puede ser aplicado a CL's con actividad biológica en donde el conocimiento de las Relaciones Estructura Actividad, SAR, junto con la generación de modelos QSAR, proporcionará una gran cantidad de evidencias que pueden colaborar en el desarrollo de nuevos CL's con actividades biológicas específicas, así como utilizar los modelos generados para el escaneo de bases de datos en la búsqueda automatizada de cristales líquidos con potencial actividad biológica.

AGRADECIMIENTOS

MV agradece al CONACYT y a la Dirección Apoyo a la Investigación y al Posgrado de la Universidad de Guanajuato por los apoyos otorgados para proyectos de investigación. MRH y TOV agradecen al CONACYT por sendas becas para la realización de su tesis de licenciatura.

REFERENCIAS

- Boyd, D. B. and Lipkowitz, X. B. (1982), *J. Chem. Ed.* 59, (4), 269.
- Berereton, R.G. (1990) *Chemometrics: applications of mathematics and statistics to laboratory systems*, Ellis Horwood, New York.
- ChandrasekhaR, S., (1992), *Liquid crystals*, Second Edition, Centre for Liquid Crystal Research, Bangalore, Cambridge University Press.
- Clark, David E., (2000), *Evolutionary Algorithms in Molecular Design, Methods and Principles in Medicinal Chemistry*, Volume 8, Wiley, Federal Republic of Germany.
- Collings, P.J., Hird, M., (1997) *Introduction to liquid crystals. Chemistry and Physics*, Taylor and Francis, London, p. 193.
- Cramer, R. D., III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. (1988) *Quantitative Structure-Activity Relationships*. Vol. 7, Issue 1, 18-25.
- Demus, D., (1998), *Handbook of liquid crystals* Vol. 1. Fundamentals, Wiley-VCH, New York.
- Friedel, G., (1922) *Ann. Physique*, 18, 273.
- Hansch, C., and Leo, A., (1995), *Exploring QSAR, Fundamentals and applications in chemistry and biology*, American Chemical Society, Washington D.C.
- Jackson, J. Edward, (1991), *A user's guide to principal components*, Wiley, United States of America.
- Katritzky Alan R., Petrukhin Ruslan, and Tatham Douglas, (2001), *J. Chem. Inf. Comput. Sci.* 41, 679-685.
- Lehmann, O.H., (1889) *Physikal Chem.*, 4, 462.
- Liquid Crystals Database Version 3.4 (1998), LCI Publisher GmbH, Eichenstr, 3, D-20259 Hamburg, Germany.

- Mayo, S. L.; Olafson, B. D. And Goddard, W. A. (1990), *J. Phys. Chem.* 94, (26), 8897.
- Pla, Laura E., (1986), *Análisis Multivariado: Método de Componentes Principales*, Serie de Matemática, Monografía No. 27, O.E.A., Washington, D.C.
- Primas, Hans, (1981) *Chemistry, Quantum Mechanics, and Reductionism: Perspectives in Theoretical Chemistry*, Springer Verlag, Berlin.
- Reinitzer, F., (1888), *Monatsch Chem.*, 1888, 9, 421.
- ScienceDaily. (2007), *Pharmaceuticals*, <http://www.sciencedaily.com/releases/2007/09/070906135516.htm>.
- Spartan 06 V112 (2006), *Spartan for Windows and Linux*, Wavefunction, Inc., Irvine, Calif. (E.U.).
- Stewart, (1989), J.J.P., *Computational Chem* 10, (2), 209.
- Talete srl, (2006) Dragon for Windows (software for Molecular Descriptor Calculations) Version 5.4, <http://www.talete.mi.it/>. Thiemann, T. and Vill, V., (1997), *Liquid Crystals*, Vol. 22, No. 4, 519-523.
- Todeschini, R.; Weinheim, V., (2000) *Handbook of Molecular Descriptors*, Edited by Mannhold, R., Kubinyi, H., Timmerman, H., *Methods and Principles in Medicinal Chemistry*, Volume II, wiley-vch, Federal Republic of Germany.
- Todeschini, R.; Ballabio, D.; Consonni, V.; Mauri, A.; Pavan, M., (2004), *MobyDigs*-version 1.0-2004-Copyright talete srl.
- Todeschini R.; Ballabio, D.; Consonni, V.; Mauri, A.; Pavan, M., (2004), *MobyDigs User Manual Version 1-204*, Milano Chemometrics and QSAR Research Group, Talete srl, Milano, Italy.
- Woltman, S., Crawford, G., Jay, G., (2007), Editors, liquid crystals, *Frontiers in Biomedical Applications*, World Scientific Publishing Co. Singapore.