



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

**DESCUBRIMIENTO DE ESTRUCTURAS DE CRISTALES Y
CLUSTERS MEDIANTE ALGORITMOS GENÉTICOS**

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en el
área de Ciencias Físicas.

Víctor E. Bazterra

Director de tesis: Dra. Marta B. Ferraro

Lugar de Trabajo: Departamento de Física, FCEyN, Universidad de Buenos Aires.

Buenos Aires, Agosto 2006.

Resumen

Esta tesis está dedicada a predecir estructuras de cristales orgánicos. En la misma se presenta una nueva aplicación capaz de generar las mejores estructuras cristalinas asociadas a una dada molécula flexible, para un número arbitrario de moléculas en la celda asimétrica y para cualquier grupo espacial cristalográfico. Esta aplicación incluye un conjunto de programas altamente integrados encargados de la generación de potenciales intermoleculares, de la búsqueda de los posibles cristales asociados a una molécula utilizando algoritmos genéticos (GA) distribuidos (en paralelo) y del análisis, procesamiento y almacenamiento de las estructuras más importantes generadas por el GA. Este *paquete* computacional puede ser utilizado para la predicción de estructuras cristalinas de una gran variedad de moléculas orgánicas sin ninguna restricción, ni por simetría ni por el tamaño de la celda unidad. Como ejemplo se presenta la predicción para los cristales de L-alanina, DL-alanina, imidazolidina-2,4-diona, N-(2-metil-4,5-dinitrofenil) y ácido 2-(acetiloxi)benzoico (aspirina). También como parte de esta tesis se desarrollaron los primeros pasos para extender la aplicación al modelado de las estructuras de clusters atómicos. Como resultado preliminar se informa de la existencia de una serie nueva de estructuras estables para el Si_{36} , que presentan una energía de confinamiento significativamente mayor a cualquier otra estructura propuesta previamente en la literatura.

PALABRAS CLAVES: Predicción de estructuras cristalinas, algoritmos genéticos paralelos, potenciales empíricos y búsqueda comparativa de cristales.

Abstract

This thesis is dedicated to the problem of crystal structure prediction. It is described a new distributed computing framework for crystal structure prediction that is capable of performing crystal structure searches for flexible molecules with an arbitrary number of molecules in the asymmetric cell and for any space group. The distributed computing framework includes a series of tightly integrated computer programs for generating the force fields, sampling possible crystal structures using a distributed parallel genetic algorithm (GA) and analyzing, sorting and archiving the most relevant crystal structures found by the GA. This framework can be used to predict the crystal structures of a large variety of organic molecules without any symmetry or unit cell size constraints. As an example, the results of the application of this new method for the prediction of the crystal structures of L-alanine, DL-alanine, imidazolidine-2,4-dione, N-(2-methyl-4,5-dinitrophenyl)acetamide y 2-(acetyloxy)benzoic acid (aspirine) are reported. As a part of this thesis the first stages for extending this application to the problem of structure prediction for atomic clusters are reported. As preliminaries results, a new series of stable isomers for Si_{36} was found with higher binding energy than any other previously reported structure in the literature.

KEYWORDS: Crystal structure prediction, parallel genetic algorithms, force fields and crystal similarity search.

A Flor

Agradezco sinceramente a la *Dra. Marta B.*

Ferraro, directora de esta tesis, por su apoyo y

guía permanente y su invaluable afecto.

Mi agradecimiento al *Dr. Julio C. Facelli* por su
interés en mi formación científica, su amistad y
su constante colaboración.

Los resultados de esta tesis se obtuvieron utilizando las siguientes computadoras: Arches metacluster, parcialmente financiado por NIH (subsidio NCRR 1 S10RR17214-01) y TeraGrid, en el marco del proyecto MCA 05T018.

Agradecimientos

Quiero agradecer a las siguientes personas e Instituciones, cuyos aportes fueron indispensables en la creación de esta tesis.

Universidad de Buenos Aires.

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Center for High Performance Computing (CHPC), University of Utah

A María Florencia Rodríguez Brasco (mi esposa y Jefe de Trabajos Prácticos del Departamento de Química Orgánica, FCEyN, UBA) por el intercambio de ideas en general y en particular por su ayuda en la corrección y edición.

A Matthew Thorley (Network assistant, CHPC) por su participación en el desarrollo de servidores de páginas web y de bases de datos. Su colaboración ha sido una de las más influyentes en los últimos dos años, impulsando el uso de estándares en la programación y para el intercambio de datos.

A Martin Čuma (Scientific applications, CHPC) por su contribución en la creación de métodos de paralelización adaptativos, y en el mantenimiento de compiladores y paquetes de códigos para aplicaciones de alta performance.

A Thomas Cheatham III (Biomolecular modeling, CHPC) por las horas de cómputos cedidas y fundamentalmente por su guía en la búsqueda de paquetes computacionales para simulaciones en gran escala. También Tom me sugirió el uso de antechamber y GAFF como potencial empírico para evaluar la energía de los cristales.

A Samuel T. Liston (Digital communication and visualization, CHPC) por su ayuda y recomendaciones para la visualización de las estructuras cristalinas mostradas en esta tesis. En particular, Sam ha creado un muro de visualización que permite proyectar en tres dimensiones las diferentes estructuras utilizando luz polarizada.

A Brian Haymore (Lead, computation cluster administrator, CHPC) por su aporte en la creación y mantenimiento del metacluster Arches. Además por su extensas explicaciones de cómo funcionan internamente los cluster, lo que ha permitido el refinamiento de todos mis programas.

A Anita Orendt (Molecular sciences, CHPC) y María Florencia Rodríguez Brasco porque me han ayudado en muchos de los aspectos químicos y biológicos relacionados con los cristales orgánicos.

A Ofelia Oña (Ayudante de primera del Departamento de Física, FCEyN, UBA) por las largas discusiones sobre todos los aspectos de este trabajo. Su visión crítica fue muy importante para el perfeccionamiento de todos los algoritmos desarrollados.

A Wayne Bradford (System administrator, CHPC) por su servicios como administrador de sistemas en el CHPC, y sus explicaciones relacionadas a sistemas y seguridad informática.

A Ron Price (Software Engineer, CHPC) por su contribución en lo que será parte del futuro de estos proyectos, permitiendo sus implementaciones en GRID.

A Cristina Caputo (Profesora del Departamento de Física, FCEyN, UBA) por su amistad e interés en mi formación y por sus aportes críticos a muchas de las ideas desarrolladas aquí.

A la familia Facelli, en particular a Ana, por su cariño y consejos en mi estadía en Salt Lake City.

A mis amigos en Utah: Alvaro, Angélica, Luciano, Ximena, William, Mauricio, Carolina, Yudith, Lucas, Toto, Cecilia, Martin, Ron, Iris y Matthew; y también a mis amigos en Buenos Aires: Leo (a la distancia), Tito, Andrea, Mirta, Federico y Gabriel, los cuales aportaron con su presencia y discusiones a esta tesis.

A mi suegra Norma por su cariño y generosidad al recibirme en su casa cada vez que me encontraba en Buenos Aires.

A toda mi familia, que siempre se emocionó y me apoyó en mis vocaciones. En especial a mi mamá Nelly, que siempre quiso que fuera *doctor* aunque no me aclaró que se refería a medicina y no a física.

Finalmente, a mi esposa Florencia por su infinidad de aportes a este trabajo y su apoyo durante estos años a pesar de los grandes períodos de separación.

Índice de figuras

2.1. Efecto sobre la superficie de energía al aplicar la optimización local.	18
2.2. Representación del operador <i>cut and splice</i>	19
2.3. Ejemplo de un cluster molecular y una celda unidad en P1 que genera un cristal mal formado (a), en donde las moléculas de la celda unidad se superponen con sus imágenes (b).	21
3.1. Representación esquemática del genoma de un cristal	25
3.2. Representación esquemática del cristal generado por el genoma.	26
3.3. Diagrama de flujo para el MGA	28
4.1. Estructuras de las moléculas de la tabla 4.1. Se muestran además los ángulos diedros que confieren flexibilidad a estas moléculas.	41
5.1. Fragmento del cristal compuesto por 16 moléculas de imidazolidina-2,4-diona.	45

5.2. Superposición entre el fragmento de referencia (en tonos de grises) y los centros de masas del fragmento de comparación (puntos rojos). Se puede observar que esta superposición se logra a través del alineado de las moléculas centrales de ambos fragmentos. 48

6.1. Superposición entre la estructura experimental y la predicha (en color rosa) para la molécula de L-alanina. El RMS de la superposición es de 0.30 Å . . . 58

6.2. Superposición entre la estructura experimental y la predicha (en color rosa) para las moléculas de DL-alanina. El RMS de la superposición es de 0.40 Å . 59

6.3. Superposición entre la estructura experimental y la predicha (en color rosa) para la molécula de imidazolidina-2,4-diona. El RMS de la superposición es de 0.33 Å 63

6.4. Distribución de los primeros 100 cristales generados por MGAC en función de la energía medida respecto a la energía del mínimo global para la molécula de imidazolidina-2,4-diona. 64

6.5. Superposición entre la estructura experimental y la predicha (en color rosa) para la molécula de N-(2-dimetil-4,5-dinitrofenil)acetamida. El RMS de la superposición es de 0.16 Å 69

6.6. Distribución de los primeros 200 cristales generados por MGAC en función de la energía medida respecto a la energía del mínimo global para la molécula de N-(2-dimetil-4,5-dinitrofenil)acetamida 70

6.7. Evolución en promedio de las mejores estructuras correspondientes a los grupos espaciales $P2_1/c$, $P2_1$ y $P1$	71
6.8. Distribución de los primeros 393 cristales generados por MGAC en función de la energía medida en referencia a la energía del mínimo global para el ácido 2-(acetiloxi)benzoico	75
6.9. Primeras dos conformaciones en orden de energía creciente para la molécula de aspirina.	76
6.10. Superposición entre la estructura experimental y la predicha (en color rosa) para el ácido 2-(acetiloxi)benzoico. El RMS de la superposición es de 0.29 Å	80
6.11. Distribuciones de los cristales generados por el MGAC para las restricciones $P2_1$ y $P2_1/c$ para los casos $\tau = 0$ en azul y $\tau = 180$ en rojo.	82
6.12. Número de cristales evaluados por molécula para cada año de esta tesis.	87
7.1. Estructuras optimizadas y energía de confinamiento por átomo para Si_n ($n = 4-9$).	94
7.2. Estructuras de cluster Si_n ($n=4-9$) extraídas de la referencia [1]	95
7.3. Comparación entre la estructura hallada por el MGAC y (*) las estructuras de la referencia [2].	96
A.1. Esquema del operador de recombinación a un punto	105
A.2. Esquema del operador de recombinación a dos puntos	105
A.3. Esquema del operador de recombinación a N puntos	106
A.4. Esquema del operador de recombinación uniforme	106

A.5. Esquema del operador de media aritmética 107

A.6. Esquema del operador de media geométrica 107

A.7. Esquema del operador inversión 107

Índice de tablas

2.1. Programas y metodologías empleadas en la predicción de cristales orgánicos. (*) CSD proviene de la base de datos CSD (Cambridge Structural Database).	12
4.1. Lista de moléculas estudiadas a lo largo de la tesis.	40
4.2. Optimización local utilizando GAFF de las estructuras cristalinas experimentales.	42
6.1. Comparación de las estructuras y los factores R_{wp} de las estructuras optimizadas localmente de L-alanina y DL-alanina para los potenciales CHARMM y AMBER	55
6.2. Calidad de las estructuras cristalinas predichas para la L y DL-alanina. . . .	57
6.3. Primeras 20 estructuras cristalinas en orden de energía creciente para la molécula de imidazolidina-2,4-diona. La estructura de la segunda fila coincide con la estructura cristalina conocida experimentalmente para esta molécula.	61
6.4. Calidad de la estructura cristalina predicha para molécula de imidazolidina-2,4-diona.	62

6.5. Primeras 20 estructuras cristalinas en orden de energía creciente para la molécula de N-(2-metil-4,5-dinitrofenil)acetamida. La estructura de la tercera fila coincide con la estructura cristalina conocida experimentalmente para esta molécula.	67
6.6. Calidad de la estructura cristalina predicha para la molécula de N-(2-metil-4,5-dinitrofenil)acetamida.	68
6.7. Primeras 20 estructuras cristalinas en orden de energía creciente para el ácido 2-(acetiloxi)benzoico generadas por el MGAC restringidas a los grupos espaciales $P2_1$ y $P2_1/c$ y a la conformación B . La energía se calcula respecto al mínimo global encontrado en la primera corrida. La estructura de la segunda fila coincide con la estructura cristalina conocida experimentalmente para esta molécula.	78
6.8. Calidad de la estructura cristalina predicha para el ácido 2-(acetiloxi)benzoico. (*) Esta estructura se obtuvo al restringir la posición del ácido a la conocida experimentalmente.	79
6.9. Diferencias de energías entre las conformaciones de la aspirina A y B	83
6.10. Comparación entre el mínimo global de la referencia [3] y el primer cristal de la tabla 6.7.	84
6.11. Calidad de las estructuras predichas. (*) Estructura predicha para la aspirina cuando se restringió la posición del ácido a la conocida experimentalmente. .	85

7.1. Energía de confinamiento por átomo para los isómeros de Si_{36} . Todos los valores se calcularon tomando la energía del átomo de silicio como -102.155 eV y -101.428 eV para las aproximaciones LSDA y B3PW91 respectivamente. 97

B.1. Primeras 100 estructuras cristalinas en orden de energía creciente para la molécula de imidazolidina-2,4-diona. La estructura de la segunda posición coincide con la estructura del cristal conocida experimentalmente para esta molécula. 110

B.2. Primeras 189 estructuras cristalinas en orden de energía creciente para la molécula de N-(2-dimetil-4,5-dinitrofenil)acetamida. La estructura de la tercera posición coincide con la estructura cristalina conocida experimentalmente para esta molécula. 114

B.3. Primeras 393 estructuras cristalinas en orden de energía creciente para el ácido 2-(acetiloxi)benzoico. 124

B.4. Primeras 199 estructuras cristalinas en orden de energía creciente para el ácido 2-(acetiloxi)benzoico generadas por el MGAC restringidas a los grupos espaciales $P2_1$ y $P2_1/c$ y a la conformación B . La energía se calcula respecto al mínimo global encontrado en la primera corrida. La estructura de la segunda fila coincide con la estructura cristalina conocida experimentalmente para esta molécula. 128

Índice general

1. Introducción	1
1.1. Importancia de la predicción de estructuras	1
1.2. Obstáculos en la predicción de estructuras cristalinas	3
1.2.1. Polimorfismo	3
1.2.2. Precisión en la evaluación de la energía	4
1.2.3. Optimización global	5
1.3. Obstáculos en la predicción de clusters	6
1.3.1. Evaluación de la energía	6
1.3.2. Optimización global	6
1.4. Objetivos de la tesis	8
2. Predicción de estructuras cristalinas	10
2.1. Métodos existentes	10
2.2. Complejidad asociada	13
2.3. Algoritmos Genéticos aplicados a problemas estructurales	15

2.4. Proposición de un nuevo método	21
3. Algoritmo Genético Modificado para Cristales	23
3.1. Introducción	23
3.2. Genoma de un cristal	23
3.3. Descripción del algoritmo	27
3.4. Operadores Genéticos	29
3.5. Implementación del MGAC	32
4. Cálculo de la energía.	34
4.1. Cálculo de la energía utilizando GAFF	34
4.2. Optimización local de la energía	37
4.3. Validación en estructuras cristalinas conocidas	39
5. Análisis de las estructuras cristalinas	44
5.1. Comparación de la estructura tridimensional de dos cristales.	44
5.2. Colección de los cristales más significativos	50
5.3. Procesamiento y clasificación de los cristales	51
6. Resultados	54
6.1. Cristales de L-alanina y de DL-alanina	54
6.2. Cristales de imidazolidina-2,4-diona	60
6.3. Cristales de N-(2-metil-4,5-dinitrofenil)acetamida	65
6.4. Límites del método: el ejemplo de la aspirina	73

6.5. Calidad de las estructuras predichas	85
6.6. Eficiencia de la metodología MGAC	86
7. Extendiendo MGAC a clusters atómicos	88
7.1. Motivaciones	88
7.2. Algoritmo genético para clusters atómicos	90
7.3. Resultados	92
Conclusiones	99
A. Operadores de recombinación	103
A.1. Recombinación a un punto	103
A.2. Recombinación a dos puntos	103
A.3. Recombinación a N puntos	103
A.4. Recombinación uniforme	104
A.5. Recombinación media aritmética	104
A.6. Recombinación de media geométrica	104
A.7. Inversión	104
B. Material suplementario	108
B.1. Cristales de imidazolidina-2,4-diona	108
B.2. Cristales de N-(2-dimetil-4,5-dinitrofenil)acetamida	110
B.3. Cristales del ácido 2-(acetiloxi)benzoico	115
B.4. Cristales del ácido 2-(acetiloxi)benzoico (conformación B)	124

Capítulo 1

Introducción

1.1. Importancia de la predicción de estructuras

El objetivo de esta tesis es la predicción de estructuras en sistemas atómicos y moleculares con muchos grados de libertad, utilizando Algoritmos Genéticos (Genetic Algorithms o GA) [4]. En particular esta tesis está dedicada a predecir estructuras de cristales moleculares y se incluye una pequeña contribución para el caso de clusters atómicos, empleando siempre GA.

Existen varias razones para justificar la necesidad de predecir estas estructuras. En el caso de los cristales moleculares, las mismas están relacionadas con la nueva disciplina de *Ingeniería de Cristales*, cuyo objetivo es el diseño de materiales sólidos con propiedades específicas [5]. Productos farmacéuticos, agroquímicos, pigmentos, materiales altamente energéticos, etc., son ejemplos de sólidos cristalinos orgánicos de gran importancia para diferentes aplicaciones. El modelado de las estructuras moleculares es bien conocido, y de-

pendiendo del tamaño del compuesto, existe un grupo de metodologías prescriptas (empíricas, semi-empíricas e incluso ab-initio) que pueden ser utilizadas. Esto no es así cuando se desea predecir sus estructuras cristalinas, para lo cual existen diferentes metodologías que se encuentran aún en desarrollo [6, 7, 8].

La evidencia experimental muestra que algunas sustancias pueden existir en un número de formas cristalográficas diferentes, fenómeno conocido con el nombre de *polimorfismo* [9]. Este fenómeno es la principal dificultad presente en la Ingeniería de Cristales, debido a que diferentes polimorfos pueden exhibir distintas propiedades físicas, como solubilidad, morfología, densidad, presión de vapor, color, sensibilidad a determinados agentes, etc. Por lo tanto, aprovechar y controlar la producción de cada polimorfo puede ser una tarea difícil [10]. Es por eso que es necesario el desarrollo de metodologías que permitan conocer las formas más probables en las que un compuesto orgánico cristaliza [11].

En el caso de clusters atómicos y moleculares, el estudio de sus estructuras y propiedades físicas es un área extremadamente activa debido a su importancia, tanto en ciencia básica como en aplicaciones tecnológicas [12, 13]. Esta importancia está fundada no sólo en sus propiedades intrínsecas, sino también en la posición central que los clusters ocupan en el estudio de la materia condensada. Los clusters existen en una gran variedad de número de partículas, desde moléculas hasta micro-cristales. Constituyen un nuevo tipo de materiales conocido como *nanopartículas*, que pueden presentar propiedades muy diferentes de aquellas presentes en moléculas o en fragmentos sólidos macroscópicos.

El problema de predicción de cristales orgánicos comparte muchas similitudes con

el problema de predicción de clusters atómicos. Ambos tienen en común la necesidad de conocer la estructura de la superficie de energía definida en un espacio multi-dimensional de configuración. En esta superficie se encuentra la información geométrica para las distintas conformaciones estables. Es esta relación entre dichos problemas que permite el uso de GA en la predicción de sus estructuras.

1.2. Obstáculos en la predicción de estructuras cristalinas

1.2.1. Polimorfismo

La mayor parte de los métodos de predicción de estructuras en cristales orgánicos se basa en la optimización de la energía del cristal para un conjunto de moléculas rígidas. La energía está dada principalmente por las interacciones intermoleculares. La estructura de mínima energía se considera como la termodinámicamente más estable, en otras palabras, se ignoran las contribuciones a la energía libre debidas a la entalpía y la entropía vibracional. En todo caso, hay razones para creer que el proceso de cristalización se encuentra bajo control cinético y no termodinámico [6, 10, 14]. Como consecuencia, el polimorfo que se obtiene bajo una dada condición no es necesariamente la conformación termodinámicamente más estable.

Por lo tanto, lo único que puede afirmarse es que toda estructura cristalina obtenida experimentalmente tiene que pertenecer al grupo de polimorfos cuyas energías libres difieren

razonablemente poco entre ellas [15].

1.2.2. Precisión en la evaluación de la energía

En la predicción de cristales se hace indispensable el uso de potenciales clásicos para evaluar la energía de una dada conformación cristalina. Las razones de esto están dadas tanto por consideraciones de costo computacional así como por falencias en la representación cuántica de sistemas cristalinos. Por ejemplo, es bien conocido que la teoría de la funcional de la densidad (Density Functional Theory o DFT) [16] no reproduce apropiadamente las contribuciones por dispersión en las interacciones intermoleculares [17]. En especial este problema es importante en interacciones tales como las de Van der Waals o en complejos unidos por enlace de hidrógenos de la forma $C - H \cdots O$. Esto hace que en el presente estado de la teoría, su uso no se vea justificado. Por último, los métodos que incluyen correlación electrónica son por ahora totalmente imprácticos debido a sus altísimos costos computacionales asociados.

De todas formas, los potenciales clásicos presentan un error del mismo orden de magnitud que la diferencia de energía libre entre diferentes polimorfos. Por lo tanto el criterio para discernir qué estructura es termodinámicamente más estable no se encuentra en la práctica bien definido. Es por eso que el efecto de despreciar la influencia de la entropía vibracional no se considera tan grande, ya que introduce sólo un error marginal frente al producido en el cálculo de las energías [18].

1.2.3. Optimización global

La complejidad numérica de encontrar una cantidad de mínimos representativos, incluyendo el mínimo global, puede llegar a ser muy grande. En particular, se estima que el número de mínimos locales crece al menos exponencialmente con los grados de libertad del sistema [19, 20]. Por lo tanto, hacer un muestreo del espacio de configuración para obtener un número representativo de mínimos puede ser una tarea no trivial [18]. Uno de los mayores problemas que pueden encontrarse en las presentes metodologías es la carencia del uso de métodos de optimización en paralelo. El uso de estos métodos, en conjunto con modernas computadoras paralelas, permitirían ampliar el tamaño del muestreo respecto a la mayor parte de los métodos que se emplean actualmente. Este problema se ve acentuado cuando las moléculas pertenecientes al cristal presentan cierto nivel de *flexibilidad*. Esto significa que las mismas pueden existir en una gran variedad de conformaciones que pueden ser energéticamente accesibles. Las distintas conformaciones pueden traer como consecuencia diferentes arreglos espaciales de las moléculas, es decir diferentes polimorfos. Por lo tanto, en la búsqueda de polimorfos, se encuentran acopladas tanto las posiciones de las moléculas dentro del cristal como las diferentes conformaciones que éstas pueden presentar. Se ha identificado al tratamiento de moléculas flexibles como uno de los mayores problemas en la predicción de estructuras cristalinas [8]. Esto se debe en gran medida a que la mayor parte de las presentes metodologías intentan predecir a los cristales utilizando un conjunto de conformaciones moleculares rígidas previamente establecidas [8]. Por lo tanto aparece la necesidad de un proceso integral, en donde tanto los grados de libertad intra como

intermoleculares sean considerados conjuntamente como parte del proceso de optimización.

1.3. Obstáculos en la predicción de clusters

1.3.1. Evaluación de la energía

El estudio de clusters de hasta 10 átomos se realiza en general utilizando métodos cuánticos tales como DFT [1]. Aunque es común el uso de potenciales clásicos para calcular la energía en clusters, la mayor parte de estas aplicaciones utilizan a los mismos como modelo simplificado para las interacción entre átomos. En estas aplicaciones se estudia la relación entre la geometría de los clusters y los parámetros que definen sus interacciones [21, 22]. En la actualidad, se acepta que todo estudio realista sobre estructura de clusters requiere el uso de métodos cuánticos [23]. Por lo tanto, la optimización global y sistemática en clusters de más de 10 átomos es una tarea compleja que consume mucho tiempo, debido a la cantidad de cálculos requeridos para evaluar la energía total en forma precisa.

1.3.2. Optimización global

En la optimización global de clusters atómicos se presenta otra gran dificultad, el extraordinario aumento del número de estructuras estables (mínimos locales) con el número de átomos. En particular, se ha establecido que este crecimiento es al menos exponencial [19, 20], evidenciando el grado de dificultad que el problema puede presentar.

Esto se ve complementado con el hecho de que el problema de optimización global en

clusters, donde la energía está dada por términos de dos cuerpos, es NP-difícil [24]. La condición NP-difícil es una clase de complejidad definida en la *teoría de la complejidad computacional* [25, 26]. Esto implica que no habría algoritmo determinista que realizara la optimización global para todas las posibles interacciones en un tiempo acotado por un polinomio del número de átomos. Esto no implica que toda optimización global en clusters no pueda ser realizada en un tiempo polinomial, sino que existen casos particulares en donde las interacciones son tales que resulta extremadamente difícil hallar el mínimo global.

Debido a la complejidad inherente al problema, todo método de optimización global necesitará evaluar un gran número de estructuras para muestrear significativamente el espacio de configuración de los clusters. Por lo tanto, la opción de usar en estos métodos cálculos ab-initio o DFT para la evaluación de la energía u otras propiedades significativas, está limitada a unas pocas estructuras plausibles. Es por este motivo que una de las estrategias más comunes en el estudio de clusters es la construcción de estructuras observadas en otros nanosistemas (jaulas, alambres, fullerenos, etc.), para luego ser optimizadas localmente hasta obtener nuevas estructuras estables. A esta estrategia se la llamará *búsqueda por patrones*, en donde se entiende por patrones al conjunto de estructuras construidas para la optimización [2]. El alto costo de los cálculos ab-initio desalienta por lo general el estudio extensivo del espacio de configuraciones, dejando abierta la posibilidad de que geometrías completamente diferentes de las seleccionadas sean candidatos válidos de isómeros estables de las especies en cuestión [27, 28].

1.4. Objetivos de la tesis

El objetivo de esta tesis es principalmente la utilización de GA en la predicción de estructuras de cristales orgánicos. Se hizo especial hincapié en el desarrollo de la generalización del GA para acoplar tanto a los grados de libertad intramoleculares (moléculas flexibles) como a los intermoleculares. Esto involucró el uso de algoritmos paralelos adaptativos para incrementar la capacidad de muestreo a través del uso de computadoras paralelas. En el proceso, se implementaron los programas necesarios para el cálculo de potenciales clásicos, así como el desarrollo de algoritmos y bases de datos para la extracción y conservación de aquellos cristales más significativos.

En esta tesis además se realizó una implementación de GA para la predicción de clusters de silicios. Aunque el uso de GA para esta clase de clusters se encuentra bien establecida, estos métodos evalúan la energía de los clusters utilizando *tight binding* basada en la funcional de la densidad (Density Functional Based Tight Binding) [29]. Esto hace que la dificultad numérica limite el número de átomos que estos métodos pueden alcanzar. Por lo tanto, en esta tesis se reporta el uso de GA con el método semi-empírico MSINDO [30, 31, 32] para evaluar la energía y así reducir los costos computacionales. El objetivo principal en este caso es verificar si existen estructuras de más baja energía en silicios de tamaño mediano que no son halladas por la búsqueda por patrones. Esto establecería la necesidad o no del uso de algoritmos de optimización global más sofisticados que la búsqueda por patrones.

Esta tesis está diagramada en la siguiente forma. En el capítulo 2 se realiza un resumen de las metodologías existentes para la predicción de cristales incluyendo en detalle el

desarrollo de GA en esta área. En el capítulo 3 se describe el algoritmo genético modificado para cristales. En el capítulo 4 se introduce el programa *antechamber*, que es el generador de potencial empírico GAFF y el método con que se evalúa la energía. En el capítulo 5 se desarrollan los métodos de análisis y preservación de la estructura tridimensional de los cristales. En el capítulo 6 se presentan los resultados de la predicción para varios cristales. En el capítulo 7 se muestra una adaptación del algoritmo para la predicción de clusters atómicos utilizando el método semi-empírico MSINDO.

Capítulo 2

Predicción de estructuras cristalinas

2.1. Métodos existentes

La creación de metodologías para la predicción de estructuras cristalinas es relativamente reciente. Como punto de partida se puede considerar el comentario realizado por Maddox en 1988 [33], que consideraba como *un continuo escándalo para la ciencias físicas* que fuese imposible la predicción de inclusive el más simple de los cristales, conociendo sólo sus composiciones químicas. Comentarios similares pueden encontrarse a lo largo de la literatura al comienzo de la década del 90 [34, 35]. Por ejemplo, Gavezzotti [14] enunció que *no es posible* la predicción de estructuras cristalinas, citando como principales dificultades las relacionadas con el fenómeno de polimorfismo. Casi simultáneamente aparecieron las primeras propuestas para la predicción de cristales orgánicos, todas muy diversas. Hoy existen al menos 14 programas dedicados a este problema (ver tabla 2.1). En la actualidad se piensa que la contundente negativa realizada por Gavezzotti en su artículo puede relajarse

diciendo que: para algunos cristales, la predicción es posible [6, 36, 37]. Aunque existe una gran variedad de métodos, todos ellos tienen en común implícitamente o explícitamente cuatro componentes generales:

1. el modelo molecular, el cual puede permitir o no la flexibilidad conformacional;
2. el método de búsqueda u optimización global;
3. cálculo de la energía u otro criterio para caracterizar el cristal;
4. programas para el análisis y conservación de las estructuras más importantes.

A los modelos moleculares en general se los clasifica en *rígidos o flexibles*. Un modelo rígido se basa en el uso de moléculas con conformaciones generadas en forma convencional, por lo general mediante cálculos ab-initio o DFT; mientras que un modelo flexible permite el control por parte de la optimización, de la rotación rígida entre fragmentos moleculares unidos por enlaces simples. En algunos casos, a este último método se lo complementa con la relajación de toda la molécula, cada vez que se calcula la energía total del cristal.

Los métodos de búsqueda o de optimización global designan las formas con que las distintas metodologías muestrean el espacio de configuraciones. Estos métodos pueden ser clasificados en diferentes grupos, como se ilustra en la tabla 2.1. Puede observarse que existe una buena variedad de métodos desarrollados en torno a la búsqueda en grilla en forma sistemática (6 métodos) y en el uso de templado simulado (4 métodos). También es fácil notar que es mayor el número de modelos rígidos (8 métodos) que flexibles.

Programa y referencias	Modelo molecular	Método	Ranking
Cerius2 [38]	Rígido	Templado Simulado Monte Carlo [39, 18]	Potenciales clásicos + cargas atómicas
CRYSCA [40]	Flexible	Búsqueda aleatoria [40]	Potenciales clásicos + cargas atómicas
CRYSTALG [40]	Flexible	Monte Carlo sobre familia de conformaciones [41]	Potenciales clásicos + cargas atómicas
DMAREL [42, 43]	Rígido	Búsqueda en grilla [44]	Potenciales clásicos + multipolos atómicos
FlexCryst [45]	Rígido	Búsqueda en grilla	Potenciales estadísticos deducidos de CSD*
MOLPAK [43]	Rígido	Búsqueda en grilla	Potenciales clásicos + cargas atómicas
MGAC [8, 46, 47, 48, 49]	Flexible	Algoritmo Genético Modificado [8]	Potenciales clásicos + cargas atómicas
MPA [50]	Flexible	Búsqueda en grilla	Potenciales clásicos + cargas atómicas + sitios extra
MSP [51]	Rígido	Templado Simulado Monte Carlo	Potenciales clásicos + cargas atómicas
Packstar [6]	Rígido	Templado Simulado Monte Carlo	Grupos de contactos del CSD*
PMC [52]	Flexible	Búsqueda en grilla adaptada por simetría	Potenciales clásicos + cargas atómicas
UPACK [53]	Flexible	Búsqueda aleatoria y en grilla	Potenciales clásicos + cargas atómicas; Energía libre
RANCEL [54]	Rígido	Algoritmos Genéticos	Potenciales clásicos + grupos de contactos del CSD*
Zip-Promet [55]	Rígido	Construcción a través de dímeros y capas [56]	Potenciales clásicos + cargas atómicas

Tabla 2.1: Programas y metodologías empleadas en la predicción de cristales orgánicos. (*) CSD proviene de la base de datos CSD (Cambridge Structural Database).

La energía o la propiedad que caracteriza el cristal se calcula bajo diferentes premisas y sujeta a varias aproximaciones. La única regularidad que se puede observar es la ausencia de métodos cuánticos para el cálculo de la energía. Este hecho está justificado por el tiempo de cómputo que demanda el cálculo y las dificultades asociadas a las teorías electrónicas para calcular fuerzas intermoleculares (ver sección 1.2.2). La mayor parte de los cálculos de la energía se basan en el uso de potenciales clásicos, los cuales requieren el ajuste de parámetros libres en función de observables asociados a un grupo representativo de moléculas y cristales moleculares conocidos previamente. En general, las interacciones electrostáticas se representan por cargas puntuales localizadas en las posiciones atómicas, aunque los métodos más elaborados utilizan desarrollos multipolares. Alternativamente, existen métodos que clasifican a los cristales en función de algunas regularidades geométricas observadas en cristales conocidos, que pueden encontrarse en la base de datos CSD (Cambridge Structural Database) [57].

2.2. Complejidad asociada

Se sabe que la búsqueda de mínimos significativos de energía en sistemas complejos, tales como proteínas y clusters atómicos o moleculares, puede presentar una gran dificultad. Para ambos problemas es posible probar su condición de ser *NP*-difícil [24, 58], por lo que no habría un algoritmo determinista que resolviese todas las posibles instancias del problema en tiempo polinomial. Se encontraron dificultades similares en la predicción de cristales, en donde es posible hallar un gran número de estructuras cuyas energías difieren en unos

pocos kJ/mol [6]. Esta dificultad crece en general con el número de dimensiones del espacio de configuración.

En el comienzo de la búsqueda, por lo general el grupo espacial G y el número de moléculas Z' en la celda asimétrica son desconocidos. Por lo tanto, toda optimización global del espacio de configuraciones cristalinas requiere una optimización para cada elección de G y Z' . Aunque en principio esto requeriría una búsqueda sistemática sobre los 230 posibles grupos espaciales [59], estudios sobre CSD [18] revelan que 92,7% de los cristales orgánicos corresponden sólo a 18 grupos espaciales y que además, en un 91,7% de los cristales sólo tienen una molécula por celda asimétrica ($Z' = 1$). Por lo tanto, esta información sostiene el argumento de que sólo se necesitaría buscar en un grupo reducido de simetrías con $Z' = 1$ para tener una alta probabilidad de predecir la estructura correcta de los cristales. Se llamará a esta forma de optimización *búsqueda con simetrías*.

Alternativamente, se podría fijar el grupo espacial en P1 (grupo con ausencia de simetría) y probar con diferentes números de moléculas Z por celda unidad. Este método está apoyado por los argumentos dados por Gao y William [60] que muestran que en promedio el valor de Z es de 3,85, y que inclusive en los grupos cúbicos, el promedio es de 6,7. En general se considera que incluyendo todos los valores de $Z \leq 8$ se contempla la mayoría de los cristales observados. Se llamará a esta forma alternativa de optimización como *búsqueda sin simetrías*.

La búsqueda sin simetrías por lo general requiere menos optimizaciones, pero deben realizarse sobre un espacio de configuración de mayor dimensión que en la búsqueda con

simetrías. Debido a esto, en general la búsqueda con simetrías es la más comúnmente implementada. Sin embargo, este método ignora la posibilidad de la existencia de otros grupos espaciales, o del hecho de que un cristal pueda presentar más de una molécula por celda asimétrica. Esto puede ser un gran problema en la predicción de polimorfos, ya que algunos estudios sugieren que el 28 % de las moléculas que presentan polimorfismo tienen un cristal de $Z' = 1$ y otro con $Z' > 1$ [61]. Esto contrasta fuertemente con el argumento a favor de la búsqueda con simetrías, en donde la condición $Z' > 1$ sólo aparece en un 8,3 %, por lo que esta metodología no sería la más apropiada en el caso de la búsqueda de polimorfos. Por lo tanto, en la actualidad no existe una indicación precisa de cuál búsqueda es mejor, ya que este criterio puede variar según sean las circunstancias y el sistema al cual se aplica.

2.3. Algoritmos Genéticos aplicados a problemas estructurales

Se designa como Algoritmos Genéticos (Genetic Algorithms o GAs) a la familia de técnicas de optimización global inspirada en el principio de selección natural. Estos métodos se basan en la generación y selección de las mejores propuestas para resolver un problema dado. Por cada problema es necesario definir una codificación que transforme a toda proposición en una cadena finita de símbolos. En el campo de GAs, estas cadenas son designadas con el nombre de *genomas*. Las diferentes propuestas son agrupadas en *poblaciones* que for-

man las distintas *generaciones*. En cada generación, las diferentes propuestas, codificadas en cada genoma, compiten entre ellas para definir su permanencia y la producción de nuevas propuestas para la siguiente generación. Operadores análogos a los de recombinación, mutación y selección natural son empleados para crear la siguiente generación combinando los genomas existentes en la población presente. El objetivo es diseñar un mecanismo de búsqueda capaz de explorar y analizar el espacio multi-dimensional de parámetros, determinando aquellas regiones de este espacio en donde se encuentran las mejores propuestas.

Los GAs fueron empleados por primera vez por Hartke en la optimización de pequeños clusters de silicio [62] y por Xiao y William para la optimización de clusters moleculares [63] a comienzos de los 90. En ambos casos la geometría de los clusters se codificó en cadenas binarias. Luego, Hartke aplicó el método utilizando potenciales clásicos [64, 65] a clusters de gases nobles [66], de agua [67, 68] y de Lennard-Jones [69]. En el caso de clusters de Lennard-Jones conteniendo entre 10 y 150 átomos, Hartke mostró empíricamente que el tiempo requerido para encontrar el mínimo global crece aproximadamente como el cubo del número de átomos ($t \approx 0,05 \times N^{2,8}$ donde t es tiempo de cómputo medido en segundos) [69].

Zeiri [70, 71, 72] introdujo el uso directo de coordenadas cartesianas para la codificación de los clusters. Este método permite representar a los mismos en términos de variables continuas (números de coma flotante), eliminado la necesidad de transformar sus estructuras respectivas en sus representaciones binarias. Además, esta generalización de los GAs permitió el diseño de un mayor número de operadores genéticos, aumentando la variedad

con que la información presente en las generaciones anteriores se utiliza para crear nuevas estructuras.

El siguiente paso en el desarrollo de GAs fue dado por Deaven, Ho [73, 74] y Gregurick, et. al. [66] quienes introdujeron el uso de optimizaciones locales sobre los clusters en cada generación. Esto hace que los GAs no trabajen sobre la superficie de energía original del cluster, sino sobre el conjunto discreto de todos los mínimos locales de la energía (ver figura 2.1). El uso combinado de codificar con números reales, nuevos operadores genéticos y de la optimización local en cada generación mostró ser tan superior a los métodos previos [75, 76], que todas las siguientes implementaciones de GAs tomaron al mismo como punto de partida. Es por eso que a esta metodología se la designa especialmente con el nombre de *Algoritmos Genéticos Modificados* (Modified Genetic Algorithms o MGA).

Otra innovación desarrollada por Deaven y Ho [73, 74] es la creación de operadores de recombinación aplicados directamente a la estructura tridimensional del cluster. Hasta ese momento, todos los operadores genéticos utilizados eran de carácter genérico y no específico a la optimización en clusters. Los operadores *genéricos o genotípicos* se aplican directamente sobre el genoma de los clusters (tanto binarios como reales) sin tener en cuenta lo que estos expresan. En cambio, los operadores propuestos por Deaven y Ho pertenecen a otra clase de operadores conocida con el nombre de *físicos o fenotípicos*, ya que estos se aplican a lo que el genoma expresa (en este caso la estructura tridimensional del cluster) y no al genoma en sí. Un ejemplo de este último tipo de operadores es el *cut and splice*, que intercambia dos mitades pertenecientes a dos clusters para crear otros dos (ver figura 2.2).

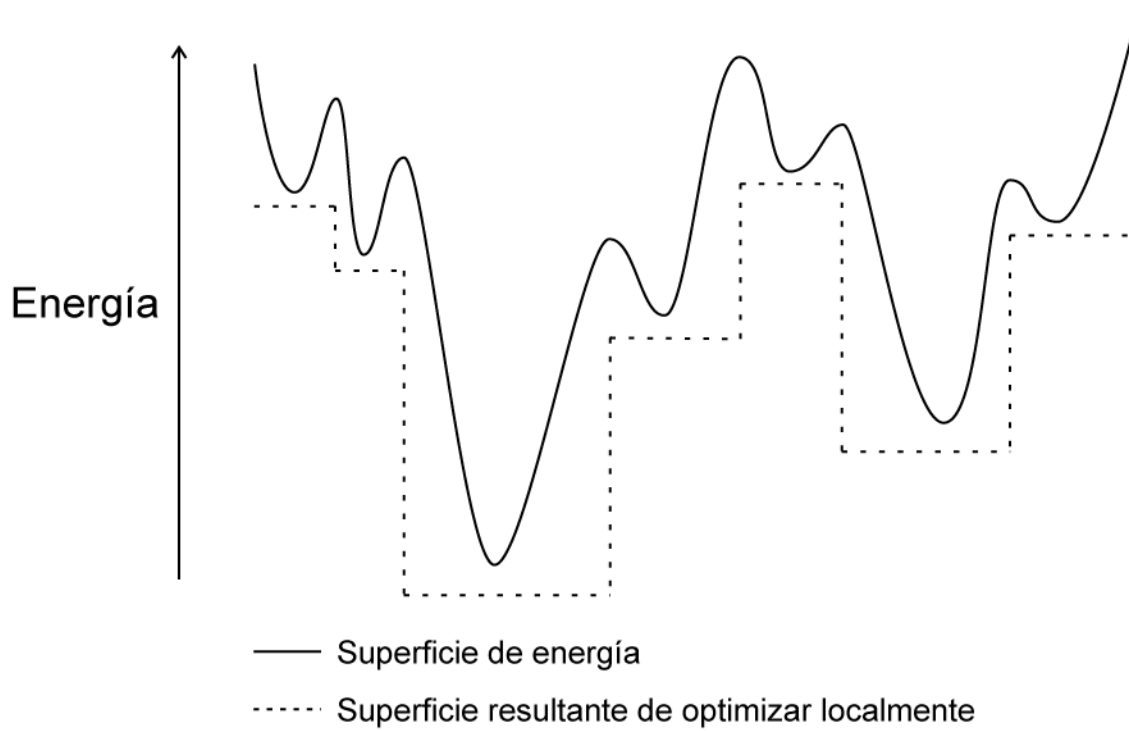


Figura 2.1: Efecto sobre la superficie de energía al aplicar la optimización local.

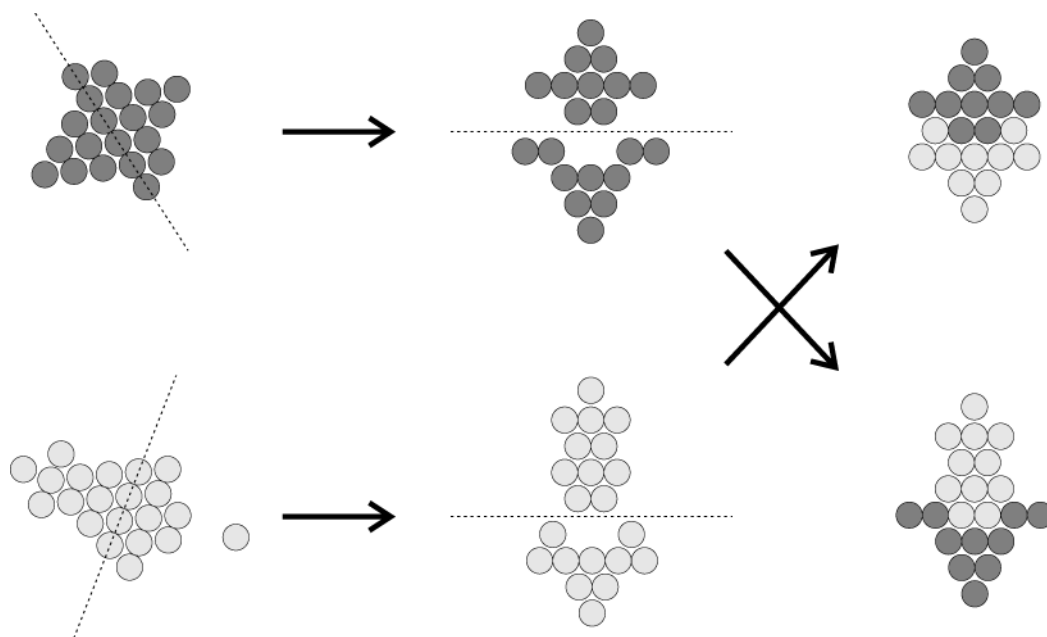


Figura 2.2: Representación del operador *cut and splice*.

En el caso de cristales, sólo se han desarrollado tres estrategias que utilizan GA. La primera fue creada por Woodley et. al. [77] y fue diseñada para ser utilizada en cristales inorgánicos mientras que las otras dos, centradas en cristales orgánicos, fueron desarrolladas en forma independiente y al mismo tiempo por Motherwell [54] y Bazterra et. al. [8, 46, 47, 48, 49]

En el caso de Bazterra et al., el método fue inspirado en considerar al cristal como un cluster molecular que se repite periódicamente. Primero se realizó una extensión del MGA desarrollado por White et. al. [76] utilizado en clusters moleculares, para el caso de cristales moleculares [46]. Este algoritmo se basaba en la búsqueda sin simetrías utilizando operadores genéticos genéricos y sólo fue diseñado para los cristales de benceno, naftaleno y antraceno [47, 48]. El siguiente paso fue la inclusión de una interfase con CHARMM [78]

para la evaluación y optimización de la energía del cristal, además de implementarse la primera versión del algoritmo en paralelo. Esto dió lugar a la posibilidad de extender el algoritmo a todo tipo de moléculas, de incluir una mayor variedad de potenciales clásicos y de implementar la búsqueda con simetría [49].

La experiencia adquirida en estas implementaciones mostró algunos problemas si las optimizaciones globales son realizadas sin considerar que los parámetros que definen un cristal se encuentran relacionados entre sí. Esto se debe a que puede existir una combinación de valores para estos parámetros incompatibles con la definición de cristal. Por ejemplo, para el caso de simetría P1, los valores de las coordenadas de las moléculas deben estar acotados dentro de la celda unidad. En el caso de que esto no se cumpla, existe la posibilidad de superponer en la misma posición moléculas de la celda unidad con las moléculas *imágenes*, producto de la repetición periódica de la celda unidad (ver figura 2.3). Estos cristales mal formados introducen toda clase de inestabilidades numéricas al ser optimizados localmente, por lo que en general deben ser reemplazados por otros cristales generados aleatoriamente. Empíricamente se observó que el número de estructuras reemplazadas (mal formadas) era tan grande, que hacía que el método no se distinguiera demasiado de los método por búsqueda aleatoria. Este fenómeno era particularmente importante (se reemplazaban un gran número de estructuras) para el caso de la búsqueda con simetrías, en especial para aquellos casos altamente simétricos.

Estos problemas mostraron la necesidad de un nuevo algoritmo que previniera la creación de cristales mal formados. El mismo constituye uno de los pilares de esta tesis y fue utilizado

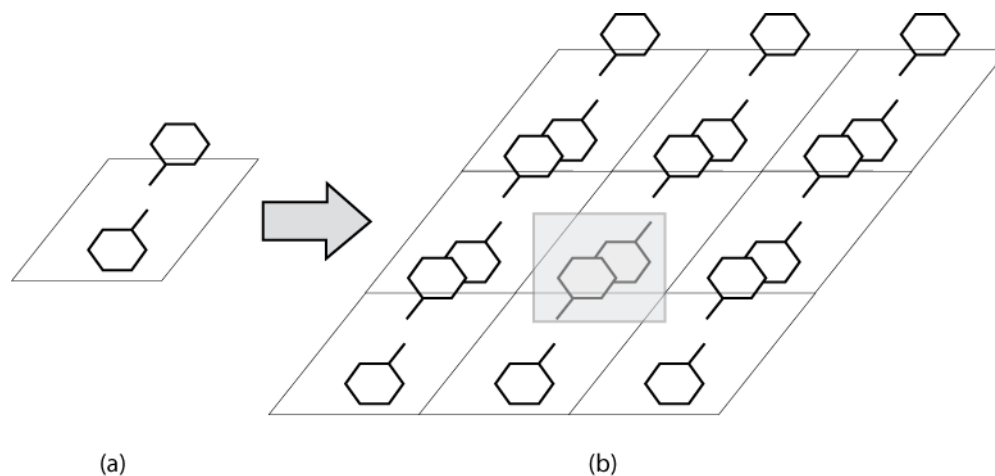


Figura 2.3: Ejemplo de un cluster molecular y una celda unidad en P1 que genera un cristal mal formado (a), en donde las moléculas de la celda unidad se superponen con sus imágenes (b).

por primera vez en el experimento CSP2004 [8].

2.4. Proposición de un nuevo método

A diferencia de otros métodos como templado simulado, el uso de GA en la predicción de estructuras cristalinas ha sido escaso. En particular, los únicos dos métodos asociados con el uso de GAs son el desarrollado por Motherwell [54] y el desarrollado en esta tesis [8, 46, 47, 48, 49]. Esto contrasta enormemente con la predicción sobre otros sistemas, como en el caso de clusters atómicos y moleculares, en donde se ha probado que el uso de GAs es el método más eficiente, resultando superior a otras alternativas como templado simulado y dinámica molecular [22, 20].

Por lo tanto, aquí se presenta un método basado en GAs para la predicción de estructuras

cristalinas, en donde se exige además que se verifiquen las siguientes condiciones:

1. que permita la búsqueda con o sin simetría
2. que acople tanto grados de libertad intramoleculares como intermoleculares en la optimización global;
3. la implementación computacional debe ser en paralelo para permitir un muestreo apropiado y eficiente del espacio de configuraciones;
4. que exista un procedimiento sistemático para el cálculo de potenciales clásicos que modelen las interacciones entre moléculas;
5. que proponga algoritmos eficientes para el procesamiento y preservación del creciente número de estructuras que se generan.

En los siguientes tres capítulos se describe la implementación de esta proposición en detalle.

Capítulo 3

Algoritmo Genético Modificado para Cristales

3.1. Introducción

En este capítulo se describen los algoritmos genéticos modificados para cristales (Modified Genetic Algorithm for crystals o MGAC), que es el método de optimización global que se desarrolló en esta tesis.

3.2. Genoma de un cristal

El primer paso para definir el algoritmo es describir cómo se codifica la información del cristal. Un cristal se describe mediante los parámetros de su celda unidad y la posición de todos los átomos dentro de la misma. La *celda unidad* se define como el bloque de átomos

que al ser repetido periódicamente forma el cristal. En el caso que el cristal presentase más simetrías que la traslacional (simetrías de punto), las posiciones de los átomos solamente necesitan ser definidas en una porción de la celda unidad conocida con el nombre de *celda asimétrica*. El resto de las posiciones dentro de la celda unidad están dadas por las transformaciones de simetría que presenta el cristal. La combinación entre simetrías de punto y traslacionales forman lo que se conoce por *grupos espaciales* de los cristales [59].

Por lo tanto, para definir un cristal completamente es necesario conocer los parámetros de la celda unidad, las posiciones de los átomos dentro de la celda asimétrica y por último el grupo espacial al cual pertenece. En el caso de cristales moleculares es posible reducir el problema de definir la posición de todos los átomos, si se considera que existen porciones de la molécula que son rígidas. Esto se debe a que las moléculas orgánicas están unidas a través de enlaces covalentes, que definen buena parte de sus geometrías y que además son 1 a 3 órdenes de magnitud más fuertes que el resto de las fuerzas intermoleculares. Esto implica que la presencia de varias moléculas en un mismo entorno no produce cambios significativos en la geometría de las mismas. La única excepción que se observa a esta regla es en el caso de uniones covalentes simples, en donde es relativamente fácil rotar fragmentos de la molécula que se encuentren conectados a través de estos enlaces. Por lo tanto, para especificar las coordenadas de todos los átomos, sólo se necesita conocer la posición del centro de masa de todas las moléculas, sus orientaciones y los valores de los ángulos diedros para cada una de las uniones simples.

Si se consideran N moléculas presentes en la celda asimétrica, el genoma del cristal se

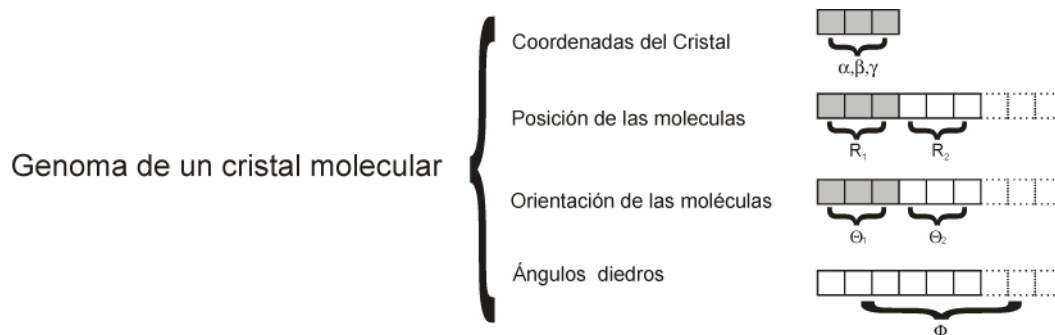


Figura 3.1: Representación esquemática del genoma de un cristal

define dando los parámetros $\{\alpha, \beta, \gamma\}$ para la celda unidad, la posición de los centros de masa de las moléculas $\{\mathbf{R}_1, \dots, \mathbf{R}_N\}$, sus orientaciones $\{\Theta_1, \dots, \Theta_N\}$ y los valores de los ángulos diedros $\{\Phi_1, \dots, \Phi_{NM}\}$ de los enlaces simples de cada molécula (suponiendo que hay M ángulos diedros por molécula). Cada tipo de parámetros se mantiene en diferentes *sectores* del genoma (ver figura 3.1). Esto permite que los operadores genéticos intercambien información entre categorías similares de parámetros y no se mezcle información que codifica, por ejemplo, ángulos con distancias.

Como puede observarse, los valores a, b, c (las dimensiones de la celda unidad) no son parte del genoma y no forman el conjunto de parámetros que son optimizados por el algoritmo genético. Esto se debe a que estos parámetros quedan aproximadamente definidos según la extensión espacial del cristal, que depende de las interacciones entre las moléculas. Los valores de a, b, c son elegidos de manera tal que el cristal resultante encierre al grupo de N moléculas en su celda asimétrica. Luego estos ejes se alargan para que la distancia

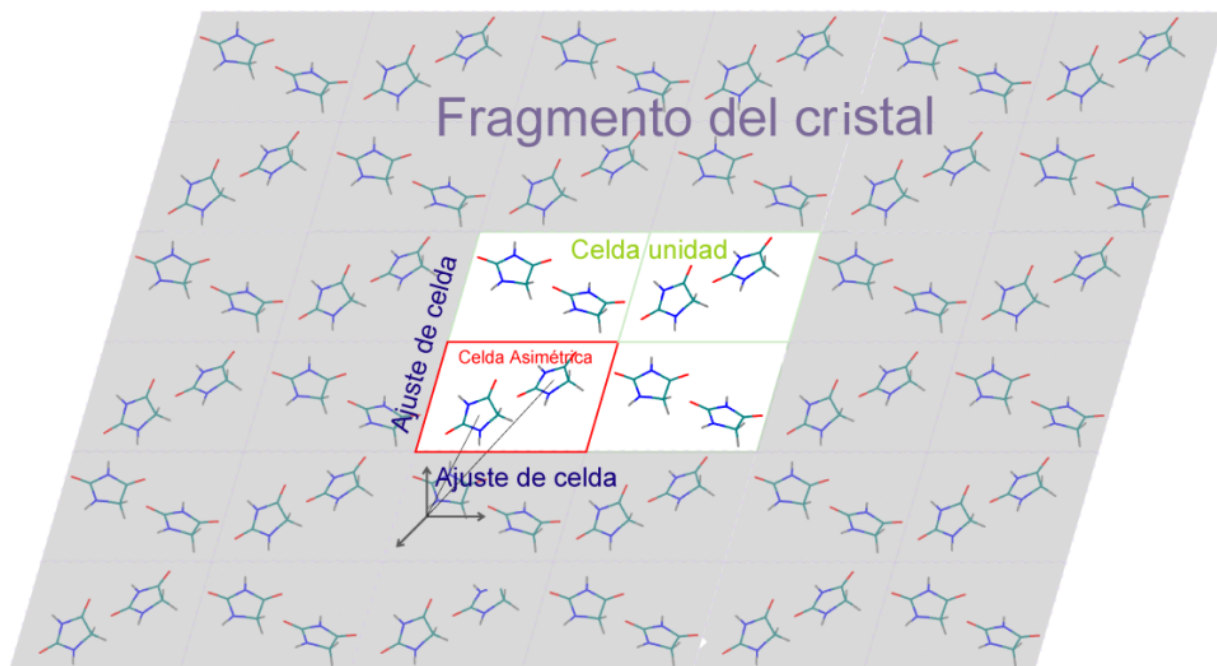


Figura 3.2: Representación esquemática del cristal generado por el genoma.

mínima entre las moléculas en la celda asimétrica y sus imágenes sea mayor a una distancia definida de antemano (por lo general 3 \AA). Esta elección de a, b, c evita el problema de generar cristales mal formados, por lo que el número de estructuras reemplazadas se reduce dramáticamente. Finalmente, todos estos valores terminan de ajustarse cuando el cristal es optimizado localmente, ver figura 3.2. A este procedimiento se lo conoce con el nombre de *fitcell*.

3.3. Descripción del algoritmo

El primer paso en el algoritmo consiste en crear aleatoriamente una población o generación inicial. Para esto las posiciones de las moléculas son generadas eligiendo puntos aleatorios de un cubo de un dado tamaño, que dependerá de la máxima extensión de las moléculas y el número de ellas presente en la celda asimétrica. En forma similar, las orientaciones de estas moléculas son elegidas también aleatoriamente, así como la posición de los ángulos diedros sobre los enlaces simples. En este último caso es posible definir limitaciones en la selección de los ángulos. Esto es útil cuando las moléculas presentan enlaces simples sobre ciclos cerrados, por lo que sus ángulos diedros no pueden tomar valores arbitrarios sin romper el ciclo. Por lo tanto, cada parámetro tiene asociado un intervalo de valores que son utilizados en el proceso de inicialización aleatoria, y como se describirá más adelante, en la definición del operador de mutación.

Una vez producido el cluster molecular, el cristal se forma ajustando la celda unidad para que este conjunto de moléculas esté contenido dentro de la celda asimétrica, para luego ser optimizado localmente. Esta operación se repite sobre todos los individuos presentes en la población inicial.

El siguiente paso es la elección del 50% de los cristales de la población inicial a través del método estocástico conocido con el nombre de *Roulette Wheel* [4]. Cada cristal presenta una probabilidad de ser elegido proporcional a

$$p_i \propto \frac{E_{max} - E_i}{E_{max} - E_{min}} \quad (3.1)$$

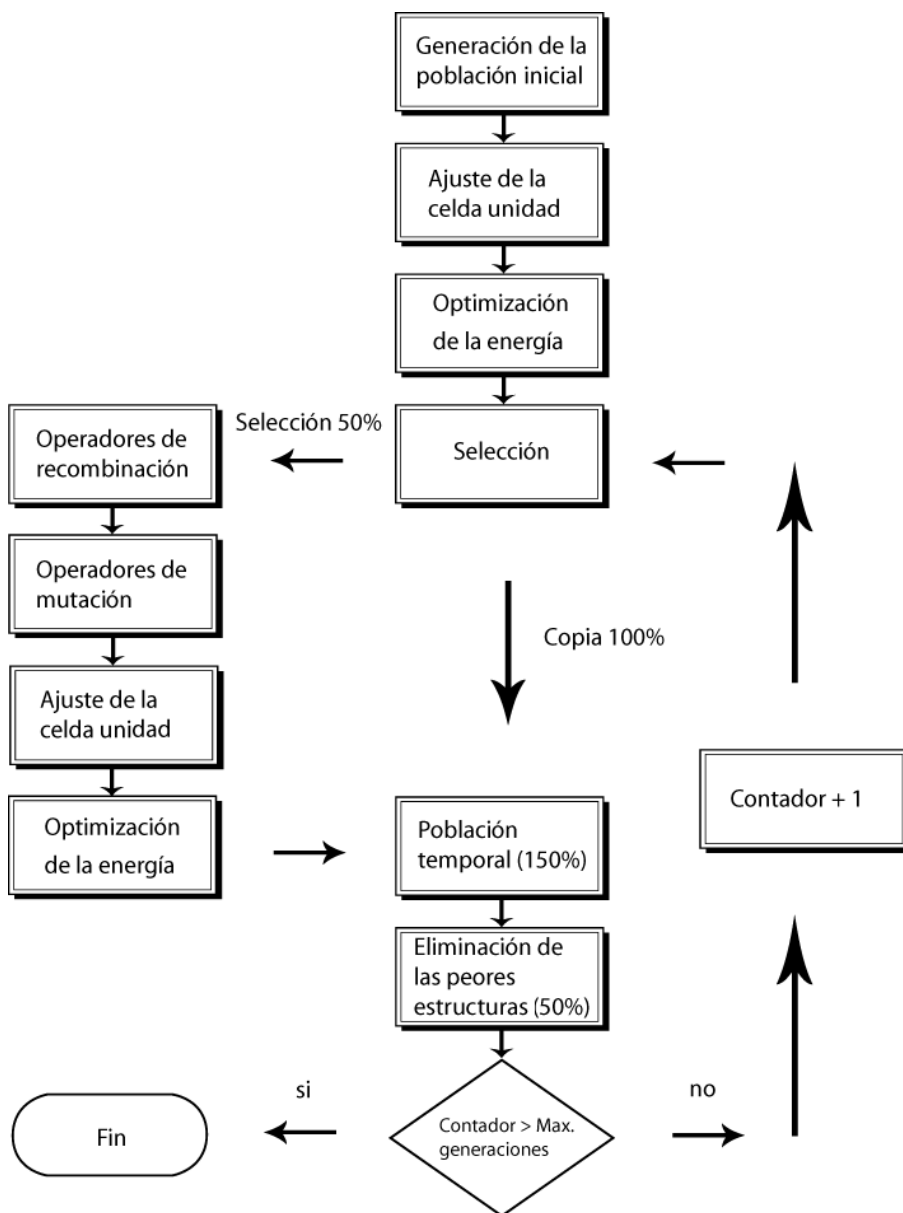


Figura 3.3: Diagrama de flujo para el MGA

donde E_i es la energía total del cristal y E_{max} , E_{min} son la máxima y la mínima energía presentes en la población. Todas estas energías son calculadas una vez que los cristales han sido optimizados localmente. Se puede observar que los cristales de menor energía son elegidos con mayor frecuencia. Estos cristales son utilizados luego para aplicarles los operadores de recombinación y mutación para la producción de nuevos cristales. A estas estructuras resultantes se les ajusta la celda unidad usando el proceso de fitcell y se las optimiza localmente para luego introducirlas en una población temporal compuesta por estas estructuras más las estructuras de la población inicial. Esta población es 50% más grande que la inicial (100% de la inicial y 50% de las nuevas estructuras), por lo que se reduce hasta el mismo tamaño de la población inicial eliminando los cristales de mayor energía. La población resultante define una nueva generación, que puede ser considerada como la población inicial y repetir el proceso descrito arriba por un número predefinido de generaciones, ver figura 3.3.

3.4. Operadores Genéticos

La función de los operadores genéticos es el intercambio y la creación de la información presente en cada generación. Existen dos tipos de operadores: de recombinación y de mutación [4]. Los operadores de *recombinación* toman dos estructuras para crear otras dos a través del intercambio de la información presente en las estructuras a la entrada del mismo, mientras que los operadores de mutación toman una sola estructura a su entrada, que se modifica para crear una nueva estructura.

Los operadores de recombinación utilizados para los cristales son: *recombinación a un punto*, *a dos puntos*, *a N puntos*, *uniforme*, *media aritmética*, *media geométrica* e *inversión*. Estos operadores han probado ser de gran eficiencia en la predicción de clusters moleculares [76]. Los detalles de implementación para cada uno de los operadores están dados en el apéndice A.

Cada operador se aplica por separado en cada sector del genoma del cristal (ver figura 3.1), de forma tal que la información intercambiada pertenece a la misma categoría. Esto evita que las coordenadas de los centros de masa moleculares, que se miden en unidades de distancia, se mezclen con los ángulos diedros de las moléculas que se miden en grados.

El tipo de operador aplicado se elige al azar entre los operadores previamente nombrados. En el caso de los ángulos de la celda unidad, el operador de *inversión* es eliminado de la lista de operadores enumerada arriba. Esto se debe a que el mismo puede dar lugar a cambios en el sistema cristalino, produciendo por lo general cristales mal formados, ya que cada grupo espacial se encuentra definido dentro de un tipo de sistema cristalino en particular.

La mutación es implementada por un operador conocido con el nombre de *mutación gaussiana*. Este operador modifica a un número dado de parámetros en el genoma sumando los valores resultantes de una variable aleatoria con distribución normal. Esta distribución tiene como valor medio cero y una desviación estándar igual a un sexto del intervalo asociado a cada parámetro. La misión de este operador es la de producir nuevas estructuras que contengan información no presente en la población. Esto permitiría en principio la construcción de estructuras que no serían posibles con sólo utilizar operadores de recombi-

nación.

La frecuencia con que se aplican los operadores de recombinación y mutación no es la misma. Esto se debe a las diferentes funciones que se les adjudica. Por un lado los operadores de recombinación *explotan* la información existente en la generación previa, mientras que la mutación *explora* nuevas estructuras, introduciendo nueva información en las futuras generaciones. Es por eso que generalmente todas las implementaciones de GAs acostumbran a asignar una mayor frecuencia a los operadores de recombinación que a los operadores de mutación. De esta forma se privilegia una búsqueda con *memoria*, la cual aprende de las estructuras previamente generadas para construir las nuevas generaciones. Por ejemplo, para la mayor parte de las simulaciones realizadas en esta tesis, se asignó una probabilidad de 1.0 para la aplicación de operadores de recombinación y 0.001 para la mutación. Esto implica que entre todas las estructuras elegidas para ser modificadas por los operadores genéticos, 100 % son modificadas por los operadores de recombinación, mientras que sólo el 0.1 % del total de todos los parámetros presentes en el genoma son modificados por mutación.

Es importante notar que el cálculo de energía es fundamental, ya que es el parámetro que guía la búsqueda para encontrar nuevos cristales. Si las estructuras no son correctamente ordenadas en función de sus energías, puede suceder que la frecuencia con que se eligen los cristales no privilegie la creación del cristal experimental. Este sería el caso en el que el cristal experimental no se encuentra entre los mínimos principales de la superficie de energía. Por lo tanto, el cálculo de la energía debe ser lo suficientemente preciso como para

generar una superficie de energía en donde la estructura experimental sea prácticamente su mínimo global.

3.5. Implementación del MGAC

Para aumentar el número de estructuras que se evalúan por cada molécula, lo que implica una mayor eficiencia para muestrear los posibles cristales, el MGAC se implementa utilizando un esquema de *paralelización global*. El mismo permite la optimización simultánea de un dado número de cristales pertenecientes a una misma generación usando APGA (Adaptive Parallel Genetic Algorithm) [79]. El APGA fue diseñado para aprovechar en forma eficiente los recursos de clusters de computadoras heterogéneas. Además el MGAC contiene un número de herramientas que permite su ejecución simultánea y en forma distribuida sobre varios clusters de computadoras.

El MGAC tiene como interfase un archivo de entrada. La sintaxis del mismo es en XML [80] y se define a través de un esquema para XML [81] llamado mgacML. A su vez, los archivos de salida están definidos en CML [82], que es un dialecto de XML diseñado para la transferencia y preservación de información química. Ésto permite que tanto los archivos de entrada como de salida puedan ser validados en función de sus esquemas, tal que es posible detectar en forma automática si los archivos han sido corrompidos o no. Este proceso de validación es importante para detectar errores que aparecen a menudo en ambientes distribuidos.

La implementación del MGAC está escrita en C++, en donde se utiliza BASH [83] para

realizar múltiples ejecuciones, MPICH [84] para la implementación del APGA, GALib [85] para el desarrollo de los algoritmos genéticos y xerces-c [86] para leer, escribir y validar archivos en XML. Más información sobre el MGAC puede encontrarse en la referencia [87].

Capítulo 4

Cálculo de la energía.

4.1. Cálculo de la energía utilizando GAFF

Como se observa en las secciones 1.2 y 2.1 es indispensable el uso de potenciales empíricos en la predicción de estructuras cristalinas. En esta tesis se eligió al GAFF (General Amber Force Field), un potencial clásico con el suficiente número de parámetros para modelar las interacciones entre la mayoría de las moléculas orgánicas y de interés farmacológico compuestas por C, H, N, O, S, P y halógenos [88]. La forma funcional de este potencial está dada por una simple fórmula para el cálculo de la energía dada por

$$E = \sum_{uniones} k_r(r - r_0)^2 + \sum_{ángulos} k_\theta(\theta - \theta_0)^2 + \sum_{diedros} \frac{v_n}{2} [1 + \cos(n\phi - \gamma_n)] + \sum_{I < J} \left[\frac{A_{IJ}}{R_{IJ}^{12}} - \frac{B_{IJ}}{R_{IJ}^6} + \frac{q_I q_J}{\epsilon R_{IJ}} \right] \quad (4.1)$$

en donde r es la longitud de cada enlace covalente, θ es el ángulo formado entre dos

uniones covalentes entre tres átomos, ϕ es el valor del ángulo diedro entre dos planos formados por tres uniones covalentes entre cuatro átomos. Los índices I y J designan aquellos átomos que no están *unidos covalentemente*, siendo así R_{IJ} la distancia entre ellos. En la práctica estos átomos se definen como aquellos que se encuentran separados por cuatro o más uniones covalentes o que no pertenecen a la misma molécula.

Los primeros dos términos modelan las distancias y los ángulos entre uniones. Los parámetros r_0 y θ_0 corresponden, respectivamente, a las distancias y ángulos de la estructura de equilibrio. Se utiliza un potencial cuadrático definido por las constantes k_r y k_θ para modelar todo movimiento que se aparte del equilibrio.

Los términos v_n y γ_n son los coeficientes que modelan los ángulos diedros. Como su contribución a la energía es periódica, la forma más general de escribir este término es a través de una serie de Fourier.

Los términos A_{IJ} y B_{IJ} modelan las contribuciones de Lennard-Jones (LJ) producto de las interacciones dipolo-dipolo generadas por auto-inducción. Estos términos no dependen de cómo los átomos se encuentran unidos, y se extienden más allá de una molécula, siendo una de las contribuciones a las interacciones intermoleculares. Los parámetros q_I y q_J modelan las interacciones electrostáticas y forman la segunda contribución a las interacciones intermoleculares.

A diferencia de otros potenciales genéricos tales como DREIDING [89] o UFF [90], GAFF ha sido diseñado con un gran número de tipos de átomos (35 en total) para describir una mayor variedad de ambientes químicos que pueden estar presentes en moléculas

orgánicas. Esto tiene como consecuencia un aumento considerable del número de parámetros necesarios para describir todas las posibles interacciones. Es por eso que GAFF incorpora a su vez un conjunto de reglas empíricas y heurísticas para estimar aquellos parámetros que no han sido definidos explícitamente dentro del potencial. Esto permite alcanzar un compromiso entre el número de parámetros necesarios para definir el potencial y el nivel con que se describe a las moléculas.

Debido a que GAFF está definido por un conjunto de parámetros y reglas, la aplicación consistente del mismo requiere de un generador de parámetros para cada tipo de molécula. Este generador es implementado en el programa *antechamber* creado por Junmei Wang [91]. El mismo toma como entrada las coordenadas atómicas de una molécula y genera la serie de parámetros necesarios para modelarla.

Como se observó al comienzo, las contribuciones electrostáticas al potencial están dadas por las interacciones Coulombianas de un conjunto de cargas puntuales ubicadas en las posiciones atómicas. Las mismas son ajustadas de manera tal que el potencial total resultante sea el mismo que el de la molécula en fase gaseosa, calculada en general a través de algún método ab-initio o DFT. En el caso de GAFF, el cálculo del potencial electrostático de una molécula aislada se realiza a nivel HF/6-31G* o AM1-BCC según cuál sea el tamaño de la molécula. El ajuste se obtiene utilizando el programa RESP (Restrained ElectroStatic Potential) [92], mientras que para el cálculo del potencial electrostático a nivel HF/6-31G* o AM1-BCC se utiliza el programa Gaussian [93] o *MOPAC6* [94], respectivamente.

Una vez que el potencial se encuentra definido, es posible calcular la energía de cualquier

crystal utilizando el programa CHARMM [78]. Para esto se creó una extensión del programa antechamber llamada *charmmgen*, que permite generar el potencial y la topología de una molécula de forma tal que puedan ser leídos por CHARMM.

El cálculo de la energía procede luego de la siguiente forma. Primero el programa genera un fragmento esférico del cristal con un radio dado. En esta tesis se utilizó un radio igual a tres veces el módulo del mayor de los ejes que definen la celda unidad. Esto garantiza que al menos tres de estas celdas sean incluídas en cada dirección. Luego se realiza el cálculo de las contribuciones intra e intermoleculares sumando estas últimas, hasta un cierto radio de corte (normalment hasta 14 Å) para el caso de las contribuciones de Van Der Waals. Las contribuciones electrostáticas se calculan utilizando sumas de Ewald que emplean el algoritmo PME (Particle Mesh Ewald) [95] para la suma sobre el espacio recíproco.

4.2. Optimización local de la energía

Se entiende como *optimización local* a los procedimientos para hallar mínimos sobre la superficie de energía. Aunque existe una gran variedad de métodos propuestos, todos estos comparten una estructura general una vez que se define cómo se calcula la energía y/o su gradiente.

Una *optimización local* de un potencial $V(x)$ es la composición l_V^n de un operador l_V , siendo n el número de veces que se aplica l_V tal que cumpla con las siguientes condiciones:

1. el operador $l_V(x)$ depende del valor de la función y su gradiente en el punto x .

2. para cada mínimo local x_a de V se cumple $l_V(x_a) = x_a$.
3. existe un entorno sobre cada mínimo local Λ_a tal que $l_V(\Lambda_a) \subseteq \Lambda_a$.
4. existe n_0 y un conjunto $\Gamma_a \supseteq \Lambda_a$ tal que $l_V^n(\Gamma_a) \subseteq \Lambda_a$ para todo $n \geq n_0$.

Los métodos de optimización local son básicamente iterativos. Como tales existe un conjunto de condiciones que detiene el proceso una vez cumplidas. Estas condiciones son elegidas de manera tal que el método se detenga cerca o sobre un mínimo local. El conjunto Λ_a representa aquellos puntos que cumplen estas condiciones, es decir que se consideran suficientemente cerca del mínimo. El conjunto Γ_a representa todos los puntos iniciales en donde la optimización converge a aquellos puntos que se consideran suficientemente cerca del mínimo local. A este conjunto se lo conoce por *zona de atracción* de un mínimo.

Existen varios algoritmos que utilizan el valor de la función, sus primeras derivadas o sus primeras y segundas derivadas. En general, los algoritmos más eficientes son aquellos basados en la primera derivada de la función, considerando que la misma puede ser calculada analíticamente y no por diferencias finitas. La versión de memoria limitada del algoritmo Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) [96, 97, 98, 99, 100, 101] desarrollada por Nocedal se considera el método más eficiente para grandes sistemas y se basa en la aproximación de cuasi-Newton-Raphson. El mismo utiliza una aproximación de la inversa de la matriz Hessiana, que se construye usando información de los gradientes de un número dado de puntos de las iteraciones previas del algoritmo.

Otras dos alternativas comúnmente utilizadas son el método de Gradiente Conjugado (GC) [102, 103] y el método de inversión directa del espacio iterativo (GIIDS) [104, 105,

[106]. Se considera en general que L-BFGS es más eficiente que los métodos de gradiente conjugado, pero algunas implementaciones sofisticadas de GIIDS pueden ser competitivas [107].

Todos estos métodos trabajan calculando una serie de direcciones por las cuales se lleva a cabo la optimización. Sobre estas direcciones se emplean algoritmos de búsqueda para encontrar aquellos puntos que hacen mínima esta función. En otras ocasiones se utiliza la aproximación de radio de confianza [103]. En el mejor de los casos, el número de direcciones para alcanzar el mínimo es igual a la dimensión del dominio de la función, por lo que la complejidad en tiempo crece al menos proporcionalmente con el número de átomos.

En el caso de CHARMM, la optimización se realiza utilizando el algoritmo ABNR (adapted basis Newton-Raphson) [108]. Este método, aunque anterior a L-BFGS, comparte la misma estrategia básica. Se basa en el método de Newton-Raphson, pero aplicado a un subespacio expandido por las coordenadas de los últimos desplazamientos realizados por el algoritmo.

4.3. Validación en estructuras cristalinas conocidas

En esta sección se valida el uso de GAFF para la evaluación de energía en cristales, midiendo su habilidad para reproducir las estructuras de cristales conocidos experimentalmente. Tomando a un conjunto de cristales, se extrae a la molécula de cada uno de ellos y su geometría se utiliza para generar todos los parámetros asociados a GAFF. Luego se ejecuta una optimización local tomando como punto de partida a las estructuras experi-

Código	Comercial, CSD/CSP, IUPAC, InChI, Aplicación
001	LALNIN L-alanina InChI=1/C3H8NO2/c1-2(4)3(5)6/h2H,1,4H3,(H,5,6)/t2-/m0/s1
002	DLALNI01 DL-alanina InChI=1/C3H8NO2/c1-2(4)3(5)6/h2H,1,4H3,(H,5,6)/t2-/m0/s1, InChI=1/C3H8NO2/c1-2(4)3(5)6/h2H,1,4H3,(H,5,6)/t2-/m1/s1
003	Molecule IIIX imidazolidina-2,4-diona InChI=1/C3H4N2O2/c6-2-1-4-3(7)5-2/h1H2,(H2,4,5,6,7)
004	Molecule X N-(2-metil-4,5-dinitrofenil)acetamida InChI=1/C9H11N3O5/c1-5-3-8(11(14)15)9(12(16)17)4-7(5)10-6(2)13/h3-4H,1-2H3,(H, 10,13)(H,14,15)(H,16,17) Aspirina, ácido acetilsalicílico.
005	ACSALA01 ácido 2-(acetiloxi)benzoico InChI=1/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12) Analgésico, antipirético.

Tabla 4.1: Lista de moléculas estudiadas a lo largo de la tesis.

mentales. Las diferencias entre los mínimos locales resultantes y las estructuras iniciales dan una medida de la calidad del potencial.

En la tabla 4.1 se muestra la lista de moléculas estudiadas a lo largo de esta tesis. En la tabla se ilustra el nombre utilizado en esta tesis, su nombre comercial, su designación en la base de datos CSD, su nomenclatura IUPAC, su identificación InChI y su aplicación práctica si es que ésta existe.

Para la comparación entre estructuras se eligió un conjunto de parámetros para analizar la influencia del potencial tanto en las interacciones intra como intermoleculares. Primero se muestran los parámetros de la celda unidad. Los mismos son indicadores directos de la extensión espacial del cristal, que depende de cómo se modelan las interacciones inter-

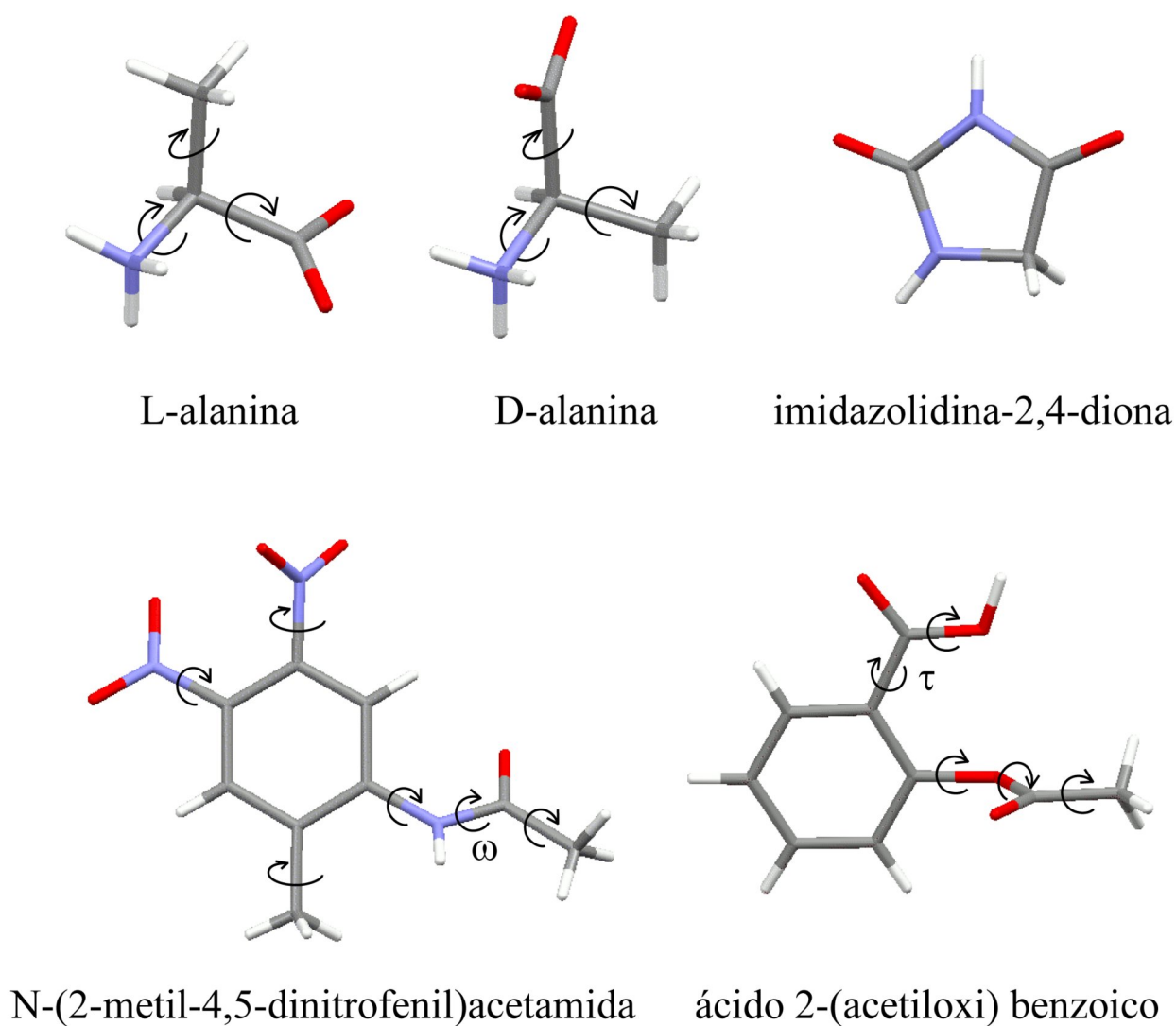


Figura 4.1: Estructuras de las moléculas de la tabla 4.1. Se muestran además los ángulos diedros que confieren flexibilidad a estas moléculas.

Molécula	a Å	b Å	c Å	α °	β °	γ °	volumen Å^3	densidad g/cm^3	mRMS Å	RMS Å
003 (exp.)	9.354	12.176	7.228	90.0	104.59	90.0	797	1.70		
003 (opt.)	8.692 (-7.1%)	11.952 (-1.8%)	7.598 (4.7%)	90.0	101.85 (-2.6%)	90.0	772 (-3.1%)	1.72 (3.0%)	0.053	0.34
004 (exp.)	12.569	4.853	17.2663	90.0	99.16	90.0	1040	1.53		
004 (opt.)	12.5547 (-0.12%)	4.7992 (-1.1%)	17.3526 (+0.45%)	90.0	100.87 (+1.78%)	90.0	1027 (-1.3%)	1.55 (1.3%)	0.076	0.16
005 (exp.)	11.430	6.591	11.395	90.0	95.68	90.0	854	1.40		
005 (opt.)	11.419 (-0.10%)	6.319 (-4.1%)	11.313 (-0.72%)	90.0	94.67 (-1.0%)	90.0	813 (-4.8%)	1.47 (5.0%)	0.072	0.29

Tabla 4.2: Optimización local utilizando GAFF de las estructuras cristalinas experimentales.

moleculares. Otros indicadores son la densidad y el volumen de la celda unidad, ver tabla 4.2.

Para las diferencias entre cristales a nivel de las posiciones atómicas se utiliza la distancia media mínima entre átomos de dos fragmentos cristalinos compuestos por un número dado de moléculas (root-mean-square deviation o RMS). Para más detalle de cómo se implementó esta opción véase el capítulo 5. En nuestro caso se eligió construir fragmentos de 16 moléculas [6], lo cual se muestra en la tabla 4.2. Para estudiar las conformaciones resultantes de las moléculas luego de la optimización, se utiliza nuevamente la raíz cuadrada de la media de la diferencia entre átomos que sólo pertenecen a las moléculas centrales de los fragmentos cristalinos (mRMS), para más detalles ver el capítulo 5.

El promedio de los RMS para todas las predicciones es 0.26 ± 0.05 Å. En el caso de la dimensión de la celda unidad, el promedio del error relativo de sus longitudes es de $2.24 \pm 0.83\%$, mientras que para los ángulos es de $1.79 \pm 0.46\%$. En el caso del volumen de la celda unidad y las densidades de los cristales, el promedio del error relativo es de $3.0 \pm 1.0\%$ y $3.1 \pm 1.1\%$ respectivamente.

La calidad de reproducción del GAFF para estas estructuras es muy alta, si se lo compara con el RMS 0.22 ± 0.015 Å alcanzado por uno de los mejores métodos existentes en la actualidad para evaluar la energía de cristales con moléculas flexibles [109]. Esto es especialmente interesante si se considera que este último método tiene un costo computacional mucho mayor que GAFF.

Capítulo 5

Análisis de las estructuras cristalinas

5.1. Comparación de la estructura tridimensional de dos cristales.

La aplicación del MGAC a una dada molécula produce generalmente una gran colección de estructuras cristalinas, que representa una muestra de mínimos locales en el espacio de configuración. En esta muestra se favorecen aquellas estructuras de menor energía, con el propósito de encontrar aquellos cristales que sean o se encuentren energéticamente cerca del mínimo global. En este procedimiento pueden crearse estructuras geoméricamente similares. El objetivo es extraer de esta colección los primeros n cristales (típicamente $n = 100$ o 200) de menor energía. Para esto se necesita crear un algoritmo que compare la estructura tridimensional de dos cristales para detectar y eliminar aquellos que son similares.

La estructura de un cristal puede ser representada por un fragmento finito del mismo

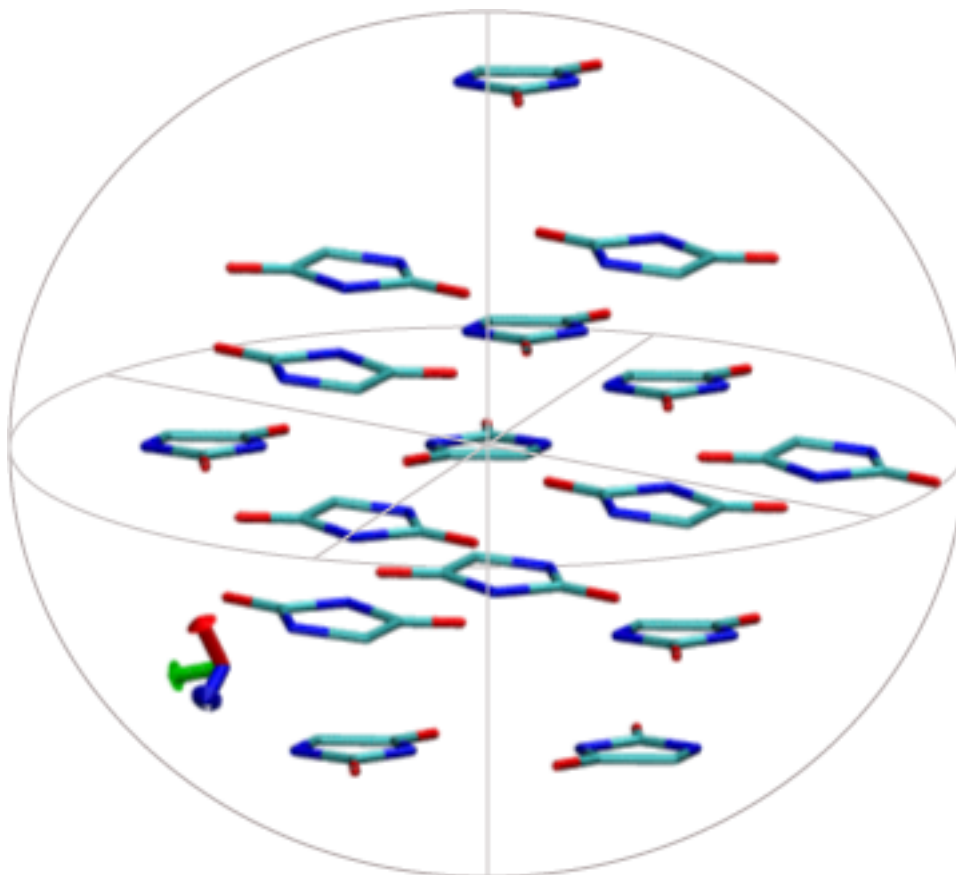


Figura 5.1: Fragmento del cristal compuesto por 16 moléculas de imidazolidina-2,4-diona.

[110]. En el caso de cristales moleculares, este fragmento está dado por una agrupación de moléculas, cuyas posiciones son extraídas del cristal. Para esto, se toma primero como molécula central a una de las pertenecientes a la celda asimétrica. Luego se construye el fragmento agregando las primeras $n - 1$ moléculas más cercanas a la central. La distancia entre moléculas está dada por la distancia entre sus centros de masa. En esta construcción no se incluyen átomos de hidrógeno. El resultado es una agrupación aproximadamente esférica de n moléculas como se muestra en la figura 5.1.

Para comparar dos cristales, que designaremos como de referencia y de comparación, es

necesario primero generar dos fragmentos de tamaño n y m , siendo $m > n$, asociados a los mismos. El tamaño de ambos fragmentos debe ser suficientemente grande para representar los alrededores de cada molécula central. Por lo general, n y m están fijados en 16 y 24 moléculas respectivamente.

Luego, la comparación se realiza aplicando la rotación y traslación que mejor superponga a las moléculas centrales de cada fragmento. Esto se realiza minimizando la raíz cuadrada de la distancia cuadrática media (root-mean-square o RMS) entre las coordenadas atómicas de cada molécula utilizando el algoritmo de Kabsch [111, 112]. Si la diferencia resultante está por debajo de un dado límite (por lo general 0.5 Å), la comparación continúa buscando la mejor rotación y traslación aunque ahora para los fragmentos.

Para la alineación entre fragmentos, es necesario primero asociar para cada molécula del fragmento de referencia la correspondiente molécula en el fragmento de comparación. Esto se debe a que un mismo cristal puede ser creado utilizando diferentes simetrías y/o número de moléculas en su celda asimétrica. Por lo tanto el orden con que las moléculas se generan en cada fragmento no es necesariamente el mismo. Para esta asignación, se aplican primero las mismas transformaciones que se utilizaron en la alineación de las moléculas centrales, sobre los centros de masa de las moléculas pertenecientes al fragmento de comparación. Esta operación, en el caso de que los cristales sean similares, resulta en la superposición de los centros de masa de las moléculas del fragmento de comparación con los de las moléculas del fragmento de referencia, ver figura 5.2. La molécula en el fragmento de referencia que se encuentre más cerca de cada centro de masa es asignada a la correspondiente molécula del

fragmento de comparación. Luego, la comparación entre coordenadas de cada fragmento es posible y los fragmentos son alineados utilizando nuevamente el algoritmo de Kabsch. Si la diferencia entre estos está por debajo de un dado límite (por lo general 1.0 Å) se considera entonces que ambos cristales son similares.

Por el contrario, si la diferencia resultante de la superposición entre las moléculas centrales o entre los fragmentos está por arriba de los límite tolerados, entonces sucede una de las tres alternativas siguientes: las moléculas centrales o fragmentos son similares pero difieren por una inversión, los cristales son similares, pero el de comparación tiene más de una molécula en su celda asimétrica o los cristales no son similares.

En el primer caso, el hecho de que ambos fragmentos difieran por una inversión implica que los cristales son enantiómeros. Esto significa que sus estructuras tridimensionales son imágenes especulares. Por lo tanto si sólo se utilizan rotaciones y traslaciones para la superposición de los fragmentos cristalinos, el algoritmo resultante es sensible a la estereoquímica del cristal, es decir que el mismo no identifica como similares a dos cristales que son enantiómeros. Ahora no todos los grupos espaciales permiten la posibilidad de que existan enantiómeros. Aquellos grupos espaciales designados como *centro simétricos* producen cristales que son simétricos ante inversión, por lo que no pueden producir enantiómeros. En estos casos es posible encontrar cristales similares en los cuales sus moléculas en la celda asimétrica se encuentran separados por una inversión, por lo que sus fragmentos no pueden ser alineados solo por rotaciones y traslaciones, por lo que se consideraría erróneamente a estos cristales como no similares.

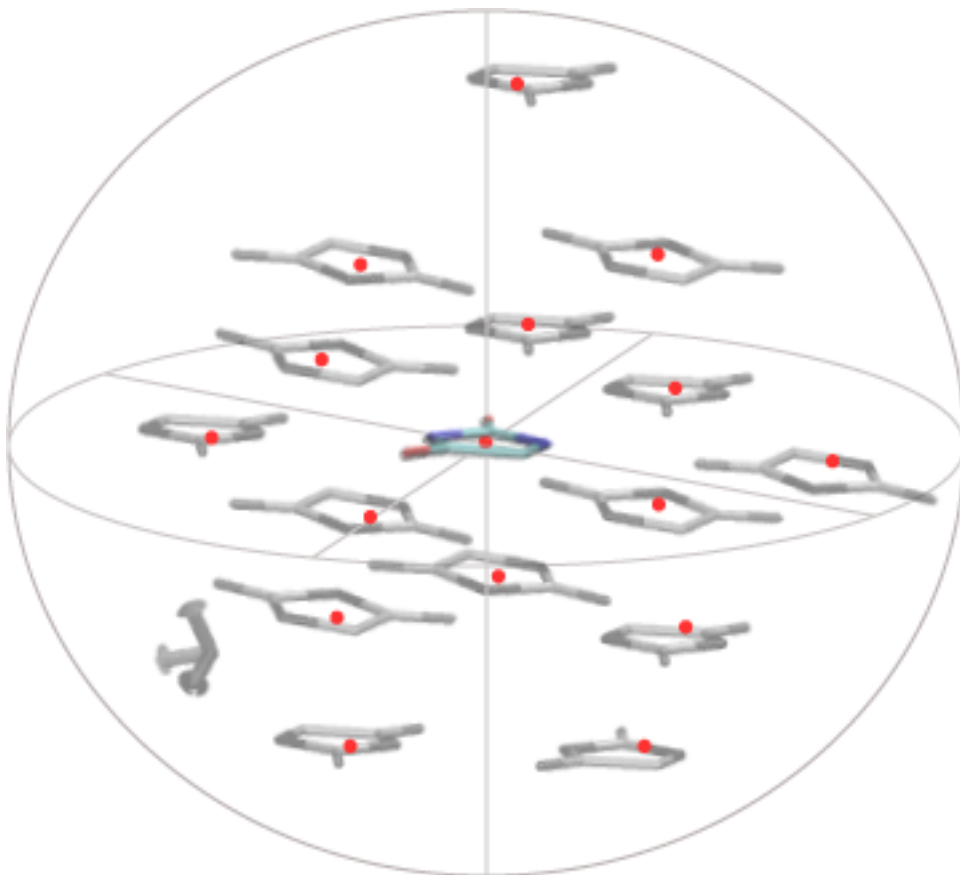


Figura 5.2: Superposición entre el fragmento de referencia (en tonos de grises) y los centros de masas del fragmento de comparación (puntos rojos). Se puede observar que esta superposición se logra a través del alineado de las moléculas centrales de ambos fragmentos.

Para resolver este problema, se invierte el fragmento de comparación en el caso de que no sea posible alinear las moléculas centrales de los fragmentos o a los fragmentos entre sí. Ésto hace que el algoritmo sea en general insensible a la estereoquímica del cristal. Además de esta forma se resuelve el problema de reconocer cristales que pertenezcan a grupos espaciales centro simétricos, y considera en general idénticos a diferentes enantiómeros. Este es un comportamiento deseable en la comparación entre cristales, debido al hecho de que las interacciones entre moléculas sólo dependen de la distancia entre átomos, por lo que la energía de los cristales es invariante ante inversión. Como consecuencia, cada vez que se obtiene un cristal, en un grupo espacial no centrosimétrico, se sabe automáticamente que puede tener un enantiómero de la misma energía.

En el caso que el cristal de comparación presente más de una molécula en su celda asimétrica es posible crear varios fragmentos cristalinos asociados al mismo. Este problema no se presenta en el caso del cristal de referencia, ya que por convención, el fragmento de referencia se construye alrededor de la primera molécula asimétrica. Por lo tanto existe la posibilidad de que dos cristales sean similares aunque sus moléculas en la celda asimétrica no estén enumeradas en el mismo orden. Para reconocer esta situación es necesario intentar el alineamiento, como se describe arriba, entre el fragmento de referencia y todos los posibles fragmentos de comparación por cada vez que una superposición fracasa. Sólo cuando todas estas opciones son exhaustivamente eliminadas y en ningún caso es posible la superposición entre fragmentos se considera que los cristales no son similares.

5.2. Colección de los cristales más significativos

Una vez que se conoce cómo comparar dos cristales, es posible la creación de un filtro que solamente seleccione los primeros n cristales de menor energía. Este filtro está compuesto internamente por una lista que contiene a las estructuras deseadas una vez que éste es aplicado.

El filtro comienza leyendo las estructuras resultantes de haber aplicado el MGAC. Inicialmente la lista interna del filtro se encuentra vacía, por lo que el primer cristal leído es incluido en la misma. De aquí en adelante todo cristal leído, candidato a ser incluido en la lista, se compara con un subconjunto de estructuras preexistentes de la lista. Este subconjunto se define como aquellas estructuras dentro de la lista cuyas energías no difieran en más que una dada energía (por lo general 4 kcal/mol) de la energía de la estructura candidata. Esto permite acotar el número de comparaciones por cristal candidato entre aquellos cristales cercanos en energía y hace que el tiempo requerido para construir la lista sea independiente a su capacidad.

En el caso de que ya exista un cristal en la lista similar al cristal candidato y la energía de este último sea menor que la de la estructura preexistente en la lista, entonces el cristal de la lista se reemplaza por el cristal candidato. En caso contrario, el cristal candidato es rechazado y el contenido de la lista permanece intacto. En el caso de que la lista se encuentre en su máxima capacidad y que la energía del cristal candidato sea mayor que la mayor de las energías presentes en el filtro, entonces el cristal candidato se rechaza sin realizar ninguna comparación.

La implementación del comparador y filtro de cristales fue realizada como una extensión del módulo conocido por CCTBX (Computation Crystallography Toolbox)[113]. A esta extensión se la llama MOLCRY , ya que se refiere a rutinas aplicadas a cristales moleculares. MOLCRY fue implementado usando un modelo híbrido [114] que combina C++ y Python [115]. Se utiliza C++ en aquellas secciones del código que son computacionalmente intensas, por lo que se requiere velocidad y eficiencia. Python se utiliza para integrar las diferentes partes del código en una aplicación coherente. Además, Python es el lenguaje elegido para el desarrollo de un conjunto de aplicaciones para procesar y clasificar automáticamente aquellos cristales que son generados en las simulaciones.

5.3. Procesamiento y clasificación de los cristales

En intentos previos de predicción de cristales se evidenció la necesidad de automatizar y estandarizar el procesamiento por el cual los cristales son procesados y clasificados. El mayor inconveniente proviene del hecho del gran número de estructuras que pueden generarse para cada molécula.

Es por eso que se creó una aplicación que permite el análisis de los cristales e implementa una interfase a una base de datos donde son archivados. El desarrollo se hizo en Python bajo la arquitectura REST [116], que se encuentra implementada sobre HTTP [117] utilizando el paquete computacional Django [118].

Antes de guardar un cristal se envía un pedido a la aplicación definida por su URL (Universal Resource Locator) a través de un programa cliente. Cuando este pedido se recibe,

se crea una nueva lista en la base de datos, a la cual se asigna un número de indentificación. Este número puede ser utilizado para acceder y operar sobre esta lista. Es posible ahora el agregado de cristales a través del envío de archivos escritos en el formato CML [82]. Cuando se recibe un cristal, el programa toma todos los cristales previamente guardados en la base de datos asociada a una lista, para luego filtrar al nuevo cristal utilizando MOLCRY. Una vez que el filtrado se completa, los cristales resultantes son guardados en la base de datos.

La base de datos se implementa por un servidor de MySQL [119] con un simple esquema para guardar la información. Cada lista de cristales se representa por una tabla en la base de datos. La estructura de esta tabla está dada por un campo que describe la lista y otro campo en formato de texto que contiene a los cristales en formato CML. Para reconocer qué cristales se encuentran guardados en cada lista, es necesario identificar a los mismos con un código. Este código tiene que ser una única cadena de caracteres que refleje la composición química del cristal. Además debe generarse automáticamente para reconocer y nombrar todo posible cristal orgánico.

Se utiliza InChi (IUPAC International Chemical Identifier) [120] para nombrar a cada tipo de molécula orgánica. Ejemplos de esta nomenclatura aplicados a las moléculas relacionadas con esta tesis pueden verse en la tabla 4.1. Lo que hace a un cristal químicamente único es la proporción del número de moléculas presentes en el mismo. Por lo que si un cristal tiene moléculas del tipo $\{A, B, C, \dots\}$ en una proporción de $\{1 : 3 : 2 : \dots\}$ al mismo se lo identificará con la cadena

```
[{'InChi de A':1},{'InChi de C':2}, {'InChi de B':3}, ... ]
```


en donde el orden de cada molécula está dado por el orden ASCII de sus respectivos códigos en InChI.

Finalmente, esta aplicación permitir acceder a los resultados en la base de datos a través de un navegador de internet. Esto se debe a que la salida de la misma es en XHTML [121], por lo que puede transformarse en un formato que sea legible tanto para el usuario como para otra aplicación. Las listas de cristales y sus contenidos pueden ser leídas como cualquier otra página de internet.

Capítulo 6

Resultados

6.1. Cristales de L-alanina y de DL-alanina

El MGAC se aplicó por primera vez a una molécula flexible en la predicción de los cristales de L-alanina y DL-alanina [49]. Estas moléculas fueron elegidas debido a que constituyen un sistema adecuado para comenzar a construir un método de predicción. Presentan un reducido número de átomos (solamente 6 átomos pesados), tres uniones simples que confieren a la molécula una gran flexibilidad y existen varios potenciales empíricos que permiten modelar sus interacciones (ver figura 4.1)¹. La molécula de alanina tiene un centro quiral y puede existir como dos enantiómeros designados por L y D. La L-alanina es uno de los 20 aminoácidos utilizados en la biosíntesis de proteínas.

Mientras que el cristal de L-alanina está completamente constituido por moléculas de

¹No se contaba todavía con el potencial GAFF ni con los programas asociados antechamber y charmmgen, al momento en que se realizó esta predicción.

Molécula	a	b	c	α	β	γ	R_{wp}
	Å	Å	Å	°	°	°	
001 (exp.)	6.032	12.343	5.784	90.0	90.0	90.0	0.000
001 (AMBER)	5.758	12.323	5.694	90.0	90.0	90.0	0.412
001 (CHARMm)	5.521	12.483	5.404	90.0	90.0	90.0	1.265
002 (exp.)	12.026	6.032	5.829	90.0	90.0	90.0	0.000
002 (AMBER)	12.768	5.667	5.687	90.0	90.0	90.0	0.655
002 (CHARMm)	12.742	5.554	5.402	90.0	90.0	90.0	1.401

Tabla 6.1: Comparación de las estructuras y los factores R_{wp} de las estructuras optimizadas localmente de L-alanina y DL-alanina para los potenciales CHARMm y AMBER

L-alanina, el cristal DL-alanina es el producido por la mezcla racémica de L- y D- alanina. Tanto el cristal de L-alanina como el de DL-alanina se conocen experimentalmente [122, 123]. Se pretendió probar si el MGAC era capaz de predecir estos cristales. Debido a que L-alanina y DL-alanina cristalizan en los grupos espaciales $P2_12_12_1$ y $Pna2_1$ respectivamente, las simulaciones fueron restringidas a estas simetrías.

Para el cálculo de la energía se analizaron dos tipos de potenciales empíricos diseñados para esta molécula, CHARMm [124] y AMBER [125]. El análisis constó de optimizaciones locales tomando como estructuras de referencia los cristales experimentales de L-alanina y DL-alanina y calculando las diferencias entre estas estructuras y las resultantes de las optimizaciones para ambos potenciales, ver tabla 6.1.

A diferencia de las comparaciones realizadas en el capítulo 4, en la tabla 6.1 se utiliza el factor R_{wp} y no el RMS para la comparación entre estructuras². El factor R_{wp} es una medida de las diferencias entre los espectros de rayos X de polvo calculados teóricamente

²Todavía no se contaba con los programas para la comparación de estructuras cristalinas, al momento en que esta predicción se realizó.

para cada estructura; $y_i(exp)$ y $y_i(opt)$ son las intensidades de los espectros simulados para la estructura experimental y optimizada, y R_{wp} se calcula como

$$R_{wp} = \sqrt{\frac{\sum_i w_i [y_i(exp) - y_i(opt)]^2}{\sum_i w_i y_i^2(exp)}} \quad (6.1)$$

en donde w_i es un parámetro de peso dado por $w_i = y_i(exp)^{-1}$ y $w_i = 0$ en el caso que $y_i(exp) = 0$. Los espectros fueron calculados utilizando el programa PLATON [126]. Analizando la tabla 6.1 se eligió el potencial AMBER para trabajar en estos sistemas, debido a que las estructuras experimentales optimizadas, tanto para la L-alanina como para la DL-alanina, presentan con este potencial las menores desviaciones respecto a sus estructuras experimentales.

Se realizaron tres simulaciones en los grupos espaciales $P2_12_12_1$ y $Pna2_1$ de 30 generaciones y 20 cristales por generación. Esto dió como resultado un total de 960 cristales que se evaluaron para cada una de estas moléculas, de donde se extrajo la estructura de menor energía para cada simetría. Estas estructuras representan las predicciones para cada uno de los sistemas. En ambos casos las estructuras predichas fueron similares a la estructuras conocidas experimentalmente. En las figuras 6.1 y 6.2 se muestra la superposición entre las estructuras predichas y experimentales. Estas superposiciones se realizan sobre fragmentos de 16 moléculas extraídas de los cristales como se describió en el capítulo 5. La comparación de estas superposiciones resultan en un RMS de 0.30 Å y 0.40 Å para los cristales predichos de L- y DL-alanina respectivamente. La comparación completa para estos cristales se reporta en la tabla 6.2.

Molécula	a	b	c	α	β	γ	volumen	densidad	mRMS	RMS
	Å	Å	Å	°	°	°	Å ³	g/cm ³	Å	Å
001 (exp.)	6.032	12.343	5.784	90.0	90.0	90.0	431	1.38		
001 (pred.)	5.703	12.357	5.774	90.0	90.0	90.0	407	1.46	0.081	0.30
	(-5.5 %)	(+0.11 %)	(-0.17 %)				(-5.6 %)	(+5.8 %)		
002 (exp.)	12.026	6.032	5.829	90.0	90.0	90.0	423	1.40		
002 (pred.)	12.785	5.664	5.684	90.0	90.0	90.0	412	1.44	0.17	0.40
	(+6.3 %)	(-6.1 %)	(-2.5 %)				(-2.6 %)	(+2.8 %)		

Tabla 6.2: Calidad de las estructuras cristalinas predichas para la L y DL-alanina.

La calidad de reproducción de los cristales de L- y DL-alanina no es buena, tal como se ve en la tabla 6.2. Comparando las tablas 6.1 y 6.2 se ve que las estructuras reproducidas por el MGAC se asemejan a las resultantes de las optimizaciones locales de las estructuras experimentales. Por lo tanto es el potencial empírico el que introduce mayormente las deficiencias en las reproducciones y no el procedimiento con que los cristales se construyen. Esta deficiencia en el potencial y el hecho de que el mismo no puede ser transferido a otras moléculas impulsó la búsqueda de una alternativa más general al problema de generar los potenciales.

En estas predicciones se utilizó la versión inicial del MGAC [46], que optimizaba los ejes de la celda unidad. Como se explica al final de la sección 2.3, esta característica generaba una gran cantidad de cristales mal formados debido a que los mismos pueden presentar parámetros incompatibles con la definición de cristal dentro de sus genomas. Este problema no era tan grave para los casos de los grupos espaciales $P2_12_12_1$ y $Pna2_1$ que forman cristales ortorómbicos. El problema se acentuó cuando se intentó aplicar el MGAC a otras simetrías asociadas a cristales triclinicos y monoclinicos, para los que la frecuencia con que se producían cristales mal formados crecía enormemente.

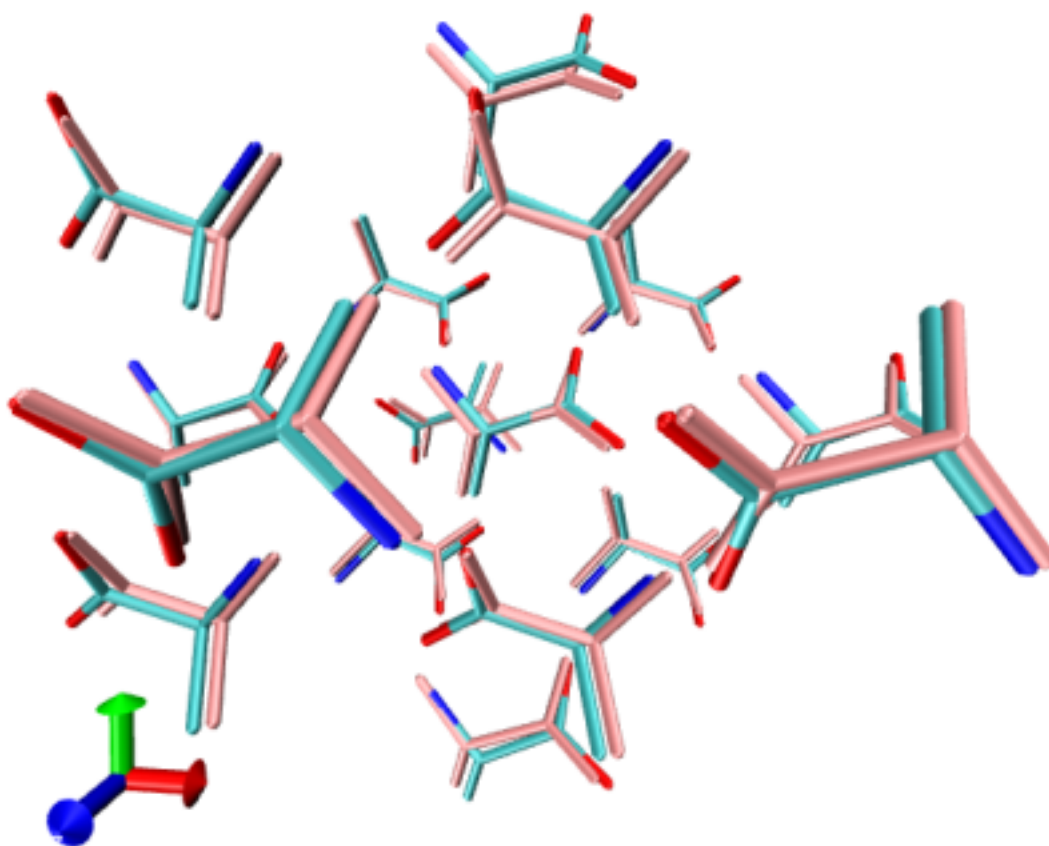


Figura 6.1: Superposición entre la estructura experimental y la predicha (en color rosa) para la molécula de L-alanina. El RMS de la superposición es de 0.30 Å

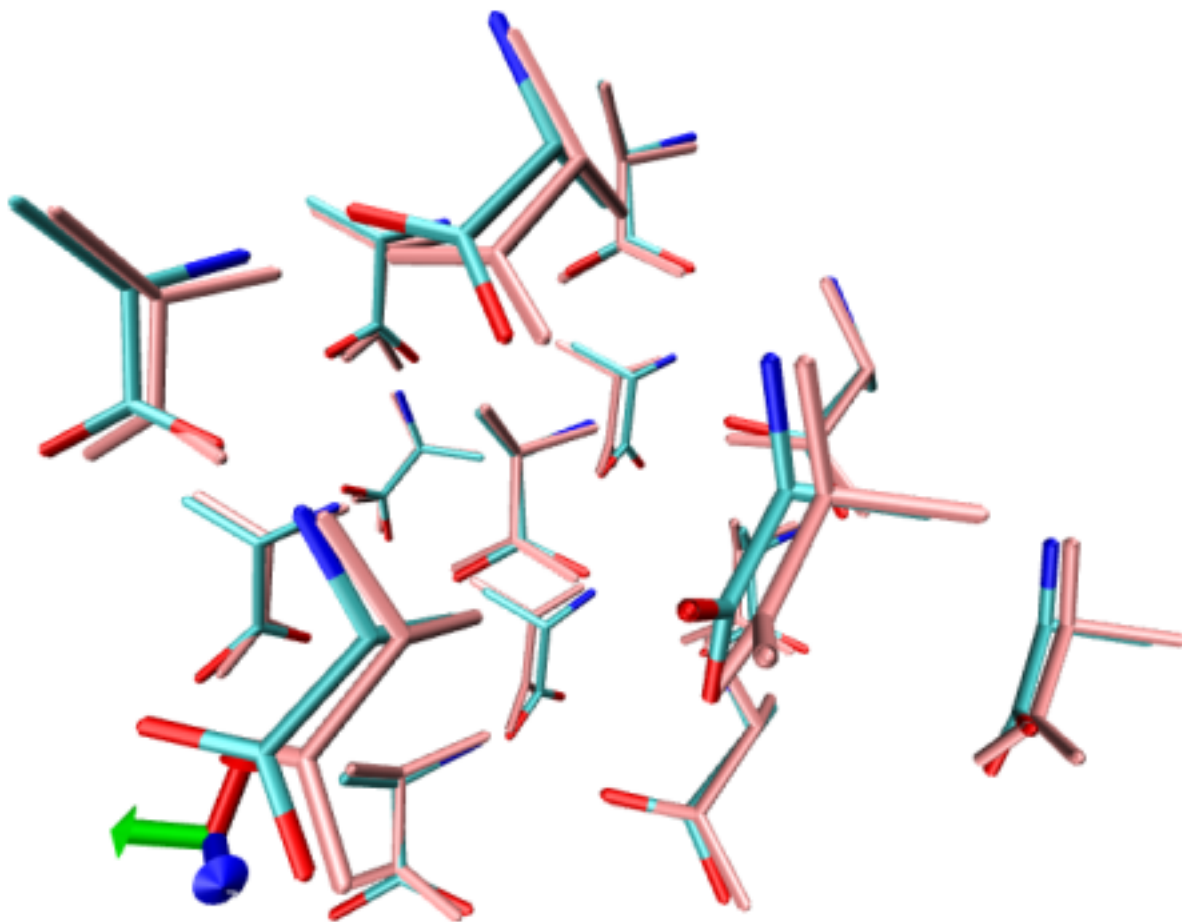


Figura 6.2: Superposición entre la estructura experimental y la predicha (en color rosa) para las moléculas de DL-alanina. El RMS de la superposición es de 0.40 Å

6.2. Cristales de imidazolidina-2,4-diona

La predicción del cristal de imidazolidina-2,4-diona (ver figura 4.1) fue propuesta en la competencia CSP2004 [8], la prueba a ciegas que evalúa las distintas metodologías disponibles para la predicción de estructuras cristalinas [8]. Este fue el primer caso en donde el MGAC no optimizó las dimensiones de la celda unidad, sino que las mismas fueron calculadas con el procedimiento *fitcell* (ver capítulo 3). Esto permitió la extensión del MGAC para búsquedas restringidas sobre un mayor número de grupos espaciales. Como primer paso se implementó la búsqueda sobre los grupos espaciales en los cuales los cristales orgánicos cristalizan con mayor frecuencia.

Para esta molécula se realizaron dos predicciones para los grupos espaciales más comunes ($P1$, $P-1$, $P2_1$, $C2$, Pc , Cc , $P21/c$, $C2/c$, $P2_12_12_1$, $Pca2_1$, $Pna2_1$, $Pbcn$, $Pbca$, $Pnma$) con una y dos moléculas por celda asimétrica. Cada simulación constó de 60 generaciones con 30 cristales por cada generación. Esto dió un número total de 52080 cristales que fueron evaluados para esta molécula.

La energía se calculó usando GAFF³. Las cargas atómicas se calcularon usando el programa RESP, calculando el potencial electrostático de la molécula a un nivel B3LYP/6-31G*.

De las estructuras generadas se extrajeron las 100 de menor energía sin incluir cristales mal formados o similares. Luego estas estructuras se procesaron para la detección de simetrías escondidas utilizando el algoritmo ADDSYM del programa PLATON [126]. En la

³La versión específica de GAFF que se utilizó para esta molécula fue la distribuida por AMBER versión

orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
1	6.064	6.346	11.746	89.96	117.72	89.99	400	P-1	P21/c	0.0000
2	8.769	12.087	7.598	90.00	101.27	90.00	790	C2c		0.0822
3	3.934	12.032	8.439	90.00	100.45	90.00	393	P21c		0.1755
4	15.953	12.056	10.444	90.00	128.33	90.00	1576	C2c		0.1949
5	12.076	10.44	6.289	90.00	90.00	90.00	793	Pbca	Pbcm	0.2589
6	10.33	6.357	12.35	90.00	90.00	90.00	811	P212121	Pbca	0.3084
7	6.19	10.324	6.347	90.00	89.98	90.00	406	Cc	Cmc21	0.3107
8	17.627	6.119	7.598	90.00	104.91	90.00	792	C2	C2/c	0.3453
9	4.996	9.076	9.875	116.28	102.17	91.44	389	P-1		0.3481
10	17.618	6.113	7.596	90.01	104.95	89.99	790	P-1	C2/c	0.3521
11	3.838	6.065	17.093	90.00	99.72	90.00	392	P21	P21/c	0.3956
12	6.189	6.368	11.675	89.97	117.92	89.97	407	P-1	P21/c	0.4177
13	6.369	6.174	10.345	90.00	90.00	90.00	407	Pca21	Pnma	0.4194
14	6.281	20.322	8.85	90.00	134.19	90.00	810	P21c		0.4418
15	6.378	6.619	9.557	90.00	90.00	90.00	403	P21	P212121	0.4541
16	3.925	23.937	8.518	90.00	99.81	90.00	789	P21c		0.4733
17	10.397	6.339	12.058	90.00	90.00	90.00	795	P212121	Pnma	0.5154
18	3.828	6.052	8.637	90.00	98.86	90.00	198	P21	P21	0.5908
19	8.7	6.085	7.585	90.00	99.29	90.00	396	P21	P21/c	0.5916
20	6.113	20.95	6.249	90.00	90.00	90.00	800	P21c	Pbcm	0.6042

Tabla 6.3: Primeras 20 estructuras cristalinas en orden de energía creciente para la molécula de imidazolidina-2,4-diona. La estructura de la segunda fila coincide con la estructura cristalina conocida experimentalmente para esta molécula.

Molécula	a Å	b Å	c Å	α °	β °	γ °	volumen Å^3	densidad g/cm^3	mRMS Å	RMS Å
003 (exp.)	9.354	12.176	7.228	90.0	104.59	90.0	797	1.67		
003 (pred.)	8.769 (-6.2%)	12.087 (-0.73%)	7.598 (+5.0%)	90.0	101.27 (-3.2%)	90.0	790 (-0.88%)	1.69 (1.1%)	0.045	0.33

Tabla 6.4: Calidad de la estructura cristalina predicha para molécula de imidazolidina-2,4-diona.

tabla 6.3 se muestran las primeras veinte estructuras, mientras que la información de las 100 primeras estructuras ordenadas según energía creciente se encuentra en la tabla B.1 en el apéndice B.

En esta prueba a ciegas, no se obtuvo acceso a información de cómo la molécula cristaliza. Una vez que se presentaron los distintos cristales candidatos, se dió a conocer el cristal experimental por parte de los coordinadores del CSP2004 [8], que se comparó con los generados por MGAC. El cristal que aparece en segundo lugar en orden de energía presentó una estructura similar a la experimental. En la figura 6.3 se muestra la superposición entre la estructura predicha y experimental. Esta superposición se realizó sobre fragmentos de 16 moléculas extraídas de los cristales como se describió en el capítulo 5. La comparación de esta superposición resulta en un RMS de 0.33 \AA . La comparación detallada para este cristal puede verse en la tabla 6.4.

El intervalo de la energía para los primeros 100 cristales es aproximadamente de 1.35 kcal/mol, por lo que existen al menos 74 cristales por kcal/mol. Esta distribución de cristales no es uniforme como se muestra en la figura 6.4. En la misma puede verse el rápido crecimiento del número de cristales a medida que la energía crece respecto a la del mínimo global. Esta distribución da una medida de la dificultad inherente a la predicción de estructuras

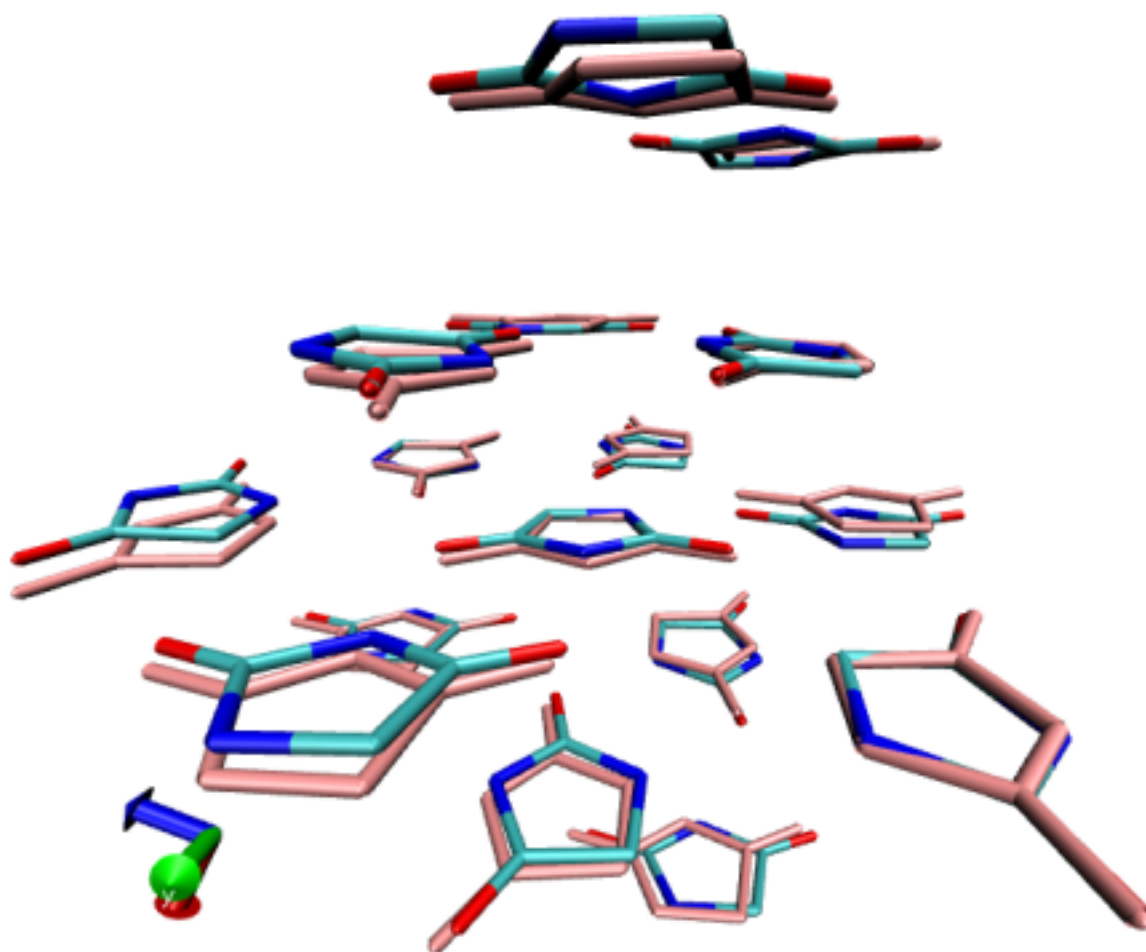


Figura 6.3: Superposición entre la estructura experimental y la predicha (en color rosa) para la molécula de imidazolidina-2,4-diona. El RMS de la superposición es de 0.33 Å

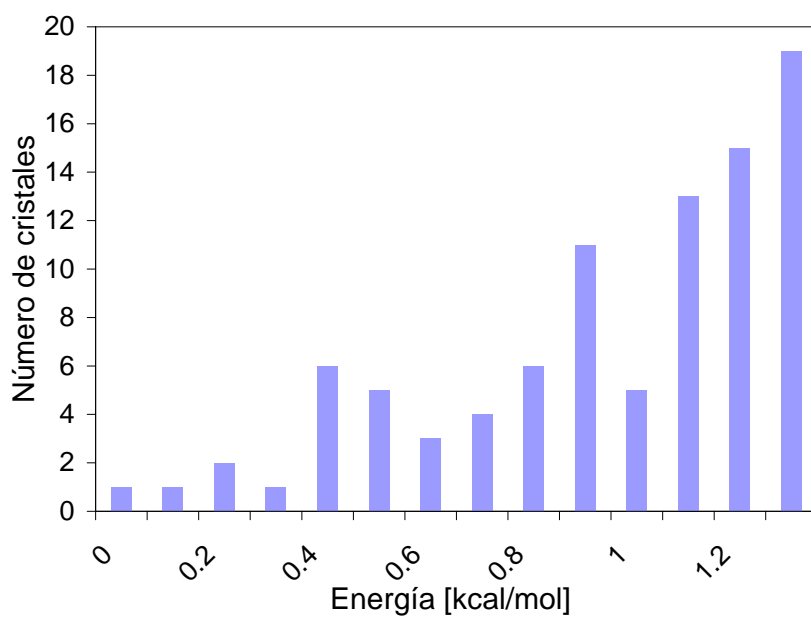


Figura 6.4: Distribución de los primeros 100 cristales generados por MGAC en función de la energía medida respecto a la energía del mínimo global para la molécula de imidazolidina-2,4-diona.

cristalinas. Antes de alcanzar al mínimo global, todo programa de optimización necesita discriminar entre un conjunto de cientos de cristales que presentan pequeñas diferencias de energía respecto al mínimo global. Para peor, todas estas estructuras son potenciales candidatos a ser el cristal o algún polimorfo hallado experimentalmente, debido a imprecisiones del potencial o al hecho de que el proceso de cristalización no es necesariamente controlado termodinámicamente.

6.3. Cristales de N-(2-metil-4,5-dinitrofenil)acetamida

La predicción del cristal de N-(2-metil-4,5-dinitrofenil)acetamida fue otro de los desafíos (targets) propuestos como parte del CSP2004 [8], sirviendo esta molécula como modelo de molécula flexible. La estructura cristalina de esta molécula fue la más difícil de predecir por las distintas metodologías dedicadas a la predicción de cristales. Ninguno de los 18 participantes en CSP2004 [8] pudo encontrar su estructura experimental entre los primeros tres candidatos, que fue el criterio utilizado en CSP2004 [8] para considerar exitosa una predicción. Además, solamente cuatro de los participantes hallaron estructuras similares a la experimental dentro de sus listas extendidas (estas listas tienen un tamaño de más de 135 cristales).

Esta molécula presenta seis ángulos diedros que involucran enlaces simples (ver figura 4.1). El primer intento de predecir con el MGAC los posibles cristales para esta molécula durante el CSP2004 [8] fracasó y no se encontró ninguna estructura similar a la experimental dentro de la lista de los 100 cristales de menor energía. Una vez conocida la estructura

experimental se verificó que, por error, no se permitió que varíe uno de los ángulos diedros ω durante la optimización. Esto trajo como resultado que se produjeran cristales candidatos con conformaciones moleculares muy diferentes a la experimental. Predicciones posteriores en las que se permitió la variación de este ángulo produjeron cristales en los que sus moléculas presentan conformaciones similares a las presentes en la estructura experimental, pero la aparición de una estructura cristalina semejante a la experimental resultó elusiva.

Es por eso que se decidió aumentar la capacidad de muestreo del MGAC de forma de incrementar la probabilidad de predecir la estructura experimental para este cristal. Como consecuencia fue necesario desarrollar programas para comparar y procesar las estructuras resultantes en la lista de estructuras más significativas, como se describió en el capítulo 5.

Para esta molécula se volvió a intentar la predicción de igual forma que en la prueba a ciegas, esto es sin utilizar ninguna información experimental, pero ahora permitiendo la optimización de todos sus ángulos diedros asociados a enlaces simples. Para esto se realizaron 20 simulaciones para los grupos espaciales más comunes ($P1$, $P-1$, $P2_1$, $C2$, Pc , Cc , $P21/c$, $C2/c$, $P2_12_12_1$, $Pca2_1$, $Pna2_1$, $Pbcn$, $Pbca$, $Pnma$) con una y dos moléculas por celda asimétrica. Cada simulación constó de 130 generaciones con un número de 30 cristales por cada generación. Esto da un total de 560 simulaciones del MGAC en las cuales se evaluaron 1108800 cristales para esta molécula en un período de tres semanas. Todas estas corridas utilizaron un total de aproximadamente 90.000 horas de CPU que fueron provistas por CHPC Arches Metacluster [127] y NCSA Teragrid cluster [128]. Luego se extrajeron las primeras 200 estructuras diferentes en orden de energía creciente de todas las estructuras

orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
1	4.842	14.142	15.118	90.00	102.53	90.00	1011	P21/c		0.0000
2	14.865	8.097	16.749	90.00	90.00	90.00	2016	Pbca		0.0929
3	12.554	4.799	19.406	90.00	118.58	90.00	1027	P21	P21/c	0.2726
4	4.849	9.659	11.953	109.99	96.84	96.94	514	P-1		0.4338
5	8.262	12.971	9.590	90.00	95.65	90.00	1023	P21/c		0.4841
6	10.247	11.064	9.470	90.00	105.71	90.00	1034	P21/c		0.5678
7	5.030	13.115	15.861	90.00	107.79	90.00	996	P21/c		0.5738
8	4.915	14.318	7.319	90.00	102.79	90.00	502	P21		0.5743
9	8.105	16.598	8.179	90.00	112.34	90.00	1018	P21/c		0.6767
10	3.955	16.617	7.809	90.00	98.58	90.00	508	P21		0.8289
11	9.037	15.783	7.481	90.00	112.19	90.00	988	P21/c		0.8587
12	8.422	9.536	15.112	90.00	121.81	90.00	1031	Pc	P21/c	0.8911
13	11.821	4.825	19.397	90.00	111.96	90.00	1026	P21	P21/c	0.8988
14	7.324	14.988	9.524	90.00	99.36	90.00	1032	Pc	P21/c	0.9806
15	8.310	8.748	14.823	94.40	95.98	100.99	1047	P-1		0.9967
16	8.420	8.814	14.460	86.82	82.59	79.64	1046	P-1		1.0105
17	9.408	14.723	14.990	90.00	90.00	90.00	2076	P212121		1.0661
18	4.773	14.640	15.048	90.00	92.83	90.00	1050	P21/c		1.0955
19	9.201	7.995	14.330	90.00	100.15	90.00	1038	P21		1.0977
20	4.828	14.921	14.448	90.00	95.12	90.00	1037	P21		1.1162

Tabla 6.5: Primeras 20 estructuras cristalinas en orden de energía creciente para la molécula de N-(2-metil-4,5-dinitrofenil)acetamida. La estructura de la tercera fila coincide con la estructura cristalina conocida experimentalmente para esta molécula.

Molécula	a	b	c	α	β	γ	volumen	densidad	mRMS	RMS
	Å	Å	Å	°	°	°	Å ³	g/cm ³	Å	Å
004 (exp.)	12.569	4.853	19.672	90.0	119.95	90.0	1040	1.53		
004 (pred.)	12.554	4.799	19.406	90.0	118.58	90.00	1027	1.55	0.076	0.16
	(-0.12 %)	(-1.1 %)	(+1.4 %)		(-1.1 %)		(-1.3 %)	(1.3 %)		

Tabla 6.6: Calidad de la estructura cristalina predicha para la molécula de N-(2-metil-4,5-dinitrofenil)acetamida.

generadas por el MGAC. De esta lista, unas 11 colapsadas fueron eliminadas, quedando 189 estructuras que se procesaron con el algoritmo ADDSYM del programa PLATON en búsqueda de simetrías escondidas.

La energía se calculó usando GAFF⁴. Las cargas atómicas se calcularon usando el método RESP, calculando el potencial electrostático de la molécula a un nivel HF/6-31G*.

En la tabla 6.5 se muestran las primeras 20 estructuras en orden de energía creciente. La información completa para las 189 estructuras está dada en la tabla B.2 del apéndice B. Una estructura similar a la experimental fue hallada en tercer lugar en orden de energía. En la figura 6.5 se muestra la superposición entre la estructura predicha y la experimental. Esta superposición se realizó sobre fragmentos de 16 moléculas extraídas de los cristales como se describió en el capítulo 5. La comparación de esta superposición resulta en un RMS de 0.16 Å. La comparación detallada para este cristal puede verse en la tabla 6.6.

Se observa en la tabla 6.5 que la estructura experimental se detecta a través de una simulación restringida al grupo espacial $P2_1$ con $Z' = 2$ y no al $P2_1/c$ y $Z' = 1$, como se conoce experimentalmente. Sólo cuando se aplica el algoritmo ADDSYM, se encuentra que la estructura predicha por el MGAC es más simétrica y en realidad ésta corresponde al

⁴En este caso se utiliza la version 1.25

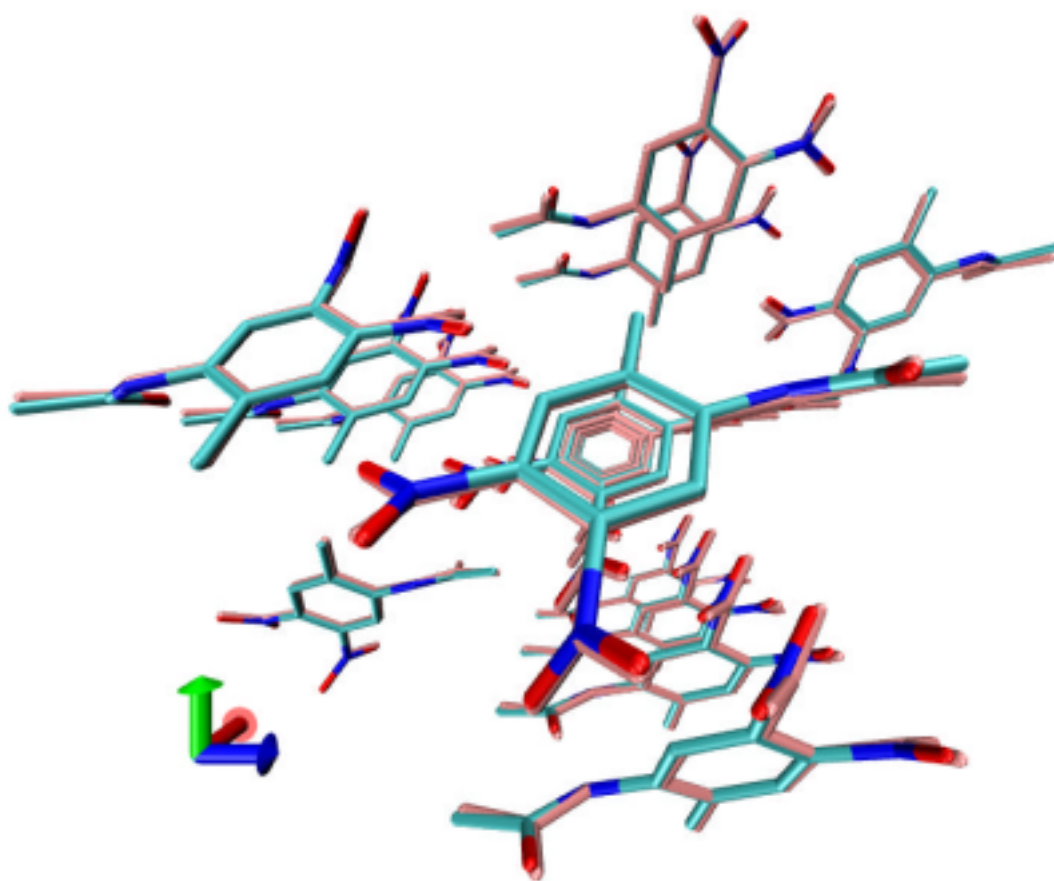


Figura 6.5: Superposición entre la estructura experimental y la predicha (en color rosa) para la molécula de N-(2-dimetil-4,5-dinitrofenil)acetamida. El RMS de la superposición es de 0.16 Å

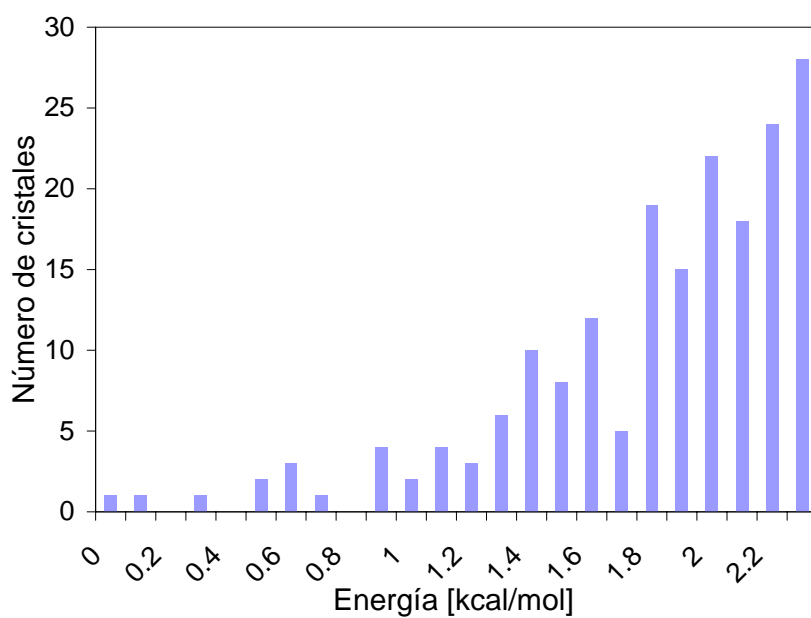


Figura 6.6: Distribución de los primeros 200 cristales generados por MGAC en función de la energía medida respecto a la energía del mínimo global para la molécula de N-(2-dimetil-4,5-dinitrofenil)acetamida

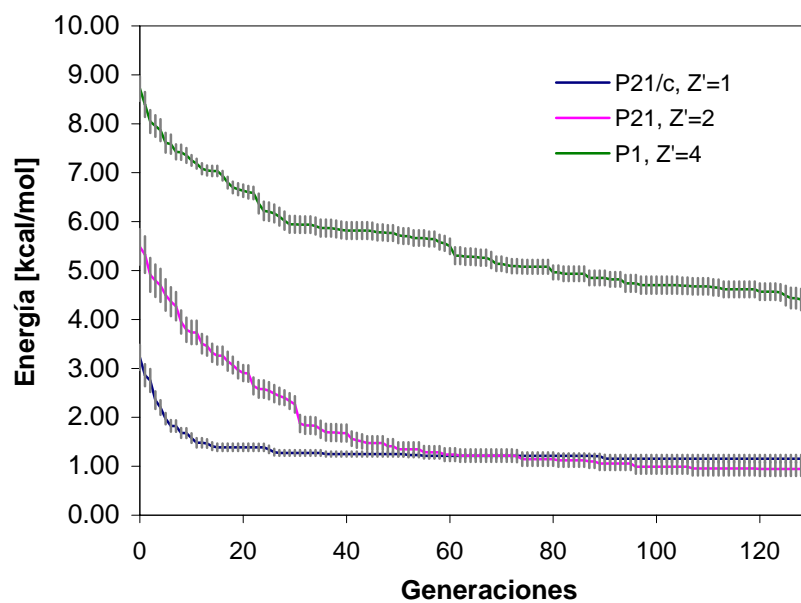


Figura 6.7: Evolución en promedio de las mejores estructuras correspondientes a los grupos espaciales $P2_1/c$, $P2_1$ y $P1$.

grupo $P2_1/c$ y $Z' = 1$. La ausencia sistemática de estructuras similares a la experimental en búsquedas restringidas en $P2_1/c$ sugiere que es más difícil para el método hallar a este cristal en su propio grupo espacial que en búsquedas menos restrictivas como $P2_1$. Para entender este comportamiento, tres series de 20 simulaciones restringidas a $P2_1/c$ con $Z' = 1$, $P2_1$ con $Z' = 2$ y $P1$ con $Z' = 4$ fueron ejecutadas por el MGAC. Se debe notar que en cualquiera de estas tres opciones es posible generar una estructura similar a la experimental. Para cada una de estas series se promediaron las menores energías alcanzadas por cada una de las 20 corridas en cada generación. Estos promedios y sus desviaciones estándares se muestran en la figura 6.7. Es interesante ver que las búsquedas en $P2_1$, aunque comiencen con energías mayores que las correspondientes a $P2_1/c$, luego de 70 generaciones encuentran en promedio los cristales con las menores energías. Este comportamiento es inusual debido a que se espera que sea más difícil la búsqueda en el espacio de mayor dimensión asociado al caso $P2_1$ que para el caso $P2_1/c$, que presenta un espacio de búsqueda de menores dimensiones. Claramente, ésto se cumple para el caso de $P1$, en donde la dificultad asociada a esta búsqueda puede verse por la lenta convergencia del MGAC. Esto implica que el MGAC es más eficiente, por lo menos para esta molécula, en búsquedas sobre cristales con dos y no con una molécula por celda asimétrica. Una posible explicación para este fenómeno es que en búsquedas más restrictivas como la correspondiente a $P2_1/c$, la zona de atracción asociada a la estructura experimental ocupa un hipervolumen menor en el espacio de búsqueda que para el caso $P2_1$, haciendo que este último caso sea mucho más fácil de resolver.

El intervalo de la energía para los primeros 189 cristales es aproximadamente de 2.3

kcal/mol, por lo que existen al menos 82 cristales por kcal/mol. De igual forma que para el caso de los cristales de imidazolidina-2,4-diona, esta distribución de cristales no es uniforme (ver figura 6.6). Por lo tanto se puede conjeturar que la existencia del orden de cientos de cristales por kcal/mol respecto al mínimo global es una propiedad persistente en estos problemas. Además sugiere que todo método de búsqueda que intente disminuir el número de estructuras muestradas, concentrándose en aquellas de menor energía, puede llegar a ser atrapado en las mismas en su evolución hacia el mínimo global. Esto refuerza la necesidad de realizar varias corridas independientes que generen distintas instancias del MGAC. De esta forma, si una de las simulaciones se estanca, aún existe la posibilidad de que alguna otra prosiga.

6.4. Límites del método: el ejemplo de la aspirina

La molécula del ácido 2-(acetiloxi)benzoico, comúnmente conocida como aspirina, es una candidata ideal para el estudio de metodologías relacionadas con la predicción de estructuras cristalinas, ya que la misma ha sido cristalizada en una gran variedad de condiciones debido a su importancia tanto terapéutica como comercial. Hasta el momento sólo se ha encontrado un único cristal para esta molécula, por lo que parece difícil la existencia de algún polimorfo estable. Además, predicciones de las posibles estructuras cristalinas para esta molécula se han realizado anteriormente en las referencias [3, 129].

La aspirina presenta cinco ángulos diedros sobre enlaces simples como se ilustra en la figura 4.1. Éstos fueron considerados variables para ser modificadas durante la optimización.

Se realizaron 20 simulaciones para los grupos espaciales más comunes ($P1$, $P-1$, $P2_1$, $C2$, Pc , Cc , $P21/c$, $C2/c$, $P2_12_12_1$, $Pca2_1$, $Pna2_1$, $Pbcn$, $Pbca$, $Pnma$) con una y dos moléculas por celda asimétrica. Cada simulación constó de 130 generaciones con un número de 30 cristales por cada generación. Esto da un total de 560 simulaciones del MGAC en las cuales se evaluaron 1108800 cristales para esta molécula en un período de tres semanas. Se utilizaron un total de aproximadamente 90.000 horas de CPU que fueron provistos por CHPC Arches Metacluster [127] y NCSA Teragrid cluster [128]. Luego se extrajeron las primeras 400 estructuras diferentes en orden de energía creciente de todas las generadas por el MGAC. De esta lista, 7 estructuras colapsadas fueron eliminadas, quedando 393 que se procesaron con el algoritmo ADDSYM del programa PLATON en búsqueda de simetrías escondidas. La información completa de las 393 estructuras está dada en la tabla B.3 del apéndice B.

La energía se calculó usando GAFF⁵. Las cargas atómicas se calcularon usando el método RESP, calculando el potencial electrostático de la molécula a nivel HF/6-31G*.

El intervalo de la energía para los primeros 393 cristales es de aproximadamente 3.3 kcal/mol, por lo que existen al menos 119 cristales por kcal/mol. De igual forma que para el caso de los cristales de imidazolidina-2,4-diona y N-(2-metil-4,5-dinitrofenil)acetamida, esta distribución de cristales no es uniforme (ver figura 6.8).

Ninguna de estas estructuras resultó ser similar a la experimental. Como se explicará a lo largo de esta sección, este fracaso en la predicción mostrará claramente algunas de las limitaciones del método propuesto.

⁵En este caso se utiliza la versión 1.25.

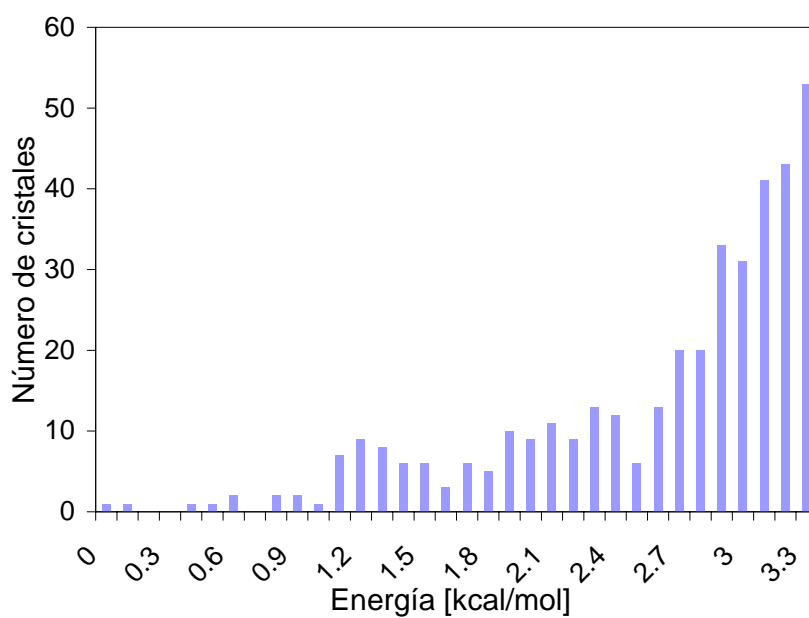


Figura 6.8: Distribución de los primeros 393 cristales generados por MGAC en función de la energía medida en referencia a la energía del mínimo global para el ácido 2-(acetiloxi)benzoico

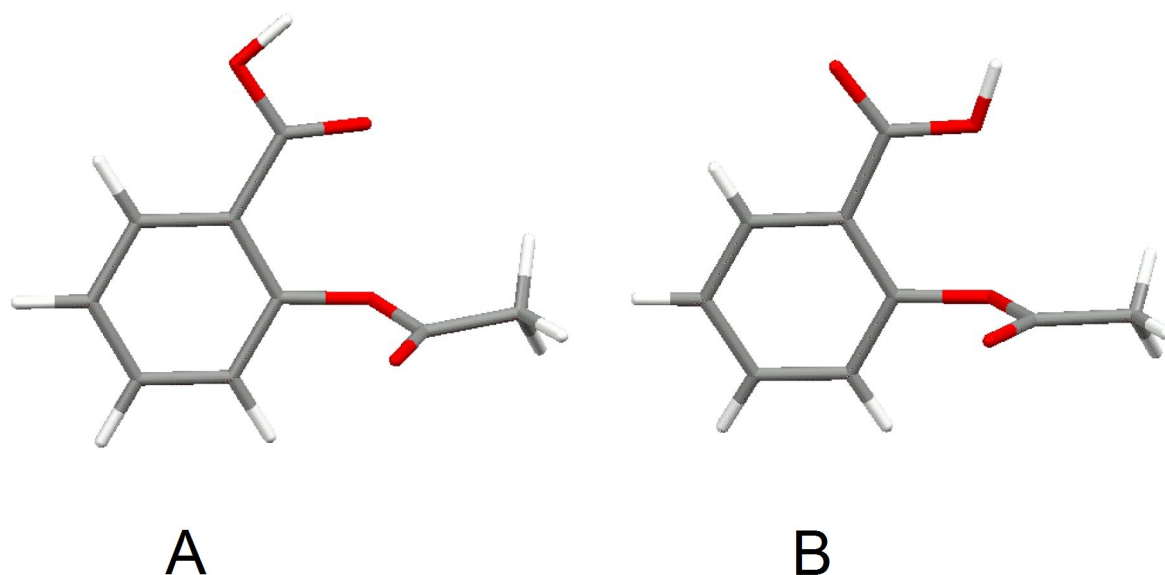


Figura 6.9: Primeras dos conformaciones en orden de energía creciente para la molécula de aspirina.

Del estudio de las conformaciones que la aspirina puede adquirir se conoce que las dos conformaciones de menor energía están definidas según la orientación del ácido carboxílico respecto al plano definido por el anillo aromático. Las dos posibilidades son: $\tau = 0$ que define la conformación *A*, o $\tau = 180$ que define la conformación *B* (ver la figura 6.9). Aunque la conformación *A* es la más estable en fase gaseosa, la conformación *B* es la que se observa en el cristal experimental.

Estudiando todos los cristales generados por el MGAC para esta molécula se tiene que sólo 5 de las 393 presentan la conformación *B*, mientras que el resto se encuentra en la conformación *A*⁶. Esto sugiere dos posibles escenarios para explicar el fracaso en la

⁶En el caso de los cristales con dos moléculas por celda asimétrica se consideró que el cristal presentaba la conformación *B* cuando ambas moléculas se encontraban en esa conformación. Esto se debe a que ambas

predicción: que exista alguna tendencia en los operadores genéticos a no crear moléculas con la conformación B , privilegiando la creación de cristales con la conformación A y por lo tanto, imposibilitando la reproducción de un cristal similar al experimental; o que exista una tendencia a favor de la conformación A por ser de menor energía. El primero de estos casos sería el más grave, ya que por diseño no debería haber ninguna tendencia que no sea establecida por las energías de las estructuras. El segundo caso implicaría que el GAFF falla en reproducir algún aspecto de las interacciones entre las moléculas.

Para discernir entre estos dos escenarios, se realizaron dos series de 20 corridas restringidas a $P2_1/c$ con $Z' = 1$ y $P2_1$ con $Z' = 2$, fijando la posición del ácido carboxílico en $\tau = 180$. Se permitió al resto de los ángulos diedros variar libremente durante la optimización. Se debe notar que en cualquiera de estos dos grupos espaciales es posible generar una estructura similar a la experimental. Luego, se extrajeron las primeras 200 estructuras diferentes en orden de energía creciente, de donde se eliminó a una estructura colapsada, quedando 199 estructuras que se procesaron con el algoritmo ADDSYM del programa PLATON en búsqueda de simetrías escondidas. En la tabla 6.7 se muestran las primeras 20 de estas estructuras agrupadas por energía creciente. La información completa para todas las 199 estructuras está dada en la tabla B.4 del apéndice B.

En la tabla 6.7 puede verse que todas las estructuras generadas con $\tau = 180$ presentan una energía de al menos 1.2 kcal/mol respecto a al mínimo global hallado en la primera corrida. Además, la estructura con número de orden 2 en energía de la tabla 6.7 *es similar al cristal experimental*. En la figura 6.10 se muestra la superposición entre esta estructura moléculas deben presentar la conformación B para reproducir el cristal experimental.

orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
1	12.371	6.301	11.280	90.00	112.58	90.00	812	P21	P21/c	1.2007
2	11.419	6.319	11.313	90.00	94.69	90.00	814	P21	P21/c	1.3854
3	3.924	8.143	25.201	90.00	94.68	90.00	803	P21	P21/c	1.6780
4	7.869	4.180	25.893	90.00	105.96	90.00	819	P21	P21/c	1.9419
5	13.528	3.921	16.117	90.00	108.20	90.00	812	P21	P21/c	2.2447
6	8.473	8.847	13.068	90.00	124.84	90.00	804	P21	P21/c	2.2616
7	11.190	5.123	14.519	90.00	97.26	90.00	826	P21	P21/c	2.4092
8	11.479	6.289	12.459	90.00	112.05	90.00	834	P21		2.4354
9	7.835	4.116	24.195	90.00	94.07	90.00	778	P21		2.4611
10	8.549	9.703	11.783	90.00	124.35	90.00	807	P21	P21/c	2.6052
11	10.108	5.464	15.664	90.00	111.15	90.00	807	P21	P21/c	2.6918
12	6.931	12.827	9.125	90.00	95.81	90.00	807	P21	P21/c	2.8390
13	7.069	15.984	9.002	90.00	126.14	90.00	821	P21/c		2.8647
14	18.894	4.372	9.841	90.00	90.00	90.00	813	P21	Pca21	2.9014
15	12.979	5.278	14.607	90.00	122.16	90.00	847	P21	P21/c	2.9143
16	4.021	25.598	8.287	90.00	106.09	90.00	820	P21/c		2.9477
17	12.294	4.914	14.879	90.00	110.70	90.00	841	P21	P21/c	3.0735
18	5.292	24.804	7.250	90.00	120.53	90.00	820	P21/c		3.2497
19	8.822	10.258	12.306	90.00	130.78	90.00	843	P21	P21/c	3.3263

Tabla 6.7: Primeras 20 estructuras cristalinas en orden de energía creciente para el ácido 2-(acetiloxi)benzoico generadas por el MGAC restringidas a los grupos espaciales $P2_1$ y $P2_1/c$ y a la conformación B . La energía se calcula respecto al mínimo global encontrado en la primera corrida. La estructura de la segunda fila coincide con la estructura cristalina conocida experimentalmente para esta molécula.

Molécula	a	b	c	α	β	γ	volumen	densidad	mRMS	RMS
	Å	Å	Å	°	°	°	Å ³	g/cm ³	Å	Å
005 (exp.)	11.430	6.591	11.395	90.0	95.68	90.0	854	1.40		
005 (pred.)*	11.419	6.319	11.313	90.0	94.69	90.00	814	1.47	0.072	0.29
	(-0.10%)	(-4.1%)	(-0.72%)		(-1.5%)		(-4.7%)	(4.8%)		

Tabla 6.8: Calidad de la estructura cristalina predicha para el ácido 2-(acetiloxi)benzoico. (*) Esta estructura se obtuvo al restringir la posición del ácido a la conocida experimentalmente.

y la experimental. Esta superposición se realizó sobre fragmentos de 16 moléculas extraídas de los cristales como se describió en el capítulo 5. La comparación de esta superposición resulta en un RMS de 0.29 Å. La comparación detallada para este cristal puede verse en la tabla 6.8.

Utilizando las tablas B.3 y B.4 se puede construir el histograma para las estructuras de la primera corrida filtrando aquellas que no fueron generadas bajo la restricción $P2_1$ y $P2_1/c$ y el histograma generado para las corridas con $\tau = 180$. Debido a que prácticamente no se generaron conformaciones del tipo B en la primera corrida, se puede decir que cada histograma representa la distribución de cristales para $\tau = 0$ y $\tau = 180$ generados por el MGAC restringido a los grupos espaciales $P2_1$ y $P2_1/c$. En la figura 6.11 se muestran estos histogramas marcando además la posición en donde se encuentra la estructura experimental. Se observa que la distribución asociada a la conformación B se halla desplazada en energía en aproximadamente 1 kcal/mol por arriba de la distribución de cristales con conformación A . Se concluye que MGAC eliminó a las conformaciones del tipo B en la primera corrida debido exclusivamente a consideraciones energéticas. Esto ocurre porque existe una gran cantidad de estructuras con energías por debajo de la estructura experimental. Además es muy baja la probabilidad de que el MGAC encuentre la estructura experimental durante su

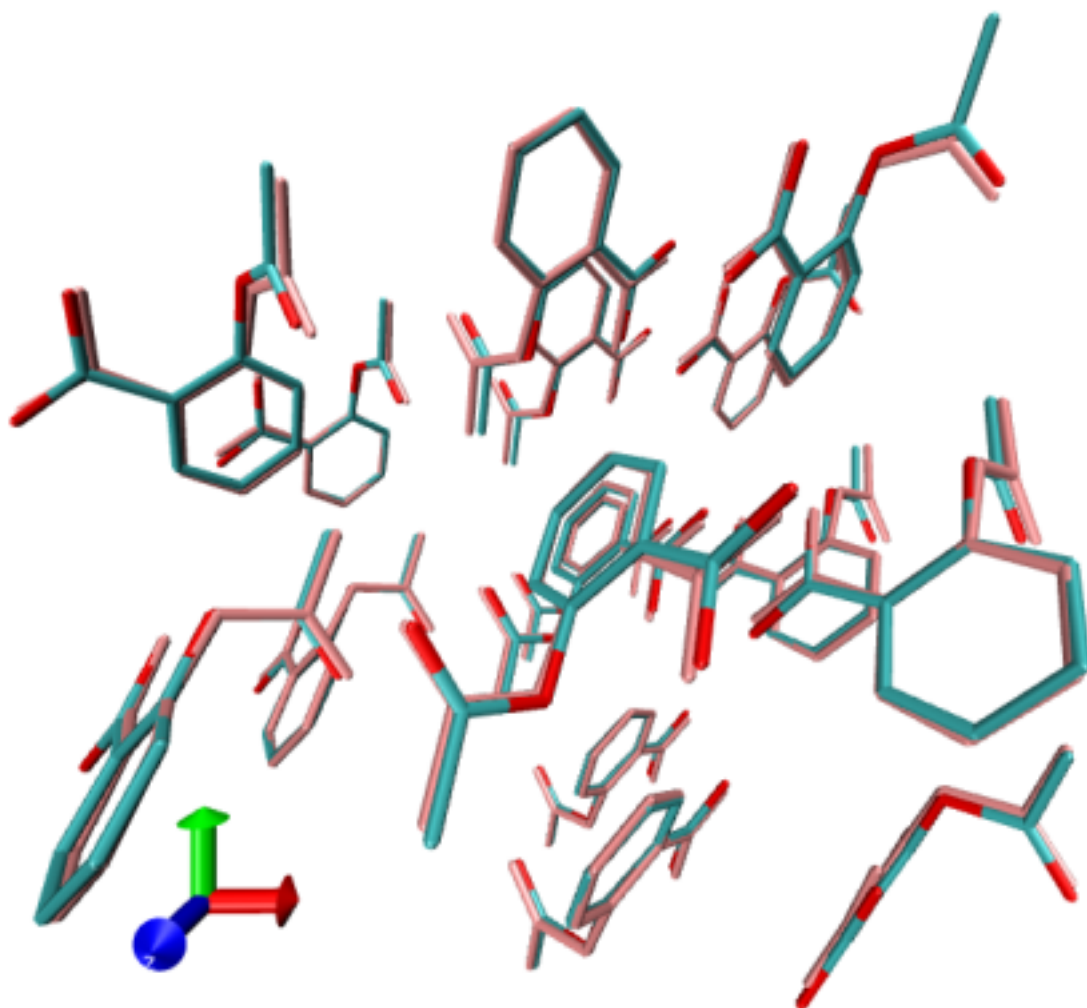


Figura 6.10: Superposición entre la estructura experimental y la predicha (en color rosa) para el ácido 2-(acetiloxi)benzoico. El RMS de la superposición es de 0.29 Å

evolución, ya que en el rango de energías en donde se encuentra la estructura experimental existen muchas más estructuras con $\tau = 0$ que con $\tau = 180$. Por lo tanto, es mucho más probable que se generen estructuras con $\tau = 0$, reduciendo enormemente la habilidad del MGAC de muestrear cristales con moléculas en la conformación necesaria para predecir la estructura experimental.

Como el fracaso de predecir el cristal de la aspirina tiene origen en el cálculo de la energía y no en cómo trabaja MGAC, se puede concluir o que el potencial MGAC falla en reproducir el orden correcto en energías para los distintos cristales, o que el proceso de cristalización de la aspirina no es termodinámico. Esta última posibilidad de que la estructura experimental sea un metaestable de la aspirina, estaría en directa contradicción con la estabilidad observada en la práctica para esta estructura.

Entre los estudios teóricos previos realizados para esta molécula, solamente el trabajo de Ouvrard y Price [3] consideró la existencia de ambas conformaciones *A* y *B*. En este trabajo se calcula la energía como la suma de dos contribuciones: una intramolecular, que está dada por el perfil de energía de las distintas conformaciones que la molécula puede adquirir, y otra intermolecular, que está dada por las interacciones entre las moléculas. La contribución intramolecular se calculó con métodos ab-initio y DFT, mientras que para la contribución intermolecular se utilizaron potenciales empíricos. Como resultado se reportó la existencia de cristales con ambas conformaciones, incluyendo un cristal similar al experimental como la segunda estructura en orden de energía creciente. Por lo tanto, este trabajo sugiere fuertemente que la existencia de 28 estructuras encontradas por el MGAC con menor energía

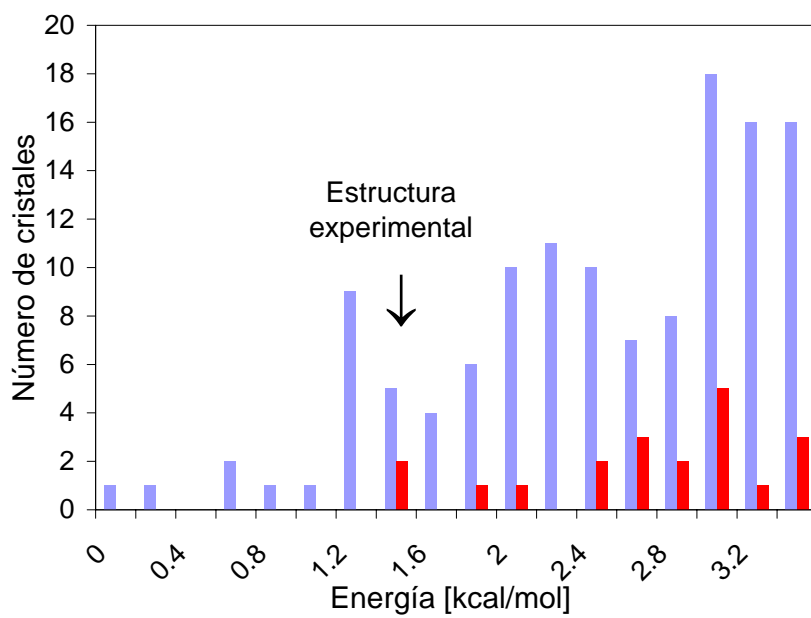


Figura 6.11: Distribuciones de los cristales generados por el MGAC para las restricciones $P2_1$ y $P2_1/c$ para los casos $\tau = 0$ en azul y $\tau = 180$ en rojo.

GAFF	HF/6-31G(d,p)	B3LYP/6-31G(d,p)	MP2/6-31G(d,p)
kcal/mol	kcal/mol	kcal/mol	kcal/mol
1.78	1.10	0.83	0.68

Tabla 6.9: Diferencias de energías entre las conformaciones de la aspirina *A* y *B*

que la experimental es un producto artificial creado por el GAFF.

En la tabla 6.9 se informa la diferencia de energía entre las conformaciones *A* y *B*, tanto para GAFF como para tres métodos no empíricos HF/6-31G(d,p), B3LYP/6-31G(d,p) y MP2/6-31G(d,p) [3]. Esta información permite discriminar cómo GAFF evalúa la contribución intramolecular a la energía de los cristales. Se puede observar que el GAFF sobrestima la diferencia entre ambas conformaciones por 0.92 kcal/mol respecto a lo dado por B3LYP/6-31G(d,p). Este último método fue utilizado por Ouvrard y Price para el cálculo de la energía. Por lo tanto el GAFF estaría sobrevaluando la energía de los cristales por aproximadamente 0.92 kcal/mol para la conformación *B* respecto a la misma estructura encontrada por Ouvrard y Price. Esta diferencia puede mayoritariamente explicar el desplazamiento aproximado de 1 kcal/mol entre las distribuciones para ambas conformaciones observadas en la figura 6.11. Esta idea se ve reforzada si se agrega el hecho de que la primera estructura en la tabla 6.7 es similar al mínimo global encontrado por Ouvrard y Price. En la tabla 6.10 se comparan los parámetros cristalográficos y sus diferencias de energía respecto al cristal experimental para esta estructura. Esto significa que fijando la conformación de la molécula en *B* se reproducen las dos principales estructuras encontradas por Ouvrard y

	a	b	c	α	β	γ	simetría	energía
	Å	Å	Å					kcal/mol
Referencia [3]	12.124	6.696	11.475	90.00	111.5	90.00	P21/c	-0.0478
MGAC	12.371	6.301	11.280	90.00	112.58	90.00	P21/c	-0.1847
Diferencia	(2.0 %)	(-5.9 %)	(-1.7 %)		(1.0 %)			

Tabla 6.10: Comparación entre el mínimo global de la referencia [3] y el primer cristal de la tabla 6.7.

Price⁷.

Por lo tanto se concluye que el fracaso en predecir el cristal de aspirina se debió a deficiencias del potencial empírico GAFF, mostrando que la precisión del método propuesto está principalmente limitada por la forma en que se calcula la energía.

Un procedimiento para detectar problemas del potencial empírico en el cálculo de las contribuciones intramoleculares, es comparar las diferencias de energía para las principales conformaciones calculadas con GAFF y algún método ab-initio o DFT para la molécula de interés. Un fuerte indicador de que el potencial introduciría errores inaceptables para las predicciones es si se observan grandes desviaciones entre el GAFF y los restantes métodos. Este proceso puede ser implementado sin conocimiento previo de cómo la molécula cristaliza.

⁷Ouvrard y Price sugieren que la ausencia experimental de esta estructura se debe a que su constante elástica mínima es de 0.15 GPa, por lo que es fácilmente deformable creando dificultades en su cristalización. Esto contrasta con la estructura experimental, cuya constante elástica mínima es de 4.72 GPa

Molécula	a Å	b Å	c Å	α °	β °	γ °	volumen Å^3	densidad g/cm^3	mRMS Å	RMS Å
001 (exp.)	6.032	12.343	5.784	90.0	90.0	90.0	431	1.38		
001 (pred.)	5.703 (-5.5%)	12.357 (+0.11%)	5.774 (-0.17%)	90.0	90.0	90.0	407 (-5.6%)	1.46 (+5.8%)	0.081	0.30
002 (exp.)	12.026	6.032	5.829	90.0	90.0	90.0	423	1.40		
002 (pred.)	12.785 (+6.3%)	5.664 (-6.1%)	5.684 (-2.5%)	90.0	90.0	90.0	412 (-2.6%)	1.44 (+2.8%)	0.17	0.40
003 (exp.)	9.354	12.176	7.228	90.0	104.59	90.0	797	1.67		
003 (pred.)	8.769 (-6.2%)	12.087 (-0.73%)	7.598 (+5.0%)	90.0	101.27 (-3.2%)	90.0	790 (-0.88%)	1.69 (1.1%)	0.045	0.33
004 (exp.)	12.569	4.853	19.672	90.0	119.95	90.0	1040	1.53		
004 (pred.)	12.554 (-0.12%)	4.799 (-1.1%)	19.406 (+1.4%)	90.0	118.58 (-1.1%)	90.00	1027 (-1.3%)	1.55 (1.3%)	0.076	0.16
005 (exp.)	11.430	6.591	11.395	90.0	95.68	90.0	854	1.40		
005 (pred.)*	11.419 (-0.10%)	6.319 (-4.1%)	11.313 (-0.72%)	90.0	94.69 (-1.5%)	90.00	814 (-4.7%)	1.47 (4.8%)	0.072	0.29

Tabla 6.11: Calidad de las estructuras predichas. (*) Estructura predicha para la aspirina cuando se restringió la posición del ácido a la conocida experimentalmente.

6.5. Calidad de las estructuras predichas

En la tabla 6.11 se comparan las estructuras experimentales con las predichas en esta tesis. Comparando a esta última tabla con las optimizaciones locales de las estructuras experimentales (tablas 4.2 y 6.1) se encuentra que las desviaciones son producto del potencial empírico y no del método de búsqueda. El promedio de los RMS para todas las predicciones es 0.30 ± 0.04 Å. En el caso de la dimensión de la celda unidad, el promedio del error relativo de sus longitudes es de $2.68 \pm 0.66\%$, mientras que para los ángulos es de $1.93 \pm 0.64\%$. En el caso del volumen de la celda unidad y las densidades de los cristales, el promedio del error relativo es de $3.0 \pm 0.93\%$ y $3.2 \pm 0.94\%$ respectivamente.

El cristal peor reproducido es el de DL-alanina, que fue discutido en la sección 6.1. La mejor predicción corresponde a la molécula de N-(2-metil-4,5-dinitrofenil)acetamida. La diferencia entre la estructura predicha y la experimental es de 0.16 Å, contrastando

fuertemente con las predicciones realizadas en CSP2004 [8], en donde la estructura predicha más cercana a la experimental presentó un RMS de 0.5 Å.

6.6. Eficiencia de la metodología MGAC

El análisis de los cristales de las moléculas de L-, DL-alanina e imidazolidina-2,4-diona se realizó manualmente para separar y distinguir entre estructuras similares. Este proceso solía requerir más de una semana por molécula y no pudo ser realizado para más de 50.000 cristales por molécula, limitando el número de cristales generados por el MGAC que podían ser analizados. Luego de la creación de MOLCRY esta limitación desapareció, permitiendo en la actualidad procesar del orden de un millón de estructuras por molécula. Para ilustrar este punto se puede tomar el caso de caso de N-(2-metil-4,5-dinitrofenil)acetamida. En este sistema se generaron más de un millón de estructuras, y se extrajeron las 200 estructuras de menor energía, geoméricamente diferentes, en aproximadamente 240 minutos utilizando MOLCRY en una PC AMD Athlon 64 X2 2.2 Ghz.

Una forma de cuantificar el progreso de la capacidad de muestreo a lo largo de esta tesis es graficando el número de cristales que pudieron ser evaluados por molécula para cada año. En la figura 6.12 puede observarse que el número de estructuras evaluadas por molécula ha crecido casi exponencialmente en los últimos años⁸. En principio, con la presente metodología es posible un mayor muestreo si existen razones que lo justifiquen.

⁸El año 2005 se dedicó completamente al desarrollo de las metodologías para la comparación y procesamiento de los cristales. Esta es la razón por la cual el mismo no se reporta en la figura 6.12.

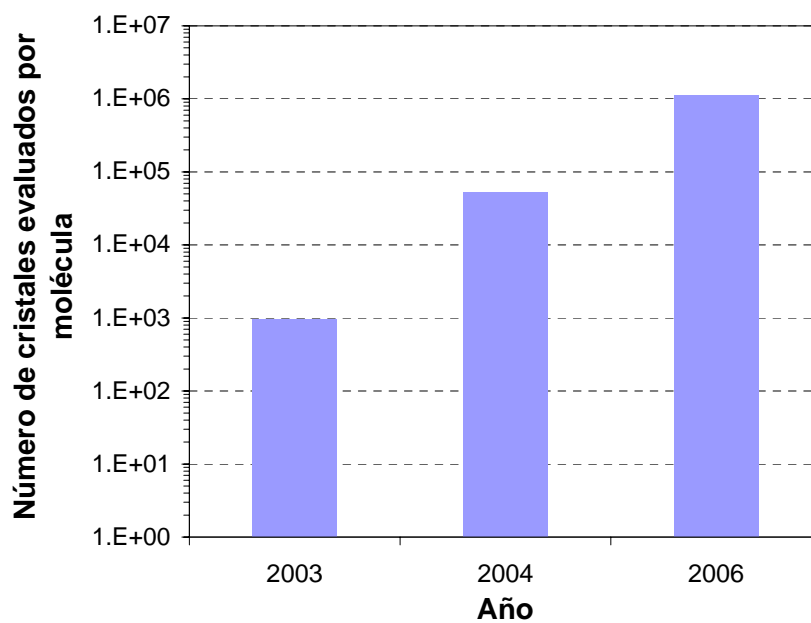


Figura 6.12: Número de cristales evaluados por molécula para cada año de esta tesis.

Capítulo 7

Extendiendo MGAC a clusters atómicos

7.1. Motivaciones

Como se describe en la sección 2.3, los algoritmos genéticos se han aplicado para el modelado de una gran variedad de clusters. También se ha reportado la utilización de diferentes aproximaciones para el cálculo de la energía [22, 29, 130, 131]. Estas aproximaciones y sus costos computacionales asociados son los que determinan el máximo tamaño (número de átomos) y precisión con que estas simulaciones se realizan. La mejor de estas aproximaciones se define por el cálculo de la energía bajo la teoría de la Funcional de la Densidad (Density Functional Theory o DFT). Debido a su alto costo computacional, el uso combinado de GA y DFT se restringe para el caso de clusters de, por ejemplo, hasta 10 átomos de

silicio [132, 133, 134]. Todo intento de analizar clusters de silicios de mayor tamaño usando DFT se ha basado en búsquedas por patrones, como se describe en la sección 1.3 y en la referencia [2].

Una alternativa que se utiliza en este capítulo es el uso de métodos semi-empíricos para la evaluación de la energía. A esta combinación de MGAC y métodos semi-empíricos se la llama MGAC/MSINDO. Debido a su bajo costo computacional es posible extender el uso de GA para sistemas con un gran número de átomos. Esta idea es similar a la desarrollada en las referencias [29, 130, 131] que utilizan métodos de *tight binding*. A lo largo de este capítulo se presentan dos resultados: en el primero se aplica el MGAC/MSINDO a clusters de silicios Si_n ($n = 4-9$) como proceso de validación sobre clusters pequeños y luego se aplica el método sobre el cluster de Si_{36} , uno de los sistemas más grandes estudiados previamente bajo la estrategia de búsqueda por patrones. El objetivo principal es realizar comparaciones sobre el Si_{36} para determinar si las estructuras predichas utilizando búsqueda por patrones son suficientes para describir estos sistemas, o si existen estructuras que no son encontradas por esa metodología. Para que estas comparaciones resulten independientes de la forma en que se calcula la energía, todas las estructuras producidas como resultado de la predicción MGAC/MSINDO fueron optimizadas localmente usando DFT. La existencia de estructuras no encontradas por la búsqueda por patrones justificaría el desarrollo de metodologías como algoritmos genéticos para clusters atómicos de tamaño mediano.

7.2. Algoritmo genético para clusters atómicos

Se implementó el algoritmo genético para clusters atómicos tomando como modelo el algoritmo desarrollado por Johnston y Roberts [22]. El genoma para este sistema se construye con un vector que contiene las coordenadas cartesianas de todos los átomos. Las mismas se ordenan como una sucesión de las coordenadas atómicas individuales, esto es $\{x_1, y_1, z_1, \dots, x_n, y_n, z_n\}$ en donde x_i, y_i, z_i son las coordenadas del i -ésimo átomo. La dimensión de este vector es $3n$, en donde n es el número de átomos en el cluster.

La primera generación se obtiene proponiendo el valor de las coordenadas atómicas en forma aleatoria. Estos valores se encuentran limitados dentro de un intervalo que define la extensión espacial del cluster. Es por eso que estos límites son estimados en función del tamaño esperado para el cluster. Para los clusters resultantes se calculan las distancias entre todo par de átomos, las cuales luego son comparadas con un conjunto de reglas que garantizan que las distancias entre átomos se encuentren entre valores compatibles con enlaces covalentes. En el caso de que esto último no se cumpla, la estructura se reemplaza por otra estructura aleatoria. Este procedimiento se repite hasta que todas las estructuras iniciales sean compatibles con estas reglas.

Las reglas que definen la permanencia de un cluster en la población están diseñadas con el objetivo de eliminar aquellas estructuras que son evidentemente no físicas. Estas reglas son básicamente dos: la primera es que si la distancia entre dos átomos es menor a una dada distancia r_1 , la estructura es rechazada; la segunda regla dice que si un átomo se encuentra a una distancia mayor que r_2 respecto a cualquier otro átomo dentro del

cluster, la estructura también se rechaza. En general $r_1 = 0,5 \text{ \AA}$ mientras que a r_2 se la elige en función de los tipos de átomos que componen el cluster (en el caso de clusters de silicios se fijó $r_2 = 3,2 \text{ \AA}$, ya que se encuentra por encima de la distancia entre átomos de silicio enlazados covalentemente).

Una vez obtenida la población inicial, el algoritmo prosigue en forma similar a la descrita en el capítulo 3. Para el caso de clusters pequeños se eligieron como operadores genéticos los mismos que se utilizan para los cristales. En cambio, para el caso de Si_{36} se utilizaron como operadores de recombinación a los *cut and splice* y *two cuts and splice* definidos en la referencia [22]. Este cambio se debe a la evidencia de que estos operadores exploran el espacio de configuraciones de clusters de forma más eficiente que otros propuestos previamente [22]. Utilizando la información presente en cada genoma, cada estructura candidata es optimizada localmente. Todos los cálculos de energía se realizaron con el programa MSINDO [30, 31, 32]. Para los sistemas que contenían hasta 9 átomos se utilizaron poblaciones de 20 individuos, mientras que para el caso del Si_{36} se utilizaron poblaciones de 40 individuos. Para pequeños clusters los algoritmos genéticos convergieron a poblaciones finales poco diversas en sólo 30 generaciones, mientras que para el Si_{36} esta convergencia se produjo luego de 200 generaciones. Se consideró que las poblaciones convergían cuando la desviación estándar de la energía de las estructuras presentes en una misma población alcanzaba 0.1 eV. Luego, las estructuras fueron clasificadas manualmente, seleccionando aquellas de menores energías y geoméricamente diferentes. Las geometrías de estos clusters resultantes se optimizaron localmente a nivel B3PW91 utilizando los pseudo potenciales de Stuttgart y

una base de funciones extendidas con orbitales difusos s , p [135] y orbitales de polarización d definidos por Sadlej [136]. Para el caso de Si_{36} se utilizó una base de menor tamaño LanL2DZ, en conjunto con el pseudo potencial de Los Alamos [137], con la intención de reducir el costo computacional debido al tamaño de estos clusters. Además, todas las estructuras se optimizaron utilizando la aproximación de densidad de spin local (local spin density approximation o LSDA) [138] para estimar la sensibilidad de los resultados con la funcional de intercambio-correlación que se utilizó. Para confirmar cuáles isómeros son verdaderos mínimos y no puntos de transición o puntos críticos se calcularon los espectros de frecuencias vibracionales para cada una de las estructuras minimizadas. Todos los cálculos de DFT se realizaron con el programa Gaussian [93].

7.3. Resultados

La figura 7.1 muestra el conjunto de estructuras de menor energía para los clusters de Si_n ($n = 4-9$) obtenidos utilizando MGAC. La energía de confinamiento por átomo se calculó utilizando la fórmula

$$BE = \frac{E(Si_n) - nE(Si)}{n} \quad (7.1)$$

en donde $E(Si_n)$ es la energía total para un cluster de silicio de tamaño n y $E(Si)$ la energía total del átomo de silicio. No todos los isómeros encontrados se muestran en la figura 7.1, sino aquellos significativos para la comparación con trabajos de otros autores. Comparando estas estructuras con aquellas previamente publicadas (ver figura 7.2) puede

observarse que el MGAC es capaz de encontrar casi todas las estructuras conocidas para estos clusters. La única excepción es para el tercer isómero de Si_6 , que no se encontró en estas simulaciones. En contraste puede verse que el MGAC encuentra otros dos isómeros estables ubicados en orden de energía creciente entre los dos isómeros conocidos para el Si_8 . Esto sugiere la posibilidad de la existencia de isómeros no hallados previamente en sistemas tan pequeños como Si_8 . Es claro que esta aproximación reproduce correctamente las estructuras para estos clusters, validando su uso en sistemas de Si_n .

Las estructuras para el cluster de Si_{36} han sido estudiadas recientemente [2] bajo la aproximación de búsqueda por patrones. Estas estructuras se obtuvieron de la optimización local de conformaciones derivadas a través de la introducción de pequeños cambios a un conjunto de patrones estructurales. Esta restricción abre la posibilidad de que existan estructuras más estables, esto es de menor energía total o mayor energía de confinamiento, que no pueden ser derivadas de los patrones propuestos. Utilizando MGAC/MSINDO, que demostró ser útil para modelar pequeños clusters, se encontraron nuevos isómeros estables del Si_{36} . Manualmente se extrajeron los tres isómeros de menor energía, y el más estable fue optimizado localmente a nivel B3PW91 y con una base LanL2DZ de pseudopotenciales de Los Alamos. La estructura resultante difiere significativamente de los isómeros hallados en la referencia [2], que fueron derivados de tres tipos de patrones: estructuras con forma de jaulas, de cable y similares a fullerenos rellenos de átomos. A estos tres isómeros se los nombraron como *Cage 2*, *Wire 2* y *Stuff30-A*. En la figura 7.3 se muestra a los primeros tres isómeros de la referencia [2] y se incluye la estructura hallada por el MGAC.

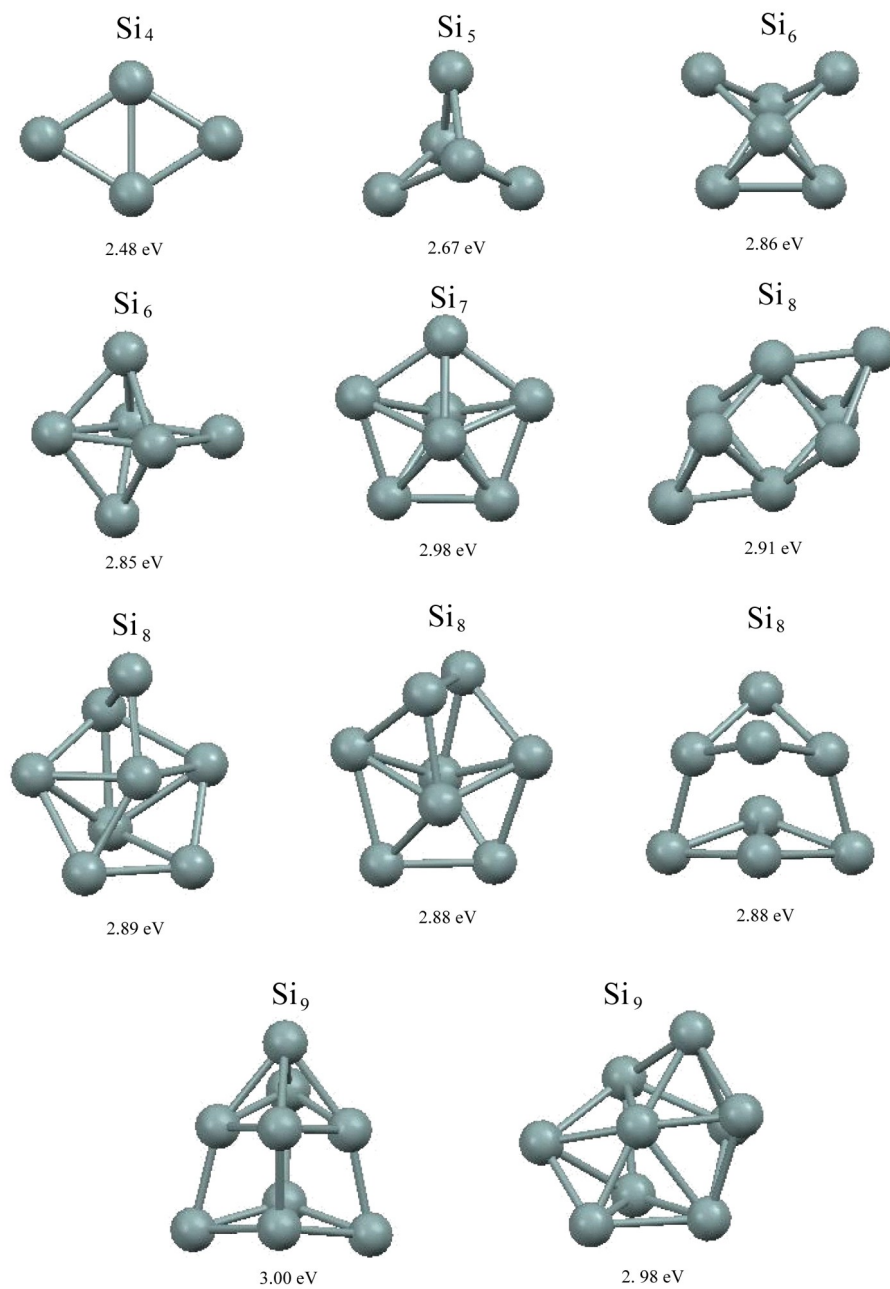


Figura 7.1: Estructuras optimizadas y energía de confinamiento por átomo para Si_n ($n = 4-9$).

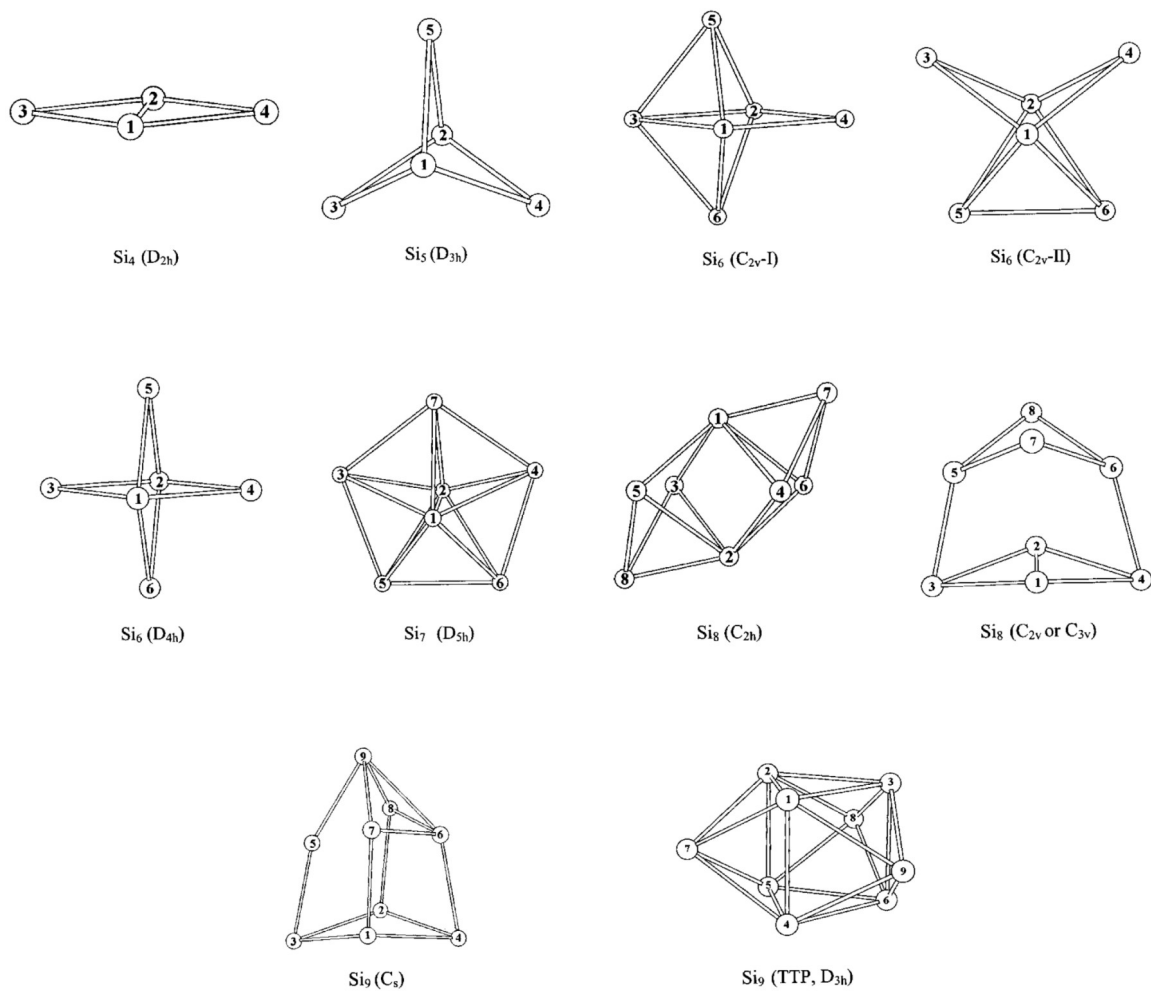
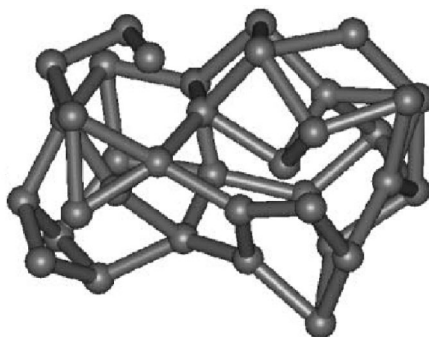
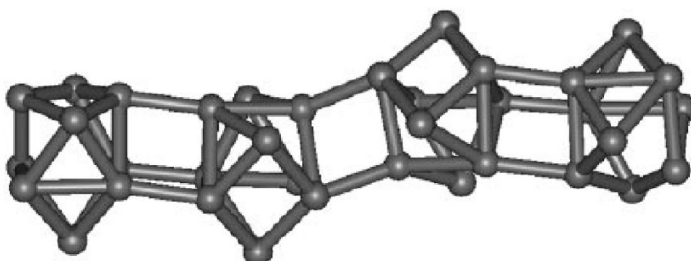


Figura 7.2: Estructuras de cluster Si_n ($n=4-9$) extraídas de la referencia [1]

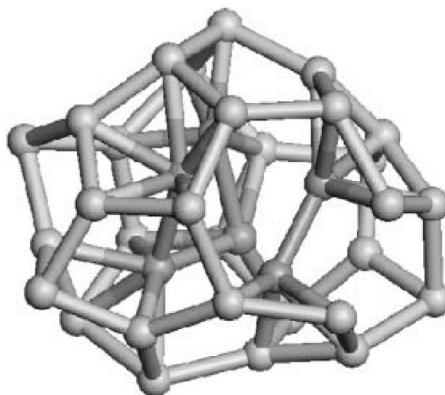
Cage 2*



Wire 2*



Stuff30-A*



MGAC
MSINDO

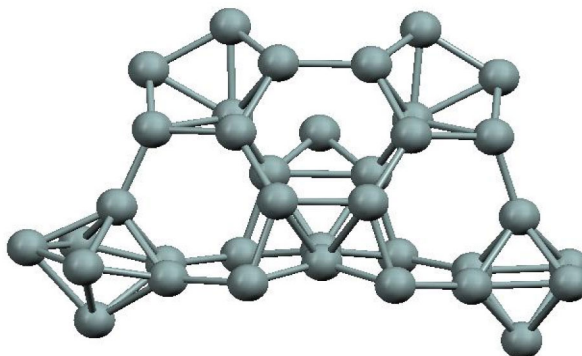


Figura 7.3: Comparación entre la estructura hallada por el MGAC y (*) las estructuras de la referencia [2].

Estructuras	LSDA	B3PW91
	eV	eV
Cage 2	4.79	3.93
Wire 2	4.95	4.18
Stuff30-A	5.00	4.16
MGAC/MSINDO	5.13	4.41
MGAC/MSINDO (opt.)	5.19	4.46

Tabla 7.1: Energía de confinamiento por átomo para los isómeros de Si_{36} . Todos los valores se calcularon tomando la energía del átomo de silicio como -102.155 eV y -101.428 eV para las aproximaciones LSDA y B3PW91 respectivamente.

En la tabla 7.1 se comparan las energías de confinamiento para todos los isómeros. Con el nombre MGAC/MSINDO se designa la estructura resultante de la optimización global utilizando MGAC, mientras que se designa por MGAC/MSINDO (opt.) a esta última luego de ser optimizada localmente utilizando DFT. Los cálculos de energía se realizaron para las funcionales de intercambio-correlación LSDA y B3PW91. Es evidente que la estructura MGAC/MSINDO optimizada es significativamente más estable que aquellas obtenidas en [2], con una diferencia de al menos 0.3 eV entre energías de confinamiento. También se obtiene que la estabilidad de estas estructuras no depende de la funcional que se utilice. La optimización local de los otros dos clusters en orden de energía creciente generados por el MGAC dan como resultado estructuras muy similares pero diferentes a la de la figura 7.3. Sus energías de confinamiento son 4.46 eV y 4.45 eV respectivamente. Por lo tanto,

esto muestra que hay al menos tres isómeros con menor energía que los propuestos por la búsqueda de patrones en la referencia [2].

Debido a que MGAC utiliza el método semi-empírico MSINDO para el cálculo de la energía y siendo la energía la que guía la búsqueda en el MGAC, no se puede garantizar que las estructuras propuestas sean mínimos globales para el Si_{36} . Además, dado que la dimensión del problema es muy grande, es posible que se requiera un mayor número de simulaciones para coleccionar un conjunto de estructuras significativas. En todo caso, el hallazgo de estructuras más estables que en el caso de la búsqueda por patrones demuestra la importancia de explorar el espacio de configuración en forma más exhaustiva.

Conclusiones

En esta tesis se propone un método para la predicción de estructuras cristalinas de moléculas orgánicas. El mismo se implementa a través de un conjunto de programas altamente integrados encargados de la generación de potenciales intermoleculares, de la búsqueda de los posibles cristales asociados a una molécula utilizando algoritmos genéticos distribuidos (en paralelo) y del análisis, procesamiento y almacenamiento de las estructuras más importantes generadas en la búsqueda. Este *paquete* computacional puede ser utilizado para la predicción de estructuras cristalinas de una gran variedad de moléculas debido a la utilización del potencial empírico GAFF, diseñado para modelar moléculas de interés farmacológico compuestas por C, H, N, O, S, P y halógenos.

El método fue aplicado en la predicción de los cristales de L-alanina, DL-alanina, imidazolidina-2,4-diona, N-(2-metil-4,5-dinitrofenil)acetamida y ácido 2-(acetiloxi)benzoico (aspirina). En todos los casos, con excepción del último, se predijeron cristales similares a los hallados experimentalmente. Estas estructuras se encontraron entre las más significativas (entre las primeras tres estructuras en orden de energía creciente) generadas por las distintas búsquedas. En el caso de la aspirina, la predicción fue posible sólo cuando

se incluyó información experimental acerca de la posición del ácido carboxílico respecto al plano definido por el anillo aromático. Esto se debió a que el potencial GAFF sobrevalúa, por aproximadamente 1 kcal/mol, la diferencia de energía entre las dos conformaciones principales de esta molécula. Esto genera un ordenamiento de las energías de los distintos cristales que posiciona a la estructura experimental en forma desfavorable respecto a los otros cristales candidatos.

Es posible detectar, aún sin conocimiento previo de cómo una molécula cristaliza, aquellos casos en los que el potencial falla al calcular la contribución intramolecular a la energía. Para esto es necesaria la comparación entre las diferencias de energía para las principales conformaciones que la molécula pueda presentar utilizando diferentes métodos, tal como se realizó para el caso de la aspirina. Grandes desviaciones entre los valores calculados por GAFF y por otros métodos serían indicadores de que GAFF no es el potencial más apropiado para el estudio de dicha molécula.

Independientemente de las dificultades acarreadas por el potencial al estudiar la aspirina, el método demostró ser eficiente en todos los casos para la búsqueda de cristales compuestos por moléculas flexibles. En particular, para N-(2-metil-4,5-dinitrofenil)acetamida y ácido 2-(acetiloxi)benzoico, que presentan gran flexibilidad, las estructuras experimentales fueron halladas como versiones más simétricas en búsquedas definidas por dos moléculas por celda asimétrica ($Z' = 2$). Esto sugiere que el método converge con una mayor eficiencia en el caso $Z' = 2$ que para $Z' = 1$. De todos los métodos para la predicción de estructuras cristalinas presentados en el capítulo 2, el MGAC es el único que posee esta propiedad.

Por ende, esta tesis muestra argumentos a favor de que el MGAC resuelve dos de los principales problemas encontrados en la búsqueda de cristales candidatos, que son la ausencia de métodos eficientes para la predicción de cristales compuestos por moléculas flexibles y/o que contienen dos moléculas en su celda asimétrica.

Otra conclusión importante que se deriva de este trabajo es la dificultad intrínseca asociada al problema de predicción de estructuras cristalinas. La misma se manifiesta en el gran número de mínimos locales por cada kcal/mol (aproximadamente 100) hallados por el MGAC con energías similares a la del mínimo global. Esta estimación sobre el número de estructuras, así como su distribución en función de la energía (figura 6.6, 6.8 y 6.11), son cotas inferiores de las verdaderas cantidades, ya que el MGAC privilegia aquellas estructuras con menor energía. Por lo tanto, existe un número apreciable de estructuras con bajas energías que forman un conjunto aceptable de candidatas a ser mínimo global. Desafortunadamente esto puede contribuir al rápido estancamiento de la evolución del GA, y de allí surge la necesidad de realizar varias simulaciones independientes del MGAC. En principio, ningún método sofisticado de búsqueda que intente muestrear más eficientemente los posibles cristales mediante información adquirida a medida que se va ejecutando, evita la necesidad de realizar extensas búsquedas para capturar a las estructuras más significativas. Este problema se ve intensificado por el hecho de que el cálculo de la energía no se encuentra bien definido, por lo que es difícil distinguir cuáles estructuras aportan información relevante para hallar el mínimo global. Para cualquier desarrollo futuro es indispensable poseer una metodología para el muestreado eficiente de los posibles cristales. Es por eso que una de las

principales contribuciones de esta tesis es la creación de un marco para implementar dicho muestreo.

Una posible extensión a lo desarrollado en este trabajo es la inclusión de información experimental, que imponga restricciones a la distancia entre átomos o sobre las conformaciones de las moléculas, para minimizar los errores asociados con el cálculo de la energía. Aunque esto no resolvería el problema por primeros principios, sí permitiría la elucidación de cristales para los cuales sólo se tiene información experimental parcial. Esto sería de gran utilidad para aquellos cristales cuyas estructuras no pueden resolverse por difracción de rayos X.

De todas formas, los resultados mostrados en esta tesis permiten concluir que, para una gran variedad de casos, es posible la predicción de estructuras cristalinas utilizando solamente información de la molécula y su conectividad. Sin embargo, será necesaria la validación del método sobre el conjunto más extenso posible de estructuras conocidas experimentalmente, para proponer estrategias adecuadas para detectar aquellos casos en que la predicción pueda fallar. Estas estrategias formarán parte del manual del código que será accesible a los usuarios que lo requieran en un futuro próximo.

Finalmente esta tesis muestra la necesidad de la búsqueda sistemática y eficiente para la predicción de las estructuras de clusters atómicos. Se mostró que la utilización combinada de métodos semi-empíricos y GA puede ser una solución eficiente para predecir las estructuras de estos sistemas. Por lo tanto, se ve justificada la implementación y estudio detallado de esta estrategia en futuros trabajos.

Apéndice A

Operadores de recombinación

A.1. Recombinación a un punto

Este operador toma a dos individuos (los padres) y genera dos nuevos (la descendencia), ver figura A.1.

A.2. Recombinación a dos puntos

Es la generalización de la operación anterior a dos puntos, ver figura A.2.

A.3. Recombinación a N puntos

Se genera al azar un número binario con una longitud igual al tamaño del genoma, a este número se lo conoce como máscara. La máscara indica que genes serán intercambiados

entre los padres, ver figura A.3.

A.4. Recombinación uniforme

Se selecciona un gen al azar y se intercambia, generando dos descendientes, ver figura A.4.

A.5. Recombinación media aritmética

Este operador actúa sobre dos padres generando un solo descendiente. Para mantener el número de individuos entre poblaciones, se incluye un segundo descendiente, que es alguno de sus padres seleccionado al azar, ver figura A.5.

A.6. Recombinación de media geométrica

Este operador actúa sobre dos padres generando un solo descendiente. Al igual que antes, se incluye un segundo descendiente, que es alguno de los padres seleccionado al azar, ver figura A.6.

A.7. Inversión

Sean dos números aleatorios $r, q \in [1, N]$ tales que $r < q$, entonces se aplica sobre un solo individuo $C_b(k) = C_a(q + r - k)$ donde $r \leq k \leq q$, ver figura A.7.

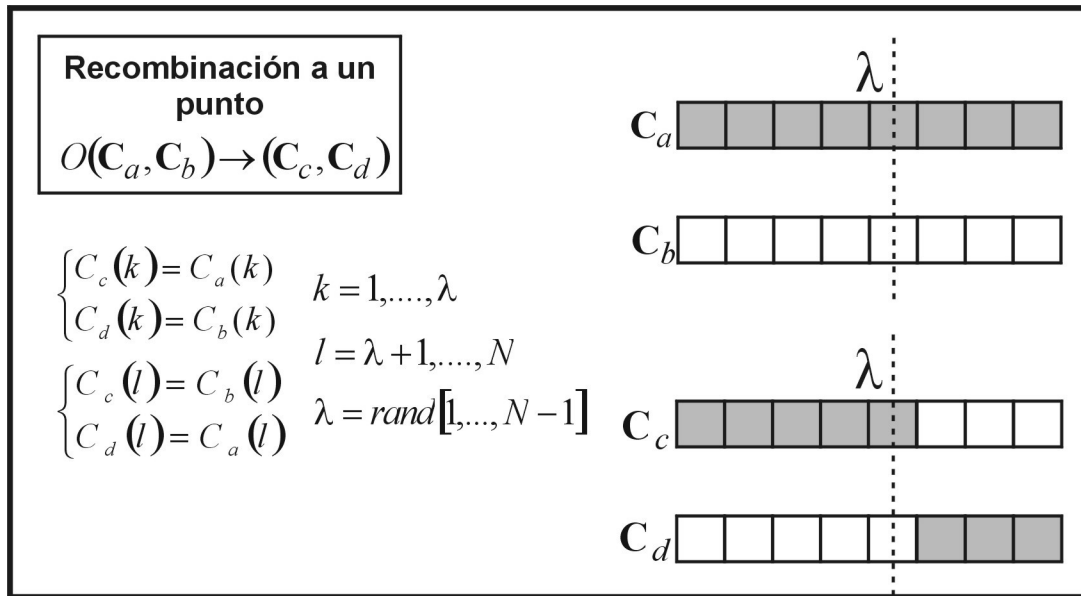


Figura A.1: Esquema del operador de recombinación a un punto

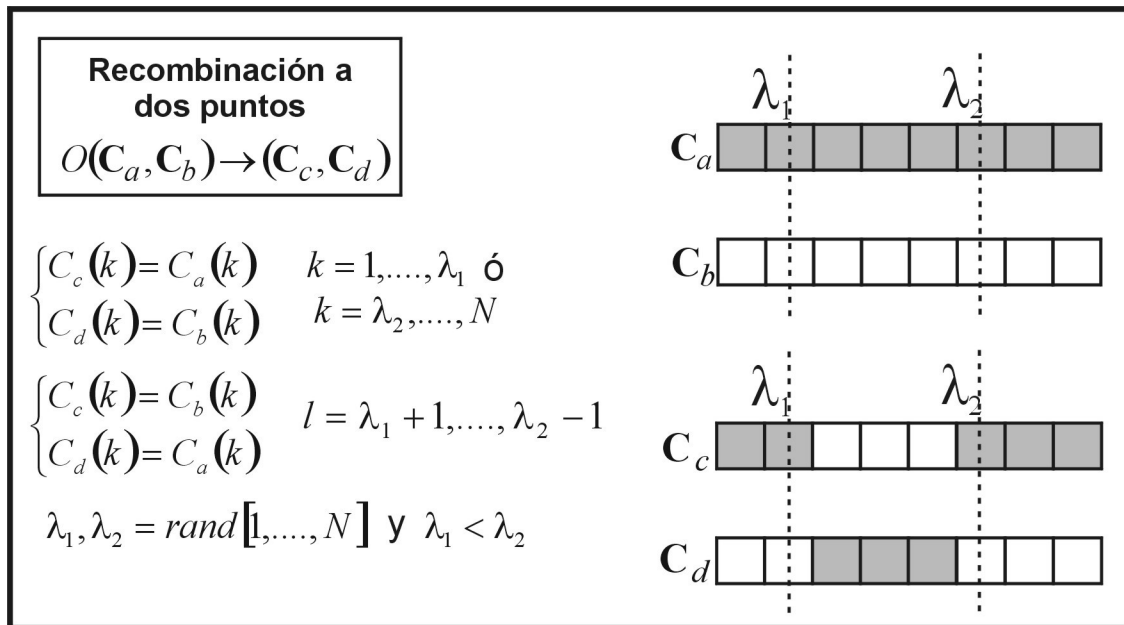


Figura A.2: Esquema del operador de recombinación a dos puntos

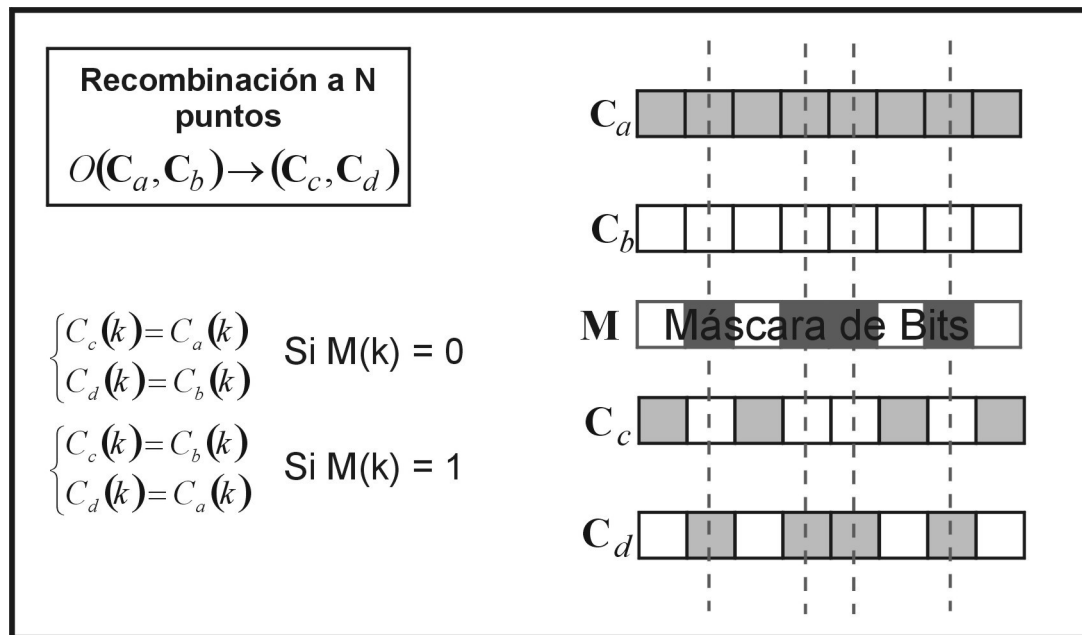


Figura A.3: Esquema del operador de recombinación a N puntos

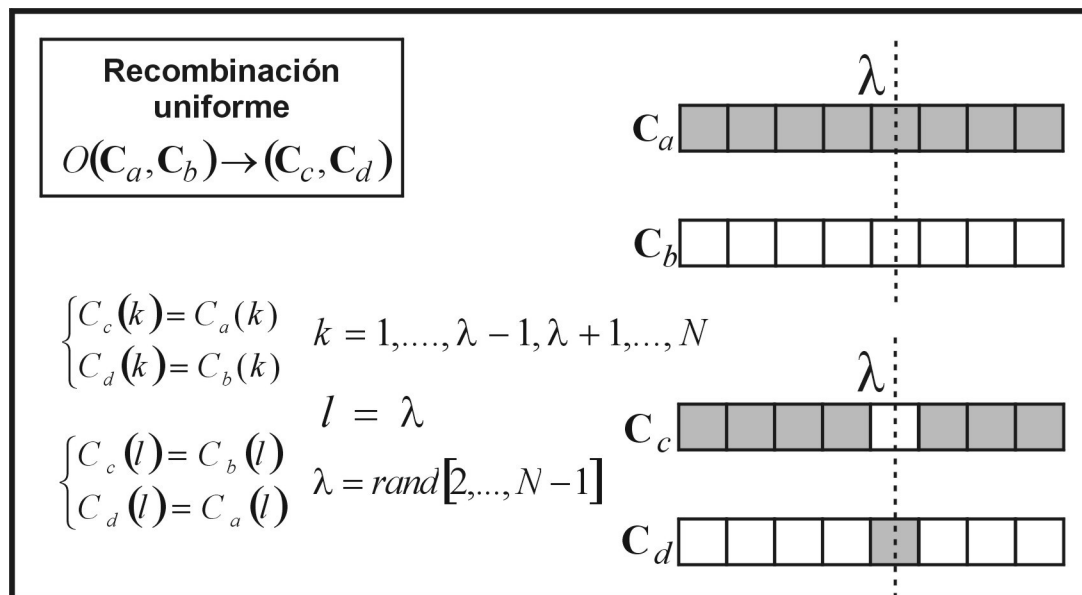


Figura A.4: Esquema del operador de recombinación uniforme

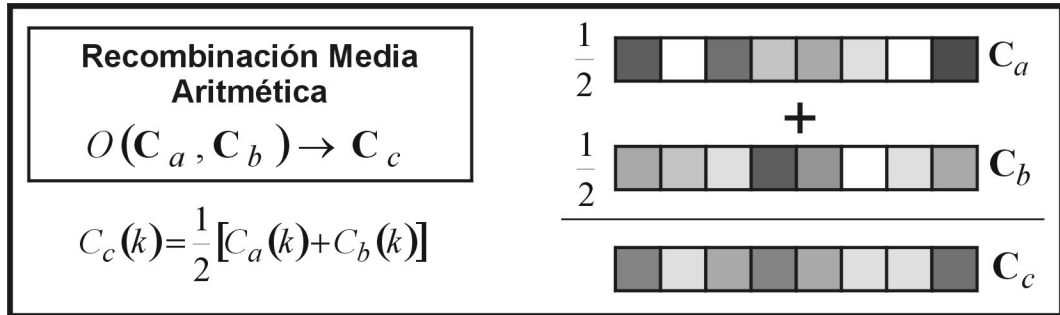


Figura A.5: Esquema del operador de media aritmética

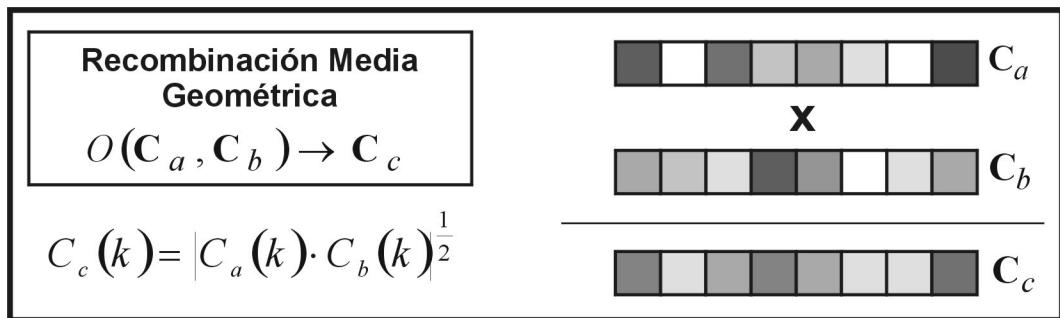


Figura A.6: Esquema del operador de media geométrica

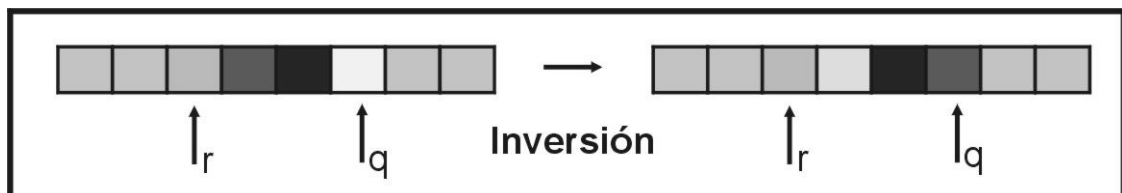


Figura A.7: Esquema del operador inversión

Apéndice B

Material suplementario

B.1. Cristales de imidazolidina-2,4-diona

orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
1	6.064	6.346	11.746	89.96	117.72	89.99	400	P-1	P21/c	0.0000
2	8.769	12.087	7.598	90.00	101.27	90.00	790	C2c		0.0822
3	3.934	12.032	8.439	90.00	100.45	90.00	393	P21c		0.1755
4	15.953	12.056	10.444	90.00	128.33	90.00	1576	C2c		0.1949
5	12.076	10.44	6.289	90.00	90.00	90.00	793	Pbca	Pbcm	0.2589
6	10.33	6.357	12.35	90.00	90.00	90.00	811	P212121	Pbca	0.3084
7	6.19	10.324	6.347	90.00	89.98	90.00	406	Cc	Cmc21	0.3107
8	17.627	6.119	7.598	90.00	104.91	90.00	792	C2	C2/c	0.3453
9	4.996	9.076	9.875	116.28	102.17	91.44	389	P-1		0.3481
10	17.618	6.113	7.596	90.01	104.95	89.99	790	P-1	C2/c	0.3521
11	3.838	6.065	17.093	90.00	99.72	90.00	392	P21	P21/c	0.3956
12	6.189	6.368	11.675	89.97	117.92	89.97	407	P-1	P21/c	0.4177
13	6.369	6.174	10.345	90.00	90.00	90.00	407	Pca21	Pnma	0.4194
14	6.281	20.322	8.85	90.00	134.19	90.00	810	P21c		0.4418
15	6.378	6.619	9.557	90.00	90.00	90.00	403	P21	P212121	0.4541
16	3.925	23.937	8.518	90.00	99.81	90.00	789	P21c		0.4733
17	10.397	6.339	12.058	90.00	90.00	90.00	795	P212121	Pnma	0.5154
18	3.828	6.052	8.637	90.00	98.86	90.00	198	P21	P21	0.5908
19	8.7	6.085	7.585	90.00	99.29	90.00	396	P21	P21/c	0.5916
20	6.113	20.95	6.249	90.00	90.00	90.00	800	P21c	Pbcm	0.6042
21	3.82	6.052	8.628	90.00	98.77	90.00	197	P21	P21	0.6085

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
22	12.384	6.307	20.585	90.00	90.00	90.00	1608	Pbca		0.6744
23	12.334	6.056	10.595	90.00	94.23	90.00	789	C2		0.6749
24	16.752	4.079	11.783	90.00	91.50	90.00	805	C2	C2/c	0.7492
25	34.469	6.152	7.478	90.00	90.00	90.00	1586	Pbcn		0.7498
26	3.832	5.981	8.735	93.94	98.95	95.60	196	P-1	P-1	0.7569
27	6.111	6.873	10.595	86.32	89.95	64.22	400	P-1		0.7872
28	9.733	6.593	6.356	90.00	90.00	90.01	408	P21	Pna21	0.7951
29	6.239	10.606	12.278	90.00	90.00	90.00	812	P212121		0.7989
30	17.042	3.984	11.908	90.00	90.92	90.00	808	P-1	C2/c	0.8001
31	6.118	6.224	10.6	89.99	90.30	90.02	404	P-1	P21/m	0.8193
32	8.07	6.07	8.256	90.00	99.18	90.00	399	C2		0.8261
33	19.493	6.07	8.067	90.00	56.70	90.00	798	C2	P2	0.8261
34	19.907	6.198	15.431	90.00	122.69	90.00	1602	C2c		0.8281
35	6.087	8.232	8.924	76.10	73.35	71.90	402	P-1		0.8294
36	8.075	6.068	8.277	90.00	99.28	90.00	400	P21	C2	0.8332
37	20.928	6.303	6.086	89.99	90.00	90.00	803	P21c	Pnma	0.8555
38	3.839	8.856	11.914	87.01	83.78	78.79	395	P-1		0.8753
39	3.819	5.977	8.766	94.45	99.05	95.14	196	P-1	P-1	0.8776
40	3.839	24.587	8.664	90.00	100.43	90.00	804	Cc		0.8843
41	7.298	7.432	7.794	104.30	101.21	94.90	398	P-1		0.9190
42	5.974	6.812	10.658	83.17	83.30	65.98	392	P-1		0.9322
43	5.196	9.086	9.669	63.78	76.31	89.40	396	P-1		0.9735
44	3.861	11.802	8.827	90.00	101.30	90.00	394	Pc		0.9919
45	6.655	6.693	9.84	105.99	104.18	96.60	401	P-1		0.9978
46	17.447	6.186	7.343	90.00	90.00	90.00	793	Pna21	Pccn	1.0336
47	17.713	3.79	14.548	90.00	123.53	90.00	814	C2		1.0391
48	8.52	6.064	8.102	90.00	108.07	90.00	398	P21	P21/c	1.0507
49	6.183	7.384	17.406	90.00	90.00	90.00	795	P212121	Pbca	1.0595
50	8.184	6.113	8.162	90.00	99.10	90.00	403	P21	P21/c	1.0641
51	6.354	7.475	9.471	74.33	76.30	71.37	405	P-1		1.0642
52	19.98	6.131	15.318	90.00	121.09	90.00	1607	C2c		1.0675
53	8.119	4.944	9.996	90.00	105.90	90.00	386	P21	P21/c	1.0727
54	5.009	5.068	8.242	87.60	78.03	72.60	195	P-1	P-1	1.0777
55	3.846	11.758	8.854	90.00	100.80	90.00	393	P21		1.0784
56	20.36	21.314	3.718	90.00	89.98	90.00	1613	Cc	Fdd2	1.0796
57	12.504	12.261	10.666	90.00	93.19	90.00	1633	C2c		1.0831
58	5.001	32.816	6.098	90.00	127.55	90.00	794	P21c		1.0959
59	3.807	22.145	9.521	90.00	101.09	90.00	788	Cc		1.1001
60	12.263	10.61	12.558	90.00	90.00	90.00	1634	Pbcn		1.1019
61	11.873	10.48	6.421	90.00	96.61	90.00	794	P21c		1.1059
62	3.813	6.007	17.581	90.00	98.75	90.00	398	P21	P21/c	1.1089
63	4.61	8.219	11.084	92.69	97.59	105.70	399	P-1		1.1146
64	10.801	12.382	6.238	90.00	107.67	90.00	795	C2	C2/c	1.1160
65	3.817	35.075	6.834	90.00	119.06	90.00	800	P21c		1.1180
66	10.472	6.416	23.713	90.00	90.00	90.00	1593	Pbca		1.1257
67	10.741	12.351	6.239	89.91	107.59	90.37	789	P-1	C2/c	1.1311
68	3.801	24.364	9	90.00	104.89	90.00	805	P21c		1.1456

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
69	11.947	10.494	6.36	90.00	90.00	90.00	797	Pbcm	Pbcm	1.1472
70	16.056	6.044	10.63	90.00	128.26	90.00	810	C2		1.1598
71	20.913	6.35	12.103	90.00	90.00	90.00	804	Pbca	Pbcm	1.1907
72	6.291	10.62	6.038	90.00	90.00	90.00	403	Pna21	Pmn21	1.1926
73	35.467	3.765	6.043	90.00	90.00	90.00	807	Pca21		1.1956
74	6.761	9.704	6.872	89.98	115.68	90.00	406		P21/c	1.2024
75	6.851	10.847	21.214	90.00	90.00	90.00	1576	Pbca		1.2030
76	6.057	20.987	6.34	90.00	90.00	90.00	806	Pbcn	Pbcm	1.2054
77	5.83	8.707	8.809	102.40	107.52	101.40	400	P-1		1.2156
78	3.716	21.721	9.871	90.00	98.11	90.00	789	Cc		1.2164
79	3.833	34.477	6.067	90.00	90.17	90.00	802	P21c		1.2200
80	10.085	3.913	11.881	90.00	122.70	90.00	395	P21	P21/c	1.2270
81	3.724	17.408	7.019	90.00	118.50	90.00	400	P21c		1.2325
82	3.806	35.367	6.106	90.00	99.60	90.00	810	P21c		1.2363
83	8.124	5.979	16.436	89.98	98.16	89.96	790	P-1	C2/c	1.2429
84	13.118	10.45	6.331	90.00	108.96	90.00	821	C2c		1.2482
85	10.569	12.326	7.474	90.00	123.60	90.00	811	C2c	C2/c	1.2537
86	12.289	6.174	21.397	90.00	94.09	90.00	1619	C2c		1.2558
87	17.066	6.181	15.181	90.00	90.00	90.00	1601	Pbcn		1.2599
88	6.084	10.579	6.34	90.00	92.89	90.00	408	Pc		1.2663
89	10.421	12.056	12.691	90.00	90.00	90.00	1594	Pbca		1.2757
90	3.754	35.11	7.099	90.00	121.20	90.00	800	P21c		1.2800
91	3.805	23.482	9.08	90.00	101.46	90.00	795	Cc		1.2859
92	6.105	6.289	10.605	89.96	90.17	90.10	407	P-1	P21/m	1.2874
93	5.129	7.237	11.714	76.11	84.77	71.50	400	P-1		1.3015
94	10.148	11.872	6.641	90.00	90.00	90.00	800	Pna21	Pccn	1.3041
95	3.9	5.811	17.779	90.03	90.00	90.00	403	P21	P212121	1.3060
96	3.941	23.681	8.69	90.00	101.00	90.00	796	P21c		1.3239
97	34.658	3.822	6.053	90.00	90.00	90.00	802	Pca21		1.3293
98	3.848	9.263	11.633	92.92	99.01	101.35	400	P-1		1.3422
99	18.283	6.209	7.479	90.00	103.78	89.97	825	P-1	C2/c	1.3433
100	6.016	6.335	10.683	89.83	96.69	89.96	404	P-1	P21/m	1.3448

Tabla B.1: Primeras 100 estructuras cristalinas en orden de energía creciente para la molécula de imidazolidina-2,4-diona. La estructura de la segunda posición coincide con la estructura del cristal conocida experimentalmente para esta molécula.

B.2. Cristales de N-(2-dimetil-4,5-dinitrofenil)acetamida

orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
1	4.842	14.142	15.118	90.00	102.53	90.00	1011	P21/c		0.0000
2	14.865	8.097	16.749	90.00	90.00	90.00	2016	Pbca		0.0929

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
3	12.554	4.799	19.406	90.00	118.58	90.00	1027	P21	P21/c	0.2726
4	4.849	9.659	11.953	109.99	96.84	96.94	514	P-1		0.4338
5	8.262	12.971	9.590	90.00	95.65	90.00	1023	P21/c		0.4841
6	10.247	11.064	9.470	90.00	105.71	90.00	1034	P21/c		0.5678
7	5.030	13.115	15.861	90.00	107.79	90.00	996	P21/c		0.5738
8	4.915	14.318	7.319	90.00	102.79	90.00	502	P21		0.5743
9	8.105	16.598	8.179	90.00	112.34	90.00	1018	P21/c		0.6767
10	3.955	16.617	7.809	90.00	98.58	90.00	508	P21		0.8289
11	9.037	15.783	7.481	90.00	112.19	90.00	988	P21/c		0.8587
12	8.422	9.536	15.112	90.00	121.81	90.00	1031	Pc	P21/c	0.8911
13	11.821	4.825	19.397	90.00	111.96	90.00	1026	P21	P21/c	0.8988
14	7.324	14.988	9.524	90.00	99.36	90.00	1032	Pc	P21/c	0.9806
15	8.310	8.748	14.823	94.40	95.98	100.99	1047	P-1		0.9967
16	8.420	8.814	14.460	86.82	82.59	79.64	1046	P-1		1.0105
17	9.408	14.723	14.990	90.00	90.00	90.00	2076	P212121		1.0661
18	4.773	14.640	15.048	90.00	92.83	90.00	1050	P21/c		1.0955
19	9.201	7.995	14.330	90.00	100.15	90.00	1038	P21		1.0977
20	4.828	14.921	14.448	90.00	95.12	90.00	1037	P21		1.1162
21	4.853	44.712	9.659	90.00	97.17	90.00	2080	P21/c		1.1186
22	4.756	15.937	14.197	90.00	94.45	90.00	1073	P21/c		1.1599
23	8.090	15.938	9.096	90.00	118.15	90.00	1034	P21/c		1.2701
24	11.059	19.412	4.826	90.00	90.00	90.00	1036	Pna21		1.2777
25	8.254	9.150	13.684	90.00	91.03	90.00	1033	Pc	P21/c	1.2832
26	28.500	9.214	8.004	90.00	90.00	90.00	2102	Pna21		1.2908
27	3.880	32.408	9.123	90.00	114.76	90.00	1042	P21/c		1.2920
28	4.702	14.553	15.179	90.00	97.66	90.00	1029	P21		1.2934
29	4.874	25.429	8.434	90.00	100.30	90.00	1028	P21/c		1.3071
30	13.164	4.770	20.056	90.00	126.08	90.00	1018	P21	P21/c	1.3078
31	4.773	15.819	13.783	90.00	90.99	90.00	1041	Pc		1.3208
32	7.861	15.086	9.401	90.00	113.51	90.00	1022	P21		1.3219
33	9.139	14.365	8.316	90.00	108.12	90.00	1038	P21/c		1.3391
34	15.015	7.677	17.452	90.00	90.00	90.00	2012	Pbca		1.3498
35	4.719	29.386	15.172	90.00	91.92	90.00	2103	P21/c		1.3704
36	3.987	8.455	16.368	75.47	88.10	81.97	529	P-1		1.3794
37	7.127	10.026	14.463	90.00	90.00	90.00	1033	P212121		1.3877
38	4.783	13.912	31.470	90.00	90.00	90.00	2094	P212121		1.3977
39	7.147	15.515	10.305	90.00	118.96	90.00	1000	Pc	P21/c	1.4161
40	29.407	4.725	15.179	90.00	90.00	90.00	2109	Pca21		1.4164
41	8.511	14.618	8.535	90.00	100.21	90.00	1045	P21		1.4415
42	7.609	9.488	7.671	90.00	110.84	90.00	518	P21		1.4508
43	7.209	15.845	18.528	90.00	90.00	90.00	2116	P212121		1.4569
44	4.868	8.454	13.135	96.91	95.54	101.16	522	P-1		1.4731
45	4.708	15.163	14.890	90.00	98.16	90.00	1052	P21/c		1.4780
46	32.436	3.834	8.269	90.00	90.00	90.00	1028	Pna21	Pca21	1.4961
47	14.521	14.913	19.216	90.00	90.00	90.00	4161	Pbca		1.5010
48	7.892	8.252	8.314	88.72	80.02	78.30	522	P-1		1.5096
49	15.900	7.506	33.528	90.00	90.00	90.00	4001	Pbcn		1.5169

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
50	9.528	9.643	11.183	87.25	84.71	81.78	1012	P-1		1.5266
51	31.829	4.769	13.858	90.00	90.00	90.00	2104	Pca21		1.5319
52	8.747	12.140	9.430	90.00	93.17	90.00	1000	P21	P21/c	1.5329
53	4.814	11.877	18.448	90.00	94.08	90.00	1052	P21/c		1.5411
54	3.908	64.918	8.199	90.00	91.41	90.00	2079	P21/c		1.5686
55	5.030	14.439	14.218	90.00	98.17	90.00	1022	Pc	P21/c	1.5712
56	4.815	13.740	15.863	90.00	98.58	90.00	1038	P21		1.5731
57	4.932	9.643	11.099	96.38	90.22	94.79	523	P-1		1.5742
58	4.679	7.573	15.304	102.85	91.34	94.34	527	P-1		1.5901
59	9.371	14.913	15.061	90.00	90.00	90.00	2105	P212121		1.6055
60	4.901	15.516	13.667	90.00	100.56	90.00	1022	P21/c		1.6070
61	4.619	7.769	30.029	90.00	92.79	90.00	1076	P21/c		1.6318
62	9.124	9.459	13.809	82.33	78.08	64.78	1054	P-1		1.6400
63	8.302	9.391	13.933	98.16	99.39	95.91	1052	P-1		1.6599
64	7.671	14.554	9.599	90.00	103.93	90.00	1040	P21/c		1.7047
65	4.862	28.671	15.176	90.00	93.76	90.00	2111	P21/c		1.7076
66	9.362	7.839	14.406	90.00	95.62	90.00	1052	P21		1.7082
67	8.944	14.272	9.287	90.00	118.61	90.00	1041	P21		1.7186
68	8.741	14.770	8.284	90.00	100.46	90.00	1052	P21/c		1.7315
69	8.237	14.179	9.942	90.00	113.98	90.00	1061	P21		1.7341
70	4.892	9.751	11.454	105.39	93.19	95.64	522	P-1		1.7446
71	12.697	4.771	17.514	90.00	101.37	90.00	1040	P21		1.7449
72	4.695	16.824	13.481	90.00	94.96	90.00	1061	Pc		1.7501
73	31.104	3.955	16.583	90.00	90.00	90.00	2040	Pca21		1.7506
74	34.119	12.872	4.789	90.00	90.00	90.00	2103	Pna21		1.7658
75	7.562	17.219	7.886	90.00	90.87	90.00	1027	P21		1.7690
76	4.770	13.974	16.935	68.30	83.02	81.22	1034	P-1		1.7705
77	10.380	13.393	8.076	90.00	113.58	90.00	1029	Pc	Cc	1.7712
78	7.578	15.299	9.391	90.00	92.94	90.00	1087	P21	P21/c	1.7781
79	4.621	7.738	14.870	90.00	95.46	90.00	529	P21		1.7802
80	4.821	7.057	15.407	77.20	85.13	84.03	507	P-1		1.7805
81	8.987	10.410	11.445	88.03	82.46	72.49	1012	P-1		1.7831
82	9.494	7.957	14.360	90.00	97.49	90.00	1076	Pc		1.7917
83	16.510	3.946	16.729	90.00	104.09	90.00	1057	P21	P21/c	1.8052
84	12.563	4.817	17.566	90.00	101.77	90.00	1041	P21		1.8139
85	32.506	4.681	14.043	90.00	90.00	90.00	2137	Pca21		1.8144
86	15.802	7.490	33.901	90.00	90.00	90.00	4012	Pbca		1.8171
87	9.129	9.474	13.909	101.28	94.44	115.14	1050	P-1		1.8204
88	7.235	9.677	14.574	90.00	96.88	90.00	1013	P21/c		1.8275
89	9.114	9.452	14.006	101.28	94.23	115.54	1050	P-1		1.8584
90	4.079	10.531	11.911	90.00	98.33	90.00	506	P21		1.8617
91	8.953	7.586	15.556	90.00	105.86	90.00	1016	P21		1.8675
92	14.555	8.227	35.689	90.00	90.00	90.00	4274	Pbcn		1.8691
93	9.475	10.540	12.252	64.59	82.13	72.76	1055	P-1		1.8691
94	4.778	31.382	13.988	90.00	90.65	90.00	2097	P21/c		1.8723
95	8.333	9.413	14.003	79.14	78.70	84.39	1056	P-1		1.8755
96	17.957	9.406	23.912	90.00	90.17	90.00	4039	C2/c		1.8827

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
97	8.036	14.112	18.223	90.00	90.00	90.00	2067	P212121		1.8918
98	22.539	12.460	7.453	90.00	90.00	90.00	2093	Pna21		1.9072
99	9.428	7.987	14.302	90.00	99.68	90.00	1062	P21		1.9137
100	16.043	15.519	16.493	90.00	90.00	90.00	4106	Pbca		1.9172
101	4.684	16.077	28.292	90.00	90.00	90.00	2131	P212121		1.9186
102	9.130	14.245	9.319	90.00	118.08	90.00	1069	P21		1.9198
103	4.885	14.498	29.210	90.00	90.00	90.00	2069	P212121		1.9268
104	4.754	13.899	31.761	90.00	90.00	90.00	2099	P212121		1.9303
105	7.963	8.468	8.713	96.68	91.69	115.04	527	P-1		1.9355
106	3.876	8.370	31.909	90.00	90.00	90.00	1035	P21	P212121	1.9397
107	4.839	14.710	14.810	90.00	95.90	90.00	1049	Pc		1.9407
108	8.253	15.111	16.894	90.00	90.00	90.00	2107	P212121		1.9545
109	4.794	14.149	31.095	90.00	90.00	90.00	2109	P212121		1.9565
110	7.663	8.772	7.853	90.00	92.33	90.00	527	P21		1.9574
111	9.106	9.361	14.429	76.88	89.80	63.38	1064	P-1		1.9582
112	4.743	35.073	12.812	90.00	90.69	90.00	2131	P21/c		1.9640
113	7.609	13.516	10.110	90.00	104.81	90.00	1005	P21	P21/c	1.9693
114	7.838	14.582	17.558	90.00	90.00	90.00	2007	P212121	Pbca	1.9721
115	14.455	16.781	17.641	90.00	90.00	90.00	4279	Pbca		1.9751
116	9.918	10.378	12.303	90.00	126.20	90.00	1022	P21	P21/c	1.9794
117	8.383	9.498	13.063	91.31	99.74	98.65	1012	P-1		1.9797
118	31.186	4.532	15.259	90.00	102.99	90.00	2101	C2/c		1.9924
119	8.160	9.269	14.529	76.85	81.88	85.06	1058	P-1		1.9943
120	13.548	15.814	19.151	90.00	90.00	90.00	4103	Pbca		2.0120
121	8.841	8.877	13.464	98.09	92.68	100.55	1026	P-1		2.0133
122	8.901	9.154	14.626	84.98	88.04	61.77	1046	P-1		2.0302
123	4.771	30.460	14.717	90.00	90.36	90.00	2138	P21/c		2.0368
124	7.405	9.083	16.074	104.25	96.40	90.06	1041	P-1		2.0464
125	4.599	15.175	15.759	90.00	105.11	90.00	1062	P21/c		2.0570
126	3.859	64.733	8.997	90.00	111.63	90.00	2089	P21/c		2.0597
127	4.824	37.136	12.282	90.00	95.28	90.00	2191	P21/c		2.0603
128	9.098	9.487	13.859	81.67	77.55	64.33	1051	P-1		2.0608
129	9.094	9.290	14.241	84.76	81.85	63.41	1065	P-1		2.0669
130	59.361	4.610	7.771	90.00	90.00	90.00	2127	Pca21		2.0671
131	4.721	14.151	15.791	90.00	94.73	90.00	1051	P21		2.0750
132	9.007	9.026	14.038	80.12	79.64	67.56	1031	P-1		2.0828
133	7.967	16.636	9.078	90.00	119.75	90.00	1045	P21/c		2.0834
134	13.859	4.735	16.939	90.00	105.75	90.00	1070	P21		2.0836
135	18.904	12.682	9.905	90.00	119.64	90.00	2064	C2/c		2.0837
136	9.157	28.118	8.131	90.00	94.96	90.00	2086	P21/c		2.0955
137	11.671	18.598	4.843	90.00	90.00	90.00	1051	P212121	P21212	2.0961
138	5.131	14.260	14.377	90.00	103.04	90.00	1025	Pc	P21/c	2.1014
139	13.585	33.210	4.701	90.00	90.00	90.00	2121	Pna21		2.1099
140	4.748	15.744	14.799	90.00	101.05	90.00	1086	Pc		2.1214
141	17.028	4.763	26.838	90.00	102.66	90.00	2124	C2/c		2.1273
142	10.281	14.009	14.564	90.00	90.00	90.00	2098	P212121		2.1282
143	4.874	17.176	25.085	90.00	90.00	90.00	2100	P212121		2.1322

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
144	8.415	8.822	14.548	90.73	98.83	102.30	1042	P-1		2.1338
145	14.763	7.846	34.955	90.00	90.00	90.00	4049	Pbca		2.1349
146	9.261	9.590	11.678	90.13	95.90	99.73	1017	P-1		2.1380
147	4.711	10.127	22.005	90.00	98.51	90.00	1038	P21/c		2.1513
148	8.294	13.387	9.418	90.00	90.00	90.00	1046	Pc	Pna21	2.1541
149	4.826	52.616	8.313	90.00	99.38	90.00	2083	P21/c		2.1545
150	4.841	29.289	15.697	90.00	107.21	90.00	2126	P21/c		2.1577
151	8.803	10.562	12.811	103.74	95.17	111.24	1058	P-1		2.1648
152	3.863	19.243	13.852	90.00	96.45	90.00	1023	P21/c		2.1677
153	13.755	32.816	4.657	90.00	90.00	90.00	2102	Pna21		2.1765
154	7.418	9.166	15.635	104.06	90.70	91.46	1031	P-1		2.1779
155	14.772	15.456	4.737	90.00	90.00	90.00	1082	Pc	Pna21	2.1783
156	7.226	10.207	13.884	90.00	90.00	90.00	1024	P21	P212121	2.1917
157	14.618	8.212	18.177	90.00	90.00	90.00	2182	Pbcn		2.1928
158	3.883	32.045	9.089	90.00	112.42	90.00	1046	Pc		2.1934
159	4.867	11.187	19.588	90.00	100.19	90.00	1050	Pc		2.1959
160	17.997	7.873	14.384	90.00	90.00	90.00	2038	Pca21		2.1988
161	8.766	9.838	12.756	75.45	84.49	86.06	1059	P-1		2.1993
162	3.947	63.234	8.853	90.00	106.83	90.00	2115	P21/c		2.2004
163	4.821	17.606	12.926	90.00	103.17	90.00	1068	P21/c		2.2018
164	8.570	10.062	13.181	107.36	99.27	94.57	1061	P-1		2.2122
165	16.006	7.522	16.763	90.00	90.00	90.00	2018	Pbcn		2.2129
166	9.025	11.797	9.534	90.00	95.41	90.00	1010	P21	P21/c	2.2134
167	24.904	16.989	4.903	90.00	90.00	90.00	2074	Pna21		2.2136
168	3.892	64.708	9.050	90.00	114.75	90.00	2070	P21/c		2.2223
169	9.572	10.349	10.907	88.34	85.83	77.17	1051	P-1		2.2238
170	4.808	14.207	30.646	90.00	90.00	90.00	2093	P212121		2.2272
171	8.764	14.222	17.077	90.00	90.00	90.00	2129	P212121		2.2284
172	14.696	8.224	35.849	90.00	90.00	90.00	4332	Pbcn		2.2330
173	3.937	16.844	32.410	90.00	90.00	90.00	2149	P212121		2.2363
174	4.877	13.022	33.540	90.00	90.00	90.00	2130	P212121		2.2484
175	4.947	7.340	14.813	98.84	95.91	102.51	514	P1		2.2488
176	4.799	7.608	14.459	90.46	98.69	96.60	518	P1		2.2549
177	9.626	8.121	15.790	90.00	119.34	90.00	1076	Pc		2.2576
178	10.497	12.433	16.054	90.00	90.00	90.00	2095	P212121		2.2577
179	4.768	10.260	11.103	91.98	91.07	101.76	531	P-1		2.2583
180	15.399	7.603	34.790	90.00	90.00	90.00	4073	Pbca		2.2627
181	4.675	14.654	31.165	90.00	90.00	90.00	2135	P212121		2.2628
182	6.336	17.811	8.961	90.00	91.50	90.00	1011	Pc	P21/c	2.2703
183	7.827	15.627	9.447	90.00	117.33	90.00	1027	P21/c		2.2739
184	11.547	11.802	15.589	90.00	93.10	90.00	2121	C2/c		2.2753
185	8.195	15.154	9.391	90.00	115.26	90.00	1055	P21		2.2757
186	4.747	14.146	31.923	90.00	90.00	90.00	2143	P212121		2.2856
187	4.739	30.823	14.651	90.00	102.11	90.00	2092	P21/c		2.2865
188	4.665	60.847	8.559	90.00	118.43	90.00	2136	P21/c		2.2874
189	4.935	10.106	21.081	94.37	93.68	97.44	1036	P-1		2.2946

Tabla B.2: Primeras 189 estructuras cristalinas en orden de energía creciente para la molécula de N-(2-dimetil-4,5-dinitrofenil)acetamida. La estructura de la tercera posición coincide con la estructura cristalina conocida experimentalmente para esta molécula.

B.3. Cristales del ácido 2-(acetiloxi)benzoico

orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
1	8.027	4.064	25.065	90.00	90.00	90.00	818	P21	P212121	0.0000
2	7.852	4.174	12.867	90.00	101.26	90.00	414	P21		0.0820
3	7.938	25.775	4.007	90.00	90.00	90.00	820	Pc	Pna21	0.3896
4	14.050	4.090	15.533	90.00	112.10	90.00	827	P21	P21/c	0.4895
5	12.607	4.186	15.795	90.00	99.97	90.00	821	P21	P21/c	0.5274
6	33.235	4.065	27.975	90.00	119.24	90.00	3298	C2/c		0.5897
7	4.076	7.909	12.071	86.49	83.88	87.78	386	P-1		0.7875
8	4.053	18.637	10.834	90.00	97.31	90.00	812	P21/c		0.7977
9	3.967	22.309	9.351	90.00	102.37	90.00	808	P21/c		0.8640
10	4.018	7.946	12.998	99.09	92.48	92.77	409	P-1		0.8876
11	26.482	3.950	18.455	90.00	122.15	90.00	1635	C2/c		0.9097
12	4.001	13.411	15.905	101.74	94.66	96.84	825	P-1		1.0195
13	4.336	10.747	17.492	90.02	90.00	90.00	815	P21	P212121	1.0291
14	4.041	18.853	22.081	90.00	90.00	90.00	1682	P212121		1.0502
15	21.845	9.151	4.114	90.00	90.00	90.00	822	Pc	Pna21	1.0504
16	6.784	7.249	8.636	101.09	90.75	94.12	416	P-1		1.0531
17	4.250	10.864	35.252	90.00	90.00	90.00	1628	P212121		1.0608
18	22.215	18.947	3.966	90.00	90.00	90.00	1669	Pna21		1.0822
19	4.062	43.734	9.969	90.00	110.53	90.00	1659	P21/c		1.1050
20	6.347	6.590	10.384	85.13	73.88	86.59	415	P-1		1.1139
21	13.145	4.118	16.411	90.00	111.29	90.00	828	P21		1.1308
22	12.941	4.049	18.517	90.00	122.08	90.00	822	P21	P21/c	1.1418
23	9.216	9.111	12.756	90.00	131.57	90.00	801	P21	P21/c	1.1436
24	4.043	42.539	9.563	90.00	95.48	90.00	1637	P21/c		1.1758
25	3.955	36.001	11.611	90.00	90.64	90.00	1653	P21/c		1.1873
26	12.999	4.046	18.909	90.00	125.01	90.00	815	P21	P21/c	1.1938
27	7.851	4.669	22.736	90.00	98.12	90.00	825	P21	P21/c	1.1939
28	13.213	4.045	16.199	90.00	103.07	90.00	843	P21		1.2081
29	6.125	6.990	19.496	95.73	92.26	95.67	825	P-1		1.2370
30	7.471	5.758	19.674	90.00	105.00	90.00	818	P21	P21/c	1.2558
31	22.285	4.058	18.525	90.00	102.82	90.00	1634	C2/c		1.2596
32	4.081	11.046	18.328	90.00	96.93	90.00	820	Pc	P21/c	1.2624
33	11.932	10.032	13.639	90.00	90.00	90.00	1633	Pbca		1.2926
34	7.163	10.651	11.392	77.81	89.59	80.29	837	P-1		1.2942
35	4.764	10.199	17.176	90.00	90.00	90.00	835	P212121		1.2942
36	22.378	4.013	18.692	90.00	90.00	90.00	1679	Pca21		1.3103
37	13.130	4.002	16.941	90.00	111.71	90.00	827	P21		1.3234
38	4.004	9.448	22.382	90.00	90.00	90.00	847	P21	P212121	1.3311
39	4.037	15.685	13.256	90.00	90.43	90.00	839	P21		1.3879
40	11.598	18.226	3.937	90.00	90.00	90.00	832	Pc	Pna21	1.3925
41	9.516	21.941	3.971	90.00	90.00	90.00	829	Pc	Pna21	1.3940
42	6.796	33.675	9.583	90.00	130.73	90.00	1662	P21/c		1.4270
43	3.981	13.174	15.833	90.00	95.80	90.00	826	Pc		1.4339
44	26.509	3.978	15.818	90.00	90.00	90.00	1668	Pca21		1.4465
45	26.911	4.096	14.977	90.00	90.00	90.00	1651	Pca21		1.4900

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
46	3.922	9.420	22.425	90.00	95.19	90.00	825	P21	P21/c	1.4944
47	28.667	3.993	15.585	90.00	112.39	90.00	1650	C2/c		1.4950
48	4.013	15.956	12.960	90.00	94.98	90.00	827	P21		1.5380
49	26.109	8.816	15.033	90.00	107.43	90.00	3301	C2/c		1.5626
50	4.416	17.918	10.966	90.00	103.71	90.00	843	P21/c		1.5797
51	4.099	12.648	17.195	110.15	95.34	96.61	823	P-1		1.6138
52	12.932	4.021	16.274	90.00	102.58	90.00	826	P21		1.6568
53	9.478	13.741	24.845	90.00	90.00	90.00	3236	Pbca		1.6785
54	6.503	6.610	20.190	90.00	106.22	90.00	833	P21	P21/c	1.6904
55	5.047	13.244	13.310	108.95	97.01	93.45	830	P-1		1.6915
56	3.924	8.188	25.063	90.00	94.08	90.00	803	P21	P21/c	1.6983
57	9.074	4.488	20.195	90.00	92.64	90.00	822	P21		1.7527
58	21.592	18.618	4.144	90.00	90.00	90.00	1666	Pna21		1.7614
59	6.793	16.756	9.585	90.00	130.28	90.00	832	P21/c		1.7642
60	6.503	11.187	11.544	97.11	91.90	90.25	833	P-1		1.7761
61	4.051	36.719	11.609	90.00	103.16	90.00	1681	P21/c		1.7934
62	8.469	5.172	9.873	90.00	104.70	90.00	418	P21		1.8155
63	9.514	18.621	9.664	90.00	94.01	90.00	1708	P21/c		1.8388
64	8.592	5.815	16.389	90.00	90.00	90.00	819	P21	P212121	1.8472
65	5.046	11.339	29.209	90.00	90.00	90.00	1671	P212121		1.8497
66	6.357	6.455	20.407	95.55	97.19	96.39	820	P-1		1.8535
67	20.817	4.637	17.938	90.00	111.39	90.00	1612	C2/c		1.8776
68	4.566	10.378	17.458	90.00	96.91	90.00	821	P21	P21/c	1.8830
69	29.621	4.092	15.363	90.00	115.23	90.00	1684	C2/c		1.8845
70	6.410	11.708	11.356	90.00	100.98	90.00	837	P21/c		1.8885
71	22.422	3.984	9.443	90.00	100.03	90.00	831	C2		1.8958
72	8.830	9.692	10.678	110.75	92.79	101.49	830	P-1		1.9099
73	11.311	5.071	14.504	90.00	90.00	90.00	832	P21	P212121	1.9192
74	9.622	4.028	22.078	90.00	103.05	90.00	833	P21	P21/c	1.9320
75	8.844	13.281	14.719	90.00	101.35	90.00	1695	P21/c		1.9375
76	4.024	25.982	8.790	90.00	114.06	90.00	839	P21/c		1.9553
77	4.862	13.597	25.059	90.00	90.00	90.00	1657	P212121		1.9661
78	4.097	24.897	16.273	90.00	92.99	90.00	1658	Cc		1.9690
79	6.043	22.159	7.927	90.00	129.46	90.00	819	P21/c		1.9819
80	3.966	26.829	15.680	90.00	94.72	90.00	1663	Cc		1.9999
81	4.492	13.285	27.662	90.00	90.00	90.00	1651	P212121		2.0046
82	35.341	4.829	9.970	90.00	90.00	90.00	1701	Pca21		2.0063
83	12.964	3.977	16.299	90.00	99.94	90.00	828	P21	P21/c	2.0230
84	17.343	10.368	4.768	90.00	90.00	90.00	857	Pc	Pna21	2.0287
85	8.735	5.003	19.951	90.00	110.57	90.00	816	P21	P21/c	2.0345
86	36.226	11.485	4.022	90.00	90.00	90.00	1673	Pna21		2.0420
87	7.924	5.071	20.978	90.00	100.69	90.00	828	P21		2.0515
88	4.028	24.936	8.049	90.00	93.69	90.00	807	P21/c		2.0550
89	16.448	4.883	21.541	90.00	92.19	90.00	1729	C2/c		2.0584
90	8.494	9.067	13.084	90.00	124.95	90.00	826	P21	P21/c	2.0796
91	4.127	18.784	21.582	90.00	90.00	90.00	1673	P212121		2.0813
92	8.688	13.332	14.701	90.00	103.52	90.00	1656	P21/c		2.1071

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
93	10.828	7.141	14.062	90.00	127.70	90.00	860	P21	P21/c	2.1261
94	9.536	20.886	9.583	90.00	118.78	90.00	1673	P21/c		2.1297
95	21.512	12.538	12.372	90.00	90.00	90.00	3337	Pbca		2.1666
96	10.796	6.948	11.925	90.00	106.22	90.00	859	P21		2.1704
97	9.122	41.255	4.434	90.00	92.57	90.00	1667	Cc		2.1736
98	7.474	5.008	11.896	90.00	107.32	90.00	425	P21		2.1741
99	4.297	42.245	10.125	90.00	112.08	90.00	1703	P21/c		2.1784
100	6.144	8.147	16.921	95.49	96.03	99.41	826	P-1		2.1925
101	18.508	8.125	21.691	90.00	90.00	90.00	3262	Pbca		2.2041
102	3.974	13.295	17.216	112.15	93.52	92.48	839	P-1		2.2179
103	3.890	23.830	9.330	90.00	107.33	90.00	826	Pc	P21/c	2.2414
104	3.924	8.232	24.997	90.00	94.07	90.00	805	P21		2.2509
105	10.162	11.427	14.365	90.00	90.00	90.00	1668	P212121	Pbca	2.2556
106	6.907	12.955	9.978	90.00	115.87	90.00	803	P21/c		2.2626
107	12.493	5.002	17.912	90.00	127.75	90.00	885	P21	P21/c	2.2652
108	7.835	10.892	9.522	90.00	94.45	90.00	810	P21	P21/c	2.2796
109	18.884	22.545	4.031	90.00	90.00	90.00	1716	Pna21		2.2799
110	9.746	9.700	10.003	90.00	115.52	90.00	853	Pc	P21/c	2.2870
111	20.436	4.787	19.506	90.00	113.94	90.00	1744	C2/c		2.2911
112	14.252	3.906	15.875	90.00	108.21	90.00	840	P21	P21/c	2.2959
113	8.013	4.702	22.685	90.00	100.09	90.00	841	P21		2.2994
114	8.740	6.701	14.643	90.00	102.23	90.00	838	Pc		2.3070
115	3.935	8.661	24.794	81.72	87.11	84.32	831	P-1		2.3077
116	4.726	9.285	10.739	68.65	78.41	78.48	426	P-1		2.3272
117	19.598	9.718	8.702	90.00	90.00	90.00	1657	Pbcn		2.3477
118	21.719	4.147	19.377	90.00	102.04	90.00	1707	C2/c		2.3552
119	5.424	13.436	12.148	90.00	110.68	90.00	828	P21/c		2.3724
120	31.346	11.857	4.578	90.00	90.00	90.00	1702	Pna21		2.3825
121	9.463	7.350	11.999	90.00	98.89	90.00	825	P21		2.3862
122	9.886	36.105	4.761	90.00	90.00	90.00	1699	Pna21		2.3868
123	16.591	4.814	21.258	90.00	99.44	90.00	1675	C2/c		2.3919
124	5.018	24.126	7.534	90.00	115.97	90.00	820	P21/c		2.3940
125	10.283	4.967	16.872	90.00	93.28	90.00	860	P21		2.3951
126	8.204	4.971	20.489	90.00	90.61	90.00	836	P21		2.4432
127	5.322	18.144	18.308	90.00	90.00	90.00	1768	P212121		2.4447
128	9.012	10.209	10.758	73.13	69.92	65.04	830	P-1		2.4747
129	6.830	7.379	8.843	92.87	111.04	102.04	403	P1	P-1	2.4811
130	7.600	14.202	7.714	90.00	90.00	90.00	833	Pc	Pna21	2.4894
131	9.083	4.138	22.891	90.00	103.47	90.00	837	P21	P21/c	2.4923
132	7.046	19.263	24.875	90.00	90.00	90.00	3376	Pbca		2.5026
133	11.690	9.459	7.532	90.00	95.71	90.00	829	P21	P21/c	2.5050
134	8.666	8.891	11.860	107.02	93.96	96.39	863	P-1		2.5076
135	4.416	6.370	7.500	109.74	100.58	100.29	189	P1		2.5212
136	6.295	40.471	6.664	90.00	93.28	90.00	1695	P21/c		2.5228
137	15.697	6.781	15.952	90.00	108.84	90.00	1607	C2/c		2.5241
138	30.971	11.884	4.685	90.00	90.00	90.00	1724	Pca21		2.5363
139	4.893	30.152	12.726	90.00	112.33	90.00	1737	P21/c		2.5368

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
140	4.082	32.674	13.594	90.00	104.75	90.00	1753	P21/c		2.5480
141	4.633	14.918	25.206	90.00	90.00	90.00	1742	P212121		2.5551
142	10.595	17.473	18.311	90.00	90.00	90.00	3390	Pbca		2.5949
143	24.219	3.872	18.363	90.00	98.62	90.00	1703	C2/c		2.5973
144	5.766	30.342	9.584	90.00	97.56	90.00	1662	P21/c		2.5980
145	22.547	4.656	16.441	90.00	91.49	90.00	1725	C2/c		2.6013
146	16.184	9.449	22.405	90.00	101.04	90.00	3363	C2/c		2.6031
147	8.179	9.645	12.199	89.92	70.84	68.07	835	P-1		2.6227
148	25.509	8.860	7.765	90.00	105.87	90.00	1688	C2/c		2.6238
149	8.931	9.059	11.545	75.53	77.97	72.53	854	P-1		2.6336
150	22.011	4.749	18.706	90.00	121.84	90.00	1661	C2/c		2.6415
151	13.237	4.904	16.560	90.00	126.79	90.00	861	P21	P21/c	2.6452
152	6.739	10.213	13.560	112.06	92.11	104.31	829	P-1		2.6455
153	4.702	10.410	17.401	90.00	90.00	90.00	852	P212121		2.6484
154	8.898	13.412	14.776	90.00	99.93	90.00	1737	P21/c		2.6496
155	9.620	14.346	12.210	90.00	102.70	90.00	1644	P21/c		2.6647
156	22.423	8.903	8.473	90.00	98.71	90.00	1672	C2/c		2.6651
157	9.977	7.559	13.957	90.00	126.53	90.00	846	P21	P21/c	2.6695
158	8.193	14.107	7.335	90.00	99.82	90.00	835	P21/c		2.6718
159	9.296	9.852	11.196	67.12	78.44	64.69	853	P-1		2.6774
160	7.582	8.235	13.765	82.20	84.34	74.70	820	P-1		2.6829
161	4.615	12.943	15.080	66.91	86.52	79.93	816	P-1		2.6874
162	4.029	7.940	12.878	79.75	87.03	88.23	405	P1		2.6899
163	23.687	9.634	17.465	90.00	121.82	90.00	3387	C2/c		2.6911
164	9.345	10.645	16.778	90.00	90.00	90.00	1669	Pbca		2.6931
165	6.395	6.997	19.079	80.31	84.02	89.53	837	P-1		2.7026
166	13.660	6.859	17.699	90.00	90.00	90.00	1658	P212121	Pbca	2.7033
167	6.724	17.762	28.131	90.00	90.00	90.00	3360	Pbca		2.7112
168	7.225	14.474	16.164	90.00	90.00	90.00	1690	P212121		2.7199
169	6.995	25.299	11.166	90.00	123.33	90.00	1651	P21/c		2.7280
170	20.172	9.632	17.601	90.00	98.38	90.00	3383	C2/c		2.7319
171	13.612	11.271	21.935	90.00	90.00	90.00	3365	Pbca		2.7407
172	19.913	13.593	6.173	90.00	90.00	90.00	1671	Pna21		2.7449
173	10.022	43.374	4.141	90.00	110.68	90.00	1684	Cc		2.7529
174	13.420	3.938	16.114	90.00	105.52	90.00	821	P21	P21/c	2.7544
175	6.731	12.312	11.617	89.99	118.06	90.00	850	P-1	P21/c	2.7668
176	8.104	3.978	25.480	90.00	95.77	90.00	817	P21		2.7772
177	6.655	7.414	17.557	98.84	90.98	97.80	847	P-1		2.7837
178	3.883	18.028	24.037	90.00	90.00	90.00	1682	P212121		2.7841
179	21.808	7.341	15.001	90.00	133.27	90.00	1749	C2/c		2.7870
180	6.732	11.930	20.456	90.00	90.00	90.00	1643	P212121		2.7930
181	4.032	15.550	26.256	90.00	90.00	90.00	1646	P212121		2.7945
182	8.012	9.869	10.465	85.33	88.25	82.63	818	P-1		2.7968
183	6.495	15.168	16.993	90.00	90.00	90.00	1674	P212121		2.7986
184	29.693	11.672	5.000	90.00	90.00	90.00	1733	Pna21		2.7989
185	5.011	11.481	30.277	90.00	90.00	90.00	1742	P212121		2.8018
186	39.254	6.940	6.337	90.00	90.00	90.00	1726	Pca21		2.8087

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
187	6.153	14.396	10.017	90.00	97.86	90.00	879	Pc	P21/c	2.8097
188	8.603	6.825	15.065	90.00	98.37	90.00	875	P21	P21/c	2.8104
189	7.805	4.140	25.748	90.00	91.92	90.00	831	P21		2.8112
190	7.678	12.128	18.365	90.00	90.00	90.00	1710	P212121	Pbca	2.8113
191	26.888	10.174	6.247	90.00	90.00	90.00	1709	Pna21		2.8132
192	3.966	8.347	12.574	97.22	93.75	99.48	406	P1		2.8137
193	6.624	11.318	11.526	90.00	105.12	90.00	834	Pc	P21/c	2.8233
194	30.004	11.751	4.975	90.00	90.00	90.00	1754	Pca21		2.8255
195	24.585	8.546	8.558	90.00	109.16	90.00	1698	C2/c		2.8304
196	8.707	14.019	14.784	90.00	104.85	90.00	1744	C2/c		2.8328
197	4.715	13.062	27.460	90.00	90.00	90.00	1691	P212121		2.8330
198	7.083	8.690	14.603	72.82	76.56	86.18	835	P-1		2.8346
199	4.794	8.165	21.328	82.99	87.53	88.60	828	P-1		2.8346
200	7.465	9.473	12.873	82.31	87.38	67.18	832	P-1		2.8354
201	8.106	10.437	10.885	77.61	87.36	68.46	836	P-1		2.8572
202	22.014	4.744	19.226	90.00	120.58	90.00	1729	C2/c		2.8608
203	11.444	13.008	11.352	90.00	90.00	90.00	1690	Pbca		2.8614
204	8.813	9.914	10.946	93.68	95.30	114.00	865	P-1		2.8617
205	8.260	9.108	12.152	81.59	87.24	64.94	819	P-1		2.8626
206	12.864	4.125	15.789	90.00	100.56	90.00	824	P21		2.8640
207	7.069	15.983	9.002	90.00	126.14	90.00	821	P21/c		2.8652
208	34.317	5.754	8.524	90.00	90.00	90.00	1683	Pca21		2.8693
209	5.730	6.866	11.554	88.19	88.59	71.92	432	P-1		2.8712
210	4.986	15.491	22.312	90.00	90.00	90.00	1723	P212121		2.8738
211	16.497	8.814	23.409	90.00	107.55	90.00	3245	C2/c		2.8742
212	8.778	8.998	10.998	82.25	77.65	84.24	839	P-1		2.8791
213	6.774	10.755	12.834	112.23	91.09	102.44	840	P-1		2.8825
214	8.282	4.952	10.627	90.00	96.34	90.00	433	P21		2.8841
215	8.895	10.668	12.184	90.00	131.04	90.00	872	P21/c		2.8871
216	13.396	11.824	22.311	90.00	101.04	90.00	3468	C2/c		2.8886
217	9.944	10.107	17.240	90.00	99.59	90.00	1708	C2/c		2.8999
218	18.893	4.374	9.838	90.00	90.00	90.00	813	Pc	Pca21	2.9049
219	10.542	23.243	7.306	90.00	102.04	90.00	1751	P21/c		2.9065
220	14.101	7.426	17.301	90.00	111.35	90.00	1687	C2/c		2.9120
221	4.126	28.127	15.196	90.00	105.06	90.00	1703	P21/c		2.9137
222	9.026	4.634	20.121	90.00	93.85	90.00	840	P21		2.9149
223	17.855	6.202	17.352	90.00	116.85	90.00	1714	C2/c		2.9171
224	9.973	4.762	11.194	90.00	94.24	90.00	530	P21	P21/c	2.9178
225	10.342	5.088	15.864	90.00	90.00	90.00	835	P21	Pca21	2.9233
226	16.909	3.893	13.658	90.00	112.70	90.00	829	C2		2.9271
227	4.371	10.800	8.893	90.00	95.52	90.00	418	P21		2.9277
228	4.559	8.290	22.408	90.00	90.00	90.00	847	P212121		2.9294
229	7.176	15.228	15.746	90.00	90.00	90.00	1721	P212121		2.9326
230	4.604	18.189	21.024	90.00	90.00	90.00	1760	P212121		2.9370
231	7.606	7.922	8.328	96.90	109.92	113.99	411	P-1		2.9389
232	15.131	28.598	4.035	90.00	90.00	90.00	1746	Pna21		2.9438
233	7.355	7.507	8.940	69.18	75.74	68.72	426	P-1		2.9473

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
234	7.199	28.235	8.614	90.00	99.45	90.00	1727	P21/c		2.9481
235	9.336	15.680	23.158	90.00	90.00	90.00	3390	Pbca		2.9496
236	10.392	7.668	12.781	90.00	124.02	90.00	844	P21	P21/c	2.9575
237	8.778	19.824	12.500	90.00	130.48	90.00	1655	P21/c		2.9602
238	25.694	4.035	15.847	90.00	90.00	90.00	1643	Pca21		2.9650
239	9.349	27.164	6.725	90.00	104.20	90.00	1656	Cc		2.9682
240	13.087	3.930	16.102	90.00	101.80	90.00	811	P21		2.9745
241	5.338	8.453	18.970	101.05	92.25	92.81	838	P-1		2.9770
242	17.486	9.712	20.760	90.00	102.79	90.00	3438	C2/c		2.9788
243	6.902	7.242	9.387	78.00	74.61	62.57	400	P-1		2.9842
244	12.821	8.202	16.375	90.00	90.00	90.00	1722	Pbca		2.9899
245	11.130	4.058	18.309	90.00	95.81	90.00	823	P21		2.9916
246	5.272	9.191	17.363	92.95	93.84	99.15	827	P-1		2.9921
247	8.393	4.559	22.838	90.00	96.66	90.00	868	P21		2.9940
248	12.580	6.707	20.387	90.00	90.00	90.00	1720	Pbca		2.9949
249	12.842	4.481	18.396	90.00	126.97	90.00	846	P21	P21/c	3.0016
250	9.383	39.454	4.802	90.00	105.95	90.00	1709	Cc		3.0042
251	4.762	8.360	20.909	84.53	89.34	89.88	828	P-1		3.0047
252	9.379	7.715	23.388	90.00	90.00	90.00	1692	Pbca		3.0081
253	15.003	11.940	18.795	90.00	90.00	90.00	3367	Pbca		3.0124
254	4.184	14.207	29.118	90.00	90.00	90.00	1731	P212121		3.0124
255	5.130	15.624	21.011	90.00	90.00	90.00	1684	P212121		3.0165
256	3.982	8.229	25.819	96.82	90.48	93.00	839	P-1		3.0176
257	6.860	17.955	7.237	90.00	115.08	90.00	807	Pc	P21/c	3.0177
258	8.943	12.472	8.634	90.00	114.88	90.00	874	Pc	P21/c	3.0188
259	17.030	16.270	13.654	90.00	114.31	90.00	3448	C2/c		3.0240
260	7.937	9.332	12.311	75.20	76.48	73.00	831	P-1		3.0252
261	8.092	8.728	12.171	80.39	84.88	80.41	834	P-1		3.0276
262	5.144	13.530	23.808	90.00	90.00	90.00	1657	P212121	Pbca	3.0277
263	18.772	9.334	18.786	90.00	95.35	90.00	3277	C2/c		3.0318
264	6.866	9.333	13.865	100.37	91.83	106.30	836	P-1		3.0347
265	20.925	12.169	13.092	90.00	96.77	90.00	3310	C2/c		3.0372
266	8.159	15.242	26.526	90.00	90.00	90.00	3299	Pbca		3.0385
267	15.460	26.714	4.231	90.00	90.00	90.00	1747	Pna21		3.0402
268	3.891	25.730	8.136	90.00	90.76	90.00	814	Pc		3.0404
269	11.641	13.303	12.084	90.00	114.18	90.00	1707	C2	C2/c	3.0448
270	6.900	7.353	8.380	96.54	91.86	97.90	418	P1	P-1	3.0491
271	8.337	4.602	12.023	90.00	116.28	90.00	414	Pc		3.0502
272	17.793	10.334	9.308	90.00	99.53	90.00	1688	C2/c		3.0518
273	21.712	34.528	4.471	90.00	90.00	90.00	3352	Cc	Fdd2	3.0531
274	28.282	15.422	3.936	90.00	90.00	90.00	1717	Pna21		3.0551
275	6.613	8.183	15.926	100.81	91.19	104.17	819	P-1		3.0557
276	8.637	9.615	10.927	91.95	109.75	95.92	847	P-1		3.0558
277	5.145	34.243	10.282	90.00	99.53	90.00	1787	P21/c		3.0565
278	8.682	8.744	11.861	96.78	95.48	111.67	821	P-1		3.0617
279	18.547	4.846	20.108	90.00	97.19	90.00	1793	C2/c		3.0644
280	35.973	3.984	11.337	90.00	90.00	90.00	1625	Pca21		3.0662

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
281	8.237	4.121	24.938	90.00	97.09	90.00	840	P21		3.0716
282	12.292	4.916	14.877	90.00	110.71	90.00	841	P21	P21/c	3.0759
283	7.558	14.717	7.706	90.00	101.45	90.00	840	P21/c		3.0763
284	7.182	10.753	11.809	96.51	100.59	103.75	859	P-1		3.0819
285	4.810	16.955	20.300	90.00	90.00	90.00	1655	P212121		3.0836
286	10.036	12.498	13.764	90.00	90.00	90.00	1727	P212121		3.0880
287	4.449	19.495	19.817	90.00	90.00	90.00	1719	P212121		3.0938
288	36.059	10.497	4.459	90.00	90.00	90.00	1688	Pna21		3.0989
289	8.470	9.806	11.022	71.93	81.59	89.63	860	P-1		3.0989
290	6.858	35.911	7.233	90.00	114.77	90.00	1617	P21/c		3.1023
291	9.113	8.892	11.362	90.00	117.44	90.00	817	Pc	P21/c	3.1023
292	3.931	8.089	25.730	90.00	91.57	90.00	818	P21		3.1027
293	31.200	4.446	24.872	90.00	90.00	90.00	3450	Pbcn		3.1031
294	33.113	4.026	12.515	90.00	90.00	90.00	1668	Pca21		3.1060
295	5.033	16.646	19.995	90.00	90.00	90.00	1675	P212121		3.1071
296	4.151	12.786	32.921	90.00	90.00	90.00	1747	P212121		3.1105
297	8.298	3.995	25.180	90.00	102.37	90.00	815	P21	P21/c	3.1112
298	14.771	5.335	21.128	90.00	90.00	90.00	1665	P212121	Pbca	3.1136
299	13.931	7.531	16.253	90.00	90.00	90.00	1705	P212121	Pbca	3.1185
300	27.057	3.872	16.299	90.00	105.21	90.00	1648	C2/c		3.1208
301	8.947	9.380	10.134	97.09	97.92	96.29	829	P-1		3.1245
302	26.589	15.731	3.991	90.00	90.00	90.00	1669	Pna21		3.1276
303	7.673	17.476	24.912	90.00	90.00	90.00	3341	Pbca		3.1277
304	5.149	10.443	16.603	105.21	92.35	96.87	853	P-1		3.1280
305	10.057	29.919	5.542	90.00	90.00	90.00	1668	Pna21		3.1290
306	14.607	12.103	18.529	90.00	90.00	90.00	3276	Pbca		3.1385
307	8.570	20.519	12.413	90.00	129.92	90.00	1674	P21/c		3.1388
308	7.612	14.136	15.528	90.00	90.00	90.00	1671	P212121		3.1433
309	9.177	10.306	9.441	90.00	107.97	90.00	849	P21/c		3.1450
310	23.997	18.645	3.828	90.00	90.00	90.00	1713	Pna21		3.1462
311	8.348	4.097	24.183	90.00	90.77	90.00	827	P21		3.1466
312	3.916	16.807	25.039	90.00	90.00	90.00	1648	P212121		3.1485
313	5.060	16.341	20.535	90.00	90.00	90.00	1698	P212121		3.1516
314	13.093	14.266	18.069	90.00	90.00	90.00	3375	Pbca		3.1520
315	7.145	14.038	17.066	90.00	90.00	90.00	1712	P212121		3.1576
316	9.163	4.454	21.134	90.00	99.65	90.00	850	P21	P21/c	3.1620
317	4.110	16.720	26.135	90.00	90.00	90.00	1796	P212121		3.1630
318	3.854	14.394	15.295	90.00	93.74	90.00	847	Pc		3.1671
319	4.049	15.105	16.030	81.76	83.53	86.75	963	P-1		3.1699
320	22.754	8.594	8.767	90.00	90.00	90.00	1715	Pbcn		3.1777
321	8.756	8.911	11.341	70.32	83.99	88.41	829	P-1		3.1805
322	21.140	9.486	17.284	90.00	102.73	90.00	3381	C2/c		3.1868
323	8.610	4.437	23.271	90.00	98.26	90.00	880	P21		3.1868
324	11.783	5.020	15.776	90.00	111.72	90.00	867	P21		3.1881
325	6.079	10.239	14.569	72.24	87.36	78.60	846	P-1		3.1902
326	9.373	14.037	6.556	90.00	101.16	90.00	846	P21/c		3.1927
327	4.795	8.283	10.721	90.08	92.01	101.02	418	P1		3.1931

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
328	13.242	7.495	8.678	90.00	90.00	90.00	861	Pc	Pna21	3.1931
329	9.320	11.745	7.780	90.00	101.56	90.00	834	P21/c		3.1933
330	25.616	11.581	5.682	90.00	90.00	90.00	1686	Pna21		3.1969
331	11.912	36.180	3.985	90.00	90.00	90.00	1717	Pna21		3.1974
332	15.694	9.888	21.543	90.00	90.00	90.00	3343	Pbca		3.1985
333	4.073	28.187	15.300	90.00	95.58	90.00	1748	P21/c		3.2006
334	17.600	10.021	18.774	90.00	90.00	90.00	3311	Pbca		3.2008
335	4.919	9.431	18.228	90.00	90.00	90.00	846	P212121		3.2008
336	16.840	8.006	24.870	90.00	90.00	90.00	3353	Pbcn		3.2023
337	10.123	12.013	13.998	90.00	90.00	90.00	1702	Pbca		3.2029
338	8.460	10.295	10.783	112.24	91.55	98.21	857	P-1		3.2048
339	5.268	15.043	21.539	90.00	90.00	90.00	1707	P212121		3.2081
340	8.149	23.653	8.938	90.00	93.16	90.00	1720	P21/c		3.2107
341	9.389	11.978	31.118	90.00	90.00	90.00	3499	Pbca		3.2120
342	16.544	4.433	11.405	90.00	90.00	90.00	837	P21	Pna21	3.2120
343	15.641	11.452	4.860	90.00	90.00	90.00	870	Pc	Pna21	3.2138
344	10.033	5.203	16.024	90.03	90.00	90.00	837	Pc	Pna21	3.2165
345	12.817	11.196	23.434	90.00	90.00	90.00	3363	Pbca		3.2180
346	7.956	14.996	7.033	90.00	95.54	90.00	835	Pc		3.2207
347	4.260	8.877	11.678	79.51	83.06	85.99	431	P1		3.2257
348	5.387	13.251	23.667	90.00	90.00	90.00	1689	P212121		3.2274
349	7.753	9.202	12.497	70.35	84.13	74.35	809	P-1		3.2277
350	17.534	7.509	15.440	90.00	123.96	90.00	1686	C2/c		3.2279
351	16.220	9.586	21.996	90.00	97.41	90.00	3392	C2/c		3.2280
352	7.556	14.481	30.383	90.00	90.00	90.00	3325	Pbca		3.2317
353	8.806	18.532	13.422	90.00	128.29	90.00	1719	P21/c		3.2328
354	14.910	28.881	3.947	90.00	90.00	90.00	1700	Pna21		3.2346
355	4.226	8.590	24.267	90.00	90.00	90.00	881	P212121		3.2362
356	6.224	10.466	25.653	90.00	90.00	90.00	1671	P212121		3.2379
357	4.858	9.457	9.823	76.93	81.86	77.29	427	P-1		3.2415
358	8.270	4.428	23.878	90.00	98.62	90.00	865	P21		3.2417
359	4.616	9.119	10.295	81.18	87.10	84.30	426	P1	P-1	3.2425
360	4.521	18.274	20.286	90.00	90.00	90.00	1676	P212121	Pbca	3.2465
361	7.995	7.739	13.524	90.00	101.32	90.00	820	P21		3.2474
362	7.514	10.567	11.189	105.70	95.18	102.55	824	P-1		3.2482
363	5.853	12.130	11.992	90.00	93.59	90.00	850	P21/c		3.2509
364	4.978	12.321	15.632	111.60	96.70	96.51	873	P-1		3.2526
365	10.487	17.333	9.676	90.00	109.66	90.00	1656	C2/c		3.2547
366	8.964	6.345	15.419	90.00	111.98	90.00	813	P21	P21/c	3.2561
367	13.244	12.081	11.024	90.00	91.97	90.00	1763	C2/c		3.2584
368	18.150	7.733	14.820	90.00	126.34	90.00	1676	C2	C2/c	3.2585
369	22.192	7.288	11.891	90.00	112.08	90.00	1782	C2		3.2596
370	9.301	5.208	17.851	90.00	98.34	90.00	856	P21		3.2602
371	10.663	7.094	11.184	90.00	91.11	90.00	846	P21	P21/c	3.2625
372	11.691	4.949	14.639	90.00	102.24	90.00	828	P21		3.2628
373	22.272	14.432	5.243	90.00	90.00	90.00	1685	Pna21		3.2633
374	4.167	32.296	13.335	90.00	105.11	90.00	1733	P21/c		3.2640

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
375	8.969	26.618	7.608	90.00	112.91	90.00	1673	P21/c		3.2652
376	7.996	9.937	11.654	88.45	86.76	69.76	867	P-1		3.2719
377	12.610	4.343	15.404	90.00	95.41	90.00	840	P21		3.2749
378	25.893	15.827	4.274	90.00	90.00	90.00	1751	Pna21		3.2761
379	8.304	8.854	13.365	104.75	96.73	114.25	839	P-1		3.2814
380	8.327	14.543	27.538	90.00	90.00	90.00	3335	Pbcn		3.2841
381	12.632	15.913	16.533	90.00	90.00	90.00	3323	Pbca		3.2850
382	8.424	5.752	18.220	90.00	104.19	90.00	856	Pc	P21/c	3.2889
383	7.265	18.642	24.602	90.00	90.00	90.00	3332	Pbcn		3.2904
384	14.378	29.652	3.980	90.00	90.00	90.00	1697	Pna21		3.2947
385	18.913	9.411	18.591	90.00	93.25	90.00	3304	C2/c		3.2997
386	7.039	9.366	12.935	87.14	77.62	83.77	828	P-1		3.3002
387	11.594	30.980	4.938	90.00	90.09	90.00	1773	Cc		3.3026
388	21.804	9.829	16.688	90.00	102.87	90.00	3487	C2/c		3.3034
389	11.454	3.951	18.571	90.00	103.06	90.00	819	P21	P21/c	3.3040
390	4.365	8.319	12.226	71.61	87.16	76.01	409	P-1		3.3057
391	15.681	11.707	4.887	90.00	90.00	90.00	897	Pc	Pca21	3.3063
392	4.832	15.183	24.673	90.00	90.00	90.00	1810	P212121		3.3068
393	12.096	4.753	15.184	90.00	110.59	90.00	817	P21		3.3091

Tabla B.3: Primeras 393 estructuras cristalinas en orden de energía creciente para el ácido 2-(acetiloxi)benzoico.

B.4. Cristales del ácido 2-(acetiloxi)benzoico (conformación B)

orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
1	12.371	6.301	11.280	90.00	112.58	90.00	812	P21	P21/c	1.2007
2	11.419	6.319	11.313	90.00	94.69	90.00	814	P21	P21/c	1.3854
3	3.924	8.143	25.201	90.00	94.68	90.00	803	P21	P21/c	1.6780
4	7.869	4.180	25.893	90.00	105.96	90.00	819	P21	P21/c	1.9419
5	13.528	3.921	16.117	90.00	108.20	90.00	812	P21	P21/c	2.2447
6	8.473	8.847	13.068	90.00	124.84	90.00	804	P21	P21/c	2.2616
7	11.190	5.123	14.519	90.00	97.26	90.00	826	P21	P21/c	2.4092
8	11.479	6.289	12.459	90.00	112.05	90.00	834	P21		2.4354
9	7.835	4.116	24.195	90.00	94.07	90.00	778	P21		2.4611
10	8.549	9.703	11.783	90.00	124.35	90.00	807	P21	P21/c	2.6052
11	10.108	5.464	15.664	90.00	111.15	90.00	807	P21	P21/c	2.6918
12	6.931	12.827	9.125	90.00	95.81	90.00	807	P21	P21/c	2.8390

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
13	7.069	15.984	9.002	90.00	126.14	90.00	821	P21/c		2.8647
14	18.894	4.372	9.841	90.00	90.00	90.00	813	P21	Pca21	2.9014
15	12.979	5.278	14.607	90.00	122.16	90.00	847	P21	P21/c	2.9143
16	4.021	25.598	8.287	90.00	106.09	90.00	820	P21/c		2.9477
17	12.294	4.914	14.879	90.00	110.70	90.00	841	P21	P21/c	3.0735
18	5.292	24.804	7.250	90.00	120.53	90.00	820	P21/c		3.2497
19	8.822	10.258	12.306	90.00	130.78	90.00	843	P21	P21/c	3.3263
20	9.721	5.468	15.569	90.00	99.54	90.00	816	P21	P21/c	3.3296
21	4.289	10.233	9.425	90.00	92.55	90.00	413	P21		3.4068
22	4.102	22.734	9.300	90.00	105.03	90.00	838	P21/c		3.4700
23	6.975	6.736	17.233	90.00	96.58	90.00	804	P21	P21/c	3.4717
24	6.807	10.953	11.093	90.00	95.51	90.00	823	P21		3.5181
25	5.270	9.486	16.466	90.00	90.81	90.00	823	P21	P21/c	3.5494
26	3.848	26.936	8.723	90.00	112.14	90.00	837	P21/c		3.5746
27	26.807	3.899	8.075	90.00	102.59	90.00	824	P21	C2	3.5892
28	8.808	8.885	11.342	90.00	109.35	90.00	838	P21/c		3.6571
29	13.487	3.981	16.223	90.00	95.15	90.00	868	P21		3.7623
30	9.411	8.466	12.638	90.00	123.76	90.00	837	P21	P21/c	3.7647
31	10.861	5.467	15.239	90.00	113.91	90.00	827	P21	P21/c	3.7786
32	4.494	9.771	19.055	90.00	90.00	90.00	837	P21	P212121	3.8030
33	11.465	6.415	15.347	90.00	130.33	90.00	860	P21	P21/c	3.8495
34	4.016	9.096	23.038	90.00	94.65	90.00	839	P21	P21/c	3.8556
35	6.104	21.465	8.365	90.00	130.71	90.00	831	P21/c		3.8870
36	4.378	9.956	19.479	90.00	93.91	90.00	847	P21	P21/c	3.8974
37	5.402	10.986	14.292	90.00	98.82	90.00	838	P21	P21/c	3.9229
38	8.543	6.469	15.228	90.00	91.53	90.00	841	P21	P21/c	3.9684
39	7.462	8.429	13.563	90.00	104.83	90.00	825	P21/c		3.9758
40	14.643	7.795	7.181	90.00	103.61	90.00	797	P21	P21/c	3.9836
41	7.158	6.860	17.567	90.00	90.25	90.00	863	P21	P21/c	3.9906
42	7.737	6.234	18.593	90.00	110.12	90.00	842	P21	P21/c	4.0249
43	4.051	9.161	22.824	90.00	92.46	90.00	846	P21		4.0297
44	14.734	3.799	16.225	90.00	114.48	90.00	827	P21	P21/c	4.0654
45	6.242	5.535	23.807	90.00	90.00	90.00	823	P21	P212121	4.0838
46	4.250	7.902	24.581	90.00	90.00	90.00	825	P21	P212121	4.0921
47	10.819	6.755	14.820	90.00	129.16	90.00	840	P21	P21/c	4.0990
48	3.854	30.256	7.631	90.00	115.73	90.00	802	P21/c		4.1256
49	6.333	6.587	21.353	90.00	105.60	90.00	858	P21	P21/c	4.1403
50	7.776	4.323	12.517	90.00	98.55	90.00	416	P21		4.1498
51	11.850	5.481	15.369	90.00	122.61	90.00	841	P21	P21/c	4.1698
52	6.845	6.043	20.285	90.00	91.03	90.00	839	P21	P21/c	4.1872
53	4.273	11.480	17.072	90.00	95.32	90.00	834	P21	P21/c	4.2081
54	12.212	4.429	15.446	90.00	96.57	90.00	830	P21	P21/c	4.2317
55	13.443	3.861	15.837	90.00	99.83	90.00	810	P21	P21/c	4.2334
56	10.918	4.106	18.547	90.00	100.14	90.00	818	P21	P21/c	4.2364
57	9.112	5.804	15.304	90.00	92.07	90.00	809	P21		4.2569
58	13.569	3.903	15.230	90.00	92.96	90.00	806	P21		4.2821
59	13.618	3.877	16.035	90.00	108.61	90.00	802	P21	P21/c	4.2891

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
60	9.534	6.605	16.183	90.00	122.45	90.00	860	P21	P21/c	4.2974
61	7.964	3.994	27.215	90.00	96.20	90.00	861	P21	P21/c	4.3556
62	4.224	13.873	14.475	90.00	96.12	90.00	843	P21		4.3949
63	3.809	7.058	14.904	90.00	91.57	90.00	401	P21		4.4176
64	11.282	4.233	17.991	90.00	103.31	90.00	836	P21	P21/c	4.4277
65	4.429	10.332	18.235	90.00	92.84	90.00	833	P21		4.4438
66	7.803	12.255	10.755	90.00	127.18	90.00	819	P21/c		4.4693
67	9.670	5.137	16.831	90.00	91.27	90.00	836	P21		4.4754
68	6.284	8.619	16.633	90.00	110.33	90.00	845	P21	P21/c	4.4896
69	11.173	5.521	13.194	90.00	91.83	90.00	813	P21	P21/c	4.5064
70	4.956	9.416	18.171	90.00	94.88	90.00	845	P21		4.5091
71	8.317	10.345	9.831	90.00	108.98	90.00	800	P21	P21/c	4.5119
72	23.910	7.556	4.588	90.00	90.00	90.00	829	P21	Pca21	4.5294
73	4.102	10.307	20.105	90.00	94.49	90.00	847	P21		4.5484
74	6.500	15.413	8.508	90.00	93.92	90.00	850	P21/c		4.5547
75	13.786	4.348	15.241	90.00	115.75	90.00	823	P21	P21/c	4.5684
76	11.126	5.498	13.665	90.00	97.86	90.00	828	P21		4.5745
77	12.846	9.004	14.050	90.00	90.00	89.97	1625	P21	C2221	4.5995
78	3.905	8.243	25.121	90.00	94.32	90.00	806	P21		4.5996
79	9.614	4.538	18.604	90.00	98.17	90.00	803	P21	P21/c	4.6127
80	6.157	13.962	9.516	90.00	92.06	90.00	818	P21		4.6220
81	7.714	4.186	24.955	90.00	90.00	90.00	806	P21	P212121	4.6264
82	4.004	17.245	12.477	90.00	100.58	90.00	847	P21/c		4.6280
83	4.168	7.830	25.049	90.00	96.30	90.00	813	P21	P21/c	4.6297
84	3.852	7.038	29.829	90.00	94.96	90.00	806	P21	P21/c	4.6430
85	13.452	4.083	17.637	90.00	120.04	90.00	839	P21	P21/c	4.6455
86	5.875	10.407	13.924	90.00	101.65	90.00	834	P21/c		4.6737
87	3.806	8.127	26.189	90.02	90.00	90.00	810	P21	P212121	4.6863
88	11.754	5.896	12.554	90.00	109.73	90.00	819	P21		4.7042
89	9.408	8.775	10.226	90.00	93.88	90.00	842	P21	P21/c	4.7063
90	14.437	7.908	7.008	90.00	90.27	90.00	800	P21	P21/c	4.7111
91	7.338	13.162	8.505	90.00	94.28	90.00	819	P21/c		4.7233
92	3.881	30.250	7.613	90.00	115.40	90.00	807	P21/c		4.7251
93	9.659	7.342	11.357	90.01	90.00	90.00	805	P21	P212121	4.7312
94	5.443	6.243	24.274	90.00	91.31	90.00	825	P21	P21/c	4.7390
95	10.472	4.283	18.306	90.00	90.00	90.00	821	P21	P212121	4.7416
96	4.533	23.713	7.691	90.00	90.00	90.00	827	P21	Pna21	4.7568
97	8.176	6.272	17.686	90.00	112.96	90.00	835	P21	P21/c	4.7582
98	7.862	3.994	26.877	90.00	91.85	90.00	843	P21		4.7685
99	9.381	12.549	7.922	90.00	113.77	90.00	853	P21/c		4.7778
100	8.007	26.630	3.825	90.00	90.00	90.01	816	P21	Pna21	4.7781
101	9.518	6.671	15.955	90.00	121.62	90.00	863	P21	P21/c	4.7790
102	8.633	13.586	7.125	90.00	101.95	90.00	818	P21/c		4.7850
103	8.093	3.927	26.448	90.00	106.83	90.00	805	P21	P21/c	4.7877
104	5.969	7.760	9.442	90.00	107.77	90.00	417	P21		4.7902
105	8.273	5.440	20.040	90.00	108.17	90.00	857	P21	P21/c	4.8065
106	14.539	3.844	16.115	90.00	114.95	90.00	817	P21	P21/c	4.8100

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
107	5.776	9.149	16.080	90.00	98.99	90.00	839	P21		4.8226
108	7.207	8.510	13.597	90.00	100.97	90.00	819	P21	P21/c	4.8392
109	7.880	25.457	4.128	90.00	90.00	90.00	828	P21	Pna21	4.8465
110	3.865	7.127	29.343	90.00	90.00	90.00	808	P21	P212121	4.8478
111	7.856	3.977	26.374	90.00	90.87	90.00	824	P21		4.8485
112	6.159	10.598	13.730	90.00	109.84	90.00	843	P21/c		4.8494
113	9.492	6.079	15.515	90.00	113.09	90.00	823	P21	P21/c	4.8611
114	8.238	10.671	9.992	90.00	90.00	90.00	878	P21	Pna21	4.8627
115	7.321	16.413	6.924	90.00	92.05	90.00	831	P21/c		4.8697
116	13.734	4.665	13.836	90.00	107.94	90.00	843	P21		4.8731
117	12.185	4.400	16.393	90.00	110.66	90.00	822	P21		4.8775
118	9.085	4.873	20.475	90.00	106.27	90.00	870	P21	P21/c	4.8856
119	8.296	5.850	18.708	90.00	111.21	90.00	846	P21	P21/c	4.9017
120	4.036	16.444	13.448	90.00	91.47	90.00	892	P21		4.9333
121	7.695	11.485	10.307	90.00	115.13	90.00	825	P21	P21/c	4.9445
122	12.876	4.186	16.222	90.00	105.82	90.00	841	P21		4.9557
123	7.851	11.737	9.645	90.00	108.87	90.00	841	P21/c		4.9557
124	8.950	7.123	13.623	90.00	101.42	90.00	851	P21		4.9697
125	12.149	4.418	16.578	90.00	109.50	90.00	839	P21		4.9720
126	11.933	6.276	12.636	90.00	117.01	90.00	843	P21		4.9830
127	4.284	9.921	20.073	90.00	94.91	90.00	850	P21		4.9931
128	10.721	9.396	8.900	90.00	102.36	90.00	876	P21	P21/c	4.9965
129	9.007	13.707	7.526	90.00	113.79	90.00	850	P21/c		5.0005
130	11.582	9.426	7.571	90.00	90.00	90.00	826	P21	Pna21	5.0074
131	12.159	4.558	18.526	90.00	126.63	90.00	824	P21	P21/c	5.0117
132	7.442	8.988	12.581	90.00	108.60	90.00	798	P21/c		5.0237
133	10.673	7.238	11.530	90.00	103.21	90.00	867	P21		5.0361
134	13.180	3.957	15.650	90.00	95.71	90.00	812	P21		5.0409
135	8.029	8.158	12.774	90.00	91.96	90.00	836	P21		5.0421
136	10.615	5.482	15.002	90.00	110.46	90.00	818	P21		5.0441
137	5.754	8.820	16.717	90.00	93.26	90.00	847	P21	P21/c	5.0472
138	7.522	15.392	7.629	90.00	106.03	90.00	849	P21		5.0561
139	7.679	6.703	16.705	90.00	98.23	90.00	851	P21		5.0656
140	10.604	7.051	11.705	90.00	108.25	90.00	831	P21		5.0680
141	9.518	8.754	13.075	90.00	129.43	90.00	841	P21	P21/c	5.0687
142	4.011	17.632	12.565	90.00	99.45	90.00	876	P21/c		5.0746
143	10.877	5.645	13.710	90.00	94.51	90.00	839	P21		5.0750
144	12.879	4.209	16.459	90.00	112.14	90.00	826	P21		5.0847
145	5.378	12.731	13.060	90.00	111.67	90.00	831	P21/c		5.1003
146	5.346	18.010	8.900	90.00	90.39	90.00	857	P21/c		5.1108
147	6.794	9.627	13.855	90.00	102.22	90.00	886	P21		5.1191
148	9.012	5.108	18.383	90.00	91.96	90.00	846	P21		5.1204
149	9.043	4.588	20.967	90.00	90.49	90.00	870	P21		5.1229
150	8.125	7.006	14.816	90.00	90.00	90.00	843	P21	P212121	5.1306
151	9.557	4.108	21.479	90.00	100.87	90.00	828	P21	P21/c	5.1317
152	8.342	3.731	13.334	90.00	98.06	90.00	411	P21		5.1392
153	7.857	7.253	15.302	90.00	98.21	90.00	863	P21		5.1396

continúa en la siguiente página

<i>continuación de la página anterior</i>										
orden	a Å	b Å	c Å	α	β	γ	volumen Å ³	simetría	addsym	energía kcal/mol
154	8.141	15.252	6.588	90.00	90.00	90.00	818	P21	Pca21	5.1434
155	12.646	5.032	12.795	90.00	90.48	90.00	814	P21		5.1444
156	5.647	8.666	17.886	90.00	99.25	90.00	864	P21	P21/c	5.1450
157	12.786	4.640	14.245	90.00	105.85	90.00	813	P21	P21/c	5.1457
158	4.189	15.450	12.829	90.00	90.79	90.00	830	P21		5.1457
159	9.381	8.402	11.313	90.00	90.90	90.00	892	P21	P21/c	5.1514
160	6.486	16.418	8.060	90.00	98.94	90.00	848	P21		5.1567
161	13.858	4.516	13.948	90.00	103.87	90.00	847	P21		5.1579
162	15.098	3.849	13.733	90.00	95.23	90.00	795	P21	P21/c	5.1585
163	10.974	7.474	11.390	90.00	116.34	90.00	837	P21		5.1600
164	9.418	8.557	12.831	90.00	126.32	90.00	833	P21	P21/c	5.1603
165	14.246	5.481	11.421	90.00	112.66	90.00	823	P21	P21/c	5.1646
166	13.567	4.010	18.589	90.00	115.25	90.00	915	P21	P21/c	5.1658
167	7.221	7.065	16.711	90.00	104.81	90.00	824	P21	P21/c	5.1763
168	7.069	6.997	17.583	90.00	100.53	90.00	855	P21		5.1821
169	9.031	7.908	12.149	90.00	103.07	90.00	845	P21		5.1846
170	11.601	5.123	14.521	90.00	100.09	90.00	850	P21		5.1995
171	7.994	29.055	4.012	90.00	92.75	90.00	931	P21/c		5.2004
172	3.940	7.986	26.401	90.00	90.64	90.00	831	P21		5.2034
173	8.415	11.311	9.204	90.00	110.58	90.00	820	P21/c		5.2047
174	13.716	3.865	16.920	90.00	93.05	90.00	896	P21	P21/c	5.2093
175	9.395	4.122	21.123	90.00	99.22	90.00	808	P21	P21/c	5.2143
176	3.953	12.829	16.642	90.00	93.14	90.00	843	P21		5.2187
177	3.855	15.200	14.009	90.00	94.65	90.00	818	P21/c		5.2253
178	10.974	4.096	18.849	90.00	90.43	90.00	847	P21		5.2289
179	8.484	6.258	15.746	90.00	90.00	90.00	836	P21	P212121	5.2309
180	5.796	8.743	16.373	90.00	95.69	90.00	826	P21	P21/c	5.2333
181	10.361	6.111	13.440	90.00	97.23	90.00	844	P21		5.2373
182	3.839	7.069	29.630	90.00	91.14	90.00	804	P21	P21/c	5.2411
183	8.441	9.409	10.894	90.00	99.15	90.00	854	P21		5.2432
184	7.122	3.786	14.986	90.00	92.80	90.00	404	P21		5.2514
185	5.753	18.989	8.731	90.00	116.59	90.00	853	P21/c		5.2612
186	15.602	4.109	13.450	90.00	107.67	90.00	822	P21	C2	5.2617
187	8.838	4.165	23.869	90.00	93.94	90.00	877	P21		5.2632
188	8.126	3.978	25.603	90.00	98.00	90.00	820	P21		5.2636
189	9.493	7.095	15.008	90.00	121.38	90.00	863	P21	P21/c	5.2670
190	6.384	10.056	13.130	90.00	90.00	90.00	843	P21	P212121	5.2703
191	15.176	3.879	15.499	90.00	116.42	90.00	817	P21		5.2837
192	9.321	4.212	10.676	90.00	103.36	90.00	408	P21		5.2871
193	8.159	3.803	27.158	90.00	98.35	90.00	834	P21		5.2927
194	14.415	3.883	16.332	90.00	94.00	90.00	912	P21		5.3009
195	3.954	14.258	15.897	90.00	90.06	90.00	896	P21		5.3173
196	9.743	11.169	8.075	90.00	108.87	90.00	832	P21/c		5.3229
197	16.261	3.933	13.364	90.00	110.33	90.00	802	P21	P21/c	5.3233
198	8.043	5.038	23.055	90.00	99.46	90.00	921	P21	P21/c	5.3248
199	5.877	6.064	23.329	90.00	97.16	90.00	825	P21		5.3419

Tabla B.4: Primeras 199 estructuras cristalinas en orden de energía creciente para el ácido 2-(acetiloxi)benzoico generadas por el MGAC restringidas a los grupos espaciales $P2_1$ y $P2_1/c$ y a la conformación B . La energía se calcula respecto al mínimo global encontrado en la primera corrida.

La estructura de la segunda fila coincide con la estructura cristalina conocida experimentalmente para esta molécula.

Bibliografía

- [1] C. Xiao, F. Hagelberg, and W. A. Lester, *Phys. Rev. B* **66**, 075425 (2002).
- [2] Q. Sun, Q. Wang, P. Jena, S. Waterman, and Y. Kawazoe, *Phys. Rev. A* **67**, 063201 (2003).
- [3] C. Ouvrard and S. L. Price, *Cryst. Growth Des.* **20**, 262 (2004).
- [4] D. E. Goldberg, *Genetic Algorithms in Search, Optimisation and Machine Learning* (Addison-Wesley, New York, 1989).
- [5] G. R. Desiraju, *Science* **278**, 404 (1997).
- [6] J. P. M. Lommerse, W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, W. T. M. Mooij, S. L. Price, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, and D. E. Williams, *Acta Crystallogr.* **56**, 697 (2000).
- [7] W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, J. P. M. Lommerse, W. T. M. Mooij,

- S. L. Price, H. Scheraga, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, and D. E. Williams, *Acta Crystall.* **B58**, 647 (2002).
- [8] G. M. Day, W. D. S. Motherwell, H. Ammon, S. X. M. Boerrigter, R. G. D. Valle, E. Venuti, A. Dzyabchenko, J. Dunitz, B. P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, G. Karamertzanis, S. L. Price, T. C. Lewis, A. Torrissi, H. Nowell, H. Scheraga, Y. Arnautova, M. U. Schmidt, B. Schweizer, and P. Verwer, *Acta Cryst. B* **61**, 511 (2005).
- [9] W. Thiel and A. A. Voityuk, *Analyst* **120**, 2435 (1995).
- [10] J. D. Dunitz and J. Bernstein, *Acc. Chem. Res.* **28**, 193 (1995).
- [11] P. Erk, H. Hengelsberg, M. F. Haddow, and R. v. Gelder, *CrstEngComm* **6**, 474 (2004).
- [12] H. Haberland, *Clusters of Atoms and Molecules* (Spring-Verlag, Berlin, 1994).
- [13] R. L. Johnston, *Atomic and Molecular Clusters* (Taylor and Francis, London, 2002).
- [14] A. Gavezzotti, *Acc. Chem. Res.* **27**, 309 (1994).
- [15] G. M. Day, J. Chisholm, N. Shan, W. D. S. Motherwell, and W. Jones, *Cryst. Growth Des.* **4**, 1372 (2004).
- [16] F. Jensen, *Introduction to Computational Chemistry* (John Wiley and Sons, Denmark, 2001).

- [17] P. Hobza, H. L. Selzle, and E. W. Schlag, *J. Chem. Phys.* **93**, 5893 (1993).
- [18] R. J. Gdanitz, *Theoretical Aspects and Computer Modeling of the Molecular Solid State* (, 1997), chap. Ab Initio Prediction of Possible Molecular Crystal Structure.
- [19] F. H. Stillinger, *Phys. Rev. E* **59**, 48 (1999).
- [20] D. J. Wales, *Energy Landscape* (Cambridge University Press, Cambridge, 2003).
- [21] D. J. Wales and H. A. Scheraga, *Science* **285**, 1368 (1999).
- [22] R. L. Johnston and C. Roberts, *Soft Computing Approaches in Chemistry* (Springer-Verlag, Heidelberg, 2003), chap. Genetic Algorithms for the Geometry Optimization of Clusters and Nanoparticles, p. 161.
- [23] N. Nisnith, T. Bredow, and K. Jug, *J. Comput. Chem.* **25**, 1255 (2004).
- [24] L. T. Wille and J. Vennik, *J. Phys. A: Math. Gen.* **18**, L419 (1985).
- [25] M. R. Garey and D. S. Johnson, *Computers and Intractability* (W. H. Freeman and Company, New York, 1979).
- [26] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, second ed. (MIT Press and McGraw-Hill Company, USA, 2001).
- [27] V. E. Bazterra, O. B. Oa, M. C. Caputo, M. B. Ferraro, P. Fuentealba, and J. C. Facelli, *Phys. Rev. A* **69**, 053202 (2004).

- [28] O. Oña, Modelo estructural de clusters atómicos, Tesis de Licenciatura del Dpto. Física, Universidad de Buenos Aires, 2004.
- [29] I. Rata, A. A. Shvartsburg, M. Horoi, T. Frauenheim, K. W. M. Siu, and K. A. Jackson, *Phys. Rev. Lett.* **85**, 546 (2000).
- [30] B. Ahlswede and K. Jug, *J. Comput. Chem.* **20**, 563 (1999).
- [31] B. Ahlswede and K. Jug, *J. Comput. Chem.* **20**, 572 (1999).
- [32] T. Bredow, G. Geudtner, and K. Jug, *J. Comput. Chem.* **22**, 861 (2001).
- [33] J. Maddox, *Nature* **335**, 201 (1988).
- [34] F. C. Hawthorne, *Nature* **345**, 297 (1990).
- [35] P. J. Fagan and M. D. Ward, *Scientific American* **267**, 28 (1992).
- [36] T. Beyer, T. Lewis, and S. L. Price, *CrystEngComm* **44**, 1 (2001).
- [37] A. Gavezzotti, *CrystEngComm* **4**, 343 (2002).
- [38] Accelrys Inc., San Diego, USA, *Cerius 2*, 1997.
- [39] H. R. Karfunkel, F. J. J. Leusen, and R. J. Gdanitz, *J. Comput. Aided Mater. Des.* **1**, 177 (1993).
- [40] M. U. Schmidt and U. Englert, *J. Chem. Soc. Dalton Trans.* , 2077 (1996).
- [41] J. Pillardy, Y. A. Arnautova, C. Czaplewski, K. D. Gibson, and H. A. Scheraga, *Proc. Nat. Acad. Sci.* **98**, 12351 (2001).

- [42] S. L. Price, D. J. Willock, M. Leslie, and G. M. Day, <http://www.ucl.ac.uk/uc-ca17p/dmarelmanual/dmarel.html> (2001).
- [43] D. J. Willock, S. L. Price, M. Leslie, and C. R. A. Catlow, *J. Comp. Chem.* **16**, 628 (1995).
- [44] B. P. van Eijck, W. T. M. Mooij, and J. Kroon, *Acta Crystallogr. B* **51**, 99 (1995).
- [45] D. W. M. Hofmann and T. Lengauer, *Acta Crystallogr. A* **53**, 225 (1997).
- [46] V. E. Bazterra, Modelado de cristales orgánicos por medio del algoritmo genético modificado, Tesis de Licenciatura del Dpto. Física, Universidad de Buenos Aires, 2001.
- [47] V. E. Bazterra, M. B. Ferraro, and J. C. Facelli, *J. Chem. Phys.* **116**, 5984 (2004).
- [48] V. E. Bazterra, M. B. Ferraro, and J. C. Facelli, *J. Chem. Phys.* **116**, 5992 (2004).
- [49] V. E. Bazterra, M. B. Ferraro, and J. C. Facelli, *Int. J. Quant. Chem.* **96**, 312 (2004).
- [50] D. E. Williams, *MPA/MPG, Molecular Packing Analysis and Molecular Packing Graphics* (Chemistry Department, University of Louisville, Louisville, KY 40292, USA, 1996).
- [51] Accelrys Inc., San Diego, USA, *MS Modelling release 3.0.1*, 2004.
- [52] A. V. Dzyabchenko, *PMC, version 2004* (Karpov Institute of Physical Chemistry, Moscow, 2004).

- [53] B. P. van Eijck and J. Kroon, *Acta Crystallogr. B* **56**, 535 (2000).
- [54] W. D. S. Motherwell, *Mol. Cryst. Liq. Cryst.* **356**, 559 (2001).
- [55] A. Gavezzotti, *Zip-Promet, A program for the Generation of Crystal Structures from Molecular Structure* (University of Milano, Milan, 2000).
- [56] A. Gavezzotti, *J. Am. Chem. Soc.* **113**, 4622 (1991).
- [57] The cambridge structural database, <http://www.ccdc.cam.ac.uk/products/csd/>.
- [58] J. T. Ngo, J. Marks, and M. Karplus, *The Protein Folding Problem and Tertiary Structure Prediction* (Birkhäuser, USA, 1994), chap. Computational Complexity, Protein Structure Prediction and the Levinthal Paradox, p. 433.
- [59] G. Burns and A. M. Glazer, *Space Groups for Solid State Scientists* (Academic Press, Londres, 1978).
- [60] D. Gao and D. E. Williams, *Acta Crystallogr. A* **55**, 621 (1999).
- [61] A. Gavezzotti and G. Filippini, *J. Am. Chem. Soc.* **117**, 12299 (1995).
- [62] B. Hartke, *J. Phys. Chem.* **97**, 9973 (1993).
- [63] Y. Xiao and D. E. Williams, *Chem. Phys. Lett.* **215**, 17 (1993).
- [64] B. Hartke, *Chem. Phys. Lett.* **240**, 560 (1995).
- [65] B. Hartke, *Chem. Phys. Lett.* **258**, 144 (1996).

- [66] S. K. Gregurick, M. H. Alexander, and B. Hartke, *J. Chem. Phys.* **104**, 2684 (1996).
- [67] B. Hartke, M. Schütz, and H. J. Werner, *Chem. Phys.* **239**, 561 (1998).
- [68] B. Hartke, *Z. Phys. Chem.* **214**, 1251 (2000).
- [69] B. Hartke, *J. Comput. Chem.* **20**, 1752 (1999).
- [70] Y. Zeiri, *Phys. Rev. E* **51**, 1995 (1999).
- [71] Y. Zeiri, *Comput. Phys. Commun.* **103**, 28 (1997).
- [72] Y. Zeiri, *J. Phys. Chem. A* **102**, 2785 (1998).
- [73] D. M. Deaven and K. M. Ho, *Phys. Rev. Lett.* **75**, 288 (1995).
- [74] D. M. Deaven, N. Tit, J. R. Morris, and K. M. Ho, *Chem. Phys. Lett* **256**, 195 (1996).
- [75] J. A. Niesse and H. R. Mayne, *J. Comput. Chem.* **18**, 1233 (1996).
- [76] R. P. White, J. A. Niesse, and H. R. Mayne, *J. Chem. Phys.* **108**, 2208 (1998).
- [77] S. M. Woodley, P. D. Battle, J. D. Galec, and C. R. A. Catlowa, *Phys. Chem. Chem. Phys.* **1**, 2535 (1999).
- [78] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 178 (1983).
- [79] V. E. Bazterra, M. Cuma, M. B. Ferraro, and J. C. Facelli, *J. Parallel Distrib. Comput.* **65**, 48 (2005).

- [80] Extensible markup language, <http://www.w3.org/XML/>.
- [81] Xml schema, <http://www.w3.org/XML/Schema>.
- [82] Chemical markup language, <http://www.xml-cml.org/>.
- [83] Gnu bourne again shell, <http://www.gnu.org/software/bash/>.
- [84] Implementation of the message-passing interface, <http://www-unix.mcs.anl.gov/mpi/mpich/>.
- [85] Galib: A c++ library of genetic algorithm components, <http://lancet.mit.edu/ga/>.
- [86] Xerces c++ parser, <http://xml.apache.org/xerces-c/>.
- [87] Dingo project, <http://www.chpc.utah.edu/~baites/dingo>.
- [88] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *J. Comput. Chem.* **25**, 1157 (2004).
- [89] S. L. Mayo, B. D. Olafson, and W. A. G. III, *J. Phys. Chem.* **94**, 8897 (1990).
- [90] A. K. Rappé, C. J. Casewit, K. Colwell, W. A. Goddard, and W. M. Skiff, *J. Am. Chem. Phys.* **114**, 10024 (1992).
- [91] Antechamber & gaff, <http://amber.scripps.edu/antechamber/antechamber.html>.
- [92] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, *J. Chem. Phys.* **97**, 10269 (1993).

- [93] M. J. Frisch *et al.*, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.
- [94] J. J. P. Stewart, *Mopac Manual* (Frank J. Seiler Research Laboratory, United State Air Force Academy CO 80840, 1990).
- [95] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).
- [96] C. G. Broyden, *J. Inst. Math. Appl.* **6**, 75 (1970).
- [97] R. Fletcher, *Comput. J.* **13**, 317 (1970).
- [98] D. Goldfarb, *Math. Comput.* **24**, 23 (1970).
- [99] D. F. Shanno, *Math. Comput.* **24**, 647 (1970).
- [100] J. Nocedal, *Math. Comput.* **35**, 773 (1980).
- [101] D. Liu and J. Nocedal, *Math. Programming* **45**, 503 (1989).
- [102] W. H. Press and *et. al.*, *Numerical Recipes in C, second edition* (Cambridge University Press, Cambridge, 1992).
- [103] R. Fletcher, *Parctical Methods of Optimization: Volume 1 - Unconstrained Optimization* (Wiley, New York, 1980).
- [104] P. Pulay, *Chem. Phys. Lett.* **73**, 393 (1980).
- [105] P. Pulay, *J. Comput. Chem.* **3**, 556 (1982).

- [106] P. Császár and P. Pulay, *J. Mol. Struct. (Theochem)* **114**, 31 (1984).
- [107] O. Farkas and H. B. Schlegel, *Phys. Chem. Chem. Phys.* **4**, 11 (2002).
- [108] J. W. Chu, B. L. Trout, and B. R. Brooks, *J. Chem. Phys.* **119**, 12708 (2003).
- [109] P. G. Karamertzanis and S. L. Price, *J. Chem. Theory Comput.* **2**, 1184 (2006).
- [110] J. A. Chisholm and S. Motherwell, *J. Appl. Cryst.* **38**, 228 (2005).
- [111] W. Kabsch, *Acta Crystallogr. A* **32**, 922 (1976).
- [112] W. Kabsch, *Acta Crystallogr. A* **34**, 827 (1978).
- [113] Computational crystallography toolbox, <http://cctbx.sourceforge.net/>.
- [114] R. W. Grosse-Kunstleve and P. D. Adams, State of the toolbox: an overview of the computational crystallography toolbox (cctbx), http://cci.lbl.gov/publications/download/iucrcompcomm_jan2003.pdf.
- [115] The python programming language, <http://www.python.org/>.
- [116] R. T. Fielding, Architectural styles and the design of network-based software architectures, <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>.
- [117] Http - hypertext transfer protocol, <http://www.w3.org/Protocols/>.
- [118] Django: The web framework for perfectionists with deadline, <http://www.djangoproject.com/>.

- [119] Mysql: The world's most popular open source database, <http://www.mysql.com/>.
- [120] The iupac international chemical identifier, <http://www.iupac.org/inchi/>.
- [121] The extensible hypertext markup language, <http://www.w3.org/TR/xhtml1/>.
- [122] R. Destro, R. E. March, and R. Bianchi, *J. Phys. Chem.* **92**, 966 (1988).
- [123] M. S. Nandhini, R. V. Krishnakumar, and S. Natarajan, *Acta Crystallogr. C* **57**, 614 (2001).
- [124] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.* **117**, 5179 (1995).
- [125] A. D. J. MacKerell *Encyclopedia of Computational Chemistry* Vol. 1 (John Wiley and Sons, Chichester, 1998), chap. Protein Force Fields, p. 271.
- [126] A. L. Spek, *A multipurpose crystallographic tool*, Utrecht University, Utrecht, The Netherlands.
- [127] Arches opteron cluster, <http://www.chpc.utah.edu/docs/manuals/hardware/>.
- [128] Ncsa teragrid ia-64 linux cluster, <http://www.ncsa.uiuc.edu/UserInfo/Resources-/Hardware/TGIA64LinuxCluster/>.
- [129] R. S. Payne, R. C. Rowe, R. J. Roberts, M. H. Charlton, and R. Docherty, *J. Comput. Chem.* **20**, 262 (1999).

-
- [130] K.-M. Ho, A. A. Shvartsburg, B. Pan, Z.-Y. Lu, C.-Z. Wang, J. G. Wacker, J. L. Fye, and M. F. Jarrold, *Nature* **352**, 695 (1998).
- [131] S. Yoo, N. Shao, C. Koehler, T. Fraunhaum, and X. C. Zeng, *J. Chem. Phys.* **124**, 164311 (2006).
- [132] O. Oña, V. E. Bazterra, M. C. Caputo, J. C. Facelli, P. Fuentealba, and M. B. Ferraro, *Phys. Rev. A.* **72**, 053205 (2005).
- [133] O. Oña, V. E. Bazterra, M. C. Caputo, J. C. Facelli, P. Fuentealba, and M. B. Ferraro, *Phys. Rev. A.* **73**, 053203 (2006).
- [134] O. Oña, V. E. Bazterra, M. C. Caputo, M. B. Ferraro, and J. C. Facelli, *Mater. Res. Soc. Symp. Proc.* **894**, 277 (2006).
- [135] G. Igell-Mann, H. Stoll, and H. Presuss, *Mol. Phys.* **65**, 1321 (1988).
- [136] A. J. Sadlej, *Collect. Czech. Chem. Commun.* **53**, 1995 (1998).
- [137] P. J. Hay and W. R. Wadt, *Nature* **345**, 297 (1990).
- [138] P. J. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **2**, 816 (1980).

Índice alfabético

- Algoritmo de Kabsch, 46
- Algoritmos Genéticos, 1, 15
 - generación, 16
 - genoma, 15
 - población, 15
- Algoritmos Genéticos Modificados, 17
- APGA, 32
- clusters atómicos, 1
 - búsqueda por patrones, 7
- cristales moleculares, 1
 - búsqueda con simetrías, 14
 - búsqueda sin simetrías, 14
 - celda asimétrica, 24
 - celda unidad, 23
 - genoma, 25
 - grupos espaciales, 24, 47
 - moléculas flexibles, 5, 11
 - moléculas rígidas, 11
- estándares computacionales
 - CML, 32, 52
 - esquema para XML, 32
 - InChi, 52
 - MPI, 33
 - REST, 51
 - XHTML, 53
 - XML, 32
- fitcell, 26
- Ingeniería de Cristales, 1
- lenguajes de programación
 - BASH, 32
 - C++, 32, 51
 - Python, 51
- módulo

- GALib, 33
- módulos
 - CCTBX, 51
 - MOLCRY, 51
 - Django, 51
 - MPICH, 33
 - xerces-c, 33
- MGAC, 23
- mgacML, 32
- modelos rígidos o flexibles, 11
- nanopartículas, 2
- operadores genéticos, 16, 29
 - de mutación, 29
 - de mutación gaussiana, 30
 - de recombinación, 29
 - a dos puntos, 30, 103
 - a N puntos, 30, 104
 - a un punto, 30, 103
 - inversión, 30, 104
 - media aritmética, 30, 104
 - media geométrica, 30, 104
 - uniforme, 30, 104
 - físicos o fenotípicos, 17
 - cut and splice, 17
 - genéricos o genotípicos, 17
 - optimización global, 11
 - optimización local, 37
 - ABNR, 39
 - GC, 38
 - GIIDS, 39
 - L-BFGS, 38
 - zona de atracción, 38
- Polimorfismo, 2
- potencial clásico, 4
 - AMBER, 55
 - CHARMm, 55
 - DREIDING, 35
 - GAFF, 34
 - UFF, 35
- programas
 - antechamber, 36
 - CHARMM, 37

charmmgen, 37

Gaussian, 36

MOPAC6, 36

PLATON, 60

RESP, 36

RMS, 42, 46

 a 16 moléculas, 42

 mRMS, 42

servidores

 MySQL, 52