ESTUDIO DE LA PARTICIPACIÓN DE PARÁMETROS VIBRACIONALES EN RELACIONES ESTRUCTURA-OLOR.

WILLIAM ERNESTO RODRÍGUEZ CÓRDOBA

LABORATORIO DE CROMATOGRAFÍA, CIBIMOL
FACULTAD DE CIENCIAS,
ESCUELA DE QUÍMICA
UNIVERSIDAD INDUSTRIAL DE SANTANDER
BUCARAMANGA,
2005

ESTUDIO DE LA PARTICIPACIÓN DE PARÁMETROS VIBRACIONALES EN RELACIONES ESTRUCTURA-OLOR.

WILLIAM ERNESTO RODRÍGUEZ CÓRDOBA

Trabajo de grado para optar al título de
Magíster en Química

Directores:
ELENA STASHENKO;
Química, Ph. D.
JAIRO R. MARTÍNEZ
Químico, Ph. D.

LABORATORIO DE CROMATOGRAFÍA, CIBIMOL
FACULTAD DE CIENCIAS,
ESCUELA DE QUÍMICA
UNIVERSIDAD INDUSTRIAL DE SANTANDER
BUCARAMANGA,
2005

TITULO
ESTUDIO DE LA PARTICIPACIÓN DE PARÁMETROS VIBRACIONALES EN RELACIONES ESTRUCTURA-OLOR.*

WILLIAM ERNESTO RODRÍGUEZ CORDOBA**

PALABRAS CLAVES
QSAR, PCA, Olfacción, Clustering, Vibracional.

RESUMEN

La olfacción, es de nuestros cinco sentidos, el menos comprendido. Desde el punto de vista químico, la olfacción inicia con la interacción entre las moléculas odorantes y los receptores, y es de vital importancia conocer como las propiedades químicas de los odorantes determinan el olor que percibimos

Un entendimiento de la olfacción a un nivel molecular, facilitaría la predicción de los olores, lo cual es de interés en la industria de fragancias. Turin ha revivido una idea Antigua, la cual postula que el olor de las moléculas se encuentra relacionado con el espectro vibracional, "Teoría vibracional". Esta idea ha sido muy poco fundamentada, ya que parece estar en total desacuerdo con la teoría establecida, "la teoría de enlace", la cual predice que la olfacción es el resultado de el enlace de las moléculas odorantes a un pequeño numero de los 350 o más receptores olfativos humanos, los cuales son proteínas-G acoplada.

EL análisis de las relaciones estructura-olor (SOR) usando métodos asistidos por computadores y técnicas de reconocimientos de patrones, pueden proveer un práctico acercamiento a el estudio de los odorantes. El corazón de este trabajo es encontrar un conjunto de descriptores moleculares, los cuales, a partir de relaciones de discriminación pueden ser establecidos. El análisis de las relaciones estructura-olor, utilizando el análisis de componentes principales y técnicas de agrupamiento, fueron realizadas sobre el diverso conjunto de datos de las 60 moléculas odorantes. Estas moléculas fueron divididas en 4 categorías: Floral, Frutal, Verde y Almizcle.

Los cálculos de la energía, optimización de la geometría, momento dipolar y las frecuencias, fueron realizadas con el software Gaussian03. Usando un conjunto de base 6-31G(d) con el método de Hartree-Fock. Diferentes subespacios fueron construidos a partir de los descriptores disponibles, donde se clasificaron las 60 moléculas de acuerdo con su nota olfativa.

*Tesis

**Facultad de Ciencias, Escuela de Química, E. E. STASHENKO, J. R. MARTÍNEZ

TITLE
STUDY OF THE PARTICIPATION OF VIBRATIONAL PARAMETERS IN STRUCTURE-ODOR RELATIONSHIPS.*

WILLIAM ERNESTO RODRÍGUEZ CORDOBA**

KEYWORDS
QSAR, PCA, Olfaction, Clustering, Vibrational.

ABSTRACT

Olfaction is the least well understood of our five senses. Since it is a chemical sense, initiated by the interaction between odour molecules and receptors, it is of interest to know how the chemical properties of odorants determine the odour we perceive. An understanding of olfaction at the molecular level would facilitate odour prediction, and hence is of interest to the fragance industry. Turin has in recent years revived the decades-old idea that the smell of a molecule is related to its vibrational spectrum, the "vibrational theory". His idea have found little support, as they appear to contradict the established "binding theory" that olfaction results from the specific binding of an odour molecule to a small number of the 350 or so human olfactory receptor, which are the largest family of G-protein-coupled receptors.

Analysis of structure-odor relations (SOR) using computer assisted methods and pattern recognition techniques can provide a practical approach to the study of odorants. The heart of this approach is finding a set of molecular descriptors from which discriminating relationships can be found. Structure-odour relationships analyses using principal component analysis and hierarchical clustering were carried out on diverse data set of 60 odorants molecules. These molecules were divided into four odour categories: floral, fruity, green and musk.

Energy, geometry optimization, dipole moment and frecuencies calculations were performed with the sotware Gaussian03, using the 6-31 G(d) basis set with the Hartree-Fock method. Different subspaces were constructed from the available descriptors, using constitutional, topological, electronics, geometrics, charge-partial surface area, and vibrational descriptors and the 60 molecules were classified according to the olfactory note.

*Tesis

**Facultad de Ciencias, Escuela de Química, E. E. STASHENKO, J. R. MARTÍNEZ

CONTENIDO

LISTA DE FIGURAS

LISTA DE TABLAS

LISTA DE ANEXOS

ABREVIATURAS, ACRÓNIMOS Y SIMBOLOGÍA

Å	Angstroms
AM1	*Austin Model* 1 (Modelo de Austin 1)
cAMP	*Cyclic Adenosyne Monophosphate* (Adenosinmonofosfato cíclico)
CPSA	*Charged Partial Surface Area* (Área superficial parcialmente cargada)
d	Descriptor
d1	Número de átomos
d2	Número de átomos de C
d3	Número de átomos de H
d4	Número relativo de átomos de C
d5	Número relativo de átomos de H
d6	Número de anillos
d7	Número de anillos de benceno
d8	Número de enlaces
d9	Número de enlaces sencillos
d10	Número de enlaces dobles
d11	Número de enlaces triples
d12	Número de enlaces aromáticos
d13	Número relativo de enlaces sencillos
d14	Número relativo de enlaces dobles
d15	Número relativo de enlaces triples
d16	Número relativo de enlaces aromáticos
d17	Peso molecular
d18	Índice de Wiener
d19	Índice de Randic (orden 0)
d20	Índice de Randic (orden 1)
d21	Índice de Randic (orden 2)
d22	Índice de Randic (orden 3)
d23	Índice de Kier&Hall (orden 0)

d24	Índice de Kier&Hall (orden 1)
d25	Índice de Kier&Hall (orden 2)
d26	Índice de Kier&Hall (orden 3)
d27	Índice de forma de Kier (orden 1)
d28	Índice de forma de Kier (orden 2)
d29	Índice de forma de Kier (orden 3)
d30	Contenido de información (orden 0)
d31	Contenido de información complementaria (orden 0)
d32	Contenido de información estructural (orden 0)
d33	Contenido de información de enlace (orden 0)
d34	Contenido de información (orden 1)
d35	Contenido de información complementaria (orden 1)
d36	Contenido de información estructural (orden 1)
d37	Contenido de información de enlace (orden 1)
d38	Contenido de información (orden 2)
d39	Contenido de información complementaria (orden 2)
d40	Contenido de información estructural (orden 2)
d41	Contenido de información de enlace (orden 2)
d42	3D-Índice de Wiener
d43	3D-Índice de Randic (orden 0)
d44	3D-Índice de Randic (orden 1)
d45	3D-Índice de Randic (orden 2)
d46	3D-Índice de Randic (orden 3)
d47	3D-Índice de Kier&Hall (orden 0)
d48	3D-Índice de Kier&Hall (orden 1)
d49	3D-Índice de Kier&Hall (orden 2)
d50	3D-Índice de Kier&Hall (orden 3)
d51	3D-Índice de forma de Kier (orden 1)
d52	3D-Índice de forma de Kier (orden 2)
d53	3D-Índice de forma de Kier (orden 3)
d54	3D-Contenido de información (orden 0)

d55	3D-Contenido de información complementaria (orden 0)
d56	3D-Contenido de información estructural (orden 0)
d57	3D-Contenido de información de enlace (orden 0)
d58	3D-Contenido de información (orden 1)
d59	3D-Contenido de información complementaria (orden 1)
d60	3D-Contenido de información estructural (orden 1)
d61	3D-Contenido de información de enlace (orden 1)
d62	3D-Contenido de información (orden 2)
d63	3D-Contenido de información complementaria (orden 2)
d64	3D-Contenido de información estructural (orden 2)
d65	3D-Contenido de información de enlace (orden 2)
d66	Momento de inercia A
d67	Momento de inercia B
d68	Momento de inercia C
d69	Área de la superficie molecular
d70	Volumen Molecular
d71	TMSA *Total molecular surface area* [Empirical PC]
d72	PPSA-1 *Partial positive surface area* [Empirical PC]
d73	PPSA-2 *Total charge weighted* PPSA [Empirical PC]
d74	PPSA-3 *Atomic charge weighted* PPSA [Empirical PC]
d75	PNSA-1 *Partial negative surface area* [Empirical PC]
d76	PNSA-2 *Total charge weighted* PNSA [Empirical PC]
d77	PNSA-3 *Atomic charge weighted* PNSA [Empirical PC]
d78	DPSA-1 *Difference in* CPSAs (PPSA1-PNSA1) [Empirical PC]
d79	DPSA-2 *Difference in* CPSAs (PPSA2-PNSA2) [Empirical PC]
d80	DPSA-3 *Difference in* CPSAs (PPSA3-PNSA3) [Empirical PC]
d81	FPSA-1 *Fractional* PPSA (PPSA-1/TMSA) [Empirical PC]
d82	FPSA-2 *Fractional* PPSA (PPSA-2/TMSA) [Empirical PC]
d83	FPSA-3 *Fractional* PPSA (PPSA-3/TMSA) [Empirical PC]
d84	FNSA-1 *Fractional* PNSA (PNSA-1/TMSA) [Empirical PC]
d85	FNSA-2 *Fractional* PNSA (PNSA-2/TMSA) [Empirical PC]

d86	FNSA-3 *Fractional* PNSA (PNSA-3/TMSA) [Empirical PC]
d87	E(HF)
d88	ZPE
d89	E(HF)+ZPE
d90	Momento dipolar
d91	Energía HOMO
d92	Energía LUMO
E(HF)	Energía de Hartree-Fock
eV	Electrón-Voltio
E	Energía total
GC-MS	*Gas Chromatography linked to Mass Spectrometry* (Cromatografía de gases acoplada a espectrometría de masas)
GTP	*Guanosine Triphosphate* (Guanosintrifosfato)
HOMO	*Highest Occupied Molecular Orbital* (Orbirtal molecular ocupado más alto)
IETS	*Inelastic Electron Tunneling Spectroscopy* (Espectroscopía de tunelamiento de electrones inelástica)
IP_3	*Inositol Triphosphate* (Inositoltrifosfato)
IR	Infrarrojo
LUMO	*Lowest Unoccupied Molecular Orbital* (Orbirtal molecular no ocupado más bajo)
MMFF	*Merck Molecular Force Field* (Campo de fuerza molecular de Merck)
PC	*Principal Component* (Componente principal)
PCA	*Principal Component Analysis* (Análisis de componentes principales)
RHF	*Restricted Hartree-Fock* (Hartree–Fock restringido)
OBP	*Odorant Binding Proteins* (proteínas enlazantes de odorantes)
PEST	*Property-Encoded Surface Translation* (Traducción de propiedades codificadas de superficie)
PVC	*Poly (vinyl chloride)* (Cloruro de polivinilo)
QSAR	*Quantitative Structure Activity-Relationship* (Relaciones cuantitativas estructura-actividad)

SANH	*Sequential Agglomerative Hierarchical Non-Overlapping* (Jerarquías de no solapamiento por aglomeración secuencial)
SOR	*Odor-Structure Relationship* (Relaciones estructura-olor)
STP	Condiciones estándar de temperatura y presión
TAE	*Transferable Atom Equivalente* (Átomos equivalentes transferibles)
U	Energía potencial
ZPE	*Zero point energy* (Energía del punto cero)

1. INTRODUCCIÓN

Entre los cinco sentidos existentes, el olfato, a pesar de ser ampliamente estudiado durante las últimas décadas, es el menos comprendido. La olfacción se inicia por medio de la interacción entre las moléculas odorantes y los receptores. Debido a esto, es de particular interés conocer cómo las propiedades químicas de los odorantes producen el olor percibido. El entendimiento de la olfacción a un nivel molecular facilitaría la predicción de los olores, objetivo importante de la industria de fragancias.

Para poder elucidar un mecanismo detallado de la olfacción, es necesario investigar la interacción entre las moléculas odorantes y los receptores olfativos a un nivel molecular. Desafortunadamente, las estructuras tridimensionales de los receptores olfativos no se encuentran disponibles hasta el momento. Los estudios realizados en el campo de las fragancias, la ciencia alimenticia y algunas áreas de farmacología [1], han utilizado métodos empíricos como las relaciones cuantitativas estructura-olor (QSAR). Estos estudios se volverán indudablemente útiles a medida que nuestro conocimiento sobre la estructura del receptor aumente y las técnicas de modelamiento molecular sean más realistas y completas. Entre tanto, la mayoría de trabajos proceden examinando únicamente las estructuras de los odorantes. No está claro, cuántos odorantes se han diseñado usando únicamente QSAR o como una herramienta principal para guiar su síntesis. Este parece ser el mejor lugar para establecer lo que, quizás, es el hecho más sorprendente de la relación "estructura-olor" (SOR): *no se han encontrado dos odorantes que posean exactamente el mismo olor*. Sin embargo, cientos de estos compuestos han sido descritos en la literatura y sus SORs han sido extensamente estudiadas [2].

Los análisis de las relaciones estructura-olor, usando métodos asistidos por computadores y técnicas de reconocimiento de patrones, pueden proveer un acercamiento al estudio de los compuestos odorantes. Un método que ha sido ampliamente estudiado en la industria de fragancias, es el 3D-QSAR. Sin embargo la

mayoría de estos métodos como CoMFA están basados en la “teoría de enlace”, la cual se encuentra ampliamente aceptada como el mecanismo de interacción entre las moléculas odorantes y los receptores olfativos. Bajo esta teoría, los odorantes son reconocidos y enlazados por los receptores, mientras que las moléculas que poseen un olfactóforo –un arreglo tridimensional de grupos funcionales asociados con un olor en particular– producen respuestas similares a los receptores, debido a que se enlazan en forma análoga. Existe una teoría competitiva que es totalmente diferente de la anterior, la “teoría vibracional”.

La teoría de vibración es una hipótesis antigua, la cual fue propuesta por primera vez por Dyson en 1938 [3]. Sin embargo, después de varios años de mantenerse en el olvido, fue retomada y modificada por Luca Turín en 1996 [4], quien propuso un mecanismo de tunelamiento de electrones como método de detección de las frecuencias de vibración de los compuestos odorantes, argumentado que existe una mejor correlación entre el olor de una molécula y su espectro vibracional, que entre el olor y la forma molecular.

Desde hace varios años, los investigadores han realizado un gran número de esfuerzos para correlacionar la estructura molecular con el carácter del olor, posteriormente, inspirados por el desarrollo de la teoría “llave y seguro” en otras áreas de la biología, empezaron a sugerir que la forma molecular podría determinar el olor. Por ejemplo, algunas moléculas que huelen a alcanfor son esféricas; sin embargo, la afirmación inversa, todas las moléculas esféricas huelen a alcanfor, no es cierta. Evidentemente, existen otros factores, los cuales deben ser considerados. Theimer y Davis [5] han discutido la importancia de áreas de secciones transversales y energías libres de desorción en relación con el olor. Amoore [6] ha reportado que la forma molecular, tamaño y naturaleza electrónica de las moléculas pueden ser relacionadas con la calidad del olor. Dravnieks y Laffort [7] han relacionado el olor con fuerzas intermoleculares. Sin embargo, a pesar de que existen bases de datos de cientos y miles de compuestos odorantes, no existe una relación clara entre la forma y el olor.

El corazón de este trabajo es encontrar un conjunto de descriptores moleculares, a partir de los cuales se pueda realizar la clasificación de 60 compuestos pertenecientes a las familias odorantes verde, almizcle, frutal y floral. Con este objetivo, se estudiaran diferentes clases de descriptores moleculares como: los descriptores electrónicos, topológicos, geométricos, topográficos, CPSA y constitucionales.

Para poder realizar la clasificación de los compuestos fragantes, se utilizaron dos metodologías no supervisadas, el análisis de componentes principales y las técnicas de agrupamiento. Estas técnicas fueron seleccionadas de acuerdo con los excelentes resultados que han presentado en el estudio de un gran número de compuestos químicos. A partir del método de PCA se busca establecer diferentes subespacios por medio del estudio individual de cada clase de descriptores, así como por la combinación de los mismos, para posteriormente poder establecer cuáles descriptores realizan la mejor clasificación y qué relaciones estructura-olor pueden ser encontradas.

Además, se espera estudiar la participación de los parámetros vibracionales en las relaciones “estructura-olor”, por medio de la estadística multivariable con base en pruebas de similitud entre moléculas, donde se espera contribuir de una manera eficaz en la gran investigación que se ha venido desarrollando sobre este tipo de estudios tan interesante e importante, y de paso, revisar la validez de los postulados propuestos por Luca Turín [4].

2. ESTADO DEL ARTE

Las impresiones de olor y sabor son el resultado de la interacción directa de algunos compuestos químicos con el sistema periférico de receptores. Los compuestos que poseen propiedades físicas como sabor, estimulan las papilas gustativas presentes en la lengua; por otro lado, los odorantes son compuestos volátiles que son transportados con el aire inhalado hacia el epitelio olfativo ubicado encima de las dos cavidades nasales, justamente debajo y entre los ojos [8].

Los compuestos con sabor no están "limitados" por el peso molecular o polaridad. Sin embargo, los odorantes, son moléculas pequeñas con pesos moleculares menores de 300 Da, que deben poseer además ciertas propiedades moleculares, tales como solubilidad parcial en agua, alta presión de vapor, baja polaridad, lipofilicidad y tensoactividad. Otra diferencia importante entre estos compuestos es su comportamiento sensorial; el gusto está compuesto de cuatro sensaciones, a saber: dulce, ácido, salado y amargo, las cuales son percibidas solamente en concentraciones altas, mientras que el olfato, es capaz de distinguir entre un número prácticamente infinito de compuestos químicos, muchos de los cuales están presentes en muy bajas concentraciones [9].

En esencia, el olor es una sensación primaria para los humanos así como para los animales. Desde un punto de vista evolutivo, es uno de los sentidos más antiguos. El olor (u olfacción) permite a los vertebrados y a otros organismos vivos con receptores olfativos identificar la comida, los compañeros, los depredadores, proporcionando tanto un placer sensual (el olor a flores y perfume), como advertencias de peligro (comida en descomposición, peligros químicos).

Para conocer y entender de una manera más profunda cómo se lleva a cabo la olfacción, se debe estudiar su fisiología, además de las diferentes teorías (hipótesis) de olfacción, que han surgido en años anteriores.

2.1 FISIOLOGÍA GENERAL DE LA OLFACCIÓN

La anatomía básica de la nariz y el sistema olfativo han sido sujeto de estudio desde muchos años atrás. En los mamíferos, la detección inicial de los olores se lleva a cabo en la parte posterior de la nariz, en la pequeña región llamada epitelio olfativo (**Figura 1**). En un adulto humano esta región posee un tamaño de 1-2 cm^2 y contiene aproximadamente 50 millones de células receptoras sensoriales primarias (neuronas). En el otro extremo de las neuronas, los axones se proyectan directamente a través del plato cribiforme hacia el bulbo olfativo en el cerebro, donde se unen por medio de la sinapsis con células secundarias conocidas como células mitrales.

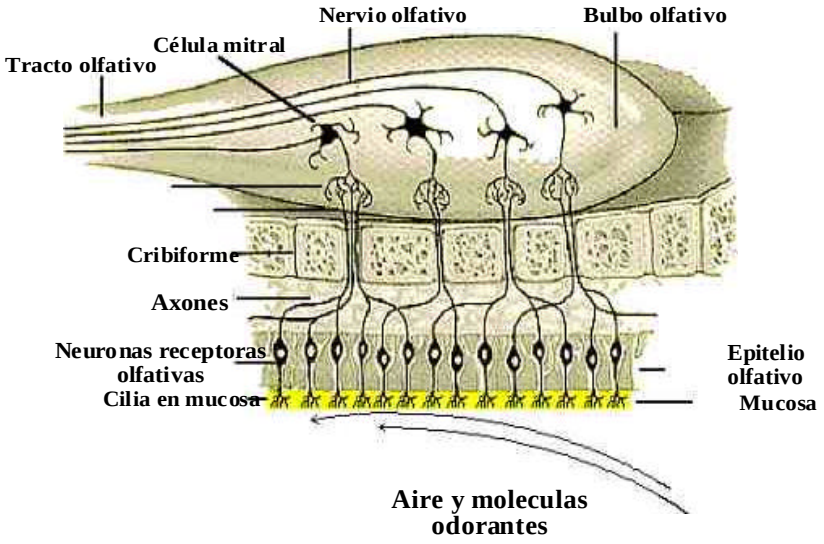


Figura 1. Esquema general de la región olfativa.

La estimulación sensorial es iniciada por el contacto directo de una molécula odorante con una parte de la membrana receptora en la cilia. Cuando esto sucede, el potencial estático de la membrana celular es interrumpido por una despolarización parcial y el estímulo es transformado en potenciales de acción. Esta estimulación, la cual contiene toda la información sobre las propiedades moleculares de un odorante en una forma codificada, es transmitida a través de la sinapsis. La sinapsis es complicada y consiste

en una simple célula mitral a la cual convergen 500 axones olfativos y desde aquí son transmitidas las señales a través del tracto olfativo hacia niveles más altos del sistema nervioso central, donde son decodificadas en percepciones olfativas. La descripción semántica es el resultado del análisis de la señal y su comparación con un olor patrón conocido.

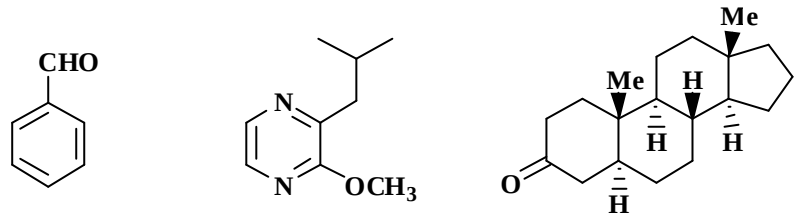
Recientemente, se han desarrollado investigaciones en la bioquímica de la olfacción, debido al descubrimiento del aumento del nivel de adenosinmonofosfato cíclico (cAMP), cuando la cilia es expuesta a ciertos odorantes [10]. Además, fue demostrado, que los odorantes que no afectan el nivel de cAMP, inducen un rápido cambio en los niveles de inositoltrifosfato (IP_3); esto demuestra la existencia de otros mecanismos de transducción, donde se encuentran involucrados los dos mensajeros secundarios. El mecanismo conocido muestra que un incremento en la cantidad de cAMP o IP_3 causa la apertura de canales de iones, donde el paso de iones positivos hacia la célula promueve un decrecimiento en el voltaje a través de la membrana celular, lo cual resulta en la generación de un impulso nervioso, que es posteriormente interpretado por el cerebro. El papel de estos mensajeros secundarios en la olfacción, ha sido resumido por Breer [11] y la diversidad de conductancia en la membrana olfativa por Dionne y Dubin [12].

Los descubrimientos anteriores, donde el olor induce los niveles de cAMP e IP_3, son acordes con el descubrimiento de que las proteínas-G se encuentran involucradas en la transducción olfativa. Las proteínas-G median con el guanosintrifosfato (GTP), el cual actúa como un intermediario entre los dos receptores, adenilato ciclasa y la fosfolipasa C, enzimas intracelulares responsables por la producción de cAMP e IP_3, respectivamente [13]. Por otro lado, en órganos no olfativos el tipo de proceso de transducción anterior está vinculado a proteínas receptoras las cuales se encuentran insertadas en la membrana celular y la cruzan en siete lugares. Es así, como fue postulado que los receptores de odorantes debían pertenecer a la familia de los receptores con siete hélices de las proteínas-G acopladas. Guiados por esta idea, Buck y Axel [14] descubrieron la familia de proteínas transmembranosas, las cuales se suponía

eran los receptores del olor, además de algunos de los genes que codifican estos receptores.

Los experimentos realizados por estos científicos sugirieron que la diferenciación de los olores involucra un gran número de receptores diferentes, donde cada uno de ellos es capaz de asociarse con un pequeño número de odorantes, al contrario de la hipótesis precedente, donde se creía que cada uno de los receptores de un pequeño número de ellos, era capaz de interaccionar con múltiples moléculas odorantes. Actualmente, el número del tipo de receptores está estimado en 1000.

En contraste con las proteínas receptoras de odorantes, se ha demostrado la existencia en la mucosa olfativa de una clase de proteínas llamadas proteínas enlazantes de odorantes (OBPs). La primera proteína de esta familia descubierta en 1979 por Steven Price y colaboradores [15], enlazaba el compuesto químico anisol; posteriormente, se ha encontrado un gran número de estas proteínas para otros compuestos odoríferos como benzaldehído (olor a almendra-cereza), 2-isobutil-3-metoxipirazina (olor a pimiento verde) y 5-α-androst-16-en-ona (olor a orina) (**Figura 2**).



Benzaldehído 2-isobutil-3-metoxipirazina 5-α-androst-16-en-ona

Figura 2. Ejemplo de algunas sustancias odorantes, para las cuales se han encontrado las proteínas enlazantes.

La afinidad de estas proteínas con los odorantes y las feromonas sugiere su papel importante en la percepción olfativa. Sin embargo, a pesar de la información acumulada durantes los años anteriores, su función aún no se conoce con exactitud. Una hipótesis

sugiere que las OBPs transportan las moléculas odorantes a través de la barrera acuosa presente en la mucosa hacia los receptores en la membrana olfativa. La base de esta hipótesis está sujeta a la similitud de las OBPs con una familia numerosa de proteínas portadoras de moléculas hidrofóbicas en varios fluidos biológicos llamadas lipocalinas, las cuales fueron caracterizadas por primera vez en los humanos en el año 2000 [16]. Una segunda hipótesis, es que las OBPs juegan un papel de filtro o buffer; atrapando varias moléculas odorantes cuando entran en la nariz en una concentración alta, ya que pueden inactivar los receptores olfativos debido a la saturación de los mismos.

En resumen, los descubrimientos en la bioquímica de la olfacción en los últimos años incluyen la identificación de proteínas y mensajeros secundarios, los cuales deben jugar algún papel en la olfacción, y es con base en estos estudios bioquímicos que se han propuesto varias teorías (hipótesis) de olfacción.

2.2 TEORÍAS DE OLFACCIÓN

La investigación en el área olfativa siempre ha iniciado con preguntas como las siguientes: ¿Cómo reconocemos y discriminamos entre miles de olores?, ¿Cuáles propiedades moleculares determinan el olor de un compuesto?, ¿Por qué en algunos casos, compuestos tan diferentes en estructura poseen el mismo olor y compuestos con estructuras similares poseen olores totalmente distintos?

Antes de 1980, el estudio bioquímico de la olfacción había sido virtualmente abandonado; debido a esto, los químicos se vieron forzados a postular mecanismos de olfacción basados en correlaciones observadas entre el olor y las propiedades moleculares. Las primeras teorías de olfacción eran solamente especulativas y, frecuentemente, suponían mecanismos físicos asombrosos como, por ejemplo, la emisión de rayos por parte de las moléculas odorantes, siendo éstos los que el ser humano podía percibir y describir [17].

Esfuerzos para correlacionar la estructura molecular con el carácter del olor son tan antiguos como la química sintética. Muchas teorías de SORs han sido propuestas en el pasado, pero debido a los avances científicos en la biología y, en especial, al descubrimiento de los receptores olfativos [18], en la actualidad, se mantienen vigentes dos tipos de teorías. El primer tipo de teoría, está basado en fragmentos de forma molecular u "odótopos" y el segundo, se fundamenta en vibraciones moleculares [4].

2.2.1 Teoría basada en la forma: odótopos. Varios tipos de enlaces como "enzima-sustrato" y "ligando-receptor" están basados en el reconocimiento molecular entre proteína y ligando. El reconocimiento depende de las interacciones que se presenten entre los dos sistemas, que pueden ser tanto atractivas, como repulsivas. Las interacciones atractivas son de naturaleza electrostática ya que ocurren entre cargas (dipolo, dipolo inducido, átomos capaces de formar enlaces débiles electrónicos), mientras que las repulsivas pueden ser de naturaleza electrostática o mecano-cuántica. Cualquier cambio en la estructura molecular (con algunas excepciones) afecta las características superficiales, involucradas en interacciones atractivas o repulsivas, y esto, en su turno, afecta lo que se llama forma molecular.

En 1946, Linus Pauling [19] indicó que el olor específico era debido a la forma molecular y el tamaño del compuesto. Esta idea fue retomada posteriormente por Moncrieff [20] y Amoore [6]. Amoore consideraba que todas las sensaciones de olores estaban basadas en la combinación de un número limitado de olores primarios, los cuales eran detectados por diferentes receptores en la nariz; originalmente, fueron sugeridos siete olores primarios, a saber: etéreo, alcanforáceo, mohoso, floral, menta, picante y pútrido. En la actualidad, esta teoría "sobrevive" en su forma modificada y refinada, la teoría de los llamados "odótopos".

De acuerdo con la teoría de odótopos [21], los receptores "enlazan" o reconocen solo una parte estructural característica (odótopo) y es, de esta manera, como el cerebro interpreta una estructura única para toda la molécula.

La principal evidencia de esta teoría está basada en estudios realizados tanto *in vivo,* como *in vitro*, donde se ha observado con una notable excepción, que diferentes receptores responden a más de un odorante [22].

De acuerdo con la teoría de odótopos, el olor de una molécula se debe a un patrón o característica, por ejemplo, la excitación relativa de un número de *N* receptores a los cuales se enlaza. Sin embargo, si se supone, que el receptor es del tipo de dos pasos ("encendido" y "apagado"); este esquema presenta un gran número de posibles combinaciones, dependiendo de la cantidad de odótopos que pueda presentar una molécula odorante.

2.2.2 Teoría vibracional. La idea que la nariz opera como un espectroscopio vibracional fue primero propuesta por Dyson [3], quien sugirió que los órganos olfativos podían detectar las vibraciones moleculares. Si todo el rango vibracional puede ser detectado (0-4000 cm^{-1}), así como lo hacen la espectroscopía de Raman e IR, la detección de los grupos funcionales podía ser explicada, ya que cada grupo posee una vibración característica definida, usualmente, por encima de 1000 cm^{-1}. Esta teoría fue modificada en una serie de artículos por Wright [23].

Lo atractivo de esta teoría, en principio, es que los espectros vibracionales "comparten" las siguientes tres propiedades con el sistema olfativo humano:

1. No existen dos espectros moleculares iguales, particularmente, en la región llamada "la región de huella digital"; 2. Muchos grupos funcionales se identifican fácilmente por sus frecuencias específicas vibracionales; 3. Un sistema que utiliza una propiedad física como la vibración, no dependerá de si una molécula se ha olido o no anteriormente y siempre se obtienen los mismos resultados, *i.e.* no está basada en un repertorio de moléculas existentes o establecidas.

Sin embargo, existían varias dificultades las cuales asediaron la teoría de vibración y, finalmente, causaron su fallecimiento hace veinte años.

Entre estas dificultades se pueden nombrar algunas, a saber: 1. Los enantiómeros poseen un espectro de vibración idéntico en solución, pero algunas veces poseen diferentes olores [24]. Wrigth se opuso a ésto, haciendo énfasis que mientras los espectroscopios del laboratorio eran aquirales y, de esta manera, incapaces de distinguir entre los enantiómeros, un receptor basado en proteínas sería intrínsecamente quiral y respondería así diferentemente a los enantiómeros. 2. No se encontró ningún mecanismo plausible para un espectroscopio basado en proteínas. Wrigth supuso que los receptores eran sensores de vibración mecánicos, y que solamente podían percibir vibraciones excitadas por movimientos térmicos a temperatura ambiente; además, posteriormente restringió su investigación de correlación entre estructura y olor a la región por debajo de 600 cm^{-1} del espectro vibracional.

Esta situación cambió en 1996 cuando Luca Turín [4] propuso que el tunelamiento de electrones podría ser un mecanismo viable para que las proteínas actuaran como un espectroscopio vibracional. El espectro vibracional (100-4000 cm^{-1}) está compuesto por dos "mitades". Por encima de 1700 cm^{-1}, los modos vibracionales en su mayoría, se deben a la tensión de pares de átomos, *e.g.* C=O, C≡N, S-H, C-H, N-H y O-H. Por debajo de 1700 cm^{-1}, los modos de vibración son complejos e involucran tres o más átomos. La teoría vibracional explica nuestra habilidad para reconocer la presencia de los modos de vibración de tensión de los grupos funcionales presentes en los odorantes en la "mitad" superior del espectro; por ejemplo, el hecho que infaliblemente distingamos los grupos SH y OH. Por otro lado, la otra "mitad" del espectro determinará el olor de diferentes odorantes que posean el mismo grupo funcional.

La teoría vibracional puede ser probada en principio. Si las energías vibracionales e intensidades pueden ser calculadas exactamente, los espectros de las moléculas, así como son percibidas por nuestra nariz, podrían ser determinados y comparados.

2.3 UN "ESPECTROSCOPIO" BIOLÓGICO

Por tunelamiento de electrones se entiende el paso de electrones de una región "clásicamente" permitida a otra, a través de una región dónde se encuentra "clásicamente" prohibida la existencia del electrón (**Figura 3**). La región o barrera prohibida es aquella donde la energía potencial, U, es mayor que la energía clásica total, E. Si la partícula se mueve de una región permitida a la otra, debe sufrir tunelamiento a través de la barrera de potencial. Sin embrago, no existe ninguna prohibición en la mecánica cuántica para que esto ocurra.

Si U y E son ambos grandes (varios eV) y la distancia d es muy pequeña (del orden de 1 nm), hay una probabilidad finita que el electrón incidente emerja a la región "clásicamente" permitida II. Si las energías incidentes y emergentes son las mismas, este fenómeno físico se denomina "tunelamiento elástico"; por el contrario, si las energías incidentes y emergente son diferentes se llama "tunelamiento inelástico".

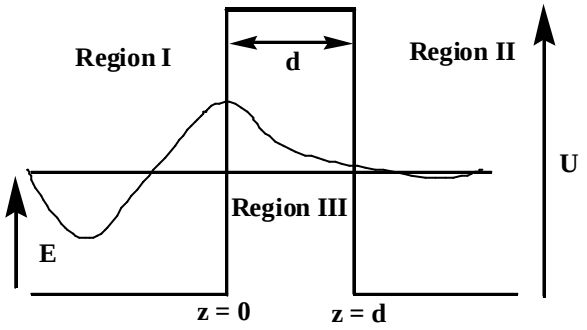


Figura 3. Representación gráfica del tunelamiento de electrones. U: energía potencial, E: energía clásica total, d: distancia.

La espectroscopía de tunelamiento de electrones inelástica (IETS) es una forma no-óptica de la espectroscopía vibracional [25], donde el fenómeno físico transcurre cuando los electrones sufren tunelamiento a través de una distancia estrecha entre electrodos

metálicos. Cuando no se encuentra una molécula en esta distancia, los electrones cruzan la brecha manteniendo constante su energía y la corriente de tunelamiento es proporcional al solapamiento entre los estados electrónicos llenos y vacíos en los metales. Por otro lado, si una molécula se encuentra entre los electrodos, los electrones serán esparcidos por las cargas parciales presentes en los átomos que constituyen la molécula, perdiendo energía y excitando uno de los modos vibracionales de la molécula. Cuando esto pasa, los electrones pueden seguir un camino indirecto, primero, excitando los modos vibracionales moleculares y, posteriormente, sufriendo tunelamineto hacia el segundo metal pero, teniendo una energía más baja. El nuevo camino de tunelamiento causa un aumento en la conductancia de la unión.

A pesar de que en la biología no encontramos conductores metálicos, la transferencia de electrones es ubicua. El desarrollo de la IETS con proteínas involucra la adición y la sustracción de electrones con energías bien definidas en ambos lados del odorante, el cual actúa como distancia de tunelamiento (**Figura 4**).

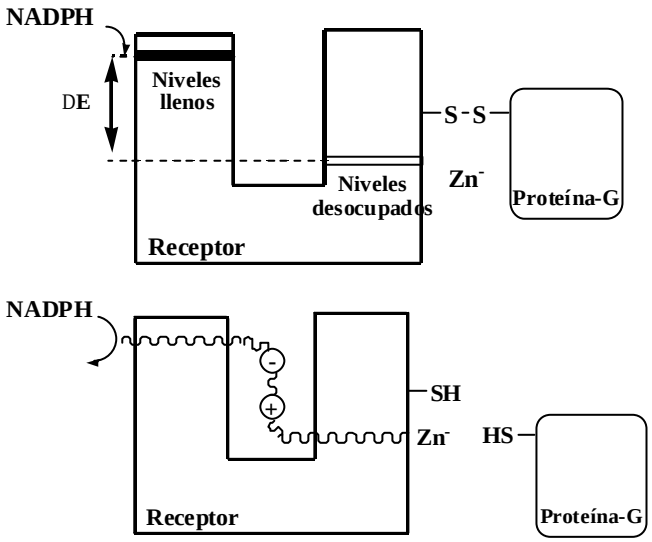


Figura 4. Representación del mecanismo de transducción propuesto por Luca Turin [4].

Si una molécula está presente entre la fuente de electrones y el aceptor y, además, la molécula vibra, entonces (tomando la energía del cuanto vibracional como E) el

tunelamiento indirecto puede ocurrir si existe un nivel de energía en el donor con energía E por encima del aceptor. Después del tunelamineto, la molécula tendrá una energía vibracional más alta en E. En otras palabras, el tunelamineto de electrones ocurre solamente cuando la energía vibracional molecular E iguala la diferencia de energía entre el nivel de energía del donor y el nivel de energía del aceptor. Entonces, el receptor opera como un espectrómetro que permite detectar energías bien definidas, E.

Turin y Yoshii han publicado en un artículo reciente [26] los pros y los contras de las teorías de odótopos y vibración. Sus argumentos se analizan a continuación:

2.4 TEORÍA DE ODÓTOPOS *Vs* TEORÍA DE VIBRACIONES

En un campo tan amplio como el de relaciones “estructura-olor”, las observaciones pueden dar el soporte a cualquier teoría o hipótesis. A continuación, se revisan algunas de estas observaciones, las cuales, potencialmente, pueden habilitar o contradecir cualquiera de las dos teorías.

2.4.1 Olfacción de grupos químicos. La habilidad de los humanos para poder reconocer los diferentes grupos funcionales ha sido objeto de muy poco estudio [27], el caso de los tioles (-SH-) es un caso familiar, pero otros grupos funcionales como los nitrilo (-CN), isonitrilo (-NC), oxima (-NOH), nitro (NO_2) y aldehído (CHO), solamente pueden ser identificados, una vez el carácter de olor que confiere el grupo funcional, sea conocido. Por ejemplo, cuando el grupo nitrilo es usado químicamente para reemplazar al grupo formilo, imparte un carácter metálico a cualquier olor, las oximas proporcionan un carácter verde-alcanforáceo e isonotrilos un carácter metálico intenso y desagradable.

2.4.1.1 Grupos funcionales como odótopos. La teoría de odótopos solamente puede explicar lo expuesto anteriormente, suponiendo que cada grupo funcional es un odótopo. De acuerdo con los estudios “estructura-olor”, esta hipótesis se explica con base en factores electrónicos y no según factores estéricos. La idea central se basa en un mecanismo de reconocimiento, el cual, debe ser sensible a la distribución electrónica (energías de orbitales, densidad de carga, etc.) del grupo funcional y no a su tamaño, debido a que varios grupos funcionales poseen tamaños similares.

El caso más llamativo y el mejor conocido, es el del grupo -SH (tiol, mercaptano) que imparte a cualquier molécula, sin tener en cuenta su forma, un carácter olfativo llamado apropiadamente "sulfuráceo". Lo interesante desde el punto de vista de reconocimiento molecular, es que este grupo funcional, junto con los nombrados anteriormente, constituye el grupo de odótopos pequeños capaces de formar uno o hasta dos puentes de hidrógeno. De acuerdo con esto, es difícil de entender cómo se detectan con absoluta exactitud los diferentes grupos funcionales. Por ejemplo, el hecho que los alcoholes (–OH) nunca huelan como tioles (–SH), a cualquier concentración, no concuerda con el mecanismo conocido de reconocimiento molecular.

Otro factor interesante se presenta con el grupo tioéter (–S–), el cual puede ser reemplazado por el enlace doble carbono-carbono (–C=C–), con un pequeño cambio en el olor sin carácter sulfuráceo. Ello sugiere, que las propiedades electrónicas del azufre no son suficientes para el reconocimiento molecular (**Figura 5**).

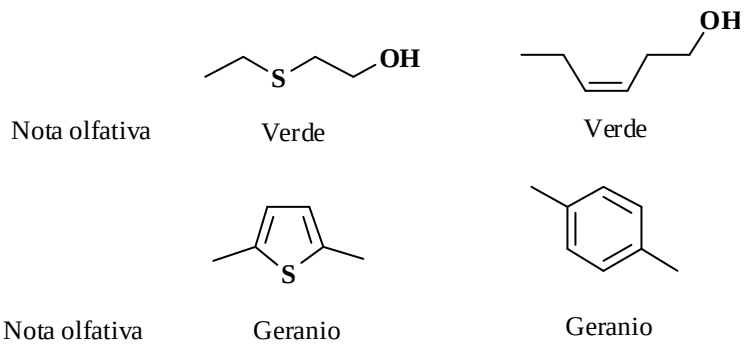


Figura 5. El reemplazo del enlace C=C con un átomo de azufre en algunos compuestos no cambia sus propiedades olfativas.

El tamaño pequeño de este grupo de odótopos restringe el número de interacciones repulsivas. Por consiguiente, las moléculas pequeñas deben enlazarse con varios grados de afinidad a muchos receptores de odótopos; por ende, las moléculas pequeñas deberían poseer olores similares, particularmente, a concentraciones altas. Ése no es el caso. Por ejemplo, moléculas pequeñas como el metilnitrilo y los metilnitratos huelen absolutamente diferente en cualquier concentración en que se encuentren.

2.4.1.2 Grupos funcionales y la teoría vibracional. En contraste, la diferencia de olor de los grupos funcionales es una característica "natural" de la teoría de vibración. Cualquier espectroscopista de IR sabe, que el reconocimiento de los grupos funcionales por medio de sus frecuencias de vibración es inequívoco.

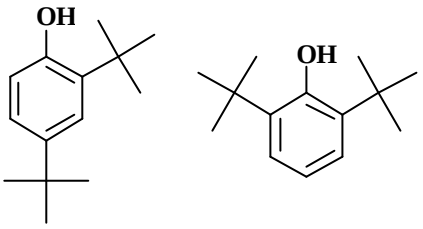
El ejemplo más relevante es nuevamente el grupo –SH, el cual posee una vibración única de tensión alrededor de 2550 cm^{-1}. Una predicción de la teoría de vibración es que cualquier otro grupo, que posea la misma frecuencia de vibración, deberá poseer un olor sulfuráceo. Éste es el caso en la mayoría de los boranos, donde el enlace terminal B-H en estos compuestos tiene una frecuencia de vibración en el mismo rango de los tioles, como fue comprobado en 1912 por Stock [28], quien fue el primero en sintetizar varios de estos compuestos. Sin embrago, existe muy poco en común, químicamente hablando, entre los grupos BH y SH.

Por otro lado, la teoría vibracional explica las similitudes del olor entre las moléculas. El ejemplo bien conocido es el reemplazo de grupos nitrilo por el grupo aldehído; las similitudes odoríferas se explican de acuerdo con la proximidad de sus frecuencias vibracionales de tensión [4].

2.4.1.3 Grupos funcionales impedidos. Las moléculas pueden ser diseñadas, en principio, para establecer, si los grupos funcionales se perciben como odótopos o por sus vibraciones. Supongamos, por ejemplo, que un grupo funcional que posee un olor distintivo se encuentra presente en una molécula, pero está "obstaculizado" de tal

manera, que sea inaccesible para el reconocimiento molecular. De acuerdo con la teoría vibracional y el mecanismo de tunelamiento de electrones, la molécula, sin embargo, debe poseer olor, mientras que describiéndola con base en la teoría del odótopo, no lo tendría.

Como primera aproximación, se puede considerar a los fenoles estéricamente impedidos. La presencia de un grupo OH en un anillo de benceno sustituido proporciona un olor característico a la molécula, *i.e.* "el olor fenólico". Una vez más, si se supone que el grupo OH es un odótopo, mientras menos accesible sea éste para su reconocimiento molecular, debería perder más su nota olfativa. Esta idea se prueba fácilmente comparando el olor de los derivados de di-*ter*-butil fenol, los cuales son comercialmente accesibles. Los resultados contradicen la teoría del odótopo. Ya que el 2,6-di-*ter*-butil fenol, donde el grupo OH está fuertemente impedido, huele similar al derivado 2,4-, en el cual el grupo OH está más accesible para su reconocimiento molecular (**Figura 6**).



A B

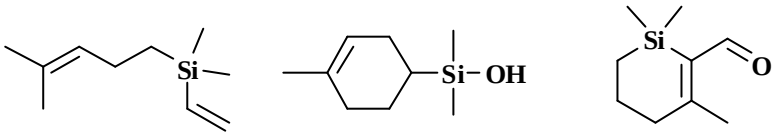
Figura 6. Estructuras moleculares de los compuestos A: 2,4-di-*ter*-butilfenol y B: 2,6-di-*ter*-butilfenol.

2.4.2 Moléculas isostéricas. Una prueba contundente de la teoría vibracional contra la teoría del odótopo se lograría, comparando el olor de moléculas idénticas de acuerdo con su composición de átomos, forma, peso, distribución de electrones y demás propiedades físicas, que sólo difieren en las frecuencias de sus vibraciones. Esto, sin embrago, es imposible de alcanzar, pero se puede lograr un acercamiento bastante

apropiado, sustituyendo uno de los elementos, *e.g.* Ni por Fe, dentro de un metaloceno, Si por el C, o por la substitución isotópica (D por H) en un odorante normal.

En la literatura existen varios ejemplos de moléculas que poseen formas similares, pero olores muy diferentes. Quizás, la serie de compuestos más conocida fue investigada por Wannagat y colaboradores [29], donde el carbono fue reemplazado por Si, Ge y Sn.

2.4.2.1 Compuestos de silicio. Debido a la polaridad alta e inestabilidad del grupo Si-H, sólo pueden sustituirse por Si aquellos átomos de carbono, que están unidos a otros carbonos. La geometría del enlace Si-C es "tetraédrica", pero a pesar de la similitud geométrica global, los compuestos de silicio difieren de sus compuestos padre de carbono. El enlace Si-C es de longitud 1.8 Å comparado con 1.5 Å para el enlace C-C; además, el enlace Si-C es más polar.



A B C

Figura 7. Estructuras moleculares de A: Sila-linalool, B: Sila-terpineol y C: Sila-ciclocitral.

En todos los casos de sustitución, se presenta un cambio en el olor, algunas veces muy sutil o casi imperceptible, mientras que en otras, totalmente marcado, por ejemplo, la sustitución en el linalool, cuyo olor se describe como "floral-leñoso con una nota cítrica", para obtener el linalool-silano (**A**), resulta en el cambio de olor a "jacinto, dulce". Análogamente, el terpineol-silano (**B**) huele más "como a *muguet*" y menos a lila, como en el compuesto padre, mientras que, el carvomenteno-silano (**C**) huele "similar" al compuesto precursor.

De acuerdo con la teoría vibracional, la gran diferencia de olor, sin importar la similitud en cuanto a la forma se refiere, es fácilmente explicada debido a que cualquier vibración molecular será afectada por los cambios en la masa de acuerdo con la serie C-Si-Ge-Sn.

2.4.2.2 Sustitución isotópica. La sustitución isotópica es, en principio, la mejor manera de obtener compuestos isostéricos que difieren "sólo" en las vibraciones moleculares. El "sólo" en la frase anterior ilustra el hecho que existen diferencias sutiles en las propiedades físicas y químicas, comparadas con el compuesto padre. La hidrofobicidad será ligeramente distinta debido a la pequeña diferencia en el tamaño, así como la polarizabilidad de la nube de electrones, que rodea los núcleos más pesados. Además, el rango de conformaciones que el compuesto adquirirá durante el movimiento térmico, será diferente, porque las masas alteradas responderán de manera diferente a las excitaciones térmicas. No obstante, estos efectos son muy pequeños. En contraste, los efectos en las vibraciones moleculares pueden ser grandes, la substitución de H por D reduce las frecuencias de tensión, por ejemplo, para CH de 3000 a 2200 cm^{-1} [26].

Es importante tener en cuenta que estos experimentos requieren que la sustitución isotópica se lleve a cabo en: a. Protones no intercambiables, de otra manera, el D se intercambiaría rápidamente por H; b. Los hidrógenos a intercambiar no deben estar involucrados en enlaces tipo puente de hidrógeno, ya que éstos serán afectados ligeramente por la sustitución.

Un ejemplo claro, se realizó con el cálculo de los espectros vibracionales de la acetofenona y la acetofenona-d_8 (**Figura 8**), lo cual sugirió que las diferencias de olor debían ser perceptibles. En los espectros vibracionales se puede apreciar la ausencia de la vibración de flexión C-H alrededor de 1400 cm^{-1} de la acetofenona-d_8, mientras que la vibración de tensión C-H se encuentra en la región ~2200 cm^{-1}, lo cual, normalmente indicaría la presencia de un grupo nitrilo (CN) [26].

La diferencia de olor entre los dos compuestos fue marcada, la acetofenona-d_8 huele más a fruta y tiene menor carácter de tolueno, que la acetofenona no sustituida, además posee un olor muy fuerte a almendra amarga. Esto último es particularmente interesante, porque el espectro de acetofenona-d_8 está estrechamente relacionado con el del benzonitrilo, un odorante con olor fuerte a almendra amarga.

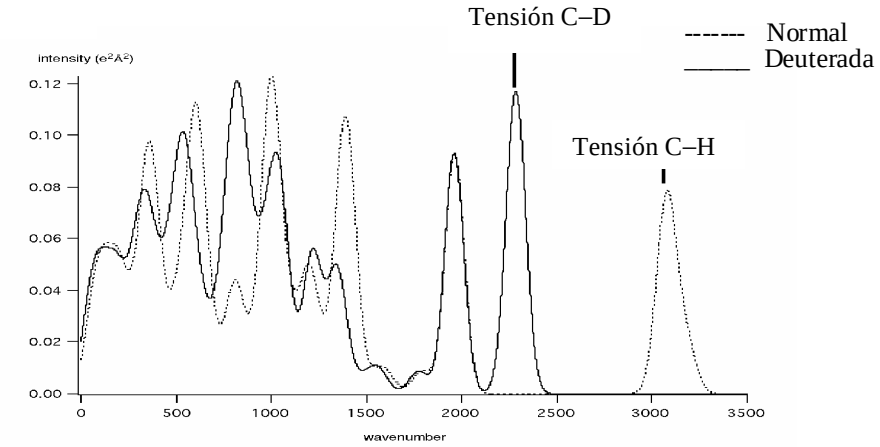


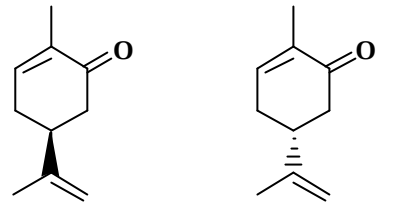
Figura 8. Espectros vibracionales de la acetofenona y acetofenona-d_8.

En resumen, la evidencia disponible de los experimentos de intercambio isotópico parece ser incoherente con la teoría del odótopo y se encuentra en buena concordancia con la teoría vibracional. Para que estos resultados puedan explicarse por medio de la teoría del odótopo, se deben postular factores adicionales; por ejemplo, hay que tener en cuenta una sensibilidad diferencial muy alta de los receptores del odótopo a pequeños cambios en la hidrofobicidad del odorante.

2.4.3 Enantiómeros. Muchos pares de enantiómeros de odorantes, poseen olores similares, pero existen varios ejemplos de enantiómeros que huelen completamente diferente. En una recopilación de 277 pares de enantiómeros conocidos [30] se encontró que el 59% de ellos poseían olores similares (un descriptor en común), el 5% "idénticos" (todos los descriptores en común), el 17% diferentes (ningún descriptor en común) y el resto 19% son desconocidos. Aun cuando todos los enantiómeros

desconocidos posean olores completamente diferentes, la mayoría de enantiómeros olerán similarmente.

Un ejemplo, de pares de enantiómeros que huelen distintamente, son las R- y S-carvonas: la R-carvona huele a menta y la S-carvona a alcaravea (**Figura 9**) [4].



S-(+)-Carvona R-(-)-Carvona

Figura 9. Estructuras de los dos enantiómeros de la carvona.

Las diferencias en el olor de los enantiómeros han sido consideradas en el pasado como la mayor evidencia en contra de la teoría vibracional olfativa, debido a que los espectros infrarrojo de enantiómeros en solución sondeados con luz no polarizada son idénticos. En contraste, si la absorción IR de un sólido regular (cristal) es sondeada con luz polarizada, entonces, el espectro depende de la orientación relativa de los dipolos moleculares en el cristal respecto al plano de la luz polarizada.

Turin [4] ha argumentado, que "el espectroscopio biológico" ejemplifica el último caso mencionado, ya que posee una característica inusual; presenta efectos de polarización. Por lo tanto, el olor de las carvonas puede ser explicado por medio de este efecto. De acuerdo con esto, una parte de la molécula puede ser "invisible" al receptor debido a una orientación desventajosa de un grupo dipolar.

La inspección de las configuraciones más estables calculadas de los enantiómeros de la carvona, muestran que si el anillo de ciclohexeno y su sustituyente isopropenílico se enlazan al receptor en una orientación fija con respecto al eje principal de la molécula, los dos grupos carbonilo se orientan con un ángulo de 120 grados entre si (**Figura 10**).

Además, los grupos carbonílicos poseen las cargas parciales más grandes, por lo tanto, deben dominar la dispersión de electrones y el carácter del olor. Si en la S-carvona (A), el carbonilo se alinea correctamente con respecto a la ruta del electrón y es detectado, entonces en la R-carvona (B) el carbonilo está aproximadamente orientado con los ángulos adecuados en el camino de los electrones, de tal manera, que son dispersados y no alcanzan el sitio del receptor (**Figura 10**). El resto de la molécula se reconoce idénticamente en ambos casos, porque ningún otro dipolo predominante está presente.

Como existen varios odorantes que poseen olor a menta y no contienen un grupo carbonilo, se puede sospechar que la R-carvona es el isómero donde el grupo C=O no es detectado. Esta predicción fue elucidada por Weyerstahl [31] en un experimento donde el grupo C=O se reemplazó por el C-OH en el (-)-carveol y, como resultado, se obtuvo el correspondiente alcohol, el cual sí retuvo el olor a menta. En contraste, el mismo reemplazo en la (+)-carvona alteró el carácter del olor completamente.

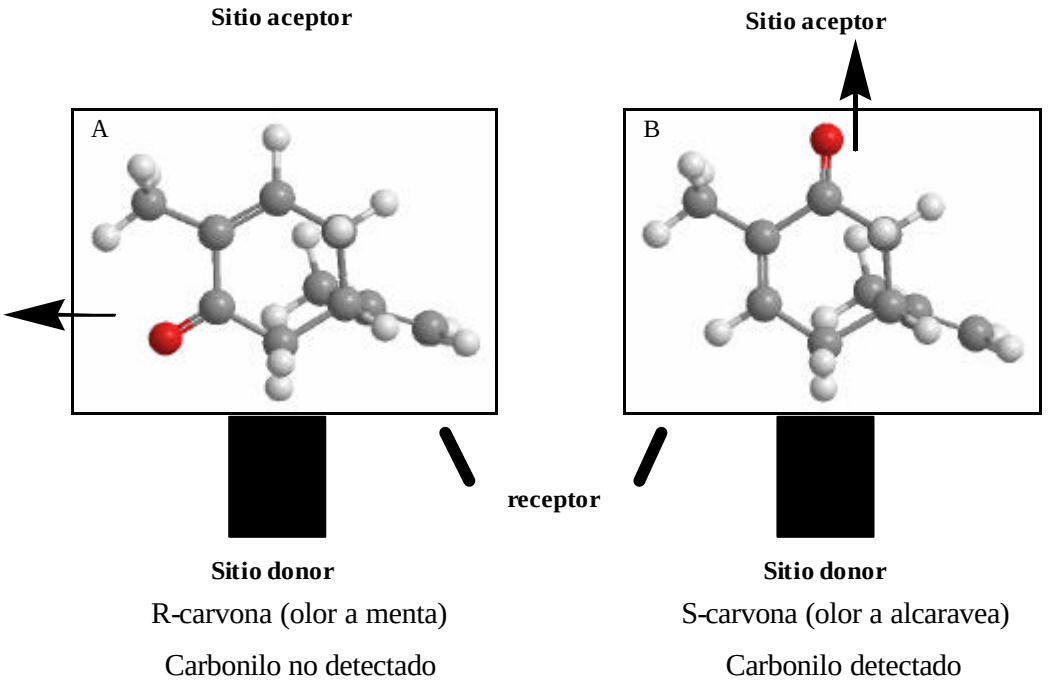


Figura 10. Diferenciación en la recepción de los enantiómeros de la carvona.

2.5 MÉTODOS DE ESTADÍSTICA MULTIVARIABLE

Los análisis de datos por medio de la estadística multivariable, básicamente el análisis de componentes principales (PCA) [32-36], el análisis de factores [37-39] y el análisis de agrupamiento [39-42], se han convertido en una herramienta poderosa en varias investigaciones químicas, debido a su alto poder predictivo y su capacidad de obtención de resultados, los cuales no son accesibles por otros métodos.

2.5.1. Análisis de componentes principales. El análisis de componentes principales (PCA) es uno de los métodos más comúnmente utilizados para proveer una descripción condensada y establecer métodos de variación en conjuntos de datos con muchas variables.

El PCA consiste principalmente en la rotación y transformación de los *n* ejes iniciales, donde cada eje está representado por una variable, en un nuevo conjunto de ejes llamados componentes principales, PC. Esta transformación es realizada de tal forma que los nuevos ejes obtenidos permanezcan en la dirección de la máxima varianza de los datos, pero además, cumplan el requisito fundamental de la ortogonalidad. Esto quiere decir, que todos los componentes principales son ortogonales, uno respecto al otro, y no se encuentran correlacionados.

Usualmente se presenta el caso donde el número de las nuevas variables, *p*, necesarias para describir la mayoría de información, es menor que el número de las variables originales, *n*. De esta manera, el PCA se puede utilizar como un método para reducir la dimensionalidad de los datos. Por otro lado, el análisis de componentes principales puede revelar aquellas variables, o combinación de variables, que determinan una estructura inherente en los datos, la cual puede ser interpretada en términos químicos o fisicoquímicos.

La obtención de los diferentes componentes principales a partir de un conjunto de datos que contiene varias variables, se puede llevar a cabo por medio de la combinación lineal de las variables originales como se muestra a continuación:

$$PC_1 = a_{1,1}n_1 + a_{1,2}n_2 + \ldots\ldots\ldots a_{1,n}n_n$$
$$PC_2 = a_{2,1}n_1 + a_{2,2}n_2 + \ldots\ldots\ldots a_{2,n}n_n \qquad (1)$$
$$PC_q = a_{q,1}n_1 + a_{q,2}n_2 + \ldots\ldots\ldots a_{q,n}n_n$$

Los coeficientes $a_{i,j}$, o factores de aporte, representan la contribución de cada variable original (n_1-n_n) a cada componente principal (1-q). Por otro lado, el signo particular asociado con cada coeficiente indica si la variable realiza un aporte negativo o positivo, mientras que su magnitud indica el grado de contribución a cada componente.

Como los PC son nuevas variables, es posible calcular valores individuales de estos componentes para cada uno de los objetos (compuestos o muestras) presentes en el conjunto de datos originales, para producir un nuevo conjunto de datos (reconstruidos pero relacionados). Los números presentes en este nuevo conjunto de datos son conocidos como valores de componentes principales (*scores*). Este proceso se muestra en forma de diagrama en la **Figura 11**.

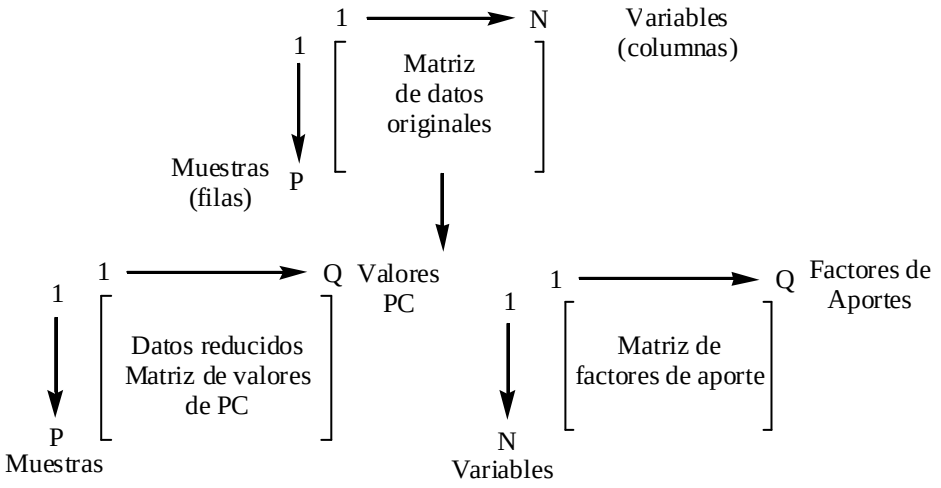


Figura 11. Ilustración del proceso de análisis de componentes principales.

La utilidad del análisis de componentes principales en la reducción de la dimensionalidad, radica en el hecho que los factores principales son generados de tal manera que expliquen la máxima cantidad de la varianza. Por lo general, en la mayoría de los resultados obtenidos por medio del PCA los tres primeros componentes principales contienen más del 80% de la información original.

En la **Figura 12** se puede observar un ejemplo del PCA en la industria alimenticia, donde se aprecia la gráfica de los valores de cada compuesto, representados por los dos primeros PC derivados de un conjunto de datos de 15 variables para 34 muestras de jugos de frutas [43].

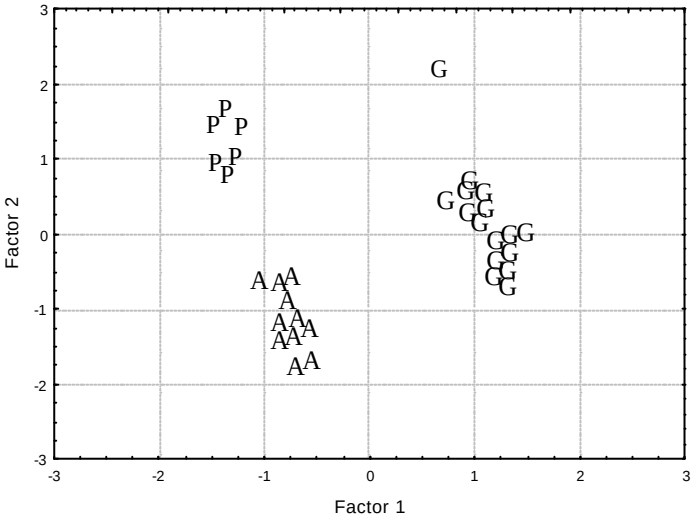


Figura 12. Dispersión de los valores de los componentes principales para las muestras de jugos. A: Manzana; P: Piña; G: Uva.

Algunas de las variables incluidas fueron el pH, fenoles totales, azúcares reducidos, nitrógeno total, contenido de glucosa y el número de formol. Entre las 34 muestras de jugo, se encontraban 17 de uva, 11 de manzana y 6 de piña. El primer PC está relacionado con el grado de azúcar y separa las muestras de jugo de uva de las de jugos de manzana y de piña. El segundo componente principal, el cual separa las muestras de manzana con las de piña, se encuentra altamente correlacionado con la relación glucosa: fructosa, nitrógeno total y el número de formol.

En este ejemplo es posible intentar dar algún “significado” químico al primer PC, ya que describe el grado de azúcar, sin embargo, siempre debe tenerse en cuenta que los PCs son construcciones matemáticas, los cuales no poseen necesariamente significado físico.

Los análisis de componentes principales, también han sido utilizados analizando perfiles de espectros digitalizados [44]. Un ejemplo de esta aplicación se encuentra en el estudio realizado por Ian y James, quienes han estudiado el espectro infrarrojo de 21 muestras de los polímeros acrílicos, PVC, estireno y nylon. Cada espectro fue normalizado en la banda de absorción más intensa con el fin de remover efectos del grosor de las muestras. La matriz resultante de 21 x 216 fue sometida al análisis de componentes principales. Los resultados obtenidos por medio del PCA mostraron que los tres primeros componentes principales contenían más del 91% de la varianza total de los datos originales.

En la **Figura 13** se representan los tres factores principales, donde se puede evidenciar la clasificación de los cuatro polímeros. El primer componente principal separa los polímeros acrílicos del resto. El PC_2 provee dos particiones, tanto el nylon como el PVC son separados del acrílico y del estireno y el PC_3 permite la separación entre el estireno y los demás.

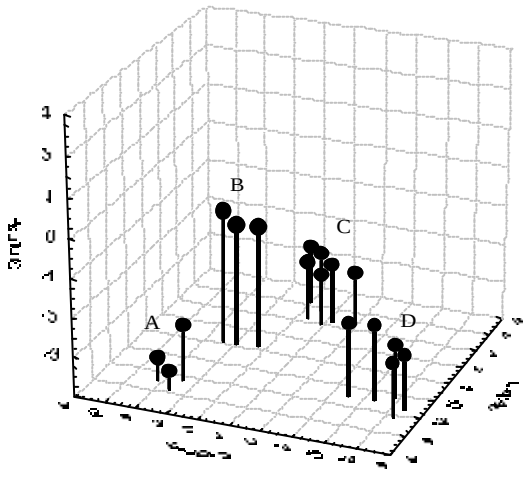


Figura 13. Gráfica de dispersión en tres dimensiones derivada de los espectros de 21 polímeros. A: Nailon; B: Estireno; C: PVC y D: Polímeros acrílicos.

Existen algunas publicaciones donde se ha realizado la clasificación de compuestos o muestras, especialmente a partir de conjuntos de datos obtenidos por medio de GC-MS [45,46]. En la **Figura 14** se puede observar la gráfica en tres dimensiones de los tres primeros componentes principales calculados a partir del análisis de GC-MS de muestras de aromas naturales. Las diferentes muestras designadas pertenecientes a distintos tipos de aroma de naranja, fueron provistas por seis diferentes casas fabricantes de sabores. Estos aromas de naranja pudieron ser clasificados en nueve categorías, como se indica en la gráfica de acuerdo con la simbología establecida.

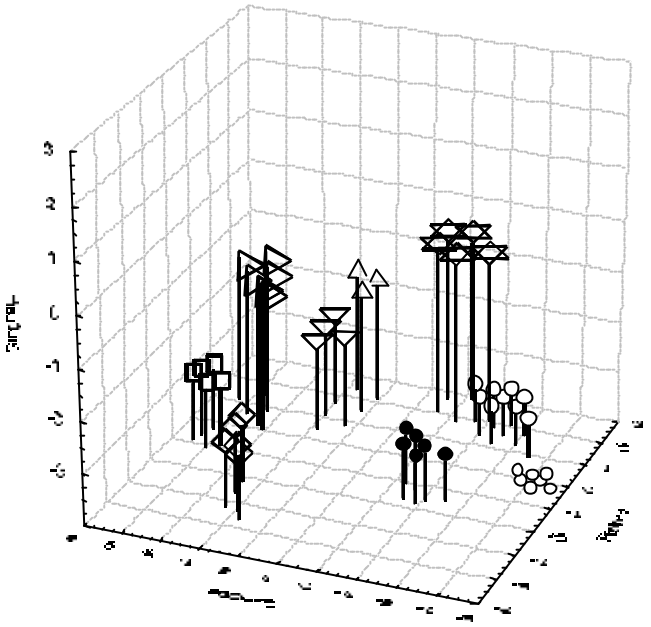


Figura 14. Clasificación de los aromas de naranja.

Una metodología para facilitar el diseño de nuevos compuestos odorantes ha sido publicada recientemente [47]. En esta investigación se realizó la clasificación de algunos compuestos pertenecientes a la familia Almizcle. En la **Figura 15** se observan las dos clases estructurales principales presentes en el estudio.

Todos los compuestos utilizados en este estudio fueron tomados a partir de la literatura especializada. La escogencia de los compuestos que no presentan olor se realizó con

base en las estructuras de los compuestos de la familia Almizcle, buscando una similitud estructural. Para la descripción de cada uno de los compuestos se utilizó la metodología de átomos equivalentes transferibles de Breneman (TAE) por medio de la cual se obtuvieron los descriptores de traducción de propiedades codificadas de superficie (PEST). La escogencia de los descriptores se debe especialmente a que éstos se correlacionan con la forma molecular y con las propiedades electrónicas, así como con los modos de interacción molecular.

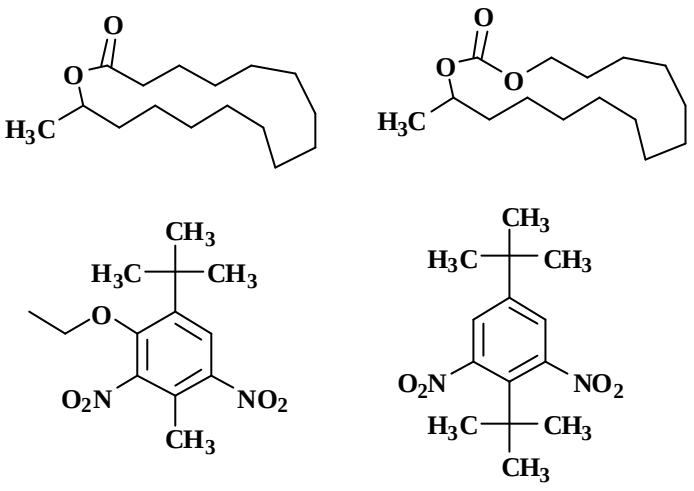


Figura 15. Compuestos que presentan estructuras similares pero se diferencian según su nota olfativa.

A partir de los resultados obtenidos por medio del análisis de componentes principales, se puede apreciar la separación se los compuestos Almizcle de los que no presentan olor, (**Figura 16**). En esta gráfica se representan los dos primeros PC, los cuales contienen el 52% de la varianza total. Por otro lado, cuando se clasificaron los diferentes compuestos de acuerdo con su estructura en 1: No Almizcle; 2: Compuestos aromáticos nitro no Almizcle; 3: Almizcle y 4; Compuestos aromáticos nitro Almizcle; se puede observar una clasificación de las cuatro clases de compuestos por medio de la representación de los dos primeros componentes principales (**Figura 17**).

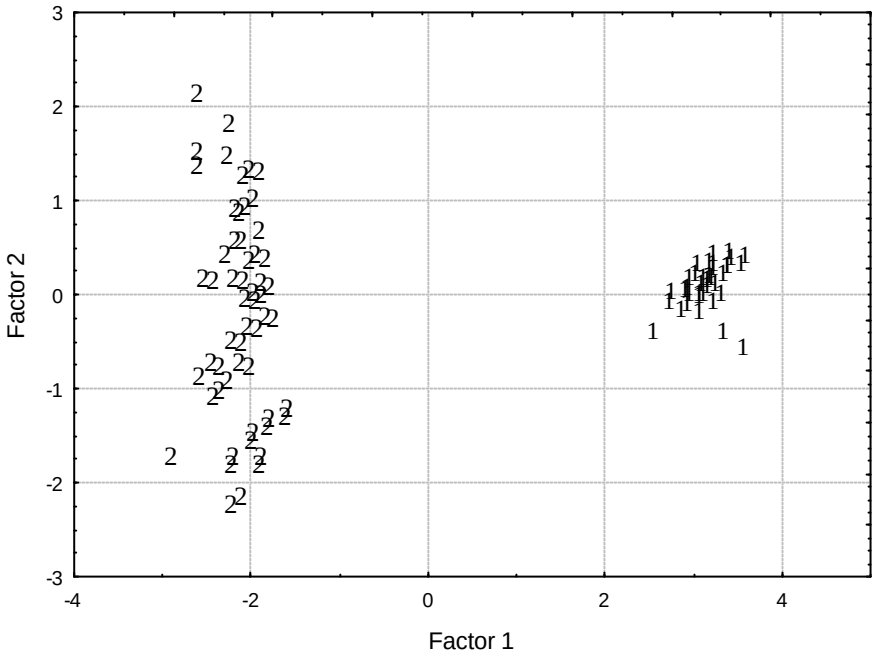


Figura 16. Dispersión de los dos primeros PC de 163 macrociclos 1: no Almizcle, 2: Almizcle [47].

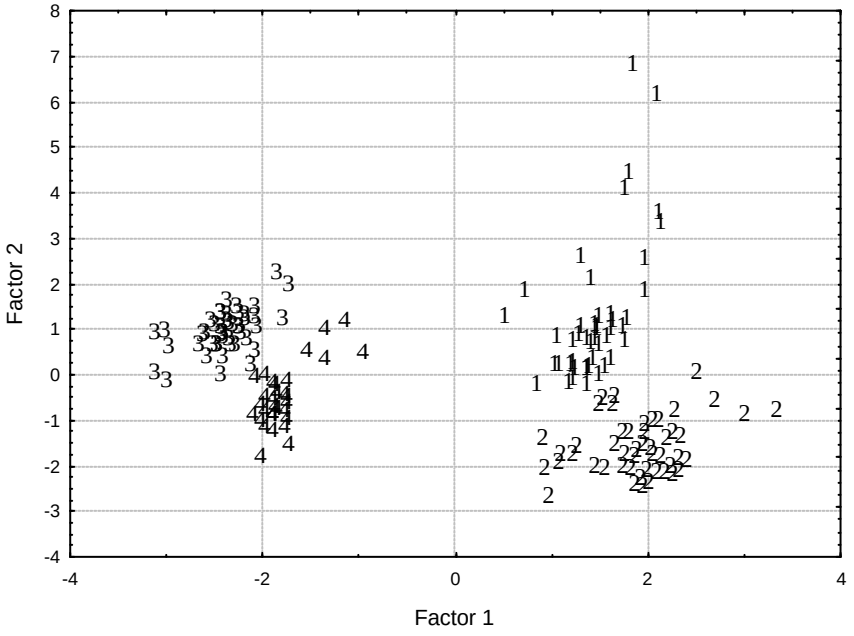


Figura 17. Dispersión de los dos primeros PC de 331 compuestos [47].

El método de análisis por medio de componentes principales es una técnica con un amplio rango de aplicaciones en muchas áreas de investigación, debido a su versatilidad y excelentes resultados. Por otro lado, permite encontrar relaciones entre puntos (compuestos, muestras etc.), por lo que ha sido utilizada en la clasificación de un gran número de compuestos químicos, especialmente, en la industria alimenticia, donde a partir de estos resultados se ha diseñado la síntesis nuevos compuestos con aplicaciones industriales.

2.5.2. Métodos de agrupamiento. El agrupamiento es una técnica de análisis de datos donde se busca identificar subgrupos homogéneos, los cuales pueden presentarse debido a un modelo patrón o a medidas de similitud. Entre los varios usos de los métodos de agrupamiento, una primera motivación que ha incrementado el uso de estas técnicas, es su utilización en la selección y diseño de estructuras químicas pertinentes en descubrimientos farmacéuticos.

Los métodos de agrupamiento son técnicas exploratorias, y se usan frecuentemente en análisis preliminares de datos que presentan una mediana o alta dimensionalidad [41,42]. Por otro lado, una característica de los métodos de agrupamiento es que son procesos no supervisados, ésto quiere decir que no se busca reproducir grupos predefinidos a partir de los datos iniciales.

Debido al interés creciente en el uso de los métodos de agrupamiento en el reconocimiento de modelos patrones, procesamientos de imágenes y recuperación de información, las técnicas de agrupamiento poseen una rica historia en otras disciplinas tales como la arqueología, astronomía, biología, ciencia de la computación, electrónica, ingenierías, ciencia de la información, medicina, psiquiatría, psicología, geología, geografía, y mercadeo. Además, se puede observar la importancia y la naturaleza interdisciplinaria de los métodos de agrupamiento, la cual se hace evidente a través del gran número de artículos reportados.

El proceso general de agrupamiento involucra las siguientes etapas:

1. Generación de los descriptores apropiados para cada compuesto del conjunto de datos;
2. Selección de una medida apropiada de similitud;
3. Uso de un método apropiado para realizar el agrupamiento;
4. Análisis de los resultados.

Para la primera etapa, los descriptores pueden incluir valores de propiedades biológicas, índices topológicos y fragmentos estructurales, entre otros. El uso de diferentes descriptores, así como su forma de representación, han sido ampliamente estudiados por varios autores [48,49]. En la segunda etapa, las diferentes medidas utilizadas para describir la similitud entre compuestos han sido discutidas y resumidas por Downs y Willett [50]. Sin embargo, existe reportada en la literatura especializada una gran variedad de medidas de distancias [51], entre las cuales, una de las medidas ampliamente utilizada es la distancia Euclideana, la cual puede usarse a menudo para reflejar la desigualdad entre dos moléculas.

Para el análisis de resultados no se dispone de un modelo general o específico acerca de la visualización y análisis de los mismos. Sin embargo, existen varias publicaciones que se enfocan en la implementación de sistemas que utilizan las técnicas de agrupamiento, y proveen detalles acerca del análisis de resultados.

2.5.2.1 Técnicas de agrupamiento. En la **Figura 18** se aprecian los diferentes métodos de agrupamiento existentes. De acuerdo con la taxonomía presentada en la gráfica, podemos observar una primera clasificación general de las técnicas de agrupamiento en dos clases, los métodos jerárquicos y los no jerárquicos.

Los métodos jerárquicos producen un *cluster* después de cada paso, presentando una relación padre-hijo establecida entre grupos a cada nivel de interacción. En contraste, si

el conjunto de datos se analiza para producir una única partición de los compuestos, el resultado es no jerárquico.

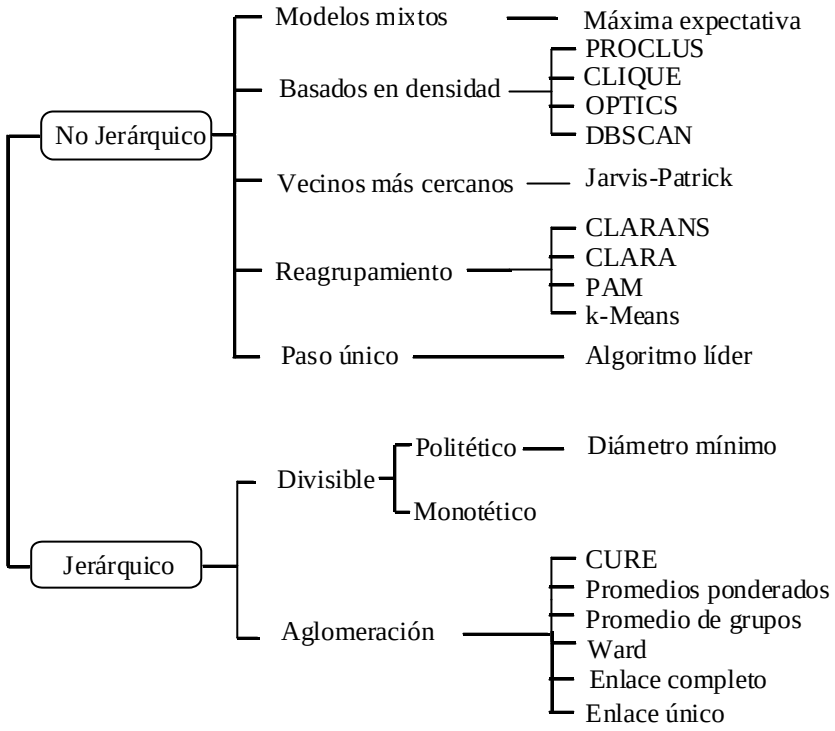


Figura 18. Clasificación de los métodos de agrupamiento [41].

Si un método jerárquico empieza con todos los compuestos representados como grupos individuales (cada compuesto representa un grupo) y consecutivamente fusiona los diferentes grupos hasta poder representar todos los compuestos en un solo grupo; el método se denomina *aglomeración*. Por el contrario, si el método jerárquico inicia con todos lo compuestos representados en un solo grupo y posteriormente divide cada grupo en dos hasta que todos los compuestos se representen individualmente, la técnica de agrupamiento es divisible. Por otro lado, si en cada una de las divisiones, un solo descriptor es utilizado para determinar cómo se divide el grupo, el método es monotético; de otra forma, si se utilizan más de dos descriptores el método es politético. Los métodos no jerárquicos comprenden un amplio rango de diferentes técnicas para la construcción *clusters*. A continuación se hace referencia a las diferentes metodologías empleadas por algunos de estos métodos. Por ejemplo, en el método de paso único, la

partición se realiza por medio de un análisis aleatorio único de los datos, en el cual cada compuesto es analizado una sola vez. Posteriormente, de acuerdo con este único análisis se determina en qué grupo debe ser clasificado cada uno de los compuestos. El método de reagrupamiento consiste en el continuo movimiento de los compuestos de un grupo a otro, donde se busca incrementar la estimación inicial de cada grupo. Mientras que el análisis de vecinos más cercanos es una técnica que se basa principalmente en cada uno de los compuestos a diferencia de las otras técnicas no jerárquicas. En este método cada uno de los alrededores de los compuestos es examinado en términos de la proximidad entre compuestos, y de acuerdo con las distancias presentadas, se establecen los diferentes grupos. Entre los diferentes métodos de vecinos más cercanos, el método de Jarvis-Patrick es el único utilizado en aplicaciones químicas.

Otras técnicas no jerárquicas incluyen los métodos basados en la densidad y los métodos probabilísticos. Los métodos basados en la densidad consideran la distribución de los descriptores a través de los datos, generando patrones de alta y baja densidad, los cuales, una vez identificados, pueden ser utilizados para separar los compuestos en diferentes grupos. Los agrupamientos basados en probabilidades generan grupos no solapados donde le es asignado a cada uno de los compuestos una probabilidad en el rango de 0 a 1. Este valor de probabilidad significa la pertenencia o no del compuesto a cada uno de los diferentes grupos.

Una vez resumidas de una manera general las diferentes técnicas presentes en la metodología de agrupamiento, se presentan los métodos “clásicos”, los cuales incluyen a manera general los métodos jerárquicos. La clasificación descrita en la **Figura 18** es comúnmente utilizada por la comunidad científica. Sin embargo, existen varias posibles clasificaciones, entre las cuales se pueden considerar la paramétrica y la no paramétrica. Los métodos paramétricos requieren que se realicen comparaciones basadas en las distancias, mientras que los métodos no paramétricos realizan varias suposiciones acerca de los datos base. En otras palabras, no adaptan los parámetros dados y, en general, solamente necesitan una matriz de proximidad (matriz de distancias).

El termino proximidad es utilizado para describir similitud o desigualdad en adición a las medidas de distancia.

Debido a la importancia de la similitud en la definición de un grupo en la mayoría procedimientos de agrupamiento, es esencial una medida de la semejanza entre dos modelos dibujados en el mismo espacio. Por esta razón, y gracias a la variedad de características y escalas, la medida de distancia (o medidas) debe escogerse cuidadosamente. Es más común calcular la desigualdad entre dos modelos usando una distancia definida de acuerdo con las características del espacio.

La distancia métrica más popular es la Euclideana, definida por la ecuación:

$$d_{AB} = \left[\sum_{j} (x_{1j} - x_{2j})^2 \right]^{1/2} \quad (2)$$

Donde X_{1j} es el valor de la j-esima variable medida en el objeto i-esimo. La distancia de Euclides es comúnmente utilizada para evaluar la proximidad de objetos representados en espacios de dos o tres dimensiones. Además, funciona muy bien en conjuntos de datos que poseen grupos “compactos” o “aislados”.

2.5.3. Métodos jerárquicos de aglomeración. Los métodos jerárquicos comúnmente utilizados, son aquellos que pertenecen a la familia de las jerarquías de no solapamiento por aglomeración secuencial (SAHN). El proceso completo involucrado en los métodos de aglomeración puede ser resumido por un algoritmo de cuatro etapas, así.

1. Calcular la matriz inicial de distancias entre todos los objetos (compuestos);
2. Buscar de los menores valores en la matriz de distancias, y posteriormente agrupar estos objetos en un mismo grupo;

3. Calcular una nueva matriz de distancias teniendo en cuenta que los nuevos grupos obtenidos en la segunda etapa han formando un nuevo conjunto de objetos, los cuales han reemplazando los datos originales;
4. Repetir las etapas 2 y 3 hasta la obtención de un solo grupo.

Las diferencias entre los métodos SANH radican en la forma cómo se define la proximidad entre grupos en la etapa 1 y en cómo se representan los pares de compuestos similares en un mismo grupo en la etapa 3. Las diferentes medidas de distancias entre objetos generalmente utilizan la fórmula de Lance-Williams [52], la cual se encuentra definida en términos de la fórmula general:

$$d[k,(i,j)] = a_i d[k,i] + a_j d[k,j] + b d[i,j] + g\left|d[k,i] - d[k,j]\right| \qquad (3)$$

Donde $d[k,(i,j)]$ es la proximidad entre el *cluster* k y el *cluster* (i,j), éste último formado por la unión de los grupos i y j. Los diferentes valores de los coeficientes a_i, a_j, b_i, y g_i definen cada uno de los métodos SAHN, algunos de los cuales son representados en la **Tabla 1**.

Tabla 1. Valores de los parámetros para algunos métodos SAHN [41].

Método	a_i	a_j	b_i	g_i
Enlace único	0.5	0.5	0	-0.5
Enlace completo	0.5	0.5	0	0.5
Promedio de grupos	$\frac{N_i}{N_i + N_j}$	$\frac{N_j}{N_i + N_j}$	$\frac{-N_i \times N_j}{(N_i + N_j)^2}$	0
Ward	$\frac{N_i + N_k}{N_i + N_j + N_k}$	$\frac{N_j + N_k}{N_i + N_j + N_k}$	$\frac{-N_k}{N_i + N_j + N_k}$	0

[a]Los parámetros N_i, N_j y N_k = Número de compuestos en los grupos i, j y k, respectivamente.

En el agrupamiento por medio del método de enlace único, la proximidad entre dos grupos es la distancia entre los puntos más cercanos de cada grupo (Uno para cada grupo). En contraste, en el agrupamiento por enlace completo, la proximidad se representa por la máxima distancia entre los pares de objetos más lejanos en cada grupo. Estos dos métodos de agrupamiento representan los extremos de los métodos SAHN. En la mitad, se encuentran los métodos de promedios de grupos y el de Ward. En el primer método la medida de proximidad entre grupos es el promedio aritmético de las distancias entre todos los pares de objetos, mientras que en el método de Ward, la proximidad se representa por medio de la varianza entre los grupos, donde la varianza se encuentra definida por la suma de los errores cuadrados de los diferentes grupos (4):

$$e^2 = \left[\sum_{j=1}^{N} (x(r) - x(c))^2 \right]^2 \qquad (4)$$

Donde: N = Número de compuestos; $x(r)$ es el vector que representa a cada uno de los compuestos y $x(c)$ es el vector, que representa al grupo central (centroide). Además, se puede observar que el error cuadrado, e^2, para cada grupo es la suma del cuadrado de las distancias Euclideanas al centroide para todos los N en cada grupo.

La operación de un algoritmo jerárquico se ilustra usando datos bidimensionales de la **Figura 19**. En esta figura, aparecen siete objetos, *e.g.* A, B, C, D, E, F, y G, los cuales forman tres grupos.

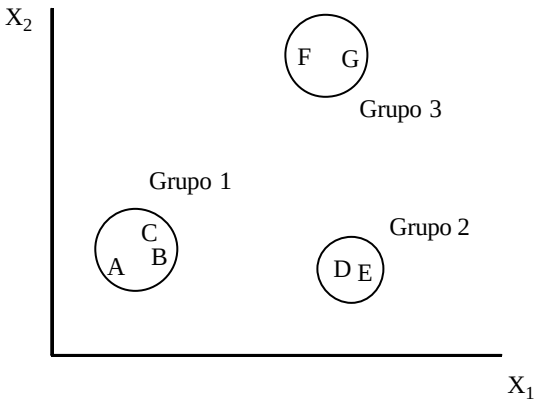


Figura 19. Clasificación de siete objetos en tres grupos diferentes.

Un algoritmo jerárquico produce un dendrograma, que representa los grupos a diferentes niveles de similitud. En la **Figura 20 (a)** se puede observar el dendograma obtenido para los siete objetos anteriores por medio del método de enlace único. El dendrograma puede romperse en cualquier nivel para producir diferentes agrupamientos de los datos. Sin embargo, dependiendo del los conocimientos previos de los objetos (compuestos) estudiados, la selección del nivel de partición puede realizarse en forma segmentada y no lineal como se menciono anteriormente. Por ejemplo, en la **Figura 20 (b)** se observan tres grupos, los cuales no pueden ser obtenidos a partir de una línea recta a través del dendograma. El uso de una línea recta conduciría a la obtención de dos grupos conformados por los elementos [8,3,1,2] y [4,5,6,7] o, a la obtención de cuatro grupos conformados por [8], [3,1,2], [4,5] y [6,7].

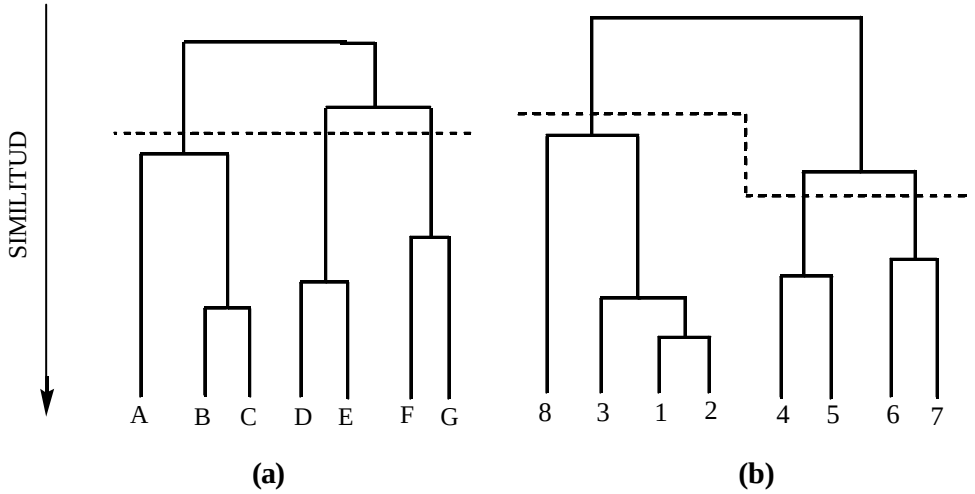


Figura 20. (a) Dendograma obtenido por medio del método de enlace único y (b) ilustración de cómo se puede realizar una partición segmentada.

En el agrupamiento jerárquico, cada nivel define una partición del conjunto de datos en diferentes grupos. Sin embargo, no existe una información asociada que indique cuál nivel es el mejor en términos de la división de los datos en un número “natural” de grupos presentes y cuál de estos contiene los compuestos más apropiados. Varios métodos y criterios han sido propuestos para tratar de esclarecer esta situación. Milligan y Cooper publicaron la primera comparación de la selección de los diferentes niveles

jerárquicos usando datos psicológicos [53]. En esta publicación, fueron estudiados treinta métodos diferentes de selección para un pequeño conjunto de datos. Quince años después, Wild y Blankey [54] publicaron un artículo donde se presentó la comparación de los niveles de selección en los métodos jerárquicos utilizando un conjunto de datos químicos, en el cual, ocho de los métodos presentados ya habían sido estudiados por Milligan y Cooper; sin embargo, el noveno fue un método recientemente publicado por Kelley, Garner y Sutcliffe [55]. El estudio realizado por Wild y Blankey concluyó que el criterio de relación de varianza y el método de Kelley presentaron los mejores resultados, siendo este último el que presentó una mejor eficiencia computacional.

Los métodos de agrupamiento han sido utilizados ampliamente durante los últimos 25 años. En la literatura se pueden encontrar varios ejemplos del desarrollo de estas metodologías, especialmente, en el área de la química donde se puede concluir que los resultados obtenidos son promisorios. Sin embargo, se debe tener en cuenta que este proceso es subjetivo, ya que el mismo conjunto de datos a menudo necesita ser dividido en diferentes niveles de acuerdo con las diferentes aplicaciones buscadas. Esta subjetividad hace que el proceso de agrupamiento sea difícil y la razón principal se debe a la no existencia de una sola metodología general que resuelva cada uno de los problemas de agrupamiento. Por otro lado, una posible solución radica en realizar las diferentes particiones con base en el conocimiento químico previo sobre los compuestos a analizar. Este conocimiento es usado implícitamente o explícitamente en una o más fases de la metodología establecida.

En resumen, la metodología de agrupamiento es interesante, viable y desafiante. Además posee un gran potencial en aplicaciones como el reconocimiento de modelos patrón, segmentación de imágenes y clasificación. Sin embargo, es posible aprovechar este potencial sólo después de realizar cuidadosamente un gran número de decisiones de selección de acuerdo con los objetos bajo estudio.

3. OBJETIVOS

3.1 OBJETIVO GENERAL

Establecer las relaciones estructura-olor entre los compuestos odorantes pertenecientes a las familias olfativas Floral, Frutal, Verde y Almizcle, por medio de los análisis de componentes principales y de agrupamientos, e involucrando parámetros vibracionales dentro de los descriptores moleculares.

3.2 OBJETIVOS ESPECÍFICOS

3.2.1 Calcular los descriptores moleculares electrónicos, así como las frecuencias de vibración de 60 compuestos odorantes, utilizando el método de campo auto consistente de Hartree–Fock restringido RHF, a un nivel de teoría HF 6–31 G(d), bajo las condiciones estándar (STP) 298,15 K y 1.0 atm, por medio del paquete computacional *Gaussian03*. Con el objetivo de estudiar por medio de los métodos no supervisados si los modos vibracionales y los descriptores electrónicos realizan la clasificación de los diferentes compuestos odorantes

3.2.2 Calcular los descriptores estructurales y topológicos de 60 moléculas odorantes, por medio del programa computacional *Moldes.* Para poder aplicar los diferentes métodos multivariables y establecer cuales descriptores o combinación de los mismos, representan de manera adecuada los diferentes grupos odorantes bajo estudio.

3.2.3 Realizar un estudio comparativo de los diferentes subespacios, para establecer cuáles permiten realizar la mejor clasificación de sustancias odoríferas. Así

mismo, aplicar el análisis de componentes principales a los descriptores moleculares y representar cada uno de los compuesto odorantes en uno de los subespacios construidos con los componentes principales.

3.2.4 Realizar por medio del análisis de agrupamiento un estudio detallado de los diferentes tipos de descriptores, para establecer qué conjunto de variables conduce a los mejores resultados de clasificación de los compuestos odorantes.

3.2.5 Establecer las relaciones estructura olor presentes en los compuestos fragantes, por medio de los diferentes análisis de estadística multivariable establecidos.

3.2.6 Estudiar la participación de los parámetros vibracionales en las relaciones “estructura-olor”, por medio de la estadística multivariable con base en pruebas de similitud entre moléculas, donde se espera realizar la verificación de la teoría de Luca Turin.

4. METODOLOGÍA Y RESULTADOS

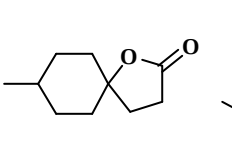
El desarrollo computacional que se ha llevado a cabo durante los últimos años permitió, que la adquisición de datos teóricos fuera más fácil, rápida y confiable. Es así como el cálculo de los espectros vibracionales y los diferentes descriptores moleculares pueden obtenerse usando una gran variedad de paquetes de software semiempíricos o *ab initio*.

Para lograr los objetivos de esta investigación es necesario observar que la ejecución del presente proyecto giró en torno de dos pasos metodológicos, la adquisición de los datos y el tratamiento de los mismos.

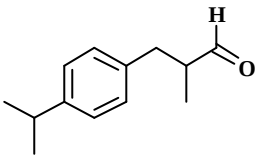
4.1 COMPUESTOS ODORANTES SELECCIONADOS

En la **Figura 21** se presentan las 60 moléculas odorantes escogidas como objetos de trabajo, las cuales comprenden 15 compuestos de cada una de las familias Floral, Frutal, Verde y Almizcle. Cabe mencionar, que todos los compuestos utilizados en este estudio fueron tomados de los reportes presentes en la literatura de las estructuras químicas y la clasificación del olor [2, 8, 30].

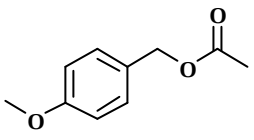
GRUPO 1
OLOR: FLORAL



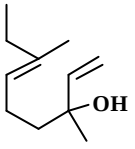
Metil laitona
1F



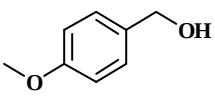
Ciclamen aldehído
2F



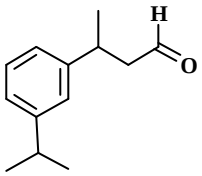
Acetato de 4-metoxibencilo
3F



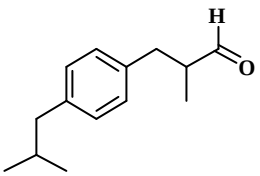
Etil linalool
4F



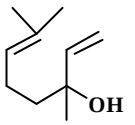
(4-Metoxifenil)metanol
5F



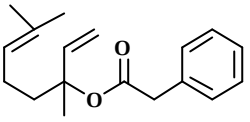
Florhidral
6F



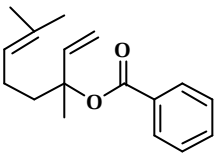
Silvial
7F



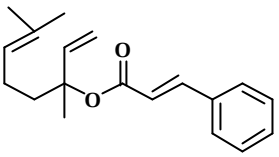
Linalool sintético
8F



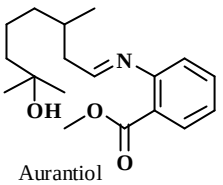
Cinamato de linalil-fenilo
9F



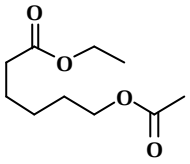
Benzoato de linalilo
10F



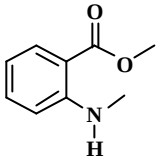
Cinamato de linalilo
11F



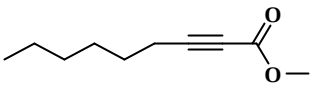
Aurantiol
12F



Berriflor
13F

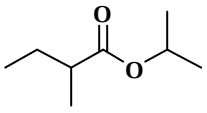


Dimetil antranilato
14F

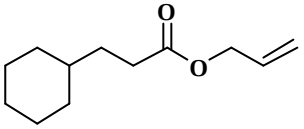


2-noninato de metilo
15F

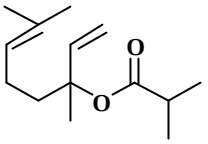
GRUPO 2
OLOR: FRUTAL



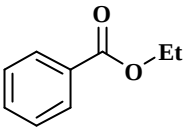
2-Metil-Butirato de isopropilo
1Fr



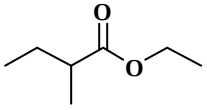
3-Ciclohexilpropanoato de alilo
2Fr



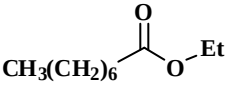
Isobutirato de linalilo
3Fr



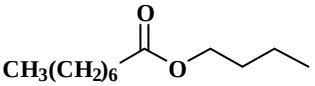
Benzoato de etilo
4Fr



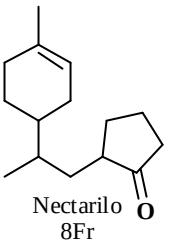
2-Metilbutanoato de etilo
5Fr



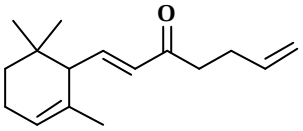
Octanoato de etilo
6Fr



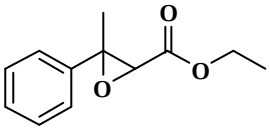
Octanoato de Butilo
7Fr



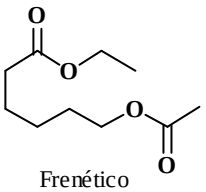
Nectarilo
8Fr



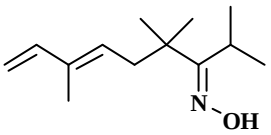
Cetona V
9Fr



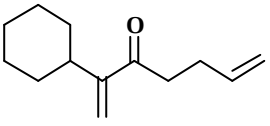
Fresa
10Fr



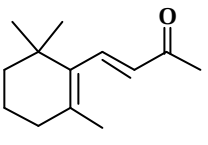
Frenético
11Fr



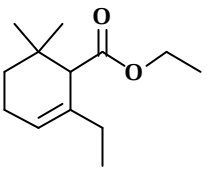
Labienoxima
12Fr



2-Ciclohexilhepta-1,6-dien-3-ona
13Fr

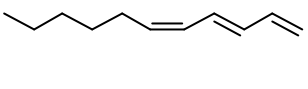


β-Ionona
14Fr

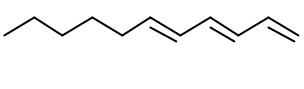


Givescona
15Fr

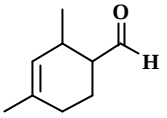
GRUPO 3
OLOR: VERDE



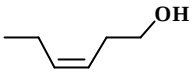
cis-Undecatrieno
1V



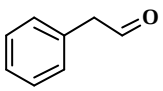
Undecatrieno
2V



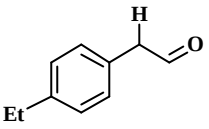
Ciclal C
3V



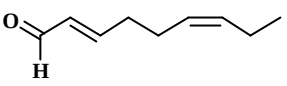
cis-Hex-3-en-1-ol
4V



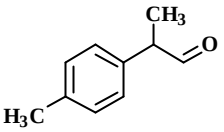
2-Fenilacetaldehído
5V



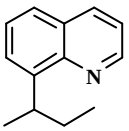
2-(4-Etilfenil)acetaldehído
6V



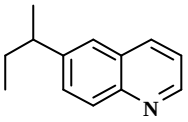
Nonadienal
7V



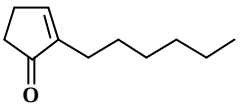
2-*p*-Tolilpropanal
8V



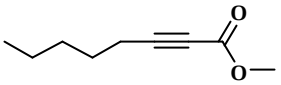
8-Butilquinolina
9V



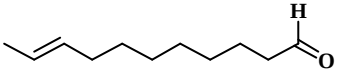
6-Butilquinolina
10V



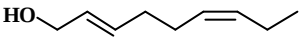
iso-Jasmona B
11V



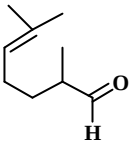
2-Octinoato de metilo
12V



iso-Aldehído C
13V

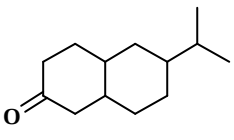


Nonadienol
14V

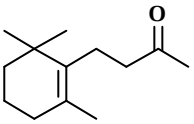


Melonal
15V

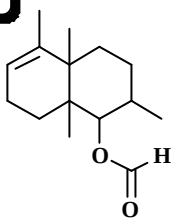
GRUPO 4
OLOR: ALMIZCLE



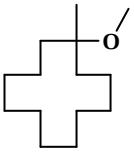
Decatona
1A



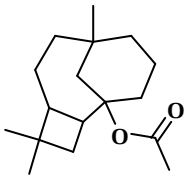
β-Dihidroionona
2A



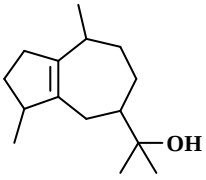
Formato de Oxioctalina
3A



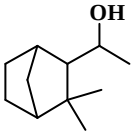
Madrox^{R}
4A



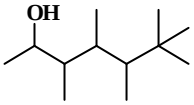
Vetinal
5A



Guaiol
6A



Camecol
7A



Koinool
8A

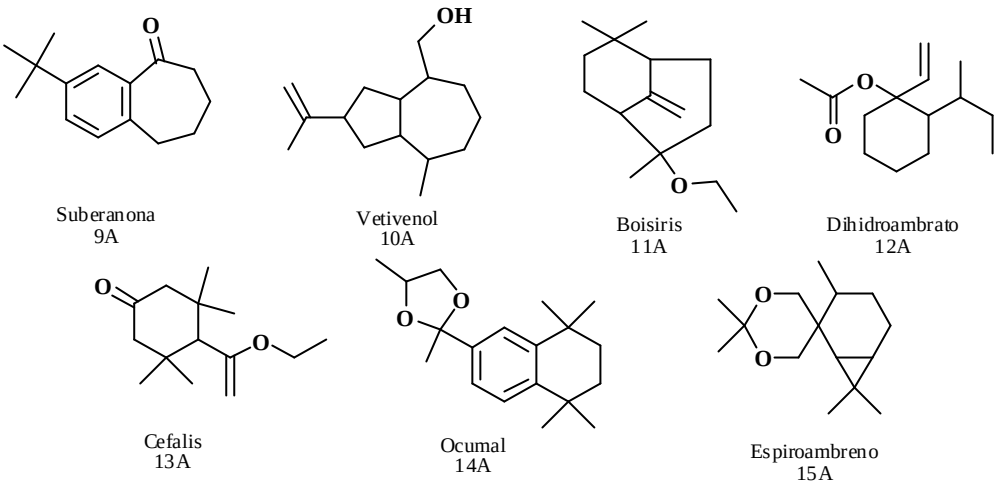


Figura 21. Compuestos odorantes seleccionados.

4.2 OBTENCIÓN DE DATOS

La obtención de los descriptores moleculares se realizó por medio de varias etapas utilizando diferentes paquetes computacionales dependiendo del tipo de descriptor deseado. No obstante, las 60 moléculas odorantes fueron sometidas a algunos tratamientos previos antes de realizar los cálculos computacionales “robustos”.

4.2.1 Obtención de los descriptores moleculares electrónicos y los modos vibracionales. Antes iniciar los cálculos computacionales a un nivel de teoría *ab initio* de las 60 moléculas odorantes, se realizó un tratamiento previo de los compuestos, con el fin de disminuir el tiempo computacional.

4.2.1.1 Análisis conformacional. Con el objetivo de obtener los confórmeros más estables, se realizó como primer paso el análisis conformacional de las 60 moléculas odorantes. Los diferentes compuestos fragantes fueron dibujados en dos dimensiones

inicialmente con el *software* Spartan Pro 1.0.5., para posteriormente realizar el análisis conformacional a cada una de estas moléculas individualmente. Este análisis fue desarrollado con el método de mecánica molecular, a un nivel de teoría MMFF, con el paquete computacional mencionado anteriormente.

Una vez terminados los análisis conformacionales, se procedió a identificar los confórmeros de menor energía, los cuales fueron seleccionados como estructuras moleculares de partida para los posteriores cálculos. Las energías de los confórmeros más estables se resumen en la **Tabla 2**.

Tabla 2. Energía de los confórmeros más estables.

Compuesto	Energía confórmero, kcal/mol	Compuesto	Energía confórmero, kcal/mol	Compuesto	Energía confórmero, kcal/mol	Compuesto	Energía confórmero, kcal/mol
1A	19,046	**1F**	-116,926	**1V**	7,964	**1Fr**	-5,897
2A	24,802	**2F**	26,092	**2V**	5,818	**2Fr**	-3,252
3A	39,100	**3F**	-106,359	**3V**	9,863	**3Fr**	18,002
4A	34,008	**4F**	15,394	**4V**	-62,382	**4Fr**	26,717
5A	120,462	**5F**	-69,160	**5V**	-15,421	**5Fr**	-8,653
6A	45,074	**6F**	24,751	**6V**	-28,894	**6Fr**	-142,607
7A	51,556	**7F**	22,116	**7V**	19,405	**7Fr**	-4,468
8A	45,311	**8F**	17,738	**8V**	34,709	**8Fr**	26,315
9A	57,528	**9F**	57,074	**9V**	42,455	**9Fr**	40,084
10A	34,175	**10F**	55,252	**10V**	38,109	**10Fr**	17,187
11A	67,112	**11F**	75,441	**11V**	1,213	**11Fr**	-19,661
12A	18,843	**12F**	139,554	**12V**	-4,472	**12Fr**	52,754
13A	43,761	**13F**	-19,659	**13V**	8,041	**13Fr**	25,985
14A	65,521	**14F**	45,753	**14V**	8,406	**14Fr**	29,490
15A	78,461	**15F**	-4,713	**15V**	10,901	**15Fr**	34,262

4.2.1.2 Optimización de la geometría. Existen muchos procedimientos matemáticos sistemáticos (algoritmos) para encontrar un mínimo local de una función de varias variables. Estos procedimientos encontrarán un mínimo local en la energía en las

proximidades de la geometría inicialmente supuesta. El proceso de obtención de tal mínimo se denomina *optimización de la geometría* o *minimización de la energía*.

Para facilitar los cálculos y hacer más rápida su obtención, antes de proceder con la optimización de la geometría por medio del *software Gaussian03* utilizando el método de campo auto consistente de Hartree–Fock restringido RHF, con un conjunto de base HF 6-31 G(d), se realizó inicialmente una preoptimización utilizando *Gaussian03* con el método semiempírico a un nivel de teoría AM1. Los resultados obtenidos por medio de los cálculos *ab initio* se resumen en la **Tabla 3** (no se incluyen las frecuencias).

4.2.2 Obtención de los descriptores moleculares estructurales y topológicos. El cálculo de los descriptores moleculares estructurales y topológicos se realizó a partir de las geometrías optimizadas de las diferentes moléculas fragantes, obtenidas por medio de los cálculos *ab initio*. Para este propósito se utilizó el programa computacional *Moldes* (Mati Karelson, Universidad de Tartu). Los resultados obtenidos se presentan en el **Anexo 1**[1].

[1] Los diferentes anexos se encuentran en el disco titulado “anexos de tesis de maestría”, proporcionado junto con el documento escrito.

Tabla 3. Descriptores obtenidos por medio de los cálculos *ab initio.*

	E(HF), Hartrees	ZPE, Hartrees	E(HF)+ZPE, Hartrees	Momento bipolar, Debye	E. HOMO, Hartrees	E. LUMO, Hartrees		E(HF), Hartrees	ZPE, Hartrees	E(HF)+ZPE, Hartrees	Momento dipolar, Debye	E. HOMO, Hartrees	E. LUMO, Hartrees
1A	-579,987	0,354	-579,633	3,599	-0,385	0,164	**1V**	-426,984	0,280	-426,703	0,543	-0,285	0,109
2A	-579,956	0,348	-579,607	2,955	-0,322	0,170	**2V**	-426,987	0,280	-426,707	0,657	-0,284	0,110
3A	-731,698	0,395	-731,304	4,659	-0,337	0,185	**3V**	-423,815	0,227	-423,587	2,965	-0,341	0,163
4A	-621,304	0,433	-620,870	1,154	-0,398	0,211	**4V**	-309,024	0,184	-308,840	1,974	-0,347	0,184
5A	-809,700	0,457	-809,242	2,109	-0,390	0,196	**5V**	-382,462	0,148	-382,314	3,100	-0,335	0,127
6A	-658,011	0,413	-657,599	1,351	-0,316	0,184	**6V**	-460,534	0,208	-460,325	3,331	-0,324	0,128
7A	-503,035	0,316	-502,719	1,658	-0,399	0,206	**7V**	-423,786	0,225	-423,561	4,304	-0,350	0,107
8A	-544,383	0,391	-543,992	1,588	-0,402	0,200	**8V**	-460,532	0,208	-460,324	3,372	-0,322	0,129
9A	-654,550	0,340	-654,210	3,214	-0,319	0,102	**9V**	-555,486	0,267	-555,219	1,893	-0,300	0,091
10A	-657,995	0,415	-657,580	1,183	-0,341	0,196	**10V**	-555,488	0,267	-555,222	2,296	-0,303	0,092
11A	-658,002	0,414	-657,588	0,692	-0,337	0,187	**11V**	-501,900	0,288	-501,612	3,501	-0,360	0,122
12A	-693,854	0,385	-693,469	2,018	-0,358	0,180	**12V**	-498,646	0,230	-498,415	2,300	-0,391	0,134
13A	-693,827	0,385	-693,443	4,256	-0,339	0,172	**13V**	-503,041	0,310	-502,731	3,291	-0,343	0,162
14A	-885,539	0,467	-885,073	1,521	-0,310	0,146	**14V**	-424,939	0,250	-424,689	2,220	-0,342	0,174
15A	-732,802	0,419	-732,383	1,873	-0,348	0,212	**15V**	-424,971	0,248	-424,723	3,010	-0,336	0,160
1F	-537,765	0,267	-537,498	4,825	4,825	-0,420	**1Fr**	-462,018	0,249	-461,770	2,139	-0,421	0,199
2F	-577,635	0,299	-577,336	3,318	3,318	-0,318	**2Fr**	-615,799	0,326	-615,472	2,230	-0,376	0,172
3F	-610,266	0,220	-610,046	0,810	0,810	-0,308	**3Fr**	-693,839	0,380	-693,459	1,920	-0,328	0,179
4F	-503,011	0,309	-503,337	1,543	1,543	-0,326	**4Fr**	-496,391	0,185	-496,206	2,187	-0,340	0,102
5F	-458,467	0,179	-458,288	1,645	1,645	-0,304	**5Fr**	-422,980	0,219	-422,761	2,215	-0,423	0,198
6F	-577,634	0,299	-577,335	3,171	3,171	-0,323	**6Fr**	-540,085	0,311	-539,775	1,701	-0,429	0,200
7F	-616,672	0,330	-616,342	2,576	2,576	-0,309	**7Fr**	-618,151	0,372	-617,779	1,906	-0,431	0,198
8F	-463,979	0,278	-463,700	1,589	1,589	-0,326	**8Fr**	-656,863	0,389	-656,475	3,067	-0,328	0,156
9F	-845,296	0,407	-844,890	4,026	4,026	-0,323	**9Fr**	-694,687	0,390	-694,297	3,936	-0,342	0,109
10F	-806,285	0,377	-805,909	1,949	1,949	-0,329	**10Fr**	-687,142	0,256	-686,886	1,597	-0,331	0,134
11F	-883,154	0,412	-882,741	4,415	4,415	-0,315	**11Fr**	-688,659	0,298	-688,362	2,203	-0,433	0,194
12F	-976,390	0,455	-975,935	5,062	5,062	-0,316	**12Fr**	-634,879	0,363	-634,516	0,726	-0,316	0,150
13F	-688,649	0,296	-688,353	6,676	6,676	-0,430	**13Fr**	-578,768	0,326	-578,442	2,688	-0,358	0,111
14F	-551,404	0,203	-551,201	3,427	3,427	-0,283	**14Fr**	-578,776	0,323	-578,453	3,333	-0,327	0,107
15F	-537,680	0,261	-537,419	2,759	2,759	-0,391	**15Fr**	-654,836	0,355	-654,481	1,677	-0,335	0,180

4.3 TRATAMIENTO DE LOS DATOS OBTENIDOS

Uno de los aspectos más importantes en el análisis estadístico es la consideración previa de los datos a analizar. Por esta razón, antes de iniciar con el análisis estadístico multivariable, se realizó un tratamiento previo de los datos obtenidos por medio de una observación minuciosa de los mismos, la cual constó de tres etapas, a saber: el análisis de la distribución de datos, el escalamiento y la reducción de datos.

4.3.1 Distribución de datos. La estadística se encuentra involucrada con el tratamiento de un número finito de muestras, las cuales pueden ser descritas por una o más variables. Los valores de estas variables han sido calculados individualmente para cada muestra.

La mayoría de los análisis estadísticos suponen que los datos obtenidos se comportan de acuerdo con la distribución normal. La distribución normal o *Gaussiana*, es la más importante de las distribuciones consideradas en la estadística debido a su amplio rango de aplicaciones prácticas; entre ellas, cabe destacar, que la mayoría de mediciones de características físicas junto con sus errores aleatorios y variaciones naturales pueden ser aproximadas por medio de la distribución normal.

Un problema comúnmente encontrado en la distribución de datos es la presencia de puntos anómalos. Si un valor anómalo es encontrado en algunos de los descriptores (datos fisicoquímicos), puede ser debido a un error en la medida o cálculo de dicho parámetro para un compuesto particular, en el caso de variables obtenidas por medio de cálculos computacionales los puntos anómalos pueden indicar que alguna suposición básica ha sido violada o que el método utilizado no era el apropiado para ese compuesto. Sin embrago, sin importar cual tratamiento matemático sea adoptado, la inspección visual de la distribución de datos durante el análisis, es la manera más efectiva de identificar los valores anómalos.

Una vez determinados los puntos anómalos, se debe proceder a analizarlos cuidadosamente para determinar la causa de tal anomalía, y de esta manera, poder tomar la decisión de eliminarlo o no, sin embargo, no todos los puntos anómalos pueden ser descartados arbitrariamente, ya que pueden contener una buena cantidad de información.

En esta investigación se encontraron dos variables que presentaron puntos anómalos en los histogramas: el número de triples enlaces (d11) y el número relativo de triples enlaces (d15). El análisis de estos descriptores es tratado más adelante en la sección de Reducción de Datos (4.3.3). En la **Figura 22** se pueden observar algunos histogramas obtenidos a partir de los descriptores moleculares, así como los histogramas que presentan valores anómalos.

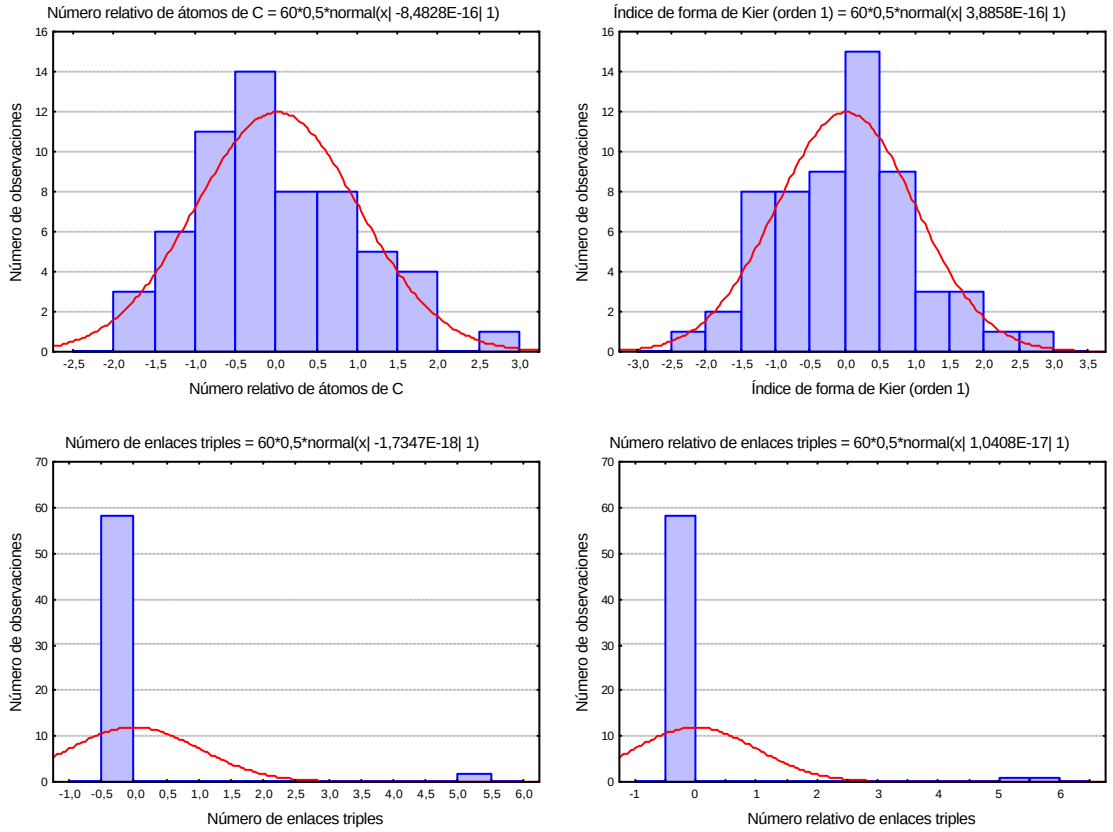


Figura 22. Histogramas de algunos descriptores moleculares.

4.3.2 Escalamiento de datos. La comparación de valores con diferentes unidades puede realizarse transformando los datos originales por medio de un proceso llamado escalamiento. El objetivo de los métodos de escalamiento es reducir cualquier descompensación debida exclusivamente a las unidades utilizadas para expresar cualquier variable.

Uno de los métodos comúnmente utilizado es la normalización o escalamiento de rango, donde el valor mínimo de la variable se establece como cero y los demás valores son divididos por la diferencia entre los valores máximo y mínimo de la variable, según lo expresa la ecuación (6):

$$X'_{ij} = \frac{X_{ij} - X_j(\min)}{X_j(\max) - X_j(\min)} \qquad (6)$$

En esta ecuación X'_{ij} es el valor normalizado para la fila *i* (compuesto *i*) de la variable *j*. La normalización puede llevarse a cabo sobre cualquier rango particular; sin embargo, este método es muy sensible a los valores anómalos.

Otro método de escalamiento menos sensible a los valores anómalos se conoce como autoescalamiento o estandarización. A partir de éste se obtienen valores que presentan un promedio con valor cero y una desviación estándar con un valor igual a uno. El autoescalamiento se lleva a cabo por medio de la sustracción del valor promedio de cada variable a cada uno de los valores individuales de la misma variable y, posteriormente, dividiendo este valor por la desviación estándar de la distribución de la variable, según la ecuación (7):

$$X'_{ij} = \frac{X_{ij} - \overline{X}_j}{s_j} \qquad (7)$$

Nuevamente, X'_{ij} representa el valor autoescalado para la fila *i* (compuesto *i*) de la variable *j*; $\overline{X}_j$ es el promedio de la variable *j*; y σ_j es su desviación estándar.

Para el escalamiento de los descriptores obtenidos anteriormente, se utilizó el método de autoescalamiento o estandarización. En el **Anexo 2** se pueden observar los valores de las variables una vez estandarizados, junto con los de los promedios y de las desviaciones estándar.

4.3.3 Reducción de datos. Ya se ha discutido la importancia del tratamiento previo de datos, así como de la distribución de frecuencias y de los motivos por los cuales es importante escalar los datos, pero quizás una de las preguntas más importantes que debe aclararse desde el inicio del tratamiento de los datos es ¿qué información contienen los datos? Se establece que un conjunto de datos contendrá tantas piezas de información como número de variables; sin embargo, en algunos casos el conjunto de datos puede presentar un cierto grado de redundancia.

Los métodos de reducción de datos se basan en la identificación de las variables que presentan redundancia y su posterior eliminación. La reducción de datos no debe ser confundida con la reducción de dimensiones, en la cual un conjunto de datos con alta dimensionalidad se procesa hasta obtener un nuevo conjunto de datos con menor dimensionalidad, usualmente con el propósito de poder graficar los datos.

El primer paso utilizado en la reducción de datos en la presente investigación, fue la búsqueda de variables que presentan valores constantes o cercanamente constantes. Esta situación puede presentarse debido a que el descriptor no fue seleccionado apropiadamente, dicho de otra forma, el descriptor no aporta información significativa.

Por medio de la observación de los descriptores obtenidos a través de los cálculos computacionales, se pudo establecer la presencia de dos descriptores que presentaron valores cercanamente constantes, *i.e.* el número de enlaces triples y el número relativo de enlaces triples. Estos descriptores son los mismos que presentaron puntos anómalos en sus respectivos histogramas. Se puede observar que el valor casi constante de estos descriptores es debido a la presencia de triples enlaces solamente en dos de las 60

moléculas bajo estudio. Para poder determinar si estos descriptores podían ser removidos, se procedió a realizar los posteriores análisis observando detalladamente su comportamiento.

El siguiente criterio de eliminación de descriptores utilizado fue la correlación entre variables. De esta manera, se calculó la matriz de correlación de los 92 descriptores, donde la correlación entre un par de descriptores particulares se puede determinar por la intersección de la fila y la columna correspondiente a cada descriptor. La diagonal de la matriz está compuesta por valores unitarios, ya que representan la correlación de cada descriptor con si mismo. Por otro lado, como la matriz de correlación es simétrica se muestra solamente la mitad inferior (**Anexo 3)**.

La inspección de la matriz de correlación permitió identificar los pares de descriptores correlacionados entre si, tomando como factores indicadores de correlación alta valores mayores de 0,75 ($r > 0,75$). Una vez identificados los pares de descriptores correlacionados, se presentaron dos problemas en la decisión de cuál de los descriptores eliminar. El primero fue la verificación de la correlación con el fin de establecer si la correlación era "real" o era causada por algún "efecto punto-grupo" debido a un valor anómalo. Esta verificación se llevó a cabo realizando gráficas de dispersión entre cada par de descriptores correlacionados, donde los efectos debidos a los puntos anómalos pudieron ser visualizados (**Figura 23**).

En la **Figura 23**, las gráficas de dispersión del número de átomos contra el número de enlaces (A) y el 3D-índice de Kier y Hall (orden 0) contra la energía HF (B) muestran claramente que los descriptores presentan una alta interdependencia. Por otro lado, en la gráfica (C) se observan dos "nubes" de puntos separadas, donde la línea de correlación ha sido ajustada entre los dos grupos; lo anterior muestra que los descriptores no se encuentran realmente correlacionados. Por último, en la gráfica (D) se presenta un caso particular, donde a pesar del alto valor del coeficiente de correlación ($r = 1,00$) no existe ningún tipo de interdependencia entre los descriptores graficados. Es importante subrayar, que estos últimos descriptores son los mismos que presentaron valores

anómalos en los histogramas (Número de enlaces triples y número relativo de enlaces triples) así como valores casi constantes.

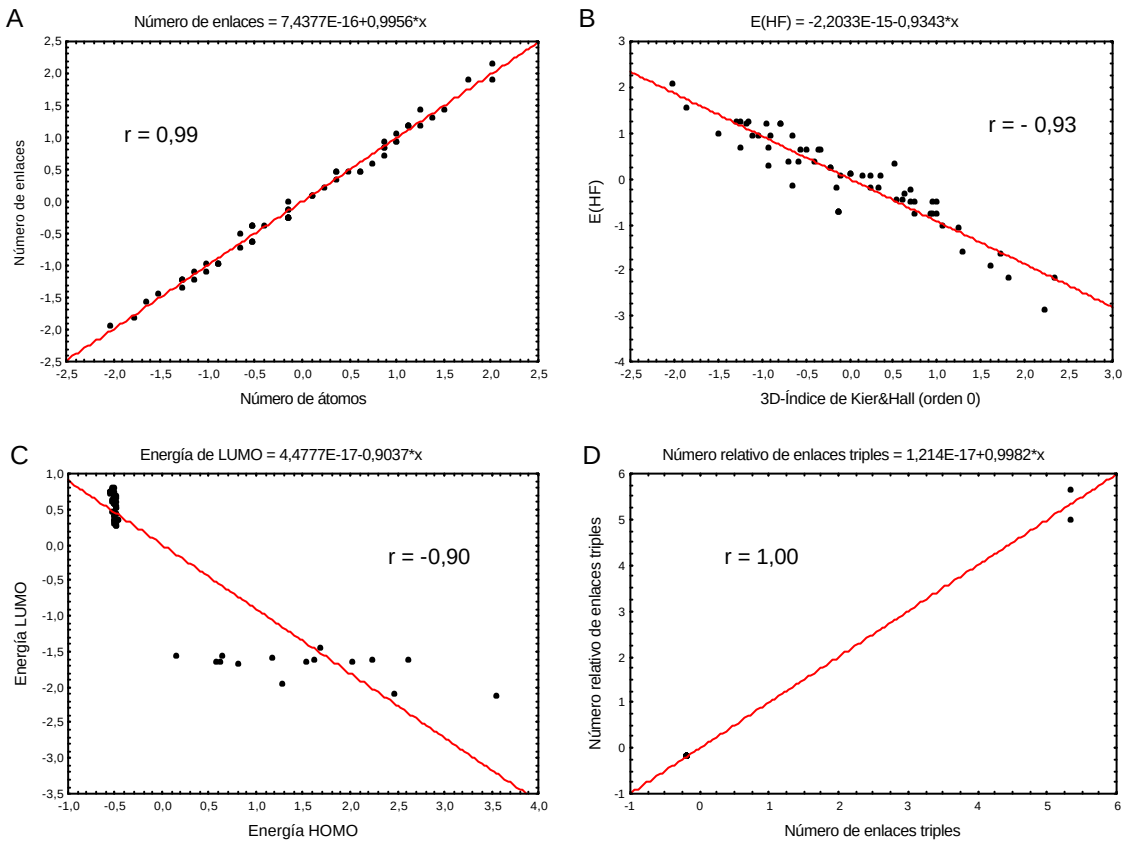


Figura 23. Gráficas de correlación entre diferentes descriptores.

El segundo problema, una vez establecidas cuáles correlaciones eran reales, fue la selección del descriptor que sería eliminado. Para solucionar esta dificultad se realizaron dos tratamientos distintos. El primer tratamiento (Método A) fue eliminar el descriptor que presentara el mayor número de correlaciones con otros descriptores. Este método condujo a una matriz en la cual el máximo número de parámetros fue retenido pero las correlaciones ínterparámetros se mantuvieron bajas. El segundo tratamiento (Método B), donde el objetivo fue reducir el tamaño el la matriz, se basó en la retención de los descriptores con el mayor número de correlaciones y posterior eliminación de los descriptores correlacionados con éstos.

Durante los tratamientos anteriormente descritos se encontraron algunos descriptores correlacionados, donde cada uno de los descriptores presentó el mismo número de correlaciones con otros descriptores. La escogencia de cuál de estos descriptores sería eliminado se realizó con base en la desviación de la distribución normal. Por lo tanto, el descriptor, cuya distribución de frecuencia se ajustara mejor a la distribución normal, fue retenido. Para este análisis se utilizó la prueba de *Shapiro-Wilk W* [56].

La prueba de *Shapiro-Wilk W* es quizás la prueba preferida para evaluar la normalidad debido a sus excelentes propiedades y resultados comparados con un amplio rango de pruebas alternativas. Entre mayor sea el valor del criterio estadístico *W*, más pronunciada será la tendencia de los valores hacia una distribución normal. En la **Tabla 4**, se observan los resultados obtenidos en el análisis de la distribución normal para algunos descriptores seleccionados por medio de los métodos A y B.

Tabla 4. Valores de *W*.

	Descriptores	*W*
1	Índice de forma de Kier (orden 2)	0,913
	Índice de forma de Kier (orden 3)	0,873
2	Contenido de información estructural (orden 1)	0,935
	Contenido de información de enlace (orden 1)	0,933
3	Momento de inercia B	0,895
	Momento de inercia C	0,881
4	PNSA-2	0,761
	FNSA-2	0,838
5	PNSA-3	0,941
	DPSA-3	0,903
	FNSA-3	0,956

Una vez eliminados los descriptores que presentaban el mismo número de correlaciones y teniendo en cuenta que los descriptores topológicos y los topográficos se encontraban altamente correlacionados, se obtuvieron dos conjuntos de datos reducidos. A partir del método A se obtuvo un conjuntos de datos de 23 descriptores (véase, **Tabla 5)**, mientras que a partir del Método B, un conjunto de datos de 20 descriptores (véase, **Tabla 6)**.

Tabla 5. Descriptores seleccionados por medio de la reducción de variables utilizando el Método A.

Mol.	d4	d6	d7	d10	d11	d12	d14	d15	d16	d26	d28	d36	d39	d66	d67	d72	d74	d81	d83	d85	d86	d90	d91	d92
1A	-0,241	1,197	-0,573	-0,584	-0,184	-0,784	-0,801	-0,184	-0,718	1,188	-0,722	-0,783	0,512	-0,143	-0,362	0,254	-0,573	0,721	-0,563	0,674	-0,031	0,843	-0,530	0,594
2A	-0,241	0,039	-0,573	0,162	-0,184	-0,784	0,019	-0,184	-0,718	0,458	-0,600	-0,428	0,309	-0,556	-0,085	0,323	-0,590	0,791	-0,617	0,754	0,314	0,318	-0,494	0,620
3A	-0,119	1,197	-0,573	0,162	-0,184	-0,438	-0,209	-0,184	-0,500	2,131	-0,669	1,086	0,076	-0,788	-0,133	0,017	0,113	0,112	0,141	-0,524	-0,286	1,708	-0,503	0,690
4A	-1,153	0,039	-0,573	-1,329	-0,184	-0,784	-1,577	-0,184	-0,718	0,819	0,852	-1,283	3,632	-0,789	-0,035	1,216	0,522	2,026	0,496	1,580	1,420	-1,151	-0,538	0,807
5A	-0,226	2,356	-0,573	-0,584	-0,184	-0,438	-0,991	-0,184	-0,531	2,582	-0,747	0,356	1,889	-0,935	-0,252	1,116	0,364	1,151	-0,117	0,481	1,082	-0,372	-0,533	0,741
6A	-0,343	1,197	-0,573	-0,584	-0,184	-0,784	-0,909	-0,184	-0,718	1,522	-0,655	0,188	1,123	-0,796	-0,147	0,922	0,830	1,254	0,681	0,832	0,744	-0,990	-0,491	0,684
7A	-0,686	1,197	-0,573	-1,329	-0,184	-0,784	-1,577	-0,184	-0,718	1,192	-1,484	-0,821	0,157	-0,368	1,842	-0,331	0,275	1,461	1,823	1,228	1,008	-0,740	-0,539	0,784
8A	-1,612	-1,120	-0,573	-1,329	-0,184	-0,784	-1,577	-0,184	-0,718	0,702	-0,520	-0,970	2,344	-0,630	0,581	0,105	-0,355	0,986	0,038	0,659	0,694	-0,797	-0,540	0,758
9A	1,185	1,197	1,718	-0,584	-0,184	1,291	-0,801	-0,184	0,764	0,432	-0,720	0,605	0,371	-0,612	-0,368	0,173	-0,740	0,055	-1,069	0,273	0,189	0,529	-0,493	0,308
10A	-0,343	1,197	-0,573	-0,584	-0,184	-0,784	-0,909	-0,184	-0,718	1,745	-0,296	0,620	0,480	-0,858	-0,143	1,045	1,755	1,169	1,518	0,693	0,871	-1,128	-0,505	0,737
11A	-0,343	1,197	-0,573	-0,584	-0,184	-0,784	-0,909	-0,184	-0,718	1,665	-0,804	0,414	1,143	-0,716	-0,095	0,932	0,197	1,389	0,028	1,020	1,583	-1,528	-0,503	0,699
12A	-0,526	0,039	-0,573	0,162	-0,184	-0,438	-0,140	-0,184	-0,489	0,868	-0,004	0,657	0,429	-0,911	0,128	0,574	0,145	0,594	-0,146	0,114	0,890	-0,446	-0,515	0,665
13A	-0,526	0,039	-0,573	0,162	-0,184	-0,784	-0,140	-0,184	-0,718	0,537	-0,582	0,508	1,371	-0,863	0,075	0,108	-0,206	0,054	-0,383	-0,414	0,050	1,379	-0,504	0,631
14A	0,443	2,356	1,718	-1,329	-0,184	1,291	-1,577	-0,184	0,357	1,915	-0,465	1,444	2,315	-0,937	-0,972	2,039	2,033	0,874	0,501	0,003	1,016	-0,852	-0,487	0,512
15A	-0,556	2,356	-0,573	-1,329	-0,184	-0,784	-1,577	-0,184	-0,718	2,190	-1,097	-0,066	1,646	-0,852	-0,006	0,917	0,778	1,324	0,674	0,670	1,026	-0,565	-0,509	0,811
1F	-0,343	1,197	-0,573	-0,584	-0,184	-0,438	-0,586	-0,184	-0,403	0,376	-1,349	-1,006	-0,202	-0,003	0,801	-0,899	-0,762	-0,056	0,127	0,029	-1,176	1,844	2,479	-2,081
2F	0,918	0,039	1,718	-0,584	-0,184	1,291	-0,679	-0,184	0,995	-0,258	-0,281	0,116	0,177	0,073	-0,669	-0,232	-0,526	-0,618	-0,807	-0,131	-0,396	0,614	1,608	-1,615
3F	0,758	0,039	1,718	-0,584	-0,184	1,636	-0,428	-0,184	1,840	-0,871	-0,438	0,035	-1,035	-0,006	-0,200	-1,267	0,957	-1,435	1,881	-1,589	-1,543	-1,432	0,160	-1,566
4F	-0,686	-1,120	-0,573	0,162	-0,184	-0,784	0,277	-0,184	-0,718	-0,475	-0,180	0,312	-0,597	-0,126	-0,268	0,245	0,167	0,514	0,203	0,499	0,632	-0,834	0,583	-1,650
5F	0,758	0,039	1,718	-1,329	-0,184	1,291	-1,577	-0,184	2,023	-1,062	-1,061	-0,667	-1,229	1,384	1,250	-1,904	1,037	-1,433	3,442	-0,903	-1,784	-0,751	0,642	-1,548
6F	0,918	0,039	1,718	-0,584	-0,184	1,291	-0,679	-0,184	0,995	-0,141	-0,281	0,116	0,014	-0,594	-0,169	-0,179	-0,578	-0,542	-0,877	-0,083	-0,460	0,494	1,523	-1,639
7F	0,758	0,039	1,718	-0,584	-0,184	1,291	-0,756	-0,184	0,848	-0,243	0,085	0,455	0,309	-0,396	-0,740	0,396	0,078	-0,060	-0,450	0,110	0,355	0,009	1,180	-1,575
8F	-0,659	-1,120	-0,573	0,162	-0,184	-0,784	0,475	-0,184	-0,718	-0,840	-0,573	0,096	-0,760	0,277	0,045	-0,152	-0,070	0,327	0,280	0,476	0,500	-0,796	0,610	-1,652
9F	0,991	0,039	1,718	0,907	-0,184	1,636	0,382	-0,184	0,736	0,276	1,055	2,261	0,069	-0,789	-1,245	1,001	0,597	-0,861	-1,033	-1,562	0,401	1,192	2,018	-1,639
10F	1,133	0,039	-0,573	3,143	-0,184	0,599	2,628	-0,184	0,174	0,127	0,688	2,136	-0,485	-0,957	-0,989	0,733	0,315	-0,811	-1,025	-1,293	0,786	-0,503	0,818	-1,666
11F	1,328	0,039	-0,573	3,889	-0,184	0,599	2,893	-0,184	0,094	0,281	1,332	2,661	-0,353	-0,755	-1,362	0,977	0,502	-1,653	-1,603	-2,496	0,277	1,509	2,242	-1,602
12F	-0,081	0,039	-0,573	2,398	-0,184	0,599	1,355	-0,184	0,028	0,547	1,544	3,000	0,401	-0,886	-1,489	1,929	3,900	-0,860	0,936	-3,011	-0,413	2,037	2,616	-1,606
13F	-1,489	-1,120	-0,573	0,162	-0,184	-0,092	0,277	-0,184	-0,128	-0,846	1,469	-0,458	-0,404	1,095	-1,229	-0,367	1,706	-1,273	1,376	-2,531	-3,125	3,353	3,548	-2,127
14F	0,534	0,039	1,718	-0,584	-0,184	1,982	-0,328	-0,184	2,460	-0,871	-0,859	-0,054	-1,198	-0,417	1,778	-1,761	-0,285	-1,829	0,794	-1,607	-1,077	0,703	1,671	-1,452
15F	-0,343	-1,120	-0,573	-0,584	5,340	-0,438	-0,513	5,006	-0,379	-0,922	1,272	-0,816	-0,465	0,672	-0,928	0,145	0,287	0,095	0,176	0,075	-1,418	0,158	1,286	-1,947

Continuación Tabla 5.

Mol.	d4	d6	d7	d10	d11	d12	d14	d15	d16	d26	d28	d36	d39	d66	d67	d72	d74	d81	d83	d85	d86	d90	d91	d92
1V	0,226	-1,120	-0,573	0,907	-0,184	-0,784	1,502	-0,184	-0,718	-0,968	1,803	-0,443	-0,597	0,314	-0,609	0,066	-1,270	-0,014	-1,619	0,753	1,330	-1,649	-0,473	0,340
2V	0,226	-1,120	-0,573	0,907	-0,184	-0,784	1,502	-0,184	-0,718	-0,968	1,803	-0,443	-0,597	2,690	-0,861	0,268	-1,097	-0,032	-1,615	0,743	1,332	-1,556	-0,472	0,347
3V	0,116	0,039	-0,573	0,162	-0,184	-0,784	0,818	-0,184	-0,718	-0,381	-1,229	-0,634	-1,361	0,386	2,169	-1,184	-1,301	0,095	-0,122	0,713	-0,441	0,326	-0,505	0,589
4V	-1,404	-1,120	-0,573	-0,584	-0,184	-0,784	0,019	-0,184	-0,718	-1,534	0,088	-1,361	-1,229	3,125	2,516	-1,415	-0,270	0,435	2,147	0,896	-0,841	-0,482	-0,509	0,684
5V	2,570	0,039	1,718	-0,584	-0,184	1,291	0,113	-0,184	2,506	-1,193	-1,217	-1,613	-1,163	1,560	2,531	-2,732	-1,856	-2,698	-0,381	-0,220	-1,388	0,436	-0,502	0,422
6V	1,651	0,039	1,718	-0,584	-0,184	1,291	-0,328	-0,184	1,665	-0,762	-0,783	-0,600	-0,984	0,644	0,424	-1,500	-1,108	-1,198	-0,277	0,069	-0,689	0,625	-0,495	0,429
7V	0,116	-1,120	-0,573	0,907	-0,184	-0,784	2,171	-0,184	-0,718	-1,226	1,282	-0,779	-0,811	1,045	-0,059	-0,768	-0,771	-0,282	-0,225	0,480	-0,575	1,419	-0,510	0,333
8V	1,651	0,039	1,718	-0,584	-0,184	1,291	-0,328	-0,184	1,665	-0,720	-1,067	-0,600	-0,954	0,448	0,766	-1,704	-1,598	-1,706	-1,022	-0,179	-0,891	0,658	-0,494	0,431
9V	1,997	1,197	1,718	0,162	-0,184	2,328	0,339	-0,184	2,023	0,014	-0,845	0,477	-1,004	-0,560	0,589	-0,586	-1,203	-0,529	-1,165	0,432	1,320	-0,548	-0,481	0,260
10V	1,997	1,197	1,718	0,162	-0,184	2,328	0,339	-0,184	2,023	0,015	-0,845	0,477	-1,004	0,002	-0,292	-0,477	-1,121	-0,680	-1,277	0,342	0,983	-0,219	-0,483	0,264
11V	-0,098	0,039	-0,573	0,162	-0,184	-0,784	0,339	-0,184	-0,718	-0,270	-0,192	-0,761	-0,109	0,454	-0,556	0,430	-0,410	0,772	-0,525	0,959	0,347	0,763	-0,516	0,400
12V	-0,269	-1,120	-0,573	-0,584	5,340	-0,438	-0,380	5,655	-0,337	-1,106	0,790	-0,949	-0,892	2,318	-0,700	-0,283	-0,131	-0,193	0,009	-0,005	-1,798	-0,217	-0,534	0,456
13V	-0,686	-1,120	-0,573	0,162	-0,184	-0,784	0,277	-0,184	-0,718	-0,631	2,383	-0,561	0,236	2,667	-1,116	0,801	0,349	0,565	-0,177	0,645	-0,086	0,593	-0,506	0,585
14V	-0,625	-1,120	-0,573	0,162	-0,184	-0,784	0,722	-0,184	-0,718	-1,184	1,423	-0,698	-0,485	0,605	0,008	-0,295	0,603	0,370	1,366	0,561	-0,464	-0,282	-0,506	0,638
15V	-0,625	-1,120	-0,573	0,162	-0,184	-0,784	0,722	-0,184	-0,718	-1,014	-0,149	-0,617	-0,923	0,371	0,506	-0,615	-1,175	-0,076	-0,818	0,515	-0,532	0,363	-0,502	0,576
1Fr	-1,612	-1,120	-0,573	-0,584	-0,184	-0,438	-0,428	-0,184	-0,352	-1,043	-0,589	-1,409	-0,078	-0,095	1,952	-0,758	-1,134	0,394	-0,286	0,492	0,222	-0,348	-0,551	0,754
2Fr	-0,450	0,039	-0,573	0,162	-0,184	-0,438	0,113	-0,184	-0,449	0,041	0,542	0,028	0,278	-0,330	-0,806	0,387	0,367	0,030	-0,061	-0,264	-0,385	-0,273	-0,525	0,631
3Fr	-0,526	-1,120	-0,573	0,907	-0,184	-0,438	0,633	-0,184	-0,483	-0,263	0,485	1,104	0,289	-0,961	0,353	0,688	-0,178	0,715	-0,556	0,271	1,133	-0,527	-0,498	0,662
4Fr	1,491	0,039	-0,573	0,162	-0,184	1,291	1,160	-0,184	1,892	-1,103	-0,858	-0,525	-1,442	0,597	0,909	-1,993	-1,085	-2,115	-0,150	-1,021	-0,891	-0,308	-0,505	0,309
5Fr	-1,698	-1,120	-0,573	-0,584	-0,184	-0,438	-0,271	-0,184	-0,302	-1,091	-0,566	-1,540	-0,668	0,115	2,609	-1,021	-1,018	0,480	0,315	0,644	0,041	-0,285	-0,552	0,748
6Fr	-1,489	-1,120	-0,573	-0,584	-0,184	-0,438	-0,650	-0,184	-0,423	-0,733	1,484	-1,078	0,348	1,854	-0,946	0,653	0,410	0,636	0,096	0,398	-0,465	-0,705	-0,556	0,759
7Fr	-1,404	-1,120	-0,573	-0,584	-0,184	-0,438	-0,801	-0,184	-0,471	-0,398	2,452	-0,846	1,328	-0,670	-0,887	1,503	0,733	1,026	-0,177	0,569	0,478	-0,538	-0,557	0,748
8Fr	0,116	1,197	-0,573	0,162	-0,184	-0,784	-0,175	-0,184	-0,718	1,308	-0,176	0,362	0,085	-0,664	-0,760	1,071	0,112	0,770	-0,576	0,582	0,542	0,410	-0,498	0,557
9Fr	0,507	0,039	-0,573	1,652	-0,184	-0,784	1,227	-0,184	-0,718	0,638	0,232	1,443	-0,098	-0,825	-0,924	1,070	0,092	0,094	-0,994	-0,006	0,298	1,118	-0,506	0,340
10Fr	1,112	1,197	1,718	-0,584	-0,184	1,636	-0,619	-0,184	1,414	-0,185	-0,905	0,427	-0,532	-0,102	-0,623	-0,980	0,451	-1,886	0,339	-2,493	-1,129	-0,790	-0,499	0,454
11Fr	-1,489	-1,120	-0,573	0,162	-0,184	-0,092	0,277	-0,184	-0,128	-0,846	1,469	-0,458	-0,404	-0,738	-0,140	-0,505	1,273	-0,613	1,701	-1,738	-2,107	-0,295	-0,558	0,731
12Fr	-0,729	-1,120	-0,573	0,907	-0,184	-0,784	0,753	-0,184	-0,718	-0,135	0,167	1,114	0,278	-0,755	-0,259	0,579	1,542	0,364	1,256	0,131	0,923	-1,500	-0,491	0,531
13Fr	0,305	0,039	-0,573	0,907	-0,184	-0,784	0,958	-0,184	-0,718	0,179	0,132	0,223	0,286	0,013	-0,684	0,503	-0,419	0,426	-0,814	0,548	0,733	0,100	-0,515	0,349
14Fr	0,305	0,039	-0,573	0,907	-0,184	-0,784	0,958	-0,184	-0,718	0,176	-0,683	0,261	-0,159	-0,554	-0,030	0,152	-0,633	0,396	-0,720	0,565	0,105	0,627	-0,497	0,330
15Fr	-0,491	0,039	-0,573	0,162	-0,184	-0,438	-0,024	-0,184	-0,471	0,360	-0,340	0,285	0,329	-0,769	0,305	0,570	-0,028	0,987	-0,094	0,594	0,765	-0,725	-0,502	0,666

Tabla 6. Descriptores seleccionados por medio de la reducción de variables utilizando el Método B.

Mol.	d4	d6	d7	d10	d11	d12	d14	d15	d16	d23	d28	d66	d67	d74	d75	d78	d83	d86	d90	d91	d92
1A	-0,241	1,197	-0,573	-0,584	-0,184	-0,784	-0,801	-0,184	-0,718	0,143	-0,722	-0,143	-0,362	-0,573	-0,725	0,533	-0,563	-0,031	0,843	-0,530	0,594
2A	-0,241	0,039	-0,573	0,162	-0,184	-0,784	0,019	-0,184	-0,718	0,364	-0,600	-0,556	-0,085	-0,590	-0,782	0,613	-0,617	0,314	0,318	-0,494	0,620
3A	-0,119	1,197	-0,573	0,162	-0,184	-0,438	-0,209	-0,184	-0,500	1,054	-0,669	-0,788	-0,133	0,113	-0,109	0,063	0,141	-0,286	1,708	-0,503	0,690
4A	-1,153	0,039	-0,573	-1,329	-0,184	-0,784	-1,577	-0,184	-0,718	0,704	0,852	-0,789	-0,035	0,522	-1,936	1,843	0,496	1,420	-1,151	-0,538	0,807
5A	-0,226	2,356	-0,573	-0,584	-0,184	-0,438	-0,991	-0,184	-0,531	1,724	-0,747	-0,935	-0,252	0,364	-0,993	1,331	-0,117	1,082	-0,372	-0,533	0,741
6A	-0,343	1,197	-0,573	-0,584	-0,184	-0,784	-0,909	-0,184	-0,718	0,945	-0,655	-0,796	-0,147	0,830	-1,152	1,252	0,681	0,744	-0,990	-0,491	0,684
7A	-0,686	1,197	-0,573	-1,329	-0,184	-0,784	-1,577	-0,184	-0,718	-0,351	-1,484	-0,368	1,842	0,275	-1,593	0,474	1,823	1,008	-0,740	-0,539	0,784
8A	-1,612	-1,120	-0,573	-1,329	-0,184	-0,784	-1,577	-0,184	-0,718	0,507	-0,520	-0,630	0,581	-0,355	-1,049	0,565	0,038	0,694	-0,797	-0,540	0,758
9A	1,185	1,197	1,718	-0,584	-0,184	1,291	-0,801	-0,184	0,764	0,604	-0,720	-0,612	-0,368	-0,740	0,025	0,124	-1,069	0,189	0,529	-0,493	0,308
10A	-0,343	1,197	-0,573	-0,584	-0,184	-0,784	-0,909	-0,184	-0,718	0,748	-0,296	-0,858	-0,143	1,755	-1,030	1,292	1,518	0,871	-1,128	-0,505	0,737
11A	-0,343	1,197	-0,573	-0,584	-0,184	-0,784	-0,909	-0,184	-0,718	0,986	-0,804	-0,716	-0,095	0,197	-1,300	1,328	0,028	1,583	-1,528	-0,503	0,699
12A	-0,526	0,039	-0,573	0,162	-0,184	-0,438	-0,140	-0,184	-0,489	0,751	-0,004	-0,911	0,128	0,145	-0,481	0,671	-0,146	0,890	-0,446	-0,515	0,665
13A	-0,526	0,039	-0,573	0,162	-0,184	-0,784	-0,140	-0,184	-0,718	0,948	-0,582	-0,863	0,075	-0,206	-0,000	0,085	-0,383	0,050	1,379	-0,504	0,631
14A	0,443	2,356	1,718	-1,329	-0,184	1,291	-1,577	-0,184	0,357	2,330	-0,465	-0,937	-0,972	2,033	-0,414	1,788	0,501	1,016	-0,852	-0,487	0,512
15A	-0,556	2,356	-0,573	-1,329	-0,184	-0,784	-1,577	-0,184	-0,718	1,247	-1,097	-0,852	-0,006	0,778	-1,231	1,285	0,674	1,026	-0,565	-0,509	0,811
1F	-0,343	1,197	-0,573	-0,584	-0,184	-0,438	-0,586	-0,184	-0,403	-0,690	-1,349	-0,003	0,801	-0,762	-0,312	-0,560	0,127	-1,176	1,844	2,479	-2,081
2F	0,918	0,039	1,718	-0,584	-0,184	1,291	-0,679	-0,184	0,995	0,005	-0,281	0,073	-0,669	-0,526	0,684	-0,497	-0,807	-0,396	0,614	1,608	-1,615
3F	0,758	0,039	1,718	-0,584	-0,184	1,636	-0,428	-0,184	1,840	-0,642	-0,438	-0,006	-0,200	0,957	1,058	-1,479	1,881	-1,543	-1,432	0,160	-1,566
4F	-0,686	-1,120	-0,573	0,162	-0,184	-0,784	0,277	-0,184	-0,718	-0,327	-0,180	-0,126	-0,268	0,167	-0,495	0,419	0,203	0,632	-0,834	0,583	-1,650
5F	0,758	0,039	1,718	-1,329	-0,184	1,291	-1,577	-0,184	2,023	-1,496	-1,061	1,384	1,250	1,037	0,562	-1,749	3,442	-1,784	-0,751	0,642	-1,548
6F	0,918	0,039	1,718	-0,584	-0,184	1,291	-0,679	-0,184	0,995	0,005	-0,281	-0,594	-0,169	-0,578	0,615	-0,423	-0,877	-0,460	0,494	1,523	-1,639
7F	0,758	0,039	1,718	-0,584	-0,184	1,291	-0,756	-0,184	0,848	0,328	0,085	-0,396	-0,740	0,078	0,269	0,187	-0,450	0,355	0,009	1,180	-1,575
8F	-0,659	-1,120	-0,573	0,162	-0,184	-0,784	0,475	-0,184	-0,718	-0,650	-0,573	0,277	0,045	-0,070	-0,422	0,075	0,280	0,500	-0,796	0,610	-1,652
9F	0,991	0,039	1,718	0,907	-0,184	1,636	0,382	-0,184	0,736	1,617	1,055	-0,789	-1,245	0,597	1,786	-0,037	-1,033	0,401	1,192	2,018	-1,639
10F	1,133	0,039	-0,573	3,143	-0,184	0,599	2,628	-0,184	0,174	1,294	0,688	-0,957	-0,989	0,315	1,537	-0,133	-1,025	0,786	-0,503	0,818	-1,666
11F	1,328	0,039	-0,573	3,889	-0,184	0,599	2,893	-0,184	0,094	1,822	1,332	-0,755	-1,362	0,502	3,223	-0,718	-1,603	0,277	1,509	2,242	-1,602
12F	-0,081	0,039	-0,573	2,398	-0,184	0,599	1,355	-0,184	0,028	2,219	1,544	-0,886	-1,489	3,900	2,364	0,424	0,936	-0,413	2,037	2,616	-1,606
13F	-1,489	-1,120	-0,573	0,162	-0,184	-0,092	0,277	-0,184	-0,128	-0,124	1,469	1,095	-1,229	1,706	1,511	-0,983	1,376	-3,125	3,353	3,548	-2,127
14F	0,534	0,039	1,718	-0,584	-0,184	1,982	-0,328	-0,184	2,460	-0,923	-0,859	-0,417	1,778	-0,285	1,147	-1,907	0,794	-1,077	0,703	1,671	-1,452
15F	-0,343	-1,120	-0,573	-0,584	5,340	-0,438	-0,513	5,006	-0,379	-0,591	1,272	0,672	-0,928	0,287	-0,036	0,130	0,176	-1,418	0,158	1,286	-1,947

Continuación Tabla 6.

Mol.	d4	d6	d7	d10	d11	d12	d14	d15	d16	d23	d28	d66	d67	d74	d75	d78	d83	d86	d90	d91	d92
1V	0,226	-1,120	-0,573	0,907	-0,184	-0,784	1,502	-0,184	-0,718	-0,788	1,803	0,314	-0,609	-1,270	0,065	0,022	-1,619	1,330	-1,649	-0,473	0,340
2V	0,226	-1,120	-0,573	0,907	-0,184	-0,784	1,502	-0,184	-0,718	-0,788	1,803	2,690	-0,861	-1,097	0,176	0,129	-1,615	1,332	-1,556	-0,472	0,347
3V	0,116	0,039	-0,573	0,162	-0,184	-0,784	0,818	-0,184	-0,718	-1,149	-1,229	0,386	2,169	-1,301	-0,586	-0,658	-0,122	-0,441	0,326	-0,505	0,589
4V	-1,404	-1,120	-0,573	-0,584	-0,184	-0,784	0,019	-0,184	-0,718	-2,021	0,088	3,125	2,516	-0,270	-0,980	-0,658	2,147	-0,841	-0,482	-0,509	0,684
5V	2,570	0,039	1,718	-0,584	-0,184	1,291	0,113	-0,184	2,506	-1,858	-1,217	1,560	2,531	-1,856	1,213	-2,698	-0,381	-1,388	0,436	-0,502	0,422
6V	1,651	0,039	1,718	-0,584	-0,184	1,291	-0,328	-0,184	1,665	-1,114	-0,783	0,644	0,424	-1,108	0,595	-1,449	-0,277	-0,689	0,625	-0,495	0,429
7V	0,116	-1,120	-0,573	0,907	-0,184	-0,784	2,171	-0,184	-0,718	-1,247	1,282	1,045	-0,059	-0,771	-0,008	-0,598	-0,225	-0,575	1,419	-0,510	0,333
8V	1,651	0,039	1,718	-0,584	-0,184	1,291	-0,328	-0,184	1,665	-1,039	-1,067	0,448	0,766	-1,598	1,043	-1,815	-1,022	-0,891	0,658	-0,494	0,431
9V	1,997	1,197	1,718	0,162	-0,184	2,328	0,339	-0,184	2,023	-0,206	-0,845	-0,560	0,589	-1,203	0,375	-0,631	-1,165	1,320	-0,548	-0,481	0,260
10V	1,997	1,197	1,718	0,162	-0,184	2,328	0,339	-0,184	2,023	-0,206	-0,845	0,002	-0,292	-1,121	0,625	-0,661	-1,277	0,983	-0,219	-0,483	0,264
11V	-0,098	0,039	-0,573	0,162	-0,184	-0,784	0,339	-0,184	-0,718	-0,553	-0,192	0,454	-0,556	-0,410	-0,731	0,673	-0,525	0,347	0,763	-0,516	0,400
12V	-0,269	-1,120	-0,573	-0,584	5,340	-0,438	-0,380	5,655	-0,337	-0,915	0,790	2,318	-0,700	-0,131	0,123	-0,278	0,009	-1,798	-0,217	-0,534	0,456
13V	-0,686	-1,120	-0,573	0,162	-0,184	-0,784	0,277	-0,184	-0,718	-0,483	2,383	2,667	-1,116	0,349	-0,373	0,799	-0,177	-0,086	0,593	-0,506	0,585
14V	-0,625	-1,120	-0,573	0,162	-0,184	-0,784	0,722	-0,184	-0,718	-1,170	1,423	0,605	0,008	0,603	-0,521	0,009	1,366	-0,464	-0,282	-0,506	0,638
15V	-0,625	-1,120	-0,573	0,162	-0,184	-0,784	0,722	-0,184	-0,718	-0,956	-0,149	0,371	0,506	-1,175	-0,164	-0,406	-0,818	-0,532	0,363	-0,502	0,576
1Fr	-1,612	-1,120	-0,573	-0,584	-0,184	-0,438	-0,428	-0,184	-0,352	-0,899	-0,589	-0,095	1,952	-1,134	-0,711	-0,267	-0,286	0,222	-0,348	-0,551	0,754
2Fr	-0,450	0,039	-0,573	0,162	-0,184	-0,438	0,113	-0,184	-0,449	-0,142	0,542	-0,330	-0,806	0,367	0,148	0,235	-0,061	-0,385	-0,273	-0,525	0,631
3Fr	-0,526	-1,120	-0,573	0,907	-0,184	-0,438	0,633	-0,184	-0,483	0,924	0,485	-0,961	0,353	-0,178	-0,589	0,810	-0,556	1,133	-0,527	-0,498	0,662
4Fr	1,491	0,039	-0,573	0,162	-0,184	1,291	1,160	-0,184	1,892	-1,250	-0,858	0,597	0,909	-1,085	1,291	-2,156	-0,150	-0,891	-0,308	-0,505	0,309
5Fr	-1,698	-1,120	-0,573	-0,584	-0,184	-0,438	-0,271	-0,184	-0,302	-1,297	-0,566	0,115	2,609	-1,018	-0,884	-0,393	0,315	0,041	-0,285	-0,552	0,748
6Fr	-1,489	-1,120	-0,573	-0,584	-0,184	-0,438	-0,650	-0,184	-0,423	-0,402	1,484	1,854	-0,946	0,410	-0,506	0,744	0,096	-0,465	-0,705	-0,556	0,759
7Fr	-1,404	-1,120	-0,573	-0,584	-0,184	-0,438	-0,801	-0,184	-0,471	0,244	2,452	-0,670	-0,887	0,733	-0,752	1,523	-0,177	0,478	-0,538	-0,557	0,748
8Fr	0,116	1,197	-0,573	0,162	-0,184	-0,784	-0,175	-0,184	-0,718	0,695	-0,176	-0,664	-0,760	0,112	-0,545	1,090	-0,576	0,542	0,410	-0,498	0,557
9Fr	0,507	0,039	-0,573	1,652	-0,184	-0,784	1,227	-0,184	-0,718	0,998	0,232	-0,825	-0,924	0,092	0,348	0,678	-0,994	0,298	1,118	-0,506	0,340
10Fr	1,112	1,197	1,718	-0,584	-0,184	1,636	-0,619	-0,184	1,414	-0,115	-0,905	-0,102	-0,623	0,451	1,927	-1,654	0,339	-1,129	-0,790	-0,499	0,454
11Fr	-1,489	-1,120	-0,573	0,162	-0,184	-0,092	0,277	-0,184	-0,128	-0,124	1,469	-0,738	-0,140	1,273	0,524	-0,636	1,701	-2,107	-0,295	-0,558	0,731
12Fr	-0,729	-1,120	-0,573	0,907	-0,184	-0,784	0,753	-0,184	-0,718	0,637	0,167	-0,755	-0,259	1,542	-0,199	0,545	1,256	0,923	-1,500	-0,491	0,531
13Fr	0,305	0,039	-0,573	0,907	-0,184	-0,784	0,958	-0,184	-0,718	-0,100	0,132	0,013	-0,684	-0,419	-0,302	0,533	-0,814	0,733	0,100	-0,515	0,349
14Fr	0,305	0,039	-0,573	0,907	-0,184	-0,784	0,958	-0,184	-0,718	0,246	-0,683	-0,554	-0,030	-0,633	-0,391	0,299	-0,720	0,105	0,627	-0,497	0,330
15Fr	-0,491	0,039	-0,573	0,162	-0,184	-0,438	-0,024	-0,184	-0,471	0,527	-0,340	-0,769	0,305	-0,028	-0,934	0,876	-0,094	0,765	-0,725	-0,502	0,666

4.4 ANÁLISIS NO SUPERVISADOS

Después del tratamiento previo y la reducción de los diferentes descriptores moleculares obtenidos a partir de las 60 moléculas odorantes, se procedió a realizar el análisis no supervisado para los diferentes conjuntos de datos, utilizando el paquete computacional Statistica 6.0. Entre los métodos de análisis no supervisados se utilizaron dos, en especial, *i.e.* el análisis de componentes principales y el análisis de agrupamiento.

4.4.1 Análisis de componentes principales (PCA). El análisis de componentes principales es un método utilizado para realizar la transformación de las variables originales en nuevas variables llamadas componentes principales (PC). Cada componente principal es una combinación lineal de las variables originalmente medidas. El uso de este procedimiento es análogo al de encontrar un nuevo sistema de coordenadas, el cual es mejor en cuanto a la información contenida se refiere, que el sistema obtenido de las variables originalmente medidas. Usualmente, solamente dos o tres PC son necesarios para explicar la mayoría de la información presente en el conjunto de datos, siempre y cuando exista un número grande de variables intercorrelacionadas. Por medio del uso del análisis de componentes principales en un sistema de datos de multivariables, se puede realizar la reducción de dimensiones, la clasificación de muestras e identificación de grupos.

Para poder iniciar la investigación de las SORs, los descriptores moleculares obtenidos fueron sometidos al análisis de componentes principales, y, posteriormente, se representaron las 60 moléculas odorantes en cada uno de los subespacios construidos con los primeros PC. Por otro lado, se aplicaron diferentes criterios de selección a los descriptores con el objetivo de analizar los diferentes subespacios para cada tipo de descriptores obtenidos, de esta forma, se busca establecer cuáles subespacios permiten realizar la mejor clasificación de los compuestos fragantes.

Los descriptotes moleculares se dividieron de acuerdo con su clasificación, en 6 categorías, a saber: descriptores constitucionales, topológicos, geométricos, CPSA, electrónicos y vibracionales. Además, se trabajó también con sus combinaciones.

4.4.2 Agrupamiento por medio de métodos jerárquicos. Las técnicas de análisis de grupos han sido diseñadas para encontrar diferentes grupos en un conjunto de datos, donde cada grupo es similar al otro de acuerdo con un conjunto de descriptores dados. Estos métodos han sido ampliamente utilizados en aplicaciones químicas, particularmente, para encontrar grupos de compuestos con estructura o propiedades físicas similares.

Dos métodos de agrupamiento se han convertido en herramienta de uso popular en el manejo de datos químicos, debido en su mayor parte a los excelentes resultados: uno jerárquico (método de Ward) [57] y uno no jerárquico (método de Jarvis–Patrick) [58]. Recientes publicaciones han mostrado que el método de Ward es superior en la separación de compuestos activos de los no activos [49] y en la predicción de propiedades [59].

4.4.2.1 Escogencia de las variables. En esencia, el objetivo de los algoritmos de las técnicas de agrupamiento es agrupar los objetos similares representados en un espacio *n*-dimensional definido por *n*-variables medidas sobre cada uno de los objetos (en nuestro caso, compuestos).

La escogencia inicial de las variables medidas y utilizadas para describir cada uno de los objetos constituye una etapa fundamental en el análisis de agrupamiento. En la mayoría de los casos, el número de variables es determinado empíricamente y a menudo excede el número “real” requerido para establecer una clasificación exitosa. Debido a esta razón, cuando no se presenta una clasificación adecuada, se debe realizar un tratamiento

previo de los datos iniciales con el objetivo de aumentar la eficiencia de los resultados a obtener, así como para eliminar las variables redundantes.

Otra técnica adecuada es realizar inicialmente sobre los datos originales, un análisis de componentes principales para producir un nuevo conjunto de variables estadísticamente independientes y, posteriormente, realizar el análisis de agrupamiento sobre los primeros factores principales que describen la mayor parte de la varianza de los compuestos.

Con base en lo mencionado anteriormente, se inició el estudio de las técnicas de agrupamiento utilizando el método de Ward, el cual es un método jerárquico de aglomeración que inicialmente toma cada uno de los compuestos como grupo independiente, y seguidamente se conforman diferentes grupos de acuerdo con la distancia Euclidena entre los centroides, hasta formar un solo grupo que contenga todos los compuestos bajo estudio.

La clasificación por medio del análisis de agrupamiento se inició, al igual que el desarrollo del análisis de componentes principales, con el estudio de los descriptores separados, para finalmente estudiar la combinación de los mismos.

4.4.2.2 Determinación del número de grupos. Por medio de los dendogramas obtenidos, el siguiente paso es la determinación del número de grupos. Una característica de los análisis de agrupamientos es que la determinación del número de grupos es subjetiva. Para efectuar dicha determinación se examinan dos tipos de métodos. El primero es el método de Kelley y la segunda metodología consiste en fijar el número de grupos manualmente.

Metodología de Kelley. Kelley *et al*. [55] han descrito el cálculo de una medida, la cual es utilizada para la selección del nivel óptimo en los dendogramas. Esta metodología consta de varias etapas, las cuales se describen a continuación:

1. Determinación de la matriz de distancias;
2. Cálculo de la dispersión de cada grupo en cada uno de los niveles presentes en el dendograma;

La dispersión de un *cluster m*, el cual contiene *N* miembros, se determina por medio de:

$$Dispersión_m = \frac{\left(\sum_{k=1}^{N}\sum_{i=1,i>k}^{N} dist(i,k)\right)}{N(N-1)/2} \qquad (8)$$

Donde *i* y *k* son miembros del grupo *m*. El promedio de la dispersión está dado por:

$$AvDisp_i = \frac{\left(\sum_{m=1}^{Numc_i} dispersión_m\right)}{Numc_i} \qquad (9)$$

Donde $Numc_i$ es el número de grupos en el nivel *i* (excluyendo los grupos que contienen un solo miembro)

3. Normalización del promedio de dispersión;

La normalización se realiza para igualar cualquier ponderamiento del número de grupos y del promedio de dispersión en la función de penalidad.

$$AvDisp(norm)_i = \left(\frac{N-2}{Max(AvDisp)-Min(AvDisp)}\right)\left(AvDisp_i - Min(AvDisp)\right)+1 \qquad (10)$$

Donde *Max(AvDisp)* y *Min(AvDisp)* son los valores máximo y mínimo, respectivamente, en el conjunto $\{ AvDisp_1, AvDisp_2 ,\ldots\ldots AvDisp_{N-1}\}$

4. Función de penalidad.

Para cada nivel del *cluster, i,* el valor de penalidad, *Pi,* puede ser calculado por:

$$P_i = AvDisp(norm)_i + nclusti \qquad (11)$$

Donde $nclust_i$ es el número total de grupos en el nivel i (incluyendo los grupos individuales)

5. Definición del límite de corte;

El mínimo valor de la función de penalidad en el conjunto $\{P_1, P_2, \ldots. P_{N-1}\}$, escogido como nivel de corte.

$$P_{icorte} = Min(P) \qquad (12)$$

De esta manera, el *icorte* representa un estado donde los grupos se encuentran poblados lo más altamente posible, pero manteniendo simultáneamente la menor dispersión. Entre más pequeña sea la dispersión, más similares son los miembros de cada grupo.

Una vez establecida la metodología del análisis de componentes principales, así como del análisis de agrupamiento, el siguiente paso es la realización de los métodos no supervisados.

5. DISCUSIÓN DE LOS RESULTADOS

5.1 DESCRIPTORES SELECCIONADOS POR MEDIO DE LOS MÉTODOS A Y B

Inicialmente se sometieron al agrupamiento por análisis de componentes principales y al análisis de agrupamiento los descriptores seleccionados por medio de los Métodos A y B (**Tablas 5 y 6, respectivamente)**.

5.1.1 PCA de los descriptores seleccionados por medio de los Métodos A y B. Los PCA de estos descriptores revelaron que el descriptor d15 (Número relativo de enlaces triples) podía ser eliminado sin perder información significativa en los resultados, por otro lado, el descriptor d11 (Número de enlaces triples) se comporta como un indicador, ya que identifica dos subgrupos de datos con valores diferentes, 0 para 58 compuestos y 1 para los dos restantes.

- **Método A**

En la **Tabla 7** se observan los resultados obtenidos por medio del PCA. El criterio popular ampliamente aceptado de utilizar factores con valores propios mayores de uno [60], conduce a una solución de seis componentes principales. Esta solución explica solamente el 86% de la varianza total de los datos. El primer componente principal explica el 27%, el segundo 20%, el tercero 17%, el cuarto el 11%, el quinto 6% y el sexto 5% de la varianza total.

Tabla 7. Resultados del PCA sobre los descriptores seleccionados por el Método A.

Factor	Valor propio	% Total de varianza	% de Contribución
1	6,251348	27,17978	27,1798
2	4,488800	19,51652	46,6963
3	3,856721	16,76835	63,4646
4	2,602163	11,31375	74,7784
5	1,387785	6,03385	80,8122
6	1,209318	5,25791	86,0702
7	0,863162	3,75288	89,8230
8	0,777192	3,37909	93,2021
9	0,513342	2,23192	95,4340
10	0,256656	1,11590	96,5499
11	0,229966	0,99985	97,5498
12	0,155434	0,67580	98,2256
13	0,136336	0,59276	98,8184
14	0,067575	0,29380	99,1122
15	0,056776	0,24685	99,3590
16	0,041771	0,18161	99,5406
17	0,033992	0,14779	99,6884
18	0,023775	0,10337	99,7918
19	0,020544	0,08932	99,8811
20	0,014023	0,06097	99,9421
21	0,007015	0,03050	99,9726
22	0,004216	0,01833	99,9909
23	0,002091	0,00909	100,0000

Debido a que el PCA se realizó basado en la matriz de correlación, los resultados de los factores de aporte pueden ser interpretados como correlaciones entre las respectivas variables con cada factor. De acuerdo con los diferentes valores obtenidos para los factores de aporte, como se puede observar en el **Anexo 4** y en la **Tabla 8**, se establece, que el primer factor se encuentra mayormente correlacionado con las variables d81 (FPSA-1) y d39 [Contenido de información complementaria (orden 2)] (altas Correlaciones positivas) y d16 (Número relativo de enlaces aromáticos) y d12 (Número de enlaces aromáticos) (altas correlaciones negativas). El factor 2 con d36 [Contenido de información estructural (orden 1)] y d67 (Momento de inercia B), el Factor 3, d6 (Número de anillos) y d28 [Índice de forma de Kier (orden 2)] y el Factor 4, d83 (FPSA-3).

En la **Tabla 8** se presentan los valores de correlación mayores de 0.7 (marcados con un asterisco).

Tabla 8. Resumen de los factores de aporte (Método A).

Descriptores	Factor 1	Factor 2	Factor 3	Factor 4
d4	-0,651859	0,168110	0,469813	0,438478
d6	0,132326	0,180552	0,849153*	-0,017299
d7	-0,656344	-0,044500	0,592345	-0,032904
d10	-0,143944	0,692553	-0,415205	0,475784
d11	-0,060509	-0,090220	-0,306710	-0,227780
d12	-0,769419*	0,138039	0,508415	-0,001328
d14	-0,269938	0,414491	-0,556398	0,584203
d16	-0,828613*	-0,083074	0,459995	-0,003455
d26	0,589792	0,370441	0,617106	-0,067237
d28	0,118373	0,390347	-0,706164*	-0,046311
d36	-0,032753	0,888686*	0,233297	0,136497
d39	0,740119*	0,231108	0,327168	-0,231281
d66	-0,304605	-0,492657	-0,535162	-0,028792
d67	-0,165755	-0,755983*	0,133589	0,031228
d72	0,699521	0,654860	0,026283	-0,028837
d74	0,227569	0,579651	-0,019481	-0,666270
d81	0,955274*	-0,092755	0,019983	-0,054308
d83	0,021376	-0,248751	-0,021851	-0,799594*
d85	0,643182	-0,552766	0,055943	0,307791
d86	0,652729	0,179129	0,328125	0,499586
d90	-0,362408	0,349707	-0,171427	-0,092488
d91	-0,475109	0,545173	-0,140066	-0,404595
d92	0,566015	-0,448301	0,104838	0,323641

La tabla anterior muestra que la variable d81, además de encontrarse altamente correlacionada con el Factor 1, no presenta valores significativos de correlación para los factores restantes. Similarmente, pero en menor proporción, se puede concluir lo mismo para la variable d36 respecto al Factor 2 y la variable d28 para el Factor 3.

Lo mencionado anteriormente, se puede observar más claramente en las gráficas de dispersión de los factores de aporte presentadas en la **Figura 24**. El valor más alto que pueden tomar las factores de aporte (correlación variable-factor) es 1, valor que se encuentra representado por el círculo unidad representado en cada gráfica; entre más cerca se encuentre la variable al círculo, mejor es su representación por el sistema de

coordenadas. En estas gráficas se observa, que la variable d81, se encuentra muy cerca al Factor 1, así mismo, la variable d36 al Factor 2 y la variable d28 al Factor 3.

En la **Figura 25** aparecen las gráficas de dispersión de los factores de coordenadas de las 60 moléculas en los tres primeros factores principales. En la gráfica bidimensional de los Factores 1 y 2, los cuales son los que poseen mayor información, se aprecia la separación de los compuestos de la familia Almizcle y Floral, a excepción de los compuestos 4F, 8F y 9A. Por lo tanto, el Factor 1, el cual representa la relación entre el área superficial accesible al solvente positivamente cargada (FPSA1) es el encargado de esta separación, además, los compuestos del grupo Almizcle se encuentran localizados en la región positiva del Factor 1, mientras que los compuestos de la familia Floral se encuentran en la región negativa. Por otro lado, en la representación de los Factores 1 y 3, además de presentarse la separación anteriormente mencionada, se registra la separación de la familia Almizcle de la Verde, junto con un pequeño agrupamiento de algunos compuestos de la familia Frutal. El Factor 3 representa el número de anillos presentes en los compuestos. Por último, en la gráfica de los Factores 2 y 3 no se apreciar alguna diferenciación de los compuestos odorantes.

En resumen, a partir de la metodología de eliminación de los descriptores, utilizando los criterios descritos en el Método A, se obtiene la clasificación principalmente de un solo grupo, el grupo Almizcle, como lo muestra la gráfica tridimensional de los tres primeros PC; además, se aprecia el agrupamiento de algunos compuestos del grupo Verde y Frutal.

De acuerdo con los resultados obtenidos, se establece, que la clasificación de las familias fragantes no fue completa. De esta manera, se procedió a analizar los descriptores seleccionados por medio del Método B.

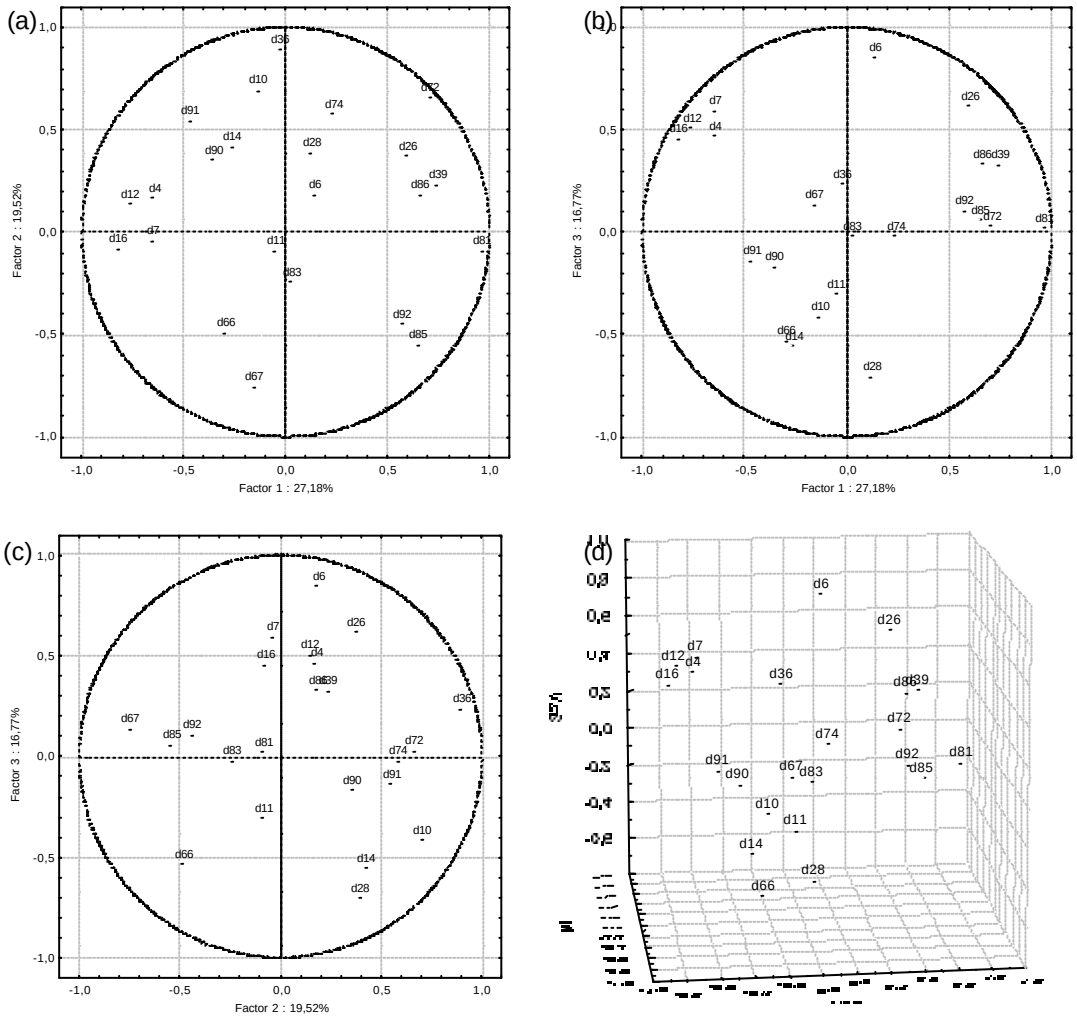


Figura 24. Gráficas de los factores de aporte en las tres dimensiones (Método A).

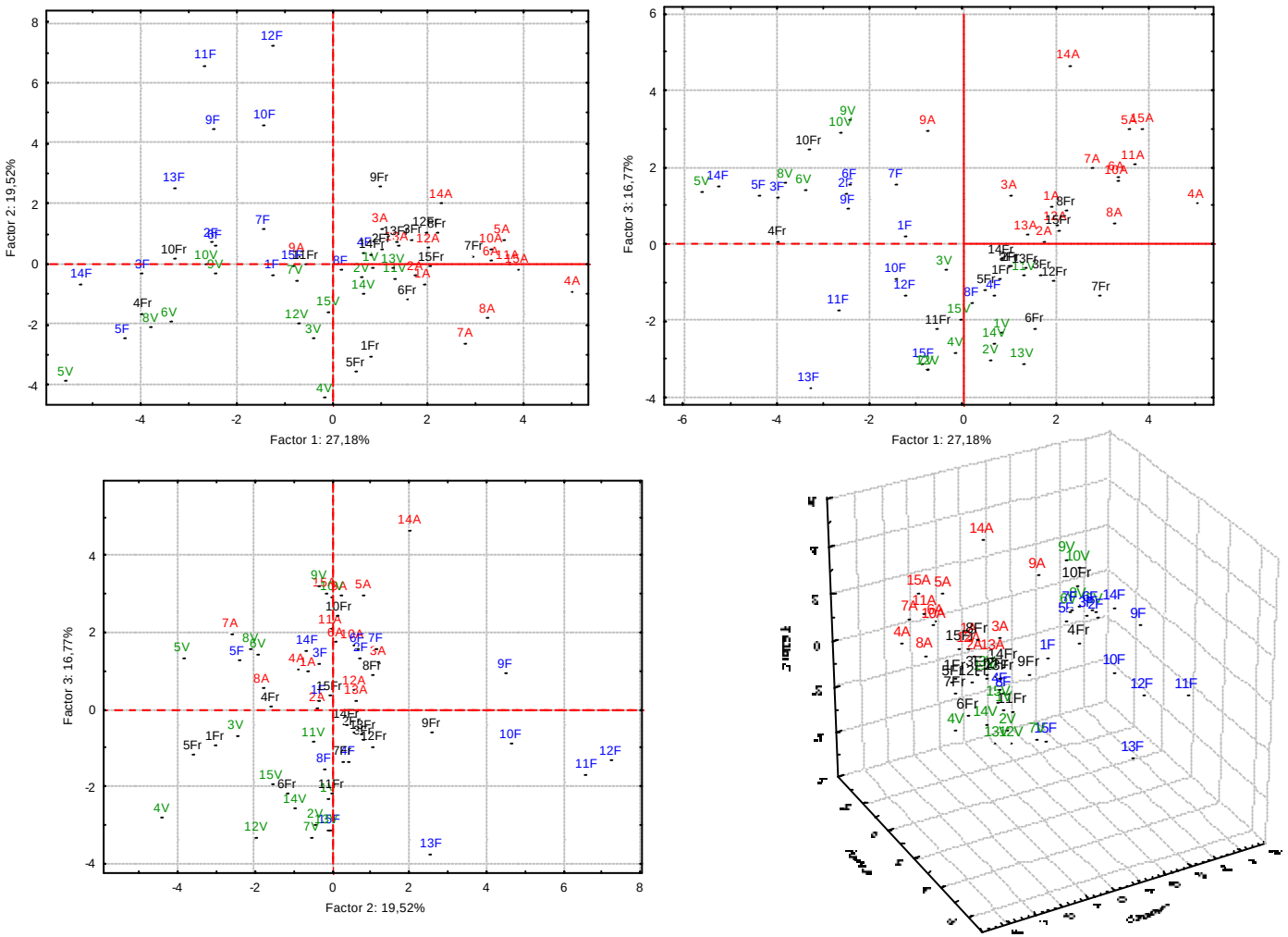


Figura 25. Representación de las 60 moléculas bajo estudio en el subespacio formado por los tres primeros PC (63% de la información), luego de la selección de descriptores según el Método A.

- **Método B**

Los resultados del PCA sobre los descriptores obtenidos por el Método B se resumen en la **Tabla 9**. Estos resultados muestran que solamente los seis primeros PC poseen un valor superior de la unidad, lo cual conduce a una solución de seis componentes principales. Esta solución explica solamente el 87% de la varianza total de los datos originales; sin embargo, se presenta un pequeño aumento de este valor, con referencia a los resultados obtenidos en el Método A. El primer componente principal explica el 27%, el segundo 19%, el tercero 16%, el cuarto el 12%, el quinto 7% y el sexto 5% de la varianza total.

Tabla 9. Resultados del PCA sobre los descriptores seleccionados por el Método B.

Factor	Valor propio	% Total de varianza	% de Contribución
1	5,369767	26,84883	26,8488
2	3,873829	19,36914	46,2180
3	3,246699	16,23349	62,4515
4	2,438485	12,19242	74,6439
5	1,304232	6,52116	81,1651
6	1,098505	5,49252	86,6576
7	0,772539	3,86269	90,5203
8	0,674636	3,37318	93,8935
9	0,459703	2,29851	96,1920
10	0,234669	1,17335	97,3653
11	0,158068	0,79034	98,1557
12	0,124321	0,62160	98,7773
13	0,065531	0,32766	99,1049
14	0,058881	0,29440	99,3993
15	0,043188	0,21594	99,6153
16	0,028404	0,14202	99,7573
17	0,021587	0,10794	99,8652
18	0,017692	0,08846	99,9537
19	0,007678	0,03839	99,9921
20	0,001587	0,00794	100,0000

La **Tabla 10** muestra los valores de correlación entre las variables seleccionadas y los primeros cuatro Factores. Las variables que presentan los mayores valores de correlación variable-factor son: Factor 1, d16 (Número relativo de enlaces aromáticos), d12 (Número de enlaces aromáticos), d78 (DPSA-1), d75 (PNSA-1), d7 (Número de anillos de benceno), d4 (Número relativo de átomos de C), Factor 2, d10 (Número de enlaces dobles), d67 (Momento de inercia B), Factor 3, d6 (Número de anillos), d66 (Momento de inercia A) y d23 [Índice de Kier&Hall (orden 0)] y Factor 4, d83 (FPSA-3), d74 (PPSA-3). En el **Anexo 5** se presentan los factores de aporte de todos los componentes principales.

Tabla 10. Resumen de los factores de aporte (Método B).

Descriptores	Factor 1	Factor 2	Factor 3	Factor 4
d4	-0,756944*	-0,052729	0,373331	-0,383963
d6	-0,050402	-0,250021	0,810504*	0,085187
d7	-0,762305*	-0,326358	0,352283	0,071016
d10	-0,118128	0,817133*	-0,016091	-0,430485
d11	-0,007788	0,068934	-0,346652	0,218040
d12	-0,864579*	-0,129264	0,346807	0,042206
d14	-0,187075	0,650224	-0,280163	-0,573218
d16	-0,886987*	-0,297960	0,184755	0,014733
d23	0,279015	0,498036	0,762180*	0,147585
d28	0,192953	0,683802	-0,382281	0,035927
d66	-0,167842	-0,168416	-0,772396*	-0,049069
d67	-0,090449	-0,733469*	-0,278739	-0,101728
d74	0,198399	0,452053	0,239639	0,718910*
d75	-0,821021*	0,478110	0,006714	-0,030284
d78	0,847195*	0,217177	0,374569	0,098316
d83	0,096103	-0,281836	-0,212716	0,772553*
d86	0,544489	0,018167	0,563499	-0,434590
d90	-0,376098	0,415004	-0,101060	0,094270
d91	-0,505882	0,543054	-0,009816	0,456174
d92	0,586703	-0,438543	0,027664	-0,386509

A partir de las magnitudes de los factores de aporte se logra apreciar que las variables, que se encuentran mayormente correlacionadas con cada uno de los tres primeros

factores, corresponden a descriptores constitucionales; además, el Factor 1, el cual contiene la mayor cantidad de información, se encuentra representado por cuatro descriptores constitucionales y dos descriptores de área superficial parcialmente cargada.

En la **Figura 26** se presentan las gráficas de dispersión de los factores de aporte de las variables en los factores principales que contienen la mayor información. En estas gráficas se observa la presencia de un grupo conformado por los cuatro descriptores constitucionales d4, d7, d12 y d16, los cuales corresponden a las variables que presentan valores de correlación variable-factor altos con el Factor 1. Además, se aprecia que estas variables se encuentran cerca del Factor 1 y del círculo unidad, lo que se podía esperar de acuerdo con las magnitudes de los factores de aporte. De la misma manera, se presenta un comportamiento similar de la variable d10 respecto al Factor 2 y de la variable d6, respecto al Factor 3.

Las gráficas de dispersión de los factores de coordenadas de las 60 moléculas en los tres primeros factores principales muestran un patrón de separación similar al obtenido en el análisis anterior (**Figura 27**).

A partir de los resultados de clasificación obtenidos por medio del análisis de componentes principales en los dos tratamientos anteriores, se observa una clasificación deficiente de las cuatro familias odorantes bajo estudio. En ambos tratamientos, solamente se aprecia una ligera clasificación de la familia Almizcle, principalmente en las gráficas de dispersión del primer y tercer factor principal, al igual que en las gráficas tridimensionales.

Por otro lado, se establece que estas clasificaciones se presentan debido principalmente a las variables que realizan los mayores aportes en los primeros componentes principales y, como característica común en ambos resultados, las variables de mayor aporte son los descriptores constitucionales.

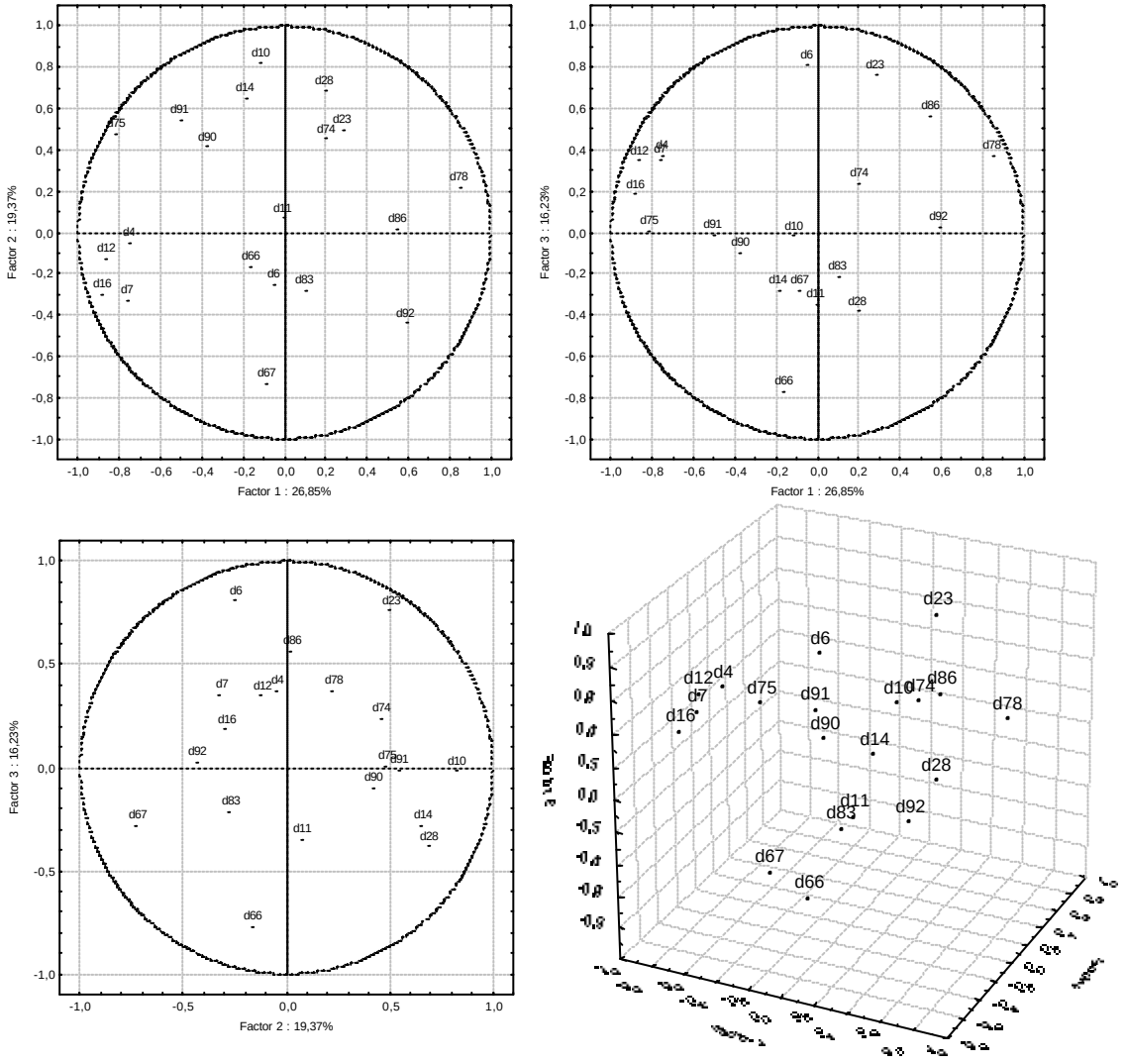


Figura 26. Gráficas de los factores de aporte en las tres dimensiones (Método B).

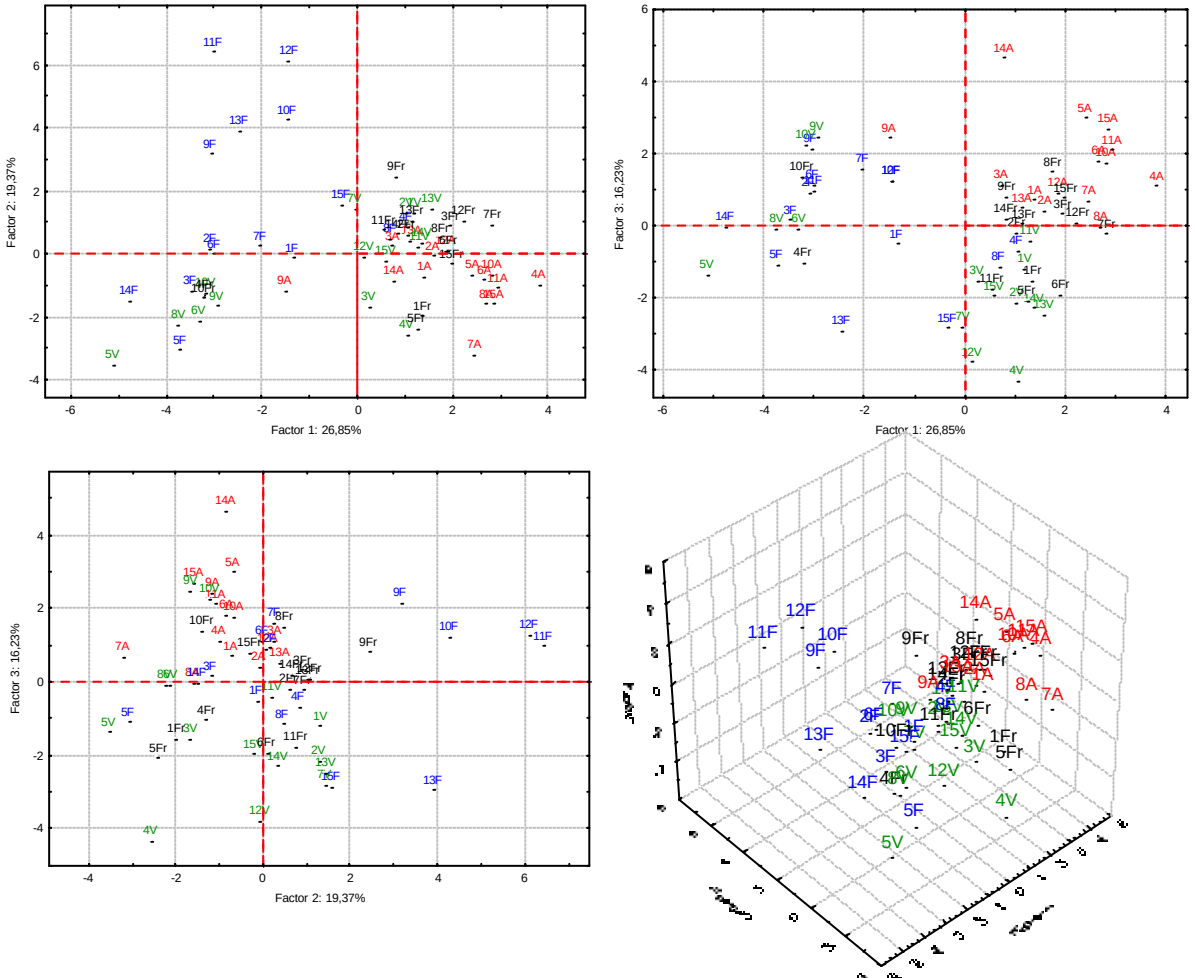


Figura 27. Representación de las 60 moléculas bajo estudio en el subespacio formado por los tres primeros PC (63% de la información), luego de la selección de descriptores según el Método B.

Con base en las observaciones anteriores, y continuando con el estudio de clasificación de los compuestos fragantes, se procedió a separar los diferentes descriptores de acuerdo con su clasificación, para posteriormente someter individualmente cada grupo de descriptores a la clasificación por medio del análisis de componentes principales y al análisis de agrupamiento. Es importante subrayar que el tratamiento previo realizado sobre cada tipo de descriptores fue el mismo utilizado en el análisis de los 92 descriptores moleculares; esto es, que para poder eliminar los descriptores se utilizaron los criterios establecidos tanto en el Método A, como en el Método B.

Antes de realizar el PCA sobre los diferentes tipos de descriptores, se desarrolló el análisis de agrupamiento sobre los descriptores seleccionados por medio de los Métodos A y B. De esta manera, se efectuó una comparación de los resultados obtenidos en los diferentes análisis no supervisados.

5.1.2 Análisis de agrupamiento de los descriptores seleccionados por medio de los Métodos A y B. El análisis de agrupamiento se llevo a cabo utilizando el método de Ward basado en la distancia Euclideana sobre las diferentes matrices obtenidas por medio de los Métodos A y B (**Tablas 5 y 6, respectivamente**).

- **Método A**

El dendograma obtenido a partir de las variables seleccionadas por medio del Método A, utilizando la metodología de Ward, se presenta en la **Figura 28**. En esta figura se aprecia que las moléculas presentes en la izquierda del dendograma, se encuentran separadas de los otros grupos. Estas moléculas pertenecen en su mayoría a la familia Floral, excepto por los compuestos 5V, 6V, 8V, 9V, 10V, 4Fr y 9A (**Figura 28)**.

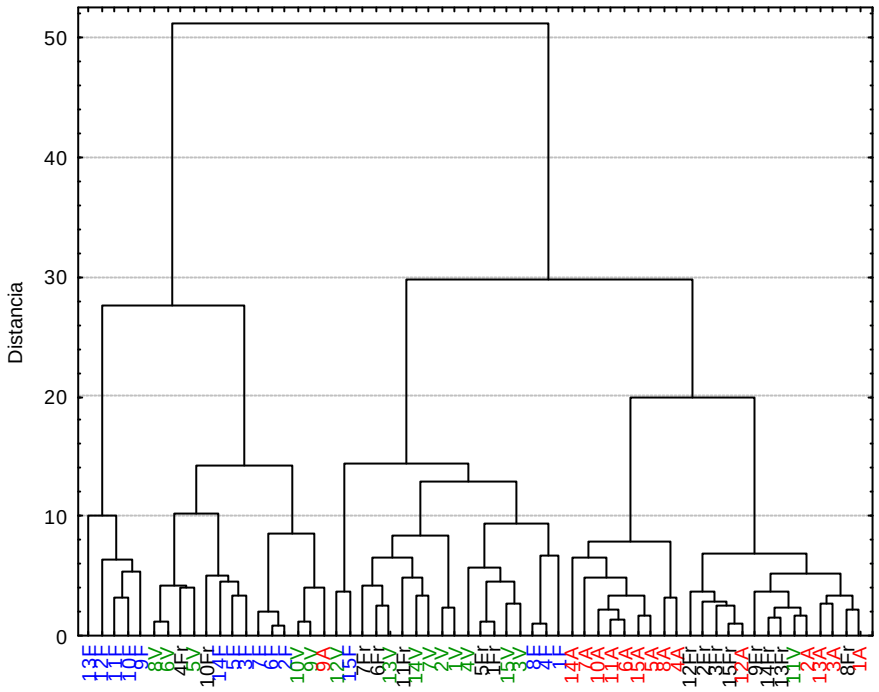


Figura 28. Dendograma obtenido a partir de las variables seleccionadas por medio del Método A, utilizando el método de Ward.

A partir del dendograma (**Figura 28)** y de la matriz de distancia, obtenida por medio del análisis de agrupamiento, utilizando la distancia Euclediana (**Anexo 6**), se procedió a determinar el número de grupos.

Usando la metodología Kelley, se obtuvieron un total de 15 grupos. El segundo acercamiento utilizado fue fijar el número de grupos manualmente. Como el número de familias olfativas presentes en este estudio es de cuatro, se estableció que el número de grupos fuera 4, 5 ó 6, para poder realizar comparaciones entre los diferentes grupos; estos resultados se resumen en la **Tabla 11**.

Tabla 11. Número de grupos determinados por medio de los métodos de Kelley y manual.

Método A														
	Grupos													
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**	**C_{10}**	**C_{11}**	**C_{12}**	**C_{13}**	**C_{14}**
Floral	4	0	3	3	0	1	0	0	0	2	0	0	0	0
Verde	0	3	0	0	2	1	3	2	3	0	0	0	0	1
Frutal	0	1	1	0	0	0	3	0	2	0	0	0	4	4
Almizcle	0	0	0	0	1	0	0	0	0	0	7	2	1	4

Grupos establecidos por medio del índice de Kelley

	Grupos				
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**
Floral	5	6	4	0	0
Verde	0	5	9	0	1
Frutal	0	2	5	0	8
Almizcle	0	1	0	9	5

Número de grupos establecido manualmente

En la **Tabla 12** se reportan las moléculas pertenecientes a cada uno de los diferentes grupos. De acuerdo con estos resultados, se observa, que los grupos C_1 y C_4 obtenidos manualmente, están compuestos por moléculas pertenecientes a una sola de las cuatro familias odorantes, así: Floral y Almizcle, respectivamente.

Tabla 12. Moléculas pertenecientes a cada uno de los grupos determinados.

Grupos	**Moléculas**	**Grupos**	**Moléculas**
C_1	9-12F	**C_8**	1V, 2V
C_2	5V, 6V, 8V, 4Fr	**C_9**	3V, 4V, 15V, 1Fr, 5Fr
C_3	3F, 5F, 14F, 10Fr	**C_{10}**	4F, 8F
C_4	2F, 6F, 7F	**C_{11}**	5-7A, 10A, 11A, 14A, 15A
C_5	9V, 10V, 9A	**C_{12}**	4A, 8A
C_6	12V, 15F	**C_{13}**	12A, 2Fr, 3Fr, 12Fr, 15Fr
C_7	7V, 13V, 14V, 6Fr, 7Fr, 11Fr	**C_{14}**	1-3A, 13A, 8Fr, 9Fr, 13Fr, 14Fr, 11V
Unitarios	13F, 1F		

Grupos establecidos por el índice de Kelley

Grupos	**Moléculas**
C_1	9-13F
C_2	9A, 5V, 6V, 8-10V, 2,F, 3F, 5-7F, 14F, 4Fr, 10Fr
C_3	1-4V, 7V, 12-15V, 1F, 4F, 8F, 15F, 1Fr, 5-7Fr, Fr
C_4	4-8A, 10A,11A, 14A, 15A
C_5	1-3A, 12A,13A, 11V, 2Fr, 3Fr, 8Fr, 9Fr, 12-15Fr

Grupos establecidos manualmente

- **Método B**

En la **Figura 29** se presenta el dendograma obtenido a partir de las variables seleccionadas por el Método B, utilizando el método de Ward. Al igual que en el resultado anterior, se aprecia que las moléculas presentes en la izquierda del dendograma, se encuentran separadas de los otros grupos. Estas moléculas pertenecen, en su mayoría, a la familia Floral, pero además, se observa un pequeño grupo de compuestos pertenecientes a la familia Verde.

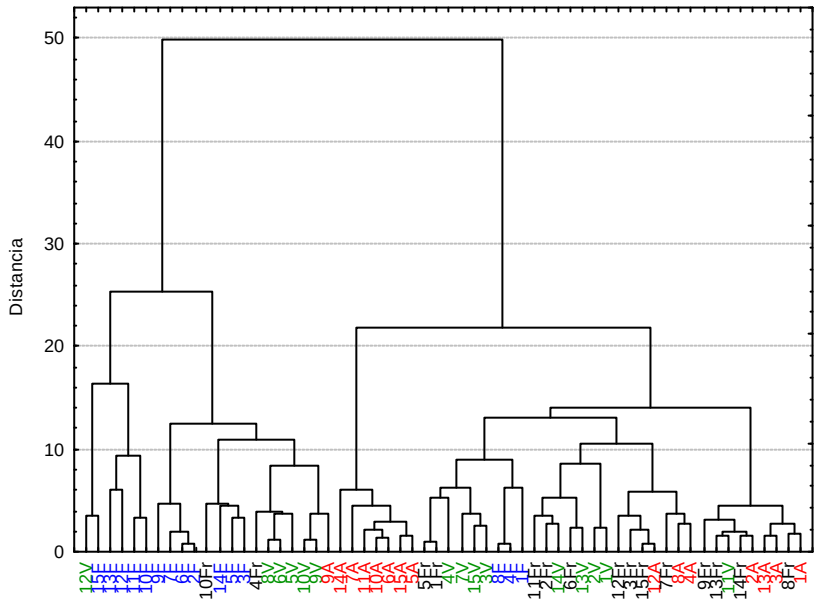


Figura 29. Dendograma obtenido a partir de las variables seleccionadas por el Método B, utilizando el método de Ward.

En el **Anexo 7** se presenta la matriz de distancias, obtenida por medio del análisis de agrupamiento, utilizando la distancia Euclediana. Los resultados de la determinación del número de grupos, así como las moléculas pertenecientes a cada grupo, se presentan en las **Tablas 13 y 14**.

Tabla 13. Número de grupos determinados por medio del método de Kelley y manualmente.

Método A														
	Grupos													
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**	**C_{10}**	**C_{11}**	**C_{12}**	**C_{13}**	**C_{14}**
Floral	1	2	2	4	3	0	0	0	0	2	0	0	0	0
Verde	1	0	0	0	0	3	2	0	4	0	2	2	0	1
Frutal	0	0	0	0	1	1	0	0	2	0	3	0	4	4
Almizcle	0	0	0	0	0	0	1	7	0	0	0	0	3	4

Grupos establecidos por medio del índice de Kelley

	Grupos				
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**
Floral	1	4	7	0	3
Verde	1	0	5	0	9
Frutal	0	0	2	0	13
Almizcle	0	0	1	7	7

Número de grupos establecido manualmente

Tabla 14. Moléculas pertenecientes a cada uno de los grupos determinados.

Grupos	Moléculas	Grupos	Moléculas
C_1	12V, 15F	**C_8**	5-7A, 9-11A, 14A, 15A
C_2	12F, 13F	**C_9**	3V, 4V, 7V, 15V, 1Fr, 5Fr
C_3	10F, 11F	**C_{10}**	4F, 8F
C_4	2F, 6F, 7F, 9F	**C_{11}**	13V, 14V, 2Fr, 6Fr, 11Fr
C_5	3F, 5F, 14F, 10Fr	**C_{12}**	1V, 2V
C_6	5V, 6V, 8V, 4Fr	**C_{13}**	4A, 8A, 12A, 3Fr, 7Fr, 12Fr, 15Fr
C_7	9V, 10V, 9A	**C_{14}**	1-3A, 13A, 8Fr, 9Fr, 13Fr, 14Fr, 11V
Unitarios	1F		

Grupos establecidos por el índice de Kelley

Grupos	Moléculas
C_1	15F, 12V
C_2	10-13F
C_3	9A, 5V, 6V, 8-10V, 4Fr, 2F, 3F, 5-7F, 9F, 14F, 10Fr
C_4	5-7A, 10A, 11A, 14A, 15A
C_5	1-4A, 8A, 12A, 13A, 1-3Fr, 5-9Fr, 11-15Fr, 1-4V, 7V, 11V, 13-15V, 1F, 4F, 8F

Grupos establecidos manualmente

Se observa, que por medio del análisis de agrupamiento, utilizando los diferentes Métodos A y B, solamente se obtiene la clasificación de algunas moléculas de las

familias Almizcle y Floral. Estos resultados se asemejan a los obtenidos por medio del análisis de componentes principales, realizados en la sección anterior (5.1.1).

5.2 CLASIFICACIÓN DE LOS DIFERENTES DESCRIPTORES MOLECULARES

Después de realizar la clasificación de los diferentes descriptotes moleculares en 6 categorías, a saber: descriptores constitucionales, topológicos, geométricos, CPSA, electrónicos y vibracionales, se procedió a realizar el análisis de componentes principales. Sin embargo, los resultados obtenidos por medio del PCA no fueron satisfactorios, salvo el caso de la clasificación de los compuestos odorantes utilizando los descriptores electrónicos y topológicos, obtenidos por la reducción de variables basada en el Método A.

5.2.1 PCA de los descriptores electrónicos. En el presente estudio, los descriptores electrónicos están constituidos por 6 descriptores (d87-d92). La reducción de descriptores por medio de los Métodos A y B condujo a una misma matriz de 60 x 4, como resultado final. En la **Tabla 15** se presentan los valores de los descriptores obtenidos.

De acuerdo con los valores propios obtenidos por el PCA (**Tabla 16**), solamente los dos primeros factores principales presentan valores mayores que uno. Estos resultados conducen a una solución de dos componentes principales. La solución, en este caso, solamente explica el 81% de la varianza total. Por otro lado, la gráfica de los valores propios *vs* factores principales sugiere la utilización de los tres primeros factores principales (**Figura 30**).

Tabla 15. Descriptores electrónicos obtenidos por medio de la reducción de datos utilizando los Métodos A y B.

	d88	d90	d91	d92		d88	d90	d91	d92		d88	d90	d91	d92
1A	0,485	0,843	-0,530	0,594	**6F**	-0,194	0,494	1,523	-1,639	**11V**	-0,332	0,763	-0,516	0,400
2A	0,421	0,318	-0,494	0,620	**7F**	0,189	0,009	1,180	-1,575	**12V**	-1,056	-0,217	-0,534	0,456
3A	1,003	1,708	-0,503	0,690	**8F**	-0,456	-0,796	0,610	-1,652	**13V**	-0,063	0,593	-0,506	0,585
4A	1,482	-1,151	-0,538	0,807	**9F**	1,152	1,192	2,018	-1,639	**14V**	-0,809	-0,282	-0,506	0,638
5A	1,783	-0,372	-0,533	0,741	**10F**	0,775	-0,503	0,818	-1,666	**15V**	-0,836	0,363	-0,502	0,576
6A	1,227	-0,990	-0,491	0,684	**11F**	1,223	1,509	2,242	-1,602	**1Fr**	-0,826	-0,348	-0,551	0,754
7A	0,017	-0,740	-0,539	0,784	**12F**	1,762	2,037	2,616	-1,606	**2Fr**	0,146	-0,273	-0,525	0,631
8A	0,954	-0,797	-0,540	0,758	**13F**	-0,230	3,353	3,548	-2,127	**3Fr**	0,821	-0,527	-0,498	0,662
9A	0,312	0,529	-0,493	0,308	**14F**	-1,399	0,703	1,671	-1,452	**4Fr**	-1,625	-0,308	-0,505	0,309
10A	1,255	-1,128	-0,505	0,737	**15F**	-0,674	0,158	1,286	-1,947	**5Fr**	-1,200	-0,285	-0,552	0,748
11A	1,239	-1,528	-0,503	0,699	**1V**	-0,431	-1,649	-0,473	0,340	**6Fr**	-0,052	-0,705	-0,556	0,759
12A	0,879	-0,446	-0,515	0,665	**2V**	-0,436	-1,556	-0,472	0,347	**7Fr**	0,719	-0,538	-0,557	0,748
13A	0,875	1,379	-0,504	0,631	**3V**	-1,096	0,326	-0,505	0,589	**8Fr**	0,924	0,410	-0,498	0,557
14A	1,903	-0,852	-0,487	0,512	**4V**	-1,644	-0,482	-0,509	0,684	**9Fr**	0,943	1,118	-0,506	0,340
15A	1,306	-0,565	-0,509	0,811	**5V**	-2,087	0,436	-0,502	0,422	**10Fr**	-0,734	-0,790	-0,499	0,454
1F	-0,598	1,844	2,479	-2,081	**6V**	-1,333	0,625	-0,495	0,429	**11Fr**	-0,215	-0,295	-0,558	0,731
2F	-0,196	0,614	1,608	-1,615	**7V**	-1,127	1,419	-0,510	0,333	**12Fr**	0,604	-1,500	-0,491	0,531
3F	-1,186	-1,432	0,160	-1,566	**8V**	-1,343	0,658	-0,494	0,431	**13Fr**	0,141	0,100	-0,515	0,349
4F	-0,069	-0,834	0,583	-1,650	**9V**	-0,599	-0,548	-0,481	0,260	**14Fr**	0,105	0,627	-0,497	0,330
5F	-1,701	-0,751	0,642	-1,548	**10V**	-0,602	-0,219	-0,483	0,264	**15Fr**	0,506	-0,725	-0,502	0,666

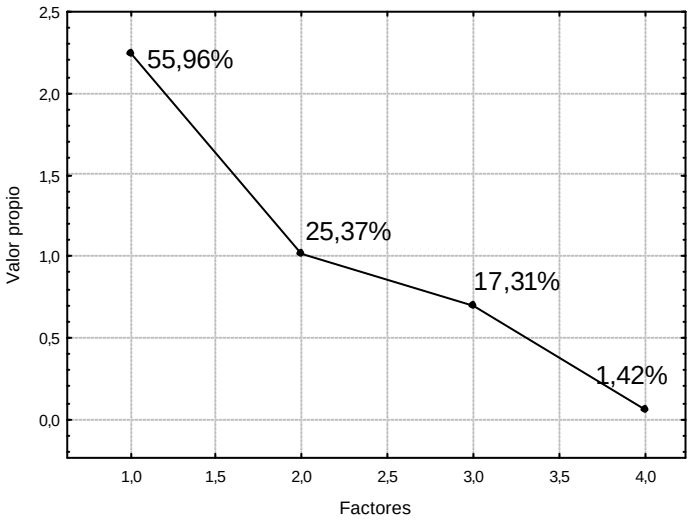


Figura 30. Gráfica de dispersión de los valores propios obtenidos por el PCA sobre los descriptores electrónicos.

En este caso, los tres primeros factores principales retenidos contienen el 98% de la varianza total de los datos originales. El primer factor principal explica el 56%, el segundo el 25% y el tercero el 17%, de la información total.

Tabla 16. Resultados del PCA sobre los descriptores electrónicos.

Factor	Valor propio	% Total de varianza	% de Contribución
1	2,238451	55,96128	55,9613
2	1,012318	25,30794	81,2692
3	0,692375	17,30939	98,5786
4	0,056856	1,42139	100,0000

En la **Tabla 17** se presentan los factores de aporte calculados para cada uno de los descriptores electrónicos. Los descriptores d91 y d92 poseen valores de correlación altos con el Factor 1 (0,967) y (-0.907), mientras que con los otros dos factores, éstos son significativamente más bajos. De la misma manera, el descriptor d88 posee una correlación excepcionalmente alta con el Factor 2 y no presenta valores significativos con los descriptores restantes. Por otro lado, el descriptor d90, se encuentra relacionado con el Factor 3, pero además, presenta un valor alto (0.688) con el Factor 1, lo anterior se aprecia claramente en las gráficas de los factores de aporte.

Tabla 17. Factores de aporte de los descriptores electrónicos.

Descriptor	Factor 1	Factor 2	Factor 3	Factor 4
d88	-0,070326	0,993385*	0,086911	-0,026201
d90	0,688826	0,092630	-0,717564*	-0,045170
d91	0,966923*	0,092314	0,161393	0,174617
d92	-0,907791*	0,091657	-0,379310	0,153746

En la **Figura 31** se representan las gráficas de los factores de aporte de los tres componentes principales. En estas gráficas se aprecia como los Factores 1 y 2 son adecuados para la representación de los descriptores d91, d92 y d88, de acuerdo con su

cercanía a los diferentes factores, así como con su magnitud. Por otro lado, la gráfica de los factores de aporte representados por los Factores 1 y 3 muestra, que la variable d90 se encuentra en el cuarto cuadrante, relacionada con los Factores 3 y 1.

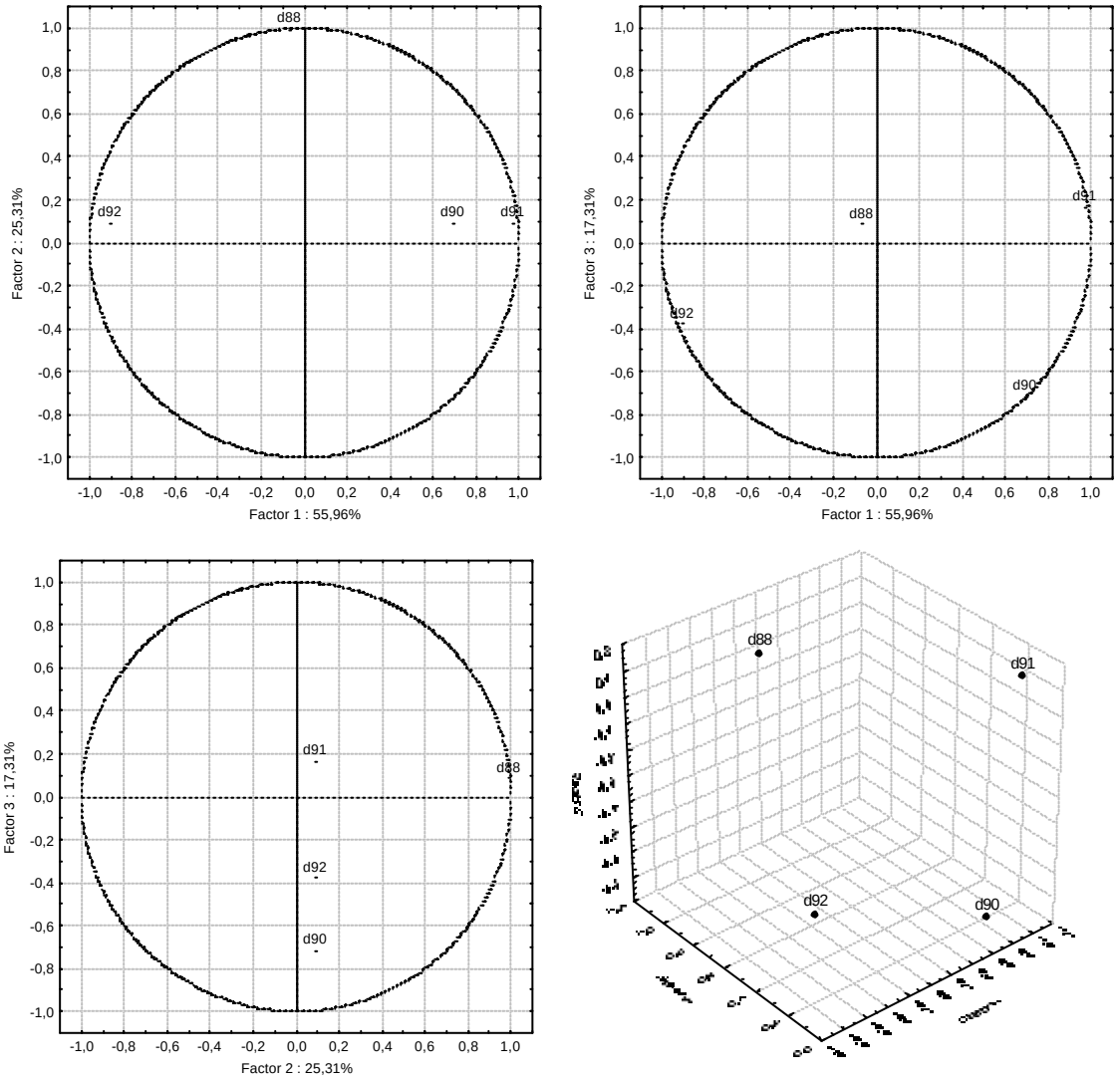


Figura 31. Gráficas de los factores de aportes de las variables de los tres componentes principales (Descriptores electrónicos).

De acuerdo con la **Tabla 17**, junto con las observaciones realizadas sobre las gráficas de dispersión de los factores de coordenadas de las moléculas en las tres dimensiones

(**Figura 32**), se establece, que el primer factor principal se encuentra relacionado con las energías de HOMO y LUMO. Este PC separa la familia Floral de los grupos Almizcle, Frutal y Verde. El segundo factor se encuentra relacionado con la energía del punto cero (ZPE) y separa la familia Verde del grupo Almizcle, mientras que el tercer factor principal relacionado con el momento dipolar no influye en la separación de grupos fragantes.

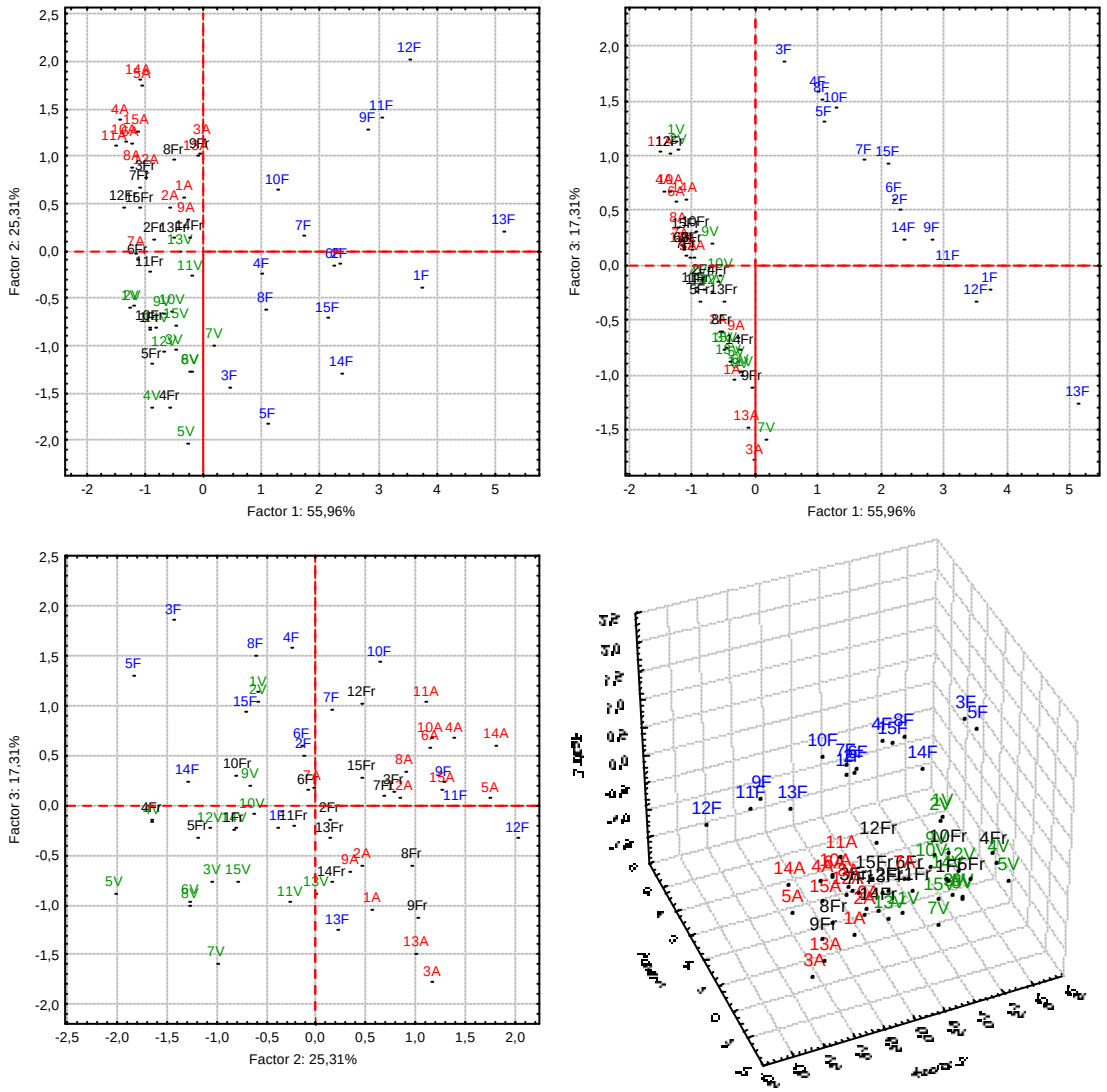


Figura 32. Gráficas de dispersión de los factores de coordenadas de las moléculas fragantes en las tres dimensiones (Descriptores electrónicos).

5.2.2 PCA de los descriptores topológicos. La reducción de variables utilizando el Método A condujo a una matriz de 60 x 4. En la **Tabla 18** se resumen los valores de los descriptores resultantes. Por otro lado, los resultados obtenidos por la reducción de variables utilizando, el Método B, no fueron adecuados de acuerdo con la clasificación; debido a ésto se omiten en esta discusión.

Tabla 18. Descriptores topológicos resultantes de la reducción de variables (Método A).

	d26	d28	d32	d39		d26	d28	d32	d39		d26	d28	d32	d39
1A	1,188	-0,722	-0,607	0,512	**6F**	-0,141	-0,281	0,099	0,014	**11V**	-0,270	-0,192	-0,527	-0,109
2A	0,458	-0,600	-0,058	0,309	**7F**	-0,243	0,085	0,396	0,309	**12V**	-1,106	0,790	-0,017	-0,892
3A	2,131	-0,669	0,992	0,076	**8F**	-0,840	-0,573	-0,602	-0,760	**13V**	-0,631	2,383	-0,461	0,236
4A	0,819	0,852	-0,793	3,632	**9F**	0,276	1,055	1,965	0,069	**14V**	-1,184	1,423	-0,882	-0,485
5A	2,582	-0,747	0,824	1,889	**10F**	0,127	0,688	1,637	-0,485	**15V**	-1,014	-0,149	-0,961	-0,923
6A	1,522	-0,655	0,148	1,123	**11F**	0,281	1,332	2,025	-0,353	**1Fr**	-1,043	-0,589	-0,887	-0,078
7A	1,192	-1,484	-1,507	0,157	**12F**	0,547	1,544	3,267	0,401	**2Fr**	0,041	0,542	0,386	0,278
8A	0,702	-0,520	-1,149	2,344	**13F**	-0,846	1,469	0,679	-0,404	**3Fr**	-0,263	0,485	1,125	0,289
9A	0,432	-0,720	0,594	0,371	**14F**	-0,871	-0,859	0,012	-1,198	**4Fr**	-1,103	-0,858	-0,716	-1,442
10A	1,745	-0,296	0,148	0,480	**15F**	-0,922	1,272	0,226	-0,465	**5Fr**	-1,091	-0,566	-1,121	-0,668
11A	1,665	-0,804	0,148	1,143	**1V**	-0,968	1,803	-1,020	-0,597	**6Fr**	-0,733	1,484	-0,436	0,348
12A	0,868	-0,004	0,843	0,429	**2V**	-0,968	1,803	-1,020	-0,597	**7Fr**	-0,398	2,452	-0,000	1,328
13A	0,537	-0,582	0,843	1,371	**3V**	-0,381	-1,229	-1,027	-1,361	**8Fr**	1,308	-0,176	0,318	0,085
14A	1,915	-0,465	2,002	2,315	**4V**	-1,534	0,088	-1,835	-1,229	**9Fr**	0,638	0,232	0,949	-0,098
15A	2,190	-1,097	-0,265	1,646	**5V**	-1,193	-1,217	-1,947	-1,163	**10Fr**	-0,185	-0,905	0,788	-0,532
1F	0,376	-1,349	-0,578	-0,202	**6V**	-0,762	-0,783	-0,949	-0,984	**11Fr**	-0,846	1,469	0,679	-0,404
2F	-0,258	-0,281	0,099	0,177	**7V**	-1,226	1,282	-0,947	-0,811	**12Fr**	-0,135	0,167	0,905	0,278
3F	-0,871	-0,438	0,151	-1,035	**8V**	-0,720	-1,067	-0,949	-0,954	**13Fr**	0,179	0,132	0,129	0,286
4F	-0,475	-0,180	-0,337	-0,597	**9V**	0,014	-0,845	-0,205	-1,004	**14Fr**	0,176	-0,683	0,129	-0,159
5F	-1,062	-1,061	-1,115	-1,229	**10V**	0,015	-0,845	-0,205	-1,004	**15Fr**	0,360	-0,340	0,617	0,329

Los valores propios obtenidos por el análisis de componentes principales y cuya gráfica se presenta en la **Figura 33**, muestran que más del 96% de la varianza total puede ser explicada por los tres primeros componentes principales.

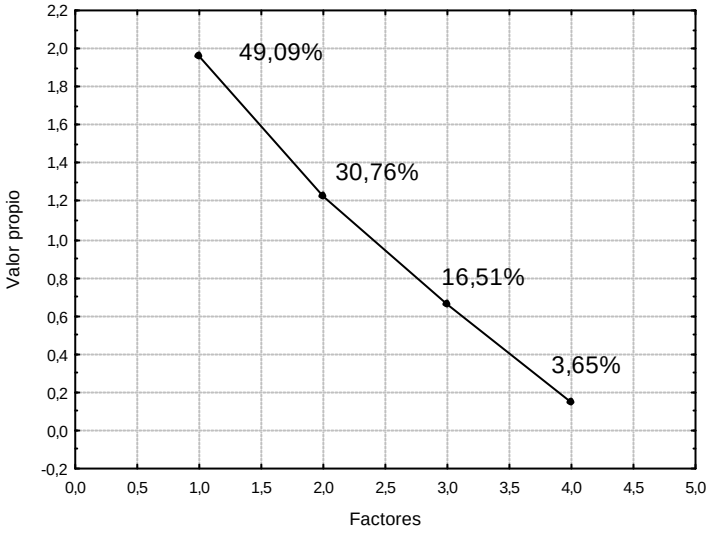


Figura 33. Gráfica de los valores propios *vs* PC (Descriptores topológicos).

La transformación de los 24 descriptores originales en tres nuevas combinaciones lineales representa una considerable reducción de variables, además, se debe tener en cuenta que la mayoría de la información original es retenida.

La **Figura 34** presenta la gráfica de los factores de aporte representados por los tres primeros factores principales. De acuerdo con ésta, el descriptor d26 [Índice de Kier&Hall (orden 3)] se encuentra altamente correlacionado con el Factor 1 (0,914). Igualmente, el descriptor d28 [Índice de forma de Kier (orden 2)] se encuentra correlacionado con el Factor 2 (-0,946). Por otro lado, el descriptor d32 [contenido de información estructural (orden 0)], presenta valores de correlación significativos con los tres primeros factores.

En el **Anexo 8** se reportan los valores de los factores de aporte de los descriptores topológicos obtenidos por medio de la reducción de variables utilizando el Método A.

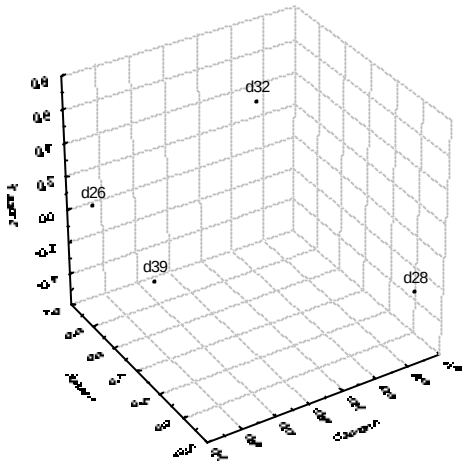


Figura 34. Gráfica en tres dimensiones de los factores de aporte (Descriptores topológicos).

La gráfica en tres dimensiones de los factores de coordenadas para las 60 moléculas muestra la clasificación de los compuestos pertenecientes a las familias Almizcle y Verde (**Figura 35**). Sin embargo, se observa un agrupamiento de algunos compuestos de la familia Frutal, agrupamiento que no se presentó en el PCA de los descriptores electrónicos.

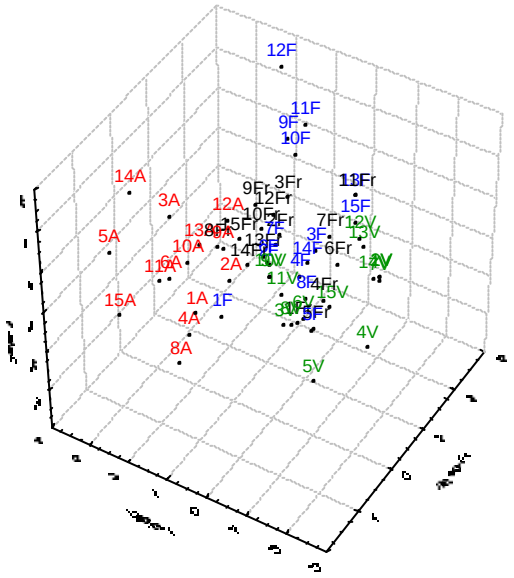


Figura 35. Gráfica de dispersión de los factores de coordenadas obtenidos por medio del PCA sobre los descriptores topológicos (Método A).

En el **Anexo 9** se presentan las gráficas de dispersión en dos dimensiones de los factores de coordenadas de los tres primeros componentes principales.

A partir de esta primera clasificación de los tipos de descriptores se pudo establecer la importancia de los descriptores electrónicos y topológicos, ya que éstos fueron los que arrojaron los mejores resultados en cuanto a la diferenciación de las familias; además, se observa, que los descriptores topológicos y topográficos condujeron a un mismo resultado de clasificación, por lo tanto se prescindió de uno de estos tipos de descriptores (los topográficos). Por último, se concluye, que los resultados menos favorables se obtuvieron a partir de los descriptores constitucionales.

5.2.3 Combinación de dos tipos de descriptores. Continuando con el estudio de los diferentes descriptores moleculares, se procedió a realizar la combinación de dos tipos de descriptores, antes de someterlos a la reducción de variables por los Métodos A y B.

La combinación de los descriptores electrónicos con los topológicos y geométricos, arrojó los mejores resultados. En estos resultados se aprecia la separación de tres familias odorantes, *e.g.* la familia Almizcle, la Verde y la Floral, principalmente, en las gráficas de los dos primeros factores principales.

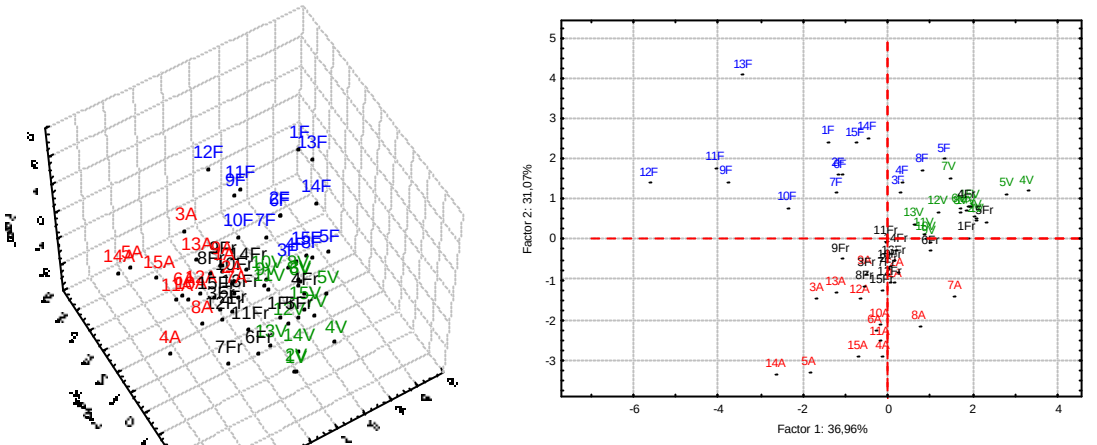
5.2.3.1 PCA de los descriptores electrónicos con descriptores topológicos. Por medio de la combinación de los descriptores electrónicos y topológicos se obtuvo un total de 30 descriptores. Una vez realizados los tratamientos previos de datos para reducir el número de variables, se obtuvieron para los Método A y B, 8 y 7 descriptores, respectivamente. En la **Tabla 19** se presentan los resultados del PCA para cada uno de los métodos. Además, se observa que los tres componentes principales explican el 82% de la varianza total para el Método A y el 83% para el Método B.

Tabla 19. Resultados del PCA sobre la combinación de los descriptores electrónicos y topológicos.

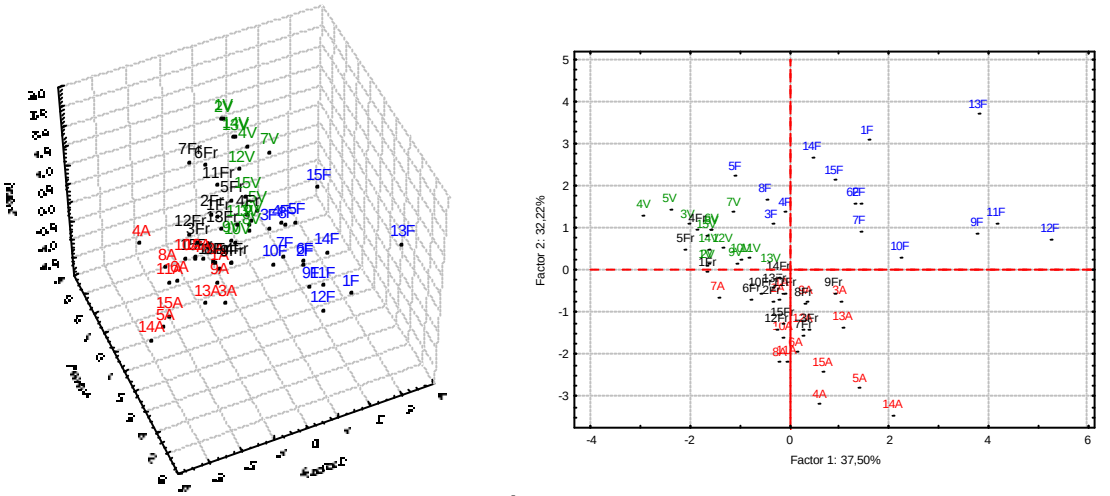
Método A				Método B			
Factor	**Valor propio**	**% Total de varianza**	**% de Contribución**	**Factor**	**Valor propio**	**% Total de varianza**	**% de Contribución**
1	2,956952	36,96190	36,9619	**1**	2,624783	37,49690	37,4969
2	2,485404	31,06755	68,0295	**2**	2,255199	32,21713	69,7140
3	1,110457	13,88071	81,9102	**3**	0,961590	13,73699	83,4510
4	0,691794	8,64743	90,5576	**4**	0,692223	9,88890	93,3399
5	0,537364	6,71705	97,2746	**5**	0,376907	5,38439	98,7243
6	0,138735	1,73419	99,0088	**6**	0,055141	0,78772	99,5120
7	0,054706	0,68383	99,6927	**7**	0,034157	0,48796	100,0000
8	0,024587	0,30734	100,0000				

En la **Figura 36** aparecen las gráficas en tres dimensiones de los dos métodos utilizados, así como la gráfica de los dos primeros componentes principales los cuales explican el 68 y 69% de la varianza total, respectivamente.

En el **Anexo 10** se reportan las tablas de los factores de aporte de cada variable. El Factor 1 se encuentra relacionado con la energía HF y con el contenido de información estructural (orden 0), el Factor 2 con el índice de Kier y Hall (orden 3), la energía de LUMO y el contenido de información complementaria (orden 2), y el Factor 3 con el índice de forma de Kier (orden 2) en cuanto al Método A se refiere. Para el Método B, el Factor 1 se encuentra relacionado con la energía HF e índice de Kier y Hall (orden 0), el Factor 2 con la energía de LUMO, el contenido de información complementaria (orden 2) y el Factor 3, con el índice de forma de Kier (orden 2).



Método A



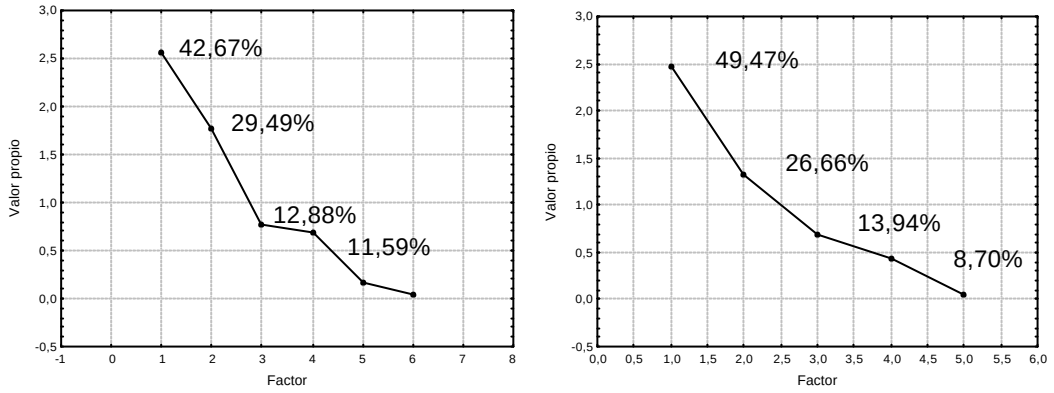
Método B

Figura 36. Gráficas de dispersión de los valores de los PC obtenidas por el PCA sobre la combinación de los descriptores electrónicos y topológicos.

5.2.3.2 PCA de los descriptores electrónicos con los geométricos. A partir de la combinación de los descriptores electrónicos con los geométricos se obtienen un total de 11 descriptores. La reducción de variables utilizando los Métodos A y B arroja una matriz de 60 x 6 para cada método.

Las gráficas de los valores propios *vs* factores muestran la diferencia en los resultados obtenidos para cada método. Para el Método A, los tres primeros factores principales

explican el 85%, mientras que los tres primeros factores principales de los resultados por el Método B explican el 90% de la varianza total.

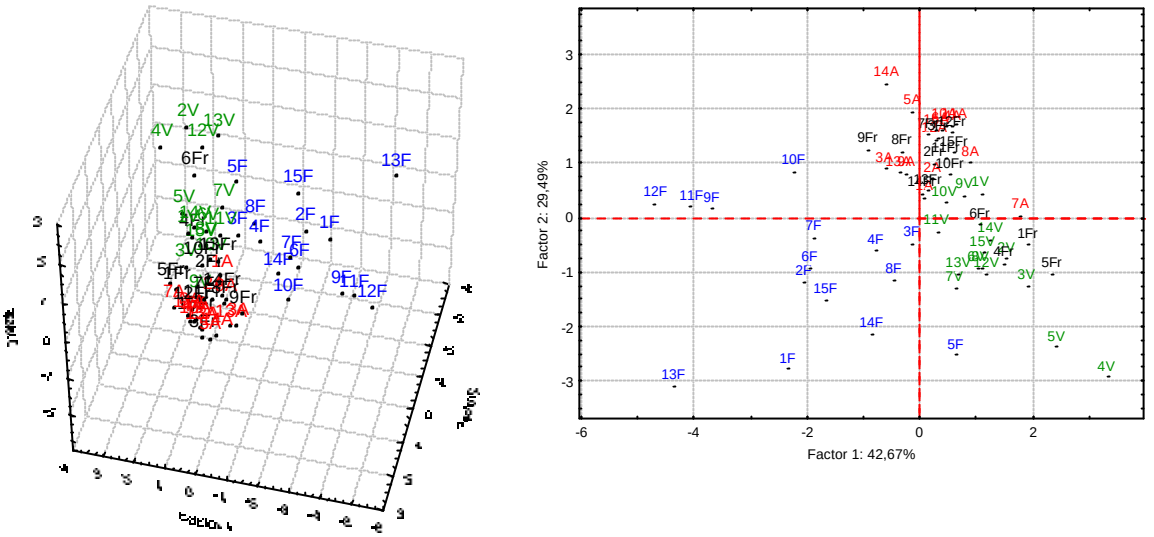


Método A Método B

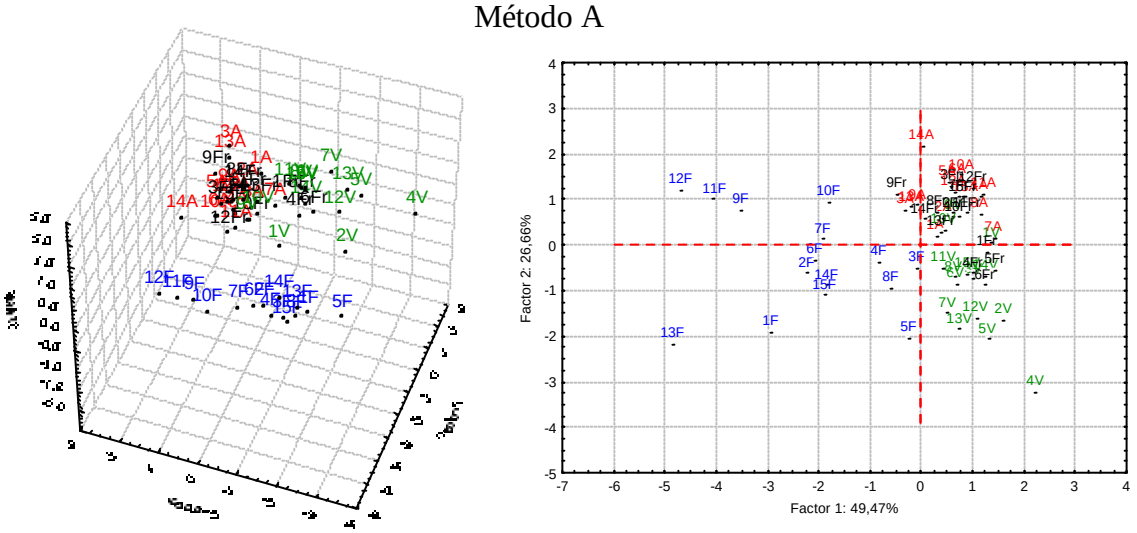
Figura 37. Gráficas de los valores propios *vs* factores principales (Descriptores electrónicos y geométricos).

Los resultados de los factores de aporte obtenidos a partir del Método A, muestran que el Factor 1 se encuentra relacionado con las energías del HOMO y del LUMO, el Factor 2 con el momento de inercia A y el volumen molecular. Para el Método B, el Factor 1 se encuentra relacionado con las energías del HOMO y del LUMO, el Factor 2 con el momento de inercia A y el Factor 3 con el momento dipolar. En el **Anexo 11** se pueden observar los resultados de los dos PCA.

En las gráficas de tres dimensiones de los factores de coordenadas de las moléculas se puede evidenciar la separación de las familias Almizcle, Floral y Verde (**Figura 38**). Sin embargo, en la gráfica obtenida por medio del Método B se observa una marcada diferencia en la separación de las moléculas del grupo Floral. Esta separación adicional se debe a la contribución del momento dipolar en el Factor 3, contribución que no se realiza por el Método A.



Método A



Método B

Figura 38. Gráficas de dispersión de los valores de los PC, obtenidos por el PCA sobre la combinación de los descriptores electrónicos y geométricos.

La clasificación de los grupos fragantes de acuerdo con la combinación de los descriptores electrónicos con los CPSA, presenta un ligero solapamiento entre las familias Almizcle y Verde, mientras que los resultados de la combinación de los descriptores electrónicos con los constitucionales no presenta clasificación alguna, salvo para algunos compuestos del grupo Almizcle (**Anexo 12)**.

5.2.4 Combinación de tres y cuatro tipos de descriptores. Con base en los resultados anteriores se procedió a realizar la combinación de tres y cuatro tipo de descriptores. La reducción de datos al igual que en los tratamientos anteriores se realizó por medio de los Métodos A y B.

El PCA realizado sobre la matriz obtenida por la combinación de tres tipos de descriptores, principalmente la combinación de los descriptores electrónicos, topológicos y geométricos arrojaron buenos resultados. Sin embargo, como se podía esperar de acuerdo con los resultados obtenidos en los PCA de dos tipos de descriptores, cuando se involucran los descriptores CPSA las moléculas de los grupos Almizcle, Frutal y Verde presentan un solapamiento considerable. Por medio de la combinación de los descriptores electrónicos, topológicos y geométricos se obtuvo una matriz de 60 x 35. Una vez aplicados los tratamientos previos indicados en los Métodos A y B a los 35 descriptores originales, se produjo una considerable reducción de variables que produjo dos matrices de 60 x 9 y 60 x 8 respectivamente, para cada método.

El posterior análisis de componentes principales para cada una de las matrices obtenidas a partir de la reducción de variables, condujo a los resultados mostrados en la **Tabla 20**. Los tres primeros PC contienen el 78 y el 80% de la varianza total, así como valores propios mayores que uno, lo que conduce a una solución de tres componentes principales.

Tabla 20. Resultados de los PCA de la combinación de tres tipos de descriptores.

	Método A				Método B		
Factor	**Valor propio**	**% Total de varianza**	**% de Contribución**	**Factor**	**Valor propio**	**% Total de varianza**	**% de Contribución**
1	2,937171	32,63523	32,6352	**1**	2,650891	33,13614	33,1361
2	2,647470	29,41634	62,0516	**2**	2,391235	29,89044	63,0266
3	1,479090	16,43434	78,4859	**3**	1,401323	17,51654	80,5431
4	0,715670	7,95189	86,4378	**4**	0,701156	8,76445	89,3076
5	0,532623	5,91803	92,3558	**5**	0,412562	5,15702	94,4646
6	0,350865	3,89849	96,2543	**6**	0,280373	3,50466	97,9693
7	0,183030	2,03367	98,2880	**7**	0,124763	1,55954	99,5288
8	0,118209	1,31343	99,6014	**8**	0,037697	0,47121	100,0000
9	0,035872	0,39858	100,0000				

En la **Tabla 21** se presentan los valores de los factores de aporte de las variables para los tres primeros componentes principales. Estos resultados muestran que para el Método A las variables que aportan mayor información a cada uno de los factores son: Factor 1, d32 [contenido de información estructural (orden 0)] y d67 (momento de inercia B), Factor 2, d26 [índice de Kier y Hall (orden 3)] y d92 (energía del LUMO) y para el Factor 3, d28 (índice de forma de Kier (orden 2)). Por otro lado, a partir del Método B las variables con mayor contribución son: Factor 1, d23 [índice de Kier y Hall (orden 0)] y d67 (momento de inercia B), Factor 2 d39 [contenido de información complementaria (orden 2)], d91 (energía del HOMO) y d92 (energía del LUMO) y Factor 3, d28 [índice de forma de Kier (orden 2)].

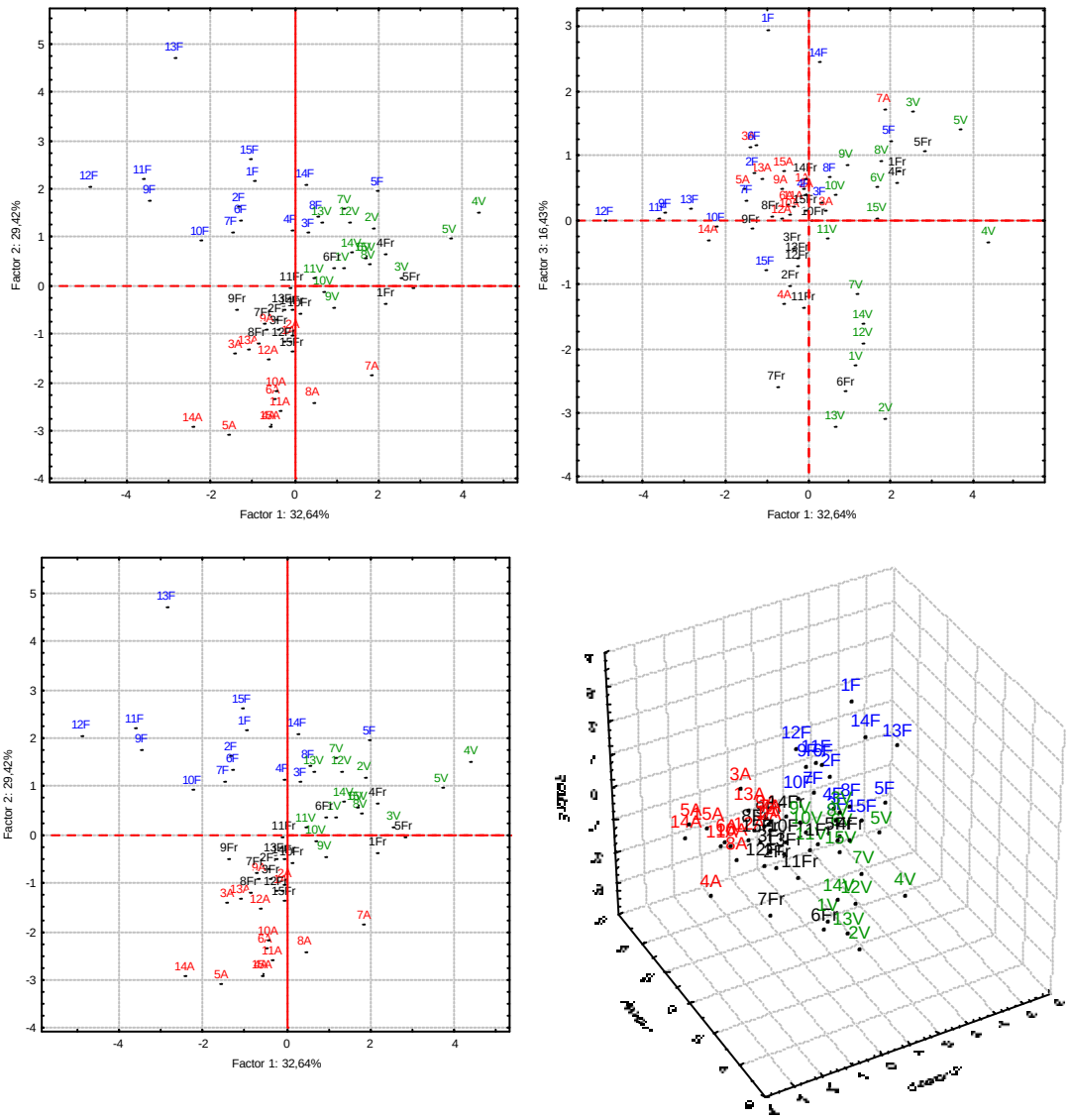
Tabla 21. Valores de los factores de aporte (Descriptores electrónicos, topológicos y geométricos).

	Método A				Método B		
	Factor 1	**Factor 2**	**Factor 3**		**Factor 1**	**Factor 2**	**Factor 3**
d26	-0,525010	-0,726414*	0,229203	**d23**	-0,826018*	0,472044	-0,117847
d28	-0,279135	0,344140	-0,842250*	**d28**	-0,358924	-0,187917	0,849414*
d32	-0,884234*	-0,040226	-0,043580	**d39**	-0,496088	0,702673*	0,035771
d39	-0,457273	-0,684790	-0,176691	**d66**	0,559624	-0,502034	0,527049
d66	0,603891	0,529663	-0,367091	**d67**	0,767978*	-0,024338	-0,493454
d67	0,742578*	-0,059134	0,538334	**d90**	-0,361156	-0,543718	-0,213549
d90	-0,381598	0,482039	0,297399	**d91**	-0,589007	-0,724914*	-0,219689
d91	-0,574607	0,695359	0,289483	**d92**	0,461869	0,752056*	0,222503
d92	0,448058	-0,728615*	-0,295375				

En las gráficas de dispersión de los factores de coordenadas se observa principalmente la separación de tres familias odorantes: Almizcle, Floral y Verde (**Figuras 39 y 40**). Igualmente, se aprecia en las gráficas de los Factores 3 y 2, así como en las gráficas en tres dimensiones, un agrupamiento de los compuestos de la familia Frutal en una zona entre los grupos Almizcle y Verde.

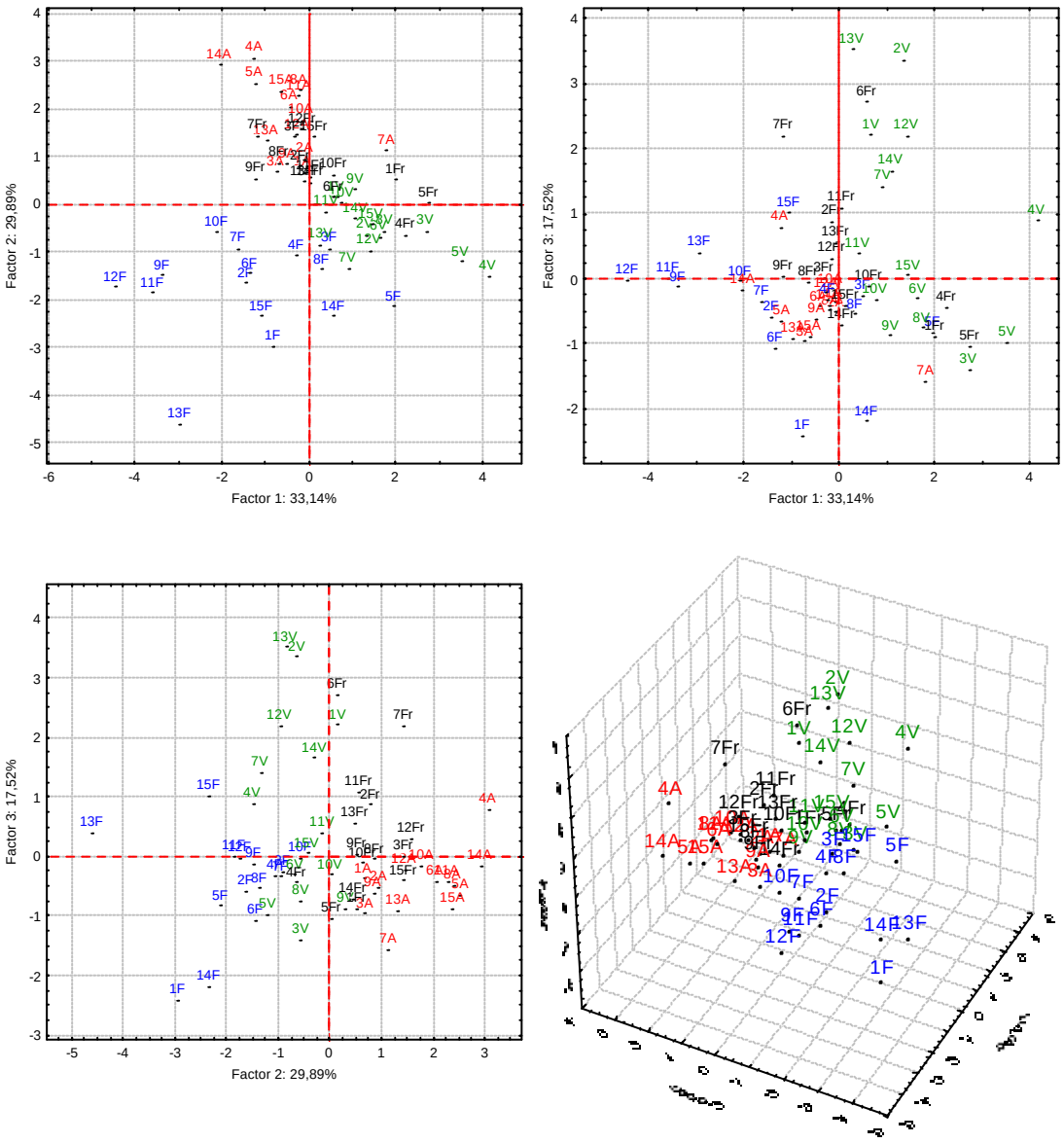
El primer componente principal separa en ambos casos los grupos Almizcle y Verde, exceptuando los compuestos 7A y 8A en el Método A y solamente al compuesto 7A en

el Método B. El segundo factor separa especialmente al grupo Almizcle del grupo Floral, mientras que el tercer factor no contribuye de manera apreciable a la separación de las familias fragantes.



Método A

Figura 39. Gráficas de dispersión de los factores de coordenadas (Descriptores electrónicos, topológicos y geométricos).



Método B

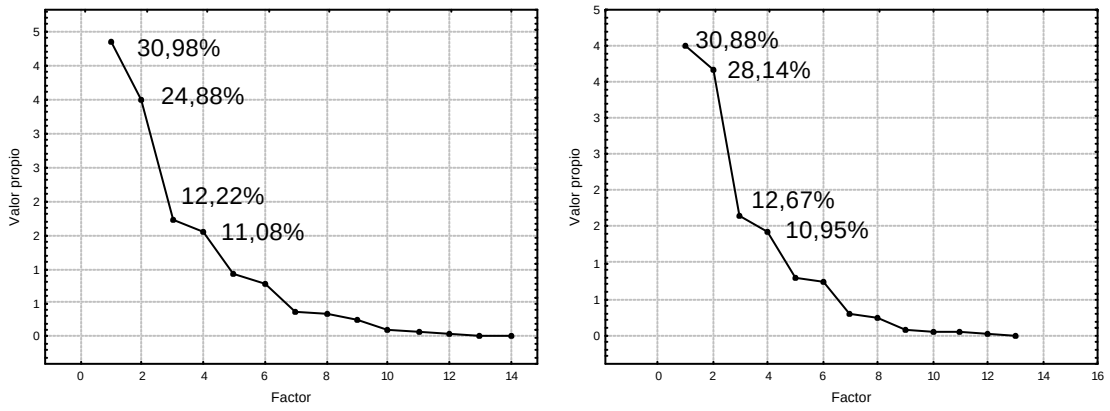
Figura 40. Gráficas de dispersión de los factores de coordenadas (Descriptores electrónicos, topológicos y geométricos).

Una vez establecidos los mejores resultados por medio de la clasificación de las moléculas fragantes, utilizando la combinación de tres tipos de descriptores, se procedió a estudiar el efecto de la inclusión de un cuarto tipo de descriptores. De esta manera, se agregaron los descriptores CPSA y los constitucionales individualmente a la combinación de los descriptores electrónicos, topológicos y geometricos.

Los resultados obtenidos por medio del PCA, después de reducir el número de variables, fueron favorables cuando se adicionaron los descriptores CPSA. Sin embargo, los resultados fueron muy similares a los obtenidos por la combinación de tres tipos de descriptores, a pesar de presentar un mayor acercamiento de las familias Verde y Almizcle. Por otro lado, la inclusión de los descriptores constitucionales no contribuye de manera positiva a la clasificación de los compuestos odorantes.

Por medio de la combinación de las cuatro clases de descriptores mencionados anteriormente (electrónicos, topológicos, geométricos y CPSA) se obtuvo un total de 51 descriptores. Después de reducir el número de variables de acuerdo con los Métodos A y B, se obtuvieron un total de 14 y 13 descriptores, respectivamente.

Los resultados obtenidos por medio del PCA (**Anexo 13)** muestran que se obtuvieron dos soluciones de cuatro componentes principales para cada uno de los métodos. En la **Figura 41** se presentan estos resultados por medio de las gráficas de los valores propios *vs* factores principales; además, se aprecia una disminución en la varianza total explicada por los tres primeros PC respecto a los resultados obtenidos en el análisis anterior.



Método A Método B

Figura 41. Gráficas de los valores propios *vs* componentes principales (Descriptores electrónicos, topológicos, geométricos y CPSA).

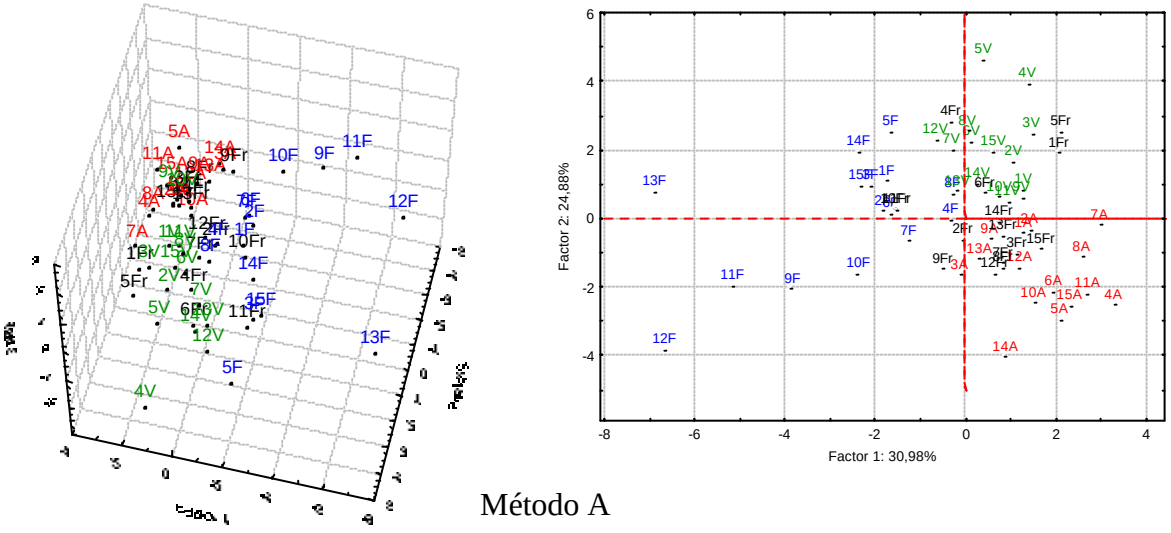
En el **Anexo 14** se presentan las gráficas de los factores de aporte representados por los tres primeros factores. De acuerdo con estas gráficas, así como con los valores de los factores de aporte presentados en **Tabla 22**, se establecen los descriptores que se encuentran mayormente correlacionados con cada uno de los tres primeros factores principales, a saber: Factor 1, d76 (PNSA-2) y d91 (energía HOMO); Factor 2, d26 [índice de Kier y Hall (orden3)] y d66 (Momento de inercia A); Factor 3, d83 (FPSA-3), Factor 4, d28 [índice de forma de Kier (orden 2)]. Por medio del Método B se obtuvieron los siguientes resultados: Factor 1, d91 (energía del HOMO) y d81 (FPSA-1); Factor 2, d70 (volumen molecular); Factor 3, d83 (FPSA-3) y Factor 4, d28 [índice de forma de Kier (orden 2)].

Tabla 22. Factores de aporte de las variables de los descriptores electrónicos, topológicos, geométricos y CPSA.

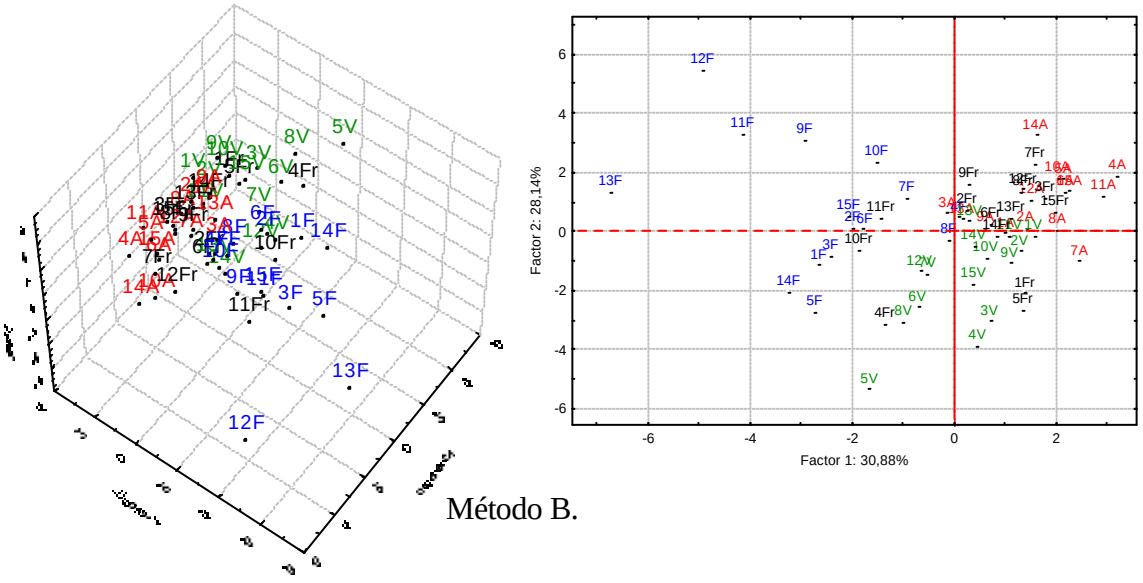
	Método A					Método B			
	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**		**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**
d26	0,277901	-0,818023*	0,179529	0,260681	**d28**	-0,180190	0,436171	-0,122464	0,796798*
d28	-0,353350	-0,126863	-0,474134	-0,736719*	**d66**	-0,179025	-0,576672	-0,233275	0,671643
d37	-0,454254	-0,677307	0,315193	-0,030269	**d67**	0,171755	-0,778715*	-0,153435	-0,409939
d39	0,345988	-0,763459*	-0,186529	-0,021299	**d70**	0,018296	0,935651*	0,140630	-0,195597
d66	-0,053597	0,739256*	-0,377491	-0,308432	**d74**	-0,227522	0,722873*	-0,621139	-0,087827
d67	0,411562	0,598055	0,091248	0,560564	**d76**	0,773737*	-0,469385	-0,026605	0,088978
d74	-0,408970	-0,635584	-0,589187	0,205326	**d78**	0,644646	0,690745	-0,110694	0,101534
d76	0,870672*	0,278101	-0,047294	-0,046238	**d81**	0,792110*	0,368839	-0,208712	0,047627
d77	0,757018*	-0,329808	0,276117	-0,104340	**d83**	-0,128932	-0,096327	-0,947360*	-0,247428
d83	-0,065627	0,063094	-0,743065*	0,642021	**d86**	0,740400*	0,402397	0,302426	-0,111440
d84	-0,656661	0,497510	0,292758	0,102613	**d90**	-0,625135	0,073036	0,313779	-0,004934
d90	-0,603755	0,054214	0,294705	0,020442	**d91**	-0,817449*	0,299556	0,044466	-0,118487
d91	-0,846798*	-0,080841	0,077795	0,160313	**d92**	0,776452*	-0,177264	-0,048506	0,124237
d92	0,787816*	-0,031506	-0,113001	-0,159369					

La representación gráfica en tres dimensiones, así como la gráfica de los dos primeros componentes principales de los factores de coordenadas de las moléculas para cada uno de los métodos utilizados se presentan en la **Figura 42**. En estas gráficas se evidencia la clasificación de tres de los grupos odorantes estudiados, las familias Almizcle, Verde y Floral. Principalmente cuando se representan por medio de los dos primeros PC. Sin embargo, se aprecia un pequeño solapamiento de los grupos Verde y Almizcle en los resultados obtenidos a partir del Método B.

De acuerdo con las gráficas de dispersión obtenidas para el primer y segundo método, el Factor 1 separa los compuestos de la familia Floral (exceptuando los compuestos 8F y 4F) de las familias Verde y Almizcle. El Factor 2 separa la familia Almizcle de la Verde, mientras que el Factor 3 no muestra alguna contribución en la separación de los grupos odorantes. Además, es importante apreciar, que la separación realizada por el Factor 2 es más efectiva, usando el Método A, que para el Método B; ya que no se presenta el solapamiento de algunos compuestos del grupo Almizcle con los compuestos del grupo Verde, así como aparece en las gráficas.



Método A



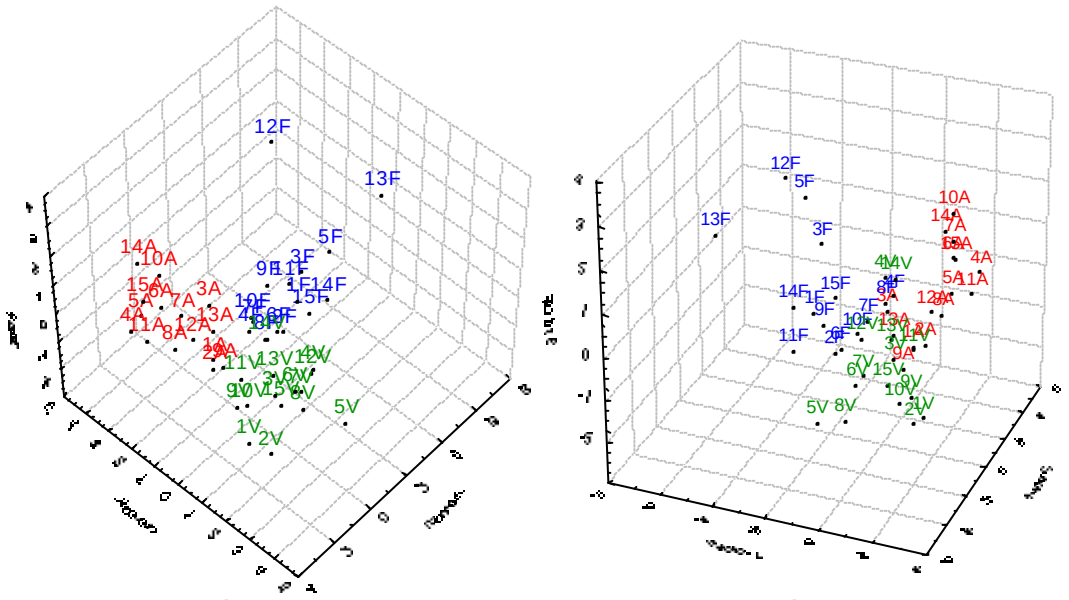
Método B.

Figura 42. Graficas de dispersión de los tres primeros PC obtenidas por el PCA sobre los descriptores electrónicos, topológicos, geométricos y CPSA.

En resumen, a partir de los diferentes subespacios obtenidos por medio de la combinación de los diferentes tipos de descriptores, se logró establecer que los descriptores constitucionales no fueran variables adecuadas para realizar la clasificación de los compuestos fragantes, utilizando el método de componentes principales.

Por otro lado, a partir de la combinación de los descriptores electrónicos, topológicos, geométricos y CPSA, junto con la combinación de los descriptores electrónicos, topológicos y geométricos, se puede realizar la clasificación de las 60 moléculas bajo estudio en tres familias odorantes, *i.e.* Almizcle, Verde y Floral.

Este resultado, puede ser observado más claramente si en las matrices resultantes de la combinación de los descriptores mencionados anteriormente, se eliminan los compuestos pertenecientes al grupo Frutal y, posteriormente, se realiza el análisis de componentes principales (**Figura 43**). Cabe destacar, que los resultados de la clasificación obtenidos por medio de la reducción de variables utilizando el Método A, son mejores en comparación con el Método B.



Método A Método B

Figura 43. Gráficas de dispersión de los tres primeros PC eliminado los compuestos de la familia Frutal (Descriptores electrónicos, topológicos, geométricos y CPSA).

5.2.5 Análisis de agrupamiento sobre los descriptores estructurales y electrónicos. El análisis de agrupamiento se inició con el estudio de todos los descriptores obtenidos, omitiendo los modos vibracionales. Sin embargo, los resultados obtenidos en la clasificación de las familias fragantes, no fueron adecuados, ya que se presentaron demasiados compuestos de diferentes familias agrupados en grupos similares (**Figura 44**).

Con base en lo anterior, se decidió llevar a cabo el análisis de agrupamiento utilizando los diferentes conjuntos de descriptores obtenidos por medio de la reducción de datos descritos en el análisis de componentes principales realizado anteriormente, con el objetivo de mejorar la clasificación de los compuestos fragantes. De esta manera, se inició el estudio a través de la clasificación de cada una de las clases de descriptores.

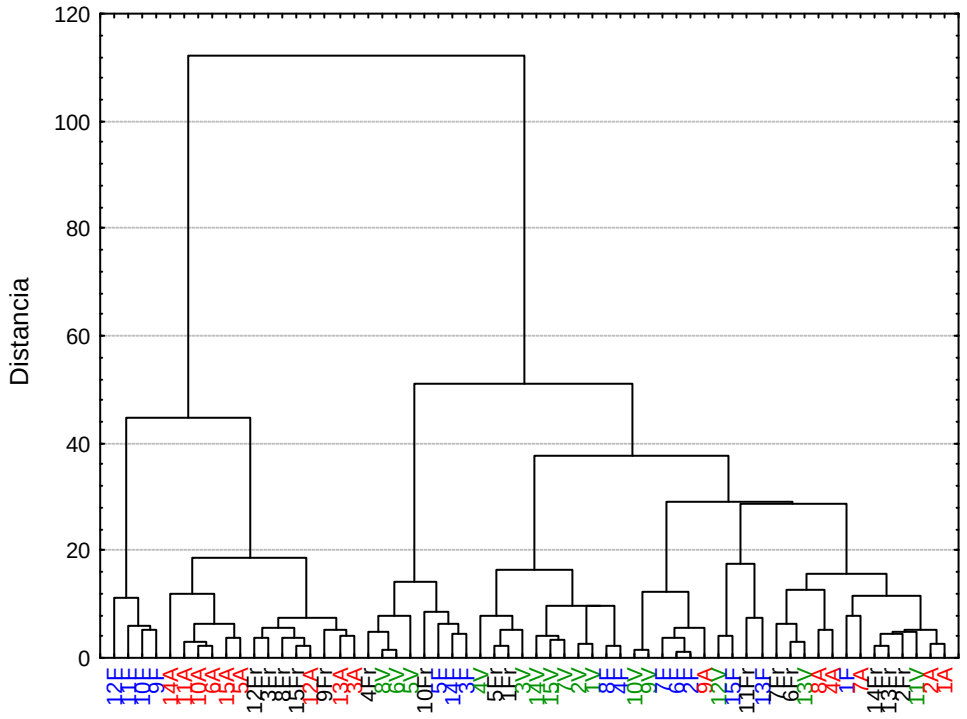


Figura 44. Dendograma obtenido por medio del agrupamiento de los descriptores estructurales y electrónicos.

5.2.6 Análisis de agrupamiento sobre los diferentes tipos de descriptores. De acuerdo con los dendogramas obtenidos para cada tipo de descriptores (**Anexo 15**), los métodos de agrupamiento realizados sobre los descriptores electrónicos proporcionaron la mejor diferenciación de las familias fragantes, ya que se observa la separación de tres familias odorantes, especialmente, la familia Floral, que constituye un solo grupo. Por otro lado, la mayoría de las moléculas pertenecientes a los grupos Almizcle y Verde, conforman dos grupos individuales, junto con algunos compuestos del grupo Frutal (**Figura 45**). En el **Anexo 16,** se presenta la matriz de las distancias de Euclidean.

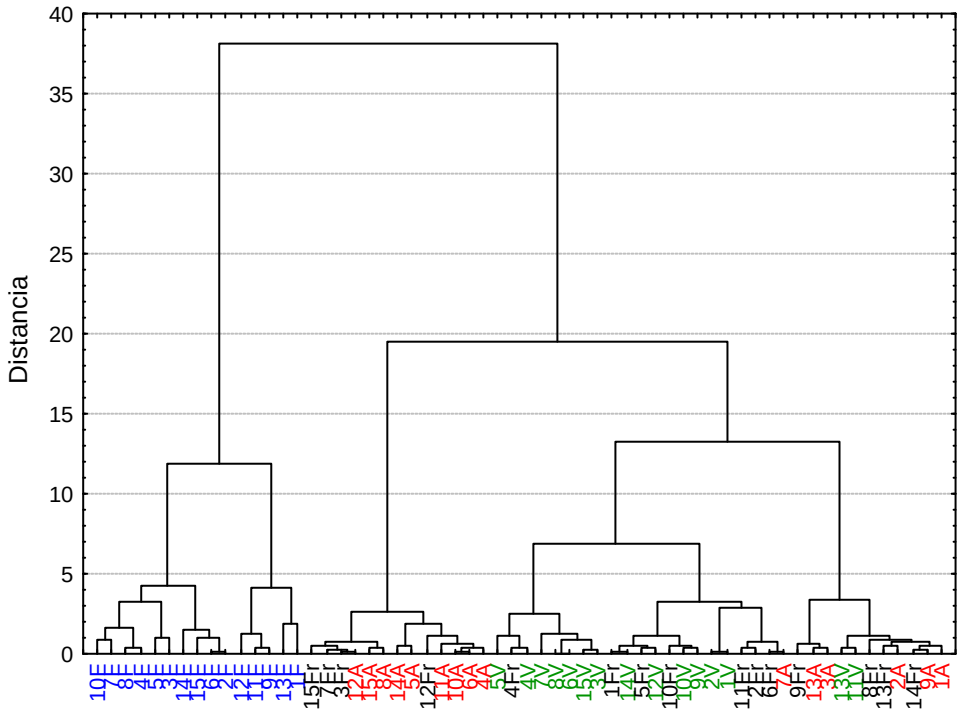


Figura 45. Dendograma obtenido por medio del agrupamiento de los descriptores electrónicos.

Los resultados obtenidos para determinar el número de grupos por medio de la metodología Kelley y el análisis manual se resumen en la **Tabla 23**.

Tabla 23. Tabla de contingencia del dendograma obtenido a partir de los descriptores electrónicos.

Método A														
	Grupos													
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**	**C_{10}**	**C_{11}**	**C_{12}**	**C_{13}**	**C_{14}**
Floral	4	2	4	3	0	0	0	0	0	0	0	0	0	0
Verde	0	0	0	0	0	0	0	2	5	4	2	0	0	2
Frutal	0	0	0	0	3	0	1	1	0	3	0	3	1	3
Almizcle	0	0	0	0	3	2	4	0	0	0	0	1	2	3

Grupos establecidos por medio del índice de Kelley

	Grupos			
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**
Floral	15	0	0	0
Verde	0	0	13	2
Frutal	0	4	7	4
Almizcle	0	9	1	5

Número de grupos establecido manualmente

Las moléculas pertenecientes a cada uno de los diferentes grupos se reportan en la **Tabla 24**. De acuerdo con estos resultados, se observa que el grupo C_1 obtenido manualmente, está compuesto por moléculas pertenecientes a la familia Floral. Mientras que los grupos C_2 y C_3 están conformados en su mayoría por compuestos pertenecientes a las familias odorantes Almizcle y Verde, respectivamente.

Tabla 24. Moléculas pertenecientes a cada uno de los grupos determinados.

Grupos	Moléculas	Grupos	Moléculas
C_1	4F, 8F, 7F, 10F	**C_8**	4V, 5V, 4Fr
C_2	3F, 5F	**C_9**	3V, 6-8V
C_3	2F, 6F, 14F, 15F	**C_{10}**	9V, 10V, 12V, 14V, 1Fr, 5Fr, 10Fr
C_4	9F, 11F, 12F	**C_{11}**	1V, 2V
C_5	8A, 12A, 15A, 3Fr, 7Fr, 15Fr	**C_{12}**	2Fr, 6Fr, 11Fr, 7A
C_6	5A, 14A	**C_{13}**	3A, 13A, 9Fr
C_7	4A, 6A, 10A, 11A, 12Fr	**C_{14}**	11V, 13V, 1A, 2A, 9A, 8Fr, 13Fr, 14Fr
Unitarios	13F, 1F		

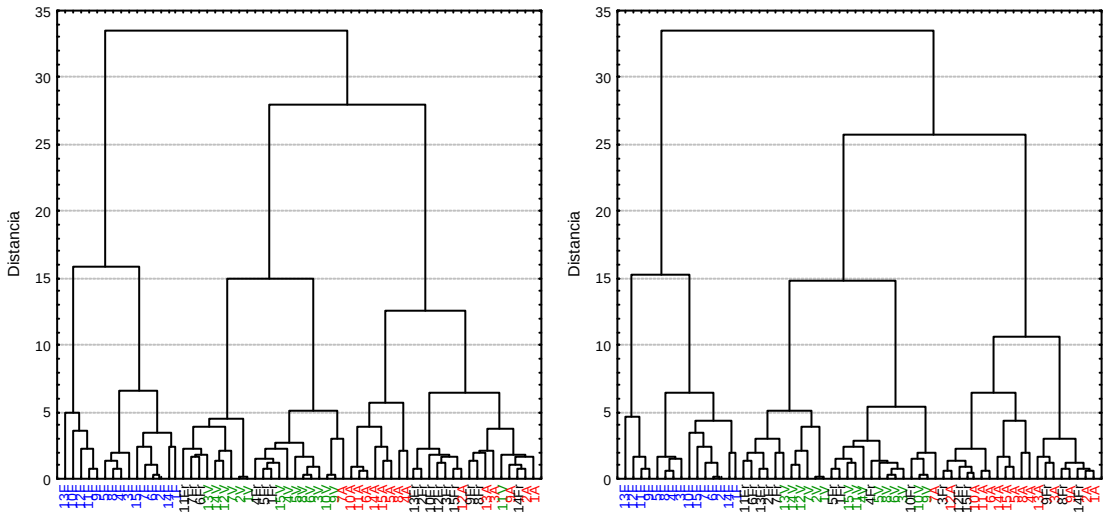
Grupos establecidos por medio del índice de Kelley

Grupos	Moléculas
C_1	1-15F
C_2	4-6A, 8A, 10-12A, 14A, 15A, 3Fr, 7Fr, 12Fr, 15Fr
C_3	1-10V, 12V, 14V, 15V, 1Fr, 2Fr, 4-6Fr, 10V, 11V, 7A
C_4	1-3A, 9A, 13A, 11V, 13V, 8Fr,9Fr, 13Fr, 14Fr

Grupos establecidos manualmente

5.2.7 Combinación de dos tipos de descriptores. Continuando con el estudio de las diferentes clases de descriptores, se procedió a realizar el análisis de agrupamiento sobre la combinación de dos tipos de descriptores. Al igual, como se presentó en el PCA realizado en la sección anterior, los resultados más promisorios, se presentaron cuando se combinaron los descriptores electrónicos con las otras clases, exceptuando la combinación de los descriptores electrónicos con los descriptores constitucionales.

En el **Anexo 17** se presentan los dendogramas obtenidos para cada combinación de descriptores reducidos por medio de los Métodos A y B. Por medio de la observación de los resultados, se aprecia que la mejor clasificación se obtiene a partir de la combinación de los descriptores electrónicos y topológicos (**Figura 46**).



(A) **(B)**

Figura 46. Dendogramas obtenidos por la combinación de los descriptores electrónicos y topológicos (Métodos A y B).

La **Tabla 25** presenta los resultados de la determinación del número de grupos de los dendogramas obtenidos por la combinación de los descriptores electrónicos y topológicos (Métodos A y B).

Tabla 25. Tabla de contingencia del dendograma obtenido a partir de los descriptores electrónicos y topológicos (Métodos A y B).

Método A												
	Grupos											
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**	**C_{10}**	**C_{11}**	**C_{12}**
Floral	4	4	6	0	0	0	0	0	0	0	0	0
Verde	0	0	0	4	2	6	2	0	0	0	0	1
Frutal	0	0	0	3	0	3	0	0	0	0	6	3
Almizcle	0	0	0	0	0	0	1	3	3	2	1	5

Método B													
	Grupos												
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**	**C_{10}**	**C_{11}**	**C_{12}**	**C_{13}**
Floral	3	4	4	2	0	0	0	0	0	0	0	0	0
Verde	0	0	0	0	1	3	2	7	2	0	0	0	0
Frutal	0	0	0	0	5	0	0	3	1	3	0	0	3
Almizcle	0	0	0	0	0	0	0	0	1	4	3	2	5

Grupos establecidos por medio del índice de Kelley

Familia odorante	**Grupos**								
	Método A					**Método B**			
	C_1	**C_2**	**C_3**	**C_4**	**C_5**	**C_1**	**C_2**	**C_3**	**C_4**
Floral	15	0	0	0	0	15	0	0	0
Verde	0	14	0	0	1	0	6	9	0
Frutal	0	6	0	6	3	0	5	4	6
Almizcle	0	1	8	1	5	0	0	1	14

Grupos establecidos manualmente

A partir de la selección de grupos establecidos manualmente, se observa la clasificación de las moléculas de la familia Floral en un solo grupo para ambos métodos. Por otro lado, los grupos C_3 y C_4, obtenidos por el Método A, están conformados por 8 compuestos con nota olfativa Almizcle y 6 compuestos de la familia Frutal, respectivamente, mientras el grupo C_4, obtenido a partir del Método B presenta 14 compuestos de la misma familia pero incluye 6 compuestos de la familia Frutal. En la **Tabla 26** se reportan las moléculas pertenecientes a cada uno de los diferentes grupos.

Tabla 26. Moléculas pertenecientes a cada uno de los grupos determinados (Métodos A y B).

Método A		Método B	
Grupos	**Moléculas**	**Grupos**	**Moléculas**
C_1	9-12F	C_1	9F, 11F,12F
C_2	3-5F, 8F	C_2	3-5F, 8F
C_3	1F, 3F, 6F, 7F, 14F, 15F	C_3	2F, 6F, 7F, 15F
C_4	7V, 12-14V, 6Fr, 7Fr, 11Fr	C_4	1F, 14F
C_5	1V, 2V	C_5	13V, 2Fr, 6Fr, 7Fr, 11Fr, 13Fr
C_6	3-6V, 8V, 15V, 1Fr, 4Fr, 5Fr	C_6	7V, 12V, 14V
C_7	9V, 10V, 7A	C_7	1V, 2V
C_8	6A, 10A, 11A	C_8	3-6V, 8V, 11V, 15V, 1Fr, 4Fr, 5Fr
C_9	5A, 14A, 15A	C_9	9V, 10V, 7A, 10Fr
C_{10}	4A, 8A	C_{10}	6A, 10-12A, 3Fr, 12Fr, 15Fr
C_{11}	12A, 2Fr, 3Fr, 10Fr, 12Fr, 13Fr, 15Fr	C_{11}	5A, 14A, 15A
C_{12}	8Fr, 9Fr, 14Fr, 1-3A, 9A, 13A, 11V	C_{12}	4A, 8A
Unitarios	13F	C_{13}	1-3A, 9A, 13A, 8Fr, 9Fr, 14Fr
		Unitarios	13F, 10F

Grupos establecidos por medio del índice de Kelley

Método A		Método B	
Grupos	**Moléculas**	**Grupos**	**Moléculas**
C_1	1-15F	C_1	1-15F
C_2	1-10V, 12-15V, 1Fr, 4-7Fr, 11Fr, 7A	C_2	1V, 2V, 7V, 12-14V, 2Fr, 6Fr, 7Fr, 11Fr, 13Fr
C_3	4-8A, 10A, 11A, 14A, 15A	C_3	3-6V, 8-11V, 15V, 1Fr, 4Fr, 5Fr, 10Fr, 7A
C_4	2Fr, 3Fr, 10-13Fr, 15Fr, 12A	C_4	1-6A, 8-15A, 3Fr, 8Fr, 9Fr, 12Fr, 14Fr, 15Fr
C_5	1-3A, 9A, 13A, 8Fr, 9Fr, 14Fr, 11V		

Grupos establecidos por medio del índice de Kelley

De forma general, se evidencia, que a partir de los dendogramas anteriores se obtiene la diferenciación de las familias Floral, Verde y Almizcle, mientras que los compuestos de la familia Frutal se encuentran dispersos entre los grupos Verde y Almizcle.

5.2.8 Combinación de tres y cuatro tipos de descriptores. Para terminar con el estudio del agrupamiento a partir de los diferentes descriptores, se realizó el agrupamiento por medio del análisis de la combinación de tres y cuatro tipos de descriptores.

En el caso de la combinación de tres clases de descriptores, los mejores resultados se obtuvieron por la combinación de los descriptores electrónicos, topológicos y geométricos (**Figura 47**).

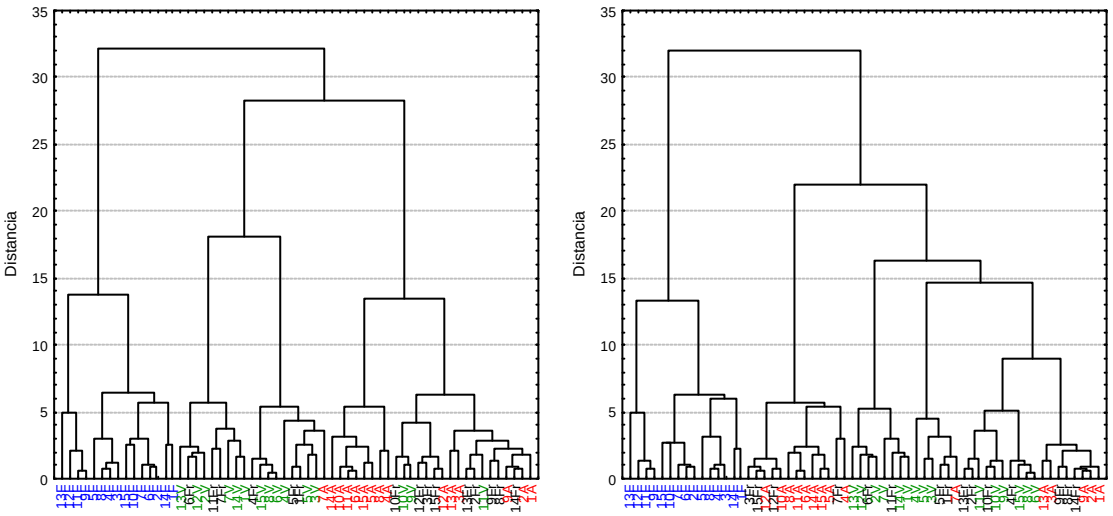


Figura 47. Dendogramas obtenidos por la combinación de dos los descriptores electrónicos, topológicos y geométricos (Métodos A y B).

La **Tabla 27** presenta los resultados de la determinación del número de grupos en los dendogramas obtenidos por la combinación de los descriptores electrónicos, topológicos y geométricos (Métodos A y B).

Tabla 27. Tabla de contingencia del dendograma obtenido a partir de los descriptores electrónicos, topológicos y geométricos (Métodos A y B).

Método A													
	Grupos												
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**	**C_{10}**	**C_{11}**	**C_{12}**	**C_{13}**
Floral	3	4	5	2	0	0	0	0	0	0	0	0	0
Verde	0	0	0	0	3	3	3	2	0	0	2	0	1
Frutal	0	0	0	0	1	2	1	2	0	0	1	3	5
Almizcle	0	0	0	0	0	0	0	1	6	2	0	1	5
Método B													
	Grupos												
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**	**C_{10}**	**C_{11}**	**C_{12}**	**C_{13}**
Floral	3	5	4	2	0	0	0	0	0	0	0	0	0
Verde	0	0	0	0	0	0	0	3	3	3	3	3	0
Frutal	0	0	0	0	3	0	1	1	1	2	3	1	3
Almizcle	0	0	0	0	2	6	1	0	0	1	0	0	5

Grupos establecidos por medio del índice de Kelley

Familia odorante	**Grupos**								
	Método A				**Método B**				
	C_1	**C_2**	**C_3**	**C_4**	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**
Floral	15	0	0	0	15	0	0	0	0
Verde	0	12	0	3	0	0	6	3	6
Frutal	0	6	0	9	0	4	2	2	7
Almizcle	0	1	8	6	0	9	0	1	5

Grupos establecidos manualmente

A partir de los resultados anteriores, se observa que el número de grupos obtenidos por medio de la metodología de Kelley es alto, comparado con el número de grupos establecidos manualmente. Además, en el Método A, los compuestos 13F y 4V, representan un grupo cada uno, de igual forma, esto ocurre con el compuesto 13F en el Método B.

Por otro lado, los grupos C_1-C_4 obtenidos por medio de los Métodos A y B, utilizando el índice de Kelley, están conformados solamente por los compuestos de la familia Floral, a excepción de la molécula 13F. Conjuntamente, estos resultados presentan un grupo que contiene solamente compuestos de la familia Almizcle, para cada uno de los métodos, C_9 y C_6 respectivamente.

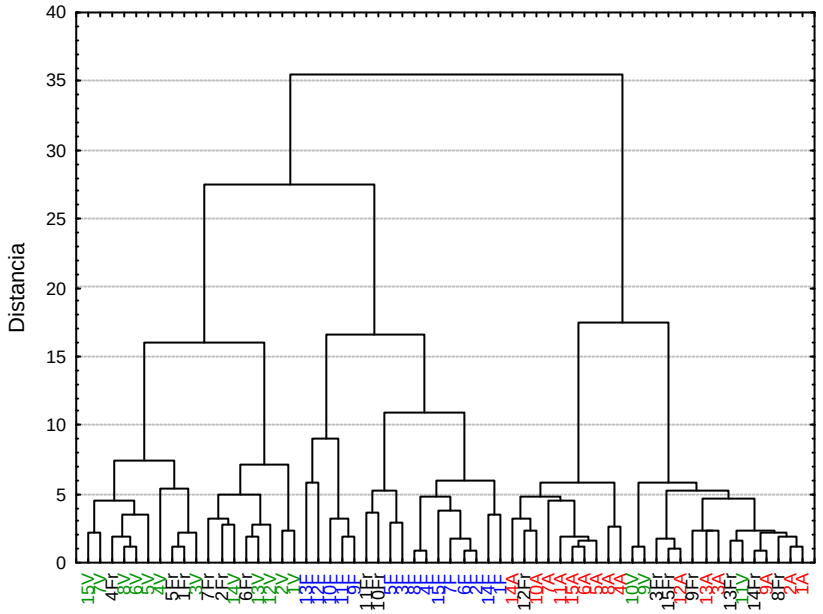
De acuerdo con los grupos obtenidos manualmente para el Método A, se lleva acabo la clasificación de 15, 12 y 8 compuestos pertenecientes a las familias Floral, Verde y Almizcle, respectivamente. Mientras que para el Método B sólo se presenta la clasificación de los 15 compuestos "florales" y 8 moléculas con nota olfativa "almizcle". Los compuestos pertenecientes a la familia Verde se clasifican en 3 grupos mezclados con compuestos de las familias Frutal y Almizcle.

En la **Tabla 28** se reportan las moléculas pertenecientes a cada uno de los diferentes grupos.

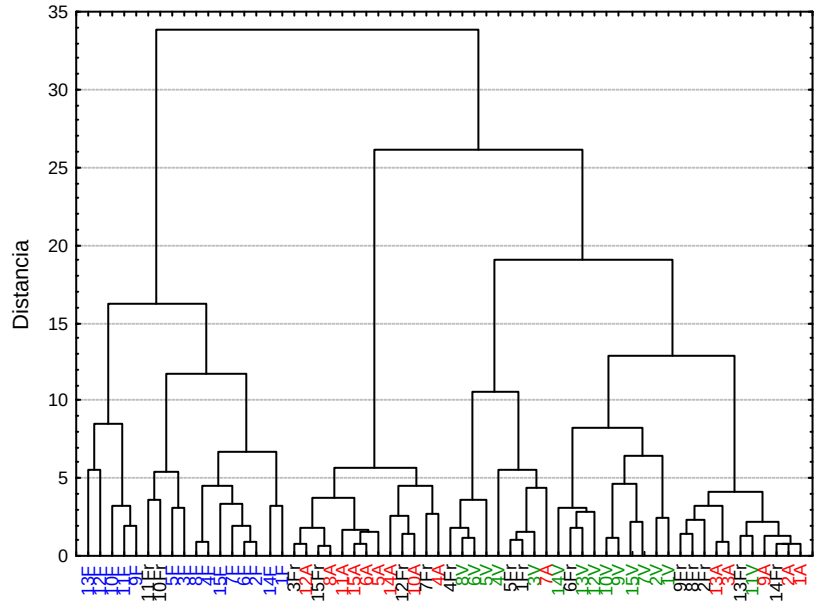
Tabla 28. Moléculas pertenecientes a cada uno de los grupos determinados (Métodos A y B).

Método A		**Método B**	
Grupos	**Moléculas**	**Grupos**	**Moléculas**
C_1	9F, 11F, 12F	C_1	9F, 11F, 12F
C_2	3-5F, 8F	C_2	2F, 6F, 7F, 10F, 15F
C_3	2F, 6F, 7F, 10F, 15F	C_3	3-5F, 8F
C_4	1F, 14F	C_4	1F, 14F
C_5	2V, 12V, 13V, 6Fr	C_5	10A, 12A, 3Fr, 12Fr, 15Fr
C_6	1V, 14V, 7V, 7Fr, 11Fr	C_6	5A, 6A, 8A, 11A, 14A, 15A
C_7	6V, 8V, 15V, 4Fr	C_7	4A, 7Fr
C_8	3-5V, 1Fr, 5Fr, 7A	C_8	2V, 12V, 13V, 6Fr
C_9	5A, 6A, 10A, 11A, 14A, 15A	C_9	1V, 7V, 14V, 11Fr
C_{10}	4A, 8A	C_{10}	3-5V, 1Fr, 5Fr, 7A
C_{11}	9V, 10V, 10Fr,	C_{11}	9-11V, 2Fr, 10Fr, 13Fr
C_{12}	12A, 3Fr, 12Fr, 15Fr	C_{12}	6V, 8V, 15V, 4Fr,
C_{13}	1-3A, 9A, 13A, 11V, 2Fr, 8Fr, 9Fr, 13Fr, 14Fr	C_{13}	1-3A, 9A, 13A, 8Fr, 9Fr, 14Fr
Unitarios	13F, 4V	**Unitarios**	13F
Grupos establecidos por medio del índice de Kelley			
Método A		**Método B**	
Grupos	**Moléculas**	**Grupos**	**Moléculas**
C_1	1-15F	C_1	1-15F
C_2	1-8V, 12-15V, 1Fr, 4-7Fr, 11Fr, 7A	C_2	4-6A, 8A, 10-12A, 14A, 15A, 3Fr, 7Fr, 12Fr, 15Fr
C_3	4-6A, 8A, 10A, 11A, 14A, 15A	C_3	1V, 2V, 7V, 12-14V, 6Fr, 11Fr
C_4	1-3A, 9A, 12A,13A, 9-11V, 2Fr, 3Fr, 8-10Fr, 12-15Fr	C_4	3-5V, 1Fr, 5Fr, 7A
		C_5	1-3A, 9A, 13A, 6V, 8-11V, 15V, 2Fr, 4Fr, 8-10Fr, 13Fr, 14Fr
Grupos establecidos manualmente			

Por último, en la **Figura 48** se presentan los dendogramas obtenidos cuando se realizó la combinación de los descriptores electrónicos, topológicos, geométricos y CPSA.



(A)



(B)

Figura 48. Dendogramas obtenidos por la combinación de los descriptores electrónicos, topológicos, geométricos y CPSA (Métodos A y B).

La **Tabla 29** presenta los resultados de la determinación del número de grupos en los dendogramas obtenidos por la combinación de los descriptores electrónicos y topológicos (Métodos A y B).

Tabla 29. Tabla de contingencia del dendograma obtenido a partir de los descriptores electrónicos, topológicos, geométricos y CPSA (Métodos A y B).

Método A

	Grupos								
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**
Floral	0	0	0	0	0	0	2	2	2
Verde	2	2	1	1	2	2	0	0	0
Frutal	0	1	2	2	1	0	0	0	0
Almizcle	0	0	0	0	0	0	0	0	0

Familia odorante	**C_{10}**	**C_{11}**	**C_{12}**	**C_{13}**	**C_{14}**	**C_{15}**	**C_{16}**	**C_{17}**
Floral	3	0	0	0	0	0	0	0
Verde	0	0	0	0	2	0	0	1
Frutal	0	1	0	0	0	2	1	3
Almizcle	0	1	4	2	0	1	2	3

Método B

	Grupos								
Familia odorante	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**
Floral	3	2	2	3	0	0	0	0	0
Verde	0	0	0	0	0	0	0	0	2
Frutal	0	0	0	0	2	0	1	1	1
Almizcle	0	0	0	0	2	4	2	1	0

Familia odorante	**C_{10}**	**C_{11}**	**C_{12}**	**C_{13}**	**C_{14}**	**C_{15}**	**C_{16}**	**C_{17}**
Floral	0	0	0	0	0	0	0	0
Verde	1	3	2	2	2	0	0	1
Frutal	2	1	0	0	0	3	0	2
Almizcle	0	0	0	0	0	0	2	3

Grupos establecidos por medio del índice de Kelley

Familia odorante	**Grupos**								
	Método A				**Método B**				
	C_1	**C_2**	**C_3**	**C_4**	**C_1**	**C_2**	**C_3**	**C_4**	**C_5**
Floral	0	15	0	0	15	0	0	0	0
Verde	12	0	0	3	0	0	5	9	1
Frutal	6	2	1	6	2	4	3	1	5
Almizcle	0	0	9	6	0	9	1	0	5

Grupos establecidos manualmente

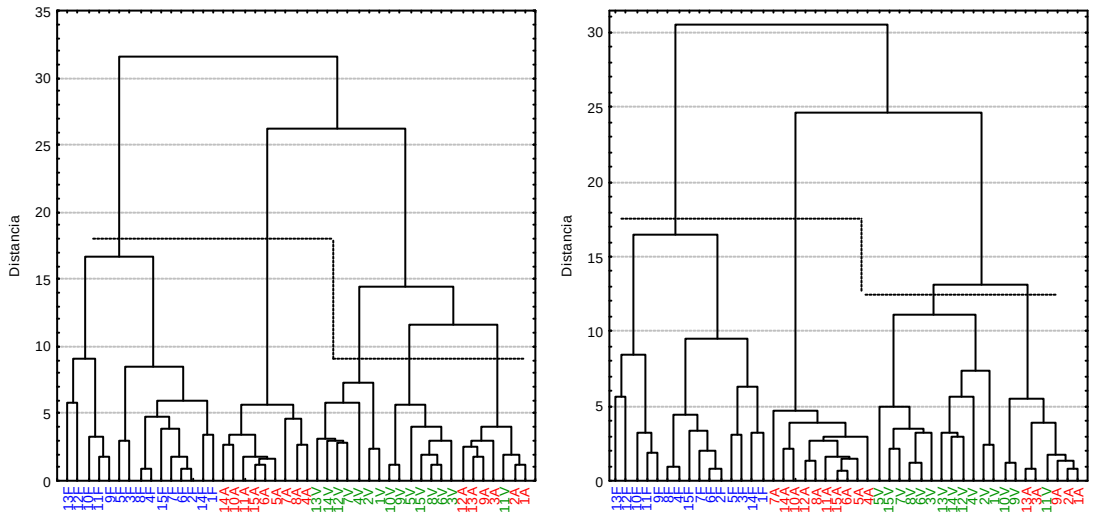
Los resultados obtenidos a partir de la metodología de Kelley, muestran un incremento del número de grupos comparados con el análisis de agrupamiento realizado sobre la combinación de tres tipos de descriptores. Además, se observa un aumento en el número de grupos, con un solo elemento. En la **Tabla 30** se pueden observar las moléculas pertenecientes a cada uno de los diferentes grupos.

Tabla 30. Moléculas pertenecientes a cada uno de los grupos (Métodos A y B).

Método A		**Método B**	
Grupos	**Moléculas**	**Grupos**	**Moléculas**
C_1	7V, 15V	C_1	9-11F
C_2	6V, 8V, 4Fr	C_2	3F, 5F
C_3	1Fr, 5Fr, 3V	C_3	4F, 8F
C_4	2Fr, 7Fr, 14V	C_4	2F ,5F ,7F
C_5	12V, 13V, 6Fr	C_5	8A, 12A, 3Fr, 15Fr
C_6	1V, 2V	C_6	5A ,6A , 11A, 15A
C_7	9F, 11F	C_7	10A, 14A, 12Fr
C_8	3F, 5F	C_8	4A, 7Fr
C_9	4F, 8F	C_9	6V, 8V, 4Fr
C_{10}	2F, 6F, 7F	C_{10}	3V, 1Fr, 5Fr
C_{11}	10A, 12Fr	C_{11}	12-14V, 6Fr
C_{12}	5A, 6A, 11A, 15A	C_{12}	9V, 10V
C_{13}	4A, 8A	C_{13}	7V, 15V
C_{14}	9V, 10V	C_{14}	1V, 2V
C_{15}	3Fr, 15Fr, 12A	C_{15}	2Fr, 8Fr, 9Fr
C_{16}	3A, 13A, 9Fr	C_{16}	3A, 13A
C_{17}	1A, 2A, 9A, 8Fr, 13Fr, 14Fr, 11V	C_{17}	1A, 2A, 9A, 11V, 13Fr, 14Fr
Grupos unitarios	4V, 5V, 1F, 10F, 12-15F, 10Fr, 11Fr, 7A, 14A	**Grupos unitarios**	1F, 12-15F, 10Fr, 11Fr, 4V, 5V, 7A
Grupos establecidos por medio del índice de Kelley			
Método A		**Método B**	
Grupos	**Moléculas**	**Grupos**	**Moléculas**
C_1	1-8V, 12-15V, 1Fr, 2Fr, 4-7Fr	C_1	1-15F, 10Fr, 11Fr
C_2	1-15F, 10Fr, 11Fr	C_2	4-6A, 8A, 10-12A, 14A, 15A, 3Fr, 4Fr, 7Fr, 12Fr, 15Fr
C_3	4-8A, 10A, 11A, 14A, 15A, 12Fr	C_3	7A, 1Fr, 4Fr, 5Fr, 3-6V, 8V
C_4	1-3A, 9A, 12A, 13A, 9-11V, 3Fr, 8Fr, 9Fr, 13Fr-15Fr	C_4	1V, 2V, 7V, 9V, 10V, 12-15V
		C_5	1-3A, 9A, 13A, 2Fr, 8Fr, 9Fr, 13Fr, 14Fr, 11V
Grupos establecidos manualmente			

De igual manera que los resultados anteriores, podemos apreciar la clasificación general de las familias Fbral, Almizcle y Verde, siendo los resultados obtenidos por medio del Método A superiores en cuanto a la clasificación se refiere.

Con el objetivo visualizar de una manera más clara la clasificación de las familias odorantes Almizcle, Floral y Verde, se realizó el análisis de agrupamiento sobre la combinación de los descriptores electrónicos, topológicos, geométricos y CPSA, eliminando los compuestos de la familia Frutal en las matrices obtenidas por medio de los Métodos A y B (**Figura 49**).



Método A **Método B**

Figura 49. Dendogramas obtenidos por la combinación de los descriptores electrónicos, topológicos, geométricos y CPSA, eliminando la familia Frutal (Métodos A y B).

Los dendogramas anteriores, muestran la clasificación de las tres familias odorantes en cinco grupos para el Método A. Los compuestos de la familia Floral se clasifican en un solo grupo, los compuestos Almizcle en dos grupos e incluyen el compuesto 11V y la familia Verde en dos grupos. Por otro lado, de acuerdo con el dendograma obtenido por el Método B, los compuestos se clasifican en cuatro grupos, el primero corresponde a la

familia Floral, el segundo y cuarto están conformados por las moléculas Almizcle, a excepción de los compuestos 9-10V, incluidos en el cuarto grupo y el tercero constituye los compuestos de la familia Verde.

De esta manera, se puede concluir, que por medio del análisis de agrupamiento realizado sobre los diferentes tipos de descriptores, se puede realizar la clasificación de las familias Floral, Almizcle y Verde. Este resultado es similar al obtenido en la sección anterior (5.2.4) donde se realizó el análisis de componentes principales.

Continuando con el estudio de clasificación de los compuestos fragantes, el siguiente objetivo fue poder establecer si los modos vibracionales podían ser utilizados como descriptores en la clasificación de las 60 moléculas. Para poder alcanzar este objetivo, el siguiente paso consistió en realizar el PCA y el análisis de grupos sobre las frecuencias obtenidas por medio de los cálculos *ab initio*.

5.2.9 PCA de los modos vibracionales. El espectro vibracional (100-4000 cm^{-1}) está conformado especialmente por dos regiones. La primera de ellas es la región por encima de 1700 cm^{-1}. En esta región los modos vibracionales son debidos casi exclusivamente a las tensiones de un par de átomos, por ejemplo en los enlaces C=O, S–H, C≡N, C–H, N–H y O–H. En la segunda región, por debajo de 1700 cm^{-1}, la mayoría de las vibraciones son complejas e involucran tres o más átomos. El rango vibracional para las moléculas odorantes es típicamente de 50 a 3400 cm^{-1}, sin embrago, Turín [4] ha argumentado que por debajo de 600 cm^{-1} el olfato humano no es sensible a las vibraciones moleculares.

Por medio de los cálculos computacionales se obtuvieron los diferentes modos vibracionales (frecuencias) de las 60 moléculas bajo estudio. En la **Figura 50** se presenta un ejemplo de los espectros infrarrojo obtenidos para algunos compuestos odorantes.

Para realizar el análisis de los modos vibracionales se trabajó con los valores de frecuencias comprendidas en el intervalo 10-4130 cm^{-1}. La razón de escogencia de todo el espectro vibracional como rango de trabajo, se debe a la presencia de bandas con intensidad apreciable en algunos espectros infrarrojo por debajo de 600 cm^{-1}. Una vez establecido el rango de trabajo, se procedió a realizar una tabla donde se ordenaron los valores de frecuencia en rangos de 10 cm^{-1}, junto con su intensidad respectiva. Como resultado se obtuvo una matriz de datos de 60 x 442 (**Anexo 18**), en la cual se puede apreciar, que existen ciertos valores de frecuencias con más de un valor en el intervalo establecido.

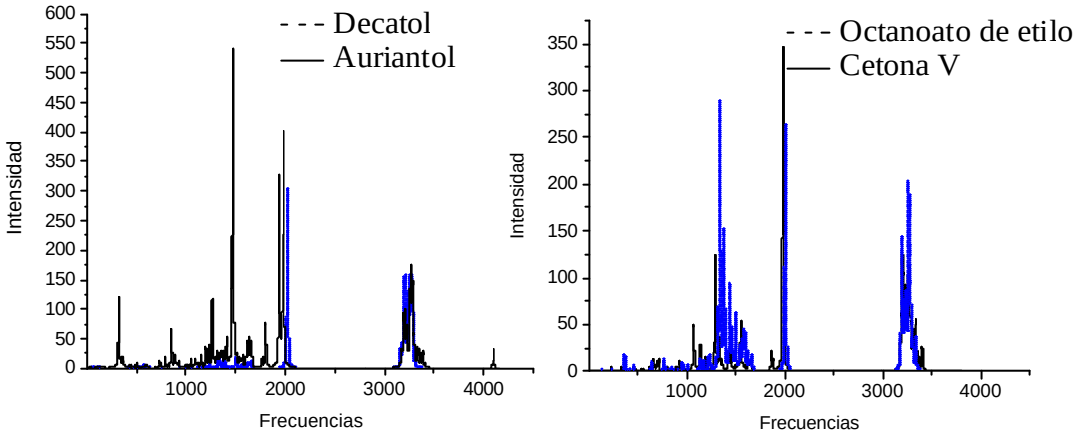


Figura 50. Espectros infrarrojo obtenidos por medio de los cálculos computacionales.

Para verificar si la clasificación de las 60 moléculas bajo estudio se podía realizar de acuerdo con sus modos vibracionales, se realizó inicialmente el PCA de las frecuencias obtenidas teóricamente sin un previo tratamiento de los valores iniciales.

En la **Tabla 31** se puede observar que a partir de los resultados obtenidos por medio del PCA se obtienen 59 factores principales, donde los tres primeros factores principales solamente contienen el 12 % de la varianza total de los datos originales.

Tabla 31. Resultados del PCA de los modos vibracionales.

Factor	Valor propio	% Total de varianza	% de Contribución	Factor	Valor propio	% Total de varianza	% de Contribución
1	22,17136	5,016144	5,0161	**31**	6,45642	1,460729	75,0053
2	17,01401	3,849323	8,8655	**32**	6,11482	1,383443	76,3887
3	15,63584	3,537521	12,4030	**33**	5,98943	1,355075	77,7438
4	15,04260	3,403302	15,8063	**34**	5,80922	1,314304	79,0581
5	14,47721	3,275386	19,0817	**35**	5,60755	1,268677	80,3268
6	13,58940	3,074525	22,1562	**36**	5,54126	1,253679	81,5804
7	12,64905	2,861775	25,0180	**37**	5,29540	1,198055	82,7785
8	12,49400	2,826697	27,8447	**38**	5,18435	1,172929	83,9514
9	12,06312	2,729213	30,5739	**39**	4,88892	1,106091	85,0575
10	11,81351	2,672739	33,2466	**40**	4,84196	1,095465	86,1530
11	11,69644	2,646254	35,8929	**41**	4,68545	1,060057	87,2130
12	10,95253	2,477947	38,3708	**42**	4,33780	0,981403	88,1944
13	10,69539	2,419772	40,7906	**43**	4,25537	0,962754	89,1572
14	10,35664	2,343132	43,1337	**44**	4,11397	0,930763	90,0880
15	10,13063	2,291999	45,4257	**45**	3,97584	0,899512	90,9875
16	9,91063	2,242224	47,6680	**46**	3,84102	0,869010	91,8565
17	9,40104	2,126933	49,7949	**47**	3,59616	0,813612	92,6701
18	9,23399	2,089139	51,8840	**48**	3,52764	0,798109	93,4682
19	9,06835	2,051663	53,9357	**49**	3,42654	0,775236	94,2434
20	8,96123	2,027428	55,9631	**50**	3,33603	0,754759	94,9982
21	8,73198	1,975562	57,9387	**51**	3,15407	0,713591	95,7118
22	8,52697	1,929179	59,8679	**52**	3,02688	0,684815	96,3966
23	8,39933	1,900301	61,7682	**53**	2,93518	0,664067	97,0607
24	8,23190	1,862420	63,6306	**54**	2,73435	0,618632	97,6793
25	8,07305	1,826482	65,4571	**55**	2,44100	0,552263	98,2316
26	7,66970	1,735226	67,1923	**56**	2,42722	0,549144	98,7807
27	7,48909	1,694365	68,8867	**57**	2,09180	0,473257	99,2540
28	7,07535	1,600757	70,4874	**58**	1,68428	0,381059	99,6350
29	6,81365	1,541550	72,0290	**59**	1,61318	0,364972	100,0000
30	6,69886	1,515579	73,5445				

La representación gráfica de los PC que poseen la mayor información (**Figura 51**) muestra una ligera diferencia entre dos grupos de odorantes: el grupo Almizcle y grupo el Verde. Sin embargo, las moléculas de las familias Floral y Frutal se encuentran mezcladas entre los grupos Verde y Almizcle. Además, se observa que los compuestos 12F, 11F, 9F y 14F se comportan como puntos anómalos.

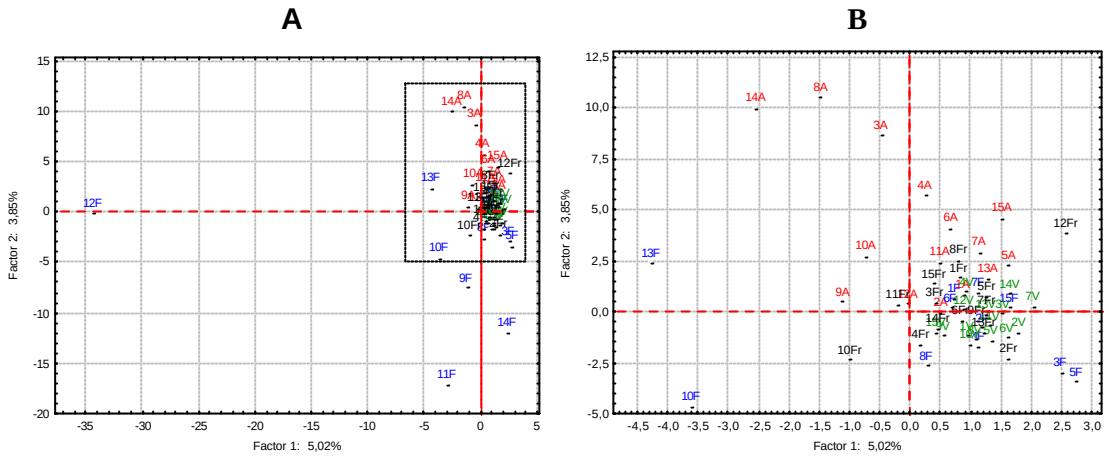


Figura 51. Representación de los dos primeros PC. A: Gráfica de dispersión de las 60 moléculas odorantes; B: zona del recuadro ampliada.

Con base en este primer análisis de componentes principales de los modos vibracionales, se pudo establecer que las frecuencias podían servir como descriptores en la clasificación de las familias odorantes. Sin embargo, se concluyó, que era necesario realizar un tratamiento previo de los modos vibracionales, antes de realizar el PCA, con el objetivo de reducir el número de modos vibracionales y obtener una mejor clasificación de los compuestos fragantes.

De acuerdo con lo mencionado anteriormente, se establecieron diferentes tratamientos de los modos vibracionales. A continuación, se presentan los resultados de algunos de los tratamientos realizados sobre la tabla de frecuencias, antes de someter los datos a la clasificación por medio del análisis de componentes principales.

Promedios de intensidades. El primer tratamiento realizado, con el objetivo de disminuir el número de descriptores, fue tomar el promedio de los valores de intensidad siempre y cuando las frecuencias presentaran más de un valor en el rango establecido (10 cm^{-1}). Por medio de este tratamiento se obtuvo una matriz de datos de 60 x 240.

Los resultados obtenidos por medio de PCA arrojaron una solución de 59 factores principales donde los tres primeros factores principales solamente contienen el 15% de la varianza total.

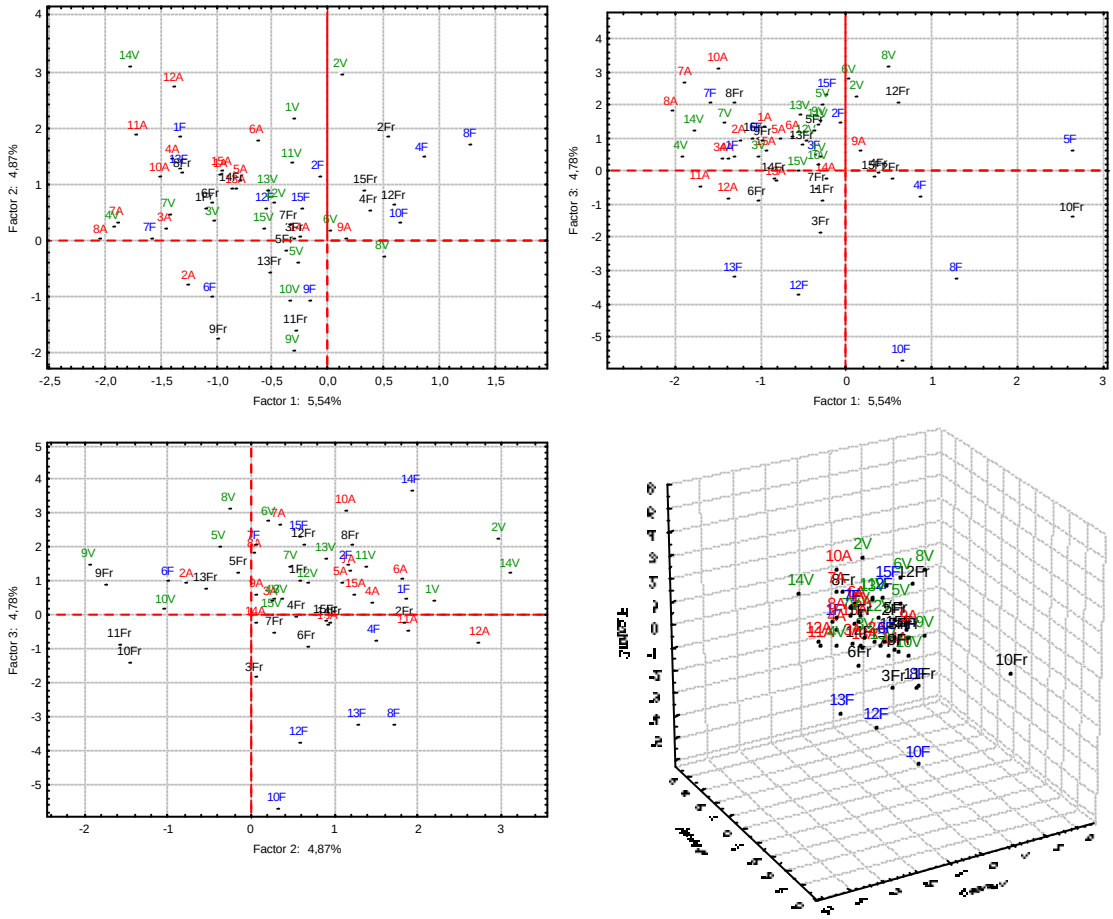


Figura 52. Gráficas de dispersión de los valores de coordenadas de los 60 compuestos (Promedios de intensidades).

La **Figura 52** presenta las gráficas en dos y tres dimensiones de los tres primeros componentes principales ampliadas en la zona de mayor concurrencia de las moléculas. Sin embargo, como se puede apreciar, el agrupamiento por medio del PCA de la matriz de promedios no presenta alguna diferenciación adecuada de las cuatro familias odorantes.

Después de este primer tratamiento de datos, donde no se observa ninguna clasificación de las 60 moléculas odoríferas, se decidió realizar otra matriz donde se retener los valores de intensidad más altos presentes en un mismo rango de frecuencias, con el fin de establecer que la reducción del número de frecuencias no era el único factor predominante en la clasificación de las moléculas fragantes.

Retención del mayor valor de intensidad presente en un mismo rango de frecuencias. Por medio del tratamiento previo de los datos de frecuencias iniciales donde se retuvo el mayor valor de intensidad, se obtuvo una segunda matriz de datos 60 x 240. Sin embargo, los resultados no fueron superiores al tratamiento anterior e igualmente no se obtuvo una buena clasificación de las moléculas odorantes.

En la **Figura 53** se representa la gráfica en tres dimensiones de los tres primeros componentes principales, los cuales representan el 15% de la varianza total. Además, al igual que en el caso anterior, se obtienen los mismos cuatro compuestos de la familia Floral con una clasificación anómala (14F, 11F, 9F y 3F).

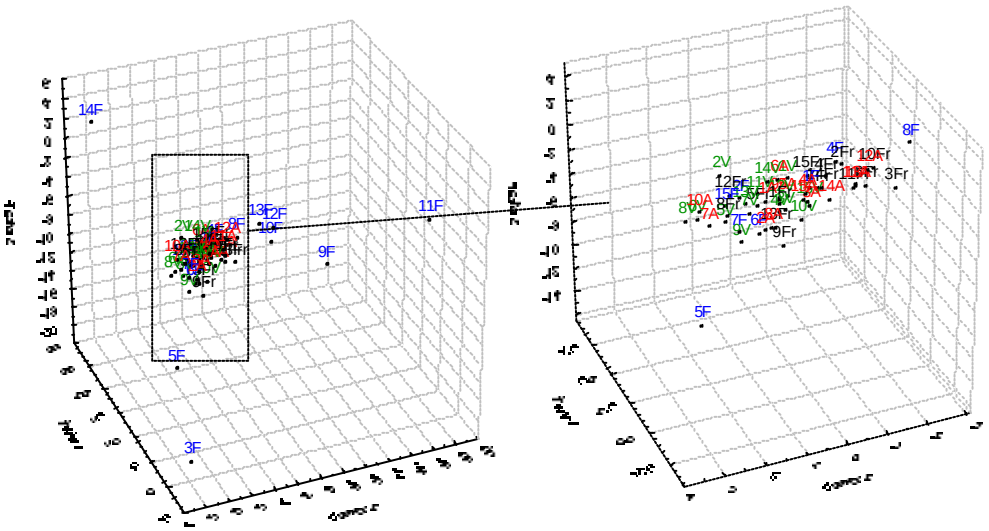


Figura 53. Gráfica en tres dimensiones de los tres primeros PC.

En el **Anexo 19**, se muestran las gráfica bidimensionales de los tres primeros factores principales, junto con las valores propios de cada uno de los PC (**Anexo 20**).

De acuerdo con los resultados obtenidos en los dos análisis de componentes principales anteriores de los modos vibracionales, se estableció, que además de necesitar tener en cuenta las frecuencias que presentaban los mayores valores de intensidad; ya que estas frecuencias son las que diferencian una familia odorante de otra, según la teoría de Turín, se debe retener el mayor número posible de frecuencias y no agrupar indiscriminadamente las variables cuando presentan más de un valor de frecuencia en el mismo rango de trabajo.

Los siguientes tratamientos desarrollados para las frecuencias se realizaron con base en algunas observaciones sobre los espectros vibracionales obtenidos anteriormente. La primera observación se basa en ciertos rangos de frecuencias específicos, ya que una característica especial de los espectros de las 60 moléculas odorantes es que se presentan ciertos rangos donde los valores de las intensidades no son superiores a algún valor específico.

Con base en esta primera observación, y manteniendo la idea de retener las frecuencias que presentaran valores de intensidad altos, se procedió a asumir valores de intensidad determinados como valores límites representativos de información en los espectros IR. Esto es, que las frecuencias, que presentaban valores de intensidades por debajo de este límite, podían ser despreciadas o simplemente no aportaban suficiente información para realizar la clasificación de las moléculas fragantes.

Con el objetivo de disminuir el número de variables, sin eliminar las frecuencias con valores altos de intensidad, a diferencia de los tratamientos anteriores, el siguiente paso fue agrupar en un mayor rango de frecuencia (mayor de 10 cm^{-1}) los modos vibracionales que presentaron valores por debajo del límite de información establecido y manteniendo intactas las frecuencias sobrantes.

Para el primer tratamiento de los datos, se estableció un valor límite de intensidades inicial igual a 10. Posteriormente, fue aumentado en 10 unidades sucesivamente hasta encontrar la mejor clasificación de las moléculas fragantes.

De los resultados obtenidos se observó que si el valor límite era superior o igual que 80, los compuestos de los diferentes grupos odorantes, en especial, el grupo Almizcle y el Frutal, empezaba a superponerse. Por otro lado, el valor límite de intensidad, que presentó la mejor clasificación de los compuestos, fue de 60.

Valor límite de intensidad 60. Utilizando como límite de intensidad el valor 60, se obtuvo una matriz de frecuencias de 60 x 186. En el **Anexo 21** se aprecia que número de frecuencias se ha reducido considerablemente (de 442 a 186), pero se han retenido las variables que poseen las bandas más intensas.

De acuerdo con los resultados de los valores propios de cada componente principal (**Anexo 22**), cuyos valores se encuentran graficados en la **Figura 54**, solamente el 16% de la varianza total de los datos originales puede ser representado por los tres primeros factores principales.

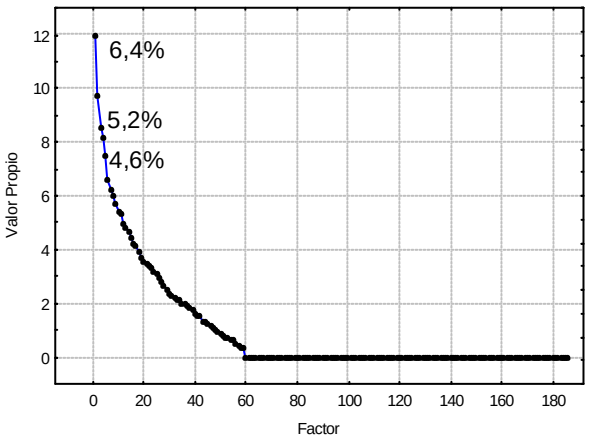


Figura 54. Gráfica de valores propios *vs* factores principales.

La **Figura 55** muestra las gráficas de las coordenadas de las moléculas en dos y tres dimensiones, donde se presenta principalmente, la clasificación de los grupos Almizcle y Verde, junto con el solapamiento de algunas moléculas pertenecientes al grupo Frutal entre los dos grupos anteriores. Sin embargo, en la gráfica de Factor 1 *vs* Factor 3 y Factor 3 *vs* Factor 2, los cuales representan el 11 y el 9.8%, respectivamente, se observa un pequeño agrupamiento de algunas moléculas pertenecientes al grupo Frutal, al igual que en la gráfica de los tres componentes principales. Por otro lado, cabe mencionar que las moléculas 12F, 11F, 13 F, 3F 5F, 10Fr y especialmente 14F, se encuentran separadas del grupo principal.

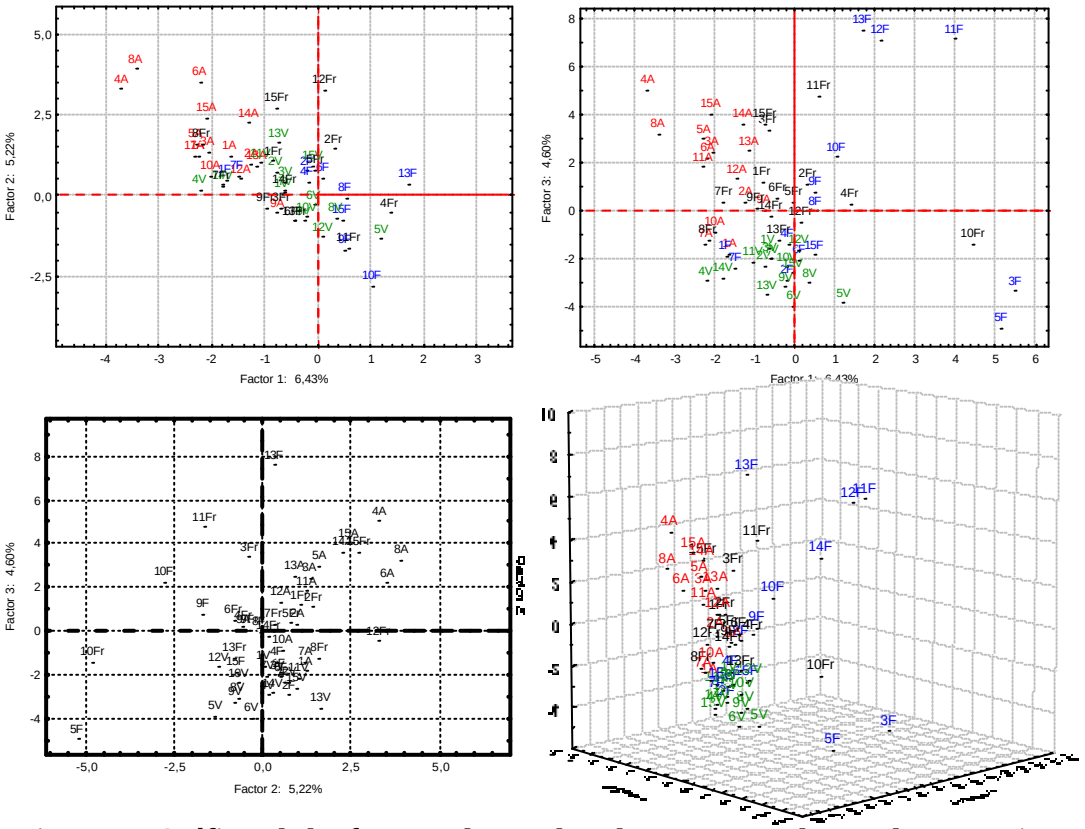


Figura 55. Gráficas de los factores de coordenadas representadas por los tres primeros componentes principales (Valor límite de intensidad 60).

Pensando en disminuir el número de variables y aumentar el porcentaje de varianza representado por los primeros factores principales, se procedió a determinar si la

presencia de las frecuencias que exhiben valores de intensidad menores 60 era un factor determinante en la clasificación de los compuestos fragantes. Como resultado de la eliminación de estas variables, se obtuvo la matriz de 60 x 139.

Los resultados de los valores propios arrojados por el PCA sobre esta matriz (**Tabla 32**), muestran que el porcentaje total de varianza representada por los tres primeros PC aumentó solamente de 16 a 17%. Sin embargo la reducción del número de variables fue considerable.

Tabla 32. Valores propios y porcentaje de varianza.

Factor	Valor propio	% Total de varianza	% de Contribución	Factor	Valor propio	% Total de varianza	% de Contribución
1	9,932937	7,145998	7,1460	**31**	1,664124	1,197211	84,1313
2	7,358844	5,294132	12,4401	**32**	1,612886	1,160349	85,2917
3	6,688143	4,811614	17,2517	**33**	1,556500	1,119784	86,4115
4	6,022763	4,332923	21,5847	**34**	1,539075	1,107248	87,5187
5	5,853431	4,211102	25,7958	**35**	1,434170	1,031777	88,5505
6	5,501888	3,958193	29,7540	**36**	1,317748	0,948020	89,4985
7	5,101838	3,670387	33,4243	**37**	1,303556	0,937810	90,4363
8	4,694553	3,377376	36,8017	**38**	1,269983	0,913657	91,3500
9	4,420216	3,180011	39,9817	**39**	1,179052	0,848239	92,1982
10	4,212930	3,030885	43,0126	**40**	1,146387	0,824739	93,0230
11	4,065547	2,924854	45,9375	**41**	1,069836	0,769666	93,7926
12	3,924579	2,823438	48,7609	**42**	0,939463	0,675872	94,4685
13	3,858121	2,775627	51,5365	**43**	0,900184	0,647615	95,1161
14	3,641980	2,620130	54,1567	**44**	0,829222	0,596563	95,7127
15	3,441262	2,475728	56,6324	**45**	0,785645	0,565212	96,2779
16	3,270879	2,353150	58,9855	**46**	0,738333	0,531175	96,8091
17	3,108893	2,236613	61,2222	**47**	0,649513	0,467275	97,2763
18	3,034788	2,183300	63,4055	**48**	0,567912	0,408569	97,6849
19	2,887137	2,077077	65,4825	**49**	0,545993	0,392801	98,0777
20	2,677936	1,926573	67,4091	**50**	0,501308	0,360653	98,4384
21	2,648876	1,905666	69,3148	**51**	0,404925	0,291313	98,7297
22	2,494665	1,794723	71,1095	**52**	0,382813	0,275405	99,0051
23	2,458975	1,769047	72,8785	**53**	0,365941	0,263267	99,2684
24	2,244371	1,614655	74,4932	**54**	0,258509	0,185978	99,4543
25	2,182081	1,569843	76,0630	**55**	0,249540	0,179525	99,6339
26	2,064544	1,485284	77,5483	**56**	0,212181	0,152648	99,7865
27	2,012878	1,448114	78,9964	**57**	0,179234	0,128945	99,9154
28	1,924082	1,384232	80,3807	**58**	0,079697	0,057336	99,9728
29	1,826323	1,313901	81,6946	**59**	0,037830	0,027216	100,0000
30	1,722982	1,239555	82,9341				

La gráfica de los dos primeros factores principales (**Figura 56**) muestra, al igual que en el caso anterior, la separación de dos familias, Almizcle y Verde, mientras que algunos compuestos de la familia Frutal se agrupan en un pequeño *cluster* (8Fr, 6Fr, 12Fr, 14Fr y 9Fr); además, como punto común con el análisis anterior, los compuestos 14F y 11F se comportan de manera anómala (el compuesto 14F no se representa en la **Figura 56**.)

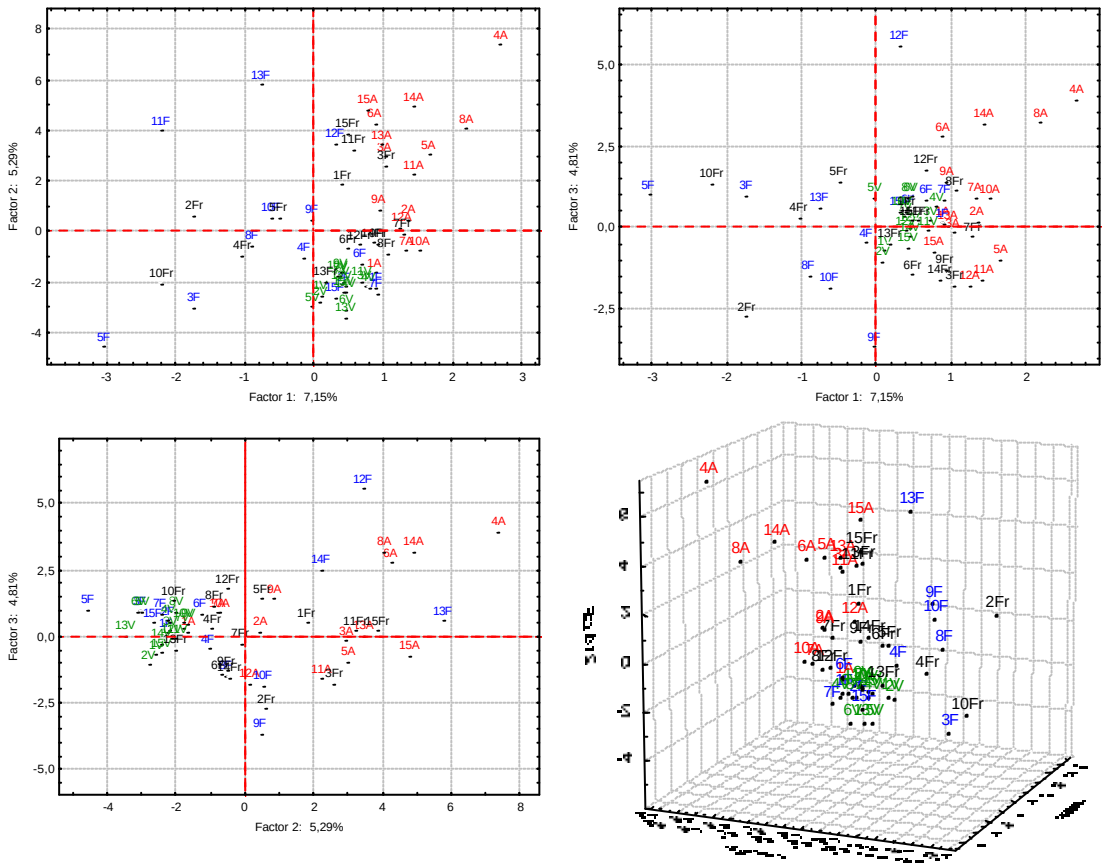


Figura 56. Gráficas de los factores de coordenadas representadas por los tres primeros componentes principales.

Para concluir con los análisis de las frecuencias se realizó un último tratamiento. La idea general de este tratamiento se basó en observaciones sobre los valores de la desviación estándar. De esta manera, la tabla de frecuencias e intensidades fue sometida a un análisis estadístico para calcular la varianza y la desviación estándar junto con otras propiedades estadísticas.

Por medio de la observación de los datos obtenidos en el análisis estadístico, se puede inferir que las variables que presentan un valor de intensidad alto, también presentan un valor de varianza y desviación estándar alto. Teniendo esto en cuenta, se eliminaron las variables de acuerdo con los valores de desviación estándar. El análisis se desarrolló de igual manera que el anterior, tomando valores límites y eliminando las variables, cuyos valores se encontraran por debajo de este valor. Los valores límite se tomaron a partir de 3 y este valor se fue aumentando una unidad en los siguientes tratamientos.

De acuerdo con los resultados obtenidos, se observó que la diferenciación de los compuestos fragantes cundo se tomaron valores por encima de 16, presentaron un solapamiento y no se logró una clasificación adecuada de las moléculas. El valor límite, que presentó la mejor clasificación, fue 14. Una vez eliminadas las variables, se obtuvo una matriz de 60 x 110.

En la **Tabla 33** se presentan los valores propios de los factores principales donde se observa un aumento en la varianza total representada por los tres primeros PC.

Las gráficas de los tres primeros factores principales muestran una clasificación similar a la obtenida en el análisis de PCA anterior (**Figura 57**); sin embargo, se obtiene una mejor clasificación en cuanto al grupo Almizcle se refiere, ya que se observan dos grupos de estos compuestos con un mayor agrupamiento (el compuesto 14F no se representa en las gráficas).

En la **Tabla 34**, se presentan los factores de aporte de las variables de los cuatro primeros componentes principales. Los valores con un asterisco representan los aportes de cada variable con un valor superior a 0.7. El factor 1 posee el mayor número de frecuencias con estos valores (1240 cm^{-1}, 1480 cm^{-1}, 1640 cm^{-1}, 1700 cm^{-1}, 1780 cm^{-1}, 1800 cm^{-1}, 1960 cm^{-1} y 3940 cm^{-1}). Mientras que el factor 3 y el factor 4 solamente presentan un valor 1440 cm^{-1} (-0,713296) y 1570 cm^{-1} (-0,785029), respectivamente.

Tabla 33. Resultados del análisis de componentes principales.

Factor	Valor propio	% Total de varianza	% de Contribución	Factor	Valor propio	% Total de varianza	% de Contribución
1	8,214809	7,468008	7,4680	**31**	1,259273	1,144794	86,8560
2	6,837501	6,215910	13,6839	**32**	1,215670	1,105155	87,9612
3	6,160139	5,600126	19,2840	**33**	1,174810	1,068009	89,0292
4	5,558930	5,053573	24,3376	**34**	1,075980	0,978164	90,0073
5	4,688367	4,262152	28,5998	**35**	1,023223	0,930202	90,9375
6	4,241263	3,855694	32,4555	**36**	0,965944	0,878131	91,8157
7	3,949394	3,590358	36,0458	**37**	0,907998	0,825453	92,6411
8	3,770730	3,427936	39,4738	**38**	0,808170	0,734700	93,3758
9	3,650023	3,318202	42,7920	**39**	0,754830	0,686209	94,0620
10	3,501056	3,182778	45,9747	**40**	0,706948	0,642680	94,7047
11	3,348275	3,043886	49,0186	**41**	0,686809	0,624372	95,3291
12	3,081847	2,801679	51,8203	**42**	0,606543	0,551403	95,8805
13	3,063557	2,785052	54,6054	**43**	0,567613	0,516012	96,3965
14	2,744538	2,495035	57,1004	**44**	0,510769	0,464335	96,8608
15	2,693958	2,449052	59,5494	**45**	0,492741	0,447947	97,3088
16	2,616135	2,378304	61,9277	**46**	0,436947	0,397225	97,7060
17	2,550342	2,318492	64,2462	**47**	0,382211	0,347464	98,0535
18	2,382316	2,165742	66,4120	**48**	0,350463	0,318603	98,3721
19	2,335692	2,123357	68,5353	**49**	0,313673	0,285157	98,6572
20	2,187391	1,988538	70,5239	**50**	0,282249	0,256590	98,9138
21	2,044393	1,858539	72,3824	**51**	0,238901	0,217183	99,1310
22	1,959841	1,781673	74,1641	**52**	0,213742	0,194311	99,3253
23	1,853376	1,684887	75,8490	**53**	0,183551	0,166864	99,4922
24	1,810314	1,645740	77,4947	**54**	0,145324	0,132113	99,6243
25	1,733309	1,575736	79,0705	**55**	0,140201	0,127455	99,7517
26	1,595612	1,450556	80,5210	**56**	0,113570	0,103246	99,8550
27	1,535628	1,396026	81,9170	**57**	0,097689	0,088808	99,9438
28	1,469096	1,335542	83,2526	**58**	0,032294	0,029358	99,9732
29	1,361302	1,237547	84,4901	**59**	0,029532	0,026848	100,0000
30	1,343197	1,221088	85,7112				

El comportamiento anómalo del compuesto 14F se debe principalmente a los valores de intensidad altos que presentan las frecuencias anteriormente mencionadas en el Factor 1, excluyendo a la frecuencia 1640 cm^{-1}, mientras que la clasificación del compuesto 11F depende del Factor 3 (frecuencia 1440 cm^{-1}) y el Factor 2 el cual, aunque no muestra valores por encima de 0.7 exhibe valores por encima de 0.5 en las frecuencias 1300 cm^{-1}, 1320 cm^{-1}, 1440 cm^{-1}, 1490 cm^{-1}, 1850 cm^{-1} correspondientes a las frecuencias de mayor intensidad en el espectro IR del Cinamato de linalilo 11F (**Figura 58**).

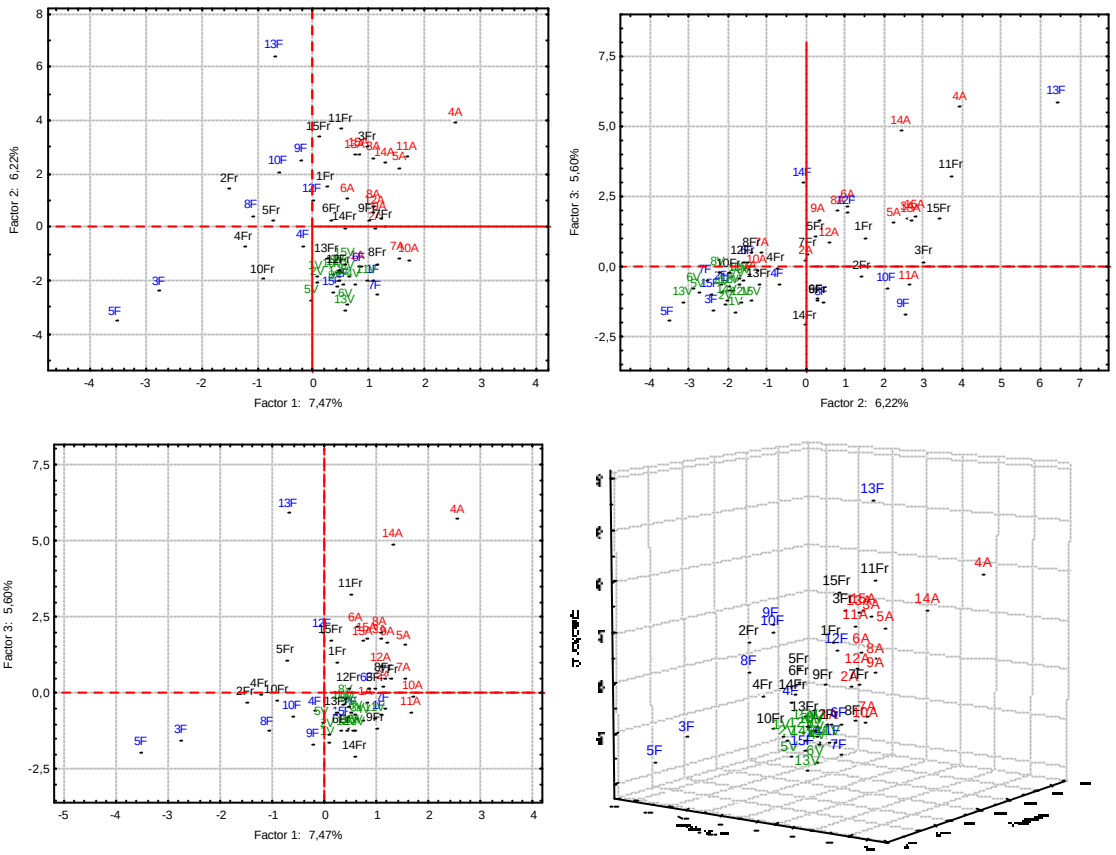
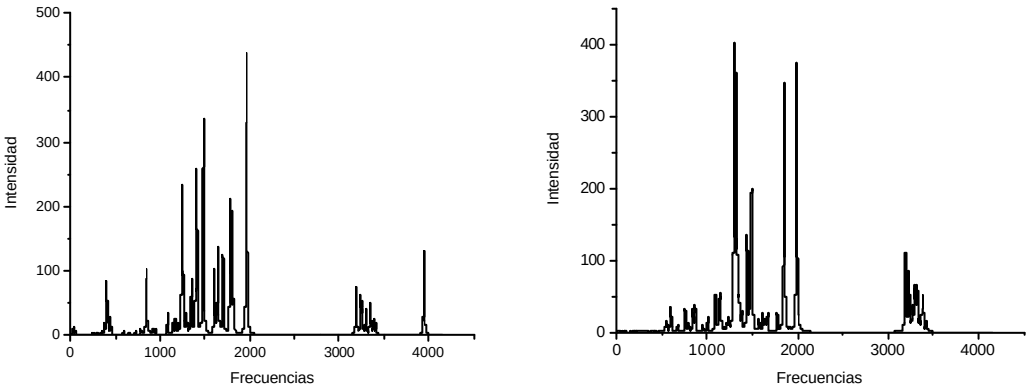


Figura 57. Gráficas de los factores de coordenadas representados por los tres primeros factores (Reducción basada en la desviación estándar).



(a) (b)

Figura 58. Espectros IR del antranilato de dimetilo 14F (a) y el Cinamato de Linalilo 11F (b).

Tabla 34. Factores de aporte de las variables en los cuatro primeros componentes principales.

	Factor 1	Factor 2	Factor 3	Factor 4		Factor 1	Factor 2	Factor 3	Factor 4		Factor 1	Factor 2	Factor 3	Factor 4		Factor 1	Factor 2	Factor 3	Factor 4
290	0,073099	-0,079160	0,030311	0,008751	**1390**	-0,042996	0,412765	0,443576	-0,349541	**1960**	-0,932192*	-0,004344	0,158210	0,224250	**V85**	0,010730	0,102635	0,115853	0,113118
360	-0,017625	-0,047217	-0,077272	0,021506	**1400**	-0,102367	-0,079497	-0,031360	-0,668610	**1970**	0,011223	0,075506	0,079983	-0,061693	**3230**	0,084509	0,459882	-0,279512	0,032971
400	-0,585649	-0,153175	-0,038905	-0,080830	**1410**	-0,395934	0,368921	0,420355	-0,266809	**1980**	0,030130	0,046399	0,009258	0,141927	**V87**	0,081077	0,090446	-0,081755	0,238931
800	-0,056460	0,076328	-0,034069	0,010719	**V32**	0,070374	0,110720	0,082452	0,176839	**1990**	0,001601	0,421561	-0,320012	-0,057788	**3240**	-0,098802	-0,151368	-0,001705	0,286815
1100	0,020792	-0,080972	-0,000860	0,056280	**1420**	0,074054	0,044502	0,039546	0,104265	**2000**	-0,032894	-0,022749	0,055622	-0,355074	**V89**	0,071899	-0,102411	-0,021201	0,101382
1140	-0,142776	0,443519	-0,124709	0,164925	**1430**	-0,159794	-0,089146	-0,043832	-0,487111	**2010**	0,067014	0,106339	0,010905	0,076496	**3250**	0,020058	-0,119902	0,053218	0,250856
1170	-0,263637	-0,230777	-0,051233	-0,632414	**V35**	-0,129819	-0,118048	-0,083259	-0,669981	**2020**	0,092708	-0,378013	-0,100928	0,057931	**V91**	0,196484	0,166987	0,251927	0,225150
1190	-0,090911	-0,192814	-0,083581	-0,070852	**1440**	-0,002662	-0,097850	-0,059100	0,033625	**2030**	0,017848	0,305102	0,271133	-0,249518	**V92**	0,146072	0,247718	0,335489	0,227237
1200	0,063319	0,245151	0,281405	-0,260805	**V37**	-0,083158	0,569272	-0,713296*	-0,017215	**2060**	0,048651	-0,100652	-0,038240	0,065539	**3260**	0,258523	-0,307116	-0,126675	0,295173
1220	0,008931	0,076073	0,032161	-0,042833	**1450**	-0,050658	0,140748	-0,018857	0,032692	**2550**	0,024256	-0,148973	-0,082977	0,055167	**V94**	0,132880	-0,174356	-0,012065	0,145559
1230	0,013892	0,485827	0,301124	-0,381893	**1460**	0,032694	0,220402	0,138878	-0,260687	**3120**	0,037473	-0,144714	-0,054766	0,031027	**V95**	0,123771	0,063605	0,247798	0,073935
1240	-0,873012*	0,012516	0,190354	0,197430	**1470**	-0,042174	0,141305	0,159918	-0,186225	**3130**	0,056633	-0,202122	-0,062216	0,088733	**V96**	0,057672	0,001851	0,024810	0,114411
1250	-0,298189	0,147720	0,142696	0,137460	**V41**	-0,040982	-0,098664	-0,010978	-0,159627	**3140**	0,064857	-0,143263	0,000058	-0,035220	**3270**	0,247391	0,232600	0,337249	-0,146059
1260	-0,020003	0,241756	0,335938	0,123557	**1480**	-0,892125*	0,184069	-0,065874	0,242172	**3150**	0,043247	-0,230778	-0,062117	-0,002576	**V98**	0,132710	0,217878	0,276782	-0,073318
1290	0,064628	0,217666	-0,243322	0,125376	**1490**	-0,049118	0,528233	-0,640666	-0,017673	**3170**	0,106131	0,101387	0,164445	0,132192	**V99**	0,160676	0,206910	0,322487	0,132404
V16	-0,171206	0,124027	-0,038881	0,075611	**1500**	0,008988	0,158114	0,086708	0,170264	**3180**	0,142559	0,113659	-0,231265	0,295110	**V100**	0,095177	0,087772	0,001129	0,093348
1300	-0,046827	0,545925	-0,693220	-0,103172	**1530**	0,031674	0,258230	0,001786	0,011096	**3190**	-0,213517	0,110652	-0,335979	0,221713	**3280**	0,249085	-0,016708	0,207476	-0,161986
V18	-0,023369	0,557285	-0,606587	0,037790	**1560**	0,110561	0,398546	0,328324	-0,129811	**V74**	0,181633	-0,300486	-0,089025	0,236803	**V102**	0,095358	0,298070	0,479137	-0,020830
1310	0,066637	0,308833	0,165812	0,106113	**V47**	0,062439	-0,058182	-0,024435	-0,023138	**V75**	0,105684	-0,181850	0,001753	0,166958	**3290**	-0,171993	0,592514	0,221233	-0,026967
1320	-0,097594	0,584129	-0,697440	-0,092180	**1570**	-0,055403	0,258515	0,142995	-0,785029*	**3200**	0,186865	-0,180151	-0,129620	-0,134669	**V104**	0,081360	0,279881	0,065166	0,046734
1330	0,011799	0,188201	-0,018981	0,076805	**1640**	-0,915970*	0,015758	0,167557	0,248647	**V77**	0,193636	0,225662	0,127791	0,258233	**3300**	0,118465	0,435893	0,012894	0,209245
V22	0,021845	0,015065	-0,063366	0,010046	**1700**	-0,718118*	-0,178323	-0,015406	-0,407113	**V78**	0,164187	0,214169	0,414651	0,164423	**V106**	0,091358	0,157990	0,240218	0,024530
1340	-0,053888	0,421716	0,328982	-0,277121	**1780**	-0,929185*	-0,015431	0,153984	0,222202	**3210**	0,286330	0,123918	0,252255	0,233486	**3310**	-0,025559	0,233390	-0,122210	-0,097917
1350	-0,335803	0,268505	0,122045	0,266868	**1800**	-0,874740*	0,023883	0,168499	0,186916	**V80**	0,220165	0,258903	0,373129	0,030393	**3320**	0,042441	0,185427	-0,117454	-0,117125
1360	0,082992	0,179954	0,089064	0,025771	**1820**	-0,201105	-0,209943	-0,128658	-0,686777	**V81**	0,162623	0,346899	0,332732	0,118896	**3940**	-0,932192*	-0,004344	0,158210	0,224250
V26	0,028336	0,206300	0,185760	-0,273768	**1840**	0,025778	-0,003333	-0,109567	0,011022	**V82**	0,075027	-0,016631	0,089649	0,046568	**4130**	0,020850	-0,080752	0,008522	0,038887
1370	0,111433	0,035574	0,089688	0,146133	**1850**	-0,066324	0,582871	-0,633738	-0,031808	**3220**	0,208419	0,244606	0,173185	0,191556					
1380	0,043137	0,307875	0,046298	0,320310	**1940**	-0,000862	0,048782	0,104821	-0,058625	**V84**	0,067745	0,310852	0,272520	0,058229					

De acuerdo con los resultados anteriores, se aprecia que la separación de los compuestos pertenecientes a las familias Verde y Almizcle es un factor común presente en las gráficas de dispersión de los valores de los componentes en las tres dimensiones. Para poder establecer la similitud de estos compuestos, se compararon los espectros infrarrojos de estos dos grupos.

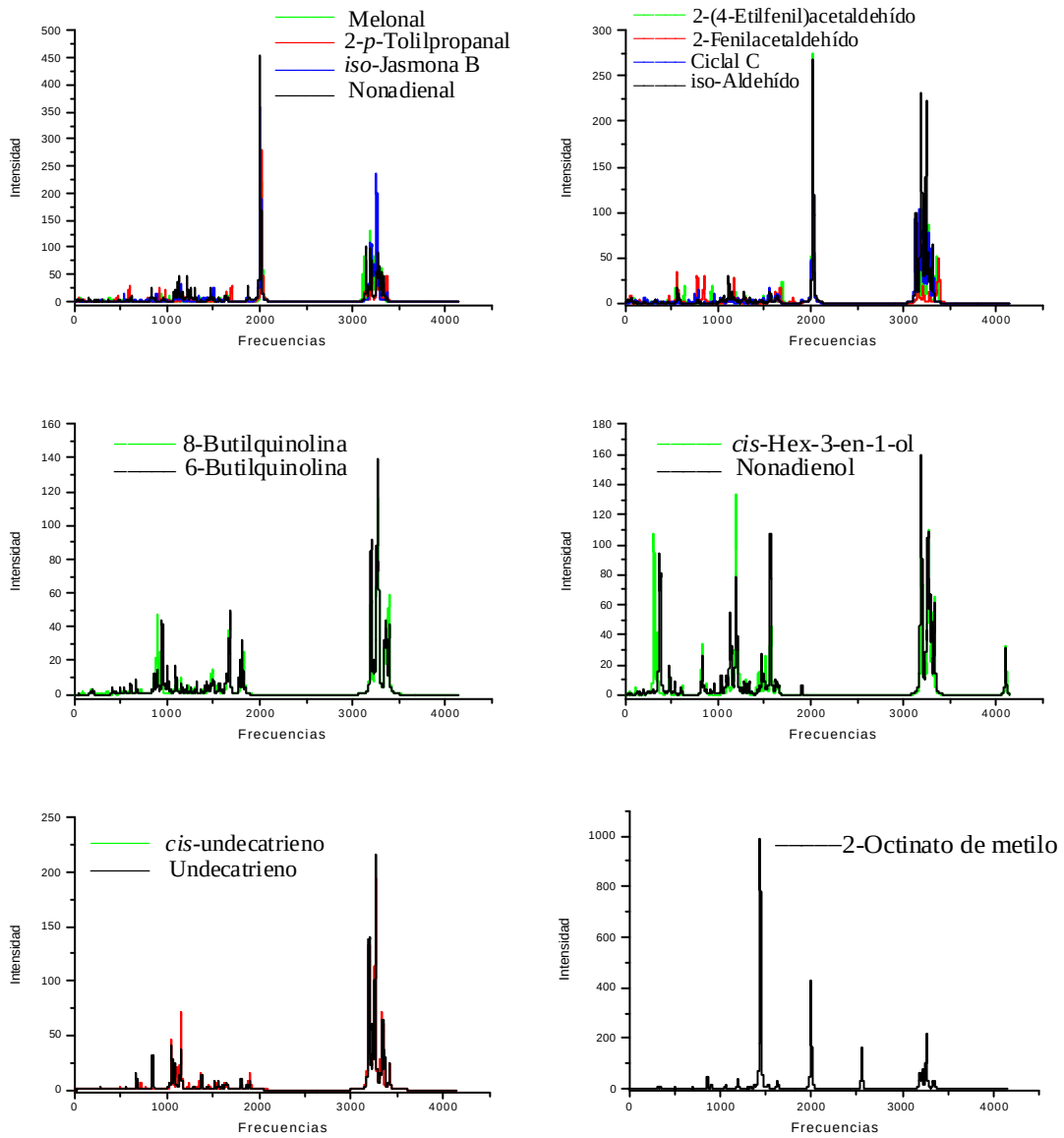


Figura 59. Espectros infrarrojo de los compuestos pertenecientes a la familia Verde.

Los compuestos que conforman la familia Verde están representados por 9 compuestos con un grupo funcional carboxílico (7 aldehídos, 1 cetona, 1 ácido carboxílico), 2 alcoholes, 2 hidrocarburos alifáticos, 2 heterociclos aromáticos, cuyos espectros infrarrojo se presentan en la **Figura 60**. De igual manera, los compuestos pertenecientes a la familia Almizcle se encuentran representados por 4 cetonas, 4 alcoholes, 3 ésteres, 2 heterociclos y 2 éteres (**Figura 60**).

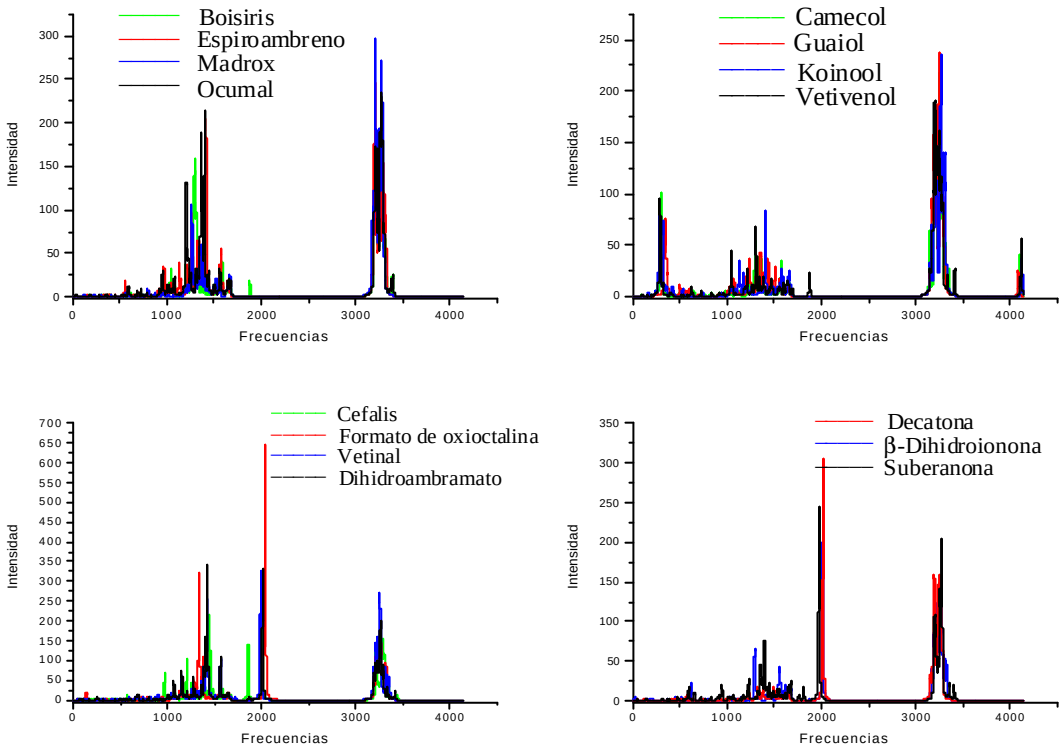


Figura 60. Espectros infrarrojo de compuestos pertenecientes a la familia Almizcle.

Se puede apreciar la similitud de los espectros infrarrojo de los compuestos con igual grupo funcional, como es el caso de los compuestos con un grupo carboxílico, donde se puede evidenciar una banda intensa común alrededor de 2000 cm^{-1}, correspondiente a la tensión C=O. Sin embargo, entre los compuestos con este mismo grupo funcional pertenecientes a distintas familias, se observa una diferencia en la intensidad de las bandas en las regiones de 1000-1700 cm^{-1}. Igualmente ello ocurre con los espectros IR

de los compuestos con un grupo –OH, aproximadamente en la región de 1300-1500 cm^{-1}.

De esta manera, se establece que la utilización de los modos vibracionales como únicos descriptores en la clasificación de los 60 compuestos odorantes bajo estudio, conduce a la separación de dos familias odorantes, *i.e.* Almizcle y Verde. Además, de acuerdo con las **Figuras 59 y 60**, se aprecia que no existe ningún patrón similar para todos los espectros infrarrojo de los compuestos pertenecientes a una misma familia; solamente se presentan espectros similares entre compuestos con un mismo grupo funcional. Sin embargo, cuando se analizan estos espectros entre compuestos pertenecientes a distintas familias, se observa que la diferencia en la clasificación se produce de acuerdo con las intensidades de las frecuencias.

Como se observa en los resultados de los valores propios obtenidos en el análisis de componentes principales anterior, los tres primeros factores principales solamente representan aproximadamente el 20% de la varianza total (**Tabla 33**). Por otro lado, se concluye que para poder explicar el 80% de la varianza, es necesario tener en cuenta los primeros 26 componentes principales.

Con el objetivo de aumentar el porcentaje de varianza explicado por los primeros factores, se realizó el análisis de los valores de contribución de cada uno de los modos vibracionales obtenidos en el análisis de PCA anterior (**Anexo 23**), teniendo en cuenta solamente los primeros 26 factores.

Como se muestra en la matriz de contribución de las variables, el mayor valor de contribución es del orden de 0.14. De esta manera, se fueron seleccionando las variables que presentaban valores de contribución superiores a cierto valor establecido, para posteriormente realizar el PCA sobre las variables seleccionadas hasta encontrar una solución donde los primeros factores principales contuvieran el 80% de la varianza total.

Inicialmente se estableció el valor de 0.01, y posteriormente este valor fue aumentado en 0.005 unidades en cada uno de los análisis subsiguientes, hasta alcanzar el objetivo trazado.

Utilizando como límite el valor de 0.09, se obtuvo un matriz de 60 x 12. La **Tabla 35** muestra los resultados del análisis de componentes principales, donde se observa que los tres primeros factores contienen el 80% de la varianza original.

Tabla 35. Resultados del análisis de componentes principales después de seleccionar los modos vibracionales que efectúan la mayor contribución a los factores.

Factor	Valor propio	% Total de varianza	% de Contribución
1	6,497892	54,14910	54,1491
2	1,969540	16,41283	70,5619
3	1,147166	9,55972	80,1217
4	1,035247	8,62706	88,7487
5	0,800950	6,67458	95,4233
6	0,203740	1,69783	97,1211
7	0,174192	1,45160	98,5727
8	0,105494	0,87912	99,4519
9	0,041782	0,34818	99,8000
10	0,017770	0,14808	99,9481
11	0,006226	0,05188	100,0000

De la tabla anterior se obtiene una solución de 4 factores, los cuales contienen el 89% de la información de las variables originales. Sin embargo, la representación de los 60 compuestos odorantes por medio de los tres primeros factores, no muestran una clasificación adecuada de acuerdo con su nota olfativa, ya que se presenta un agrupamiento de las diferentes moléculas odorantes (**Figura 61**).

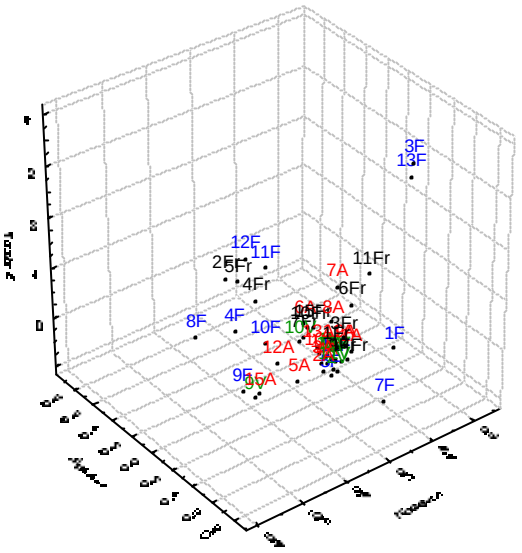


Figura 61. Representación de las 60 moléculas bajo estudio en el subespacio formado por los tres primeros PC (80% de la información).

A partir del resultado anterior, se puede establecer, que la utilización de los modos vibracionales como descriptores de las 60 moléculas bajo estudio, no presentan resultados favorables en su diferenciación, salvo en el análisis anterior, donde se aprecia la clasificación de dos familias odorantes: Almizcle y Verde, sin olvidar que los tres primeros componentes principales contienen muy poca información en relación con las variables originales.

5.2.10 Combinación de los descriptores moleculares con los modos vibracionales. Para poder completar el estudio de la clasificación de las moléculas odorantes, se procedió a realizar el análisis de los componentes principales de todos los descriptores obtenidos en el presente estudio.

A través de la combinación de todos los descriptores moleculares con las frecuencias, se obtuvo una matriz de 60 x 534. Los resultados alcanzados, una vez se realizó la clasificación por medio de PCA, muestran la clasificación de dos grupos fragantes,

principalmente, en la gráfica de los dos primeros componentes principales, el grupo Almizcle y el grupo Verde (**Figura 62**).

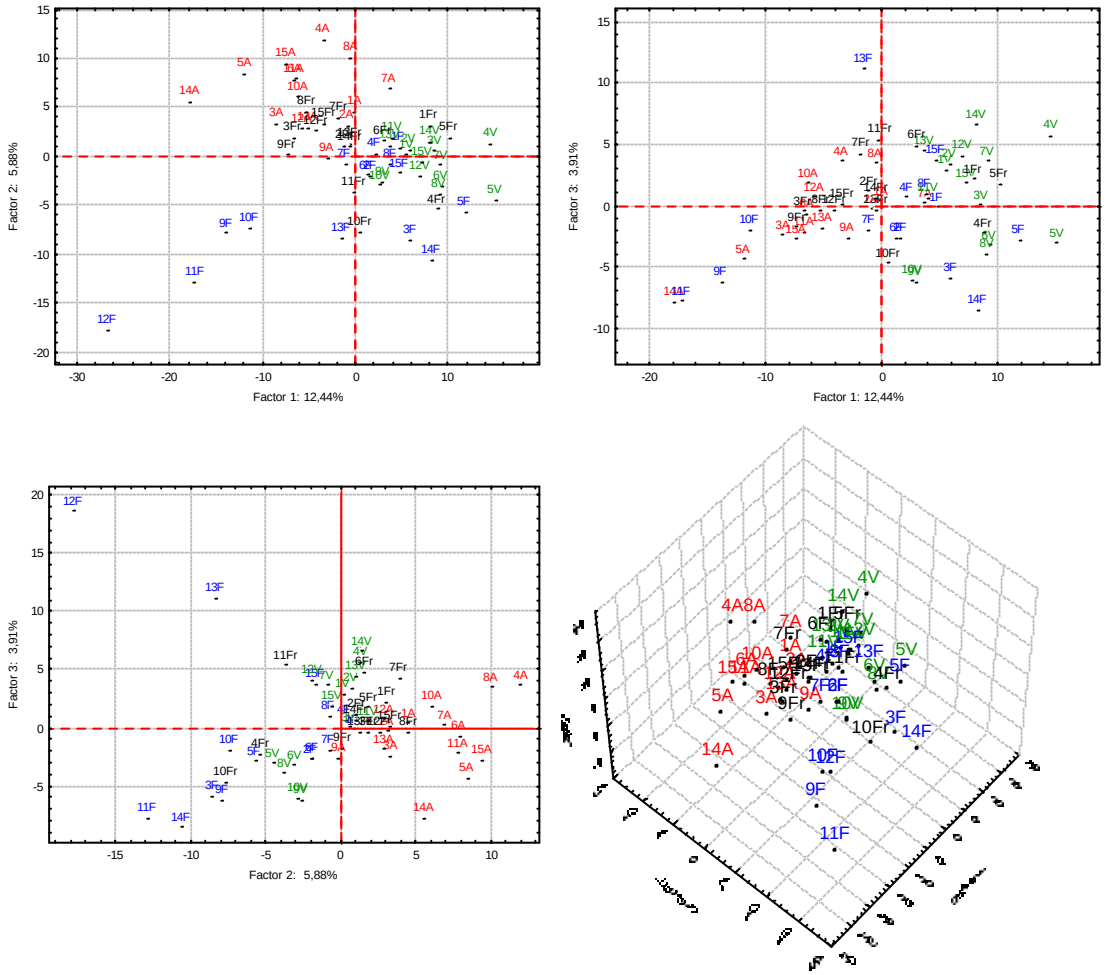


Figura 62. Gráficas de dispersión de las 60 moléculas odorantes, obtenidos por medio del PCA de todos los descriptores moleculares.

A partir de los resultados presentados en la **Tabla 36,** se aprecia, que aún cuando la transformación de los 534 descriptores originales en tres nuevas combinaciones lineales representa una gran reducción de variables, se pierde un gran porcentaje de la información original; ya que los tres componentes principales explican solamente el 22% de la varianza total de los datos originales.

Tabla 36. Valores propios obtenidos por medio del PCA.

Factor	Valor propio	% Total de varianza	% de Contribución	Factor	Valor propio	% Total de	% de Contribución
1	66,41838	12,43790	12,4379	**31**	6,83570	1,28009	78,1780
2	31,38654	5,87763	18,3155	**32**	6,51737	1,22048	79,3985
3	20,89873	3,91362	22,2291	**33**	6,37802	1,19439	80,5929
4	18,85065	3,53008	25,7592	**34**	6,25394	1,17115	81,7641
5	16,56346	3,10177	28,8610	**35**	5,92661	1,10985	82,8739
6	14,70044	2,75289	31,6139	**36**	5,82030	1,08994	83,9639
7	13,83220	2,59030	34,2042	**37**	5,65764	1,05948	85,0233
8	13,41437	2,51205	36,7163	**38**	5,54653	1,03868	86,0620
9	13,21205	2,47417	39,1904	**39**	5,26017	0,98505	87,0471
10	12,95444	2,42592	41,6163	**40**	5,15263	0,96491	88,0120
11	12,35104	2,31293	43,9293	**41**	4,88920	0,91558	88,9276
12	11,50191	2,15391	46,0832	**42**	4,71694	0,88332	89,8109
13	11,28443	2,11319	48,1964	**43**	4,61403	0,86405	90,6749
14	10,99402	2,05880	50,2552	**44**	4,31428	0,80792	91,4828
15	10,79841	2,02217	52,2774	**45**	4,09387	0,76664	92,2495
16	10,38766	1,94525	54,2226	**46**	4,00145	0,74934	92,9988
17	9,96042	1,86525	56,0879	**47**	3,78620	0,70903	93,7078
18	9,77721	1,83094	57,9188	**48**	3,65289	0,68406	94,3919
19	9,71641	1,81955	59,7383	**49**	3,52972	0,66100	95,0529
20	9,44136	1,76805	61,5064	**50**	3,42951	0,64223	95,6951
21	9,22364	1,72727	63,2337	**51**	3,32229	0,62215	96,3173
22	9,13477	1,71063	64,9443	**52**	3,09990	0,58051	96,8978
23	8,87584	1,66214	66,6064	**53**	3,09217	0,57906	97,4768
24	8,61574	1,61344	68,2199	**54**	2,85039	0,53378	98,0106
25	8,54483	1,60016	69,8200	**55**	2,52379	0,47262	98,4833
26	7,91015	1,48130	71,3013	**56**	2,48298	0,46498	98,9482
27	7,78653	1,45815	72,7595	**57**	2,15655	0,40385	99,3521
28	7,74992	1,45130	74,2108	**58**	1,74299	0,32640	99,6785
29	7,24273	1,35632	75,5671	**59**	1,71692	0,32152	100,0000
30	7,10674	1,33085	76,8979				

Con base en los resultados anteriores, se decidió investigar la combinación de los modos vibracionales con los descriptores obtenidos por medio de la reducción de datos cuando se combinaron cuatro o menos tipos de descriptores. El motivo por el cual se tomaron como base estos descriptores, se debe a los excelentes resultados obtenidos en la clasificación de las familias odorantes.

Una vez realizada la clasificación por medio del PCA, se observa, que los resultados fueron similares; ya que en todos los análisis se presentó un mismo patrón de clasificación. Cabe mencionar, que los resultados obtenidos cuando se utilizaron las

frecuencias retenidas en el análisis basado en la desviación estándar, siempre fueron superiores a los obtenidos por los análisis basados en el valor de la intensidad. Debido a esta razón se presentan a continuación solamente estos resultados.

Los siguientes resultados se obtuvieron por medio de la combinación de los modos vibracionales obtenidos cuando se impuso un valor límite de 14 en la desviación estándar, con los descriptores resultantes de la combinación de los cuatro tipos de descriptores en el Método A.

Tabla 37 Resultados del análisis de componentes principales.

Factor	Valor propio	% Total de varianza	% de Contribución	Factor	Valor propio	% Total de	% de Contribución
1	10,05833	8,111553	8,1116	**31**	1,39331	1,123639	86,9872
2	8,81654	7,110114	15,2217	**32**	1,33877	1,079654	88,0669
3	7,29684	5,884545	21,1062	**33**	1,24906	1,007306	89,0742
4	6,55493	5,286236	26,3924	**34**	1,19761	0,965814	90,0400
5	5,39744	4,352771	30,7452	**35**	1,11039	0,895476	90,9355
6	4,53412	3,656551	34,4018	**36**	1,03581	0,835330	91,7708
7	4,39788	3,546680	37,9484	**37**	1,00813	0,813010	92,5838
8	4,20008	3,387161	41,3356	**38**	0,87243	0,703570	93,2874
9	4,03705	3,255687	44,5913	**39**	0,84464	0,681161	93,9685
10	3,83022	3,088887	47,6802	**40**	0,80553	0,649618	94,6181
11	3,63531	2,931702	50,6119	**41**	0,73371	0,591700	95,2098
12	3,46625	2,795360	53,4072	**42**	0,71173	0,573975	95,7838
13	3,37906	2,725048	56,1323	**43**	0,61728	0,497807	96,2816
14	3,21669	2,594104	58,7264	**44**	0,54433	0,438976	96,7206
15	2,88363	2,325509	61,0519	**45**	0,51931	0,418797	97,1394
16	2,82032	2,274451	63,3264	**46**	0,50137	0,404330	97,5437
17	2,66634	2,150276	65,4766	**47**	0,44625	0,359877	97,9036
18	2,59031	2,088960	67,5656	**48**	0,36975	0,298183	98,2018
19	2,50697	2,021747	69,5873	**49**	0,36674	0,295759	98,4976
20	2,34461	1,890814	71,4782	**50**	0,34345	0,276977	98,7745
21	2,20844	1,781003	73,2592	**51**	0,32544	0,262454	99,0370
22	2,11734	1,707530	74,9667	**52**	0,25923	0,209060	99,2460
23	1,97901	1,595976	76,5627	**53**	0,22155	0,178669	99,4247
24	1,88388	1,519255	78,0819	**54**	0,18594	0,149954	99,5747
25	1,82159	1,469024	79,5509	**55**	0,16518	0,133211	99,7079
26	1,72579	1,391768	80,9427	**56**	0,13922	0,112271	99,8201
27	1,65775	1,336898	82,2796	**57**	0,12767	0,102958	99,9231
28	1,55729	1,255880	83,5355	**58**	0,05672	0,045744	99,9688
29	1,46011	1,177511	84,7130	**59**	0,03863	0,031153	100,0000
30	1,42670	1,150568	85,8636				

En la **Tabla 37** se muestran los resultados obtenidos por medio de PCA sobre la matriz de 60 x 124. De acuerdo con estos valores, se presenta un ligero aumento en la varianza total descrita por los tres primeros PC, en comparación con los resultados conseguidos en los análisis anteriores, donde solamente se utilizaron como variables las frecuencias.

Los tres primeros factores principales contienen más del 21% de la varianza total de los datos originales. Las gráficas de dispersión de los factores de coordenadas de las moléculas permiten apreciar una separación de dos grupos odorantes, *i.e.* Almizcle y Verde. También se observa el comportamiento anómalo de los compuestos 14F y 11F por las mismas razones explicadas anteriormente.

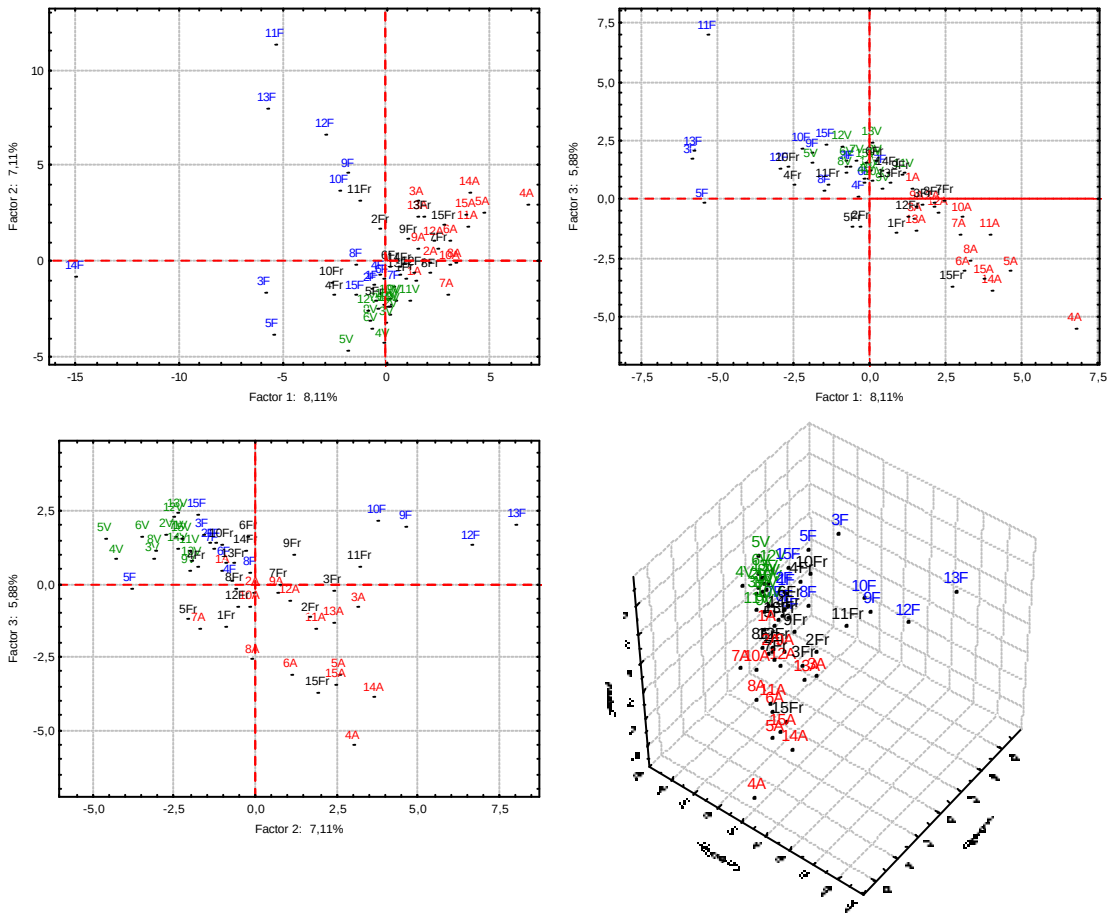


Figura 63. Gráficas de dispersión de las 60 moléculas odorantes, obtenidos por medio del PCA de los modos vibracionales y los resultados basados en la desviación estándar.

Estas gráficas muestran que los tres factores principales contribuyen con la separación de dos de los grupos odorantes, principalmente, el grupo Almizcle y el grupo Verde. Sin embargo, también se aprecian dos pequeños grupos, el primero conformado por seis compuestos de la familia Floral, mientras que los compuestos restantes se encuentran dispersos en la gráfica y el segundo grupo, por cinco compuestos de la familia Frutal; estos dos pequeños grupos se encuentran justamente encima del grupo Verde. El hecho, que no se separen adecuadamente todos los grupos odorantes, se puede explicar de acuerdo con los valores obtenidos de los factores de aporte de las variables, los cuales solamente presentan valores de correlación mayores de 0,7 para las variables FNSA-1 en el primer PC y en la frecuencia 1640 cm^{-1} para el Factor 3.

5.2.11 Análisis de agrupamiento sobre los modos vibracionales. Con la idea de observar a través de la metodología de agrupamiento si las frecuencias podían servir como descriptores de los diferentes compuestos fragantes en la clasificación de las cuatro familias odorantes bajo estudio, se realizó el análisis de agrupamiento sobre la matriz de frecuencias (60 x 442) obtenida por medio de los cálculos computacionales.

El dendograma obtenido por medio del método de Ward (**Figura 64**), muestra solamente la agrupación de unas cuantas moléculas pertenecientes a los grupos Almizcle y Frutal. Debido a esto, se procedió a estudiar las diferentes matrices obtenidas a través de la reducción de variables en la sección anterior.

Los dendogramas resultantes por la clasificación de las matrices reducidas por medio de la retención de los valores promedios para cada rango de variable, así como la retención de los valores mayores, se pueden apreciar en el **Anexo 24**.

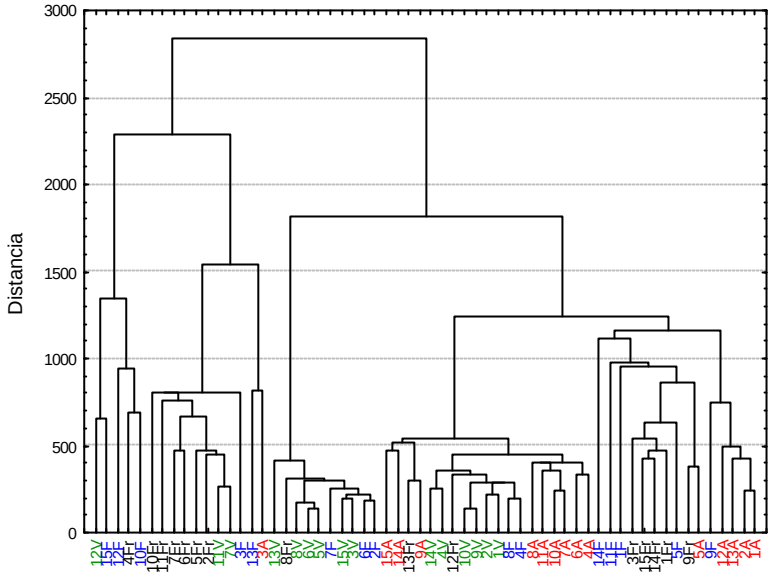


Figura 64. Dendograma obtenido por medio del agrupamiento de todos los valores de frecuencia.

De igual forma, se realizaron los análisis de agrupamiento sobre las matrices reducidas cuando se impuso valores límites en la intensidad y en la desviación estándar, junto con la combinación de estas matrices con los diferentes descriptores. Sin embargo, los dendogramas obtenidos a través de estos análisis fueron iguales (**Figura 65)**.

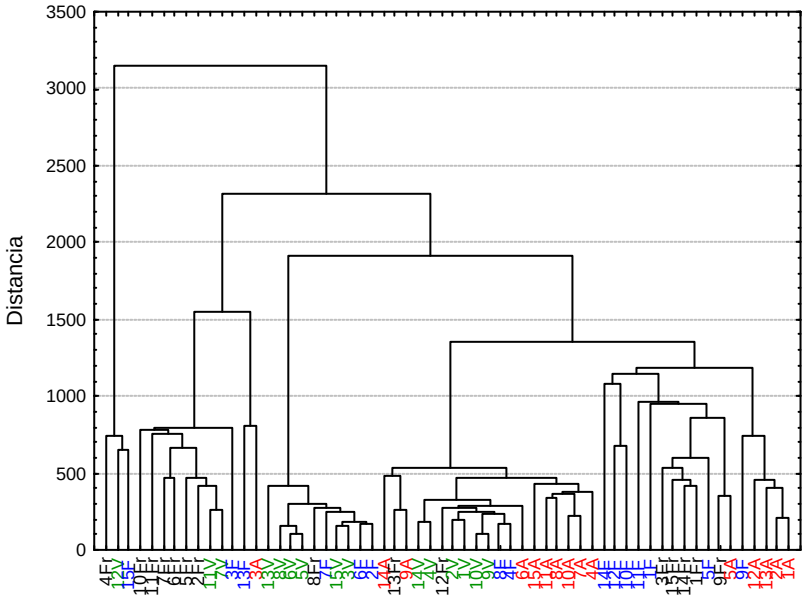


Figura 65. Dendograma obtenido por medio del método de Ward.

Por medio de los resultados obtenidos utilizando el análisis de agrupamiento sobre los modos vibracionales, así como de la combinación de las frecuencias con diferentes descriptores moleculares, al igual que en el caso del análisis de componentes principales, establecen que la utilización de los diferentes modos vibracionales no conducen a una adecuada clasificación de los compuestos odorantes bajo estudio.

En resumen, se establece que la clasificación de los 60 compuestos odorantes bajo estudio en el presente trabajo, se puede realizar por medio del análisis de componentes principales así como por medio del análisis de agrupamiento utilizando el método de Ward.

La utilización de diferentes descriptores moleculares como los descriptores electrónicos, topológicos, geométricos y CPSA, conducen a la clasificación de tres familias odorantes, a saber: Almizcle, Floral y Verde, como se observa en los resultados obtenidos.

6. RELACIONES ESTRUCTURA-OLOR

6.1 DESCRIPTORES ELECTRÓNICOS

De acuerdo con los resultados de clasificación de los compuestos odorantes bajo estudio, utilizando los descriptores electrónicos obtenidos por medio de los cálculos computacionales, se observa, que la representación de las moléculas fragantes por medio del subespacio formado por los tres primeros PC (99% de la información), permite clasificar las familias Almizcle, Verde y Floral (**Figura 32**).

Por otro lado, con base en los resultados de los factores de aporte de los descriptores (**Tabla 17**), se establece, que el Factor 1 representa la relación entre las energías del HOMO y del LUMO, mientras que el Factor 2, representa la energía del punto cero.

Como el Factor 1 separa las moléculas de la familia Floral de los compuestos restantes, debe existir una relación entre estos compuestos con los valores obtenidos para los orbitales HOMO y LUMO. De igual manera, de acuerdo con la separación de los compuestos de las familias Almizcle y Verde, por medio del Factor 2, debe existir alguna relación con los valores de la energía del punto cero y estos compuestos.

Si se observa la **Tabla 3**, donde se presentan los valores de los descriptores electrónicos, se aprecia, que los compuestos de la familia Floral son los únicos que presentan valores positivos de energía del HOMO y valores negativos de energía del LUMO, motivo por el cual este factor realiza la separación de estas moléculas.

Por otro lado, los valores obtenidos de energía del punto cero para los compuestos de la familia Almizcle son los valores más altos, y se encuentran en el rango de 0.316-0.467, mientras que los valores calculados para las moléculas de la familia Verde presentan valores entre 0.148-0.310. La frontera de separación es el Factor 2, ya que los compuestos odorantes tipo Almizcle se encuentran en el segundo cuadrante, donde el

Factor 2 adquiere valores positivos y el Factor 1 valores negativos, mientras que los compuestos odorantes de la familia Verde, se encuentran en el tercer cuadrante, a excepción del compuesto 13V, que presenta el valor de energía de punto cero más elevado (0.310) y se clasifica junto con los compuestos de la familia Almizcle (**Figura 33**).

Por último, los valores de ZPE de las moléculas de nota Frutal, que se clasifican con los compuestos de la familia Verde o con los de la familia Almizcle, se encuentran entre los rangos encontrados para estos compuestos.

6.2 DESCRIPTORES TOPOLÓGICOS

La gráfica de dispersión de los factores de coordenadas de las 60 moléculas (**Figura 35)** muestra la clasificación de los compuestos pertenecientes a las familias Almizcle y Verde. La separación de los compuestos pertenecientes a estas familias odorantes, se realiza de acuerdo con el Factor 1, donde los compuestos de la familia Almizcle se encuentran en la región de los valores positivos, mientras que los compuestos del grupo Verde se ubican en la región de los valores negativos.

El Factor 1 se encuentra relacionado con el índice de Kier y Hall (orden 3) o índice de conectividad de valencia, que de acuerdo con su definición, considera la presencia de heteroátomos y la hibridación de los átomos presentes en la molécula. A partir de la magnitud de los valores obtenidos para los compuestos de las familias Almizcle y Verde, se aprecia que los valores de los compuestos del grupo Almizcle son los valores más altos, 3.376-6.306, mientras que el rango de los valores determinados para los compuestos de la familia Verde es 0.698-2.808 (**Anexo 1**).

6.3 DESCRIPTORES ELECTRÓNICOS Y TOPOLÓGICOS

Método A

Los resultados obtenidos por medio de la clasificación de la combinación de los descriptores electrónicos y topológicos, cuando se utilizó el Método A, muestran la clasificación de tres familias odorantes a saber: Almizcle, Floral y Verde; además, el agrupamiento de la mayoría de los compuestos del grupo Frutal en el origen de la gráfica de dispersión de las coordenadas de las 60 moléculas, representadas por los dos primeros factores (68% de la información) (**Figura 36**).

De acuerdo con los factores de aporte, el Factor 1 se encuentra relacionado con la energía HF y con el contenido de información estructural (orden 0). El Factor 1, separa los compuestos de la familia Verde de las familias Almizcle, Frutal y Floral. Sin embargo, esta separación no es completa, ya que los compuestos 8F, 5F, 7A, 8A, 1Fr, 4-6Fr, se encuentran mezclados con los de la familia Verde.

Los magnitud de la energía HF, muestran que los compuestos que se encuentran en la región de valores positivos del Factor 1 presentan los valores de energía HF más altos entre -309,024 y -555,488, a excepción de los compuestos 14 y 15F. Esta clasificación corresponde a las moléculas de la familia Verde junto con los compuestos mencionados anteriormente. Inversamente, estos compuestos presentan los menores valores obtenidos para el descriptor d32 [Contenido de información estructural (orden 0)].

Por otro lado, el Factor 2 separa los compuestos de la familia Almizcle de los compuestos Verde y Floral. Este factor se encuentra mayormente relacionado con el índice de Kier y Hall (orden 3), sin embrago, este descriptor también presenta un valor de correlación significativo con el Factor 1 (-0.5).

La observación de los valores obtenidos para este descriptor, muestra que los compuestos del grupo Almizcle presentan los valores más altos, junto con la mayoría de

los compuestos de la familia Frutal. Estos compuestos se clasifican en la región de los valores negativos del segundo factor (**Figura 36**).

Método B

Los resultados obtenidos por el PCA de los descriptores electrónicos y topológicos realizan una clasificación similar a la establecida por medio del Método A (**Figura 36**).

Al igual que en el caso anterior, el Factor 1 se encuentra relacionado con la energía HF, pero a diferencia del Método A, presenta un valor de correlación con el Factor 1 negativo. El Factor 1, separa los compuestos de la familia Verde de la mayoría de las moléculas pertenecientes a las familias Almizcle, Frutal y Floral, a excepción del compuesto 13V; sin embargo, en contraste con el método anterior, estos compuestos se clasifican en la zona negativa del primer factor.

El Factor 2 separa los compuestos de la familia Almizcle de los compuestos de las familias Verde y Floral. Este factor se encuentra mayormente relacionado con el contenido de información complementaría (orden 2). La observación sobre los valores de este descriptor para los diferentes compuestos odorantes, establece que los compuestos presentes en la región de valores negativos del Factor 2 corresponden a los mayores valores y se encuentran entre el rango 50.79-138.14.

6.4 DESCRIPTORES ELECTRÓNICOS, TOPOLÓGICOS Y GEOMÉTRICOS

Método A

Los resultados del PCA muestran que el primer componente principal separa el grupo Verde, pero de igual forma como se menciono en el caso de los descriptores electrónicos y topológicos, no se realiza una separación completa, ya que se pueden observar la presencia de los compuestos 3F, 14F 8F, 5F, 7A, 8A, 1Fr, 4-6Fr. Este factor se encuentra mayormente relacionado con el descriptor d32 [contenido de información estructural (orden 0)]. Además, se observa la misma relación establecida entre los

compuestos presentes en el primer cuadrante del subespacio conformado por los dos primeros componente principales (62% de la información) y los valores calculados para del descriptor d32, mencionada en el caso anterior.

El segundo factor, el cual se encuentra correlacionado con los descriptores d92 (energía del LUMO) y d26 [índice de Kier y Hall (orden 3)], separa especialmente al grupo Almizcle del grupo Floral. La separación de la familia Floral, se debe especialmente a los valores de energía del LUMO calculados para estas moléculas, ya que son los únicos valores que presentan un valor negativo.

Por otro lado, los compuestos que se clasifican en la región de los valores negativos del segundo factor, presentan los valores del descriptor d26 más altos. Estos compuestos corresponden a la familia Almizcle, junto con la mayoría de los compuestos de la familia Frutal.

Método B

El patrón de separación encontrado en la **Figura 40**, es similar al encontrado en el Método A, sin embargo, se puede apreciar una inversión de las coordenadas de los compuestos separados por el Factor 2, en comparación con las gráficas del método anterior.

Los factores de aporte presentados en la **Tabla 21**, muestran que los descriptores que se encuentran correlacionados con el Factor 1 son d23 [índice de Kier y Hall (orden 0)] y d67 (momento de inercia B). Por medio de la observación de los valores obtenidos para el descriptor d23, de las moléculas pertenecientes a la familia Verde, se aprecia, que estos compuestos presentan los valores más bajos, junto con los valores de algunos compuestos de las familias Frutal y Floral.

El segundo factor, el cual se encuentra correlacionado con los descriptores d92 (energía del LUMO), d91 (energía del HOMO) y d39 [contenido de información complementaria (orden 2)], realiza la misma separación encontrada en el Método A.

Por otro lado, el cambio de cuadrante de los compuestos de la familia Floral y Almizcle, puede ser debido a la inversión del signo de los factores de aporte de los descriptores d91 y d92 (**Tabla 21**).

6.5 DESCRIPTORES ELECTRÓNICOS, TOPOLÓGICOS, GEOMÉTRICOS Y CPSA

Los resultados de la representación de los 60 compuestos fragantes por medio de los tres primeros factores principales (68% de la información), cuando se utilizó la combinación de los descriptores electrónicos, topológicos, geométricos y CPSA, presentan la clasificación de las familias Almizcle, Floral y Verde.

Sin embargo, los resultados cuando se utilizó la matriz de variables reducida por medio del Método A, son superiores a los obtenidos a partir del Método B. Debido a lo anterior, solamente se discuten las relaciones “estructura-olor” para los resultados obtenidos por medio de esta metodología.

La **Figura 42** muestra que el primer componente separa los compuestos de la familia Floral, exceptuando los compuestos 8F y 4F, de las familias Verde y Almizcle, principalmente. Estas moléculas se clasifican en la región de valores negativos del primer componente. El Factor 1 se encuentra mayormente correlacionado con los descriptores d76 (PNSA-2), mayor valor positivo y d91 (energía HOMO) mayor valor negativo. Sin embargo, el descriptor d91 presenta valores de aporte significativamente bajos sobre los Factores 2 y 3. A partir de la **Tabla 3,** se establece la relación entre los compuestos de la familia Floral con los valores de la energía del HOMO, cuyos valores

son positivos, a diferencia de los valores determinados para los demás compuestos odorantes.

Por otro lado, el Factor 2 separa la familia Almizcle de la verde, y se encuentra mayormente correlacionado con los descriptores d26 [índice de Kier y Hall (orden3)], mayor valor negativo y d66 (Momento de inercia A) mayor valor positivo. Las magnitudes calculadas para estos descriptores, muestran que las moléculas pertenecientes a la familia Almizcle presentan los valores más altos y se encuentran entre el rango 3,376-6,036, mientras que los valores de los descriptores de la familia Verde, se encuentran por debajo de 2,808. De la misma manera, los valores del momento de inercia A entre estas familias difieren de una manera notoria, ya que los valores pertenecientes a los compuesto de la familia Verde son los más altos, y se encuentra entre el rango 0.064-0.216, mientras que los valores arrojados por los compuestos de la familia Almizcle se encuentran entre 0.018-0.057.

Además, en la **Figura 42** se observa, que los compuestos de la familia Verde presentan factores de coordenada positivos respecto al segundo factor, mientras que los compuestos de la familia Almizcle presentan valores negativos.

Por último, aún cuando el descriptor d83 (FPSA-3) presenta un valor de correlación alto con el Factor 3 y que además no presenta valores significativos de correlación los otros primeros factores, no influye en la clasificación de los compuestos fragantes.

Con base en los resultados anteriores, se establece que en los diferentes subespacios establecidos a partir de los descriptores, las energías del HOMO y del LUMO son descriptores adecuados para realizar la clasificación de los compuestos de la familia Floral. Por otro lado, la separación de los compuestos pertenecientes a las familias Almizcle y Verde puede efectuarse de acuerdo con varios descriptores, entre éstos, se pueden mencionar el índice de Kier y Hall (orden 3), el momento de inercia A y el contenido de información complementaria (orden 2).

6.6 MODOS VIBRACIONALES

Como se mencionó en la sección anterior, los resultados obtenidos por medio del análisis de componentes principales utilizando los modos vibracionales, conducen a la clasificación de los compuestos pertenecientes a las familias Verde y Almizcle.

Los resultados de la **Figura 57**, muestran que el Factor 2 realiza la separación de las familias fragantes mencionadas anteriormente, sin embargo, los valores de los factores de aporte de cada una de las frecuencias **(Tabla 34),** exponen que este factor no se encuentra correlacionado con ninguno de los modos vibracionales. Por tal razón, no se puede establecer una relación entre la nota olfativa y el espectro infrarrojo de las moléculas pertenecientes a cada una de los grupos fragantes.

7. CONCLUSIONES

- La clasificación de los 60 compuestos odorantes por medio del análisis de componentes principales cuando se utiliza una sola clase de descriptores, presenta buenos resultados con los descriptores electrónicos y topológicos. Igualmente, cuando se realiza la combinación de dos o más clases de descriptores, se observa que los resultados de clasificación son satisfactorios cuando se involucran los descriptores electrónicos y no se tienen en cuenta los descriptores constitucionales.

- La combinación de los descriptores, electrónicos y topológicos, junto con los descriptores electrónicos y geométricos, conduce a la clasificación de tres familias odorantes, Verde, Almizcle y Floral. En ambos casos, se aprecia un patrón de clasificación general.

- La combinación de los descriptores electrónicos, topológicos y geométricos, conduce a la clasificación de las familias Almizcle, Verde y Floral, por medio de la representación gráfica de los dos primeros componentes principales. En este caso, el primer componente principal separa los grupos Almizcle y Verde, mientras que el segundo factor separa especialmente al grupo Almizcle del grupo Floral.

- El subespacio representado por los dos primeros componentes principales obtenido por medio de la combinación de los descriptores electrónicos, topológicos, geométricos y CPSA, permite la diferenciación de las familias fragantes Almizcle, Verde y Floral. El patrón de separación de los diferentes compuestos es similar a los obtenidos en los análisis anteriores.

- Por medio de las gráficas de clasificación obtenidas de la combinación de diferentes tipos de descriptores, se establece, que los resultados adquiridos a través de la reducción de variables utilizando la metodología A, son superiores a los obtenidos cuando se reducen las variables por medio de la metodología B.

- Al igual que en el caso del análisis de componentes principales, el análisis de agrupamiento sobre la combinación de los descriptores anteriormente establecidos, produce la separación de las familias almizcle, verde y floral. Por lo tanto, se puede establecer que el método de Ward es adecuado en la clasificación de los compuestos odorantes bajo estudio.

- La implementación de la metodología de Kelley utilizada en la selección del número de grupos presentes en los diferentes dendogramas, produce un número elevado de estos, al igual que grupos representados por un solo compuesto, comparada con la metodología manual.

- El análisis de componentes principales por medio de la utilización de los modos vibracionales como descriptores de las 60 moléculas fragantes, conducen a la clasificación de dos familias, *i.e.* Almizcle y Verde. Sin embargo, los tres primeros componentes solamente contienen el 20% de la información de los datos originales. Si se busca aumentar el porcentaje de información de los datos retenidos, se pierde la diferenciación alcanzada anteriormente.

- Los dendogramas obtenidos a través del análisis de agrupamiento por medio del método de Ward, cuando se utilizan los modos vibracionales como descriptores de los 60 compuestos odorantes, no producen una clasificación adecuada de las familias fragantes.

- Los espectros vibracionales obtenidos para cada uno de los 60 compuestos odorantes, no presentan alguna relación “estructura-olor” con las moléculas pertenecientes a cada uno de las familias Almizcle, Verde, Frutal y Floral.

- Se establece, que los descriptores de energías del HOMO y del LUMO son descriptores adecuados para realizar la clasificación de los compuestos de la familia Floral. Además, existe una relación entre los valores de estos descriptores con su nota olfativa.

- De igual manera, los compuestos pertenecientes a las familias Almizcle y Verde presentan una relación con los descriptores índice de Kier&Hall (orden 3), el momento de inercia A y el contenido de información complementaria (orden 2).

8. RECOMENDACIONES

- Introducir nuevos descriptores moleculares en la investigación de la clasificación de los compuestos odorantes por medio del análisis de componentes principales y del análisis de agrupamiento.

- Incluir en el estudio de clasificación otras familias odorantes, y de esta manera determinar las diferentes relaciones estructura-olor presentes en una amplia selección de compuestos fragantes.

- Utilizar diferentes conjuntos de base en la optimización de la geometría de los compuestos fragantes, para establecer una comparación entre los resultados obtenidos por diferentes métodos computacionales.

- Introducir diferentes métodos de agrupamiento en el desarrollo del análisis de grupos y estudiar la diferencia entre los resultados obtenidos por medio de las técnicas jerárquicas y no jerárquicas.

- Realizar el cálculo del número de grupos de los diferentes dendogramas obtenidos por medio se otros métodos de parado (*stopping rules*), y de esta manera, poder determinar cual método de selección de grupos presenta los mejores resultados en el estudio de clasificación de los compuestos fragantes.

9. BIBLIOGRAFÍA

[1] BALBES L. M., MASCARELLA S. W., BOYD D. B, "A perspective of modern methods in computer-aided drug design", *Rev. Comput. Chem.*, **1994**, 5, 337-379.

[2] ROSSITER, K. J., "Structure-Odor Relationships", *Chem. Rev.*, **1996**, 96, 3201-3240.

[3] DYSON G.M. "The scientific basis of odour", *Chem. Ind.*, **1938**, 57, 647-651.

[4] TURIN L., "A spectroscopic mechanism for primary olfactory reception", *Chem. Senses,* **1996**, 21, 773–791.

[5] THEIMER E., DAVIES J. T., "Olfaction, Musk odor, and molecular prperties", *J. Agric. Food Chem*. **1967**, 15, 6-14.

[6] AMOORE J. E., "Molecular Basis of Odor", **1970**, Springfield IL: Thomas.

[7] DRAVNIEKS A., LAFFORT P., "Physicochemical basis of quantitative and qualitative odor discrimination in humans. In olfaction and taste", Schneider, D., Ed. Wissens-Verlag-MBH: Stuttgart, FRG, 1972, pp. 14-148.

[8] JOHN C. LEFFINGWELL "Olfaction", *Leffingwell Reports*, **2002**, 2, 1-34.

[9] OHLOFF, G., "Scent and Fragrances", Springer-Verlag, Berlin, Heidelberg, **1994**, pp. 1-9.

[10] MORRISON E. E., MORAN D. T., "In hand book of olfaction and gustation", Doty, R. L., Ed. Marcel Dekker, Inc., New York, 1995, pp. 75-101.

[11] BREER H., "The molecular basis of smell and transduction", Wiley; Chichester, **1993**, pp. 97-114.

[12] DIONNE V. E., DUBIN A. E., "Transduction diversity in olfaction", *J. Exp. Biol.*, **1994**, 194, 1-21.

[13] SHEPHERD G. M., "Discrimination of molecular signals by the olfactory receptor neuron", *Neuron*, **1994**, 13, 771-790.

[14] BUCK, L., AXEL, R., "A novel multigene family may encode odorant receptors- a molecular basis for odour recognition", *Cell*, **1991**, 65, 175-187.

[15] GOLDBERG S., TURPIN J., PRICE S., "Anisole binding protein from olfactory epithelium evidence for a role in transduction", *Chem. Senses & Flavour*, **1979**, 4, 207.

[16]. LACAZETTE E, GACHON A. M., PITIOT G., “A novel human odorant-binding protein gene family resulting from genomic duplicons at 9q34: differential expression in the oral and genital spheres”, *Hum. Mol. Genet.*, **2000**, 2, 289-301.

[17] KLOPPING H. L. “Olfactory Theories and the Odors of Small Molecules”, *J. Agr. Food Chem.*, **1971**, 19, 999-1004.

[18] BUCK L., “Identification and analysis of a multigene family encoding odorant receptors: implications for mechanisms underlying olfactory information processing”, *Chemical Senses*, **1993**, 18, 203-208; NGAI J, DOWLING M. M., BUCK L, AXEL R., CHESS A., “The family of genes encoding odorant receptors in the channel catfish”, *Cell*, **1993**, 72, 657-666; RAMING K, KRIEGER J, STROTMANN J, BOEKHOFF I, KUBICK S, BAUMSTARK C., BREER H., “Cloning and expression of odorant receptors”, *Nature*, **1993**, 361, 353-356.

[19] PAULING L., “Molecular architecture and Biological Reactions”, *Chem. Eng. News*, **1946**, 24, 1375.

[20] MONCRIEFF R. W., Chemical Senses, **1946,** New York John Wiley,

[21] MORI K., SHEPHERD G. M., “Emerging principles of molecular signal processing by mitral / tufted cells in the olfactory bulb”, *Semin. Cell Biol.*, **1994**, 5, 65-74.

[22] MALNIC B., HIRONO J., SATO T., BUCK L. B. “Combinatorial receptor codes for odors”, *Cell*, **1999**, 96, 713–723.

[23] WRIGHT R. H., “Odor and molecular vibration: neural coding of olfactory information”, *J. Theor. Biol.*, **1977**, 64, 473–502.

[24] BOELENS M. H., VAN GEMERT L. J., “Sensory properties of optical isomers”, *Perf. Flav*., **1993**, 18.

[25] JAKLEVIC R. C., LAMBE J, “Molecular vibration spectra by electron tunneling”, *Phys. Rev. Lett*., **1966**, 17, 1139-1140.

[26] TURIN L., YOSHII F., “Structure–odor relations: a modern perspective. In: Handbook of olfaction and Gustation” (Doty, R., ed.)., New York, Marcel Dekker, **2002**.

[27] KLOPPING H. L., “Olfactory theories and the odors of small molecules”, *J. Agric. Food Chem.*, **1971**, 19, 999-1004.

[28] STOCK A, MASSENEZ C., "Boron Hydrides", *Berichte*, **1913**, 45, 3539-3568.

[29] WANNAGAT U., DAMRATH V, HUCH V, VEITH M., HARDER U., "Sila-Riechstoffe und Riechstofffisostere XII. Geruchsvergleiche homologer organoelementverbindungen der vierten hauptgruppe (C,Si, Ge, Sn)", *J. Organom. Chem.*, **1993**, 443, 153-165.

[30] http://www.leffingwell.com/chirality/chirality.htm, consultada el 15 julio del 2004.

[31] WEYERSTAHL P., "Odor and structure", *Journal für Pratische Chemie-Chemiker Zeitung*, **1994**, 336, 95-109.

[32] ADAMS M. J., "Chemometrics in analytical spectroscopy", RSC Analytical Spectroscopy Monographs, 1995, pp. 54-115.

[33] VAN DE WATERBEEMD H., TAYAR E., CARRUPT P. A., TESTA B., *J. computer-aided Molec. Design*, **1989**, 3, 111-132.

[34] CALVO R. A., PARTRIDGE M., JABRI M. A., "A Comparative Study of Principal Component Analysis Techniques", Department of Electrical Engineering Universida de Sydney, Australia, **2006**, pp. 1-5.

[35] TIPPING M. E., BISHOP C. M., "Probabilistic Principal Component Analysis", *J. Royal Statist. Soc.*, **1999**, 61, *pp.* 611–622.

[36] THORNHILL N.F., SHAH S.L., HUANG B., VISHNUBHOTLA A., "Spectral principal component analysis of dynamic process data", *Control Engineering Practice*, **2002**, 10, 833–846.

[37] RITTER G. L., LOWRY S. R., ISENHOUR T. L., "Factor Analysis of the Mass Spectra of Mixtures", *Analytical Chemistry*, **1976**, 48, 591-595.

[38] LI-CHAN E., NAKAI S:, WOOD D. E., *Journal of Food Science*, **1987**, 52, 31-41.

[39] BERRY M. W., "Survey of Text Mining *Clustering, Classification, and Retrieval"*, Springer-Verlag New York Berlin Heidelberg, **2003**, pp, 1-22.

[40] HÄRDLE W., SIMAR L., "Applied Multivariate Statistical Analysis", Method and data technologies, Berlin and Louvain-la-Neuve, **2003**, pp. 275-298.

[41] DOWNS G., BARNARD J., "Clustering methods and their uses in computational chemistry", *Rev. Comput. Chem.*, K. B. Lipkowitz and D. B. Boyd, Eds, Jhon Wiley and sons, 2002, 18, pp. 1-40.

[42] JAIN A.K., MURTY M.N., FLYNN P.J., "Data Clustering: A Review", *ACM Computing Surveys*, **1999**, 31, pp. 264-323.

[43] DIZY M., MARTIN-ALVAEZ P., CABEZUDO M., *J.Sci. Food Agric*., **1992**, 60, 47-53.

[44] COWE I., McNICOL W., "The use of principal components in the analysis of Near-Infrared spectra", *Appl. Spectrosc.*, **1985**, 39, 257-266.

[45] LIN J. C., NAGY S., KLIM M., *Food Chemistry*, **1993**, 47, 235-245.

[46] HÉBERGER K., "Evaluation of polarity indicators and stationary phases by principal component analysis in gs–liquid chromatography", *Chemometrics and Intelligent Laboratory Systems*, **1999**, 47, 41–49.

[47] LAVINE B. DAVIDSIN C., "Electronic van der Waals surface property descriptors and genetic algorithms for developing structure-activity correlations in olfafactory databases", *J. Chem. Inf. Sci*., **2003**, 43, 1890-1905.

[48] BROWN R. D., MARTIN Y. C., "Use of structure-activity data to compare Structure-Based clustering methods and descriptors for use in compound selection", *J. Chem. Inf. Comput. Sci.*, **1996**, 36, 572-584.

[49] BROWN R. D. , MARTIN Y. C. "The Information Content of 2D and 3D Structural descriptors Relevant to Ligand-Receptor Binding", *J. Chem. Inf. Comput. Sci.* **1997**, 37, 1-9.

[50] DOWN G. M., WILLETT P., "Similarity Searching in data bases of chemical structures", *Rev. Comput. Chem.*, K. B. Lipkowitz and D. B. Boyd, Eds, VCH publishers, New york **1995**, Vol. 7, pp. 1-66.

[51] WILLETT P., "Chemical Similarity Searching", *J. Chem. Inf. Comput. Sci*. **1998**, 38, 983-996.

[52] LANCE G. N., WILLIAMS W. T., "A general theory of classificatory sorting strategies 1. Hieralchical systems", *Computer J*., **1967**, 9, 373.

[53] MILLIGAN G. W., COOPER M. C., "An examination of procedures for determining the number of clusters in a data set", *Psycometrika*, **1985**, 50, 159.

[54] WILD D. J., BLANKLEY C. J., "Comparison of 2D Fingerprint Types and Hierarchy Level Selection Methods for Structural Grouping Using Ward's Clustering", *J. Chem. Inf. Comput. Sci.* **2000**, 40, 155-162.

[55] KELLEY L. A., GARDNER S. P., SUTCLIFFE M. J.," An Automated Approach For clustering An Ensemble Of NMR-Derived Protein Structures Into Conformationally-Related Subfamilies", *Protein Eng.*, **1996**, 9, 1063.

[56] MILLER J. C., MILLER J. N., "Estadística para química analítica" Addison-Wesley Iberoamericana, Londres, Inglaterra, 1988, pp. 40-63.

[57] WARD J. H., "Hierarchical Grouping to Optimize an Objective Function", *J. Am. Stat. Assoc*. **1963**, 58, 236-244.

[58] JARVIS R. A., PATRICK E. A., "Clustering using a Similarity Measure Based on Shared Near Neighbors", *IEEE Trans. Comput*. **1973**, 22, 1025-1034.

[59] BAYADA D. M., HAMERSMA H., VAN GEERESTEIN V. J., "Molecular Diversity and Representativity in Chemical Databases", *J. Chem. Inf.Comput. Sci*. **1999**, 39, 1-10.

[60] PERES-NETO PEDRO R., JACKSON DONALD A., SOMERS KEITH M., "How many principal components? Stopping rules for determinig the number of non-trivial axes revisited", *Comput. Statist. Data Anal.*, **2004**, (article in press).