

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Optimización multiobjetivo aplicado en diseño molecular asistido por computadora para la selección de solventes de extracción

Víctor Alfonso Roa Gómez

Departamento de Ingeniería Química y Ambiental

Facultad de Ingeniería

Universidad Nacional de Colombia

Bogotá, Colombia

2023

Optimización multiobjetivo aplicado en diseño molecular asistido por computadora para la selección de solventes de extracción

Víctor Alfonso Roa Gómez

Tesis presentada como requisito parcial para optar al título de:

Magíster en Ingeniería Química

Director:

Juan Carlos Serrato Bermúdez, MSc. Dr. Ing.

Línea de Investigación:

Optimización de procesos en Ingeniería Química

Grupo de Investigación:

Grupo de Investigación en Procesos Químicos y Bioquímicos.

Universidad Nacional de Colombia

Facultad de Ingeniería

Departamento de Ingeniería Química y Ambiental

Bogotá, Colombia

2023

La angustia es una expresión de la perfección de la naturaleza humana, es el anhelo en la vida terrena por una vida más alta. (...) Aquel que mira dentro del abismo, siente vértigo. La ansiedad es el vértigo de la libertad.

Søren Kierkegaard

Si miras mucho tiempo dentro del abismo, el abismo también mira dentro de ti (...) No hay razón para buscar el sufrimiento, pero si éste llega y trata de meterse en tu vida, no temas; míralo a la cara y con la frente bien levantada.

Friedrich Nietzsche

Si deseas empezar y desarrollar algo grandioso, no necesitas millones de dólares de capitalización. Necesitas suficiente pizza y Coca-Cola en la nevera, una PC barata y trabajo y dedicación para realizar tu idea.

John Carmack.

Declaración de obra original

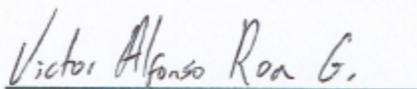
Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.



Víctor Alfonso Roa Gómez

Fecha: 11/09/2023

Agradecimientos

El autor desea expresar su profundo agradecimiento a todos aquellos que hicieron posible la realización de este proyecto, entre las que se encuentran, de manera especial:

- A Dios por su infinita bondad, por darme la salud, sabiduría y perseverancia para cumplir con la realización de esta tesis. Por enseñarme que, a pesar de todas las dificultades, ha estado siempre acompañándome y guiándome para cumplir esta meta.
- A los Ingenieros Juan Carlos Serrato y Jonatan Gómez, por su dirección, su paciencia, supervisión, inspiración y apoyo a lo largo de todo este proyecto.
- A mis padres, Víctor Manuel y Sonia Liliana, a mis hermanos Lina María y Jesús David, por ser los principales promotores de mis sueños, por su apoyo incondicional permanente, por cada consejo y guía en el transcurso de mi vida.
- A Paola Morales, por su apoyo espiritual y emocional brindado a lo largo de este proyecto. Por su amor, cuidado, esfuerzo, ánimo y por brindarme esa paz mental tan necesaria en esta etapa de mi vida. Te amo mi corazón, gracias por todo tu apoyo.
- A la Universidad Nacional de Colombia, por permitirme los medios físicos para la realización de este proyecto investigativo. Agradecimientos especiales a Ricardo Cortés, Oscar Camelo, Edgar Martínez, por todo su apoyo en el laboratorio, por la gestión realizada para hacer posible el desarrollo experimental de este proyecto.
- A mis amigos, Fredy Alejandro Guevara, Willian Quintero, Juan Sebastián Rodríguez, Miguel Flórez, Diana Guzmán, Jesús Quintero, Andrés Felipe Martínez, Cesar García y María Fernanda Gutiérrez por su soporte, apoyo y consejo a lo largo del proyecto.
- A Alexandra Elbakyan, por su aporte en que el conocimiento científico sea libre.
- A John Elder y Codemy por su orientación en aspectos relacionados con Python.
- Finalmente quiero agradecer especialmente a Wolfgang Amadeus Mozart, Led Zeppelin, Pink Floyd, Metallica, Queen, Radiohead, Pearl Jam, Korn, Linkin Park, Slipknot, Disturbed, Rammstein y Two Steps From Hell, por brindar la música que alimentaba mis ánimos al momento de desarrollar y escribir esta tesis.

Resumen

El objetivo general de esta tesis es proponer una metodología de selección de solventes de extracción empleando como herramienta el diseño molecular asistido por computadora (CAMD) y la optimización multiobjetivo como criterio de decisión secuencial. El diseño metodológico de la metodología CAMD cuenta con la formulación del problema de optimización a resolver (Prediseño), la implementación del algoritmo (Diseño) y la validación experimental del modelo (Post-diseño) para el caso de estudio de la extracción de los sistemas ácido láctico-agua y ácido acético-agua.

La metodología CAMD propuesta formula y resuelve un problema de optimización multiobjetivo de tres objetivos (ambiental, fisicoquímico y económico), para un espacio de búsqueda conformado por compuestos alifáticos, aromáticos y heterocíclicos, e integrando aspectos como la disponibilidad de las moléculas en el mercado como restricciones del problema de optimización. La solución de dicho problema de optimización se realizó empleando el algoritmo genético MOHAEA, el cual permite contar con flexibilidad para la representación de individuos, empleando la optimización de Pareto y la distancia de aglomeración como pautas guía del algoritmo. La metodología propuesta es una extensión de trabajos previos, de autores como Serrato [1] y Rodríguez [2], [3].

Palabras clave: Optimización multiobjetivo, extracción líquido-líquido, diseño asistido por computadora, diseño de producto.

Abstract

Multi-objective optimization applied in computer-aided molecular design for the selection of extraction solvents.

The general objective of this thesis was to propose a methodology for the choice of extraction solvents using computer-aided molecular design (CAMD) as a tool and multi-objective optimization as a sequential decision criterion. The methodological design of the CAMD methodology includes the formulation of the optimization problem to be solved (Pre-design), the implementation of the algorithm (Design) and the experimental validation of the model for the case study of the extraction of the lactic acid-water systems and acetic acid-water.

The proposed CAMD methodology formulates and solves a multi-objective optimization problem with three objectives (environmental, physicochemical, and economic), for a search space made up of aliphatic, aromatic, and heterocyclic compounds, and integrating aspects such as the availability of molecules in the market as optimization problem constraints. The solution of this optimization problem was conducted using the genetic algorithm MOHAEA, which allows flexibility for the representation of individuals, using Pareto optimization and the agglomeration distance as guidelines for the algorithm. The proposed methodology is an extension of earlier works, by authors such as Serrato [1] and Rodríguez [2], [3].

Keywords: Multi-objective Optimization, liquid-liquid extraction, computer-aided molecular design, product design

Contenido

Declaración de obra original	VII
Agradecimientos	IX
Resumen	X
Abstract	XI
Contenido	XII
Lista de figuras	XV
Lista de tablas	XVIII
Lista de símbolos y abreviaturas	XXII
Introducción	XXIX
Objetivos	XXXIV
1. Marco teórico	1
1.1 Extracción líquido-líquido.....	2
1.1.1 Metodologías de selección de solventes de extracción	7
1.2 Diseño molecular asistido por computadora (CAMD)	7
1.2.1 Fundamento matemático del diseño CAMD	11
1.2.2 Etapas del diseño CAMD	13
1.2.3 Aplicación del diseño CAMD en el diseño de procesos y productos	14
1.2.4 Uso del diseño CAMD en extracción líquido-líquido	19
1.3 Optimización en un solo objetivo (SOO)	22
1.3.1 Formulación canónica y conceptos claves de optimización	22
1.3.2 Clasificación del problema de optimización	24
1.3.3 Algoritmo genético (GA) como técnica de optimización estocástica	28
1.3.4 Uso de algoritmos genéticos en problemas de ingeniería química	35
1.4 Optimización multiobjetivo (MOO)	36
1.4.1 Formulación canónica y conceptos claves en optimización multiobjetivo	37
1.4.2 Métricas de desempeño para optimizaciones multiobjetivo	42
1.4.3 Enfoques existentes en el manejo de múltiples funciones en optimización multiobjetivo.....	44
1.4.4 Aplicaciones multiobjetivo para resolución de problemas ingenieriles, ambientales y de diseño CAMD	46

1.5 Optimización en múltiples objetivos (MaOO)	48
1.5.1 Extrapolación de algoritmos multiobjetivos en optimización MaOO.....	51
2. Marco metodológico	53
2.1 Planteamiento metodológico del problema	53
2.2 Prediseño: Formulación del problema.....	56
2.2.1 Formulación de las funciones objetivo y restricciones.....	57
2.2.2 Reglas de construcción molecular y representación estructural.....	60
2.2.3 Metodología de actualización del algoritmo de construcción molecular aplicado en compuestos monocíclicos heteroatómicos	65
2.2.4 Extrapolación del algoritmo de construcción molecular para aplicación en compuestos policíclicos heteroatómicos.....	73
2.2.5 Aplicabilidad de grupos heterocíclicos en moléculas de uso industrial.....	79
2.2.6 Impacto de la aromaticidad en la construcción de compuestos cíclicos	79
2.2.7 Modelos de estimación de propiedades físicas, termodinámicas y ambientales de sustancias puras	80
2.2.8 Estimación del coeficiente de actividad con el modelo UNIFAC-DMD	82
2.2.9 Evaluación de disponibilidad y costo del solvente mediante bases de datos ..	83
2.2.10 Determinación del impacto de propiedades en las funciones objetivo.....	85
2.2.11 Algoritmo de optimización HAEA/MOHAEA	88
2.2.12 Operadores genéticos considerados.....	96
2.2.13 Mecanismo de elección de padres y manejo de restricciones del algoritmo...	99
2.2.14 Representación de resultados mediante frentes de Pareto.....	102
2.3 Diseño: Desarrollo de la herramienta CAMD	102
2.3.1 Construcción de la herramienta CAMD.....	103
2.3.2 Condiciones operacionales de las corridas realizadas.....	107
2.3.3 Elección de solventes obtenidos en la ejecución de la herramienta CAMD para validación experimental.....	107
2.4 Post-diseño: Validación experimental de moléculas	109
2.4.1 Cuantificación de ácido final en fase acuosa	110
2.4.2 Cuantificación de agua final en fase orgánica.....	110
2.4.3 Cuantificación de ácido en fase orgánica y de agua en fase acuosa	111
2.4.4 Comparación entre modelo teórico y resultados experimentales	111
3. Resultados y discusión	112
3.1 Resultados asociados a la etapa de prediseño	112
3.1.1 Grupos heterocíclicos encontrados producto de la actualización del algoritmo de construcción molecular.....	113
3.1.2 Impacto del carácter aromático en la construcción de compuestos heterocíclicos	135
3.1.3 Aplicabilidad de grupos heterocíclicos establecidos en moléculas de uso industrial.....	136
3.1.4 Impacto de propiedades estudiadas sobre las funciones objetivo.....	138
3.2 Resultados asociados a la etapa de diseño	142
3.2.1 Convergencia y desempeño del algoritmo evolutivo	143
3.2.2 Desempeño de operadores genéticos y cumplimiento de restricciones	145
3.2.3 Frentes de Pareto obtenidos por el diseño CAMD	147
3.2.4 Comparación de resultados respecto a estudios de línea base	150
3.2.5 Candidatos de extracción obtenidos para validación experimental	156
3.3 Resultados asociados a la etapa de post-diseño	158

3.3.1 Validación experimental: Sistema Ácido láctico-Agua (LA/W).....	159
3.3.2 Validación experimental: Sistema Ácido acético-Agua (AA/W)	162
3.3.3 Comparación de parámetros experimentales teóricos obtenidos.....	164
4. Conclusiones y recomendaciones	167
4.1 Conclusiones.....	167
4.2 Recomendaciones y trabajo futuro	169
Anexos.....	171
A. Reglas de representación molecular mediante notación SMILES/SMARTS....	173
B. Tablas de Cayley y equivalencias rotacionales de estructuras.....	179
C. Estructuras moleculares evaluadas con aplicabilidad industrial.....	183
D. Parámetros del modelo UNIFAC-DMD	191
E. Manual de operación del programa CAMD	199
F. Registro fotográfico de las extracciones líquido-líquido realizadas.....	207
G. Método de cromatografía HPLC y curvas de calibración.....	211
H. Método de titulación de Karl Fischer y volumetrías ácido-base.....	217
I. Grupos estructurales heterocíclicos añadidos al modelo CAMD.....	223
Bibliografía	226

Lista de figuras

	Pág.
Figura 1-1: Representación del equilibrio líquido-líquido (ELL) para distintos tipos de sistemas. Adaptado de Gmehling [15].	4
Figura 1-2: Etapas metodológicas del diseño CAMD. Adaptado de Roughton [45].	13
Figura 1-3: Clasificación de problemas de optimización [74] [75].	26
Figura 1-4: Analogías de conceptos de genética y biología empleados en algoritmos genéticos. Adaptado de [84].	29
Figura 1-5: Etapas del algoritmo genético [69] [91].	32
Figura 1-6: Frente de Pareto y dominancia de Pareto en un problema de minimización con dos objetivos. Adaptado de Kumar [70] y Subía [126].	40
Figura 1-7: Ubicación de soluciones pertenecientes al frente de Pareto respecto al espacio de las variables de diseño y el espacio de las funciones objetivo [127].	41
Figura 1-8: Regiones de diseño convexas y no convexas en el espacio de las funciones objetivo [127].	41
Figura 2-1: Metodología desarrollada en el presente trabajo.	55
Figura 2-2: Representación SMILES del benceno en forma Kekulé y representación aromática, obtenida del algoritmo de identificación de aromaticidad.	63
Figura 2-3: Detección de aromaticidad del ciclobutadieno y ciclooctatetraeno.	63
Figura 2-4: Representación SMILES de los isómeros estructurales del xileno.	64
Figura 2-5: Representación del 1H-pirrol con la representación SMILES.	65
Figura 2-6: Metodología de tres filtros propuesta para compuestos cíclicos.	66
Figura 2-7: Ejes de simetría y grupos diedrales para ciclos (grafos) conformados por 3, 4, 5 y 6 átomos.	68
Figura 2-8: Viabilidad de sustitución del R-átomo respecto al tipo de enlace.	72
Figura 2-9: Tipos de conformaciones en estructuras policíclicas.	73
Figura 2-10: Ejes de simetría y grupos diedrales para policíclicos tipo espiro.	74
Figura 2-11: Ejes de simetría y grupos de simetría para policíclicos fusionados.	75
Figura 2-12: Configuraciones estructurales generalizadas para policíclicos espiro.	76
Figura 2-13: Configuraciones estructurales generalizadas para compuestos policíclicos, con configuración fusionada.	77
Figura 2-14: Viabilidad de sustitución de R-átomos en policíclicos fusionados.	78
Figura 2-15: Compuestos evaluables por el modelo de Hukkerikar según las contribuciones de primer, segundo y tercer orden [66]–[68].	81

Figura 2-16: Escala de colores para las asociaciones entre relaciones FO vs P, de acuerdo con el modelo de correlaciones bivariados de Spearman/Pearson.....	87
Figura 2-17: Distancia de hacinamiento según distancia Manhattan [2].	90
Figura 2-18: Representación gráfica de los operadores genéticos empleados en trabajos anteriores (Serrato [1], Rodríguez [2], [3]) y en este trabajo.	97
Figura 2-19: Tipos de selección mediante torneos: determinístico vs probabilístico. ..	99
Figura 2-20: Manejo de restricciones no cumplidas, de acuerdo con la estrategia propuesta por Rodríguez [2], [3] y variante propuesta en el presente trabajo.....	101
Figura 2-21: Frentes de Pareto para un problema de optimización de 2 y 3 objetivos, aplicando MOHAEA en funciones de prueba: (a) ZDT2, (b) DTLZ4. [246]	102
Figura 2-22: Interfaz gráfica de usuario (GUI) base, propuesta por Serrato [1] y Rodríguez [2], [3].	104
Figura 2-23: Interfaz gráfica de usuario (GUI) del programa CAMD desarrollado en el presente trabajo.	104
Figura 2-24: Esquema del algoritmo empleado en el programa CAMD.	105
Figura 3-1: Configuraciones estructurales generalizadas para compuestos monocíclicos de 3 a 6 átomos, obtenidas del Filtro I.	114
Figura 3-2: Viabilidad de sustitución del R-átomo en policíclicos espiro y estructuras viables respecto a valencia (Filtro II).	126
Figura 3-3: Viabilidad de sustitución del R-átomo en policíclicos fusionados.....	127
Figura 3-4: Verificación de la viabilidad de valencia estructural en estructuras policíclicas fusionadas mediante Filtro II.	128
Figura 3-5: Media de las funciones objetivos empleadas a lo largo de las generaciones, promediada para 50 ejecuciones del programa.	144
Figura 3-6: Desviación estándar de las funciones objetivos empleadas a lo largo de las generaciones, promediada para 50 ejecuciones del programa.	144
Figura 3-7: Frente de Pareto para el sistema LA/W y AA/W, para el conjunto de funciones objetivo 1 (Y vs ε vs K).	149
Figura 3-8: Frente de Pareto para el sistema LA/W y AA/W, para el conjunto de funciones objetivo 2 (S vs ε vs K).	149
Figura 3-9: Frente de Pareto para el sistema LA/W, para el conjunto de funciones objetivo 3 (S vs L vs K).	150
Figura 3-10: Conjuntos solución de moléculas obtenidas en los estudios de Serrato [1], Rodríguez [2], [3] y el presente trabajo.....	151
Figura D-1: Disponibilidad de parámetros de interacción y grupos funcionales – modelo UNIFAC-DMD (6ta revisión, última versión: septiembre 2021) [232].	193
Figura E-1: Interfaz gráfica del programa <i>Multi-Objective Resolution Interface Toted Up with Aromatics</i> (MORITA).	200
Figura F-1: Montaje realizado para las extracciones realizadas en el sistema LA/W. con distintos solventes. Izquierda: Ciclohexano (CyC6), Centro: Metil Isobutil Cetona (MIBK), Derecha: Acetato de etilo (AE).	207
Figura F-2: Montaje realizado para las extracciones realizadas en el sistema AA/W. con distintos solventes. Izquierda: Acetato de etilo (AE), Centro: Metil Isobutil Cetona (MIBK), Derecha: Ciclohexano (CyC6).	208

Figura F-3:	Fase acuosa obtenida en el sistema LA/W. Izquierda: Acetato de Etilo (AE), Centro: Ciclohexano (CyC6), Derecha: Metil Isobutil Cetona (MIBK).....	209
Figura F-4:	Fase acuosa obtenida para el sistema AA/W. Izquierda: Acetato de Etilo (AE), Centro: Ciclohexano (CyC6), Derecha: Metil Isobutil Cetona (MIBK).....	209
Figura F-5:	Fase orgánica obtenida para el sistema LA/W. Izquierda: Metil Isobutil Cetona (MIBK), Centro: Ciclohexano (CyC6), Derecha: Acetato de Etilo (AE).....	210
Figura F-6:	Fase orgánica obtenida para el sistema LA/W. Izquierda: Metil Isobutil Cetona (MIBK), Centro: Ciclohexano (CyC6), Derecha: Acetato de Etilo (AE).....	210
Figura G-1:	Montaje experimental HPLC para la determinación de concentración final de ácido en fase acuosa.....	211
Figura G-2:	Curva de calibración para el sistema ácido láctico-agua (LA/W).....	214
Figura G-3:	Curva de calibración para el sistema ácido acético-agua (AA/W).	215
Figura H-1:	Montaje experimental Karl Fischer para determinación de la concentración final de agua en la fase orgánica.	217
Figura H-2:	Flujograma metodológico para la determinación de agua mediante el método de Karl Fischer.	219
Figura H-3:	Registro fotográfico de las titulaciones de Karl Fisher realizada para los sistemas LA/W y AA/W.....	220

Lista de tablas

	Pág.
Tabla 1-1: Características deseables en un solvente de extracción [8][17][18][19].	6
Tabla 1-2: Métodos de estimación de propiedades aplicables al diseño CAMD [35] [22] [36] [37].	10
Tabla 1-3: Aplicaciones CAMD enfocados al diseño de productos [45] [53] [54].	16
Tabla 1-4: Aplicaciones CAMD enfocados al diseño de procesos para diversas operaciones unitarias [46] [47] [48] [49] [50] [51] [52].	17
Tabla 1-5: Condiciones de optimalidad de Karuch-Kuhn-Tucker para minimizaciones con una sola función objetivo [71] [72].	24
Tabla 1-6: Definición de óptimo de Pareto [123], [124], [125].	38
Tabla 1-7: Definición de óptimo débil de Pareto [123], [124], [125].	38
Tabla 1-8: Definición de dominancia de Pareto [123].	39
Tabla 1-9: Criterios de desempeño en optimizaciones multiobjetivo [122], [131].	43
Tabla 1-10: Métricas e indicadores para evaluación del desempeño de algoritmos de optimización multiobjetivo [135].	43
Tabla 1-11: Métodos de solución de problemas multiobjetivo con articulación a priori de preferencias [69], [123], [136], [137], [138], [139].	45
Tabla 1-12: Aplicaciones de problemas de optimización multiobjetivo para distintos enfoques en ingeniería [121], [124], [150], [162], [163], [164].	47
Tabla 2-1: Funciones objetivos empleadas en este trabajo y anteriores [1], [2], [3].	58
Tabla 2-2: Factores de normalización kWS , $kBFC$ y $kLD50$ reportados por Rodríguez [2], [3] para la evaluación de la función objetivo de impacto ambiental (\mathcal{E}).	58
Tabla 2-3: Restricciones consideradas en el problema CAMD a optimizar [1], [2], [3].	59
Tabla 2-4: Propiedades físicas y ambientales consideradas en este proyecto [66]–[68] y anteriores [1], [2], [3].	82
Tabla 2-5: Categorías de compra de sustancias presentes en ZINC [235], [236].	84
Tabla 2-6: Grupos funcionales considerados en el estudio del impacto de propiedades vs funciones objetivo, de acuerdo con categorías clasificatorias.	86
Tabla 2-7: Algoritmo de optimización de un solo objetivo HAEA [103], [151] [104].	89
Tabla 2-8: Asignación de distancia de hacinamiento (DH) para un frente de Pareto, mediante NSGA-II [247].	91
Tabla 2-9: Clasificación rápida no dominada para determinación de rango de no dominancia (RND), mediante NSGA-II [247].	91

Tabla 2-10: Selección de mejores individuos [103], [151] [104].....	94
Tabla 2-11: Algoritmo de optimización multiobjetivo MOHAEA [152], [246].....	95
Tabla 2-12: Operadores genéticos empleados en este trabajo y anteriores [1], [2], [3]. .	98
Tabla 2-13: Condiciones iniciales empleadas en el presente trabajo.	107
Tabla 2-14: Criterios de elección de solvente para validación experimental.....	108
Tabla 3-1: Aplicación del teorema de enumeración de Pólya a estructuras monocíclicas n-atómicas para determinación de estructuras no isomorfas (ENI).....	113
Tabla 3-2: Grupos válidos para compuestos monocíclicos n-atómicos considerando valencias atómicas, obtenidos mediante sustituciones de heteroátomos (Filtro II).	115
Tabla 3-3: Aplicabilidad de los modelos de estimación de propiedades para las estructuras evaluadas de compuestos monocíclicos en el Filtro III.....	117
Tabla 3-4: Resumen metodología de tres filtros: Compuestos monocíclicos.....	118
Tabla 3-5: Conjunto universal de estructuras construibles de compuestos policíclicos tipo espiro y fusionados, obtenidos mediante permutaciones.....	120
Tabla 3-6: Aplicación del teorema de enumeración de Pólya a estructuras policíclicas de tipo espiro y fusionadas para determinación de estructuras no isomorfas (ENI).....	121
Tabla 3-7: Validación de resultados de estructuras no isomorfas (ENI), mediante iteración.....	122
Tabla 3-8: Cálculo de $ENI \times P$ para evaluación de estructuras policíclicas (Filtro II).	124
Tabla 3-9: Grupos válidos para compuestos policíclicos espiro n-atómicos considerando valencias atómicas, obtenidos mediante sustituciones de heteroátomos (Filtro II).	130
Tabla 3-10: Grupos válidos para compuestos policíclicos fusionados n-atómicos considerando valencias atómicas, obtenidos mediante sustituciones de heteroátomos (Filtro II).....	131
Tabla 3-11: Aplicabilidad de los modelos de estimación de propiedades para las estructuras evaluadas de compuestos policíclicas espiro y fusionadas en el Filtro III....	132
Tabla 3-12: Resumen metodología de tres filtros: Compuestos policíclicos.	134
Tabla 3-13: Carácter aromático de las moléculas empleadas en el estudio de aplicabilidad.	135
Tabla 3-14: Aplicabilidad de grupos heterocíclicos para las moléculas de uso industrial seleccionadas.....	137
Tabla 3-15: Comparación del error relativo medio obtenido para el modelo de estimación de Hukkerikar [66]–[68], la evaluación realizada por Rodríguez [244], [245] y el presente trabajo.	139
Tabla 3-16: Resultados de la prueba de normalidad de Kolmogórov-Smirnov para las relaciones entre propiedades y funciones objetivo. Sin segregación de familias funcionales.	140
Tabla 3-17: Resultados del estudio estadístico no paramétrico aplicado para para las relaciones entre propiedades y funciones objetivo. Sin segregación de familias funcionales.	141
Tabla 3-18: Desempeño de las tasas de los operadores genéticos a lo largo de las generaciones, durante 50 ejecuciones del programa.	146
Tabla 3-19: Cumplimiento de las restricciones en las mejores soluciones.	147

Tabla 3-20: Conjunto de solventes óptimos, obtenidos por Serrato [1] y Rodríguez [2], [3], para los sistemas LA/W y AA/W.....	151
Tabla 3-21: Listado parcial de moléculas candidato obtenidas en el presente estudio, para los sistemas LA/W y AA/W.....	156
Tabla 3-22: Mejores solventes obtenidos por el modelo CAMD para el sistema LA/W.	157
Tabla 3-23: Mejores solventes obtenidos por el modelo CAMD para el sistema AA/W.	157
Tabla 3-24: Comportamiento del Ácido Láctico en función del solvente candidato adicionado, en fase acuosa, medido por HPLC.....	160
Tabla 3-25: Agua obtenida en la fase orgánica, en función del solvente candidato adicionado, para el sistema LA/W.....	160
Tabla 3-26: Reparto de ácido en fase orgánica y agua en fase acuosa, en función del solvente candidato adicionado, para el sistema LA/W.....	161
Tabla 3-27: Comportamiento del Ácido Acético en función del solvente candidato adicionado, en fase acuosa, medido por HPLC.....	162
Tabla 3-28: Agua obtenida en la fase orgánica, en función del solvente candidato adicionado, para el sistema AA/W.....	163
Tabla 3-29: Reparto de ácido en fase orgánica y agua en fase acuosa, en función del solvente candidato adicionado, para el sistema AA/W.	164
Tabla 3-30: Comparación de parámetros experimentales y teóricos obtenidos por el modelo CAMD.....	165
Tabla A-1: Valencias normales para átomos orgánicos – Modelo SMILES.	173
Tabla A-2: Simbología de los distintos tipos de enlace en el modelo SMILES.....	174
Tabla A-3: Ejemplos de estructuras aplicando el método de representación SMILES..	176
Tabla A-4: Operadores lógicos del lenguaje SMARTS y jerarquía de operadores.....	178
Tabla A-5: Ejemplos de búsqueda de subestructuras aplicando los distintos operadores del lenguaje SMARTS.....	178
Tabla B-1: Tabla de Cayley para el grupo diedral D_3 (polígono: triángulo).....	179
Tabla B-2: Tabla de Cayley para el grupo diedral D_4 (polígono: cuadrado).....	179
Tabla B-3: Tabla de Cayley para el grupo diedral D_5 (polígono: pentágono).....	180
Tabla B-4: Tabla de Cayley para el grupo diedral D_6 (polígono: hexágono).....	180
Tabla B-5: Distribución de estructuras policíclicas espiro y fusionadas en clases, según la cantidad de movimientos rotacionales válidos.....	181
Tabla B-6: Tabla de Cayley para las estructuras policíclicas de Clase I, con 4 movimientos rotacionales en el grupo diedral.....	181
Tabla B-7: Tabla de Cayley para las estructuras policíclicas de Clase II, con 2 movimientos rotacionales en el grupo diedral.....	181
Tabla C-1: Estructuras moleculares evaluadas respecto a aplicabilidad industrial.....	183
Tabla D-1: Parámetros de subgrupo del modelo UNIFAC [232].	194
Tabla D-2: Parámetros de interacción de UNIFAC ($amk [K]$) [231], [232], [260].....	195
Tabla E-1: Listado de herramientas y funcionalidades presentes en el programa <i>Multi-Objective Resolution Interface Toted Up with Aromatics</i> (MORITA).	203
Tabla G-1: Datos iniciales establecidos en la detección de ácido acético y láctico mediante cromatografía HPLC.....	212
Tabla G-2: Datos obtenidos para la calibración del sistema LA/W.....	214

Tabla G-3: Datos obtenidos para la calibración del sistema AA/W.....	216
Tabla G-4: Bondad de ajuste de la regresión lineal para las curvas de calibración	216
Tabla H-1: Datos obtenidos para la determinación de la deriva.	220
Tabla H-2: Datos obtenidos en la verificación del método de Karl Fischer.	221
Tabla H-3: Condiciones experimentales de las titulaciones realizadas.	221

Lista de símbolos y abreviaturas

Símbolos con letras latinas

Símbolo	Definición	Unidad SI
A	Disponibilidad en el mercado.	[adimensional]
a_i	Actividad del componente i .	[adimensional]
A_{mk}, B_{mk}, C_{mk}	Constantes de dependencia de temperatura de la interacción entre subgrupos m y k .	[adimensional]
a_{mk}	Parámetros de interacción entre subgrupos m y k .	[adimensional]
BFC	Factor de bioconcentración.	[adimensional]
C	Concentración.	[mol/L]
$C_{available}$	Compuesto disponible en base de datos ZINC.	[adimensional]
$C_{hard-to-acquire}$	Compuesto difícil de conseguir en base de datos ZINC.	[adimensional]
C_i	Contribución de una propiedad a los grupos de primer orden de Hukkerikar.	[adimensional]
C_j	Contribución de una propiedad a los grupos de segundo orden de Hukkerikar.	[adimensional]
C_k	Contribución de una propiedad a los grupos de segundo orden de Hukkerikar.	[adimensional]
$C1(m), C2(n)$	Reglas de formación de estructuras factibles y selección de compuestos	[adimensional]
D	Radio de distribución de soluto.	[adimensional]
$DH(i)$	Distancia de hacinamiento total de un individuo.	[adimensional]
DH_i	Distancia de hacinamiento de un individuo i en una población respectiva.	[adimensional]
E	Extracto (fase rica en solvente y pobre en soluto).	[adimensional]
\mathcal{E}	Impacto ambiental.	[adimensional]
$\%E$	Porcentaje de extracto.	[adimensional]
\mathcal{F}	Frente de Pareto con rangos de no dominación	[adimensional]
F	Mezcla de alimentación.	[adimensional]
$f(\bar{X}), F(x)$	Función matemática, dependiente de parámetro X .	[adimensional]
f_i	Fugacidad del componente i .	[Pa]
f_i^0	Fugacidad estándar del componente i .	[Pa]
$fitness(i)$	Función de aptitud del algoritmo genético.	[adimensional]
$g_i(x)$	Funciones de restricciones de desigualdad.	[adimensional]
G_f	Energía libre de Gibbs estándar de formación.	[J/mol]
$h_j(x)$	Funciones de restricciones de igualdad.	[adimensional]
I	Indicador o métrica para evaluación de desempeño	[adimensional]
K	Capacidad o poder de extracción.	[adimensional]
K_i	Coefficiente de distribución del componente i .	[adimensional]

Símbolo	Definición	Unidad SI
k_{LD50}	Índice ambiental (constante) para LD_{50} .	[adimensional]
k_{LC50}	Índice ambiental (constante) para LC_{50} .	[adimensional]
k_{BFC}	Índice ambiental (constante) para BFC .	[adimensional]
k_{WS}	Índice ambiental (constante) para W_S .	[adimensional]
L	Pérdida de solvente.	[adimensional]
$\mathcal{L}(x, \mu, \lambda)$	Lagrangeano de la función objetivo.	[adimensional]
LD_{50}^{sp}	Dosis letal al 50% en la especie sp .	[kg/kg]
LC_{50}^{sp}	Concentración letal al 50% en la especie sp .	[kg/kg]
m	Número de objetivos del problema	[adimensional]
NG_k	Número de diferentes grupos estructurales presentes en el componente k .	[adimensional]
$NT(i)_k$	Número de apariciones de un grupo estructural en el componente k .	[adimensional]
MW_i	Peso molecular del componente i .	[adimensional]
N_H	Número de hijos.	[adimensional]
N_i	Número de ocurrencias de grupos de primer orden.	[adimensional]
N_j	Número de ocurrencias de grupos de segundo orden.	[adimensional]
N_k	Número de ocurrencias de grupos de tercer orden.	[adimensional]
N_H	Número de hijos.	[adimensional]
$N_{H,min}$	Números de hijos mínimo.	[adimensional]
N_{OG}	Número de operadores genéticos.	[adimensional]
N_P	Población inicial aleatoria de tamaño.	[adimensional]
NC	Número de componentes.	[adimensional]
NGT	Número de grupos totales.	[adimensional]
NP	Número total de propiedades objetivos.	[adimensional]
p	Parámetro de aleatoriedad para la selección mediante torneo.	[adimensional]
$P(j)$	Propiedades objetivas consideradas como restricciones	[adimensional]
Q_k	Contribución de superficie del subgrupo k para el modelo UNIFAC / UNIFAC-DMD.	[adimensional]
q_i	Contribución total de superficie de todos los subgrupos k presentes en la molécula para el modelo UNIFAC / UNIFAC-DMD.	[adimensional]
R	Refinado (fase rica en soluto y pobre en solvente).	[adimensional]
R_k	Contribución de volumen del subgrupo k para el modelo UNIFAC / UNIFAC-DMD.	[adimensional]
r_i	Contribución total de superficie de todos los subgrupos i presentes en la molécula para el modelo UNIFAC / UNIFAC-DMD.	[adimensional]
r_j	Contribución total de superficie de todos los subgrupos j presentes en la molécula para el modelo UNIFAC / UNIFAC-DMD.	[adimensional]
$RND(i)$	Rango de no dominancia para el individuo i .	[adimensional]
$\mathbb{R}^2, \mathbb{R}^3, \mathbb{R}^n$	Conjunto de números reales de orden 2, 3 o n .	[adimensional]
S	Soluto.	[adimensional]
\mathcal{S}	Selectividad.	[adimensional]

Símbolo	Definición	Unidad SI
$SG(i)_k$	Número de identificación de grupo para el grupo i presente en el componente k .	[adimensional]
$SGT(j)$	Grupo preseleccionado para el grupo j .	[adimensional]
s_k	Producto entre el segmento ponderado molar θ_m y la dependencia de la temperatura τ_{mk} .	[adimensional]
T	Solvente de extracción o extractante.	[adimensional]
T_b	Punto normal de ebullición.	[K]
T_m	Punto normal de fusión.	[K]
V	Volumen.	[mL]
V_m	Volumen molar de líquido (a 298 K).	[mL]
$W_C(i)$	Valor de ponderación, dependiente de si el individuo i cumple o no la restricción C .	[adimensional]
W_S	Solubilidad en agua.	[g/L]
\bar{X}	Variable de entrada de función objetivo $f(\bar{X})$	[g/mol]
x_i	Concentración de la fase líquida del componente i .	[g/mol]
x_j	Concentración de la fase líquida del componente j .	[g/mol]
\mathcal{Y}	Producto KS.	[adimensional]
Z	Espacio factible.	[adimensional]

Símbolos con letras griegas

Símbolo	Definición	Unidad SI
β_{ik}	Producto entre la relación de superficie e_{ki} y la dependencia de la temperatura τ_{mk} .	[adimensional]
γ_i	Coefficiente de actividad del componente i .	[adimensional]
$\gamma_{i,j}^\infty$	Coefficiente de actividad del componente i en la mezcla $i - j$ a dilución infinita.	[adimensional]
δ	Parámetro de tasa de aprendizaje del algoritmo HAEA/MOHAEA.	[adimensional]
δ	Error o tolerancia definida.	[adimensional]
e_{ki}	Relación entre la contribución de superficie de todos los subgrupos k presentes en la molécula i y la contribución total de superficie, para el modelo UNIFAC / UNIFAC-DMD.	[adimensional]
e_{mi}	Relación entre la contribución de superficie de todos los subgrupos m presentes en la molécula i y la contribución total de superficie, para el modelo UNIFAC / UNIFAC-DMD.	[adimensional]
λ	Parámetro de población inicial del algoritmo HAEA/MOHAEA.	[adimensional]
λ_i	Multiplicador de Lagrange para restricciones de igualdad.	[adimensional]
μ_i	Multiplicador de Lagrange para restricciones de desigualdad.	[adimensional]
θ_i	Segmento ponderado molar para el compuesto i para el modelo UNIFAC.	[adimensional]
θ_k	Segmento ponderado molar para el compuesto k .	[adimensional]

Símbolo	Definición	Unidad SI
$v_k^{(i)}$	Número de subgrupos de tipo k presentes en una molécula de especie i .	[adimensional]
ρ_i	Densidad del componente i .	[g/mL]
τ_{mk}	Parámetro de dependencia de la temperatura de la interacción entre grupos m y k para el modelo UNIFAC / UNIFAC-DMD.	[adimensional]
ϕ_i	Área fraccional de molécula i en el modelo UNIFAC.	[adimensional]
ϕ_i'	Corrección del área fraccional de la molécula i para el modelo UNIFAC-DMD.	[adimensional]
ω	Identificador del carácter aromático, antiaromático o no de una molécula	[adimensional]
Ω	Conjunto de restricciones	[adimensional]

Operadores matemáticos

Operador	Definición
$\{\}$	Conjunto vacío.
$[a, b]$	Intervalo cerrado entre a y b .
\neq	Diferente o no igual.
$/$	División.
$=$	Igualdad.
$f(x)$	Función f dependiente de x .
$*, \times$	Multiplicación.
$*$	Operador comodín en Teorema del Esquema.
$>$	Operador de comparación Mayor que.
\geq	Operador de comparación Mayor que o igual.
\leq	Operador de comparación Menor o igual.
\nlessgtr	Operador de comparación Menor o igual negado.
$<$	Operador de comparación Menor que.
$x < y$	Operador de dominancia (x domina a y).
∇	Operador nabla.
$\&$	Operador lógico AND.
\neg	Operador lógico NOT (Negación).
\parallel	Operador lógico OR.
$x \in A$	Operador pertenece a (x pertenece al conjunto A).
$x \subset A$	Operador subconjunto de (x es subconjunto de conjunto A).
$x \subseteq A$	Operador subconjunto o igual a (x es subconjunto o igual a A).
$-$	Resta.
$+$	Suma.
Σ	Operador sumatorio.
$A \cup B$	Unión de conjuntos A y B .
$\min F(x)$	Operador minimizar función $F(x)$.
$\max F(x)$	Operador maximizar función $F(x)$.
M^T	Operador de transposición de la matriz M .
\forall	Cuantificador universal (para todo).

Subíndices

Subíndice	Definición
<i>aq</i>	Fase acuosa.
<i>C</i>	Cancerígeno.
<i>f</i>	Condición final.
<i>i, j</i>	Índices relacionados a especie.
<i>l, m</i>	Índice relacionado a subgrupos.
<i>NC</i>	No cancerígeno.
<i>org</i>	Fase orgánica.
<i>P</i>	Solvente problema que contiene el soluto objetivo.
<i>S</i>	Agente de separación, solvente a diseñar.
<i>T</i>	Soluto objetivo, compuesto a separar.
<i>0</i>	Condición inicial.
*	En la condición óptima.

Superíndices

Superíndice	Definición
<i>org</i>	Fase orgánica.
<i>aq</i>	Fase acuosa.
<i>I</i>	Fase líquida I.
<i>II</i>	Fase líquida II.
<i>acid</i>	Ácido.
<i>C</i>	Contribución combinatorial.
<i>DM</i>	<i>Daphnia Magna</i> .
<i>exp</i>	Experimental.
<i>FM</i>	<i>Fathead Minnow</i> .
<i>R</i>	Contribución residual.
<i>Solv</i>	Solvente.
<i>W</i>	Agua.
∞	Condición a dilución infinita.

Abreviaturas

Abreviatura	Definición
AA/W	Sistema ácido acético / agua.
ACO	Optimización mediante Colonia de Hormigas (<i>Ant Colony Optimization</i>)
AGE	Algoritmo de evolución guiada por aproximación (<i>Approximation-Guided Evolution</i>).
A-NSGA	Algoritmo genético alternativo con enfoque de clasificación rápida no dominada (<i>Alternative Non-dominated Sorting Genetic Algorithm</i>).
ASCII	Código Estándar estadounidense para el Intercambio de Información (<i>American Standard Code for Information Interchange</i>).
CAMD	Diseño molecular asistido por computadora (<i>Computer Aided Molecular Design</i>).

Abreviatura	Definición
CAMPD	Diseño molecular y de proceso asistido por computadora (<i>Computer Aided Molecular and Process Design</i>).
CAPEC	Enumeración y clasificación de patrones de ataque comunes (<i>Common Attack Pattern Enumeration and Classification</i>).
CFC	Compuestos Clorofluorocarbonados
CI	Índice de conectividad (<i>Connectivity index</i>).
COSMO-RS	Método conductor como modelo de cribado para disolventes reales (<i>COnductor like Screening Model for Real Solvents</i>).
DDBST	Dortmund Data Bank Software & Separation Technology.
ELL	Equilibrio líquido-líquido.
EPA	Agencia de protección ambiental de Estados Unidos (<i>Environmental Protection Agency</i>).
DMSO	Dimetilsulfóxido
DSSTox	Estructura distribuida – toxicidad buscable (<i>Distributed Structure – Searchable Toxicity</i>).
FAHP	Proceso de jerarquía analítica difusa (<i>Fuzzy Analytic Hierarchy Process</i>).
FO	Función objetivo.
GA	Algoritmo genético (<i>Genetic Algorithm</i>)
GAMS	Sistema general de modelado algebraico (<i>General Algebraic Modeling System</i>)
GC, GC+	Modelo de contribución de grupos (<i>Group Contribution model</i>).
GmbH	Empresa con responsabilidad limitada (<i>Gesellschaft mit beschränkter Haftung</i>).
GUI	Interfaz Gráfica de Usuario (<i>Graphical User Interface</i>).
HAEA	<i>Hybrid Adaptative Evolutionary Algorithm</i> .
HPLC	Cromatografía líquida de alta eficiencia (<i>High Performance Liquid Chromatography</i>).
ICAS	Sistema integrado asistido por computadora (<i>Integrated Computer Aided System</i>).
IPP	Índice del precio al Productor estadounidense.
IUPAC	Unión Internacional de Química Pura y Aplicada (<i>International Union of Pure and Applied Chemistry</i>).
KKT	Condiciones de Karuch-Kuhn-Tucker (<i>KKT-conditions</i>)
LA/W	Sistema ácido láctico / agua.
LIQ	Laboratorio de Ingeniería Química de la Universidad Nacional de Colombia (Sede Bogotá).
LP	Programación lineal (<i>Linear Programming</i>)
MACE-gD	Entropía cruzada de muchos objetivos basada en descomposición generalizada (<i>Generalized Decomposition-based Many Objective Cross-Entropy</i>)
MaOO	Optimización en múltiples objetivos (<i>Many-Objective Optimization</i>).
MI	Programación entera (<i>Integer Programming</i>)
MILP	Programación mixto-entera lineal (<i>Mixed Integer Linear Programming</i>).
MINLP	Programación mixta-entera no lineal (<i>Mixed Integer Non-Linear Programming</i>).
MIMO	Control con múltiples entradas y múltiples salidas (<i>multi-input multi-output control</i>)

Abreviatura	Definición
MOEA	Algoritmo evolutivo multiobjetivo (<i>Multi-Objective Evolutionary Algorithm</i>).
MOEA/D	Algoritmo evolutivo multiobjetivo basado en descomposición (<i>Multi-Objective Evolutionary Algorithm based on Decomposition</i>).
MOGA	Algoritmo genético multiobjetivo (<i>Multi-Objective Genetic Algorithm</i>)
MOHAEA	<i>Multi-Objective Hybrid Adaptative Evolutionary Algorithm</i> .
MOO	Optimización multiobjetivo (<i>Multi-Objective Optimization</i>).
MORITA	Interfaz de resolución multiobjetivo equipada con aromáticos (<i>Multi-Objective Resolution Interface Toted Up with Aromatics</i>).
NLP	Programación no-lineal (<i>Non-Linear Programming</i>).
NSGA	Algoritmo genético con enfoque de clasificación rápida no dominada (<i>Non-dominated Sorting Genetic Algorithm</i>).
OG	Operador Genético.
OM	Micrografía óptica (<i>Optical Microscopy</i>).
QP	Programación cuadrática (<i>Quadratic Programming</i>).
PA	Propiedad Ambiental.
PF	Propiedad física.
PSE	Ingeniería de sistemas de procesos (<i>Process System Engineering</i>).
PSO	Optimización mediante enjambre de partículas (<i>Particle Swarm Optimization</i>).
QSPR	Relación cuantitativa estructura propiedad (<i>Quantitative Structure Properties Relationship</i>).
QSAR	Relación cuantitativa estructura actividad (<i>Quantitative Structure Activity Relationship</i>).
RE	Restricción.
SMILES	Especificación de introducción lineal molecular simplificada (<i>Simplified Molecular Input Line Entry Specification</i>).
SOO	Optimización en un solo objetivo (<i>Single Objective Optimization</i>).
SPEA	<i>Strength Pareto Evolutionary Algorithm</i>
TSP	Problema del agente viajero (<i>Travelling Salesman Problem</i>).
TS	Templado simulado.
UNIFAC-DMD	Método UNIFAC-Dortmund.
UNIFAC	Método de coeficientes de actividad de grupos funcionales de UNQUAC (<i>UNQUAC Functional-group Activity Coefficients</i>).
UNQUAC	Método Universal Cuasi-químico (<i>UNiversal QUAsiChemical</i>).
U-NSGA	Algoritmo genético unificado con enfoque de clasificación rápida no dominada (<i>Unified Non-dominated Sorting Genetic Algorithm</i>).
USD	Dólar estadounidense.
VEGA	Verificación mediante algoritmo genético (<i>VERification through Genetic Algorithm</i>)
ZINC	<i>ZINC is not comercial</i> .

Introducción

El diseño de procesos químicos y bioquímicos se realiza de acuerdo con las especificaciones particulares de un producto que responde a las necesidades del mercado. Usualmente, al evaluar si un proceso es adecuado, la rentabilidad es el criterio que justifica la elección de cierta tecnología de producción. No obstante, existen variables adicionales como el impacto ambiental del proceso, la disponibilidad de materias primas y las propiedades fisicoquímicas de las sustancias. Estas variables tienen implicaciones significativas, más allá del impacto financiero, y que usualmente no son consideradas en las etapas preliminares del diseño conceptual del proceso.

En la mayoría de las industrias químicas y bioquímicas, las operaciones de separación asociadas a la purificación o recuperación de productos a baja concentración presentan un alto impacto ambiental y económico. Si se considera la gran diversidad de estrategias de separación existentes, es posible agrupar las tecnologías en cuatro categorías: operaciones gas-líquido, separaciones con membrana, operaciones de adsorción y extracción líquido-líquido. Algunos procesos industriales, como la producción de ácido láctico por vía fermentativa, emplean agentes de extracción para favorecer la separación del producto orgánico de la fase acuosa. Por tal motivo, la selección del solvente de extracción es una decisión clave, ya que debe asegurar la separación de acuerdo con diferencias en la solubilidad de los compuestos. Esta elección del solvente puede ser limitante, pues es necesario considerar su disponibilidad, el costo asociado a su uso y el impacto ambiental y en la salud que puede ocasionar su incorporación en el proceso de separación.

Existen muchas metodologías de selección de solventes de extracción, como el uso de bases experimentales y/o reglas heurísticas, de acuerdo con clasificaciones de solventes por grupos funcionales; o mediante enumeración molecular [4]–[6]. En estos métodos se consideran las contribuciones de las propiedades físicas y químicas de una molécula como

la sumatoria de las contribuciones individuales de los enlaces y grupos funcionales asociados a dicha molécula, mediante el uso de modelos de contribución de grupos, como UNIFAC o UNIFAC-Dortmund (UNIFAC-DMD). Adicionalmente, el crecimiento de la potencia computacional y el acceso a ordenadores potentes ha facilitado la solución de este problema de diseño de productos mediante el diseño molecular asistido por computadora (CAMD, por sus siglas en inglés).

El diseño CAMD es un enfoque muy prometedor destinado a la solución de problemas de diseño de productos químicos. Este enfoque integra modelos de predicción de propiedades fisicoquímicas con algoritmos eficientes para diseñar, evaluar y seleccionar moléculas óptimas, a través de la formulación de un problema de optimización mixto-entero no lineal (MINLP) restringido a características propias de la molécula, como la valencia y restricciones estructurales particulares. El mayor beneficio del diseño CAMD respecto a los enfoques tradicionales, es la reducción del trabajo experimental y la posibilidad de exploración de un espacio de compuestos más grande.

El desarrollo y solución de problemas de optimización MINLP presenta un alto grado de complejidad, conforme a la interacción de las variables relacionadas y la cantidad de objetivos a considerar. La gran mayoría de problemas de optimización en ingeniería química considera un único objetivo para el diseño de procesos y productos, como el costo o ganancia del proceso. No obstante, son pocos los modelos que consideran de manera simultánea varios objetivos, pues se pueden presentar conflictos entre ellos. A pesar de la gran variedad de metodologías para seleccionar solventes de extracción, no se evidencian referencias en la literatura que consideren tanto las propiedades fisicoquímicas que favorezcan la separación como el impacto ambiental del solvente en el proceso y/o su facilidad o dificultad de obtención. Por tal motivo, es imperativo desarrollar una nueva metodología multiobjetivo que permita evaluar y seleccionar eficazmente un solvente de extracción, considerando estos aspectos conflictivos.

A partir de la década de los 60, el avance tecnológico en las ciencias de la informática y la computación ha desarrollado nuevos mecanismos de solución de problemas numéricos de alta complejidad, tales como las optimizaciones multiobjetivo de carácter combinatorial, los cuales han sido objeto de estudio de la computación evolutiva. Durante el transcurso de los años 70, surgió un nuevo paradigma en los métodos de optimización estocásticos,

conocidos como algoritmos genéticos. Estos son inspirados en los conceptos darwinianos de evolución biológica y genética mendeliana, siendo empleados en problemas de alta complejidad matemática, en los que el cálculo analítico es insuficiente y las funciones objetivo pueden contener un número no determinado de soluciones óptimas. Las herramientas de algoritmos genéticos son extrapolables a optimizaciones multiobjetivo, en donde varias funciones pueden presentar “conflictos de intereses” entre sí, empleando un factor de dominancia entre las variables de optimización definidas en el problema.

Considerando los aspectos antes mencionados, a nivel local¹ se han propuesto dos metodologías para diseñar disolventes óptimos, aplicables al estudio de la extracción de ácido láctico. El primero de estos trabajos, desarrollado por Serrato² [1] aborda el problema de diseñar compuestos con valores óptimos de selectividad y coeficientes de distribución, empleando una optimización de un solo objetivo, en donde la función a maximizar es un producto entre la selectividad y el coeficiente de distribución de los compuestos diseñados. Este trabajo, generó un conjunto de moléculas factibles matemáticamente (con valores de selectividad y coeficiente de distribución apropiados), más no obtuvo moléculas disponibles en catálogos de proveedores de productos químicos. Adicionalmente, las moléculas obtenidas contaban con altos niveles de toxicidad.

Posteriormente, el trabajo de Rodríguez³ [2], [3] toma como línea base el algoritmo de Serrato, reformulándolo e integrando el aspecto de toxicidad y disponibilidad de mercado, mediante un enfoque de optimización multiobjetivo, en el espacio de búsqueda de sustancias establecido en el trabajo anterior. Este trabajo, contrastó resultados respecto al estudio de Serrato, excluyendo aquellas especies químicas no disponibles en el mercado. No obstante, no consideró el espacio de moléculas heterocíclicas y aromáticas.

¹ Por nivel local, refiérase al Grupo de Investigación en Procesos Químicos y Bioquímicos del Departamento de Ingeniería Química y Ambiental, Universidad Nacional de Colombia (Sede Bogotá)

² Ingeniero químico, PhD., profesor titular del Departamento de Ingeniería Química y Ambiental, Universidad Nacional de Colombia (Sede Bogotá).

³ Ingeniero químico, MSc. en Ingeniería de sistemas y computación, Universidad Nacional de Colombia (Sede Bogotá)

El enfoque del presente trabajo busca profundizar en el carácter multiobjetivo del problema con tres funciones a considerar (aspectos fisicoquímicos, de impacto ambiental y económicos), evaluando si los resultados de este nuevo modelo multiobjetivo distan de los resultados obtenidos por Serrato [1] y Rodríguez [2], [3]. Adicionalmente, este trabajo contempla ampliar el espacio de búsqueda, integrando el espectro de sustancias alifáticas, aromáticas y heterocíclicas, con el fin de identificar nuevos solventes aplicables al caso de la extracción de los sistemas ácido láctico-agua y ácido acético-agua, junto con la evaluación experimental de los resultados obtenidos mediante el enfoque multiobjetivo. El desarrollo de esta estrategia de selección de solventes, implementada en un software computacional y validada para un caso de estudio industrial, permitirá incrementar el interés en la aplicación de la extracción líquido-líquido como operación, ya que favorece la generación de sustancias que cumplan una visión holística de la problemática de separación, considerando de manera simultánea reducir costos de solvente, mejorar eficiencias de separación, reducir la afectación ambiental, entre otros objetivos de interés e identificando los posibles conflictos entre los objetivos antes mencionados.

Teniendo en cuenta lo anterior, el objetivo general de esta tesis es: *“Proponer una metodología de selección de solventes de extracción empleando como herramienta el diseño molecular asistido por computadora (CAMD) y la optimización multiobjetivo como criterio de decisión secuencial”*. El desarrollo metodológico es de tipo mixto (cuantitativo/cualitativo) y cuenta con las tres etapas principales del diseño CAMD: I.) Prediseño, en donde se formula el problema CAMD a resolver, mediante una optimización multiobjetivo basado en algoritmo genético empleando el algoritmo MOHAEA. II.) Diseño, en donde se realiza la implementación del algoritmo en lenguaje Python y III.) Post-diseño, en donde se realiza la validación experimental del modelo de selección de solvente para los sistemas ácido láctico-agua y ácido acético-agua.

Para presentar los resultados de esta investigación, se dividió el documento en tres capítulos, adicionales a la introducción, objetivo, conclusiones y anexos.

En el capítulo 1, **Marco teórico**, se presenta el estado del arte, en donde se realiza una revisión bibliográfica referente a la extracción líquido-líquido, las metodologías de extracción de solventes, enfocadas en el diseño molecular asistido por computadora (CAMD). Adicionalmente, se realizó una revisión de la teoría matemática de optimización,

focalizándola en fundamentos de optimización mono/multiobjetivo y algoritmos genéticos evolutivos requeridos para el abordaje de este problema de investigación.

En el capítulo 2, **Marco metodológico**, considera la metodología empleada para llevar a cabo los objetivos propuestos. En este capítulo, se describen las actividades realizadas, los modelos de evaluación de propiedades fisicoquímicas, ambientales y económicas de las sustancias, la formulación del problema de optimización multiobjetivo y las etapas de construcción e implementación de la metodología MOHAEA-CAMD en el software computacional. Adicionalmente, se hace mención del diseño experimental propuesto para la validación de las moléculas respuesta.

En el capítulo 3, **Resultados y discusión**, se presentan los resultados y análisis de las diferentes corridas experimentales del algoritmo evolutivo multiobjetivo con el modelo CAMD integrado. Se analiza el comportamiento del algoritmo evolutivo (convergencia y desempeño de operadores genéticos), junto con una comparación con los resultados de solventes obtenidos bajo los modelos de Serrato [1] y Rodríguez [2], [3].

Finalmente, se presentan las conclusiones obtenidas en la investigación realizada y las recomendaciones a considerar para el desarrollo de futuras investigaciones relacionadas con el tema. Como anexos, se presenta información complementaria relacionada con la investigación.

Objetivos

Objetivo general

Proponer una metodología de selección de solventes de extracción empleando como herramienta el diseño molecular asistido por computadora (CAMD) y la optimización multiobjetivo como criterio de decisión secuencial.

Objetivos específicos

- Actualizar el algoritmo de construcción molecular para la creación de aromáticos y heterocíclicos, propuesto por Serrato [1] y Rodríguez [2], [3], incorporando la selección de indicadores fisicoquímicos y ambientales.
- Determinar el impacto de las propiedades más relevantes de un solvente en los aspectos fisicoquímicos, ambientales y económicos del diseño.
- Desarrollar el problema de optimización multiobjetivo para la selección de solventes de extracción líquido-líquido, considerando los indicadores mencionados.
- Validar el modelo con el caso de estudio del ácido láctico y ácido acético, mediante la evaluación experimental de parámetros fisicoquímicos de los posibles solventes obtenidos en la optimización.

1. Marco teórico

El marco teórico desarrollado a continuación, busca introducir los conceptos básicos necesarios para el entendimiento de la investigación desarrollada. Inicialmente se parte de una descripción del fenómeno físico presente en la extracción líquido-líquido, junto con una formulación matemática soportada en la termodinámica química y la teoría del equilibrio químico. Adicionalmente, se reseñan las principales problemáticas encontradas al momento de realizar la selección del solvente de extracción, desde el punto de vista fisicoquímico, económico y ambiental. Esto con el fin de contextualizar al lector en los fenómenos básicos asociados con la metodología de selección de solvente a solucionar.

Posteriormente, se realiza una revisión bibliográfica de las principales metodologías de selección de solventes de extracción existentes, junto con sus principales características y limitaciones. Entre las metodologías descritas, se profundiza en los antecedentes del diseño molecular asistido por computadora (CAMD), enfocándose en los parámetros más significativos empleados y en las distintas aplicaciones al problema de selección de solvente, realizado en estudios previos.

A continuación, se hace una descripción de los fundamentos de la teoría formal de la optimización matemática, realizando una breve clasificación de la extensa variedad de modelos de optimización existentes, de acuerdo con la cantidad de funciones objetivo, la cantidad de restricciones existentes, el tipo de variables de optimización presentes y técnicas de solución empleadas. Se introducen conceptos matemáticos tales como las condiciones necesarias y suficientes de un problema de optimización, el criterio de optimalidad de Pareto, las formas de representación gráfica y la complejidad computacional de un algoritmo, entre otras definiciones.

Finalmente, se revisan las principales técnicas de optimización multiobjetivo desarrolladas hasta el momento, haciendo énfasis en las técnicas metaheurísticas y bio-inspiradas. Particularmente, se ahonda en los algoritmos genéticos multiobjetivo: operadores genéticos, secuencia de cálculo y criterios de parada, haciendo una diferenciación sustancial con los algoritmos genéticos empleados en optimización de un solo objetivo. Se incluye una revisión de aspectos relacionados con optimización en múltiples objetivos (*Many-Objective Optimization*, por sus siglas en inglés), teoría matemática de amplio interés actualmente.

1.1 Extracción líquido-líquido

La extracción líquido-líquido, también conocida como extracción por solvente, es una operación de separación en donde existe un reparto de un soluto entre dos fases líquidas inmiscibles (normalmente una orgánica y una acuosa), con distinto grado de solubilidad, y que se encuentran en contacto entre sí mediante una interfaz [7]. Este reparto del soluto⁴ se origina al poner en contacto un agente de separación (solvente⁵ de extracción o extractante, T), con una mezcla líquida (el alimento, F), quien contiene el soluto (S) de interés. Como productos, se obtienen dos fases líquidas inmiscibles, fácilmente separables por decantación: Extracto⁶ (E), y Refinado⁷ (R). Luego de alcanzado un cierto grado de separación (máximo el dado por el equilibrio), es posible incrementar el reparto de los componentes mediante la adición de etapas de extracción adicionales [8].

La extracción con solventes es usada en numerosas industrias químicas para producir gran diversidad de compuestos químicos, desde fármacos y biomedicamentos [7], pasando por polímeros [9], hasta compuestos orgánicos pesados y metales [10][11]. Incluso, en algunos procesos fermentativos, tales como la producción de ácido láctico empleando *Lactobacillus*

⁴ Entiéndase como **Soluto** como aquella sustancia, usualmente sólida (pero puede ser un gas u otro líquido), que se disuelve en una solución.

⁵ Entiéndase como **Solvente** como aquella sustancia química en la que se disuelve un soluto, generando una solución.

⁶ Entiéndase como **Extracto** a aquella fase rica en solvente y pobre en soluto, obtenida producto de la extracción líquido-líquido.

⁷ Entiéndase como **Refinado** a aquella fase líquida residual de donde se separó el soluto, contiene mayor cantidad de soluto y menor cantidad de solvente.

delbrueckii, es posible realizar una intensificación del proceso de separación, acoplando el sistema de separación a la etapa fermentativa [12][13]. Por tal razón, la extracción líquido-líquido es una operación alternativa útil cuando la separación por destilación es ineficaz o muy difícil; siendo útil en mezclas con puntos de ebullición muy cercanos o con presencia de sustancias termolábiles, ya que esta operación unitaria se basa en la separación de componentes a partir de diferencias estructurales, a diferencia de la destilación, en donde la volatilidad relativa de los componentes actúa como criterio de separación [14].

La extracción por solventes se rige termodinámicamente por el equilibrio líquido-líquido (ELL) [15][16], en donde las fugacidades en diferentes fases líquidas deben ser idénticas, de acuerdo con la ecuación (1-1).

$$f_i^I = f_i^{II} \quad \forall i = 1, 2, \dots, NC \quad (1-1)$$

Ambas fugacidades pueden describirse usando coeficientes de actividad o coeficientes de fugacidad [16]. Usando un modelo de coeficientes de actividad, la ecuación (1-1) se convierte en la ecuación (1-2).

$$(x_i \gamma_i f_i^0)^I = (x_i \gamma_i f_i^0)^{II} \quad (1-2)$$

Debido a que el coeficiente de fugacidad estándar, f_i^0 , es el mismo en las dos fases líquidas, la ecuación (1-2) se reduce a la ecuación (1-3).

$$x_i^I \gamma_i^I = x_i^{II} \gamma_i^{II} \quad (1-3)$$

El producto $x_i \gamma_i$ es conocido como actividad (a_i), y representa un criterio de isoactividad para la condición de equilibrio termodinámico en sistemas líquido-líquido. Para el caso de sistemas ternarios, representados en diagramas triangulares, es posible representar la región de dos fases líquidas mediante la curva binodal [7][15], la cual cuenta con las concentraciones x_i^I y x_i^{II} conectadas por el equilibrio mediante las líneas de reparto (*tie-lines*, en inglés). Entre los puntos de la línea es posible definir el coeficiente de distribución K_i entre las dos fases, mediante la ecuación (1-4).

$$K_i = \frac{x_i^I}{x_i^II} \quad (1-4)$$

Existen tres tipos principales de sistemas líquido-líquido: a.) los sistemas cerrados, en donde solo un par binario muestra una brecha de miscibilidad, surgiendo un punto crítico C, en donde ambas fases líquidas cuentan con la misma concentración, b.) los sistemas abiertos, en donde dos pares binarios muestran miscibilidad parcial y c.) los sistemas en donde se encuentra una brecha de miscibilidad o isla en el sistema ternario [15]. Dichos casos son presentados en la **Figura 1-1**. Cabe resaltar que la solubilidad mutua de las especies en una mezcla ternaria en fase líquida se verá afectada apreciablemente por la temperatura, más no de manera significativa por la presión dado el carácter de los solventes como fluidos incompresibles.

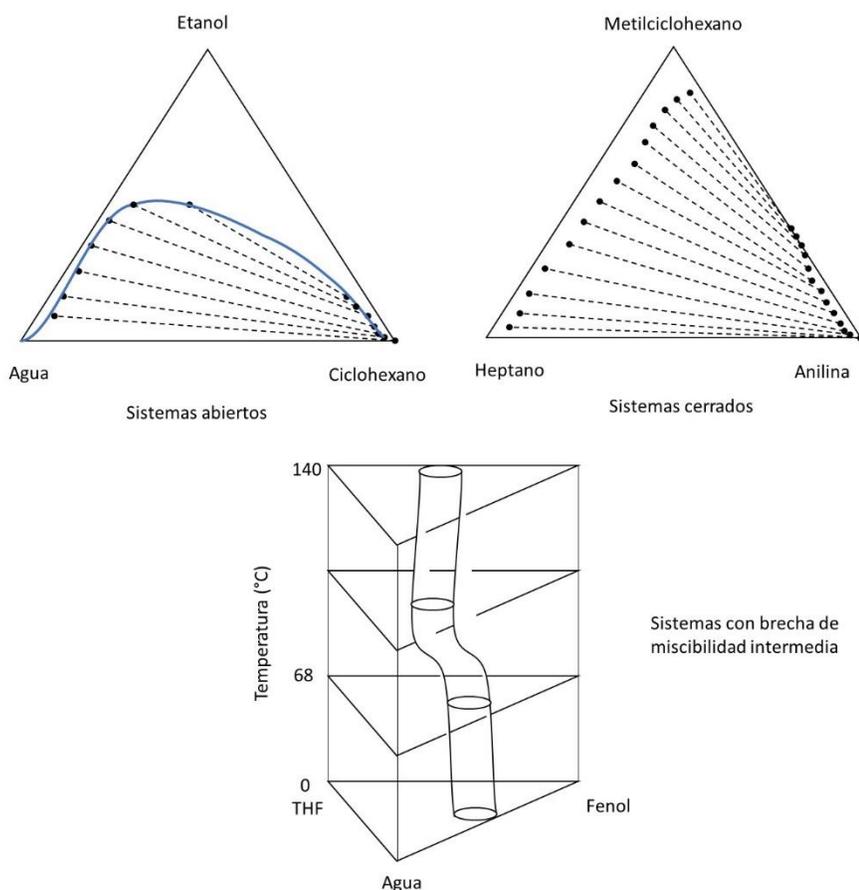


Figura 1-1: Representación del equilibrio líquido-líquido (ELL) para distintos tipos de sistemas. Adaptado de Gmehling [15].

Con el fin de realizar una selección del solvente que favorezca la separación de las fases, se cuenta con una amplia variedad de compuestos factibles, los cuales deben cumplir con ciertas características para ser considerados como potenciales agentes de extracción [8], [17], [18]. Dichas características pueden estar sujeto a enfoques operativos, económicos y medioambientales, por lo cual pueden actuar como principios de preferencia de un solvente en específico en detrimento de otro, al momento de tomar la decisión.

La **Tabla 1-1** presenta una recopilación de los atributos principales con los cuales debe contar una sustancia para ser considerada como un potencial agente de extracción. No obstante, estas características pueden ser priorizadas una respecto a otra, en función de los fines específicos de la separación en el marco del proceso industrial. Por tal razón, la predilección de un potencial solvente sobre otro con características extractivas similares estará condicionado a criterios objetivos, sujetos al proceso químico.

Debido a la gran cantidad de requisitos a cumplir por el compuesto para asegurar una separación eficiente, la noción del solvente más apropiado para cierto sistema debe buscar cumplir la mayor cantidad de características posibles, motivo por el cual es fundamental que la metodología de selección considere la mayor cantidad posible de características deseables de solvente.

Considerando la amplia cantidad de opciones de solventes de extracción presentes para una mezcla soluto/solvente específica, se han desarrollado estrategias de selección teniendo en cuenta las características antes mencionadas. En la Sección 1.1.1 se realiza una recopilación de las principales técnicas de selección del solvente de extracción, conforme a lo descrito por la literatura.

Tabla 1-1: Características deseables en un solvente de extracción [8][17][18][19].

Enfoque	Característica del solvente
Aspectos propios de la operación	<i>Capacidad de extracción:</i> (K): Para seleccionar un solvente óptimo T , el soluto objetivo S debe ser removido del alimento F en el que se encuentra. Para ello, es necesario conocer la distribución de S entre T y F . La forma en que se cuantifica la capacidad de extracción es mediante el coeficiente de distribución K_i .
	<i>Reactividad química:</i> En algunos casos de extracción, es deseable que el solvente sea inerte químicamente frente a los demás componentes del sistema. No obstante, en operaciones de extracción reactiva, pueden darse reacciones débiles.
	<i>Solubilización:</i> El solvente a considerar debe contar con una baja solubilidad en agua.
	<i>Selectividad (S):</i> El solvente debe ser eficiente para separar los componentes de una solución binaria, sin extraer otros componentes en ningún grado.
	<i>Pérdidas de solvente (L) y facilidad de recuperación:</i> Es fundamental que la cantidad de solvente perdido en la fase refinada sea pequeña o que su recuperación sea tan simple y rentable como sea posible.
	<i>Propiedades químicas:</i> Se suelen preferir solventes con diferencias significativas en sus propiedades (principalmente viscosidad y densidad) respecto a los compuestos involucrados en el sistema, con el fin de favorecer el proceso de extracción.
Aspectos económicos	<i>Disponibilidad:</i> El solvente debe ser fácilmente accesible en cantidad suficiente para realizar la separación.
	<i>Bajo costo:</i> El costo del solvente preferiblemente debe ser bajo.
Aspectos medioambientales	<i>Baja toxicidad:</i> El solvente debe de ser de baja toxicidad y/o no tóxico, no contaminar suelos, recursos acuíferos o tener un impacto negativo al medio ambiente.

1.1.1 Metodologías de selección de solventes de extracción

La selección de un solvente específico para una separación particular es un problema directamente relacionado con el equilibrio de fase. La identificación del solvente (o mezcla de solventes), su concentración, temperatura y presión de operación determinan las propiedades del proceso de separación en sí (selectividad, poder solvente, pérdida de solvente, etc.) [20]. Por tal razón, se han desarrollado diversas metodologías para realizar la selección del solvente de extracción teniendo en cuenta las propiedades de interés.

Un enfoque clásico para seleccionar un solvente es examinar una lista de solventes potenciales con respecto a las propiedades de interés (por ejemplo, densidad, viscosidad, punto de ebullición, peso molecular, etc.). En este enfoque, se aborda el problema de selección directamente en la separación, donde se determina la relación solvente/alimentación que permita alcanzar una cierta recuperación. La solución a este problema permitirá identificar el caudal, la selectividad y pérdida de solvente, el coeficiente de actividad del solvente (tanto en fase orgánica como acuosa) y la recuperación de este, una vez finalizada la operación de separación. Ese enfoque, aunque válido, cuenta como desventaja principal la realización de múltiples experimentos por ensayo y error, con el respectivo gasto de recursos para determinar el solvente apropiado [21], [22].

Por otra parte, el enfoque de selección inverso consiste fundamentalmente en sintetizar estructuras moleculares, basadas en un modelo termodinámico de contribución de grupos, partiendo previamente de una especificación de las propiedades deseadas del solvente. En general, los grupos funcionales son los bloques de construcción en la síntesis molecular, aplicando modelos termodinámicos tales como UNIFAC o UNIFAC-DMD para predecir los coeficientes de actividad en fase líquida de los compuestos presentes en una mezcla determinada. Este enfoque generó el desarrollo de nuevas estrategias de selección de solventes, tales como el diseño CAMD, el cual se describe en la Sección 1.2.

1.2 Diseño molecular asistido por computadora (CAMD)

El diseño molecular asistido por computadora (CAMD, por sus siglas en inglés), es una aproximación a la investigación molecular empleando modelos de química teórica [23]. El diseño CAMD es, junto al diseño de experimentos, una herramienta fundamental del diseño de productos químicos, el cual, de acuerdo con Cisternas [24], se define como el

procedimiento de determinación de una sustancia o un conjunto de ellas, con unas ciertas características que permitan suplir una necesidad identificada por el problema de diseño. Estos compuestos serán sintetizados para uno o varios propósitos, por lo que deben cumplir con una serie de requisitos estructurales específicos y unas propiedades características, asociadas a aspectos fisicoquímicos, ambientales y económicos, junto con la interacción de otros componentes presentes en el proceso químico.

Para realizar un diseño de productos apropiado, Moggridge [25] propone un procedimiento simplificado de cuatro etapas de diseño, centradas en aspectos tales como las necesidades identificadas en el mercado, las ideas de posibles productos que puedan suplir dicha necesidad, la selección del producto más apropiado y el proceso de manufactura y evaluación del producto. Este enfoque de diseño se enfoca en el producto a diferencia del diseño de procesos, en el cual se establecen aspectos tales como las condiciones de operación de equipos, los flujos de entrada de materias primas, las condiciones de reacción y separación, los reciclos y otras mejoras del proceso (Intensificación, integración energética, optimización del proceso, etc.). No obstante, estos enfoques de diseño no son mutuamente excluyentes, por lo que es posible, por ejemplo, plantear modelos para intensificación de procesos a partir de diseños CAMD [26].

La línea limítrofe entre el diseño de procesos y el diseño de productos tiende a ser más difusa en tiempos recientes, donde autores como Eden [27], han propuesto sistemas de solución de problemas de diseño simultáneos, integrando metodologías CAMD y CAMPD (*Computer Aided Molecular and Process Design*, por sus siglas en inglés), mediante una formulación general del problema de optimización en términos de variables continuas y discretas relacionadas tanto al diseño de proceso (propiedades fisicoquímicas de las moléculas, bloques constructores para el diseño molecular) como de producto (recuperación, conversión, costo, rendimiento, rentabilidad, etc.). Considerando la mejora en la disponibilidad y eficiencia de los recursos computacionales, los problemas de diseño CAMD han tenido un enorme auge como herramienta de diseño de productos en los últimos 40 años [28].

Tradicionalmente, la ingeniería de sistemas de procesos (PSE, *Process System Engineering*, por sus siglas en inglés) ha empleado modelos matemáticos para la

estimación de propiedades de sustancias, lo cual ha creado una oportunidad enorme para el sector académico e industrial, al generar metodologías y herramientas sistemáticas flexibles y precisas, que permitan suplir la necesidad de sintetizar un compuesto y/o de emplearlo en una etapa de diseño [29]. Tanto las metodologías CAMD como CAMPD buscan proporcionar herramientas para la toma de decisiones, identificación, diseño y manufactura de productos químicos. No obstante, la metodología CAMD tiene enormes oportunidades para la formulación de nuevos tipos de compuestos, como el diseño de compuestos electrolíticos, solventes líquidos iónicos y diseño de mezclas, reactivas o no. Para ello, el diseño CAMD debe mejorar los modelos de predicción de propiedades, conformación de moléculas y eficiencia computacional de algoritmos de optimización asociados [30]. Se han empleado una gran variedad de algoritmos de optimización estocásticos, tales como algoritmos genéticos evolutivos [31], métodos de cálculo basados en química cuántica [32] y otras técnicas de descomposición [33] para la resolución de una amplia gama de problemas de diseño de productos.

Teniendo en cuenta lo anterior, el problema a resolver utilizando una metodología de diseño CAMD puede ser formulado como: *“Dado un conjunto de bloques constructores y unos objetivos específicos de propiedades, determine la molécula o la estructura molecular que reúna dichas propiedades”* [34]. Respecto a esto, también existe un problema inverso que puede ser planteado, correspondiente a la predicción de propiedades en donde la identidad de las moléculas químicas es conocida de forma previa. En este problema inverso, se requiere obtener el valor estimado de las propiedades objetivo a partir de la estructura molecular, empleando bloques constructores y modelos de contribución de grupos (GC). El diseño CAMD puede contemplar múltiples niveles de complejidad molecular en función del problema y la cantidad de propiedades consideradas [21].

Entre las características importantes del diseño CAMD aplicadas en la mayoría de los sistemas de modelamiento se encuentran: la capacidad de construcción de estructuras moleculares a partir de átomos y/o fragmentos de moléculas o mediante la alteración de una estructura preexistente, partiendo de una base de datos. Los modelos pueden ser desarrollados en distintas representaciones, ya sea como vectores que describan los tipos de enlaces, grafos orientados, representaciones esféricas o representaciones en forma condensada (tipo SMILES, InChI, InChIKey, etc.) [23]. Esta es una enorme ventaja, ya que

permite desarrollar tanto análisis de propiedades basadas en la estructura como en análisis geométrico, conformacional y estereoquímica de las moléculas construidas.

Los modelos de estimación de propiedades aplicados en diseños CAMD permiten establecer las propiedades del producto a partir de la descripción estructural, considerando información fisicoquímica, limitaciones estructurales, comportamiento de los enlaces y geometría molecular. Es de gran importancia establecer de forma apropiada los modelos de propiedades, ya que es necesario balancear la precisión y factibilidad estructural con la eficiencia computacional, al momento de identificar las propiedades de las sustancias [30] [34]. En general, para la estimación de propiedades, se aplican procesos de cálculo iterativo basados en solución de modelos de ecuaciones cuasi-empíricas y validación de modelos generados a partir de datos experimentales. La **Tabla 1-2** describe una clasificación general de los distintos tipos de modelos de estimación de propiedades.

Tabla 1-2: Métodos de estimación de propiedades aplicables al diseño CAMD [35] [22] [36] [37].

Selección de grupos básicos	Restricciones de factibilidad	Métodos de estimación de propiedades
Grupos UNIFAC.	Restricciones basadas en grupos de UNIFAC.	QSPR / QSAR.
Grupos Marrero-Gani.	Restricciones basadas en regla del octeto.	Métodos basados en contribución de grupos (GC). Joback GC. Constantinou – Gani GC. Marrero-Gani GC. GC Posicional.
Grupos base para índice de conectividad (CI).	Restricciones basadas en matriz de adyacencia.	Índice de conectividad (CI). <ul style="list-style-type: none"> • Constantinou – Gani CI. • Marrero-Gani CI. - Simulación molecular. - Índices topológicos.

Dada la enorme variedad de especies químicas existentes, ha sido necesario establecer relaciones semi-empíricas cuantitativas estructura-propiedad (modelo QSPR, *Quantitative Structure Properties Relationship*, por sus siglas en inglés) y estructura-actividad (modelo QSAR, *Quantitative Structure Activity Relationship*, por sus siglas en inglés), las cuales son modelos matemáticos que establecen nexos entre propiedades y estructuras moleculares, validados mediante análisis estadístico. Los modelos de relación QSPR y QSAR suelen ser integrados con eficientes algoritmos de optimización numérica, al momento de aplicar la metodología de diseño CAMD [38].

El modelo QSPR más comúnmente empleado es el modelo de contribución de grupos (GC), el cual asume que las propiedades de una molécula pueden ser predichas mediante la suma de las contribuciones de varias subestructuras moleculares, conocidas como grupos. Dada su simplicidad y uso intuitivo, los modelos de contribución de grupos permiten representar una gran cantidad de diversos compuestos químicos, en función de las combinaciones factibles realizada entre una lista de diversos grupos químicos, lo cual, desde una perspectiva del diseño CAMD es fundamental para abarcar la amplia gama de compuestos químicos existentes.

Entre las ventajas de los modelos de contribución de grupos y las relaciones QSPR y QSAR se encuentra la formulación de una amplia gama de propiedades más allá de los aspectos químicos de las moléculas, ya que estos modelos de estimación pueden ser extrapolados fácilmente a aspectos ambientales [39] [40], de salud y seguridad [41]. No obstante, los modelos de contribución de grupos cuentan con ciertas limitaciones, tales como la distinción de las propiedades en isómeros estructurales y la ausencia de parámetros para la predicción de propiedades de grupos complejos [22] [38].

1.2.1 Fundamento matemático del diseño CAMD

El problema asociado al diseño CAMD es una aplicación específica del problema de optimización de una o múltiples funciones, dado un conjunto de restricciones. La descripción canónica general de los problemas de optimización se describe con más detalle en la Sección 1.3.1.

Matemáticamente, Gani *et al.* [42] formulan el modelo matemático de un diseño CAMD, empleado para predicción de propiedades, de acuerdo con el problema (1-5).

$$\begin{aligned}
 & \text{Determine } NG_k, SG(i)_k, NT(i)_k, \quad i = 1, \dots, NG \quad k = 1, \dots, NC \\
 & \text{sujeto a:} \\
 & P_{LB}(j) < (P(j) < P_{UB}(j) \quad j = 1, \dots, NP \\
 & C1(m) = s \quad m = 1, \dots, MC1 \\
 & C2(n) = s \quad n = 1, \dots, MC2
 \end{aligned} \tag{1-5}$$

Donde:

$SG(i)_k$ es el número de identificación de grupo para el grupo i presente en el componente k y perteneciente al conjunto de grupos preseleccionados $SGT(j), j = 1, \dots, NGT$.

$NT(i)_k$ es el número de apariciones de un grupo estructural $SG(i)_k$ en el componente k .

NG_k es el número de diferentes grupos estructurales presentes en el componente k .

$P(j)$ son las propiedades objetivas consideradas como restricciones del problema de diseño CAMD.

NP es el número total de propiedades objetivos contempladas.

$C1(m), m = 1$ y $C2(n), n = 1$ representan las reglas $C1$ de formación de estructuras químicamente factibles y las reglas $C2$ asociadas al diseño final y selección de compuestos.

NC es el número de compuestos formados que satisfacen las restricciones de las reglas $C1$ y $C2$.

Para el caso específico de un diseño CAMD aplicado para la evaluación del problema de selección de solvente, Odele y Macchietto [43] proponen un modelo matemático simplificado, equivalente a la formulación canónica de un problema de maximización clásica, en donde la función objetivo representará al indicador de rendimiento (recuperación de solvente, selectividad, o una combinación de estas). Esta función objetivo dependerá de dos variables principalmente: a.) las condiciones de operación de la extracción (composiciones en el equilibrio de fase, relación solvente/alimento, temperatura, presión, entre otras) y b.) el número y tipo de grupos estructurales presentes en la molécula. Adicionalmente, la maximización puede contar con restricciones asociadas a requerimientos que las moléculas diseñadas deben cumplir (por ejemplo, balance de materia y energía, relaciones entre propiedades físicas, criterios de factibilidad estructural, restricciones lógicas y otras especificaciones de diseño).

Este modelo de maximización clásica entra en la categoría de problemas de optimización MINLP (*Mixed Integer Non-Linear Programming*, por sus siglas en inglés), las cuales tienen gran variedad de estrategias de solución matemáticas. La Sección 1.3 describe con mayor detalle las consideraciones matemáticas para resolver este tipo de problemas.

1.2.2 Etapas del diseño CAMD

El objetivo de emplear técnicas CAMD es identificar los compuestos capaces de desarrollar una serie de tareas específicas, sujetos al estudio y generación computacional de compuestos con un conjunto de características respecto al tipo de componente (familia química), propiedades físicas y químicas [44]. Para ello, el diseño CAMD se divide en tres etapas principales [45], descritas por el flujograma presentado por la **Figura 1-2**. A continuación, se realiza una breve descripción de las particularidades asociadas a las etapas metodológicas del diseño CAMD.

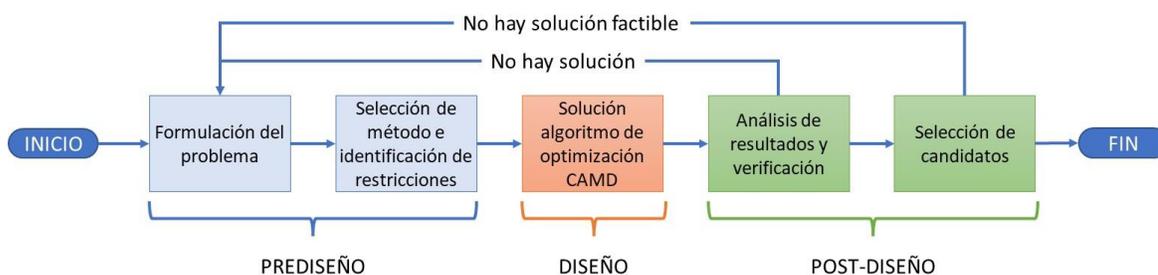


Figura 1-2: Etapas metodológicas del diseño CAMD. Adaptado de Roughton [45].

- **Pre-diseño:** El método CAMD inicia por la definición de los objetivos globales del diseño, los cuales consisten en una función general que la molécula candidato debe cumplir mediante sus características. Basado en este objetivo global, la formulación de las restricciones de diseño, de acuerdo con el conocimiento de las propiedades de interés de los compuestos candidatos, es clave para acotar y enfocar el diseño de compuestos hacia aquellos que tienen una aplicabilidad real a nivel experimental [44]. Entre los aspectos considerados en dicha base teórica inicial se encuentran la identificación de propiedades físicas, químicas y termodinámicas para sustancias puras y mezclas, interacciones con equilibrio de fases, costos, disponibilidad en el mercado, entre otros aspectos. Es fundamental obtener un rango aproximado de operabilidad de las propiedades de los componentes del sistema [37]. Entre las principales preguntas a considerar en la etapa de Pre-diseño se encuentran: ¿Qué información estructural se requiere? ¿Cuáles son las propiedades por considerar como objetivos del diseño? ¿Como será la representación estructural usada en el algoritmo de optimización? [45].

- **Diseño:** El método de solución empleado por el diseño CAMD es un híbrido de un algoritmo de generación y prueba de tipos de componentes, en donde todas las moléculas factibles son generadas a partir de un conjunto de bloques constructores, con los cuales se pueden evaluar las propiedades de los compuestos, una vez estos han sido generados [44]. Con el fin de evitar una explosión combinatorial, se han desarrollado técnicas de generación y evaluación de compuestos de manera sucesiva que incrementan el detalle molecular sólo para los candidatos más prometedores [37]. La etapa de diseño es la etapa dominante, influenciando tanto la formulación del problema como las opciones para el análisis en la etapa del Post-diseño. Entre las principales preguntas a considerar en la etapa de diseño se encuentran: ¿Qué estructuras moleculares deben ser generadas y representadas? [45].

- **Post-diseño:** En la etapa de Post-diseño, se responden las preguntas formuladas como hipótesis de estudio para el estudio CAMD, analizando los compuestos obtenidos con respecto a las propiedades y el comportamiento considerado en la etapa del diseño. Algunas propiedades consideradas como criterios de evaluación de compuestos son: precio, disponibilidad, restricciones legislativas, rendimiento dentro del proceso e impacto ambiental. En esta etapa, se involucran modelos predictivos, bases de datos y, si es posible, simulaciones con el fin de obtener una descripción de las capacidades de los componentes diseñados para el caso particular [44]. Entre las principales preguntas a resolver en esta etapa se encuentran: ¿El comportamiento de los solventes diseñados es el esperado? ¿Es posible verificar el comportamiento del solvente diseñado mediante simulación o validación experimental?

1.2.3 Aplicación del diseño CAMD en el diseño de procesos y productos

Durante los últimos 40 años, el uso de diseño CAMD ha ido en auge, conforme las necesidades de diseño de procesos y productos aumentan, considerando la transformación de la industria química en sí misma desde una óptica centrada en el proceso a una centrada en el producto [28]. A medida que las herramientas

computacionales han mejorado en su capacidad de procesamiento y almacenamiento, se ha aplicado esta metodología de diseño de forma más exhaustiva.

Por una parte, se han realizado diversos estudios [46], [47], [48], [49], [50], [51], [52] focalizados en el diseño y selección de solventes para cumplir con las demandas operativas de diversas operaciones de separación en procesos, empleando el diseño CAMD para la obtención de moléculas candidatos, según los objetivos requeridos por la operación unitaria. Por otro lado, también existe literatura enfocada [31], [45], [53], [54] en el diseño de productos y la determinación de las características más apropiadas para nuevos materiales a partir de consideraciones estructurales, tales como polímeros y medicamentos.

La **Tabla 1-3** describe los principales estudios de diseños CAMD aplicados a diseño de productos, de acuerdo con lo reportado en literatura. Se empleó como criterio de selección que la investigación en cuestión contara con las tres etapas del diseño CAMD descritas, y se tuvo preferencia sobre los estudios que hayan realizado validación experimental con los candidatos obtenidos computacionalmente.

De manera similar, la **Tabla 1-4** enlista los principales estudios de diseño CAMD aplicados a diseño de procesos, clasificados según la operación de separación objeto del diseño CAMD exceptuando extracción líquido-líquido, la cual se detallará a profundidad más adelante en la Sección 1.2.4.

Cabe resaltar que, en algunos estudios [47] [48] [49], los autores realizan una comparación de los compuestos obtenidos por su investigación mediante el software ProCAMD/ICAS, desarrollado por Gani [55]. Esta herramienta computacional permite la solución de problemas establecidos de diseño CAMD, predicción de propiedades, modelamiento de producto y proceso, empleando la base de datos CAPEC. Un Toolbox de la herramienta, ICAS (*Integrated Computer Aided System*, por sus siglas en inglés) permite realizar la aplicación y evaluación de diseños de moléculas CAMD en simulaciones de procesos.

Tabla 1-3: Aplicaciones CAMD enfocados al diseño de productos [45] [53] [54].

Enfoque	Aplicación	Especificaciones del modelo CAMD	Ref.
Diseño de productos	Formulación de excipientes para proteínas hidrofóbicas Diseño de solventes iónicos aplicados en bioseparaciones (destilación extractiva + fermentación extractiva in-situ)	<ul style="list-style-type: none"> - <u>Propiedades contempladas:</u> Coeficiente de fugacidad y actividad, la presión de vapor, entre otras propiedades físicas. - <u>Función Objetivo:</u> Modelo de regresión lineal multivariable de datos de propiedades calculadas vs experimentales. - <u>Restricciones:</u> Parámetros de solubilidad (predicho y objetivo) del solvente. - <u>Modelo de contribución de grupos (GC):</u> UNIFAC-IL. - <u>Caso de estudio:</u> Sistemas Acetona-Metanol y Etanol-Agua. - <u>Algoritmo de optimización:</u> Representación de grafos para la captura de topología bidimensional y tridimensional, integrado con algoritmo genético. - <u>Moléculas obtenidas:</u> Sistema Acetona-Metanol: 1-octil-4-metilpiridinium triflorometano sulfonato. Sistema Etanol-Agua: 1,3-dimetilimidazolium dimetilfosfato y 1,3-dimetilimidazolium tetrafluoroborato. - <u>Verificación experimental:</u> Espectrofotometría UV, cromatografía SEC, electroforesis con gel de poliácridamida (SDS-PAGE), espectroscopia de masas por intercambio de hidrógeno-deuterio (HX-MS) 	[45]
	Diseño de una estructura polimérica.	<ul style="list-style-type: none"> - <u>Propiedades contempladas:</u> Densidad, temperatura de transición vítrea, coeficiente de expansión térmica, capacidad calorífica específica y módulo de Young del polímero. - <u>Función Objetivo:</u> Relación de complejidad de la cadena principal y las cadenas laterales. - <u>Restricciones:</u> Máximo 17 grupos en la cadena principal. Y 15 grupos en la cadena lateral. - <u>Modelo de contribución de grupos (GC):</u> Van Krevelen. - <u>Caso de estudio:</u> No aplica. - <u>Algoritmo de optimización:</u> Algoritmo genético, con los operadores genéticos clásicos. - <u>Moléculas obtenidas:</u> Polietileno tereftalato (PET), el poliviniliden propileno (PVP) y el policarbonato de bisfenol-A (PC). - <u>Verificación experimental:</u> No contempla. 	[53]
	Diseño de un detergente personal basado en emulsiones	<ul style="list-style-type: none"> - <u>Propiedades contempladas:</u> Costo, Inflamabilidad, habilidad de formación de espuma, solubilidad, cuidado de piel, propiedades mecánicas, estabilidad, toxicidad. - <u>Función Objetivo:</u> Combinación lineal de expectativas de los consumidores de emulsiones de lavado (económicas, solubilidad, estabilidad y seguridad). - <u>Restricciones:</u> Límites máximos permitidos en las propiedades contempladas. - <u>Modelo de contribución de grupos (GC):</u> UNIFAC. - <u>Caso de estudio:</u> Detergente personal basado en emulsiones (ingredientes activos + emulsificante + aditivos). - <u>Algoritmo de optimización:</u> Algoritmo genético. - <u>Moléculas obtenidas:</u> Principio activo: Laurel sulfato de sodio, óxido de octil esaetileno. Fase continua: Agua, Fase dispersa: Aceite de jojoba. Emulsificantes: Cocomidopril betaina. Aditivos: Aceite de rosa (aroma), C117200 (pigmento), triclosán (agente antimicrobiano), policuaternium 7 (estabilizador), benzoato de sodio (preservante). - <u>Verificación experimental:</u> Se evaluó un análisis de estabilidad de la mezcla de componentes obtenidos, sujeto a desviaciones hidrofílicas-lipofílicas para cada surfactante. 	[54]

Tabla 1-4: Aplicaciones CAMD enfocados al diseño de procesos para diversas operaciones unitarias [46] [47] [48] [49] [50] [51] [52].

Enfoque	Operación unitaria	Aplicación	Especificaciones del modelo CAMD	Ref.
Diseño de procesos	Destilación extractiva	Determinación de agente de extracción para el sistema acetona-metanol	<ul style="list-style-type: none"> - <u>Propiedades contempladas:</u> Puntos de ebullición y entalpía de vaporización. - <u>Función Objetivo:</u> Maximización de la fuerza impulsora de la operación de destilación extractiva. - <u>Restricciones:</u> Restricciones termodinámicas, balance de materia y energía. - <u>Modelo de contribución de grupos (GC):</u> Hukkerikar y UNIFAC. - <u>Caso de estudio:</u> Sistema acetona-metanol. - <u>Algoritmo de optimización:</u> LINDOGlobal (Solver MINLP, disponible en el software GAMS). - <u>Moléculas obtenidas:</u> DMSO (dimetilsulfóxido), agua y etilenglicol. - <u>Verificación experimental:</u> Validación mediante simulación del proceso a partir de datos experimentales. 	[46]
		Diseño de solvente de extracción en cinco casos de estudio.	<ul style="list-style-type: none"> - <u>Propiedades contempladas:</u> Volatilidad relativa, punto de ebullición y de congelación. - <u>Función Objetivo:</u> Sumatoria de funciones sigmoideas para cada componente (valor de la propiedad calculado vs requerido), aplicando pesos a las propiedades. - <u>Restricciones:</u> Restricciones estructurales y tamaño de las moléculas. - <u>Modelo de contribución de grupos (GC):</u> UNIFAC. - <u>Caso de estudio:</u> 5 sistemas: etanol/agua, acetona/metanol, etanol/etilacetato, metanol/metilacetato, benceno/ciclo hexano. - <u>Algoritmo de optimización:</u> Messy GA empleando el software SolvGen. - <u>Moléculas obtenidas:</u> Depende del sistema considerado. - <u>Verificación experimental:</u> Se realizaron verificaciones experimentales de las propiedades de las moléculas candidato, de acuerdo con el sistema, constatándose buenos resultados para los candidatos obtenidos. Adicionalmente, se realizó el estudio comparativo de los compuestos obtenido con la herramienta ProCAMD. 	[47]
		Diseño de solvente para purificación del bioetanol con criterios multiobjetivo.	<ul style="list-style-type: none"> - <u>Propiedades contempladas:</u> Punto de ebullición y flash, coeficiente de partición octanol/agua, toxicidad LC50. - <u>Función Objetivo:</u> 2 funciones, asociadas a seguridad inherente y costo total anualizado. - <u>Restricciones:</u> Restricciones económicas, de seguridad, geométricas y operativas. - <u>Modelo de contribución de grupos (GC):</u> Gani. - <u>Caso de estudio:</u> Purificación de bioetanol. - <u>Algoritmo de optimización:</u> Optimización libre de derivadas (DFO), con un algoritmo combinado de templado simulado (para sistemas de destilación extractiva compleja), acoplado a un algoritmo genético multiobjetivo (diseño de columnas térmicamente acopladas). Se reportan las soluciones óptimas de Pareto. - <u>Moléculas obtenidas:</u> Dietilenglicol, etilenglicol, dipropilenglicol, dimetilformamida. - <u>Verificación experimental:</u> Se realizó verificación de las condiciones de operación obtenidas mediante simulación del proceso, empleando el software comercial Aspen Plus. Adicionalmente, se realizó el estudio comparativo de los compuestos obtenido con la herramienta ProCAMD. 	[48]

Cont. Tabla 1-4.

Enfoque	Operación unitaria	Aplicación	Especificaciones del modelo CAMD	Ref.
Diseño de procesos	Destilación extractiva	Identificación y aplicación de solventes de extracción para una red de síntesis sistemática de destilación extractiva	<ul style="list-style-type: none"> - <u>Propiedades contempladas:</u> coeficientes de actividad, temperatura mínima de ebullición, azeotropía. - <u>Función Objetivo:</u> Función combinada que integra selectividad mínima posible y capacidad de arrastre. - <u>Restricciones:</u> Regla del octeto y estructurales, cantidad de grupos funcionales. - <u>Modelo de contribución de grupos (GC):</u> UNIFAC. - <u>Caso de estudio:</u> Sistema acetona-metanol. - <u>Algoritmo de optimización:</u> Algoritmo de relajación sucesiva SR-MINLP, convertir el problema de MINLP a NLP. - <u>Moléculas obtenidas:</u> DMSO, etilenglicol, isopropanol, etanol. - <u>Verificación experimental:</u> No se realizó verificación experimental de los compuestos obtenidos con la herramienta ProCAMD. 	[49]
Diseño de procesos	Destilación extractiva	Diseño óptimo de solventes para destilación extractiva, desde un punto de vista multiobjetivo.	<ul style="list-style-type: none"> - <u>Propiedades contempladas:</u> Temperatura de ebullición y fusión del solvente. - <u>Función Objetivo:</u> Selectividad, potencial de extracción de un solvente. - <u>Restricciones:</u> Estructurales, límites de propiedades físicas, regla del octeto y regla de enlace modificada. - <u>Modelo de contribución de grupos (GC):</u> UNIFAC-Dortmund (UNIFAC-DMD). - <u>Caso de estudio:</u> Sistema n-Hexano-Metanol. - <u>Algoritmo de optimización:</u> Algoritmo genético. - <u>Moléculas obtenidas:</u> Pareto óptimos: butane-1,4-diamina, DMSO (dimetilsulfóxido), 1,2-etanediol, 1,4-butanediol, glicerol y agua. - <u>Verificación experimental:</u> Se realizó verificación de los resultados mediante la simulación del proceso, empleando cada uno de los solventes, en el simulador Aspen Plus. 	[50]
	Extracción reactiva	Diseño óptimo de solventes para procesos de destilación extractiva	<ul style="list-style-type: none"> - <u>Propiedades contempladas:</u> Biocompatibilidad, Punto de fusión y ebullición normalizado. - <u>Función Objetivo:</u> Maximizar la producción de acetato de hexilo. - <u>Restricciones:</u> Estructurales, límites en las propiedades contempladas, funciones de penalización en caso del no cumplimiento de restricciones. - <u>Modelo de contribución de grupos (GC):</u> UNIFAC. - <u>Caso de estudio:</u> Reacción de esterificación del ácido acético con 1-hexanol, para producción de acetato de hexilo y agua. - <u>Algoritmo de optimización:</u> Algoritmo Genético, con operadores genéticos clásicos. - <u>Moléculas obtenidas:</u> 1,1-dicloro-2-metilpropano en términos globales, Isoctano como compuesto con mayor conversión. - <u>Verificación experimental:</u> No se realizó verificación experimental de los resultados obtenidos, ni una validación mediante simulación del proceso. 	[56]
	Cristalización	Diseño de solvente para cristalización de ácidos carboxílicos	<ul style="list-style-type: none"> - <u>Propiedades contempladas:</u> Solubilidad del soluto, potencial de recuperación del soluto, morfología del cristal, toxicidad del solvente, puntos de fusión, flash y ebullición del solvente. - <u>Función Objetivo:</u> Potencial de recuperación de cristal. - <u>Restricciones:</u> Estructurales, límites en los puntos de fusión y ebullición de solventes, toxicidad, seguridad en flash point, restricción de solubilidad. - <u>Modelo de contribución de grupos (GC):</u> Gani. - <u>Caso de estudio:</u> Obtención de ácido sebácico. - <u>Algoritmo de optimización:</u> Método de descomposición basado en soluciones. - <u>Moléculas obtenidas:</u> 2-metoxietanol. - <u>Verificación experimental:</u> Se realizó caracterización del cristal, mediante micrografía óptica (OM) y difracción de rayos X en polvo. 	[52]

1.2.4 Uso del diseño CAMD en extracción líquido-líquido

Se han identificado una gran variedad de estudios [57], [58], [59], [60], [61] [62], que relacionan al diseño CAMD como una herramienta computacional válida para la determinación del solvente en operaciones de extracción líquido-líquido. Conforme ha incrementado la potencia computacional disponible para resolver los problemas de optimización asociados, se ha aumentado la cantidad de propiedades evaluadas, la interacción entre los diseños CAMD y CAMPD [57], junto con los enfoques operacional, ambiental, sanitario y de seguridad de forma simultánea [41], [48], [58], [59], [63]. Adicionalmente, se han resuelto los problemas de optimización mediante modificaciones de algoritmos genéticos [62], [61], [64], templado simulado [60], búsqueda tabú [65], entre otros métodos de optimización estocásticos.

Skiborowski [57] desarrolla un diseño molecular y de procesos simultáneo (CAMPD, por sus siglas en inglés), empleando modelos termodinámicos avanzados, tales como las proporcionadas por las predicciones termodinámicas basadas en COSMO-RS, en donde se pueden evaluar solventes alternativos para procesos de extracción líquido-líquido, no solo en función de su selectividad y capacidad, sino también teniendo en cuenta la cantidad mínima de solvente fresco y la energía empleada para la recuperación del solvente. Adicionalmente, a diferencia de los métodos clásicos de selección de solventes, los cuales se encuentran enfocados en la determinación de parámetros en una sola etapa (distribución de solutos y selectividad del solvente a dilución infinita), la estimación de propiedades en diseños CAMPD puede integrar la evaluación de candidatos a solventes en procesos multietapa, con ciclos cerrados que consideren la recuperación del solvente, mejorando la estimación de los costos operativos del proceso.

Patel [58], en su tesis doctoral, propone incorporar el diseño inherentemente seguro en el diseño de procesos peligrosos relacionados con operaciones con solventes (tales como la extracción líquido-líquido), asociando el diseño molecular a técnicas de diseño seguro, con el fin de poder seleccionar compuestos químicos más seguros, y escoger unos parámetros de diseño más apropiados para las restricciones de diseño. Para ello, parte de la base de datos presente en el software ProCAMD/ICAS, para llevar la simulación de los escenarios peligrosos un paso más adelante, empleando guías de la Agencia de Protección Ambiental de Estados Unidos (*EPA*, por sus siglas en inglés). Se resuelve el problema de

optimización del diseño CAMD tanto para el criterio operativo, como el de seguridad ambiental, con un enfoque en estos dos objetivos. Los modelos de contribución de grupos considerados son construidos a partir de técnicas de predicción de propiedades QSPR. Para aplicar la metodología propuesta, se tomó como caso de estudio el sistema ácido acético-agua.

Ooi [59] propone una estrategia de identificación de solventes mediante el modelo CAMD empleando consideraciones en dos objetivos, asociados a la capacidad de extracción del soluto y la recuperación de solvente de manera simultánea. Se hace un énfasis en la efectividad de la operación unitaria, junto con aspectos de seguridad, salud e impacto ambiental. Para la resolución del problema de optimización aplica un proceso de jerarquía analítica difusa (FAHP, por sus siglas en inglés) empleando un enfoque multiobjetivo consistente con pesos balanceados para los parámetros de ambos objetivos, empleando los modelos de estimación de propiedades de Hukkerikar [66]–[68]. Se evalúan el caso de estudio para diseñar un solvente para la industria del aceite de palma a partir de la extracción de aceite residual desde la fibra de palma, a partir de tres rutas de extracción diferentes (evaporación normal, evaporación al vacío y cristalización por enfriamiento).

Ourique [60] propone una metodología para la identificación de mezclas de sustancias mediante CAMD empleando una representación molecular de grafos y templado simulado para la optimización, ilustrando la aplicación mediante dos casos de estudio (selección de solvente para recuperación de n-butanol en un sistema de destilación extractiva y la búsqueda de un nuevo refrigerante para bombas en reemplazo de los clorofluorocarbonados (CFC).

En su tesis doctoral, Heintz [62], [61] propone una herramienta CAMPD para la determinación de propiedades objetivo de mezclas para obtener moléculas sustitutas de solventes para procesos de impresión litográfica, reemplazo de parafinas cloradas en la industria de plastificantes y alternativas para la extracción natural de antioxidantes. Para ello, aplica una modificación al algoritmo genético respecto a los operadores genéticos tradicionales de cruce y mutación, con el fin de integrar operadores sustituyentes de grupos cíclicos. El enfoque del problema de optimización es de una única función objetivo, que evalúa la suma ponderada de las propiedades.

Con respecto a los sustituyentes de grupos cíclicos aromáticos, Devillers [64] propone adicionar parámetros como los índices de constantes π , las habilidades de la molécula como aceptor de enlaces de hidrógeno (HBA) y donador de enlaces de hidrógeno (HBD) y otros parámetros de resonancia e inductivos de moléculas cíclicas. Adicionalmente, presenta propuestas para modificación de operadores relacionados con este tipo de compuestos, tales como adición, delección y escisión de las estructuras cíclicas.

Lin [65] realiza una implementación del algoritmo de búsqueda tabú para la resolución mediante método CAMD para identificación de metales de transición empleables como catalizadores. Se usaron los índices de conectividad como métodos de predicción de propiedades para el problema de optimización. El diseño buscó minimizar la diferencia entre los valores promedio de las propiedades, asegurándose el cumplimiento de las valencias y propiedades para distintos tipos de enlaces, mediante la definición de restricciones sobre la búsqueda de vecindades durante 200 iteraciones. Se evaluaron dos casos de estudio, asociados al diseño de un catalizador de molibdeno para una reacción de epoxidación con 8 y 10 grupos básicos.

De acuerdo con lo mencionado en la sección introductoria del presente trabajo, y considerando los nuevos antecedentes bibliográficos antes referenciados, se toman como referentes de línea base los estudios realizados a nivel local⁸, aplicados al estudio de metodologías de diseño CAMD de solventes óptimos, desarrollados por Serrato [1] y Rodríguez [2], [3], junto con las variantes propuestas por Medina-Mora [19]. En estos trabajos, se aborda el problema de diseño de solventes para sistemas binarios asociados a efluentes de fermentación (ácido láctico-agua y ácido acético-agua), mediante la optimización de un solo objetivo de diversas funciones de optimización empleando un algoritmo genético modificado, enfocado a aspectos operacionales, ambientales y de disponibilidad de mercado. Dado que el presente trabajo será una extensión generalizada para problemas de optimización multiobjetivo, con integración de grupos aromáticos y

⁸ Por nivel local, refiérase al Grupo de Investigación en Procesos Químicos y Bioquímicos del Departamento de Ingeniería Química y Ambiental, Universidad Nacional de Colombia (Sede Bogotá)

heterocíclicos, las diferencias presentadas respecto a los modelos presentados por estos dos autores es descrita con amplitud en las Secciones 2.2.1, 2.2.3, 2.2.7, 2.2.12 y 2.2.13.

1.3 Optimización en un solo objetivo (SOO)

La optimización, en términos simples, consiste en hacer las cosas mejor, generalmente a partir de un estado inicial haciendo ajustes a partir de la información ganada en el proceso. En términos formales, la optimización es una rama de la matemática enfocada en la determinación e identificación del mejor elemento, entre un conjunto de elementos disponibles, con respecto a un criterio o un conjunto de ellos. Para ello, el problema a resolver consiste en hallar el valor máximo o mínimo de una función $f(\bar{X})$, seleccionando de manera sistemática las variables de entrada \bar{X} , pertenecientes a un conjunto permitido o espacio de búsqueda [69].

La presente sección hace una descripción brevísima sobre conceptos clásicos [70] de optimización en un solo objetivo (formulación canónica, clasificación y técnicas de solución matemática), con el fin de sugerir un marco introductorio al tipo de optimización que se estudiará en el presente documento (Optimización multiobjetivo, descrita en la Sección 1.4). Adicionalmente, la optimización en un solo objetivo brinda fundamentos matemáticos básicos para la optimización en múltiples objetivos, estado de arte actual de vanguardia, la cual se introducirá en la Sección 1.5.

1.3.1 Formulación canónica y conceptos claves de optimización

Los problemas de optimización [71] son descritos por una función objetivo (también llamada función de costos, o función de utilidad) f , definida en el subconjunto $D \subseteq \mathbb{R}^n$ y un conjunto de restricciones $\Omega \subseteq D$. Los elementos \bar{x} pertenecientes al conjunto de restricciones Ω son llamados soluciones factibles. Se busca encontrar entonces una solución $\bar{x} \in \mathbb{R}^n$ que satisfaga la ecuación (1-6) (para el problema de minimización sin restricciones), o la ecuación (1-7) (en el caso del problema de minimización restringida). Por definición, los valores de las restricciones Ω se encuentran acotadas y típicamente son representadas por las funciones $g_i(x)$ para las restricciones de desigualdad y $h_i(x)$ para las restricciones de igualdad.

$$f(\bar{x}) \leq f(x^*), \quad x^* \in \mathbb{R}^n \quad (1-6)$$

$$f(\bar{x}) \leq f(x^*), \quad x^* \in \Omega \quad (1-7)$$

De forma análoga, el problema de maximización puede ser descrito tanto para el problema sin restricciones como restringido, de acuerdo las ecuaciones (1-8) y (1-9), respectivamente. No obstante, la formulación canónica del problema de optimización se describe en términos de la minimización de la función objetivo. Para convertir un problema de maximización al formato canónico, basta aplicar la ecuación (1-10).

$$f(\bar{x}) \geq f(x^*), \quad x^* \in \mathbb{R}^n \quad (1-8)$$

$$f(\bar{x}) \geq f(x^*), \quad x^* \in \Omega \quad (1-9)$$

$$\min_{\bar{x}} f(\bar{x}) = \max_{\bar{x}} -f(\bar{x}) \quad (1-10)$$

Una solución x^* del problema de minimización es llamado minimizador. El valor de f al ser evaluado en x^* es llamado mínimo global. Como una caracterización de minimizadores es posible considerar también los minimizadores locales los cuales satisfacen la ecuación (1-11), para cada $\delta > 0$.

$$f(\bar{x}) \leq f(x), \quad x \in \Omega, \quad \|x - \bar{x}\| \leq \delta \quad (1-11)$$

La existencia de al menos un minimizador para una función es garantizada por el Teorema de Weierstrass [71], el cual indica que si la función f es continua en un intervalo compacto (cerrado y acotado) $[a, b]$, entonces hay al menos dos puntos x_1, x_2 pertenecientes a $[a, b]$ donde f alcanza valores extremos absolutos, es decir $f(x_1) \leq f(x) \leq f(x_2)$, para cualquier $x \in [a, b]$. El teorema de Weierstrass se cumple para funciones definidas con valores en los números reales, y puede ser generalizado para funciones con codominio distinto de \mathbb{R} .

Tanto para la solución de problemas de minimización con restricciones como sin restricciones, existen una serie de condiciones de optimalidad, conocidas como condiciones de Karuch-Kuhn-Tucker (*KKT-conditions* en inglés) [72], descritas por las ecuaciones (1-12) y (1-13) de la **Tabla 1-5**, las cuales son de dos tipos: condiciones necesarias y condiciones suficientes, donde $g_i(x)$ y $h_j(x)$ representan a cada una de las condiciones de desigualdad e igualdad del problema. Por su parte, los valores μ_i y λ_i

corresponden a los multiplicadores de Lagrange para las restricciones de desigualdad e igualdad, respectivamente; mientras que $\mathcal{L}(x, \mu, \lambda)$ representa al Lagrangeano de la función objetivo [71] [72].

Tabla 1-5: Condiciones de optimalidad de Karuch-Kuhn-Tucker para minimizaciones con una sola función objetivo [71] [72].

<u>Condición Necesaria:</u>	
$\nabla f(x^*) + \sum_{i=1}^m \mu_i \nabla g_i(x) + \sum_{j=1}^l \lambda_j \nabla h_j(x) = 0$	(1-12)
<u>Condición Suficiente:</u>	
Sea:	
$\mathcal{L}(x, \mu, \lambda) = f(x) + \sum_{i=1}^m \mu_i g_i(x) + \sum_{j=1}^l \lambda_j h_j(x)$	
Luego	(1-13)
$s^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \mu^*) s \geq 0$	
Donde	
$[\nabla_x g_i(x^*), \nabla_x h_j(x^*)]^T s = 0$	
Con $s \neq 0$.	

1.3.2 Clasificación del problema de optimización

Los problemas de optimización pueden ser clasificados de múltiples maneras, de acuerdo con la presencia (o ausencia) de restricciones, la cantidad de funciones objetivo y la naturaleza de las funciones presentes en el problema [73]. La **Figura 1-3** presenta una clasificación general de las distintas formas de clasificar el problema de optimización, junto con un listado de algoritmos y métodos de optimización más comunes utilizados en su solución.

- Inicialmente, es posible dividir los problemas con respecto a las funciones de restricción, existen problemas no restringidos (*unconstrained*) y restringidos (*constrained*), los cuales pueden tener una o varias restricciones de igualdad y/o desigualdad en función de las necesidades definidas en la formulación del problema.

En el caso particular de problemas de optimización relacionados con el diseño CAMD, descritos en la Sección 1.2.1, la ecuación (1-5) contempla la formulación de restricciones $P(j)$ relacionadas con las propiedades objetivas de las moléculas, junto con las restricciones $C1(m)$ y $C2(n)$ asociadas a las reglas de factibilidad estructural y selección de compuestos.

- Por otra parte, los problemas de optimización pueden dividirse en optimización de un solo objetivo (más conocidas por su término en inglés: *single-objective* o *single-criterion*) debido a que solo cuenta con una sola función objetivo a ser optimizada. Si el problema diseñado involucra múltiples funciones objetivo, es llamado problema de optimización multiobjetivo (o multicriterio), en donde los objetivos a minimizar (o maximizar) deben ser resueltos de manera simultánea. La solución del problema de optimización de un solo objetivo cuenta con una serie de ventajas respecto a la optimización multiobjetivo, tales como: menor complejidad de modelamiento matemático, mayor facilidad para implementación y menores tiempos de solución computacional. No obstante, suelen considerarse los problemas de optimización de un solo objetivo como una descripción parcial de los problemas reales de ingeniería, dada la interconectividad de criterios de optimalidad, muchas veces en conflicto entre sí (por ejemplo, para el caso de la determinación de solventes de extracción, los aspectos económicos, operativos y ambientales pueden tener conflictos entre sí). Tradicionalmente, se acota el problema de optimización multiobjetivo a 2 o 3 funciones objetivo, mientras que para problemas con más de 3 funciones objetivos existe una subcategoría conocida como optimización en múltiples objetivos (más conocida por su nombre en inglés: *Many-Objective Optimization* o *MaOO*). La Sección 1.4 describe con detalle aspectos clave respecto a la optimización multiobjetivo, de suma importancia para el desarrollo de esta tesis, mientras que la Sección 1.5 presenta el estado del arte actual asociado a optimización en múltiples objetivos.
- Con respecto al tipo de variables que presentan los problemas de optimización, se pueden clasificar en problemas que involucran solamente variables continuas y en problemas con variables tanto continuas como discretas (mixto entero, MI). Respecto a la naturaleza de las expresiones matemáticas de las funciones objetivo y las restricciones, los problemas pueden ser clasificados como lineales (LP) o no lineales

(NLP). Integrando estos dos aspectos, se obtienen problemas de optimización mixto entero lineal (MILP), mixto entero no lineal (MINLP), dentro de los cuales se agrupan gran cantidad de problemas de optimización en ingeniería. Además, si las funciones objetivo se comportan de forma cuadrática, el problema es denominado programación cuadrática (QP). Esta clasificación es muy útil desde un punto de vista computacional, debido a que permite establecer algoritmos de solución específicos para cada tipo de problema, teniendo en consideración la complejidad computacional y, por ende, el tiempo de solución numérica del problema empleando un algoritmo en particular [74].

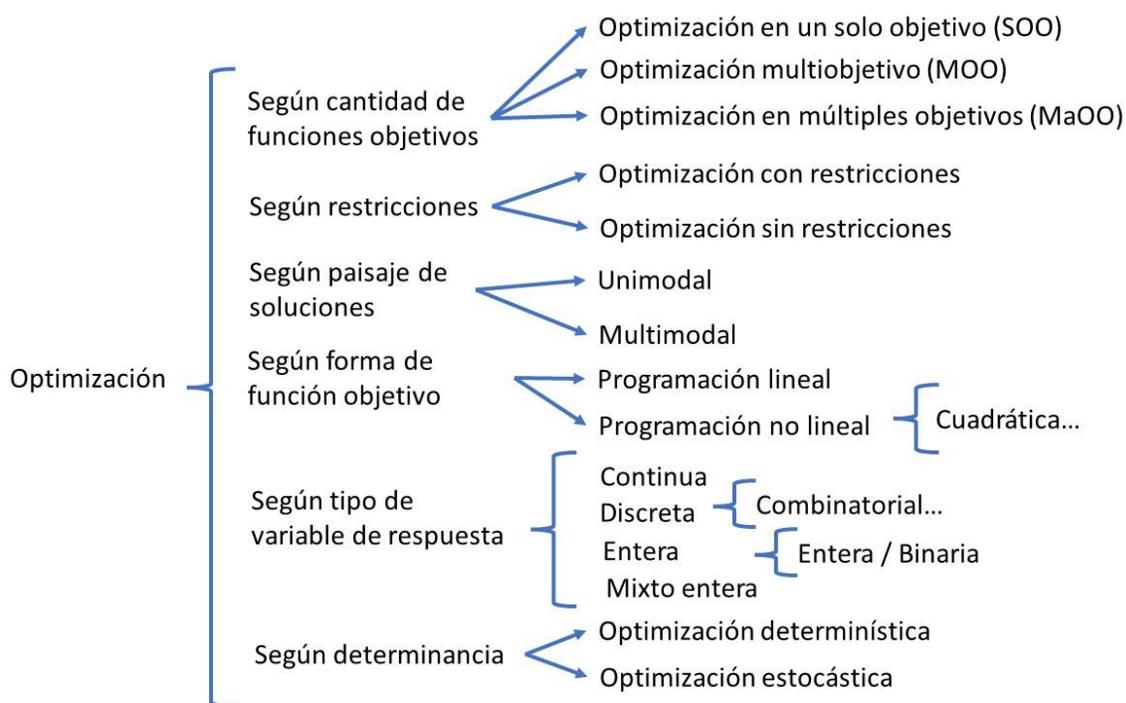


Figura 1-3: Clasificación de problemas de optimización [74] [75].

De acuerdo con el problema, las distintas técnicas de solución presentadas en la **Figura 1-3** pueden agruparse en dos enfoques principales: 1.) El enfoque determinístico, en donde las condiciones iniciales del problema generarán los resultados sin afectación de la aleatoriedad, azar o incertidumbre del modelo de optimización, mediante métodos de

búsqueda basado en la evaluación de gradientes⁹, matrices hessianas¹⁰ o evaluación de la función objetivo¹¹, y 2.) El enfoque estocástico, el cual contempla aleatoriedad en el modelo aplicado, basándose en técnicas bio-inspiradas. Para el desarrollo de este trabajo de investigación, se enfocará el estudio en técnicas de solución de problemas estocásticos, integrados en el área de conocimiento conocido como computación evolutiva.

La computación evolutiva es la disciplina de la inteligencia artificial que implementa modelos computacionales para la resolución de problemas de optimización inspirados en la evolución natural [76], [77]. Entre los problemas abordados por la computación evolutiva se encuentra la búsqueda de información, modelamiento, optimización combinatoria, entre otros; los cuales son trabajados a partir de una (o varias) poblaciones conformadas por unas posibles soluciones candidatas, interactuantes entre sí [76].

La computación evolutiva se inspira en la teoría biológica existente, tales como la ley de la supervivencia del más apto de Darwin, la teoría de herencia de Mendel y el evolucionismo de Welsmann [77]. A partir de estas leyes principales, se establece la existencia de tres paradigmas asociados a la computación evolutiva: La programación evolutiva, las estrategias evolutivas y los algoritmos genéticos [76], [77]. Dado el interés específico del presente trabajo en la solución del problema de diseño CAMD mediante la aplicación del algoritmo genético como herramienta de solución, la Sección 1.3.3 profundiza en ello.

A pesar de que la computación evolutiva abarca una gran parte de las estrategias de solución de problemas de optimización estocástica, existe otro tipo de herramientas de optimización metaheurística, de carácter bio-inspirado, tales como el templado simulado (TS), el enjambre de partículas (PSO), la colonia de hormigas (ACO), entre otros algoritmos [78]. Uno de los grandes retos de la computación evolutiva a futuro es la integración de este conjunto de algoritmos meméticos no poblacionales o de búsqueda local, con el fin

⁹ Ejemplo de algoritmos de optimización determinísticos basados en gradientes son los métodos Cuasi-Newton, gradiente conjugado, punto interior, descenso del gradiente y ascenso a la colina.

¹⁰ El método de Newton-Raphson y sus aproximaciones respectivas que usa diferencias finitas es el claro ejemplo de algoritmos de optimización determinísticos basados en matrices hessianas.

¹¹ El método de Nelder-Mead es un algoritmo de optimización basado en el concepto de simplex, que busca un modo aproximado de la solución sin la evaluación de gradientes o matrices hessianas.

de mejorar el comportamiento de los algoritmos paradigmáticos, mediante la hibridación o paralelización de estos [79].

La evolución de las soluciones a valores óptimos mediante la aplicación de las técnicas evolutivas antes mencionadas depende en gran medida de la implementación de aspectos tales como la representación apropiada de los individuos (soluciones) del problema a resolver, el establecimiento de operadores que afecten y/o modifiquen a los individuos (cruce, mutación, entre otros), la descripción de una función de aptitud y un mecanismo de selección [80]. Todas estas consideraciones son profundizadas en la Sección 1.3.3 para el caso específico de los algoritmos genéticos.

1.3.3 Algoritmo genético (GA) como técnica de optimización estocástica

Los algoritmos genéticos (GAs) son algoritmos de búsqueda basados en los mecanismos de la genética y la selección natural [81]. Fueron desarrollados en 1975 por Holland [82], [83] a partir del concepto de “*la supervivencia del más apto*”. Dicho autor contaba con un doble propósito en la construcción de esta técnica de optimización estocástica: mejorar el entendimiento de los procesos naturales de adaptación y diseñar sistemas artificiales que tengan propiedades similares en los sistemas naturales. Desde entonces, los algoritmos genéticos han sido estudiados a profundidad por gran cantidad de investigadores [81], [82], [83], [84], [85], [86], [87], [88], [89], [90].

La idea básica de los GAs es considerar que las características genéticas de una población dada potencialmente contienen la solución (la mejor posible o una mejor a la actual) a un problema adaptativo dado. Esta solución no se encuentra activa porque la combinación genética se divide en distintos factores (genes), por lo que la asociación de diferentes genomas puede conducir a la solución óptima. El algoritmo genético de Holland es especialmente efectivo, ya que considera la mutación (la cual mejora pocas veces las características del individuo) y la recombinación genética (también llamada cruce o *crossover*) como operaciones claves para mejorar la capacidad del algoritmo para acercarse y eventualmente encontrar la solución óptima [91]. Dada la inspiración de los algoritmos genéticos en conceptos de genética mendeliana y teoría evolutiva darwiniana

[92], buena parte de la terminología existente para ciertos conceptos base son prestados de estas teorías mediante analogías equivalentes. La **Figura 1-4** presenta una equivalencia entre conceptos relacionados a genética/biología y su adaptación al algoritmo genético.

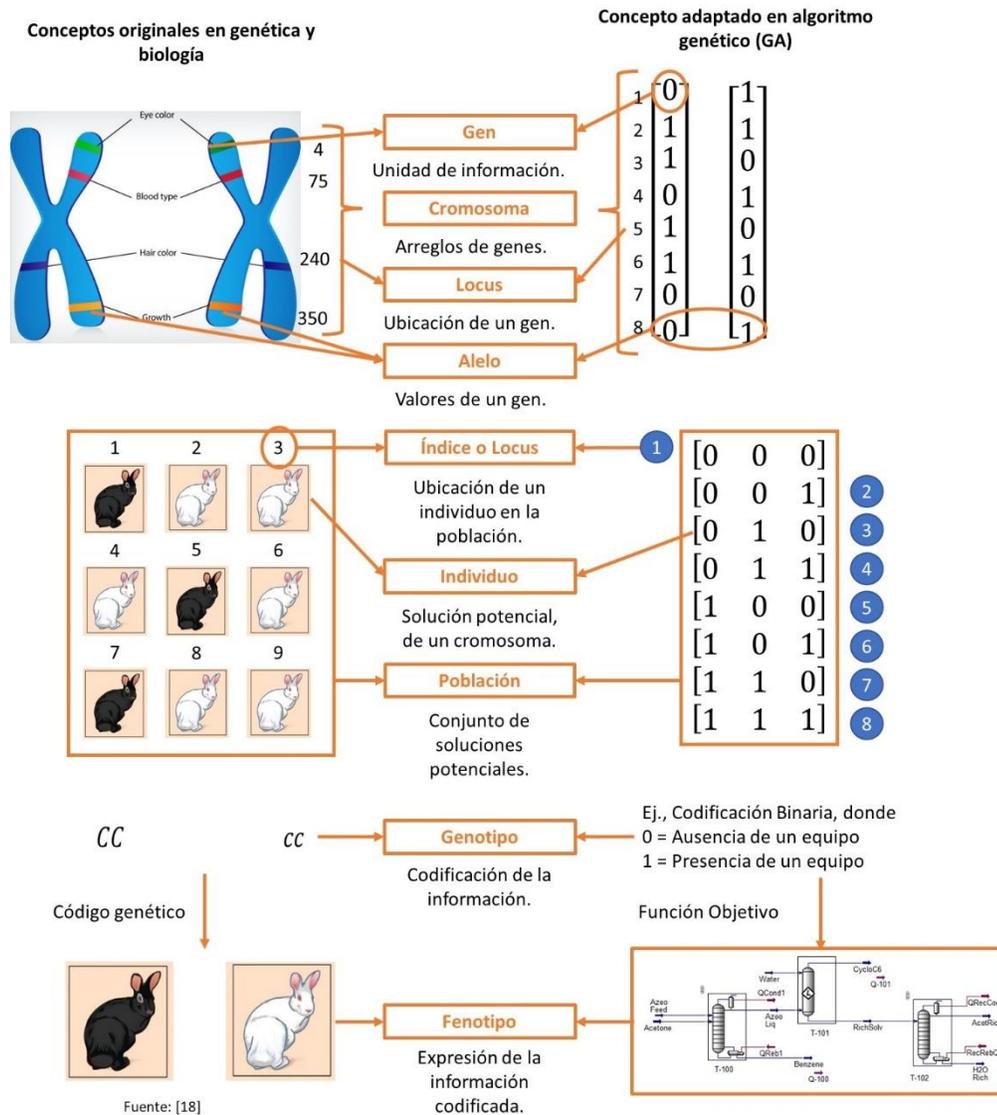


Figura 1-4: Analogías de conceptos de genética y biología empleados en algoritmos genéticos. Adaptado de [84].

Entre las ventajas del algoritmo genético [91] se encuentran las características descritas a continuación.

- La capacidad del algoritmo de ser paralelizado, es decir, de operar distintas etapas del código secuencial en multihilo, con el fin de emplear procesamiento múltiple [93].
- Facilidad de evaluación del espacio de soluciones de manera amplia, mediante operadores de exploración (como el cruce) y de explotación (como la mutación).
- Facilidad en identificar el óptimo global. Según las características del algoritmo, pueden llegar a ser resistentes a quedar atrapados en óptimos locales.
- La posibilidad de ser aplicado a problemas cuyo panorama de aptitud (*fitness*) sea complejo, tales como problemas con funciones a trozos, funciones discontinuas o multimodales.
- El GA solo emplea evaluaciones de funciones. No contempla la evaluación de gradientes o matrices hessianas, lo cual permite resolver problemas cuyas funciones no son derivables, suavizadas o que son discontinuas.
- Fácilmente modificable para diferentes tipos de problemas. La versatilidad del GA permite resolver problemas de gran complejidad, donde se manejan espacios de búsqueda muy grandes y difíciles de comprender, típicos en problemas ingenieriles a gran escala, con gran variedad de variables de optimización.
- El GA maneja relativamente bien funciones con ruido. No es necesario realizar un filtrado previo a la función objetivo, aunque es recomendable.
- La posibilidad de integrar variantes con múltiples poblaciones de individuos y operadores genéticos adicionales, según las necesidades del problema de optimización [94].
- Los algoritmos genéticos operan mediante reglas de transición probabilísticas, asociadas a los operadores genéticos y el método de selección. No existen reglas determinísticas en su construcción, ya que la generación aleatoria de los resultados favorece los procesos de exploración y explotación del espacio de búsqueda.

Por su parte, los algoritmos genéticos cuentan con algunas limitaciones [91], tales como las mencionadas a continuación.

- La necesidad de identificar apropiadamente la función de aptitud. En algunos casos típicos en ingeniería, el establecimiento matemático de la función de

aptitud puede ser altamente compleja, siendo necesario acoplar el GA a un modelo de caja negra (*black-box*), tales como el representado en modelos basados en simulaciones.

- El establecimiento claro de la representación de los individuos en el problema. La representación adecuada para los individuos de la población condicionara la manera de efectuar las operaciones genéticas (cruce y mutación), la forma de expresar la aptitud del individuo y el empaquetamiento de la información genética. Existen gran variedad de representaciones posibles para individuos, siendo los más comunes las representaciones enteras, binarias y reales [95] [96]¹².
- Requiere elegir una gran cantidad de parámetros, tales como el tamaño de la población, las tasas de mutación y cruce, el método de selección y su fuerza. No obstante, se han implementado variantes del algoritmo genético clásico que buscan reducir la cantidad de parámetros a definir, llegando a ser con parámetros adaptativos y auto-sincronizables.
- Dificultad para encontrar el óptimo global exacto en una sola ejecución del algoritmo. Usualmente, los algoritmos genéticos deben contar con múltiples ejecuciones del problema para la identificación de puntos óptimos.
- Mayor complejidad en la configuración del algoritmo, respecto a los métodos determinísticos.

El algoritmo genético cuenta con una serie de etapas establecidas [69], [91], descritas por la **Figura 1-5**. A continuación, se realiza una breve descripción de las características de cada etapa del algoritmo evolutivo clásico desarrollado por Holland [82] [83]. No obstante, conforme ha ido profundizándose en este método estocástico, diversos autores han propuesto modificaciones o variantes en una o más etapas del proceso evolutivo, con el fin de mejorar el rendimiento, la robustez, evitar la convergencia prematura y reducir los tiempos de procesamiento. Cada variante cuenta con ajustes específicos, siendo los más

¹² De acuerdo con el problema de optimización a resolver mediante el algoritmo genético, representaciones tales como grafos, arboles de decisión, collares e incluso redes pueden ser más apropiadas. El diseño de la representación más apropiada debe contemplar elementos tales como la redundancia, el escalado de los alelos y la localidad, que describe la modificación de distancias correspondientes entre genotipos y fenotipos.

reconocidos los algoritmos VEGA [97], NSGA-I [98], NSGA-II [99], MOGA [100], SPEA-I [101], SPEA-II [102] y HAEA [103], [104].

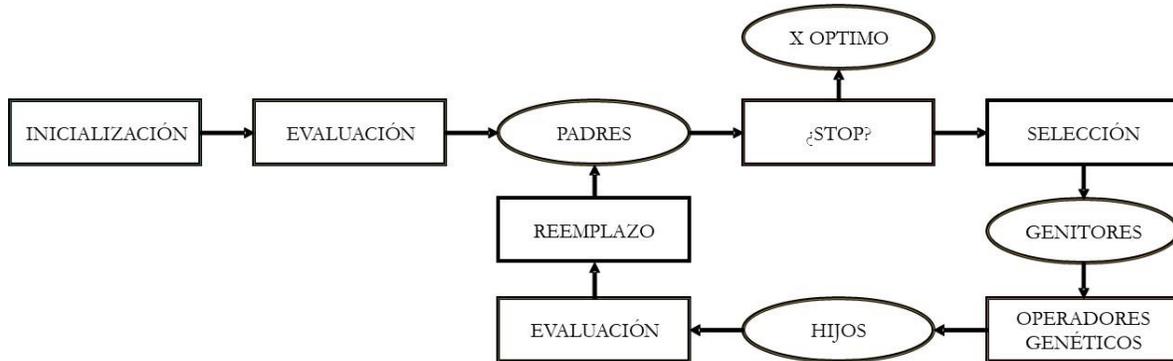


Figura 1-5: Etapas del algoritmo genético [69] [91].

1. *Inicialización, creación de población inicial y establecimiento de parámetros del algoritmo genético:* Se generan aleatoriamente una población inicial constituida por posibles soluciones del problema (individuos) con una representación establecida. Adicionalmente, se fijan los parámetros de ejecución del algoritmo genético, tales como el tamaño de la población, el número de iteraciones, la tasa de cruce, la tasa de mutación, la función de aptitud, la técnica de selección y los operadores genéticos a emplear.
2. *Evaluación de la aptitud de la población inicial:* Partiendo de los individuos generados en la etapa anterior, se evalúan los genotipos de los individuos en una función de aptitud, la cual indicará la capacidad del individuo a adaptarse. En problemas de minimización, un valor menor de aptitud representará a un individuo con mejor fenotipo.
3. *Selección de parejas de cruzamiento:* Una vez establecida la aptitud de los individuos, es necesario establecer un procedimiento de selección de individuos padres para aplicar los operadores genéticos. Aunque el objetivo de la etapa de selección es garantizar que los mejores individuos presentes en una población tengan mayor probabilidad de reproducirse, la selección debe ser suficientemente sensible para “*dar la oportunidad*” de reproducirse a individuos

con baja aptitud¹³. Por tal razón, la selección condiciona las operaciones de cruce y mutación en las generaciones.

Existen dos clasificaciones de las técnicas de selección de individuos, según si se establecen constantes las probabilidades de selección entre generaciones (selecciones estáticas) o no (selecciones dinámicas) [85]. Entre los métodos de selección más comúnmente aplicados se encuentran la selección elitista, la selección proporcional por ruleta, selección mediante torneo y la selección de estado uniforme.

4. *Aplicación de operadores genéticos:* En el algoritmo genético clásico, existen dos operadores genéticos: cruce y mutación. La operación de cruce consiste en tomar dos genotipos y producir nuevos genotipos mediante el mezclado de los genes encontrados en los originales. Por su parte, la mutación corresponde en modificar los genes.

El efecto del cruce es muy importante debido a que las características de los distintos padres son transmitidas a sus hijos, generando un efecto de exploración del espacio de búsqueda de soluciones; mientras que la mutación suple el efecto de explotación del espacio de búsqueda, al generar individuos significativamente similares a los padres [91]. Tanto el cruce como la mutación cuentan con mecanismos para obtener genomas nuevos mediante la creación de nuevos individuos (hijos) a partir de un conjunto de individuos presentes en una población inicial (padres).

5. *Evaluación de la aptitud de individuos hijos y reemplazo poblacional para conservación de cantidad de individuos mediante método de selección:* Una vez generada una cantidad de hijos igual al número de padres, es necesario reducir

¹³ El grado de influencia que ejerce la etapa de selección en la reproducción de los mejores individuos de una generación se conoce como **Presión de Selección**. De acuerdo con la técnica de selección de individuos escogida, la presión de selección puede ser ajustada según se requiera.

la población al tamaño original. Para ello, es posible aplicar cualquiera de las técnicas descritas en la etapa 2.

Cabe resaltar que en el caso de AG generacionales, es necesario contar con un aprendizaje incremental, donde se recuerde lo aprendido por la población en generaciones anteriores. Para ello, se parte del concepto de población traslapable¹⁴, la cual indica la brecha generacional existente entre padres e hijos en una generación dada. Una brecha generacional alta suele mejorar el desempeño del algoritmo genético [85].

6. *Cambio generacional y criterio de parada*: Una vez actualizada la población luego de la selección, se efectúa un incremento en el contador de iteraciones, iniciando una generación nueva y repitiendo nuevamente las etapas anteriores. El criterio de parada del algoritmo genético suele ser alcanzar el número de generaciones fijado inicialmente. No obstante, es posible culminar el algoritmo genético anticipadamente cuando no se observa variación en los genotipos obtenidos para los individuos de la población, no hay cambios en la aptitud del mejor individuo durante un número especificado de generaciones, o cuando el tiempo de cómputo ha superado un límite definido [85], [91].

Siguiendo los conceptos provenientes de la teoría de Darwin, se espera la mejora de la aptitud para adaptarse de los individuos de la población conforme la cantidad de generaciones (iteraciones) aumenta. Esto fue verificado matemáticamente por Holland mediante la formulación del Teorema del Esquema¹⁵ [81], [105], [106], el cual provee un modelo formal para justificar la efectividad del proceso de búsqueda del algoritmo genético.

¹⁴ Entiéndase como **población traslapable** (*overlapping*) a aquellas poblaciones de individuos padres e hijos, en donde existe competencia en la selección para su supervivencia. Usualmente, los individuos padres cuentan con un periodo de vida de una sola generación, por lo que poner en competencia en términos de aptitud a los padres preexistentes y los hijos generados, puede mejorar el aprendizaje poblacional.

¹⁵ El **teorema del esquema** consiste en una fundamentación matemática para el análisis del impacto del proceso de selección y los operadores de cruce y mutación en un algoritmo genético

1.3.4 Uso de algoritmos genéticos en problemas de ingeniería química

El desarrollo y aplicación del algoritmo genético para la solución de problemas en ingeniería química es de uso común logrando obtener soluciones factibles aún con enunciados de gran complejidad. A continuación, se describirán algunos ejemplos aplicados de algoritmos genéticos en problemas de interés para la ingeniería química, no relacionados de manera directa con el diseño de solventes mediante metodología CAMD. Por su parte, la **Tabla 1-3** y la **Tabla 1-4** descritas en la Sección 1.2.3, junto con la Sección 1.2.4, presentan casos de estudio específicos de problemas de ingeniería química resueltos con algoritmos genéticos y otras herramientas de optimización.

El campo de la ingeniería química ofrece problemas de optimización y de búsqueda desafiantes, tales como los relacionados con el control de procesos químicos para asegurar la operabilidad en los distintos fenómenos de transporte. El establecimiento de una estrategia de control en plantas complejas de carácter no-lineal, con múltiples entradas y salidas (MIMO, *multi-input multi-output* por sus siglas en inglés), cuenta con una gran variedad de problemas interesantes para la aplicación de algoritmos genéticos con un enfoque industrial, partiendo del establecimiento de objetivos de producción. La sintonización de parámetros de controladores, sensores y elementos finales de control empleando algoritmos genéticos ha demostrado mayor rapidez de respuesta respecto a otros métodos (basados en errores cuadráticos, por ejemplo). Adicionalmente, la aplicación de algoritmos genéticos permite al modelo de sintonización adaptarse a los cambios en tiempo real [107].

Uno de los ejemplos clásicos que relacionan algoritmos genéticos e ingeniería química es el establecimiento de la programación de operaciones dentro de una planta química (*scheduling*). Los procesos de alto valor agregado se producen de forma rutinaria en industrias de procesamiento de materias primas, en los que se conectan varios reactores

establecido. Partiendo de una representación binaria de individuos, un esquema se construye utilizando un nuevo símbolo (*) para representar un gene cuyo valor es comodín. Por ejemplo.: el esquema 11*00* representa a los individuos 111001, 111000, 110001, 110000.

y se alimentan distintos lotes de productos químicos, recuperándose los productos deseados al final de la línea de producción, gracias a la transformación de materias primas mediante cambios físicos y/o reacciones químicas y la concentración de productos a un grado de pureza comercial. Considerando los procesos discontinuos con una serie de productos a sintetizar, existe un orden óptimo que minimiza el tiempo de permanencia de los reactivos dentro del diagrama de flujo de proceso. Este problema es análogo al conocido problema del vendedor viajero (TSP, *Travelling Salesman Problem* por sus siglas en inglés), el cual es un problema de complejidad NP-completo, donde el tamaño del problema aumenta exponencialmente con la cantidad de parámetros considerados, pero no es infinito. La utilización de algoritmos genéticos asociados a este tipo de problemas ha demostrado ser entre 100 y 1800 veces más rápido que la determinación de soluciones mediante búsqueda exhaustiva [108].

Otras aplicaciones que asocian la aplicación de algoritmos genéticos para la solución de problemas ingenieriles son el diseño de redes de reacción [109], la predicción de estructuras proteínicas ternarias [110], el diseño de sistemas térmicos eficientes mediante redes de intercambio de calor [111], la selección y optimización de mezclas fluidas para la utilización de calor de bajo grado en ciclos orgánicos Rankine [112], la optimización de catalizadores para un sistema reactivo específico [113], el desarrollo de secuencias cuaternarias de destilación para el establecimiento de mapas de consumo energético [113], modelos matemáticos para la remoción de contaminantes [114], el modelamiento no-lineal del comportamiento cromatográfico [110] y el modelamiento molecular de proteínas para el diseño de principios activos en fármacos y drogas [115]. Por la gran diversidad de aplicaciones de esta herramienta de optimización, es posible considerar el enorme potencial en la solución de problemas de larga escala relacionados con la ingeniería química, a partir de un modelo matemático claramente establecido.

1.4 Optimización multiobjetivo (MOO)

Muchos de los problemas de optimización del mundo real requieren de la solución simultánea de un número establecido de funciones objetivo, las cuales pueden estar en conflicto (por ejemplo, maximizar los beneficios económicos generados en un proceso químico y, simultáneamente, minimizar el impacto ambiental generado por el proceso). Por

este motivo, la aplicación de los métodos de optimización antes mencionados, siguiendo un enfoque en un solo objetivo, es insuficiente para la determinación de las mejores soluciones posibles [116] [117].

El problema de optimización multiobjetivo (también llamado optimización multicriterio [70], [118], problema vectorial de optimización [119], [120] o problema de múltiple rendimiento) puede definirse de manera informal [121] como:

“Un vector de variables de decisión que satisface restricciones y optimiza a una función vectorial donde los elementos representados corresponden a las funciones objetivo. Estas funciones forman una descripción matemática del criterio de rendimiento donde usualmente se encuentran en conflicto una entre otra. Luego, el término “optimizar” significa encontrar al menos una solución que pueda dar los valores de todas las funciones objetivo de forma aceptable para el tomador de la decisión.”

La Sección 1.4.1 profundizará en una definición matemática canónica para el problema de optimización multiobjetivo, incorporando las definiciones óptimo de Pareto, frente de Pareto y dominancia de soluciones. Para la evaluación de los resultados obtenidos en la solución del problema multiobjetivo, se establecen unas métricas (convergencia y diversidad), descritas a profundidad en la Sección 1.4.2. Adicionalmente, la Sección 1.4.3 describe los principales métodos de solución estocástica multiobjetivo, basados en algoritmos genéticos, los cuales permiten resolver problemas de distinto tipo, aplicados a ingeniería química. Una breve descripción de los problemas formulados y resueltos reportados en la literatura con un enfoque multiobjetivo es descrita en la Sección 1.4.4.

1.4.1 Formulación canónica y conceptos claves en optimización multiobjetivo

Un problema de optimización multiobjetivo puede ser formulado canónicamente de acuerdo con la ecuación (1-14), en donde $x = (x_1, x_2, \dots, x_n)$ es un n-vector de variables de decisión, X es un espacio de decisión n-dimensional y m es el número de objetivos del problema a ser optimizado [122].

$$\min F(x) = (f_1(x), f_2(x), \dots, x_m(x))^T \quad x \in X \subset \mathbb{R}^n \quad (1-14)$$

En contraste con la optimización en un solo objetivo, una solución de un problema multiobjetivo es más un concepto que una definición [123], [124]. Típicamente, no hay una única solución global, siendo necesario determinar un conjunto de puntos que se ajusten a una definición determinada de lo que es un óptimo. El concepto predominante para definir un punto óptimo fue definido por Vilfredo Pareto en el año 1.906, siendo conocido como óptimo de Pareto, que es descrito en la **Tabla 1-6**.

Tabla 1-6: Definición de óptimo de Pareto [123], [124], [125].

Un punto, $x^* \in X$, es un óptimo de Pareto (también llamado Pareto-óptimo) si y solo si no hay ningún otro punto existente $x \in X$, tal que $F(x) \leq F(x^*)$ y $F_i(x) < F_i(x^*)$ para al menos una función objetivo.

Todos los puntos Pareto-óptimos se deben encontrar en los límites del espacio factible Z . Dependiendo del algoritmo, es posible establecer soluciones que no puedan ser Pareto-óptimas, pero pueden satisfacer otros criterios, haciéndolos significativos para las aplicaciones prácticas. Para estas consideraciones, es útil la definición de óptimo débil de Pareto, descrita por la **Tabla 1-7**.

Tabla 1-7: Definición de óptimo débil de Pareto [123], [124], [125].

Un punto, $x^* \in X$, es un óptimo débil de Pareto si y solo si no hay ningún otro punto existente $x \in X$, tal que $F(x) < F(x^*)$.

La diferencia principal entre un óptimo débil de Pareto y un óptimo de Pareto radica en la forma de mejora de las funciones objetivos respecto a las soluciones. Un punto es un óptimo débil de Pareto, si no existe otro punto que mejore todas las funciones objetivo simultáneamente. En contraste, un punto es Pareto-óptimo si no hay otro punto que mejore en al menos una función objetivo sin empeorar respecto a otra. A partir de las definiciones,

es posible establecer que la relación entre estas dos definiciones de optimalidad no es bidireccional¹⁶.

Para cualquier problema dado, hay una cantidad infinita de puntos Pareto-óptimos, los cuales constituyen el conjunto denominado Frente óptimo de Pareto (también llamado Frente de Pareto). De acuerdo con la representación gráfica utilizada (en \mathbb{R}^2 o \mathbb{R}^3), el frente óptimo de Pareto puede verse reflejado como una curva o superficie de puntos óptimos. La **Figura 1-6** presenta un ejemplo del Frente óptimo de Pareto obtenido en la minimización de dos funciones objetivo.

En el contexto de las optimizaciones multiobjetivo, las soluciones óptimas también son conocidas como soluciones no dominadas. Esta denominación, está fundamentada en la noción de dominancia, descrita por la **Tabla 1-8**. Como puede observarse, un punto Pareto-óptimo respecto al espacio de soluciones es casi considerado lo mismo que un punto no-dominado.

Tabla 1-8: Definición de dominancia de Pareto [123].

Un vector de funciones objetivo, $F(x^*) \in Z$, es no dominado si y solo si no existe otro vector $F(x) \in Z$, tal que $F(x) \leq F(x^*)$ con al menos un $F_i(x) < F_i(x^*)$. En otro caso, $F(x^*)$ es dominado.

Teniendo en cuenta la descripción de dominancia de Pareto descrita anteriormente, dados dos soluciones $x \in X$ y $x^* \in X$, existen tres posibles casos para sus vectores objetivos correspondientes [126]. La **Figura 1-6** describe el comportamiento de distintas soluciones respecto al punto solución C , para la minimización de dos funciones objetivo F_1 y F_2 conflictivas entre sí.

- Caso 1: $F_i(x) \leq F_i(x^*) \forall i$. En este escenario, el vector $F(x)$ domina a $F(x^*)$ y x domina a x^* .

¹⁶ Todo punto Pareto-óptimo es un óptimo débil de Pareto, pero un punto débil de Pareto no es Pareto-óptimo.

- Caso 2: $F_i(x^*) \leq F_i(x) \forall i$. En este escenario, el vector $F(x^*)$ domina a $F(x)$ y x^* domina a x .
- Caso 3: $F_i(x^*) \not\leq F_i(x)$ y $F_i(x) \not\leq F_i(x^*)$. En este escenario, se dice que el vector objetivo $F(x)$ es indiferente de $F(x^*)$ y viceversa, lo cual es equivalente a decir que x es indiferente de x^* y viceversa.

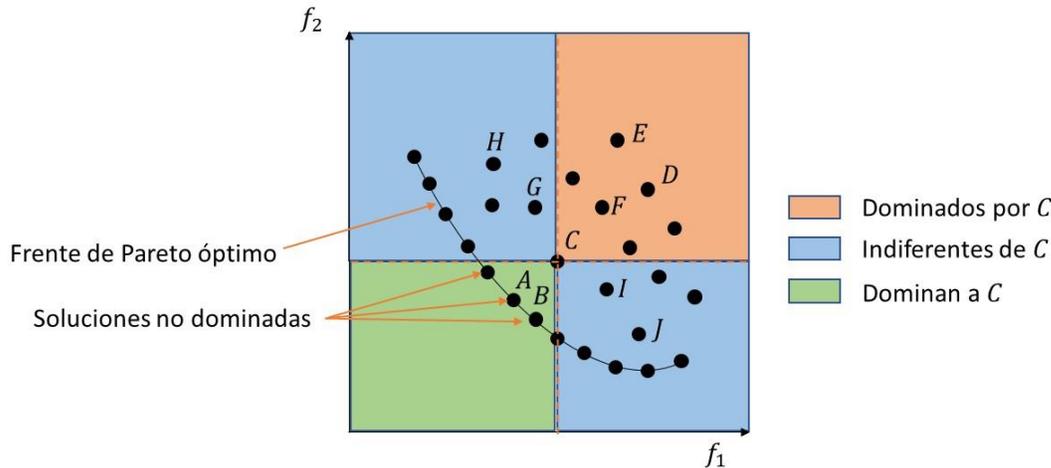


Figura 1-6: Frente de Pareto y dominancia de Pareto en un problema de minimización con dos objetivos. Adaptado de Kumar [70] y Subía [126].

En optimización multiobjetivo, los óptimos de Pareto suelen encontrarse en el contorno o en los puntos del dominio donde son tangentes dos o más funciones objetivo, tal como se observa en la **Figura 1-7**. Es por esta razón, que resulta importante representar el problema de optimización en el espacio de las funciones objetivo (o espacio de solución). Adicionalmente, dadas las características del problema multiobjetivo, los contornos de las funciones pueden ser o no ser convexos¹⁷, hecho que puede incrementar la dificultad para la identificación de los mínimos de Pareto [127]. La **Figura 1-8** presenta los escenarios de convexidad¹⁸ para un problema de optimización de dos funciones objetivo.

¹⁷ Entiéndase como **función convexa** a aquella función real, definida sobre un intervalo (a, b) en donde existe una cuerda que une dos puntos cualesquiera de la función, quedando por encima de la función.

¹⁸ La **convexidad** es una característica particular de un polígono geométrico. Una parte C de un espacio vectorial real es convexo si para cada par de puntos C, el segmento está totalmente incluido en C, siendo posible ir de cualquier punto a otro de C en línea recta, sin salirse del mismo.

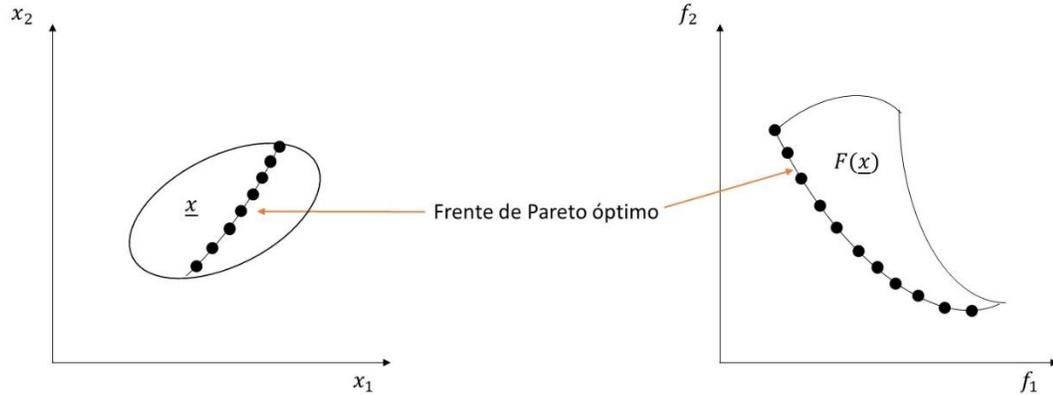


Figura 1-7: Ubicación de soluciones pertenecientes al frente de Pareto respecto al espacio de las variables de diseño y el espacio de las funciones objetivo [127].

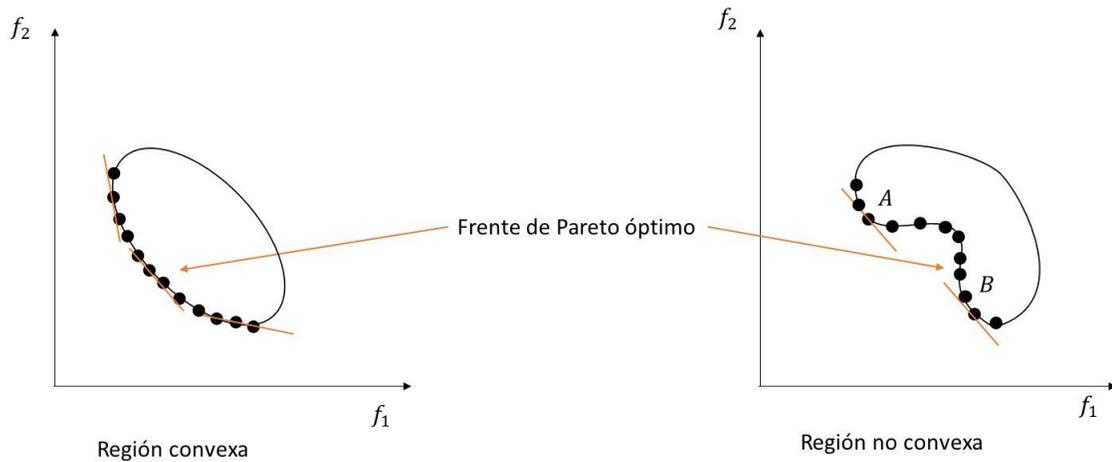


Figura 1-8: Regiones de diseño convexas y no convexas en el espacio de las funciones objetivo [127].

De manera similar a las optimizaciones en un solo objetivo, es posible establecer condiciones necesarias y suficientes para la identificación del óptimo de un problema de optimización multiobjetivo [128]. Esto se realiza mediante extrapolación aproximada de las condiciones de Karuch-Kuhn-Tucker para este tipo de problemas de optimización, tal como proponen Giorgi *et al* [129].

Por su parte, el manejo de restricciones en optimización multiobjetivo se puede realizar mediante distintos mecanismos, los cuales provienen principalmente de algoritmos de

optimización evolutiva, tales como MOGA, NSGA-II y VEGA [130]. Estos mecanismos son: 1.) Usar la dominancia de Pareto como criterio de selección de los individuos, 2.) Emplear la jerarquización de Pareto para asignar una aptitud a los individuos y 3.) Dividir la población principal en subpoblaciones que serán evaluadas cada una con respecto a una restricción del problema o a la función objetivo original. La utilización de un mecanismo para el manejo de restricciones estará sujeto a las características específicas del algoritmo evolutivo utilizado para la solución del problema multiobjetivo [130].

1.4.2 Métricas de desempeño para optimizaciones multiobjetivo

En optimizaciones de un solo objetivo, el enfoque principal es la convergencia del algoritmo. Para ello, es posible establecer métricas de desempeño, de acuerdo con el grado de avance de las aptitudes de los individuos de la población a lo largo de las generaciones, dada la existencia de una relación próxima entre la aptitud de un individuo y la función objetivo¹⁹. Por su parte, en las optimizaciones multiobjetivo, deben establecerse dichas métricas de desempeño en función de criterios n –dimensionales, establecidas según las n funciones objetivos definidas, lo que se ve afectado por el efecto Hughes (también conocido como maldición de la dimensionalidad²⁰). Por tal razón, la incorporación de la dominancia y del frente óptimo de Pareto son necesarias para contar con una solución ecuánime al problema planteado [122]. Esto, sumado al reto de asegurar tanto convergencia como diversidad en las soluciones, hacen a las optimizaciones multiobjetivo un desafío para la determinación de las soluciones que satisfagan todos los criterios del problema [131]. Las métricas de desempeño (convergencia y diversidad) se precisan conforme a lo descrito por la **Tabla 1-9**.

¹⁹ En problemas de un solo objetivo no restringidos, la aptitud suele ser la misma función objetivo. No obstante, este comportamiento puede diferir en problemas con restricciones.

²⁰ La **maldición de la dimensionalidad** contempla la problemática existente en analizar y organizar datos de espacios multidimensionales superiores a las dimensiones del espacio físico, descrito solo por tres dimensiones.

Tabla 1-9: Criterios de desempeño en optimizaciones multiobjetivo [122], [131].

Criterio	Descripción
Convergencia	Consiste en que tan distante se encuentra el conjunto de soluciones óptimas encontradas por el algoritmo respecto al frente óptimo de Pareto teórico (o conocido).
Diversidad	Se refiere a la distribución y dispersión de las soluciones encontradas. La distribución es la distancia relativa de las soluciones, mientras que la dispersión asocia al rango de valores cubierto por las soluciones.

Con el fin de establecer una forma de cuantificación del desempeño de los algoritmos de optimización multiobjetivo, diversos autores [132], [133], [134] han establecido y recopilado una serie de indicadores y métricas que buscan evaluar cuantitativamente la convergencia del algoritmo hacia el frente de Pareto y la diversidad de las soluciones a lo largo del mismo. Castro [135] presenta una recopilación de las principales métricas de desempeño identificadas en algoritmos de optimización multiobjetivo, las cuales son descritas en la **Tabla 1-10**. La utilización de un indicador/métrica u otro en particular, depende de las características del algoritmo de optimización empleado para la solución del problema y puede estar sujeta al criterio de quien desarrolla la optimización.

Tabla 1-10: Métricas e indicadores para evaluación del desempeño de algoritmos de optimización multiobjetivo [135].

Indicador	Nombre de la métrica o indicador
I_{R2}	Indicador R.
$I_{\bar{H}}$	Indicador de Hipervolumen.
I_{HC}	Indicador de recubrimiento del hipercubo.
I_O	Indicador del vector de objetivos.
I_l	Indicador unario ϵ .
I_P	Número de puntos de Pareto.
I_W	Promedio de la mejor combinación de pesos.
I_D	Distancia al frente de Pareto.
I_{PF}	Fracción del frente de Pareto cubierta.
I_{ER}	Promedio del error.

Cont. Tabla 1-10.

Indicador	Descripción
I_{ONVG}	Vector no dominado generado.
I_{GD}	Distancia generacional.
I_{ME}	Máximo error al Frente de Pareto.
I_{CD}	Indicador de desviación Chi-cuadrado.
I_S	Espaciado.
I_{MS}	Máximo espaciamiento.
I_{MD}	Mínima distancia entre dos soluciones.
I_{CE}	Error de cobertura.
I_{DU}	Desviación de una distribución uniforme.
I_{OS}	Espaciamiento de Pareto.
I_A	Precisión.
I_{NDC}	Número de soluciones distintas.
I_{CL}	Clúster.

1.4.3 Enfoques existentes en el manejo de múltiples funciones en optimización multiobjetivo

De manera similar a los problemas de optimización en un solo objetivo, es posible emplear algoritmos estocásticos en la solución de problemas multiobjetivo, siempre que se establezca una forma para el manejo de las múltiples funciones objetivo. Existen dos enfoques principales para el manejo de los conflictos entre las múltiples funciones objetivos presentes en el problema: un enfoque clásico y un enfoque evolutivo [136]. En el enfoque clásico, existen métodos generadores para la toma de decisiones según una preferencia *a priori* de un objetivo respecto a otro. La **Tabla 1-11** presenta algunos métodos generadores empleados para la solución de problemas de optimización multiobjetivo, siguiendo el enfoque clásico.

Tabla 1-11: Métodos de solución de problemas multiobjetivo con articulación a priori de preferencias [69], [123], [136], [137], [138], [139]

Criterio de solución	Descripción del método generador / toma de decisión
<i>Pay-off matrix</i>	Obtener el valor óptimo de un objetivo sin considerar otros, y valores del resto de objetivos en esa solución.
<i>Trade-off Method</i>	Determinar lo que está dispuesto a empeorar un objetivo por mejorar en una unidad otro objetivo.
<i>Sumas ponderadas</i>	Multiplicar cada objetivo por peso o factor no negativo y reducir el problema a optimización en un solo objetivo.
<i>Método ϵ-constraint</i>	Optimizar uno de los objetivos e incorporar el resto como restricciones paramétricas.
<i>Simplex Multiobjetivo</i>	Obtener un conjunto eficiente a partir de combinaciones lineales de puntos extremos adyacentes.
<i>Goal attainment Method</i>	Resolver un problema de objetivos múltiples como un problema lineal que equilibre pros y contras de los objetivos en conflicto.
<i>Método lexicográfico</i>	Se realiza un ordenamiento de las funciones objetivo según un orden de importancia establecido, para después resolver la optimización multiobjetivo un objetivo a la vez. Las funciones objetivo no evaluadas son integradas como restricciones.

A pesar de facilitar el manejo de las funciones objetivo, el uso de un enfoque clásico para el manejo de las funciones objetivo en problemas multiobjetivo se suele visualizar como una mala praxis, para la capacidad computacional de cálculo actual [136], [137], [138]. Por tal razón, el enfoque evolutivo busca representar de manera holística el espacio de búsqueda para este tipo de problemas, lo que hace necesario adaptar los algoritmos evolutivos empleados en problemas de un solo objetivo, siguiendo los conceptos mencionados en las Secciones 1.4.1 y 1.4.2 [136], [140], [141], [142], [143]. Para la validación de los algoritmos evolutivos multiobjetivos empleados, se emplean una serie de funciones de prueba (*benchmarking*), las cuales son conjuntos de funciones objetivos cuya tendencia del frente de Pareto es ampliamente conocido, e incluso en algunos casos obtenible analíticamente. Coello [144] recopila una larga lista de funciones de prueba para problemas multiobjetivo, con comportamientos variados en \mathbb{R}^2 y \mathbb{R}^3 .

Respecto a algoritmos genéticos, se han construido algoritmos específicos para optimizaciones multiobjetivo, partiendo de las bases dadas por algoritmos para optimizaciones en un solo objetivo, tales como VEGA [145], SPEA2 [102], MOEA [51], NSGA-II [146], [147], [147], [148], [149], [150] y HAEA/MOHAEA [151], [135], [152], siendo algunos aplicados a diversos casos con un conflicto entre dos o más funciones objetivo (por ejemplo, minimizar el impacto ambiental y maximizar la ganancia económica de un proceso industrial). Adicionalmente, se han propuesto modificaciones para asegurar el desempeño de los algoritmos multiobjetivo, como la selección de los individuos considerando vecindades o nichos [153], o generando un orden de preferencia por semejanza a una solución ideal establecida [154].

Por otra parte, la optimización multiobjetivo también ha generado interés en otros algoritmos estocásticos bio-inspirados, tales como enjambre de partículas (PSO) [155], [156], estrategias evolutivas [157], búsqueda esparcida (*scatter search*) [158], algoritmos de optimización difusa [159] y colonia de abejas [160], [161]. No obstante, el uso de algoritmos genéticos en problemas multiobjetivo es significativamente mayor, debido a las ventajas en el manejo de funciones objetivos.

1.4.4 Aplicaciones multiobjetivo para resolución de problemas ingenieriles, ambientales y de diseño CAMD

Dada la gran cantidad variedad de problemas de optimización multiobjetivo disponibles en la literatura [121], [124], [150], [162], [163], [164], se realizó una recopilación de las principales aplicaciones identificadas para este tipo de problemas, resueltos mediante la aplicación de un algoritmo genético que permitiera de manera holística el respectivo problema a resolver. La **Tabla 1-12** recopila de manera sintética algunas aplicaciones asociadas a problemáticas de ingeniería química, diseño CAMD y búsqueda de óptimos ambientales para problemas de diseño de procesos y productos.

Tabla 1-12: Aplicaciones de problemas de optimización multiobjetivo para distintos enfoques en ingeniería [121], [124], [150], [162], [163], [164].

Enfoque	Aplicación	Especificaciones de la optimización	Ref.
Matemática aplicada	Optimización multiobjetivo usando un micro algoritmo genético	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Seis problemas de optimización en dos objetivos. - <u>Algoritmo de optimización:</u> Micro-GA, NSGA-II, PAES. - <u>Métricas empleadas:</u> Tasa de error, espacio distribuido, M1, Métrica O. 	[165]
Ingeniería de Telecomunicaciones	Optimización multiobjetivo en la planificación de centrales telefónicas	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Problema de tres objetivos; cantidad y ubicación optima de centrales en Asunción (Paraguay), Costo de implementación de la central de telecomunicación. - <u>Algoritmo de optimización:</u> SPEA2. - <u>Métricas empleadas:</u> No describe 	[166]
	Aplicación de la Programación Multiobjetivo en la Optimización del Tráfico Generado por un IDS/IPS	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Modelos de tráfico en redes LAN Ethernet, - <u>Algoritmo de optimización:</u> Modelos estocásticos autorregresivos. - <u>Métricas empleadas:</u> No describe. 	[167]
Ingeniería industrial e investigación de operaciones	Metodología Multiobjetivo para optimizar un ambiente Job Shop	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> tiempo total del proceso, costo de energía y accidentalidad laboral. - <u>Algoritmo de optimización:</u> NSGA-II. - <u>Métricas empleadas:</u> No describe. 	[168]
Ingeniería de sistemas y computación	Metodología multiobjetivo para problemas de <i>scheduling</i> en el ahorro de energía en sistemas computacionales distribuidos.	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> agenda para procesamiento con menor tiempo, energía consumida. - <u>Algoritmo de optimización:</u> MOCeII, IBEA, NSGA-II - <u>Métricas empleadas:</u> Épsilon, IGD, distancia entre soluciones. 	[169]
Ingeniería ambiental	Análisis multiobjetivo para control de la exposición a ozono	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Modelo de dos objetivos, pasado en una simulación de precursores de contaminantes y funciones de costos reducidos por emisión, acotado al norte de Italia. - <u>Algoritmo de optimización:</u> Redes neuronales - <u>Métricas empleadas:</u> No describe. 	[170]
	Optimización multiobjetivo de un sistema de energías renovables para cumplimiento de demanda energética	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Costos de producción, pico de la demanda, intermitencia de la fuente renovable - <u>Algoritmo de optimización:</u> Optimización global. - <u>Métricas empleadas:</u> No describe. 	[171]
	Optimización termo-económica y ambiental usando algoritmos genéticos multiobjetivo	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Tres objetivos termo económica, tecnológico y ambiental para el proceso de generación de electricidad en Termocesar (Cesar, Colombia), basados en una función de pesos ponderados. - <u>Algoritmo de optimización:</u> MOGA. - <u>Métricas empleadas:</u> No describe. 	[172]
	Uso del método <i>e-constraint</i> para problemas multiobjetivo aplicados a problemas medioambientales.	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Se aplicó al caso de estudio de la cadena de valor del etanol y del hidrógeno, enfocado en el diseño de la red optima que optimice simultáneamente el valor presente neto, el costo total de la tecnología y el rendimiento medioambiental. - <u>Algoritmo de optimización:</u> Método <i>e-constraint</i>. - <u>Métricas empleadas:</u> Área del espacio completa vs reducido, hipervolumen, equidistancia de soluciones, numero de soluciones factibles. 	[173]

Cont. **Tabla 1-12.**

Enfoque	Aplicación	Especificaciones de la optimización	Ref.
Ingeniería ambiental	Toma de decisiones multicriterio para la gestión de recursos en microrredes asistidas por energías renovables	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> 10 objetivos asociados a la operación, planeamiento y diseño de las microrredes, para la gestión de recursos renovables. - <u>Algoritmo de optimización:</u> NSGA-II. - <u>Métricas empleadas:</u> No describe. 	[174]
	Problema de despacho de mercancías con enfoque medioambiental y económico.	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> costo del combustible, emisión de NOx, con restricción relacionada al balance de potencia eléctrica. - <u>Algoritmo de optimización:</u> NSGA-II y SPEA-II. - <u>Métricas empleadas:</u> Distancia generacional por convergencia, propagación de la diversidad, métricas combinadas de convergencia y diversidad. 	[175]
Diseño de procesos químicos	Red de optimización basada en algoritmo genético multiobjetivo para el diseño de una planta de procesamiento batch	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Costos de inversión, flexibilidad del diseño, número de equipos seleccionados y periodo de ciclo para alcanzar estado estacionario, para dos procedimientos de procesamiento, con 7 productos finales 10 materias primas y 10 operaciones unitarias. - <u>Algoritmo de optimización:</u> MOGA. - <u>Métricas empleadas:</u> No describe. 	[176]
	Optimización de secuencias Petlyuk usando algoritmo genético multiobjetivo con restricciones	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> carga calórica y número de etapas de separación, para un sistema integrado de columnas de destilación con secuencias tipo Petlyuk, simulador en Aspen Plus. - <u>Algoritmo de optimización:</u> MOGA. - <u>Métricas empleadas:</u> No describe. 	[177]
Diseño CAMD	Diseño multiobjetivo para sistemas de síntesis de productos con integración de solventes.	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Rendimiento de solventes (volatilidad relativa, selectividad, coeficiente de distribución, pérdidas en refinado y solvente), toxicidad para los casos de separación n-Butanol-Agua (extracción LL), Aire-Acetona (absorción de gases). - <u>Algoritmo de optimización:</u> Templado simulado. - <u>Métricas empleadas:</u> No describe. 	[178]
	Evolución de moléculas usando optimización multiobjetivo para la determinación de bioactividad de fármacos y ligandos.	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Problema de tres objetivos: bioactividad del compuesto (solubilización), penetración de barrera sangre-cerebro y capacidad de formación de enlaces. - <u>Algoritmo de optimización:</u> MOGA. - <u>Métricas empleadas:</u> No describe. 	[179]
	Diseño CAMD para la optimización simultánea del rendimiento, aspectos de salud y seguridad para el proceso de endulzamiento de gas.	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Maximización del parámetro λ asociado a la suma ponderada de propiedades. - <u>Algoritmo de optimización:</u> Optimización difusa (fuzzy) mediante una aproximación min-max. - <u>Métricas empleadas:</u> No describe. 	[180]
	Aplicación del análisis multicriterio en optimización de solventes para procesos de extracción líquido-líquido a escala micro.	<ul style="list-style-type: none"> - <u>Función Objetivo:</u> Capacidad de extracción, aspectos toxicológicos, eco toxicológicos e impacto ambiental. - <u>Algoritmo de optimización:</u> Técnica para orden de preferencia por similitud con la solución ideal (TOPSIS). - <u>Métricas empleadas:</u> No describe. 	[181]

1.5 Optimización en múltiples objetivos (MaOO)

La optimización en múltiples objetivos (*Many-Objective Optimization*, por su nombre en inglés) se refiere a aquellos problemas de optimización multiobjetivo que cuentan con más de tres objetivos a resolver de manera simultánea [122]. Actualmente el área de la

optimización en múltiples objetivos es uno de los campos de mayor desarrollo investigativo y aplicación en el área de la computación evolutiva. A pesar de que es un área relativamente nueva, es importante mencionar que algunos estudios relacionados con esta área han sido formulados desde los 90s. Uno de los primeros algoritmos que aplicaron un enfoque en múltiples objetivos es el algoritmo MOGA, el cual evaluó cuatro objetivos a un problema de optimización de un sistema de turbinas de gas combustible [182].

Considerando que los problemas de optimización en múltiples objetivos se encuentran inmersos dentro de las optimizaciones multiobjetivo (en donde el número de objetivos $m \geq 4$ para la ecuación (1-11)), la formulación canónica y consideraciones mencionadas en la Sección 1.4.1 son aplicables en cierta medida de forma similar para las optimizaciones en múltiples objetivos. No obstante, dado el incremento de la cantidad de las funciones objetivo, las optimizaciones en múltiples objetivos establecen una serie de retos adicionales respecto a su equivalente multiobjetivo [122].

Considerando que la gran mayoría de algoritmos de optimización multiobjetivo emplean el concepto de dominancia de Pareto, descrito en la Sección 1.4.1, para la identificación de las mejores soluciones, existe una relación proporcional entre el número de objetivos del problema y la porción de población inicial generada aleatoriamente que son soluciones no dominadas. Esto afecta en la generación de nuevas soluciones en cada generación, ralentizando el proceso general de búsqueda de soluciones factibles [183], [184], [185]. Para resolver este inconveniente, algunos estudios proponen modificaciones en el uso de la dominancia de Pareto. Sato [186] propone un cálculo de dominancia parcial, entre soluciones usando un subconjunto de los objetivos, que son intercambiados luego de un número fijo de generaciones. Esta modificación ha permitido obtener mejores resultados en convergencia para la solución del problema del agente viajero (TSP), respecto al algoritmo convencional NSGA-II. Otros autores, como Aguirre y Tanaka [187], particionan el espacio de búsqueda general en subespacios, realizando la búsqueda evolutiva sobre cada subespacio, obteniendo buen rendimiento para problemas de entre 4 a 10 objetivos.

Un problema relacionado con la alta dimensionalidad de los frentes de Pareto es el aumento en la población necesaria para representarlo adecuadamente. Papadimitriou [188] y Erlebach [189] evalúan distintas suposiciones para establecer un conjunto Pareto-

óptimo aproximado, cuyo tamaño pueda ser determinado por un polinomio dependiente de la longitud de la cadena de la solución.

Otro gran problema identificado en los algoritmos de optimización en múltiples objetivos es la tendencia a ser costosos computacionalmente. En algunos casos, las matrices de rendimiento empleadas (tales como el hipervolumen) requieren un consumo de recursos de cómputo exponencial a medida que incrementa el número de objetivos. Con el fin de mejorar la velocidad de cálculo sin afectar la precisión del algoritmo, se han establecido algoritmos para el cálculo del hipervolumen de manera aproximada, útiles en optimizaciones en múltiples objetivos [190], [191], [192].

Por otra parte, algunas medidas de diversidad (tales como la distancia de apiñamiento o *crowding distance*) pueden convertirse en costosas computacionalmente, dada la necesidad de identificación de soluciones vecinales en espacios de alta dimensionalidad. Para resolver esto, Deb y Jain [184] proponen el uso de puntos de referencia basados en puntos aproximados generados a partir de un conjunto de soluciones Pareto-óptimas distribuidas ampliamente a lo largo del frente de Pareto. Con el fin de reducir el costo computacional de los indicadores y métricas de desempeño, diversos autores [193], [194], [195], [196], [197], [198] han creado indicadores y métricas específicas para optimizaciones en múltiples objetivos, tomando como referencia los indicadores presentados en la Sección 1.4.2.

El incremento del número de objetivos considerados por el problema también tiene un efecto importante en la visualización de los resultados y la toma de decisiones [122]. Dado que los problemas de más de tres dimensiones son difíciles de representar gráficamente, se han establecido alternativas de visualización de las soluciones obtenidas mediante la representación tridimensional parcial de tres de los objetivos enfatizados, con el fin de reducir los objetivos analizados [199]. También es frecuente el uso de diagramas de

coordenadas paralelas²¹ [200], mapas de calor²² [201], [202], diagramas radiales [202], [203] y mapas geodésicos [204].

1.5.1 Extrapolación de algoritmos multiobjetivos en optimización MaOO

Con el fin de resolver problemas de optimización en múltiples objetivos a partir de la ampliación de las características de los algoritmos de optimización multiobjetivo, diversos autores [205], [206], [207] han explorado una variedad de algoritmos, partiendo de la construcción de problemas con dos o tres funciones objetivo, para posteriormente añadir extensiones o variantes para la adición de mayor cantidad de funciones objetivo [122]. Uno de los métodos más populares utilizados es el dado por algoritmo NSGA-II [146], el cual se toma como referencia para la comparación de nuevos algoritmos. Entre las modificaciones de NSGA-II consideradas en la literatura para problemas en múltiples objetivos se encuentran el uso de un archivo de ϵ -dominancia, ajuste adaptativo en el tamaño de población y la división del espacio de búsqueda en hiper-cajas [205], [206], [207].

Por otra parte, Bringmann [192] propone un algoritmo que emplea una noción formal de aproximación, denominado algoritmo AGE (*Approximation-Guided Evolution*, por sus siglas en inglés). Este algoritmo establece un ajuste basado en aproximaciones aditivas y la adición del hipervolumen cubierto para funciones de prueba (benchmarking). A pesar de su buen rendimiento para problemas con múltiples objetivos, el algoritmo AGE no es eficiente para problemas de optimización en un solo objetivo o multiobjetivo menores a 3 funciones objetivo. A partir de este algoritmo, Wagner y Neumann [208] proponen una mejora basada en una manera meticulosa para la determinación de las aproximaciones, denominada AGE-II. Esto es alcanzado bajo el uso de la ϵ -dominancia, permitiendo mejorar las aproximaciones aditivas deseadas para problemas de prueba de hasta 20

²¹ Los diagramas de coordenadas paralelas son gráficas bidimensionales que permiten representar a las funciones objetivo en el eje x y a los valores de las soluciones en el eje y.

²² Para la representación de los resultados mediante un mapa de calor, las columnas representan a la función objetivo, mientras que las filas representan a las soluciones, empleando valores relativos asociadas a una paleta de colores definida.

funciones objetivos. A diferencia de su antecesor, AGE-II tiene buen rendimiento para problemas con pocos objetivos.

Otro algoritmo que ha presentado buenos resultados para optimizaciones en múltiples objetivos es el denominado MOEA/D [209]. Este algoritmo contempla un método de descomposición para trasladar el problema a un conjunto de problemas de optimización escalares, los cuales pueden ser resueltos simultáneamente empleando algoritmos evolutivos tradicionales. Se han realizado estudios comparativos para la evaluación de este algoritmo con funciones de pruebas en múltiples objetivos, observando buenos resultados. Giagkiozis [210] propone un algoritmo denominado MACE-gD, que sigue el concepto de descomposición integrando un método de entropía cruzada como una estimación de la distribución de las soluciones respecto al frente de Pareto teórico. Este algoritmo es útil en el establecimiento de las soluciones de Pareto si la geometría del frente de Pareto es conocida previamente.

Jain y Deb [184] proponen el algoritmo NSGA-III, el cual usa un conjunto de puntos de referencia basados en aproximación para optimizaciones en múltiples objetivos. La cantidad de puntos de referencia es similar al tamaño de la población, permitiendo asegurar diversidad de los puntos de referencia asociados a una distribución uniforme sobre el hiperplano normalizado. Este algoritmo ha mostrado un rendimiento superior respecto a MOEA/D y NSGA-II siendo evaluado para 15 objetivos, permitiendo preservar la diversidad de las soluciones a partir de puntos de referencia bien distribuidos [184]. Posteriormente, los mismos autores han propuesto una variante denominada A-NSGA-III, con el fin de modificar adaptativamente la cantidad de puntos de referencia, dependiendo la proximidad de los individuos en distintas partes del frente no dominado actual. Finalmente, establecieron una extensión unificada del algoritmo NSGA-III denominado U-NSGA-III, el cual es capaz de resolver problemas de un solo objetivo, multiobjetivo y en múltiples objetivos, siendo uno de los algoritmos más eficientes hasta la fecha [211].

2. Marco metodológico

La metodología propuesta para diseñar un método para la selección de solventes de extracción, empleando como herramienta el diseño molecular asistido por computadora (CAMD) y la optimización multiobjetivo como criterio de decisión secuencial, parte del trabajo desarrollado por Serrato [1] y Rodríguez [2], [3], que conforma la línea base; y por ello se considera el presente estudio como una extensión de los anteriores.

Este capítulo describe la metodología realizada para aplicar cada una de las etapas del diseño molecular asistido por computadora (CAMD), haciendo énfasis en el establecimiento de parámetros y especificaciones requeridas para la identificación del conjunto de moléculas (solventes) óptimos para los casos de estudio ácido láctico/agua (LA/W) y ácido acético/agua (AA/W).

2.1 Planteamiento metodológico del problema

La metodología aplicada para resolver el problema se enmarca en la propuesta por Harper y Gani [44], aplicada por Serrato [1] y por Rodríguez [2], con las tres etapas del método CAMD, de forma que sea posible formular, resolver y validar el modelo de selección de solvente que permita representar compuestos aromáticos y heterocíclicos. La **Figura 2-1** presenta un diagrama de flujo con la metodología desarrollada en esta investigación. A continuación, se describen brevemente las etapas del diseño CAMD, las cuales se detallan en las Secciones 2.2, 2.3 y 2.4.

1. **Prediseño:** En esta etapa, se realiza inicialmente la formulación matemática del problema de diseño CAMD, estableciendo las funciones objetivo, restricciones y variables de optimización, representadas en un conjunto de ecuaciones consistentes. Para realizar la solución del problema formulado, es necesario integrar un algoritmo de construcción molecular, un modelo de estimación de

propiedades y un algoritmo de optimización multiobjetivo. Respecto al algoritmo de construcción molecular, se establecieron reglas de construcción para individuos alifáticos, heterocíclicos y aromáticos, empleando la representación SMILES y un identificador de aromaticidad molecular, para los individuos generables para el algoritmo de optimización. Por su parte, los modelos de estimación de propiedades permiten identificar los atributos de los individuos, en términos de propiedades físicas, termodinámicas, ambientales, coeficiente de actividad, disponibilidad en el mercado y costo del solvente. Posterior a la formulación matemática, es necesario la integración de los parámetros del algoritmo de optimización multiobjetivo empleado (HAEA/MOHAEA), junto con la definición de los operadores genéticos considerados, el mecanismo de selección de individuos, la representación de la dominancia de Pareto y la evaluación de métricas de convergencia y diversidad entre respuestas obtenidas por el algoritmo evolutivo.

2. **Diseño:** En esta etapa, se realiza la construcción de la herramienta CAMD, en donde se integran simultáneamente el algoritmo de construcción molecular, el modelo de estimación de propiedades y el algoritmo de optimización. La construcción del código fuente fue realizada en el lenguaje de programación Python, integrando las bases de datos requeridas para la estimación de propiedades e implementando una interfaz gráfica de usuario para el programa (GUI). Una vez construida la herramienta, se realizó la ejecución de corridas experimentales, mediante el establecimiento de las condiciones de realización de los experimentos en las corridas CAMD, junto con la definición de criterios para elegir los solventes candidatos a validar experimentalmente en la siguiente etapa.
3. **Post-diseño:** En esta etapa se realizó la selección de los candidatos obtenidos computacionalmente, con el fin de evaluarlos de manera experimental mediante técnicas analíticas (Cromatografía HPLC, valoración de Karl Fisher y valor ácido) y comparar los indicadores de K y S entre los modelos teórico (CAMD) y experimental.

De esta manera, se valida que el modelo desarrollado en la herramienta computacional describe de forma apropiada el comportamiento real de los solventes evaluados en los sistemas ácido láctico/agua (LA/W) y ácido acético/agua (AA/W), con el fin de determinar si son capaces de resolver el problema de extracción o no.

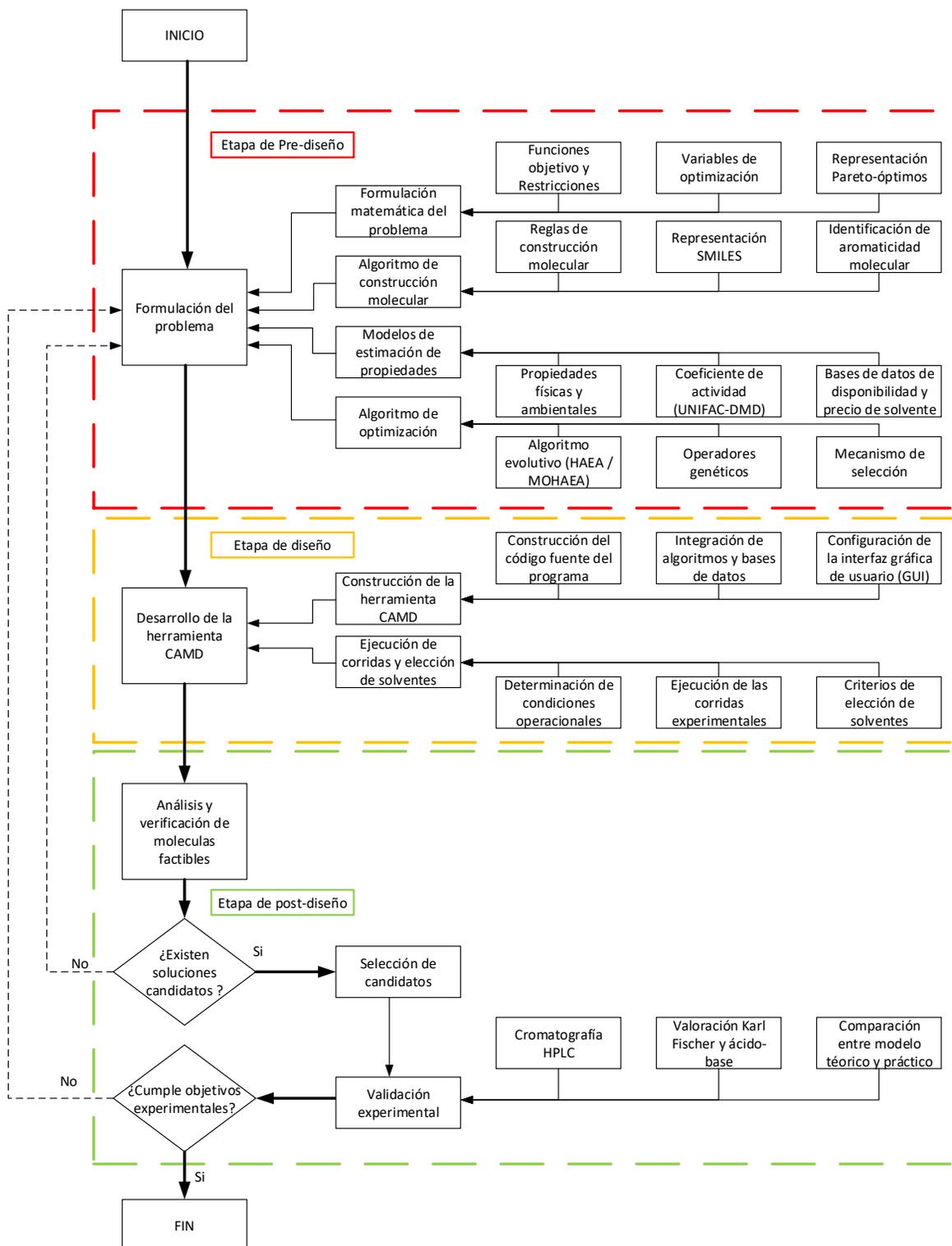


Figura 2-1: Metodología desarrollada en el presente trabajo.

2.2 Prediseño: Formulación del problema

Previo al desarrollo de la herramienta e implementación del diseño CAMD, fue necesario realizar la identificación del problema de optimización, mediante la formulación de las funciones objetivo y restricciones a considerar (descritas en la Sección 2.2.1). A partir del planteamiento del problema a resolver, es necesario establecer los algoritmos integrados requeridos para la solución del problema de diseño CAMD.

De acuerdo con lo mencionado previamente en la Sección 2.1, el modelo CAMD cuenta con tres partes claramente diferenciables: el algoritmo de construcción molecular, modelos de estimación de propiedades y un algoritmo de optimización. A continuación, se describirán brevemente los aspectos asociados a estas partes del modelo.

Con respecto al algoritmo de construcción molecular, es necesario establecer las características estructurales de las moléculas objeto del estudio, junto con la definición de una forma de representación integrable con facilidad dentro del algoritmo de optimización. El algoritmo de construcción molecular incluye tanto moléculas alifáticas como aromáticas y heterocíclicas, permitiendo identificar y diferenciar una molécula en aspectos estructurales para evaluar sus propiedades posteriormente. Las reglas contempladas para la construcción molecular son descritas a profundidad en la Sección 2.2.2, junto con la definición de la representación SMILES.

Dado que en los estudios previos de Serrato [1] y Rodríguez [2], la representación de estructuras aromáticas y heterocíclicas está limitada para algunos conjuntos de compuestos específicos, fue necesario formular una metodología de actualización del algoritmo de construcción molecular para esos compuestos. Esta metodología es descrita detalladamente en la Sección 2.2.3 para compuestos monocíclicos heteroatómicos, siendo extrapolada posteriormente para grupos policíclicos heteroatómicos en la Sección 2.2.4. La estrategia para la determinación de la aplicabilidad de estos grupos heterocíclicos y el impacto de la aromaticidad en su construcción es descrita en las Secciones 2.2.5 y 2.2.6, respectivamente.

Con respecto a la evaluación de las propiedades moleculares, se emplearon expresiones para la estimación de propiedades físicas, termodinámicas y ambientales de sustancias puras, detalladas en la Sección 2.2.7. Además, se actualizó el modelo de estimación del coeficiente de actividad basado en métodos de contribución de grupos, descritos a detalle en la Sección 2.2.8. Para la evaluación de la disponibilidad y costo del solvente, definidas en la Sección 2.2.9, se requirió de la integración de una serie de bases de datos, las cuales reportan estas características no asociadas a la estructura molecular. Con todos los modelos anteriores, es posible establecer relaciones de impacto entre las propiedades y las funciones objetivo, las cuales fueron obtenidas mediante una metodología de análisis estadístico multivariable, descrita en la Sección 2.2.10.

Finalmente, se utilizó un algoritmo de optimización multiobjetivo, que cuenta con la versatilidad para realizar la implementación numérica en un tiempo de cálculo razonable. Para ello, se implementó el algoritmo MOHAEA, descrito en la Sección 2.2.11. Este algoritmo adaptativo, desarrollado por Gómez [103], [151], [104], permite analizar la dominancia entre dos soluciones en un número de funciones objetivo cualesquiera, introducir operadores genéticos específicos más allá del cruce y la mutación, entre otros aspectos. Los operadores genéticos sugeridos para la implementación del algoritmo en el modelo CAMD son descritos en la Sección 2.2.12, mientras que la noción de dominancia de Pareto y evaluación de la distancia entre dos soluciones factibles se describe con detalle en la Sección 2.2.13. Los aspectos relacionados con los mecanismos de selección de individuos y el manejo de restricciones dentro del algoritmo evolutivo son descritos en la Sección 2.2.14.

Adicionalmente, para cada una de las etapas antes mencionadas se presentan las diferencias entre los modelos de Serrato [1] y Rodríguez [2], [3], con respecto a la metodología propuesta y desarrollada en el presente documento.

2.2.1 Formulación de las funciones objetivo y restricciones

Partiendo de los estudios de línea base, se presenta un listado de funciones objetivo considerados en los trabajos de Serrato [1], Rodríguez [2], [3], y en este, las cuales son presentadas en la **Tabla 2-1**. Por su parte, la **Tabla 2-3** presenta un resumen de las

restricciones consideradas en el problema de optimización multiobjetivo para resolver el problema de selección de solvente.

Tabla 2-1: Funciones objetivas empleadas en este trabajo y anteriores [1], [2], [3].

Función Objetivo (FO)	Ecuación	Modelo
FO1: Maximizar el producto KS (y)	$\max y = KS = \frac{\gamma_{T,P}^{\infty} \gamma_{P,S}^{\infty} MW_T}{\gamma_{T,S}^{\infty} \gamma_{T,S}^{\infty} MW_S} \quad (2-1)$	S, R
FO2: Minimizar el impacto ambiental (ε)	$\begin{aligned} \min \varepsilon = & k_{WS} \log(W_S) + k_{BFC} \log(BFC) \\ & - k_{LD50} \log(LD_{50}) \\ & - \log(LC_{50}^{FM}) - \log(LC_{50}^{DM}) \end{aligned} \quad (2-2)$	R, A
FO3: Maximizar el poder de extracción (K)	$\max K = \frac{\gamma_{T,P}^{\infty} MW_P}{\gamma_{T,S}^{\infty} MW_S} \quad (2-3)$	A
FO4: Maximizar la selectividad (S)	$\max S = \frac{\gamma_{P,S}^{\infty} MW_T}{\gamma_{T,S}^{\infty} MW_P} \quad (2-4)$	A
FO5: Minimizar la pérdida de solvente (L)	$\min L = \frac{1}{\gamma_{S,P}^{\infty}} \frac{MW_S}{MW_P} \quad (2-5)$	A

S = Modelo de Serrato [1], R = Modelo de Rodríguez [2], [3], A = Actual estudio.

En el caso particular de la función objetivo ambiental (FO2), propuesta por Rodríguez [2], [3], se toman los factores de normalización k_{WS} , k_{BFC} y k_{LD50} para las variables W_S , BFC y LD_{50} obtenidos en dicho estudio a partir de la evaluación de estas propiedades para 50.000 compuestos generados aleatoriamente. La **Tabla 2-2** presenta los factores de normalización obtenidos por Rodríguez [2], [3]. En el presente estudio, se mantienen dichos factores, para la aplicación de esa función objetivo.

Tabla 2-2: Factores de normalización k_{WS} , k_{BFC} y k_{LD50} reportados por Rodríguez [2], [3] para la evaluación de la función objetivo de impacto ambiental (ε)

Variable	Factor de normalización	Valor reportado
W_S	k_{WS}	5.235
BFC	k_{BFC}	1.272
LD_{50}	k_{LD50}	2.771

Tabla 2-3: Restricciones consideradas en el problema CAMD a optimizar [1], [2], [3].

Restricción (RE)	Ecuación	Modelo
RE1: Punto de fusión (T_f)	$T_f < 293K$ (2-6)	S, R, A
RE2: Punto de ebullición (T_b)	$T_b < 573K$ (2-7)	S, R, A
RE3: Energía libre de Gibbs de formación estándar (G_f)	$G_f < 100 \text{ kJ/mol}$ (2-8)	S, R, A
RE4: Pérdida de solvente (L)	$L < 0.1$ (2-9)	S, R, A
RE5: Disponibilidad en el mercado (A)	$A = i \in \{C_{available} \cup C_{HTA}\}$ (2-10)	R, A

S = Modelo de Serrato [1], R = Modelo de Rodríguez [2], [3], A = Actual estudio.

Con respecto a las restricciones del problema de optimización multiobjetivo, se contemplaron como punto de partida las propuestas por Serrato [1] para la consideración de propiedades físicas y termodinámicas (Restricciones RE1-RE4, de la **Tabla 2-3**). Adicionalmente, se contempló la disponibilidad en el mercado, de acuerdo con la restricción recomendada por Rodríguez [2], [3] (Restricción RE5). Para que el diseño se realizara bajo las mismas condiciones, se fijaron los mismos valores de las restricciones propuestos por Serrato [1] y Rodríguez. [2], [3], variando exclusivamente en la cantidad de funciones objetivos empleadas para la solución del problema de optimización.

Las condiciones establecidas por las restricciones son definidas de acuerdo con consideraciones termodinámicas relacionadas con los sistemas a evaluar en los casos de estudio (ácido láctico/agua – LA/W y ácido acético/agua – AA/W). Las restricciones RE1 y RE2 aseguran que tanto el solvente como el soluto se encuentren en fase líquida. Por su parte, la restricción RE3 permite evaluar compuestos factibles termodinámicamente²³. Por otra parte, existen dos restricciones (RE4 y RE5) adicionales relacionadas con condiciones operativas esperadas en el estudio, con el fin de asegurar una pérdida máxima de solvente permitida y la disponibilidad en el mercado del solvente candidato (respecto a las bases

²³ La energía libre de Gibbs de formación es un indicador de la estabilidad del compuesto, siendo un valor negativo sinónimo de factibilidad termodinámica de la formación del compuesto. No obstante, se sugiere dar una holgura en la restricción para los compuestos antes mencionados.

de datos integradas al modelo). A pesar de que el modelo actual considera el precio del solvente, como se menciona en la Sección 2.2.9, no se integró restricción asociada a esta característica.

2.2.2 Reglas de construcción molecular y representación estructural

El proceso de construcción molecular de los individuos involucrados en la optimización CAMD es fundamental para la representación genotípica de las moléculas (individuos) dentro del algoritmo evolutivo a emplear. De acuerdo con las características asociadas al problema de extracción líquido-líquido objeto del presente estudio, las moléculas construidas deben cumplir con las siguientes reglas de diseño:

- El número de aristas de enlace de cada nodo del subgrupo funcional debe ser igual a la valencia del subgrupo. Por ejemplo, el número de enlaces de un subgrupo CH₃- (1) debe ser uno, ya que esta es la valencia respectiva, mientras que el subgrupo CH=CH (6) debe contar con dos enlaces.
- La metodología CAMD permite establecer un número máximo de nodos de grupos funcionales por molécula, donde ningún individuo puede contener más grupos funcionales de los permitidos. El máximo número de nodos predeterminado por molécula en el modelo de Serrato [1] y el modelo de Rodríguez [2], [3] es de diez. Esto se hace para acotar la dimensión del problema a un conjunto de datos factible y tratable matemáticamente.
- El número de nodos de subgrupos fuertes no debe exceder un máximo permitido de tres. Un subgrupo es fuerte si no pertenece al grupo alquilo (grupo principal 1, subgrupos 1 a 4) o alqueno (grupo principal 2, subgrupos 5 a 8).
- En el caso de compuestos aromáticos, es necesario verificar que los compuestos construidos cumplan con la regla de Hückel.
- En el caso de compuestos heterocíclicos, debe respetarse la valencia respectiva del heteroátomo (-O-, -N-, -S-).
- Por estabilidad molecular, se establece el tamaño máximo de los ciclos de individuos cíclicos en un valor máximo de 6 átomos, ya sean carbonados o con los heteroátomos antes mencionados (-O-, -N-, -S-).

Para la representación de la estructura molecular de una manera compacta, se seleccionó la notación SMILES (“*Simplified Molecular Input Line Entry Specification*”, por sus iniciales en inglés). Esta notación de línea, propuesta en 1988 por Weininger [212], que usa caracteres imprimibles (ASCII), es empleada ampliamente para introducir y representar estructuras moleculares sin ambigüedades. Las representaciones SMILES no contienen espacios, y generalmente, suelen omitir los átomos de hidrógeno de las estructuras [213].

Entre las ventajas del uso del modelo SMILES [214] se encuentran:

- SMILES se considera como una construcción lingüística (idioma), de vocabulario simple y pocas reglas gramaticales. Para el desarrollo de este proyecto, se emplearon solo 4 de las 5 reglas gramaticales básicas, debido a la ausencia de iones dentro de las estructuras generadas. El detalle de las reglas SMILES empleadas se presenta en el **Anexo A**.
- SMILES contiene la misma información que puede encontrarse en una base de datos tabular. Las representaciones de estructuras SMILES pueden utilizarse como cadenas de texto o palabras en el vocabulario de programas diseñados para el almacenamiento de información química.
- La notación de línea de una molécula es compacta, ya que permite describir grafos químicos en función de nodos (átomos) y arcos (enlaces).
- Debido a que la representación molecular se realiza con una notación compacta, el almacenamiento de estructuras moleculares se realiza en un menor espacio, respecto a una base de datos tabular. Esta característica es clave para el desarrollo de la herramienta del presente trabajo.
- Usando la representación SMILES, el nombre de una molécula es sinónimo de su estructura. No existe una misma representación SMILES que origine dos moléculas diferentes.
- Pueden existir múltiples representaciones SMILES para caracterizar una misma molécula. Por tal razón es necesario considerar algoritmos de canonicalización que permitan representar a dicha molécula en una representación SMILES única (canónica). Existen gran variedad de estos algoritmos disponibles en la literatura [215].

- Es posible representar estructuras aromáticas, ya sea de forma directa o en forma Kekulé. Para convertir una representación SMILES de estructura directa a forma Kekulé, existen algoritmos de “*kekulización*” disponibles en literatura [215].

Adicional a las reglas de representación en notación SMILES descritas en el **Anexo A**, para la representación de grupos funcionales aromáticos y heterocíclicos es necesario tener en cuenta aspectos como la aromaticidad de las moléculas y la representación de aspectos quirales, isotópicos o tautoméricos.

Con respecto a la aromaticidad, la representación SMILES utiliza una versión extendida de la regla de Hückel para identificar iones y moléculas aromáticas. Para calificar como aromática una molécula, todos los átomos en el anillo deben tener hibridación sp^2 y el número de electrones p "en exceso" disponibles deben satisfacer el criterio $4N + 2$ de la regla de Hückel. Debe considerarse que solamente los átomos de C, N, O, P, S, As, y Se pueden considerarse como aromáticos. En el presente trabajo, se acotó la generación de moléculas aromáticas con átomos de C, N, O y S. Adicionalmente, los dobles enlaces exocíclicos no rompen la aromaticidad. De acuerdo con la molécula, se pueden presentar dos escenarios:

- Compuestos aromáticos: Para estos compuestos, un ingreso de una entrada y/o la generación de una molécula en forma Kekulé, debe conducir a la detección de aromaticidad, empleando el criterio de Hückel. Como resultado, el algoritmo de representación molecular debe realizar una conversión estructural a una representación aromática. La **Figura 2-2** presenta el caso del benceno, a modo de ejemplo.
- Compuesto antiaromáticos: De acuerdo con la IUPAC, los compuestos antiaromáticos, son sustancias que contienen $4N$ electrones π (con $N \neq 0$) en un sistema cíclico planar, o casi plano, de enlaces sencillos y dobles alternados [216]. En este caso, el modelo SMILES debe buscar una estructura que conserve la hibridación sp^2 implícita, mediante el recuento de hidrógenos implícitos en la estructura. La **Figura 2-3** presenta el caso del ciclobutadieno y el ciclooctatetraeno, a modo de ejemplo.

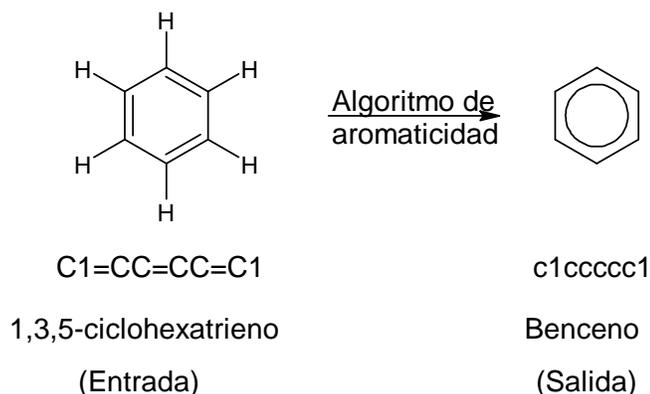


Figura 2-2: Representación SMILES del benceno en forma Kekulé y representación aromática, obtenida del algoritmo de identificación de aromaticidad.

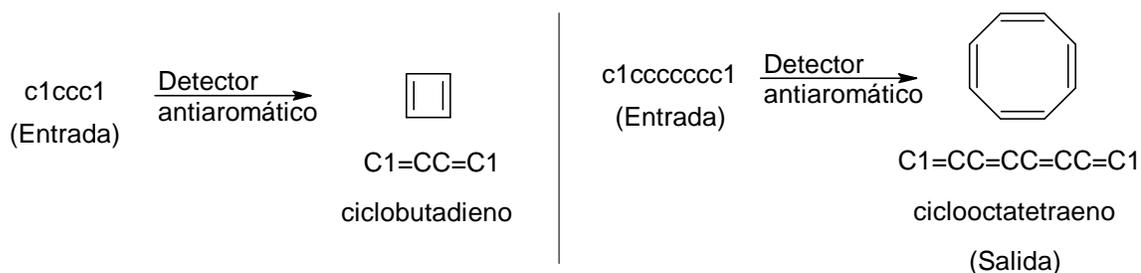


Figura 2-3: Detección de aromaticidad del ciclobutadieno y ciclooctatetraeno.

Debe precisarse que el fin del algoritmo de detección de aromaticidad en la notación SMILES solo es para la representación de la información química, por lo que no tendrá injerencia en el modelo de estimación de propiedades físicas ni en el algoritmo de optimización multiobjetivo.

En lo que respecta a la información quiral, debe precisarse que, aunque existen notaciones SMILES específicas asociadas a quiralidad parcial y total [214], no es de interés del algoritmo de optimización ni del modelo multiobjetivo CAMD, generar moléculas con diferencias en aspectos especulares. Esto, debido a la imposibilidad del modelo de contribución de grupos seleccionado de representar moléculas de manera diferencial con isomería espacial (estereoisómeros, diastereoisómeros, mezclas racémicas, formas meso, isomería cis-trans, etc.).

De acuerdo con lo anteriormente mencionado, cualquier molécula cuya valencia sea distinta de cero, no puede ser representada adecuadamente por el modelo de contribución de grupos (isótopos [217], iones [217], [218]²⁴ y tautomerías²⁵). Por otra parte, a pesar de que las isomerías estructurales generarán representaciones SMILES diferentes, para efectos prácticos el modelo de contribución de grupos asociará el mismo conjunto y cantidad de grupos funcionales, por lo que no serán distinguibles entre sí para la estimación de propiedades físicas. La **Figura 2-4** presenta a modo de ejemplo el caso de los isómeros del xileno (o-xileno, m-xileno y p-xileno).

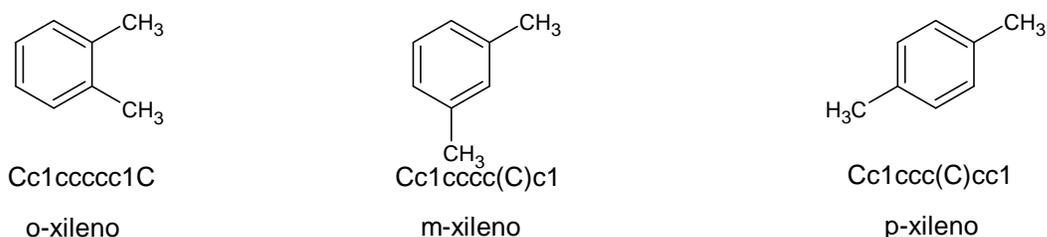


Figura 2-4: Representación SMILES de los isómeros estructurales del xileno.

Por otra parte, para evitar confusiones es necesario precisar en algunas moléculas aromáticas, particularmente el nitrógeno aromático. Los tres tipos comunes de nitrógeno aromático pueden especificarse con el símbolo de nitrógeno aromático (*n*). Se toma como ejemplo el caso de piridina, piridina-N-óxido y pirrol en donde el nitrógeno pirrolil presente en el 1H-pirrol se puede expresar como [nH] para distinguir este tipo de nitrógeno de un

²⁴ Las moléculas con carga no neutra (isótopos, aniones, cationes, complejos, zwitteriones, etc.) no puede ser representadas por el modelo de contribución de grupos empleado en este trabajo (UNIFAC / UNIFAC-DMD, descrito en la Sección 2.2.8) ya que estos modelos contemplan moléculas de valencia cero. Para estos casos, se pueden emplear los modelos GC+ de Mostafa para iones y Van Krevelen para radicales poliméricos.

²⁵ Las tautomerías (isómeros diferentes solo en la posición de un grupo funcional) cuentan con migración de átomos dados por el equilibrio químico (por ejemplo, tautomerías cadena-anillo como la presentada en la glucosa y tautomerías ceto-enol como las presentadas en el acetaldehído y propanona) tampoco pueden ser representados por el modelo de contribución de grupos empleado en este trabajo (UNIFAC / UNIFAC-DMD) ya que los grupos funcionales deben encontrarse fijos en la molécula, sin variaciones dinámicas asociadas al equilibrio producto de migraciones de hidrógeno o formación de puentes de hidrógeno.

piridil-N. Las representaciones SMILES válidas son 1H-pirrol incluyen las configuraciones descritas por la **Figura 2-5**.

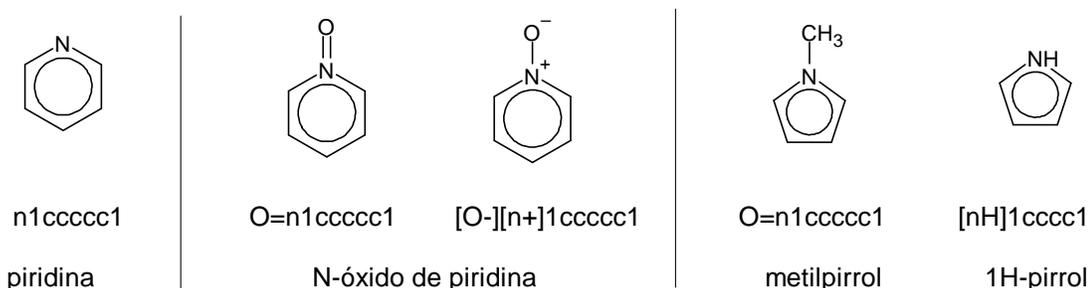


Figura 2-5: Representación del 1H-pirrol con la representación SMILES.

Todas estas consideraciones contempladas para los aromáticos y heterocíclicos fueron tenidas en cuenta para la representación SMILES de este tipo de moléculas, de forma que sea posible realizar la representación de los individuos generados en el algoritmo CAMD. También se verificó la aplicación de los operadores genéticos y la creación de nuevas moléculas.

Considerando el hecho de que los estudios de Serrato [1] y Rodríguez [2], [3] cuentan con algunas estructuras aromáticas (limitadas a las obtenibles mediante los métodos de estimación de propiedades de las Secciones 2.2.7 y 2.2.8), fue necesario modificar la construcción de moléculas para contemplar grupos heterocíclicos. Por tal razón, se desarrolló una metodología de actualización del algoritmo de construcción molecular que integrara tanto las sustancias aromáticas como heterocíclicas, de manera que estas estructuras fueran utilizables por el algoritmo CAMD. Esta metodología propuesta es descrita con detalle en la Sección 2.2.3.

2.2.3 Metodología de actualización del algoritmo de construcción molecular aplicado en compuestos monocíclicos heteroatómicos

La finalidad de la presente sección es describir las modificaciones realizadas en lo que respecta al algoritmo de construcción molecular para incluir representaciones estructurales para moléculas aromáticas y heterocíclicas complejas. Para ello, se propone una metodología para el estudio de las distintas configuraciones estructurales de los compuestos químicos de carácter heterocíclico, desde moléculas cicladas factibles con 3

a 6 átomos, con al menos una sustitución (o varias) de carbono por un heteroátomo (N, S, O), que integre enlaces sencillos dobles y triples en la configuración estructural de los ciclos.

La metodología propuesta cuenta con tres filtros principales, descritos por la **Figura 2-6** que permiten determinar los grupos a adicionar al modelo CAMD: I.) Ausencia de simetrías estructurales con otros grupos, II.) Cumplimiento de valencia válida para cada átomo presente en la estructura, y III.) Plausibilidad de determinación de propiedades del grupo heterocíclico con los modelos de estimación empleados en el modelo CAMD (Hukkerikar y UNIFAC-DMD).

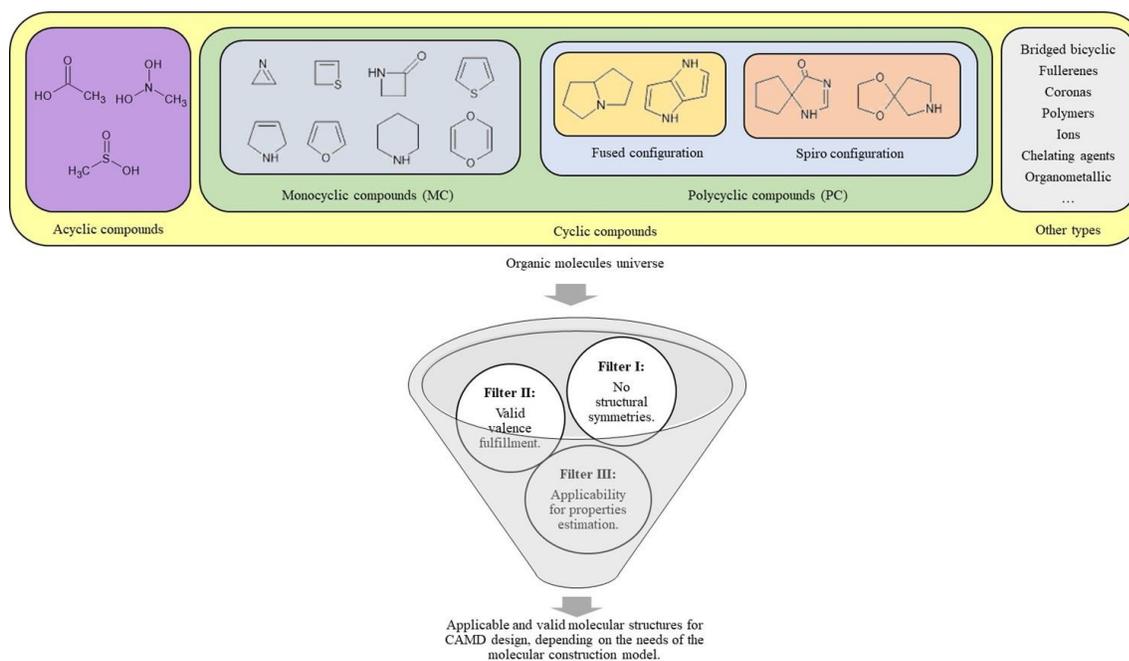


Figura 2-6: Metodología de tres filtros propuesta para compuestos cíclicos.

A continuación, se realiza la descripción de las características de cada uno de los filtros propuestos para la actualización del algoritmo de construcción molecular, haciendo énfasis en compuestos monocíclicos y policíclicos.

Filtro I. Ausencia de simetrías estructurales con otros grupos.

En esta etapa, se formuló una estrategia para evaluar las simetrías estructurales para compuestos cíclicos que permitiera identificar estructuras no repetidas por simetría (isomorfias²⁶) de forma simplificada, a partir de una función generadora, con el fin de evitar la realización de una segregación manual. Para ello, fue necesario integrar algunos aspectos de la teoría de grafos, teoría de grupos y análisis combinatorio.

Para cada polígono regular conformado por n lados, se tomaron los grupos diedrales²⁷, D_n , con el fin de determinar los distintos tipos de simetrías posibles. Con el fin de evaluar las moléculas orgánicas de ciclos formados por entre 3 y 6 átomos ($3 \leq n \leq 6$), fue necesario determinar los distintos tipos de simetría geométrica de las estructuras moleculares, asumiendo que la estereoquímica de los compuestos representados por ellas es planar. No se contemplaron quiralidades estructurales asociadas a la orientación y ángulos de enlace de manera tridimensional. La **Figura 2-7** presenta los ejes de simetría y los grupos diedrales para cada una de las operaciones de rotación y reflexión (R) factibles en las estructuras cíclicas evaluadas.

²⁶ En teoría de grafos, dos grafos son isomorfos si tienen el mismo número de vértices y los vértices de cada grafo se pueden numerar de 1 hasta n de modo que dos vértices del segundo grafo están unidos por una arista si y solo si los dos vértices del primer grafo que tienen los mismos números están unidos por una arista.

²⁷ Un grupo diedral (D_n) es el grupo de simetría de un polígono regular de n lados, incluyendo rotaciones y reflexiones. Un polígono regular con n lados cuenta con n simetrías rotacionales y n simetrías de reflexión. Si n es par, hay $\frac{n}{2}$ ejes de simetría conectados con los puntos medios de lados opuestos y $\frac{n}{2}$ ejes de simetría conectando vértices opuestos.

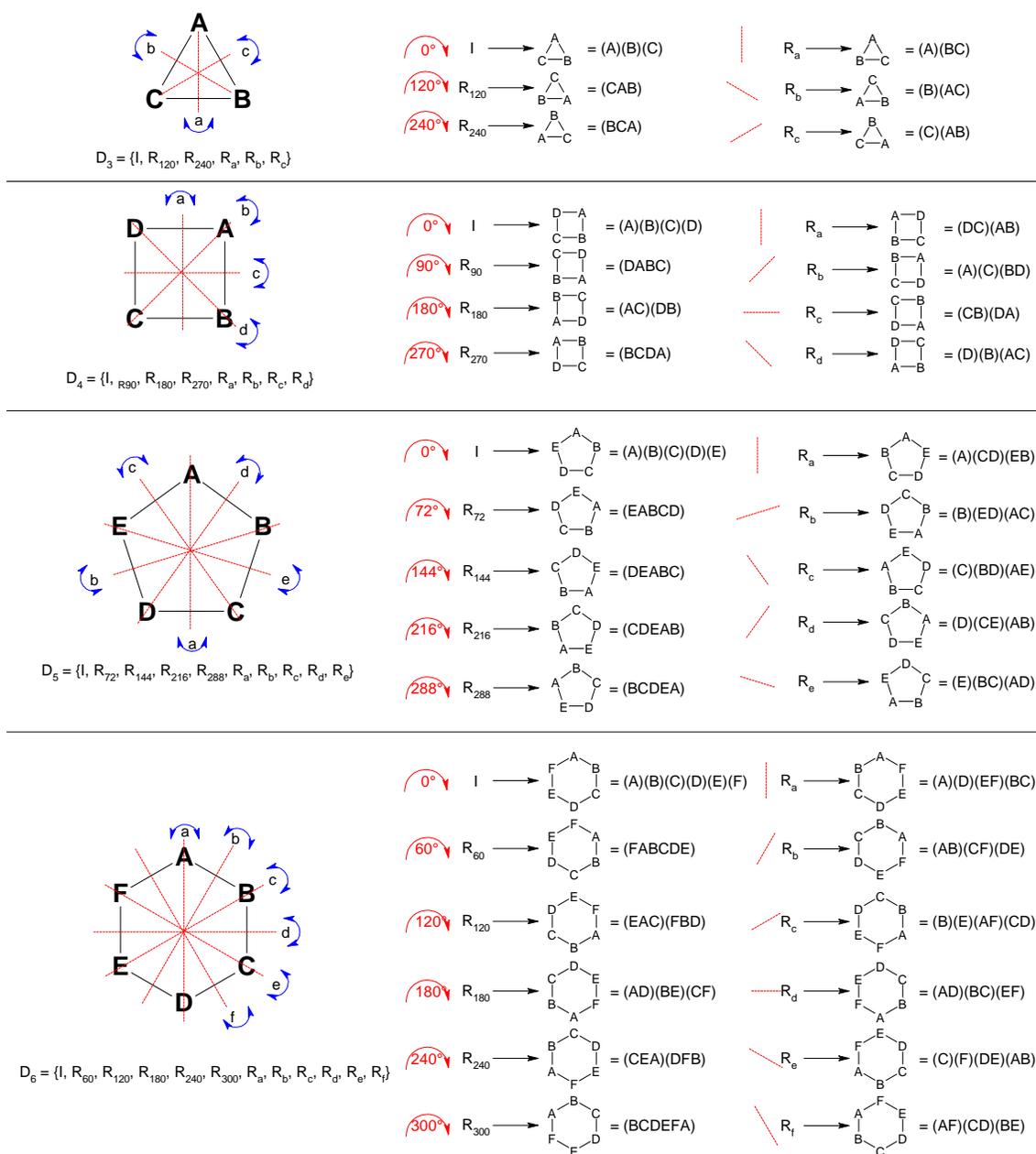


Figura 2-7: Ejes de simetría y grupos diedrales para ciclos (grafos) conformados por 3, 4, 5 y 6 átomos.

De acuerdo con la teoría matemática de grupos, los distintos tipos de movimientos rotacionales de moléculas se pueden agrupar en un conjunto finito denominado grupo, denotado G . Para el uso de los movimientos estructurales, se define un operador denominado acción de grupo (denotada $R_i \circ R_j$, $\circ: G \rightarrow G$, donde $R_i, R_j \in G$). Este operador de aridad dos, consiste en la realización de los movimientos R_i y R_j de manera

consecutiva²⁸. Cada uno de los movimientos estructurales (denotados g , donde $g \in G$) cumplen con las siguientes cuatro propiedades matemáticas²⁹, descritas a continuación [219], [220], [221], [222]:

- Clausura: $R_i \circ R_j = R_k$, donde $R_k \in G$.
- Identidad (I): Existe un movimiento estructural I , tal que $I \circ R_i = R_i \circ I = R_i$
- Inverso ($R_i \rightarrow R_i^{-1}$): Existe un movimiento R_i^{-1} , tal que $R_i R_i^{-1} = R_i^{-1} R_i = I$
- Asociatividad: Para R_i, R_j, R_k , se cumple que $R_i \circ (R_j \circ R_k) = (R_i \circ R_j) \circ R_k$

La realización de dos movimientos moleculares de forma consecutiva es equivalente a realizar un único movimiento molecular específico. Dichas equivalencias de movimientos, a partir de la acción de dos operandos, R_i y R_j , puede ser descrita mediante la Tabla de Cayley. Esta matriz permite establecer cómo se comportan y generan las rotaciones posibles en un ciclo de n -átomos, a partir de sus grupos diedrales³⁰, lo cual es de gran utilidad para el análisis de rotaciones estructurales. El **Anexo B** presenta las tablas de Cayley para las rotaciones equivalentes para los ciclos de 3 a 6 átomos, considerando los elementos (rotaciones estructurales) presentes en los grupos diedrales D_3, D_4, D_5 y D_6 .

El problema de identificación del número de moléculas cíclicas no isomorfas de n -átomos con tres tipos diferentes de enlaces (sencillo, doble y triple), es análogo al problema de conteo de collares coloreados no equivalentes, el cual se resuelve en teoría de grupos

²⁸ La connotación formal matemática del operador de acción de un grupo G sobre un conjunto no vacío X contempla dos acciones importantes: i.) la acción del grupo simétrico S_X sobre los elementos de X , mediante permutación de sus elementos y, ii.) la acción del grupo lineal $GL_n(\mathbb{R})$ sobre los puntos del espacio n -dimensional \mathbb{R}^n , mediante la multiplicación de una matriz por un vector. Se utilizó una aproximación aplicada de la definición formal para la resolución del Filtro I.

²⁹ La propiedad conmutativa no es aplicable, es decir, los movimientos estructurales no se comportan como un grupo abeliano (Ab). Esto se puede evidenciar en la tabla de Cayley, en donde, un grupo es abeliano, si y solo si su representación matricial de la tabla de Cayley es una matriz simétrica respecto a la diagonal principal.

³⁰ La tabla de Cayley se comporta como una abstracción de las posibles permutaciones de todos los elementos del grupo, siendo un “esqueleto de identidad” asociado al tamaño del ciclo orgánico. Por tal razón, ninguna fila o columna de una tabla de Cayley puede contar con el mismo elemento (rotación estructural) dos veces. Un ejemplo análogo de este comportamiento puede identificarse en estructuras matriciales como los cuadrados mágicos y los Sudokus.

mediante el conteo de órbitas³¹ no isomorfas, mediante el teorema de Cauchy-Frobenius³², descrito por la ecuación (2-11).

$$\#Órbitas (grupos no isomorfos) = \frac{1}{|G|} \sum_{g \in G} |Fix(g)| \quad (2-11)$$

Donde:

$Fix(g) := \{x \in X: g \circ x = x\}$, es el conjunto de puntos fijos bajo $g \in G$.

$|G|$ corresponde a la cardinalidad del grupo G . En términos prácticos, $|G|$ será igual a la cantidad de elementos pertenecientes al grupo diedral respectivo D_n .

Para efectos prácticos, la determinación de la cantidad de grupos puede determinarse directamente mediante la generación de un polinomio indicador de ciclos, basado en el Teorema de Enumeración de Pólya, el cual es una generalización del teorema de Cauchy-Frobenius para generación de funciones generatrices con coeficientes finitos [223], [224], [225]. El teorema de enumeración de Pólya, integra una función generadora multivariante llamada índice de ciclo $Z(x_1, x_2, \dots, x_n)$, descrito por la Ecuación (2-12).

$$Z(x_1, x_2, \dots, x_n) = \frac{1}{|G|} \sum_{g \in G} x_1^{b_1(g)} x_2^{b_2(g)} \dots x_n^{b_n(g)} \quad (2-12)$$

Donde:

n es el número de elementos de X , es decir, la cantidad de vértices pertenecientes a un k – ciclo en la representación condensada de una rotación estructural específica, de acuerdo con la **Figura 2-7**.

³¹ Entiéndase como **órbita** de un elemento X como el conjunto de elementos en los cuales x puede ser movido por los elementos de G . En términos informales, las órbitas corresponden a las clases de orientaciones estructurales equivalentes formadas por la acción (rotación estructural) entre dos o más grupos.

³² También conocido ampliamente como Lema de Burnside, debido a un error histórico en la atribución del autor de la expresión.

b_1, b_2, \dots, b_k son los números de k – ciclos del elemento del grupo g como una permutación de X , es decir la cantidad de repeticiones de un k – ciclo en la representación condensada de una rotación estructural, de acuerdo con la **Figura 2-7**.

k – ciclo, son los distintos grupos de tamaño k sobre los cuales se puede descomponer una rotación específica descritos en la representación condensada, de acuerdo con la **Figura 2-7**.

A partir del índice de ciclo $Z(x_1, x_2, \dots, x_n)$, es posible enumerar todas las configuraciones estructurales generalizadas para compuestos monocíclicos a partir de una única función generadora, $p(X)$, generada a partir de la sustitución $x_1 = x_2 = \dots = x_n = x = 3$, donde se hacen explícitos los tres tipos de enlaces químicos posibles dentro de la estructura monocíclica (enlaces sencillos, dobles y triples).

Filtro II. Cumplimiento de valencia válida para cada átomo presente en la estructura.

Al momento de sustituir los átomos generales R por átomos de carbono o heteroátomos (N, O, S), algunas de las configuraciones estructurales generalizadas obtenidas en el Filtro I pueden ser directamente no factibles, debido al incumplimiento de las valencias internas asociadas a estos átomos. Con el fin de establecer una manera hábil y extrapolable para determinar cuáles estructuras generalizadas no son factibles por incumplimiento de valencia, se simplificó el análisis combinatorio en dos pasos: evaluación de los tipos de enlace presentes en una estructura generalizada y la determinación de la validez de la sustitución para cada tipo de enlace.

Se evaluaron los 6 tipos de enlaces posibles entre un R-átomo central y dos átomos R^1 y R^2 colindantes dentro de un ciclo (denotados como A, B, C, D y E), para todas las sustituciones de R-átomo central posibles (C, N, O, S), tal como se presenta en la **Figura 2-8**. Las configuraciones resaltadas en rojo incumplen con los criterios de valencia atómica o generan cargas sobre el átomo central (imposibilitando la evaluación posterior de propiedades, necesaria para la aplicación del modelo CAMD), por lo cual son rechazadas esas combinaciones. Se identificaron 12 configuraciones válidas con los criterios anteriores, indicados con color verde. Posteriormente, se realiza el conteo de la cantidad de tipos de enlaces presentes en cada estructura generalizada que superó el Filtro I.

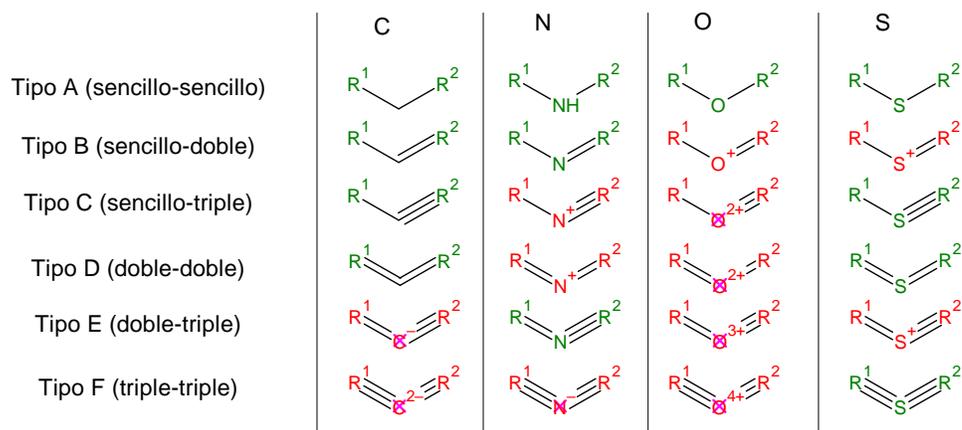


Figura 2-8: Viabilidad de sustitución del R-átomo respecto al tipo de enlace.

Filtro III. Plausibilidad de determinación de propiedades del grupo heterocíclico con los modelos de estimación empleados en el modelo CAMD.

Con el fin de que las estructuras sean aplicables en el algoritmo CAMD del presente trabajo, las estructuras que superaron el filtro II fueron evaluadas en sus propiedades, utilizando los modelos de Hukkerikar y UNIFAC-DMD, descritos a detalle en las Secciones 2.2.7 y 2.2.8. Dado que para determinar el valor numérico de una propiedad es necesario aplicar la estructura molecular del grupo a una molécula de valencia cero, se evaluaron las estructuras anteriores como moléculas individuales, conformadas exclusivamente por estos grupos heterocíclicos, sin ramificaciones ni interacciones con otros grupos alifáticos previamente existentes. Se tomó como criterio de aceptación para el filtro III la validez de la aplicación de ambos modelos de propiedades, más no la diferencia entre el valor de la propiedad para la molécula y el valor reportado experimentalmente.

La metodología de tres filtros antes descrita, permite obtener de manera sistemática los grupos heterocíclicos aplicables al modelo CAMD, considerando exclusivamente aspectos estructurales, de valencia y aplicabilidad de modelos de propiedades. Además, dado que es aplicable para evaluación de estructuras cíclicas, sin importar el tamaño del ciclo o su complejidad, es fácilmente extrapolable a grupos policíclicos (con más de un ciclo en su estructura), requiriendo ajustes menores en los criterios tomados en los filtros anteriores. Esta extrapolación del algoritmo de construcción molecular para grupos policíclicos se realiza en la Sección 2.2.4.

2.2.4 Extrapolación del algoritmo de construcción molecular para aplicación en compuestos policíclicos heteroatómicos

A partir de la metodología de creación de grupos monocíclicos, descrita en la Sección 2.2.3, es posible construir una extrapolación hacia compuestos cíclicos de mayor complejidad. Tal es el caso de los compuestos policíclicos, conformados por dos o más anillos conectados dentro de la misma molécula. Las estructuras químicas policíclicas pueden dividirse en tres tipos de configuraciones: I.) Conformaciones tipo espiro, donde un mismo átomo pertenece a dos o más ciclos, II.) Conformaciones fusionadas, donde dos anillos comparten dos átomos adyacentes y, III.) Conformaciones tipo puente, donde dos átomos no adyacentes son compartidos por dos o más ciclos. La **Figura 2-9** presenta una representación general de los tres tipos de compuestos policíclicos antes mencionados.

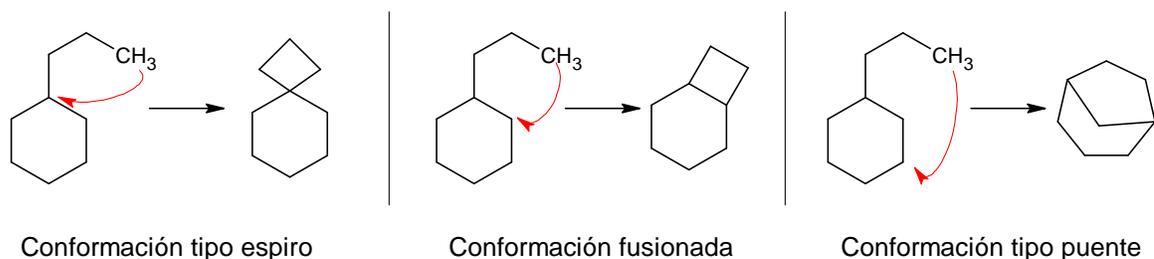


Figura 2-9: Tipos de conformaciones en estructuras policíclicas.

Se realizó el estudio combinatorio para la evaluación de los compuestos policíclicos, acotados para máximo 2 ciclos fusionados entre sí, donde los ciclos internos estén compuestos por entre 3 a 6 átomos, unidos por enlaces covalentes sencillos, dobles y triples. Considerando la enorme cantidad de configuraciones factibles generables para compuestos policíclicos de tipo puente y la poca frecuencia de este tipo de compuestos en listados de solventes relacionados con extracción líquido-líquido (debido a los altos costos de síntesis y dificultad estérica de activa participación en la separación), se decidió enfocar el estudio en la extrapolación de grupos policíclicos para conformaciones tipo espiro y conformaciones fusionadas. A continuación, se describirán los ajustes requeridos en la metodología de 3 filtros, descrita en la Sección 2.2.3, para compuestos policíclicos tipo espiro y fusionados.

Filtro I. Ausencia de simetrías estructurales con otros grupos.

Aplicando los conceptos de teoría de grafos, teoría de grupos y análisis combinatorio mencionados previamente en la Sección 2.2.3, se aplicó el método de enumeración de Pólya para identificar las estructuras repetidas por simetría (isoformas). Las **Figura 2-10** y **Figura 2-11** presentan los ejes de simetría presentes en las conformaciones tipo espiro y fusionadas.

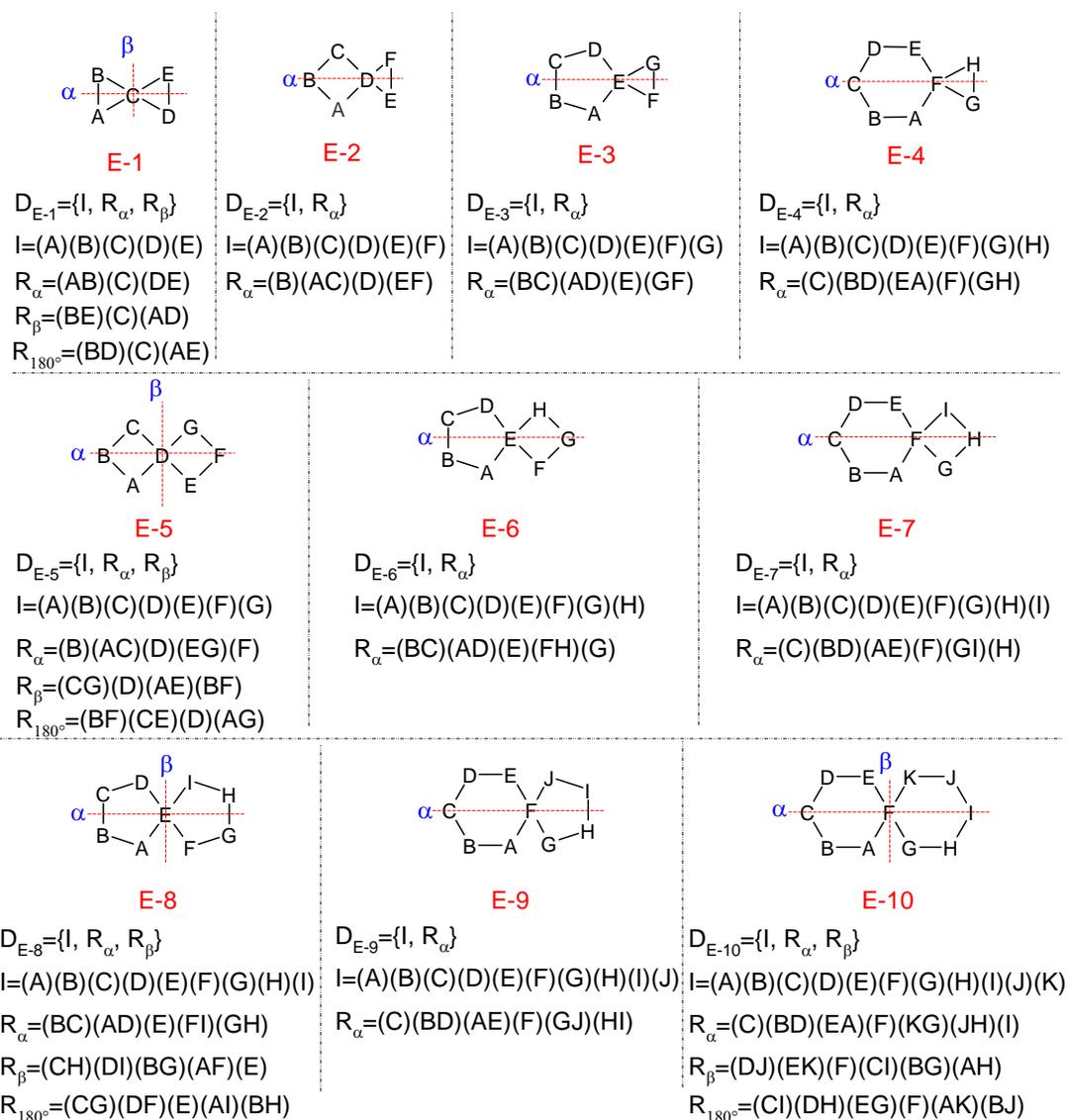


Figura 2-10: Ejes de simetría y grupos diedrales para policíclicos tipo espiro.

A diferencia de las estructuras monocíclicas, las estructuras policíclicas con configuración tipo espiro y fusionadas cuentan con una menor cantidad de ejes de simetría, debido al incremento de la complejidad molecular, lo cual implica que las operaciones de rotación estructural carezcan de sentido por falta de simetría entre dos estructuras cíclicas con distinto número de átomos. Por esta misma razón, la cantidad de movimientos estructurales para compuestos policíclicos tipo espiro y fusionados pueden contar con uno o dos ejes de simetría, en función de si los ciclos relacionados cuentan con el mismo número de átomos o no.

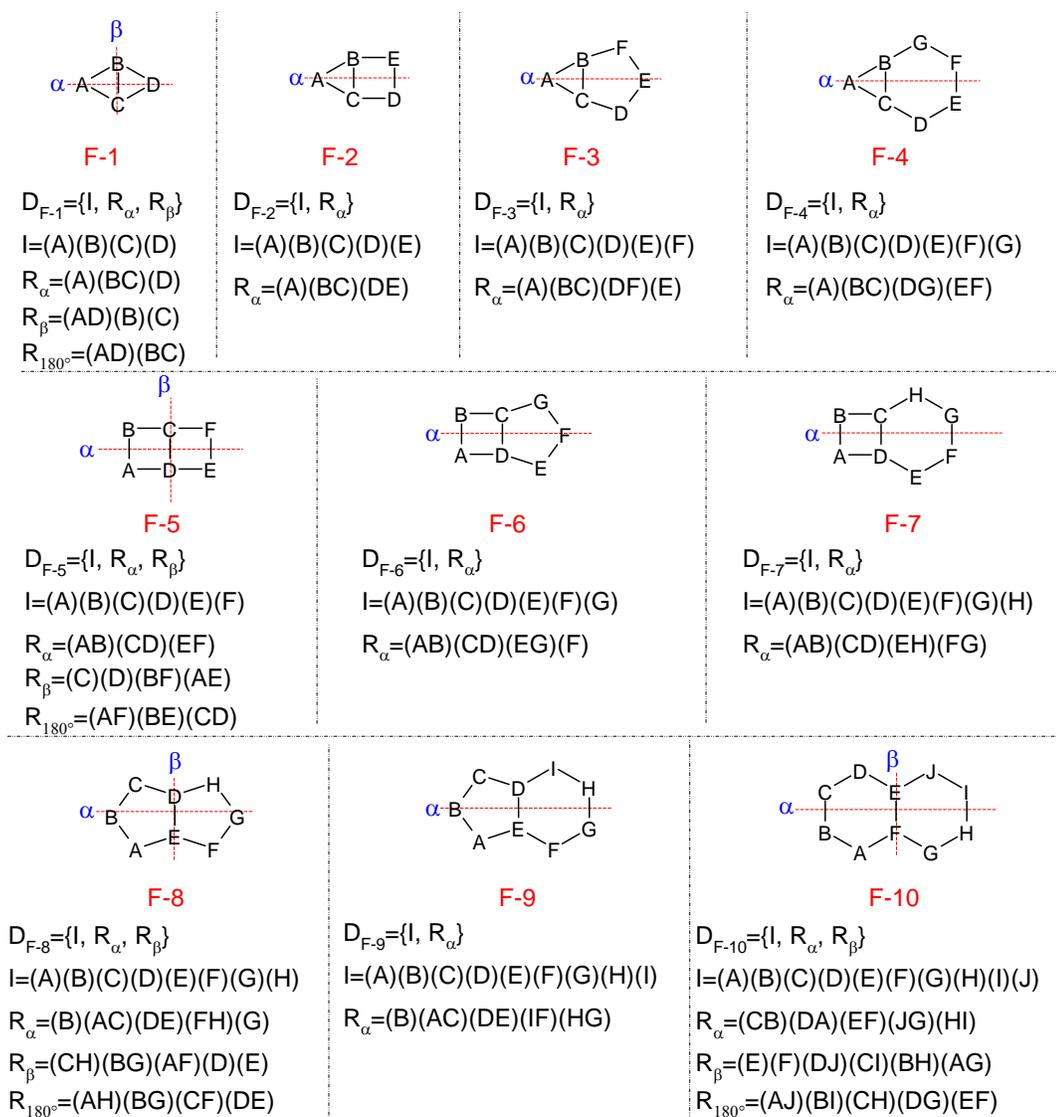


Figura 2-11: Ejes de simetría y grupos de simetría para policíclicos fusionados.

Filtro II. Cumplimiento de valencia válida para cada átomo presente en la estructura.

Dadas la gran cantidad de estructuras generalizadas policíclicas factibles en el Filtro I, es necesario plantear un Filtro II más riguroso respecto al presentado por la **Figura 2-8** en la Sección 2.2.3 para compuestos monocíclicos.

Para el caso de los policíclicos tipo espiro, es necesario establecer una consideración adicional, con el fin de asegurar el cumplimiento de la valencia del espiroátomo³³. Dado que la valencia del espiroátomo debe ser compartida por ambos anillos, el número de escenarios restringidos aumenta, respecto a lo observado en compuestos monocíclicos en la **Figura 2-8**. Por tanto, se realizó la evaluación de los tipos de enlace posibles entre un R-átomo central, dos átomos R_1^1 y R_1^2 (pertenecientes al anillo 1, indicado por el subíndice) y dos átomos R_2^1 y R_2^2 (pertenecientes al anillo 2), considerando la presencia de 2 conjuntos de tres átomos con enlaces colindantes (uno por cada anillo). De esta manera existen 21 configuraciones³⁴ para la sustitución del R-átomo central, siendo este un espiroátomo, tal como se presenta en la **Figura 2-12**.

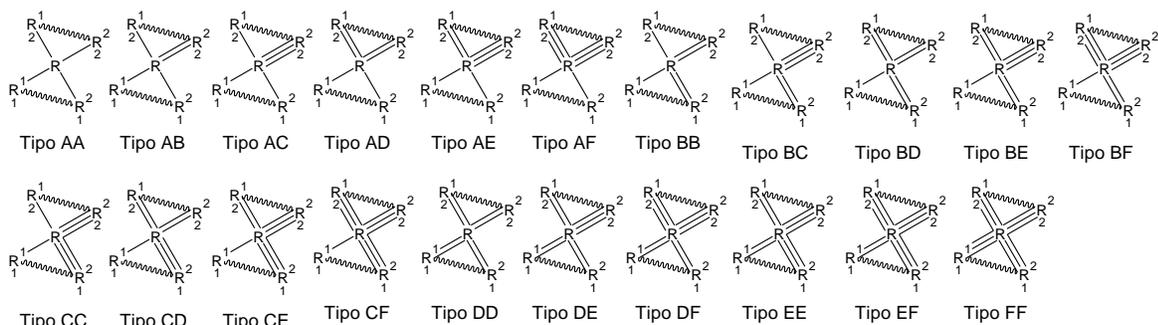


Figura 2-12: Configuraciones estructurales generalizadas para policíclicos espiro.

Para el caso los policíclicos fusionados, se establecieron por simetría (Filtro I) 108 configuraciones estructurales de compuestos policíclicos con configuración fusionada, las

³³ Entiéndase como **espiroátomo** como aquel átomo perteneciente a dos anillos diferentes dentro de un compuesto policíclico tipo espiro.

³⁴ Estas configuraciones están representadas por AA, AB, AC, AD, ..., DF, EE, EF y FF, donde la primera y segunda letra corresponden al tipo de enlace en el anillo 1 y 2, respectivamente.

cuales son presentadas por la **Figura 2-13**. Se consideró el hecho de que dos (o más) estructuras monocíclicas deben contar con al menos un tipo de enlace coincidente entre sí, con el fin de realizar la fusión respectiva.

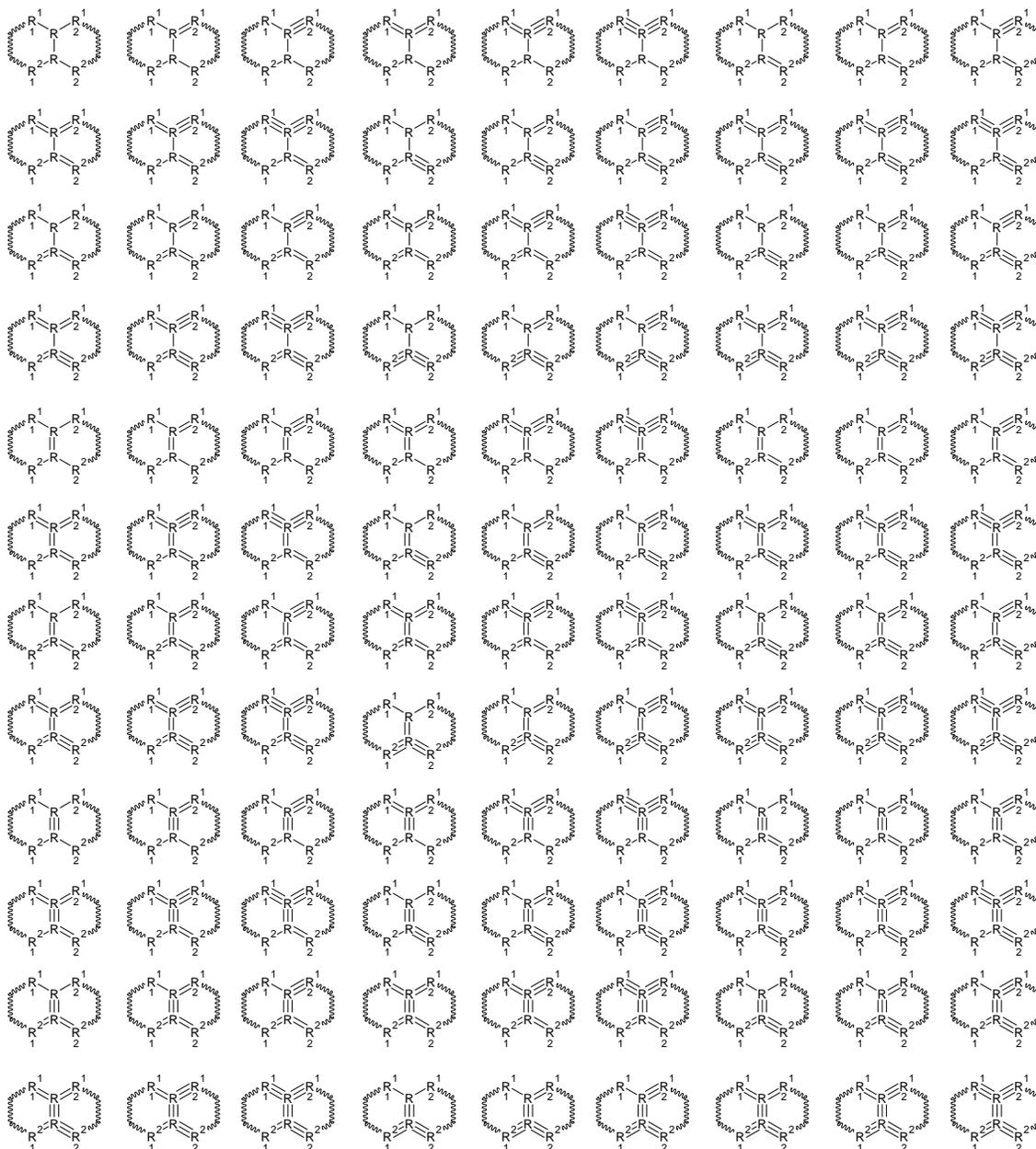


Figura 2-13: Configuraciones estructurales generalizadas para compuestos policíclicos, con configuración fusionada.

Considerando los criterios de viabilidad de valencia de los dos R-átomos presentes en los ciclos fusionados (Filtro II), se establecieron todas las sustituciones de R-átomo centrales posibles (C, N, O, S) de forma similar a los realizados para grupos policíclicos tipo espiro. Debido a la extensa cantidad de estructuras generalizadas a evaluar (108 en total), a modo de ejemplo se presenta en la **Figura 2-14** la evaluación de viabilidad de valencia para el primer grupo de policíclicos fusionados. Las variaciones sobre la fila indican las sustituciones sobre el átomo superior y las columnas indican sustituciones sobre el átomo inferior.

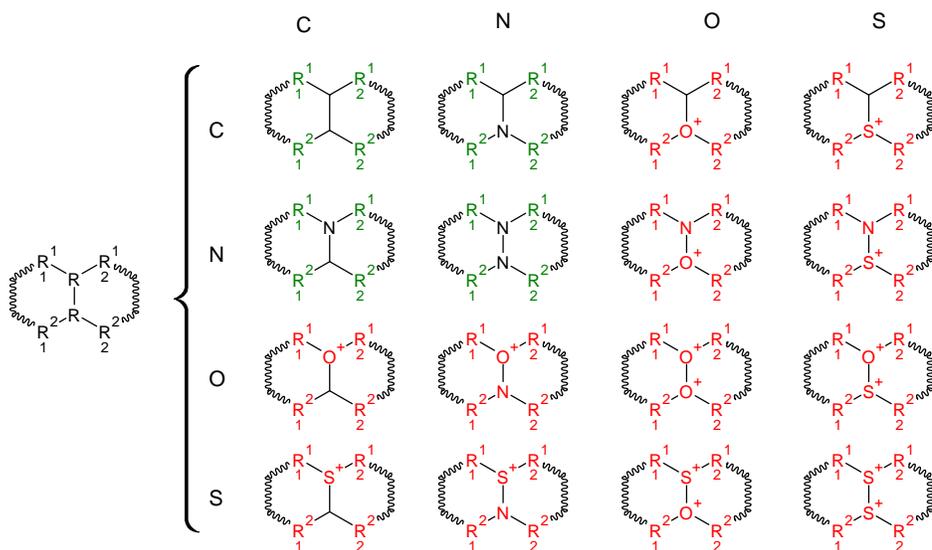


Figura 2-14: Viabilidad de sustitución de R-átomos en policíclicos fusionados.

Filtro III. Plausibilidad de determinación de propiedades del grupo heterocíclico con los modelos de estimación empleados en el modelo CAMD.

De forma análoga a las estructuras monocíclicas, descritas por la metodología de la Sección 2.2.3, se determinó la aplicabilidad de los modelos de Hukkerikar y UNIFAC-DMD, para las estructuras policíclicas tipo espiro y fusionadas que superaron los Filtros I y II. Dadas las características de los grupos policíclicos, se realizó la estimación de sus propiedades, considerando el grupo como una molécula de valencia cero, conformada exclusivamente por los grupos policíclicos, sin ramificaciones ni interacciones con otros grupos alifáticos. Se consideró como criterio de aceptación para el Filtro III la validez de la aplicación de ambos modelos de propiedades, más no la diferencia entre el valor de la propiedad para la molécula y el reportado experimentalmente.

2.2.5 Aplicabilidad de grupos heterocíclicos en moléculas de uso industrial

Posterior a la construcción de grupos heterocíclicos (monocíclicos y policíclicos), fue necesario determinar la aplicabilidad real de los nuevos grupos adicionados al modelo CAMD en moléculas de uso industrial con características estructurales asociadas a estas familias orgánicas. Para ello, se empleó un listado de 162 moléculas³⁵ asociadas a diversas industrias químicas, cuyas estructuras son presentadas en el **Anexo C**.

2.2.6 Impacto de la aromaticidad en la construcción de compuestos cíclicos

La aromaticidad es uno de los aspectos más relevantes de los compuestos cíclicos, ya que está asociada a la estabilidad estructural de la molécula. Por tal razón, se determinó una forma de evaluación del carácter aromático de las moléculas generadas con los grupos adicionados al modelo de construcción molecular (heterocíclicos). Para ello, fue necesario tener en cuenta la consideración de estabilidad asociada al cumplimiento de la regla de Hückel. Esto se desarrolló mediante la atribución de una propiedad adicional para las moléculas cíclicas (monocíclicas o policíclicas), conocida como *carácter aromático* (CA). Esta nueva propiedad, exclusiva de este tipo de moléculas, fue definida con un valor entero: 1 si es aromática, -1 si es antiaromática o 0 si la molécula no es ninguna de las anteriores.

Para determinar el carácter aromático o antiaromático de una molécula en particular, se emplearon las definiciones descritas en la Sección 2.2.2, junto con los algoritmos de canonicalización y kekulización, utilizados para el reconocimiento del subgrafo de deslocalización (DS). Dicho subgrafo, contiene a aquellos átomos y enlaces que han sido identificados como “aromáticos”, y que pueden estar conectados o no, pero que están contenidos dentro del grafo o representación SMILES asociado a la molécula completa [226], [227].

³⁵ Se emplearon 84 moléculas monocíclicas (4, 9, 34 y 37 para ciclos de tres, cuatro, cinco y seis átomos, respectivamente) y 78 moléculas policíclicas (12 con configuración espiro y 66 con configuración fusionada).

Se empleó el listado de moléculas presentes en el estudio de aplicabilidad, reportado por el **Anexo C** para probar los algoritmos de aromaticidad y la determinación de la propiedad CA. Posteriormente, se acoplaron los algoritmos aromáticos al modelo CAMD, para determinar la tasa de compuestos aromáticos, antiaromáticos y no aromáticos obtenidos en las corridas experimentales.

2.2.7 Modelos de estimación de propiedades físicas, termodinámicas y ambientales de sustancias puras

Una vez formulado el problema de optimización y establecido el algoritmo de construcción molecular, es necesario describir los modelos de estimación de las diversas propiedades de las moléculas creadas, requeridas para la evaluación de la aptitud de los individuos en las funciones objetivo, definidas en la Sección 2.2.1.

De acuerdo con lo establecido por Rodríguez [2], la estimación de las propiedades físicas y termodinámicas fue desarrollada empleando el modelo de contribución de grupos (GC+) propuesto por Hukkerikar [66]–[68], el cual realiza la predicción de una propiedad X en función de las contribuciones C_i , C_j y C_k (primer, segundo y tercer orden, respectivamente), tal como se presenta en la ecuación (2-13). La **Tabla 2-4** presenta las propiedades físicas y termodinámicas X estimadas, propuestas en este trabajo usando el modelo de Hukkerikar [66]–[68].

$$f(X) = \sum_i N_i C_i + \sum_j N_j C_j + \sum_k N_k C_k \quad (2-13)$$

El modelo de Hukkerikar, descrito por la ecuación (2-13), parte del modelo de contribución de grupos propuesto por Marrero-Gani [228], en donde cada orden del modelo permite la estimación de propiedades de algunas estructuras particulares, tal como reporta la **Figura 2-15**. Para la aplicación de cada nivel, existen reglas particulares a cumplir, dadas las características de las moléculas evaluadas [66]–[68].



Figura 2-15: Compuestos evaluables por el modelo de Hukkerikar según las contribuciones de primer, segundo y tercer orden [66]–[68].

Con respecto a las propiedades físicas, se mantienen las mismas definidas por el estudio de Rodríguez [2], empleando el modelo de estimación de Hukkerikar [66]–[68]. Por otra parte, el estudio de Serrato [1] contempló el uso de distintos modelos específicos para cada propiedad evaluada, tales como el método de contribución de Constantinou-Gani (para la determinación de la energía libre de Gibbs, los puntos de ebullición y de fusión) y el modelo GCVOL-OL-60 [229] para el cálculo de la densidad. Adicionalmente, Serrato [1] contempló el método de Sheldon, Adjiman y Cordiner [230] para la estimación de propiedades como índice de refracción, momento dipolar y el parámetro de solubilidad de Hildebrand.

Aunque solo Rodríguez [2] contempló en su modelo al conjunto de propiedades ambientales (PA1-PA5), no incluyó las contribuciones de tercer orden, ya que estas se enfocan en la estimación de propiedades para aromáticos y heterocíclicos. Por tanto, en este trabajo, se incluyeron esas contribuciones para estimar las propiedades físicas y termodinámicas de las sustancias.

Tabla 2-4: Propiedades físicas y ambientales consideradas en este proyecto [66]–[68] y anteriores [1], [2], [3].

Tipo	Propiedad	Ecuación f(x)	Constantes
Propiedad Física (PF)	PF1: Punto normal de ebullición, (T_b)	$\exp\left(\frac{T_b}{T_{b0}}\right)$ (2-14)	$T_{b0} = 244,5166 K$
	PF2: Punto normal de fusión, (T_m)	$\exp\left(\frac{T_m}{T_{m0}}\right)$ (2-15)	$T_{m0} = 143,5706 K$
	PF3: Energía estándar de formación, (G_f)	$G_f - G_{f0}$ (2-16)	$G_{f0} = -1,3385 \frac{kJ}{mol}$
	PF4: Volumen molar de líquido a 298K (V_m)	$V_m - V_{m0}$ (2-17)	$V_{m0} = 0,0160 \frac{cc}{kmol}$
	PF5: Densidad (ρ_i)	$\rho_i = \frac{MW}{V_m}$ (2-18)	–
Propiedad Ambiental (PA)	PA1: Concentración letal (50%) en <i>Fathead Minnow</i> (LC_{50}^{FM})	$LC_{50}^{FM} = -\log LC_{50}^{FM} - FM_0$ (2-19)	$FM_0 = 2,1949 \frac{mol}{h}$
	PA2: Concentración letal (50%) en <i>Daphnia Magna</i> (LC_{50}^{DM})	$LC_{50}^{DM} = -\log LC_{50}^{DM} - DM_0$ (2-20)	$DM_0 = 2,9717 \frac{mol}{h}$
	PA3: Dosis letal (oral) en ratas (LD_{50})	$LD_{50} = -\log LD_{50} - A_{LD50} - B_{LD50}$ (2-21)	$A_{LD50} = 1,9372 \frac{mol}{kg}$ $B_{LD50} = 0,0016 \frac{mol}{kg}$
	PA4: Solubilidad en agua, (W_s)	$W_s = \log W_s - A_{WS} - B_{WS}MW$ (2-22)	$A_{WS} = 4,5484 \frac{mg}{L}$ $B_{WS} = 0,3411 \frac{mg}{L}$
	PA5: Factor de bioconcentración (BFC)	$\log BFC$ (2-23)	–

2.2.8 Estimación del coeficiente de actividad con el modelo UNIFAC-DMD

Para la estimación del coeficiente de actividad, se empleó el modelo de UNIFAC-Dortmund (UNIFAC-DMD), el cual es una extensión del UNIFAC original. En el desarrollo de este modelo CAMD se emplean los datos de la Base de Datos de Dortmund, la cual contiene el conjunto de datos termofísicos más grande del mundo [231], [232].

Los modelos UNIFAC y UNIFAC-DMD son empleados para la predicción de los coeficientes de actividad de sustancias no-electrolíticas en mezclas no ideales. Este fue publicado por primera vez en 1975 por Fredenslund, Jones y Prausnitz [233] y se encuentra

en permanente revisión experimental [234]. Parte del concepto de que una mezcla líquida puede ser considerada como una solución de las unidades estructurales que forman las moléculas, más que de una solución de las moléculas mismas [16]. Dichas unidades, llamadas subgrupos, son actualizadas de forma periódica, de acuerdo con la evaluación experimental de dos parámetros requeridos para su descripción: el volumen relativo R_k y el área superficial relativa Q_k . Adicional a estos parámetros, los coeficientes de actividad también dependen de las interacciones entre subgrupos, los cuales son categorizados respecto a un grupo principal. Todos los subgrupos pertenecientes a un mismo grupo principal se consideran idénticos con las interacciones de grupo, por lo que los parámetros que caracterizan dichas interacciones se identifican con pares de parámetros a_{mk} .

La gran ventaja del modelo UNIFAC/UNIFAC-DMD es que permite representar un gran número de moléculas, a partir de una combinación relativamente pequeña de subgrupos previamente establecidos. Cuando es posible construir una molécula a partir de la combinación de más de un conjunto de subgrupos, se opta por aquel conjunto que contiene el menor número de subgrupos.

El **Anexo D** hace una breve descripción de las ecuaciones que rigen a los modelos UNIFAC y UNIFAC-DMD, junto con la enumeración detallada de los parámetros de interacción a_{mk} y los parámetros de subgrupos R_k y Q_k para todos los grupos y subgrupos disponibles en la última versión del modelo a la fecha (6ta revisión, septiembre 2016) [231].

2.2.9 Evaluación de disponibilidad y costo del solvente mediante bases de datos

Uno de los principales inconvenientes en el desarrollo del algoritmo CAMD para la selección de solventes es el hecho de que algunos compuestos candidatos obtenidos pueden no encontrarse disponibles en el mercado, ya sea por consideraciones propias de la molécula³⁶ o por consideraciones de proceso³⁷. El modelo de Rodríguez [3] contempla la disponibilidad de consecución de las moléculas candidatos generadas por el algoritmo

³⁶ Baja estabilidad, facilidad de degradación en compuestos de menor complejidad estérica, formación de tautomerías, entre otros aspectos.

³⁷ Dificil consecución, elevado precio, baja pureza o restricciones dentro de las vías químicas para la síntesis de dicho componente, entre otros aspectos.

CAMD en el mercado químico, mediante dos bases de datos (ZINC y EPA DSSTox), para orientar al algoritmo de optimización en la evaluación de compuestos óptimos factibles y adquiribles en el mercado industrial. A continuación, se realiza una breve descripción de dichas bases de datos.

- ZINC (acrónimo de “ZINC is not comercial”) es una base de datos de acceso público para proyección virtual, contiene millones de compuestos adquiribles con información detallada sobre la estructura y los proveedores [235], de acuerdo con la Universidad de California, San Francisco y el Instituto Nacional de Ciencias Médicas Generales [236]. Actualmente, la base de datos cuenta con más de 1380 millones de sustancias [235], [236]. La base de datos ZINC cuenta con un conjunto de 6 categorías para la clasificación de los compuestos, de acuerdo con su facilidad de compra. Dichas categorías están descritas en la **Tabla 2-5**.
- EPA DSSTox es una base de datos creada por la Agencia de Protección Ambiental de los Estados Unidos (EPA), la cual pretende ser un punto de acceso a información de alta calidad para bioensayos, considerando datos fisicoquímicos de compuestos y sus correspondientes estructuras químicas [237]. Se encuentra disponible en línea, con alrededor de 720.000 compuestos químicos de relevancia en estudios ambientales.

Tabla 2-5: Categorías de compra de sustancias presentes en ZINC [235], [236].

Categoría	Descripción	Tasa de éxito
En stock.	Listo para enviar y entrega prevista en 2 semanas	95%
Agente de adquisiciones.	Disponible a través de agentes de compras, entrega en 2 semanas	95%
Fabricación bajo demanda.	Entrega típicamente dentro de 8 a 10 semanas	70%
Boutique.	El costo puede ser alto, pero es probable que sea más barato que hacerlo usted mismo.	70%
Anotado.	En catálogos, pero actualmente no a la venta	-
Espera OK.	En stock + Agente + Bajo demanda	-

A pesar de que Rodríguez [3] identifica en su modelo la disponibilidad de consecución de un cierto componente, no integra el costo intrínseco para la obtención de cierto solvente. Este trabajo integra tanto el criterio de disponibilidad en el mercado como el precio de solvente como aspectos a considerar en la solución del problema.

La determinación del costo de una sustancia química en el mercado siempre es un reto en sí mismo, ya que mucha de la información específica es de pago. La evaluación de precio del solvente está sujeta a la disponibilidad en el mercado, y está regida por las leyes de oferta y demanda para una región establecida. En este trabajo se emplea la base de datos de precios de productos químicos de la Universidad de Texas A&M [238], la cual cuenta con citaciones de precios actualizados de productos químicos de revistas comerciales tales como Chemical Week [239] y la revista ICIS Chemical Business [240]. Se realizó una recopilación de los precios actualizados disponibles en dicha base de datos, para los solventes con disponibilidad reportada por ZINC y EPA DSSTox, integrando así al modelo CAMD no solo la disponibilidad del solvente sino también su valor en el mercado. Considerando la facilidad de adquisición de información económica relacionada, se toman como referencia los precios en dólares americanos (USD) para una 1 tonelada del solvente químico considerado.

Considerando que el precio en las bases de datos económicas se encuentra reportado para distintos periodos, se efectuó un ajuste al valor presente (año 2022), empleando el índice del precio al productor estadounidense (IPP, por sus siglas en inglés) para diversos tipos de productos químicos, reportado por la oficina estadounidense de estadísticas laborales (*US. Bureau Labor Statistics*) [241].

2.2.10 Determinación del impacto de propiedades en las funciones objetivo

Para la determinación de los impactos de las propiedades sobre las funciones objetivos estudiadas, se realizó un análisis estadístico multivariable, enfocado en establecer relaciones de dependencia entre las funciones objetivo y las propiedades físicas/ambientales. Para ello, se consideró el uso de una base de datos de 28.677 sustancias, suministrada por Müller [242], [243], conformada con estructuras de distintas familias químicas (alifáticas, aromáticas y heterocíclicas, descritas en las categorías de la

Tabla 2-6), a las cuales se les determinaron las propiedades antes mencionadas y se evaluaron las funciones objetivo con los valores de dichas propiedades. Se evaluaron los casos de estudio de las funciones objetivo y sus propiedades para los sistemas LA/W y AA/W.

Considerando que Rodríguez [244], [245] realizó una evaluación de los comportamientos de las propiedades en función de los grupos funcionales, se tomó este enfoque como punto de referencia, con el identificar la variación del error relativo promedio, respecto a la aplicación del modelo de Hukkerikar [66]–[68], usando solo las contribuciones de primer y segundo orden. La adición de la base de datos proporcionada por Müller [242], [243] permite tomar una muestra más amplia de estructuras heterocíclicas (monocíclicas y policíclicas), las cuales no fueron consideradas previamente por Rodríguez [244], [245].

Con el fin de profundizar el tipo de relación entre propiedades y funciones objetivo, se desea resolver de manera estadística la hipótesis: *¿Existe significancia estadística en las relaciones entre una cierta propiedad P y una cierta función objetivo FO?* La resolución de este interrogante para cada una de las funciones objetivo (γ , ε , K , S , L), respecto a una propiedad en particular se evaluó empleando un análisis estadístico multivariado con correlaciones bivariadas, considerando como la muestra de estudio las moléculas de la base de datos de Müller [242], [243].

Tabla 2-6: Grupos funcionales considerados en el estudio del impacto de propiedades vs funciones objetivo, de acuerdo con categorías clasificatorias.

Categorías clasificatorias	Grupos Funcionales
(1.) Alifáticos y acíclicos	(1.1.) Alcanos. (1.7.) Esteres. (1.2.) Alquenos. (1.8.) Éteres. (1.3.) Alcoholes. (1.9.) Nitrilos. (1.4.) Aldehídos. (1.10.) Amidas. (1.5.) Cetonas. (1.11.) Aminas. (1.6.) Ácidos carboxílicos. (1.12.) Halogenuros.
(2.) Aromáticos y cíclicos	(2.1.) Compuestos alicíclicos. (2.2.) Arenos (hidrocarburos aromáticos).
(3.) Compuestos heterocíclicos	(3.1.) Heterocíclicos nitrogenados. (3.2.) Heterocíclicos oxigenados. (3.3.) Heterocíclicos sulfonados.

Con el fin de establecer si la muestra seleccionada cumple con los requisitos de la estadística paramétrica, el conjunto de datos asociados a la relación FO vs P fue objeto de estudio mediante la prueba de normalidad de Kolmogórov-Smirnov (válida para un tamaño de muestra $N > 50$). Esta prueba estadística permite determinar la bondad de ajuste frente a la distribución normal. En función de los resultados de la prueba de normalidad, se empleó el modelo de correlación bivariado³⁸ para realizar el estudio de la relación entre propiedades y funciones objetivo, determinando la fuerza y dirección de asociación entre dos variables aleatorias ordinales.

Tanto el coeficiente de correlación con rho de Spearman como el coeficiente de correlación lineal de Pearson se encuentran acotados bajo el intervalo $[-1,1]$, donde el signo del coeficiente representa una correlación negativa o positiva entre las dos variables y la proximidad al valor límite refleja la fuerza de la correlación. La **Figura 2-16** propone una escala de colores para las asociaciones entre dos variables, de acuerdo con el coeficiente de correlación de Spearman/Pearson.

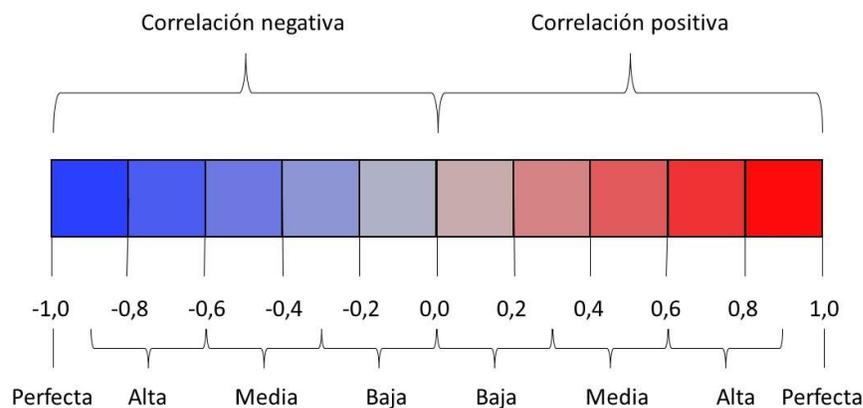


Figura 2-16: Escala de colores para las asociaciones entre relaciones FO vs P, de acuerdo con el modelo de correlaciones bivariados de Spearman/Pearson.

³⁸ Dependiendo de los resultados obtenidos en la prueba de normalidad de Kolmogórov-Smirnov, se aplicó el modelo de correlación de r de Pearson ($r_{Pearson}$) para datos que cumplan con los preceptos de la estadística paramétrica. En caso contrario, se aplicó el modelo de correlación de rho de Spearman ($\rho_{Spearman}$) para datos que se ajusten a los supuestos de la estadística no-paramétrica.

Con el fin de asegurar la no dependencia al azar de los datos obtenidos para el estudio de las asociaciones entre variables, se evaluó la significancia estadística de los pares de datos FO vs. P empleando como criterio de aceptación un valor de α de 0,05 bilateral (probabilidad de cometer un error estadístico tipo I), en ambas colas (significancia estadística del 95%).

Adicionalmente, el estudio estadístico incluye la determinación del tamaño del efecto estadístico, utilizando el criterio d de Cohen. De esta manera, se cuenta con una justificación del grado de afectación de la función objetivo como consecuencia de la propiedad física evaluada. El criterio de Cohen establece cuatro niveles cuantitativos de tamaño de efecto: muy grande (p superiores a 0,5), grande (p entre 0,3 y 0,5, inclusive), medio (p entre 0,1 y 0,3, inclusive) y bajo (p menores a 0,1).

Finalmente, se evaluó el poder estadístico ($1 - \beta$), con el fin de establecer la probabilidad de no cometer un error estadístico de tipo II, es decir, determinar la tasa de falsos negativos (β). Este indicador estadístico permitirá establecer que tan generalizable es la relación determinada por el estudio para un conjunto de pares de variables FO vs P .

2.2.11 Algoritmo de optimización HAEA/MOHAEA

Para la solución del problema de optimización, se emplearon los algoritmos evolutivos HAEA, desarrollado por Gómez [103], [151], [104], junto con su variante multiobjetivo MOHAEA, [152], [246]. HAEA es un algoritmo genético adaptativo que a diferencia de los algoritmos clásicos modifica la probabilidad de los operadores genéticos teniendo en cuenta cuál de ellos mejora en mayor medida la aptitud. Las probabilidades iniciales de los operadores se definen como $1/N_{OG}$, donde N_{OG} es el número de operadores considerados. Las probabilidades para un individuo específico se heredan a su prole, cuál linaje de sucesión. Un individuo debe tener $N_H = 1$, donde N_H corresponden al número de hijos (criterio del hijo único). En caso de que el operador implique descendencia $N_H > 1$, se aplica un torneo para seleccionar al mejor hijo (criterio del hijo único más apto). Otra ventaja de HAEA es la versatilidad de incorporación de nuevos operadores genéticos

personalizados, adaptados al problema particular estudiado. La **Tabla 2-7** presenta el algoritmo HAEA, de acuerdo con la propuesta de Gómez [103], [151] [104].

Tabla 2-7: Algoritmo de optimización de un solo objetivo HAEA [103], [151] [104].

<p>Algorithm 1 HAEA # input: Tamaño de la población λ, condición de parada. HAEA (λ, terminationCondition) 1: $t0 = 0$ 2: $P_0 = \text{InitPopulation}(\lambda)$ 3: while (terminationCondition(t, P_t) is false) do 4: $P_{t+1} = \{\}$ 5: for each ind $\in P_t$ do 6: rates = GetRates(ind) 7: $\delta = \text{Random}(0,1)$ # Tasa de aprendizaje. 8: operator = ChooseOperator(operators, rates) 9: parents = ParentSelection($P(t)$, ind) 10: offspring = ApplyOperator(operator, parents) 11: child = Best (offspring, ind) # Mejores hijos de acuerdo con la aptitud. 12: if (child_{fitness} < ind_{fitness}) then 13: rates[oper] = $(1.0+\delta) * \text{rates}[\text{oper}]$ # Recompensa. 14: else 15: rates[oper] = $(1.0-\delta) * \text{rates}[\text{oper}]$ # Castigo. 16: NormalizeRates(rates) 17: SetRates(child, rates) 18: $P_{t+1} = P_{t+1} \cup \{\text{child}\}$ 19: $t = t + 1$</p>

De acuerdo con los criterios de desempeño para la evaluación del comportamiento de un algoritmo de optimización multiobjetivo, es necesario establecer criterios de convergencia y diversidad, tal como se menciona en la Sección 1.4.2. La implementación propuesta con el algoritmo MOHAEA en el presente trabajo integra ambas nociones de desempeño.

En MOHAEA, el desempeño de los individuos se realiza empleando solo la aptitud física de los individuos de una población, la cual requiere clasificar la población de acuerdo con criterios de no dominancia, determinando la distancia de hacinamiento de los individuos en cada nivel. La aptitud de un individuo i corresponde entonces a la suma del rango de no dominancia (RND) y la distancia de hacinamiento (DH), tal como se observa en la Ecuación (2-24). Esto permite obtener una medida de desempeño que unifique los criterios de desempeño del algoritmo (convergencia y diversidad, descritos en la Sección 1.4.2), independiente del número de funciones objetivo que presente el problema.

$$fitness(i) = DH(i) + RND(i) \quad (2-24)$$

- *Distancia de hacinamiento (DH)*: Se asigna en función de la medida de distancia de hacinamiento presentada en el algoritmo NSGA-II. Este indicador es una estimación de la densidad poblacional de soluciones que rodean una solución particular en un frente de Pareto específico. La distancia de hacinamiento de un individuo (DH_i) se calcula mediante la distancia Manhattan formada por el individuo y los vecinos más próximos pertenecientes al mismo frente de Pareto, tal como se representa en la **Figura 2-17**. En proximidad al individuo, $DH_i \rightarrow 1$ (valor máximo) para cada n-dimensión. El algoritmo de la **Tabla 2-8** describe el proceso para calcular la distancia de hacinamiento de los individuos de un frente.

Luego de calcular todas los DH_i para cada individuo de la población, se determina la contribución total de la distancia de hacinamiento (DH) a la aptitud, empleando la ecuación (2-25). El valor de $DH(i)$ se encuentra acotado en el intervalo (0,1), lo que permite que individuos de un mismo frente de Pareto tengan aptitudes en rangos acotados que no se superpongan con los rangos de otros frentes de Pareto.

$$DH(i) = \frac{1}{2 + DH_i} \quad (2-25)$$

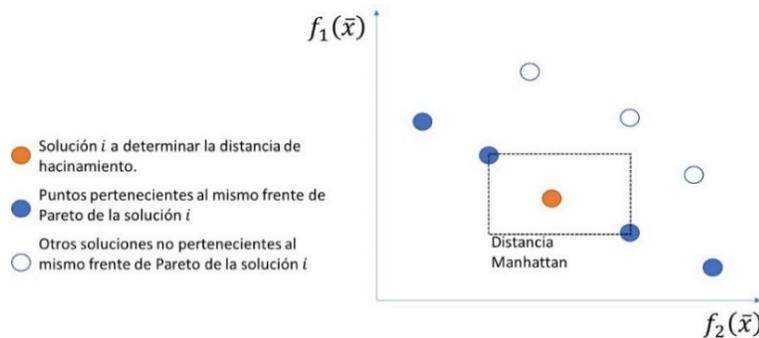


Figura 2-17: Distancia de hacinamiento según distancia Manhattan [2].

- *Rango de no dominancia (RND)*: El rango de no dominancia de la población se define de acuerdo con el enfoque de clasificación rápida no dominada (*Fast Non-Dominated Sorting*) presentada en el algoritmo NSGA-II, descrito en la **Tabla 2-9**. La finalidad de este algoritmo es extraer los diferentes niveles de no dominancia de

una población. La dominancia de Pareto (operador $<$) se emplea para separar cada frente en niveles: el primer nivel (el cual contiene los individuos no dominados), el segundo nivel (con individuos dominados exclusivamente por individuos del primer nivel) y así sucesivamente. El resultado del algoritmo NSGA-II es una lista de rangos de no dominación \mathcal{F} , donde cada rango es una lista de individuos clasificados en el respectivo frente de Pareto.

Tabla 2-8: Asignación de distancia de hacinamiento (DH) para un frente de Pareto, mediante NSGA-II [247].

Algoritmo 3: ASSIGN_CROWDING_DISTANCES(J) # input: Frente de Pareto J.	
1:	$l_J = J $ # Número de individuos en J.
2:	for each i in J do
3:	$J[i]_{\text{crowd-dist}} = 0$ # Distancia de inicialización.
4:	for each m in problem_objectives do
5:	SORT(J, m) # Clasificar J respecto a los valores de m objetivos.
6:	$J[1]_{\text{crowd-dist}} = J[1]_{\text{crowd-dist}} + 1$ # Distancia máxima para puntos en frontera.
7:	$J[l_J]_{\text{crowd-dist}} = J[l_J]_{\text{crowd-dist}} + 1$
8:	for i=2 to (i-1) do
9:	$J[i]_{\text{crowd-dist}} = J[i]_{\text{crowd-dist}} + (J[i+1].m - J[i-1].m) / (J[l_J].m - J[1].m)$

Tabla 2-9: Clasificación rápida no dominada para determinación de rango de no dominancia (RND), mediante NSGA-II [247].

Algoritmo 2: NON_DOMINATED_SORT(P) # input: Población P	
1:	$F = \{\}$
2:	for each p in P do
3:	$S_P = \{\}$ # Conjunto de soluciones dominadas por p
4:	$n_P = 0$ # Conteo de dominación, número de soluciones que domina p
5:	for each p in P do
6:	if p < q then # Si p domina a q
7:	$S_P = S_P \cup \{q\}$ # q está adicionado a las soluciones dominadas por p
8:	else if q < p then # Si q domina p
9:	$n_P = n_P + 1$ # Incremento en el contador de dominación de p
10:	if $n_P = 0$ then
11:	$p_{\text{rank}} = 1$
12:	$F_1 = F_1 \cup \{p\}$ # p pertenece al frente de Pareto 1
13:	i=1 # Inicialización del contador del frente
14:	while $F_1 \neq \{\}$
15:	$Q = \{\}$ # Inicialización del siguiente frente
16:	for each p in F_i
17:	for each q in S_P
18:	$n_q = n_q - 1$
19:	if $n_q = 0$ then # Si todos los miembros que dominan q han sido eliminados
20:	$q_{\text{rank}} = i + 1$
21:	$Q = Q \cup \{q\}$

```
22:   i = i + 1
23:   F1 = Q
24:   return F
```

Teniendo en cuenta todo lo anteriormente mencionado, se procede a describir el algoritmo MOHAEA, mediante las etapas de un algoritmo genético clásico (de acuerdo con lo expuesto en la Sección 1.3.3), con ligeras modificaciones. A continuación, se realiza una descripción breve de las etapas de cálculo en MOHAEA, la cual corresponde a la representación matemática descrita por el algoritmo descrito en la **Tabla 2-11**.

- **Inicialización del algoritmo:** Este algoritmo inicia con la creación de una población inicial aleatoria de tamaño N_p y asignando las probabilidades de los operadores genéticos de manera uniforme para cada individuo de la población ($1/N_{OG}$).
- **Selección I – Elección de operador genético:** De manera análoga a HAEA, el individuo evoluciona empleado un operador genético seleccionado mediante ruleta, de acuerdo con la probabilidad del operador, usando el método SELECCIONAR_OPERADOR. Es decir, la variación de los intervalos de la ruleta de operadores genéticos estará sujeta a los cambios en sus respectivas probabilidades.
- **Aplicación del operador genético y obtención de descendencia:** De acuerdo con la aridad³⁹ del operador pueden suceder dos escenarios:
 1. Si el operador es de aridad 1, éste se aplica directamente al individuo.
 2. Si el operador es de aridad superior a 1, se selecciona el número de individuos faltantes, mediante la estrategia de selección (ruleta), con el método SELECT_PARENTS.

Luego de que se ha generado la descendencia de un individuo padre, los hijos son agregados a la población de descendientes (N_H). La selección y aplicación de un operador para cada individuo, junto con la determinación de la descendencia, debe realizarse para **todos** los individuos de la población. A pesar de que la población de descendientes sea variable por la aridad de las operaciones efectuadas, se

³⁹ Entiéndase **aridad** de un operador genético como la cantidad de individuos requeridos por este operador para generar descendencia.

cumple que $N_{H,min} = N_P$, gracias a los criterios del hijo único y del criterio del hijo único más apto, antes mencionados.

- **Asignación de aptitud de individuos en población combinada:** Se unifican los individuos de la población de padres e hijos en una población combinada, para posteriormente asignar los valores de aptitud de los individuos mediante la ecuación (2-24), considerando tanto el valor del rango de no dominancia RND (calculado con el algoritmo de la **Tabla 2-9**), como la distancia de hacinamiento DH (obtenido con la **Tabla 2-8** y la ecuación (2-25)).
- **Selección II – Determinación de los individuos de la siguiente generación:** Se determinan las familias⁴⁰ presentes en la población combinada. Empleando el operador BEST_BY_FAMILY se determina el mejor individuo de cada familia con mejor aptitud, el cual es añadido en la población de individuos de la siguiente generación. Esta población con los mejores individuos seleccionados contará con un tamaño de λ .
- **Actualización de probabilidades de operadores genéticos para la siguiente generación:** Se realiza de la misma manera como se aplica en el algoritmo HAEA:
 1. Se genera un parámetro δ con un valor aleatorio acotado en intervalo $[0,1]$
 2. Dependiendo del individuo seleccionado de la familia, se suma o resta el parámetro δ al valor actual de la probabilidad del operador. Si el individuo seleccionado es hijo, δ es sumado; mientras que, si el padre es el individuo seleccionado, δ es restado.
 3. Se realiza la normalización de las probabilidades actualizadas de cada uno de los operadores genéticos, asignándole las tasas actualizadas de los operadores genéticos al mejor individuo de la familia, quien ha evolucionado a la siguiente generación. El algoritmo descrito en la **Tabla 2-10** describe como aplicar el operador BEST_BY_FAMILY a un conjunto de familias.
- **Terminación de la generación actual y criterio de finalización del algoritmo MOHAEA:** Luego de que todos los pasos anteriores han sido completados, la nueva población reemplaza a la generación actual, se realiza un incremento unitario del valor del contador de la generación, iterando hasta cuando el criterio de

⁴⁰ En el ámbito del algoritmo HAEA/MOHAEA, entiéndase **familias** como el conjunto de individuos conformados por un padre y los hijos descendientes a partir de él.

terminación TERMINATION_CONDITION sea alcanzado (número de generaciones a realizar, previamente establecidas en la inicialización del algoritmo).

Tabla 2-10: Selección de mejores individuos [103], [151] [104].

Algoritmo 4: BEST_BY_FAMILY (families, operator_rates)	
# Input: Familias, mapa de tasas de operadores genéticos (OG)	
1:	new_population = {}
2:	for each f in families do
3:	SORT ($f_{children}$, 'fitness') # Clasificación de descendencia según la aptitud.
4:	best_child = $f_{children}[1]$ # Hijo con la mejor aptitud.
5:	go_rates = operators_rates [$f_{parents}$] # Índices de OG del padre.
6:	oper = $f_{parents}$ # Operador genético usado para generar la descendencia.
7:	δ = Random (0,1) # Tasa de aprendizaje.
8:	best_individual = {}
9:	if FITNESS (best_child) < FITNESS ($f_{parents}$) then # ¿Es el mejor hijo?
10:	go_rates[oper] = (1.0 + δ) * go_rates[oper] # Recompensa.
11:	best_individual = best_child
12:	else
13:	go_rates[oper] = (1.0 - δ) * go_rates[oper] # Castigo.
14:	best_individual = $f_{parents}$
15:	go_rates = NORMALIZE (go_rates)
16:	operators_rates[best_individual] = ind_rates
17:	new_population = new_population U {best_individual}
18:	return new_population

Finalizada la aplicación del algoritmo MOHAEA, se obtendrán una generación final, la cual puede variar entre experimentos, fundamentalmente por las condiciones iniciales de la población de origen, el comportamiento específico de cada operador genético, la estrategia de selección (I y II) y el manejo de restricciones a lo largo de todas las generaciones. La Sección 2.2.12 describe los operadores genéticos contemplados, su respectiva aridad, junto con una ilustración de los cambios de representación molecular entre padres e hijos. Por su parte, la Sección 2.2.13 describe con detalle el manejo de las restricciones en la implementación y la Sección 2.2.14 presenta la forma de representación gráfica de los resultados de la optimización, asociadas a los frentes de Pareto.

La **Tabla 2-11** presenta el algoritmo de optimización multiobjetivo MOHAEA empleado en el desarrollo del presente proyecto, a partir de las modificaciones y consideraciones descritas anteriormente, Los operadores genéticos empleados por las funciones *ChooseOperator* y *ApplyOperator* son descritas con detalle en la Sección 2.2.12. Por su

parte, los mecanismos de elección de padres empleados por la función *ParentSelection* son detallados por la Sección 2.2.13.

Tabla 2-11: Algoritmo de optimización multiobjetivo MOHAEA [152], [246].

<p>Algoritmo 5 MOHAEA MOHAEA (λ, terminationCondition) # <i>input: Tamaño de población λ, criterio de parada.</i></p>	
1:	$t_0 = 0$
2:	$P_0 = \text{InitPopulation}(\lambda)$
3:	while (terminationCondition(t , $P_{(t)}$) is false) do
4:	$P_{t+1} = \{\}$
5:	UpdateDominanceCount($P_{(t)}$)
6:	for each $ind \in P_t$ do
7:	rates = GetRates(ind)
8:	$\delta = \text{Random}(0,1)$
9:	operator = ChooseOperator(<i>operators</i> , rates)
10:	parents = ParentSelection($P(t)$, ind)
11:	offspring = ApplyOperator(operator, parents)
12:	child, rates = Best*(offspring, ind , rates, oper)
13:	NormalizeRates(rates)
14:	SetRates(child, rates)
15:	$P_{t+1} = P_{t+1} \cup \{\text{child}\}$
19:	$t = t + 1$
<p>UpdateDominanceCount(P_t)</p>	
1:	for each $ind \in P_t$ do
2:	$ind_{<} = 0$
3:	for each $x \in P_t$ do
4:	if $x < ind$ then
5:	$ind_{<} = ind_{<} + 1$
<p>Best*(offspring, ind, rates, oper)</p>	
1:	$x = \text{KNN}(ind, \text{offspring})$
2:	UpdateFitness(x , ind)
3:	if ($x_{fitness} > ind_{fitness}$) then
4:	$x = ind$ <i># La descendencia no es mejor que el padre.</i>
5:	if ($x_{fitness} < ind_{fitness}$) then
6:	rates[oper] = $(1.0 + \delta) \times \text{rates}[\text{oper}]$ <i># Recompensa.</i>
7:	else
8:	rates[oper] = $(1.0 - \delta) \times \text{rates}[\text{oper}]$ <i># Castigo.</i>
9:	$x_{z*} = ind_{z*}$ <i># Asignación de puntos de referencia.</i>
10:	return x , rates
<p>UpdateFitness(x, ind)</p>	
1:	$ind_{fitness} = d_1(f(ind))$
2:	$x_{fitness} = d_1(f(x))$
3:	if ($x < ind$) & ($\neg(ind < x)$) then
4:	$ind_{fitness} = ind_{fitness} + 1$ <i># Penalización por no dominancia de ind</i>
5:	if ($ind < x$) & ($\neg(x < ind)$) then

6:	$X_{fitness} = X_{fitness} + 1$
----	---------------------------------

Penalización por no dominancia de x

2.2.12 Operadores genéticos considerados

En el algoritmo genético, los descendientes son generados mediante el uso de operadores genéticos sobre los individuos padres, en una generación (iteración) específica. Tradicionalmente, los operadores genéticos son el cruce y la mutación. No obstante, la propuesta de nuevos operadores genéticos implementados para el algoritmo CAMD puede ser considerada, tanto para operadores unarios, binarios o ternarios, y está acorde con las necesidades del problema de optimización en sí.

En el trabajo de Serrato [1], la optimización de un solo objetivo (empleando el algoritmo HAEA) contempló los operadores genéticos descritos en la **Tabla 2-12**, representados gráficamente por la **Figura 2-18**. Por su parte, Rodríguez [2], [3] propone la adición de nuevos operadores genéticos para el algoritmo con dos objetivos. No obstante, estos trabajos aplican los operadores genéticos para moléculas alifáticas y/o aromáticos sin cambiar la naturaleza estructural (cíclico o no) de los compuestos antes y después de efectuado el operador genético.

En el presente trabajo, los operadores genéticos se dividieron en dos clases:

- *Operadores genéticos de naturaleza simple*: Estos operadores actúan para un único tipo de moléculas, generando individuos hijos de una naturaleza específica a partir de individuos padre de la misma naturaleza. Por ejemplo, un cruce entre dos padres alifáticos para generar dos individuos hijos alifáticos.
- *Operadores genéticos de múltiple naturaleza*: Estos operadores genéticos actúan en moléculas de distinta naturaleza, permitiendo que la(s) molécula(s) padre(s) generen individuos hijos de distinta naturaleza. Por ejemplo, mutar una molécula alifática para generar una molécula heterocíclica.

La naturaleza de una molécula es una característica estructural intrínseca, pudiendo ser alifática, cíclica, aromática, antiaromática o heterocíclica. Para facilitar el uso de los operadores genéticos previamente establecidos, se incorporó a la molécula un atributo identificador ϖ que permita determinar la naturaleza de una molécula particular, la cual podrá variar o no en función del operador genético empleado. Adicionalmente, se

modificaron los operadores genéticos **OG1–OG6** (descritos en la **Tabla 2-12**), con el fin de que sean operables tanto para especies de distinta naturaleza (diferentes valores de ϖ).

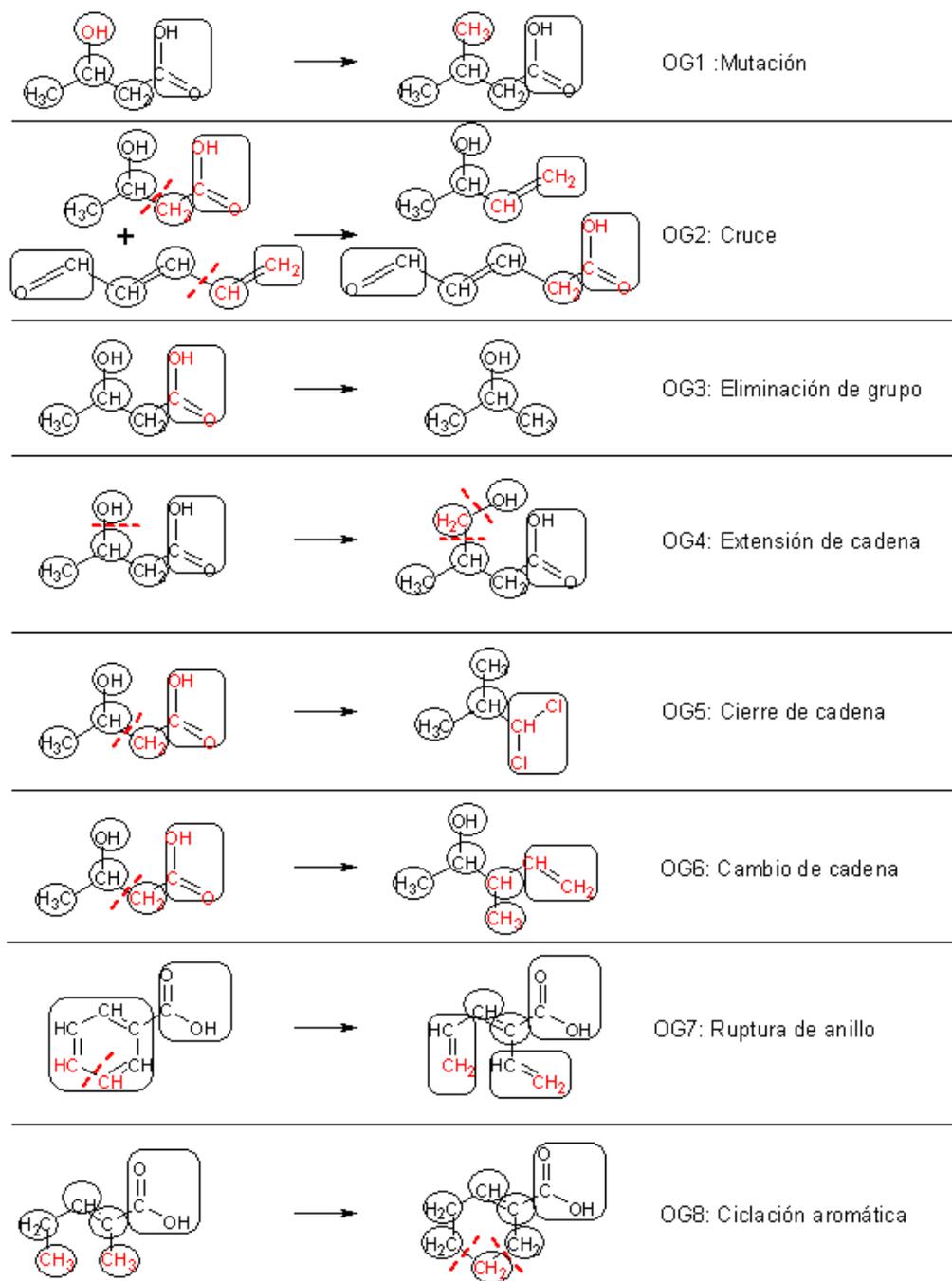


Figura 2-18: Representación gráfica de los operadores genéticos empleados en trabajos anteriores (Serrato [1], Rodríguez [2], [3]) y en este trabajo.

Tabla 2-12: Operadores genéticos empleados en este trabajo y anteriores [1], [2], [3].

Operador (OG) [Aridad]	Descripción del operador genético	Modelo
OG1: Mutación [U]	Selección de una molécula para sustituir un subgrupo del individuo por otro subgrupo con la misma valencia, seleccionado aleatoriamente.	S, R, A
OG2: Cruce [B]	Selección de dos individuos de la población, división de cada individuo en dos partes, y construcción de dos nuevos individuos a partir de las partes divididas.	S, R, A
OG3: Eliminación de grupos [U]	Selección de un individuo de la población, selección de un nodo de subgrupo del individuo y eliminación del subgrupo contiguo de valencia 1, reemplazándolo por un subgrupo alquilo de valencia n-1.	S, R, A
OG4: Extensión de cadena [U]	Selección de un individuo de la población, división del individuo y reunión de las partes, mediante la adición de un subgrupo CH _n adicional.	S, R, A
OG5: Cierre de cadena [U]	Selección de un individuo de la población, división del individuo en dos partes, las cuales una se descarta y la otra se une con un subgrupo terminal seleccionado.	S, R, A
OG6: Cambio de cadena [B]	Selección de una molécula individual de la población, generación aleatoria de otra molécula individuo, cruzamiento entre las dos moléculas. Uno de los individuos se descarta y el otro es el resultado del operador genético.	R, A
OG7: Ruptura de anillo [U]	Se selecciona una molécula individual, con naturaleza aromática o heterocíclica. Se escoge un punto de corte. Se divide la molécula en dos partes diferentes.	A
OG8: Ciclación aromática [U]	Se selecciona una molécula individual, con naturaleza alifática. Se toman los dos puntos extremos de la molécula. Se retira un grupo CH en cada extremo. Se evalúa la valencia de cada molécula y se completa la molécula con un grupo funcional que permita cerrar el ciclo.	A

Aridad: U=Unario, B=Binario. *Modelo:* S = Modelo de Serrato [1], R = Modelo de Rodríguez [2], [3], A = Actual estudio.

2.2.13 Mecanismo de elección de padres y manejo de restricciones del algoritmo

El mecanismo de selección de individuos es clave tanto en la elección de los individuos que participan en los operadores genéticos, como en la selección de los mejores hijos que sobreviven en la generación. Tanto el modelo de Serrato [1] como el de Rodríguez [2], [3] incorporan una selección mediante torneo, donde se realiza la selección de cuatro individuos moleculares de la población de manera aleatoria, eligiendo al individuo con mejor aptitud perteneciente a la muestra tomada. Con el fin de contemplar la propagación en las generaciones de individuos con menor aptitud, evitando el elitismo, se propone modificar el mecanismo de selección del algoritmo genético de un torneo determinístico por un torneo probabilístico.

En este tipo de torneos, se incorpora un parámetro de probabilidad p , acotado entre 0 y 1, generado aleatoriamente a partir de una función de probabilidad (uniforme). Este parámetro permite invertir el resultado del individuo seleccionado, si $p > 0.5$, siendo ganador del torneo aquel individuo que no clasificaría en el torneo determinístico. Si un operador genético requiere la selección de n individuos de la población, se requerirán n torneos para su selección. Adicionalmente, se reduce el tamaño del torneo, pasando a seleccionar dos individuos en vez de 4. La **Figura 2-19** presenta de manera ilustrativa las diferencias entre el torneo determinístico (recomendado por los modelos de Serrato [1] y Rodríguez [2], [3]) y el torneo probabilístico, aplicado en el presente trabajo.

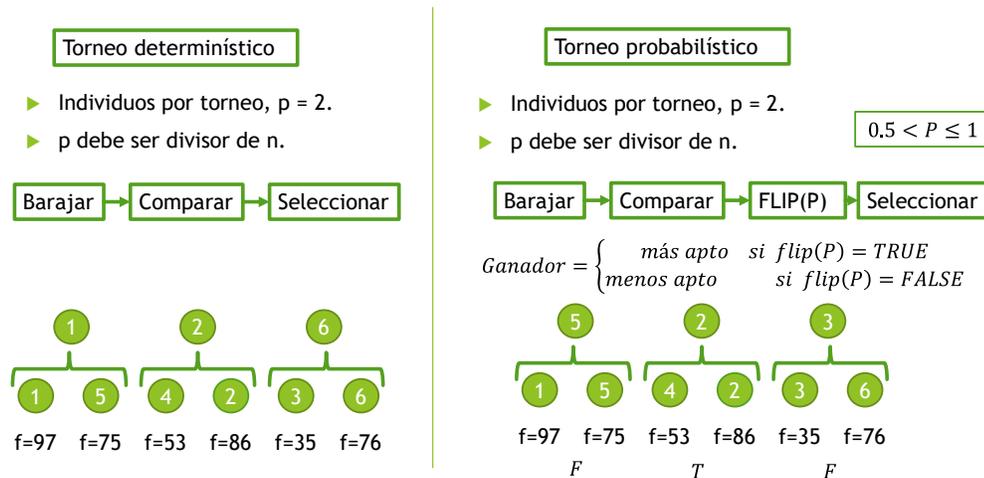


Figura 2-19: Tipos de selección mediante torneos: determinístico vs probabilístico.

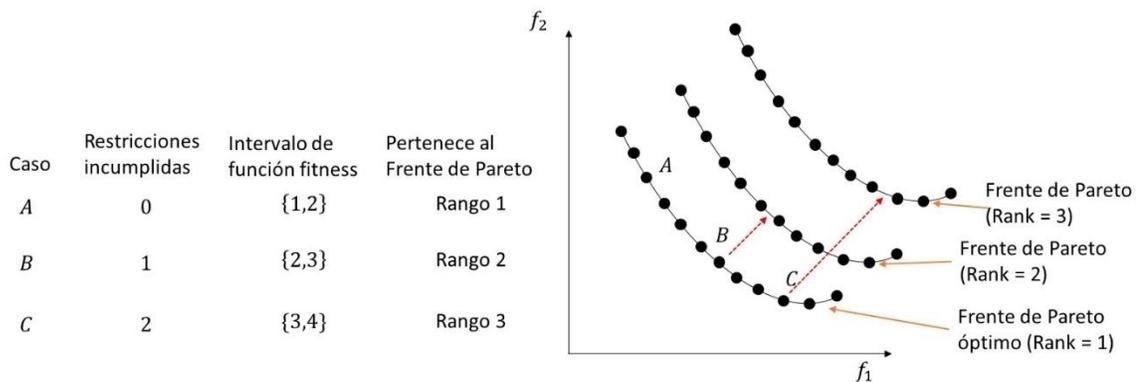
En cuanto al manejo de las restricciones, es necesario establecer una manera para asociar sobre la función de aptitud un efecto penalizador sobre los individuos, en caso de que no se cumplan una (o varias) restricciones del problema de optimización (económicas, operativas, ambientales, etc.). Para ello, tanto Serrato [1] como Rodríguez [2], [3] proponen una adición de un parámetro a la ecuación (2-24), que contemple las restricciones violadas.

Para la solución del problema de optimización multiobjetivo, el algoritmo MOHAEA aplica una visión orientada a la dominación de restricciones del algoritmo NSGA-II, haciendo que los individuos factibles cuenten con un mejor rango de no dominación que aquellos individuos que no son factibles. No obstante, se condiciona la participación de los individuos presentes en proximidades a la región de factibilidad en la generación de descendientes. Por tal razón, Rodríguez [2], [3] modifica MOHAEA para introducir una visión que permita que cualquier individuo invariable tenga un peor rango de no dominación respecto a los individuos factibles del mismo rango de Pareto. De esta manera, el rendimiento de un individuo se ve reducido en un frente de Pareto por cada restricción violada, implicando un incremento unitario al valor de aptitud, de acuerdo con la ecuación (2-26) en donde $W_C(i)$ es un valor de ponderación de las restricciones C , definido previamente por defecto como 1.0, lo cual corresponde a un rango en el valor de aptitud. Por ende, el valor de $W_C(i)$ puede tomar valores enteros, en función de si se viola o no la restricción C , tal como presenta la ecuación (2-27). El peso asociado a cada restricción es asignado por el usuario, por única vez al inicio del algoritmo, de acuerdo con que tan estricta deba ser una restricción y cuanto deba afectar el rango una cierta restricción, haciendo que una restricción de poca importancia pueda tener una penalización de 1.0 en un rango ($W_C = 1.0$) mientras que otra restricción de gran importancia tenga efecto en dos o más rangos de Pareto ($W_C = 2.0$ o más). La **Figura 2-20** presenta de manera ilustrativa el efecto del incumplimiento de una restricción para dos soluciones factibles presentes en el frente de Pareto (rango 1), empleando el manejo de restricciones de Rodríguez [2], [3].

$$fitness(i) = RND(i) + DH(i) + \sum_{C=1}^C W_C(i) \quad (2-26)$$

$$W_C(i) = \begin{cases} 0, & \text{si el individuo } i \text{ cumple la restricción } c \\ W_C, & \text{si el individuo } i \text{ no cumple la restricción } c \end{cases} \quad (2-27)$$

Para el presente trabajo, se adiciona al manejo de restricciones de Rodríguez [2], [3] el efecto de incumplimiento de las restricciones a lo largo de las generaciones, haciéndolo variable en función de la cantidad de individuos que no cumplan una c restricción en cada iteración. De esta manera, el algoritmo MOHAEA cuenta con una forma de sintonizar la penalización de las restricciones de manera similar a como se ajustan los operadores genéticos, mediante el parámetro δ , tal como se menciona en la Sección 2.2.11. Esta modificación de la estrategia de manejo de restricciones permite ajustar el efecto negativo creado por la generación de individuos que tiendan a fallar la misma restricción definida. El criterio para establecer si el parámetro W_c debe ser actualizado, mediante un incremento unitario, consiste en una tasa de individuos que violan la restricción superior al 50%. En caso contrario, el valor del parámetro W_c descenderá unitariamente, hasta alcanzar un mínimo de 0.



Manejo de restricciones

Según Rodríguez

1. Fijar de $W_c = 1,0$ para cada restricción
2. Mantener W_c constante durante todas las generaciones

Este estudio

1. Establecimiento de $W_c = 1,0$ para cada restricción
2. Conteo de restricciones incumplidas en el frente óptimo
 - Si $W_c(i) > \frac{N_{ind}}{2} \rightarrow W_c + 1$
 - Si $W_c(i) \leq \frac{N_{ind}}{2} \rightarrow W_c - 1$
3. Actualización de W_c al finalizar cada generación

Figura 2-20: Manejo de restricciones no cumplidas, de acuerdo con la estrategia propuesta por Rodríguez [2], [3] y variante propuesta en el presente trabajo.

2.2.14 Representación de resultados mediante frentes de Pareto

Con el fin de analizar las respuestas obtenidas por el modelo de optimización multiobjetivo, se construyeron una serie de perfiles que permiten analizar la ubicación de un conjunto de soluciones factibles en el espacio. Estas soluciones están regidas por la definición de dominancia y proximidad suministrada en la Sección 2.2.11. La **Figura 2-21** presenta a modo de ejemplo dos perfiles de frentes de Pareto para 2 y 3 funciones objetivo, empleando el algoritmo MOHAEA, en donde cada objetivo se encuentra vinculado a un eje del diagrama. La evaluación del modelo de optimización multiobjetivo del presente trabajo para el modelo CAMD estará sujeta exclusivamente a la formulación de 2 o 3 funciones objetivo.

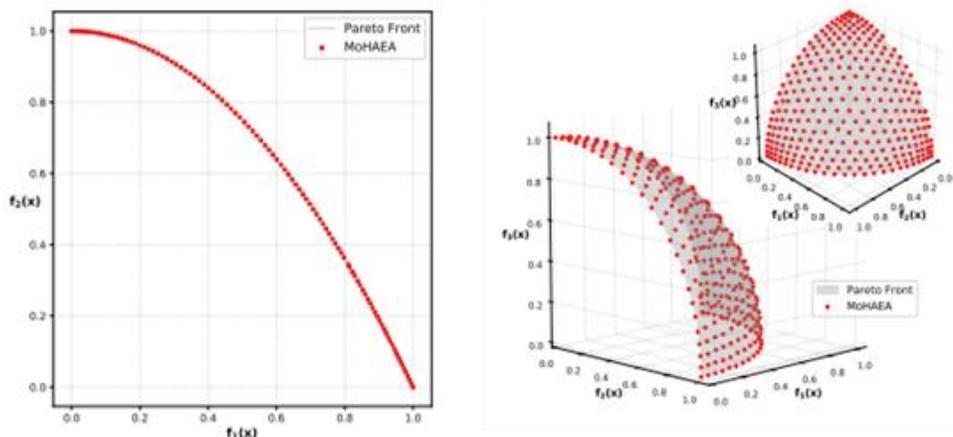


Figura 2-21: Frentes de Pareto para un problema de optimización de 2 y 3 objetivos, aplicando MOHAEA en funciones de prueba: (a) ZDT2, (b) DTLZ4. [246]

2.3 Diseño: Desarrollo de la herramienta CAMD

Posterior a la formulación del problema de optimización, descrito en la Sección 2.2, se realizó la implementación computacional para resolver el problema de selección de solventes de extracción líquido-líquido, acoplado numéricamente todas las partes antes detalladas. Para ello, fue necesario acoplar los modelos de construcción molecular descritos en las Secciones 2.2.2 a 2.2.6, los modelos de estimación de propiedades definidos en las Secciones 2.2.7 a 2.2.8, las bases de datos descritas en la Sección 2.2.9 y el algoritmo de optimización multiobjetivo MOHAEA, descrito en la Sección 2.2.11. Adicionalmente, se integraron los operadores genéticos y los mecanismos de selección

asociados al algoritmo evolutivo, descritos por las Secciones 2.2.12 y 2.2.13, respectivamente.

Con el fin de establecer una secuencia descriptiva del desarrollo de la herramienta CAMD, la Sección 2.3.1 menciona las consideraciones asociadas a las modificaciones del programa, partiendo de la línea base dada por los códigos fuente de Serrato [1] y Rodríguez [2], [3]. Por su parte, en la Sección 2.3.2 se mencionan las condiciones de operación de los experimentos (corridas) realizados en el estudio de casos de estudio. Finalmente, en la Sección 2.3.3 se describen los criterios de elección de solventes para validación experimental, a partir de la lista de candidatos obtenidos por la ejecución de las corridas experimentales.

2.3.1 Construcción de la herramienta CAMD

Se parte como línea base de los códigos fuente de Serrato [1] y Rodríguez [2], [3], con previa autorización de los autores. Ambos modelos se encuentran desarrollados en lenguaje Java y cuentan con una interfaz gráfica (GUI), presentada en la **Figura 2-22**, para interacción del usuario con el programa, sin necesidad de manipular el código directamente. No obstante, el enfoque actual en el área de la programación considera la simplificación del código, su legibilidad y su capacidad de procesamiento con mayor eficiencia, junto con una interfaz interactiva más amigable con el usuario y una mayor facilidad de reporte de resultados. Por tanto, se consideró necesario reescribir el algoritmo CAMD a lenguaje de programación Python, incorporando una nueva interfaz gráfica en dicho lenguaje, presentada en la **Figura 2-23**. A esta nueva interfaz se le asignó un nombre: *Multi-Objective Resolution Interface Toted Up with Aromatics* (MORITA por sus siglas en inglés, Interfaz de resolución multiobjetivo equipada con aromáticos)⁴¹. El esquema del algoritmo empleado por el programa se describe mediante el flujograma de la **Figura 2-24**. El **Anexo E** presenta una descripción detallada del programa mediante su manual de operación y recopilación de resultados para un caso de ejemplo.

⁴¹ El nombre del programa surgió mediante creación de anagramas de las palabras clave. Se consideró no integrar el término CAMD para evitar enmascaramiento con otros programas de características similares. Además, la analogía de moléculas y átomos respecto a la fruta mora, junto con la gran variedad de especies existentes, facilitó el desarrollo del concepto gráfico del programa.

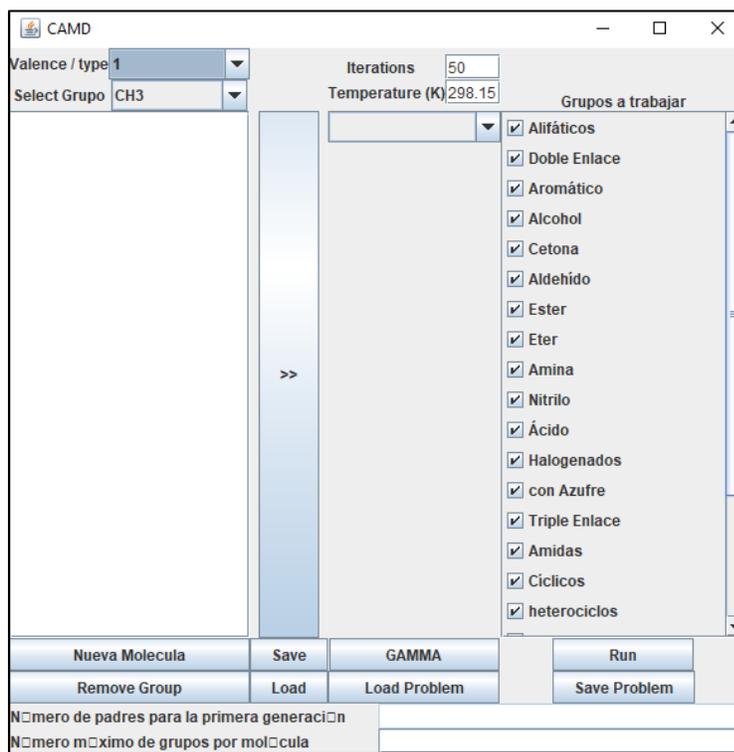


Figura 2-22: Interfaz gráfica de usuario (GUI) base, propuesta por Serrato [1] y Rodríguez [2], [3].

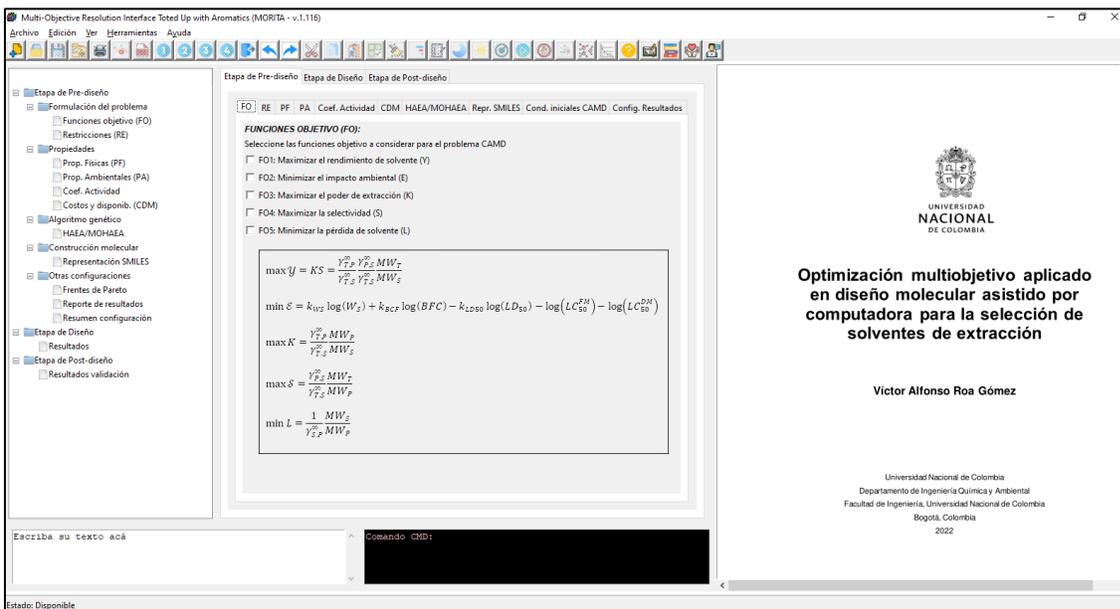


Figura 2-23: Interfaz gráfica de usuario (GUI) del programa CAMD desarrollado en el presente trabajo.

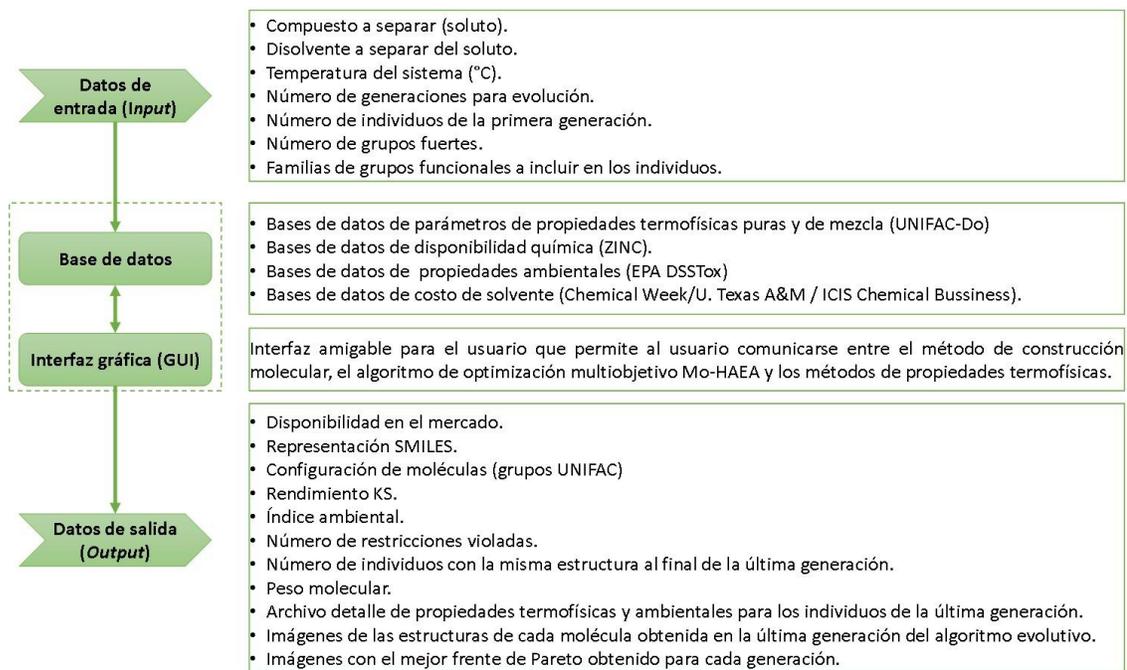


Figura 2-24: Esquema del algoritmo empleado en el programa CAMD.

Para la incorporación de los datos de las moléculas, se contemplan las mismas reglas propuestas en el modelo de Rodríguez [2], [3], descritas a continuación:

- Los átomos de la molécula no se introducen directamente. Deben introducirse los subgrupos de UNIFAC-DMD tal como los define DDBST GmbH [248] [249]. Por ejemplo, la representación del agua es '16', ya que este subgrupo UNIFAC-DMD corresponde para este compuesto.
- Un punto '.' representa el enlace simple que une dos subgrupos. Los enlaces dobles y triples deben representarse como los subgrupos UNIFAC-DMD definidos para ello. A modo de ejemplo, la representación del etanol es '1.2.14', donde los números representan los grupos CH3, CH2 y OH, respectivamente. El propileno se representará como '5.1', siendo el 5 el número para el subgrupo CH2=CH.
- Para subgrupos ramificados, una subestructura entre paréntesis indica que corresponde a una rama que emerge del subgrupo anterior. En el caso del isobutanol, la configuración corresponderá a '1.3(1).2.14', siendo la subestructura (1) es una rama que emerge del subgrupo 3.

Entre las modificaciones estructurales realizadas al programa, respecto a la línea base suministrada por el modelo de Rodríguez [2], [3], se integraron los siguientes aspectos:

- Establecimiento de una interfaz más amigable para la ejecución del problema CAMD, empleando las librerías *tkinter*.
- Manejo simplificado del código, mediante bibliotecas matemáticas específicas de ciencias de datos, como *NumPy* y *Pandas*.
- Integración de una ventana para presentación de resultados de manera gráfica, usando las librerías *matplotlib* y *PIL*.
- Aplicación de la representación SMILES canónica, con facilidad de manejo de moléculas aromáticas y heterocíclicas (verificación de estructuras factibles, kekulización, identificación de aromaticidad de una molécula particular, entre otros aspectos) utilizando la librería *Rdkit*, común en bioinformática.
- Visualización de resultados de las estructuras en formato PDF mediante un reporte de las características de las moléculas encontradas en el diseño experimental realizado, mediante la incorporación de la librería *tkPDFViewer*.
- Versatilidad de exportación de las moléculas candidatos obtenidas, mediante archivos MOL (.mol), mediante la librería *OechemTk*.
- Manipulación de los directores del sistema operativo para abrir, guardar y almacenar información, junto con la impresión de resultados en un reporte estándar mediante las librerías *os* y *win32api*.
- Integración de las bases de datos empleadas en el estudio mediante vínculo con MongoDB y facilidad de actualización de este.
- Integración de una terminal CMD intrínseca al programa, que permita la ejecución de este mediante comando en Python, mediante la librería *terminal*. De esta manera será más fácil actualizar las bases de datos presentes en el programa para futuros estudios. Dichos comandos se describen en el **Anexo E**.
- Asociación de manual de ayuda y guías de usuario para la fácil comprensión y ejecución del programa.
- El programa fue desarrollado siguiendo el paradigma de la programación orientada a objetos (POO), con documentación de clases y métodos en el código fuente.

2.3.2 Condiciones operacionales de las corridas realizadas

Para la evaluación de los resultados y su posterior comparación con los modelos de Serrato [1] y Rodríguez [2], [3], se realizaron 50 experimentos (es decir, corridas del algoritmo) para analizar y comparar la reproducibilidad de las moléculas obtenidas como posibles solventes de extracción líquido-líquido, para dos sistemas binarios específicos: ácido láctico-agua (LA/W) y ácido acético-agua (AA/W). Las corridas del algoritmo CAMD, emplearon las condiciones iniciales descritas en la **Tabla 2-13**, para cada uno de los sistemas antes mencionados. De esta manera es posible identificar el comportamiento de los criterios de desempeño del algoritmo multiobjetivo, sin influencia de la aleatoriedad.

Tabla 2-13: Condiciones iniciales empleadas en el presente trabajo.

Condición inicial	Sistema LA/W	Sistema AA/W
Número de ejecuciones (Experimentos).	1 a 50	51 a 100
Compuesto a separar (soluto).	Ácido láctico	Ácido acético
Disolvente del sistema.	Agua	Agua
Temperatura del sistema (K).	298	298
Número de generaciones.	50	50
Población inicial, (Np).	50	50
Número máximo de grupos fuertes por molécula.	3	3
Número máximo de grupos funcional por molécula.	10	10

2.3.3 Elección de solventes obtenidos en la ejecución de la herramienta CAMD para validación experimental

Posterior a la implementación computacional del modelo CAMD y la obtención de un conjunto de moléculas candidatas producto de múltiples evaluaciones de la herramienta, se realizó una clasificación de sustancias de acuerdo con los criterios descritos en la **Tabla 2-14**, con el fin de determinar cuáles moléculas serían empleadas para la validación experimental del modelo multiobjetivo. El desarrollo de la validación experimental se encuentra enmarcado en la etapa del Post-diseño, contando con una metodología descrita con detalle en la Sección 2.4.

Tabla 2-14: Criterios de elección de solvente para validación experimental.

Criterio	Especificación
Criterio 1	Disponibilidad de la sustancia candidato en el mercado: Las sustancias candidato deben encontrarse en el mercado ⁴² con fácil disponibilidad, a altas concentraciones y a las cantidades requeridas para realizar la serie de experimentos.
Criterio 2	Costo de la sustancia candidata en el mercado ⁴³ : El precio de la sustancia seleccionada debe ser bajo, con el fin de facilitar su reposición.
Criterio 3	Estabilidad estructural de la molécula candidato. La sustancia seleccionada debe ser estable estructuralmente, no debe degradarse producto de la temperatura, fuentes de luz u otras condiciones de operación. Esto facilitará el almacenamiento de la sustancia durante el periodo que se requiera en la validación experimental.
Criterio 4	Presencia de la sustancia entre los posibles candidatos del algoritmo de Serrato [1], con fines comparativos respecto a este modelo.
Criterio 5	Presencia de la sustancia entre los posibles candidatos del algoritmo de Rodríguez [2], con fines comparativos respecto a este modelo.
Criterio 6	No reactividad con los componentes presentes en el sistema. La sustancia seleccionada para realizar la validación experimental no debe interferir en el equilibrio de fases producto de interacciones con alguno de los componentes de la mezcla binaria del caso de estudio ⁴⁴ .

⁴² La disponibilidad de la sustancia, descrita en el Criterio 1, se enmarca en el mercado colombiano, con el fin de asegurar la fácil obtención de las sustancias a nivel local o regional. No obstante, cabe aclarar que la disponibilidad descrita en el modelo CAMD puede variar respecto a esta, dado que la base de datos ZINC se encuentra construida para el mercado estadounidense.

⁴³ El costo de la sustancia, descrito en el Criterio 2, se enfoca al mercado colombiano, con el fin de reducir los gastos asociados al proyecto en aspectos experimentales. No obstante, los costos asociados a los compuestos en el modelo CAMD, dependen del mercado estadounidense.

⁴⁴ Para evaluar la no reactividad de componente, se empleará como indicador de reacción los cambios de pH en el sistema ácido/agua dados por la adición del solvente, reportados en literatura para los solventes evaluados. Adicionalmente, se tendrá en cuenta si la literatura reporta precipitaciones, efervescencia o cambios de color de la mezcla, producto de la adición del solvente.

2.4 Post-diseño: Validación experimental de moléculas

La validación del modelo construido en la herramienta CAMD busca verificar de manera experimental que los compuestos obtenidos por el modelo computacional como potenciales solventes de extracción, cuentan con un comportamiento real apropiado en casos de estudios definidos, bajo condiciones controladas. Para ello, es necesario considerar el balance de materia de la extracción líquido-líquido, la selección de métodos analíticos para determinación de concentraciones y la evaluación de indicadores experimentales.

En la extracción líquido-líquido, existen tres componentes que coexisten durante la separación: agua, ácido y solvente. Las ecuaciones (2-28) a (2-30) se obtienen mediante la aplicación del balance de materia en las condiciones iniciales y finales de la operación.

$$BM \text{ del ácido: } V_0^{acid} C_0^{acid} = V_{aq,f}^{acid} C_{aq,f}^{acid} + V_{org,f}^{acid} C_{org,f}^{acid} \quad (2-28)$$

$$BM \text{ del agua: } V_0^w C_0^w = V_{aq,f}^w C_{aq,f}^w + V_{org,f}^w C_{org,f}^w \quad (2-29)$$

$$BM \text{ del solvente: } V_0^{Solv} C_0^{Solv} = V_{aq,f}^{Solv} C_{aq,f}^{Solv} + V_{org,f}^{Solv} C_{org,f}^{Solv} \quad (2-30)$$

Donde:

aq y *org* son la fase acuosa y orgánica, respectivamente.

C_0 y C_f son las concentraciones inicial y final, respectivamente.

V_0 y V_f son los volúmenes inicial y final, respectivamente.

Dado que las condiciones iniciales de volumen y concentración (indicadas con color azul) son conocidas a partir de la definición del experimento, es necesario establecer métodos de cuantificación de concentración y volumen, una vez alcanzado el equilibrio termodinámico. Para la medición de los volúmenes (iniciales y finales), se emplearon elementos de medición disponibles en laboratorio (material volumétrico). Por su parte, la cuantificación de concentraciones de ácido y agua se realizó mediante los métodos analíticos descritos por las Secciones 2.4.1 a 2.4.3.

Para la verificación experimental de los resultados obtenidos por el modelo CAMD, se realizó un montaje extracción líquido-líquido de etapa simple, para los dos casos de estudio

considerados: ácido láctico/agua (LA/W) y ácido acético/agua (AA/W). Para todos los experimentos de extracción se estableció una relación entre ácido/agua de 90 g ácido/ L solución, valor dentro de los rangos típicos de concentraciones de efluentes de fermentadores [1]. Para cada uno de los casos de extracción, se evaluaron experimentalmente tres moléculas candidato elegidas bajo los criterios de la Sección 2.3.3.

Se realizaron ensayos triplicados para cada una de las extracciones en cuestión, para cada escenario de solvente y caso de estudio. También se utilizaron triplicados durante la aplicación de los métodos analíticos. Se aseguró que el sistema ternario ácido-agua-solvente alcanzara el equilibrio de fases previo a la determinación de las condiciones finales. Entre la información reportada al inicio del experimento se encuentra la cantidad de solvente adicionada, junto con los volúmenes de la fase orgánica y acuosa obtenidos luego de hecha la separación de las fases vía decantación. El **Anexo F** presenta el registro fotográfico de las extracciones líquido-líquido realizadas experimentalmente.

2.4.1 Cuantificación de ácido final en fase acuosa

La cuantificación de la concentración final de ácido en la fase acuosa ($C_{aq,f}^{acid}$) se realizó por medio de cromatografía líquida de alta eficiencia (HPLC, por sus siglas en inglés). El equipo utilizado fue un Thermo Scientific Ultimate 3000, con un detector IR; una columna KC-811 Shodex, con agua desionizada con 0.1% de H_3PO_4 como fase móvil, el flujo de trabajo fue de 1.0 mL/min, temperatura del horno de 40°C. El software de procesamiento para la integración de las muestras fue Chromeleon V.7.0 ®.

Para la aplicación de este método analítico es necesario previamente establecer las curvas de calibración para cada uno de los sistemas considerados (LA/W y AA/W). La descripción detallada de la construcción de las curvas para el método HPLC, junto con los cromatogramas obtenidos para ambos sistemas, se describe en el **Anexo G**.

2.4.2 Cuantificación de agua final en fase orgánica

La cuantificación de la concentración final del agua en la fase orgánica ($C_{org,f}^w$) se obtiene por medio del método de titulación potenciométrica de Karl Fischer, empleando el equipo disponible en el Laboratorio de Combustibles y Lubricantes del Laboratorio de Ingeniería

Química (LIQ), de la Universidad Nacional de Colombia. El montaje consta de un titulador automático Mettler Toledo T50, el cual cumple con los requisitos de la norma técnica ASTM-D4377-00, junto con una balanza analítica Sartorius CPA225D (capacidad de pesaje 40 – 220 g, repetibilidad de 0.02 - 0.1 mg y linealidad de 0.03 - 0,2 mg).

Para la aplicación de este método analítico es necesario realizar previamente la estandarización del reactivo de Karl Fischer (libre de piridina, grado analítico) empleado en la titulación, usando como estándar agua o tartrato ácido de sodio. La descripción detallada del método de Karl Fischer efectuado se presenta en el **Anexo H**.

2.4.3 Cuantificación de ácido en fase orgánica y de agua en fase acuosa

Para la determinación de la concentración final de ácido en la fase orgánica ($C_{org,f}^{acid}$) se empleó una titulación ácido-base, dependiendo del ácido empleado en el sistema. Por su parte, De manera similar, la concentración final de agua en la fase acuosa ($C_{aq,f}^W$) fue calculada mediante el mismo método experimental. El registro fotográfico y las condiciones experimentales usadas en las volumetrías ácido-base se presentan en el **Anexo H**.

2.4.4 Comparación entre modelo teórico y resultados experimentales

Una vez alcanzado al equilibrio líquido-líquido y evaluadas las concentraciones finales de ácido y agua para cada una de las fases, se determina el coeficiente de distribución experimental (K^{exp} , descrita por la ecuación (2-31)) para determinar cuánto se obtuvo de ácido en la fase orgánica como resultado de la extracción. También, se determina la selectividad experimental (S^{exp} , descrita por la ecuación (2-32)), la cual permite identificar cuánta agua se disolvió en el solvente. Con estos parámetros experimentales, se realiza la comparación respectiva con los resultados de los valores de K y S , obtenidos por el algoritmo de optimización multiobjetivo.

$$K^{exp} = \frac{C_{org,f}^{acid}}{C_{aq,f}^{acid}} \quad (2-31)$$

$$S^{exp} = \frac{C_{org,f}^W}{C_{aq,f}^W} \quad (2-32)$$

3. Resultados y discusión

Este capítulo presenta los resultados obtenidos en el desarrollo de la presente tesis, segmentada en las tres etapas del diseño CAMD (Prediseño, diseño y Post-diseño), enfocada en los cuatro objetivos establecidos.

3.1 Resultados asociados a la etapa de prediseño

La finalidad de la presente sección es presentar y analizar los resultados asociados a la actualización del algoritmo de construcción molecular y la evaluación de impactos de propiedades sobre las funciones objetivo, mediante la aplicación de modelos de estimación de propiedades. Esto con el fin de identificar diferencias apreciables entre el presente modelo y los estudios de línea base de Serrato [1] y Rodríguez [2], [3]. Los resultados obtenidos en esta sección se enmarcan en los objetivos 1 y 2 del presente trabajo.

Con respecto al algoritmo de construcción molecular, la Sección 3.1.1 realiza un mayor énfasis en la construcción de aromáticos y compuestos heterocíclicos, estableciendo unos nuevos grupos aplicables al modelo CAMD. Se aplica la metodología de tres filtros, descrita en la Sección 2.2.3 y 2.2.4, para compuestos monocíclicos y policíclicos, respectivamente. Adicionalmente, la Sección 3.1.2 determina el impacto del carácter aromático en la construcción de nuevas estructuras moleculares, mientras que la Sección 3.1.3 evalúa que tan aplicables son estos nuevos grupos en moléculas típicas de uso industrial.

Con respecto a los modelos de estimación de propiedades, la Sección 3.1.4 presenta el análisis estadístico multivariable de la influencia de las propiedades definidas en las Secciones 2.2.7 y 2.2.8, con respecto a las funciones objetivos establecidas en la Sección 2.2.1. Se presentan los indicadores estadísticos asociados a la correlación multivariable, descritos en la Sección 2.2.10, buscando establecer si existe efectos similares en compuestos de una misma familia química o si son efectos intrínsecos a cada molécula.

3.1.1 Grupos heterocíclicos encontrados producto de la actualización del algoritmo de construcción molecular

La determinación de grupos heterocíclicos fue desarrollada para grupos monocíclicos y policíclicos mediante la metodología de tres filtros descrita en la Sección 2.2.3 y 2.2.4, respectivamente. A continuación, se describen los resultados obtenidos para cada filtro.

Compuestos monocíclicos - Filtro I:

Inicialmente, se identificó el conjunto universal (P_U) de estructuras monocíclicas generalizables que se pueden construir para ciclos con una cantidad menor a 6 átomos (repetidas o no). Esto se realiza mediante permutación, siguiendo la expresión 3^n , donde n corresponde al número de átomos del ciclo evaluado. De esta manera, se determinaron 1.080 estructuras iniciales divididas en: 27 para ciclos de 3 átomos, 81 para ciclos de 4 átomos, 243 para ciclos de 5 átomos y 729 para ciclos de 6 átomos.

Con el fin de identificar las estructuras no isomorfas, se evaluaron los ejes de simetría y grupos diedrales de los ciclos, descritos por **Figura 2-7**. Por su parte, el **Anexo B** presenta las tablas de Cayley para las rotaciones equivalentes, a partir de los movimientos de la **Figura 2-7**. La **Tabla 3-1** presenta los resultados de aplicar el teorema de enumeración de Pólya (ecuación (2-12)) en los ciclos para la determinación del índice de ciclo $Z(x_1, x_2, \dots, x_n)$ y la función generadora de $p(X)$ para ciclos de 3, 4, 5 y 6 átomos. Cabe resaltar que se verificó el cumplimiento del teorema de enumeración de Pólya mediante conteo manual y clasificación de las estructuras según similitudes asociadas a simetría.

Tabla 3-1: Aplicación del teorema de enumeración de Pólya a estructuras monocíclicas n-atómicas para determinación de estructuras no isomorfas (ENI).

Ciclo	Grupo diedral (D_n)	$Z(D_n)$	$p(X)$	ENI	P_U	%
(MC3) $n = 3$	$\left\{ \begin{matrix} I, R_{120}, R_{240} \\ R_a, R_b, R_c \end{matrix} \right\}$	$\frac{1}{6}(X_1^3 + 2X_3 + 3X_1X_2)$	$\frac{1}{6}(X^3 + 2X + 3X^2)$	10	27	37,04%
(MC4) $n = 4$	$\left\{ \begin{matrix} I, R_{90}, R_{180} \\ R_{270}, R_a, R_b, \\ R_c, R_d \end{matrix} \right\}$	$\frac{1}{8}(X_1^4 + 2X_4 + 3X_2^2 + 2X_1^2X_2)$	$\frac{1}{8}(X^4 + 3X + 2X^2 + 2X^3)$	21	81	25,96%
(MC5) $n = 5$	$\left\{ \begin{matrix} I, R_{72}, R_{144} \\ R_{216}, R_{288}, R_a, \\ R_b, R_c, R_d, R_e \end{matrix} \right\}$	$\frac{1}{10}(X_1^5 + 4X_5 + 5X_1X_2^2)$	$\frac{1}{10}(X^5 + 4X + 5X^3)$	39	243	16,05%
(MC6) $n = 6$	$\left\{ \begin{matrix} I, R_{60}, R_{120} \\ R_{180}, R_{240}, R_{300} \\ R_a, R_b, R_c, R_d, R_e \end{matrix} \right\}$	$\frac{X_1^6 + 2X_6 + 3X_1^2X_2^2 + X_2^3 + 2X_3^2}{12}$	$\frac{X^6 + 5X + 3X^3 + 3X^2}{12}$	92	729	12,62%
Total Estructuras Monocíclicas				162	1.080	15,00%

En total, la aplicación del Filtro I permitió identificar 162 estructuras no isomorfas (15,00% del total) para compuestos monocíclicos con menos de 6 átomos, reportadas en la **Figura 3-1**. El átomo genérico R es asignable como carbono o un heteroátomo, en función de la valencia del átomo sustituyente y del cumplimiento de la regla del Octeto. Estas estructuras generalizadas iniciales se toman como punto de partida para la determinación de grupos de los Filtros II y III.

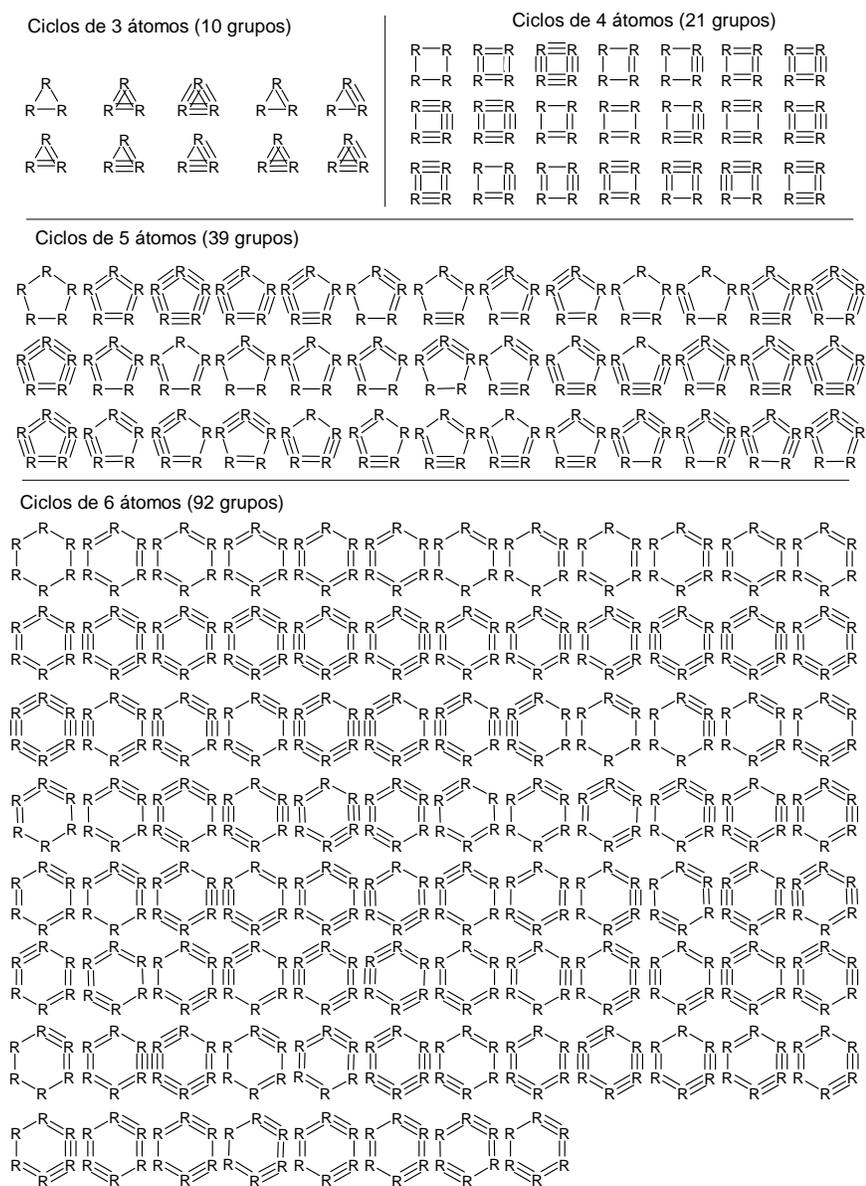


Figura 3-1: Configuraciones estructurales generalizadas para compuestos monocíclicos de 3 a 6 átomos, obtenidas del Filtro I.

Compuestos monocíclicos - Filtro II:

A partir de las 162 estructuras no isomórficas (*ENI*) que superaron el Filtro I, se realizó una permutación con repetición (*P*) para determinar los posibles átomos asignables para una molécula cíclica conformada por *n*-átomos y, por ende, la cantidad de sustituciones de heteroátomos posibles para una cierta estructura cíclica *n*-atómica, considerando las valencias válidas del átomo (*u* átomos) a reemplazar. De forma similar a realizado en el conjunto universal, el valor de *P* puede ser calculado a partir de la expresión 4^n , donde *n* representa la cantidad de átomos de la estructura cíclica.

De esta manera, el producto $ENI \times P$ representa la cantidad de combinaciones totales cuya validez de valencia debe ser analizada, limitada exclusivamente para estructuras no isomorfas y para una combinación de átomos establecida (C, N, O, S). Teniendo en cuenta lo anterior, el código script en Python del Filtro II evalúa la valencia y clasifica las estructuras entre válidos y no válidos, para cada grupo de estructuras no isomorfas. La **Tabla 3-2** presenta estos resultados para compuestos monocíclicos conformados por 3 a 6 átomos, obtenidos a partir sustituciones sencillas o múltiples en las estructuras cíclicas no isomorfas obtenidas en el Filtro I. Fue necesario revisar el isomorfismo de las estructuras una vez aplicado el Filtro II, dado que, al hacer la modificación de los átomos, algunos de las estructuras se repetían.

Tabla 3-2: Grupos válidos para compuestos monocíclicos *n*-atómicos considerando valencias atómicas, obtenidos mediante sustituciones de heteroátomos (Filtro II).

Ciclo	<i>P</i>	<i>ENI</i>	$ENI \times P$	¿Superan Filtro II? (valencia atómica)		¿Superan Filtro II y son no isomorfas? (%)*	Repetición respecto a estructuras válidas (%)
				SI (%)*	NO (%)*		
MC3 (n = 3)	64	10	640	142 (22,19%)	498 (77,81%)	72 (11,25%)	49,30%
MC4 (n = 4)	256	21	5.376	721 (13,41%)	4.655 (86,59%)	299 (5,56%)	58,53%
MC5 (n = 5)	1.024	39	39.936	3.158 (7,91%)	36.778 (92,09%)	1.378 (3,45%)	56,36%
MC6 (n = 6)	4.096	92	376.832	17.451 (4,63%)	359.381 (95,37%)	6.846 (1,82%)	60,77%
Sustituciones válidas	5.440	162	422.784	21.472 (48,14%)	401.312 (51,86%)	8.595 (2,03%)	59,97%

* Porcentaje respecto al producto $ENI \times P$.

De esta manera se determinó que solamente 8.595 estructuras no isomorfas monocíclicas cumplen con los criterios de valencia establecidos por el modelo CAMD, correspondientes al 2,03% del total del producto $ENI \times P$. Esta proporción de estructuras válidas no es uniforme respecto a la cantidad de átomos que conforman el ciclo, ya que la proporción respecto a tamaño del ciclo va disminuyendo a medida que el ciclo es más grande. No obstante, la tasa de aceptación respecto al ciclo más pequeño evaluado (MC3) es 11,25%, lo cual indica que gran parte de las estructuras no cumplen con los criterios de valencia propuestos para el modelo CAMD (carga cero dentro de las estructuras).

Con respecto a los tiempos de evaluación del Filtro II, se observó un comportamiento rápido del Filtro II, obteniéndose las estructuras válidas en ordenes de segundos a menos de 5 minutos (caso MC6). Por tal razón, y considerando de la cantidad de estructuras a evaluar, dadas por el producto $ENI \times P$, es posible mencionar que el algoritmo es eficiente para la determinación de las estructuras válidas. Por otra parte, respecto a la tasa de almacenamiento de resultados, se observa que las estructuras son almacenables en ficheros de bajo peso, menores a 5Mb. A pesar de que el número de evaluaciones realizada aumenta conforme el número de átomos en el ciclo incrementa, fue posible optimizar el código de evaluación, con el fin de reducir los tiempos de evaluación y procesamiento de resultados, junto con un consumo mínimo de recursos de disco duro.

El Filtro II desarrollado permite explorar de forma eficiente todas las combinaciones de estructuras heterocíclicas posibles, partiendo de las consideraciones estructurales establecidas. Las 8.595 estructuras obtenidas serán remitidas al Filtro III, en donde se podrá cuantificar cuantas estructuras son representables por los modelos de estimación de propiedades.

Compuestos monocíclicos - Filtro III:

Para las 8.595 estructuras que superaron el Filtro II, se evaluó la aplicabilidad de propiedades mediante los modelos de estimación de propiedades descritos en la Sección 2.2.7 y 2.2.8. La **Tabla 3-3** presenta la aplicabilidad observada de las estructuras monocíclicas evaluadas con respecto a cada modelo de estimación de propiedades.

Tabla 3-3: Aplicabilidad de los modelos de estimación de propiedades para las estructuras evaluadas de compuestos monocíclicos en el Filtro III.

Tamaño del compuesto monocíclico	Número de estructuras evaluadas por modelos de estimación de propiedades				
	Por ningún modelo	Solo con modelo de Hukkerikar	Solo por el modelo UNIFAC-DMD	Por ambos modelos	Total
n = 3	8 (11,11%)	26 (36,11%)	16 (22,22%)	22 (30,56%)	72
n = 4	32 (10,70%)	134(44,82%)	78 (26,09%)	55 (18,39%)	299
n = 5	101 (7,33%)	710 (51,52%)	382 (27,72%)	185 (13,43%)	1.378
n = 6	310 (4,52%)	3.951 (57,71%)	1946 (28,43%)	639 (9,33%)	6.846
Total	456 (5,31%)	4.816 (56,03%)	2.422 (28,18%)	901 (10,48%)	8.595

De esta manera, se obtuvieron 901 estructuras que son aplicables con ambos modelos, lo cual representa un 4,196% del total de moléculas que superaron el Filtro II y un 0,213% de todas las moléculas consideradas por el producto $ENI \times P$ construido a partir de las todas las estructuras generalizadas establecidas en el Filtro I. Estas estructuras moleculares encontradas son aplicables en el diseño CAMD, dado que pueden ser empleadas tanto para estimación de las propiedades (requeridas en la determinación de las funciones objetivo y la evaluación de restricciones) mediante el modelo de Hukkerikar, como para la determinación de los coeficientes de actividad mediante el modelo UNIFAC-DMD.

La baja cantidad de estructuras que son evaluables con los dos modelos antes descritos se debe principalmente a tres factores: 1.) la gran complejidad de determinación experimental de parámetros experimentales para ciertos tipos de enlaces dentro del anillo heterocíclico, 2.) la variabilidad de los comportamientos de las propiedades de los enlaces dentro del anillo heterocíclico, a medida que este incrementa su tamaño, y 3.) la dificultad de síntesis de estructuras orgánicas heterocíclicas específicas, a medida de que incrementa el tamaño del ciclo orgánico.

Considerando que tanto los parámetros del modelo de Hukkerikar como el modelo UNIFAC-DMD se encuentran en constante actualización, se espera a futuro una mayor cantidad de estructuras monocíclicas sean representables por ambos modelos, considerando las limitantes estructurales definidas establecidos previamente.

Compuestos monocíclicos – Resumen y análisis:

La **Tabla 3-4** presenta un resumen del proceso de filtrado de grupos en compuestos monocíclicos sustituidos por heteroátomos. Debido a su extensión, la representación estructural de los grupos heteroatómicos monocíclicos válidos añadidos al modelo CAMD es remitida en medio digital, tal como se menciona en el **Anexo I**. Dichas moléculas corresponden al listado de estructuras que superaron los tres filtros establecidos por la metodología desarrollada.

Tabla 3-4: Resumen metodología de tres filtros: Compuestos monocíclicos.

Criterio de filtrado de estructuras		Compuestos monocíclicos n-atómicos				
		n = 3	n = 4	n = 5	n = 6	Total
Conjunto universal (P_U) de estructuras monocíclicas construibles, contadas mediante permutación 3^n		27	81	243	729	1.080
Filtro 1: ¿La configuración estructural está repetida por simetría con otra?	Si (%)	17 (62,96)	60 (74,07)	204 (83,95)	637 (87,38)	918
	No (%)	10 (37,04)	21 (25,93)	39 (16,05)	92 (12,62)	162
Estructuras no isomorfas (ENI) *		10	21	39	92	162
Permutaciones factibles (P) por sustitución de R-átomos**		64	256	1.024	4.096	5.440
Producto ENI x PNI		640	5.376	39.936	376.832	422.784
Filtro 2: ¿La configuración representa una estructura válida respecto a valencia atómica?	No (%)	498 (77,81)	4.655 (86,59)	36.778 (92,09)	359.381 (95,37)	401.312
	Si (%)	42 (22,19)	721 (13,41)	3.158 (7,91)	17.451 (4,63)	21.472
Estructuras válidas al Filtro II y no isomorfas, evaluadas en modelos de propiedades ***		72	299	1.378	6.846	8.595
Filtro 3: ¿Son válidos ambos modelos para determinación de propiedades? ****	Si (%)	22 (30,56)	55 (18,39)	185 (13,43)	639 (9,33)	901
	No (%)	50 (69,44)	244 (81,61)	1.193 (86,57)	6.207 (90,67)	7.694
Grupos válidos adicionados al modelo CAMD		22	55	185	639	901

* Representadas por las **Figura 3-1** y verificadas mediante el Teorema Enumerador de Pólya.

** Las sustituciones consideradas en el estudio pueden ser de uno o varios R-átomos por C y heteroátomos (N, O, S).

*** Se emplearon los modelos de Hukkerikar y UNIFAC-DMD para evaluar a las estructuras moleculares.

**** La validez de la estructura molecular propuesta radica en el cumplimiento de las valencias de los átomos involucrados (incluyendo la no presencia de cargas en la molécula).

A partir de los resultados obtenidos en la **Tabla 3-4**, es posible establecer que se ha construido una metodología para establecimiento de nuevo grupos funcionales, aplicables al problema de diseño CAMD. Estos grupos funcionales son válidos desde el punto de vista estructural y valencia, y además cuentan con propiedades obtenibles en los modelos

de estimación de propiedades empleados por el diseño CAMD. El modelo presentado cuenta con la ventaja de ser fácilmente extrapolable para compuestos monocíclicos heteroatómicos con tamaños de ciclo superiores a 6 átomos, siguiendo la metodología propuesta. No obstante, se identifica que existe una relación directa entre el tamaño del ciclo y la cantidad de grupos estructurales que son capaces de superar los tres filtros de la metodología, por los motivos antes mencionados. Por esta razón y, considerando que los compuestos ciclados de gran tamaño (compuestos con más de seis átomos dentro del anillo) no cuentan con un interés significativo de carácter práctico para realizar las extracciones de los sistemas LA/W y AA/W, es posible mencionar que el modelo de construcción molecular representa aceptablemente una proporción apropiada de los compuestos monocíclicos heteroatómicos de hasta 6 átomos.

Adicionalmente, la metodología de tres filtros desarrollada permite establecer una hoja de ruta para la determinación de parámetros experimentales futuros, que requieren identificarse por los modelos de estimación de propiedades utilizados por este estudio (UNIFAC-DMD y Hukkerikar). Hasta el momento, la determinación experimental de estos parámetros ha sido realizada en función de las necesidades encontradas por la comunidad científica en diversos estudios de estructuras orgánicas. No obstante, dado que la metodología realizada contempla el conjunto universal de estructuras válidas para un tamaño de ciclo y tipos de átomos válidos dentro del ciclo, es posible cuantificar cuales son los grupos faltantes con parámetros por identificar por los modelos de estimación de propiedades. Esto permitirá fortalecer dichos modelos, permitiendo su generalización para cualquier configuración de átomos heterocíclicos monoatómicos.

Finalmente, el establecimiento de las funciones generadoras dadas por el teorema de enumeración de Pólya permite cuantificar de forma rápida la cantidad de estructuras válidas geométricamente bajo los supuestos estructurales. Se deja para estudios posteriores la aplicación del teorema de enumeración de Pólya para ciclos de tamaños superiores a 6 átomos y un conjunto de compuestos mayor a los estudiados (C, N, O, S).

Compuestos policíclicos – Filtro I:

Siguiendo la estrategia anterior, se identificó el conjunto universal (P_U) de estructuras policíclicas generalizables que se pueden construir para compuestos policíclicos tipo espiro y fusionada (repetidas o no). Considerando que el tamaño de ciclo máximo presente

para estructuras policíclicas es de 6 átomos, se obtienen 10 estructuras tipo espiro y 10 estructuras fusionadas, construidas por dos ciclos conectados. Según los tipos de ciclos involucrados, cada estructura contará con una cantidad de átomos y enlaces variables, por lo que las permutaciones también cambiarán, tal como se presenta en la **Tabla 3-5**.

Tabla 3-5: Conjunto universal de estructuras construibles de compuestos policíclicos tipo espiro y fusionados, obtenidos mediante permutaciones.

Tipo	Nro. Átomos	Nro. Enlaces	Permutaciones Totales (3^n)
<i>E</i> - 1	4	5	729
<i>E</i> - 2	5	6	2.187
<i>E</i> - 3	6	7	6.561
<i>E</i> - 4	7	8	19.683
<i>E</i> - 5	6	7	6.561
<i>E</i> - 6	7	8	19.683
<i>E</i> - 7	8	9	59.049
<i>E</i> - 8	8	9	59.049
<i>E</i> - 9	9	10	177.147
<i>E</i> - 10	10	11	531.441
Total Policíclicos Espiro			882.090

Tipo	Nro. Átomos	Nro. Enlaces	Permutaciones Totales (3^n)
<i>F</i> - 1	4	5	243
<i>F</i> - 2	5	6	729
<i>F</i> - 3	6	7	2.187
<i>F</i> - 4	7	8	6.561
<i>F</i> - 5	6	7	2.187
<i>F</i> - 6	7	8	6.561
<i>F</i> - 7	8	9	19.683
<i>F</i> - 8	8	9	19.683
<i>F</i> - 9	9	10	59.049
<i>F</i> - 10	10	11	177.147
Total Policíclicos Fusionados			294.030

Una vez identificados los movimientos simétricos factibles, descritos por las **Figura 2-10** y **Figura 2-11**, junto con las tablas de Cayley presentadas en el **Anexo B** para las rotaciones equivalentes, se planteó la hipótesis de evaluar cada una de los tipos de estructuras siguiendo la estrategia proporcionada por el Teorema de Enumeración de Pólya, con el fin de evaluar las estructuras generalizadas de forma eficiente la cantidad de estructuras no isomorfas obtenibles para cada tipo de estructura.

A continuación, presentan las ecuaciones de índice de ciclo $Z(x_1, x_2, \dots, x_n)$ y función generadora $p(X)$ para estructuras policíclicas con configuración espiro y fusionadas, presentadas en la **Tabla 3-6**, siguiendo la misma metodología previamente aplicada en compuestos monocíclicos. Es necesario precisar que, tanto en el caso de estructuras espiro como fusionadas, se sustituye la definición de grupo diedral (D_n) por la de grupo

simétrico (S_n), dado que los primeros aplican exclusivamente para polígonos regulares, tal como es el caso de los grupos monocíclicos.

Tabla 3-6: Aplicación del teorema de enumeración de Pólya a estructuras policíclicas de tipo espiro y fusionadas para determinación de estructuras no isomorfas (ENI).

Tipo	Grupo simétrico (S_n)	$Z(S_{E-n})$	$p(X)$	ENI (según Pólya)
E - 1	$\{I, R_\alpha, R_\beta, R_{180}\}$	$\frac{1}{4}(X_1^5 + 3X_1X_2^2)$	$\frac{1}{4}(X^5 + 3X^3)$	81
E - 2	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^6 + X_1^2X_2^2)$	$\frac{1}{2}(X^6 + X^4)$	405
E - 3	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^7 + X_1X_2^3)$	$\frac{1}{2}(X^7 + X^4)$	1.134
E - 4	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^8 + X_1^2X_2^3)$	$\frac{1}{2}(X^8 + X^5)$	3.402
E - 5	$\{I, R_\alpha, R_\beta, R_{180}\}$	$\frac{1}{4}(X^7 + X_1^3X_2^2 + 2X_1X_2^3)$	$\frac{1}{4}(X^7 + X^5 + 2X^4)$	648
E - 6	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^8 + X_1^2X_2^3)$	$\frac{1}{2}(X^8 + X^5)$	3.402
E - 7	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^9 + X_1^3X_2^3)$	$\frac{1}{2}(X^9 + X^6)$	10.206
E - 8	$\{I, R_\alpha, R_\beta, R_{180}\}$	$\frac{1}{4}(X_1^9 + 3X_1X_2^4)$	$\frac{1}{4}(X^9 + 3X^5)$	5.103
E - 9	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^{10} + X_1^2X_2^4)$	$\frac{1}{2}(X^{10} + X^6)$	29.889
E - 10	$\{I, R_\alpha, R_\beta, R_{180}\}$	$\frac{1}{4}(X_1^{11} + X_1^3X_2^4 + 2X_1X_2^5)$	$\frac{1}{4}(X^{11} + X^7 + 2X^6)$	45.198
Total estructuras no isomórficas encontradas (Policíclicos tipo espiro)				99.468

Tipo	Grupo simétrico (S_n)	$Z(S_{E-n})$	$p(X)$	ENI (según Pólya)
F - 1	$\{I, R_\alpha, R_\beta, R_{180}\}$	$\frac{1}{4}(X_1^4 + 2X_1^2X_2 + X_2^2)$	$\frac{1}{4}(X^4 + 2X^3 + X^2)$	36
F - 2	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^5 + X_1X_2^2)$	$\frac{1}{2}(X^5 + X^3)$	135
F - 3	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^6 + X_1^2X_2^2)$	$\frac{1}{2}(X^6 + X^4)$	405
F - 4	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^7 + X_1X_2^3)$	$\frac{1}{2}(X^7 + X^4)$	1.134
F - 5	$\{I, R_\alpha, R_\beta, R_{180}\}$	$\frac{1}{4}(X_1^6 + X_2^3 + X_1^2X_2^2 + X_2^3)$	$\frac{1}{4}(X^6 + X^3 + X^4 + X^3)$	216
F - 6	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^7 + X_1X_2^3)$	$\frac{1}{2}(X^7 + X^4)$	1.134
F - 7	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^8 + X_2^4)$	$\frac{1}{2}(X^8 + X^4)$	3.321
F - 8	$\{I, R_\alpha, R_\beta, R_{180}\}$	$\frac{1}{4}(X_1^8 + 2X_1^2X_2^3 + X_2^4)$	$\frac{1}{4}(X^8 + 2X^5 + X^4)$	1.782
F - 9	$\{I, R_\alpha\}$	$\frac{1}{2}(X_1^9 + X_1X_2^4)$	$\frac{1}{2}(X^9 + X^5)$	9.963
F - 10	$\{I, R_\alpha, R_\beta, R_{180}\}$	$\frac{1}{4}(X_1^{10} + 2X_2^5 + X_1^2X_2^4)$	$\frac{1}{4}(X^{10} + 2X^5 + X^6)$	15.066
Total estructuras no isomórficas encontradas (Policíclicos fusionados)				33.192

Con el fin de verificar la aplicación del Teorema de Pólya realizada, se planteó realizar el conteo de estructuras simétricas en los grupos de menor tamaño, con el fin de contrastar los resultados. Tal como puede observarse, la cantidad de estructuras no isomorfas (*ENI*) obtenidas según en Teorema de Pólya es lo suficientemente grande como para hacer inviable la enumeración manual. Por tanto, se realizó un código en Python que evaluara las equivalencias simétricas entre dos o más estructuras, a partir de un listado de estructuras base, en este caso, dado por la cantidad total de estructuras generables (es decir, el conjunto universal obtenido por la permutación 3^n).

Una vez realizada la verificación de resultados respectiva, usando la herramienta antes descrita, se observó una diferencia entre los valores estimados de manera analítica (empleando la ecuación) y el método sistemático mediante código iterativo. Se observa que el Teorema de Pólya tiene un error de subdimensionamiento del conteo de alrededor del 66%, respecto a los valores obtenidos en la verificación, lo cual se presenta en la **Tabla 3-7** para compuestos policíclicos espiro y fusionados.

Tabla 3-7: Validación de resultados de estructuras no isomorfas (*ENI*), mediante iteración.

Tipo	Diferencia Pólya (Error)	% P_U	ENI (obtenido mediante iteración)
<i>E</i> - 1	+ 135 (62,50%)	29,63	216
<i>E</i> - 2	+ 729 (64,29%)	51,85	1.134
<i>E</i> - 3	+ 2.268 (66,67%)	51,85	3.402
<i>E</i> - 4	+ 6.561 (65,85%)	50,62	9.963
<i>E</i> - 5	+ 1.053 (61,90%)	25,93	1.701
<i>E</i> - 6	+ 6.561 (65,85%)	50,62	9.963
<i>E</i> - 7	+ 19.440 (65,57%)	50,21	29.646
<i>E</i> - 8	+ 9.963 (66,13%)	25,51	15.066
<i>E</i> - 9	+ 59.049 (66,39%)	50,21	88.938
<i>E</i> - 10	+ 88.209 (66,12%)	25,10	133.407
Total Policíclicos Espiro			293.436

Tipo	Diferencia Pólya (Error)	% P_U	ENI (obtenido mediante iteración)
<i>F</i> - 1	+ 45 (55,56%)	33,33	81
<i>F</i> - 2	+ 270 (66,67%)	55,56	405
<i>F</i> - 3	+ 729 (64,29%)	51,85	1.134
<i>F</i> - 4	+ 2.268 (66,67%)	51,85	3.402
<i>F</i> - 5	+ 432 (66,67%)	29,63	648
<i>F</i> - 6	+ 2.268 (66,67%)	51,85	3.402
<i>F</i> - 7	+ 6.885 (67,46%)	51,85	10.206
<i>F</i> - 8	+ 3.321 (65,08%)	25,93	5.103
<i>F</i> - 9	+ 19.926 (66,67%)	50,62	29.889
<i>F</i> - 10	+ 30.132 (66,67%)	25,51	45.198
Total Policíclicos Fusionados			99.468

La diferencia entre los resultados obtenidos mediante el Teorema de Enumeración de Pólya y el método iterativo realizado por computador reflejan que el teorema se encuentra

limitado por diferencias estructurales adicionales a las presentadas en compuestos monocíclicos, por lo que la aplicabilidad de este teorema se encuentra limitada para compuestos policíclicos.

Se identificaron dos causas que pueden justificar las discrepancias obtenidas:

- En primer lugar, dado que las estructuras policíclicas (tanto espiro como fusionadas) no son regulares geoméricamente, es posible que un enlace de la molécula se superponga a un eje de simetría y/o de rotación. Esto implica que, a medida que se hace el conteo mediante el Teorema de Enumeración de Pólya, existen ciertas estructuras que se enmascaran y pasan como isomorfos cuando en realidad no lo son. Este inconveniente ocurre solamente en compuestos policíclicos, dado que la representación estructural de los compuestos monocíclicos está relacionada directamente con polígonos regulares y por ende con grupos diedrales específicos ya establecidos.
- En segundo lugar, dada las características estructurales de los compuestos policíclicos, se identificó un mayor impacto de la cantidad y ubicación de los enlaces comunes a dos ciclos, en la cantidad de estructuras generadas. Dado que Teorema de Enumeración de Pólya solo considera movimientos rotacionales y de simetría respecto a puntos pivotes (en este caso particular, los átomos), pueden generarse aún más estructuras que no pueden ser contadas por la ecuación respectiva.

Teniendo en cuenta las razones antes mencionadas, se concluye que no es posible aplicar de forma directa el Teorema de Pólya para la enumeración eficiente de compuestos policíclicos, dada su incapacidad de identificar ciertas estructuras específicas. Por lo anterior, las ecuaciones de índice de ciclo $Z(x_1, x_2, \dots, x_n)$ no permiten obtener funciones generadoras $p(X)$ que reflejen de forma apropiada la cantidad real de estructuras no isomorfos presentes en un tipo particular de compuesto policíclico. Esto pudo verificarse mediante la búsqueda de la serie numérica ENI obtenida en la plataforma OEIS (*Enciclopedia en línea de Secuencias Enteras*, por sus siglas en inglés) [250], la cual reporta ausencia de polinomio generador para el conjunto de números obtenido. Por tal razón, es necesario proponer una modificación al teorema en cuestión para poder determinar una forma de conteo específica para compuestos policíclicos espiro y funcionales. Se deja para estudios posteriores el establecimiento de las ecuaciones de

índice de ciclo $Z(x_1, x_2, \dots, x_n)$ y las funciones generadoras $p(X)$ para cada uno de los grupos policíclicos espiro y fusionados estudiados.

Por todo lo antes mencionado, se decide continuar con la evaluación de la metodología de tres filtros empleando el método iterativo para el caso particular de compuestos policíclicos. De esta manera, la aplicación del Filtro I de forma iterativa permitió identificar 392.904 estructuras no isomorfas (*ENI*), correspondientes a 293.436 estructuras de tipo espiro (74,68%) y 99.468 estructuras fusionadas (25,32%), las cuales pasan al Filtro II.

Compuestos policíclicos - Filtro II:

A partir de las 392.904 estructuras no isomorfas (*ENI*) que superaron el Filtro I, se realizó una permutación con repetición (*P*) para determinar los posibles átomos asignables para una molécula policíclica de cierto tipo (espiro o fusionada), conformada por *n*-átomos, y, por ende, la cantidad de sustituciones de heteroátomos posibles considerando las valencias válidas del átomo (*u* átomos) a reemplazar.

Nuevamente, el producto $ENI \times P$ representa la cantidad de combinaciones totales cuya validez de valencia debe ser analizada, siguiendo los mismos criterios estructurales antes mencionados. La **Tabla 3-8** presenta los resultados obtenidos para el Filtro II en compuestos policíclicos espiro y fusionados, respectivamente.

Tabla 3-8: Cálculo de $ENI \times P$ para evaluación de estructuras policíclicas (Filtro II).

Ciclo	<i>P</i>	<i>ENI</i>	<i>ENI</i> × <i>P</i>	Ciclo	<i>P</i>	<i>ENI</i>	<i>ENI</i> × <i>P</i>
<i>E</i> - 1	1.024	216	221.184	<i>F</i> - 1	256	81	20.736
<i>E</i> - 2	4.096	1.134	4.644.864	<i>F</i> - 2	1.024	405	414.720
<i>E</i> - 3	16.384	3.402	55.738.368	<i>F</i> - 3	4.096	1.134	4.644.864
<i>E</i> - 4	65.536	9.963	652.935.168	<i>F</i> - 4	16.384	3.402	55.738.368
<i>E</i> - 5	16.384	1.701	27.869.184	<i>F</i> - 5	4.096	648	2.654.208
<i>E</i> - 6	65.536	9.963	652.935.168	<i>F</i> - 6	16.384	3.402	55.738.368
<i>E</i> - 7	262.144	29.646	7.771.521.024	<i>F</i> - 7	65.536	10.206	668.860.416
<i>E</i> - 8	262.144	15.066	3.949.461.504	<i>F</i> - 8	65.536	5.103	334.430.208
<i>E</i> - 9	1.048.576	88.938	93.258.252.288	<i>F</i> - 9	262.144	29.889	7.835.222.016
<i>E</i> - 10	4.194.304	133.407	559.549.513.728	<i>F</i> - 10	1.048.576	45.198	47.393.538.048
∑	5.936.128	293.436	665.923.092.480	∑	1.484.032	293.436	56.351.261.952

A diferencia del caso monocíclico, la cantidad de sustituciones de heteroátomos posibles es significativamente grande como para hacer inviable una validación manual o usando representaciones Mol, estableciendo sustituciones parciales de R-átomos. Por tal razón, fue necesario analizar una estrategia para reducir el número de evaluaciones para identificar las estructuras factibles, a partir de la extensión de estructuras ENI obtenida en el Filtro I.

La enorme extensión de estructuras ENI se encuentra justificada producto de la mayor cantidad de puntos de variación estructural (en el caso de los policíclicos espiro condicionados a que el heteroátomo central sea un mismo tipo de átomo para ambos ciclos, mientras que en el caso de los policíclicos fusionados debe cumplirse que coexistan dos heteroátomos colindantes, con el mismo tipo de enlace en común). Esto genera que las sustituciones de los R-átomos con los átomos (C, N, O, S) junto con las distintas relaciones asociadas a los tres tipos de enlace, sea significativamente superior la cantidad de estructuras a evaluar por valencia atómica, respecto al caso monocíclico.

Por tal razón, con el fin de facilitar el proceso de filtrado mediante análisis de la valencia de los átomos de la estructura, se subdividió el problema en dos partes: Análisis de la valencia en los heteroátomos centrales (es decir, átomos compartidos por dos ciclos en la configuración espiro y fusionada) y análisis de la valencia de átomos proximales (pertenecientes a un solo ciclo de la molécula policíclica, es decir los demás átomos del compuesto policíclico). De esta manera, se analizó por separado la validez de las sustituciones de los heteroátomos centrales y posteriormente, para los grupos válidos por valencia, se evaluaron las combinaciones con los demás átomos de la molécula.

Para el caso de los heteroátomos centrales, se realizó el filtrado de configuraciones posibles simétricas generalizadas, obteniéndose 129 configuraciones de compuestos policíclicos (21 para policíclicos tipo espiro y 108 para policíclicos fusionados), los cuales fueron presentado por la **Figura 2-12** y la **Figura 2-13**, respectivamente.

Evaluando la posibilidad de sustitución del espiroátomo por un heteroátomo, se identificaron 6 configuraciones válidas respecto a valencia molecular, indicados con color verde en la **Figura 3-2**, lo cual representa el 7,14% de las 84 configuraciones evaluadas.

Las configuraciones resaltadas en rojo incumplen los criterios de valencia atómica o generan cargas sobre el átomo central, por lo que fueron rechazadas.

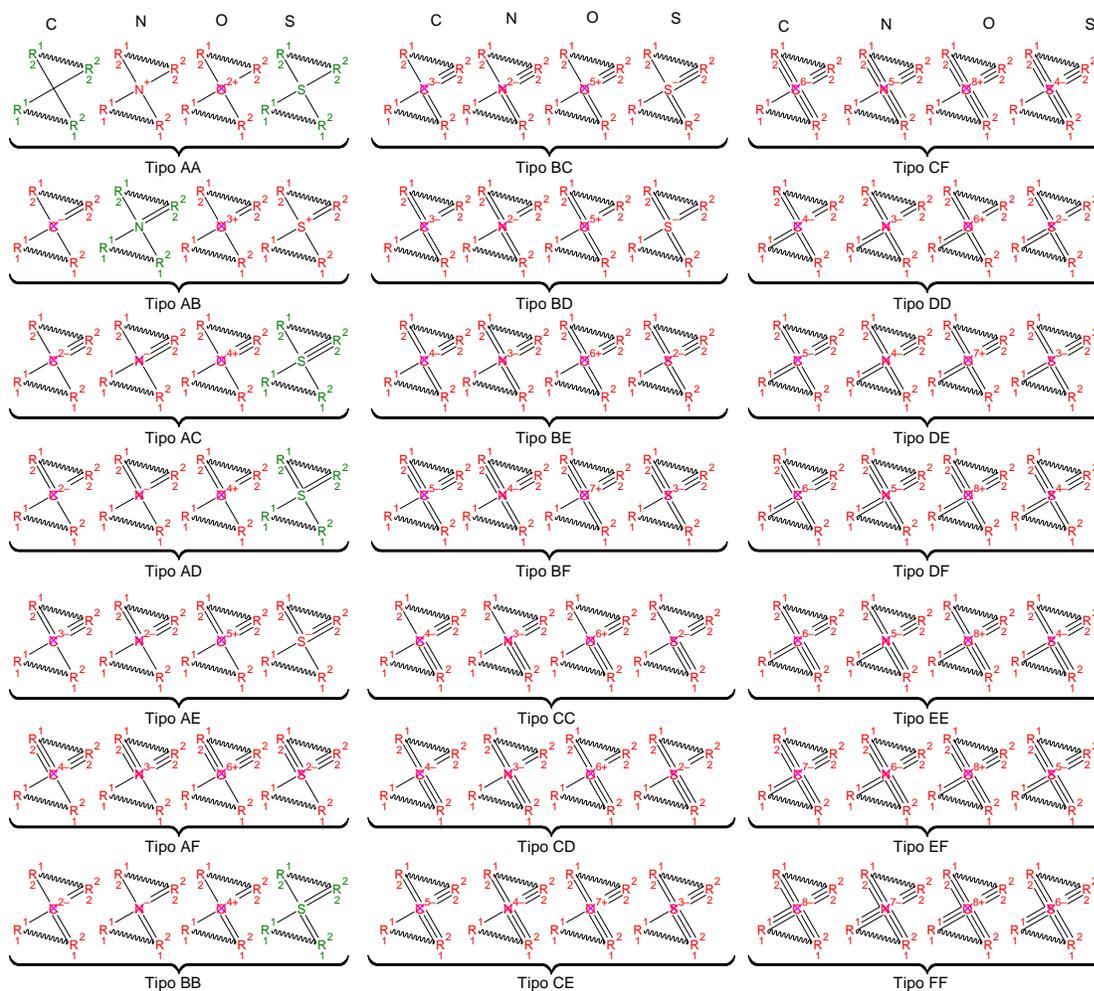


Figura 3-2: Viabilidad de sustitución del R-átomo en policíclicos espiro y estructuras viables respecto a valencia (Filtro II).

Para el caso de compuestos policíclicos, dado que es necesario evaluar la viabilidad de dos heteroátomos en conjunto, cada evaluación de las estructuras requerirá analizar una matriz de viabilidad de valencia para cada uno de los heteroátomos de las estructuras de la **Figura 2-13**. Con el fin de facilitar la visualización de los resultados, la figura **Figura 3-3** presenta los arreglos viables de estructuras policíclicas de forma matricial, de forma similar al arreglo matricial 4x4 presentado en la **Figura 2-14**. Se obtuvieron 79 estructuras válidas por valencia de 1728 configuraciones estudiadas (108 estructuras x 16 combinaciones), lo

cual representa solo el 4,57% de total evaluado. La **Figura 3-4** presenta el listado de las configuraciones policíclicas fusionadas que cumplen la valencia con los dos heteroátomos centrales.

	COL 1	COL 2	COL 3	COL 4	COL 5	COL 6	COL 7	COL 8	COL 9
Fila 1	Estruct. 1	Estruct. 2	Estruct. 3	Estruct. 4	Estruct. 5	Estruct. 6	Estruct. 7	Estruct. 8	Estruct. 9
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C 1 1 N 1 1 O S	C 1 1 N O S 1 1	C N 1 1 O S	C N 1 1 O S	C N O S 1 1	C N O S 1 1	C N O S 1 1	C 1 1 N 1 1 O S	C 1 1 N O S 1 1
Fila 2	Estruct. 10	Estruct. 11	Estruct. 12	Estruct. 13	Estruct. 14	Estruct. 15	Estruct. 16	Estruct. 17	Estruct. 18
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C N 1 O S	C N O S 1 1	C N O S	C N O S	C N O S 1 1	C N O S 1 1	C N O S 1 1	C N O S 1 1	C N O S
Fila 3	Estruct. 19	Estruct. 20	Estruct. 21	Estruct. 22	Estruct. 23	Estruct. 24	Estruct. 25	Estruct. 26	Estruct. 27
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C N 1 O S	C N O S 1 1	C N O S	C N O S 1 1	C N O S	C N O S	C N O S 1 1	C N O S 1 1	C N O S 1 1
Fila 4	Estruct. 28	Estruct. 29	Estruct. 30	Estruct. 31	Estruct. 32	Estruct. 33	Estruct. 34	Estruct. 35	Estruct. 36
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C N O S	C N O S 1 1	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S
Fila 5	Estruct. 37	Estruct. 38	Estruct. 39	Estruct. 40	Estruct. 41	Estruct. 42	Estruct. 43	Estruct. 44	Estruct. 45
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C 1 1 N O S 1 1	C N 1 1 O S	C N O S	C N O S 1 1	C N O S 1 1	C N O S	C N O S 1 1	C N O S 1 1	C N O S 1 1
Fila 6	Estruct. 46	Estruct. 47	Estruct. 48	Estruct. 49	Estruct. 50	Estruct. 51	Estruct. 52	Estruct. 53	Estruct. 54
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C N O S 1 1	C N O S	C N O S	C N O S 1 1	C N O S 1 1	C N O S 1 1	C N O S 1 1	C N O S	C N O S
Fila 7	Estruct. 55	Estruct. 56	Estruct. 57	Estruct. 58	Estruct. 59	Estruct. 60	Estruct. 61	Estruct. 62	Estruct. 63
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C N O S 1 1	C N O S 1 1	C N O S	C N O S 1 1	C N O S 1 1	C N O S	C N O S	C N O S	C N O S
Fila 8	Estruct. 64	Estruct. 65	Estruct. 66	Estruct. 67	Estruct. 68	Estruct. 69	Estruct. 70	Estruct. 71	Estruct. 72
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S
Fila 9	Estruct. 73	Estruct. 74	Estruct. 75	Estruct. 76	Estruct. 77	Estruct. 78	Estruct. 79	Estruct. 80	Estruct. 81
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C N O S 1 1	C N O S 1 1	C N O S	C N O S	C N O S	C N O S	C N O S 1 1	C N O S 1 1	C N O S 1 1
Fila 10	Estruct. 82	Estruct. 83	Estruct. 84	Estruct. 85	Estruct. 86	Estruct. 87	Estruct. 88	Estruct. 89	Estruct. 90
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S
Fila 11	Estruct. 91	Estruct. 92	Estruct. 93	Estruct. 94	Estruct. 95	Estruct. 96	Estruct. 97	Estruct. 98	Estruct. 99
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S
Fila 12	Estruct. 100	Estruct. 101	Estruct. 102	Estruct. 103	Estruct. 104	Estruct. 105	Estruct. 106	Estruct. 107	Estruct. 108
	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S	X C N O S
	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S	C N O S

Figura 3-3: Viabilidad de sustitución del R-átomo en policíclicos fusionados.

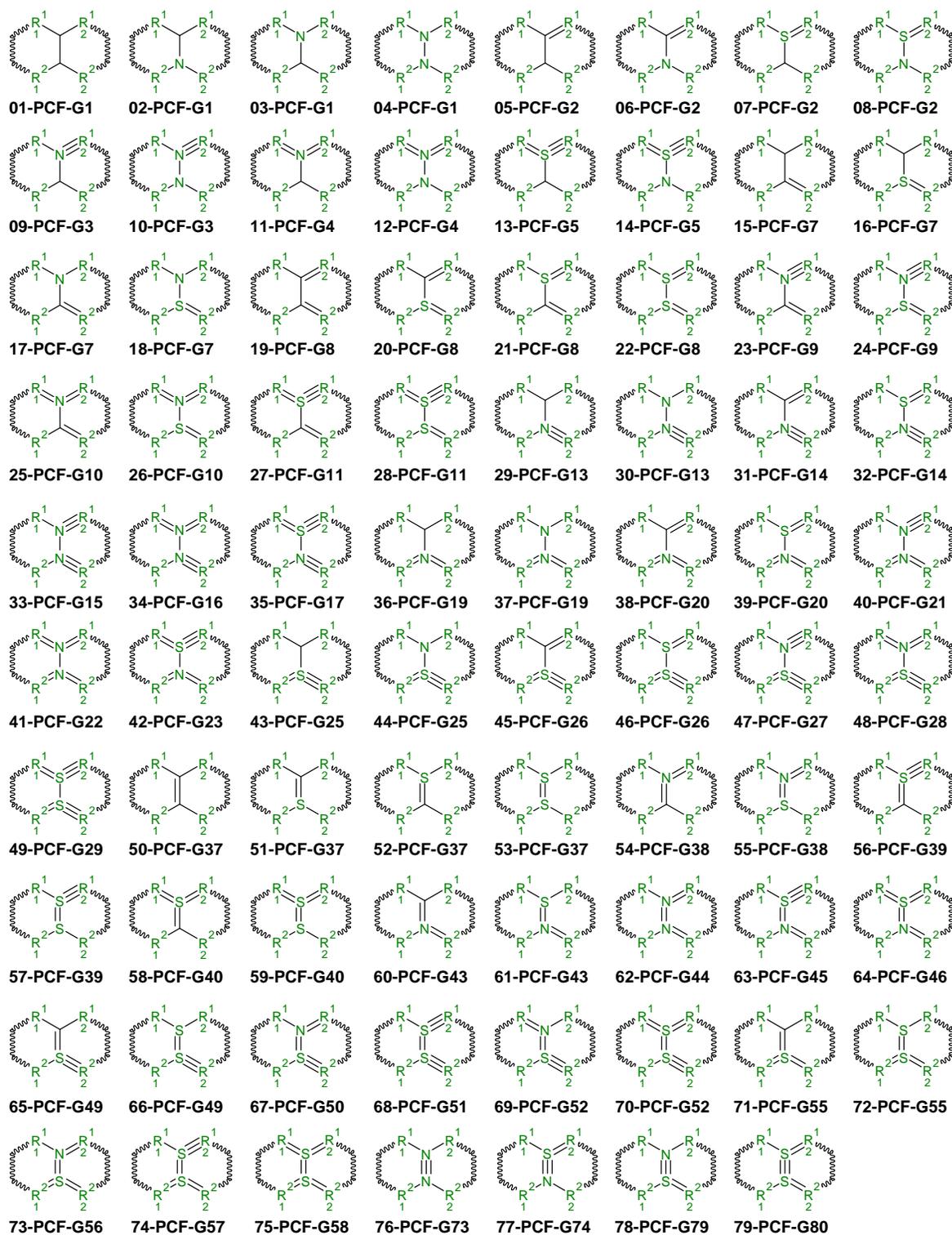


Figura 3-4: Verificación de la viabilidad de valencia estructural en estructuras policíclicas fusionadas mediante Filtro II.

De esta manera, el problema global de evaluación de valencia de estructuras descrito por las **Tabla 3-8** se reduce significativamente a evaluar los átomos proximales en cada tipo de configuración (policíclica espiro o fusionada, tomando como base estructural las configuraciones que cumplen la valencia sobre el heteroátomo central, descritas por la **Figura 3-2** (para compuestos espiro) y la **Figura 3-4** (para compuestos fusionados).

Con respecto a los heteroátomos periféricos, pertenecientes a un solo ciclo dentro de la estructura policíclica, es posible establecer rápidamente el cumplimiento de la valencia (o no), siguiendo la misma figura presentada para la validación de valencia de R-átomos en compuestos monocíclicos, en **Figura 2-8**.

De esta manera, el Filtro II es adaptado fácilmente a las necesidades de las estructuras policíclicas, permitiendo la eficiente evaluación de la validez de las valencias atómicas para cada una de las combinaciones de compuestos policíclicos determinados por la **Tabla 3-9**. En este punto, fue necesario verificar nuevamente la generación de estructuras no isomorfas, similar al Filtro I, las cuales pueden generarse una vez se realiza las sustituciones de los R-átomos por los respectivos heteroátomos centrales.

Siguiendo esta estrategia, empleó el mismo script de Python empleado para compuestos monocíclicos, con modificaciones enfocadas a permitir la determinación de las estructuras policíclicas válidas para cada uno de los tipos de configuraciones policíclicas evaluadas. La **Tabla 3-9** presenta los resultados obtenidos para este tipo de estructuras policíclicas en el Filtro II. Cabe resaltar, que considerando la cantidad de evaluaciones requerida para cada uno de los grupos policíclicos, dadas por el producto $ENI \times P$ respectivo, la evaluación por fuerza bruta (es decir, haciendo la verificación individual de cada una de las estructuras policíclicas), es inviable en términos de tiempo de cálculo, a menos de que se incluyan estrategias de programación avanzadas⁴⁵, que permitan verificar la valencia

⁴⁵ Entre las múltiples estrategias de programación empleadas en este trabajo para la aplicación del Filtro II está el aprovechamiento de múltiples núcleos dentro del procesador mediante paralelización del código, la disminución de llamados y recursividades de las funciones establecidas, el uso de unidades de procesamiento gráfico (GPU) y la optimización del recurso de memoria RAM y disco duro.

estructural en tiempos de evaluación razonables⁴⁶ y capacidades de almacenamiento viables⁴⁷.

Tabla 3-9: Grupos válidos para compuestos policíclicos espiro n-atómicos considerando valencias atómicas, obtenidos mediante sustituciones de heteroátomos (Filtro II).

Ciclo	P	ENI	ENI × P	¿Superan Filtro II? (valencia atómica)		¿Superan Filtro II y son no isomorfos? (%)*	Repetición respecto a estructuras válidas (%)
				SI (%)*	NO (%)*		
E – 1	1.024	216	221.184	1.194 (0,5398)	219.990 (99,4602)	380 (0,1718)	68,17%
E – 2	4.096	1.134	4.644.864	7.772 (0,1673)	4.637.092 (99,8327)	3.952 (0,0851)	49,15%
E – 3	16.384	3.402	55.738.368	52.954 (0,0950)	55.685.414 (99,9050)	24.396 (0,0438)	53,93%
E – 4	65.536	9.963	652.935.168	285.476 (0,0437)	652.649.692 (99,9563)	156.256 (0,0239)	45,26%
E – 5	16.384	1.701	27.869.184	28.184 (0,1011)	27.841.000 (99,8989)	10.920 (0,0392)	61,25%
E – 6	65.536	9.963	652.935.168	277.548 (0,0425)	652.657.620 (99,9575)	133.536 (0,0205)	51,89%
E – 7	262.144	29.646	7.771.521.024	1.646.696 (0,0212)	7.769.874.328 (99,9788)	n.d.**	n.d.**
E – 8	262.144	15.066	3.949.461.504	1.094.530 (0,0277)	3.948.366.974 (99,9723)	412.806 (0,0105)	62,28%
E – 9	1.048.576	88.938	93.258.252.288	10.978.372 (0,0118)	93.247.273.916 (99,9882)	n.d.**	n.d.**
E – 10	4.194.304	133.407	559.549.513.728	37.596.384 (0,0067)	559.511.917.344 (99,9933)	n.d.**	n.d.**
Sustituc. válidas**	5.936.128	293.436	665.923.092.480	51.969.110 (0,0078)	665.871.123.370 (99,9922)	742.246 (0,0001%)	98,57%

* Porcentaje respecto al producto ENI × P.

** n.d. = no determinado. Debido a la enorme extensión de estructuras generadas, demora en el tiempo de cálculo (estimado en semanas) y la dificultad de almacenar dicha cantidad de estructuras, se desestimó realizar la evaluación de los grupos E – 7, E – 9 y E – 10. No obstante, considerando las tasas de aceptación de estructuras obtenidas en los otros grupos policíclicos espiro, es apropiado considerar que la cantidad de estructuras aceptadas en dichos grupos puede rondar el 1% - 3% de estructuras evaluadas en el producto ENI × P, para cada grupo en particular.

⁴⁶ El tiempo de cálculo dependerá significativamente de los requerimientos del computador en donde se evalúe. Como referencia a los tiempos promedios reportados, el cálculo del tiempo fue realizado a partir de los tiempos de ejecución en un computador AMD Ryzen 9 III Gen, con 6Gb de RAM y 24 núcleos de procesamiento. Adicionalmente, este equipo contaba con una GPU integrada RTX 3070, la cual cuenta a su vez con 6.204 núcleos de procesamiento exclusivo. Se implementaron técnicas de código limpio, de acuerdo con lo mencionado. Estos resultados de tiempo pueden diferir significativamente (en horas o días), en computadores con especificaciones más antiguas a la referencia antes mencionada.

⁴⁷ Debido a la generación masiva de estructuras con bajo peso de almacenamiento (alrededor de 2Kb), fue necesario comprimir los resultados en archivos tipo WinRAR (*.rar), con el fin de reducir el gasto de espacio en almacenamiento en disco. Se empleó una tasa de compresión de entre 20%-25%, para cada grupo, con el fin de aprovechar más eficientemente el almacenamiento de datos.

Tabla 3-10: Grupos válidos para compuestos policíclicos fusionados n-atómicos considerando valencias atómicas, obtenidos mediante sustituciones de heteroátomos (Filtro II).

Ciclo	<i>P</i>	<i>ENI</i>	<i>ENI</i> × <i>P</i>	¿Superan Filtro II? (valencia atómica)		¿Superan Filtro II y son no isomorfas? (%)*	Repetición respecto a estructuras válidas (%)
				SI (%)*	NO (%)*		
<i>F</i> – 1	256	81	20.736	348 (1,6782)	20.388 (99,3218)	207 (0,9983)	40,52%
<i>F</i> – 2	1.024	405	414.720	3.118 (0,7518)	411.602 (99,2482)	2.272 (0,5478)	27,13%
<i>F</i> – 3	4.096	1.134	4.644.864	3.402 (0,0732)	4.641.462 (99,9268)	2.124 (0,0457)	37,57%
<i>F</i> – 4	16.384	3.402	55.738.368	28.928 (0,0519)	55.709.440 (99,9481)	20.720 (0,0372)	28,37%
<i>F</i> – 5	4.096	648	2.654.208	3.481 (0,1312)	2.650.727 (99,8688)	1.187 (0,0447)	65,90%
<i>F</i> – 6	16.384	3.402	55.738.368	115.024 (0,2064)	55.623.344 (99,7936)	95.632 (0,1716)	16,86%
<i>F</i> – 7	65.536	10.206	668.860.416	206.932 (0,0309)	668.653.484 (99,9691)	167.894 (0,0251)	18,87%
<i>F</i> – 8	65.536	5.103	334.430.208	395.492 (0,1183)	334.034.716 (99,8817)	311.640 (0,0932)	21,20%
<i>F</i> – 9	262.144	29.889	7.835.222.016	557.504 (0,0071)	7.834.664.512 (99,9929)	429.872 (0,0055)	22,89%
<i>F</i> – 10	1.048.576	45.198	47.393.538.048	994.212 (0,0021)	47.392.543.836 (99,9979)	324.670 (0,0007)	67,34%
Sustituciones validas	1.484.032	293.436	56.351.261.952	2.308.441 (0,0041)	56.348.953.511 (99,9959)	1.356.218 (0,0024)	41,25%

* Porcentaje respecto al producto *ENI* × *P*.

De esta manera, se determinó que solamente 742.246 estructuras no isomorfas policíclicas tipo espiro cumplen con los criterios de valencia establecidos por el modelo CAMD, correspondientes al 0,00011% del total *ENI* × *P*. De forma similar, se solamente 1.356.218 estructuras no isomorfas policíclicas tipo fusionadas cumplen con los criterios de valencia en cuestión, correspondientes al 0,00241% del total *ENI* × *P*. Se puede observar una tasa de aceptación de estructuras significativamente baja (menor al 1% del total *ENI* × *P* en todos los grupos), siendo este comportamiento independiente de los tamaños de ciclo involucrados. tanto para estructuras policíclicas espiro como fusionadas. Esto indica que el espacio de exploración de estructuras policíclicas a evaluar cuenta con un impacto significativo de la asignación de valencia cero. Esto ocurre, en parte, se debe a que durante el proceso de sustitución de los R-átomos en las estructuras generalizadas por átomos (C, N, O, S), se estructuras repetidas, por lo aplicación de una revisión del no isomorfismo estructural permite simplificar significativamente el conjunto de estructuras que pasarán al Filtro III.

Compuestos policíclicos - Filtro III:

Para las 2.098.464 estructuras policíclicas que superaron el Filtro II, se evaluó la aplicabilidad de propiedades mediante los modelos de estimación de propiedades descritos en la Sección 2.2.7 y 2.2.8. La **Tabla 3-11** presenta la aplicabilidad observada de las estructuras policíclicas evaluadas con respecto a cada modelo de estimación de propiedades, para los tipos espiro y fusionadas respectivamente.

Tabla 3-11: Aplicabilidad de los modelos de estimación de propiedades para las estructuras evaluadas de compuestos policíclicos espiro y fusionadas en el Filtro III.

Tipo de compuesto policíclico espiro	Por ningún modelo	Solo con modelo de Hukkerikar	Solo por el modelo UNIFAC-DMD	Por ambos modelos	Total
<i>E</i> - 1	67 (17,63%)	161 (42,32%)	107 (28,16%)	45 (11,84%)	380
<i>E</i> - 2	733 (18,55%)	1.761 (44,56%)	1.174 (29,71%)	284 (7,19%)	3.952
<i>E</i> - 3	4679 (19,18%)	11.630 (47,67%)	6.455 (26,46%)	1.632 (6,69%)	24.396
<i>E</i> - 4	31.645 (20,25%)	70.068 (44,84%)	46.712 (29,89%)	7.831 (5,01%)	156.256
<i>E</i> - 5	2.284 (20,92%)	5.883 (53,87%)	2.255 (20,65%)	498 (4,56%)	10.920
<i>E</i> - 6	30.858 (23,11%)	73.607 (55,12%)	24.389 (18,26%)	4.682 (3,51%)	133.536
<i>E</i> - 7	n.d.**	n.d.**	n.d.**	n.d.**	n.d.**
<i>E</i> - 8	103.779 (25,14%)	235.416 (57,03%)	66.142 (16,02%)	7.469 (1,81%)	412.806
<i>E</i> - 9	n.d.**	n.d.**	n.d.**	n.d.**	n.d.**
<i>E</i> - 10	n.d.**	n.d.**	n.d.**	n.d.**	n.d.**
Total	174.045 (23,45%)	398.526 (53,69%)	147.234 (19,84%)	22.441 (3,02%)	742.246

Tipo de compuesto policíclico fusionado	Por ningún modelo	Solo con modelo de Hukkerikar	Solo por el modelo UNIFAC-DMD	Por ambos modelos	Total
<i>F</i> - 1	24 (11,59%)	82 (39,61%)	61 (29,47%)	40 (19,32%)	207
<i>F</i> - 2	317 (13,95%)	934 (41,11%)	729 (32,09%)	292 (12,85%)	2.272
<i>F</i> - 3	307 (14,45%)	912 (42,94%)	636 (29,94%)	269 (12,66%)	2.124
<i>F</i> - 4	3.060 (14,77%)	9.881 (47,69%)	4.964 (23,96%)	2.815 (13,59%)	20.720
<i>F</i> - 5	176 (14,83%)	613 (51,64%)	288 (24,26%)	110 (9,27%)	1.187
<i>F</i> - 6	14.909 (15,59%)	50.149 (52,44%)	22.054 (23,06%)	8.520 (8,91%)	95.632
<i>F</i> - 7	31.279 (18,63%)	89.772 (53,47%)	32.455 (19,33%)	14.388 (8,57%)	167.894
<i>F</i> - 8	66.940 (21,48%)	89.772 (55,61%)	45.469 (14,59%)	25.928 (8,32%)	311.640
<i>F</i> - 9	96.549 (22,46%)	173.303 (57,94%)	61.086 (14,21%)	23.170 (5,39%)	429.872
<i>F</i> - 10	80.096 (24,67%)	190.875 (58,79%)	37.011 (11,40%)	16.688 (5,14%)	324.670
Total	293.657 (21,65%)	765.558 (56,45%)	204.753 (15,10%)	92.220 (6,80%)	1.356.218

** n.d. = no determinado. Dado que los grupos *E* - 7, *E* - 9 y *E* - 10 no fueron evaluados en el Filtro II por las consideraciones mencionadas en la **Tabla 3-9**, no es posible cuantificar la aplicabilidad de dichas moléculas dentro de los modelos de estimación de propiedades en el Filtro III.

A partir de los resultados obtenidos en la **Tabla 3-11**, es posible establecer que la metodología construida para establecimiento de nuevos grupos funcionales es aplicable al problema de diseño CAMD para la determinación de propiedades, para estructuras policíclicas, tanto espiro como fusionadas.

Cabe resaltar el hecho de que la dificultad de evaluación de las estructuras policíclicas dentro del modelo CAMD no está dado por el modelo CAMD en sí, ni por el algoritmo de optimización utilizado (MOHAEA), sino por la dificultad de representación de las propiedades que requiere el modelo CAMD, es decir, por las limitaciones asociadas a los modelos de UNIFAC-DMD y Hukkerikar. Por esta razón, conforme dichos modelos sean actualizados, el porcentaje de estructuras válidas que pueden ingresar al modelo CAMD aumentará. Se sugiere para futuros estudios tomar como punto de partida las estructuras obtenidas en el filtro II para reevaluar la aplicabilidad de los modelos de estimación de propiedades para los nuevos grupos de parámetros incorporados a estos modelos.

Compuestos policíclicos – Resumen y análisis:

La **Tabla 3-12** presenta un resumen del proceso de filtrado de grupos en compuestos policíclicos sustituidos por heteroátomos. Debido a su extensión, la representación estructural de los grupos heteroatómicos policíclicos válidos (espiro y fusionados) añadidos al modelo CAMD es remitida en medio digital, tal como se menciona en el **Anexo I**.

A partir de los resultados obtenidos en la **Tabla 3-12**, es posible establecer que la metodología para establecida previamente para compuestos monocíclicos heteroatómicos es fácilmente extrapolable a compuestos policíclicos, bajo ligeras variaciones. Nuevamente, la creación de grupos policíclicos (espiro y fusionados) de más de 2 ciclos conectados es también extrapolable, siguiendo la metodología propuesta. No obstante, es importante considerar que, al ser los compuestos policíclicos de mayor complejidad, el efecto entre el tamaño de los ciclos de la estructura policíclica y la cantidad de grupos estructurales que son capaces de superar los tres filtros es aún más significativa, respecto a los compuestos monocíclicos. Sin embargo, la menor cantidad de movimientos rotacionales para las estructuras policíclicas ejerce un mayor condicionamiento en el Filtro II, junto con la baja tasa de configuraciones validas respecto a la valencia en el Filtro II, genera un mayor descarte de estructuras no válidas, sin importar el tamaño o tipo de compuesto policíclico heteroatómicos considerado.

Por otra parte, es importante considerar la viabilidad de adicionar los compuestos policíclicos de mayor complejidad a la evaluación del modelo CAMD y, por ende, para la verificación experimental propuesta. Como fue mencionado anteriormente, la adición de

estos grupos tendrá un consumo significativo de recursos de memoria y tiempo de cálculo, lo cual puede hacer reducir el interés de adicionar estos compuestos al modelo CAMD.

Tabla 3-12: Resumen metodología de tres filtros: Compuestos policíclicos.

Criterio de filtrado de estructuras		Compuestos policíclicos		
		Espiro	Fusionados	Total
Configuraciones estructurales iniciales evaluadas (P_U)		10	10	20
Total combinaciones construibles a partir de las estructuras iniciales evaluadas		882.090	1.011.366	1.893.456
Filtro 1: ¿La configuración estructural está repetida por simetría con otra?	Si (%)	782.622 (88,72)	978.174 (96,72)	1.760.796
	No (%)	99.468 (11,27)	33.192 (3,28)	132.660
Estructuras no isomorfas (ENI)*		99.468	33.192	132.660
Permutaciones factibles (P) por sustitución de R-átomos **		2.448.540.472.320	204.045.039.360	2.652.585.511.680
Producto ENI x P		665.923.092.480	56.351.261.952	722.274.354.432
Filtro 2: ¿La configuración representa una estructura válida respecto a valencia atómica?	No (%)	665.871.123.370 (99,9922)	56.348.953.511 (99,9959)	722.220.076.881
	Si (%)	51.969.110 (0,0078)	2.308.441 (0,0041)	54.277.551
Estructuras moleculares evaluadas en modelos de propiedades. ***		742.246	1.356.218	2.098.464
Filtro 3: ¿Son válidos ambos modelos para determinación de propiedades? ****	Si (%)	22.441 (3,02%)	92.220 (6,80%)	114.661
	No (%)	719.805 (96,97%)	1.263.998 (93,20%)	1.983.803
Grupos válidos adicionados al modelo CAMD		22.441	92.220	114.661

* Representadas por las **Figura 2-12** y la **Figura 2-13**. Estas estructuras son una generalización de todos los casos contados con el Teorema Enumerador de Pólya.

** Las sustituciones consideradas en el estudio pueden ser de uno o varios R-átomos por C y heteroátomos (N, O, S). En el caso de las sustituciones de ciclos según tamaño de anillo puede de 3, 4, 5 o 6 átomos.

*** Se emplearon los modelos de Hukkerikar y UNIFAC-DMD para evaluar a las estructuras moleculares.

**** La validez de la estructura molecular propuesta radica en el cumplimiento de las valencias de los átomos involucrados (incluyendo la no presencia de cargas en la molécula).

Teniendo en cuenta lo anterior, es posible concluir que, aunque el modelo de determinación de grupos policíclicos es extrapolable fácilmente (bajo ciertos ajustes, a partir de la metodología planteada para el caso de grupos monocíclicos), la integración de los grupos funcionales policíclicos al modelo CAMD tendrá un carácter práctico muy reducido, respecto al espacio de exploración estructural definido, requiriendo un consumo de recursos significativo (tiempo y memoria) para la ejecución de las corridas experimentales bajo el modelo CAMD propuesto. Además, si se considera la baja tasa de representación de compuestos policíclicos en modelos de solventes de extracción, los

limitados parámetros de grupos funcionales policíclicos asociados al modelo UNIFAC-DMD y la extensa variabilidad relacionada al tercer orden del modelo de Hukkerikar, es posible justificar la no adición de los grupos funcionales policíclicos dentro del programa de diseño CAMD realizado en el presente estudio.

3.1.2 Impacto del carácter aromático en la construcción de compuestos heterocíclicos

El carácter aromático de los compuestos está asociado a la capacidad de la estructura molecular de compuesto de distribución de electrones π deslocalizados a lo largo de orbitales sp^2 hibridados, lo cual confiere una estabilidad adicional a la estructura cíclica. Estas estructuras siguen la regla de Hückel ($4n+2$ electrones π). Por su parte, los compuestos antiaromáticos suelen fallar la regla de Hückel, teniendo $4n$ electrones π en su sistema conjugado. La **Tabla 3-13** identifica la cantidad de moléculas cíclicas que cumplen con alguno de estos criterios, segregadas por tipo de compuesto heterocíclico, de acuerdo con las moléculas establecidas en el estudio de aplicabilidad, reportadas en el **Anexo C**.

Dentro de las 162 moléculas evaluadas, existe un comportamiento diferenciado respecto moléculas monocíclicas y policíclicas, por lo cual es necesario hacer una discriminación de los resultados de carácter aromático según el tipo de molécula ciclada.

Tabla 3-13: Carácter aromático de las moléculas empleadas en el estudio de aplicabilidad.

Tipos de compuestos	Carácter aromático			Total moléculas evaluadas
	Antiaromático (CA=-1)	No aromático ni antiaromático (CA=0)	Aromático (CA=1)	
Monocíclicos con ciclos de 3 átomos	1 (25,00%)	3 (75,00%)	0 (0,00%)	4 (2,47%)
Monocíclicos con ciclos de 4 átomos	1 (11,11%)	8 (88,89%)	0 (0,00%)	9 (5,56%)
Monocíclicos con ciclos de 5 átomos	0 (0,0%)	23 (67,65%)	11 (32,35%)	34 (20,99%)
Monocíclicos con ciclos de 6 átomos	0 (0,0%)	30 (81,08%)	7 (18,92%)	37 (22,84%)
Policíclicos con configuración espiro	0 (0,00%)	12 (100,00%)	0 (0,00%)	12 (7,41%)
Policíclicos con configuración fusionada	0 (0,00%)	66 (100,00%)	0 (0,00%)	66 (40,74%)
Total	2 (1,23%)	142 (87,65%)	18 (11,11%)	162 (100,00%)

Con respecto a las 84 moléculas monocíclicas (el 51,85% del total de la muestra estudiada), se identifica un carácter antiaromático más apreciable para moléculas

monocíclicas con tamaños pequeños (en ciclos conformados por 3 y 4 átomos), efecto que va reduciéndose a medida que incrementa el tamaño del ciclo. La rareza del comportamiento antiaromático se debe a que la combinación energética de $4n$ electrones π es desfavorable en términos de estabilidad, por lo que puede verse compensada en los compuestos de bajo tamaño de ciclo por la rigidez estructural asociada a la tensión conformacional. Por su parte, el carácter aromático en las moléculas monocíclicas tiende a ser más favorable en moléculas con tamaño de ciclos grandes (con 5 y 6 átomos), gracias a que esta extensión de la cadena cíclica favorece aún más la formación de enlaces conjugados (alternancia entre enlaces sencillos y dobles dentro del anillo).

Por su parte, las 78 moléculas policíclicas evaluadas (el 48,15% del total de la muestra estudiada) presentó dificultades en el establecimiento del carácter aromático y antiaromático mediante los algoritmos de canonicalización y kekulización (disponibles en la librería Rdkit, de acuerdo a lo mencionado en la Sección 2.2.2), debido principalmente a las limitaciones de la regla de Hückel en la determinación de los electrones π en enlaces conjugados dentro de cadenas cicladas.

No obstante, tanto el carácter aromático como antiaromático de las moléculas son relativamente poco frecuentes en las estructuras establecidas dentro de la muestra del estudio de aplicabilidad, representando el 11,11% y el 1,23%, respectivamente. Por tal razón, es de esperarse que la exista una baja tasa de compuestos aromáticos y antiaromáticos dentro de los compuestos solución generados por el algoritmo CAMD.

3.1.3 Aplicabilidad de grupos heterocíclicos establecidos en moléculas de uso industrial

Se evaluó la aplicación de los grupos heterocíclicos en 162 moléculas descritas en la Sección 2.2.5, cuyas estructuras moleculares están descritas en el **Anexo C**. Se identificaron los compuestos moleculares cuya representación completa fue posible empleando tanto los grupos heterocíclicos construidos como los grupos antes establecidos por los estudios base de Serrato [1] y Rodríguez [2], [3]. La **Tabla 3-14** reporta los compuestos cuya representación estructural fue completa, siendo posible evaluar 62 moléculas con los nuevos grupos heterocíclicos, un 38,27% del total de la muestra evaluada.

Tabla 3-14: Aplicabilidad de grupos heterocíclicos para las moléculas de uso industrial seleccionadas.

Heterocíclico	Nro. de moléculas evaluadas	Nro. de moléculas representables	Moléculas representadas
Monocíclicos con ciclos de 3 átomos	4 (2,47%)	2 (50,00%)	M2, M4.
Monocíclicos con ciclos de 4 átomos	9 (5,56%)	4 (44,44%)	M5, M8, M10, M11.
Monocíclicos con ciclos de 5 átomos	34 (20,99%)	12 (35,29%)	M14, M15, M16, M17, M18, M22, M24, M26, M29, M39, M42, M46.
Monocíclicos con ciclos de 6 átomos	37 (22,84%)	14 (86,48%)	M48, M49, M50, M52, M58, M64, M67, M70, M74, M76, M80, M82, M83, M84.
Policíclicos con configuración espiro	12 (7,41%)	5 (41,67%)	M87, M89, M93, M95, M96.
Policíclicos con configuración fusionada	66 (40,74%)	25 (37,88%)	M97, M98, M100, M104, M106, M107, M108, M110, M111, M116, M117, M118, M124, M125, M136, M137, M138, M139, M140, M141, M142, M143, M144, M151, M157.
Total	162 (100,00%)		62 (38,27%)

En concordancia con los resultados obtenidos en el desarrollo del Filtro III en la metodología descrita en la Sección 3.1.1, la **Tabla 3-14** se evidencia las limitaciones presentadas en los modelos de estimación de propiedades para evaluar moléculas de aplicabilidad significativa de manera industrial.

En este aspecto, las estructuras monocíclicas se observan mejor representadas por los modelos UNIFAC-DMD y Hukkerikar que en el caso de las estructuras policíclicas. Esto puede atribuirse al hecho de la adición de la contribución de tercer orden en el modelo de Hukkerikar, mejora significativamente la identificación de grupos funcionales en estructuras cíclicas, aromáticas y heterocíclicas.

Por otra parte, considerando el hecho de que UNIFAC-DMD no cuenta con modelo grupos estructurales desarrollados para este tipo de compuestos, habrá una reducción de las posibles moléculas representadas correctamente. De esta manera, los cálculos evaluados permiten trazar una primera visión de que grupos estructurales son necesarios incorporar para la evaluación más amplia de estructuras heterocíclicas.

Cabe destacar que, debido al tamaño reducido de la muestra de estructuras industriales tomadas, es posible que los resultados varíen para la aplicación de los modelos de estimación de propiedades para moléculas heterocíclicas (monocíclicas y policíclicas). No

obstante, se identifica una necesidad identificar parámetros experimentales validados para estructuras de este tipo, de manera que la tasa de representación de estos grupos en el modelo CAMD (y en otros tipos de modelos asociados a estimación de propiedades) aumente.

3.1.4 Impacto de propiedades estudiadas sobre las funciones objetivo

Para el desarrollo del impacto de las propiedades estimadas por el modelo de propiedades, se partió de una base de moléculas de 28.677 sustancias, de distintas familias químicas (alifáticas, aromáticas y heterocíclicas), las cuales fue tomada como referencia de Müller [242], [243]. Se tomó como punto de referencia el enfoque realizado por Rodríguez [244], [245], quien aplicó el modelo de estimación a las mismas propiedades determinadas con el modelo de Hukkerikar [66]–[68]. No obstante, a diferencia del estudio en cuestión, la base de datos de Müller [242], [243] proporcionó un conjunto de datos que integraba tanto moléculas alifáticas como aromáticas y heterocíclicas. Para el establecimiento de los valores de las propiedades experimentales, se emplearon información presente en las bases de datos PubChem [251], ZINC [235], [236] y DSSTox [237].

Partiendo de lo anterior, se realizó la estimación de las propiedades mencionadas en la Sección 2.2.7, integrando la tercera contribución de Hukkerikar [66]–[68], enfocado a estructuras heterocíclicas y policíclicas. Integrar esta contribución, reduce el error relativo medio de los datos pronosticados respecto a lo reportado en el estudio de Rodríguez [244], [245], quien aplicó el modelo de estimación de propiedades empleando solo la primera y segunda contribución. La **Tabla 3-15** presenta una comparativa entre el valor del error relativo medio reportado por Hukkerikar [66]–[68], el error relativo medio presentado por Rodríguez [244], [245] y el error obtenido en el presente estudio, para cada una de las propiedades evaluadas. Dado que la estimación de propiedades de Rodríguez [244], [245] fue realizada para el sistema LA/W, se presentan los errores estimados para este sistema en particular.

Tabla 3-15: Comparación del error relativo medio obtenido para el modelo de estimación de Hukkerikar [66]–[68], la evaluación realizada por Rodríguez [244], [245] y el presente trabajo.

Propiedad evaluada	Fuente del modelo de estimación de propiedades			Error Relativo Promedio (ARE)*		
	Hukkerikar [66]–[68]	Rodríguez [244], [245]	Este trabajo	Hukkerikar [66]–[68]	Rodríguez [244], [245]	Este trabajo
Punto de Fusión Normal	3.510	108	28.677	5,07%	12,12%	7,23%
Punto de Ebullición	5.183	108	28.677	1,44%	3,13%	2,56%
Energía de libre de Gibbs estándar	749	108	28.677	5,24%	5,25%	5,24%
Densidad	1.056	108	28.677	2,03%	3,92%	2,74%
LC50 Fatthed Minnow	809	108	28.677	0,48%	0,57%	0,52%
LC50 Daphnia Magna	320	108	28.677	0,49%	0,84%	0,69%
LD50 (rata)	5.995	108	28.677	0,35%	0,49%	0,41%
Solubilidad en agua	4.681	108	28.677	0,71%	0,86%	0,77%
Factor de bioconcentración	662	108	28.677	0,60%	0,44%	0,53%

De acuerdo con los errores obtenidos en la **Tabla 3-15**, se observa una reducción del error relativo promedio respecto a la evaluación realizada por Rodríguez [244], [245]. Esta reducción puede explicarse por la adición de la contribución de tercer orden, la cual tendrá un efecto correctivo en los resultados de las propiedades evaluadas para aquellas moléculas obtenidas cuya característica presente compuestos heterocíclicos. Adicionalmente, tomar un conjunto de estructuras más amplia reducir el efecto en el error de emplear una distinta fuente de base de datos respecto al modelo original.

Una vez observado el efecto de la adición de la tercera contribución del modelo de Hukkerikar [66]–[68] se evaluó el coeficiente de actividad empleando el modelo de UNIFAC-DMD. Considerando que dicho modelo no presentó diferencias respecto a los parámetros utilizados por Rodríguez [244], [245], es de esperarse que las diferencias observadas entre dicho estudio y el presente estén sujetas exclusivamente a la variabilidad del conjunto de datos empleado. Por tal razón, se consideró no comparar el error asociado a este modelo con el estudio de Rodríguez [244], [245].

Luego de evaluar cada una de las propiedades requeridas para la evaluación de las funciones objetivos estudiadas, descritas en la Sección 2.2.7, se procede a evaluar cada función para los dos sistemas en estudio (LA/W y AA/W), a partir de las 28.677 moléculas

evaluadas de la base datos de Müller [242], sin considerar clasificación en familias estructurales empleado representación SMARTS para filtrar cada estructura en la familia correspondiente. Considerando que una estructura puede contar con más de un grupo de los asignados en la Sección 2.2.10, la sumatoria de las moléculas evaluadas no corresponde al tamaño de la muestra establecida por la base de datos de Müller [242], [243].

Una vez las funciones objetivo han sido calculadas empleando las distintas propiedades estimadas, surge la inquietud asociada a si la relación de interacción entre propiedades y funciones objetivo cumple con la distribución normal. La **Tabla 3-16** presenta un resumen de la prueba de normalidad de Kolmogórov-Smirnov, siguiendo la tendencia de las relaciones entre variables dependientes (valores de las funciones objetivo) y las variables independientes (propiedades), siguiendo un modelo de correlaciones bivariadas. Se observa un comportamiento estadístico que no se ajusta a una distribución normal entre los datos, por lo que el uso de estadísticos no paramétricos para el estudio de las relaciones es válido y aplicable.

Tabla 3-16: Resultados de la prueba de normalidad de Kolmogórov-Smirnov para las relaciones entre propiedades y funciones objetivo. Sin segregación de familias funcionales.

FO	Indicador estadístico	Propiedades físicas					Propiedades ambientales				
		PF1 (T_b)	PF2 (T_f)	PF3 (G_f)	PF4 (V_m)	PF5 (ρ)	PA1 (LC_{50}^{FM})	PA2 (LC_{50}^{DM})	PA3 (LD_{50})	PA4 (W_s)	PA5 (BFC)
FO1: y	Z K-S*	0.251	0.184	0.069	0.145	0.175	0.138	0.041	0.266	0.136	0.169
	N	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677
FO2: ε	Z K-S*	0.021	0.044	0.068	0.056	0.199	0.093	0.111	0.04	0.64	0.142
	N	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677
FO3: K	Z K-S*	0.272	0.255	0.170	0.059	0.03	0.095	0.066	0.065	0.035	0.103
	N	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677
FO4: s	Z K-S*	0.075	0.042	0.012	0.044	0.074	0.035	0.005	0.112	0.0103	0.123
	N	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677
FO5: L	Z K-S*	0.108	0.122	0.139	0.332	0.158	0.109	0.272	0.030	0.003	0.053
	N	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677	28.677
	Significancia	0.000	0.000	0.019	0.006	0.001	0.009	0.009	0.0039	0.000	0.010
	Significancia	0.0223	0.009	0.035	0.078	0.000	0.0018	0.0196	0.0001	0.001	0.001
	Significancia	0.007	0.000	0.000	0.001	0.0260	0.0155	0.051	0.0504	0.0001	0.015
	Significancia	0.0415	0.005	0.009	0.003	0.009	0.012	0.039	0.0072	0.073	0.011
	Significancia	0.0178	0.026	0.006	0.059	0.015	0.000	0.011	0.074	0.006	0.085

Datos con significancia asintótica (bilateral),

* Con corrección de significación de Lilliefors.

Z K-S: Estadístico Z de la prueba de Kolmogórov-Smirnov.

La prueba de Kolmogórov-Smirnov implica normalidad si $Z > \text{Significancia} > 0.05$.

Convención de color:

Celdas color verde = La relación PF-FO si cuenta con una distribución normal.

Celdas color naranja = La relación PF-FO no cuenta con una distribución normal.

Una vez establecida la no normalidad de las relaciones entre propiedades y funciones objetivo, se realizó el análisis estadístico para las parejas de variables propiedad – función

objetivo relacionados, los cuales son presentados por la **Tabla 3-17**. Se determinaron los coeficientes de correlación de Spearman ($\rho_{Spearman}$), el valor de la significancia, tamaño de efecto y poder estadístico, siguiendo los criterios mencionados en la Sección 2.2.10 (valor de α de 0,05 bilateral, mismos niveles cuantitativos del criterio de Cohen y reporte del poder estadístico para determinar qué tan generalizables son los resultados). Se empleó la misma escala de colores para el análisis del impacto de las propiedades, de acuerdo con la **Figura 2-16**.

Este estudio estadístico entre las propiedades físicas/ambientales estudiadas y las funciones objetivo, permite inferir el grado de afectación de un objetivo específico (selectividad, poder de extracción, pérdida de solvente, impacto ambiental) respecto a las características intrínsecas de la propiedad de un solvente cualesquiera. Este resultado permite establecer si la propiedad tiene un efecto muy grande, mediano o pequeño, tanto directa como inversamente proporcional en el comportamiento de las funciones de interés y permite presentar un indicio de la viabilidad de aplicación de cierto solvente, en función de las dimensiones y valores de sus propiedades. Queda como recomendación para trabajos posteriores ahondar el efecto de las propiedades sobre las funciones objetivo, haciendo segmentación entre distintos grupos funcionales.

Tabla 3-17: Resultados del estudio estadístico no paramétrico aplicado para para las relaciones entre propiedades y funciones objetivo. Sin segregación de familias funcionales.

FO	Índice estad.	Propiedades físicas									
		PF1 (T_b)		PF2 (T_r)		PF3 (G_r)		PF4 (V_m)		PF5 (ρ)	
		LA	AA	LA	AA	LA	AA	LA	AA	LA	AA
FO1 (γ)	$\rho_{Spearman}$	-0,795**	0,467**	0,514**	-0,081**	0,062**	-0,405**	-0,391**	-0,222**	0,595**	0,474**
	Signif.	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
	p	0,897	0,467	0,514	0,081	0,062	0,405	0,391	0,222	0,595	0,474
	$1 - \beta$	0,999	1	1	0,980	0,788	1	1	1	1	1
FO2 (ϵ)	$\rho_{Spearman}$	-0,097**	0,452**	0,634**	-0,081**	0,103**	-0,009	-0,386**	-0,057**	0,219**	0,044**
	Signif.	0,001	0,001	0,001	0,001	0,001	0,561	0,001	0,001	0,001	0,004
	p	0,097	0,497	0,583	0,081	0,103	0,009	0,386	0,057	0,219	0,044
	$1 - \beta$	0,999	1	1	0,980	0,999	0,625	1	0,680	1	0,509
FO3 (K)	$\rho_{Spearman}$	-0,289**	-0,397**	0,143**	-0,602**	-0,208**	-0,252**	0,122**	0,394**	-0,129**	-0,309**
	Signif.	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
	p	0,289	0,397	0,143	0,602	0,208	0,252	0,122	0,394	0,129	0,309
	$1 - \beta$	1	1	1	1	1	1	0,999	1	0,999	1
FO4 (S)	$\rho_{Spearman}$	0,018	0,001	-0,525**	-0,158**	-0,171**	0,059**	-0,012	-0,218**	-0,458**	-0,277**
	Signif.	0,243	0,986	0,001	0,001	0,001	0,001	0,437	0,001	0,001	0,001
	p	0,018	0,001	0,525	0,158	0,171	0,059	0,012	0,218	0,458	0,277
	$1 - \beta$	0,517	0,986	1	1	1	0,726	0,564	1	1	1
FO5 (L)	$\rho_{Spearman}$	0,021	-0,406**	-0,214**	0,535**	0,257**	0,453**	0,456**	0,216**	-0,006	0,138**
	Signif.	0,163	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,686	0,001
	p	0,021	0,406	0,214	0,535	0,257	0,453	0,456	0,216	0,006	0,138
	$1 - \beta$	0,498	1	1	1	1	1	1	1	0,708	1

Cont. **Tabla 3-17**

FO	Índice estad.	Propiedades ambientales									
		PA1 (LC_{50}^{FM})		PA2 (LC_{50}^{DM})		PA3 (LD_{50})		PA4 (W_5)		PA5 (BFC)	
		LA	AA	LA	AA	LA	AA	LA	AA	LA	AA
FO1 (γ)	$\rho_{Spearman}$	-0,795**	0,467**	0,514**	-0,081**	0,062**	-0,405**	-0,391**	-0,222**	0,595**	0,474**
	Signif.	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
	p	0,897	0,467	0,514	0,081	0,062	0,405	0,391	0,222	0,595	0,474
	$1 - \beta$	0,999	1	1	0,980	0,788	1	1	1	1	1
FO2 (ϵ)	$\rho_{Spearman}$	-0,097**	0,452**	0,634**	-0,081**	0,103**	-0,009	-0,386**	-0,057**	0,219**	0,044**
	Signif.	0,001	0,001	0,001	0,001	0,001	0,561	0,001	0,001	0,001	0,004
	p	0,097	0,497	0,583	0,081	0,103	0,009	0,386	0,057	0,219	0,044
	$1 - \beta$	0,999	1	1	0,980	0,999	0,625	1	0,680	1	0,509
FO3 (K)	$\rho_{Spearman}$	-0,289**	-0,397**	0,143**	-0,602**	-0,208**	-0,252**	0,122**	0,394**	-0,129**	-0,309**
	Signif.	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
	p	0,289	0,397	0,143	0,602	0,208	0,252	0,122	0,394	0,129	0,309
	$1 - \beta$	1	1	1	1	1	1	0,999	1	0,999	1
FO4 (δ)	$\rho_{Spearman}$	0,018	0,001	-0,525**	-0,158**	-0,171**	0,059**	-0,012	-0,218**	-0,458**	-0,277**
	Signif.	0,243	0,986	0,001	0,001	0,001	0,001	0,437	0,001	0,001	0,001
	p	0,018	0,001	0,525	0,158	0,171	0,059	0,012	0,218	0,458	0,277
	$1 - \beta$	0,517	0,986	1	1	1	0,726	0,564	1	1	1
FO5 (L)	$\rho_{Spearman}$	0,021	-0,406**	-0,214**	0,535**	0,257**	0,453**	0,456**	0,216**	-0,006	0,138**
	Signif.	0,163	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,686	0,001
	p	0,021	0,406	0,214	0,535	0,257	0,453	0,456	0,216	0,006	0,138
	$1 - \beta$	0,498	1	1	1	1	1	1	1	0,708	1

** La correlación es significativa a nivel 0.01 (bilateral).

* La correlación es significativa a nivel 0.05 (bilateral).

Convención de color:

- Verde oscuro: Relaciones entre variables significantes (<0.01), con tamaño de efecto muy grande ($p > 0.5$), generalizable y predecible ($(1 - \beta) > 0.8$).

- Verde claro: Relaciones entre variables significantes (<0.01), con tamaño de efecto grande ($0.3 < p < 0.5$), generalizable y predecible ($(1 - \beta) > 0.8$).

- Naranja: Relaciones entre variables significantes (<0.01), con tamaño de efecto medio ($0.1 < p < 0.3$), generalizable y predecible ($(1 - \beta) > 0.8$).

- Azul: Relaciones entre variables significantes (<0.01), con tamaño de efecto medio ($p < 0.1$), generalizable y predecible ($(1 - \beta) > 0.8$).

- Gris: Relaciones entre variables que no son significantes o no generalizables.

3.2 Resultados asociados a la etapa de diseño

La finalidad de la presente sección es presentar los resultados asociados a la construcción de la herramienta CAMD, el comportamiento del algoritmo de construcción molecular, evaluación de propiedades y optimización multiobjetivo empleados. Se verifica el cumplimiento de restricciones para los casos de estudio LA/W y AA/W. Adicionalmente, se realizan las comparaciones respecto a los estudios de la línea base (Serrato [1] y Rodríguez [2], [3]), y la elección de los solventes a evaluar experimentalmente para la validación del modelo. Los resultados descritos en esta sección están enmarcados en el objetivo 3 del presente trabajo.

Con respecto al comportamiento del algoritmo genético, la Sección 3.2.1 reporta los resultados asociados a la indicadores de desempeño del algoritmo MOHAEA aplicado para los casos de estudio LA/W y AA/W, teniendo en cuenta las consideraciones descritas en las Secciones 2.2.11, 2.2.12 y 2.2.13. En el marco de las optimizaciones multiobjetivo, se evalúa tanto la convergencia del algoritmo como la diversidad de las soluciones. Adicionalmente, la Sección 3.2.2 reporta el análisis del desempeño de los operadores

genéticos específicos empleados en las corridas experimentales del modelo, partiendo de las definiciones descritas para los operadores dadas en la Sección 2.2.12. Adicionalmente, la Sección 3.2.2 describe el cumplimiento de las restricciones definidas en la Sección 2.2.1, a lo largo de las corridas experimentales efectuadas.

Con respecto al comportamiento del modelo multiobjetivo propuesto respecto a los estudios de línea base de Serrato [1] y Rodríguez [2], [3] la Sección 3.2.4 establece comparación entre los resultados obtenidos en 50 corridas experimentales del modelo CAMD, para ambos casos de estudio. Cabe resaltar que, aunque los sistemas evaluados son los mismos, la obtención de una molécula u otra como posible solvente candidato, depende del tipo de función (o funciones) objetivo consideradas y la cantidad de restricciones especificadas en la corrida experimental. Se resaltan las principales diferencias y ventajas del presente modelo respecto a los anteriores estudiados, de acuerdo con los resultados obtenidos. Finalmente, la Sección 3.2.5 presenta el proceso de elección de solventes a evaluar experimentalmente, utilizando los criterios de elección experimental establecidos en la Sección 2.3.3.

3.2.1 Convergencia y desempeño del algoritmo evolutivo

Al algoritmo MOHAEA, empleado durante el desarrollo de la metodología CAMD propuesta en este trabajo, se le aplicaron indicadores de convergencia y desempeño que permiten determinar el rendimiento del algoritmo evolutivo, durante las ejecuciones del programa CAMD (50 experimentos). Se evaluaron medidas de tendencia central y dispersión para los datos obtenidos, siendo la media (**Figura 3-5**) y la desviación estándar (**Figura 3-6**) computada para cada valor objetivo a lo largo de cada una de las generaciones.

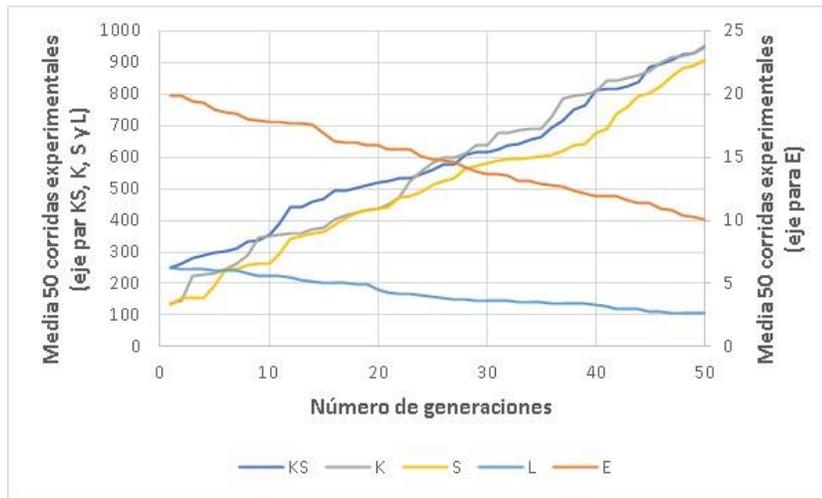


Figura 3-5: Media de las funciones objetivos empleadas a lo largo de las generaciones, promediada para 50 ejecuciones del programa.

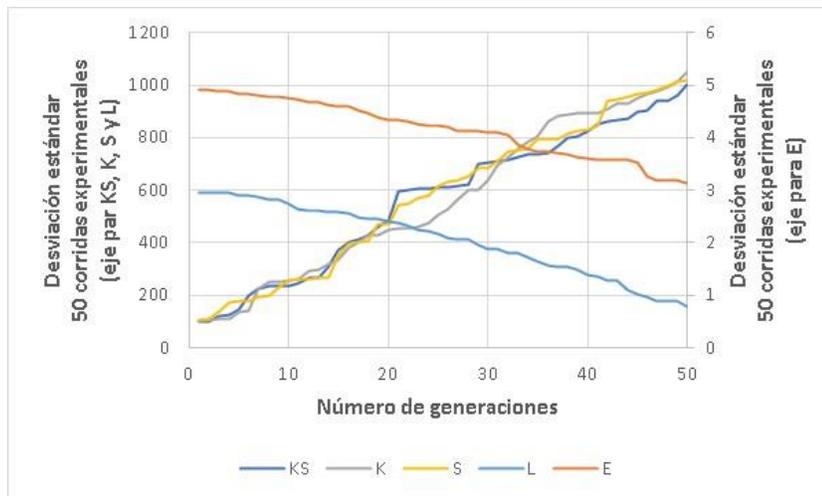


Figura 3-6: Desviación estándar de las funciones objetivos empleadas a lo largo de las generaciones, promediada para 50 ejecuciones del programa.

Tal como se observa en las **Figura 3-5** y **Figura 3-6**, es posible evidenciar que, a medida que transcurren las generaciones, el valor del mejor solvente (o en este caso particular, lo mejores solventes, tienden al óptimo matemático, es decir al frente óptimo de Pareto. Adicionalmente, el hecho de que la desviación estándar disminuya a lo largo de las generaciones da un indicio de la convergencia del algoritmo de optimización hacia el conjunto de soluciones óptima y permite establecer que tanta diversidad tienen las soluciones entre sí.

Cabe aclarar que los resultados anteriores corresponden al estudio realizado para el conjunto de funciones objetivo 1 (Y vs \mathcal{E} vs K), los cuales corresponden al estudio de las funciones de optimización de referencia, dadas por los trabajos de Serrato y Rodríguez. De acuerdo con lo que se mencionará a profundidad posteriormente en la Sección 3.2.3, el estudio estará dado por conjuntos de funciones objetivos, de acuerdo con el énfasis que se desea analizar. Para cada uno de los anteriores enfoques, se obtuvieron resultados similares a los observados en las **Figura 3-5** y **Figura 3-6**.

3.2.2 Desempeño de operadores genéticos y cumplimiento de restricciones

Los operadores genéticos implementados en la metodología CAMD realizado en este trabajo fueron evaluadas en función de las tasas de los operadores genéticos de cada generación a lo largo de 50 ejecuciones del programa, usando la metodología descrita en la Sección 2.2.12. El promedio de las tasas de operadores genéticos para cada operador, a lo largo de las generaciones, es descrito por la **Tabla 3-18**. Considerando que algunos operadores genéticos han sido introducidos en los trabajos previos de Serrato [1] y Rodríguez [2], [3], se efectúa la comparación del comportamiento reportado de los operadores de estos trabajos respecto a los obtenidos en el actual estudio.

De acuerdo con los operadores genéticos contemplados en el presente estudio, podemos observar cómo, conformen avancen las generaciones, los operadores van teniendo una tendencia a distribuir la tasa de aceptación de cada operación en función de las necesidades de convergencia del modelo. Esto se debe a la versatilidad del algoritmo evolutivo MOHAEA en ajustar las tasas según sea necesario. En este aspecto, podemos indicar que esta característica particular del algoritmo evolutivo desarrollado por Gómez [252], [253] permite asegurar la certeza de poder explorar de forma completa el espacio de búsqueda de moléculas.

Tabla 3-18: Desempeño de las tasas de los operadores genéticos a lo largo de las generaciones, durante 50 ejecuciones del programa.

Operador genético	Tasa de aplicación promedio		
	Máximo	Mínimo	Promedio
Mutación	0,196	0,176	0,181
Cruce	0,134	0,123	0,127
Eliminación de grupo	0,147	0,104	0,132
Extensión de cadena	0,153	0,129	0,138
Cierre de cadena	0,194	0,148	0,153
Cambio de cadena	0,164	0,124	0,146
Ruptura de anillo	0,174	0,117	0,154
Ciclación aromática	0,123	0,116	0,121

Con respecto al cumplimiento de restricciones del problema de optimización, realizadas durante la ejecución del algoritmo evolutivo, se establecen los resultados de las restricciones obtenidas en la **Tabla 3-19**. Se realizó el comparativo con los porcentajes de error en las mejores soluciones presentados por Rodríguez [2], [3] en sus experimentos, asociados al sistema ácido láctico-agua (LA/W), donde es posible evidenciar que el problema desarrollado en el presente estudio cuenta con menor tasa de cumplimiento respecto a su equivalente desarrollado por Rodríguez [2], [3]. Esto es posible justificarlo considerando la incorporación de estructuras heterocíclicas, las cuales difícilmente cumplen restricciones como pérdida de solvente y disponibilidad en el mercado. Adicional a esto, el hecho de que el problema desarrollado sea de multiobjetivo (con tres objetivos simultáneos) aumenta significativamente la complejidad de una molécula candidato de cumplir con varias restricciones de forma simultánea. Por tanto, es posible que estructuras que cumplieran con las restricciones en el estudio de Rodríguez [2], [3], ya no cumplan con el mismo tipo de restricciones en el presente estudio.

Tabla 3-19: Cumplimiento de las restricciones en las mejores soluciones.

Nombre de la restricción	Estudio de Rodríguez [2], [3]		Este estudio	
	Compuestos que cumplen restricciones	% de cumplimiento	Compuestos que cumplen restricciones	% de cumplimiento
Punto de fusión	495	74,10%	326	70,14%
Punto de ebullición	641	95,96%	604	86,45%
Energía libre de Gibbs estándar	667	99,85%	685	97,26%
Perdida del solvente	169	25,30%	121	17,35%
Disponibilidad en el mercado	538	80,54%	439	65,37%

3.2.3 Frentes de Pareto obtenidos por el diseño CAMD

La presente sección describe el desarrollo realizado en los distintos frentes de Pareto obtenidos y los análisis asociados a los mismos. Considerando que la problemática estudiada corresponde a una optimización multiobjetivo, se evaluaron las siguientes opciones de representación de tres funciones objetivo de las descritas en la Sección 2.2.1.

- *Conjunto de funciones objetivo 1 (Y vs \mathcal{E} vs K):* Este conjunto de funciones objetivo, corresponde, de forma simplificada, al sistema de funciones propuesto por los estudios de Serrato [1] y Rodríguez [2], [3]. Este frente de Pareto se emplea como referencia para los estudios comparativos con las otras funciones.
- *Conjunto de funciones objetivo 2 (\mathcal{S} vs \mathcal{E} vs K):* Este conjunto de funciones objetivo busca evaluar la estrategia de separación desde un enfoque operativo (Y , con K y \mathcal{S}), sin interesar la cantidad de solvente perdido como producto de la extracción en el sistema (L).
- *Conjunto de funciones objetivo 3 (\mathcal{S} vs L vs K):* Este conjunto de funciones objetivo pretende evaluar la estrategia de separación ignorando los efectos que tiene la selección de un solvente u otro en el impacto ambiental (\mathcal{E}).
- *Conjunto de funciones objetivo 4 (\mathcal{E} vs L vs S):* Este conjunto de funciones objetivo pretende analizar los efectos del sistema, sin considerar que tanto poder de

extracción tiene la selección del solvente. Desde el punto de vista práctico, la aplicación de este conjunto de funciones no cuenta con ningún interés operativo, dado que no se cumple con el motivo intrínseco de ser de la operación de separación. Por tal razón, no fue considerado en el estudio.

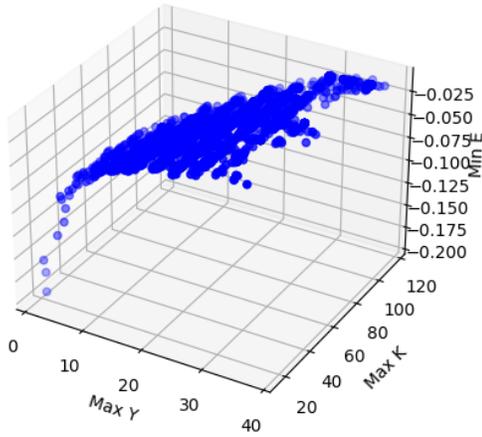
- *Conjunto de funciones objetivo 5 (\mathcal{E} vs L vs K):* Este conjunto de funciones objetivo contempla analizar los efectos de la aplicación de un solvente en la extracción sin considerar que tan selectivo es para la separación del compuesto de interés. Desde el punto de vista práctico, la aplicación de este conjunto de funciones no cuenta con ningún interés, dado que se pierde el fin por el cual se utiliza el solvente de extracción. Teniendo en cuenta lo anterior, este conjunto no fue considerado en el estudio.

De acuerdo con lo anterior, se establece analizar el comportamiento del conjunto de funciones objetivo 1, 2 y 3, para ambos sistemas evaluados (LA/W y AA/W). Con el fin de poder filtrar un poco los resultados, respecto a estructuras candidato que cuenten con alto costo de consecución, se limitó la presentación de los resultados a un 30% de los solventes más baratos, de acuerdo con las bases de datos empleadas, descritas en la Sección 2.2.9.

Con respecto al cumplimiento de restricciones, se tuvo en consideración las mismas consideraciones de los estudios de Serrato [1] y Rodríguez [2], [3], ajustando los límites de las variables en función de los requerimientos de las estructuras heterocíclicas ingresadas al modelo. Dado que este tipo de estructuras cuentan con tamaños estructurales superiores a las moléculas, fue necesario suavizar ligeramente los límites en las variables de optimización, con el fin de que las estructuras heterocíclicas no fueran descartadas rápidamente por el modelo de optimización.

Las **Figura 3-7** a **Figura 3-8** presentan los frentes de Pareto obtenidos durante el desarrollo de las corridas experimentales del algoritmo CAMD, para los sistemas LA/W y AA/W, indicando el enfoque de tres objetivos desarrollado en la realización de la figura. Para cualquier enfoque, los datos de los frentes de Pareto corresponden al valor promedio de las 50 corridas experimentales desarrolladas.

Frente de Pareto Y vs E vs K - Sistema LA/W



Frente de Pareto Y vs E vs K - Sistema AA/W

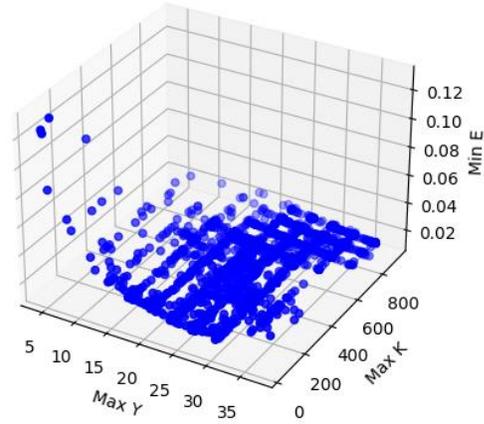
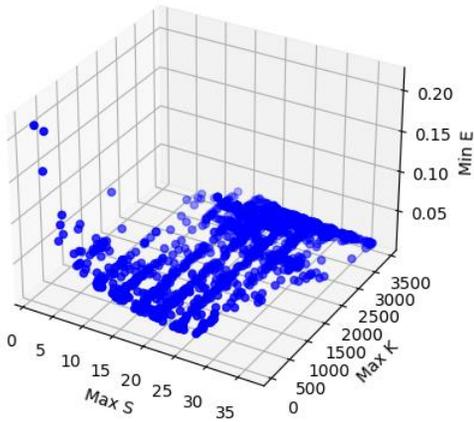


Figura 3-7: Frente de Pareto para el sistema LA/W y AA/W, para el conjunto de funciones objetivo 1 (Y vs E vs K).

Frente de Pareto S vs E vs K - Sistema AA/W



Frente de Pareto S vs E vs K - Sistema LA/W

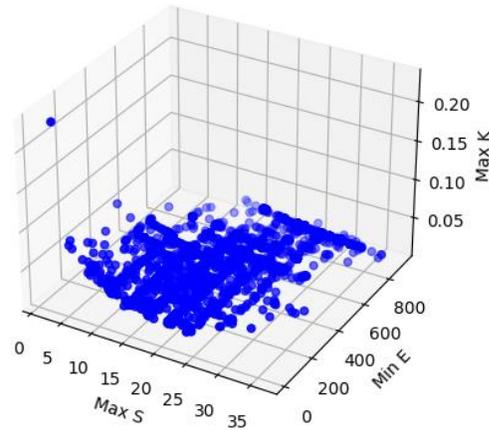
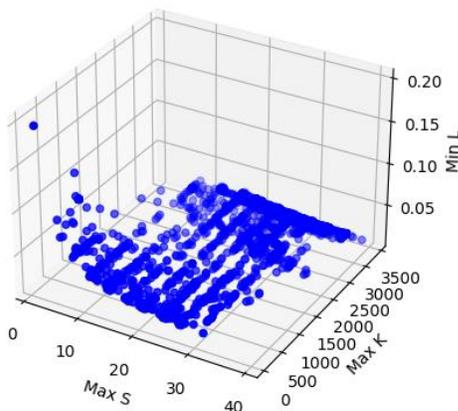


Figura 3-8: Frente de Pareto para el sistema LA/W y AA/W, para el conjunto de funciones objetivo 2 (S vs E vs K).

Frente de Pareto S vs L vs K - Sistema LA/W



Frente de Pareto S vs L vs K - Sistema AA/W

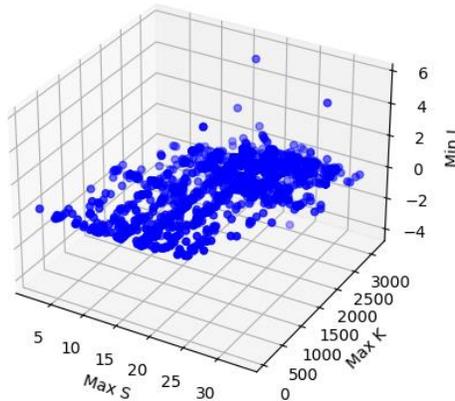


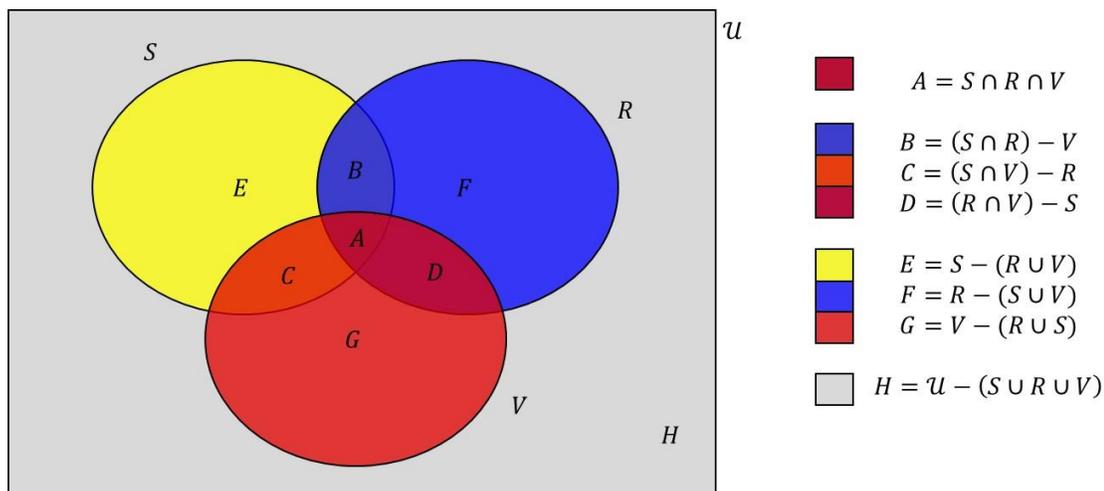
Figura 3-9: Frente de Pareto para el sistema LA/W, para el conjunto de funciones objetivo 3 (S vs L vs K).

Obtener los frentes de Pareto para cada terna de funciones objetivos evaluadas permite proporcionar una visión completa del conjunto de soluciones óptimas, permitiendo tomar decisiones para la selección de solventes, analizando el compromiso entre los diferentes objetivos. Adicionalmente, para el caso de las estructuras cuyos grupos funcionales están involucrados en los compuestos solución del frente de Pareto, permite identificar qué tipo de estructuras son favorables como solventes, facilitando la separación mediante extracción.

3.2.4 Comparación de resultados respecto a estudios de línea base

Los resultados obtenidos por el algoritmo CAMD desarrollado en este trabajo son comparados respecto a las metodologías de un solo objetivo de los estudios de la línea base de Serrato [1] y Rodríguez [2], [3], con el fin de destacar las mejoras introducidas en la optimización. Con el fin de facilitar la comparación de resultados, la **Figura 3-10** presenta los conjuntos conformados por los compuestos solución obtenidos por el estudio de Serrato [1] (conjunto S), el estudio de Rodríguez [2], [3] (conjunto R) y el presente trabajo (conjunto V). Adicionalmente, se identifica los subconjuntos de pertenencia de una molécula solución en términos de los estudios en donde fue obtenida como respuesta.

Figura 3-10: Conjuntos solución de moléculas obtenidas en los estudios de Serrato [1], Rodríguez [2], [3] y el presente trabajo.



Partiendo de los subconjuntos antes descritos, la **Tabla 3-20** reporta los solventes candidatos presentes en los estudios del modelo de Serrato [1] y el modelo de Rodríguez [2], [3], en los sistemas LA/W y AA/W. Esto permite establecer la incidencia de la aplicación de una metodología de un solo objetivo y/o multiobjetivo en la resolución del problema de selección de solvente. La principal diferencia entre los tres estudios radicó en el conjunto de funciones objetivos seleccionadas, ya que las restricciones y los métodos de estimación de propiedades se mantuvieron de manera similar. Esto con el fin de establecer un punto de comparación aceptable entre los estudios anteriores.

Es importante resaltar el hecho de que, al ser tres estudios con características diferentes, en términos de la cantidad de funciones objetivos y forma de relacionarse entre sí, se espera que los resultados obtenidos en los tres estudios sean muy diferentes entre sí. Esto se debe a que los espacios exploratorios de búsqueda de las estructuras son diferentes en función del objetivo establecido.

Tabla 3-20: Conjunto de solventes óptimos, obtenidos por Serrato [1] y Rodríguez [2], [3], para los sistemas LA/W y AA/W.

Nro	Nombre	SMILES	PM (g/mol)	Formula	Estudio
1	Ethane-1,1,1-tricarbaldehyde	O=CC(C)(C=O)C=O	114,09	C5H6O3	Serrato
2	3-Formyl-6-methoxy-3,6-dimethylheptanoic acid	O=CC(C)(CC=O)C=O	128,12	C6H8O3	Serrato
3	Propanedial	O=CCC=O	72,06	C3H4O2	Serrato

4	2-Formyl-2-methyl-5-oxopentanoic acid	<chem>CC(CCC=O)(C=O)C(=O)O</chem>	158,15	C7H10O4	Serrato
5	2-Methyl-1,3-dioxopropan-2-yl acetate	<chem>O=CC(C)(C=O)OC(=O)C</chem>	144,12	C6H8O4	Serrato
6	(Methoxymethyl)(methyl)propanedial	<chem>O=CC(C)(C=O)COC</chem>	130,14	C6H10O3	Serrato
7	2,5-Dimethylhexane-1,2,5-tricarbaldehyde	<chem>CC(C)(CCC(C)(CC=O)C=O)C=O</chem>	198,25	C11H18O3	Serrato
8	2-Methyl-1,4-dioxobutan-2-yl acetate	<chem>CC(C=O)(CC=O)OC(=O)C</chem>	158,15	C7H10O4	Serrato
9	2-Methyl-1,6-dioxohexan-2-yl pentanoate	<chem>O=C(CCCC)OC(C)(CCCC=O)C=O</chem>	228,28	C12H20O4	Serrato
10	2-Formyl-2,4,4-trimethyl-5-oxopentyl acetate	<chem>CC(CC(C)(C)C=O)(COC(=O)C)C=O</chem>	241,25	C11H18O4	Serrato
11	3,6,6-Trimethyl-1,7-dioxoheptan-3-yl acetate	<chem>CC(=O)CC(C)(CCC(C)(C)C=O)CC=O</chem>	226,31	C13H22O3	Serrato
12	Methylpropanedial	<chem>O=CC(C)C=O</chem>	86,08	C4H6O2	Serrato
13	3-(Acetyloxy)-3-methyl-4-oxobutanoic acid	<chem>CC(C=O)(CC(O)=O)OC(=O)C</chem>	174,15	C7H10O5	Serrato
14	2-Methoxy-2-methylpentanedial	<chem>CC(C=O)(CCC=O)OC</chem>	144,16	C7H12O3	Serrato
15	2-Formyl-2-methyl-3-oxopropanoic acid	<chem>O=CC(C)(C=O)C(=O)O</chem>	130,09	C5H6O4	Serrato
16	2-(Acetyloxy)-2-methyl-5-oxopentanoic acid	<chem>CC(CCC=O)(OC(=O)C)C(=O)O</chem>	188,17	C8H12O5	Serrato

Cont. **Tabla 3-20.**

Nro	Nombre	SMILES	PM (g/mol)	Formula	Estudio
17	2-Methyl-6-oxo-1-propoxyhexan-2-yl acetate	<chem>O=C(C)OC(C)(COCCC)CCCC=O</chem>	230,3	C12H22O4	Serrato
18	2-Methyl-1-(3-oxopropoxy)hexan-2-yl acetate	<chem>O=CCCOCC(C)(CCCC)OC(C)=O</chem>	230,3	C12H22O4	Serrato
19	2-Methylbutanedial	<chem>CC(CC=O)C=O</chem>	100,11	C5H8O2	Serrato
20	Acetyl propylcarbamate	<chem>CCCNC(=O)OC(=O)C</chem>	145,15	C6H11NO3	Serrato
21	2-(3-Methoxy-3-methylbutyl)-2-methylbutanedial	<chem>O=CCC(C)(CCC(C)(C)OC)C=O</chem>	200,27	C11H20O3	Serrato
22	2-(Acetyloxy)-6-oxohexan-2-yl pentanoate	<chem>O=C(CCCC)OC(C)(CCCC=O)OC(=O)C</chem>	258,31	C13H22O5	Serrato
23	2-Formyl-5-methoxy-2,5-dimethylhexyl acetate	<chem>COC(C)(C)CCC(C)(COC(C)=O)C=O</chem>	230,3	C12H22O4	Serrato
24	4-Butoxypentane-1,4-diyl diacetate	<chem>O=C(C)OC(C)(CCCOC(C)=O)OCCCC</chem>	260,32	C13H24O5	Serrato
25	Methoxypropanedial	<chem>O=CC(C=O)OC</chem>	102,08	C4H6O3	Serrato
26	2,4-Dimethylpentane-1,2,4-tricarbaldehyde	<chem>O=CCC(C)(CC(C)(C)C=O)C=O</chem>	184,23	C10H16O3	Serrato
27	3-(Acetyloxy)-3-methylbutyl 3-oxopropanoate	<chem>O=C(C)OC(C)(C)CCOC(=O)CC=O</chem>	216,23	C10H16O5	Serrato
28	2-(But-3-en-1-yl)-2-methoxy-3-methylpentanedial	<chem>O=CCC(C)C(CCC=C)(C=O)OC</chem>	198,25	C11H18O3	Serrato
29	3-(3,3-Dimethyl-4-oxobutoxy)-3-oxopropanoic acid	<chem>CC(C)(C=O)CCOC(=O)CC(=O)O</chem>	202,2	C9H14O5	Serrato
30	3-Formyl-6-methoxy-3,6-dimethylheptanoic acid	<chem>OC(=O)CC(C)(CCC(C)(C)OC)C=O</chem>	216,27	C11H20O4	Serrato
31	2-[(2-Methylhexan-2-yl)oxy]ethyl 3-oxopropanoate	<chem>O=C(OCCOC(C)(C)CCCC)CC=O</chem>	230,3	C12H22O4	Serrato
32	2-(Ethylamino)-2-oxoethyl acetate	<chem>CCNC(=O)COC(=O)C</chem>	145,15	C6H11NO3	Serrato
33	2-Formyl-2-methylbutyl 3-(acetyloxy)-3-methylbutanoate	<chem>CCC(C)(COC(=O)CC(C)(C)OC(C)=O)C=O</chem>	258,31	C13H22O5	Serrato
34	1-(2-Methoxyethoxy)-2-methylhexan-2-yl acetate	<chem>COCCOCC(C)(CCCC)OC(C)=O</chem>	232,31	C12H24O4	Serrato

35	Methoxyacetaldehyde	O=CCOC	74,07	C3H6O2	Serrato
36	Dimethylpropanedial	O=CC(C)(C)C=O	100,11	C5H8O2	Serrato
37	2-(Hydroxymethyl)-5-methoxy-2,5-dimethylhexanal	CC(C)(CCC(C)(CO)C=O)OC	188,26	C10H20O3	Serrato
38	3-Formyl-3,5,5-trimethyl-6-oxohexanoic acid	CC(CC(C)(C)C=O)(CC(=O)O)C=O	200,23	C10H16O4	Serrato
39	Hydroxy(methyl)propanedial	O=CC(C)(O)C=O	102,08	C4H6O3	Serrato
40	Methoxy(methyl)propanedial	O=CC(C)(C=O)OC	116,11	C5H8O3	Serrato
41	2-(Hydroxymethyl)-2,5,5-trimethylhexanedial	CC(C=O)(CCC(C)(C)C=O)CO	186,24	C10H18O3	Serrato
42	2,2-Dimethoxypropanal	O=CC(C)(OC)OC	118,13	C5H10O3	Serrato
43	4-(Ethylamino)-4-oxobutyl acetate	O=C(C)OCCCC(=O)NCC	173,2	C8H15NO3	Serrato
44	6-Butoxy-6-hydroxy-2-methylheptan-2-yl acetate	OC(C)(CCCC(C)(C)OC(C)=O)OCCCC	260,36	C14H28O4	Serrato
45	3-(Acetyloxy)-3,5,5-trimethyl-6-oxohexanoic acid	O=C(C)OC(C)(CC(C)(C)C=O)CC(O)=O	230,25	C11H18O5	Serrato

Cont. **Tabla 3-20.**

Nro	Nombre	SMILES	PM (g/mol)	Formula	Estudio
46	Hydroxy(methyl)propanedial	<chem>O=CC(C)(O)C=O</chem>	102,09	C4H6O3	Serrato
47	Butanamidomethyl acetate	<chem>O=C(C)OCNC(=O)CCC</chem>	159,18	C7H13NO3	Serrato
48	2-[(2,2-Dimethyl-3-oxopropoxy)methyl]-2-methylbutanoic acid	<chem>CC(CC)(COCC(C)(C)C=O)C(=O)O</chem>	216,27	C11H20O4	Serrato
49	1-[3-(Acetyloxy)-3-methylpentanamido]-2-methylpropan-2-yl acetate	<chem>CC(C)(CNC(=O)CC(C)(CC)OC(C)=O)OC(=O)C</chem>	287,35	C14H25NO5	Serrato
50	2-hydroxy-4,4,6,6-tetramethyl-7-oxoheptan-2-yl acetate	<chem>O=C(C)OC(O)(C)CC(C)(C)CC(C)(C)C=O</chem>	244,32	C13H24O4	Serrato
51	6-oxoheptanal	<chem>CC(=O)CCCCC=O</chem>	128,17	C7H12O2	Rodríguez
52	4-Oxopentyl acetate	<chem>CC(=O)CCCOC(=O)C</chem>	144,17	C7H12O3	Rodríguez
53	2,2-dichlorobutan-1-ol	<chem>C(CC(Cl)(Cl)CO</chem>	143,01	C4H8Cl2O	Rodríguez
54	4,4-Dichlorobutanal	<chem>C(CC(Cl)Cl)C=O</chem>	140,99	C4H6Cl2O	Rodríguez
55	2,6-Heptanedione	<chem>CC(=O)CCCC(=O)C</chem>	128,17	C7H12O2	Rodríguez
56	3-Methylidenehept-1-en-4-one	<chem>CCCC(=O)C(=C)C=C</chem>	124,18	C8H12O	Rodríguez
57	2,3-Heptanedione	<chem>CCCCC(=O)C(=O)C</chem>	128,17	C7H12O2	Rodríguez
58	2,2-Dichloroethanol	<chem>C(C(Cl)Cl)O</chem>	114,96	C2H4Cl2O	Rodríguez
59	Nona-5,7-Dien-2-One	<chem>CC=CC=CCCC(=O)C</chem>	138,21	C9H14O	Rodríguez
60	3,3-Dichloropropanol	<chem>C(CO)C(Cl)Cl</chem>	128,98	C3H6Cl2O	Rodríguez
61	Heptanedial	<chem>C(CCC=O)CCC=O</chem>	128,17	C7H12O2	Rodríguez
62	6-oxoheptanal	<chem>CC(=O)CCCCC=O</chem>	128,17	C7H12O2	Rodríguez
63	N-(2,2-dichloroethyl)-N-methylbutanamide	<chem>CCCC(=O)N(C)CC(Cl)Cl</chem>	191,70	C9H18ClNO	Rodríguez
64	Methyl Valerate	<chem>CCCCC(=O)OC</chem>	116,16	C6H12O2	Rodríguez
65	Adipaldehyde	<chem>C(CCC=O)CC=O</chem>	114,14	C6H10O2	Rodríguez
66	a-keto heptanal	<chem>CCCCCC(=O)C=O</chem>	128,17	C7H12O2	Rodríguez
67	4-methyl-4-pentenal	<chem>C=C(C)CCC=O</chem>	98,14	C6H10O	Rodríguez
68	2-(4-Chlorobutoxy)ethan-1-ol	<chem>C(CCOCCO)CCl</chem>	152,62	C6H13ClO2	Rodríguez
69	Oct-6-en-4-one	<chem>CC=CCC(=O)CCC</chem>	126,2	C8H14O	Rodríguez
70	2,2-dichloropentanal	<chem>CCCC(C=O)(Cl)Cl</chem>	155,02	C5H8Cl2O	Rodríguez
71	Methyl 4-hydroxybutanoate	<chem>COC(=O)CCCCO</chem>	118,13	C5H10O3	Rodríguez
72	Butaneperoxoic Acid	<chem>CCCC(=O)OO</chem>	104,1	C4H8O3	Rodríguez
73	3,3-Dichloropropanal	<chem>C(C=O)C(Cl)Cl</chem>	126,97	C3H4Cl2O	Rodríguez
74	4,4-Dichlorobutanal	<chem>C(CC(Cl)Cl)C=O</chem>	140,99	C4H6Cl2O	Rodríguez
75	5-methylhex-5-enal	<chem>C=C(C)CCCC=O</chem>	112,17	C7H12O	Rodríguez
76	3-Methylenepentanal	<chem>CCC(=C)CC=O</chem>	98,14	C6H10O	Rodríguez
77	2-Hexanone	<chem>CCCCC(=O)C</chem>	100,16	C6H12O	Rodríguez
78	4-Acetoxybutyraldehyde	<chem>CC(=O)OCCCC=O</chem>	130,14	C6H10O3	Rodríguez

A partir de las moléculas identificadas en los estudios anteriores es posible mencionar que no se encontraron moléculas que se generaran, de forma simultánea, en este estudio y en los estudios de Serrato [1] y/o Rodríguez [2], [3] (intersecciones de los conjuntos presentados en la **Figura 3-10**). Esto puede darse, como se mencionó previamente, producto del carácter multiobjetivo del presente estudio, respecto a los anteriores. Por este motivo, las contribuciones específicas de una molécula candidato pueden encontrarse sujetas a la existencia dentro del frente de Pareto óptimo, lo cual puede incidir en la incorporación de una molécula candidato en la lista final de moléculas soluciones.

Teniendo en cuenta lo anterior, el presente estudio obtuvo las moléculas descritas en la **Tabla 3-21**. Con respecto a los grupos funcionales de las moléculas candidato generadas, existe una marcada tendencia hacia aldehídos, cetonas, aminas cuaternarias y estructuras alifáticas de baja extensión de cadena. Entre los hallazgos importantes del presente estudio se encuentra la verificación matemática de la poca viabilidad de las estructuras aromáticas/heterocíclicas para resolver este tipo de sistemas. No obstante, es necesario resaltar que el presente estudio puede variar en sus resultados a futuro, en la medida de los cambios paramétricos y la robustez de los modelos de estimación de propiedades, para estos de este tipo de moléculas.

Por otra parte, es importante que el enfoque de los tres trabajos en cuestión (los estudios de Serrato [1], Rodríguez [2], [3] y este estudio) se encuentran más allá de las diferencias presentadas en la cantidad de funciones objetivo. Tanto el estudio de Serrato [1], como el de Rodríguez [2], [3] buscan solucionar el problema de selección de solventes con un enfoque clásico, usando una ecuación basada en peso (caso del estudio de Serrato [1]) o una función bi-objetivo (caso del estudio de Rodríguez [2], [3]). A pesar de que el presente estudio pueda generar grupos estructurales similares, según la configuración aplicada al modelo, el nivel de complejidad y enfoque del programa MORITA desarrollado busca ser el punto de partida para estudios más complejos en el área del diseño molecular.

Tabla 3-21: Listado parcial de moléculas candidato obtenidas en el presente estudio, para los sistemas LA/W y AA/W.

Sistema	Nro	Nombre	SMILES	PM (g/mol)	Formula
LA/W	1	2-Pentanol	CCCC(C)O	88,15	C5H12O
	2	Ácido beta-hidroxiisovalerico	C(C(C)(C)O)C(=O)O	118,13	C5H10O3
	3	Ácido isovalérico	CC(C)CC(=O)O	102,13	C5H10O2
	4	Ácido isobutírico	CC(C)C(=O)O	88,11	C4H8O2
	5	2-Methyl-2-Butanol	CCC(C)(C)O	88,15	C5H12O
	6	Ácido 3-Hidroxi-butírico	C(C(C)O)C(=O)O	104,1	C4H8O3
	7	Ácido 3-Hidroxi-3methylpentanoico	C(C(CC)(C)O)C(=O)O	132,16	C6H12O3
	8	Ácido Trimetilacético	CC(C)(C)C(=O)O	102,13	C5H10O2
	9	Ácido 2-Metilbutanoico	CCC(C)C(=O)O	102,13	C5H10O2
	10	t-butanol	CC(C)(C)O	74,12	C4H10O
	11	Ácido 2,2-Dimetilbutírico	CCC(C)(C)C(=O)O	116,16	C6H12O2
AA/W	1	Ácido Propiónico	CCC(=O)O	74,08	C3H6O2
	2	Ácido Propanedioico	C(C(=O)O)C(=O)O	104,06	C3H4O4
	3	Ácido 2-oxobutírico	CCC(=O)C(=O)O	102,09	C4H6O3
	4	Ácido 2-oxopropanoico	CC(=O)C(=O)O	88,06	C3H4O3
	5	Ácido 3-oxobutanoico	CC(=O)CC(=O)O	102,09	C4H6O3
	6	Ácido 3-Hidroxi-butanoico	C(C(C)O)C(=O)O	104,1	C4H8O3
	7	Ácido Piválico	CC(C)(C)C(=O)O	102,13	C5H10O2
	8	Ácido Neopentanoico	C(C)(C)(C)C(=O)O	102,13	C5H10O2
	9	Ácido beta-hidroxiisovalerico	C(C(C)(C)O)C(=O)O	118,13	C5H10O3
	10	t-butanol	CC(C)(C)O	74,12	C4H10O
	11	Ácido 3-Hidroxi-butírico	C(C(C)O)C(=O)O	104,1	C4H8O3

3.2.5 Candidatos de extracción obtenidos para validación experimental

El listado de los posibles solventes candidatos obtenidos por el algoritmo CAMD para el sistema LA/W es reportado por la **Tabla 3-22**, realizando previa evaluación e identificación de los criterios mencionados en la Sección 2.3.3. Por su parte, la **Tabla 3-23** presenta el listado de los posibles solventes candidatos obtenidos por el algoritmo CAMD para el sistema AA/W.

Para cada sustancia se asignó un valor de 1 por cada criterio que cumpliera la molécula candidato, para luego hacer la sumatoria de los criterios cumplidos por cada molécula

candidato. En caso de igualdad entre dos o más moléculas, se efectuó el desempate respectivo por orden de criterios (primero C1, luego C2 y así sucesivamente).

Tabla 3-22: Mejores solventes obtenidos por el modelo CAMD para el sistema LA/W.

Sistema LA/W								
#	SMILES molécula	Criterio por considerar						Total
		C1	C2	C3	C4	C5	C6	
1	<chem>CCCC(C)O</chem>	1	1	0	1	1	1	5
2	<chem>C(C(C)(C)O)C(=O)O</chem>	1	0	1	1	0	1	5
3	<chem>CC(C)CC(=O)O</chem>	0	1	0	1	0	0	2
4	<chem>CC(C)C(=O)O</chem>	1	1	0	0	1	1	4
5	<chem>CCC(C)(C)O</chem>	0	1	0	1	0	1	3
6	<chem>C(C(C)O)C(=O)O</chem>	1	0	0	0	0	1	2
7	<chem>C(C(C(C)(C)O)C(=O)O</chem>	1	1	0	1	1	0	4
8	<chem>CC(C)(C)C(=O)O</chem>	0	0	1	0	0	1	2
9	<chem>CCC(C)C(=O)O</chem>	0	0	0	0	0	1	1
10	<chem>CC(C)(C)O</chem>	0	1	1	1	0	1	4
11	<chem>CCC(C)(C)C(=O)O</chem>	0	0	1	1	1	1	4

Tabla 3-23: Mejores solventes obtenidos por el modelo CAMD para el sistema AA/W.

Sistema LA/W								
#	SMILES molécula	Criterio por considerar						Total
		C1	C2	C3	C4	C5	C6	
1	<chem>CCC(=O)O</chem>	1	1	1	0	0	0	3
2	<chem>C(C(=O)O)C(=O)O</chem>	0	0	0	0	1	1	2
3	<chem>CCC(=O)C(=O)O</chem>	0	0	1	0	1	0	2
4	<chem>CC(=O)C(=O)O</chem>	0	1	0	0	0	1	2
5	<chem>CC(=O)CC(=O)O</chem>	0	0	1	0	0	0	1
6	<chem>C(C(C)O)C(=O)O</chem>	1	1	0	0	1	0	3
7	<chem>CC(C)(C)C(=O)O</chem>	0	1	0	1	1	1	4
8	<chem>C(C)(C)(C)C(=O)O</chem>	0	1	1	1	0	1	4
9	<chem>C(C(C)(C)O)C(=O)O</chem>	1	1	0	0	0	0	2
10	<chem>CC(C)(C)O</chem>	0	0	1	1	1	0	3
11	<chem>C(C(C)O)C(=O)O</chem>	1	0	0	0	0	0	1

Una vez obtenida las listas de las mejores solventes obtenidas para ambos sistemas, es posible seleccionar las mejores moléculas candidatos para desarrollar la etapa de post-diseño. No obstante, existieron ciertas consideraciones que requirieron establecer cierta variación a la selección del solvente, descritas a continuación.

- El desarrollo del modelo de disponibilidad de moléculas, asociado a las bases de datos ZINC y EPA DSSTox, descritas en la Sección 2.2.9, se encuentra más relacionado con el mercado químico estadounidense que con el mercado de disponibilidad química colombiana. Por esta razón, muchas de las estructuras que se obtienen en el modelo pueden aparecer como disponibles en el mercado, cuando en la práctica colombiana se encuentran restringidos, no disponibles o requieren mayores recursos para su consecución y transporte.
- Sumado a lo anterior, el presente estudio se realizó durante un periodo excepcional originado por la Pandemia de COVID-19 (2020-2022), la cual dificultó significativamente la consecución de solventes raros, a pesar de que se encuentren comercialmente disponibles. Producto de esta situación, las cadenas de suministro de productos químicos nacionales e internacionales fueron afectadas significativamente, lo cual restringió aún más la obtención de los solventes óptimos obtenidos por el modelo CAMD.
- Por otra parte, considerando el interés del estudio en realizar la validación del modelo CAMD construido a partir de moléculas obtenidas, es necesario contemplar la limitación de recursos económicos y disponibilidad de tiempo para conseguir estructuras complejas para realizar la validación experimental del modelo CAMD.

De acuerdo con lo anterior, se seleccionaron tres sustancias por sistema evaluado para validar el modelo matemático de la herramienta CAMD construida, empleando la metodología experimental de la etapa de Post-diseño, descrita en la Sección 2.4. Estas moléculas, al encontrarse dentro del espacio de soluciones, son identificables por el modelo y, por ende, se obtienen parámetros K y S teóricos a partir de la ejecución del modelo CAMD. Tanto para el caso del sistema LA/W y AA/W, las sustancias seleccionadas fueron: Acetato de Etilo (AE), Ciclohexano (Cy-C6), Metil Isobutil cetona (MIBK).

3.3 Resultados asociados a la etapa de post-diseño

La finalidad de la presente sección es presentar los resultados de la validación de la metodología CAMD desarrollada, aplicándola a dos sistemas: 1.) ácido láctico/agua (LA/W) y 2.) ácido acético/agua (AA/W), sistemas de gran interés por su presencia en múltiples efluentes de procesos fermentativos. Para realizar la validación experimental del modelo

en ambos sistemas, las Secciones 3.3.1 (Sistema LA/W) y 3.3.2 (Sistema AA/W) emplearon la metodología de validación experimental descrita en la Sección 2.4, usando los métodos analíticos descritos en las Secciones 2.4.1, 2.4.2 y 2.4.3.

Por su parte, la comparación de los indicadores comparativos descritos en la Sección 2.4.4 fue realizada entre el valor teórico obtenido mediante la herramienta CAMD y los datos de validación experimental. Los resultados descritos en esta sección se enmarcan en el objetivo 4 del presente trabajo.

Teniendo en cuenta las consideraciones mencionadas respecto a los candidatos de extracción obtenidos por el modelo, se tomaron para validar el modelo tres solventes identificados en los dos sistemas evaluados: Acetato de etilo (AE), Ciclohexano (CyC6) y Metil isobutil cetona (MIBK)

3.3.1 Validación experimental: Sistema Ácido láctico-Agua (LA/W)

A partir de los compuestos identificados como potenciales solventes de extracción para el sistema LA/W, descritos en la Sección 3.3.1, se procedió a realizar las extracciones respectivas de manera experimental. Posterior de la realización de estas, y una vez dado un tiempo de estabilización de 12h con el fin de asegurar haber alcanzado el equilibrio líquido-líquido, se separaron las fases acuosa y orgánica, obteniendo así una serie de alícuotas a las cuales se deseaba identificar la concentración de ácido y agua remanente. Esto con el fin de asegurar el balance de materia dado por las ecuaciones (2-28) a (2-30).

Con respecto a la concentración final de ácido en la fase acuosa ($C_{aq,f}^{acid}$) se obtuvieron los valores a partir de la identificación del área en corridas cromatográficas de alta eficiencia (HPLC), para la fase acuosa, una vez separada en la extracción, siguiendo los métodos establecidos en la Sección 2.4.1. La **Tabla 3-24** presenta los resultados para las alícuotas obtenidas para el sistema LA/W, una vez alcanzado el equilibrio líquido-líquido y separada la fase acuosa. Considerando la cantidad de ácido presente en la muestra original, se identificó la pérdida de ácido en la fase acuosa promedio para cada solvente estudiado, obteniendo que el acetato de etilo es el solvente con el cual se pierde menos ácido láctico en la fase acuosa producto de la extracción.

Tabla 3-24: Comportamiento del Ácido Láctico en función del solvente candidato adicionado, en fase acuosa, medido por HPLC.

Solvente	Ensayo	C_0^{acid} [g/L]	$C_{aq,f}^{acid}$ [g/L]
Acetato de Etilo (AE)	LA/W-1	18,02	5,719
	LA/W-2		
	LA/W-3		
Ciclohexano (Cy-C6)	LA/W-1	18,02	8,109
	LA/W-2		
	LA/W-3		
Metil isobutil cetona (MIBK)	LA/W-1	18,02	6,492
	LA/W-2		
	LA/W-3		

Con respecto a la cantidad de agua en la fase orgánica ($C_{org,f}^w$), se realizó la titulación potenciométrica usando el método de Karl Fischer, siguiendo los métodos establecidos en la Sección 2.4.2. La **Tabla 3-25** presenta los resultados obtenidos en la determinación del agua en la fase orgánica resultante, una vez alcanzado el equilibrio y separada esta fase. Considerando la cantidad de agua presente en la muestra original, se identificó que cantidad de agua tiende a la fase orgánica promedio para cada solvente estudiado, obteniendo que el Acetato de Etilo es el solvente que genera mayor cantidad de agua en la fase orgánica, producto de la extracción.

Tabla 3-25: Agua obtenida en la fase orgánica, en función del solvente candidato adicionado, para el sistema LA/W.

Molécula solvente empleada	Ensayo	% Agua en fase orgánica*	Concentración final de agua ($C_{org,f}^w$) [g/L]	Indicadores estadísticos
Acetato de Etilo (AE)	1	3,685%	0,4717	$\bar{x} = 3,121$ $\sigma = 0,00514$ $r = 0,00468$
	2	2,999%	0,3839	
	3	2,679%	0,3429	
Ciclohexano (Cy-C6)	1	0,585%	0,0749	$\bar{x} = 0,362$ $\sigma = 0,00193$ $r = 0,0228$
	2	0,255%	0,0326	
	3	0,246%	0,0315	
Metil isobutil cetona (MIBK)	1	1,947%	0,2492	$\bar{x} = 1,821$ $\sigma = 0,00266$ $r = 0,00391$
	2	1,515%	0,1939	
	3	2,000%	0,2560	

* Valor promedio obtenido mediante titulación de Karl Fischer de medidas por triplicado.

Como criterio de aceptación de los resultados, la diferencia entre los resultados sucesivos debe ser menor al valor de r , donde $r = 0.0032(\bar{x}^{0.333})$, siendo \bar{x} el valor promedio de las mediciones.

Finalmente, se determinaron las concentraciones de ácido en la fase orgánica ($C_{org,f}^{acid}$) y de agua en la fase acuosa ($C_{aq,f}^w$), mediante titulaciones ácido-base, empleando los métodos descritos en la Sección 2.4.3, presentadas en la **Tabla 3-26**. De acuerdo con lo anterior, se obtiene que la adición del solvente MIBK generará una mayor cantidad de ácido láctico en la fase orgánica, mientras que la adición del solvente Ciclohexano ocasionará mayor cantidad de agua en la fase orgánica.

Tabla 3-26: Reparto de ácido en fase orgánica y agua en fase acuosa, en función del solvente candidato adicionado, para el sistema LA/W.

Molécula solvente empleada	Ensayo	Concentración inicial				Concentración final (fase orgánica)		Concentración final (fase acuosa)	
		C_0^{acid} [g/L]	C_0^w [g/L]	Vol. LA/W [mL]	Vol. S [mL]	$C_{org,f}^{acid}$ [g/L]	$C_{org,f}^w$ [g/L]	$C_{aq,0}^w$ [g/L]	$C_{aq,f}^{acid}$ [g/L]
Acetato de Etilo (AE)	LA/W-1	72,064	3,2	87,084	37,9	183,71	5,995	2,58	5,7190
	LA/W-2	72,064	3,2	87,084	37,9	183,71	4,969	3,57	5,7190
	LA/W-3	72,064	3,2	87,084	37,9	183,71	4,792	3,74	5,7190
Ciclohexano (Cy-C6)	LA/W-1	72,064	3,2	90,54	34,5	186,76	8,109	0,00	8,1090
	LA/W-2	72,064	3,2	90,54	34,5	158,28	7,505	0,15	8,1090
	LA/W-3	72,064	3,2	90,54	34,5	158,28	6,506	0,69	8,1090
Metil isobutil cetona (MIBK)	LA/W-1	72,064	3,2	95,866	29,1	172,01	6,492	2,69	6,4920
	LA/W-2	72,064	3,2	95,866	29,1	222,55	5,454	3,37	6,4920
	LA/W-3	72,064	3,2	95,866	29,1	222,55	6,816	2,48	6,4920

A partir de los resultados promedio obtenidos para las concentraciones finales de ácido láctico y agua determinadas en cada una de las fases, se verificó la validez del balance de materia descrito por las ecuaciones (2-28) a (2-30), estableciendo una comparación entre las condiciones iniciales (parte izquierda de la ecuación) y condiciones finales luego de realizada la extracción con un solvente candidato específico (parte derecha de la ecuación), encontrándose un error del 4,76% de error, lo cual se encuentra dentro del margen aceptable asociado.

Teniendo en cuenta todas las consideraciones experimentales anteriormente mencionadas, es posible asegurar que el modelo teórico CAMD construido es capaz de obtener solventes candidatos apropiados para el sistema LA/W, considerando propiedades termodinámicas, ambientales y otras consideraciones.

Además, es posible asegurar que los tres solventes obtenidos (Acetato de Etilo, Ciclohexano y Metil isobutil cetona) son válidos desde un punto de vista experimental para realizar una extracción aceptable del sistema LA/W, considerando la cantidad de solvente

empleada, la pérdida del solvente candidato generado posterior a la separación y la cantidad de agua remanente en la fase orgánica. No obstante, para realizar la comparación de resultados del modelo CAMD respecto a los resultados experimentales obtenidos, es necesario establecer indicadores tales como el coeficiente de reparto (K) y la selectividad (S) de la extracción, lo cual se analiza con en la Sección 3.3.3.

3.3.2 Validación experimental: Sistema Ácido acético-Agua (AA/W)

Se siguió el mismo procedimiento de cuantificación de concentraciones, realizado previamente para el sistema LA/W en la Sección 3.3.1, considerando los potenciales solventes de extracción para el sistema AA/W, descritos en la **Tabla 3-22** y siguiendo los métodos descritos por las Secciones 2.4.1, 2.4.2, 2.4.3.

Con respecto a la concentración final de ácido en la fase acuosa ($C_{aq,f}^{acid}$) la **Tabla 3-27** presenta los resultados para el sistema AA/W, una vez alcanzado el equilibrio líquido-líquido y separadas la fase acuosa. Se determinó que el Ciclohexano es el solvente con el cual se pierde menos ácido acético en la fase acuosa producto de la extracción, considerando la cantidad de ácido presente en la muestra original.

Tabla 3-27: Comportamiento del Ácido Acético en función del solvente candidato adicionado, en fase acuosa, medido por HPLC.

Solvente	Ensayo	C_0^{acid} [g/L]	Área	Área Prom.	$C_{aq,f}^{acid}$ [g/L]
Acetato de Etilo (AE)	AA/W-1	12,01	9,523	7,931	10,209
	AA/W-2		7,030		
	AA/W-3		7,240		
Ciclohexano (Cy-C6)	AA/W-1	12,01	6,734	7,494	9,571
	AA/W-2		9,005		
	AA/W-3		6,743		
Metil isobutil cetona (MIBK)	AA/W-1	12,01	9,050	9,939	13,141
	AA/W-2		11,807		
	AA/W-3		8,959		

La **Tabla 3-28** presenta los resultados obtenidos en la determinación del agua en la fase orgánica resultante ($C_{org,f}^w$), una vez alcanzado el equilibrio y separada esta fase. Considerando la cantidad de agua presente en la muestra original, se identificó que cantidad de agua tienda a la fase orgánica promedio para cada solvente estudiado,

obteniendo que el Acetato de Etilo es el solvente que genera mayor cantidad de agua en la fase orgánica, producto de la extracción.

Tabla 3-28: Agua obtenida en la fase orgánica, en función del solvente candidato adicionado, para el sistema AA/W.

Molécula solvente empleada	Ensayo	% Agua en fase orgánica*	Concentración final de agua ($C_{org,f}^w$) [g/L]	. Indicadores estadísticos
Acetato de Etilo (AE)	1	2,887%	0,3695	$\bar{x} = 3,109$ $\sigma = 0,0211$ $r = 0,00467$
	2	3,132%	0,4009	
	3	3,308%	0,4234	
Ciclohexano (Cy-C6)	1	0,181%	0,0232	$\bar{x} = 0,402$ $\sigma = 0,00398$ $r = 0,0236$
	2	0,861%	0,1102	
	3	0,163%	0,0209	
Metil isobutil cetona (MIBK)	1	1,527%	0,1955	$\bar{x} = 1,664$ $\sigma = 0,00129$ $r = 0,00379$
	2	1,682%	0,2153	
	3	1,784%	0,2284	

* Valor promedio obtenido mediante titulación de Karl Fischer de medidas por triplicado.

Como criterio de aceptación de los resultados, la diferencia entre los resultados sucesivos debe ser menor al valor de r , donde $r = 0.0032(\bar{x}^{0.3333})$, siendo \bar{x} el valor promedio de las mediciones.

Finalmente, se determinaron las concentraciones finales de ácido en la fase orgánica ($C_{org,f}^{acid}$) y de agua en la fase acuosa ($C_{aq,f}^w$), descritas por la **Tabla 3-29**, para cada uno de los solventes candidatos empleados en la extracción, una vez alcanzado el equilibrio y separadas las fases acuosa y orgánica. De acuerdo con lo anterior, se obtiene que la adición del solvente MIBK generará una mayor cantidad de ácido acético en la fase orgánica, mientras que la adición del solvente Acetato de Etilo ocasionará mayor cantidad de agua en la fase acuosa.

Tabla 3-29: Reparto de ácido en fase orgánica y agua en fase acuosa, en función del solvente candidato adicionado, para el sistema AA/W.

Molécula solvente empleada	Ensayo	Concentración inicial				Concentración final (fase orgánica)		Concentración final (fase acuosa)	
		C_0^{acid} [g/L]	C_0^w [g/L]	Vol. AA/W [mL]	Vol. S [mL]	$C_{org,f}^{acid}$ [g/L]	$C_{org,f}^w$ [g/L]	$C_{aq,0}^w$ [g/L]	$C_{aq,f}^{acid}$ [g/L]
Acetato de Etilo (AE)	AA/W-1	48,040	3,2	78,483	46,5	71,860	0,3695	7,9310	2,76
	AA/W-2	48,040	3,2	78,483	46,5	71,805	0,4009	7,9310	2,74
	AA/W-3	48,040	3,2	78,483	46,5	71,765	0,4234	7,9310	2,73
Ciclohexano (Cy-C6)	AA/W-1	48,040	3,2	70,574	54,4	73,655	0,0232	7,4940	2,55
	AA/W-2	48,040	3,2	70,574	54,4	73,474	0,1102	7,4940	2,51
	AA/W-3	48,040	3,2	70,574	54,4	73,660	0,0209	7,4940	2,56
Metil isobutil cetona (MIBK)	AA/W-1	48,040	3,2	77,812	47,2	79,172	0,1955	9,9390	3,38
	AA/W-2	48,040	3,2	77,812	47,2	79,135	0,2153	9,9390	3,36
	AA/W-3	48,040	3,2	77,812	47,2	79,111	0,2284	9,9390	3,36

A partir de los resultados promedio obtenidos para las concentraciones finales de ácido acético y agua determinadas en cada una de las fases, se verificó la validez del balance de materia descrito por las ecuaciones (2-28) a (2-30), encontrándose un error del 3,45% de error, lo cual se encuentra dentro del margen aceptable asociado.

Teniendo en cuenta todas las consideraciones experimentales anteriormente mencionadas, es posible asegurar que el modelo teórico CAMD construido es también capaz de obtener solventes candidatos apropiados para el sistema AA/W, considerando propiedades termodinámicas, ambientales y otras consideraciones. Además, es posible asegurar que los tres solventes obtenidos (Acetato de etilo, Ciclohexano y Metil Isobutil Cetona) son válidos desde un punto de vista experimental para realizar una extracción aceptable del sistema AA/W, considerando la cantidad de solvente empleada, la pérdida del solvente candidato generado posterior a la separación y la cantidad de agua remanente en la fase orgánica. No obstante, para realizar la comparación de resultados del modelo CAMD respecto a los resultados experimentales obtenidos, es necesario establecer indicadores tales como el coeficiente de reparto (K) y la selectividad (S) de la extracción, lo cual se analiza con mayor profundidad en la Sección 3.3.3.

3.3.3 Comparación de parámetros experimentales teóricos obtenidos

Una vez establecidas las validaciones experimentales de los casos de estudio, se obtuvieron los valores de K y S experimentales, empleando las ecuaciones (2-31) y (2-32).

Considerando que el modelo CAMD construido genera como resultado el valor de K y S teórico para la molécula solvente seleccionada, es posible realizar la determinación de un error relativo del modelo CAMD construido respecto los datos experimentales obtenidos. Los resultados de los parámetros K y S experimentales y teóricos para las moléculas solventes seleccionadas, junto con el error asociado al modelo CAMD, se presentan en la **Tabla 3-30**.

Tabla 3-30: Comparación de parámetros experimentales y teóricos obtenidos por el modelo CAMD.

Caso de estudio	Solvente estudiado	Parámetro			Error paramétrico		
		K^{CAMD}	K^{exp}	K_{error}	S^{CAMD}	S^{exp}	S_{error}
LA/W	Acetato de Etilo (AE)	57,936	32,12	0,804	1,065	1,28	0,832
	Ciclohexano (CyC6)	33,210	19,52	0,701	2,390	6,51	0,367
	Metil isobutil cetona (MIBK)	42,702	34,28	0,246	1,248	2,41	0,518
AA/W	Acetato de Etilo (AE)	15,289	9,06	0,688	0,045	0,13	0,343
	Ciclohexano (CyC6)	11,552	9,83	0,175	0,006	0,01	0,560
	Metil isobutil cetona (MIBK)	12,106	7,96	0,521	0,040	0,06	0,668

De acuerdo con los resultados obtenidos en la **Tabla 3-30**, se determina que el modelo CAMD propuesto se ajusta aceptablemente tanto para el sistema LA/W como para el sistema AA/W. De esta manera, se concluye que el modelo CAMD propuesto responde con una tasa de error absoluto menor al 2%, lo que indica que es capaz de estimar de forma apropiada la identificación de un solvente de extracción para estos sistemas en específico.

4. Conclusiones y recomendaciones

4.1 Conclusiones

Del presente proyecto de tesis se pueden concluir lo siguiente:

- Se adicionó satisfactoriamente el conjunto de grupos heterocíclicos (monocíclicos, policíclicos espiro y policíclicos fusionados) al modelo CAMD preexistente, haciendo modificaciones en la forma de búsqueda de subestructuras y adaptaciones en la representación de estructuras aromáticas y heterocíclicas, partiendo de la representación SMILES. Esto con el fin de aumentar la representación de compuestos aromáticos y heterocíclicos en los modelos preexistentes de Serrato [1] y Rodríguez.
- El desarrollo de la metodología de tres filtros permite identificar de una forma matemática a todas las estructuras válidas no isomorfas, que cumplen con los criterios de valencia neutra y que pueden ser representados apropiadamente por los modelos de estimación de propiedades integrados en el modelo CAMD (Hukkerikar y UNIFAC-DMD). Esto es útil para identificar que estructuras son posible de evaluar usando el modelo CAMD.
- Se obtuvo un impacto bajo de estructuras aromáticas (11,11%) y antiaromáticos (1,23%), dentro la muestra de estructuras de moléculas de uso industrial recopilada. Esto implica que los efectos de aromaticidad y anti-aromaticidad tendrán efectos despreciables dentro de las moléculas cíclicas incorporadas al modelo CAMD.
- Se realizó la aplicación de los modelos de propiedades de Hukkerikar y UNIFAC-DMD en la muestra de estructuras heterocíclicas de uso industrial, determinando un 38,27% de posibilidad de representación por ambos modelos. Esto indica que las estructuras heterocíclicas se encuentran restringidas en su uso en el modelo CAMD, debido a la baja proporción de estructuras evaluables, respecto al conjunto universal posible.
- El presente estudio logró establecer el impacto de las propiedades físicas y ambientales respecto a cada una de las funciones objetivo que se encuentran involucradas, integrando un análisis estadístico profundo, enfocado en la

determinación de la normalidad de las relaciones bivariadas propiedad-función objetivo. Se pudo establecer el tamaño del efecto y las variables significantes para cada una de las 50 relaciones propiedad-función objetivo.

- Se desarrolló el problema de optimización multiobjetivo para la selección de solventes para los sistemas ácido láctico/agua (LA/W) y ácido acético/agua (AA/W). Se consideraron tres conjuntos de soluciones a analizar para cada uno de los sistemas (Y vs \mathcal{E} vs K , S vs \mathcal{E} vs K y S vs L vs K), observándose los efectos de cada uno de ellos en el Frente de Pareto reportado.
- Las características del algoritmo de optimización MOHAEA permitieron que adaptar el problema de optimización original para representar estructuras aromáticas y heterocíclicas más complejas, con la adición de 2 operadores genéticos específicos para este tipo de sustancias.
- El desarrollo del problema de optimización multiobjetivo llevó a la construcción de un software (MORITA), el cual integra tanto la evaluación de grupos aromáticos y heterocíclicos, como la aplicación del problema de optimización multiobjetivo. El desarrollo de este software permitirá explorar escenarios de otros sistemas de forma más simplificada.
- A pesar de que los sistemas son los mismos, las restricciones son similares y los algoritmos de optimización son algoritmos genéticos evolutivos, los resultados encontrados no son comparables directamente respecto a los resultados proporcionados por Serrato y Rodríguez. Esto se debe al carácter multiobjetivo del problema planteado.
- A pesar de que el modelo de disponibilidad de moléculas es bastante amplio y descrito, los resultados de disponibilidad de moléculas en el mercado local (Colombia) es significativamente diferente al reportado como referencia (Estados Unidos). Esto, sumado a otras dificultades logísticas (tales como la pandemia de COVID-19 y la crisis en el sector transporte) dificultaron validar experimentalmente las estructuras del frente de Pareto.
- Los resultados experimentales obtenidos permitieron corroborar que los solventes seleccionados no se encuentran dentro del frente de Pareto óptimo, considerando el hecho de que existe una pérdida importante de ácido en la fase acuosa. No obstante, las mediciones experimentales permitieron corroborar los resultados de los parámetros K y S teóricos y experimentales.

4.2 Recomendaciones y trabajo futuro

Entre las recomendaciones a considerar para futuros trabajos relacionados a la temática de esta tesis se encuentran:

- Aplicar la metodología de optimización multiobjetivo propuesta a más casos de estudio experimentales, reportados en literatura y/o validados experimentalmente, con el fin de asegurar confiabilidad en el modelo desarrollado en este trabajo.
- Integrar al modelo de optimización multiobjetivo la posibilidad de generar solventes de extracción mixtos, consistentes en mezclas multicomponentes (solvente + cosolvente). Para ello, es necesario agregar al modelo propuesto diversas reglas de mezclado para realizar la estimación de propiedades físicas, ambientales y aspectos económicos de los solventes de tipo mezcla, a partir de las propiedades de sustancia pura empleadas.
- Resolver el problema de optimización en más de tres funciones objetivo (*Many-Objective*), contemplando aspectos tanto de diseño de producto como de diseño del proceso (metodología CAMD+CAPD), aplicando una estrategia holística que contemple modificaciones de los solventes candidato en función de las condiciones operativas del proceso. Es necesario integrar una estrategia de visualización de los resultados, dadas las limitaciones gráficas impuestas por la maldición de la dimensionalidad.
- Efectuar la evaluación de la convergencia y desempeño del algoritmo MOHAEA, aplicada en los casos de estudio desarrollados, con métricas de desempeño que permitan complementar el análisis del rendimiento del algoritmo evolutivo.
- Integrar nuevos operadores genéticos, de distinta aridad, que permitan diversificar aún más la forma en que se generan nuevos individuos y se produce descendencia en el algoritmo evolutivo.
- Integrar nuevos modelos de construcción molecular que permita generar especies químicas de mayor complejidad y/o que no se pueden generar con el modelo acá propuesto (por ejemplo: solventes iónicos, estructuras bicíclicas, proteínas, etc.)
- Acoplar el modelo CAMD construido en el presente trabajo a una herramienta de Inteligencia Artificial y/o Machine Learning, que permita gestionar nuevas características de la herramienta desarrollada.
- Integrar al programa un mecanismo para facilitar la actualización de las bases de datos empleadas (ZINC, EPA-DSSTox, A&M, ICIS Chemical Business, Chemical Week),

dado que dichas bases de datos se van actualizando de manera periódica, ocasionando desactualización de los valores de las propiedades empleadas por el presente modelo. Esto podría mejorarse utilizando Cloud Computing, ya que la mayoría de estas bases de datos cuentan con soporte en línea.

Anexos

Con el fin de comprender a mayor detalle todos los aspectos desarrollados en la presente tesis y enriquecer los resultados generados, se anexan las siguientes secciones a modo de anexo.

- A. Reglas de representación molecular mediante notación SMILES.
- B. Tablas de Cayley y equivalencias rotacionales de estructuras.
- C. Estructuras moleculares evaluadas respecto a aplicabilidad industrial.
- D. Parámetros del modelo UNIFAC-DMD.
- E. Manual de operación del programa CAMD.
- F. Registro fotográfico de las extracciones líquido-líquido realizadas.
- G. Método de cromatografía HPLC, curvas de calibración y cromatogramas.
- H. Método de titulación de Karl Fischer y volumetrías ácido-base.
- I. Grupos estructurales heterocíclicos añadidos al modelo CAMD.

A. Reglas de representación molecular mediante notación SMILES/SMARTS

Para el desarrollo de la representación molecular en notación SMILES, existen cuatro reglas de codificación, correspondientes a la especificación de átomos, enlaces, ramificaciones y estructuras cíclicas [254], [255], [256]. Existen reglas adicionales para representar varios tipos de isomería, sin embargo, considerando la falta de necesidad de representación de las propiedades ópticas de moléculas, no fue considerado en este proyecto. El presente anexo describirá brevemente dichas reglas de codificación.

Regla 1: Átomos: Los átomos están representados por sus símbolos atómicos, siendo el único conjunto de letras de uso obligatorias en SMILES. A excepción del hidrógeno, cada átomo se especifica independientemente por su símbolo entre corchetes, []. La segunda letra de los símbolos de dos caracteres debe ingresarse en minúsculas.

En caso de que el átomo pertenezca al subconjunto orgánico (B, C, N, O, P, S, F, Cl, Br e I), el símbolo puede escribirse sin corchetes solo si el número de hidrógenos unidos se ajusta a la valencia normal más baja consistente con enlaces explícitos, de acuerdo con las valencias presentadas en la **Tabla A-1**.

Tabla A-1: Valencias normales para átomos orgánicos – Modelo SMILES.

Símbolo	B	C	N	O	P	S	F, Cl, Br, I
Valencia	3	4	3, 5	2	3, 5	2, 4, 6	1

- En caso de que un átomo pertenezca a un anillo aromático, se especifica el símbolo con letra minúscula. Por ejemplo, C representará al carbono alifático y c representará al carbono aromático.

- Los átomos con valencias distintas de las “normales” y los elementos que no pertenezcan al subconjunto orgánico, deben describirse entre paréntesis.
- En caso de representar iones, la carga se especifica dentro del corchete (+ o -). Si no se especifica, se supone que el número de hidrógenos unidos y la carga es cero para un átomo entre paréntesis. Teniendo en cuenta esto, y considerando la imposibilidad de representar iones en los modelos de contribución de grupos empleados en este trabajo, se asume que ausencia de construcción de moléculas electrolíticas (aniones, cationes, zwitteriones).

Regla 2: Enlaces: En SMILES, la representación de enlaces entre dos átomos se realiza mediante los símbolos presentes en la **Tabla A-2**. Se supone que los átomos adyacentes están conectados entre sí mediante un enlace simple o aromático (suele omitirse dichos enlaces). Para representar estructuras lineales, la notación SMILES corresponde a la notación diagramática convencional, excepto que los hidrógenos y los enlaces simples generalmente son omitidos. En caso de estructuras lineales, es posible representar una misma molécula por muchas representaciones SMILES igualmente válidas.

Tabla A-2: Simbología de los distintos tipos de enlace en el modelo SMILES.

Tipo de Enlace	Simple	Doble	Triple	Aromático
Símbolo	-	=	#	,

Regla 3: Ramificaciones: Las ramificaciones de una molécula se especifican encerrándolas entre paréntesis. Se puede hacer anidado de más de una ramificación. En este caso, la conexión implícita a una expresión entre paréntesis (una rama) está siempre hacia la izquierda.

Regla 4: Estructuras cíclicas: Las estructuras cíclicas se representan rompiendo un enlace en cada anillo. Los enlaces están numerados en cualquier orden, designando enlaces de apertura de anillo (o cierre de anillo) por un dígito inmediatamente después del símbolo atómico en cada cierre de anillo. Esto deja un gráfico no cíclico conectado que se escribe como una estructura no cíclica utilizando las tres reglas descritas anteriormente. Por lo general, hay muchas descripciones diferentes, pero igualmente válidas, de la misma estructura: Se pueden escribir muchas otras notaciones para la misma estructura,

SMILES/SMARTS

derivadas de diferentes cierres de anillo. SMILES no tiene una entrada preferida en la entrada.

Un solo átomo puede tener más de un cierre de anillo. Por tal razón, es posible reutilizar los dígitos que denotan los cierres de anillo. Esta capacidad permite especificar estructuras con 10 o más anillos. Las estructuras que requieren más de 10 cierres de anillo para abrirse a la vez son extremadamente raras, por lo cual no serán considerados. En caso de ser necesario, los cierres de anillo con números más altos pueden especificarse anteponiendo un número de dos dígitos con un signo de porcentaje (%). Por ejemplo, C2%13%24 es un átomo de carbono con cierres de anillo 2, 13 y 24.

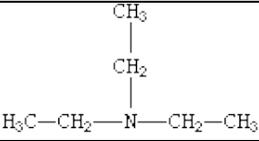
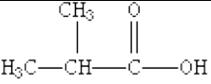
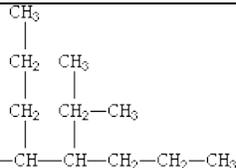
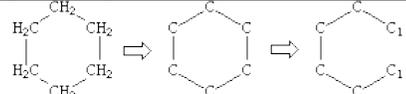
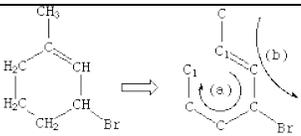
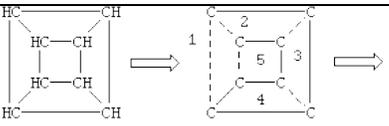
Aparte de las reglas anteriores, existe una convención específica para la representación de los átomos de hidrógeno al utilizar la notación SMILES. Normalmente, no es necesario especificarlos para la mayoría de las estructuras orgánicas. La presencia de hidrógenos se puede especificar de tres formas:

- Implícitamente, para átomos especificados sin paréntesis, a partir de supuestos de valencia normales.
- Explícitamente por recuento, entre paréntesis, por el recuento de hidrógeno suministrado; cero si no se especifica.
- Átomos tan explícitos, como átomos [H].

Hay cuatro situaciones en las que se requiere la especificación de una especificación de hidrógeno explícita: 1.) hidrógeno cargado, es decir, un protón, [H+], 2.) hidrógenos conectados a otros hidrógenos, por ejemplo, hidrógeno molecular, [H][H], 3.) hidrógenos conectados a otro átomo distinto, por ejemplo, puentes de hidrógenos; y 4.) hidrógenos isotópicos, por ejemplo, agua dura, [2H]O[2H].

De acuerdo con lo anterior, la **Tabla A-3** presenta un conjunto de moléculas de ejemplo, considerando las 4 reglas anteriormente descritas.

Tabla A-3: Ejemplos de estructuras aplicando el método de representación SMILES.

Regla	Molécula	SMILES
1	Metano (CH ₄)	C
	Fosfina (PH ₃)	P
	Amoníaco (NH ₃)	N
	Sulfuro de hidrógeno (H ₂ S)	S
	Agua (H ₂ O)	O
	Ácido clorhídrico (HCl)	Cl
2	Etano (CH ₃ CH ₃)	CC
	Formaldehído (CH ₂ O)	C=O
	Eteno (CH ₂ = CH ₂)	C=C
	Dióxido de carbono (CO ₂)	O=C=O
	Dimetil éter (CH ₃ OCH ₃)	COC
	Cianuro de hidrógeno (HCN)	C#N
	Etanol (CH ₃ CH ₂ OH)	CCO
	Hidrógeno molecular (H ₂)	[H][H]
	6-hidroxi-1,4-hexadieno (CH ₂ =CH-CH ₂ -CH=CH-CH ₂ -OH)	C=CCC=CCO C=C-C-C=C-C-O OCC=CCC=C
3	 Trietilamina	C=C-C-C=C-C-O
	 Ácido isobutírico	OCC=CCC=C
	 3-propil-4-isopropil-1-hepteno	C=CC(CCC)C(C(C)C)CCC
4	 Ciclohexano	C1CCCCC1
	 1-metil-3-bromo-ciclohexeno-1	CC1=CC(Br)CCC1 CC1=CC(CCC1)Br
	 Cubano	C12C3C4C1C5C4C3C25

El proceso de búsqueda de subestructuras, consistente en el proceso de identificación de un patrón particular (subgrafo) dentro de una molécula (grafo), es una de las tareas más importantes asociadas a la química computacional. Se utiliza prácticamente en todas las

SMILES/SMARTS

aplicaciones que requieren de una representación digital de una molécula, incluida la representación funcional, diseño de experimentos, química analítica, entre otros enfoques. En el desarrollo del presente estudio, fue necesario incorporar dicho proceso de búsqueda de subestructuras, principalmente en la identificación de estructuras factibles [257].

SMARTS es un lenguaje que permite especificar subestructuras utilizando reglas que son extensiones sencillas de la representación SMILES. Salvo excepciones, casi todas las especificaciones de SMILES son objetos SMARTS válidos. Con este lenguaje, se pueden realizar especificaciones de búsqueda de subestructuras flexibles y eficientes en términos de tiempo razonables para múltiples evaluaciones y criterios estructurales [257][258].

Teniendo en cuenta las reglas de codificación de la representación SMILES descritas previamente, es posible con el lenguaje SMARTS asignar etiquetas a los componentes del grafo (átomos, enlaces, ciclos, etc.) SMARTS permite ampliar la especificación de un gráfico usando un conjunto de operadores lógicos, y símbolos de átomos/enlaces especiales, que permiten que el proceso de búsqueda sea más general. La **Tabla A-4** presenta el conjunto de símbolos primitivos para realizar la descripción de propiedades atómicas, más allá de los utilizados en SMILES. Se describe el orden de jerarquías de los operadores, donde a menor nivel, mayor priorización en el proceso de búsqueda.

La utilización del lenguaje SMARTS en el presente estudio fue incorporada para la rápida identificación y clasificación de patrones pertenecientes a los distintos grupos funcionales de los modelos de estimación de propiedades de Hukkerikar y UNIFAC-DMD durante el Filtro III (Sección 3.1.1), en la evaluación de los caracteres aromáticos de las estructuras evaluadas (Sección 3.1.2) y en la clasificación de los distintos grupos funcionales para determinación del impacto de las propiedades sobre las funciones objetivo (Sección 3.1.4).

Con el fin de establecer ejemplos aplicados para las distintas búsquedas realizadas en este estudio, la **Tabla A-5** presenta una serie de ejemplos de aplicación del lenguaje SMARTS en la búsqueda de subestructuras y/o grupos funcionales, siguiendo las características y restricciones definidas en las reglas de construcción molecular dadas en la Sección 2.2.2.

Tabla A-4: Operadores lógicos del lenguaje SMARTS y jerarquía de operadores.

SMARTS – Operadores Lógicos				
Tipo	Nivel	Símbolo	Expresión	Significado
Agrupación	1	()	(e1)	Agrupación de átomos, máxima prioridad.
	2	[]	[e1]	Agrupación de grupos funcionales específicos.
Operadores lógicos	3	!	!e1	Negación (no e1).
	4	&	e1&e2	Y, alta precedencia (e1 Y e2).
	5	,	e1,e2	O (e1 O e2).
	6	;	e1;e2	y, baja precedencia (e1 y e2).
Átomo	7	#n	#n	Átomo de masa atómica n.
	8	h<n>	h0, ..., h4	Hidrógeno con (al menos) n enlaces.
		r<n>	rn	Átomo en anillo con (al menos) tamaño n.
		x<n>	xn	Átomo con (al menos) n enlaces totales.
		v<n>	vn	Átomo con n enlaces (valencia) a otros átomos (no hidrógenos).
	9	H<n>	H0, ..., H4	Hidrógeno con (exactamente) n enlaces.
		R<n>	Rn	Átomo en anillo con (exactamente) tamaño n.
	X<n>	Xn	Átomo con (exactamente) n enlaces totales.	
10	D<n>	D0, ..., D4	Átomo con n grado de hibridación.	
11	A	A	Cualquier átomo alifático.	
	A	A	Cualquier átomo aromático.	
	*	*	Comodín, cualquier átomo.	
Enlace	12	-	C-C	Enlace Simple (Alifático).
		:	c:c	Enlace Simple (Aromático).
=		C=C	Doble enlace.	
#		C#C	Triple enlace.	
13	~	~C	Cualquier enlace (comodín).	
	@	@C	Cualquier enlace perteneciente a anillo.	

Tabla A-5: Ejemplos de búsqueda de subestructuras aplicando los distintos operadores del lenguaje SMARTS.

SMARTS	Significado	Tipo de moléculas buscadas
C(=O)O	Un grupo carbonilo unido a un átomo de oxígeno	Moléculas que contengan un grupo carbonilo
[NH2]	Un átomo de nitrógeno unido a dos átomos de hidrógeno	Moléculas que contengan grupos amino
[C;!H0]	Un átomo de carbono que no esté unido a ningún átomo de hidrógeno	Moléculas que contengan carbonos cuaternarios
C1CCC1	Un anillo de cinco átomos de carbono en forma de ciclohexano	Moléculas que contengan ciclopentanos

B. Tablas de Cayley y equivalencias rotacionales de estructuras

Las tablas de Cayley, empleadas en la Sección 2.2.3 (monocíclicos) y la Sección 2.2.4 (policíclicos), permiten representar la operabilidad de grupos estructurales en función de los diversos tipos de rotaciones factibles (rotación y translación), necesario en la aplicación del teorema de enumeración de Pólya. Entre la **Tabla B-1** y la **Tabla B-4** se presentan las tablas de Cayley para los grupos diedrales D_3, D_4, D_5, D_6 , en compuestos monocíclicos.

Tabla B-1: Tabla de Cayley para el grupo diedral D_3 (polígono: triángulo).

0	<i>I</i>	R_{120}	R_{240}	R_a	R_b	R_c
<i>I</i>	<i>I</i>	R_{120}	R_{240}	R_a	R_b	R_c
R_{120}	R_{120}	R_{240}	<i>I</i>	R_b	R_c	R_a
R_{240}	R_{240}	<i>I</i>	R_{120}	R_c	R_a	R_b
R_a	R_a	R_c	R_b	<i>I</i>	R_{240}	R_{120}
R_b	R_b	R_a	R_c	R_{120}	<i>I</i>	R_{240}
R_c	R_c	R_b	R_a	R_{240}	R_{120}	<i>I</i>

Tabla B-2: Tabla de Cayley para el grupo diedral D_4 (polígono: cuadrado).

0	<i>I</i>	R_{90}	R_{180}	R_{270}	R_a	R_b	R_c	R_d
<i>I</i>	<i>I</i>	R_{90}	R_{180}	R_{270}	R_a	R_b	R_c	R_d
R_{90}	R_{90}	R_{180}	R_{270}	<i>I</i>	R_c	R_d	R_b	R_a
R_{180}	R_{180}	R_{270}	<i>I</i>	R_{90}	R_b	R_a	R_d	R_c
R_{270}	R_{270}	<i>I</i>	R_{90}	R_{180}	R_d	R_c	R_a	R_b
R_a	R_a	R_d	R_b	R_c	<i>I</i>	R_{180}	R_{270}	R_{90}
R_b	R_b	R_c	R_a	R_d	R_{180}	<i>I</i>	R_{90}	R_{270}
R_c	R_c	R_a	R_d	R_b	R_{90}	R_{270}	<i>I</i>	R_{180}
R_d	R_d	R_b	R_c	R_a	R_{270}	R_{90}	R_{180}	<i>I</i>

Tabla B-3: Tabla de Cayley para el grupo diedral D_5 (polígono: pentágono).

0	I	R₇₂	R₁₄₄	R₂₁₆	R₂₈₈	R_a	R_b	R_c	R_d	R_e
I	<i>I</i>	<i>R₇₂</i>	<i>R₁₄₄</i>	<i>R₂₁₆</i>	<i>R₂₈₈</i>	<i>R_a</i>	<i>R_b</i>	<i>R_c</i>	<i>R_d</i>	<i>R_e</i>
R₇₂	<i>R₇₂</i>	<i>R₁₄₄</i>	<i>R₂₁₆</i>	<i>R₂₈₈</i>	<i>I</i>	<i>R_d</i>	<i>R_e</i>	<i>R_a</i>	<i>R_b</i>	<i>R_c</i>
R₁₄₄	<i>R₁₄₄</i>	<i>R₂₁₆</i>	<i>R₂₈₈</i>	<i>I</i>	<i>R₇₂</i>	<i>R_b</i>	<i>R_c</i>	<i>R_d</i>	<i>R_e</i>	<i>R_a</i>
R₂₁₆	<i>R₂₁₆</i>	<i>R₂₈₈</i>	<i>I</i>	<i>R₇₂</i>	<i>R₁₄₄</i>	<i>R_e</i>	<i>R_a</i>	<i>R_b</i>	<i>R_c</i>	<i>R_d</i>
R₂₈₈	<i>R₂₈₈</i>	<i>I</i>	<i>R₇₂</i>	<i>R₁₄₄</i>	<i>R₂₁₆</i>	<i>R_c</i>	<i>R_d</i>	<i>R_e</i>	<i>R_a</i>	<i>R_b</i>
R_a	<i>R_a</i>	<i>R_c</i>	<i>R_e</i>	<i>R_b</i>	<i>R_d</i>	<i>I</i>	<i>R₂₁₆</i>	<i>R₇₂</i>	<i>R₂₈₈</i>	<i>R₁₄₄</i>
R_b	<i>R_b</i>	<i>R_d</i>	<i>R_a</i>	<i>R_c</i>	<i>R_e</i>	<i>R₁₄₄</i>	<i>I</i>	<i>R₂₁₆</i>	<i>R₇₂</i>	<i>R₂₈₈</i>
R_c	<i>R_c</i>	<i>R_e</i>	<i>R_b</i>	<i>R_d</i>	<i>R_a</i>	<i>R₂₈₈</i>	<i>R₁₄₄</i>	<i>I</i>	<i>R₂₁₆</i>	<i>R₇₂</i>
R_d	<i>R_d</i>	<i>R_a</i>	<i>R_c</i>	<i>R_e</i>	<i>R_b</i>	<i>R₇₂</i>	<i>R₂₈₈</i>	<i>R₁₄₄</i>	<i>I</i>	<i>R₂₁₆</i>
R_e	<i>R_e</i>	<i>R_b</i>	<i>R_d</i>	<i>R_a</i>	<i>R_c</i>	<i>R₂₁₆</i>	<i>R₇₂</i>	<i>R₂₈₈</i>	<i>R₁₄₄</i>	<i>I</i>

Tabla B-4: Tabla de Cayley para el grupo diedral D_6 (polígono: hexágono).

0	I	R₆₀	R₁₂₀	R₁₈₀	R₂₄₀	R₃₀₀	R_a	R_b	R_c	R_d	R_e	R_f
I	<i>I</i>	<i>R₆₀</i>	<i>R₁₂₀</i>	<i>R₁₈₀</i>	<i>R₂₄₀</i>	<i>R₃₀₀</i>	<i>R_a</i>	<i>R_b</i>	<i>R_c</i>	<i>R_d</i>	<i>R_e</i>	<i>R_f</i>
R₆₀	<i>R₆₀</i>	<i>R₁₂₀</i>	<i>R₁₈₀</i>	<i>R₂₄₀</i>	<i>R₃₀₀</i>	<i>I</i>	<i>R_b</i>	<i>R_c</i>	<i>R_d</i>	<i>R_e</i>	<i>R_f</i>	<i>R_a</i>
R₁₂₀	<i>R₁₂₀</i>	<i>R₁₈₀</i>	<i>R₂₄₀</i>	<i>R₃₀₀</i>	<i>I</i>	<i>R₆₀</i>	<i>R_c</i>	<i>R_d</i>	<i>R_e</i>	<i>R_f</i>	<i>R_a</i>	<i>R_b</i>
R₁₈₀	<i>R₁₈₀</i>	<i>R₂₄₀</i>	<i>R₃₀₀</i>	<i>I</i>	<i>R₆₀</i>	<i>R₁₂₀</i>	<i>R_d</i>	<i>R_e</i>	<i>R_f</i>	<i>R_a</i>	<i>R_b</i>	<i>R_c</i>
R₂₄₀	<i>R₂₄₀</i>	<i>R₃₀₀</i>	<i>I</i>	<i>R₆₀</i>	<i>R₁₂₀</i>	<i>R₁₈₀</i>	<i>R_e</i>	<i>R_f</i>	<i>R_a</i>	<i>R_b</i>	<i>R_c</i>	<i>R_d</i>
R₃₀₀	<i>R₃₀₀</i>	<i>I</i>	<i>R₆₀</i>	<i>R₁₂₀</i>	<i>R₁₈₀</i>	<i>R₂₄₀</i>	<i>R_f</i>	<i>R_a</i>	<i>R_b</i>	<i>R_c</i>	<i>R_d</i>	<i>R_e</i>
R_a	<i>R_a</i>	<i>R_f</i>	<i>R_e</i>	<i>R_d</i>	<i>R_c</i>	<i>R_b</i>	<i>I</i>	<i>R₃₀₀</i>	<i>R₂₄₀</i>	<i>R₁₈₀</i>	<i>R₁₂₀</i>	<i>R₆₀</i>
R_b	<i>R_b</i>	<i>R_a</i>	<i>R_f</i>	<i>R_e</i>	<i>R_d</i>	<i>R_c</i>	<i>R₆₀</i>	<i>I</i>	<i>R₃₀₀</i>	<i>R₂₄₀</i>	<i>R₁₈₀</i>	<i>R₁₂₀</i>
R_c	<i>R_c</i>	<i>R_b</i>	<i>R_a</i>	<i>R_f</i>	<i>R_e</i>	<i>R_d</i>	<i>R₁₂₀</i>	<i>R₆₀</i>	<i>I</i>	<i>R₃₀₀</i>	<i>R₂₄₀</i>	<i>R₁₈₀</i>
R_d	<i>R_d</i>	<i>R_c</i>	<i>R_b</i>	<i>R_a</i>	<i>R_f</i>	<i>R_e</i>	<i>R₁₈₀</i>	<i>R₁₂₀</i>	<i>R₆₀</i>	<i>I</i>	<i>R₃₀₀</i>	<i>R₂₄₀</i>
R_e	<i>R_e</i>	<i>R_d</i>	<i>R_c</i>	<i>R_b</i>	<i>R_a</i>	<i>R_f</i>	<i>R₂₄₀</i>	<i>R₁₈₀</i>	<i>R₁₂₀</i>	<i>R₆₀</i>	<i>I</i>	<i>R₃₀₀</i>
R_f	<i>R_f</i>	<i>R_e</i>	<i>R_d</i>	<i>R_c</i>	<i>R_b</i>	<i>R_a</i>	<i>R₃₀₀</i>	<i>R₂₄₀</i>	<i>R₁₈₀</i>	<i>R₁₂₀</i>	<i>R₆₀</i>	<i>I</i>

Con respecto a los grupos policíclicos, tanto para estructuras policíclicas tipo espiro como fusionadas, se identificaron dos clases de estructuras, en función de la cantidad de movimientos rotacionales presentes en los grupos diedrales de la estructura (Clase I: 2 movimientos y Clase II: 4 movimientos). Cualquier estructura policíclica perteneciente a la

misma clase presentará una tabla de Cayley idéntica. La **Tabla B-5** presenta la distribución de las estructuras policíclicas tipo espiro (E-1 a E-10) y fusionadas (F-1 a F-10), descritas en la Sección 2.2.4, respecto a la Tabla de Cayley correspondiente. Las **Tabla B-6** y la **Tabla B-7** presentan las tablas de Cayley correspondientes a las clases I y II, respectivamente.

Cabe resaltar que tanto las clases I como la II presentan tablas de Cayley simétricas, por lo que es posible considerar a todos los grupos estructurales policíclicos como abelianos, contando con un carácter conmutativo adicional, respecto a los grupos monocíclicos.

Tabla B-5: Distribución de estructuras policíclicas espiro y fusionadas en clases, según la cantidad de movimientos rotacionales válidos.

Clase de estructura	Tipo de policíclico	Estructuras válidas a la clase
Clase I: 4 movimientos	Espiro	E-1, E-5, E-8, E-10
	Fusionado	F-1, F-5, F-8, F-10
Clase II: 2 movimientos	Espiro	E-2, E-3, E-4, E-6, E-7, E-9
	Fusionado	F-2, F-3, F-4, F-6, F-7, F-9

Tabla B-6: Tabla de Cayley para las estructuras policíclicas de Clase I, con 4 movimientos rotacionales en el grupo diedral.

0	<i>I</i>	R_α	R_β	R_{180}
<i>I</i>	<i>I</i>	R_α	R_β	R_{180}
R_α	R_α	<i>I</i>	R_{180}	R_β
R_β	R_β	R_{180}	<i>I</i>	R_α
R_{180}	R_{180}	R_β	R_α	<i>I</i>

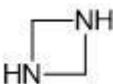
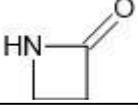
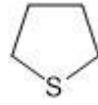
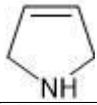
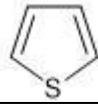
Tabla B-7: Tabla de Cayley para las estructuras policíclicas de Clase II, con 2 movimientos rotacionales en el grupo diedral.

0	<i>I</i>	R_α
<i>I</i>	<i>I</i>	R_α
R_α	R_α	<i>I</i>

C. Estructuras moleculares evaluadas con aplicabilidad industrial

La **Tabla C-1** enlista todas las moléculas empleadas en la evaluación de aplicabilidad industrial, usando la metodología de las Secciones 2.2.5 y 2.2.6. Se presenta el nombre IUPAC, la estructura molecular y su equivalencia en la representación SMILES.

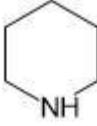
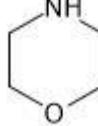
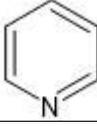
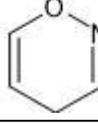
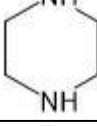
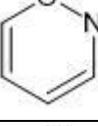
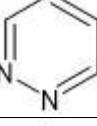
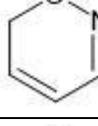
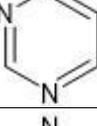
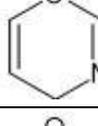
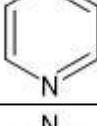
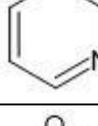
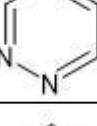
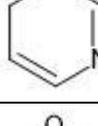
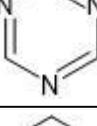
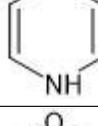
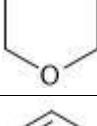
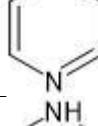
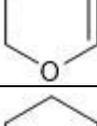
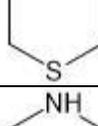
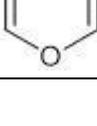
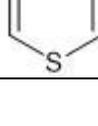
Tabla C-1: Estructuras moleculares evaluadas respecto a aplicabilidad industrial.

Monocíclicos con ciclos de 3 átomos ($N = 4$)			
	M1: Aziridina. SMILES: C1CN1		M2: Oxirano. SMILES: C1CO1
	M3: 2H-Azirina. SMILES: C1=NC1		M4: Tiirano. SMILES: C1CS1
Monocíclicos con ciclos de 4 átomos ($N = 9$)			
	M5: Azetidina. SMILES: C1CNC1		M6: 2H-Oxete. SMILES: C1=COC1
	M7: 2,3-Dihidroazate. SMILES: C1=NCC1		M8: Tietano. SMILES: C1CSC1
	M9: Azete. SMILES: c1cnc1		M10: 2H-Tiete. SMILES: C1=CSC1
	M11: 1,3-Diazetidina. SMILES: C1NCN1		M12: Azetidín-2-ona (β -lactama). SMILES: O=C1CCN1
	M13: Oxetano. SMILES: C1COC1		
Monocíclicos con ciclos de 5 átomos ($N = 34$)			
	M14: Pirrolidina. SMILES: C1CCNC1		M15: Tetrahidrotiofeno. SMILES: C1CCSC1
	M16: 3-Pirrolina. SMILES: C1=CCNC1		M17: Tiofeno. SMILES: c1ccsc1
	M18: 2-Pirrolina. SMILES: C1=CNCC1		M19: Oxazol. SMILES: c1cocn1

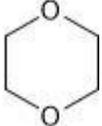
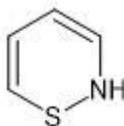
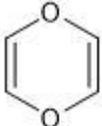
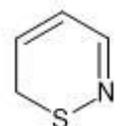
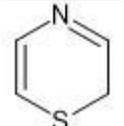
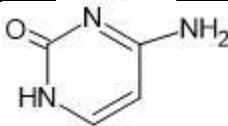
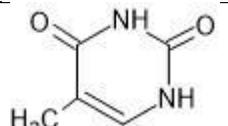
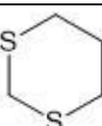
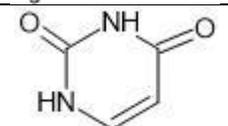
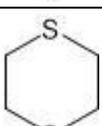
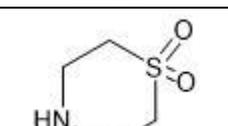
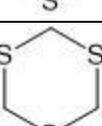
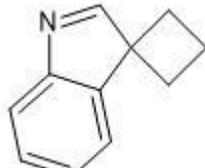
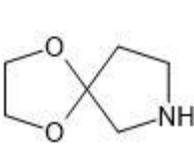
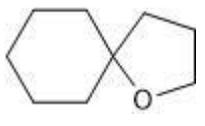
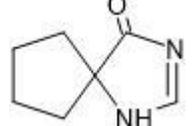
Cont. Tabla C-1.

	M20: 2H-Pirrol. SMILES: C1=CCN=C1		M21: Isoxazol. SMILES: c1cnoc1
	M22: 1H-Pirrol. SMILES: c1cc[nH]c1		M23: Isotiazol. SMILES: c1cnsc1
	M24: Pirazolidina. SMILES: C1CNNC1		M25: Tiazol. SMILES: c1cscn1
	M26: Imidazolidina. SMILES: C1CNCN1		M27: 1,2-Oxatolano. SMILES: C1COSC1
	M28: 2-Pirazolina. SMILES: C1=NNCC1		M29: 1,3-Oxatolano. SMILES: C1CSCO1
	M30: 2-Imidazolina. SMILES: C1=NCCN1		M31: 1,2,5-Oxadiazol. SMILES: c1cnon1
	M32: Pirazol. SMILES: c1cn[nH]c1		M33: 1,2,3-Oxadiazol. SMILES: c1conn1
	M34: Imidazol. SMILES: c1c[nH]cn1		M35: 1,3,4-Tiadiazol. SMILES: c1nncs1
	M36: 1,2,4-Triazol. SMILES: c1nc[nH]n1		M37: 1,2,5-Tiadiazol. SMILES: c1cnsn1
	M38: 1,2,3-Triazol. SMILES: c1c[nH]nn1		M39: Sulfolano. SMILES: O=S1(=O)CCCC1
	M40: Tetrazol. SMILES: c1nnn[nH]1		M41: 2,4-Tiazolidinediona. SMILES: O=C1CSC(=O)N1
	M42: Tetrahidrofurano. SMILES: C1CCOC1		M43: Succinimida. SMILES: O=C1CCC(=O)N1
	M44: Furano. SMILES: c1ccoc1		M45: 2-Oxazolidona. SMILES: O=C1NCCO1
	M46: 1,3-Dioxolano. SMILES: C1COCO1		M47: Hidranioina. SMILES: O=C1CNC(=O)N1

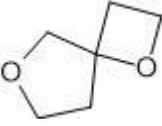
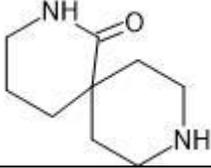
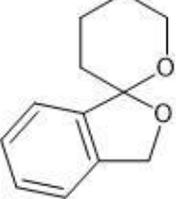
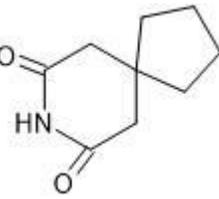
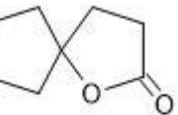
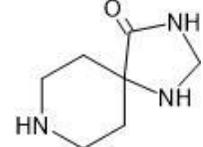
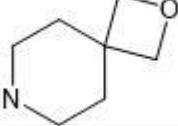
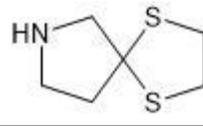
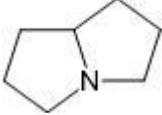
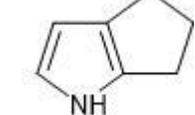
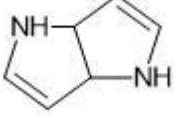
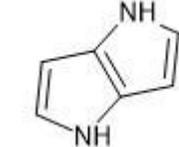
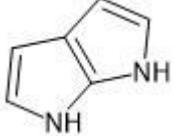
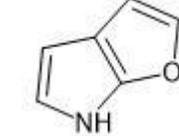
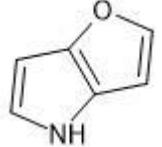
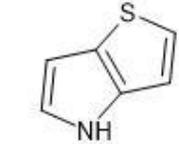
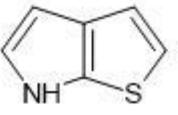
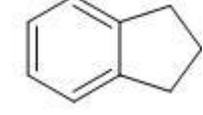
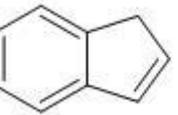
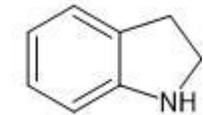
Cont. Tabla C-1.

Monocíclicos con ciclos de 6 átomos (N = 38)			
	M48: Piperidina. SMILES: C1CCNCC1		M49: Morfolina. SMILES: C1COCCN1
	M50: Piridina. SMILES: c1cncc1		M51: 4H-1,2-Oxazina. SMILES: C1=CON=CC1
	M52: Piperazina. SMILES: C1CNCCN1		M53: 2H-1,2-Oxazina. SMILES: c1cco[nH]c1
	M54: Piridazina. SMILES: c1cnnc1		M55: 6H-1,2-Oxazina. SMILES: C1=CCON=C1
	M56: Pirimidina. SMILES: c1cncnc1		M57: 4H-1,3-Oxazina. SMILES: C1=COC=NC1
	M58: Pirazina. SMILES: c1cnccn1		M59: 2H-1,3-Oxazina. SMILES: C1=COCN=C1
	M60: 1,2,4-Triazina. SMILES: c1cnncn1		M61: 6H-1,3-Oxazina. SMILES: C1=CN=COC1
	M62: 1,3,5-Triazina. SMILES: c1nncn1		M63: 4H-1,4-Oxazina. SMILES: c1cocc[nH]1
	M64: Tetrahidropirano. SMILES: C1CCOCC1		M65: 2H-1,4-Oxazina. SMILES: C1=COCC=N1
	M66: 2H-Pirano. SMILES: C1=CCOC=C1		M67: Tiomorfolina. SMILES: C1CSCCN1
	M68: 4H-Pirano. SMILES: C1=COC=CC1		M69: 4H-1,4-Tiazina. SMILES: c1csc[nH]1

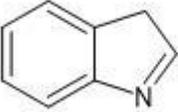
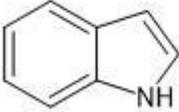
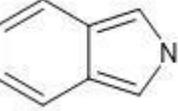
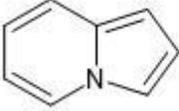
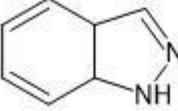
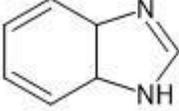
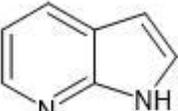
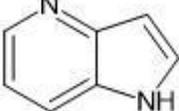
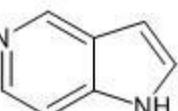
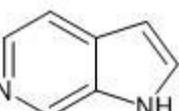
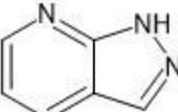
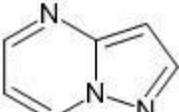
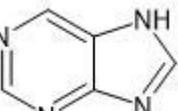
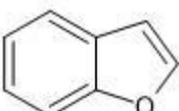
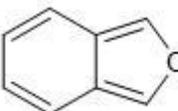
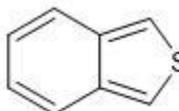
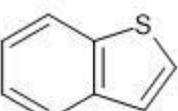
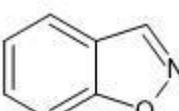
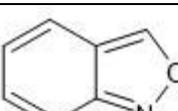
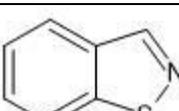
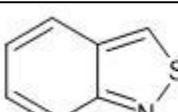
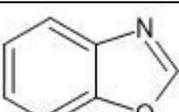
Cont. Tabla C-1.

	M70: 1,4-Dioxano. SMILES: C1COCCO1		M71: 2H-1,2- Tiazina. SMILES: c1ccs[nH]c1
	M72: 1,4-Dioxina. SMILES: c1cocco1		M73: 6H-1,2- Tiazina. SMILES: C1=CCSN=C1
	M74: Tiano. SMILES: C1CCSCC1		M75: 2H-1,4-Tiazina. SMILES: C1=CSCC=N1
	M76: 2H-Tiopirano. SMILES: C1=CCSC=C1		M77: Citosina. SMILES: Nc1cc[nH]c(=O)n1
	M78: 4H-Tiopirano. SMILES: C1=CSC=CC1		M79: Timina. SMILES: Cc1c[nH]c(=O)[nH]c1=O
	M80: 1,3-Ditiano. SMILES: C1CSCSC1		M81: Uracilo. SMILES: O=c1cc[nH]c(=O)[nH]1
	M82: 1,4-Ditiano. SMILES: C1SCCS1		M83: Dióxido de tiomorfolina. SMILES: O=S1(=O)CCNCC1
	M84: 1,3,5-Triiano. SMILES: C1SCSCS1		
Policíclicos heterocíclicos tipo espiro (N = 12)			
	M85: Espiro[ciclobutano-1,3'-indol]. SMILES: C2=Nc1cccc1C23CCC3		M86: 1,4-Dioxa-7-azaspiro [4,4]nonano. SMILES: C1CC2(CN1)OCCO2
	M87: 1-Oxaspiro[4,5]decano. SMILES: C1CCC2(CC1)CCO2		M88: 1,3-Diazaspiro[4,4]non-2-en-4-ona. SMILES: O=C1N=CNC12CCCC2

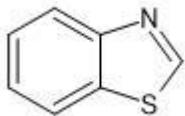
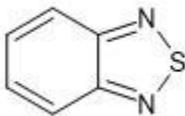
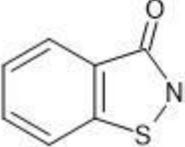
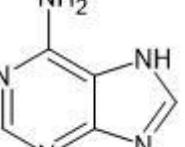
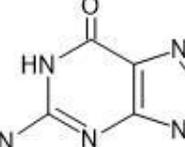
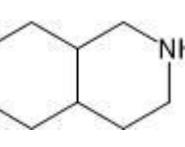
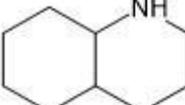
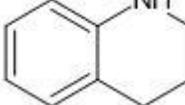
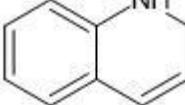
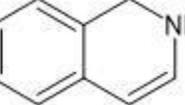
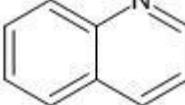
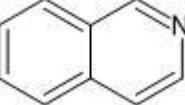
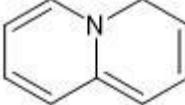
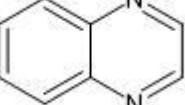
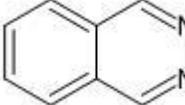
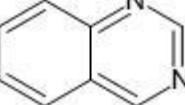
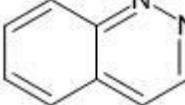
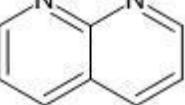
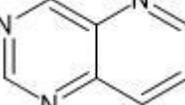
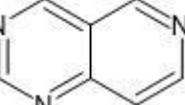
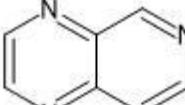
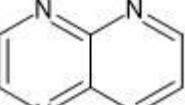
Cont. Tabla C-1.

	M89: 1,6-Dioxaspiro [3,4]octano. SMILES: C2CC1(CCO1)CO2		M90: 2,9-Diazaspiro[5,5]undecan-1-ona. SMILES: O=C1NCCCC12CCNCC2
	M91: 3',4',5',6'-Tetrahydro-3H-espiro [isobenzofuran-1,2'-piran]. SMILES: c1ccc2c(c1)COC23CCCCO3		M92: 8-Azaspiro[4,5]decan-7,9-diona. SMILES: O=C2CC1(CCCC1)CC(=O)N2
	M93: 1-Oxaspiro[4,4]nonan-2-ona. SMILES: O=C2CCC1(CCCC1)O2		M94: 1,3,8-Triazaspiro[4,5]decan-4-ona. SMILES: O=C1NCNC12CCNCC2
	M95: 2-Oxa-7-azaspiro [3,5]nonano. SMILES: C1CC2(CCN1)COC2		M96: 1,4-Ditia-7-azaspiro [4,4]nonano. SMILES: C1CC2(CN1)SCCS2
Policíclicos heterocíclicos tipo fusionado (N = 66)			
	M97: Hexahidro-1H-pirrolizina. SMILES: C2CC1CCCN1C2		M98: 1,4,5,6-Tetrahidrociclopental[b]pirrol. SMILES: c2cc1CCCc1[nH]2
	M99: 1,3-a,4,6a-Tetrahidropirrolo[3,2-b]pirrol. SMILES: C2=CC1NC=CC1N2		M100: 1,4-Dihidropirrolo[3,2-b]pirrol. SMILES: c2cc1[nH]ccc1[nH]2
	M101: 1,6-Dihidropirrolo[2,3-b]pirrol. SMILES: c2cc1cc[nH]c1[nH]2		M102: 6H-Furo[2,3-b]pirrol. SMILES: c2cc1ccoc1[nH]2
	M103: 4H-Furo[3,2-b]pirrol. SMILES: c2cc1occc1[nH]2		M104: 4H-Tieno[3,2-b]pirrol. SMILES: c2cc1sccc1[nH]2
	M105: 6H-Tieno[3,2-b]pirrol. SMILES: c2cc1ccsc1[nH]2		M106: 2,3-Dihidro-1H-indeno. SMILES: c2ccc1CCCc1c2
	M107: Indeno. SMILES: C2=Cc1cccc1C2		M108: Indolina. SMILES: c2ccc1NCCc1c2

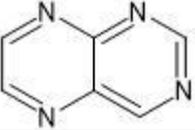
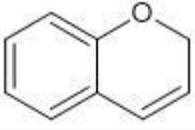
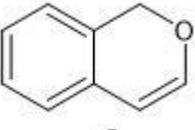
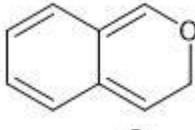
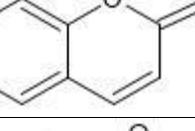
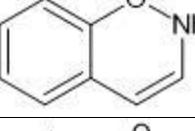
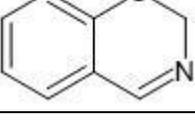
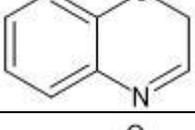
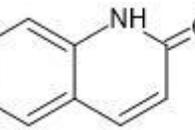
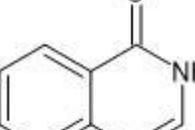
Cont. Tabla C-1.

	M109: 3H-Indol. SMILES: C2=Nc1cccc1C2		M110: 1H-Indol. SMILES: c2ccc1[nH]ccc1c2
	M111: 2H-Isoindol. SMILES: c2ccc1c[nH]cc1c2		M112: Indolizina. SMILES: c2ccn1cccc1c2
	M113: 1H-Indazol. SMILES: C2=CC1C=NNC1C=C2		M114: Benzimidazol. SMILES: C2=CC1N=CNC1C=C2
	M115: 7-Azaindol. SMILES: c2cnc1[nH]ccc1c2		M116: 4-Azaindol. SMILES: c2cnc1cc[nH]c1c2
	M117: 5-Azaindol. SMILES: c2cc1[nH]ccc1cn2		M118: 6-Azaindol. SMILES: c2cc1cc[nH]c1cn2
	M119: 7-Azaindazol. SMILES: c2cnc1[nH]ncc1c2		M120: Pirazolo[1,5-a]pirimidina. SMILES: c2cnc1ccnn1c2
	M121: Purina. SMILES: c2ncc1[nH]cnc1n2		M122: Benzofurano. SMILES: c2ccc1occc1c2
	M123: Isobenzofurano. SMILES: c2ccc1cocc1c2		M124: Benzo[c]tiofeno. SMILES: c2ccc1csc1c2
	M125: Benzo[b]tiofeno. SMILES: c2ccc1sccc1c2		M126: Benzo[d]isoxazol. SMILES: c2ccc1oncc1c2
	M127: Benzo[c]isoxazol. SMILES: c2ccc1nocc1c2		M128: Benzo[d]isotiazol. SMILES: c2ccc1sncc1c2
	M129: Benzo[c]isotiazol. SMILES: c2ccc1nsc1c2		M130: Benzo[d]oxazol. SMILES: c2ccc1ocnc1c2

Cont. Tabla C-1.

	M131: Benzo[d]tiazol. SMILES: c2ccc1scnc1c2		M132: Benzo[c][1,2,5]tiadiazol. SMILES: c2ccc1nsnc1c2
	M133: 1,2-Benzisotiazol-3(2H)-ona. SMILES: O=c1[nH]sc2ccccc12		M134: Adenina. SMILES: Nc1ncnc2nc[nH]c12
	M135: Guanina. SMILES: Nc2nc1[nH]cnc1c(=O)[nH]2		M136: Decahidroisoquinolina. SMILES: C2CCC1CNCCC1C2
	M137: Decahidroquinolina. SMILES: C2CCC1NCCCC1C2		M138: 1,2,3,4-Tetrahidroquinolina. SMILES: c2ccc1NCCCc1c2
	M139: 1,2-Dihidroquinolina. SMILES: C2=Cc1cccc1NC2		M140: 1,2-Dihidroisoquinolina. SMILES: C2=Cc1cccc1CN2
	M141: Quinolina. SMILES: c2ccc1ncccc1c2		M142: Isoquinolina. SMILES: c2ccc1cnccc1c2
	M143: 4H-Quinolizina. SMILES: C2=CCn1cccc1=C2		M144: Quinoxalina. SMILES: c2ccc1nccnc1c2
	M145: Ftalazina. SMILES: c2ccc1cncc1c2		M146: Quinazolina. SMILES: c2ccc1ncnc1c2
	M147: Tinnolina. SMILES: c2ccc1ncccc1c2		M148: 1,8-Naftiridina. SMILES: c2cnc1ncccc1c2
	M149: Pirido[3,2-d]pirimidina. SMILES: c2cnc1cncnc1c2		M150: Pirido[4,3-d]pirimidina. SMILES: c2cc1ncnc1cn2
	M151: Pirido[3,4-b]pirazina. SMILES: c2cc1nccnc1cn2		M152: Pirido[2,3-b]pirazina. SMILES: c2cnc1nccnc1c2

Cont. Tabla C-1.

	M153: Pteridina. SMILES: <chem>c2cnc1ncncc1n2</chem>		M154: 2H-Cromeno. SMILES: <chem>C2=Cc1cccc1OC2</chem>
	M155: 1H-Isocromeno. SMILES: <chem>C2=Cc1cccc1CO2</chem>		M156: 3H-Isocromeno. SMILES: <chem>C2=c1cccc1=COC2</chem>
	M157: 2H-Cromen-2-ona. SMILES: <chem>O=c2ccc1cccc1o2</chem>		M158: 2H-Benzo[e][1,2]oxazina. SMILES: <chem>c2ccc1o[nH]ccc1c2</chem>
	M159: 2H-Benzo[e][1,3]oxazina. SMILES: <chem>C1=NCOc2cccc12</chem>		M160: 2H-Benzo[b][1,4]oxazina. SMILES: <chem>C2=Nc1cccc1OC2</chem>
	M161: Quinolin-2(1H)-ona. SMILES: <chem>O=c2ccc1cccc1[nH]2</chem>		M162: Isoquinolin-1(2H)-ona. SMILES: <chem>O=c1[nH]ccc2cccc12</chem>

D. Parámetros del modelo UNIFAC-DMD

El presente anexo reporta los parámetros empleados por el modelo UNIFAC-DMD para determinación de los coeficientes de actividad y otras propiedades relacionadas con la estructura molecular de compuestos químicos, determinada mediante modelos de contribución de grupos (GC+).

El modelo UNIFAC se basa en la ecuación UNIQUAC, para los cuales los coeficientes de actividad se obtienen mediante la ecuación (D-1), donde se define una contribución combinatorial y residual para el modelo de actividad, definidas a su vez por las ecuaciones (D-2) y (D-3), respectivamente.

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \quad (D-1)$$

$$\ln \gamma_i^C = 1 - \phi_i + \ln \phi_i - 5q_i \left(1 - \frac{\phi_i}{\theta_i} + \ln \frac{\phi_i}{\theta_i} \right) \quad (D-2)$$

$$\ln \gamma_i^R = q_i \left(1 - \sum_k \left(\theta_k \frac{\beta_{ik}}{s_k} - e_{ki} \ln \frac{\beta_{ik}}{s_k} \right) \right) \quad (D-3)$$

Donde adicionalmente se requieren definir las ecuaciones auxiliares, descritas por las ecuaciones (D-4) a (D-12).

$$\phi_i = \frac{r_i}{\sum_j r_j x_j} \quad (D-4) \quad q_i = \sum_k v_k^{(i)} Q_k \quad (D-5) \quad s_k = \sum_m \theta_m \tau_{mk} \quad (D-6)$$

$$\theta_i = \frac{q_i}{\sum_j q_j x_j} \quad (D-7) \quad \beta_{ik} = \sum_m e_{mi} \tau_{mk} \quad (D-8) \quad \tau_{mk} = \exp \left(-\frac{a_{mk}}{T} \right) \quad (D-9)$$

$$r_i = \sum_k v_k^{(i)} R_k \quad (D-10) \quad \theta_k = \frac{\sum_i x_i q_i e_{ki}}{\sum_i x_j q_j} \quad (D-11) \quad e_{ki} = \frac{v_k^{(i)} Q_k}{q_i} \quad (D-12)$$

En el modelo modificado UNIFAC-DMD la contribución combinatorial al coeficiente de actividad, descrita originalmente por la ecuación (D-2) cambia a la ecuación (D-13),

integrando un parámetro empírico de corrección ϕ'_i , definido por la ecuación (D-14), empleado en la optimización de coeficientes de actividad a dilución infinita para sistemas alcano/alcano, alcano/alcohol y alcohol/alcohol [259].

$$\ln \gamma_i^C = 1 - \phi'_i + \ln \phi'_i - 5q_i \left(1 - \frac{\phi_i}{\theta_i} + \ln \frac{\phi_i}{\theta_i} \right) \quad (\text{D-13})$$

$$\phi'_i = \frac{r_i^{3/4}}{\sum_j r_j^{3/4} x_j} \quad (\text{D-14})$$

Adicionalmente, el modelo UNIFAC-DMD modifica la ecuación de los parámetros de interacción y su dependencia a la temperatura (ecuación (D-9)), reemplazándola por la ecuación (D-15). El cálculo de los grupos estructurales R_k y Q_k difiere respecto al modelo original UNIFAC para algunos sistemas alcano/alcano, alcano/alcohol y alcohol/alcohol, por lo que se toman los parámetros ajustados en el modelo UNIFAC-DMD a partir de la información recopilada de literatura.

$$\tau_{mk} = \exp \left(- \frac{A_{mk} + B_{mk}T + C_{mk}T^2}{T} \right) \quad (\text{D-15})$$

Para la aplicación de este modelo, el subíndice i identifica a la especie, el subíndice l identifica a los subgrupos; mientras que j y m son un índice ficticio que recorre todas las especies y subgrupos, respectivamente. La cantidad $v_k^{(i)}$ representa al número de subgrupos del tipo k que hay en una molécula de especie i . Tanto los parámetros de subgrupo R_k y Q_k como el parámetro de interacción de grupo a_{mk} se encuentra en permanente revisión experimental, siendo obtenible fácilmente de la literatura [234]. La **Figura D-1** presenta un resumen de los grupos y subgrupos para los cuales hay información disponible de los parámetros experimentales, en su versión publicada más reciente (6ta revisión) [232]. Por su parte, la **Tabla D-1** reporta los parámetros R_k y Q_k para grupos y subgrupos para modelo UNIFAC clásico, válidos también para UNIFAC-DMD, empleados para la aplicación de las ecuaciones (D-5), (D-10) y (D-12), de acuerdo con la descripción realizada en la Sección 2.2.8. Por su parte, la **Tabla D-2** presenta los parámetros de interacción (a_{mk}) entre dos subgrupos específicos (m y k).

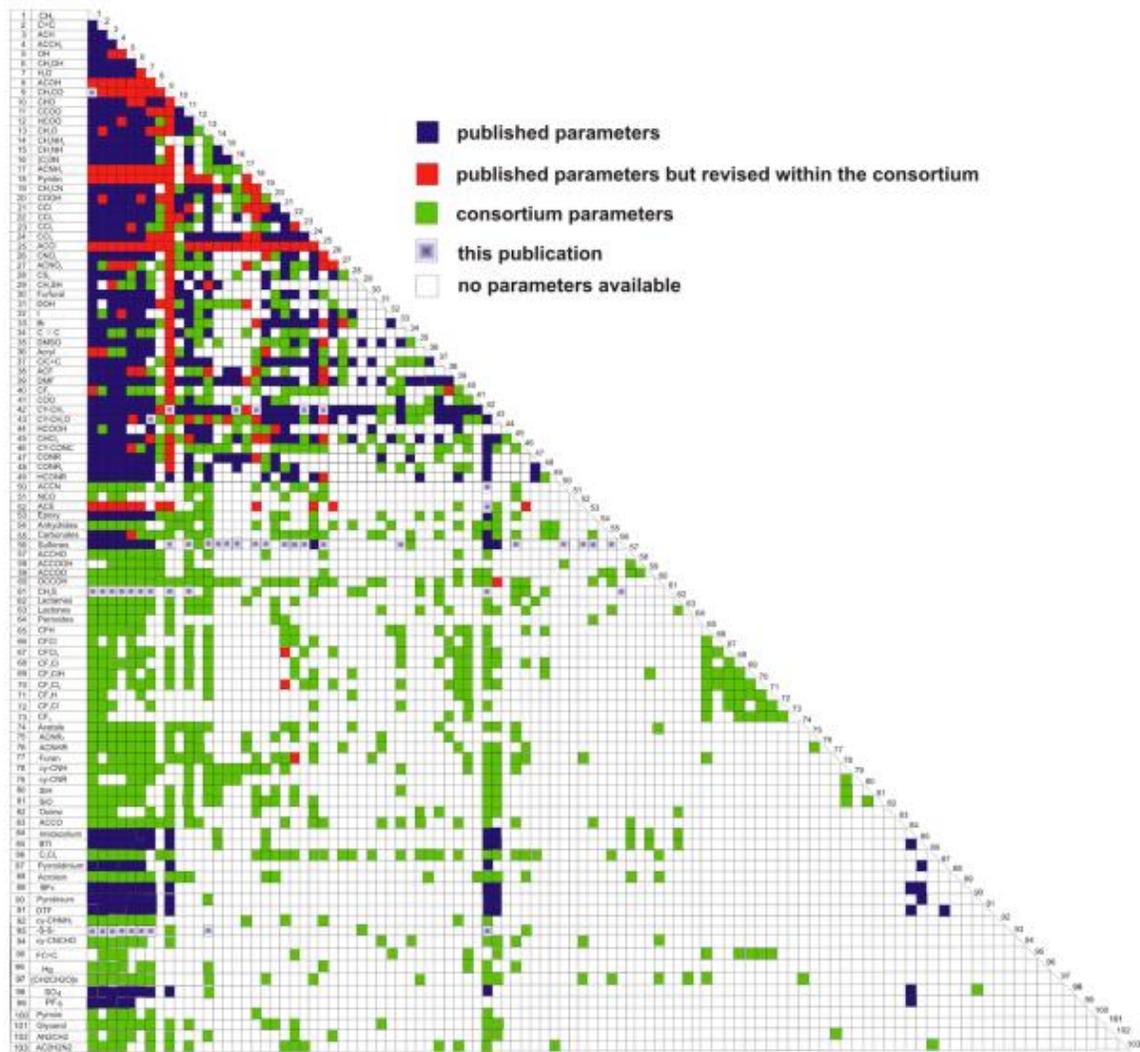


Figura D-1: Disponibilidad de parámetros de interacción y grupos funcionales – modelo UNIFAC-DMD (6ta revisión, última versión: septiembre 2021) [232].

Tabla D-1: Parámetros de subgrupo del modelo UNIFAC [232].

Grupo	S.G.	k	R _k	Q _k	Grupo	S.G.	k	R _k	Q _k
[1]CH ₂	CH ₃	1	0.9011	0.8480	[27]ACNO ₂	ACNO ₂	57	1.4199	1.1040
	CH ₂	2	0.6744	0.5400	[28]CS ₂	CS ₂	58	2.0570	1.6500
	CH	3	0.4469	0.2280	[29]CH ₃ SH	CH ₃ SH	59	1.8770	1.6760
	C	4	0.2195	0.0000		CH ₂ SH	60	1.6510	1.3680
[2]C=C	CH ₂ =CH	5	1.3454	1.1760	[30]FURFURAL	FURFURAL	61	3.1680	2.4840
	CH=CH	6	1.1167	0.8670	[31]DOH	DOH	62	2.4088	2.2480
	CH ₂ =C	7	1.1173	0.9880	[32]I	I	63	1.2640	0.9920
	CH=C	8	0.8886	0.6760	[33]BR	BR	64	0.9492	0.8320
	C=C	70	0.6605	0.4850	[34]C=C	CH=C	65	1.2920	1.0880
ACH	9	0.5313	0.4000	C=C		66	1.0613	0.7840	
[3]ACH	AC	10	0.3652	0.1200	[35]DMSO	DMSO	67	2.8266	2.4720
	ACCH ₃	11	1.2663	0.9680	[36]ACRY	ACRY	68	2.3144	2.0520
[4]ACCH ₂	ACCH ₂	12	1.0396	0.6600	[37]CLCC	CL-(C=C)	69	0.7910	0.7240
	ACCH	13	0.8121	0.3480	[38]ACF	ACF	71	0.6948	0.5240
	OH	14	1.0000	1.2000	[39]DMF	DMF	72	3.0856	2.7360
CH ₃ OH	15	1.4311	1.4320	HCON		73	2.6322	2.1200	
[7]H ₂ O	H ₂ O	16	0.9200	1.4000	[40]CF ₂	CF ₃	74	1.4060	13.800
[8]ACOH	ACOH	17	0.8952	0.6800		CF ₂	75	1.0105	0.9200
[9]CH ₂ CO	CH ₃ CO	18	1.6724	1.4880		CF	76	0.6150	0.4600
	CH ₂ CO	19	1.4457	1.1800	[41]COO	COO	77	1.3800	12.000
[10]CHO	CHO	20	0.9980	0.9480	[42]SIH ₂	SIH ₃	78	1.6035	12.632
[11]CCOO	CH ₃ COO	21	1.9031	1.7280		SIH ₂	79	1.4443	10.063
	CH ₂ COO	22	1.6764	1.4200		SIH	80	1.2853	0.7494
[12]HCOO	HCOO	23	1.2420	1.1880	SI	81	1.0470	0.4099	
[13]CH ₂ O	CH ₃ O	24	1.1450	1.0880	[43]SIO	SIH ₂ O	82	1.4838	10.621
	CH ₂ O	25	0.9183	0.7800		SIHO	83	1.3030	0.7639
	CHO	26	0.6908	0.4680	SIO	84	1.1044	0.4657	
	THF	27	0.9183	1.1000	[44]NMP	NMP	85	3.9810	32.000
[14]CNH ₂	CH ₃ NH ₂	28	1.5959	1.5440	[45]CCLF	CCL ₃ F	86	3.0356	26.440
	CH ₂ NH ₂	29	1.3692	1.2360		CCL ₂ F	87	2.2287	19.160
	CHNH ₂	30	1.1417	0.9240		HCCL ₂ F	88	2.4060	21.160
[15]CNH	CH ₃ NH	31	1.4337	1.2440		HCCLF	89	1.6493	14.160
	CH ₂ NH	32	1.2070	0.9360		CCLF ₂	90	1.8174	16.480
	CHNH	33	0.9795	0.6240		HCCLF ₂	91	1.9670	18.280
[16](C) ₃ N	CH ₃ N	34	1.1865	0.9400		CCLF ₃	92	2.1721	21.000
	CH ₂ N	35	0.9597	0.6320		CCL ₂ F ₂	93	2.6243	23.760
[17]ACNH ₂	ACNH ₂	36	1.0600	0.8160	AMH ₂	94	1.4515	12.480	
[18]PYRIDINE	C ₅ H ₅ N	37	2.9993	2.1130	AMHCH ₃	95	2.1905	17.960	
	C ₅ H ₄ N	38	2.8332	1.8330	AMHCH ₂	96	1.9637	14.880	
	C ₅ H ₃ N	39	2.6670	1.5530	AM(CH ₃) ₂	97	2.8589	24.280	
	CH ₃ CN	40	1.8701	1.7240	AMCH ₃ CH ₂	98	2.6322	21.200	
[19]CCN	CH ₂ CN	41	1.6434	1.4160	AM(CH ₂) ₂	99	2.4054	18.120	
	COOH	42	1.3013	1.2240	[47]OCCOH	C ₂ H ₅ O ₂	100	2.1226	19.040
[20]COOH	HCOOH	43	1.5280	1.5320		C ₂ H ₄ O ₂	101	1.8952	15.920
	[21]CCL	CH ₂ CL	44	1.4654	1.2640	CH ₃ S	102	1.6130	13.680
CHCL		45	1.2380	0.9520	CH ₂ S	103	1.3863	10.600	
CCL		46	1.0106	0.7240	CHS	104	1.1589	0.7480	
[22]CCL ₂	CH ₂ CL ₂	47	2.2564	1.9880	[49]MORPH	MORPH	105	3.4740	27.960
	CHCL ₂	48	2.0606	1.6840	C ₄ H ₄ S	106	2.8569	21.400	
	CCL ₂	49	1.8016	1.4480	[50]THIOPHEN	C ₄ H ₃ S	107	2.6908	18.600
[23]CCL ₃	CHCL ₃	50	2.8700	2.4100		C ₄ H ₂ S	108	2.5247	15.800
	CCL ₃	51	2.6401	2.1840	[51]NCO	NCO	109	1.0567	0.7320
[24]CCL ₄	CCL ₄	52	3.3900	2.9100	[55]SULFONES	(CH ₂) ₂ SU	118	2.6869	21.200
[25]ACCL	ACCL	53	1.1562	0.8440	[55]SULFONES	CH ₂ CHSU	119	2.4595	18.080
[26]CNO ₂	CH ₃ NO ₂	54	2.0086	1.8680	[84]IMIDAZOL	IMIDAZOL	178	2.0260	0.8680
	CH ₂ NO ₂	55	1.7818	1.5600	[85]BTI	BTI	179	5.7740	49.320
	CHNO ₂	56	1.5544	1.2480					

Tabla D-2: Parámetros de interacción de UNIFAC (a_{mk} [K]) [231], [232], [260]

m	k	a_{mk}	a_{km}																
1	2	86.02	-35.36	1	42	-450.40	-34.36	2	34	31.14	41.38	3	28	52.07	21.50	4	23	-80.25	69.90
1	3	61.13	-11.12	1	43	252.70	110.20	2	35	179.00	64.07	3	29	-10.43	28.41	4	24	-141.30	134.70
1	4	76.50	-69.70	1	44	220.30	13.89	2	36	-52.87	573.00	3	30	-64.69	157.29	4	25	-211.00	402.50
1	5	986.50	156.40	1	45	-5.87	30.74	2	37	-66.46	124.20	3	31	210.37	221.40	4	26	3629.00	-97.05
1	6	697.20	16.51	1	46	390.90	27.97	2	38	359.30	-131.70	3	32	113.30	58.68	4	27	4448.00	-127.80
1	7	1318.00	300.00	1	47	553.30	-11.92	2	39	-70.45	249.00	3	33	261.30	-154.20	4	28	-9.45	40.68
1	8	1333.00	275.80	1	48	187.00	39.93	2	40	449.40	62.40	3	34	154.26	-101.12	4	29	393.60	19.56
1	9	476.40	26.76	1	49	216.10	-23.61	2	41	48.33	1397.00	3	35	169.90	-2.50	4	30	48.49	128.80
1	10	677.00	505.70	1	50	92.99	-8.48	2	44	86.46	-16.11	3	36	383.90	-123.60	4	31	4975.00	150.64
1	11	232.10	114.80	1	51	699.13	456.19	2	46	200.20	9.76	3	37	-259.10	395.80	4	32	259.00	26.41
1	12	507.00	329.30	1	55	808.59	245.21	2	47	268.10	132.40	3	38	389.30	-237.20	4	33	210.00	1112.00
1	13	251.50	83.36	1	84	-1243.00	125.36	2	48	-617.00	543.60	3	39	245.60	-133.90	4	34	-152.55	614.52
1	14	391.50	-30.48	1	85	637.65	221.56	2	49	62.56	161.10	3	40	22.67	140.60	4	35	4284.00	-143.20
1	15	255.70	65.33	2	3	38.81	3.45	2	55	200.94	384.45	3	41	103.50	317.60	4	36	-119.20	397.40
1	16	206.60	-83.98	2	4	74.15	-113.60	2	84	-861.46	-391.81	3	42	-432.30	787.90	4	37	-282.50	419.10
1	17	920.70	1139.00	2	5	524.10	457.00	2	85	424.93	629.96	3	43	238.90	234.40	4	38	101.40	-157.30
1	18	287.77	-101.56	2	6	787.60	-12.52	3	4	167.00	-146.80	3	44	30.04	-23.88	4	39	5629.00	-240.20
1	19	597.00	24.82	2	7	270.60	496.10	3	5	636.10	89.60	3	45	-88.11	167.90	4	40	-245.39	839.83
1	20	663.50	315.30	2	8	526.10	217.50	3	6	637.35	-50.00	3	47	333.30	-86.88	4	41	69.26	615.80
1	21	35.93	91.46	2	9	182.60	42.92	3	7	903.80	362.30	3	49	-59.58	142.90	4	44	46.38	6.21
1	22	53.76	34.01	2	10	448.75	56.30	3	8	1329.00	25.34	3	50	-39.16	23.93	4	47	421.90	-19.45
1	23	24.90	36.70	2	11	37.85	132.10	3	9	25.77	140.10	3	55	360.82	47.05	4	49	-203.60	274.10
1	24	104.30	-78.45	2	12	333.50	110.40	3	10	347.30	23.39	4	5	803.20	25.82	4	50	184.90	2.85
1	25	11.44	106.80	2	13	214.50	26.51	3	11	5.99	85.84	4	6	603.25	-44.50	4	55	233.51	347.13
1	26	661.50	-32.69	2	14	240.90	1.16	3	12	287.10	18.12	4	7	5695.00	377.60	5	6	-137.10	249.10
1	27	543.00	5541.00	2	15	163.90	-28.70	3	13	32.14	52.13	4	8	884.90	244.20	5	7	353.50	-229.10
1	28	153.60	-52.65	2	16	61.11	-25.38	3	14	161.70	-44.85	4	9	-52.10	365.80	5	8	-259.70	-451.60
1	29	184.40	-7.48	2	17	749.30	2000.00	3	15	122.80	-22.31	4	10	586.80	106.00	5	9	84.00	164.50
1	30	354.55	-25.31	2	18	280.50	-47.63	3	16	90.49	-223.90	4	11	5688.00	-170.00	5	10	-203.60	529.00
1	31	3025.00	139.93	2	19	336.90	-40.62	3	17	648.20	247.50	4	12	197.80	428.00	5	11	101.10	245.40
1	32	335.80	128.00	2	20	318.90	1264.00	3	18	-4.45	31.87	4	13	213.10	65.69	5	12	267.80	139.40
1	33	479.50	-31.52	2	21	-36.87	40.25	3	19	212.50	-22.97	4	14	19.02	296.40	5	13	28.06	237.70
1	34	298.90	-72.88	2	22	58.55	-23.50	3	20	537.40	62.32	4	15	-49.29	223.00	5	14	8.64	-242.80
1	35	526.50	50.49	2	23	-13.99	51.06	3	21	-18.81	4.68	4	16	23.50	109.90	5	15	42.70	-150.00
1	36	689.00	-165.90	2	24	-109.70	160.90	3	22	-144.40	121.30	4	17	664.20	762.80	5	16	-323.00	28.60
1	37	-4.19	47.41	2	25	100.10	70.32	3	23	-231.90	288.50	4	18	52.80	49.80	5	17	-52.39	-17.40
1	38	125.80	-5.13	2	26	357.50	-2.00	3	24	3.00	-4.70	4	19	6096.00	-138.40	5	18	170.03	-132.30
1	39	485.30	-31.95	2	28	76.30	16.62	3	25	187.00	-97.27	4	20	872.30	89.86	5	19	6.71	185.40
1	40	-2.86	147.30	2	30	262.90	82.64	3	26	168.04	10.38	4	21	-114.14	122.91	5	20	199.00	-151.00
1	41	387.10	529.00	2	33	183.80	174.60	3	27	194.90	1824.00	4	22	-111.00	140.78	5	21	75.62	562.20

Cont. Tabla D-2.

<i>m</i>	<i>k</i>	a_{mk}	a_{km}																
5	22	65.28	527.60	6	15	-20.98	-370.30	7	15	168.00	-448.20	8	25	6915.00	622.30	9	50	-162.60	278.80
5	23	-98.12	742.10	6	16	53.90	-406.80	7	16	304.00	-598.80	8	27	-413.48	815.12	10	11	-110.30	185.10
5	24	143.10	856.30	6	17	489.70	-118.10	7	17	243.20	-341.60	8	28	8483.50	1421.30	10	12	766.00	-236.50
5	25	123.50	325.70	6	18	580.48	-378.24	7	18	459.00	-332.90	8	31	-687.10	838.40	10	13	304.10	-7.84
5	26	256.50	261.60	6	19	53.28	162.60	7	19	112.60	242.80	8	41	-494.20	-167.30	10	19	-106.40	224.66
5	27	157.10	561.60	6	20	-202.00	339.80	7	20	-14.09	-66.17	8	44	-659.00	-234.70	10	20	497.50	-165.50
5	28	488.90	609.80	6	21	-38.32	529.00	7	21	325.44	698.24	8	50	1005.00	810.50	10	21	751.90	-47.51
5	29	147.50	461.60	6	22	-102.54	669.90	7	22	370.40	708.69	9	10	-37.36	128.00	10	22	67.52	190.60
5	30	-120.46	521.63	6	23	-139.35	649.10	7	23	353.68	826.76	9	11	-213.70	372.20	10	23	-483.70	242.80
5	31	-318.93	267.60	6	24	-44.76	709.60	7	24	497.54	1201.00	9	12	-190.40	385.40	10	32	117.00	245.90
5	32	313.50	501.30	6	25	-28.25	612.80	7	25	133.90	-274.50	9	13	-103.60	191.10	10	34	2.21	-55.87
5	33	202.10	524.90	6	26	75.14	252.56	7	26	220.60	417.90	9	15	-174.20	394.60	10	36	-339.20	354.00
5	34	727.80	68.95	6	27	457.88	511.29	7	27	399.50	360.70	9	16	-169.00	225.30	10	37	172.40	183.80
5	35	-202.10	-25.87	6	28	-31.09	914.20	7	28	887.10	1081.00	9	17	6201.00	-450.30	10	39	-268.80	13.89
5	36	74.27	389.30	6	29	17.50	448.60	7	30	188.03	23.48	9	18	7.34	29.10	10	41	-275.50	577.50
5	37	225.80	738.90	6	30	-61.76	287.00	7	31	12.72	-137.40	9	19	481.70	-287.50	11	12	-241.80	1167.00
5	38	44.78	649.70	6	31	-119.20	240.80	7	33	777.10	79.18	9	20	669.40	-297.80	11	13	-235.70	461.30
5	39	-143.90	64.16	6	32	212.10	431.30	7	35	-139.00	-240.00	9	21	-191.69	286.28	11	15	-73.50	136.00
5	41	190.30	88.63	6	33	106.30	494.70	7	36	160.80	386.60	9	22	-130.30	82.86	11	16	-196.70	2888.60
5	42	-817.70	1913.00	6	34	-119.10	967.71	7	39	319.00	-287.10	9	23	-354.55	552.10	11	17	475.50	-294.80
5	43	-1712.80	430.06	6	35	-399.30	695.00	7	41	-197.50	284.40	9	24	-39.20	372.00	11	18	-0.13	8.87
5	44	-504.20	796.90	6	36	-5.22	218.80	7	42	-363.80	180.20	9	25	-119.80	518.40	11	19	494.60	-266.60
5	45	72.96	794.40	6	37	33.47	528.00	7	44	-452.20	832.20	9	26	137.50	-142.61	11	20	660.20	-256.30
5	46	-382.70	394.80	6	38	-48.25	645.90	7	46	835.60	-509.30	9	27	548.50	-101.50	11	21	-34.74	35.38
5	47	-248.30	517.50	6	39	-172.40	172.20	7	47	139.60	-205.70	9	28	216.14	303.66	11	22	108.85	-132.95
5	49	104.70	-61.20	6	41	165.70	171.00	7	49	407.90	-384.30	9	29	-46.28	160.60	11	23	-209.66	176.45
5	50	57.65	682.50	6	45	-52.10	762.70	7	55	-255.63	627.39	9	30	-163.70	317.50	11	24	54.57	129.49
5	55	215.81	72.19	6	48	37.63	420.00	8	9	-356.10	-133.10	9	31	71.46	135.40	11	25	442.40	-171.10
5	84	-1840.80	111.65	6	49	-59.40	-89.24	8	10	-271.10	-155.60	9	32	53.59	138.00	11	26	-81.13	129.30
5	85	56.30	122.19	6	50	-46.01	597.80	8	11	-449.40	-36.72	9	33	245.20	-142.60	11	28	183.05	243.78
6	7	-180.95	289.60	6	55	150.02	265.75	8	12	-32.52	-234.25	9	34	-246.60	443.62	11	30	202.25	-146.31
6	8	-101.70	-265.20	7	8	324.50	-601.80	8	13	-162.87	-178.55	9	35	-44.58	110.40	11	31	-101.70	152.00
6	9	23.39	108.65	7	9	-195.40	472.50	8	14	-832.97	-870.80	9	36	-63.50	114.55	11	32	148.30	21.92
6	10	306.42	-340.18	7	10	-116.00	480.80	8	17	119.90	-253.10	9	37	-34.57	-40.90	11	33	18.88	24.37
6	11	-10.72	249.63	7	11	72.87	200.80	8	18	-305.50	-341.60	9	39	-61.70	97.04	11	34	71.48	-111.45
6	12	179.70	227.80	7	12	233.87	124.63	8	20	408.90	-11.00	9	41	-18.80	123.40	11	35	52.08	41.57
6	13	-128.60	238.40	7	13	540.50	-314.70	8	22	517.27	1633.50	9	42	-588.90	992.40	11	36	-28.61	175.53
6	14	359.30	-481.65	7	14	48.89	-330.40	8	24	1827.00	10000.00	9	47	37.54	156.40	11	37	-275.20	611.30

Cont. Tabla D-2.

<i>m</i>	<i>k</i>	<i>a_{mk}</i>	<i>a_{km}</i>																
11	39	85.33	-82.12	13	39	254.80	-158.20	17	20	-396.00	493.80	20	24	183.40	631.00	23	32	86.40	21.76
11	41	560.20	-234.90	13	40	-172.51	278.15	17	21	287.00	429.70	20	25	-79.08	993.40	23	33	168.80	-46.80
11	47	151.80	-3.44	13	41	417.00	-247.80	17	22	-111.00	140.80	20	30	-208.90	570.60	23	35	363.70	-343.60
12	13	-234.00	457.30	13	42	1338.00	448.50	17	24	882.00	898.20	20	32	228.40	616.60	23	37	111.20	-149.80
12	18	-233.40	554.40	14	15	-107.20	127.40	17	25	617.50	334.90	20	33	-95.00	5256.00	23	41	255.80	-193.00
12	19	-47.25	99.37	14	16	-41.11	38.89	17	27	-139.30	134.90	20	35	-463.60	-180.20	23	44	-35.68	-196.20
12	20	-268.10	193.90	14	17	-200.70	-15.07	17	31	0.10	192.30	20	37	-11.16	898.20	23	48	565.90	-363.10
12	22	31.00	80.99	14	19	358.90	-157.30	17	39	-334.40	343.70	20	39	-228.00	-97.77	24	25	-75.97	248.40
12	23	-126.20	235.60	14	21	-82.92	131.20	17	41	-89.42	-22.10	20	41	-337.00	1179.00	24	26	490.88	-34.68
12	24	179.70	351.90	14	24	-99.81	261.10	18	19	-169.67	134.28	20	46	-322.30	-70.25	24	27	534.70	514.60
12	25	24.28	383.30	14	25	30.05	108.50	18	20	-153.70	-313.50	21	22	108.31	-84.53	24	28	132.20	-60.71
12	29	103.90	201.50	14	29	-70.14	106.70	18	22	-351.60	587.30	21	23	249.15	-157.10	24	30	546.68	-133.16
12	33	298.13	-92.26	14	35	874.19	-366.51	18	23	-114.73	18.98	21	24	62.42	11.80	24	32	247.80	48.49
12	37	-11.40	134.50	14	39	-164.00	49.70	18	24	-205.30	368.50	21	25	153.00	-129.70	24	33	146.60	77.55
12	39	308.90	-116.70	14	42	-664.40	961.80	18	25	-2.17	20.18	21	26	32.73	113.00	24	35	337.70	-58.43
12	41	-122.30	145.40	14	43	275.90	-125.20	18	27	2845.00	2475.00	21	27	86.20	1971.00	24	36	369.49	-85.15
13	14	-78.36	222.10	15	16	-189.20	865.90	18	33	-60.78	-42.71	21	28	450.09	-73.09	24	37	187.10	-134.20
13	15	251.50	-56.08	15	17	138.54	64.30	18	37	160.70	281.60	21	29	59.02	-27.94	24	38	215.20	-124.60
13	16	5422.30	-194.10	15	18	431.49	-207.66	18	38	-158.80	159.80	21	30	65.56	-39.46	24	39	498.60	-186.70
13	17	-46.39	285.36	15	19	147.10	-108.50	18	50	-136.60	221.40	21	32	2.22	179.25	24	41	256.50	335.70
13	18	213.20	-156.10	15	24	71.23	91.13	19	20	205.27	92.07	21	33	344.40	-262.30	24	43	233.10	70.81
13	19	-18.51	38.81	15	25	-18.93	102.20	19	21	4.93	54.32	21	37	-168.20	383.20	24	47	423.10	3.16
13	20	664.60	-338.50	15	31	939.07	-213.74	19	22	-152.70	258.60	21	39	6.57	-55.21	24	48	63.95	-11.30
13	21	301.14	225.39	15	38	570.90	-198.80	19	23	-15.62	74.04	21	41	63.67	182.20	24	50	108.50	-79.34
13	22	137.77	-197.71	15	39	-255.22	10.03	19	24	-54.86	491.95	22	23	0.00	0.00	24	55	585.19	75.04
13	23	-154.30	-20.93	15	41	-38.77	284.50	19	25	-4.62	363.50	22	24	56.33	17.97	25	26	132.70	132.90
13	24	47.67	113.90	15	42	448.10	1464.20	19	26	-0.52	0.28	22	25	223.10	-8.31	25	27	2213.00	-123.10
13	25	134.80	-25.15	15	43	-1327.00	1603.80	19	28	230.85	335.74	22	26	108.90	-9.64	25	33	593.40	-185.30
13	26	95.18	-94.49	16	17	287.43	-24.46	19	29	0.46	161.00	22	30	149.56	-116.21	25	35	1337.37	-334.12
13	27	155.11	220.66	16	19	1255.10	-446.86	19	31	177.50	169.60	22	32	177.60	-40.82	25	39	5143.14	-374.16
13	28	140.90	112.38	16	21	-182.91	151.38	19	33	-62.17	136.90	22	33	315.90	-174.50	25	40	309.58	33.95
13	29	-8.54	63.71	16	22	-73.85	-141.40	19	34	-203.02	329.12	22	35	215.00	-215.00	25	41	-71.18	956.10
13	30	170.10	-87.31	16	23	-352.90	-293.70	19	36	81.57	-42.31	22	37	-91.80	301.90	25	44	-209.70	161.50
13	31	-20.11	9.21	16	24	-262.00	316.90	19	37	-55.77	335.20	22	39	-160.28	397.24	25	47	434.10	7.08
13	32	-149.50	476.60	16	25	-181.90	2951.00	19	39	-151.50	150.60	22	41	-96.87	305.40	26	27	533.20	-85.12
13	33	-202.30	736.40	16	35	243.10	-257.20	19	41	120.30	-61.60	22	47	361.10	-194.70	26	28	320.20	277.80
13	34	-156.57	173.77	16	38	-196.31	116.48	19	47	16.23	119.20	23	24	-30.10	51.90	26	31	139.82	481.35
13	35	128.80	-93.51	16	39	22.05	-185.20	20	21	13.41	519.10	23	25	192.10	-0.23	26	32	304.30	64.28
13	37	240.20	-217.90	17	18	89.70	117.40	20	22	-44.70	543.30	23	28	116.61	-26.06	26	33	10.17	125.30
13	38	-273.95	167.30	17	19	-281.60	777.40	20	23	39.63	504.20	23	30	-64.38	48.48	26	34	-27.70	174.43

Cont. Tabla D-2.

m	k	a_{mk}	a_{km}
26	37	10.76	379.40
26	39	-223.10	223.60
26	41	248.40	-124.70
26	45	-218.90	844.00
26	50	-4.57	176.30
27	32	2990.00	2448.00
27	33	-124.00	4288.00
28	32	292.70	-27.45
28	37	-47.37	167.90

m	k	a_{mk}	a_{km}
28	41	469.80	885.50
29	35	31.66	85.70
29	39	78.92	-71.00
29	44	1004.20	-274.10
29	48	-18.27	6.97
30	41	43.37	-64.28
31	35	-417.20	535.80
31	39	302.20	-191.70
31	41	347.80	-264.30

m	k	a_{mk}	a_{km}
31	44	-262.00	262.00
31	47	-353.50	515.80
32	33	6.37	37.10
32	41	68.55	288.10
33	35	32.90	-111.20
33	37	-48.33	322.42
33	39	336.25	-176.26
33	41	-195.10	627.70
34	37	2073.20	631.50

m	k	a_{mk}	a_{km}
34	39	-119.80	6.70
35	39	-97.71	136.60
35	41	153.70	-29.34
36	37	-208.80	837.20
36	39	-8.80	5.15
36	41	423.40	-53.91
37	39	255.00	-137.70
37	41	730.80	-198.00
37	44	26.35	-66.31

m	k	a_{mk}	a_{km}
37	48	2429.00	148.90
38	39	-110.65	50.06
38	40	-117.17	185.60
39	40	-5.58	55.80
39	41	72.31	-28.65
40	45	111.80	-32.17
41	47	122.40	101.20
42	43	-2166.00	745.30
84	85	1517.50	-1869.90

E. Manual de operación del programa CAMD

El presente anexo presenta un manual de operación básico del programa de diseño CAMD desarrollado en marco del presente trabajo investigativo: *Multi-Objective Resolution Interface Toted Up with Aromatics* (MORITA por sus siglas en inglés, Interfaz de resolución multiobjetivo equipada con aromáticos) en su versión más reciente.

La información base asociada para descarga del programa se encuentra disponible en el repositorio GitHub del autor de la presente tesis (<https://github.com/varoag>). Para la ejecución del programa, es posible clonar el repositorio mediante el enlace <https://github.com/varoag/TesisMsc.git> o descargar el instalador *SetupMORITA.exe* ubicado en la ruta del repositorio antes mencionado, e instalar de forma tradicional.

Cualquier desarrollado posterior, actualización o corrección del código fuente Python del proyecto asociado a este software debe contar con la autorización del autor del presente trabajo, el cual puede ser consultado mediante el correo institucional (varoag@unal.edu.co).

La elaboración de la presente herramienta buscó integrar los aspectos previos desarrollados tanto en el estudio de Serrato [1] como de Rodríguez [2], [3], junto con las modificaciones y adaptaciones formuladas en el presente trabajo. Se contó con la autorización de los autores para uso del código fuente como línea base.

La interfaz gráfica del programa construido cuenta con las siguientes 9 secciones, fácilmente identificables en la **Figura E-1**. A continuación, se realizará una breve descripción de las partes del programa. Para más información, se sugiere ir a la sección

de Manual de Ayuda del programa, mediante la ruta *Menú Ayuda / Manual de Ayuda*, presionando la tecla F1 o el botón  , ubicado en la barra de herramientas.

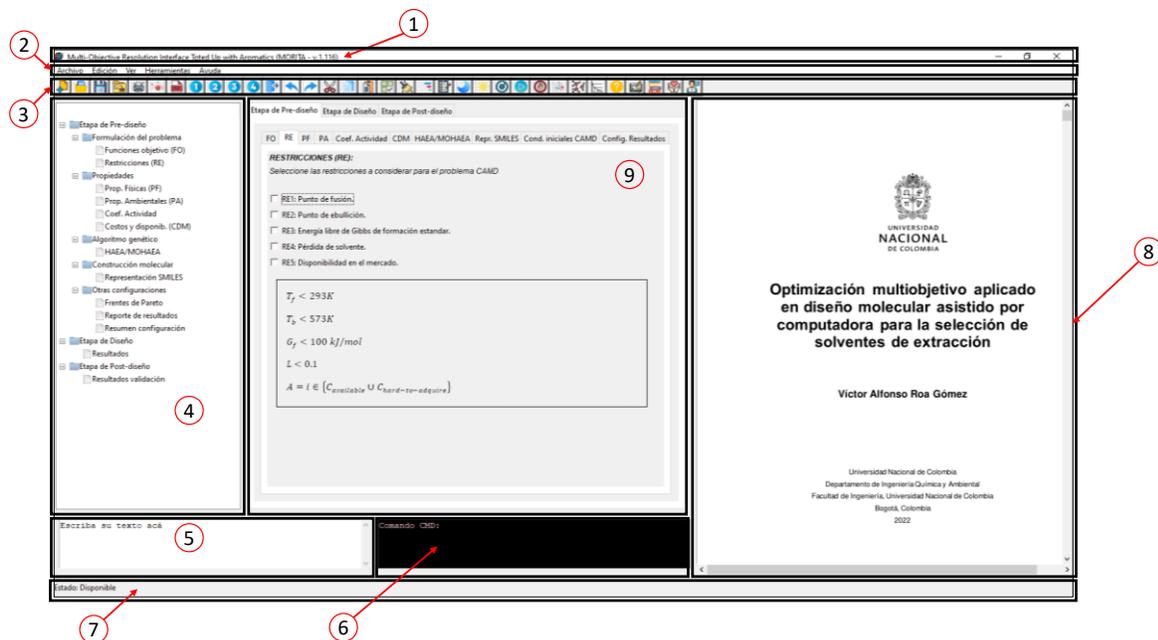


Figura E-1: Interfaz gráfica del programa *Multi-Objective Resolution Interface Toted Up with Aromatics (MORITA)*.

1. *Barra de título:* Esta situada en la parte superior de la ventana e indica el documento en donde se está trabajando. También cuenta con tres botones, para ajustar la configuración de pantalla (Minimizar, maximizar y cerrar).

2. *Barra de menús:* Esta localizada debajo de la barra de título y cuenta todas las herramientas y opciones del programa, dispuestas en menús desplegables. El programa cuenta con 5 menús principales descritos a continuación:

- *Menú Archivo:* Este menú contempla con todas las herramientas asociadas a la configuración del programa, incluyendo apertura de archivos, guardado y almacenamiento de resultados. El atajo rápido de este menú es Alt + A.
- *Menú Edición:* Este menú cuenta con todas las herramientas relacionadas con el manejo de texto, terminal y área de trabajo. El atajo rápido de este menú es Alt + E.

- *Menú Ver:* Este menú integra todas las configuraciones y personalizaciones de la interfaz gráfica del programa. El atajo rápido de este menú es Alt + V.
- *Menú Herramientas:* Este menú contempla todas las opciones para la ejecución del escenario CAMD. El atajo rápido de este menú es Alt + H.
- *Menú Ayuda:* Este menú cuenta con todas las opciones de Ayuda del programa, junto con los créditos y agradecimientos asociados al desarrollo del programa.

3. *Barra de herramientas:* Está ubicada en la parte superior del programa, debajo de la barra de menús. Permite acceder y visualizar de manera rápida las herramientas presentes en la barra de menús. De esta manera, es posible activar mediante botones las funciones específicas de la aplicación.

4. *Árbol de simulación:* Está localizado en el costado izquierdo centro del programa, siendo fácilmente identificable por la lista desplegable de carpetas y archivos, asociados a las pestañas descritas por el área de trabajo. Su finalidad es identificar de forma visual cuando las especificaciones de una respectiva pestaña son suministradas de forma correcta. En este caso, el icono presente para cierto ítem de la lista se modificará a una marca de chequeado o paloma (✓).

5. *Área de texto:* Está localizada en el costado izquierdo inferior del programa, justo debajo del árbol de simulación. Sirve como repositorio de las condiciones de operación empleadas para un experimento de diseño CAMD. Al especificar un conjunto de especificaciones en el área de trabajo y ejecutar el programa con el botón de Correr Escenario CAMD, las condiciones iniciales empleadas son escritas en formato de texto plano (.txt) en esta sección. Este archivo puede ser guardado y cargado posteriormente, como punto de partida para la ejecución del escenario.

6. *Terminal de comandos CMD:* Esta localizado en el costado inferior centro del programa, a la derecha del área de trabajo, sobre la barra de estado. Permite emplear comando de Python para la ejecución interna de las clases del programa, ajustes de configuración del sistema y otras especificaciones descritas con detalle en el Manual

de Ayuda del programa. Se integró con el fin de realizar una ejecución interna fácil y práctica de las clases, métodos y otras configuraciones escritas en el código fuente del programa.

7. *Barra de estado*: Está ubicada en la parte inferior del programa, consiste en un elemento de control gráfico que permite determinar el estado actual del programa (progreso y cumplimiento de ciertas acciones por parte de botones y menús, descripción de ejecución de los experimentos CAMD realizados y reporte de especificaciones de interés del programa).

8. *Área de resultados*: El área de resultados, ubicado en el costado derecho del programa, es el área encargada de la visualización de resultados, producto de un experimento realizado por el diseño CAMD. Reporta los resultados de un diseño realizado mediante la aplicación en formato PDF, el cual puede ser guardado a modo de reporte mediante la opción Exportar resultados como archivo PDF (.PDF), ubicado en el menú Archivo. También sirve como visor gráfico de documentación de interés (tales como el manual de ayuda, el listado de ecuaciones empleadas en el programa, el documento de tesis en sí, junto con los créditos y agradecimientos asociados a este proyecto).

9. *Área de trabajo*: Es el área principal del programa, ubicado en el centro de este. En esta área, se especifican las condiciones de interés para el desarrollo del diseño CAMD en las tres etapas metodológicas descritas: Prediseño, diseño y post-diseño. Cuenta con un conjunto de pestañas, que permiten clasificar los distintos tipos de especificaciones, permitiendo su fácil asociación con el árbol de simulación, para identificar el cumplimiento de todas las especificaciones requeridas para la ejecución del programa.

La **Tabla E-1** presenta un breve resumen de las herramientas disponibles en el programa para la ejecución de este, mediante la descripción de las tareas efectuadas por las herramientas, los iconos descriptivos presentes en la barra de herramientas y los comandos de atajo rápido empleables desde teclado.

Tabla E-1: Listado de herramientas y funcionalidades presentes en el programa *Multi-Objective Resolution Interface Toted Up with Aromatics* (MORITA).

Icono	Herramienta	Descripción de la función	Atajo rápido	Integrado en menú
	Nuevo	Crea una nueva instancia del programa.	Ctrl+N	Archivo
	Abrir	Abre un archivo de texto (.txt) inicial para asignación de las condiciones iniciales, el cual se presenta en el área de texto. También permite abrir archivos de tipo PDF, tales como resultados anteriores almacenados, los cuales se presentarán en el área de resultados.	Ctrl+O	Archivo
	Guardar	Guarda una configuración actual del experimento de diseño CAMD, en formato de archivo de texto (.txt), para su futuro uso. En caso de que el experimento ya haya sido nombrado, actualiza el contenido en el fichero de texto; en caso contrario, activa la acción de Guardar como.	Ctrl+S	Archivo
	Guardar como...	Guarda una configuración del experimento, mediante una ventana de dialogo, que requerirá nombre del archivo y una ubicación en el directorio.	Ctrl+G	Archivo
	Imprimir	Imprime los resultados obtenidos, visualizados en el área de resultados, mediante alguna de las impresoras configuradas previamente en Windows.	Ctrl+P	Archivo
	Vista Previa	Visualiza de forma rápida las moléculas generadas producto del diseño, en formato (.mol) y mediante representación SMILES (.txt)	Ctrl+W	Archivo
	Exportar resultados a PDF	Exporta los resultados presentados en el área de resultados y los almacena en formato PDF para su futura visualización.	Ctrl+F	Archivo

Cont. Tabla E-1.

Icono	Herramienta	Descripción de la función	Atajo rápido	Integrado en menú
	Mostrar resultados Objetivo 1	Muestra los resultados obtenidos en el estudio para el Objetivo 1, de acuerdo con la sección de Objetivos.	Ctrl+1	Archivo
	Mostrar resultados Objetivo 2	Muestra los resultados obtenidos en el estudio para el Objetivo 2, de acuerdo con la sección de Objetivos.	Ctrl+2	Archivo
	Mostrar resultados Objetivo 3	Muestra los resultados obtenidos en el estudio para el Objetivo 3, de acuerdo con la sección de Objetivos.	Ctrl+3	Archivo
	Mostrar resultados Objetivo 4	Muestra los resultados obtenidos en el estudio para el Objetivo 4, de acuerdo con la sección de Objetivos.	Ctrl+4	Archivo
	Salir	Permite salir del programa.	Alt+F4	Archivo
	Deshacer	Deshace la última acción realizada en el área de texto.	Ctrl+Z	Edición
	Rehacer	Rehace la última acción realizada en el área de texto.	Ctrl+Y	Edición
	Cortar	Corta el texto seleccionado.	Ctrl+X	Edición
	Copiar	Copia el texto seleccionado.	Ctrl+C	Edición
	Pegar	Pega el texto seleccionado.	Ctrl+V	Edición
	Seleccionar todo	Selecciona todo el contenido del área de texto.	Ctrl+A	Edición
	Limpiar todo	Limpia el contenido del área de texto, terminal de comandos CMD y área de trabajo.	Ctrl+L	Edición
	Mostrar / Ocultar Barra de Herramientas	Muestra u oculta la barra de herramientas, con el fin de generar mayor espacio de visualización.	Ctrl+H	Ver
	Mostrar / Ocultar Barra de Estado	Muestra u oculta la barra de estado, con el fin de generar mayor espacio de visualización.	Ctrl+U	Ver
	Ver Tema Oscuro	Visualiza la interfaz gráfica en el tema oscuro.	F11	Ver
	Ver Tema Claro	Visualiza la interfaz gráfica en el tema claro.	F12	Ver
	Reiniciar	Reinicia el estudio CAMD, limpiando el área de trabajo y el área de resultados.	F5	Herramientas

Cont. Tabla E-1.

Icono	Herramienta	Descripción de la función	Atajo rápido	Integrado en menú
	Correr escenario CAMD	Ejecuta el diseño CAMD descrito en el área de trabajo, generando el análisis de datos en el área de resultados. Adicionalmente, adiciona al área de texto una descripción resumida de las condiciones iniciales suministradas en el estudio.	F6	Herramientas
	Parar ejecución	Detiene la ejecución del programa, de acuerdo con el criterio del usuario.	F7	Herramientas
	Graficar resultados	Presentar los resultados del diseño CAMD de f	F8	Herramientas
	Resumen de componentes óptimos	Genera un resumen de resultados en un archivo de Excel (.xlsx) donde se reportan los componentes óptimos generados producto del estudio realizado.	F9	Herramientas
	Diagramas de Pareto	Visualiza los resultados del análisis multiobjetivo, mediante la presentación del diagrama de Pareto.	F10	Herramientas
	Manual de ayuda	Presenta en el área de resultados en formato PDF el manual de ayuda del programa.	F1	Ayuda
	Ecuaciones y algoritmos	Presenta en el área de resultados en formato PDF un listado resumido de las ecuaciones	F2	Ayuda
	Documento de Tesis	Presenta en el área de resultados en formato PDF el documento de tesis asociado al programa.	Ctrl+M	Ayuda
	Créditos y agradecimientos	Presenta en el área de resultados en formato PDF los créditos y agradecimientos asociados a este trabajo.	Ctrl+T	Ayuda
	Acerca de	Reporta aspectos de la versión del programa y descripción del proyecto desarrollado.	Ctrl+D	Ayuda

F. Registro fotográfico de las extracciones líquido-líquido realizadas

Durante el desarrollo de la metodología descrita en la Sección 2.4, se tomó registro fotográfico de las extracciones realizadas para los sistemas LA/W y AA/W, las cuales son reportadas por las **Figura F-1** y **Figura F-2**, respectivamente.

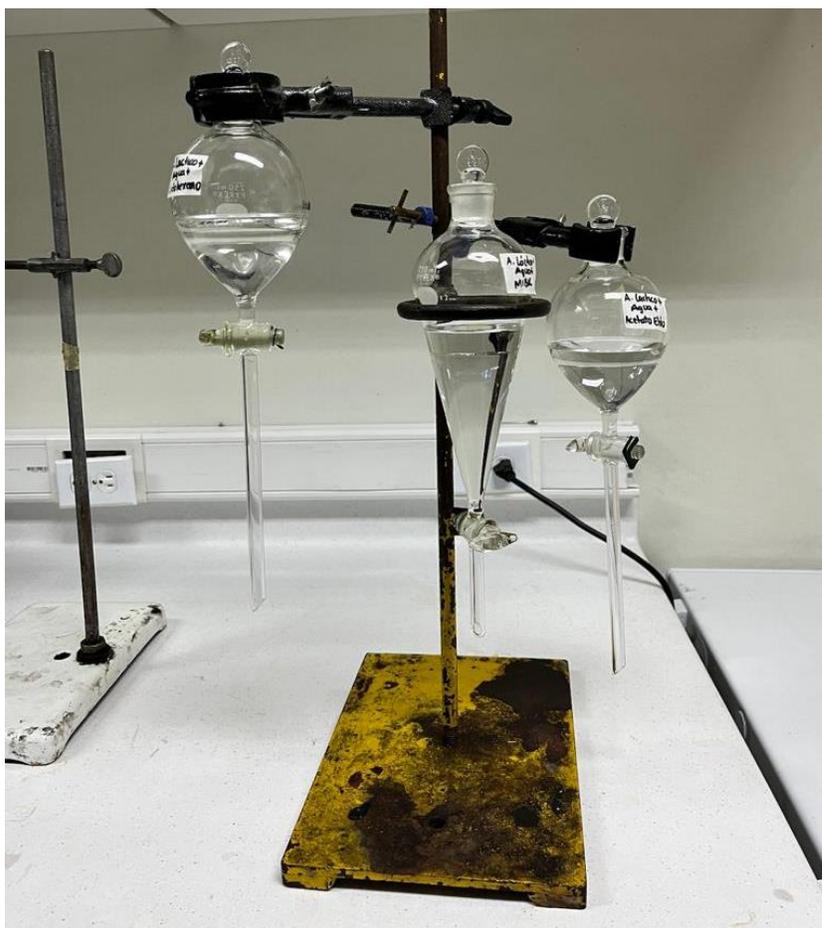


Figura F-1: Montaje realizado para las extracciones realizadas en el sistema LA/W. con distintos solventes. Izquierda: Ciclohexano (CyC6), Centro: Metil Isobutil Cetona (MIBK), Derecha: Acetato de etilo (AE).



Figura F-2: Montaje realizado para las extracciones realizadas en el sistema AA/W. con distintos solventes. Izquierda: Acetato de etilo (AE), Centro: Metil Isobutil Cetona (MIBK), Derecha: Ciclohexano (CyC6).

Una vez realizada la extracción, se realizaron las fases orgánicas y acuosa, realizando la recolección de cada una de las fases para cada sistema con los tres solventes empleados. La **Figura F-3** y la **Figura F-4** presentan las cantidades almacenadas de las fases acuosa

en los sistemas LA/W y AA/W. Por su parte, la **Figura F-5** y **Figura F-6** presentan las cantidades almacenadas de la fase orgánica en los sistemas estudiados.

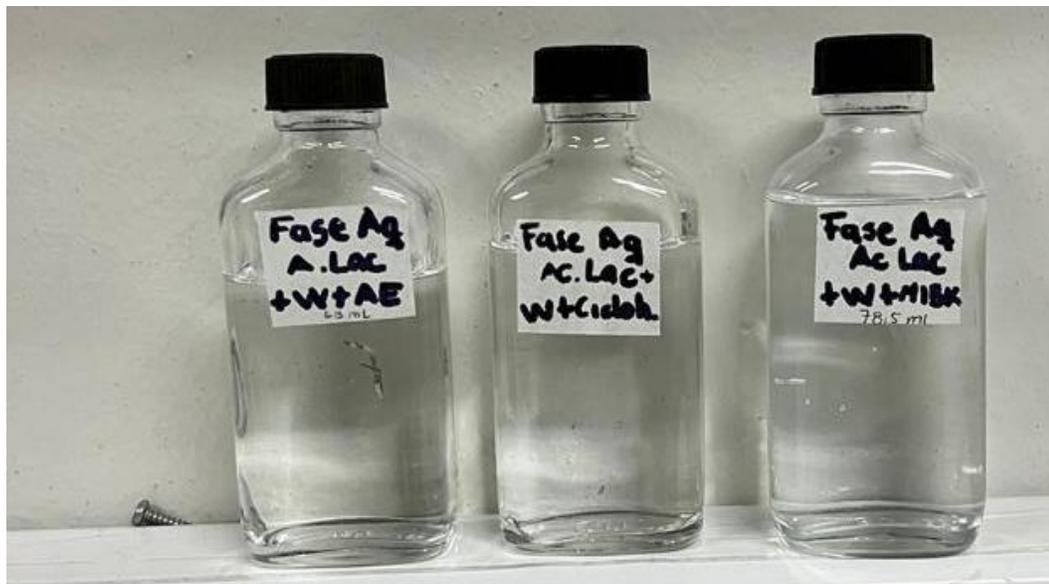


Figura F-3: Fase acuosa obtenida en el sistema LA/W. Izquierda: Acetato de Etilo (AE), Centro: Ciclohexano (CyC6), Derecha: Metil Isobutil Cetona (MIBK).



Figura F-4: Fase acuosa obtenida para el sistema AA/W. Izquierda: Acetato de Etilo (AE), Centro: Ciclohexano (CyC6), Derecha: Metil Isobutil Cetona (MIBK).

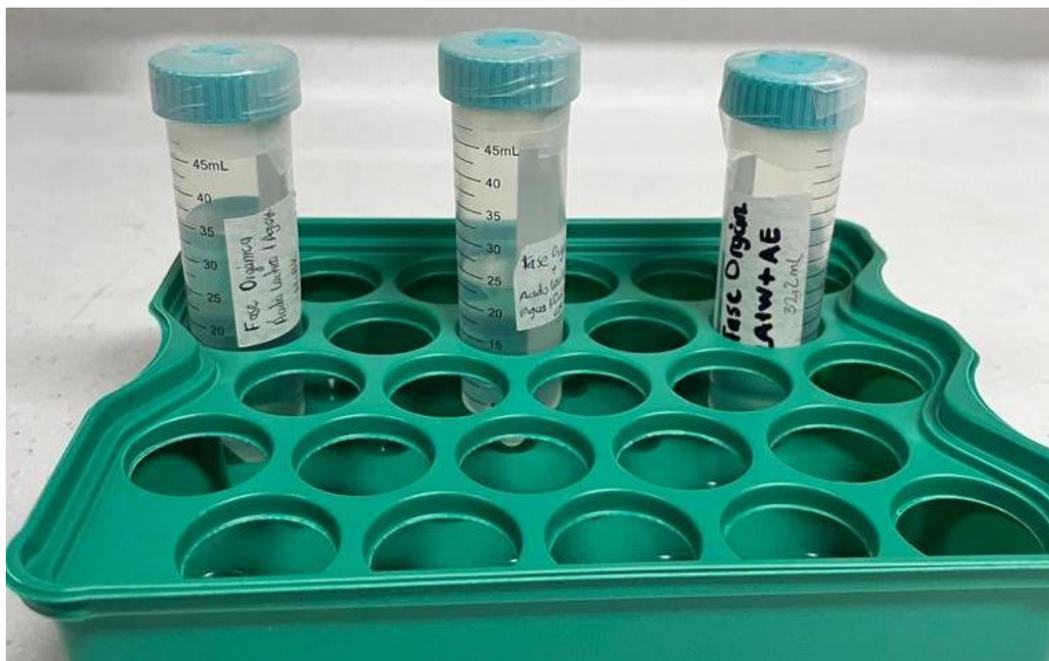


Figura F-5: Fase orgánica obtenida para el sistema LAW. Izquierda: Metil Isobutil Cetona (MIBK), Centro: Ciclohexano (CyC6), Derecha: Acetato de Etilo (AE).



Figura F-6: Fase orgánica obtenida para el sistema LAW. Izquierda: Metil Isobutil Cetona (MIBK), Centro: Ciclohexano (CyC6), Derecha: Acetato de Etilo (AE).

G. Método de cromatografía HPLC y curvas de calibración

De acuerdo con la metodología propuesta en la Sección 2.4.1, se realizó la determinación de la concentración final de ácido en la fase acuosa ($C_{aq,f}^{acid}$), se empleó el método de cromatografía líquida de alta eficacia (HPLC) empleando el equipo Thermo Scientific Ultimate 3000, con un detector IR, detallado en la **Figura G-1**.

Las características del método utilizado fueron:

- Columna: Shodex-KC811.
- Fase móvil: Agua desionizada con 0.1% H_3PO_4 .
- Flujo: 1.0 mL/min.
- Temperatura del horno: 40°C.
- Detector: Índice de refracción (IR) Shodex a una temperatura de 40°C.
- Tiempo de corrida por muestra: 15 min.



Figura G-1: Montaje experimental HPLC para la determinación de concentración final de ácido en fase acuosa.

Para identificar la calidad de la detección, la facilidad de la separación y los tiempos de separación de los ácidos de interés se preparó una mezcla de ácidos de 0,2mol/L, obteniendo los datos iniciales reportados en la **Tabla G-1**.

Tabla G-1: Datos iniciales establecidos en la detección de ácido acético y láctico mediante cromatografía HPLC.

Ácido orgánico	Concentración (g/L)	Tiempo de retención (min)
Ácido acético	12,01	Aprox. 9,2
Ácido láctico	18,02	Aprox. 7,9

Considerando que, de acuerdo con la **Tabla G-1**, los ácidos acético y láctico no se solapan durante su cuantificación, se consideró procesar las muestras de los perfiles cromatográficos en de ambos ácidos en la misma corrida experimental. No se consideró factor de dilución en los cálculos de concentración de las curvas de calibración.

Para la construcción de las curvas de calibración, se preparó una solución stock para cada ácido de 2 mM. Posteriormente, se realizaron diluciones seriadas a 10, 30, 50, 70 Y 100 g/L. Con las áreas obtenidas para cada concentración, se realizó una curva relacionando el área y la concentración, considerando para cada un punto el valor promedio de los ensayos triplicados realizados. Finalmente, se determinó el coeficiente de correlación para las curvas de calibración, siendo próximo a 1,0, lo cual evidencia un buen ajuste de datos experimentales al comportamiento lineal esperado.

A continuación, se presentan los datos y curvas de calibración determinadas para el sistema ácido láctico-agua y ácido acético-agua, que posteriormente fueron empleadas en el cálculo de las concentraciones iniciales y finales de los ácidos en los experimentos del trabajo.

Para el sistema ácido láctico – agua (LA/W), la

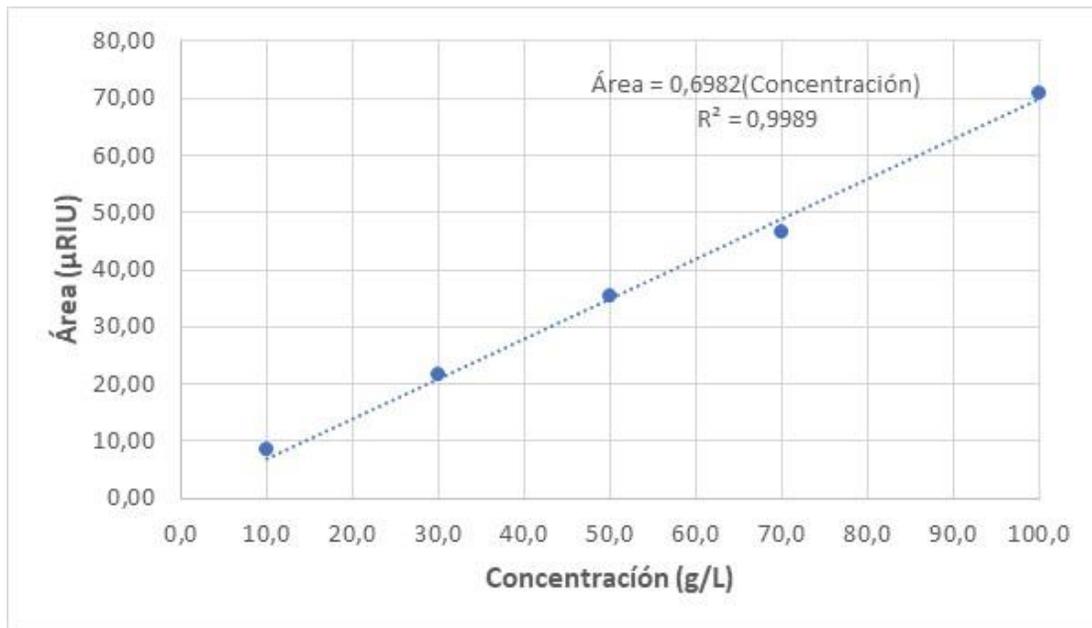


Figura G-2: Curva de calibración para el sistema ácido láctico-agua (LA/W).

Tabla G-2 presentan las áreas obtenidas para cada concentración de la curva de calibración, los cuales generaron la ecuación de calibración lineal (G-1), mediante el promedio de los ensayos triplicados de cada muestra, de acuerdo con la **Figura G-2**.

$$\text{Concentración (g/L)} = 1,4322(\text{Área}(\mu\text{RIU})) \quad (\text{G-1})$$

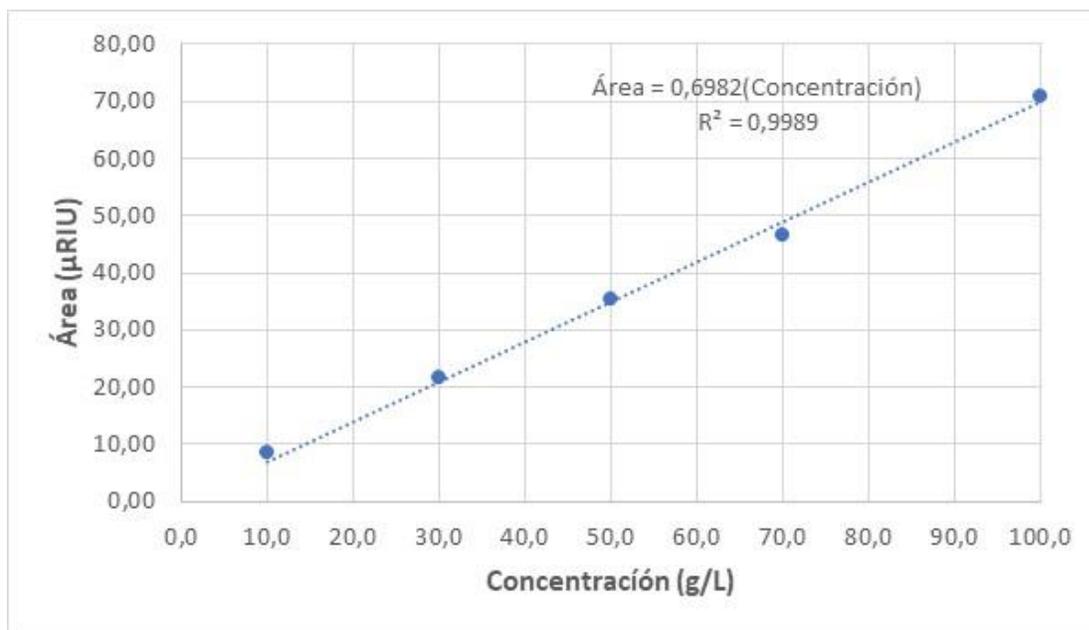


Figura G-2: Curva de calibración para el sistema ácido láctico-agua (LA/W).

Tabla G-2: Datos obtenidos para la calibración del sistema LA/W

Corrida	Concentración (g/L)	Área (µRIU)	Medidas de dispersión	
			Media (\bar{x})	Desviación Estándar (σ)
LAW-1-1	10	8,9934	8,4501	0,6119
LAW-1-2	10	8,5698		
LAW-1-3	10	7,7872		
LAW-2-1	30	22,1547	21,4919	2,0777
LAW-2-2	30	23,1573		
LAW-2-3	30	19,16360		
LAW-3-1	50	32,8570	35,4462	6,1857
LAW-3-2	50	30,9759		
LAW-3-3	50	42,5058		
LAW-4-1	70	46,4180	46,4502	0,4443
LAW-4-2	70	46,0229		

Cont. **Tabla G-2.**

Corrida	Concentración (g/L)	Área (µRIU)	Medidas de dispersión	
			Media (\bar{x})	Desviación Estándar (σ)

LAW-4-3	70	46,9097	46,4502	0,4443
LAW-5-1	100	88,6585	70,9374	16,3167
LAW-5-2	100	67,6181		
LAW-5-3	100	56,5356		

Para el sistema ácido acético – agua (AA/W), la **Tabla G-3** presentan las áreas obtenidas para cada concentración de la curva de calibración, los cuales generaron la ecuación de calibración lineal (G-2), mediante el promedio de los ensayos triplicados de cada muestra, de acuerdo con la **Figura G-3**.

$$\text{Concentración (g/L)} = 1,0810(\text{Área}(\mu\text{RIU})) \quad (\text{G-2})$$

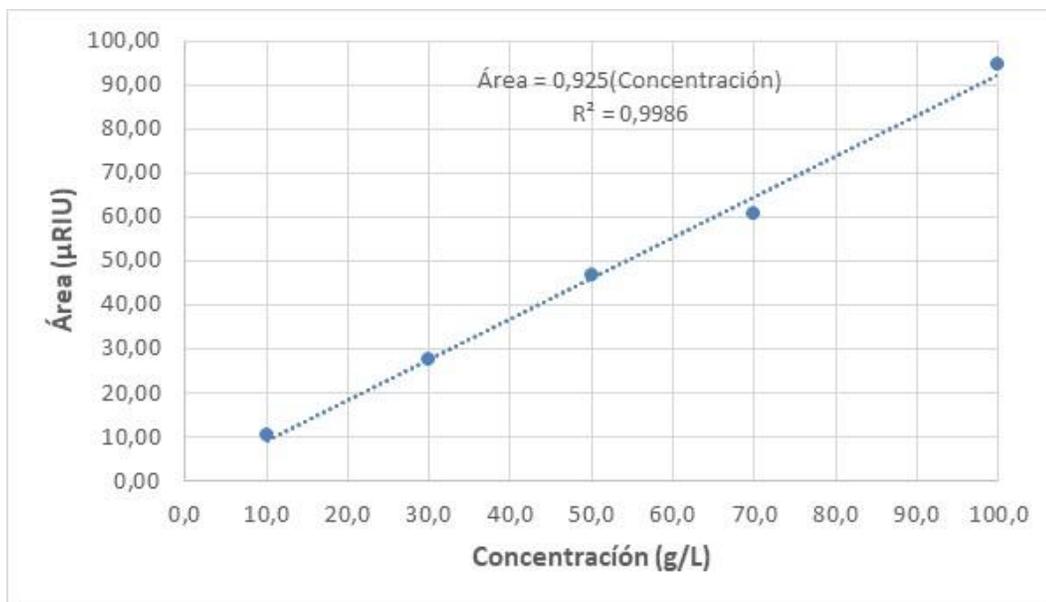


Figura G-3: Curva de calibración para el sistema ácido acético-agua (AA/W).

Tabla G-3: Datos obtenidos para la calibración del sistema AA/W

Corrida	Concentración (mM)	Área (μ RIU)	Medidas de dispersión	
			Media (\bar{x})	Desviación Estándar (σ)
AAW-1-1	10	11,8396	10,2519	1,4370
AAW-1-2	10	9,8757		
AAW-1-3	10	9,0405		
AAW-2-1	30	28,8956	27,5371	2,7914
AAW-2-2	30	29,3892		
AAW-2-3	30	24,3265		
AAW-3-1	50	43,4815	46,9607	8,4182
AAW-3-2	50	40,8399		
AAW-3-3	50	56,5608		
AAW-4-1	70	62,5251	60,8667	1,7395
AAW-4-2	70	59,0561		
AAW-4-3	70	61,0188		
AAW-5-1	100	115,3314	94,8207	19,0633
AAW-5-2	100	91,4858		
AAW-5-3	100	77,6448		

Con el fin de determinar la significancia estadística de las curvas de calibración realizadas para los sistemas LA/W y AA/W, la **Tabla G-4** se presentan los errores asociados a la regresión lineal aplicada. Considerando que el R^2 obtenido es superior a 0,99, se observa una clara dependencia estadística en una correlación lineal para las variables relacionadas (Concentración y Área). Adicionalmente, los valores de los errores SSE y RMSE identifican una baja dispersión de los datos regresados, lo suficientemente aceptable para el ajuste.

Tabla G-4: Bondad de ajuste de la regresión lineal para las curvas de calibración

Sistema	Pendiente	SSE	R^2	R^2 ajustado	RMSE
LA/W	0,6982	644,9	0.9142	0.9276	7.043
AAW	0,925	960,4	0,92887	0.9234	8.595

SSE = Squared Estimate of Errors, RMSE = Root-Mean-Square Error.

H. Método de titulación de Karl Fischer y volumetrías ácido-base

Para la determinación de agua por titulación de Karl Fisher, se empleó un titulador automático Mettler Toledo T50, descrito por **Figura H-1**. Este método aplica para determinar muestras con agua en el intervalo de 0,02 a 2% en peso de la muestra. Los mercaptanos y sulfitos interfieren con el resultado por este método en niveles superiores a 500 ppm. Por debajo de este valor, la interferencia es insignificante. El método aplicado empleó el reactivo de Karl Fischer libre de piridina.



Figura H-1: Montaje experimental Karl Fischer para determinación de la concentración final de agua en la fase orgánica.

Para determinar el contenido de agua en una muestra es necesario realizar la estandarización del reactivo de Karl Fisher empleado en la titulación, utilizando como estándar agua o tartrato ácido de sodio. El equivalente de agua del reactivo de Karl Fisher se obtiene mediante la ecuación (H-1) si se estandariza con agua o la ecuación (H-2) si se estandariza con tartrato ácido de potasio.

$$F = \frac{W}{T} \quad (\text{H-1})$$

$$F = \frac{0,1566 W}{T} \quad (\text{H-2})$$

Donde:

F = Equivalente de agua del reactivo de Karl Fischer, mg/mL.

W = Masa de agua o tartrato ácido de sodio añadida, mg.

T = Reactivo de Karl Fischer requerido para titulación del agua, mL.

El contenido de agua de la muestra problema, se obtiene por medio de la titulación con el reactivo de Karl Fischer, empleando la ecuación (H-3).

$$\text{Contenido de agua (\% en peso)} = \frac{10 CF}{w} \quad (\text{H-3})$$

Donde:

C = Reactivo de Karl Fischer necesario para titular la muestra, mL.

w = Masa de la muestra, mg.

Todos los anteriores datos requeridos para la titulación (F , W , T , C , w) son medidos directamente por el titulador automático Mettler Toledo T50.

Para la determinación de agua mediante el método de Karl Fischer, debe neutralizarse con este reactivo 60mL de metanol, el cual será utilizado como solvente de la muestra. Luego, se realiza la determinación de la deriva, es decir, la cantidad de agua que puede absorber el solvente del aire por unidad de tiempo, y se determina la concentración real del reactivo

de Karl Fischer titulando 0,1g de tartrato de sodio. Posteriormente, se añade a las cantidades especificadas en este procedimiento la muestra problema para la determinación de su contenido de agua. Puede usarse patrón de agua, tolueno o sustancias de control para asegurar la validez de los resultados. La **Figura H-2** presenta el flujograma de la metodología aplicada para la determinación de agua mediante el método de Karl Fischer.

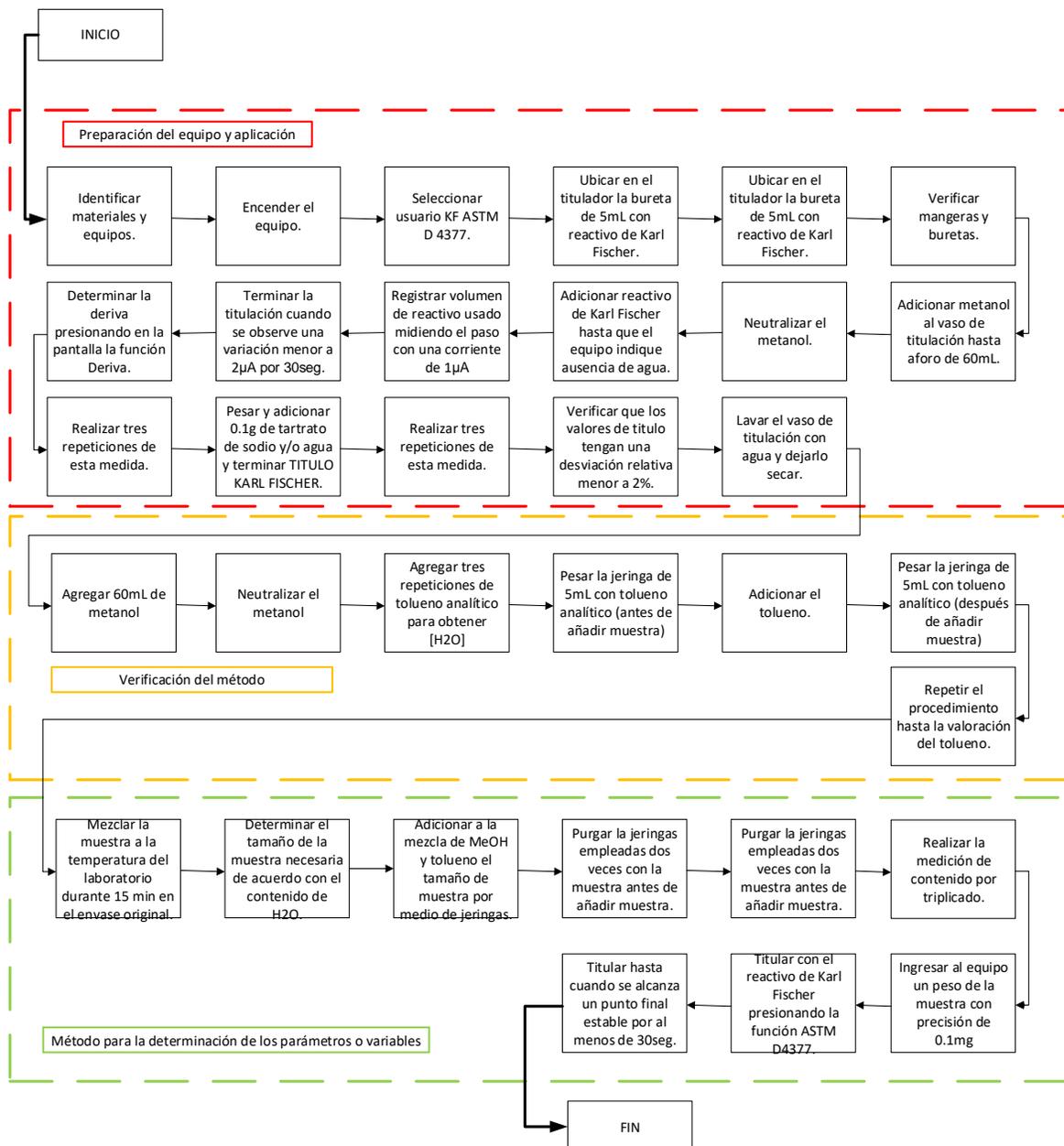


Figura H-2: Flujograma metodológico para la determinación de agua mediante el método de Karl Fischer.

La **Tabla H-1** reporta los datos obtenidos para la determinación de la deriva, en el equipo. Se realizaron dos duplicados de medición de deriva, dada la buena calidad actual del equipo de Karl Fischer. El reactivo de Karl Fischer fue suministrado estandarizado con tolueno al 99%, realizando un par de mediciones con cuyo porcentaje de agua indicaba solo un 0,02% de agua en el tolueno. Por otra parte

Tabla H-1: Datos obtenidos para la determinación de la deriva.

Tiempo de valoración (s)	Volumen añadido (mL)	Deriva reportada ($\mu\text{L}/\text{min}$)
304,200	38,75	7,643

La **Figura H-3** reporta el registro fotográfico realizado en las titulaciones de Karl Fischer para los sistemas evaluados. Por su parte, la **Tabla H-2** presenta los datos obtenidos en la verificación del método de Karl Fischer, realizada con patrón de agua y tolueno.

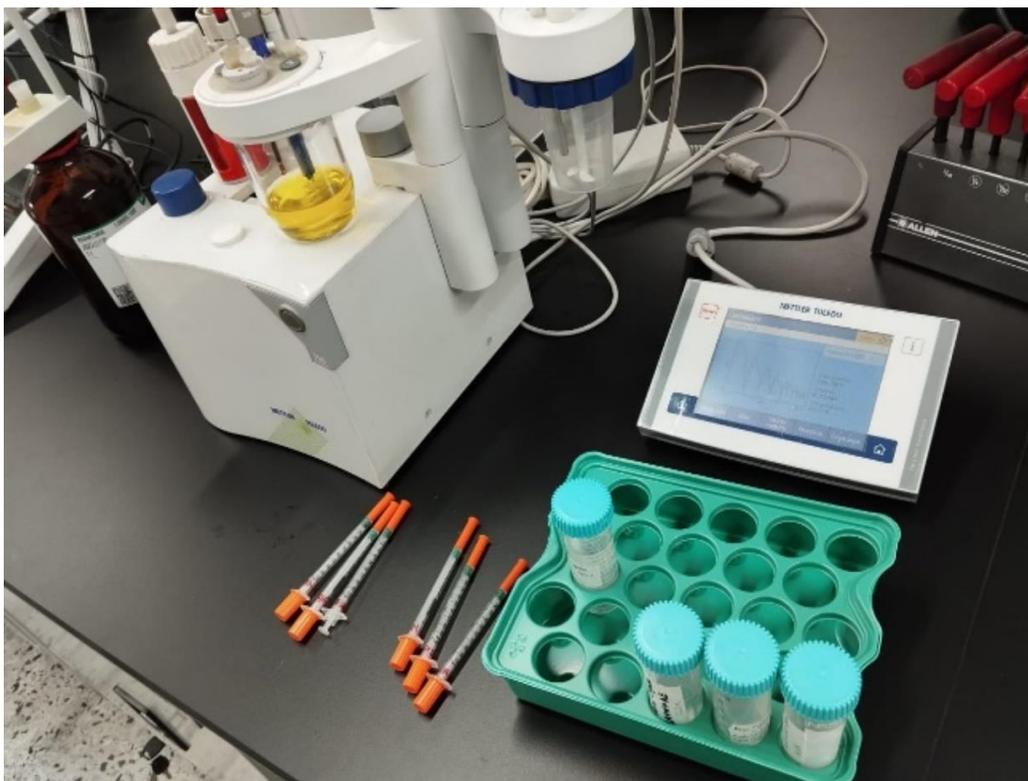


Figura H-3: Registro fotográfico de las titulaciones de Karl Fisher realizada para los sistemas LA/W y AAW.

Tabla H-2: Datos obtenidos en la verificación del método de Karl Fischer.

Sistema	Ensayo	Peso (g) de muestra de verificación	Volumen añadido de reactivo KF* (mL)	Tiempo de ejecución (min)	% de agua reportado	Promedio	Desviación estándar
LA/W con solvente AE	1	0,2164	1,59950	3,900	3,685%	3,121%	0,00514
	2	0,2763	1,65775	3,500	2,999%		
	3	0,2508	1,35025	3,660	2,679%		
LA/W con solvente CyC6	1	0,0931	0,1225	1,960	0,585%	0,362%	0,00193
	2	0,1908	0,10875	1,700	0,255%		
	3	0,2061	0,11275	1,710	0,246%		
LA/W con solvente MIBK	1	0,1383	0,55225	2,910	1,947%	1,821%	0,00266
	2	0,1894	0,60300	5,010	1,515%		
	3	0,1865	0,77875	5,830	2,000%		
AA/W con solvente AE	1	0,2580	1,49525	3,800	2,887%	3,109%	0,00211
	2	0,2609	1,64875	5,310	3,132%		
	3	0,2779	1,85100	5,460	3,308%		
AA/W con solvente CyC6	1	0,1946	0,08200	1,680	0,181%	0,402%	0,00398
	2	0,2251	0,40900	3,630	0,861%		
	3	0,2304	0,08675	1,680	0,163%		
AA/W con solvente MIBK	1	0,2593	0,80475	3,350	1,527%	1,664%	0,00129
	2	0,2065	0,72150	4,950	1,682%		
	3	0,2514	0,92250	5,200	1,784%		

Como criterio de calidad de la medida, el título del reactivo de Karl Fischer debe encontrarse entre 0,8 y 1,2.

Con respecto a las volumetrías ácido-base para la determinación de ácido láctico y ácido acético en fase orgánica, se aplicó la metodología experimental tradicional para determinación de acidez por titulación o valor ácido. La **Tabla H-3** reporta las condiciones de titulación para la determinación de estas concentraciones.

Tabla H-3: Condiciones experimentales de las titulaciones realizadas.

Sistema	Condición experimental	Cuantificación de $C_{org,f}^{acid}$
LA/W	Patrón primario	Ácido láctico 99,5%
	Titulante	NaOH
	Indicador ácido-base	Fenolftaleína
	Intervalo de viraje y color del indicador	8,2 – 9,6 (inoloro a fucsia)
AA/W	Patrón primario	Ácido acético glaciado
	Titulante	NaOH
	Indicador ácido-base	Fenolftaleína
	Intervalo de viraje y color del indicador	8,2 (inoloro) – 9,6 (rojo fucsia)

I. Grupos estructurales heterocíclicos añadidos al modelo CAMD

El presente anexo profundizar en la forma de visualización de los distintos grupos estructurales heterocíclicos identificados y añadidos al modelo CAMD. Tanto en el caso de las estructuras monocíclicas como en las estructuras policíclicas, la cantidad de estructuras obtenidas supera ampliamente una extensión escrita aceptable de un anexo al presente documento, por lo que para facilitar la visualización de las diversas estructuras, se empleó el uso del repositorio GitHub del autor de la presente tesis (<https://github.com/varoag>).

Una vez descargado el programa MORITA, en el directorio raíz principal, es posible identificar una subcarpeta llamada "Metodologia_3_Filtros" donde es posible identificar una serie de subcarpetas asociadas, descritas a continuación:

- Cálculos Combinatoriales: En esta carpeta se presentan todos los cálculos combinatoriales realizados para cada uno de los 24 grupos cíclicos evaluados (4 monocíclicos, 10 policíclicos espiro y 10 policíclicos fusionados). Toda esta información se presenta en formato de hoja de cálculo Excel (*.xlsx).
- Conjuntos Universales: En esta carpeta se reportan todas las estructuras generalizadas iniciales, empleadas como punto de partida para la evaluación de las estructuras cíclicas, en formato imagen (*.jpg y *.png) y formato molécula (*.mol). Esta carpeta cuenta con tres subcarpetas, relacionadas a compuestos monocíclicos (4 grupos: MC3, MC4, MC5 y MC6), compuestos policíclicos espiro (10 grupos: PC-E1-33, PC-E2-34, PC-E3-35, PC-E4-36, PC-E5-44, PC-E6-45, PC-E7-46, PC-E8-55, PC-E9-56, PC-E10-66) y compuestos policíclicos fusionados (10 grupos: PC-F1-33, PC-F2-34, PC-F3-35, PC-F4-36, PC-F5-44, PC-F6-45, PC-F7-46, PC-F8-55, PC-F9-56, PC-F10-66).

- Filtro I – Filtrado de grupos no isomorfos: En esta carpeta se presentan todos los resultados del proceso de filtrado de los conjuntos universales de los 24 grupos cíclicos anteriores, utilizando los criterios de simetría definidos por el cálculo de órbitas (grupos diedrales y simétricos), y validados por el Teorema Enumerador de Pólya. Estos criterios de simetría fueron sintetizados en un conjunto de vectores de entrada para cada grupo cíclico, adjuntos en formato .xlsx. A diferencia de la carpeta de conjuntos universales, las estructuras del Filtro I son clasificadas exclusivamente en formato molécula (*.mol).

Se presentan tres carpetas, una para cada tipo de grupo cíclico (monocíclico, policíclico espiro y policíclico fusionado) donde se reportan las siguientes subcarpetas:

- Carpeta con el identificador del grupo (MC, PC-E, PC-F): Datos crudos obtenidos.
- Ordenados MC/PC-E/PC-F: En esta carpeta se encuentran todas las combinaciones estructurales existentes en el conjunto universal, clasificadas por igualdad estructura (isomorfismo) dentro de subcarpetas respectivas, para todas las combinaciones de estructuras existentes. Dos estructuras dentro de la misma carpeta serán estructuras iguales, con diferencias exclusivas en la rotación espacial.
- ENI Resumidos MC/PC-E/PC-F: Una vez ordenados las distintas estructuras isomorfas, en esta carpeta se presentan en formato mol un resumen de todas las estructuras no isomorfas identificadas a partir del conjunto universal.

Adicionalmente, se anexan los distintos scripts en Python empleados para la generación de estructuras, cálculo de estructuras ENI, clasificación y resumen.

- Filtro II – Filtrado por cumplimiento de valencia: En esta carpeta se presentan todos los resultados de la identificación, filtrado por valencia y clasificación de las

estructuras que superaron el Filtro I, una vez fueron modificados los R-átomos por heteroátomos (C, N, O, S). Para facilitar la modificación atómica, se empleó un vector de permutación particular para cada uno de los 24 grupos cíclicos evaluados. A diferencia de la carpeta de conjuntos universales, las estructuras del Filtro II son clasificadas exclusivamente en formato molécula (*.mol).

Se presentan tres carpetas, una para el cálculo de cada tipo de grupo cíclico (monocíclico, policíclico espiro y policíclico fusionado) donde se reportan los resultados en dos tipos de subcarpetas.

- Valencias con el identificador del grupo (MC, PC-E, PC-F): Datos crudos obtenidos del Filtro II. Dada la configuración del Script de generación de moléculas con valencias válidas, algunos de estos grupos se encuentran vacíos, dado que la herramienta solo genera las estructuras con valencia cero. Por esta razón algunos grupos isomorfos pueden encontrarse vacíos.
- No vacíos – Valencias MC/PC-E/PC-F: Una vez generados todas las estructuras con valencias válidas, este directorio reporta la información simplificada, eliminando los directorios vacíos anteriores.
- Filtro III – Verificación de Hukkerikar y UNIFAC-DMD: En esta carpeta se presentan todos los resultados de la evaluación de las estructuras por los modelos de Hukkerikar y UNIFAC-DMD. Para cada uno de los 24 grupos, las estructuras que superaron el Filtro II son clasificadas en cuatro clases de carpetas: “Solo UNIFAC-DMD”, “Solo Hukkerikar”, “Ambos modelos”, “Ningún modelo”. A diferencia de la carpeta de conjuntos universales, las estructuras del Filtro III son clasificadas exclusivamente en formato molécula (*.mol).

A partir de esta estructuración de información, fue posible obtener los valores de estructuras que superaron cierto Filtro en particular. El autor del presente trabajo sugiere tener precaución en la apertura de estos directorios, dada la extensión de archivos molécula (*.mol) presentes. Para visualizar cada uno de los archivos molécula, es posible emplear software libre, tal como ACD Labs – ChemSketch [261].

Bibliografía

- [1] J. C. Serrato, "Diseño computacional de agentes de extracción para la separación de compuestos orgánicos en corrientes acuosas. Aplicación al ácido láctico.", Universidad Nacional de Colombia - Sede Bogotá, 2009. doi: 10.7202/1016404ar.
- [2] K. A. Rodríguez, "Inclusion of toxicity and market availability in a Computer-Aided Molecular Design methodology", 2019.
- [3] K. Rodríguez, "Computer Aided Molecular Design : State-of-the-art and environmen- tally friendly product design", p. 9, 2019.
- [4] E. Stefanies, L. Constantinou, y C. Panayiotou, "A group-contribution method for predicting pure component properties of biochemical and safety interest", *Ind Eng Chem Res*, vol. 43, núm. 19, pp. 6253–6261, 2004, doi: 10.1021/ie0497184.
- [5] M. R. Eden, S. B. Jørgensen, R. Gani, y M. M. El-Halwagi, "A novel framework for simultaneous separation process and product design", *Chemical Engineering and Processing: Process Intensification*, vol. 43, núm. 5, pp. 595–608, 2004, doi: 10.1016/j.cep.2003.03.002.
- [6] J. Marrero y R. Gani, "Group-contribution based estimation of pure component properties", *Fluid Phase Equilib*, vol. 183–184, pp. 183–208, 2001, doi: 10.1016/S0378-3812(01)00431-9.
- [7] J. Rydberg, M. Cox, C. Musikas, y G. R. Choppin, *Solvent Extraction Principles and Practice*, vol. 2004.
- [8] A. Tejada, R. M. Montesinos, y R. Guzmán, *Bioseparaciones*. 2014.
- [9] M. L. van Delden, N. J. M. Kuipers, y A. B. de Haan, "Selection and evaluation of alternative solvents for caprolactam extraction", *Sep Purif Technol*, vol. 51, núm. 2, pp. 219–231, 2006, doi: 10.1016/j.seppur.2006.02.003.
- [10] G. W. Meindersma, A. Podt, y A. B. De Haan, "Selection of ionic liquids for the extraction of aromatic hydrocarbons from

- aromatic/aliphatic mixtures”, *Fuel Processing Technology*, vol. 87, núm. 1, pp. 59–70, 2005, doi: 10.1016/j.fuproc.2005.06.002.
- [11] K. C. Sole, “Solvent extraction in the hydrometallurgical processing and purification of metals : process design and selected applications”, en *Solvent Extraction and Liquid Membranes: Fundamentals and Applications in New Materials*, 2018, pp. 141–200.
- [12] N. Tik, E. Bayraktar, y U. Mehmetoglu, “In situ reactive extraction of lactic acid from fermentation media”, *Journal of Chemical Technology and Biotechnology*, vol. 76, núm. 7, pp. 764–768, 2001, doi: 10.1002/jctb.449.
- [13] A. B. de Haan, “Affinity solvents for intensified organics extraction: Development challenges and prospects”, *Tsinghua Sci Technol*, vol. 11, núm. 2, pp. 171–180, 2006, doi: 10.1016/S1007-0214(06)70172-9.
- [14] W. L. McCabe, J. C. Smith, y P. Harriott, *Operaciones unitarias en Ingeniería Química*. 1991.
- [15] J. Gmehling, M. Kleiber, B. Kolbe, y J. Rarey, *Chemical Thermodynamics for Process Simulation*. 2019. doi: 10.1002/9783527809479.
- [16] J. M. Smith, H. C. van Ness, y M. M. Abbott, *Introducción a la Termodinámica en Ingeniería Química*. 2007. [En línea]. Disponible en: <http://librosysolucionarios.net/>
- [17] R. E. Treybal, “Azeotropes”, *A-to-Z Guide to Thermodynamics, Heat and Mass Transfer, and Fluids Engineering*, 2008, doi: 10.1615/atoz.a.azeotropes.
- [18] M. C. M. Cockrem, E. N. Lightfoot, y J. H. Flatt, “Solvent Selection for Extraction from Dilute Solution”, *Sep Sci Technol*, vol. 24, núm. 11, pp. 769–807, 1989, doi: 10.1080/01496398908049876.
- [19] J. Medina-Mora, “Optimización Multiobjetivo para la selección de solventes, aplicable a la Extracción Líquido-Líquido”, 2010.
- [20] E. Brignole y S. Pereda, *Phase equilibrium engineering principles*, vol. 3. 2013. doi: 10.1016/B978-0-444-56364-4.00006-6.
- [21] L. E. K. Achenie, R. Gani, y V. Venkatasubramanian, *Computer Aided Molecular Design: Theory and Practice*. 2003.

-
- [22] N. G. Martín, Mariano Eden, Mario R. Chemmangattuvalappil, *Tools for Chemical Product Design*. 2017.
- [23] H. Fruehbeis, R. Klein, y H. Wallmeier, "Computer-Assisted Molecular Design (CAMD) - An Overview", *ChemInform*, vol. 18, núm. 35, pp. 403–418, 1987, doi: 10.1002/chin.198735397.
- [24] L. A. Cisternas y E. D. Gálvez, "Principles for chemical products design", *Computer Aided Chemical Engineering*, vol. 21, núm. C, pp. 1107–1112, 2006, doi: 10.1016/S1570-7946(06)80194-X.
- [25] G. D. Moggridge y E. L. Cussler, "An introduction to chemical product design", *Chemical Engineering Research and Design*, vol. 78, núm. 1, pp. 5–11, 2000, doi: 10.1205/026387600527022.
- [26] A. K. Tula, D. K. Babi, J. Bottlaender, M. R. Eden, y R. Gani, "A computer-aided software-tool for sustainable process synthesis-intensification", *Comput Chem Eng*, vol. 105, pp. 74–95, 2017, doi: 10.1016/j.compchemeng.2017.01.001.
- [27] M. R. Eden, S. B. Jørgensen, R. Gani, y M. M. El-Halwagi, "A novel framework for simultaneous separation process and product design", *Chemical Engineering and Processing: Process Intensification*, vol. 43, núm. 5, pp. 595–608, 2004, doi: 10.1016/j.cep.2003.03.002.
- [28] L. Y. Ng, N. G. Chemmangattuvalappil, V. A. Dev, y M. R. Eden, *Mathematical Principles of Chemical Product Design and Strategies*, vol. 39. 2016. doi: 10.1016/B978-0-444-63683-6.00001-0.
- [29] R. Gani, "Chemical product design: Challenges and opportunities", *Comput Chem Eng*, vol. 28, núm. 12, pp. 2441–2457, 2004, doi: 10.1016/j.compchemeng.2004.08.010.
- [30] L. Y. Ng, F. K. Chong, y N. G. Chemmangattuvalappil, "Challenges and opportunities in computer-aided molecular design", *Comput Chem Eng*, vol. 81, pp. 115–129, 2015, doi: 10.1016/j.compchemeng.2015.03.009.
- [31] S. S. Y. Wong, W. Luo, y K. C. C. Chan, "EvoMD: An algorithm for evolutionary molecular design", *IEEE/ACM Trans Comput Biol Bioinform*, vol. 8, núm. 4, pp. 987–1003, 2011, doi: 10.1109/TCBB.2010.100.
- [32] J. Scheffczyk, L. Fleitmann, A. Schwarz, M. Lampe, A. Bardow, y K. Leonhard, "COSMO-CAMD: A framework for optimization-based

- computer-aided molecular design using COSMO-RS”, *Chem Eng Sci*, vol. 159, pp. 84–92, 2017, doi: 10.1016/j.ces.2016.05.038.
- [33] C. C. Solvason, “Integrated Multiscale Chemical Product Design using Property Clustering and Decomposition Techniques in a Reverse Problem Formulation”, 2011.
- [34] R. Gani, L. E. K. Achenie, y V. Venkatasubramanian, “Chapter 1: Introduction to CAMD”, en *Computer Aided Molecular Design: Theory and Practice*, 2003, pp. 3–21.
- [35] N. D. Austin, N. V. Sahinidis, y D. W. Trahan, “Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques”, *Chemical Engineering Research and Design*, vol. 116, núm., pp. 2–26, 2016, doi: 10.1016/j.cherd.2016.10.014.
- [36] M. Harini, J. Adhikari, y K. Y. Rani, “A review on property estimation methods and computational schemes for rational solvent design: A focus on pharmaceuticals”, *Ind Eng Chem Res*, vol. 52, núm. 21, pp. 6869–6893, 2013, doi: 10.1021/ie301329y.
- [37] P. M. Harper, R. Gani, P. Kolar, y T. Ishikawa, “Computer-aided molecular design with combined molecular modeling and group contribution”, *Fluid Phase Equilib*, vol. 158–160, pp. 337–347, 1999, doi: 10.1016/s0378-3812(99)00089-8.
- [38] N. D. Austin, N. v. Sahinidis, y D. W. Trahan, “Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques”, *Chemical Engineering Research and Design*, vol. 116, núm., pp. 2–26, 2016, doi: 10.1016/j.cherd.2016.10.014.
- [39] T. Martin, *User’s guide for T.E.S.T. (version 4.2) (Toxicity Estimation Software Tool) A program to estimate toxicity from molecular structure*. 2016, p. 63.
- [40] J. Song y H. H. Song, “Computer-aided molecular design of environmentally friendly solvents for separation processes”, *Chem Eng Technol*, vol. 31, núm. 2, pp. 177–187, 2008, doi: 10.1002/ceat.200700233.
- [41] J. Y. Ten, M. H. Hassim, D. K. S. Ng, y N. G. Chemmangattuvalappil, “The Incorporation of Safety and Health Aspects as Design Criteria in a Novel Chemical Product Design Framework”, *Computer Aided*

- Chemical Engineering*, vol. 39, pp. 197–220, 2016, doi: 10.1016/B978-0-444-63683-6.00007-1.
- [42] R. Gani, B. Nielsen, y A. Fredenslund, “A group contribution approach to computer-aided molecular design”, *AIChE Journal*, vol. 37, núm. 9, pp. 1318–1332, 1991, doi: 10.1002/aic.690370905.
- [43] O. Odele y S. Macchietto, “Computer Aided Molecular Design: A Novel Method for Optimal Solvent Selection”, *Fluid Phase Equilib*, vol. 00226020, núm. 3, pp. 47–54, 1993.
- [44] P. M. Harper y R. Gani, “A multi-step and multi-level approach for computer aided molecular design”, *Comput Chem Eng*, vol. 24, núm. 2–7, pp. 677–683, 2000, doi: 10.1016/S0098-1354(00)00410-5.
- [45] B. C. Roughton, “Development of Computer-Aided Molecular Design Methods for Bioengineering Applications”, 2013.
- [46] S. Cignitti, I. Rodriguez-Donis, J. Abildskov, X. You, N. Shcherbakova, y V. Gerbaud, “CAMD for entrainer screening of extractive distillation process based on new thermodynamic criteria”, *Chemical Engineering Research and Design*, vol. 147, pp. 721–733, 2019, doi: 10.1016/j.cherd.2019.04.038.
- [47] B. van Dyk y I. Nieuwoudt, “Design of solvents for extractive distillation”, *Ind Eng Chem Res*, vol. 39, núm. 5, pp. 1423–1429, 2000, doi: 10.1021/ie9904753.
- [48] N. Medina-Herrera, I. E. Grossmann, M. S. Mannan, y A. Jiménez-Gutiérrez, “An approach for solvent selection in extractive distillation systems including safety considerations”, *Ind Eng Chem Res*, vol. 53, núm. 30, pp. 12023–12031, 2014, doi: 10.1021/ie501205j.
- [49] S. Kossack, K. Kraemer, R. Gani, y W. Marquardt, “A systematic synthesis framework for extractive distillation processes”, *Chemical Engineering Research and Design*, vol. 86, núm. 7, pp. 781–792, 2008, doi: 10.1016/j.cherd.2008.01.008.
- [50] T. Zhou, Z. Song, X. Zhang, R. Gani, y K. Sundmacher, “Optimal Solvent Design for Extractive Distillation Processes: A Multiobjective Optimization-Based Hierarchical Framework”, *Ind Eng Chem Res*, vol. 58, núm. 15, pp. 5777–5786, 2019, doi: 10.1021/acs.iecr.8b04245.
- [51] J. Sun, H. Zhang, A. Zhou, Q. Zhang, y K. Zhang, “A new learning-based adaptive multi-objective evolutionary algorithm”, *Swarm Evol*

- Comput*, vol. 44, núm. December 2017, pp. 304–319, 2019, doi: 10.1016/j.swevo.2018.04.009.
- [52] A. T. Karunanithi, C. Acquah, L. E. K. Achenie, S. Sithambaram, y S. L. Suib, “Solvent design for crystallization of carboxylic acids”, *Comput Chem Eng*, vol. 33, núm. 5, pp. 1014–1021, 2009, doi: 10.1016/j.compchemeng.2008.11.003.
- [53] V. Venkatasubramanian, K. Chan, y J. M. Caruthers, “Evolutionary Design of Molecules with Desired Properties Using the Genetic Algorithm”, *J Chem Inf Comput Sci*, vol. 35, núm. 2, pp. 188–195, 1995, doi: 10.1021/ci00024a003.
- [54] M. Mattei, M. Hill, G. M. Kontogeorgis, y R. Gani, “Design of an emulsion-based personal detergent through a model-based chemical product design methodology”, *Computer Aided Chemical Engineering*, vol. 32, pp. 817–822, 2013, doi: 10.1016/B978-0-444-63234-0.50137-8.
- [55] R. Gani, “Chapter 14 Case studies in chemical product design - use of CAMD techniques”, *Computer Aided Chemical Engineering*, vol. 23, núm. 1991, pp. 435–458, 2007, doi: 10.1016/S1570-7946(07)80017-4.
- [56] K. Zhou Teng; Wang, Jiayuan; McBride, Kevin; Sundmacher, “Optimal Design of Solvents for Extractive Reaction Process”, *AIChE Journal*, vol. 61, núm. 3, pp. 857–866, 2015, doi: 10.1002/aic.
- [57] M. Skiborowski, “Process synthesis and design methods for process intensification”, *Curr Opin Chem Eng*, vol. 22, pp. 216–225, 2018, doi: 10.1016/j.coche.2018.11.004.
- [58] S. J. Patel, “Integrating Safety Issues in Optimizing Solven Selection and Porcess Design”, 2010.
- [59] J. Ooi, D. K. S. Ng, y N. G. Chemmangattuvalappil, “A Systematic Molecular Design Framework with the Consideration of Competing Solvent Recovery Processes”, *Ind Eng Chem Res*, vol. 58, núm. 29, pp. 13210–13226, 2019, doi: 10.1021/acs.iecr.9b01894.
- [60] J. E. Ourique y A. Silva Telles, “Computer-aided molecular design with simulated annealing and molecular graphs”, *Comput Chem Eng*, vol. 22, núm. SUPPL.1, pp. 0–3, 1998, doi: 10.1016/s0098-1354(98)00108-2.
- [61] J. Heintz, J. P. Belaud, N. Pandya, M. Teles Dos Santos, y V. Gerbaud, “Computer aided product design tool for sustainable product

- development”, *Comput Chem Eng*, vol. 71, pp. 362–376, 2014, doi: 10.1016/j.compchemeng.2014.09.009.
- [62] J. Heintz, “Systemic approach and decision process for sustainability in chemical engineering: Application to computer aided product design. PhD Thesis”, p. 256, 2012.
- [63] J. Y. Ten, M. H. Hassim, D. K. S. Ng, y N. G. Chemmangattuvalappil, “A molecular design methodology by the simultaneous optimisation of performance, safety and health aspects”, *Chem Eng Sci*, vol. 159, pp. 140–153, 2017, doi: 10.1016/j.ces.2016.03.026.
- [64] J. Devillers, *Genetic Algorithms in Molecular Modeling*, vol., núm. June. 2016. doi: 10.1016/b978-0-12-213810-2.x5000-2.
- [65] B. Lin, S. Chavali, K. Camarda, y D. C. Miller, “Computer-aided molecular design using Tabu search”, *Comput Chem Eng*, vol. 29, núm. 2, pp. 337–347, 2005, doi: 10.1016/j.compchemeng.2004.10.008.
- [66] A. S. Hukkerikar, B. Sarup, A. Ten Kate, J. Abildskov, G. Sin, y R. Gani, “Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis”, *Fluid Phase Equilib*, vol. 321, pp. 25–43, 2012, doi: 10.1016/j.fluid.2012.02.010.
- [67] A. S. Hukkerikar, S. Kalakul, B. Sarup, D. M. Young, G. Sin, y R. Gani, “Estimation of environment-related properties of chemicals for design of sustainable processes: Development of group-contribution+ (GC +) property models and uncertainty analysis”, *J Chem Inf Model*, vol. 52, núm. 11, pp. 2823–2839, 2012, doi: 10.1021/ci300350r.
- [68] A. S. Hukkerikar, G. Sin, J. Abildskov, B. Sarup, y R. Gani, *Development of pure component property models for chemical product-process design and analysis*, núm. September. 2013.
- [69] R. L. Haupt y S. E. Haupt, *Practical Genetic Algorithms*, vol. 2004.
- [70] R. Kumar, *Optimization: Algorithms and Applications*, vol. CRC Press, 2015.
- [71] Z. Dostál, *Optimal Quadratic Programming Algorithms: With Applications to Variational Inequalities*, vol. 23. 2009. doi: 10.1007/b138610.

- [72] J. A. Snyman y D. N. Wilke, *Practical Mathematical Optimization*, vol. 133. Cham, Switzerland, 2005. doi: 10.1007/b105200.
- [73] K.-H. Chang, "Chapter 3 - Design Optimization", en *Design Theory and Methods Using CAD/CAE*, 2015, pp. 103–210. doi: 10.1080/03772063.2020.1842159.
- [74] J. A. Caballero y I. E. Grossmann, "A review of the state of the art in optimization", *RIAI - Revista Iberoamericana de Automatica e Informatica Industrial*, vol. 4, núm. 1, pp. 5–23, 2007, doi: 10.1016/s1697-7912(07)70188-7.
- [75] X.-S. Yang, *Nature-Inspired Optimization Algorithms*. 2020.
- [76] C. A. C. Coello, "Introducción a la Computación Evolutiva", núm. 16, p. 310, 2004.
- [77] M. A. Iglesias-Solano y A. B. Iglesias-Carbonell, "La Computación Evolutiva y sus Paradigmas Paradigms of Evolutionary Computing", *Investigación y Desarrollo en TIC*, vol. 2, pp. 29–38, 2011.
- [78] N. Yusup, A. M. Zain, y S. Z. M. Hashim, "Evolutionary techniques in optimizing machining parameters: Review and recent applications (2007-2011)", *Expert Syst Appl*, vol. 39, núm. 10, pp. 9909–9927, 2012, doi: 10.1016/j.eswa.2012.02.109.
- [79] A. Menon, *Frontiers of Evolutionary Computation*, vol., núm. 2004.
- [80] C. J. Correa-Villalón, "Diseño de un Algoritmo Evolutivo para atacar Problemas NP-Duros basado en la Técnica Transgénicas", Universidad Autónoma de Aguascalientes (México), 2010. doi: -.
- [81] D. E. Goldberg, "Genetic Algorithms in Search Optimization & Machine Learning". p. 432, 1989.
- [82] J. H. Holland, "Genetic Algorithms - Computer programs that 'evolve' in ways that resemble natural selection can solve complex problems even their creators do not fully understand", *Scientific American*. pp. 66–72, 1992.
- [83] J. H. Holland, C. Langton, y S. W. Wilson, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. 1992.
- [84] M. Pelikan, *Hierarchical Bayesian Optimization Algorithm: Toward a New Generation of Evolutionary Algorithms*, vol. 53, núm. 9. 2013. doi: 10.1017/CBO9781107415324.004.

-
- [85] T. Bäck, "Evolutionary algorithms in theory and practice: Evolution strategies, evolutionary programming, genetic algorithms". p. 315, 1996.
- [86] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, vol. 1, núm. 1996.
- [87] D. A. Coley, "An Introduction to Genetic Algorithms for Scientists and Engineers". p. 185, 1999.
- [88] S. Kumar y P. J. Bentley, "Biologically Inspired Evolutionary Development", *ICES*, vol., núm., pp. 57–68, 2003, doi: 10.1007/3-540-36553-2_6.
- [89] D. Vasiljević, *Classical and Evolutionary Algorithms in the Optimization of Optical Systems*, vol., núm. 2006.
- [90] M. Melanie, *An introduction to genetic algorithms*, vol. 1999. doi: 10.1016/S0898-1221(96)90227-8.
- [91] S. N. Sivanandam y S. N. Deepa, *Introduction to Genetic Algorithms*, vol. 2008.
- [92] D. Simon, *Evolutionary Optimization Algorithms: Biologically Inspired and Population-Based Approaches to Computer Intelligence*, vol., núm. John Wiley & Sons, Inc., 2013.
- [93] W.-H. Steeb, *The nonlinear workbook: Chaos, Fractals, Cellular Automata, Neural Networks, Genetic Algorithms, Fuzzy Logic with C++, Java, SymbolicC++ and Reduce Programs*, vol. 1999.
- [94] J. Branke, *Evolutionary optimization in dynamic environment*, vol. 2002. doi: 10.1109/ICCP.2009.5284794.
- [95] C. B. Lucasius y G. Kateman, "Understanding and using genetic algorithms Part 2. Representation, configuration and hybridization", *Chemometrics and Intelligent Laboratory Systems*, vol. 25, pp. 99–145, 1994.
- [96] F. Rothlauf, *Representations for Genetic and Evolutionary Algorithms*, vol. 2006.
- [97] F. Corno, M. Sonza Reorda, y G. Squillero, "VEGA: a verification tool based on genetic algorithms", *Proceedings - IEEE International Conference on Computer Design: VLSI in Computers and Processors*, pp. 321–326, 1998, doi: 10.1109/iccd.1998.727069.

- [98] N. Srinivas y K. Deb, "Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms", núm. 0, 1386.
- [99] K. Deb, S. Agrawal, A. Pratap, y T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1917, pp. 849–858, 2000, doi: 10.1007/3-540-45356-3_83.
- [100] T. Murata y H. Ishibuchi, "MOGA: multi-objective genetic algorithms", *Proceedings of the IEEE Conference on Evolutionary Computation*, vol. 1, pp. 289–294, 1995, doi: 10.1109/icec.1995.489161.
- [101] E. Zitzler y L. Thiele, "Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach", vol. 3, núm. 4, pp. 257–271, 1999.
- [102] E. Zitzler, M. Laumanns, y L. Thiele, "SPEA2: Improving the Strength Pareto Evolutionary Algorithm", *TIK-Report 103 May*, pp. 1–21, 2001, doi: 10.1007/978-3-319-11119-3_4.
- [103] J. Gomez, "Self adaptation of operator rates in evolutionary algorithms", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3102, pp. 1162–1173, 2004, doi: 10.1007/978-3-540-24854-5_113.
- [104] J. Gómez, "Hybrid Adaptive Evolutionary Algorithm Hyper Heuristic", pp. 1–5, 2014.
- [105] C. R. Reeves y J. E. Rowe, *Genetic algorithms: Principles and Perspectives - A Guide to GA Theory*. 2002.
- [106] S. Bandyopadhyay y S. K. Pal, *Classification and Learning Using Genetic Algorithms: Applications in Bioinformatics and Web Intelligence*, vol. 2007.
- [107] L. Chambers, *Practical Handbook of Genetic Algorithms - New Frontiers Volume II*. 1995.
- [108] D. B. Hibbert, "Genetic algorithms in Chemistry", *Chemometrics and Intelligent Laboratory Systems*, vol. 19, núm. 8 SPEC., pp. 277–293, 1993, doi: 10.1016/S0010-4485(03)00002-2.
- [109] L. Elliott, D. B. Ingham, A. G. Kyne, N. S. Mera, M. Pourkashanian, y C. W. Wilson, "Genetic algorithms for optimisation of chemical kinetics

- reaction mechanisms”, *Progress in Energy and Combustion Science*, vol. 30, pp. 297–328, 2004, doi: 10.1016/j.pecs.2004.02.002.
- [110] C. B. Lucasius y G. Kateman, “Understanding and using genetic algorithms Part 1. Concepts, properties and context”, *Chemometrics and Intelligent Laboratory Systems*, vol. 19, núm. 1, pp. 1–33, 1993, doi: 10.1016/0169-7439(93)80079-W.
- [111] L. Gosselin, M. Tye-Gingras, y F. Mathieu-Potvin, “Review of utilization of genetic algorithms in heat transfer problems”, *Int J Heat Mass Transf*, vol. 52, núm. 9–10, pp. 2169–2188, 2009, doi: 10.1016/j.ijheatmasstransfer.2008.11.015.
- [112] J. G. Andreasen, U. Larsen, T. Knudsen, L. Pierobon, y F. Haglind, “Selection and optimization of pure and mixed working fluids for low grade heat utilization using organic rankine cycles”, *Energy*, vol. 73, pp. 204–213, 2014. doi: 10.1016/j.energy.2014.06.012.
- [113] Z. Kravanja y M. Bogataj, “26 European Symposium on Computer Aided Process Engineering-Elsevier”, *Computer Aided Chemical Engineering*, vol. 38, p. 588, 2016.
- [114] M. Fan, J. Hu, R. Cao, W. Ruan, y X. Wei, “A review on experimental design for pollutants removal in water treatment with the aid of artificial intelligence”, *Chemosphere*, vol. 200, pp. 330–343, 2018, doi: 10.1016/j.chemosphere.2018.02.111.
- [115] A. Maiocchi, “Genetic algorithms in molecular modelling: a review”, en *Data Handling in Science and Technology*, 2003, pp. 109–139. doi: 10.1016/S0922-3487(03)23004-5.
- [116] T. Lisboa, “Multi-Objective Optimization”, *Técnico Lisboa*. pp. 148–173.
- [117] A. Ramos, “Optimización Multicriterio”, *Universidad Pontificia Comillas*, núm. p. 16.
- [118] C. A. C. Coello, “Aplicaciones de los Algoritmos Evolutivos Multiobjetivo”, núm. 2508, 2012.
- [119] H. Massam, “Multi-criteria (MCDM) Decision Making Techniques in Planning”, *Prog Plann*, vol. 30, núm. Mcdm, pp. 1–84, 1988.
- [120] P. L. Yu, “Multiple criteria decision making: Five basic concepts”, *Handbooks in Operations Research and Management Science*, vol. 1, núm. C, pp. 663–699, 1989, doi: 10.1016/S0927-0507(89)01011-X.

- [121] C. A. C. Coello, G. B. Lamont, y D. A. Van Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems*, vol. 2007. doi: 10.1080/00949659608811725.
- [122] S. Chand y M. Wagner, “Evolutionary many-objective optimization: A quick-start guide”, *Surveys in Operations Research and Management Science*, vol. 20, núm. 2, pp. 35–42, 2015, doi: 10.1016/j.sorms.2015.08.001.
- [123] R. T. Marler y J. S. Arora, “Survey of multi-objective optimization methods for engineering”, *Structural and Multidisciplinary Optimization*, vol. 26, núm. 6, pp. 369–395, 2004, doi: 10.1007/s00158-003-0368-6.
- [124] C. Zopounidis y P. M. Pardalos, *Handbook of Multicriteria Analysis*. 2010.
- [125] D. R. Insua, “Sobre soluciones óptimas en problemas de optimización multiobjetivo”, *Trabajos de Investigación Operativa*, vol. 2, núm. 1, pp. 49–67, 1987, doi: 10.1007/BF02888810.
- [126] C. M. Subía, “Desarrollo de una Guía Metodológica sobre Computación Evolutiva y Algoritmos Genéticos, para la Optimización Evolutiva Multiobjetivo”, 2014. doi: 10.4324/9781315853178.
- [127] S. Hernández, “Del diseño convencional al diseño óptimo. Posibilidades y variantes”, *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*, vol. 9, pp. 259–270, 1993.
- [128] H. Mukai, “Algorithms for Multicriterion Optimization”, *IEEE Trans Automat Contr*, vol. 25, núm. 2, pp. 177–186, 1980, doi: 10.1109/TAC.1980.1102298.
- [129] G. Giorgi, B. Jiménez, y V. Novo, “Approximate Karush–Kuhn–Tucker Condition in Multiobjective Optimization”, *J Optim Theory Appl*, vol. 171, núm. 1, pp. 70–89, 2016, doi: 10.1007/s10957-016-0986-y.
- [130] E. Mezura-Montes y C. A. Coello Coello, “Conceptos de Optimización Multiobjetivo para el Manejo de Restricciones en Algoritmos Evolutivos: Un Estudio Comparativo”, *Proceedings of the 1st Mexican Conference on Evolutionary Computation (COMCEV 2003)*, pp. 1–12, 2003.
- [131] N. Riquelme, C. von Lüken, y B. Barán, “Performance metrics in multi-objective optimization”, en *2015 XLI Latin American Computing Conference (CLEI) Performance*, 2015.

-
- [132] B. A. Cuartas Torres, "Metodología para la optimización de múltiples objetivos basada en ag y uso de preferencias", Universidad Nacional de Colombia - Sede Medellín, 2009. [En línea]. Disponible en: <http://www.bdigital.unal.edu.co/2237/%5Cnhttp://www.bdigital.unal.edu.co/2237/1/43908352.2009.pdf>
- [133] N. Riquelme, C. Von Lücken, y B. Barán, "Performance metrics in multi-objective optimization", *Proceedings - 2015 41st Latin American Computing Conference, CLEI 2015*, vol. 1, p. 11, 2015, doi: 10.1109/CLEI.2015.7360024.
- [134] J. A. Párraga, "Clustering difuso multi-objetivo de genes basado en información biológica externa y datos de expresión génica", Universidad de Santiago de Chile, 2017. doi: 10.1088/1742-6596/134/1/012001.
- [135] J. C. Castro, "Modelo de optimización multiobjetivo para el algoritmo evolutivo HAEA (Hybrid Adaptative Evolutionary Algorithm)", Universidad Nacional de Colombia, 2020.
- [136] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, vol. John Wiley & Sons, 2001.
- [137] R. T. Marler y J. S. Arora, "The weighted sum method for multi-objective optimization: New insights", *Structural and Multidisciplinary Optimization*, vol. 41, núm. 6, pp. 853–862, 2010, doi: 10.1007/s00158-009-0460-7.
- [138] A. Singh y S. Kumar, "Multiple Objectives Mathematical Programming Using Payoff Techniques", vol. 9, núm. 1, pp. 39–46, 2012.
- [139] S. Obayashi, D. Sasaki, y A. Oyama, "Finding tradeoffs by using multiobjective optimization algorithms", *Trans Jpn Soc Aeronaut Space Sci*, vol. 47, núm. 155, pp. 51–58, 2004, doi: 10.2322/tjsass.47.51.
- [140] M. Sakawa, *Genetic algorithms and fuzzy multiobjective optimization*, vol. 1, núm. 2002.
- [141] R. Wang, R. C. Purshouse, y P. J. Fleming, "Preference-inspired co-evolutionary algorithms using weight vectors", *Eur J Oper Res*, vol. 243, núm. 2, pp. 423–441, 2015, doi: 10.1016/j.ejor.2014.05.019.
- [142] W. Wang, S. Ying, L. Li, Z. Wang, y W. Li, "An improved decomposition-based multiobjective evolutionary algorithm with a

- better balance of convergence and diversity”, *Applied Soft Computing Journal*, vol. 57, pp. 627–641, 2017, doi: 10.1016/j.asoc.2017.03.041.
- [143] Q. Zhang, W. Zhu, B. Liao, X. Chen, y L. Cai, “A modified PBI approach for multi-objective optimization with complex Pareto fronts”, *Swarm Evol Comput*, vol. 40, núm. February, pp. 216–237, 2018, doi: 10.1016/j.swevo.2018.02.001.
- [144] C. A. C. Coello, “Introducción a la Optimización Evolutiva Multiobjetivo”, *Apuntes de clase: Introducción a la optimización multiobjetivo.*, vol., núm. 4. pp. 1–144, 2012.
- [145] J. D. Schaffer, “Multiple objective optimization with vector evaluated genetic algorithms”, *The 1st international Conference on Genetic Algorithms*, núm. JANUARY 1985, pp. 93–100, 1985.
- [146] K. Deb, A. Pratap, S. Agarwal, y T. Meyarivan, “A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II”, *IEEE Transactions on Evolutionary Computation*, vol. 6, núm. 2, pp. 182–197, 2002, doi: 10.1109/4235.996017.
- [147] C. A. Correa Flórez, R. Andrés Bolaños, y A. Molina Cabrera, “Algoritmo multiobjetivo NSGA-II aplicado al problema de la mochila.”, *Scientia Et Technica*, vol. 2, núm. 39, pp. 206–211, 2008.
- [148] L. Lopez, R. A. Hincapié, y R. A. Gallego, “Planeamiento multi-objetivo de sistemas de distribución usando un algoritmo evolutivo NSGA-II”, *Revista Escuela de Ingeniería de Antioquía*, vol. 15, núm. 15, pp. 141–151, 2011.
- [149] N. Dahmani, F. Clautiaux, S. Krichen, y E. G. Talbi, “Self-adaptive metaheuristics for solving a multi-objective 2-dimensional vector packing problem”, *Applied Soft Computing Journal*, vol. 16, pp. 124–136, 2014, doi: 10.1016/j.asoc.2013.12.006.
- [150] L. Wang, A. H. C. Ng, y K. Deb, “Multi-objective Evolutionary Optimisation for Product Design and Manufacturing”, *Assembly Automation*, vol. 32, núm. 4, pp. 142–147, 2012, doi: 10.1108/aa.2012.03332daa.009.
- [151] J. Gomez, “Self adaptation of operator rates for multimodal optimization”, *Proceedings of the 2004 Congress on Evolutionary Computation, CEC2004*, vol. 2, pp. 1720–1726, 2004, doi: 10.1109/cec.2004.1331103.

-
- [152] GECCO '19, "MOHAEA: A Multi-objective Hybrid Adaptive Evolutionary Algorithm", en *GECCO '19, July 13–17, 2019*, 2019, pp. 1–3.
- [153] J. Horn, N. Nafpliotis, y D. E. Goldberg, "A niched Pareto genetic algorithm for multiobjective optimization", en *First IEEE Conference on Evolutionary Computation*, 2002, pp. 82–87. doi: 10.1109/icec.1994.350037.
- [154] C.-L. Hwang, Y.-J. Lai, y T.-Y. Liu, "A New Approach for Multiple Objective Decision Making", *Comput Oper Res*, vol. 20, núm. 8, pp. 889–899, 1993.
- [155] J. G. Vlachogiannis y K. Y. Lee, "Multi-objective based on parallel vector evaluated particle swarm optimization for optimal steady-state performance of power systems", *Expert Syst Appl*, vol. 36, núm. 8, pp. 10802–10808, 2009, doi: 10.1016/j.eswa.2009.02.079.
- [156] L. C. Cagnina, "Optimización Mono y Multiobjetivo a través de una Heurística de Inteligencia Colectiva", pp. 54–72, 2010.
- [157] A. Lara López, "Un estudio de las Estrategias Evolutivas para problemas Multiobjetivo.", pp. 23–25, 2003.
- [158] N. A. Ramírez, "Una nueva propuesta para optimización multiobjetivo basada en búsqueda dispersa (Scatter Search)", 2006.
- [159] M. Gul, E. Celik, N. Aydin, A. Taskin Gumus, y A. F. Guneri, "A state of the art literature review of VIKOR and its fuzzy extensions on applications", *Applied Soft Computing Journal*, vol. 46, pp. 60–89, 2016, doi: 10.1016/j.asoc.2016.04.040.
- [160] Y.-Z. Lu, Y.-W. Chen, M.-R. Chen, P. Chen, y G.-Q. Zeng, *Extremal Optimization: Fundamentals, Algorithms, and Applications*, vol. 2016.
- [161] S. S. Santander-Jiménez, M. A. Vega-Rodríguez, J. A. Gómez-Pulido, y J. M. Sánchez-Pérez, "Una adaptación multiobjetivo y paralela del algoritmo Artificial Bee Colony aplicada a la inferencia filogenética", 1996.
- [162] C.-K. Goh y K. C. Tan, *Evolutionary Multi-objective Optimization in Uncertain Environments: Issues and Algorithms*, vol. 186. 2009. doi: 10.1007/978-3-540-95976-2.

- [163] M. Macías Infantes, “Estudio Comparativo de Técnicas de Optimización para la Actualización de Modelos de Elementos Finitos”, 2016.
- [164] L. C. Jain y N. M. Martin, *Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications*, vol. 1. 1998.
- [165] G. Toscano, “Optimización Multiobjetivo usando un Micro Algoritmo Genético”, Universidad Veracruzana - LANIA, 2001.
- [166] C. Almeida, N. Amarilla, y B. Barán, “Optimización Multiobjetivo en la Planificación de Centrales Telefónicas”, 2003.
- [167] C. Castillo, “Aplicación de la Programación Multiobjetivo en la Optimización del Tráfico Generado por un IDS/IPS”, *Rev. Tecnol. - Journal of Technology*, vol. 11, núm. 1, pp. 41–55, 2012.
- [168] S. Ruiz, O. D. Castrillón, y W. A. Sarache, “Una metodología multiobjetivo para optimizar un ambiente job shop”, *Informacion Tecnologica*, vol. 23, núm. 1, pp. 35–46, 2012, doi: 10.4067/S0718-07642012000100005.
- [169] M. Guzek, J. E. Pecero, B. Dorronsoro, y P. Bouvry, “Multi-objective evolutionary algorithms for energy-aware scheduling on distributed computing systems”, *Applied Soft Computing Journal*, vol. 24, pp. 432–446, 2014, doi: 10.1016/j.asoc.2014.07.010.
- [170] C. Carnevale, G. Finzi, E. Pisoni, y M. Volta, “Multi-objective analysis to control ozone exposure”, en *Developments in Environmental Science*, 2007, pp. 96–108. doi: 10.1016/S1474-8177(07)06023-8.
- [171] P. S. Moura y A. T. de Almeida, “Multi-objective optimization of a mixed renewable system with demand-side management”, *Renewable and Sustainable Energy Reviews*, vol. 14, núm. 5, pp. 1461–1468, 2010, doi: 10.1016/j.rser.2010.01.004.
- [172] M. J. Bastidas, R. F. Bermúdez, G. P. Jaramillo, y F. Chejne, “Optimización termoeconómica y ambiental usando algoritmos genéticos multiobjetivo”, *Informacion Tecnologica*, vol. 21, núm. 4, pp. 35–44, 2010, doi: 10.1612/inf.tecnol.4384it.09.
- [173] P. J. Copado-Méndez, C. Pozo, G. Guillén-Gosálbez, y L. Jiménez, “Enhancing the ϵ -constraint method through the use of objective reduction and random sequences: Application to environmental problems”, *Comput Chem Eng*, vol. 87, pp. 36–48, 2016, doi: 10.1016/j.compchemeng.2015.12.016.

-
- [174] M. N. Naz, M. I. Mushtaq, M. Naeem, M. Iqbal, M. W. Altaf, y M. Haneef, "Multicriteria decision making for resource management in renewable energy assisted microgrids", *Renewable and Sustainable Energy Reviews*, vol. 71, núm. December 2016, pp. 323–341, 2017, doi: 10.1016/j.rser.2016.12.059.
- [175] R. T. F. Ah. King, K. Deb, y H. C. S. Rughooputh, "Comparison of NSGA-II and SPEA2 on the Multiobjective Environmental/Economic Dispatch Problem", *University of Mauritius Research Journal*, vol. 16, núm. 1, pp. 485–511, 2010.
- [176] L. Atmaniou *et al.*, "A multiobjective genetic algorithm optimization framework for batch plant design", *Computer Aided Chemical Engineering*, vol. 15, núm. C, pp. 400–405, 2003, doi: 10.1016/S1570-7946(03)80577-1.
- [177] C. Gutiérrez-Antonio, A. Briones-Ramírez, y A. Jiménez-Gutiérrez, "Optimization of Petlyuk sequences using a multi objective genetic algorithm with constraints", *Comput Chem Eng*, vol. 35, núm. 2, pp. 236–244, 2011, doi: 10.1016/j.compchemeng.2010.10.007.
- [178] A. I. Papadopoulos y P. Linke, "Multiobjective molecular design for integrated process-solvent systems synthesis", *AIChE Journal*, vol. 52, núm. 3, pp. 1057–1070, 2006, doi: 10.1002/aic.10715.
- [179] S. Ekins, J. D. Honeycutt, y J. T. Metz, "Evolving molecules using multi-objective optimization: Applying to ADME/Tox", *Drug Discov Today*, vol. 15, núm. 11–12, pp. 451–460, 2010, doi: 10.1016/j.drudis.2010.04.003.
- [180] L. Y. Ng, N. G. Chemmangattuvalappil, y D. K. S. Ng, "A multiobjective optimization-based approach for optimal chemical product design", *Ind Eng Chem Res*, vol. 53, núm. 44, pp. 17429–17444, 2014, doi: 10.1021/ie502906a.
- [181] P. Bigus, J. Namieśnik, y M. Tobiszewski, "Application of multicriteria decision analysis in solvent type optimization for chlorophenols determination with a dispersive liquid-liquid microextraction", *J Chromatogr A*, vol. 1446, pp. 21–26, 2016, doi: 10.1016/j.chroma.2016.03.065.
- [182] C. M. Fonseca y P. J. Fleming, "Genetic Algorithms for Multi-Objective Optimization: Formulation, discussion and generalization", en *Proceedings of the 5th International Conference on Genetic Algorithms*, 1993, pp. 416–423. doi: 10.3156/jfuzzy.9.4_471_1.

- [183] M. Garza-Fabre, G. T. Pulido, y C. A. C. Coello, "Ranking methods for many-objective optimization", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5845 LNAI, pp. 633–645, 2009, doi: 10.1007/978-3-642-05258-3_56.
- [184] K. Deb y H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints", *IEEE Transactions on Evolutionary Computation*, vol. 18, núm. 4, pp. 577–601, 2014, doi: 10.1109/TEVC.2013.2281535.
- [185] H. Ishibuchi, N. Tsukamoto, y Y. Nojima, "Evolutionary many-objective optimization: A short review", *2008 IEEE Congress on Evolutionary Computation, CEC 2008*, pp. 2419–2426, 2008, doi: 10.1109/CEC.2008.4631121.
- [186] H. Sato, "Pareto Partial Dominance MOEA in Many-Objective Optimization", *Search (Syd)*, núm. January, pp. 1–10, 2009.
- [187] H. Aguirre y K. Tanaka, "Many-objective optimization by space partitioning and adaptive ϵ -ranking on MNK-landscapes", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5467 LNCS, pp. 407–422, 2010, doi: 10.1007/978-3-642-01020-0_33.
- [188] C. H. Papadimitriou y M. Yannakakis, "On the approximability of trade-offs and optimal access of web sources", *Annual Symposium on Foundations of Computer Science - Proceedings*, pp. 86–92, 2000, doi: 10.1109/sfcs.2000.892068.
- [189] T. Erlebach, H. Kellerer, y U. Pferschy, "Approximating Multi-objective Knapsack Problems", pp. 210–211, 2001.
- [190] J. Bader y E. Zitzler, "HypE : An algorithm for fast optimization", *Evol Comput*, vol. 19, núm. 1, pp. 45–76, 2011.
- [191] L. While, P. Hingston, L. Barone, y S. Huband, "A faster algorithm for calculating hypervolume", *IEEE Transactions on Evolutionary Computation*, vol. 10, núm. 1, pp. 29–38, 2006, doi: 10.1109/TEVC.2005.851275.
- [192] K. Bringmann y T. Friedrich, "Approximating the volume of unions and intersections of high-dimensional geometric objects", *Comput Geom*, vol. 43, núm. 6–7, pp. 601–610, 2010, doi: 10.1016/j.comgeo.2010.03.004.

-
- [193] X. Cai, H. Sun, y Z. Fan, “A diversity indicator based on reference vectors for many-objective optimization”, *Inf Sci (N Y)*, vol. 430–431, pp. 467–486, 2018, doi: 10.1016/j.ins.2017.11.051.
- [194] G. Dai, C. Zhou, M. Wang, y X. Li, “Indicator and reference points co-guided evolutionary algorithm for many-objective optimization problems”, *Knowl Based Syst*, vol. 140, pp. 50–63, 2018, doi: 10.1016/j.knosys.2017.10.025.
- [195] M. Zhang y H. Li, “A reference direction and entropy based evolutionary algorithm for many-objective optimization”, *Applied Soft Computing Journal*, vol. 70, pp. 108–130, 2018, doi: 10.1016/j.asoc.2018.05.011.
- [196] J. Zou, C. Ji, S. Yang, Y. Zhang, J. Zheng, y K. Li, “A knee-point-based evolutionary algorithm using weighted subpopulation for many-objective optimization”, *Swarm Evol Comput*, vol. 47, núm. January, pp. 33–43, 2019, doi: 10.1016/j.swevo.2019.02.001.
- [197] Y. Liu, N. Zhu, K. K. Li, M. Li, J. Zheng, y K. Li, “An angle dominance criterion for evolutionary many-objective optimization”, *Inf Sci (N Y)*, núm. xxxx, 2019, doi: 10.1016/j.ins.2018.12.078.
- [198] L. Cai, S. Qu, y G. Cheng, “Two-archive method for aggregation-based many-objective optimization”, *Inf Sci (N Y)*, vol. 422, pp. 305–317, 2018, doi: 10.1016/j.ins.2017.08.078.
- [199] P. J. Fleming, R. C. Purshouse, y R. J. Lygoe, “Many-objective optimization: An engineering design perspective”, *Lecture Notes in Computer Science*, vol. 3410, pp. 14–32, 2005, doi: 10.1007/978-3-540-31880-4_2.
- [200] A. Inselberg y B. Dimsdale, “Parallel coordinates: A tool for visualizing multi-dimensional geometry”, pp. 361–378, 1990, doi: 10.1007/978-4-431-68057-4_3.
- [201] A. Pryke, S. Mostaghim, y A. Nazemi, “Heatmap visualization of population based multi objective algorithms”, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4403 LNCS, pp. 361–375, 2007, doi: 10.1007/978-3-540-70928-2_29.
- [202] D. J. Walker, R. M. Everson, y J. E. Fieldsend, “Visualizing mutually nondominating solution sets in many-objective optimization”, *IEEE*

- Transactions on Evolutionary Computation*, vol. 17, núm. 2, pp. 165–184, 2013, doi: 10.1109/TEVC.2012.2225064.
- [203] P. Hoffman y G. Grinstein, “Visualizations for High Dimensional Data Mining-Table Visualizations”, núm. August 2001, 1997.
- [204] J. B. Tenenbaum, V. De Silva, y J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction”, *Science (1979)*, vol. 290, núm. 5500, pp. 2319–2323, 2000, doi: 10.1126/science.290.5500.2319.
- [205] J. Zou, L. Fu, J. Zheng, S. Yang, G. Yu, y Y. Hu, “A many-objective evolutionary algorithm based on rotated grid”, *Applied Soft Computing Journal*, vol. 67, pp. 596–609, 2018, doi: 10.1016/j.asoc.2018.02.031.
- [206] B. Khan, S. Hanoun, M. Johnstone, C. P. Lim, D. Creighton, y S. Nahavandi, “A scalarization-based dominance evolutionary algorithm for many-objective optimization”, *Inf Sci (N Y)*, vol. 474, pp. 236–252, 2019, doi: 10.1016/j.ins.2018.09.031.
- [207] J. Zou, Y. Zhang, S. Yang, Y. Liu, y J. Zheng, “Adaptive neighborhood selection for many-objective optimization problems”, *Applied Soft Computing Journal*, vol. 64, pp. 186–198, 2018, doi: 10.1016/j.asoc.2017.11.041.
- [208] M. Wagner y F. Neumann, “A Fast Approximation-Guided Evolutionary Multi-Objective Algorithm”, pp. 687–694.
- [209] Q. Zhang, S. Member, y H. Li, “MOEA / D : A Multiobjective Evolutionary Algorithm Based on Decomposition”, vol. 11, núm. 6, pp. 712–731, 2007.
- [210] I. Giagkiozis, R. C. Purshouse, y P. J. Fleming, “Generalized Decomposition and Cross Entropy Methods for Many-Objective Optimization”, *Inf Sci (N Y)*, p. 2014, 2014, doi: 10.1016/j.ins.2014.05.045.
- [211] H. Seada y K. Deb, “U-NSGA-III : A Unified Evolutionary Algorithm for Single , Multiple , and Many-Objective Optimization”, pp. 1–30.
- [212] D. Weininger, “SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules”, *J Chem Inf Comput Sci*, vol. 28, núm. 1, pp. 31–36, 1988, doi: 10.1021/ci00057a005.

- [213] L. Hornos, "Introducción a SMILES: Dibujando moléculas en el bloc de notas", *El problema de describir una estructura molecular con caracteres comunes*, 2020.
- [214] Daylight Chemical Information Systems Inc., "SMILES - A Simplified Chemical Language", -, 2019.
- [215] N. M. O. Boyle, "Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI", pp. 1–14, 2012.
- [216] IUPAC, "Definición de compuestos anti aromáticos". <https://goldbook.iupac.org/terms/view/A00382>
- [217] A. T. M. G. Mostafa, J. M. Eakman, M. M. Montoya, y S. L. Yarbro, "Prediction of Heat Capacities of Solid Inorganic Salts from Group Contributions", pp. 343–348, 1996.
- [218] D. W. Van Krevelen, *Properties of Polymers: Their Correlation with Chemical Structure; their Numerical Estimation and Prediction from Additive Group Contributions*. 2009.
- [219] D. Saracino, *Abstract Algebra A First Course*. 2008.
- [220] Joseph A. Gallian, *Contemporary Abstract Algebra*. 2013.
- [221] D. B. Fraleigh, *A first course in abstract algebra*. 2002.
- [222] W. L. Kocay y D. L. Kreher, *Graphs, Algorithms, and Optimization*, vol. 1. Boca Raton, FL, USA: CRC Press, 2017.
- [223] P. Fernández-Gallardo y J. L. Fernández-Pérez, "La Teoría de Pólya", en *El discreto encanto de la matemática*, 2002, pp. 1133–1145.
- [224] A. R. Matamala, "PÓLYA'S COMBINATORIAL METHOD AND THE ISOMER ENUMERATION PROBLEM", *Bol. Soc. Chil. Quím.*, 2002. https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0366-16442002000200006#17
- [225] S. Pevac y G. Crundwell, "Pólya's Isomer Enumeration Method: A Unique Exercise in Group Theory and Combinatorial Analysis for Undergraduates", *J Chem Educ*, vol. 77, núm. 10, pp. 1358–1360, 2000, doi: 10.1021/ed077p1358.
- [226] R. L. Apodaca, "A Comprehensive Treatment of Aromaticity in the SMILES Language", 2021. <https://depth->

- first.com/articles/2020/02/10/a-comprehensive-treatment-of-aromaticity-in-the-smiles-language/
- [227] Daylight Chemical Information Systems Inc., “SMILES Tutorial: Conventions”.
<https://www.daylight.com/meetings/summerschool98/course/dave/smiles-convent.htm>
- [228] J. Marrero y R. Gani, “Group-contribution based estimation of pure component properties”, *Fluid Phase Equilib*, vol. 183–184, pp. 183–208, 2001, doi: 10.1016/S0378-3812(01)00431-9.
- [229] E. C. Ihmels y J. Gmehling, “Extension and revision of the group contribution method GCVOL for the prediction of pure compound liquid densities”, *Ind Eng Chem Res*, vol. 42, núm. 2, pp. 408–412, 2003, doi: 10.1021/ie020492j.
- [230] T. J. Sheldon, C. S. Adjiman, y J. L. Cordiner, “Pure component properties from group contribution: Hydrogen-bond basicity, hydrogen-bond acidity, Hildebrand solubility parameter, macroscopic surface tension, dipole moment, refractive index and dielectric constant”, *Fluid Phase Equilib*, vol. 231, núm. 1, pp. 27–37, 2005, doi: 10.1016/j.fluid.2004.12.017.
- [231] J. Gmehling, J. Lohmann, A. Jakob, J. Li, y R. Joh, “A modified UNIFAC (Dortmund) model. 4. Revision and extension”, *Ind Eng Chem Res*, vol. 37, núm. 12, pp. 4876–4882, 1998, doi: 10.1021/ie980347z.
- [232] D. Constantinescu y J. Gmehling, “Further development of modified UNIFAC (Dortmund): Revision and extension 6”, *J Chem Eng Data*, vol. 61, núm. 8, pp. 2738–2748, 2016, doi: 10.1021/acs.jced.6b00136.
- [233] A. Fredeslund, Russell L. Jones, y J. M. Prausnitz, “Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures”, vol. 21, núm. 6, 1975.
- [234] Dortmund-Databank, “Published ParametersUnifac”.
<http://www.ddbst.com/published-parameters-unifac.html>
- [235] J. J. Irwin y B. K. Shoichet, “ZINC - A free database of commercially available compounds for virtual screening”, *J Chem Inf Model*, vol. 45, núm. 1, pp. 177–182, 2005, doi: 10.1021/ci049714+.
- [236] T. Sterling y J. J. Irwin, “ZINC 15 - Ligand Discovery for Everyone”, *J Chem Inf Model*, vol. 55, núm. 11, pp. 2324–2337, 2015, doi: 10.1021/acs.jcim.5b00559.

-
- [237] EPA, “Distributed Structure-Searchable Toxicity (DSSTox) Database”, 2022. <https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database>.
- [238] Texas A&M University Libraries, “Chemical Pricing Database - Beta Version”, 2022. <https://tamu.libguides.com/c.php?g=587308&p=5694124&url=L2V2LTl5ODk3NjlvZGlvNTUzNTQvdmlldy5hc3B4>
- [239] S&P Global, “Chemical Week by S&P Global”, 2022. <https://chemweek.com/home>
- [240] Relx Inc., “ICIS Chemical Business”, 2022. https://www.icis.com/subscriber/specialpublications/#_=_
- [241] T. Group, G. All, I. N. D. Farm, P. F. F. Farm, y G. Smith, “Table 9 . Producer price indexes for commodity and service groupings and individual items , not seasonally adjusted [April 2022, Index base 1982=100, unless otherwise indicated]”, núm. April, pp. 1–58, 2022.
- [242] S. Müller, “GitHub - Simon Müller - Fragmentation Algorithm Paper”, https://github.com/simonmb/fragmentation_algorithm_paper, 2023. https://github.com/simonmb/fragmentation_algorithm_paper (consultado el 8 de abril de 2023).
- [243] S. Müller, “Flexible heuristic algorithm for automatic molecule fragmentation: Application to the UNIFAC group contribution model”, *J Cheminform*, vol. 11, núm. 1, 2019, doi: 10.1186/s13321-019-0382-3.
- [244] K. Rodríguez, “Computer Aided Molecular Design : State-of-the-art and environmentally friendly product design”, p. 9, 2019.
- [245] K. A. Rodríguez, “Inclusion of toxicity and market availability in a Computer-Aided Molecular Design methodology”, 2019.
- [246] J. Prieto y J. Gomez, “Hybrid Adaptive Evolutionary Algorithm for Multi-objective Optimization”, 2020, [En línea]. Disponible en: <http://arxiv.org/abs/2004.13925>
- [247] K. Deb, A. Pratap, S. Agarwal, y T. Meyarivan, “A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II”, *IEEE Transactions on Evolutionary Computation*, vol. 6, núm. 2, pp. 182–197, 2002, doi: 10.1109/4235.996017.

- [248] D. Constantinescu y J. Gmehling, "Further development of modified UNIFAC (Dortmund): Revision and extension 6", *J Chem Eng Data*, vol. 61, núm. 8, pp. 2738–2748, 2016, doi: 10.1021/acs.jced.6b00136.
- [249] DDBST GmbH, "Parameters of the Modified UNIFAC (Dortmund) Model", 2022.
<http://unifac.ddbst.de/PublishedParametersUNIFACDO.html>
- [250] OEIS.org, "The On-Line Encyclopedia of Integer Sequences® (OEIS®)", <https://oeis.org/>, 2023. <https://oeis.org/> (consultado el 25 de julio de 2023).
- [251] N. C. for B. I. NIH - National Library of Medicine, "PubChem: Explore Chemistry - Quickly find chemical information from authoritative sources", 2023.
- [252] J. Gómez, "Hybrid Adaptive Evolutionary Algorithm Hyper Heuristic", pp. 1–5.
- [253] J. Gomez, "Self adaptation of operator rates in evolutionary algorithms", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3102, pp. 1162–1173, 2004, doi: 10.1007/978-3-540-24854-5_113.
- [254] D. Weininger, "SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules", *J Chem Inf Comput Sci*, vol. 28, núm. 1, pp. 31–36, 1988, doi: 10.1021/ci00057a005.
- [255] L. Hornos, "Introducción a SMILES: Dibujando moléculas en el bloc de notas", *El problema de describir una estructura molecular con caracteres comunes*, 2020.
- [256] Daylight Chemical Information Systems Inc., "SMILES - A Simplified Chemical Language", -, 2019.
- [257] Inc. Daylight Chemical Information System, "SMARTS - A Language for Describing Molecular Patterns", <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>, 2019. <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html> (consultado el 8 de abril de 2023).
- [258] U. H. ZBH - Center for Bioinformatics, "SMARTS PLUS", <https://smarts.plus/>, 2023. <https://smarts.plus/> (consultado el 8 de abril de 2023).

- [259] U. Weidlich y J. Gmehling, "A Modified UNIFAC Model. 1. Prediction of VLE, hE, and $^3\infty$ ", *Ind Eng Chem Res*, vol. 26, núm. 7, pp. 1372–1381, 1987, doi: 10.1021/ie00067a018.
- [260] A. Ag, E. Fredenslund, J. Gmehling, y P. Rasmussen, *Vapor-liquid equilib using UNIFAC a group-contribution method Library of Congress Cataloging in Publication Data*. 1977.
- [261] ACD Labs, "ACD Labs ChemSketch", <https://www.acdlabs.com/resources/free-chemistry-software-apps/chemsketch-freeware/>, 2023.
<https://www.acdlabs.com/resources/free-chemistry-software-apps/chemsketch-freeware/> (consultado el 13 de julio de 2022).