



CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL
INSTITUTO POLITÉCNICO NACIONAL
UNIDAD MÉRIDA
DEPARTAMENTO DE FÍSICA APLICADA

Kaxan: Un Programa de Búsqueda de Mínimos Globales Empleando Estrategias Evolutivas

Tesis que presenta

DIEGO VICENTE MORENO RODRÍGUEZ

para obtener el grado de
DOCTOR EN CIENCIAS
en
FISICOQUÍMICA

Director de Tesis

Dr. JOSÉ GABRIEL MERINO HERNÁNDEZ

MÉRIDA, YUCATÁN, MÉXICO

ABRIL DE 2016

*We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.*

— Donald E. Knuth¹

AGRADECIMIENTOS

Si a algunas instituciones he de agradecer es al CONACyT por el apoyo de la beca de doctorado para extranjeros, al CGSTIC por permitir obtener los resultados que se presentan en este trabajo usando el clúster híbrido XIUHCOATL. También debo un sentido agradecimiento a la *Universidad de Guanajuato* que sin su diligente apoyo, de algunos directivos del posgrado en Química, no hubiera sido posible gestionar una visa de estudios en este país. A Gabriel Merino por convencerme de emprender esta aventura en tan bonito país, gracias.

A las secretarías del Cinvestav Unidad Mérida, en especial a *Zhirnay Rodríguez* y a *Maricarmen Iturralde*, quienes fueron pieza fundamental en el casi imposible proceso para llegar a Mérida y conseguir finalizar el programa de estudios. Sin su diligente y desinteresada ayuda me hubiera enterrado en el papeleo.

A tantas y tantas personas que me rodearon estos últimos años, que casi sin conocerme me trataron muy bien, entre ellas debo nombrar a Claudia Zavala, Abril, el Rafita (Rafael Islas), Juan Pech, Paco, Said, Saúl, el Pepín, Rodrigo, Jeroen... También le debo un sentido agradecimiento a Jessica y Maryel por los comentarios al primer burdo manuscrito, también a Maryel debo agradecer ayudarme a en estos difíciles momentos, sin su desinteresada ayuda yo no hubiese conseguido llegar al final. A mis hermanas Yeimy y Claudia que desde la distancia me hicieron sentir siempre en casa. A mi padre al cual le debo tanto de lo que soy. A Susana Rocha y Anais Rocha que me permitieron soñar.

Finalmente gracias a la cereza, que es la única que en todos los momentos difíciles nunca, nunca ha dejado de ser un pilar y siempre me hacía cantar.

*Las cosas, tienen
vida propia
—pregonaba el gitano
con áspero acento—,
todo es cuestión de
despertarles el
ánima.
— Cien años de
Soledad²*

RESUMEN

En este trabajo se presenta un nuevo algoritmo genético para la búsqueda de mínimos globales en cúmulos atómicos y moleculares basado en un nuevo método de cruce llamado *fragment-cut-splice*, el cual conecta los fragmentos moleculares haciendo uso de descriptores topológicos. También analizaremos varios factores que influyen en la evolución de la búsqueda usando el algoritmo, no sólo en la eficiencia para encontrar el mínimo global, sino también como mejora la diversidad durante la búsqueda en comparación a las tres estrategias de cruce más comunes en la literatura. Todas las posibles opciones del algoritmo genético, así como otras estrategias evolutivas, se implementaron en un sólo paquete computacional llamado *Kaxan* (*búsqueda* en maya), que permite realizar la búsqueda de un mínimo global, así como el posterior análisis de los resultados.

Además en este trabajo se analizan los resultados obtenidos con *Kaxan* para algunos sistemas moleculares interesantes. El primero de estos es el C_6Li_6 , un cúmulo que muestra una estructura fluxional que se compone de tres dímeros de C_2^{2-} puenteados por seis átomos de litio.

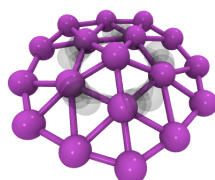
El siguiente par de sistemas estudiado son el cúmulo de B_{18}^{2-} y el B_{19}^- , una pareja que pertenece al grupo de moléculas llamadas “motores Wankel”, debido a que exhibe un movimiento similar al motor del mismo nombre. Además se estudia la estabilidad energética del dianión y su estabilidad al adicionar un átomo de litio. En el segundo se estudia el efecto de sustituir un átomo de boro por uno de carbono sobre las propiedades dinámicas.

Finalmente se presentan los resultados de la búsqueda de la geometría de más baja energía para el cúmulo del conjunto de átomos del segundo periodo de la tabla periódica (a excepción del neón), llamado periodano.

PRODUCTOS PUBLICADOS

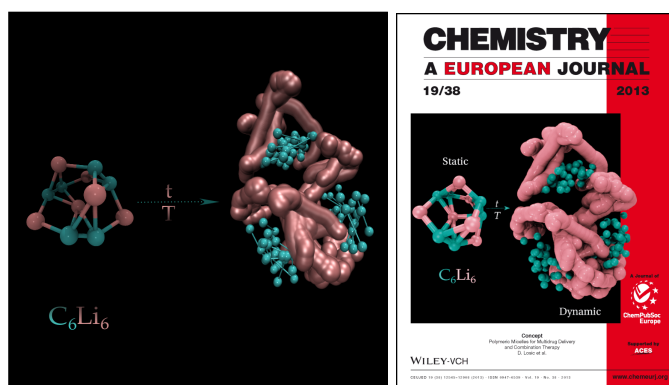
B_{18}^{2-} : a quasi-planar bowl member of the Wankel motor family

Moreno, D.; Pan, S.; Zeonjuk, L. L.; Islas, R.; Osorio, E.; Martínez-Guajardo, G.; Chattaraj, P. K.; Heine, T.; Merino, G.; *Chem. Comm.*, **2014**, *50*, 8140–8143.



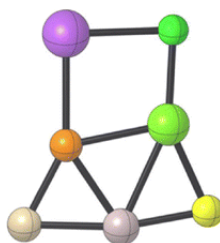
Re-examination of the C_6Li_6 structure: to be, or not to be symmetric

Moreno, D.; Martínez-Guajardo, G.; Díaz-Celaya, A.; Mercero, J. M.; de Coss, R.; Perez-Peralta, N.; Merino, G.; *Chem. Eur. J.*, **2013**, *19*, 12668–12672.



Heavy periodane

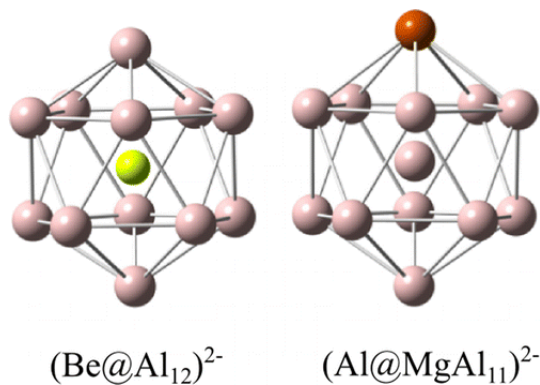
Azpiroz, J. M.; Moreno, D.; Ramirez-Manzanares, A.; Ugalde, J. M.; Mendez-Rojas, M. A.; Merino, G.; *J. Mol. Mod.*, **2012**, *19*, 1953–1958.



Heavy Periodane

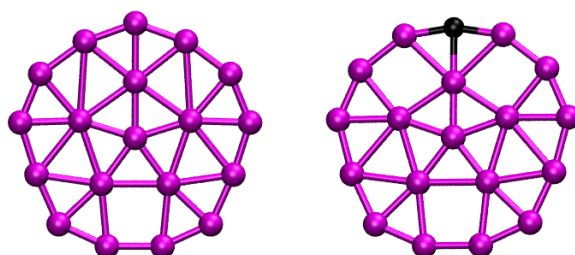
Doped aluminum cluster anions: size matters

Jimenez-Izal, E.; Moreno, D.; Mercero, J. M.; Matxain, J. M.; Audiffred, M.; Merino, G.; Ugalde, J. M.; *J. Phys. Chem. A*, **2014**, *118*, 4309–4314.



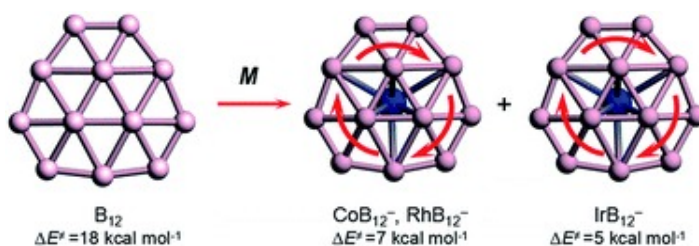
Stop rotating! one substitution halts the B₁₉⁻ motor

Cervantes-Navarro, F.; Martínez-Guajardo, G.; Osorio, E.; Moreno, D.; Tiznado, W.; Islas, R.; Donald, K. J.; Merino, G.; *Chem. Comm.*, **2014**, *50*, 10680–10682.



Structure and bonding of IrB₁₂⁻: converting a rigid boron B₁₂ platelet to a Wankel motor

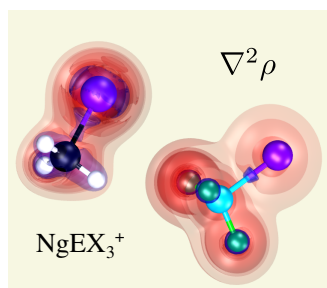
Liu, L.; Moreno, D.; Osorio, E.; Castro, A. C.; Pan, S.; Chattaraj, P. K.; Heine, T.; Merino, G.; *RSC. Adv.*, **2016**, *6*, 27177–27182.



Otros artículos publicados durante el período del doctorado son:

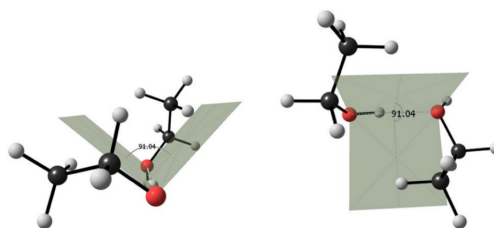
Structure and stability of noble gas bound EX_3^+ compounds (E = C, Ge, Sn, Pb; X = H, F, Cl, Br)

Pan, S.; Moreno, D.; Ghosh, S.; Chattaraj, P. K.; Merino, G.; *J. Comput. Chem.*, On line 1 Jul 2015. DOI: 10.1002/jcc.23986



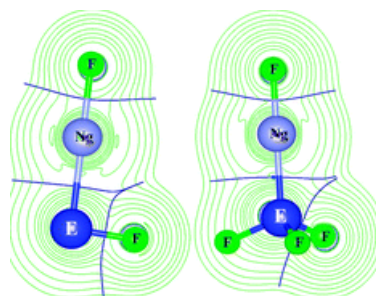
The rich and complex potential energy surface of the ethanol dimer

Vargas-Caamal, A.; Ortiz-Chi, F.; Moreno, D.; Restrepo, A.; Merino, G.; Cabellos, J. L.; *Theor. Chem. Acc.*, 2015, 134, 1–9.



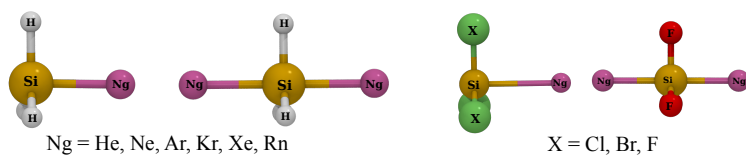
Metastable behavior of noble gas inserted tin and lead fluorides

Pan, S.; Gupta, A.; Mandal, S.; Moreno, D.; Merino, G.; Chattaraj, P. K.; *Phys. Chem. Chem. Phys.*, 2015, 17, 972–982.



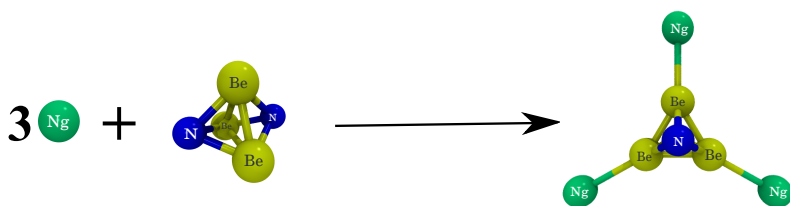
Stability of noble-gas-bound SiH_3^+ clusters

Pan, S.; Moreno, D.; Merino, G.; Chattaraj, P. K; *ChemPhysChem*, **2014**, *15*, 3554–3564.



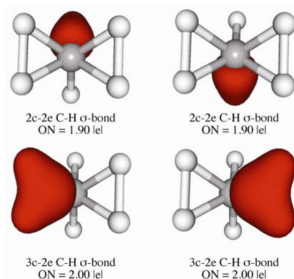
Ab Initio study on the stability of $\text{Ng}_n\text{Be}_2\text{N}_2$, $\text{Ng}_n\text{Be}_3\text{N}_2$ and NgBeSiN_2 clusters

Pan, S.; Moreno, D.; Cabellos, J. L.; Merino, G.; Chattaraj, P. K; *ChemPhysChem*, **2014**, *15*, 2618–2625.



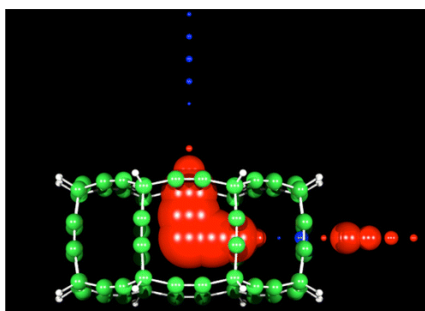
On the nature of CH_6^{2+}

Jalife, S.; Grande-Aztatzi, R.; Moreno, D.; Fernández-Herrera, M. A.; Osorio, E.; Pan, S.; Schleyer, P. v. R.; Martínez-Guajardo, G.; Merino, G.; *Ind. J. Chem.*, **2014**, *53*, 992–995.



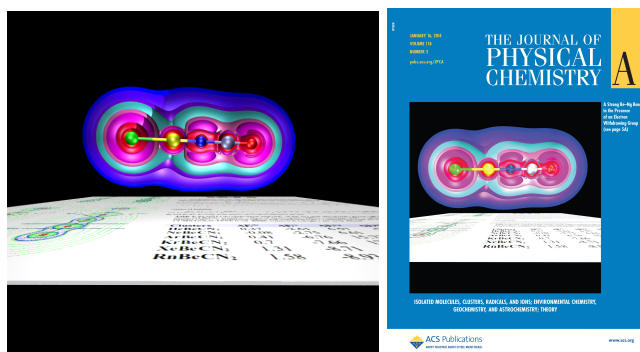
Carbo-cages: a computational study

Azpiroz, J. M; Islas, R.; Moreno, D.; Fernández-Herrera, M. A.; Pan, S.; Chattaraj, P. K; Martínez-Guajardo, G.; Ugalde, J. M.; Merino, G.; *J. Org. Chem.*, **2014**, *79*, 5463–5470.



In quest of strong Be–Ng bonds among the neutral Ng–Be complexes

Pan, S.; Moreno, D.; Cabellos, J. L.; Romero, J.; Reyes, A.; Merino, G.; Chattaraj, P. K; *J. Phys. Chem. A*, **2013**, *118*, 487–494.



O cómo supieron cuál es la virtud de los átomos
y qué podían [hacer] al cambiarse de orden,
si la misma naturaleza no les dio el modelo a crear?

Pues de muchos modos, muchos átomos,
revueltos desde un tiempo infinito por golpes
y movidos por sus propios pesos, se unieron llegando a
ensamblarse de todas las maneras y a intentar todo
lo que podían crear unidos entre sí,
así que no es de admirar que en aquel orden
cayeron y transitaron también aquel curso
por el que ahora es llevado este universo cambiante...

— Tito Lucrecio, "De Rerum Natura, liber Quintus" 3

INTRODUCCIÓN

En cúmulos y moléculas el problema de determinar las posiciones relativas de los átomos es un problema NP-complejo.⁴ Los métodos computacionales de optimización de distancias y ángulos de enlace ofrecen solo la versión local de esta búsqueda, esto es, el usuario ingresa una disposición espacial (y con suerte cercana a un mínimo) de los átomos y el programa las optimiza empleando técnicas basadas en el descenso del gradiente, garantizando la localización del punto estacionario más cercano a la estructura de partida.

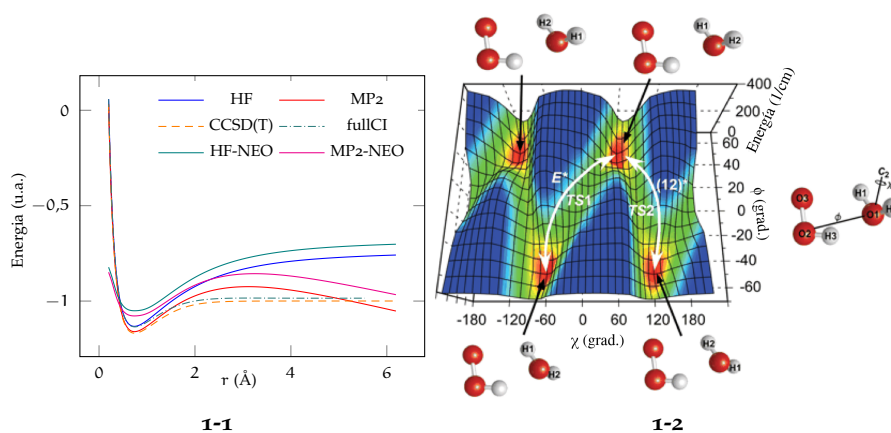


Figura 1: (1-1) Superficie de Energía Potencial (PES) de la molécula de H_2 a varios niveles de teoría. (1-2) PES reducida de H_2O-HO_2 a nivel B3LYP/aug-cc-pVDZ⁵

La superficie de energía potencial crece en complejidad conforme el número de átomos aumenta. Por ejemplo, en el caso de una molécula de dos átomos (figura 1-1), la Superficie de Energía Potencial (PES, por sus siglas en inglés) es una curva en dos dimensiones, es decir,

la energía es una función de la distancia. Pero para el caso de una molécula de seis átomos, como en el complejo $\text{H}_2\text{O}-\text{HO}_2$, la curva se convierte en una hipersuperficie de 12 dimensiones que posee un mínimo global y varios mínimos locales⁵ (figura 1-2). Para atacar este tipo de problema desarrollamos un código al que llamamos Kaxan, un programa que lleva a cabo la optimización global de estructuras moleculares. Hasta la fecha hemos implementado en Kaxan cuatro tipos de metaheurísticas, una basada en una variante del Algoritmo Genético (AG), otra en el Recocido Simulado (RS), una más con base en la Optimización con Enjambre de Partículas (PSO, por sus siglas en inglés) y otra con los fundamentos de Evolución Diferencial (ED).

En esta tesis se divide en dos partes, la primera describe únicamente el AG implementado en Kaxan, sus operadores, así como la influencia de los parámetros durante la búsqueda en cúmulos. También se presenta una nueva estrategia de cruce y el impacto que tiene esta nueva propuesta sobre el problema de optimización. El AG usado, así como la eficiencia y mejora sobre el algoritmo original se exponen en dos capítulos, en donde se resume el estado del arte en la búsqueda de mínimos globales en Química utilizando AGs y se discuten las mejoras sugeridas por otros autores. La segunda parte resume los resultados del uso de Kaxan en la exploración de la PES del cúmulo C_6Li_6 , del cual se había propuesto una estructura simétrica en forma de estrella como la más estable. Además de encontrar que la estructura de mínima energía no es simétrica, hallamos que ésta posee propiedades fluxionales interesantes, tal que los átomos de litio que puentean las unidades C_2^{2-} se mueven libremente. En el capítulo 4 se describe la exploración sistemática de la superficie de energía potencial aplicada esta vez a los cúmulos B_{18}^{2-} y CB_{18}^- . B_{18}^{2-} es el primer cúmulo no plano fluxional que se describe en la literatura. También se discuten los efectos provocados al sustituir un átomo de boro por uno de carbono en el B_{19}^- .

Finalmente en el capítulo 5 se estudia el periodano pesado, un cúmulo que no tiene referente experimental. En este ejercicio puramente teórico se estudia los patrones estructurales y el enlace químico en una molécula formada por todos los elementos del segundo periodo de la Tabla Periódica, excepto el neón.

ÍNDICE GENERAL

Resumen	v
Introducción	xiii
i METAHEURÍSTICAS PARA OPTIMIZACIÓN	1
1 METAHEURÍSTICAS PARA MÍNIMOS EN QUÍMICA	3
1.1 Algoritmo Genético	4
1.2 Algoritmo Genético en Kaxan	6
1.2.1 Población Aleatoria	6
1.2.2 Operadores Genéticos	7
1.2.3 Generación	11
2 UNA NUEVA ESTRATEGIA DE CRUZA	13
2.1 Grafos en Química	13
2.2 Descriptores de Estructura Molecular	15
2.2.1 Índices Topológicos	16
2.2.2 Relación Energía-Estructura	17
2.2.3 Algoritmo del Árbol Recubridor Mínimo	21
2.3 Un nuevo método de cruza usando fragmentos	28
2.3.1 Selección de los fragmentos	28
2.3.2 Mezcla de fragmentos	29
2.4 Resultados con cúmulos atómicos	30
2.4.1 (H ₂ O) ₆	30
2.4.2 B ₁₅ ⁻ y Al ₁₃ ⁻	32
2.4.3 Benceno y Naftaleno.	33
2.5 Diversidad del nuevo algoritmo	36
2.6 Conclusiones	42
ii RESULTADOS PARA SISTEMAS REPRESENTATIVOS	45
3 RE-EXAMINANDO LA ESTRUCTURA DEL C₆Li₆	47
3.1 Introducción	47
3.2 Estructura y Energías	48
3.3 Naturaleza del enlace	49
3.4 Estructura Simétrica vs Asimétrica	50
3.5 Fluxionalidad	51
3.6 Conclusiones	52
4 B₁₈²⁻ UN MIEMBRO QUASIPLANAR DE LA FAMILIA MOTOR DE WANKEL	55
4.1 Introducción	55
4.2 Detalles Computacionales	55
4.3 Resultados	56
4.4 Análisis del Enlace y la fluxionalidad	58
4.5 Aromaticidad	58
4.6 Una sustitución que frena el motor tipo Wankel	62

4.7	Conclusiones	66
5	PERIODANO PESADO	69
5.1	Metodología y detalles computacionales	70
5.2	¿Cuál es el isómero más estable?	70
5.3	Enlace	73
5.4	Conclusiones	73
iii	APÉNDICES	75
A	SIMILITUD MOLECULAR CON MOMENTOS ESTADÍSTICOS	77
A.1	Ultra Fast Recognition	77
B	ÍNDICES TOPOLÓGICOS	79

ÍNDICE DE FIGURAS

Figura 1	(1-1) PES de la molécula de H_2 a varios niveles de teoría. (1-2) PES reducida de H_2O-HO_2 a nivel B3LYP/aug-cc-pVDZ ⁵	xiii
Figura 2	Algoritmo genético en Kaxan	7
Figura 3	Ejemplo del cálculo de la función de aptitud para una generación con una población de 20 individuos	8
Figura 4	Ruleta representando una población de cuatro individuos con diferentes valores de aptitud (f).	9
Figura 5	Diferentes representaciones de grafos moleculares.	14
Figura 6	Isómeros del cúmulo Al_{13}^- , obtenidos a un nivel PBE0/LANL2DZ.	20
Figura 7	Descriptores topológicos en un conjunto de 30 diferentes cúmulos de Al_{13}^- ordenados por energía.	21
Figura 8	Isómeros del cúmulo C_6H_6 , obtenidos a un nivel PBE0/LANL2DZ.	22
Figura 9	Descriptores topológicos para un conjunto de isómeros del cúmulo del C_6H_6	23
Figura 10	Índices topológicos y su relación con la suma del número de coordinación de un conjunto de 90 isómeros del cúmulo Al_{13}^- . En la versión electrónica de este documento puede saber las coordenadas de cada isómero con hacer click sobre él.	24
Figura 11	Índices topológicos y su relación con la suma del número de coordinación de un conjunto de 90 isómeros del cúmulo C_6H_6 . En la versión electrónica de este documento puede saber las coordenadas de cada isómero con hacer click sobre él.	26
Figura 12	Árbol recubridor mínimo de un grafo.	28
Figura 13	Algoritmo de cruza <i>fragment-cut-splice</i> implementado en Kaxan.	29
Figura 14	Mínimos del $(H_2O)_6$	32
Figura 15	Geometría de los mínimos globales de los cúmulos Al_{13}^- y B_{15}^-	33
Figura 16	Geometría de los mínimos globales de C_6H_6 y $C_{10}H_8$	34

Figura 17	Distancia cuadrática media de la diferencia de las distancias entre mínimos, en cada algoritmo de cruza, para todos los mínimos de los resultados de la búsqueda de mínimo global usando Kaxan.	38
Figura 18	Evolución de la diversidad durante la búsqueda de mínimo global para los diferentes métodos de cruza.	40
Figura 19	Estructuras propuestas para el C_6Li_6	47
Figura 20	Isómeros del C_6Li_6	49
Figura 21	Análisis poblacional en orbitales Naturales para el isómero 20-3.	50
Figura 22	Funciones de distribución radial para C_6Li_6 a 300, 600 y 900K.	52
Figura 23	Variación temporal del desplazamiento cuadrático medio $\langle msd \rangle$ para C_6Li_6 a 300, 600 y 900K.	53
Figura 24	Mínimo global para el B_{18}^{2-}	56
Figura 25	Estructuras más estables del B_{18}^{2-}	57
Figura 26	Análisis AdPND para el B_{18}^{2-}	59
Figura 27	Orbitales moleculares π del cúmulo B_{18}^{2-}	60
Figura 28	Formas de los orbitales π de valencia del cúmulo B_{18}^{2-}	60
Figura 29	Perfil B_z^{1nd} del B_{18}^{2-}	61
Figura 30	Estructuras más estables del LiB_{18}^{2-}	62
Figura 31	Estructura de mínimo global de B_{19}^-	63
Figura 32	Isómeros optimizados de CB_{18}	64
Figura 33	Estructura de mínimo global de CB_{18} y el estado de transición relacionado a la rotación del pentágono interno. Las energías relativas están en kcal/mol.	64
Figura 34	Análisis AdPND para CB_{18}	67
Figura 35	Estructuras reportadas previamente para el periodano pesado.	69
Figura 36	Isómeros optimizados del periodano pesado	71
Figura 37	1 Distancias de enlace (Å) e índices de Wiberg (en <i>paréntesis</i>) del periodano pesado. 37-2, 37-3, 37-4 Orbitales Moleculares del periodano pesado. Isovalor 0.03 u.a.	74

ÍNDICE DE TABLAS

Tabla 1	Resultados con diferentes algoritmos de cruce para un algoritmo genético	31
Tabla 2	Número de generación en la que aparece el mínimo global para C_6H_6 con diferentes porcentajes de mutación y elitismo.	36
Tabla 3	Resultados ADEI para los isómeros 5a y 5c . .	51
Tabla 4	Energía de disociación vertical (TK y ΔSCF), EdDA y FGVE	62
Tabla 5	Energías relativas de los mínimos en el periodo pesado (kcal/mol)	72
Tabla 6	Valores numéricos de diferentes índices topológicos de un conjunto de mínimos del cúmulo Al_{13}^-	79
Tabla 7	Valores numéricos de diferentes índices topológicos de un conjunto de mínimos del cúmulo C_6H_6	81

ACRÓNIMOS

PES	Superficie de Energía Potencial
AG	Algoritmo Genético
GEGA	Algoritmo Genético de Gradiente Embebido
RS	Recocido Simulado
ED	Evolución Diferencial
IEDA	Análisis de la Descomposición de la Energía de Isomerización
NBO	Teoría de Orbitales Naturales
BOMD	Dinámica Molecular Born-Oppenheimer
NPA	Análisis de Población de Orbitales Naturales
WBI	Índice de Enlace de Wiberg
FDR	Función de Distribución Radial
ZPE	Energía de Punto Cero

AdNDP Análisis de la Descomposición Natural de la Densidad

NO Número de Ocupación

TK Teorema de Koopman

VEDE Energía de Disociación Vertical

OVSF Función de Green de Valencia-Exterior

AEDE Energía de Disociación Adiabática

HOMO Orbital molecular más alto ocupado

QSPR Relaciones Cuantitativas Estructura-Propiedad

QSAR Relaciones Cuantitativas Estructura-Actividad

DCM Desviación Cuadrática Media

Parte I

METAHEURÍSTICAS PARA OPTIMIZACIÓN

METAHEURÍSTICAS PARA BÚSQUEDAS DE MÍNIMOS GLOBALES EN QUÍMICA

Al considerar la energía como función únicamente de las coordenadas de los núcleos, surge el problema de encontrar el mejor arreglo nuclear en el espacio tridimensional que represente la menor energía (mínimo global), o en otras palabras,

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \quad (1)$$

donde $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función continua que depende de $3N-6$ dimensiones, donde N es el número de átomos y \mathbf{x} representa el conjunto de coordenadas Cartesianas.

En la literatura se ha reportado una amplia gama de minimizadores locales de f , por ejemplo, el método de Newton,⁶ el método de cuasi-Newton,^{6,7} el método de región de confianza,⁸ el método simplex,⁸ cuadrados latinos,⁸ entre otros. Sin embargo, estos algoritmos hacen uso de la información local y, por definición, caen en la cuenca más cercana al punto de inicio y no pueden escaparse de ese mínimo, no importa si es o no el más profundo de todo el espacio de búsqueda. Salir de él implica superar la barrera que lo separa de otras cuencas de f .

Los métodos de búsqueda aleatoria son técnicas aceptables para abordar estos problemas de barrera, a costa de efectuar una inmensa cantidad de ensayos de prueba para asegurar que el resultado es el mejor. En cambio, las metaheurísticas que simulan el comportamiento de sistemas naturales son capaces de escapar de una cuenca en un menor número de pasos, orientando la búsqueda en cada momento dependiendo del comportamiento de f . Aunque éstas tampoco garantizan la obtención del mínimo global, se han aplicado con éxito en problemas de optimización combinatoria, usando siempre una lógica similar: el punto de partida es una solución (o conjunto de soluciones) que típicamente no es óptima, de la cual se obtienen otras similares para elegir una o varias con las que se reinicia el proceso, el cual se detiene cuando se cumple alguna condición establecida.

Las metaheurísticas más conocidas son los algoritmos genéticos, el recocido simulado, los de enjambre (*particle swarm optimization* y *ant bee colony*), entre otros. Aunque no es posible saber si las soluciones que ofrecen estos métodos son óptimas y, en general, no es posible conocer la proximidad de las soluciones al mínimo global, permiten estudiar problemas de gran complejidad de una manera sencilla y obtener soluciones en tiempos razonables.

*For every problem,
there is one solution
which is simple,
neat, and wrong.
—H. L. Mencken*

1.1 ALGORITMO GENÉTICO

El AG es un método de búsqueda con una heurística adaptativa no determinista diseñado para simular el proceso de evolución de las especies, usando la selección natural descrita por Darwin.^{9,10} El AG propuesto originalmente por Holland,¹¹ en 1975, presenta una analogía a la estructura genética y comportamiento de los cromosomas dentro de una población de individuos y emplea los siguientes tres operadores:

- *Selección.* Supervivencia del más apto.
- *Cruza.* Forma en como se aparean los individuos.
- *Mutación.* Modificaciones aleatorias debidas a la interacción con el ambiente.

Para hallar la solución de funciones de una a varias variables, Holland construyó una estructura de datos empleando un esquema de codificación binario al que llamó *cromosoma*. Una sucesión de éstos componen a un individuo. La población inicial se generó a partir de individuos que poseen cromosomas concebidos aleatoriamente. Luego de aplicar repetidamente un operador de cruza y otro de mutación a estos cromosomas, se generan nuevos individuos que, según una función de aptitud, van reemplazando a algunos miembros de la población. La cruza combina partes de esos cromosomas y la mutación simula el proceso de como se introducen las perturbaciones aleatorias a los individuos provenientes del ambiente. Estas operaciones requieren de una selección de parejas de individuos con un sesgo hacia los más aptos. La supervivencia del más apto hace que, eventualmente, la población se mueva hacia la mejor solución.

La propuesta original de Holland se ha aplicado a muchos dominios de búsqueda como el diseño automatizado,¹² el diseño de sistemas anti-terrorismo,¹³ el procesamiento de imágenes¹⁴ *scheduling*,¹⁵ *routing*,¹⁵ la predicción de series de tiempo,¹⁶ el entrenamiento de redes neuronales,¹⁷ las estrategias de inversión,¹⁸ criptografía y seguridad,^{19,20} la teoría de juegos e inteligencia artificial en juegos,^{21,22} el alineamiento de múltiples secuencias en bioinformática,²³ entre otros. En particular, en Química se ha empleado en varios campos como: cinética de procesos químicos,²⁴⁻²⁹ mecanismos de reacción,³⁰⁻³² Qui-miometría,³³ similitud molecular,³⁴⁻⁴¹ análisis conformacional,⁴²⁻⁴⁴ relaciones cuantitativas estructura actividad^{43,45-48} y funciones de energía potencial,⁴⁹ en el estudio de las proteínas se ha enfocado el AG en el estudio del espacio conformacional,⁵⁰⁻⁵³ predicción estructural,⁵⁴⁻⁵⁷ plegamiento,⁵⁸⁻⁶¹ *docking*,⁶²⁻⁶⁶ análisis de resultados de resonancia magnética nuclear⁶⁷ y de rayos X.^{68,69}

En particular, el AG se ha empleado con éxito en la exploración de las superficies de energía potencial de cúmulos atómicos y moleculares.

And yet, in each human coupling, a thousand million sperm vie for a single egg. Multiply those odds by countless generations, against the odds of your ancestors being alive... only you, that emerged. To distill so specific a form from that chaos of improbability, like turning air to gold... that is the crowning unlikelihood. The thermo-dynamic miracle.

-Dr. Manhattan in The Watchmen

Uno de los primeros estudios reportados es el de Judtson *et al.*,⁷⁰ quienes tomaron como cromosomas a un grupo de ángulos diedros para explorar el espacio conformacional de 72 moléculas tomadas de *Cambridge Structural Database*. Más tarde, Xiao y Williams^{71,72} desarrollaron el programa GAME para optimizar pequeños cúmulos moleculares de benceno y naftaleno teniendo como cromosomas la orientación relativa entre las moléculas. Poco después, Hartke⁷³ mostró que para la búsqueda de cúmulos moleculares, el AG es más efectivo que RS. Smith⁷⁴ estudió la posibilidad de encontrar arreglos binarios de átomos en un cristal. Para este problema utilizó igual cantidad de átomos de los tipos A y B en las celdas de un cristal usando cromosomas del tipo AABAABABBAABAAA. Para cada uno de estos arreglos empleó una regla fija para distribuir los átomos en el espacio y calculó la energía del sistema con un potencial interatómico tipo Morse. Tres años después, Mestres y Scuseria⁷⁵ combinaron un potencial *tight binding* y el AG para localizar el mínimo global de pequeños cúmulos de carbono y para ello mezclaron cromosomas construidos a partir de la matriz de adyacencia. En 1995, Deaven y Ho⁷⁶ optimizaron cúmulos de carbono de hasta 60 átomos empleando un potencial *tight binding* y un AG, mostrando que es posible obtener la estructura icosaédrica del fullereno a partir de posiciones aleatorias. Una de las aportaciones de este trabajo es el uso de los operadores de cruce y mutación. El primero de ellos corta aleatoria e independientemente un par individuos, seleccionados no por una función de aptitud sino por la función de distribución de Boltzmann. De cada pareja elige también aleatoriamente una de las mitades y las une para generar una nueva estructura (*plane-cut-splice*). En el caso de la mutación, esta operación sólo se aplica a una fracción de la población. Los nuevos individuos generados por los operadores de cruce y mutación se optimizan utilizando un método basado en gradiente conjugado. Años más tarde, Gregurick y Hartke⁷⁷ propusieron un AG híbrido determinístico/estocástico (DS-GA), combinando el AG clásico con optimizaciones locales de cada individuo y lo usaron con éxito en cúmulos de B y Ar, obteniendo una considerable reducción en el número de generaciones necesarias para completar la búsqueda. En el mismo año, Niesse y Mayne⁷⁸ emplearon el DS-GA, pero cambiando las coordenadas internas por Cartesianas. Esto mejoró los tiempos en las optimizaciones locales para cúmulos de Ar de hasta 29 átomos.

En el 2003, Johnston⁷⁹ combinó el AG, el operador de cruce de Daven y Ho y las optimizaciones locales en aleaciones de metales de transición. Fue el primero en proponer que el AG combinado con “un alto nivel de sofisticación computacional” (*ab initio*) sería ideal para el estudio de cúmulos atómicos. Un año después, Alexandrova y Boldyrev^{80,81} implementaron el Algoritmo Genético de Gradiente Embebido (GEGA, por sus siglas en inglés), que combina los métodos de estructura electrónica y el AG. GEGA emplea una combinación similar a la de Johnston, pero además de las optimizaciones locales ejecuta el

He [Taniyama] was gifted with the special capability of making many mistakes, mostly in the right direction. I envied him for this and tried to imitate him, but found it quite difficult to make good mistakes.

—Goro Shimura (Nova program on Fermat's Last Theorem)

análisis de frecuencias para asegurar que la población solo contenga mínimos locales. En el caso de encontrar frecuencias imaginarias se llevan a cabo algunos pasos de optimización extra, siguiendo el primer modo normal. Esta metodología tiene la ventaja de generar buenos individuos pero a un costo computacional alto. GEGA ha sido usado para optimizar diversos cúmulos atómicos^{80-91,91-97} y moleculares.⁹⁸⁻¹⁰³

En el 2014, Kanters y Kelling¹⁰⁴ presentan CLUSTER, un programa para la búsqueda del mínimo global en cúmulos, el cual hace uso de un AG con un método de cruce similar al de Deaven y Ho, incluyendo un paso de rotación para completar la estequiometría (*plane-rot-splice*), realizando todas las cruces posibles de fragmentos moleculares contenidos en una base de datos que luego optimiza con funcionales de la densidad. Para estos autores, la cruce y el desplazamiento propio de los átomos durante el proceso que involucra seguir el gradiente provocan cierta aleatoriedad sugiriendo eliminar el operador de mutación.

1.2 ALGORITMO GENÉTICO EN KAXAN

Kaxan se construyó usando el paradigma de la programación orientada a objetos y cuenta con varias rutinas tanto en fortran 90-03 como en C++. Además de las diferentes rutinas para búsquedas estocásticas en cúmulos y moléculas, el programa usa un algoritmo de similitud molecular (apéndice A), que le permite realizar una búsqueda eficiente. Durante el trabajo de exploración de la PES, es común llevar a cabo procedimientos rutinarios que ya se han implementado en Kaxan, como lo son la reoptimización de un listado de cúmulos, el cálculo de energía, la búsqueda del mínimo desde estados de transición, y el listado de las estructuras dependiendo de su energía.

El AG, en Kaxan, se encuentra dividido en tres bloques: población aleatoria, operadores genéticos y generación, en la Figura 2 se muestra un diagrama de este algoritmo.

1.2.1 Población Aleatoria

El algoritmo inicia con ciertos parámetros impuestos por el usuario. Entre los más importantes están: la estequiometría, el tamaño de la población, el número de generaciones, el programa de estructura electrónica y el nivel de teoría. Una vez recibidos estos parámetros, el programa genera una población aleatoria de individuos (cúmulos o moléculas con la estequiometría proporcionada por el usuario) de dos tipos: individuos planos y 3D. Las posiciones de los átomos se obtienen usando números pseudo aleatorios.¹⁰⁵⁻¹⁰⁸ Estas estructuras, antes de ser aceptadas, deben cumplir dos condiciones: la distancia entre los átomos no debe ser inferior a la suma de sus radios de covalencia y no deben existir átomos o fragmentos aislados. Con esto,

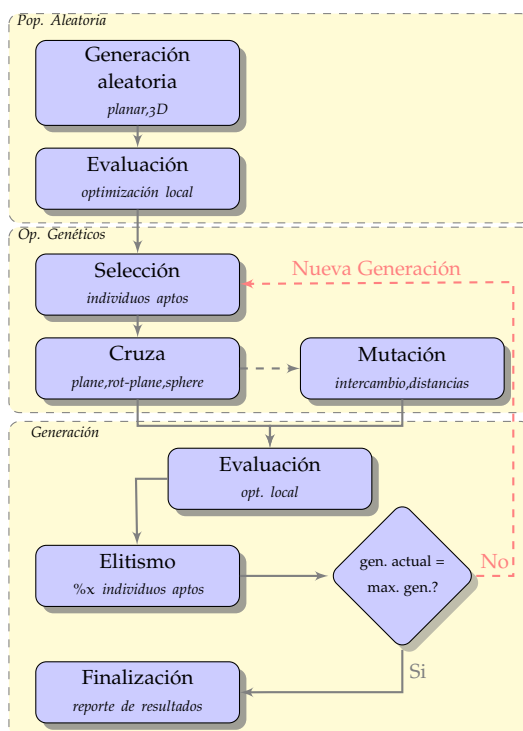


Figura 2: Esquema del AG implementado en Kaxan

se asegura que tengan mejor oportunidad de convertirse en individuos aptos.

El siguiente paso es realizar la optimización empleando un programa de estructura electrónica. Una vez optimizados, sea que lleguen a un mínimo local o no, se actualiza la nueva información de cada individuo (estructura y energía).

1.2.2 Operadores Genéticos

Aquí se ejecutan los operadores de cruce y mutación. En el primer paso de este bloque, se recibe una generación que tiene una energía electrónica asociada a la estructura. Después de descartar individuos idénticos, (Apéndice A), se seleccionan las parejas de estructuras que van a cruzarse para formar la nueva generación. Usando una función de aptitud,^{109,110} a cada individuo se le asigna un valor que representa su aptitud dentro de la población. Este valor es proporcional a su función objetivo (energía). Si denotamos la probabilidad $I_{j,i}^{\text{prob}}$ de que el individuo j en la generación i sea seleccionado, tenemos que,

$$I_{j,i}^{\text{prob}} = \frac{\text{rango}[g(I_{j,i})]}{\lambda(\lambda + 1)/2}, \quad (2)$$

donde $\text{rango}[g(I_{j,i})]$ es el rango de la función del individuo $I_{j,i}$, que se calcula ordenando la población de menor a mayor energía y se

asigna un valor dependiendo de su posición en el ordenamiento. El peor individuo tiene rango = 1, mientras que el mejor tiene rango λ , la suma de los rangos $\lambda(\lambda + 1)/2$ constituye una constante de normalización. Este valor de aptitud es proporcional al rango del individuo e invariante a la traslación y al cambio de escala. Con esta función, todos los individuos tienen una posibilidad, al menos remota, de ser seleccionados independiente de que tan grande es la diferencia de energía. Además previene el problema de convergencia prematura causada por la presencia de superindividuos ¹ que pueden surgir al aplicar la función de selección.

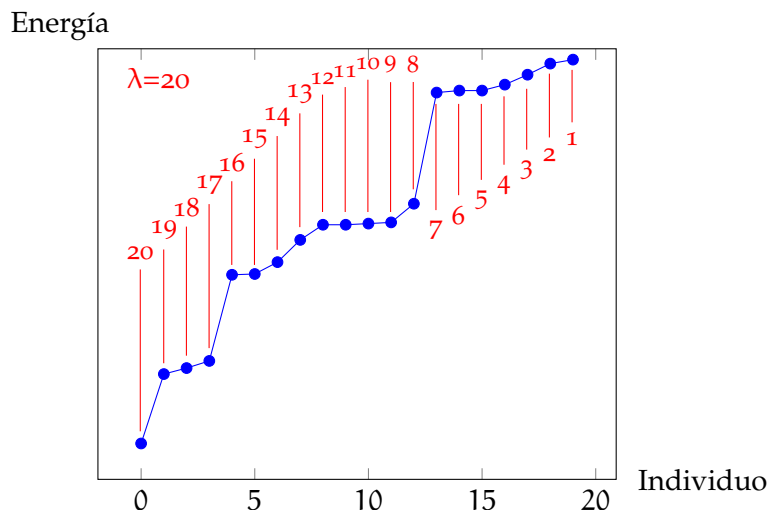


Figura 3: Ejemplo del cálculo de la función de aptitud para una generación con una población de 20 individuos

Como ejemplo, supongamos la primera generación de una búsqueda de 20 individuos (Figura 3), donde se desea calcular el valor de la función de aptitud. Después de ordenarlos de menor a mayor energía se asignan valores de rango a cada uno, según la posición del ordenamiento (Figura 3). Con estos valores se puede calcular la función de aptitud, de esta forma, el individuo número 12 tendrá un valor de aptitud de 0,047.

$$I_{12,1} = \frac{9}{20(20+1) * 0,5} = 0,047. \quad (3)$$

SELECCIÓN. El número de individuos que se van a cruzar se seleccionan usando el método de la ruleta.^{111,112} Este asigna al individuo i

¹ Los superindividuos resultan de la combinación de características óptimas de parejas de individuos y su adaptación es mucho mayor que la de los demás de la misma generación. Sin una función de aptitud, normalizada e invariante a los cambios de escala, los individuos de la generación siguiente, en su mayoría, tendrían los genes del superindividuo.

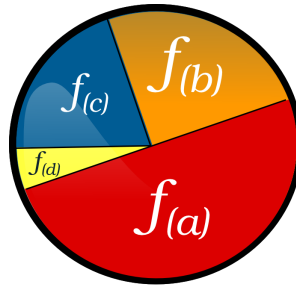
la probabilidad P_i , empleando el valor de aptitud F_i , de la siguiente manera:

$$P_i = \frac{F_i}{\sum_{j=1}^n F_j} \quad n = \text{Total de la población.} \quad (4)$$

Con P_i se construye el arreglo \mathcal{R} dado por,

$$\mathcal{R}_n = \sum_{i=1}^n P_i. \quad (5)$$

Un individuo en \mathcal{R}_x es elegido si el número aleatorio entre 0 a $\sum_{j=1}^n P_j$, está en el rango $\mathcal{R}_{x-1} < \mathcal{R}_x \leq \mathcal{R}_{x+1}$. Este proceso se repite partiendo desde la selección anterior hasta que el número de individuos sea igual a $2n$.



$$f_{(a)} > f_{(b)} > f_{(c)} > f_{(d)}$$

Figura 4: Ruleta representando una población de cuatro individuos con diferentes valores de aptitud (f).

CRUZA. El operador de cruce define cómo se mezclan los genes de las parejas seleccionadas para formar un nuevo individuo. En Kaxan se han implementado cuatro métodos de cruce de individuos *plane-cut-splice*, *plane-rot-splice*, *sphere-cut-splice* y *fragment-cut-splice*.

- *plane-cut-splice*.^{76,113} Este método genera una nueva estructura con mitades de una pareja de individuos. La forma en que se obtienen es empleando un plano aleatorio (xy , yz y xz) que pasa por el centro de masas de cada uno. Finalmente, se toma una de las mitades de cada individuo y se unen por el plano de corte.
- *plane-rot-splice*.^{76,79,114} Este procedimiento es una variante de *plane-cut-splice*, en la que el plano aleatorio, que pasa por la pareja de individuos, rota sobre un eje arbitrario hasta que la dos mitades tienen la estequiometría definida por el usuario.
- *spheric-cut-splice*. Esta cruce propuesta por Chen *et al.*¹¹⁵ consiste en crear una esfera interna común para una pareja de estructuras

(P1 y P2), tomando como punto de referencia el centro de masas de cada una. Luego los átomos contenidos en la esfera se intercambian para formar un nuevo individuo. Para delimitar esta esfera primero se ordenan los átomos de acuerdo a la distancia al centro de masa

$$r_k^1 < r_{k+1}^1 \text{ y } r_k^2 < r_{k+1}^2, k = 1, 2, \dots, N - 1, \quad (6)$$

donde r_k^1 y r_k^2 son la distancia del k -ésimo átomo al centro de masa del cúmulo P1 y P2, respectivamente, y N es el número total de átomos en el cúmulo. Luego los k átomos que pueden estar contenidos en la esfera deben satisfacer dos condiciones:

$$r_k^1 < r_{k+1}^2 \text{ y } r_k^2 < r_{k+1}^1, k = 1, 2, \dots, N - 1. \quad (7)$$

El conjunto M que contiene el número de átomos que puede intercambiarse es

$$M = \left\{ k \mid (r_k^1 < r_{k+1}^2) \cap (r_k^2 < r_{k+1}^1), k = 1, 2, \dots, N - 1 \right\}. \quad (8)$$

Finalmente, se selecciona un elemento del conjunto M y éste se guarda como m . El radio r de la esfera que define el corte de P1 y P2 es r_m .

Después de que la esfera se define, los padres pueden intercambiar los átomos de la parte interior de la esfera para producir un nuevo individuo.

- *fragment-cut-splice*. Es el tema que se aborda en el capítulo siguiente.

MUTACIÓN. Una vez que se tiene un nuevo conjunto de individuos un porcentaje se selecciona aleatoriamente para llevar a cabo una o varias mutaciones. La mutación puede ser de dos formas:

- *Reemplazo atómico*. Selecciona un número de parejas de átomos al azar y se intercambian.
- *Desplazamiento atómico*. Selecciona un número de átomos al azar que se desplazarán, también aleatoriamente, de su posición original.

En estos operadores siempre se mantiene la conectividad de los átomos y se evita que éstos se colapsen. Además de que no existan individuos duplicados. El número de átomos, así como el tipo de mutación que se lleva a cabo sobre el cúmulo seleccionado, es aleatorio.

1.2.3 Generación

Una vez que se aplican estos tres operadores se obtiene un conjunto de estructuras que se optimiza localmente usando un programa de estructura electrónica. Este es el paso que demanda más tiempo, ya que depende del número de átomos y el nivel de teoría.

Luego de reunir la información (geometrías y energías) y de eliminar las estructuras idénticas, se aplican de nuevo los operadores genéticos hasta alcanzar el número máximo de generaciones. Al finalizar cada generación, los individuos con un nivel de aptitud alto (elitismo) se adicionan en la generación siguiente, sin pasar por la cruce o mutación. El número de individuos que se adicionará dependerá del factor de elitismo que el usuario seleccione. (Figura 2)

Después de una revisión en la literatura en el uso del AG para el problema de la búsqueda de mínimo global en cúmulos, reunimos 4 factores que son de importancia en la eficiencia del AG. Estos son:

- La optimización local de los individuos.
- Una generación inicial óptima.
- El nivel de teoría adecuado en la optimización local.
- El método de cruce conveniente para cúmulos atómicos y moleculares.

Al incluir la optimización local de los individuos se reduce el número de generaciones necesarias para encontrar el mínimo global,^{76,113} pasando de cientos de generaciones, a solo 10 o 20 para un cúmulo de 15 – 20 átomos.^{86,91} En el algoritmo propuesto por Holland, la evolución del fenotipo se determina por el modelo de Darwin, donde el genotipo de los individuos se define únicamente por la combinación de los cromosomas de los padres con valor de aptitud alto. Optimizar cada estructura cambia el modelo de evolución debido a que los nuevos individuos modifican y adaptan estos cromosomas, originando nuevos patrones genéticos y como consecuencia, alcanzan niveles de diversidad más altos en un menor número de generaciones. Este tipo de régimen evolutivo donde los individuos de una especie imprimen cambios en el material genético por adaptación se formuló en 1809 por Lamarck.¹¹⁶

Otro factor importante es la generación inicial de individuos. Si bien la posición de los átomos en el espacio es un proceso aleatorio, las posiciones relativas entre los mismos deben ser las adecuadas, de esta manera se permite que las estructuras converjan a un mínimo.¹¹⁷⁻¹¹⁹

Por otro lado, el nivel de teoría seleccionado debe guardar un compromiso entre velocidad y precisión a fin de conseguir mínimos con relativa rapidez, así como geometrías más cercanas a un mínimo local. Una pobre descripción del potencial interelectrónico no reproduce efectos tales como las distorsiones Jahn-Teller⁷⁹ y en varios casos no se

alcanza el mínimo global^{91,117,120} o este no corresponde con el mínimo global a un nivel de teoría más riguroso.⁷⁹

Finalmente, un operador de cruce adecuado es un componente primordial para explorar todo el espacio de búsqueda,^{76,82,104,113} incluso algunos autores han propuesto que con el método de cruce apropiado se puede prescindir del operador de mutación.¹⁰⁴

Con estos cuatro factores, en Kaxan hemos modificado la generación aleatoria, la selección, la cruce y la mutación, para obtener mejores resultados y lograr una búsqueda amplia sin perder eficiencia durante la exploración. El AG, en Kaxan, se encuentra dividido en tres bloques: Inicialización, operadores genéticos y generación.

FRAGMENT-CUT-SPLICE. UNA NUEVA ESTRATEGIA DE CRUZA MOLECULAR PARA ALGORITMOS GENÉTICOS

En este capítulo se presenta una nueva estrategia de cruza, (*fragment-cut-splice*), que involucra la formación de cromosomas a partir de fragmentos enlazados de una pareja de individuos, en lugar de usar una construcción de cromosomas basada en la disposición espacial de los cúmulos, la cual según algunos autores^{82,104} no representa una opción adecuada, ya que la transferencia de las características genéticas de los cúmulos pueden perderse durante el proceso de cruza del tipo *plane-cut-splice*.

Se iniciará con un breve preámbulo sobre algunos tópicos de la teoría de grafos en Química, desde el punto de vista práctico y sólo nombrando las definiciones que serán usadas a lo largo del texto. Con ellas se aborda la construcción de los índices topológicos como herramienta para correlacionar la energía con las características espaciales de dos cúmulos de átomos que servirán de ejemplo para explorar dicha relación. También de forma breve se describe el algoritmo para construir árboles a partir de grafos que se empleará como herramienta para fragmentar grafos moleculares. Estos elementos se ensamblan en un nuevo método de cruza en AG para cúmulos atómicos llamado *fragment-cut-splice* y finalmente este nuevo método junto con las metodologías empleadas por otros autores son puestas a prueba con algunos resultados numéricos para evaluar su desempeño y efectos sobre el AG.

2.1 GRAFOS EN QUÍMICA

El uso de la teoría de grafos es amplia en Física, Química, Electrónica, Biología y en muchas otras ramas de la ciencia y la ingeniería.¹²¹ Un grafo muestra las relaciones entre objetos, representados por vértices y la relación existente entre estos por una arista.^{121,122} En Química, los grafos pueden representar diversos objetos como orbitales,¹²³⁻¹²⁶ átomos,¹²⁷⁻¹²⁹ enlaces,¹²⁷⁻¹²⁹ grupos de átomos,¹²⁷⁻¹²⁹ moléculas o colección de moléculas,^{127-129,131,132} además de definir las relaciones entre estos. Los grafos permiten describir enlaces, reacciones,^{133,134} mecanismos de reacción,¹²⁹ modelos cinéticos^{129,134} y cualquier interacción o transformación de los objetos químicos. En la literatura existen ejemplos del amplio uso de los grafos en Química^{127,128,131,135-137} como una importante herramienta en la exploración de aplicaciones que son de gran ayuda para la investigación. El uso más común es la

A veces la verdad no importa como debiera... pero siempre debes decirla.

– Jhon Hartigan en Sin City

I do not know whether chemists are of opinion that every chemico-graph exist or is capable of existence in nature... (should it be discovered) must admit of being stated in mathematical terms.

– J. J. Sylvester 1878¹³⁰

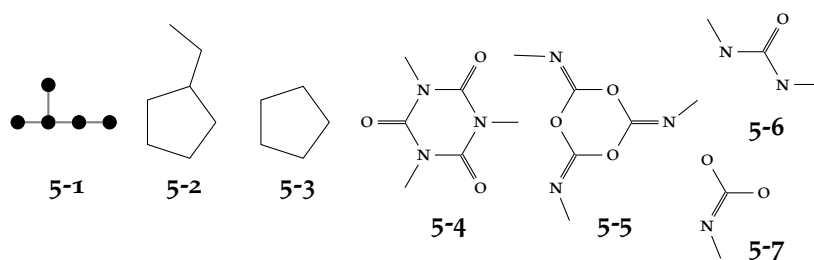


Figura 5: Diferentes representaciones de grafos moleculares.

representación de compuestos químicos como grafos moleculares, en los que los átomos se simbolizan con vértices y el enlace entre ellos con aristas.

Un grafo $G = G(V, E)$ se define como un par ordenado que consiste de dos conjuntos $V = V(G)$ y $E = E(G)$, donde los elementos del conjunto E determinan las relaciones binarias entre elementos del conjunto V . La manera gráfica de representar los elementos en V , es con vértices mientras que los elementos del conjunto E , se simbolizan vía aristas que conectan los vértices. El número de elementos en $V(G)$, $N = |V(G)|$ delimita el número de vértices N en el grafo G y el número de elementos en $E(G)$, $M = |E(G)|$, fija el número de aristas M . Una arista $e \in E(G)$ incidente con dos distintos vértices $v, v' \in V(G)$ se denota como $e_{vv'}$. Los vértices en el grafo se representan usualmente con círculos negros (figura 5-1) o por el punto de unión de dos aristas (figura 5-2), siempre y cuando el ángulo entre las mismas no sea igual a 180° .

Un subgrafo G' de G es un grafo en el que los vértices y las aristas están contenidos en G . Si $V(G')$ es un subconjunto de $V(G)$, $V(G') \subseteq V(G)$, y $E(G')$ es un subconjunto de $E(G)$, $E(G') \subseteq E(G)$, entonces el subgrafo $G' = V(G'), E(G')$ es un subgrafo del grafo $G = V(G), E(G)$. El grafo 5-3 es un subgrafo de 5-4 porque todos los vértices y aristas de 5-3 están contenidos en 5-2. El grafo G y el grafo G' tienen la misma estequiometría sí y sólo sí $V(G) = V(G')$, de lo contrario es necesario que se cumpla $E(G) = E(G')$. Un subgrafo G' se dice que es el complemento estequiométrico de G'' , si el grafo G tal que $V(G) = V(G') \cup V(G'')$ es un grafo que tiene una estequiometría antes especificada. El grafo 5-4 y el grafo 5-5 tienen la misma estequiometría porque tienen igual número y tipo de átomos. El grafo 5-6 es un ejemplo del grafo complemento estereométrico de 5-7 para la estequiometría de 5-4 y 5-5.

La matriz de adyacencia del grafo G $(A(G) = (A)_{ij})$ contiene información sobre la conectividad entre los vértices. Cada elemento $(A)_{ij}$ es 1 si los vértices i y j son adyacentes en G (conectados por una arista) y $(A)_{ij} = 0$ en el caso que no lo estén. La matriz de distancias $C(G)$ es un arreglo en el que cada elemento $(C)_{ij}$ contiene la distancia cartesiana entre el átomo i y j en un cúmulo. La matriz de distancia

If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.
– Jhon Von Neumann

del grafo $\mathbf{D} = \mathbf{D}(\mathbf{G})$ de un grafo conectado \mathbf{G} es una matriz simétrica en la que sus elementos $(\mathbf{D})_{ij}$ se determinan de la siguiente manera

$$(\mathbf{D})_{ij} = \begin{cases} d_{ij}, & \text{si } i \neq j, \\ 0, & \text{si } i = j, \end{cases} \quad (9)$$

donde d_{ij} es la longitud del camino más corto entre los vértices i y j .

La matriz de distancia-resistencia introducida por Klein y Randić es una función de distancia en grafos llamada distancia de resistencia.¹³⁸ Esta nueva distancia del grafo se utiliza para definir la matriz de distancia-resistencia $\mathbf{\Omega}$, propuesta como alternativa a la matriz de distancia \mathbf{D} . La idea principal sobre la que se construye la matriz $\mathbf{\Omega}$ es considerar que los vértices $V(\mathbf{G})$ (del grafo \mathbf{G} conexo) representan una red eléctrica. Las aristas son resistencias igual a un ohm cada una, el par de vértices i y j en \mathbf{G} es igual a la resistencia efectiva entre los vértices v_i y v_j . Si \mathbf{G} es un árbol (por ejemplo un hidrocarburo alifático acíclico), $(\mathbf{\Omega})_{ij}$ es simplemente la suma de las resistencias. Si \mathbf{G} contiene ciclos entonces se usa la ley de Kirchhoff para obtener $(\mathbf{\Omega})_{ij}$. La matriz de distancia-resistencia es el resultado de ordenar los resultados obtenidos para cada par de vértices en forma tabular.

2.2 DESCRIPTORES DE ESTRUCTURA MOLECULAR

Los índices topológicos tienen diversas aplicaciones en Química. Entre otras se emplean con frecuencia en las Relaciones Cuantitativas Estructura-Propiedad (QSPR, por sus siglas en inglés), y las Relaciones Cuantitativas Estructura-Actividad (QSAR, por sus siglas en inglés),¹³⁹⁻¹⁴³ la teoría de orbitales moleculares de compuestos conjugados,^{144,145} la estructura de hidrocarburos bencenoides,^{146,147} la enumeración de isómeros,^{131,148,149} en la percepción de la simetría constitucional,^{150,151} la codificación de compuestos químicos,¹⁵²⁻¹⁵⁴ la cinética de reacción,¹⁵⁴ y el diseño de síntesis asistida por computadora.¹⁵⁵ Además tienen una repercusión importante en la codificación de la información estructural y propiedades de moléculas.

Estos índices son descriptores numéricos derivados de la estructura molecular. A mediados de los 70's, estos índices incluían sólo información sobre la conectividad de los átomos pero a pesar de ello tienen la capacidad de describir satisfactoriamente muchas propiedades¹⁵⁶ de hidrocarburos saturados^{149,157,158} como los puntos de ebullición,¹⁵⁹⁻¹⁶⁵ la refractividad molar,^{142,163,165,166} el volumen molecular^{127,167} y el calor de formación.^{142,158,165} En la actualidad, adicional a la conectividad, la información codificada en ellos incluye la naturaleza de los átomos y la multiplicidad de los enlaces.^{142,158} Otros avances recientes incluyen la generación de índices de fragmentos moleculares¹⁶⁸ con características estereoquímicas importantes, así como parámetros electrónicos asociados a los átomos.

La industria farmacéutica es la que ha impulsado el desarrollo de nuevos descriptores moleculares,^{149,157} con énfasis en los que puedan determinar correctamente las propiedades fisicoquímicas o las propiedades biológicas, gracias a que estos descriptores dan respuesta rápida y económica a la evaluación de millones de compuestos que pueden tener una aplicación activa en nuevos fármacos.¹⁶⁹⁻¹⁷¹ Estos índices no sólo tienen relevancia en la actividad de los compuestos, característica que es ampliamente explotada en los métodos QSAR,^{158,161,165,172} sino que también los descriptores topológicos que se usan en la búsqueda de nuevos compuestos en donde se buscan fragmentos con propiedades particulares.^{149,170,173} Otro uso de estos índices moleculares¹⁵¹ es la construcción de medidas de similitud o disimilitud^{158,163} y desarrollo de métodos para generación de estructuras usando redes neuronales.^{43,149}

2.2.1 Índices Topológicos

Entre este tipo de descriptores que relacionan numéricamente la topología molecular, destacan tres por su sencilla construcción, estos son: el índice de Wiener, el de Balaban y el de Kirchhoff.

Uno de los primeros índices empleados para correlacionar la conectividad molecular con las propiedades es el índice de Wiener.¹⁵⁹ Wiener^{159,174} obtuvo correlaciones biparamétricas con los puntos de ebullición además de otros parámetros termodinámicos. Stiel y Thodos¹⁷⁵ calcularon con precisión algunas constantes críticas aplicando el índice de Wiener, mientras que Rouvray y Crafford¹⁷⁶ hallaron correlaciones con la densidad, viscosidad y tensión superficial. Papazova *et al.*¹⁷⁷ y Bonchev *et al.*¹⁷⁸ propusieron que existe una relación de este índice con los tiempos de retención en cromatografía. El índice de Wiener se define para grafos arbitrariamente conectados como la suma de cada elemento de la matriz de distancias **D** del grafo **G**:

$$W(\mathbf{G}) = \sum_{i=1}^n \sum_{j=1}^i (\mathbf{D})_{ij} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{D})_{ij}. \quad (10)$$

Como este índice se diseñó para grafos con todas sus aristas equivalentes, cada elemento $(\mathbf{D})_{ij}$ es la distancia grafo teórica en **G** entre *i* y *j*. Mihalić y Trinajstić¹⁷⁹ usaron la matriz de distancias **C** de una molécula **G** basada en las distancias cartesianas, así esta matriz topológica contiene información de la estructura 3-D de la molécula. Con esta matriz definieron un índice de Wiener como la semi-suma de la matriz **C** que puede describir apropiadamente varias propiedades moleculares.¹⁸⁰ Mihalić y Trinajstić lo llamaron índice de Wiener 3-D y se denota como 3-*W*.

De manera similar se pueden generar otros índices simplemente cambiando la matriz de distancias **D**. Por ejemplo, el índice *rWf* se

No hay que confundir el concepto de la matriz de distancia del grafo **D**, con la matriz de distancias **C**. La primera representa la distancia más corta entre los vértices de un grafo viajando por las aristas, la segunda representa la distancia euclidiana entre un par de puntos en el espacio de coordenadas *x, y, z*.

calcula empleando los elementos de la matriz Ω . Igualmente Randić¹⁸¹ propuso la matriz cociente \mathbf{D}/Ω , obtenida por dividir los elementos de la matriz de distancia con los elementos (fuera de la diagonal) correspondientes de la matriz de distancia-resistencia. Con esta nueva matriz de la ecuación 10 se obtiene el índice Wq .

El índice de Balaban, J , se define como

$$J = \frac{E}{\mu + 1} \sum_{\text{aristas}} (d_i d_j)^{-1/2} \quad (11)$$

donde E es el número de aristas en el grafo \mathbf{G} , μ es el número ciclomático de \mathbf{G} y d_i es la suma de todas las distancias desde el vértice i , (esto es, la suma de todos los elementos de la columna i en la matriz \mathbf{D}). El número ciclomático μ determina la cantidad de ciclos en el grafo \mathbf{G} y es igual al número mínimo de aristas que son necesarias remover para convertir \mathbf{G} en un grafo acíclico.^{127,128} μ está dado por

$$\mu = E - V + 1 \quad (12)$$

El índice de Balaban también puede calcularse utilizando la matriz de distancia-resistencia Ω en lugar de la matriz de distancia \mathbf{D} . En este caso se denota al índice de Balaban como J_Ω . De forma similar, como en el índice de Wiener, se define el índice $3-J$ como el índice calculado usando la ecuación 11 en donde se reemplaza la matriz \mathbf{D} por la matriz \mathbf{C} .

Con la matriz de distancia-resistencia Babić *et al.*¹⁸² reportaron un índice similar al de Wiener, pero con Ω . De esta manera el nuevo índice, al que llamaron índice de Kirchhoff (Kf), se calcula como sigue:

$$Kf = \frac{1}{2} \sum_{i=1}^V \sum_{j=1}^V (\Omega)_{ij} \quad (13)$$

donde $(\Omega)_{ij}$ representa el elemento que corresponde al vértice i y j de Ω . También Babić define el índice de Kfq como el índice calculado con la matriz cociente Ω/\mathbf{D} .

2.2.2 Relación Energía-Estructura

Como se mencionó en la sección anterior, los índices topológicos se emplean para relacionar linealmente la estructura molecular con alguna propiedad medida de forma experimental. Con el uso de estos índices, en esta sección, se intenta encontrar uno que pueda relacionar estructuras con características comunes. De esta manera, este valor numérico puede servir de guía para la construcción automática de estructuras con rasgos similares a la geometría molecular de un mínimo.

Empleando dos cúmulos (Al_{13}^- , C_6H_6), cada uno con 90 isómeros, se calcularon los 9 índices de la sección anterior. Cada uno de estos

isómeros se obtienen empleando el AG en Kaxan tal que todos ellos sean un mínimo sobre la superficie de potencial a nivel PBE0/LANL2DZ. En cada cúmulo, los isómeros se ordenaron de forma ascendente en energía, de manera que el mínimo global es el primero de la lista. Cada mínimo se le asignó un identificador (ID) que coincide con su posición en el ordenamiento ascendente de energía. De esta manera, el mínimo global tiene un ID=1 y el isómero con energía más alta tiene el ID = 90.

En la [Figura 6](#) y en la [Figura 8](#) se muestran los primeros 70 isómeros de cada cúmulo. En la [Figura 7](#) y en [Figura 9](#) se muestran los valores de los 9 índices para cada uno de los mínimos, los cuales aparecen separados en tres series (los índices de Wiener, de Kirchhoff y los de Balaban). Los índices en estos resultados se muestran como índices relativos al valor del índice del mínimo global de cada cúmulo. En el [Apéndice B](#) se encuentran los resultados numéricos de las figuras 7 y 9.

Para los isómeros de Al_{13}^- , las tres series de índices tienen tendencia a alejarse de 0 (valor del índice del mínimo) según aumenta la energía, aunque en general hay una relación directa entre el valor del índice y el ID del cúmulo. En las tres series hay cuatro estructuras en las que el valor del índice siempre está cercano a la del mínimo (2, 4, 20 y j) para el resto de los isómeros, para los demás cúmulos tienen un valor que puede acercarse o alejarse según el tipo de índice. Para 2, 4 y 20, la similitud estructural (véase [Figura 6](#)) de estos tres cúmulos explica porqué es constante el comportamiento de todos los índices. Aunque 20 parece no tener la misma estructura globular que 1, 2 y 4, éste tiene un valor de índice más cercano a 1 que 2 y 4 en todos los casos, debido a que 20 es el cúmulo que tiene una mayor similitud con el mínimo global donde sólo uno de los átomos se desplaza de uno de los extremos a una cara lateral. En el caso de 48, el cual no tiene la misma estructura globular, éste comparte las mismas características topológicas que 1 (12 átomos en los extremos con un átomo alojado al interior de la estructura).

En el caso de los cúmulos de C_6H_6 no hay una tendencia con el ID del cúmulo como en Al_{13}^- y en las tres series, los índices tienen valores negativos y positivos, pero si se localizaron cúmulos que tienen índices con valores relativos con tendencia similar al mínimo global, los cúmulos que presentan dicho comportamiento son: 2, 3, 14, 33, y 50. Todos estos mínimos tienen ciclos de 4 (3 y 14), 5 (2 y 50) y 6 (33) miembros. Todos tienen un grupo un plano de simetría que pasa por la molécula (como en el benceno) y en el caso de 2, 3, 33 y 50 tienen grupo puntual C_{2v} . Además tienen más de un doble enlace.

En las figuras [Figura 10](#) y [Figura 11](#) se muestra la relación entre el ID del cúmulo (energía), índice y el número de enlaces por cúmulo, este último lo usamos como indicador de la variación topológica de cada grafo. Para los cúmulos de Al_{13}^- , como se describió antes, hay

una relación directa entre el ID del cúmulo y el valor del índice. Sin embargo, en la [Figura 7](#) y la [Figura 9](#) hay algunos isómeros con valores cercanos a 0 que no están entre los de menor energía (ID cercanos a 1). En la [Figura 10](#), el ordenamiento es diferente, la estructura 20 aparece cerca a los primeros tres cúmulos para todos los nueve índices y el cúmulo 48, como era de esperarse, aparece separado del resto. La clasificación de los cúmulos es más clara en la [Figura 10](#). No obstante, en [10-1](#), [10-2](#), [10-3](#), [10-7](#) y [10-9](#), el ordenamiento de los cúmulos es menos disperso que en [10-4](#), los índices de Kirchhoff ([10-5](#) y [10-6](#)) y [10-9](#). Los índices de Wiener pueden clasificar cúmulos en un rango más amplio comparados con los de Balaban, además para estos cúmulos el mínimo global tiene el valor máximo y sólo en Wq la diferenciación entre cúmulos es mayor.

En la [Figura 11](#), la relación entre el ID de los cúmulos el índice y el número de enlaces se agrupa en un menor rango de valores de número de enlaces que en la [Figura 11](#), y pueden distinguirse 5 grupos, en los que tienen un número de enlaces entre 11 y 12 se clasifican los cúmulos de menor energía, pero sólo en Wq , al igual que en el caso anterior, es el que puede diferenciar cúmulos en un rango más amplio que los demás índices. Aunque en los índices de la [Figura 11](#), la clasificación de los cúmulos se mantiene constante, sólo en los índices de Wiener se consiguió una mayor clasificación que en los índices de Kirchhoff y los de Balaban, incluso en J_{Ω} ([11-9](#)) no fue posible diferenciar los cúmulos y se obtienen valores idénticos para los cúmulos 4 a 13.

Los índices en la [Figura 11](#), al igual que en Al_{13}^- , clasifican los mismos cúmulos (1, 2, 3, 14, 17, 20 y 21) con valores de índice cercanos a 0, este hecho se presentó debido a un ordenamiento respecto al número de ciclos ($20 > 21 > 1 > 2 > 3 = 4$) y en específico para el caso de los índices de Wiener los cúmulos con ciclos anidados tienen valor mayor de índice al mínimo global. Hay que destacar además que en el caso de los rotámeros (4 con 5, 6 con 7, 8 con 12, 15 con 19, 27 con 30) sólo Wq logra diferenciarlos en cada par de cúmulos. Al igual pasa con los isómeros configuracionales (1 con 55) y con los isómeros posicionales (33 con 44).

El descriptor topológico Wq describe no sólo el orden en la energía de los cúmulos, además puede diferenciar estructuras muy similares debido a que este índice presentó un rango amplio de valores, permitiendo de esta manera encontrar diferencias numéricas pequeñas que otros índices con menor rango de valores de índice no consiguen.

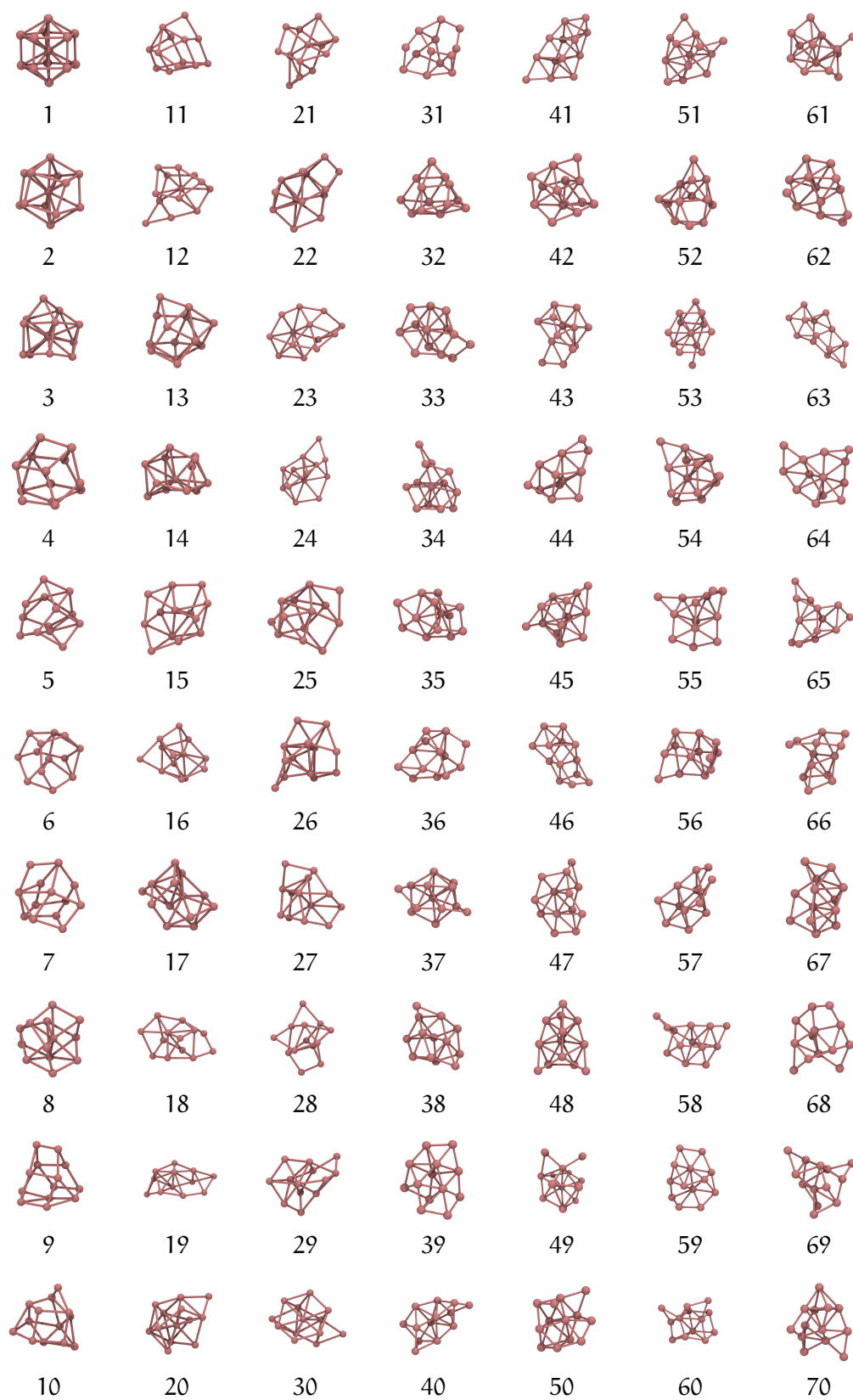


Figura 6: Isómeros del cúmulo Al_{13} , obtenidos a un nivel PBE0/LANL2DZ.

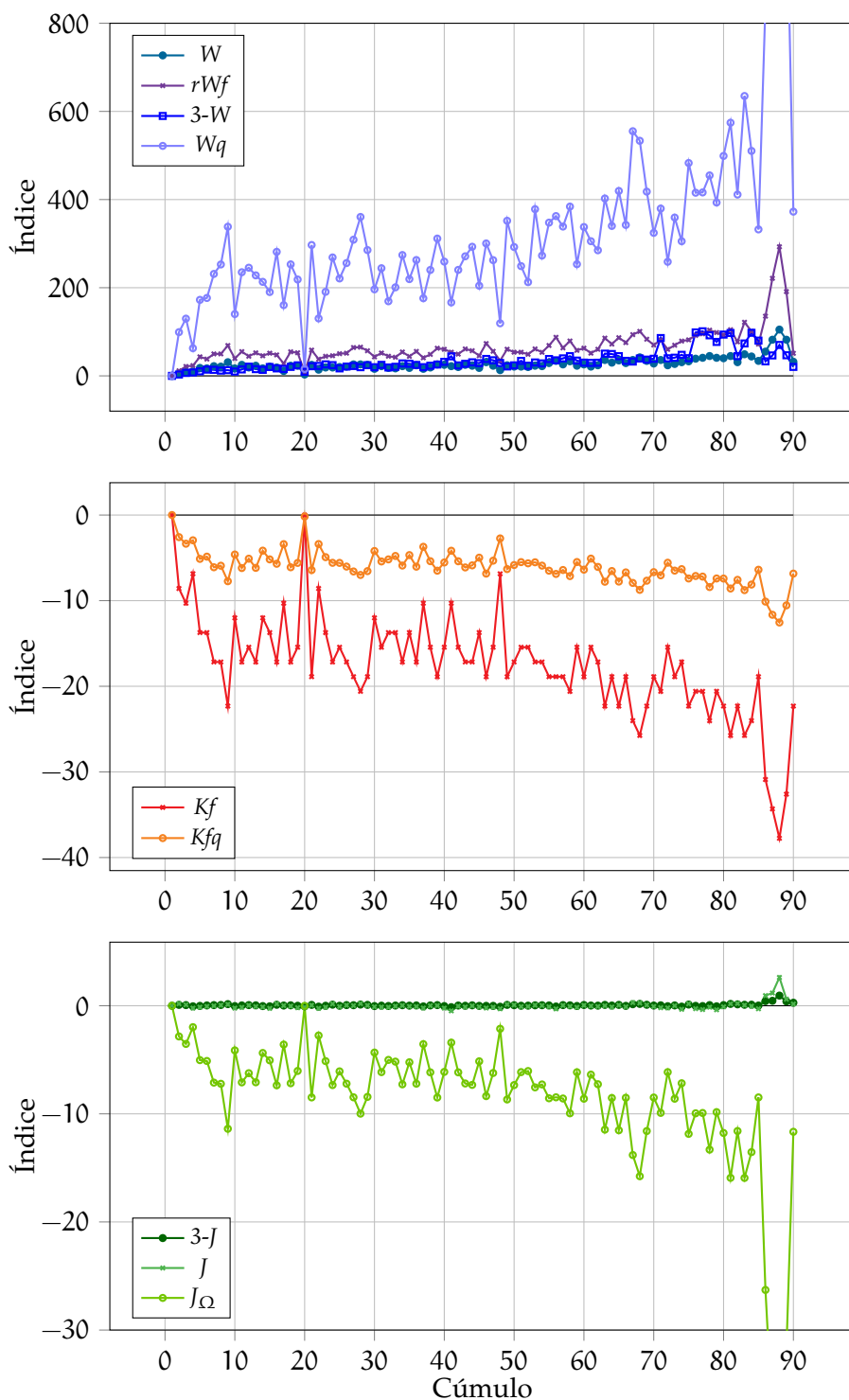


Figura 7: Descriptores topológicos en un conjunto de 30 diferentes cúmulos de Al_{13}^- ordenados por energía.

2.2.3 Algoritmo del Árbol Recubridor Mínimo

En teoría de grafos se le llama *árbol* al grafo G_T que no tiene ciclos.^{137,178} El árbol recubridor mínimo de un grafo conexo y no dirigi-

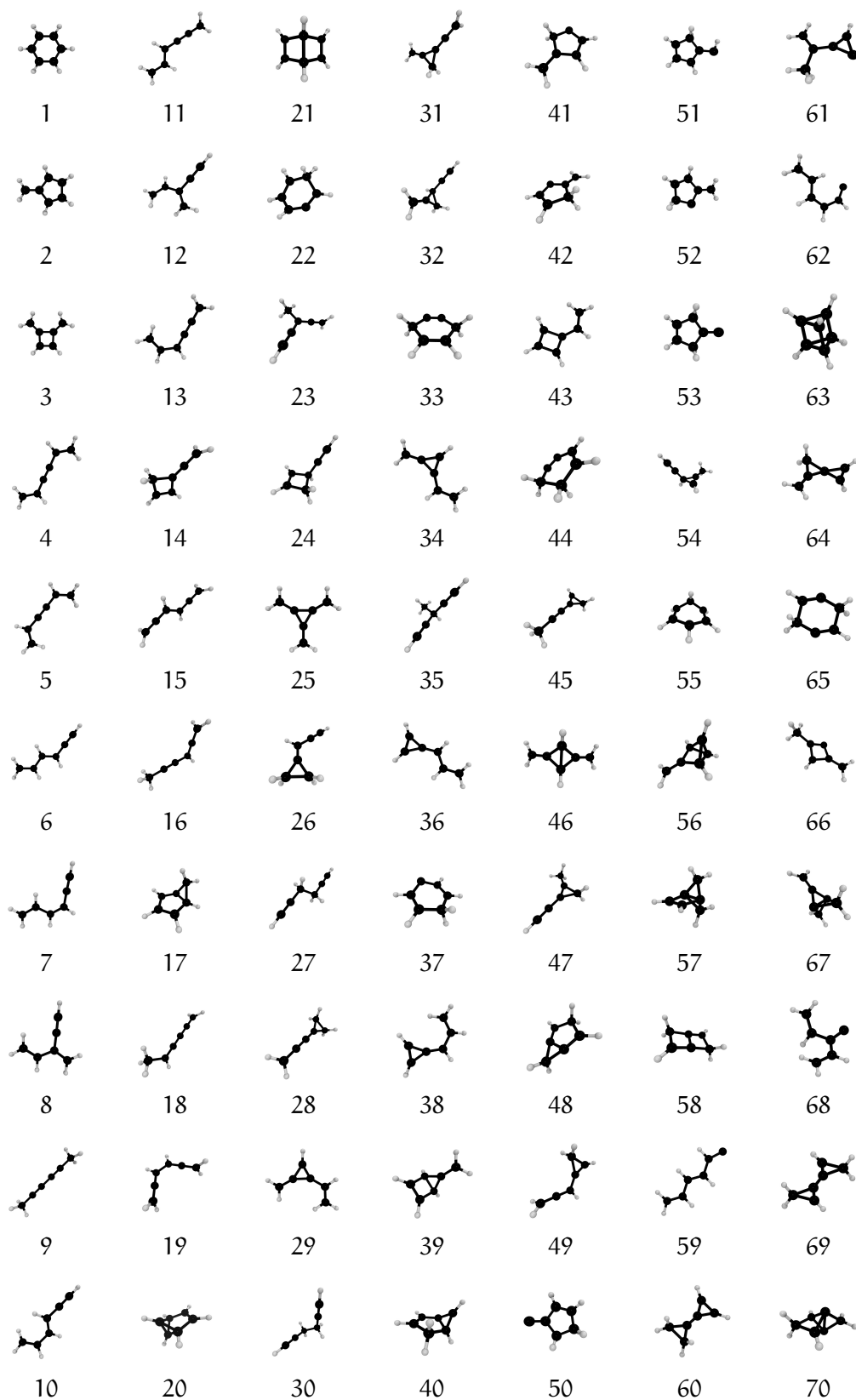


Figura 8: Isómeros del cúmulo C_6H_6 , obtenidos a un nivel PBE0/LANL2DZ.

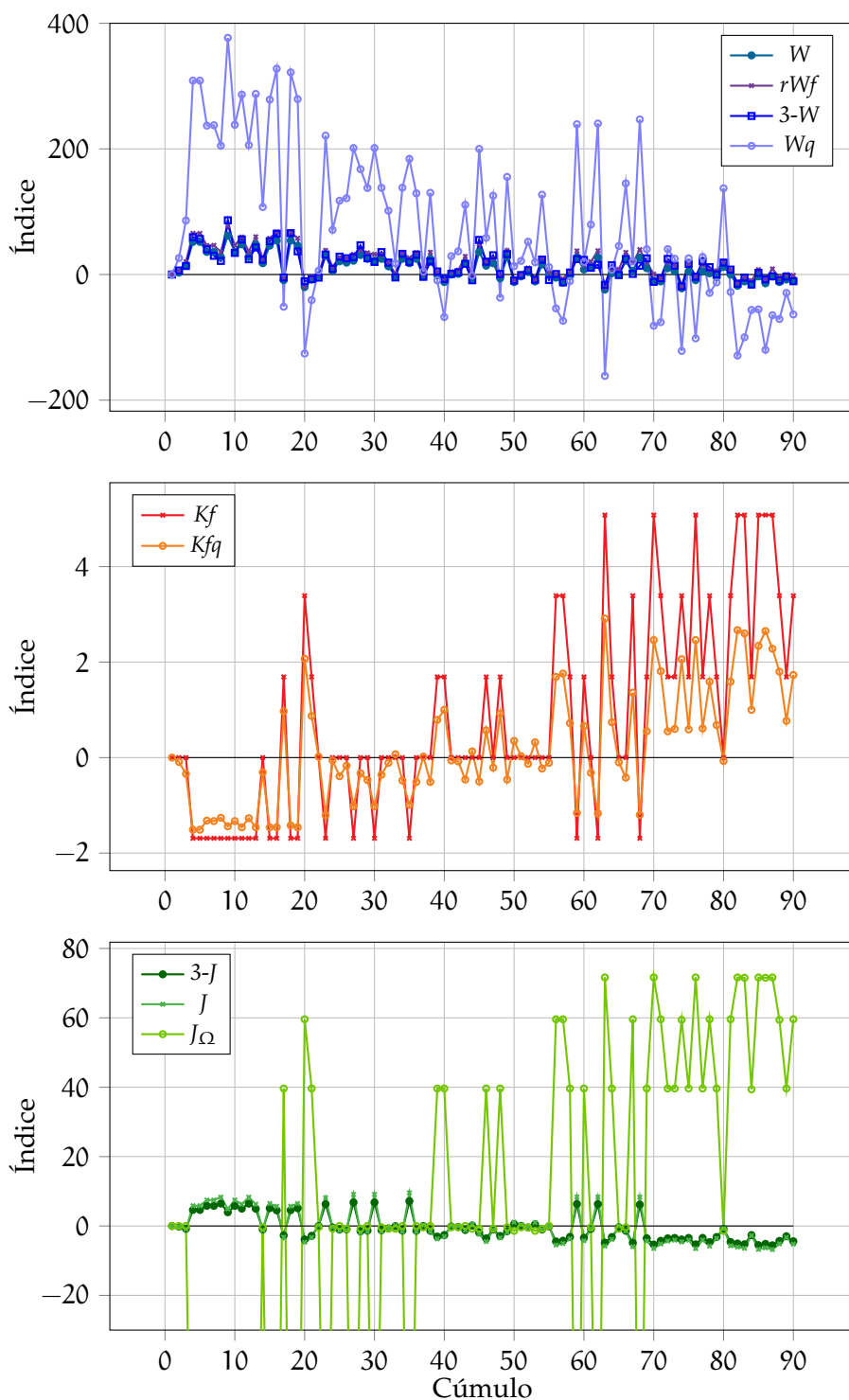


Figura 9: Descriptores topológicos para un conjunto de isómeros del cúmulo del C_6H_6 .

do \mathbf{G} , es el árbol \mathbf{G}_T que contiene todos los vértices de \mathbf{G} , en el que la suma de los pesos de las aristas de \mathbf{G}_T es mínimo. En el caso de que todos los pesos de las aristas sean iguales, puede existir un bosque

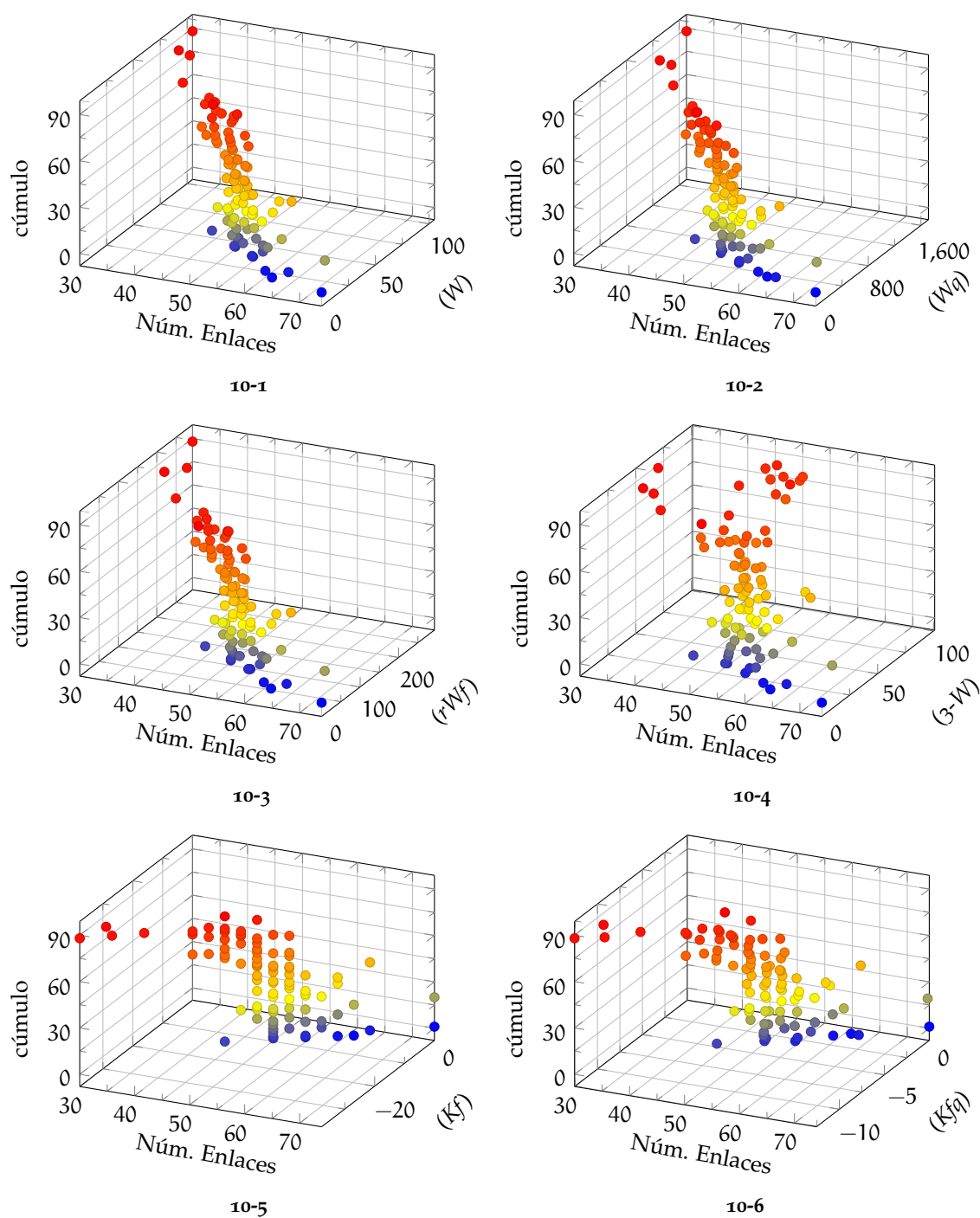
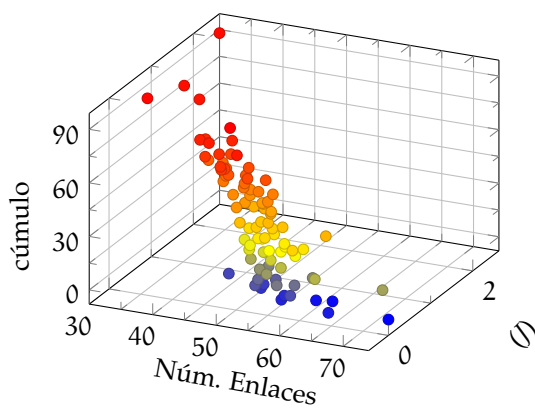
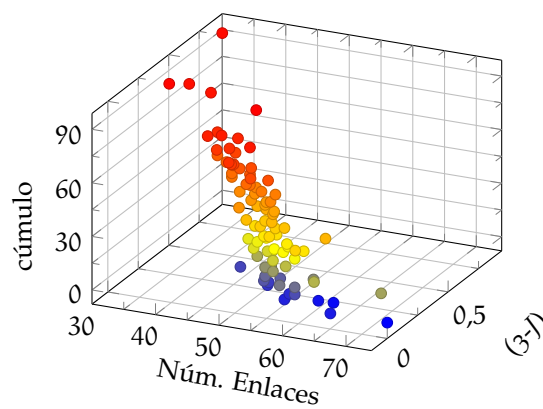


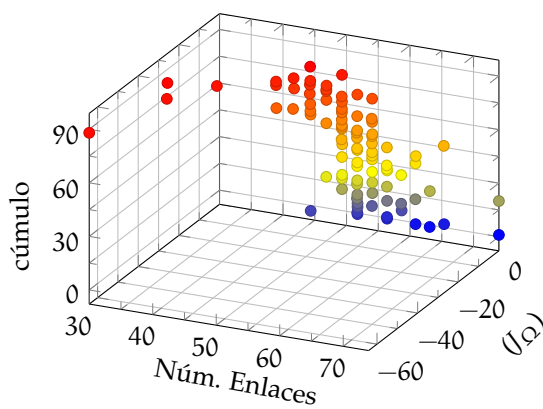
Figura 10: Índices topológicos y su relación con la suma del número de coordinación de un conjunto de 90 isómeros del cúmulo Al_{13}^- . En la versión electrónica de este documento puede saber las coordenadas de cada isómero con hacer click sobre él.



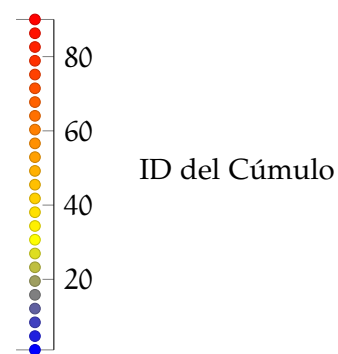
10-7



10-8



10-9



(continuación) Índices topológicos y su relación con la suma del número de coordinación de un conjunto de 90 isómeros del cúmulo Al_{13} . En la versión electrónica de este documento puede saber las coordenadas de cada isómero con hacer click sobre él.

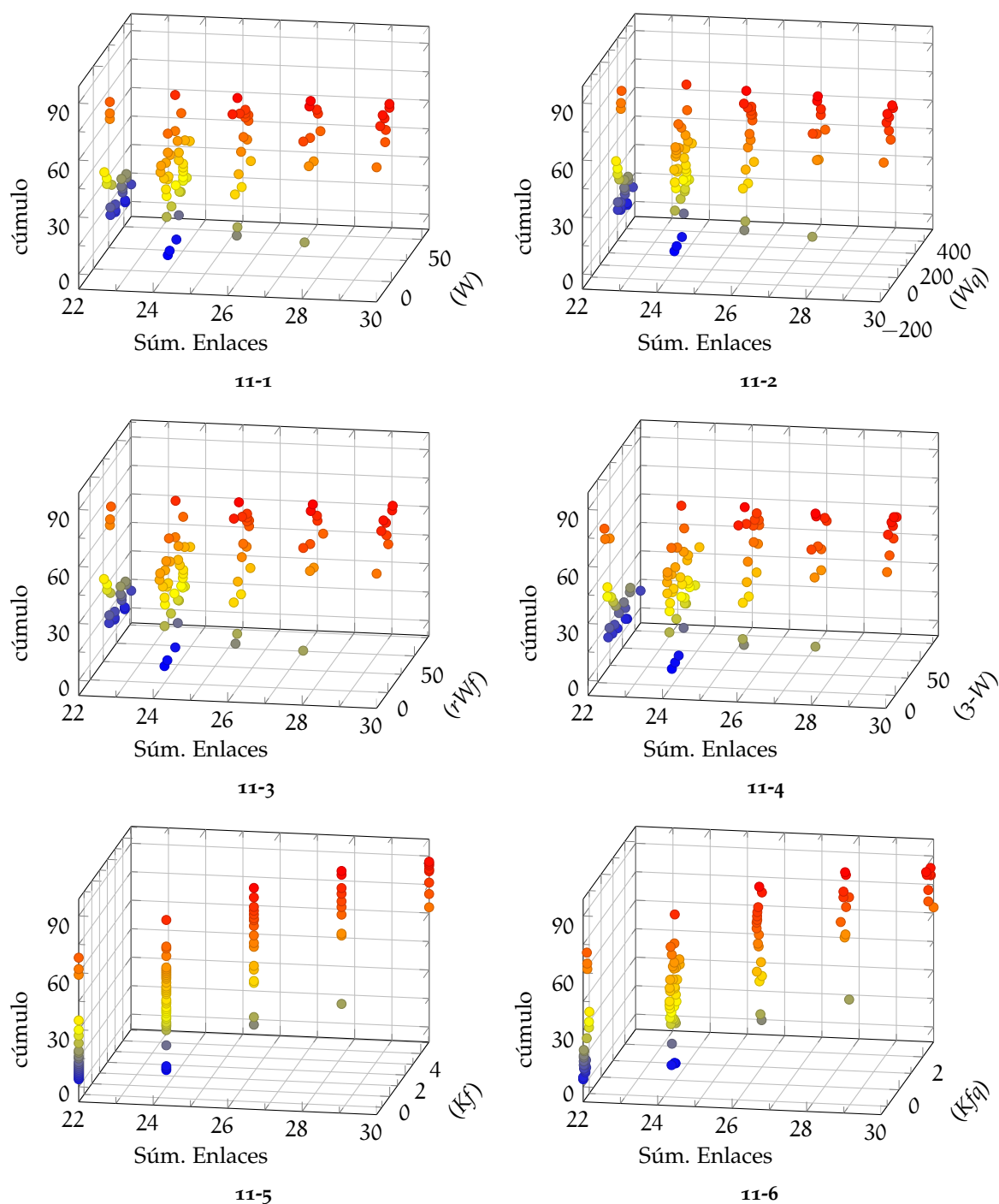
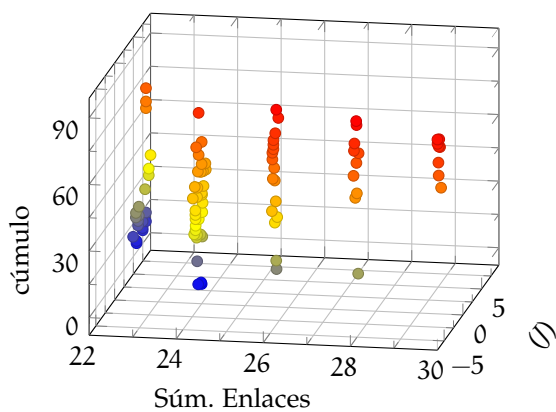
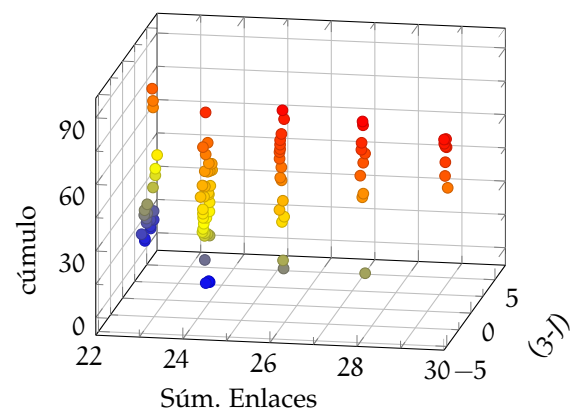


Figura 11: Índices topológicos y su relación con la suma del número de coordinación de un conjunto de 90 isómeros del cúmulo C_6H_6 . En la versión electrónica de este documento puede saber las coordenadas de cada isómero con hacer click sobre él.

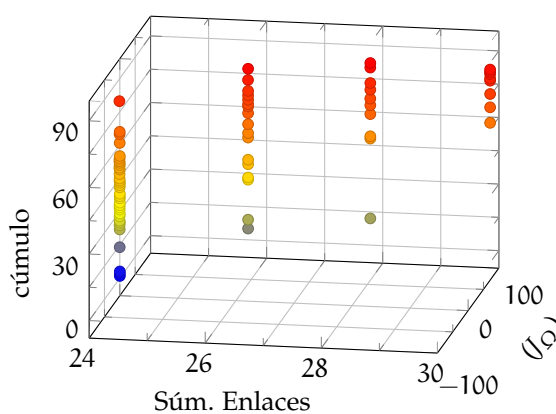
recubridor mínimo. Para obtener el árbol recubridor mínimo (o uno de ellos) de un grafo molecular, en Kaxan se usa el algoritmo *reverse-delete*.¹⁸³ Este algoritmo inicia con el grafo original y va eliminando



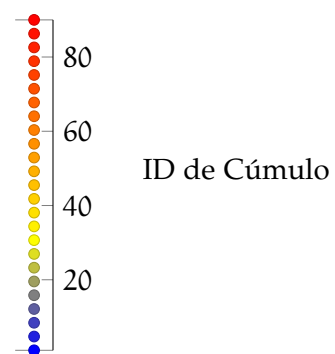
11-7



11-8



11-9



(continuación) Índices topológicos y su relación con la suma del número de coordinación de un conjunto de 90 isómeros del cúmulo C_6H_6 . En la versión electrónica de este documento puede saber las coordenadas de cada isómero con hacer click sobre él.

aristas hasta obtener el árbol de menor costo. El algoritmo trabaja de la siguiente forma:

- Inicia con un grafo G que contiene una lista de aristas E .
- Viaja en E en orden decreciente del peso de las aristas.
- Para cada arista revisa si al eliminarla desconecta el grafo.
- Elimina cualquier otra arista que no genere un grafo no conexo.

En Kaxan se usa este algoritmo para convertir los grafos moleculares en árboles, empleando como pesos de las aristas a las distancias euclidianas de la pareja de átomos enlazados. En la figura 12 se muestra un grafo y el árbol recubridor mínimo del mismo grafo.

Para ver esta animación, use Acrobat Reader® v9.0 o superior.

Figura 12: Árbol recubridor mínimo de un grafo, las aristas en gris corresponden a las eliminadas después de usar el algoritmo *reverse-delete*.

2.3 UN NUEVO MÉTODO DE CRUZA USANDO FRAGMENTOS MOLECULARES

En las últimas secciones se describen dos herramientas implementadas en *fragment-cut-splice*, ahora se describe cómo encajan dentro del algoritmo. Para abordar esta discusión la dividimos en dos partes: selección aleatoria de los fragmentos de los padres y cruce de los fragmentos. La primera explica la forma en como dos padres donan fragmentos (genes) para generar un nuevo individuo, y la última cómo estos fragmentos se ensamblan.

If you think it's simple, then you have misunderstood the problem.

– Bjarne Stroustrup

2.3.1 Selección de los fragmentos

Los fragmentos moleculares seleccionados para formar un nuevo individuo se eligen en cada padre de forma diferente. En primer lugar

se obtiene el árbol (empleando el algoritmo del árbol recubridor) del grafo molecular (conexo) del primer padre, sobre éste se selecciona un enlace (aleatoriamente) con la única condición que no tenga un átomo terminal (átomo con sólo una conexión). De esta manera siempre se tendrán dos fragmentos, tomando sólo uno de ellos para formar parte del nuevo individuo (fragmento **A**). Del siguiente padre se obtiene el grafo complemento procurando que el segundo conjunto de átomos sea conexo (fragmento **B**), en el caso que exista más de un grafo complemento se usará uno al azar.

2.3.2 Mezcla de fragmentos

Una vez obtenidos **A** y **B** se construye la matriz de conexiones perdidas, $(\mathbf{CP})_{ij}$ tomando el bloque de la matriz de adyacencia que corresponde a las i filas de los átomos del fragmento **A**, con las j columnas del fragmento complemento. De esta manera se almacena la información de la conectividad entre cada fragmento que formaban parte de cada uno de los cúmulos padres.

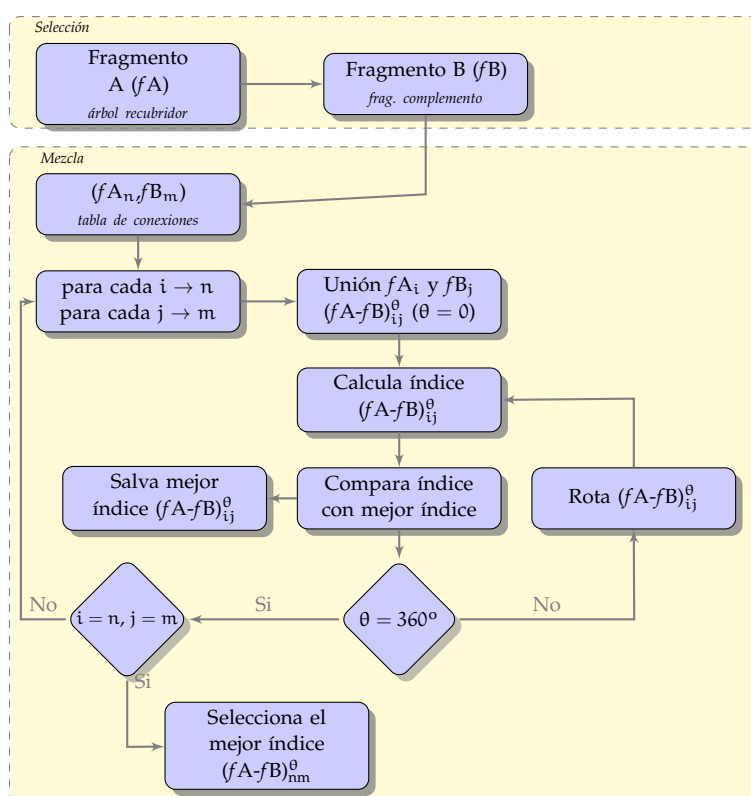


Figura 13: Algoritmo de cruce *fragment-cut-splice* implementado en Kaxan.

Para generar una unión entre **A** y **B** se toma el elemento $(\mathbf{CP})_{ij}^A$ y se crea una unión entre el átomo que corresponde al elemento $(\mathbf{CP})_{kl}^B$, donde el átomo k de **B** ocupa la posición del átomo faltante

j que se encontraba unido a i en **A**. Así se genera un nuevo enlace entre el átomo i y el átomo k. La mejor conformación de la unión de los dos fragmentos se obtiene haciendo una rotación completa de **B** sobre el enlace i-k, calculando el valor del índice topológico de cada conformación, seleccionando la conformación que tenga la menor diferencia entre el índice calculado y el índice del mejor cúmulo de la generación anterior.

Iterando sobre todas las combinaciones de las dos matrices de conexiones perdidas para **A** y **B** se obtiene el nuevo individuo para la nueva generación con el índice topológico. Durante el proceso de unir los fragmentos, lo mismo que al rotar, se lleva a cabo la revisión de las geometrías para evitar cúmulos colapsados (cúmulos con distancias entre átomos menor al radio covalente) y otra revisión para evitar cúmulos desconexos. En la figura 13 se muestra un diagrama del algoritmo.

2.4 RESULTADOS CON CÚMULOS ATÓMICOS

Para la evaluación de este nuevo método de cruza se utilizaron cinco sistemas moleculares con diferencias estructurales que los hacen interesantes para estudiar el comportamiento del AG. Los sistemas seleccionados en este estudio son: un agregado molecular ((H₂O)₆), un cúmulo con estructura globular (Al₁₃⁻), otro con mínimo global plano (B₁₅⁻) y, finalmente, dos moléculas orgánicas con forma de anillo; benceno y naftaleno).

En este estudio se evaluaron bajo las mismas condiciones *plane-cut-splice*, *plane-rot-splice*, *sphere-cut-splice* y *fragment-cut-splice*. Las búsquedas de mínimo global se llevaron a cabo con 10 generaciones por cúmulo para tres esquemas de población (1 × n, 2 × n y 3 × n, con n como el número de átomos). Cada esquema se llevó a cabo por triplicado para confirmar los resultados. En todos los cúmulos se usó el funcional PBE0¹⁸⁴ con la base LANL2DZ¹⁸⁵ empleando el programa Firefly.¹⁸⁶ La tabla 1 resume los resultados de las búsquedas de mínimo global realizadas con Kaxan para los cinco sistemas moleculares, además, se muestran las diferencias de usar un número distinto de población (columna Indiv./Gen.), las veces que aparece el mínimo global (columna Núm. Mín./Rep.), la generación en la que aparece el mínimo global (columna Mín. Gen.) y finalmente el número total de mínimos encontrados después de la búsqueda (columna Total Mín.).

2.4.1 (H₂O)₆

Los cúmulos de moléculas con enlaces de hidrógeno más estudiados son los de agua dada su vital importancia.¹⁸⁷⁻¹⁹¹ Usando los cuatro métodos de cruza se encontró el mínimo global (figura 14-1) sobre la PES, además de otros mínimos importantes. Sin embargo, *plane-rot-*

Tabla 1: Resultados con diferentes algoritmos de cruce para un algoritmo genético

Cúmulo	Método de Cruza	Indiv./Gen.	Núm. Mín./Rep. Mín.	Gen. Total	Mín.	
$(\text{H}_2\text{O})_6$	<i>plane-cut-splice</i>	18	2/3	-/8/4	139	
		36	1/3	-/-/9	233	
		54	2/3	6/5/-	323	
	<i>plane-rot-splice</i>	18	1/3	-/-/7	143	
		36	0/3	-/-/-	236	
		54	2/3	8/8/-	323	
	<i>sphere-cut-splice</i>	18	1/3	-/-/7	143	
		36	1/3	-/5/-	764	
		54	2/3	5/3/-	358	
	<i>fragment-cut-splice</i>	18	2/3	6/4/-	185	
		36	2/3	4/3/-	375	
		54	2/3	-/2/3	566	
	Al_{13}^-	<i>plane-cut-splice</i>	13	3/3	1/1/1	46
			26	3/3	0/0/1	71
			39	3/3	0/0/0	91
<i>plane-rot-splice</i>		13	3/3	1/1/1	48	
		26	3/3	0/0/0	65	
		39	3/3	0/1/0	96	
<i>sphere-cut-splice</i>		13	3/3	0/1/0	50	
		26	3/3	0/0/0	77	
		39	3/3	0/0/0	108	
<i>fragment-cut-splice</i>		13	3/3	1/1/1	117	
		26	3/3	1/1/0	242	
		39	3/3	0/1/1	340	
B_{15}^-		<i>plane-cut-splice</i>	15	3/3	3/5/7	117
			30	3/3	2/1/1	210
			45	3/3	3/1/2	114
	<i>plane-rot-splice</i>	15	3/3	4/1/3	110	
		30	3/3	2/4/2	205	
		45	3/3	2/4/4	271	
	<i>sphere-cut-splice</i>	15	3/3	-/1/1	128	
		30	3/3	1/7/6	238	
		45	3/3	1/1/2	271	
	<i>fragment-cut-splice</i>	15	3/3	5/1/1	130	
		30	3/3	1/1/1	239	
		45	3/3	1/1/1	334	
	C_{10}H_8	<i>plane-cut-splice</i>	18	0/3	-/-/-	172
			36	0/3	-/-/-	318
			54	0/3	-/-/-	459
<i>plane-rot-splice</i>		18	0/3	-/-/-	318	
		36	0/3	-/-/-	176	
		54	0/3	-/-/-	319	
<i>sphere-cut-splice</i>		18	0/3	-/-/-	181	
		36	0/3	-/-/-	327	
		54	1/3	-/5/-	468	
<i>fragment-cut-splice</i>		18	1/3	-/4/-	171	
		36	0/3	-/-/-	354	
		54	0/3	-/-/-	547	
C_6H_6		<i>plane-cut-splice</i>	12	0/3	-/-/-	122
			24	1/3	-/-/2	229
			36	0/3	-/-/-	339
	<i>plane-rot-splice</i>	12	0/3	-/-/-	116	
		24	1/3	2/-/-	234	
		36	0/3	-/-/-	337	
	<i>sphere-cut-splice</i>	12	0/3	-/-/-	118	
		24	0/3	-/-/-	237	
		36	2/3	2/-/3	328	
	<i>fragment-cut-splice</i>	12	1/3	-/6/-	109	
		24	0/3	-/-/-	213	
		36	0/3	8/-/-	313	

splice y *sphere-cut-splice* necesitaron 36 y 54 individuos y usar las 10 generaciones para llegar a **14-1**. Con *plane-cut-splice* es posible encontrar **14-1** con más frecuencia (columna Núm. Mín./Rep. tabla 1) que con los otros dos métodos debido a que es más fácil combinar los genes de los padres que con *sphere-cut-splice* o *plane-rot-splice*, ya que **14-1** presenta varios planos de simetría siendo con esto una estructura ideal para este método. Esto favorece el hecho de que al combinar

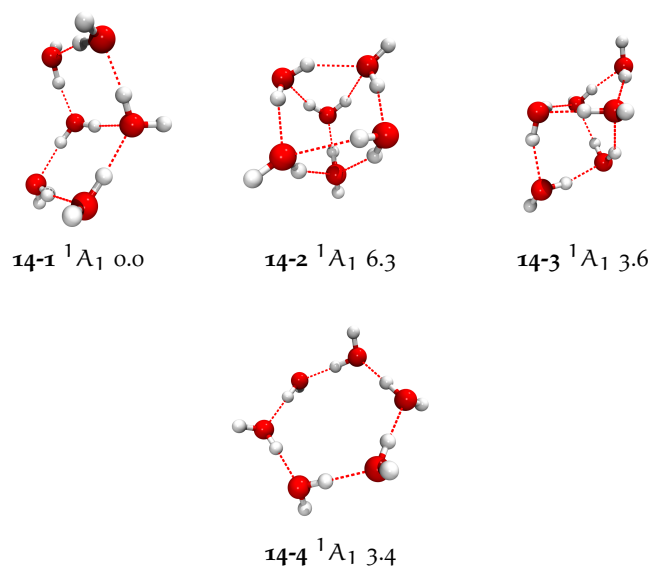


Figura 14: Mínimos de $(H_2O)_6$ encontrados con nuestro AG, energías relativas (PBE0/LANL2DZ) en kcal/mol.

genes de los padres haciendo cortes sobre los planos lleve a geometrías prometedoras en convertirse en mínimos sobre PES. También esto se ve reflejado en que aparezca **14-1** cada vez en menos generaciones. Con *fragment-cut-splice* se obtiene **14-1** con mayor frecuencia (2/3) y llegando sólo a necesitar 2 generaciones para encontrar el mínimo global. Aunque sólo en una ocasión no se encontró **14-1**, el algoritmo consiguió tocar puntos importantes sobre la PES, como los que aparecen en la figura 14. En todos los casos se encuentra el mínimo tipo caja (**14-3**) y con frecuencia el isómero tipo prisma (**14-2**). Aunque el resultado entre *plane-cut-splice* y *fragment-cut-splice* es muy similar en términos de frecuencia, ($\sim 2/3$), *plane-cut-splice* no es capaz de localizar **14-1** en las generaciones iniciales como lo hace con *fragment-cut-splice*. También es de destacar que el número final de mínimos es mayor con *fragment-cut-splice* en comparación a los otros tres algoritmos incluso llegando a tener hasta 100 mínimos de diferencia.

2.4.2 B_{15}^- y Al_{13}^- .

Estos dos sistemas fueron útiles para explorar el comportamiento de la búsqueda con sistemas de tipo planar (B_{15}^-) y geometrías globulares (Al_{13}^-) en sistemas homonucleares. En la tabla 1 se resumen los resultados para la búsqueda en estos dos tipos de cúmulos. Para el Al_{13}^- , la frecuencia con la que se encontró el mínimo global (figura 15-1) es perfecta en todos los casos (3/3). Además, el mínimo global aparece en las dos primeras generaciones (0 y 1) debido a que de la generación aleatoria de individuos se obtienen buenos candidatos para convertirse en mínimo (o en el mejor de los casos: el mínimo global)

El mínimo global para Al_{13}^- a un nivel de teoría ab-initio es el que aparece en la Figura 6 con el ID=2.

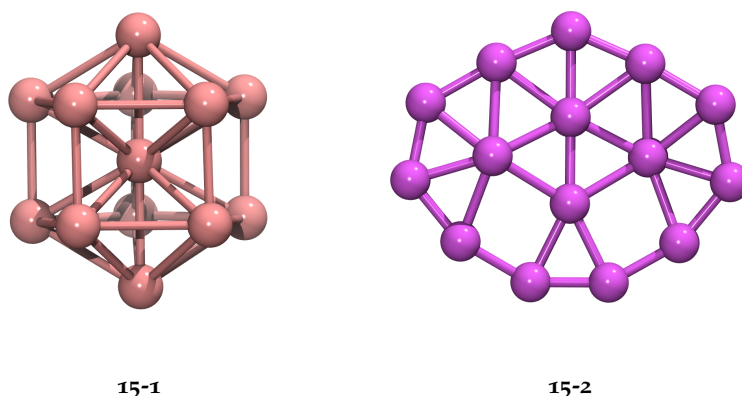


Figura 15: Cúmulos de Al_{13}^- y B_{15}^- encontrados durante la búsqueda estocástica con Kaxan, a nivel PBE0/LANL2DZ, las energías relativas en kcal/mol.

mostrando la eficiencia de estas rutinas, ya que necesitan de unos pocos individuos para encontrar a **15-1**. Este hecho es más notorio conforme aumenta el número de individuos por generación. Como era de esperarse, la probabilidad de generar **15-1** en la generación aleatoria crece directamente proporcional al tamaño de la población. Este aspecto es notorio en los cuatro casos, en donde al doblar o triplicar el número de individuos, la frecuencia de aparición del mínimo global en la generación 0 creció (columna Min. Gen. de la tabla 1). Durante las generaciones siguientes, **15-1** se mezclaron con más individuos, creando descendientes con semejanza a **15-1** afectando la diversidad estructural, hecho que se repitió conforme se crean las generaciones resultando en un número final pobre de mínimos explorados. Sin embargo, en *fragment-cut-splice* el número de mínimos encontrados es el doble (en el peor de los casos) comparado con los otros tres métodos, destacando el poder exploratorio que tiene *fragment-cut-splice*.

En el cúmulo B_{15}^- , de la misma manera que en Al_{13}^- , se obtuvo una frecuencia perfecta para encontrar el mínimo global (3/3), en la mayoría de los casos, la geometría de este cúmulo (**15-2**) se encontró en un número de generaciones mayor, en comparación a Al_{13}^- . Sin embargo, hay un efecto similar al encontrado en Al_{13}^- donde la posibilidad de generar buenos candidatos, en la generación 0 crece de manera directa con el tamaño de la población. Esto se consiguió cuando el número de individuos cambió de 15 a 30, la generación en que aparece el mínimo global, es la 1.

2.4.3 Benceno y Naftaleno.

En estas dos moléculas se tuvo un resultado pobre en la búsqueda de mínimo global, tanto para la molécula de benceno (C_6H_6) como para

Para corroborar que la conectividad del enlace C–H la causa por la que las estructuras cíclicas de seis miembros no son recurrentes durante las generaciones en la búsqueda de C_6H_6 y $C_{10}H_8$, se llevaron a cabo pruebas con 3 repeticiones para cada cúmulo empleando el algoritmo de cruza *fragment-cut-splice* y el número de generaciones mínimo estudiado para cada molécula (12 para benceno y 18 para naftaleno). En este caso, para cada cúmulo se eliminaron los átomos de hidrógeno de la búsqueda y únicamente se adicionan cuando el cúmulo va a ser optimizado localmente. Para ensamblar estos átomos de H fueron empleadas rutinas que permiten adicionar átomos de H a estructuras de carbono, estos algoritmos que agregan hidrógenos con distancias, ángulos de enlace y cantidad apropiada según la hibridación del carbono (para los dos casos la cantidad de hidrógenos fue fijada a 6 y 10 átomos respectivamente). Al finalizar el proceso de optimización y obtener la geometría, Kaxan elimina de nuevo los hidrógenos y continua la búsqueda únicamente con los átomos de carbono. En los dos casos se consiguió una perfecta frecuencia (3/3) para encontrar el mínimo global, apareciendo en generaciones muy tempranas (en la generación 0 en los tres casos para benceno y sólo en una oportunidad en la generación 0 en naftaleno, las dos restantes en la generación 1). Aunque esto no es una prueba real de que este el AG tiene la capacidad de encontrar la geometría de mínimo global para estas dos moléculas, deja claro que es posible hacer una búsqueda de moléculas orgánicas de una manera eficiente en Kaxan, además que el principal obstáculo de llegar al mínimo global se debe a la estricta conectividad imperante en moléculas con enlaces C–H.

Dos parámetros importantes que están fuera del ejercicio de búsqueda son el porcentaje de mutación de cada generación y el porcentaje de elitismo (cantidad de individuos que pasan de una generación a la siguiente sin cambio en su estructura). En los resultados de la tabla 1 se usó para todos los casos 60% en la mutación y 30% en el elitismo, esperando que fuera un buen balance entre dos aspectos importantes: la explotación y la exploración.¹

Como se muestra en los tres primeros cúmulos ($(H_2O)_6$, Al_{13}^- y B_{15}^-), el ajuste de mutación/elitismo es suficiente para encontrar rápidamente el mínimo global, pero no para C_6H_6 y $C_{10}H_8$, debido a que un porcentaje tan alto de mutación ofrece una buena exploración a costa de la explotación, hecho que se intenta amortiguar con un factor alto de elitismo. Sin embargo, la cantidad de mínimos con energías no tan favorables se ve afectada y se refleja en despreciar geometrías cíclicas (con presencia de metilenos) como candidatos a formar nuevos individuos. Para investigar este efecto, se estudiaron los efectos de variaciones del porcentaje de mutación (10 y 60) y del porcentaje de elitismo (0 y 30) con diferente número de individuos durante 10 generaciones con *fragment-cut-splice* para C_6H_6 . En la tabla 2 aparecen estos resultados, donde se encontró que existe un efecto dramático entre el

¹Estos dos términos definen el balance óptimo que una estrategia de búsqueda debe tener, la exploración da cuenta de que tan diversos son los puntos que el algoritmo alcanza durante la búsqueda, mientras que la explotación se refiere a cuanto se explora la vecindad de las soluciones óptimas.^{109,111}

porcentaje de mutación y la aparición del mínimo global, ya que en todos los casos que se usó 60 % de mutación no fue posible llegar al mínimo global. En la variación del elitismo no se descubrió un efecto tan marcado, y sólo se tuvo una ligera disminución en la generación en que aparece el mínimo global cuando el número de individuos es 24 y 36, pero con tan pocos datos es prematuro afirmar que la ausencia de individuos que se promocionen entre las generaciones sea un factor que determine la velocidad de aparición del mínimo global. De esta manera, podemos confirmar que es necesario lograr un buen compromiso entre explotación y exploración y que para lograr este equilibrio es necesario seleccionar un número adecuado de individuos por generación, además de un porcentaje de mutación adecuados, para asegurar una búsqueda exitosa con este AG.

Tabla 2: Número de generación en la que aparece el mínimo global para C_6H_6 con diferentes porcentajes de mutación y elitismo.

Población	%Elitismo/ %Muta			
	0/10	30/10	0/60	30/60
12	0	6	-	-
24	4	7	-	-
36	5	6	-	-

2.5 DIVERSIDAD DEL NUEVO ALGORITMO

En los resultados de la tabla 1 se encontró que en los cinco sistemas estudiados al emplear *fragment-cut-splice* se obtuvo un aumento en el número total de mínimos obtenidos, a pesar de que en la mayoría de los casos este número superó el doble o el triple de geometrías exploradas por los otros tres métodos de cruce. Sin embargo, este número no dice nada sobre las diferencias entre las geometrías encontradas por cada metodología de cruce. Cabe preguntarse entonces ¿qué tan diferentes son entre sí estos mínimos?, ¿es más diversa la población empleando este método de cruce?. Esta última pregunta, en particular, es quizá la más importante si se hace una lectura detallada del proceso de la cruce, como la realiza *fragment-cut-splice*, debido a que sólo permite crear nuevos individuos con características dadas por el índice topológico, por tanto, estructuralmente cercanas al mejor individuo de la generación anterior. Esta característica puede ocasionar que durante la ejecución del AG se observe una disminución en la diversidad de los mínimos o en el número de mínimos encontrados, en el peor de los casos, que se llegue a la convergencia prematura, causando que no sea posible llegar al mínimo global.

Para evaluar la diversidad¹⁹² de un conjunto de moléculas o cúmulos moleculares, se tienen varios métodos; algunos basados en los momentos estadísticos,^{132,193,194} o con representaciones moleculares basadas en esferas duras,^{195,196} gaussianas,¹⁹⁷ o también con armónicos esféricos.¹⁹⁸⁻²⁰¹ En Kaxan se utiliza el método de Ballester para diferenciar estructuras idénticas (en el Apéndice A aparece una descripción de este método) que además de prescindir del difícil proceso de superposición molecular, es fácil de calcular lo que hace que sea rápido, además este método se emplea para comparar la similitud entre dos estructuras que podría usarse como medida de la diversidad de una población de moléculas. Para evaluar la diversidad usaremos un algoritmo de similitud propuesto por Call y Boldyrev *et al.*²⁰² basado en la Desviación Cuadrática Media (DCM) de las diferencias de las distancias entre los átomos, por que este método permite diferenciar de forma directa las diferencias en las posiciones atómicas dos formas moleculares, en el método de Ballester sólo se busca diferenciar la forma molecular y no las posiciones relativas de los átomos. Para tratar de dar respuesta a las interrogantes arriba propuestas, se abordó el estudio de la diversidad desde dos puntos de vista diferentes, tal que permitan dar una visión clara del panorama. En primer lugar se exploró la diversidad de todos los individuos encontrados durante las 10 generaciones y luego se comparó la evolución de la diversidad estructural por población durante el total de la búsqueda

Para evaluar la similitud de cada individuo se empleó la DCM de las diferencias de las distancias entre los átomos de cada individuo. Este método tiene la ventaja de ser una métrica basada en las distancias interatómicas y no depende de la rotación de los cúmulos. Esta métrica requiere el ordenamiento de los átomos, según el número atómico y la distancia al centro de masas de la molécula, en ese orden. Este procedimiento calcula la DCM de la diferencia de las distancias entre pares de átomos de acuerdo a la siguiente ecuación:

$$DCM(A, B) = \sqrt{\frac{1}{\left(\frac{n^2-n}{2}\right)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n (d(a_i, a_j) - d(b_i, b_j))^2}, \quad (14)$$

donde n es el número de átomos, $d(a_i, a_j)$ es la distancia euclidiana, a_i y a_j es el i -ésimo y j -ésimo átomo en el cúmulo A, respectivamente, y b_i , b_j son el i -ésimo y j -ésimo átomo en el cúmulo B. El valor $(n^2 - n)/2$ es el número de pares únicos de átomos, que es el mismo en los dos cúmulos.

Este método es similar al utilizado para comparar dos proteínas²⁰³ que también emplea una matriz de distancias donde cada elemento de la matriz es la distancia entre dos aminoácidos, uno de cada proteína.

Measure what is measurable, and make measurable what is not so.
– Galileo Galilei

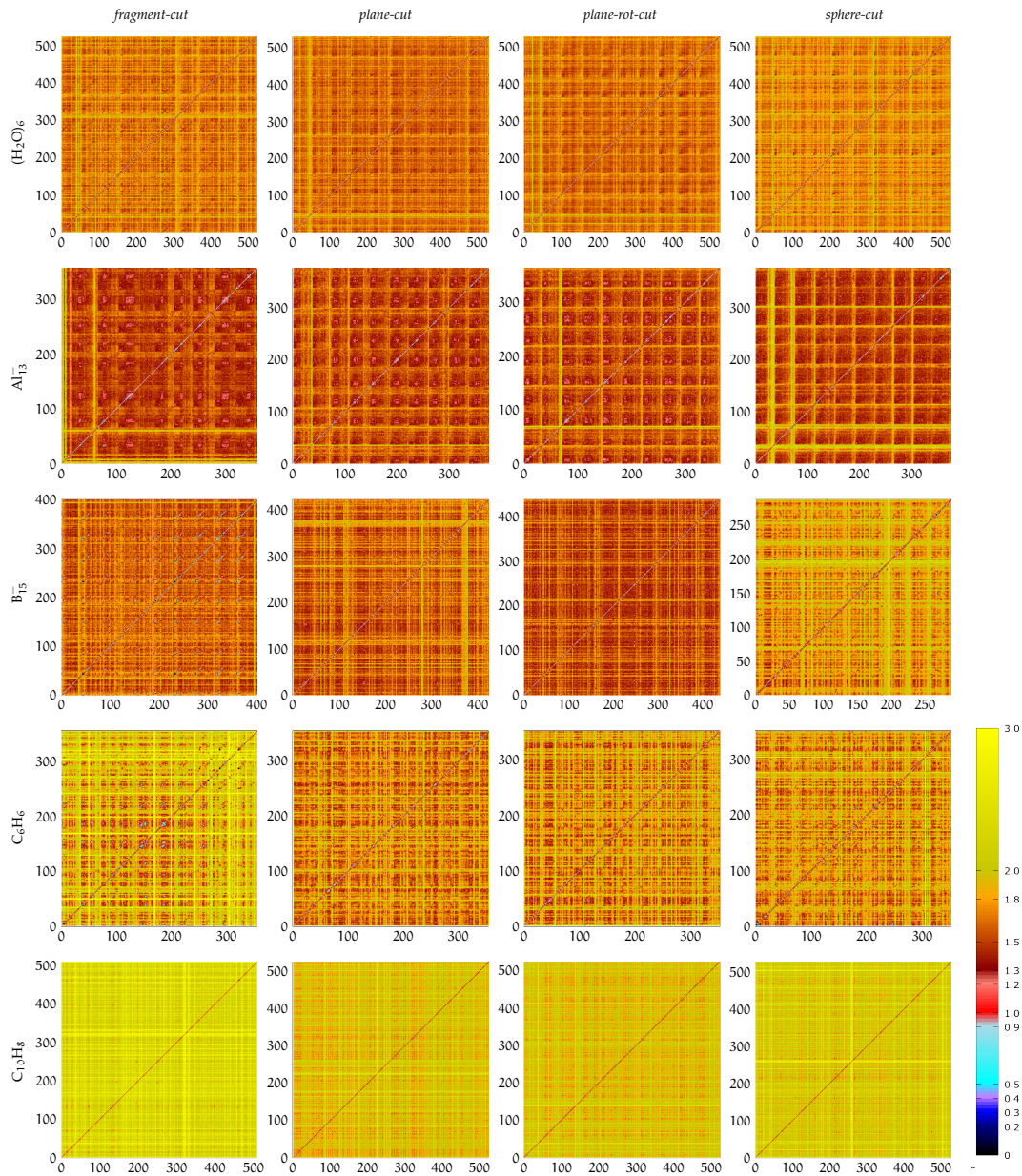


Figura 17: Distancia cuadrática media de la diferencia de las distancias entre mínimos, en cada algoritmo de cruce, para todos los mínimos de los resultados de la búsqueda de mínimo global usando Kaxan.

Luego realiza la traslación y rotación de una proteína para minimizar la DCM con la otra proteína de acuerdo a la ecuación:

$$DCM_{\text{proteína}}(A, B) = \sqrt{\frac{1}{n} \sum_{i=1}^n d(a_i, b_i)^2}, \quad (15)$$

donde n es el número de aminoácidos, a_i es el átomo de carbono primario del i -ésimo aminoácido en la proteína A y b_i es el carbono primario del i -ésimo aminoácido en la proteína B. A continuación,

reporta esta la distancia mínima, *DCM*, como la medida de similitud entre las proteínas. La capacidad de realizar esta tarea se basa en el hecho de que los aminoácidos se ordenan y se identifican de forma única por un número. En el caso de los cúmulos, los átomos no tienen jerarquía y múltiples átomos con el mismo número atómico están permitidos. La métrica de la distancia que aquí se presenta se encarga de este ordenamiento usando el número atómico y la distancia al centro de masas.

Los resultados de la [Ecuación 14](#) en los cinco sistemas moleculares estudiados, utilizando los cuatro métodos de cruce para el *AG*, se resumen en la [figura 17](#). Se emplearon los resultados de una búsqueda de mínimo global con $3 \times$ número de átomos, durante 10 generaciones, con 60% de mutación. En [17](#) se muestra la similitud entre todos los mínimos explorados ordenados por la generación en la que aparecieron (entre generaciones están ordenados por energía). La similitud calculada con la [Ecuación 14](#) para un par de estructuras varían de 0 a 3, en donde 0 indica que los dos cúmulos son iguales. Aunque la cantidad de mínimos fue diferente en cada caso, es posible diferenciar cualitativamente los resultados durante el proceso de búsqueda. Los sistemas homonucleares Al_{13}^- y B_{15}^- pueden tener un número importante de isómeros, que debe ser inferior al de los cúmulos heteronucleares, donde el número de estructuras posibles debe aumentar de acuerdo al número del tipo de átomos. En el caso de $(H_2O)_6$, hay una menor cantidad de isómeros, comparado con $C_{10}H_8$, que tienen igual número de átomos, debido a que los cúmulos de $(H_2O)_6$ poseen la tendencia de formar agregados moleculares con patrones de H_2O , contrario a $C_{10}H_8$ que no tiene un patrón limitado de enlace.

Estas características estructurales son las que determinan el orden en la diversidad de cada conjunto de cúmulos estudiados: $C_{10}H_8 > C_6H_6 > (H_2O)_6 > Al_{13}^- > B_{15}^-$. Este ordenamiento se presentó debido a características estructurales de los cúmulos escogidos y no es causada por el *AG* o alguna de sus partes. Sin embargo se pueden encontrar, por lo menos cualitativamente, diferencias entre los métodos de cruce.

En $(H_2O)_6$, a grandes rasgos, entre los mínimos encontrados con *fragment-cut-splice*, *plane-cut-splice* y *plane-rot-splice* se mantienen en valores entre 1 y 3. En *sphere-cut-splice*, hay un mayor número de individuos con valores que van de 1,5 a 3 entre las generaciones, en general más diverso que las otras tres metodologías de cruce. En el caso de Al_{13}^- , las cuatro metodologías se mantienen en valores de más baja similitud que $(H_2O)_6$ ($\sim 1 - 1,5$). En el conjunto de estructuras ensambladas con *sphere-cut-splice* hay un mayor número de moléculas con valores cercanos a 3 sobretodo en las dos primeras generaciones, aunque también las otras tres metodologías tienen valores cercanos a 3 pero en mayor medida sólo en la segunda generación para *fragment-cut-splice* y *plane-rot-cut*, en *plane-cut* el conjunto molecular es más uniforme en todas las generaciones. Para el caso de B_{15}^- hay un com-

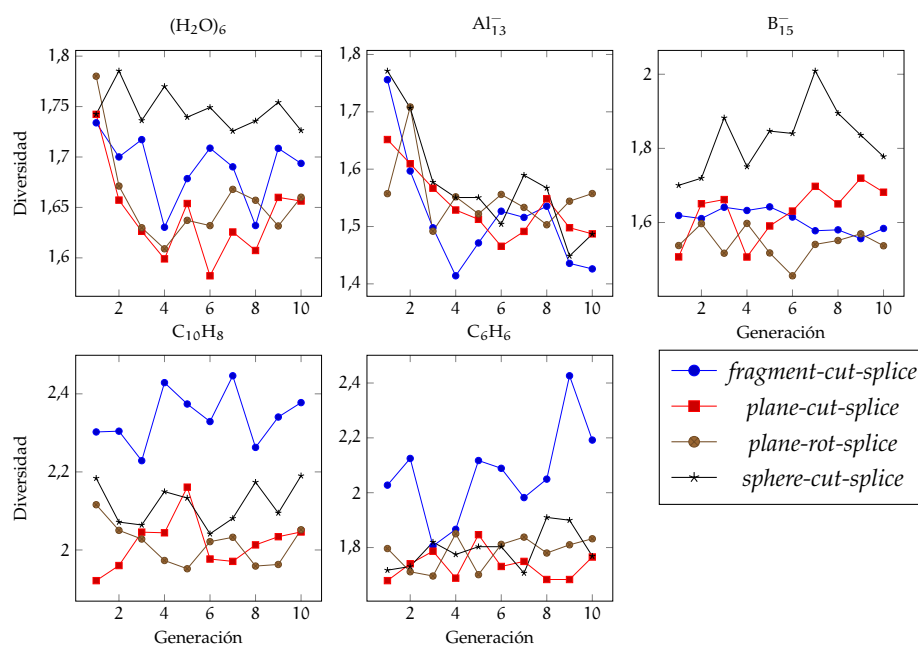


Figura 18: Evolución de la diversidad durante la búsqueda de mínimo global para los diferentes métodos de cruce. Usando 10 generaciones con $3 \times$ número de átomos como población.

portamiento parecido al de Al_{13}^- . Con *sphere-cut-splice* y *plane-cut* se obtuvo la mayor diversidad, en *fragment-cut-splice* se llegaron a tener valores cercanos a 0 ($\sim 0 - 0,5$) para estructuras similares al mínimo global, debido a que este último se halló entre las primeras generaciones. Este efecto se atenúa en los otros tres algoritmos, teniendo de nuevo a *sphere-cut-splice* como el conjunto de mínimos con mayor disimilitud. En el caso de C_6H_6 y $C_{10}H_8$ se tuvo un aumento en la diversidad en *fragment-cut-splice*, pero, en general, los resultados de los cuatro algoritmos presentan un cambio considerable entre las geometrías exploradas durante la ejecución de Kaxan, comparativamente, los otros algoritmos frente a *fragment-cut-splice* tienen un comportamiento similar en la variación estructural de los mínimos.

Si bien la diversidad de una población se conecta con el concepto de distancia entre los individuos,^{204,205} la diversidad es una medida de la población, no de un individuo frente a los demás. Hay que considerar que la similitud entre todos los individuos muestra cómo son las variaciones entre sí y no cómo es el comportamiento evolutivo de una especie. Desde el punto de vista canónico de la evolución, si una especie promueve la diversidad ésta puede tener más oportunidades para evolucionar. Típicamente los algoritmos genéticos contienen una fase de exploración seguida de una fase de explotación.²⁰⁶ Fomentar la diversidad durante el proceso entero de búsqueda puede ser contraproducente para la fase de explotación.²⁰⁷ El tipo y cantidad de diversidad requerida en múltiples momentos no es claro.¹⁰⁹ Para observar el comportamiento de la diversidad, de forma que sea

comparativa entre la población y las generaciones, utilizaremos una medida de la diversidad de la población \mathcal{D} basada en DCM de las diferencias de las distancias de cada par de átomos utilizando la siguiente ecuación^{202,208}

$$\mathcal{D} = \frac{1}{\left(\frac{|N|^2 - |N|}{2}\right)} \sum_{i=1}^{|N|-1} \sum_{j=i+1}^{|N|} \text{DMC}(A_i, A_j), \quad (16)$$

donde N es el número de individuos en la población $\frac{|N|^2 - |N|}{2}$ es el número de pares de individuos A_N . La ecuación 16 mide la diversidad de la población, sin depender del tamaño de la población o el número de átomos, esto nos permite emplearla para conjuntos atómicos variados y está mejor adaptada a sistemas químicos porque ignora la traslación, rotación y reflexión cuando se comparan dos geometrías.²⁰²

Seleccionando los resultados anteriores se obtuvo la figura 18 al emplear la ecuación 16. Comparando *fragment-cut-splice* con los otros tres algoritmos de cruce no se obtuvo una diferencia apreciable. En los resultados para $(\text{H}_2\text{O})_6$ y Al_{13}^- hay una tendencia a disminuir la diversidad en torno a las generaciones en donde aparece el mínimo global (generación ~ 4), luego de encontrar un valor mínimo la diversidad aumenta. En Al_{13}^- se tuvo una caída asintótica a valores de diversidad a 1,4 con *fragment-cut-splice* y en 1,5 en los demás algoritmos, debido a que en todos los casos el mínimo global aparece en la generación 0 ocasionando que se produzcan geometrías similares a la del mínimo global. Al no encontrar un individuo mejor, las primeras poblaciones pierden rápidamente diversidad genética causado por que cada vez más individuos son de alguna manera descendientes del mínimo global. Sin embargo en todos los casos la degradación de la diversidad genética llega a un mínimo y luego empiezan a formarse individuos con genes diferentes, pero este ciclo degradación-variabilidad genotípica ocurre más rápidamente en *plane-rot-splice* y *fragment-cut-splice*. El mínimo de la curva diversidad/generación aparece en la generación 3 y 4, respectivamente, casi la mitad del número de generaciones que en los otros dos métodos. De esta manera, permite continuar explorando otros motivos estructurales en generaciones más tempranas que en los otros dos algoritmos, efecto que se ve reflejado en la cantidad total de mínimos explorados. Aunque, *sphere-cut-splice* y *fragment-cut-splice* son las metodologías en las que hubo mayor diversidad estructural por cantidad de estructuras, *fragment-cut-splice* llega a tener de 2 a 4 veces más mínimos que (*sphere-cut-splice*). En B_{15}^- no hay una variación tan fuerte como en los casos anteriores y la diversidad se mantiene casi sin variación, a excepción de *sphere-cut-splice* que muestra una tendencia creciente en la diversidad hasta la generación 7. La razón de este comportamiento errático en la diversidad puede darse gracias a que, en particular, en los cúmulos de boro se pueden encontrar cúmulos con patrones estructurales característicos, que son constantes en los motivos planos,²⁰⁹⁻²¹¹ que en conjunto con una baja diferencia de

energía entre ellos, hace que en la población estos patrones se repitan durante la evolución del algoritmo, manteniendo la diversidad sin grandes saltos.

En los dos últimos conjuntos atómicos $C_{10}H_8$ y C_6H_6 hubo mayor diversidad de todos los otros tres cúmulos (1,8 – 2,4) gracias a sus características propias, expuestas en párrafos anteriores. Sin embargo en los dos casos la diversidad en *fragment-cut-splice* estuvo por encima de los valores más altos obtenidos con las otras tres metodologías en todas las generaciones. Es interesante notar que en estos dos casos (donde no se logró obtener el mínimo global) hay una tendencia en la diversidad (ascendente) contraria a la de los otros tres cúmulos, donde el mínimo global era una constante. Esta marcada tendencia usando *fragment-cut-splice* podría servir como parámetro para determinar, durante la ejecución del AG, si la geometría de mínimo global ha sido encontrada. Sin embargo, por carecer de más datos que puedan confirmar este efecto, sólo es materia de especulación.

2.6 CONCLUSIONES

En esta sección presentamos los aspectos técnicos del AG para búsqueda de mínimos globales de sistemas atómicos implementado en Kaxan. En este AG se presentan propuestas exitosas para la generación aleatoria de individuos, de cuatro tipos de motivos: planos, espaciales, esféricos y cíclicos. En general, Kaxan tuvo eficiencia en la creación de individuos aptos, aunque las posiciones atómicas de los cúmulos generados son aleatorias, a estas geometrías se les imponen dos condiciones: todos los átomos deben tener por lo menos un enlace y las distancias interatómicas no deben estar por debajo del 70 % de la suma de los radios de covalencia. Imponer estas dos condiciones dan como resultado excelentes individuos de partida al punto de generar, en algunos casos, el mínimo global.

Los operadores de mutación y elitismo, que forman parte del AG pueden adaptarse a gusto del usuario de una manera sencilla permitiendo así proporcionar flexibilidad para estudiar cualquier problema en el rango de átomos que los programas de Química Cuántica y el poder de cómputo permitan. Además, en Kaxan se implementaron diversas estrategias de cruce que permiten aún más abordar problemas químicos de cualquier tipo. Entre las estrategias de cruce de individuos se cuentan algunas ya reportadas por otros autores : *plane-cut-splice*, *plane-rot-splice* y *spheric-cut-splice*, además, de una totalmente original que depende de fragmentos moleculares de los padres y de la intervención de índice topológico llamada *fragment-cut-splice* aunado a ello, en este programa se implementaron varios índices topológicos conocidos (Wiener, Balaban, Kirchhoff, etc), pero el índice de Wiener (Wq) calculado usando la matriz de distancia-resistencia es el que tuvo mejores resultados en la relación estructura energía.

Empleando un conjunto de cinco cúmulos representativos de los motivos estructurales más comunes en Química ($(\text{H}_2\text{O})_6$, Al_{13}^- , B_{15}^- , C_{10}H_8 y C_6H_6) en la búsqueda de mínimo global con diferente número de población para los cuatro algoritmos de cruce, fue posible obtener la geometría de mínimo global y en especial *fragment-cut-splice* muestra eficiencia y velocidad para encontrar el mínimo global además de amplia diversidad, que se ve reflejada en el número final de cúmulos encontrados. A pesar de que *fragment-cut-splice* es un algoritmo que construye un nuevo individuo a partir de partes de los padres ensambladas bajo el criterio de la cercanía a la topología del mejor miembro de la generación anterior, la diversidad estructural no es afectada. Por el contrario, se nota que a pesar de estas restricciones, en *fragment-cut-splice* los mínimos explorados tienen una mayor variación estructural así como una abundancia relativa de especies. Comparado con las demás metodologías, *fragment-cut-splice*, en el peor de los casos, tiene el doble de individuos explorados que los otros tres algoritmos de cruce implementados, además este algoritmo explora más rápidamente motivos estructurales que le permiten llegar al mínimo global con facilidad.

Parte II

RESULTADOS PARA SISTEMAS
REPRESENTATIVOS

3.1 INTRODUCCIÓN

La historia del C_6Li_6 comienza a principios de los años setenta cuando Shimp *et al.* reportó la síntesis del hexalitiobenceno.²¹² Más tarde, Baran y Lagow²¹³ anunció la conversión de hexacloro benceno a hexalitiobenceno con más de 60% de rendimiento,²¹⁴ pero sin proponer estructura alguna. Xie y Schaefer llevaron a cabo una propuesta teórica en el que C_6Li_6 adopta una forma de estrella con grupo puntual \mathcal{D}_{6h} (Figura 19-1).²¹⁵ Glukhovtsev *et al.* también predijo la misma estructura en el mismo año.²¹⁶ Dos años después, Smith estudió algunos isómeros del C_6Li_6 , usando funcionales de la densidad e identificó un isómero más estable a 62.1 kcal/mol, Figura 19-2, que puede describirse como un trímero de C_2Li_2 , en la que los átomos de litio enlazan a los átomos de carbono de otro monómero (Figura 19).²¹⁷ Bachrach y Miller adicionaron más mínimos locales del C_6Li_6 , pero sólo se consideraron formas con anillos de seis miembros.²¹⁸ Aunque la forma \mathcal{D}_{6h} no es termodinámicamente estable, se propuso emplearla para la construcción de nuevos metalocenos o como sistemas para almacenar hidrógeno.²¹⁹⁻²²²

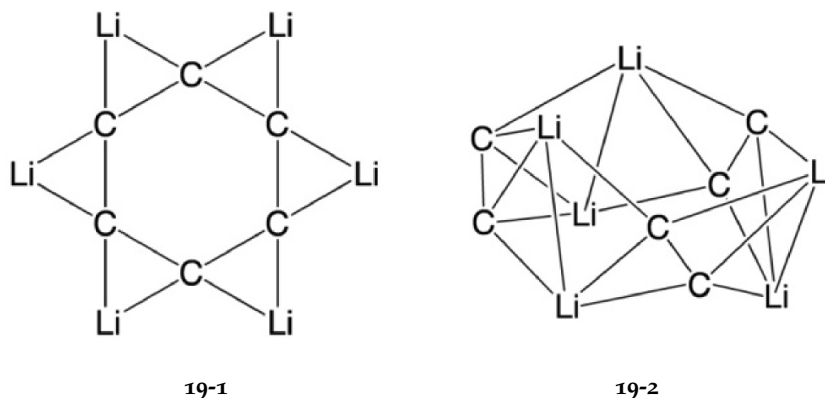


Figura 19: Estructuras Propuestas para el C_6Li_6 .^{215,216}

En este Capítulo se presentan los resultados de la búsqueda de mínimo global del C_6Li_6 usando nuestra versión del AG implementada en el programa Kaxan. Para esta búsqueda se usó el funcional PBE0 y la base DZ²²³ para los cálculos de gradiente y frecuencias, empleando el paquete computacional GAMESS (US).^{224,225} Las geometrías que se obtuvieron del AG se reoptimizaron al nivel PBE0/def2-TZVP.²²⁶ El mismo nivel se utilizó para el cálculo de frecuencias con la finalidad de caracterizar la naturaleza de todos los puntos estacionarios y cal-

cular la corrección de energía de punto cero. Las energías totales de las estructuras más estables se obtuvieron al nivel CCSD(T)^{227,228}/def2-TZVP//PBE0/def2-TZVP. La naturaleza del enlace se estudió empleando la Teoría de Orbitales Naturales (NBO, por sus siglas en inglés) propuesta por Weinhold *et al.*²²⁹ y el Análisis de la Descomposición de la Energía de Isomerización (IEDA, por sus siglas en inglés). Tanto las reoptimizaciones en PBE0 y los cálculos de energía CCSD(T) se llevaron a cabo con el programa Gaussian 03.²³⁰

Para entender la diferencia de energía entre la estructura de estrella y el mínimo global se estudió la descomposición de la energía de los isómeros empleando el IEDA,²³¹ al nivel PBE0/TZ2P//PBE0/def-2TZVP con el programa ADF 2012.^{232,233} Las simulaciones de Dinámica Molecular Born-Oppenheimer (BOMD, por sus siglas en inglés)²³⁴⁻²³⁶ se efectuaron utilizando el programa DEMON2K²³⁷ al nivel BP86^{238,239}/DZVP.²⁴⁰ La temperatura de las simulaciones BOMD se controla por un termostato Hoover chain. El comportamiento del desplazamiento cuadrático medio en función del tiempo nos permitió distinguir entre el comportamiento de tipo sólido o líquido. El desplazamiento cuadrático medio al tiempo t se calculó con la Ecuación 3.1, en donde $\mathbf{r}_i(t)$ es el vector de posición del i -ésimo átomo en el momento t y N es el número total de átomos en el sistema.

$$\langle \text{msd} \rangle = \frac{1}{N} \sum_{i=1}^N [\mathbf{r}_i(t) - \mathbf{r}_i(0)]^2. \quad (17)$$

Los mínimos locales más estables del C_6Li_6 , que se encuentran hasta 15 kcal/mol por encima del mínimo global, así como el isómero tipo estrella 20-1 se representan en la Figura 20.

3.2 ESTRUCTURA Y ENERGÍAS

Una de las características del litio es su proclividad a formar puentes^{241,242} y el C_6Li_6 no es la excepción. Aunque 20-2 (Figura 20) resultó ser un mínimo local a un nivel PBE0/def2-TZVP, es menos estable que la estructura 20-3. Esta nueva estructura asimétrica 20-3 es 6.3 kcal/mol más baja en energía que 20-2. El efecto de la redistribución de los átomos de litio entre 20-2 y 20-3, según la distancia de enlace C–C, es despreciable. En ambos isómeros las longitudes C–C son 1.250 Å, ligeramente más cortas que las que se encuentran en C_2Li_2 (1.258 Å). Un intento posterior de aumentar la simetría a \mathcal{C}_s moviendo Li_5 y Li_6 (figura 21) resulta en dos mínimos locales menos compactos, 20-4 y 20-5, que son menos estables que 20-3 por 8.1 y 13.1 kcal/mol, respectivamente. La estructura de estrella es de mayor energía que la del mínimo global por 82.9 kcal/mol. La búsqueda se extendió a los estados tripletes y se encontró que el triplete es menos estable que 20-3 por 46.6 kcal/mol (a nivel PBE0/def2-TZVP). Teniendo en cuenta que la diferencia de energía entre 20-3 y 20-1 calculada a nivel CCSD(T) varía

considerablemente a la obtenida utilizando el nivel PBE0, se llevó a cabo el análisis T1 para determinar el carácter multireferencial de cada mínimo. Sólo **20-1** tiene un carácter multireferencial ($T1=0.022$), mientras que el resto de los mínimos locales se describen correctamente por un cálculo monoreferencial.

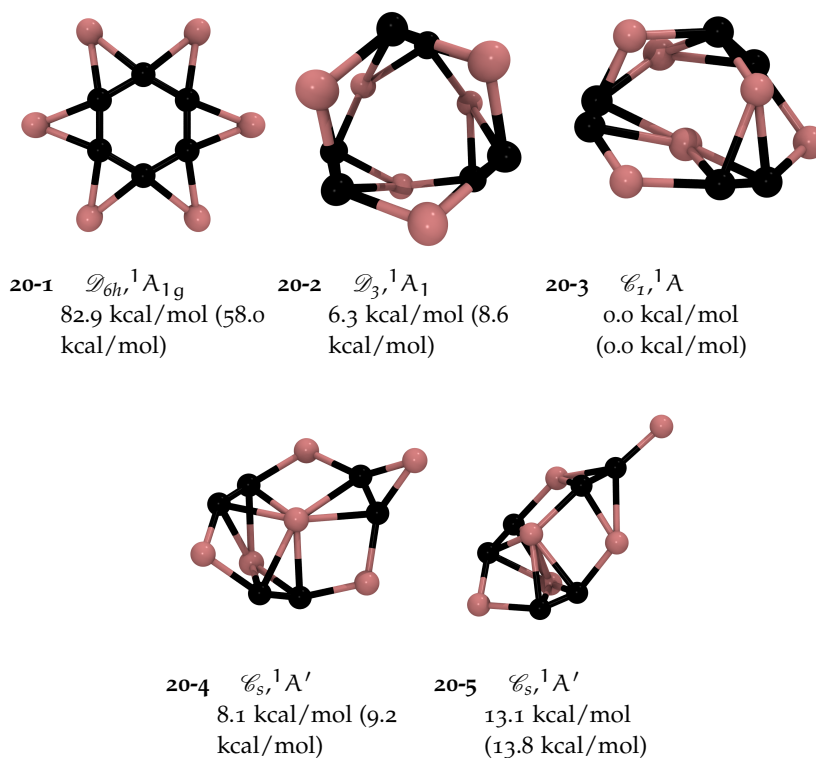


Figura 20: Isómeros del C_6Li_6 . Energías relativas incluyendo la energía de punto cero, a nivel CCSD(T)/def2-TZVP//PBE0/def2-TZVP
Energías relativas PBE0/def2-TZVP en paréntesis

3.3 NATURALEZA DEL ENLACE

El cúmulo **20-3** puede describirse como un trímero de C_2Li_2 , pero la energía de disociación calculada para la reacción $C_6Li_6 \longrightarrow 3 C_2Li_2$ es 138.3 kcal/mol. En consecuencia, los tres fragmentos C_2Li_2 poseen una interacción fuerte, lo que resulta en una nueva entidad química. Un Análisis de Población de Orbitales Naturales (NPA, por sus siglas en inglés)²⁴³ mostró que las interacciones C–Li tienen un carácter iónico importante, con cargas en cada Li en un intervalo de 0.67-0.80 e^- , a excepción de Li1, cuya carga es sólo 0.34 e^- (Figura 21). Esto reveló la naturaleza covalente multicéntrica de las interacciones C–Li en el que el puente de litio utiliza sus orbitales p para ese propósito. Por otro lado, el Índice de Enlace de Wiberg (WBI, por sus siglas en inglés)²⁴⁴ soporta la presencia de tres enlaces $C\equiv C$ triples (IEW = 2.89) en **20-3**,

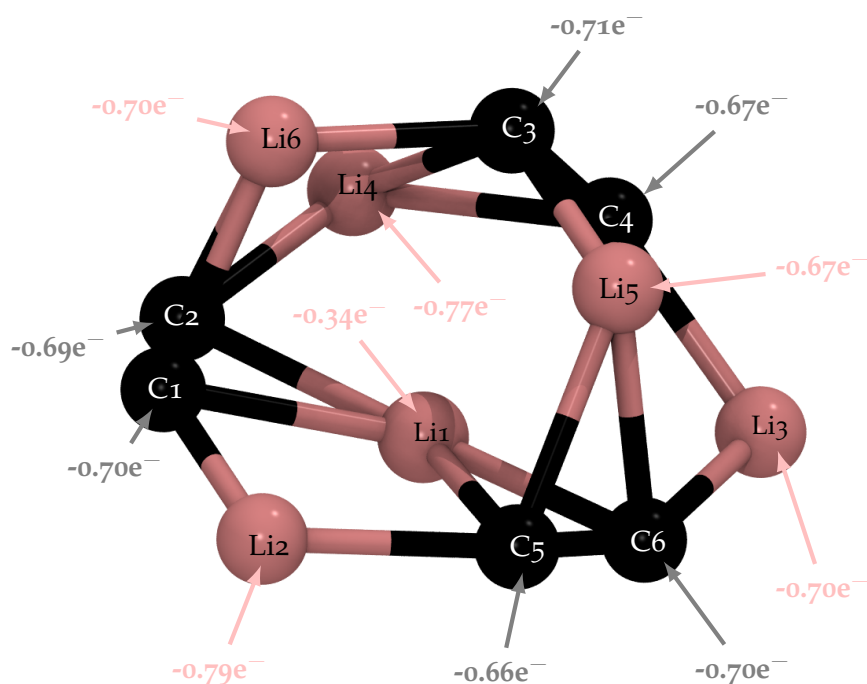


Figura 21: Análisis poblacional en orbitales Naturales para el isómero 20-3.

que corresponden a los enlaces cortos C–C. Las cargas naturales de los dímeros de carbono son $1.4 e^-$, apoyando aún más la presencia de fragmentos C_2^{2-} .

3.4 ESTRUCTURA SIMÉTRICA VS ASIMÉTRICA

Existen varios ejemplos de cúmulos de tipo estrella para el carbonolito y combinaciones de silicio-litio.^{231,245-248} Se esperaría que la forma de estrella de C_6Li_6 fuera una de las estructuras más estables. La aromaticidad es un argumento empleado para apoyar tal suposición. Sin embargo, en algunos casos la aromaticidad no es suficiente para preservar la simetría de una molécula. Una situación similar se produce en Si_6Li_6 , en el que la estructura planar \mathcal{D}_{6h} es menos estable que el mínimo global por 60.2 kcal/mol.²⁴⁸ La pregunta que aún subsiste es la razón de por qué 20-3 es más estable que 20-1. Para conocer cuantitativamente la naturaleza de las interacciones C–Li y cómo estas interacciones contribuyen a la estabilización de la estrella o estructuras asimétricas, consideramos una disociación hipotética: la reacción $C_6Li_6 \longrightarrow C_6^{6-} + Li_6^{6+}$. Con el uso de IEDA es posible racionalizar los orígenes de la preferencia isomérica en cúmulos de alta forma iónica de manera cuantitativa. La energía de isomerización, para pasar de 20-1 a 20-3, calculada a nivel PBE0/TZ2P//PBE0/def2-TZVP es -56.0 kcal/mol.

Es evidente que la energía de interacción C–Li es mayor en el sistema asimétrico por 266.2 kcal/mol (ver [Tabla 3](#)). Además, el hexágono de carbono es menos estable que el trímero C_2^{-2} por 137.3 kcal/mol. Curiosamente, la disposición hexagonal plana de átomos de litio es más estable que la encontrada en [20-3](#), pero esto no es suficiente para preservar la estructura planar. Aunque la interacción electrostática es mayor en la forma plana, la interacción covalente es más eficiente en el isómero [20-3](#). Además, la repulsión de Pauli disminuye en la forma asimétrica. En otras palabras, la transferencia de electrones desde los átomos de litio al esqueleto del carbono induce la formación de una estructura altamente iónica y muy inestable de carbono que prefiere romper los enlaces C–C dando origen a tres unidades C_2^{-2} .

Tabla 3: Resultados de IEDA con un nivel de teoría PBE0/TZ2P para los isómeros [20-1](#) y [20-3](#) con C_6^{-6} y Li_6^{+6} como fragmentos. Energías en kcal/mol

	20-1	20-3	$\Delta(20-1 \rightarrow 20-3)$
ΔE_{int}	-3835.6	-4101.8	-266.2
$E(C_6^{6-})$	308.6	171.3	-137.3
$E(Li_6^{6+})$	1890.9	2238.4	347.5
ΔE_{pauli}	463.0	256.5	-206.6
ΔE_{elstat}	-4091.6	-3937.5	154.1
ΔE_{orb}	-207.0	-420.8	-213.0

3.5 FLUXIONALIDAD

También se encontró que el estado de transición que se relaciona con la isomerización de [20-3](#) a su imagen especular, el cual está a sólo 2.0 kcal/mol, lo que indica una rotación casi libre de los fragmentos de carbono. Basado en el comportamiento fluxional de los cúmulos de litio y también una potencial redistribución de los átomos de litio en [20-3](#) inducida por la rotación de las fracciones de C_2 , se llevaron a cabo simulaciones de dinámica molecular en el sistema a diferentes temperaturas (300, 600, y 900 K). A 300 K, la estructura [20-2](#) cambia inmediatamente a [20-3](#). A la misma temperatura, la estructura [20-3](#) experimenta sólo un continuo intercambio en las posiciones Li1 y Li5/Li6. La Función de Distribución Radial (FDR) muestra una estructura con picos bien definidos a 1.25, 2.15 y 3.15 Å. Los dos primeros picos corresponden a las longitudes de enlace C–C y C–Li. El hombro a 2.65 Å y el tercer pico reflejan las distancias Li–Li a primeros y segundos vecinos, respectivamente. A 600 y 900 K, la estructura C_6Li_6 sufre una secuencia de reordenamientos rápidos que muestran claramente una completa aleatorización de los átomos de litio y la rotación

libre de los dímeros de carbono. Las funciones de distribución radiales totales muestran que, a pesar de la rotación libre de los fragmentos de carbono, las longitudes de enlace C–C permanecen inalteradas. En cambio, como consecuencia de la movilidad del litio los otros picos se reducen ampliamente (Figura 22).

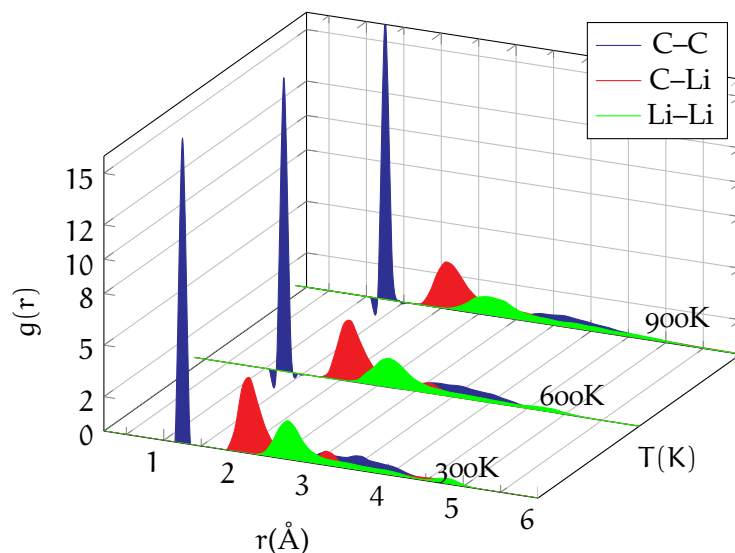


Figura 22: Funciones de distribución radial para C_6Li_6 a 300, 600 y 900K.

La fluxionalidad del C_6Li_6 también se refleja en el desplazamiento cuadrático medio, $\langle msd \rangle$. Haile²⁴⁹ propuso un método para seguir la rigidez de un sistema. Si el sistema es de tipo sólido, se esperaría que el $\langle msd \rangle$ quede esencialmente constante y cercano a cero, mientras que un sistema de tipo líquido debe exhibir un aumento casi lineal en el $\langle msd \rangle$ con el tiempo. La Figura 23 resume las gráficas de $\langle msd \rangle$ en función del tiempo. A 300 K, C_6Li_6 se comporta claramente como una molécula tipo sólido, en el que la estructura permanece casi invariante durante la simulación, pero el comportamiento de $\langle msd \rangle$ cambia drásticamente a 600 y 900 K. Como resultado de la movilidad nuclear, en particular de los átomos de Li, el C_6Li_6 puede describirse como un sistema de tipo líquido a estas temperaturas.

3.6 CONCLUSIONES

En resumen, después de la búsqueda con el AG para C_6Li_6 se encontró un nuevo mínimo global no simétrico. Este puede describirse por tres fragmentos C_2^{-2} fuertemente unidos a través de puentes de litio. Las cargas naturales confirman que los enlaces C–Li en C_6Li_6 son de tipo iónico. Los cálculos IEDA indican la formación de un fragmento de carbono altamente iónico y muy inestable que prefiere romper los enlaces C–C para producir tres unidades C_2^{-2} . A altas temperaturas, se observa la fluxionalidad del cúmulo C_6Li_6 en lugar de disociación. A 600 y 900 K, las simulaciones BOMD muestran que la movilidad de

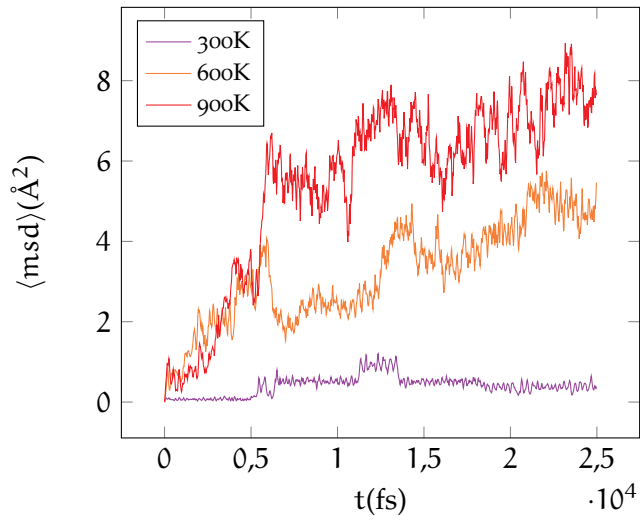


Figura 23: Variación temporal del desplazamiento cuadrático medio $\langle \text{msd} \rangle$ para C_6Li_6 a 300, 600 y 900K.

litio es alta lo que indica que el cúmulo se comporta de manera similar a un líquido.

Estos resultados sugieren que otras estructuras de carbono podrían “disolverse” en presencia de una solución concentrada de litio, lo cual es relevante para entender el comportamiento de las baterías de ion-litio con ánodos basados en materiales de carbono, como el grafito.

B_{18}^{2-} UN MIEMBRO QUASIPLANAR DE LA FAMILIA MOTOR DE WANKEL

4.1 INTRODUCCIÓN

En el 2010, los grupos de Boldyrev y Wang informaron de la detección de una rueda formada por diecinueve átomos de boro (B_{19}^-).²⁵⁰ Este cúmulo contiene un fragmento de seis átomos de boro formando un pentágono rodeado por los otros trece átomos. En estudios posteriores se encontró que el pentágono interno y el anillo exterior de boro pueden girar casi libremente en direcciones opuestas, de forma similar a un motor de tipo Wankel.²¹¹ Este tipo de comportamiento dinámico es análogo al encontrado en algunas ruedas de boro aromático (C_2B_8 , $C_3B_9^{3+}$ y $C_5B_{11}^+$) con más de un átomo de carbono en el centro.²⁵¹ Desafortunadamente, en todas esas ruedas borocarbonadas, los isómeros que tienen átomos de carbono en el anillo externo son de más baja en energía. En 2011 se reportó que el B_{13}^+ también exhibe un comportamiento dinámico similar al del B_{19}^- , con una rotación casi libre del fragmento B_3 interno.²⁵² El grupo de Alexandrova sugirió que es posible controlar la dirección de la rotación del triángulo B_3 en B_{13}^+ empleando un láser infrarrojo polarizado.^{210,253} Tai *et al.* encontraron que el $B_{20}^{-/2-}$ cuenta con una subunidad hexagonal rodeada por el anillo de B_{13} que exhibe el mismo tipo de comportamiento dinámico.²⁵⁴

En esta sección se describe a un nuevo miembro de esta familia de cúmulos. Una búsqueda estocástica reveló que la estructura más estable de B_{18}^{2-} comprende un fragmento pentagonal de seis átomos de boro, rodeados por el anillo de B_{12} , pero en contraste con el sistema B_{19}^- , tiene una forma de domo. A pesar de esto, simulaciones de BOMD indicaron que el sistema se comporta dinámicamente similar al movimiento de un “motor tipo Wankel”. Para mayor comprensión de la fluxionalidad y la estabilidad del sistema, se estudió el enlace y la aromaticidad. También se examinó la interacción de B_{18}^{2-} con un catión litio (LiB_{18}^-).

4.2 DETALLES COMPUTACIONALES

En la búsqueda del mínimo global se empleó nuestra versión del AG implementado en Kaxan (considerando estados singuletes y tripletes) para el B_{18}^{2-} y LiB_{18}^- . El funcional PBE0¹⁸⁴ y la base D95 se emplearon para el cálculo de la energía, gradiente y hessiana utilizando el programa GAMESS (US).^{224,225} Las estructuras obtenidas del AG se reoptimizaron a un nivel TPSS/def2-TZVP.^{226,255} También se obtuvieron las frecuencias

El motor diseñado por Wankel en 1924, es un motor de combustión interna que en lugar de pistones usa rotas.

de vibración al mismo nivel para caracterizar la naturaleza de todos los puntos estacionarios y calcular las correcciones de la Energía de Punto Cero (ZPE, por sus siglas en inglés). El Análisis de la Descomposición Natural de la Densidad (AdNDP, por sus siglas en inglés)^{89,256} se calculó a un nivel TPSS/def2-TZVP. Las simulaciones BOMD a nivel PBE/DZVP se obtuvieron con el programa DEMON2K²³⁷ a partir de la geometría de mínimo global del sistema B_{18}^{2-} .¹

4.3 RESULTADOS

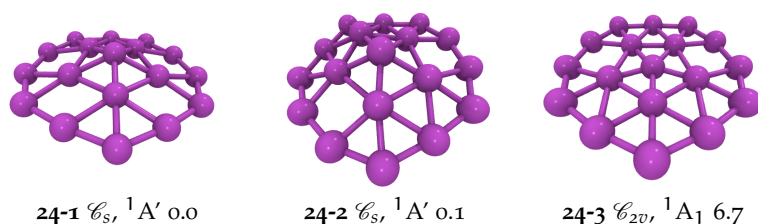


Figura 24: Mínimo global para el B_{18}^{2-} 24-1 y los estados de transición relacionados con la rotación 24-2 y la inversión 24-3. Las energías relativas están dadas en kcal/mol.

La estructura del mínimo global del B_{18}^{2-} no es plana, es una estructura de tipo domo que tiene simetría \mathcal{C}_s y se compone de un anillo de B_{12} exterior conectado a un anillo B_5 interior ligeramente fuera del plano, (Figura 24). Los centros del anillo B_{12} y el anillo B_5 están separados por ~ 0.6 Å, mientras que los átomos de B del centro están ~ 0.5 Å por encima del anillo B_5 . Por lo tanto, la profundidad del domo es de alrededor de 1.1 Å. El mínimo global es sólo 1.9 kcal/mol más estable que el segundo isómero plano, una estructura plana \mathcal{D}_{3h} (Figura 25). Por lo que, a una temperatura dada pueden existir ambos isómeros. En el 2013, los grupos de Wang y Boldyrev²⁵⁷ reportaron la existencia del cúmulo B_{18}^- en fase gaseosa usando espectroscopía fotoelectrónica además incluyen la estructura encontrada con métodos estocásticos. Ellos afirman que el isómero de más baja energía del B_{18}^- tiene simetría \mathcal{C}_{3v} , en donde la unidad central de B_3 está alrededor de 0.5 Å fuera del plano molecular lo que la hace una molécula cuasi-planar. La estructura del $B_{18}^- \mathcal{C}_{3v}$ es similar a la forma del segundo mínimo de más baja energía del B_{18}^{2-} . La unidad central de B_3 fuera del plano en el isómero \mathcal{C}_{3v} de B_{18}^- se acerca más al plano molecular en el isómero \mathcal{D}_{3h} del cúmulo B_{18}^{2-} . La carga negativa adicional en el sistema puede ayudar ligeramente a la expansión del anillo periférico de boro, permitiendo que se tenga el espacio suficiente para que la unidad B_3 se acomode en

¹ Las simulaciones se calcularon con la geometría de mínimo global 24-1, asignando velocidades aleatorias a los átomos, además la estructura se equilibró a 900K, empleando un baño térmico Nosé-Hover. Para 25 ps con 1.0 fs de tamaño de paso.

el plano. Además, el segundo isómero más estable del B_{18}^- es similar al mínimo global de B_{18}^{2-} . El isómero \mathcal{C}_s en B_{18}^- es sólo 1.6 kcal/mol menos estable que su forma \mathcal{C}_{3v} . Incluso, se identificaron ambos isómeros en el espectro fotoelectrónico. Para el caso del B_{18}^{2-} , la diferencia de energía entre los dos isómeros más estables es de 2.5 kcal/mol, calculados a nivel de teoría CCSD(T)/def2-TZVP//TPSS/def2-TZVP (Figura 25).

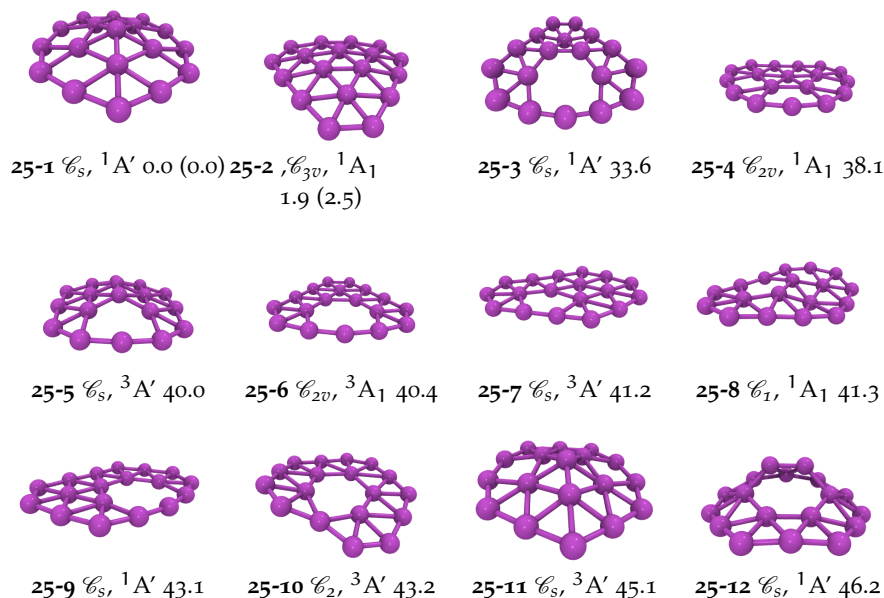


Figura 25: Estructuras más estables del B_{18}^{2-} y sus energías relativas (kcal/mol), incluyendo la ZPE, en el nivel TPSS/def2-TZVP y las energías calculadas a CCSD(T)/def2-TZVP entre paréntesis.

La característica interesante del isómero \mathcal{C}_s del cúmulo B_{18}^{2-} es que el menor modo normal de vibración (102.2 cm^{-1}) corresponde a la rotación de la unidad B_6 interior, esto es un indicador importante para que pueda presentar la rotación de estos átomos. Siguiendo este modo normal de vibración se encontró que la estructura corresponde a un estado de transición con una frecuencia imaginaria de $111,1i \text{ cm}^{-1}$. Este modo tiene relación con la rotación de la unidad B_6 interior. La diferencia de energía entre 24-1 y 24-2 es insignificante (0.1 kcal/mol) (Figura 24). Por lo tanto, una barrera tan baja sigue una rotación casi libre de la unidad B_6 dentro del anillo B_{12} . Esto coincide con los resultados de la BOMD. La unidad B_6 interior gira casi libremente dentro del perímetro del anillo de B_{12} . En la dinámica molecular este comportamiento se detectó a 900 K. La rotación de la unidad B_6 se acompaña por el rompimiento y formación simultánea de enlaces entre la unidad B_6 y el anillo periférico B_{12} , siendo este el primer sistema cuasi-planar que muestra tal fluxionalidad. También hemos localizado el estado de transición para la inversión del “domo” del isómero \mathcal{C}_s (estructura 24-3 en la Figura 24), que tiene una barrera de 6.7 kcal/mol. Esta barrera es menor que la encontrada en el sumaneno

(19.6 kcal/mol) o el coranuleno (10.2 kcal/mol), estructuras clásicas de carbono tipo domo.^{258,259}

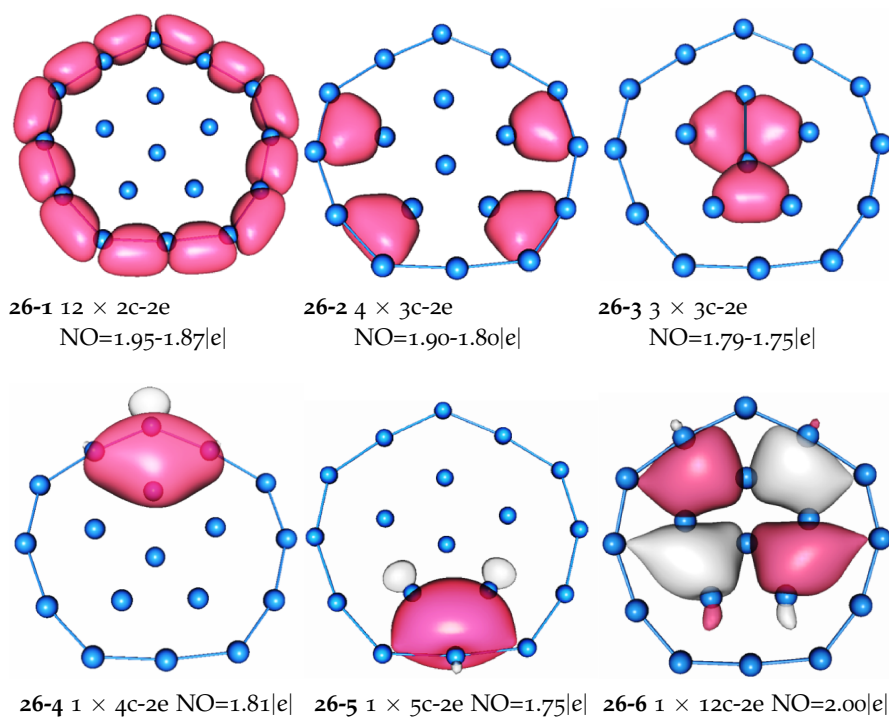
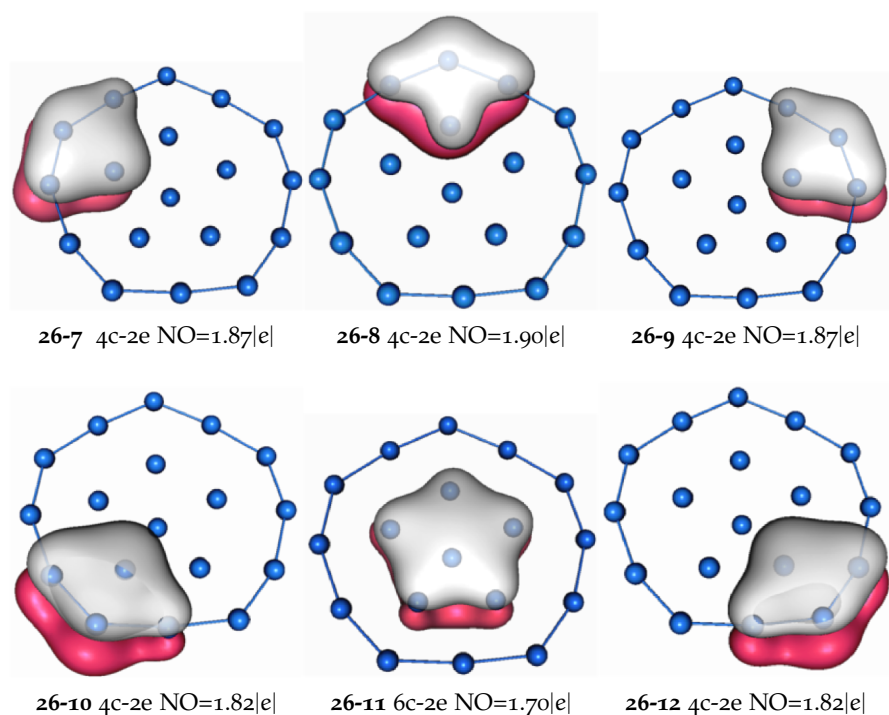
4.4 ANÁLISIS DEL ENLACE Y LA FLUXIONALIDAD

En el B_{19}^- y el B_{13}^+ la ausencia de enlaces localizados 2c-2e σ -enlazantes entre la unidad interior y el anillo exterior favorece tal rotación libre. El Número de Ocupación (NO) representa el número de electrones que participan en un enlace. Los NO para todos los enlaces en el sistema B_{18}^{2-} están cerca del valor límite de 2.0 $|e|$ por enlace. Este análisis muestra que hay 12 enlaces- σ de dos centros dos electrones (2c-2e) localizados en el anillo periférico B_{12} . Sin embargo, el resto de enlaces es de tipo deslocalizado (Figura 26). Hay 10 enlaces- σ deslocalizados, de los cuales siete están sobre tres átomos ($7 \times 3c-2e$), uno está sobre cuatro átomos (4c-2e), uno más sobre cinco átomos (5c-2e) y el último que se deslocaliza sobre 12 átomos (12c-2e). Entre estos, tres enlaces- σ 3c-2e involucran el fragmento interno B_6 , mientras que los enlaces- σ restantes deslocalizados son responsables de la unión entre la unidad B_6 y el anillo exterior B_{12} . El análisis AdNDP también muestra que hay seis enlaces- π deslocalizados: cinco se deslocalizan sobre cuatro átomos ($5 \times 4c-2e$) y uno sobre seis átomos del anillo interior (6c-2e). La densidad electrónica de cada uno de los cinco enlaces- π 4c-2e permanece localizada sobre los tres átomos de la periferia y un átomo del anillo B_5 interior. Por lo tanto, a partir del análisis AdNDP podemos inferir que no hay enlaces localizados entre el anillo B_{12} y el conjunto interno B_6 ; sólo se unen *vía* enlaces σ - multicéntricos-2e y enlaces π . Tales enlaces deslocalizados migran fácilmente de una posición a la otra durante la rotación de la unidad de B_6 ; por lo tanto, tales arreglos permiten al cúmulo exhibir un comportamiento fluxional.

4.5 AROMATICIDAD

El número total de electrones σ y π deslocalizados en B_{18}^{2-} es 20 y 12, respectivamente. Utilizando la regla de Hückel de aromaticidad, como tal, el sistema global debiera ser tanto σ - y π -antiaromático en las dos regiones (en la región interna y el anillo exterior), la conclusión es diferente. Tres enlaces- σ 3c-2e y un enlace- π 6c-2e están dentro de la unidad B_6 que satisface la regla de Hückel (para enlaces σ , $4n + 2 = 6$, $n = 1$ y para enlaces- π , $4n + 2 = 2$; $n = 0$). El fragmento interior B_6 es, por lo tanto, σ - y π -aromático. Ahora, la región entre el fragmento de B_6 y el anillo de la periferia B_{12} contiene 14 electrones σ deslocalizados ($4n + 2 = 14$; $n = 3$) y 10 electrones π deslocalizados ($4n + 2 = 10$; $n = 2$), una vez más para satisfacer la regla ($4n + 2$) (Figura 26).

Cabe señalar que la densidad electrónica de los enlaces- σ 12c-2e permanece dentro de las dos regiones, ya que la máxima densidad electrónica se encuentra sobre el área entre el fragmento B_6 y el anillo

Enlaces- σ Localizados y DeslocalizadosEnlaces- π DeslocalizadosFigura 26: Análisis AdNDP para el B_{18}^{2-} a nivel TPSS/def2TZVP.

periférico, por eso se ha considerado esta región solamente. Por lo tanto, el sistema B_{18}^{2-} puede considerarse doblemente aromático. El isómero \mathcal{C}_s del sistema B_{18}^{2-} tiene una configuración de capa cerrada

con un Orbital molecular más alto ocupado (HOMO, por sus siglas en inglés) doblemente ocupado de tipo π (Figura 27). Por lo tanto, también puede considerarse como un análogo de boro del coroneno.

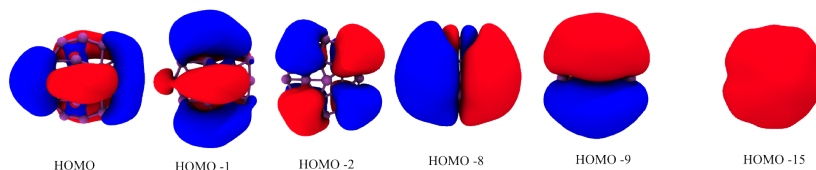


Figura 27: Orbitales moleculares π del cúmulo B_{18}^{2-} .

En el 2012, Nguyen *et al.*^{254,260,261} emplearon el modelo de una partícula en una caja circular para entender la distribución orbital en las ruedas de boro. Un sistema puede mostrar la disco-aromaticidad si sus electrones π ocupan completamente los estados propios más bajos del modelo en orden ascendente de 1σ , 1π , 1δ , 2σ , 1ϕ , 2π y así sucesivamente. Por lo tanto, los sistemas con 2, 6, 10, 12, 16, 20 ... electrones π conducen a una configuración completamente ocupada, que muestra la disco-aromaticidad. En nuestro caso, el sistema B_{18}^{2-} tiene 12 electrones π con la configuración $(1\sigma)^2 (1\pi)^4 (1\delta)^4 (2\sigma)^2$, entonces, es un sistema disco-aromático. Las formas de los orbitales π del sistema B_{18}^{2-} y las funciones de onda más bajas para una partícula en una caja circular se muestran en la Figura 28.

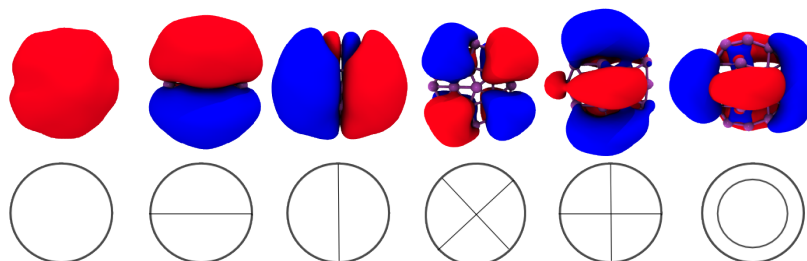


Figura 28: Formas de los orbitales π de valencia del cúmulo B_{18}^{2-} y las funciones de onda más bajas para una partícula en una caja circular.

Para demostrar la doble σ - y π -aromaticidad en B_{18}^{2-} se realizó el análisis de la componente z del campo magnético inducido (B_z^{ind}).²⁶²⁻²⁶⁴ Los valores positivos y negativos de B_z^{ind} indican el comportamiento paratrópico (antiaromaticidad) y diatrópico (aromaticidad) del sistema. Los perfiles B_z^{ind} calculados en los centros de los anillos de tres y cuatro átomos situados ya sea dentro del fragmento interior B_6 o entre el B_6 y anillo exterior B_{12} se muestran en la Figura 29. La respuesta magnética es de naturaleza altamente diatrópica dentro del plano de los anillos pequeños y aunque disminuye gradualmente con la distancia desde el centro del anillo, aún mantiene un valor alto (en sentido absoluto). Incluso a 4 Å por encima o por debajo del plano, el valor B_z^{ind} encontrado es igual o mayor (en el sentido absoluto) a -10 ppm.

Los anillos interiores (1, 2 y 3) muestran un mayor carácter diatrópico (tanto en el plano como fuera de el) que los anillos exteriores que presentan mayor aromaticidad (tanto σ y π). El anillo de cuatro miembros (8) tiene una pequeña respuesta diatrópica. Por lo tanto, los valores negativos de B_z^{ind} en el plano y perpendiculares al plano de los anillos en ambas regiones confirman que es doblemente σ concéntrica, así como la naturaleza doblemente π -aromática del sistema B_{18}^{2-} .

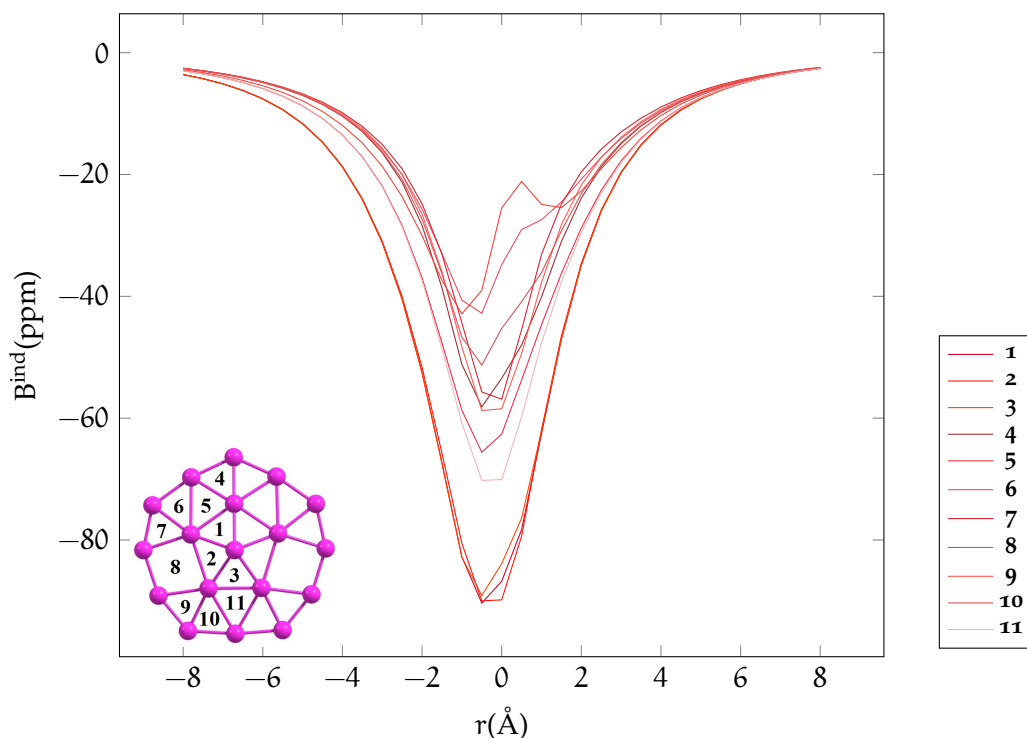


Figura 29: Perfil B_z^{ind} del B_{18}^{2-} calculado a nivel PW91/def2-TZVP. El perfil empieza en el centro geométrico de cada anillo de tres y cuatro miembros.

Por último, hemos calculado tanto Energía de Disociación Vertical (VEDE, por sus siglas en inglés) y Energía de Disociación Adiabática (AEDE, por sus siglas en inglés) con diferentes funcionales de la densidad, la Función de Green de Valencia-Exterior (OVGF, por sus siglas en inglés) y el método basado en la función de onda (Tabla 4).^{265,266} Con excepción de los valores de la VEDE calculados mediante el uso de la denominada solución de enfoque directo, el resto de los resultados muestran la naturaleza de los electrones en HOMO en el B_{18}^{2-} , que es un dianión estable respecto a la pérdida espontánea de electrones. Sin embargo, también se ha examinado la posibilidad de estabilizar el dianión con la adición de Li^+ como un contraión. En el mínimo global del LiB_{18}^- , la estructura del esqueleto de boro es idéntica a la de B_{18}^{2-} y el catión Li^+ se encuentra unido a un átomo de boro del centro de la superficie cóncava interior del anillo B_5 que tiene una distancia B–Li de 2.217 Å (Figura 30).

Tabla 4: Energía de disociación vertical (VEDE en eV) calculada por el Teorema de Koopman (TK) y Δ SCF, la OVGf y la energía de disociación adiabática (AEDE en eV). La base usada en todos los casos es def2-TZVP.

Nivel	VEDE		AEDE	OVGF
	TK	Δ SCF		
TPSS	-1.534	0.284	0.104	0.503
B3LYP	-1.072	0.306	0.135	0.525
PBE0	-0.822	0.401	0.219	0.518
M05-2X	0.075	0.543	0.365	0.542
ω B97X-D	0.683	0.450	0.223	0.570
MP2	0.202	1.044	0.855	0.480

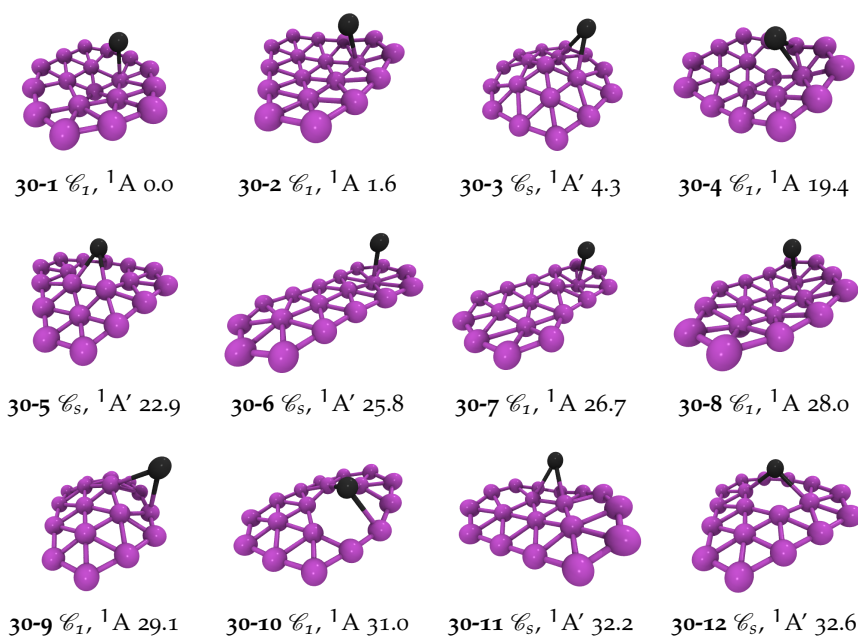
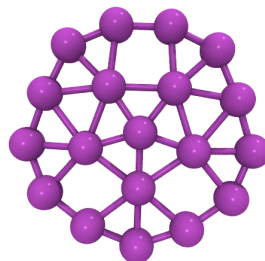


Figura 30: Estructuras más estables del LiB_{18}^- y sus energías relativas (kcal/mol), incluyendo la ZPE, en el nivel TPSS/def2-TZVP.

4.6 UNA SUSTITUCIÓN QUE FRENA EL MOTOR TIPO WANKEL

El cúmulo monoaniónico de diecinueve átomos de Boro (B_{19}^-) adopta una estructura de anillo con un átomo central rodeado por un núcleo pentagonal que a su vez está rodeado por un anillo exterior de trece átomos de boro (Figura 31). Tanto el pentágono central de seis átomos y el anillo exterior de trece átomos satisfacen independientemente la regla de Hückel ($4n + 2$) π (con dos y diez electrones, respectivamente). En el 2010, Huang *et al.* reportaron la estructura y el enlace químico de este cúmulo.²⁶⁷ En el mismo año, se reportó un comportamiento

fluxional del cúmulo B_{19}^- que muestra una rotación interna casi libre de la forma pentagonal dentro del anillo B_{13} co-planar, que es evidente a partir de los resultados de la BOMD indicando un movimiento similar al de un motor tipo Wankel.²⁶⁸



31-1

Figura 31: Estructura de mínimo global de B_{19}^- .

Una pregunta que surge de inmediato, es: ¿cómo se puede controlar, iniciar o dirigir la rotación de estos sistemas? en el 2012, Zhang *et al.* propusieron controlar la rotación del anillo exterior en B_{13}^+ en una dirección dada por la aplicación de un campo láser externo.²⁵³ Con el fin de que gire, sin un campo externo, el B_{13}^+ debe superar una barrera de sólo 0.25 kcal/mol y, dada la simetría del sistema, la rotación puede ocurrir en ambas direcciones con la misma probabilidad. Sin embargo, si un campo eléctrico se aplica en el plano de la molécula, perpendicular al eje molecular, la simetría del sistema se rompe de tal forma que el anillo prefiere girar unidireccionalmente.²¹⁰

En esta sección se describe la naturaleza del comportamiento fluxional del cúmulo neutro CB_{18} , que es el producto de una sustitución de un átomo de boro por carbono en el B_{19}^- . Encontramos que a pesar de que los anillos en B_{19}^- y CB_{18} son isoelectrónicos, y estructuralmente similares, el sistema borocarbonado neutro no exhibe el mismo comportamiento fluxional como el anión de boro puro.

El procedimiento computacional implicó el uso de nuestro AG implementado en el programa Kaxan a nivel PBE0/def2-TZVP.^{184,226} Estos cálculos se realizaron utilizando el paquete computacional Gaussian 09.²⁶⁶ Las simulaciones de BOMD y el cálculo de las propiedades electrónicas, incluyendo el campo magnético inducido, se llevaron a cabo a nivel de PBE/DZVP usando el paquete de software de Mon2K.²³⁷

Los cinco mínimos locales más estables del CB_{18} , que se encuentran a 20 kcal/mol por encima del mínimo global, se muestran en la Figura 32. En el mínimo global (32-1), el átomo de carbono se encuentra en el fragmento de anillo exterior, es decir, el átomo de carbono muestra una preferencia importante por un lugar en el anillo exterior sobre la unidad pentagonal interior o el centro de la rueda de boro, evitando la hipercoordinación extrema como lo describe Boldyrev y

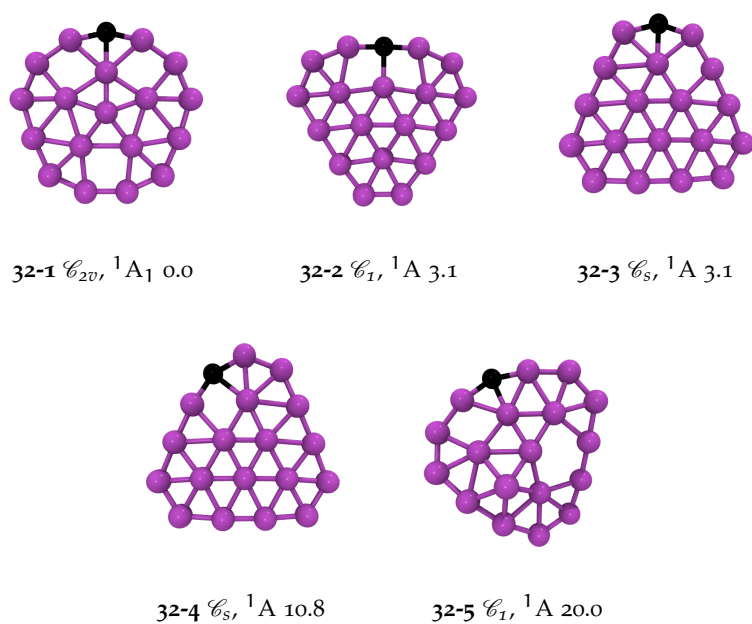


Figura 32: Isómeros optimizados de CB_{18} , hasta 20 kcal/mol. Las energías relativas están en unidades de kcal/mol. Todas las estructuras fueron optimizadas al nivel PBE0/def2-TZVP

Wang.²⁶⁹ De hecho, el átomo de carbono en las otras cuatro especies de más baja energía permanece en el anillo exterior. La diferencia de energía entre el isómero de menor energía y la siguiente forma isomérica más estable es sólo 3.1 kcal/mol. Cuando el carbono se coloca en el pentágono interno o incluso en el centro del anillo la energía relativa es 40 kcal/mol mayor a [32-1](#).

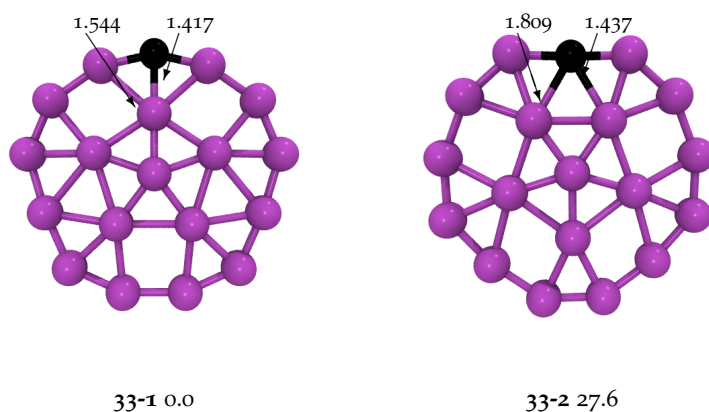


Figura 33: Estructura de mínimo global de CB_{18} y el estado de transición relacionado a la rotación del pentágono interno. Las energías relativas están en kcal/mol.

Los parámetros geométricos más relevantes de **32-1** se resumen en la **Figura 33**. La distancia C-B que conecta los dos anillos es más larga (1.544 Å) que la calculada para el anillo externo. A primera vista, el carbono no altera drásticamente la estructura de B_{19}^- . El análisis vibracional muestra que **32-1** es un mínimo y la frecuencia de vibración más pequeña corresponde a un modo vibracional de 56 cm^{-1} , asignado al movimiento fuera del plano del pentágono interior. Esto es diferente del B_{19}^- (**31-1**), donde la frecuencia más baja corresponde a la rotación del fragmento pentagonal. En el caso de B_{19}^- , la barrera a la rotación del pentágono interior y el anillo B_{13} en dirección opuesta es insignificante (menos de 0.1 kcal/mol).

La sustitución de un átomo de carbono en el anillo exterior modifica esta barrera drásticamente. Para CB_{18} , este giro en la estructura de transición (**33-2**) es de 27.6 kcal/mol menos estable que la del mínimo global, evitando la rotación de los anillos interior y exterior uno respecto al otro a bajas temperaturas (**Figura 33**). Por lo tanto, a pesar de las similitudes estructurales entre B_{19}^- y CB_{18} , parece poco probable que el último pueda mostrar cualquier carácter fluxional similar al anterior, especialmente a temperaturas relativamente bajas. Simulaciones BOMD al nivel PBE/DZVP muestran el comportamiento dinámico de B_{19}^- y la ausencia de fluxionalidad en CB_{18} . Las simulaciones se iniciaron a partir de la geometrías de equilibrio, con velocidades aleatorias asignadas a los átomos y ambos sistemas se equilibraron para 300, 600 y 900 K, utilizando un baño térmico Nosé-Hoover durante 20 ps después de lo cual se calcula un conjunto de trayectorias de 50 ps. En ambos casos, para B_{19}^- y CB_{18} , durante las simulaciones BOMD la estructura planar esencialmente se mantiene, pero sin rotación interna o externa del anillo en CB_{18} incluso para 900 K.

¿Por qué un sistema que es isoelectrónico y esencialmente isoestructural a B_{19}^- exhibe un comportamiento dinámico totalmente diferente? La respuesta está en la forma en que se unen los átomos en el cúmulo. De acuerdo al análisis AdNDP,⁸⁹ en CB_{18} hay doce electrones π de valencia con dos de ellos participando en la unión de los seis átomos de boro centrales y los otros diez siendo responsables de la unión entre los $B_{12}C$ del anillo externo y el pentágono interno B_6 (**Figura 34**). Esta distribución de electrones π es análoga a la encontrada en B_{19}^- . Los 46 electrones extra participan en el enlace- σ . Sin embargo, la distribución de electrones σ es diferente en CB_{18} comparada con B_{19}^- .

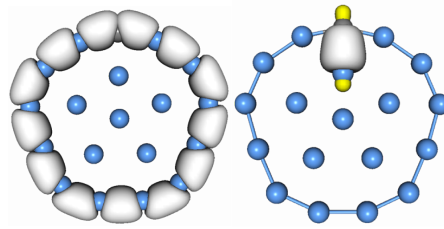
26 electrones σ están involucrados en la unión localizada y forman once enlaces B-B 2c-2e en la periferia y dos enlaces C-B 2c-2e. La diferencia radica en la presencia de un enlace C-B 2c-2e extra localizado que conecta los anillos periféricos y los anillos pentagonales internos. Esta es la razón principal para el comportamiento dinámico radicalmente diferente del cúmulo CB_{18} (**Figura 34**) en comparación con el anión de diecinueve átomos de boro.

La búsqueda de las máquinas moleculares está en curso. La reciente identificación de sistemas de anillos concéntricos de boro en el que un anillo gira con relación a otro^{209,252,254,268} proporciona un dispositivo adicional con utilidad potencial en tales máquinas. Es necesario, sin embargo, desarrollar una comprensión sistemática de cómo controlar y dirigir el movimiento de estos “motores tipo Wankel”. Este logro nos permitirá identificar las características electrónicas y geometrías clave de cúmulos de boro que son necesarias para observar la fluxionalidad y estrategias (con efectos similares) para promover la funcionalidad en otros cúmulos homonucleares o heteronucleares.

4.7 CONCLUSIONES

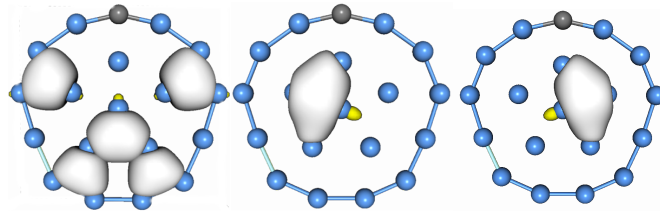
En conclusión, el B_{18}^{2-} es el quinto miembro de la familia de cúmulos tipo “motor tipo Wankel”. La unidad B_6 interior gira dentro de un anillo periférico de B_{12} con una barrera de rotación casi insignificante. El análisis del enlace químico muestra que la fracción interior B_6 y el anillo exterior se unen sólo a través de la unión 2e-multicéntrica. La ausencia de cualquier unión localizada facilita la fluxionalidad. El cúmulo B_{18}^{2-} puede considerarse como sistema doblemente concéntrico σ así como un sistema doblemente π -aromático. El análisis del campo magnético inducido confirma aún más esta doble aromaticidad. La aromaticidad en el cúmulo B_{18}^{2-} también puede justificarse utilizando el concepto de disco-aromaticidad. El cálculo de las energías electrónicas de disociación tanto vertical como adiabática revelan que es un sistema dianiónico estable. Además, el cúmulo B_{18}^{2-} en presencia de un contra-ion (Li^+), el sistema LiB_{18}^- , también es viable.

Bajo ciertas condiciones restringir la fluxionalidad en cúmulos de boro puede ser interesante. La sustitución puntual de B por un átomo de C aumenta la barrera de rotación por encima de 27 kcal/mol. El átomo de C en el CB_{18} muestra una preferencia definida para la unión en el anillo exterior en lugar de cualquiera de los seis átomos de boro del interior. La rotación, se hace imposible a bajas temperaturas. Las otras interacciones σ - y π -enlazantes en el cúmulo, que hemos identificado en este trabajo utilizando AdNDP, son casi idénticas en ambos sistemas tanto en el B_{19}^- fluxional y el rígido CB_{18} . Este análogo de B_{19}^- , con C sustituido que adopta la misma geometría básica y tiene el mismo número de electrones, promueve la localización de carga en los enlaces C–B tales que la anticipada actividad fluxional desaparece por completo. En resumen, estos resultados proporcionan una significativa pieza inicial de evidencia que la fluxionalidad en cúmulos de boro podría fácilmente desactivarse mediante una sustitución simple de C en el cúmulo, para este caso.

Enlaces- σ Localizados

34-1 $2 \times 2c-2e$ C-B NO=1.94|e|
 11 $\times 2c-2e$ B-B NO=1.95-1.83|e|

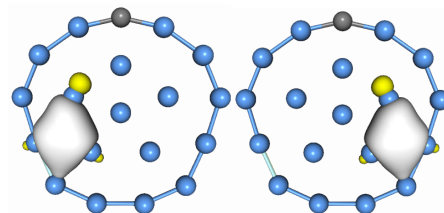
34-2 $1 \times 2c-2e$ C-B NO=1.66|e|

Enlaces- σ Deslocalizados

34-3 $5 \times 3c-2e$ B-B-B NO=1.90-1.82|e|

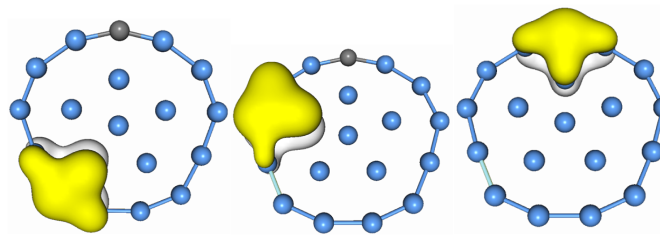
34-4 $4c-2e$ B-B-B-B NO=1.89|e|

34-5 $4c-2e$ B-B-B-B NO=1.89|e|



34-6 $4c-2e$ B-B-B-B NO=1.65|e|

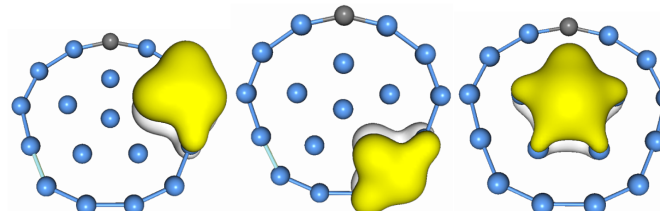
34-7 $4c-2e$ B-B-B-B NO=1.65|e|

Enlaces- π Deslocalizados

34-8 $4c-2e$ B-B-B-B NO=1.84|e|

34-9 $4c-2e$ B-B-B-B NO=1.78|e|

34-10 $4c-2e$ B-C-B-B NO=1.89|e|



34-11 $4c-2e$ B-B-B-B NO=1.78|e|

34-12 $4c-2e$ B-B-B-B NO=1.84|e|

34-13 $6c-2e$ B-B-B-B-B-B NO=1.75|e|

Figura 34: Análisis AdNDP para CB_{18} a nivel PBE0/def2-TZVP.

PERIODANO PESADO

El periodano es una molécula que contiene un átomo de cada elemento del segundo periodo de la tabla periódica a excepción del neón. No es fácil predecir la estructura más estable del LiBeBCNOF porque los potenciales arreglos no son evidentes. En 2006, Krüger sugirió que 35-1 es la forma más viable.²⁷⁰ Un año más tarde, Bera *et al.* emplearon un algoritmo de búsqueda estocástica²⁷¹ con el que encontraron 27 nuevos isómeros sustancialmente más estables que 35-1. El mínimo global 35-2 fue 122 kcal/mol más estable (al nivel B3LYP/6-31G(d)) que la estructura de Krüger. Claramente, esta molécula es difícil de imaginar en un matraz. Más allá de la estabilidad del periodano, este sistema es un buen ejemplo de la "Química sin sentido" (un término acuñado por Bera *et al.* para la búsqueda estocástica de estructuras estables donde ninguna intuición Química es aplicada en absoluto).

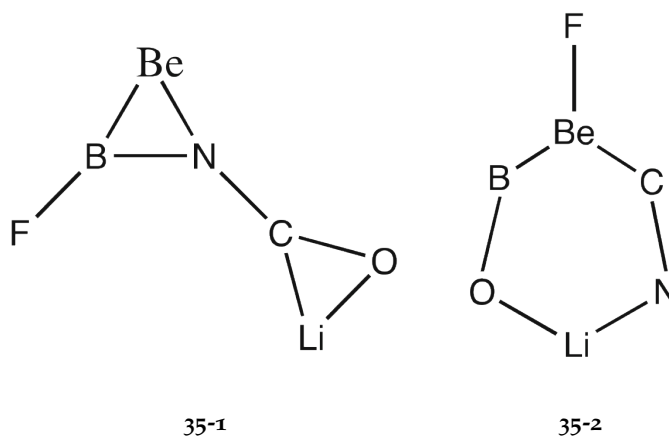


Figura 35: Estructuras reportadas previamente para el periodano pesado.

Predecir la estructura del análogo pesado del periodano, el NaMgAlSiPSCl no es una tarea fácil. A pesar de la comparación de la Química entre segunda y tercera fila de la tabla periódica, estudios tanto experimentales como teóricos ilustran diferencias estructurales entre compuestos derivados de ambas filas. Por ejemplo, la insaturación es fuertemente desfavorecida para especies de la tercera fila. Por lo tanto, predecir la forma estructural de periodano pesado, no es una tarea sencilla.

En este Capítulo, se obtiene la estructura del mínimo global del periodano pesado, empleando nuestra implementación del AG. Curiosamente, las diferencias de electronegatividad entre los elementos en la segunda y tercera fila indican que las estructuras más estables del

periodano se obtienen cuando elementos altamente electronegativos y electropositivos se unen. Pero el mínimo global del periodano es completamente diferente a su análogo de la segunda fila. Desde el punto de vista computacional, el periodano pesado también es interesante porque su PES es muy accidentada, razón por la que este sistema es atractivo para probar los límites de los algoritmos estocásticos de búsqueda.

5.1 METODOLOGÍA Y DETALLES COMPUTACIONALES

Durante la búsqueda con el AG se utilizó el funcional PBE0¹⁸⁴ y la base LANL2DZ.¹⁸⁵ Para una población inicial de 10 individuos durante 20 generaciones. La búsqueda se repitió cinco veces para asegurar que el resultado fuera reproducible. Todas las estructuras encontradas fueron reoptimizadas posteriormente con el mismo funcional, pero con la base def2-TZVP. Este mismo nivel se usó para el cálculo de frecuencias de vibración para caracterizar la naturaleza de los puntos estacionarios además de calcular la energía de punto cero.²²⁶ El funcional PBE0 proporciona resultados estructurales precisos y valores satisfactorios para propiedades termodinámicas, cinéticas y espectroscópicas en toda la tabla periódica.¹⁸⁴ Sin embargo, para garantizar que la estructura del mínimo global no es un artefacto de la metodología empleada, todas las geometrías se reoptimizaron con diferentes funcionales (BLYP, PW91, M06, y TPSSh) empleando la misma base (def2-TZVP). Todos los funcionales dieron parámetros estructurales similares. La evaluación final de las energías relativas se obtienen a nivel CCSD(T)/def2-TZVP//PBE0/def2-TZVP. Todos los cálculos de refinamiento de geometría y energía se realizaron con el programa Gaussian 09.²⁶⁶

5.2 ¿CUÁL ES EL ISÓMERO MÁS ESTABLE?

La [Tabla 5](#) reúne las diferencias de energía, incluyendo la corrección de punto cero con diferentes funcionales y la energía a nivel CCSD(T)/def2-TZVP//PBE0/def2-TZVP. En estos resultados resaltan algunas tendencias generales: (i) Los dos mínimos más estables persisten en todos los niveles de teoría. (ii) La brecha de energía entre las estructuras [36-1](#) y [36-2](#) ([Figura 36](#)) es entre 3 y 4 kcal/mol. (iii) PBE0, TPSSh y PW91 dieron diferencias de energía muy similares. Por lo tanto, el resto de la discusión, será con los resultados PBE0.

Diez mínimos locales se encuentran dentro de 15 kcal/mol del mínimo global con el funcional PBE0 y la base def2-TZVP (ver [Figura 36](#)). Es importante señalar que todas las estructuras en la [Figura 36](#) y en la [Tabla 5](#) son singlete. La estructura triplete más estable está a 18,3 kcal/mol de energía que [36-1](#) a nivel PBE0/def2-TZVP. Como en el periodano, en su congénere las formas más estables son las que involucran las

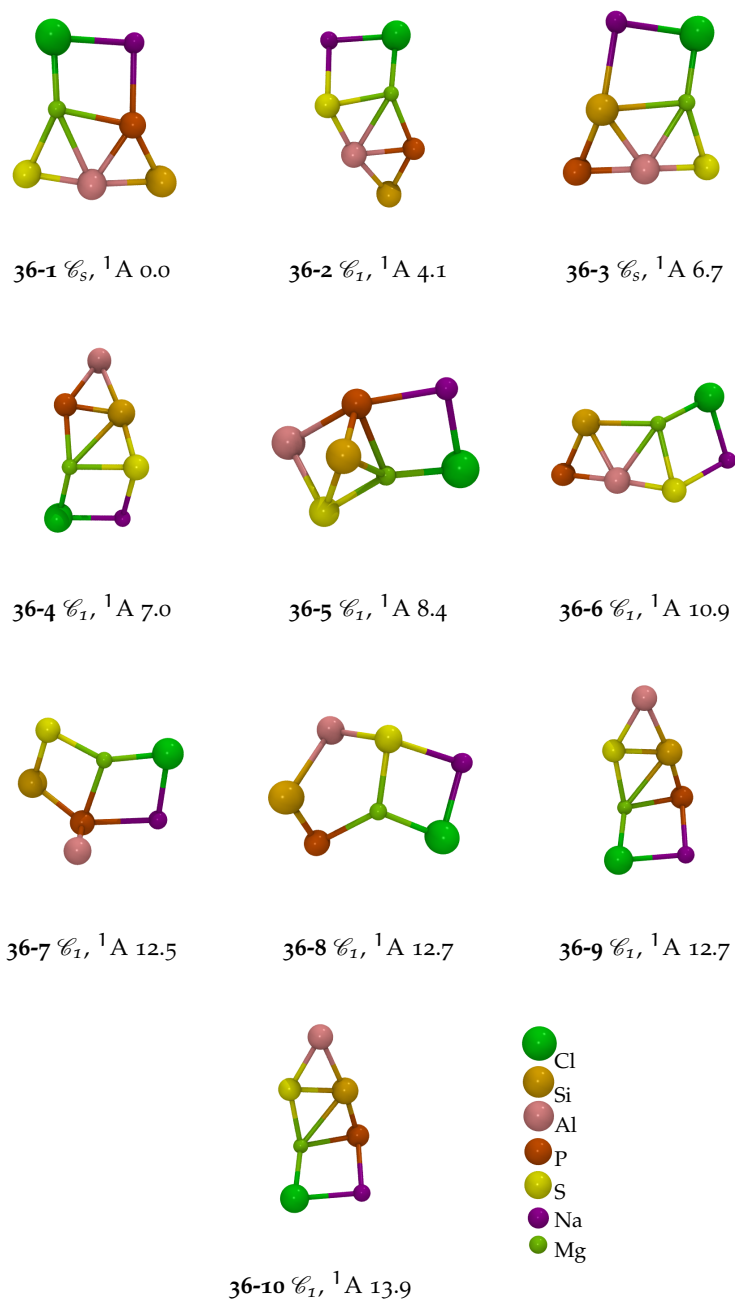


Figura 36: Isómeros optimizados del periodano pesado. Las energías relativas están en unidades de kcal/mol. Todas las estructuras fueron optimizadas al nivel PBE0/def2-TZVP.

interacciones entre átomos electronegativos y electropositivos. Los motivos estructurales más obvios son Na-Cl y Mg-Cl o una combinación de ambos (Na-Cl-Mg); de hecho, esta última tríada está presente en todos los diez mínimos. El otro motivo estructural aparente es Mg-S, que está presente en todas las estructuras de [Figura 36](#). Un fragmento que se manifiesta en cinco de los seis primeros isómeros del perio-

dano pesado es el triángulo Si-P-Al. En los 20 isómeros más estables reportados por Bera *et al.* las unidades C-N y B-O se encuentran con regularidad. Del mismo modo, los homólogos pesados del C-N y B-O, es decir, Si-P y Al-S, son recurrentes en los isómeros más estables del periodano pesado: la diada Si-P aparece entre los diez isómeros de mayor estabilidad, mientras que el motivo de Al-S emerge en ocho de ellos (36-4 y 36-7 son las únicas excepciones). Es evidente que, aunque las diferencias de electronegatividad en el tercer período no son tan grandes como las del segundo, las estructuras más estables del periodano pesado se obtienen cuando los elementos altamente electro-negativos y electropositivos se unen. Por lo tanto, algunos patrones estructurales del periodano pesado son concebibles en principio, pero todavía es difícil pensar en todos los posibles isómeros.

Nuestros resultados revelan que el isómero de más estable es de tipo planar 36-1 es completamente diferente a 35-2 (Figura 35).²⁷¹ El análisis vibracional de 36-1 muestra que el modo normal de baja energía es de 52 cm^{-1} (a nivel PBE0). Este modo representa el desplazamiento del dímero P-Mg fuera del plano molecular. Hay que tener en cuenta que 36-1 contiene un átomo de aluminio rodeado por Si-P-Mg-S, en donde el dímero Na-Cl interactúa con el P-Mg. La segunda forma más estable, 36-2, que está a 4,1 kcal/mol por encima de 36-1, tiene el mismo núcleo Si-P-Mg-S pero la unidad Na-Cl está enlazada al lado de Mg-S. El tercer isómero más estable es similar a 36-1, pero los átomos de Si y P se intercambian.

Tabla 5: Diferencias de energía de los mínimos en el periodano pesado (kcal/mol) calculadas a diferentes niveles de teoría, empleando la base def2-TZVP. La menor frecuencia vibracional (en cm^{-1}) es mostrada entre paréntesis.

Mín.	PBE0	BLYP	PW91	M06	TPSSh	CCSD(T)
36-1	0.0 (52)	0.0 (49)	0.0 (49)	0.0 (45)	0.0 (52)	0.0
36-2	4.1 (21)	3.7 (19)	4.0 (22)	3.3 (22)	4.1 (17)	3.5
36-3	6.7 (60)	8.3 (59)	6.8 (58)	8.4 (58)	6.7 (60)	7.6
36-4	7.0 (27)	5.6 (27)	6.3 (26)	6.1 (26)	7.8 (26)	7.0
36-5	8.4 (26)	8.5 (30)	7.2 (23)	7.3 (24)	9.0 (25)	8.2
36-6	10.9 (20)	11.3 (21)	10.8 (22)	11.0 (26)	11.1 (20)	11.1
36-7	12.5 (31)	7.3 (32)	11.0 (27)	^a	13.5 (31)	10.9
36-8	12.7 (18)	4.6 (17)	10.2 (18)	9.4 (28)	12.0 (20)	8.3
36-9	12.7 (16)	9.8 (10)	11.2 (13)	11.8 (16)	13.1 (15)	13.2
36-10	13.9 (18)	5.0 (19)	11.9 (17)	10.0 (27)	13.6 (19)	8.2

^a En el nivel M06/def2-TZVP el mínimo 36-7 converge a 36-5.

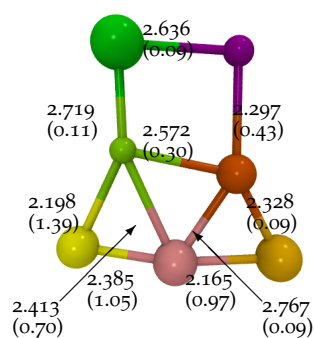
5.3 ENLACE

Los parámetros estructurales y los WBI para el isómero 36-1 se resumen en la Figura 37. Las cargas NBO de los dos elementos con un fuerte carácter catiónico (Na, 0.88 y Mg, 1.20 e^-) se intercalan entre los átomos de Cl (-0.70 e^-), P (-1.24 e^-), y S (-1.02 e^-). Los átomos restantes, Si y Al, tienen cargas positivas (+0.11 y +0.77 e^- , respectivamente). Dos fragmentos pueden diferenciarse fácilmente; el primero es un cuadrado distorsionado con P-Na-Cl-Mg que tiene un fuerte carácter iónico, mientras que el segundo es lineal (Si-Al-S) y de naturaleza covalente. La longitud del enlace Si-P (2.198 Å) es más corta que la longitud Si-P del enlace de la sililfosfina caracterizado por espectroscopía de microondas (2.25 Å).²⁷² La distancia Al-S del periodano pesado (2.165 Å) es también muy corta. Guha y Francisco reportaron una longitud de enlace Al-S de 2.249 Å en el AlSH, calculada al nivel CCSD(T)/aug-cc-pV5Z, y al que se le atribuye un carácter múltiple para la interacción Al-S.²⁷³ Por lo tanto, parece que el periodano pesado tiene enlaces múltiples en Si-P y Al-S. Claramente, los orbitales π son en su mayoría situados alrededor de las unidades P-Si y Al-S (37-2,37-3,37-4) que ratifica la naturaleza de múltiples uniones en estas interacciones. El WBI confirmó la existencia de dos unidades, una de ellas fuertemente iónica (PNaClMg) con valores WBI pequeños (<0.5) y una segunda con un fuerte carácter covalente (Si-Al-S) y valores WBI mayores de 0.97. Claramente, mientras que la interacción P-Al tiene una importante contribución covalente, el Al-Mg es más iónico (WBI=0.09).

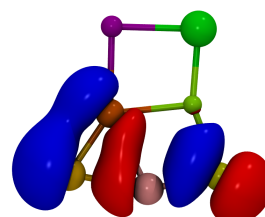
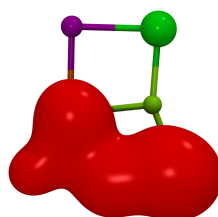
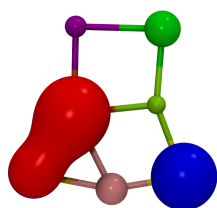
Por último, hay que tener en cuenta que la orbital HOMO-3 se deslocaliza alrededor del núcleo P-Si-Al-S, pero los parámetros geométricos indican la localización. Por lo tanto, a fin de obtener una mayor comprensión de la deslocalización electrónica en el periodano pesado, se calculó el campo magnético inducido \mathbf{B}^{ind} ^{262,264} que es un descriptor magnético de la aromaticidad. La componente z de \mathbf{B}^{ind} (B_z^{ind}) sólo indica deslocalización menor. La región más deslocalizada se encuentra en el triángulo P-Si-Al, con un valor de B_z^{ind} de -5,6 ppm.

5.4 CONCLUSIONES

Encontramos que los diez isómeros más estables del periodano que tienen patrones que implican interacciones entre átomos electronegativos y electropositivos. En principio el isómero más estable es plano, con dos fragmentos altamente reconocibles, un cuadrado fuertemente iónico PNaClMg y la triada restante de naturaleza covalente Si-Al-S. El carácter de enlace múltiple de los enlaces Si-P y Al-S es también notable.



37-1



37-2 HOMO-1

37-3 HOMO-3

37-4 HOMO-4

Figura 37: 1 Distancias de enlace (Å) e índices de Wiberg (en *paréntesis*) del periodano pesado. 37-2, 37-3, 37-4 Orbitales Moleculares del periodano pesado. Isovalor 0.03 u.a.

Parte III

APÉNDICES

SIMILITUD MOLECULAR BASADA EN LOS MOMENTOS ESTADÍSTICOS

Ultra Fast Shape Recognition (USR) es un algoritmo propuesto por Ballester,^{132,193} que usa los primeros tres momentos estadísticos de distribuciones basadas en las distancias atómicas, para comparar geometrías moleculares, tres ordenes de magnitud más rápido que las metodologías comunes.¹⁹³ También permite la identificación de la semejanza de la forma de compuestos moleculares, sin tener que llevar a cabo el requerimiento problemático de alinear moléculas, ya que este método es invariante a la orientación molecular. Este algoritmo ha sido aplicado con éxito en sistemas moleculares que van de algunas decenas de átomos hasta llegar a proteínas.^{132,193,274,275}

Los métodos de superposición son una familia de métodos comunes de comparación estructural. Un método ampliamente usado en esta categoría es ROCKS,¹⁹⁷ que calcula la similitud usando una función de *scoring* que compara la superposición de volúmenes moleculares, que sólo puede llevar a cabo, después de hacer varias iteraciones, en donde se intenta superponer las moléculas a comparar. Comparado con este método, (que es el más rápido de los métodos comerciales de comparación molecular), USR tiene una velocidad de procesamiento de 3.02¹³² veces más rápido que ROCKS. (algunos otros interesantes Benchmark pueden ser encontrados en^{132,193}).

A.1 ULTRA FAST RECOGNITION

El método propuesto por Ballester considera a la molécula como un conjunto de partículas unidas (átomos), en lugar de considerar un cuerpo sólido, USR se basa en la premisa de que la forma de una molécula está únicamente determinada por la posición relativa de sus átomos. Por lo tanto, la forma molecular se caracteriza por un conjunto de funciones de distribución unidimensionales que retienen la información de la forma tri-dimensional. USR usa la distribución de todas las distancias atómicas para obtener cuatro puntos de referencia: el centroide molecular (*ctd*), el átomo más cercano a *ctd* (*cst*), el átomo más alejado de *ctd* (*fct*) y el átomo más alejado de *fct* (*ftf*). Usar estas referencias eliminan directamente cualquier necesidad de alineamiento o traslación, ya que estas distribuciones son completamente independientes de la orientación molecular o de la posición.

Usando estos puntos de referencia se calculan 12 descriptores moleculares, basados en el cálculo de los momentos estadísticos de la distribución de posiciones atómicas para caracterizar la geometría mo-

lecular. El primer paso del cálculo es determinar el centro geométrico de la molécula usando el vector tridimensional de cada átomo. Luego se calcula el conjunto de todas las distancias euclidianas de todos los átomos al centroide molecular $(\{d_j^{ctd}\}_{j=1}^N)$, donde N es el número de átomos de la molécula considerada.

El siguiente paso del proceso es calcular los momentos de la distribución en orden de caracterizar la geometría por su forma. El primer momento (μ_1^{ctd}) es el promedio de la distancia atómica al centroide molecular, así obteniendo una estimación del tamaño molecular. El segundo momento (μ_2^{ctd}) es la varianza de estas distancias atómicas alrededor de μ_1^{ctd} . El tercer momento (μ_3^{ctd}) es la asimetría estadística de estas distancias atómicas alrededor de μ_1^{ctd} (*i.e.* una medida de la asimetría de la distribución). Para calcular los nueve restantes descriptores, sólo hay que repetir el proceso para cada una de las siguientes tres distribuciones: $\{d_j^{cst}\}_{j=1}^N$, $\{d_j^{fct}\}_{j=1}^N$ y $\{d_j^{ftf}\}_{j=1}^N$, donde *cst*, *fct* y *ftf* indican la localización desde donde son calculadas las distancias atómicas. Así de esta manera seleccionando los primeros tres momentos, de cuatro distribuciones unidimensionales:

$$\vec{M} = (\mu_1^{ctd}, \mu_2^{ctd}, \mu_3^{ctd}, \mu_1^{cst}, \mu_2^{cst}, \mu_3^{cst}, \mu_1^{fct}, \mu_2^{fct}, \mu_3^{fct}, \mu_1^{ftf}, \mu_2^{ftf}, \mu_3^{ftf})$$

Usando ese conjunto de descriptores *USR* puede mapear la forma de una molécula, con un excelente compromiso entre eficiencia y efectividad.

Usando el vector \vec{M} de cada molécula se calcula el valor de la similitud S_{ij} de dos moléculas *i*, *j* usando la siguiente ecuación:

$$S_{ij} = \left(1 + \frac{1}{12} \sum_{l=1}^{12} |M_l^i - M_l^j| \right)^{-1} \quad (18)$$

ÍNDICES TOPOLÓGICOS

Tabla 6: Valores numéricos de diferentes índices topológicos de un conjunto de mínimos del cúmulo AI_{13}^- .

Mol	W	rWf	$3-W$	Wq	Kf	Kfq	J	$3-J$	J_{Ω}
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	5.00	11.49	3.06	99.16	-8.58	-2.58	0.08	0.18	-2.84
3	9.00	21.31	6.65	129.81	-10.30	-3.35	0.06	0.09	-3.53
4	10.00	23.26	7.74	62.72	-6.86	-2.95	-0.05	-0.22	-1.98
5	18.00	43.23	10.28	172.45	-13.73	-5.12	-0.01	-0.13	-5.03
6	17.00	38.11	14.14	176.75	-13.73	-4.86	0.03	-0.06	-5.11
7	22.00	49.82	13.67	231.28	-17.16	-6.10	0.06	-0.01	-7.11
8	21.00	49.88	11.96	252.74	-17.16	-5.92	0.07	0.05	-7.22
9	31.00	68.79	12.92	338.35	-22.31	-7.73	0.15	0.13	-11.38
10	17.00	39.54	9.83	140.19	-12.01	-4.61	-0.03	-0.22	-4.13
11	25.00	55.04	14.58	235.11	-17.16	-6.19	0.03	-0.15	-7.09
12	19.00	44.45	20.90	245.36	-15.44	-5.10	0.06	0.04	-6.24
13	23.00	52.26	15.26	228.01	-17.16	-6.17	0.04	-0.08	-7.08
14	17.00	45.56	13.62	213.29	-12.01	-4.16	-0.03	-0.12	-4.37
15	21.00	51.43	20.08	190.20	-13.73	-5.18	-0.05	-0.24	-5.05
16	19.00	47.55	16.92	281.39	-17.16	-5.69	0.09	0.17	-7.36
17	11.00	28.05	15.57	160.47	-10.30	-3.40	0.02	0.01	-3.59
18	23.00	54.65	20.34	252.96	-17.16	-6.10	0.04	-0.04	-7.17
19	23.00	52.44	23.91	218.61	-15.44	-5.59	-0.00	-0.18	-6.03
20	3.00	7.92	9.25	15.56	0.00	-0.18	-0.05	-0.14	-0.01
21	25.00	58.47	21.85	296.88	-18.88	-6.43	0.08	0.05	-8.48
22	14.00	37.82	22.60	130.35	-8.58	-3.39	-0.08	-0.24	-2.75
23	19.00	44.38	26.10	190.66	-13.73	-4.92	-0.00	-0.14	-5.12
24	19.00	45.70	24.36	268.68	-17.16	-5.58	0.10	0.17	-7.34
25	20.00	50.35	16.96	221.00	-15.44	-5.59	0.00	-0.08	-6.07
26	21.00	51.49	21.08	255.90	-17.16	-6.02	0.06	0.05	-7.21
27	26.00	64.28	22.35	308.93	-18.88	-6.59	0.05	0.01	-8.47
28	26.00	65.14	19.68	360.56	-20.59	-7.00	0.11	0.20	-9.98
29	25.00	56.93	24.21	285.38	-18.88	-6.55	0.08	0.03	-8.43

-Continúa de la página anterior

Mol	W	rWf	3-W	Wq	Kf	Kfq	J	3-J	J_{Ω}
30	16.00	43.15	19.62	196.30	-12.01	-4.20	-0.03	-0.10	-4.33
31	22.00	51.82	25.00	244.24	-15.44	-5.42	0.01	-0.12	-6.14
32	18.00	44.53	18.99	169.26	-13.73	-5.18	-0.02	-0.14	-5.01
33	17.00	42.23	20.64	200.87	-13.73	-4.78	0.02	-0.04	-5.17
34	22.00	54.22	28.22	274.13	-17.16	-5.90	0.05	0.02	-7.27
35	18.00	44.12	27.41	219.51	-13.73	-4.68	0.01	-0.06	-5.23
36	24.00	55.44	24.78	262.80	-17.16	-6.03	0.03	-0.08	-7.21
37	16.00	41.04	19.70	176.06	-10.30	-3.71	-0.06	-0.21	-3.55
38	19.00	48.71	23.01	240.25	-15.44	-5.39	0.03	0.00	-6.16
39	26.00	62.69	25.71	311.48	-18.88	-6.50	0.06	0.01	-8.49
40	25.00	60.27	31.84	259.21	-15.44	-5.54	-0.03	-0.22	-6.12
41	22.00	53.72	44.26	166.57	-10.30	-4.17	-0.13	-0.47	-3.41
42	20.00	49.34	23.74	240.44	-15.44	-5.40	0.02	-0.04	-6.16
43	25.00	61.08	27.45	271.29	-17.16	-6.12	0.00	-0.12	-7.19
44	23.00	57.01	30.75	292.89	-17.16	-5.88	0.04	-0.01	-7.32
45	18.00	45.74	28.78	204.58	-13.73	-4.98	-0.01	-0.10	-5.14
46	31.00	73.34	38.40	300.39	-18.88	-6.84	-0.01	-0.21	-8.36
47	23.00	55.77	35.34	262.57	-15.44	-5.33	-0.01	-0.14	-6.22
48	13.00	34.63	29.24	119.29	-6.86	-2.73	-0.09	-0.28	-2.13
49	24.00	60.41	21.35	351.87	-18.88	-6.31	0.08	0.12	-8.67
50	21.00	53.73	23.87	292.30	-17.16	-5.84	0.05	0.07	-7.34
51	21.00	53.63	34.02	249.02	-15.44	-5.50	-0.00	-0.08	-6.14
52	20.00	48.88	23.01	212.69	-15.44	-5.64	0.01	-0.07	-6.04
53	23.00	61.14	30.42	378.21	-17.16	-5.52	0.04	0.05	-7.55
54	22.00	53.43	28.33	272.92	-17.16	-5.91	0.05	0.02	-7.28
55	29.00	68.93	38.28	347.61	-18.88	-6.50	0.03	-0.09	-8.56
56	34.00	87.48	35.30	362.49	-18.88	-6.88	-0.07	-0.29	-8.47
57	26.00	63.45	39.72	338.74	-18.88	-6.42	0.06	0.03	-8.58
58	33.00	79.12	44.68	384.09	-20.59	-7.13	0.04	-0.07	-9.95
59	23.00	57.85	34.65	253.36	-15.44	-5.49	-0.02	-0.15	-6.15
60	26.00	62.88	30.06	337.68	-18.88	-6.38	0.07	0.04	-8.60
61	21.00	51.50	29.39	305.52	-15.44	-5.09	0.02	-0.03	-6.37
62	24.00	60.30	29.80	284.97	-17.16	-6.06	0.01	-0.07	-7.26
63	36.00	85.64	50.41	402.46	-22.31	-7.79	0.08	0.00	-11.46
64	30.00	72.09	49.57	340.19	-18.88	-6.54	0.02	-0.12	-8.53
65	35.00	86.67	44.70	419.51	-22.31	-7.77	0.08	0.06	-11.52

-Continúa de la página anterior

Mol	W	rWf	$3-W$	Wq	Kf	Kfq	J	$3-J$	J_{Ω}
66	29.00	75.49	34.08	342.42	-18.88	-6.70	-0.01	-0.10	-8.51
67	36.00	93.58	32.43	555.09	-24.02	-7.93	0.14	0.29	-13.81
68	42.00	100.90	39.42	533.35	-25.74	-8.74	0.17	0.25	-15.77
69	34.00	83.13	36.86	417.97	-22.31	-7.67	0.09	0.07	-11.59
70	28.00	69.99	38.80	324.45	-18.88	-6.67	0.02	-0.07	-8.49
71	36.00	81.24	85.80	379.62	-20.59	-7.04	0.04	-0.18	-9.91
72	24.00	60.10	39.89	258.87	-15.44	-5.57	-0.04	-0.20	-6.14
73	27.00	69.56	41.21	359.19	-18.88	-6.49	0.03	0.00	-8.60
74	31.00	78.90	47.74	305.39	-17.16	-6.32	-0.08	-0.33	-7.17
75	33.00	81.72	40.15	482.93	-22.31	-7.41	0.11	0.18	-11.85
76	39.00	91.73	98.54	415.53	-20.59	-7.12	-0.01	-0.27	-9.95
77	41.00	95.50	101.11	416.18	-20.59	-7.19	-0.03	-0.35	-9.92
78	45.00	103.71	91.96	454.79	-24.02	-8.41	0.06	-0.13	-13.31
79	41.00	98.18	77.03	393.29	-20.59	-7.41	-0.05	-0.37	-9.84
80	40.00	92.74	93.62	498.99	-22.31	-7.42	0.06	-0.10	-11.77
81	45.00	104.61	97.28	574.32	-25.74	-8.58	0.17	0.18	-15.92
82	31.00	77.15	44.77	411.36	-22.31	-7.57	0.13	0.21	-11.59
83	49.00	121.60	73.54	634.92	-25.74	-8.77	0.09	0.05	-15.93
84	44.00	101.25	97.75	510.38	-24.02	-8.12	0.10	-0.04	-13.54
85	34.00	72.05	79.95	332.28	-18.88	-6.39	0.02	-0.28	-8.48
86	55.00	135.42	32.92	844.11	-30.89	-10.11	0.43	0.94	-26.29
87	82.00	221.28	46.34	1148.57	-34.32	-11.65	0.46	1.18	-38.42
88	105.00	292.89	70.11	1713.04	-37.75	-12.55	0.94	2.62	-64.12
89	82.00	191.20	46.74	1139.82	-32.60	-10.55	0.40	0.61	-31.86
90	32.00	51.15	20.33	372.60	-22.31	-6.84	0.28	0.16	-11.66

Tabla 7: Valores numéricos de diferentes índices topológicos de un conjunto de mínimos del cúmulo C_6H_6 .

Mol	W	rWf	$3-W$	Wq	Kf	Kfq	J	$3-J$	J_{Ω}
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	3.00	5.38	6.39	26.20	0.00	-0.09	-0.19	-0.14	0.00
3	14.00	20.35	13.63	85.99	0.00	-0.34	-0.81	-0.92	0.00
4	52.00	65.56	59.45	308.64	-1.69	-1.51	4.63	5.84	-119.90
5	52.00	65.57	56.54	308.67	-1.69	-1.51	4.63	5.84	-119.90
6	36.00	46.29	40.14	237.04	-1.69	-1.32	5.82	7.47	-119.90

-Continúa de la página anterior

Mol	W	rWf	3-W	Wq	Kf	Kfq	J	3-J	J_{Ω}
7	36.00	46.45	29.84	237.62	-1.69	-1.33	5.81	7.47	-119.90
8	28.00	36.19	21.66	205.06	-1.69	-1.26	6.46	8.34	-119.90
9	62.00	77.85	86.36	376.73	-1.69	-1.44	3.96	4.94	-121.58
10	36.00	46.65	34.45	238.38	-1.69	-1.33	5.80	7.47	-119.90
11	48.00	59.75	55.79	286.43	-1.69	-1.46	5.00	6.26	-119.90
12	28.00	36.46	24.28	206.07	-1.69	-1.27	6.44	8.34	-119.90
13	48.00	60.04	43.11	287.51	-1.69	-1.46	4.98	6.26	-119.90
14	18.00	25.10	23.53	107.47	-0.00	-0.31	-1.00	-1.19	-0.39
15	46.00	57.58	53.03	278.39	-1.69	-1.46	5.13	6.47	-119.90
16	54.00	68.06	64.95	327.64	-1.69	-1.46	4.50	5.67	-120.74
17	-9.00	-6.55	-4.73	-51.29	1.69	0.97	-2.72	-3.25	39.64
18	54.00	67.00	65.85	321.85	-1.69	-1.42	4.57	5.69	-120.74
19	46.00	57.79	36.90	279.20	-1.69	-1.46	5.12	6.47	-119.90
20	-20.00	-19.45	-10.80	-125.78	3.39	2.07	-3.87	-4.69	59.59
21	-8.00	-3.42	-7.28	-40.74	1.69	0.87	-2.84	-3.33	39.64
22	-2.00	0.15	-4.92	6.06	-0.00	0.02	-0.01	0.16	-0.39
23	30.00	38.47	31.83	221.12	-1.69	-1.21	6.31	8.12	-120.74
24	6.00	12.24	9.04	71.01	-0.00	-0.06	-0.44	-0.30	-0.79
25	21.00	27.26	28.45	117.53	0.00	-0.39	-1.05	-1.34	0.00
26	19.00	25.47	25.88	121.56	-0.00	-0.17	-0.97	-1.21	-0.79
27	22.00	32.22	27.61	201.34	-1.69	-1.02	6.80	9.10	-121.58
28	31.00	38.67	46.56	167.70	-0.00	-0.33	-1.49	-1.99	-0.79
29	25.00	34.53	25.89	137.72	0.00	-0.47	-1.35	-1.63	0.00
30	22.00	32.24	20.05	201.43	-1.69	-1.02	6.80	9.10	-121.58
31	25.00	31.66	35.66	138.03	-0.00	-0.36	-1.22	-1.60	-0.39
32	13.00	18.72	19.05	101.55	-0.00	-0.11	-0.68	-0.77	-0.79
33	-2.00	1.00	-4.61	17.80	-0.00	0.07	-0.05	0.16	-0.79
34	25.00	34.69	32.92	138.26	0.00	-0.48	-1.35	-1.63	0.00
35	18.00	26.39	20.10	184.10	-1.69	-0.99	7.18	9.57	-121.58
36	25.00	35.19	31.67	129.30	0.00	-0.51	-1.39	-1.67	0.00
37	-3.00	0.12	-3.66	4.63	-0.00	0.02	-0.01	0.24	-0.39
38	25.00	35.42	20.60	130.11	0.00	-0.51	-1.40	-1.67	0.00
39	-1.00	3.93	4.85	-9.29	1.69	0.79	-3.06	-3.71	39.64
40	-12.00	-9.88	-6.95	-67.68	1.69	1.00	-2.61	-3.07	39.64
41	0.00	4.62	1.05	29.25	-0.00	-0.06	-0.16	0.09	-0.39
42	1.00	6.91	3.43	36.97	-0.00	-0.08	-0.26	0.02	-0.39

-Continúa de la página anterior

Mol	W	rWf	$3-W$	Wq	Kf	Kfq	J	$3-J$	J_{Ω}
43	18.00	28.92	17.08	110.90	0.00	-0.46	-1.16	-1.21	0.00
44	-6.00	-3.03	-9.07	-1.66	-0.00	0.13	0.14	0.49	-0.79
45	36.00	48.60	55.06	199.90	-0.00	-0.50	-1.88	-2.35	-0.79
46	14.00	21.60	20.45	58.44	1.69	0.57	-3.58	-4.48	39.64
47	18.00	25.86	30.65	125.42	-0.00	-0.21	-0.98	-1.13	-0.79
48	-6.00	-2.16	0.90	-36.82	1.69	0.92	-2.88	-3.45	39.64
49	28.00	38.17	32.50	155.38	-0.00	-0.46	-1.49	-1.83	-0.39
50	-12.00	-9.39	-9.51	13.44	-0.00	0.35	0.51	1.09	-1.41
51	-3.00	1.13	-0.84	21.72	-0.00	0.03	0.03	0.35	-0.39
52	4.00	10.47	6.84	52.25	-0.00	-0.13	-0.42	-0.22	-0.39
53	-11.00	-8.14	-8.36	19.52	-0.00	0.32	0.44	0.99	-1.41
54	16.00	26.42	23.70	127.09	-0.00	-0.23	-1.01	-1.01	-0.79
55	0.00	3.29	-8.71	11.84	0.00	-0.11	-0.16	0.00	0.00
56	-5.00	0.56	0.80	-54.00	3.39	1.69	-4.46	-5.48	59.59
57	-13.00	-6.57	-12.42	-73.56	3.39	1.76	-4.25	-5.05	59.59
58	-2.00	5.37	3.05	-10.44	1.69	0.72	-3.15	-3.72	39.64
59	27.00	37.62	24.59	239.23	-1.69	-1.16	6.33	8.35	-121.27
60	8.00	14.39	23.30	20.11	1.69	0.66	-3.41	-4.24	39.64
61	12.00	20.40	10.66	79.47	-0.00	-0.32	-0.82	-0.80	-0.39
62	27.00	37.91	15.53	240.41	-1.69	-1.17	6.31	8.35	-121.27
63	-24.00	-21.43	-16.51	-161.34	5.08	2.91	-4.81	-5.80	71.65
64	2.00	8.03	14.80	7.48	1.69	0.74	-3.19	-3.87	39.64
65	0.00	8.42	-0.86	45.46	-0.00	-0.10	-0.40	0.00	-0.79
66	22.00	35.31	24.64	145.09	-0.00	-0.42	-1.39	-1.46	-0.39
67	6.00	18.38	1.00	21.14	3.39	1.36	-4.88	-5.94	59.59
68	27.00	39.71	14.04	246.84	-1.69	-1.20	6.15	8.35	-121.27
69	10.00	19.04	25.71	40.34	1.69	0.55	-3.56	-4.36	39.64
70	-10.00	0.59	-11.60	-81.53	5.08	2.46	-5.36	-6.45	71.65
71	-11.00	-6.17	-5.87	-75.87	3.39	1.81	-4.27	-5.17	59.59
72	10.00	19.09	24.87	40.49	1.69	0.55	-3.57	-4.36	39.64
73	5.00	15.21	13.99	25.23	1.69	0.60	-3.45	-4.09	39.64
74	-22.00	-19.25	-18.73	-121.56	3.39	2.06	-3.89	-4.59	59.42
75	5.00	15.38	16.86	25.79	1.69	0.59	-3.45	-4.09	39.64
76	-9.00	-2.72	-3.61	-101.65	5.08	2.46	-5.30	-6.52	71.65
77	5.00	15.32	20.82	28.05	1.69	0.61	-3.44	-4.07	39.64
78	2.00	7.58	11.71	-29.26	3.39	1.59	-4.65	-5.82	59.59

-Continúa de la página anterior

Mol	W	rWf	$3-W$	Wq	Kf	Kfq	J	$3-J$	J_{Ω}
79	-4.00	6.01	0.02	-12.56	1.69	0.68	-3.18	-3.59	39.64
80	12.00	20.89	18.98	137.26	-0.00	-0.07	-0.84	-0.81	-1.41
81	0.00	6.62	7.83	-27.82	3.39	1.59	-4.61	-5.70	59.59
82	-18.00	-12.47	-14.54	-128.98	5.08	2.67	-5.05	-6.09	71.65
83	-11.00	-3.80	-4.78	-99.90	5.08	2.60	-5.27	-6.43	71.55
84	-16.00	-9.55	-15.93	-56.40	1.69	1.00	-2.62	-2.78	39.39
85	-3.00	7.92	2.59	-55.48	5.08	2.34	-5.52	-6.75	71.65
86	-14.00	-8.43	-7.57	-119.88	5.08	2.65	-5.15	-6.30	71.55
87	-3.00	9.00	-2.88	-64.81	5.08	2.28	-5.56	-6.78	71.65
88	-12.00	-5.05	-9.19	-70.74	3.39	1.80	-4.31	-5.14	59.42
89	-8.00	-0.12	-3.22	-29.26	1.69	0.77	-2.96	-3.33	39.64
90	-10.00	-1.25	-10.65	-63.54	3.39	1.73	-4.41	-5.23	59.59

REFERENCIAS

- [1] Knuth, D. E. *Comm. ACM* **1974**, *17*, 667–673.
- [2] García Márquez, G. *Cien Años de Soledad*; Diana: México D.F., 1967.
- [3] Lucrecio, T. In *De rerum Natura*; Fiol, E. V., Ed.; Editorial Acantilado, 2012; Vol. V; liber Quitus.
- [4] Hendrickson, B. *SIAM J. Optim.* **1995**, *5*, 835–857.
- [5] Suma, K.; Sumiyoshi, Y.; Endo, Y. *Science* **2006**, *311*, 1278–1281.
- [6] Tjalling, J. *SIAM Rev.* **1995**, *37*, 531–551.
- [7] Deuflhard, P. *Newton methods for nonlinear problems: affine invariance and adaptive algorithms*; Springer Science & Business Media: Berlin, 2011; Vol. 35.
- [8] Isaacson, E. *SIAM Review* **1988**, *30*, 331–332.
- [9] Darwin, C.; de Zulueta, A. *El origen de las especies por medio de la selección natural*; Editorial CSIC-CSIC Press, 2009.
- [10] Darwin, C.; Stauffer, R. C. *Charles Darwin's natural selection: being the second part of his big species book written from 1856 to 1858*; Cambridge University Press: Cambridge, 1987.
- [11] Holland, J. In *Adaptation in Natural and Artificial Systems*; Arbor, A., Ed.; University of Michigan Press.: Michigan, 1975.
- [12] George, A.; Rajakumar, B. R.; Binu, D. *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*; ICACCI '12; ACM: New York, NY, USA, 2012; pp 174–179.
- [13] Patrascu, M. *Proceedings of the Institution of Mechanical Engineers, Part I: J. Sys. Control Eng.* **2015**, *229*, 158–168.
- [14] Parker, J. R. *Algorithms for image processing and computer vision*; John Wiley & Sons, 2010.
- [15] Maimon, O. Z.; Braha, D. *Int. J Prod. Research* **1998**, *36*, 761–784.
- [16] Wong, K.-C.; Leung, K.-S.; Wong, M.-H. *Lecture Notes in Computer Science*; Springer Berlin Heidelberg: Berlin, 2010; Vol. 6024; pp 481–490.

- [17] Prados, D. L. *Southeastcon'92, Proceedings., IEEE*; IEEE press: New York, 1992; pp 634–637.
- [18] Bauer, R. J. *Genetic algorithms and investment strategies*; John Wiley & Sons: New York, 1994; Vol. 19.
- [19] Garg, P. *Int. J. Network Sec. App.* **2010**, *1*, 34–42.
- [20] Delman, B. *Genetic algorithms in cryptography*; PhD Thesis. Rochester Institute of Technology: Rochester, 2004.
- [21] Chellapilla, K.; Fogel, D. B. *Proc. IEEE* **1999**, *87*, 1471–1496.
- [22] Van Eck, R. *Games and simulations in online learning: Research and development frameworks* **2007**, 271–307.
- [23] Gondro, C.; Kinghorn, B. *Genet. Mol. Res.* **2007**, *6*, 964–982.
- [24] Srinivas, N.; Deb, K. *Evol. Comput.* **1994**, *2*, 221–248.
- [25] Rein, G.; Lautenberger, C.; Fernandez-Pello, A. C.; Torero, J. L.; Urban, D. L. *Combust. Flame* **2006**, *146*, 95–108.
- [26] Edwards, K.; Edgar, T.; Manousiouthakis, V. *Comp. Chem. Eng.* **1998**, *22*, 239–246.
- [27] Yan, X.; Chen, D.; Hu, S.; Ding, J. J. *Chem. Ind. Eng -CHINA-* **2002**, *53*, 810–814.
- [28] Saha, B.; Reddy, P. K.; Ghoshal, A. *Chem. Eng. J.* **2008**, *138*, 20–29.
- [29] Gao, G.; Luo, J. *Comp. App. Chem.* **2005**, *22*, 915.
- [30] Elliott, L.; Ingham, D.; Kyne, A.; Mera, N.; Pourkashanian, M.; Wilson, C. *Prog. En. Comb. Sci.* **2004**, *30*, 297–328.
- [31] Balland, L.; Estel, L.; Cosmao, J.-M.; Mouhab, N. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 121–135.
- [32] Polifke, W.; Geng, W.; Döbbeling, K. *Combust. Flame* **1998**, *113*, 119–134.
- [33] Lucasius, C. B.; Kateman, G. *Proceedings of the third international conference on Genetic algorithms*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, 1989; pp 170–176.
- [34] Blaney, J. M.; Dixon, J. S. *Reviews in Computational Chemistry, Volume 5* **2007**, *5*, 299–335.
- [35] May, A. C.; Johnson, M. S. *Protein Eng.* **1994**, *7*, 475–485.
- [36] Brown, R. D.; Jones, G.; Willett, P.; Glen, R. C. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 63–70.

- [37] Brown, R. D.; Downs, G. M.; Jones, G.; Willett, P. J. *Chem. Inf. Comput. Sci.* **1994**, *34*, 47–53.
- [38] Payne, A.; Glen, R. C. *J. Mol. Graph.* **1993**, *11*, 74–91.
- [39] Fontain, E. J. *Chem. Inf. Comput. Sci.* **1992**, *32*, 748–752.
- [40] Blaney, J.; Crippen, G.; Dearing, A.; Dixon, J. *Quantum Chemistry Program Exchange* **1990**, *590*, 1.
- [41] Shenkin, P. S.; Yarmush, D. L.; Fine, R. M.; Wang, H.; Levinthal, C. *Biopolymers* **1987**, *26*, 2053–2085.
- [42] Dolata, D. P.; Walters, W. P. *J. Mol. Graph.* **1993**, *11*, 106–111.
- [43] Devillers, J. *Genetic algorithms in molecular modeling*; Academic Press, 1996.
- [44] Mannhold, R.; Kubinyi, H.; Timmerman, H.; Clark, D. E. *Evolutionary algorithms in molecular design*; John Wiley & Sons, 2008; Vol. 8.
- [45] Niculescu, S. P. *J. Mol. Struct. THEOCHEM* **2003**, *622*, 71–83.
- [46] Hoffman, B.; Cho, S. J.; Zheng, W.; Wyrick, S.; Nichols, D. E.; Mailman, R. B.; Tropsha, A. *J. Med. Chem.* **1999**, *42*, 3217–3226.
- [47] Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
- [48] Kubiny, H. *Quant. Struct.-Act. Rel.* **1994**, *13*, 285–294.
- [49] Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231–236.
- [50] Lucasius, C. B.; Blommers, M. J.; Buydens, L. M.; Kateman, G. *Davis* **1991**, *480*, 251–281.
- [51] Unger, R.; Moult, J. *J. Mol. Biol.* **1993**, *231*, 75–81.
- [52] Tuffery, P.; Etchebest, C.; Hazout, S.; Lavery, R. *J. Biomol. Struct. Dyn.* **1991**, *8*, 1267–1289.
- [53] Tuffery, P.; Etchebest, C.; Hazout, S.; Lavery, R. *J. Comput. Chem.* **1993**, *14*, 790–798.
- [54] Pedersen, J. T.; Moult, J. *Curr. Opin. Struct. Biol.* **1996**, *6*, 227–231.
- [55] Sun, S. *Protein Sci.* **1993**, *2*, 762–785.
- [56] Schulze-Kremer, S. *Machine Learning: ECML-93*; Springer: New York, 1993; pp 262–279.
- [57] Pedersen, J. T.; Moult, J. *Proteins: Struct., Funct., Bioinf.* **1995**, *23*, 454–460.

- [58] Dandekar, T.; Argos, P. *Protein Eng.* **1992**, *5*, 637–645.
- [59] Pedersen, J. T.; Moult, J. J. *Mol. Biol.* **1997**, *269*, 240–259.
- [60] König, R.; Dandekar, T. *BioSystems* **1999**, *50*, 17–25.
- [61] Judson, R. *Proc. Pacific Symp. Biocomputing; Citeseer*, 1996; pp 423–437.
- [62] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. J. *Mol. Biol.* **1997**, *267*, 727–748.
- [63] Dixon, J. *Trends in QSAR Mol. Mod.* **1993**, *92*, 412–413.
- [64] Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269–288.
- [65] Judson, R. S.; Jaeger, E.; Treasurywala, A. M. *J. Mol. Struc. THEO-CHEM* **1994**, *308*, 191–206.
- [66] Judson, R. S.; Tan, Y.; Mori, E.; Melius, C.; Jaeger, E. P.; Treasurywala, A. M.; Mathiowetz, A. J. *Comput. Chem.* **1995**, *16*, 1405–1419.
- [67] Blommers, M. J.; Lucasius, C. B.; Kateman, G.; Kaptein, R. *Biopolymers* **1992**, *32*, 45–52.
- [68] Chang, G.; Lewis, M. *Acta Cryst. D: Bio. Cryst.* **1994**, *50*, 667–674.
- [69] Buerger, M. J. *Contemporary crystallography*; McGraw-Hill: New York, 1970.
- [70] Judson, R.; Jaeger, E.; Treasurywala, A.; Peterson, M. J. *Comput. Chem.* **1993**, *14*, 1407–1414.
- [71] Xiao, Y.; Williams, D. E. *Chem. Phys. Lett.* **1993**, *215*, 17–24.
- [72] Xiao, Y.; Williams, D. *Comp. & Chem.* **1994**, *18*, 199–201.
- [73] Hartke, B. J. *Phys. Chem.* **1993**, *97*, 9973–9976.
- [74] Smith, R. W. *Comput. Phys. Commun.* **1992**, *71*, 134–146.
- [75] Mestres, J.; Scuseria, G. E. *J. Comput. Chem.* **1995**, *16*, 729–742.
- [76] Deaven, D. M.; Ho, K. M. *Phys. Rev. Lett.* **1995**, *75*, 288–291.
- [77] Gregurick, S. K.; Alexander, M. H.; Hartke, B. J. *Chem. Phys.* **1996**, *104*, 2684–2691.
- [78] Niesse, J. A.; Mayne, H. R. *J. Chem. Phys.* **1996**, *105*, 4700–4706.
- [79] Johnston, R. L. *Dalton Trans.* **2003**, 4193–4207.
- [80] Alexandrova, A. N.; Boldyrev, A. I.; Fu, Y. J.; Yang, X.; Wang, X. B.; Wang, L. S. *J. Chem. Phys.* **2004**, *121*, 5709–5719.

- [81] Alexandrova, A. N.; Boldyrev, A. I. *J. Chem. Theory Comput.* **2005**, *1*, 566–580.
- [82] Zhang, Q.; Cheng, L. *J. Chem. Inf. Model.* **2015**, *55*, 1012–1020.
- [83] Zhang, X.; Liu, G.; Ganteför, G.; Bowen, K. H.; Alexandrova, A. N. *J. Phys. Chem. Lett.* **2014**, *5*, 1596–1601.
- [84] Shen, L.; Dadras, J.; Alexandrova, A. N. *Phys. Chem. Chem. Phys.* **2014**, *16*, 26436–26442.
- [85] Alexandrova, A. N.; Nechay, M. R.; Lydon, B. R.; Buchan, D. P.; Yeh, A. J.; Tai, M.-H.; Kostrikin, I. P.; Gabrielyan, L. *Chem. Phys. Lett.* **2013**, *588*, 37–42.
- [86] Alexandrova, A. N. *Chem. Phys. Lett.* **2012**, *533*, 1–5.
- [87] Alexandrova, A. N.; Nayhouse, M. J.; Huynh, M. T.; Kuo, J. L.; Melkonian, A. V.; Chavez, G.; Hernando, N. M.; Kowal, M. D.; Liu, C.-P. *Phys. Chem. Chem. Phys.* **2012**, *14*, 14815–14821.
- [88] Huynh, M. T.; Alexandrova, A. N. *J. Phys. Chem. Lett.* **2011**, *2*, 2046–2051.
- [89] Zubarev, D. Y.; Boldyrev, A. I. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5207–5217.
- [90] Zhai, H.; Averkiev, B. B.; Zubarev, D. Y.; Boldyrev, A. I. *Angew. Chem., Int. Ed.* **2007**, *46*, 4277–4280.
- [91] Alexandrova, A. N.; Boldyrev, A. I.; Zhai, H.-J.; Wang, L.-S. *Coord. Chem. Rev.* **2006**, *250*, 2811–2866.
- [92] Shao jian Sun, G.; Hong, W.; Wang, T.; Wen, Y. *IEICE Trans. Inf. and Sys.* **2013**, *E96.D*, 2726–2732.
- [93] Cai, Z.; Cao, Q.; Du, S.; Ji, X. *Natural Computation (ICNC), 2012 Eighth International Conference on Natural Computation*; 2012; pp 1247–1250.
- [94] Dessens-Félix, M.; Pacheco-Contreras, R.; Barcaro, G.; Sementa, L.; Fortunelli, A.; Posada-Amarillas, A. *J. Phys. Chem. C* **2013**, *117*, 20967–20974.
- [95] Hong, L.; Wang, H.; Cheng, J.; Huang, X.; Sai, L.; Zhao, J. *Comp. Theor. Chem.* **2012**, *993*, 36–44.
- [96] Li, R.; Cheng, L. *Comp. Theor. Chem.* **2012**, *996*, 125–131.
- [97] Yan, S.-Y.; Zhang, W.; Zhao, Z.-X.; Lu, W.-C.; Zhang, H.-X. *Theor. Chem. Acc.* **2012**, *131*, 1–8.
- [98] Alexandrova, A. N. *J. Phys. Chem. A* **2010**, *114*, 12591–12599.

- [99] Buck, U.; Pradzynski, C. C.; Zeuch, T.; Dieterich, J. M.; Hartke, B. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6859–6871.
- [100] Kabrede, H.; Hentschke, R. *J. Phys. Chem. B* **2003**, *107*, 3914–3920.
- [101] Cai, X.; McKinney, D. C.; Lasdon, L. S. *Adv. Water Res.* **2001**, *24*, 667–676.
- [102] Niesse, J. A.; Mayne, H. R. *J. Comput. Chem.* **1997**, *18*, 1233–1244.
- [103] Savic, D. A.; Walters, G. A. *J. water res. plan. and manag.* **1997**, *123*, 67–77.
- [104] Kanters, R. P. F.; Donald, K. J. *J. Chem. Theory Comput.* **2014**, *10*, 5729–5737.
- [105] Lewis, P.; Goodman, A.; Miller, J. *IBM System Journal* **1969**, *8*, 136–143.
- [106] Bennet, F. *ACM Trans. Math. Soft.* **1986**, *12*, 362–376.
- [107] Bratley, P.; Fox, B.; Schrage, L. *A Guide to Simulation*; Springer Science & Business Media: New York, 1987.
- [108] Pierre, L. In *Random Number Generation, in Handbook of Simulation*; Jerry, B., Ed.; Jhon Wiley and Sons: Hoboken, NJ, 1998.
- [109] Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley Longman Publishing Co., Inc., 1989.
- [110] Levin, M. In *Practical Handbook of Genetic Algorithms, Complex Coding Systems*; Chambers, L., Ed.; CRC Press: New York, 1999; Vol. 3.
- [111] Holland, J. H. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*; MIT press, 1992.
- [112] Kumar, R. *Int. J. Mac. Learn. Comp.* **2012**, *2*, 365.
- [113] Deaven, D.; Tit, N.; Morris, J.; Ho, K. *Chem. Phys. Lett.* **1996**, *256*, 195–200.
- [114] Walsh, A.; Woodley, S. M. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8446–8453.
- [115] Chen, Z.; Jiang, X.; Li, J.; Li, S. *J. Chem. Phys.* **2013**, *138*, 214303.
- [116] Margulis, L. *Una revolución en la evolución*; Universitat de València: Valencia, España, 2002; Vol. 20.
- [117] Lloyd, L. D.; Johnston, R. L. *Chem. Phys.* **1998**, *236*, 107–121.

- [118] Saunders, M. J. *Comput. Chem.* **2004**, *25*, 621–626.
- [119] Wheeler, S. E.; Schleyer, P. v. R.; Schaefer, H. F. J. *Chem. Phys.* **2007**, *126*.
- [120] Avaltroni, F.; Corminboeuf, C. J. *Comput. Chem.* **2012**, *33*, 502–508.
- [121] Harary, F. *Graph Theory*; Addison–Wesley: New York, 1971.
- [122] Cvetković, D. M.; Doob, M.; Sachs, H. *Spectra of graphs: theory and application*; Academic Press: London, 1980; Vol. 87.
- [123] Gutman, I.; Trinajstić, N. *Chem. Phys. Lett.* **1972**, *17*, 535 – 538.
- [124] Graovac, A.; Gutman, I.; Trinajstić, N.; Živković, T. *Theor Chim Acta* **1972**, *26*, 67–78.
- [125] Gutman, I.; Trinajstić, N. *New Concepts II*; Topics in Current Chemistry Fortschritte der Chemischen Forschung; Springer Berlin Heidelberg: Berlin, 1973; Vol. 42; pp 49–93.
- [126] Gutman, I.; Trinajstić, N. *Chem. Phys. Lett.* **1973**, *20*, 257 – 260.
- [127] Trinajstić, N. *Chemical graph theory*; CRC press: Boca Raton, 1992.
- [128] Bonchev, D. *Chemical graph theory: introduction and fundamentals*; CRC Press: New York, 1991; Vol. 1.
- [129] Balaban, A. T. *Rev. Roum. Chim.* **1966**, *11*, 1205–1227.
- [130] Sylvester, J. J. *American Journal of Mathematics* **1878**, *1*, 64–104.
- [131] Balaban, A. T. *From chemical topology to three-dimensional geometry*; Springer Science & Business Media: New York, 2006.
- [132] Ballester, P. J.; Finn, P. W.; Richards, W. G. J. *Mol. Graph. Model.* **2009**, *27*, 836–845.
- [133] Karmakar, S.; Mandal, B. J. *Phys. Chem. A* **2014**, *118*, 7672–7682, PMID: 25120236.
- [134] Christiansen, J. *Adv. Catal.* **1953**, *5*, 311 – 353.
- [135] Rouvray, D. H.; Balaban, A. T. *Chemical applications of graph theory*; Academic Press, London: London, 1979; pp 177–221.
- [136] Gutman, I.; Polansky, O. E. *Mathematical concepts in organic chemistry*; Springer Science & Business Media: Berlin, 2012.
- [137] Trinajstić, N. *Computational chemical graph theory: characterization, enumeration, and generation of chemical structures by computer methods*; E. Horwood: New York, 1991.
- [138] Klein, D. J.; Randić, M. J. *Math. Chem.* **1993**, *12*, 81–95.

- [139] Kier, L. *Molecular connectivity in chemistry and drug research*; Elsevier: Letchworth, 2012; Vol. 14.
- [140] Voiculescu, N.; Balaban, A. T.; Niculescu-Duvaz, I.; Simon, Z. *Modeling of cancer genesis and prevention*; CRC press: Boca Raton, 1991.
- [141] Kier, L. B.; Hall, L. H. *Molecular structure description: The Electrotological State*; Academic Press: San Diego, 1999.
- [142] Devillers, J.; Balaban, A. T. *Topological indices and related descriptors in QSAR and QSPR*; CRC Press: The Netherlands, 2000.
- [143] Diudea, M. V. *QSPR/QSAR studies by molecular descriptors*; Nova Science Publishers: Huntington, New York, 2001.
- [144] Graovac, A.; Gotman, I.; Trinajstić, N. *Topological approach to the chemistry of conjugated molecules*; Springer Science & Business Media: Berlin, 2012; Vol. 4.
- [145] Dias, J. R. *Molecular orbital calculations using chemical graph theory*; Springer: Berlin, 1993; Vol. 19.
- [146] Cyvin, S. J.; Ion, G. *Lect. Notes Chem.* **1988**, *46*, 1–23.
- [147] Gutman, I.; Cyvin, S. J. *Introduction to the theory of benzenoid hydrocarbons*; Springer Science & Business Media: Berlin, 2012.
- [148] Fujita, S. *Theor Chem Acc* **2006**, *115*, 37–53.
- [149] Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics, Volume 41 (2 Volume Set)*; John Wiley & Sons: Darmstadt, 2009; Vol. 41.
- [150] Ruecker, R. C., G. *J. Chem. Inf. Comput. Sci* **1990**, *30*, 187–191.
- [151] Xu, Y.-j.; Johnson, M. J. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 181–185.
- [152] Polya, G.; Read, R. C. *Combinatorial enumeration of groups, graphs, and chemical compounds*; Springer Science & Business Media: Berlin, 2012.
- [153] Fujita, S. *Symmetry and combinatorial enumeration in chemistry*; Springer Science & Business Media: Berlin, 2012.
- [154] Ivanciuc, O. *Canonical numbering and constitutional symmetry*; Wiley-VCH: Weinheim: Chichester, 2003; pp 139–160.
- [155] Carlson, R.; Carlson, J. E. *Design and optimization in organic synthesis*; Elsevier: London, 2005; Vol. 24.
- [156] Eros, D.; Kovesdi, I.; Orfi, L.; Takacs-Novak, K.; Acsády, G.; Kéri, G. *Curr. Med. Chem.* **2002**, *9*, 1819–1829.

- [157] Todeschini, R.; Gramatica, P.; Provenzani, R.; Marengo, E. *Chemom. Intell. Lab. Syst.* **1995**, *27*, 221–229.
- [158] Consonni, V.; Todeschini, R.; Pavan, M. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 682–692.
- [159] Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- [160] Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. *J. Phys. Chem.* **1996**, *100*, 10400–10407.
- [161] Duchowicz, P. R.; Castro, E. A.; Fernández, F. M.; Gonzalez, M. P. *Chem. Phys. Lett.* **2005**, *412*, 376–380.
- [162] Krenkel, G.; Castro, E.; Toropov, A. *J. Mol. Struct. THEOCHEM* **2001**, *542*, 107–113.
- [163] Labute, P. *J. Mol. Graph. Model.* **2000**, *18*, 464–477.
- [164] Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Hicks, M. G. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 306–316.
- [165] Estrada, E.; Ramírez, A. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 837–843.
- [166] Padron, J.; Carrasco, R.; Pellon, R. *J. Pharm. Pharmaceut. Sci* **2002**, *5*, 258–265.
- [167] Estrada, E. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31–33.
- [168] Yan, X.; Yu, P. S.; Han, J. *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data; SIGMOD '04*; ACM: New York, NY, USA, 2004; pp 335–346.
- [169] van Aalten, D. M.; Bywater, R.; Findlay, J. B.; Hendlich, M.; Hooft, R. W.; Vriend, G. *J. Comput. Aided Mol. Des.* **1996**, *10*, 255–262.
- [170] Matter, H. *J. Med. Chem.* **1997**, *40*, 1219–1229.
- [171] Perminova, I. V.; Grechishcheva, N. Y.; Petrosyan, V. S. *Env. Sci. & Tech.* **1999**, *33*, 3781–3787.
- [172] Katritzky, A. R.; Gordeeva, E. V. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835–857.
- [173] Ertl, P.; Rohde, B.; Selzer, P. *J. Med. Chem.* **2000**, *43*, 3714–3717.
- [174] Wiener, H. *J. Chem. Phys.* **1947**, *15*, 766–766.
- [175] Stiel, L. I.; Thodos, G. *Amer. Inst. Chem. Eng. J.* **1962**, *8*, 527–529.
- [176] Rouvray, D. H.; Crafford, B. C. *South Afr. J. Sci.* **1976**, *72*, 74.

- [177] Papazova, D.; Dimov, N.; Bonchev, D. *J. Chromatogr. A* **1980**, *188*, 297–303.
- [178] Bonchev, D.; Mekenjan, O.; Protić, G.; Trinajstić, N. *J. Chromatogr. A* **1979**, *176*, 149–156.
- [179] Mihalić, Z.; Trinajstić, N. *J. Mol. Struc. THEOCHEM* **1991**, *232*, 65–78.
- [180] Bogdanov, B.; Mihalić, Z.; Trinajstić, N. *J. Math. Chem.* **1989**, *3*, 299.
- [181] Randić, M. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1063.
- [182] Babić, D.; Klein, D. J.; Lukovits, I.; Nikolić, S.; Trinajstić, N. *Int. J. Quantum Chem.* **2002**, *90*, 166–176.
- [183] Kruskal, J. B. *Proc. Amer. Math. Soc* **1956**, *7*, 48–50.
- [184] Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- [185] Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284–298.
- [186] Granovsky, A. A. *Firefly version 8*, *www*
<http://classic.chem.msu.su/gran/firefly/index.html>.
- [187] Pérez, C.; Muckle, M. T.; Zaleski, D. P.; Seifert, N. A.; Temelso, B.; Shields, G. C.; Kisiel, Z.; Pate, B. H. *Science* **2012**, *336*, 897–901.
- [188] Saykally, R. J.; Wales, D. J. *Science* **2012**, *336*, 814–815.
- [189] Goldman, N.; Saykally, R. J. *J. Chem. Phys.* **2004**, *120*, 4777–4789.
- [190] Mas, E. M.; Bukowski, R.; Szalewicz, K. *J. Chem. Phys.* **2003**, *118*, 4386–4403.
- [191] Mas, E. M.; Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; Wormer, P. E. S.; van der Avoird, A. *J. Chem. Phys.* **2000**, *113*, 6687–6701.
- [192] Finn, P. W.; Morris, G. M. *WiRe Comp. Mol. Sci.* **2013**, *3*, 226–241.
- [193] Ballester, P. J.; Richards, W. G. *J. Comput. Chem.* **2007**, *28*, 1711–1723.
- [194] Armstrong, M. S.; Morris, G. M.; Finn, P. W.; Sharma, R.; Richards, W. G. *J. Mol. Graph. Model.* **2009**, *28*, 368–370.
- [195] Connolly, M. *Science* **1983**, *221*, 709–713.
- [196] Connolly, M. L. *J. Am. Chem. Soc.* **1985**, *107*, 1118–1124.
- [197] Rush, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. *J. Med. Chem.* **2005**, *48*, 1489–1495.

- [198] Duncan, B. S.; Olson, A. J. *Biopolymers* **1993**, *33*, 219–229.
- [199] Ritchie, D. W.; Kemp, G. J. L. *J. Comput. Chem.* **1999**, *20*, 383–395.
- [200] Ritchie, D. W.; Kemp, G. J. *Proteins* **2000**, *39*, 178–194.
- [201] Mavridis, L.; Hudson, B. D.; Ritchie, D. W. *J. Chem. Inf. Model.* **2007**, *47*, 1787–1796.
- [202] Call, S. T.; Zubarev, D. Y.; Boldyrev, A. I. *J. Comput. Chem.* **2007**, *28*, 1177–1186.
- [203] Carugo, O.; Pongor, S. *Protein Sci.* **2001**, *10*, 1470–1473.
- [204] Collet, P. *Artificial Evolution: 5th International Conference, Evolution Artificielle, EA 2001, Le Creusot, France, October 29-31, 2001. Selected Papers*; Springer Science & Business Media, 2002; Vol. 2310.
- [205] Cantú-Paz, E.; Foster, J. A.; Deb, K.; Lawrence, D.; Roy, R.; O'Reilly, U.-M.; Beyer, H.-G.; Standish, R.; Kendall, G.; Wilson, S. *Genetic and Evolutionary Computation-GECCO 2003: Genetic and Evolutionary Computation Conference, Chicago, IL, USA, July 12-16, 2003, Proceedings*; Springer, 2003; Vol. 2723.
- [206] Eiben, A. E.; Schippers, C. *Fundamenta Informaticae* **1998**, *35*, 35–50.
- [207] Steven, G. *An Analysis of Diversity in Genetic Programming*; PhD Thesis. University of Nottingham: Nottingham, 2004.
- [208] Chi, Y.; Sun, F.; Jiang, L.; Yu, C. *Intelligent Systems (IS), 2012 6th IEEE International Conference*; IEEE press: United Kingdom, 2012; pp 361–367.
- [209] Moreno, D.; Pan, S.; Zeonjuk, L. L.; Islas, R.; Osorio, E.; Martinez-Guajardo, G.; Chattaraj, P. K.; Heine, T.; Merino, G. *Chem. Commun.* **2014**, *50*, 8140–8143.
- [210] Merino, G.; Heine, T. *Angew. Chem., Int. Ed.* **2012**, *51*, 10226–10227.
- [211] Jiménez-Halla, J.; Islas, R.; Heine, T.; Merino, G. *Angew. Chem., Int. Ed.* **2010**, *49*, 5668–5671.
- [212] Shimp, L. A.; Chung, C.; Lagow, R. J. *Inorg. Chim. Acta* **1978**, *29*, 77–81.
- [213] Baran Jr, J.; Lagow, R. J. *Am. Chem. Soc.* **1990**, *112*, 9415–9416.
- [214] Baran, J. R.; Hendrickson, C.; Laude, D. A.; Lagow, R. J. *J. Org. Chem.* **1992**, *57*, 3759–3760.

- [215] Xie, Y.; Schaefer, H. F. *Chem. Phys. Lett.* **1991**, *179*, 563–567.
- [216] Glukhovtsev, M. N.; Simkin, B. Y.; Minkin, V. I. *Zh. Org. Khim.* **1991**, *27*, 432–443.
- [217] Smith, B. J. *Chem. Phys. Lett.* **1993**, *207*, 403–406.
- [218] Bachrach, S. M.; Miller, J. V. *J. Org. Chem.* **2002**, *67*, 7389–7398.
- [219] Li, R.; Li, Q.; Liu, Z.; Li, W.; Luan, F.; Cheng, J.; Gong, B. *Mol. Phys.* **2011**, *110*, 65–74.
- [220] Guo, J. C. *J. Mol. Struct. THEOCHEM* **2010**, *953*, 139–142.
- [221] Giri, S.; Bandaru, S.; Chakraborty, A.; Chattaraj, P. K. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20602–20614.
- [222] Giri, S.; Lund, F.; Núñez, A. S.; Toro-Labbé, A. *J. Phys. Chem. C* **2013**, *117*, 5544–5551.
- [223] Dunning, T. H. *J. Chem. Phys.* **1970**, *53*, 2823–2833.
- [224] Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- [225] Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry*; Scuseria, G. E., Clifford, E., Dykstra, G. E., Grenog, F., Kim, K. S., Eds.; Elsevier: Amsterdam, 2005; pp 1167–1189.
- [226] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- [227] Čížek, J. *Adv. Chem. Phys.*; John Wiley & Sons, Inc., 2007; pp 35–89.
- [228] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479 – 483.
- [229] Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211–7218.
- [230] Frisch, M. J. et al. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, USA.
- [231] Contreras, M.; Osorio, E.; Ferraro, F.; Puga, G.; Donald, K. J.; Harrison, J. G.; Merino, G.; Tiznado, W. *Chem. Eur. J.* **2013**, *19*, 2305–2310.
- [232] te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931–967.

- [233] Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acc.* **1998**, *99*, 391–403.
- [234] Chen, W.; Hase, W. L.; Schlegel, H. *Chem. Phys. Lett.* **1994**, *228*, 436–442.
- [235] Millam, J. M.; Bakken, V.; Chen, W.; Hase, W. L.; Schlegel, H. B. *J. Chem. Phys.* **1999**, *111*, 3800–3805.
- [236] Li, X.; Millam, J. M.; Schlegel, H. B. *J. Chem. Phys.* **2000**, *113*, 10062–10067.
- [237] Koster, A. M. et al. The DeMon Developers. The deMon developers: Cinvestav, Mexico City, 2011.
- [238] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- [239] Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.
- [240] Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560–571.
- [241] von Schleyer P. Rague, *Pure Appl. Chem.* **1984**, *56*, 151.
- [242] von Schleyer P. Rague, *Pure Appl. Chem.* **1983**, *55*, 355.
- [243] Reed, A. E.; Weinstock, R. B.; Weinhold, F. J. *Chem. Phys.* **1985**, *83*, 735–746.
- [244] Mayer, I. J. *Comput. Chem.* **2007**, *28*, 204–221.
- [245] Tiznado, W.; Perez-Peralta, N.; Islas, R.; Toro-Labbe, A.; Ugalde, J. M.; Merino, G. *J. Am. Chem. Soc.* **2009**, *131*, 9426–9431.
- [246] Perez-Peralta, N.; Contreras, M.; Tiznado, W.; Stewart, J.; Donald, K. J.; Merino, G. *Phys Chem Chem Phys* **2011**, *13*, 12975–12980.
- [247] Wu, Y.; Jiang, J.; Lu, H.; Wang, Z.; Perez-Peralta, N.; Islas, R.; Contreras, M.; Merino, G.; Wu, J. I.; vonRagué Schleyer, P. *Chem. Eur. J.* **2011**, *17*, 714–719.
- [248] Santos, J. C.; Contreras, M.; Merino, G. *Chem. Phys. Lett.* **2010**, *496*, 172–174.
- [249] Haile, J. M. *Molecular Dynamics Simulation: Elementary Methods*; Jhon Wiley and Sons: New York, 1992.
- [250] Huang, W.; Sergeeva, A. P.; Zhai, H.-J.; Averkiev, B. B.; Wang, L.-S.; Boldyrev, A. I. *Nat Chem* **2010**, *2*, 202–206.
- [251] Erhardt, S.; Frenking, G.; Chen, Z.; Schleyer, P. v. R. *Angew. Chem., Int. Ed.* **2005**, *44*, 1078–1082.

- [252] Martínez-Guajardo, G.; Sergeeva, A. P.; Boldyrev, A. I.; Heine, T.; Ugalde, J. M.; Merino, G. *Chem. Commun.* **2011**, *47*, 6242–6244.
- [253] Zhang, J.; Sergeeva, A. P.; Sparta, M.; Alexandrova, A. N. *Angew. Chem., Int. Ed.* **2012**, *51*, 8512–8515.
- [254] Tai, T. B.; Ceulemans, A.; Nguyen, M. T. *Chem. Eur. J.* **2012**, *18*, 4510–4512.
- [255] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- [256] Zubarev, D. Y.; Sergeeva, A. P.; Boldyrev, A. I. In *Chemical reactivity theory: a density functional view*; Chattaraj, P. K., Ed.; CRC Press, Taylor & Francis Group, 2009; pp 439–452.
- [257] Sergeeva, A. P.; Averkiev, B. B.; Zhai, H.-J.; Boldyrev, A. I.; Wang, L.-S. *J. Chem. Phys.* **2011**, *134*, 224304.
- [258] Kalek, M.; Benediktson, P.; Vester, B.; Wengel, J. *Chem. Commun.* **2008**, 762–764.
- [259] Scott, L. T.; Hashemi, M. M.; Bratcher, M. S. *J. Am. Chem. Soc.* **1992**, *114*, 1920–1921.
- [260] Tai, T. B.; Duong, L. V.; Pham, H. T.; Mai, D. T. T.; Nguyen, M. T. *Chem. Commun.* **2014**, *50*, 1558–1560.
- [261] Tai, T. B.; Havenith, R. W. A.; Teunissen, J. L.; Dok, A. R.; Hallaert, S. D.; Nguyen, M. T.; Ceulemans, A. *Inorg. Chem.* **2013**, *52*, 10595–10600.
- [262] Merino, G.; Heine, T.; Seifert, G. *Chem. Eur. J.* **2004**, *10*, 4367–4371.
- [263] Heine, T.; Islas, R.; Merino, G. *J. Comput. Chem.* **2007**, *28*, 302–309.
- [264] Islas, R.; Heine, T.; Merino, G. *Acc. Chem. Res.* **2012**, *45*, 215–228.
- [265] Ortiz, J. V. *J. Chem. Phys.* **1988**, *89*, 6348–6352.
- [266] Frisch, M. J. et al. Gaussian 09 Revision B.01. Gaussian Inc. Wallingford CT 2009.
- [267] Huang, W.; Sergeeva, A. P.; Zhai, H.-J.; Averkiev, B. B.; Wang, L.-S.; Boldyrev, A. I. *Nat Chem* **2010**, *2*, 202–206.
- [268] Jiménez-Halla, J. O. C.; Islas, R.; Heine, T.; Merino, G. *Angew. Chem., Int. Ed.* **2010**, *49*, 5668–5671.
- [269] Averkiev, B. B.; Zubarev, D. Y.; Wang, L.-M.; Huang, W.; Wang, L.-S.; Boldyrev, A. I. *J. Am. Chem. Soc.* **2008**, *130*, 9248–9250, PMID: 18582042.
- [270] Krüger, T. *Int. J. Quantum Chem.* **2006**, *106*, 1865–1869.

- [271] Bera, P. P.; Schleyer, P. v. R.; Schaefer, H. F. *Int. J. Quantum Chem.* **2007**, *107*, 2220–2223.
- [272] Varma, R.; Ramaprasad, K. R.; Nelson, J. F. *J. Chem. Phys.* **1975**, *63*, 915–918.
- [273] Guha, S.; Francisco, J. S. *Astrophys J* **2007**, *671*, 2159–2163.
- [274] Ballester, P. J. *Shape recognition methods and systems for searching molecular databases*; Patent US8244483 B2, 2012.
- [275] Zyrianov, Y. *J. Chem. Inf. Model.* **2005**, *45*, 657–672.