

**Centro de Investigación Científica y de Educación  
Superior de Ensenada, Baja California**



---

**Maestría en Ciencias  
en Ciencias de la Computación**

---

**Análisis del desempeño de herramientas de  
acoplamiento molecular para el estudio de  
interacciones proteína-péptido**

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de  
Maestro en Ciencias

Presenta:

**Osvel Chávez Hernández**

Ensenada, Baja California, México

2019

Tesis defendida por

**Osvel Chávez Hernández**

y aprobada por el siguiente Comité

---

Dr. Carlos Alberto Brizuela Rodríguez

Codirector de tesis

---

Dra. Rosario Ivetth Corona de la Fuente

Codirector de tesis

Dr. Israel Marck Martínez Pérez

Dr. Sergio Andrés Águila Puentes



---

Dr. Ubaldo Ruiz López

Coordinador del Posgrado en Ciencias de la Computación

---

Dra. Rufina Hernández Martínez

Director de Estudios de Posgrado

*Osvel Chávez Hernández © 2019*

*Queda prohibida la reproducción parcial o total de esta obra sin el permiso formal y explícito del autor y director de la tesis*

Resumen de la tesis que presenta Osvel Chávez Hernández como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la Computación .

## **Análisis del desempeño de herramientas de acoplamiento molecular para el estudio de interacciones proteína-péptido**

Resumen aprobado por:

---

Dr. Carlos Alberto Brizuela Rodríguez

Codirector de tesis

---

Dra. Rosario Ivetth Corona de la Fuente

Codirector de tesis

El proceso de identificación de un fármaco eficaz para el tratamiento de una determinada enfermedad tiene un alto costo en tiempo y recursos. El principal objetivo en el diseño de fármacos es encontrar aquellas moléculas que puedan interactuar de manera efectiva en un determinado blanco terapéutico. Muchos de los tratamientos para las enfermedades actuales tienen como blancos terapéuticos a las proteínas que se reconocen en el proceso de la enfermedad, por lo que es necesario inhibir o modificar el funcionamiento de dichas moléculas. Los fármacos más eficientes son los que su unión en el sitio activo de su proteína objetivo, genera la pose más estable. En las últimas décadas los péptidos han sido reconocidos como una nueva clase de fármacos potenciales. Son moléculas altamente selectivas, optimizadas por millones de años de selección natural por lo que, en general, solo interactuarían con su proteína objetivo. En la actualidad contamos con procedimientos computacionales que exploran la acción de un determinado fármaco en su blanco terapéutico. Estos procedimientos computacionales son capaces de evaluar varias poses, de más de un candidato a fármaco, en la cavidad activa de la proteína blanco. Para encontrar las poses moleculares que contribuyan a la obtención de un complejo molecular estable se emplean los llamados métodos de Acoplamiento Molecular. En el presente trabajo de investigación se comparó el desempeño de cuatro herramientas de acoplamiento molecular en cuanto a su capacidad de predicción de las poses proteína-péptido. Se evaluaron las herramientas sobre 249 complejos proteína-péptido obteniendo hasta 20 poses por complejo, por herramienta. Como medida de comparación se utilizó la desviación cuadrática media entre los átomos del péptido de la pose en el cristal y los átomos del péptido en las poses obtenidas por las herramientas de acoplamiento molecular. Los resultados experimentales mostraron que el AutoDock Vina logra, en su pose de menor energía, un 96 % de casos con un error menor a 2Å. Esto se da cuando, tanto la proteína como el péptido tienen una configuración rígida. Sin embargo, este resultado cae al 17 % cuando se considera al péptido como una estructura flexible. Además, el cálculo de descriptores moleculares sencillos para las poses, no permitió a un conjunto importante de modelos de aprendizaje de máquina mejorar esta situación.

**Palabras clave: Acoplamiento Molecular, Péptido, Proteína**

Abstract of the thesis presented by Osvel Chávez Hernández as a partial requirement to obtain the Master of Science degree in Computing Sciences.

## **Performance analysis of molecular docking tools for the study of protein-peptide interactions**

Abstract approved by:

---

Dr. Carlos Alberto Brizuela Rodríguez

Thesis Co-Director

---

Dra. Rosario Ivetth Corona de la Fuente

Thesis Co-Director

The process of identifying an effective drug for the treatment of a certain disease has a high cost in time and resources. The main objective in the design of drugs is to find those molecules that can interact effectively in a certain therapeutic target. Many of the treatments for current diseases have as therapeutic targets the proteins that are recognized in the disease process, so it is necessary to inhibit or modify the function of said molecules. The most efficient drugs are those that their union to the active site of their target protein, generate the most stable pose. In recent decades, peptides have been recognized as a new class of potential drugs. They are highly selective molecules, optimized for millions of years of natural selection so they would, in general, only interact with their target protein. Currently, we have computational procedures that explore the action of a certain drug in its therapeutic target. These computational procedures are capable of evaluating the docking, of more than one drug candidate, in the active cavity of the target protein. To find the molecular docking that contribute to obtaining a stable molecular complex, the so-called Molecular Docking methods are used. In the present research work, the performance of four molecular docking tools was assessed in terms of their ability to predict protein-peptide docking. The tools were evaluated on 249 protein-peptide complexes obtaining up to 20 poses per complex, per tool. The average quadratic deviation between the peptide atoms of the pose in the crystal and the peptide atoms in the docking obtained by the molecular docking tools was used as a comparison measure. Experimental results showed that AutoDock Vina achieves, in its lowest energy pose, 96 % of cases with an error less than 2Å. This is when both, the protein and the peptide have a rigid configuration. However, this result falls to 17 % when the peptide is considered a flexible structure. In addition, the calculation of simple molecular descriptors for docking did not allow an important set of machine learning models to improve this situation.

**Keywords: Molecular Docking, Peptide, Protein**

## Dedicatoria

***A mi esposa y mi hija por todo su amor, consejos y por siempre estar a mi lado. A mis padres por mostrarme los caminos de las ciencias.***

## **Agradecimientos**

Al Centro de Investigación Científica y de Educación Superior de Ensenada por abrirme sus puertas para hacerme Maestro en Ciencias.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme el apoyo económico para realizar mis estudios de maestría. No. de becario: 638437.

A mis asesores de tesis, Dr. Carlos Alberto Brizuela Rodríguez y Dra. Rosario Ivetth Corona de la Fuente, por su guía y enseñanzas. Así como a los miembros del comité de tesis, Dr. Israel Marck Martínez Pérez y Dr. Sergio Andrés Águila Puentes por su tiempo dedicado, sus observaciones y sugerencias durante la realización de la presente investigación.

A mi esposa, Ilianet Castillo Ricardo, por su apoyo incondicional y sus fuerzas día a día; a mi hija, Camila Alvarez Castillo, con sus sonrisas.

## Tabla de contenido

|  | Página |
|--|--------|
| Resumen en español .....   | ii     |
| Resumen en inglés .....  | iii    |
| Dedicatoria .....  | iv     |
| Agradecimientos .....  | v      |
| Lista de figuras .....   | ix     |
| Lista de tablas .....  | xii    |
| <br>   |        |
| <b>Capítulo 1. Introducción</b>                                    |        |
| 1.1. Antecedentes .....  | 1      |
| 1.1.1. Péptidos como fármacos .....                                | 2      |
| 1.1.2. Acoplamiento molecular .....                                | 2      |
| 1.2. Motivación .....  | 3      |
| 1.3. Objetivos .....   | 4      |
| 1.3.1. Objetivo general .....                                      | 4      |
| 1.3.2. Objetivos específicos .....                                 | 4      |
| 1.3.3. Metodología de solución propuesta .....                     | 4      |
| 1.4. Estructura del documento de investigación .....               | 5      |
| <br>   |        |
| <b>Capítulo 2. Marco Teórico</b>                                   |        |
| 2.1. Proteínas .....   | 7      |
| 2.1.1. Aminoácidos .....   | 7      |
| 2.1.2. Cadenas de aminoácidos .....                                | 10     |
| 2.1.2.1. Estructura Primaria .....                                 | 11     |
| 2.1.2.2. Estructura Secundaria .....                               | 12     |
| 2.1.2.3. Estructura Terciaria .....                                | 13     |
| 2.1.2.4. Estructura Cuaternaria .....                              | 14     |
| 2.1.3. Interacciones con otras moléculas .....                     | 14     |
| 2.2. Acoplamiento Molecular .....                                  | 16     |
| 2.2.1. Algoritmos de búsqueda .....                                | 17     |
| 2.2.2. Funciones de puntuación .....                               | 19     |
| 2.3. Métricas para las evaluaciones .....                          | 21     |
| 2.3.1. Desviación Cuadrática Media .....                           | 22     |
| 2.3.2. Coeficiente de Correlación .....                            | 22     |
| 2.3.3. Error Absoluto Medio .....                                  | 23     |
| 2.4. Huella de interacción .....                                   | 23     |
| 2.4.1. Huella Estructural de la Interacción Proteína-ligando ..... | 23     |
| 2.5. Descriptor molecular .....                                    | 24     |
| 2.5.1. Descriptor Simple de Interacción Ligando-Receptor .....     | 24     |
| 2.6. Relación Cuantitativa Estructura-Actividad .....              | 24     |

## Tabla de contenido (continuación)

### Capítulo 3. Metodología

|        |  |    |
|--------|--|----|
| 3.1.   | Complejos químicos . . . . .                     | 26 |
| 3.1.1. | Banco de Datos de Proteínas . . . . .            | 26 |
| 3.1.2. | PeptiDB . . . . .                                | 27 |
| 3.1.3. | LEADS-PEP . . . . .                              | 27 |
| 3.1.4. | PeptideInhibitor . . . . .                       | 27 |
| 3.1.5. | DockingDB . . . . .                              | 28 |
| 3.2.   | Acoplamiento molecular . . . . .                 | 29 |
| 3.2.1. | Formas de acoplamiento molecular . . . . .       | 31 |
| 3.2.2. | AutoDock4 . . . . .                              | 31 |
| 3.2.3. | AutoDock Vina . . . . .                          | 33 |
| 3.2.4. | Dock6 . . . . .                                  | 35 |
| 3.2.5. | ClusPro . . . . .                                | 35 |
| 3.3.   | Manipulación de moléculas . . . . .              | 36 |
| 3.3.1. | AutoDock Tools 4 . . . . .                       | 37 |
| 3.3.2. | UCSF Chimera . . . . .                           | 37 |
| 3.4.   | Preparación del Acoplamiento Molecular . . . . . | 37 |
| 3.4.1. | AutoGrid4 . . . . .                              | 37 |
| 3.4.2. | Grid . . . . .                                   | 38 |
| 3.4.3. | ShowBox . . . . .                                | 38 |
| 3.4.4. | SphGen . . . . .                                 | 38 |
| 3.5.   | Descriptores moleculares . . . . .               | 38 |
| 3.5.1. | PyPlif . . . . .                                 | 38 |
| 3.5.2. | WEKA . . . . .                                   | 39 |

### Capítulo 4. Resultados

|        |  |    |
|--------|--|----|
| 4.1.   | Complejos moleculares . . . . .  | 44 |
| 4.2.   | Acoplamiento global . . . . .  | 44 |
| 4.3.   | Acoplamiento dirigido rígido . . . . .   | 47 |
| 4.4.   | Acoplamiento dirigido rígido de la DockingDB . . . . .                           | 50 |
| 4.5.   | Acoplamiento dirigido flexible de la DockingDB . . . . .                         | 53 |
| 4.6.   | Cálculo de SILIRID . . . . .   | 57 |
| 4.6.1. | Obtención de los Descriptores Moleculares . . . . .                              | 57 |
| 4.7.   | Obtención de los modelos computacionales para las poses rígidas . . . . .        | 59 |
| 4.7.1. | Poses del AutoDock4 y del Vina predicen poses del AutoDock4 y del Vina . . . . . | 59 |
| 4.7.2. | Poses del AutoDock4 predicen poses del AutoDock4 . . . . .                       | 61 |
| 4.7.3. | Poses del AutoDock Vina predicen poses del AutoDock Vina . . . . .               | 62 |
| 4.8.   | Obtención de los modelos computacionales para las poses flexibles . . . . .      | 64 |
| 4.8.1. | Poses del AutoDock4 y del Vina predicen poses del AutoDock4 y del Vina . . . . . | 64 |
| 4.8.2. | Poses del AutoDock4 predicen poses del AutoDock4 . . . . .                       | 65 |
| 4.8.3. | Poses del AutoDock Vina predicen poses del AutoDock Vina . . . . .               | 66 |
| 4.9.   | Análisis de los resultados obtenidos . . . . .                                   | 68 |



## Tabla de contenido (continuación)

### Capítulo 5. Conclusiones

|  |           |
|--|-----------|
| 5.1. Sumario . . . . .   | 73        |
| 5.2. Conclusiones . . . . .  | 73        |
| 5.3. Propuesta de trabajo futuro . . . . .                                     | 74        |
| <b>Literatura citada . . . . .</b>   | <b>76</b> |
| <b>Anexo A - Formatos para el almacenamiento de los complejos moleculares</b>  | <b>80</b> |
| <b>Anexo B - Configuración de las herramientas de acoplamiento molecular .</b> | <b>83</b> |
| <b>Anexo C - Comandos de ejecución . . . . .</b>                               | <b>89</b> |
| <b>Anexo D - Errores en tiempo de ejecución . . . . .</b>                      | <b>91</b> |
| <b>Anexo E - Tablas . . . . .</b>  | <b>94</b> |

## Lista de figuras

| Figura   | Página |
|--|--------|
| 1. Metodología general propuesta. . . . .  | 5      |
| 2. Estructura general de los aminoácidos naturales, excepto para la <i>Prolina</i> . Imagen tomada de (Nelson <i>et al.</i> , 2008). . . . .   | 8      |
| 3. Los cinco grupos en los que se encuentran clasificados los aminoácidos naturales. Las fórmulas estructurales en su estado de ionización predominante a pH 7. Las zonas sombreadas, son los grupos R de cada aminoácido. Imagen tomada de (Nelson <i>et al.</i> , 2008). . . . . | 9      |
| 4. Estereoisomería en aminoácidos, son imágenes especulares no superponibles entre sí (enantiómeros). Los dos estereoisómeros de la alanina, L y D-alanina. Imagen tomada de (Nelson <i>et al.</i> , 2008). . . . .  | 9      |
| 5. Formación de un enlace peptídico por condensación. El grupo amino de un amino ácido actúa como un nucleófilo para desplazar el grupo hidroxilo de otro aminoácido, formando un enlace peptídico. Imagen tomada de (Nelson <i>et al.</i> , 2008). . . . .                        | 10     |
| 6. Niveles estructurales en proteínas. La imagen mostrada corresponde a la hemoglobina. Imagen tomada de (Nelson <i>et al.</i> , 2008). . . . .  | 11     |
| 7. Cuatro modelos de la hélice alfa, mostrando diferentes aspectos de su estructura. Imagen tomada de (Nelson <i>et al.</i> , 2008). . . . .   | 13     |
| 8. La conformación beta de las cadenas polipeptídicas. Imagen tomada de (Nelson <i>et al.</i> , 2008). . . . .   | 14     |
| 9. Unión de PeptiDB, LEADS-PEP y PeptideInhibitor en DockingDB. . . . .  | 28     |
| 10. Representación de la caja para el acoplamiento molecular global del complejo 1apm. . . . .   | 45     |
| 11. Unión de los conjuntos del acoplamiento molecular Global que representan las poses de $RMSD \leq 2\text{\AA}$ por cada herramienta. . . . .  | 46     |
| 12. Relación entre el RMSD y la energía para el acoplamiento molecular global rígido. . . . .  | 47     |
| 13. RMSD del acoplamiento molecular global con estructuras rígidas. . . . .  | 48     |
| 14. Representación de la caja para el acoplamiento molecular local del complejo 1apm. . . . .  | 49     |
| 15. Unión de los conjuntos del acoplamiento molecular local que representan las poses de $RMSD \leq 2\text{\AA}$ por cada herramienta. . . . .   | 49     |
| 16. Relación entre el RMSD y la energía para el acoplamiento molecular dirigido rígido. . . . .  | 50     |
| 17. RMSD del acoplamiento molecular local con 10 estructuras rígidas. . . . .  | 51     |
| 18. Unión de los conjuntos del acoplamiento molecular local que representan las poses de $RMSD \leq 2\text{\AA}$ por cada herramienta en la DokingDB. . . . .  | 52     |

## Lista de figuras (continuación)

| Figura  | Página |
|---|--------|
| 19. Relación entre el RMSD y la energía para el acoplamiento molecular dirigido rígido de una muestra de la DockingDB. . . . .                                      | 53     |
| 20. RMSD del acoplamiento molecular dirigido rígido de una muestra de la DockingDB. . . . .   | 54     |
| 21. Unión de los conjuntos del acoplamiento molecular flexible local que representan las poses de $RMSD \leq 2\text{\AA}$ por cada herramienta en la DockingDB.     | 55     |
| 22. Relación entre el RMSD y la energía para el acoplamiento molecular dirigido rígido de una muestra de la DockingDB. . . . .                                      | 55     |
| 23. RMSD del acoplamiento molecular dirigido flexible de una muestra de los complejos con menos de 33 rotámeros de la DockingDB. . . . .                            | 56     |
| 24. Coeficiente de correlación de los algoritmos de predicción con las poses del AutoDock4 y del Vina prediciendo poses del AutoDock4 y del Vina. . . .             | 60     |
| 25. Error absoluto medio de los algoritmos de predicción con las poses del AutoDock4 y del Vina prediciendo poses del AutoDock4 y del Vina. . . . .                 | 61     |
| 26. Coeficiente de correlación de los algoritmos de predicción con las poses del AutoDock4 prediciendo poses del AutoDock4. . . . .                                 | 62     |
| 27. Error absoluto medio de los algoritmos de predicción con las poses del AutoDock4 prediciendo poses del AutoDock4. . . . .                                       | 63     |
| 28. Coeficiente de correlación de los algoritmos de predicción con las poses del AutoDock Vina prediciendo poses del AutoDock Vina. . . . .                         | 64     |
| 29. Error absoluto medio de los algoritmos de predicción con las poses del AutoDock Vina prediciendo poses del AutoDock Vina. . . . .                               | 65     |
| 30. Coeficiente de correlación de los algoritmos de predicción con las poses flexibles del AutoDock4 y del Vina prediciendo poses del AutoDock4 y del Vina. . . . . | 66     |
| 31. Error absoluto medio de los algoritmos de predicción con las poses flexibles del AutoDock4 y del Vina prediciendo poses del AutoDock4 y del Vina. . . . .       | 67     |
| 32. Coeficiente de correlación de los algoritmos de predicción con las poses flexibles del AutoDock4 prediciendo poses del AutoDock4. . . . .                       | 68     |
| 33. Error absoluto medio de los algoritmos de predicción con las poses flexibles del AutoDock4 prediciendo poses del AutoDock4. . . . .                             | 69     |
| 34. Coeficiente de correlación de los algoritmos de predicción con las poses flexibles del AutoDock Vina prediciendo poses del AutoDock Vina. . . . .               | 70     |

## Lista de figuras (continuación)

| Figura   | Página |
|--|--------|
| 35. Error absoluto medio de los algoritmos de predicción con las poses flexibles del AutoDock Vina prediciendo poses del AutoDock Vina. . . . .  | 71     |
| 36. Cociente de predicción de los algoritmos de predicción para: Poses del AutoDock4 y del Vina predicen poses del AutoDock4 y del Vina (Todas), Poses del AutoDock4 predicen poses del AutoDock4 (Solo AutoDock4) y Poses del Vina predicen poses del Vina (Solo Vina); para las poses rígidas.   | 72     |
| 37. Cociente de predicción de los algoritmos de predicción para: Poses del AutoDock4 y del Vina predicen poses del AutoDock4 y del Vina (Todas), Poses del AutoDock4 predicen poses del AutoDock4 (Solo AutoDock4) y Poses del Vina predicen poses del Vina (Solo Vina); para las poses flexibles. | 72     |

## Lista de tablas

| Tabla | Página   |
|-------|--|
| 1.    | Descripción general de 10 de las estructuras incluidas en el conjunto de datos de referencia DockingDB. . . . . 44                     |
| 2.    | Descripción general de las estructuras incluidas en el conjunto de datos de referencia DockingDB. . . . . 94                           |
| 3.    | Correlación de la pose de menor energía y su RMSD para el acoplamiento molecular rígido global. . . . . 99                             |
| 4.    | Correlación de la pose de menor energía y su RMSD para el acoplamiento molecular rígido local. . . . . 99                              |
| 5.    | Correlación de la pose de menor energía y su RMSD para el acoplamiento molecular rígido local con toda la DockingDB. . . . . 100       |
| 6.    | Correlación de la pose de menor energía y su RMSD para el acoplamiento molecular rígido local con los complejos flexibles. . . . . 104 |
| 7.    | Resultados de la predicción del RMSD de las poses para el acoplamiento molecular dirigido rígido. . . . . 107                          |
| 8.    | Resultados de la predicción del RMSD de las poses para el acoplamiento molecular dirigido flexible. . . . . 107                        |

# Capítulo 1. Introducción

---

## 1.1. Antecedentes

El descubrimiento y desarrollo de nuevos fármacos es un proceso altamente demandante de recursos, los más costosos están dados en términos de tiempo y dinero. En la línea de desarrollo, desde la identificación del nuevo compuesto hasta su aprobación, dura en promedio 17 años y cuesta alrededor de 800 millones de dólares (Cerqueira *et al.*, 2015).

Una de las alternativas para reducir la elevada demanda de recursos ha sido el empleo de herramientas computacionales, dando lugar a lo que actualmente se conoce como Diseño de Fármacos Asistido por Computadora (Van Drie, 2007). Esta tecnología consiste en la aplicación de algoritmos y modelos computacionales mediante el uso de computadoras potentes, que permiten describir las interacciones entre una macromolécula y otra un tanto más pequeña (Murgueitio *et al.*, 2012), de esta forma se simula y evalúan las características fisicoquímicas, farmacológicas y farmacocinéticas de este último para determinar si puede ser utilizado o no en el área clínica (Jorgensen, 2004).

La vida tal y como la conocemos está estrechamente relacionada con las formas y la composición química que presentan las estructuras macromoleculares como son las proteínas. Las proteínas de origen natural tienen variadas funciones: composición de la estructura celular, transformación de energía química en trabajo celular, reconocimiento de otras estructuras moleculares, entre otras. Estas funciones son notables no solo por su función en el mantenimiento de la vida, sino también porque están codificadas en secuencias de aminoácidos con la mayor economía de energía libre posible (Huang *et al.*, 2016). Puesto que las proteínas son los componentes macro moleculares más abundantes en los seres vivos, no resulta sorprendente que éstas sean las dianas más frecuentes de los fármacos. Inhibiendo el funcionamiento característico de las proteínas, es posible atacar estructuras más complejas como virus, bacterias, hongos, parásitos o células tumorales, que estén compuestas por esas proteínas.

### **1.1.1. Péptidos como fármacos**

La mayor parte del desarrollo de los fármacos en el siglo XX se puede clasificar como una era química, donde los fármacos eran generados a partir de productos naturales de pequeñas moléculas, por selección o por diseño molecular, mecanismos o receptores. Sin embargo, en la última parte del siglo XX comenzaron a surgir nuevas clases de terapias, basadas en proteínas de hasta 50 aminoácidos (péptidos), muy diferentes a los medicamentos con moléculas pequeñas, logrando menos efectos secundarios. Los péptidos han logrado un alto grado de selectividad, producto de millones de años de selección evolutiva, en sus interacciones, por lo que se han ajustado para interactuar con objetivos biológicos específicos. Esta especificidad y selectividad de los péptidos han llevado a que hoy en día existan más de 100 medicamentos basados en péptidos (Craik *et al.*, 2013).

### **1.1.2. Acoplamiento molecular**

Los algoritmos de Acoplamiento Molecular exploran las conformaciones de moléculas pequeñas llamados ligandos en el sitio de unión de una proteína objetivo. Con las estrategias de acoplamiento molecular, se puede calcular la especificidad de los compuestos en un objetivo particular para futuros procesos de optimización en el desarrollo de nuevos fármacos. Los métodos computacionales de acoplamiento molecular emplean un algoritmo de búsqueda en el que la conformación del ligando se evalúa recursivamente hasta que se alcanza la convergencia a la energía mínima. Por último, se emplea una función de puntuación como la suma de las energías electrostática y de Van der Waals, para calcular la afinidad del ligando en el sitio activo de la proteína objetivo. Las fuerzas motrices para estas interacciones específicas en los sistemas biológicos apuntan hacia la complementariedad entre la forma geométrica y la electrostática de las superficies del sitio de unión entre el ligando y la proteína objetivo (Pagadala *et al.*, 2017).

No siempre es sencillo encontrar el candidato a fármaco preciso para un sitio activo dado, depende del caso, del método de acoplamiento molecular utilizado y de las medidas de éxito que no son fáciles de establecer (Ciemny *et al.*, 2018). La precisión de los métodos de acoplamiento molecular con un enfoque de cuerpo rígido

es mayor para las moléculas menos complejas, compuestas por unos pocos átomos, por lo que son usados generalmente para hacer acoplamiento molecular proteína-pequeño-ligando; esto se debe a los cambios poco observables entre las formas libres y unidas del pequeño ligando. Por otro lado, la precisión de los métodos de acoplamiento molecular con un enfoque flexible es mucho menor al analizar moléculas más complejas, compuestas por cientos de átomos, necesarias para realizar acoplamiento molecular proteína-péptido; debido a que los diferentes estados conformacionales que puede adoptar un péptido afecta la afinidad al sitio de unión. Por lo que en las últimas décadas, se han desarrollado diversas herramientas y programas de acoplamiento molecular diferentes para uso académico y comercial (Pagadala *et al.*, 2017).

Según Pagadala *et al.* (2017) y Ciemny *et al.* (2018), existen una variada gama de herramientas computacionales que implementan algoritmos de acoplamiento molecular. Un par de preguntas se plantean de forma natural luego de revisar estos trabajos: ¿cuáles de estas herramientas son más eficientes en el acoplamiento molecular proteína-péptido flexible? Una vez obtenida una pose proteína-péptido, ¿será posible predecir, qué tan alejada está dicha pose de la que ocurre realmente?

Estas preguntas dan origen a la motivación del presente trabajo de investigación.

## **1.2. Motivación**

La demanda de nuevos fármacos hace que se busquen nuevas aplicaciones para los compuestos conocidos. Las ventajas que ofrece la utilización de péptidos que inhiban o modifiquen el funcionamiento de las proteínas para combatir enfermedades conocidas, hacen que se mantenga un alto interés en ellos por parte de la industria farmacéutica. Uno de los problemas encontrados en el acoplamiento molecular, es que aún no se tiene un algoritmo capaz de predecir con exactitud la unión proteína-péptido (Ciemny *et al.*, 2018). Por lo que proponemos realizar un estudio comparativo entre cuatro herramientas de acoplamiento molecular: AutoDock4, AutoDock Vina, Dock6 y ClusPro.



### **1.3. Objetivos**

#### **1.3.1. Objetivo general**

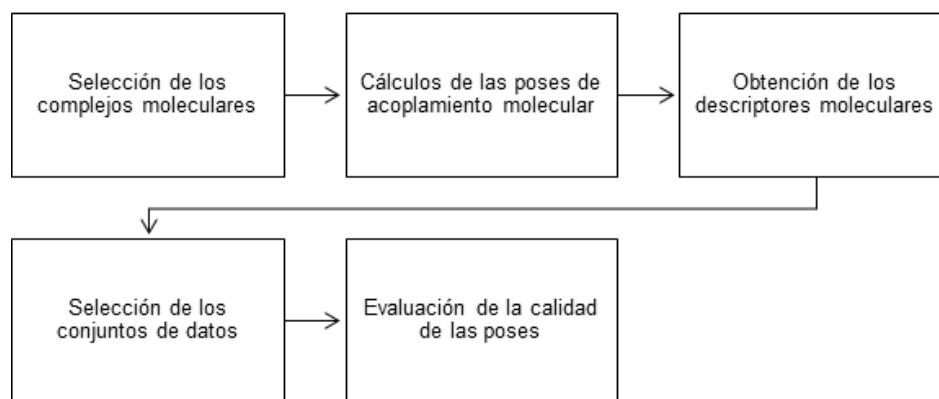
Comparar la calidad de la predicción de las poses de acoplamiento molecular de cuatro herramientas computacionales, a partir de características obtenidas de las poses proteína-péptido.

#### **1.3.2. Objetivos específicos**

1. Analizar las distintas funciones de puntuación de acoplamiento molecular en cuanto a minimización de energía de acoplamiento.
2. Analizar las distintas funciones de puntuación en cuanto a la predicción del acoplamiento molecular.
3. Evaluar diferentes descriptores moleculares para predecir la calidad de las poses obtenidas de las herramientas computacionales de acoplamiento molecular.
4. Evaluar el poder de predicción de las poses contra las predicciones obtenidas por las herramientas computacionales de acoplamiento molecular.

#### **1.3.3. Metodología de solución propuesta**

Para abordar este problema se siguió la metodología que se muestra en la Figura 1. Primero se seleccionó los complejos sobre los cuales se realizarían los cálculos de acoplamiento molecular. Una vez obtenidos los complejos se obtuvieron las poses para el péptido rígido y el péptido flexible, arrojadas por cada una de las herramientas computacionales. Luego se obtuvieron los descriptores moleculares de cada una de las poses. Ya teniendo los descriptores moleculares, se seleccionaron los conjuntos de entrenamiento y de prueba. Por último, se evaluó la calidad de las predicciones a partir de la regresión sobre el RMSD de las poses, para ello fue usado la herramienta Weka.



**Figura 1.** Metodología general propuesta.

#### 1.4. Estructura del documento de investigación

El documento de investigación está compuesto por cinco capítulos y cinco anexos, los cuales se describen a continuación.

**Capítulo 1 Introducción:** Presenta una descripción de la problemática que se abordará en la presente investigación. Se propone el objetivo general de la investigación y los objetivos específicos que darán cumplimiento al objetivo general. Se presenta la metodología para alcanzar la solución propuesta.

**Capítulo 2 Marco Teórico:** Se exponen conceptos biológicos básicos tales como proteínas y péptidos, su composición y estructuras. Por otra parte, se abordan los conceptos computacionales básicos para la comprensión del problema tratado en este trabajo como son los métodos de acoplamiento molecular, las métricas de evaluación de las poses y los descriptores moleculares.

**Capítulo 3 Metodología:** Se describe la metodología seguida para obtener los complejos moleculares de prueba, su preparación para el acoplamiento molecular, el cálculo de las poses y los descriptores moleculares de las poses.

**Capítulo 4 Resultados:** Se presentan los experimentos y resultados obtenidos, así como una comparación de los resultados para la evaluación de las herramientas de acoplamiento molecular.

**Capítulo 5 Conclusiones:** Se exponen las conclusiones a las que se llegó con el desarrollo de la presente investigación, así como algunas propuestas para la continuación del presente trabajo de investigación.

**Anexo A Formatos:** Se muestran los diferentes formatos utilizados para el almacenamiento de los complejos moleculares.

**Anexo B Configuraciones:** Se muestran los diferentes formatos de configuración de las herramientas de cálculos de las poses de acoplamiento molecular y de obtención de los descriptores moleculares.

**Anexo C Comandos:** Se muestran los comandos para la ejecución de las herramientas de cálculos de las poses de acoplamiento molecular y de obtención de los descriptores moleculares.

**Anexo D Errores:** Se muestran los errores arrojados por las herramientas de cálculos de las poses de acoplamiento molecular y de obtención de los descriptores moleculares mientras dichas herramientas realizaban sus cálculos.

**Anexo E Tablas:** Se muestran las tablas con la base de datos de complejos utilizados, los resultados de las herramientas de acoplamiento molecular y los resultados de los algoritmos de predicción.

## Capítulo 2. Marco Teórico

---

En el presente capítulo se presentarán los enunciados teóricos tales como proteínas y péptidos, los conceptos computacionales básicos, los métodos de acoplamiento molecular, las métricas de evaluación de las poses y los descriptores moleculares.

### 2.1. Proteínas

Son los complejos macro moleculares más comunes en la vida, son creadas dentro de las células y forman parte activa en todas las funciones de los seres vivos. Son producidas en gran variedad de formas y tamaño (Nelson *et al.*, 2008). Una proteína es un compuesto polimérico con una secuencia de compuestos orgánicos llamados aminoácidos. En general, la secuencia de aminoácidos determina la forma tridimensional que tendrá la proteína. Esta conformación de la proteína, determina su actividad química y por consiguiente su actividad biológica (Allouche *et al.*, 2014). Los seres vivos producen diferentes proteínas como: enzimas, hormonas, anticuerpos, transportadoras, fibras musculares, parte estructural de la lente del ojo, plumas, telarañas, cuernos, antibióticos, venenos de hongos, plantas y animales, entre muchos otros ejemplos. Las enzimas son las proteínas más especializadas y variadas, prácticamente están presente en todas las reacciones celulares cuya función es la de catalizar las sustancias intracelulares (Nelson *et al.*, 2008).

#### 2.1.1. Aminoácidos

Las proteínas son polímeros formados por aminoácidos, con cada residuo<sup>1</sup> de aminoácido unido a su vecino por un enlace covalente. Los aminoácidos, son estructuras moleculares compuestas por un grupo carboxilo ( $COO^-$ ), un grupo amino ( $H_3N^+$ ) y una cadena lateral ( $R$ ), unidos a un mismo átomo de Carbono ( $C$ ) o  $C\alpha$  (ver Figura 2). La única diferencia entre los aminoácidos radica en las cadenas laterales, que varían en estructura, tamaño y carga eléctrica; lo cual influye en la solubilidad del aminoácido. En correspondencia a las propiedades químicas de las cadenas laterales de

---

<sup>1</sup>El término residuo se refiere a la obtención de una molécula de agua al unir dos aminoácidos y al gasto de una molécula de agua para separar dos aminoácidos.

los aminoácidos, a un pH 7<sup>2</sup>, estos se pueden clasificar en cinco grupos: No polar, Polar, Aromáticos, Cargados positivos y Cargados negativos (ver Figura 3). Existen 20 aminoácidos encontrados comúnmente de forma natural, aunque existen otros no tan comunes; algunos de ellos son elementos de desecho luego de que se haya sintetizado la proteína y otros están presentes en los organismos vivos, pero no forman parte de las proteínas. A los 20 aminoácidos comunes se les han asignado para su estudio identificadores abreviados de tres letras y de una letra para indicar la composición y secuencia de los aminoácidos polymerizados en las proteínas (Nelson *et al.*, 2008):

**Alanina** Ala (A)

**Metionina** Met (M)

**Cisteína** Cys (C)

**Asparagina** Asn (N)

**Ácido aspártico** Asp (D)

**Prolina** Pro (P)

**Ácido glutámico** Glu (E)

**Glutamina** Gln (Q)

**Fenilalanina** Phe (F)

**Arginina** Arg (R)

**Glicina** Gly (G)

**Serina** Ser (S)

**Histidina** His (H)

**Treonina** Thr (T)

**Isoleucina** Ile (I)

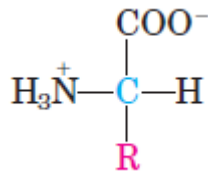
**Valina** Val (V)

**Lisina** Lys (K)

**Triptófano** Trp (W)

**Leucina** Leu (L)

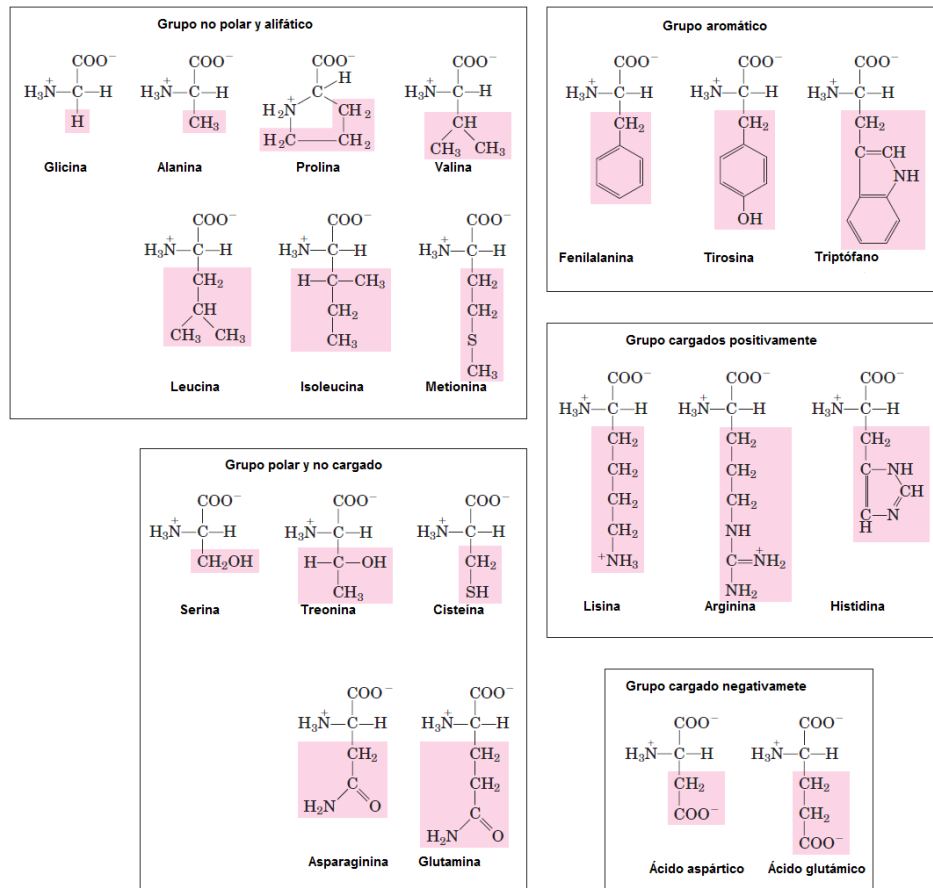
**Tirosina** Tyr (Y)



**Figura 2.** Estructura general de los aminoácidos naturales, excepto para la *Prolina*. Imagen tomada de (Nelson *et al.*, 2008).

Debido a la disposición tetraédrica de los orbitales de los enlaces alrededor del C $\alpha$ , los cuatro diferentes grupos pueden ocupar dos distintas posiciones espaciales únicas, por lo que los aminoácidos tienen dos estereoisómeros posibles: L-amino y D-

<sup>2</sup>A 25 °C, un pH menor que 7 es ácido, igual a 7 es neutro y si es mayor que 7 es básico.



**Figura 3.** Los cinco grupos en los que se encuentran clasificados los aminoácidos naturales. Las fórmulas estructurales en su estado de ionización predominante a pH 7. Las zonas sombreadas, son los grupos R de cada amino ácido. Imagen tomada de (Nelson *et al.*, 2008).

amino (ver Figura 4). Dado que son imágenes no superponibles entre si, las dos formas representan los estereoisómeros llamados enantiómeros.



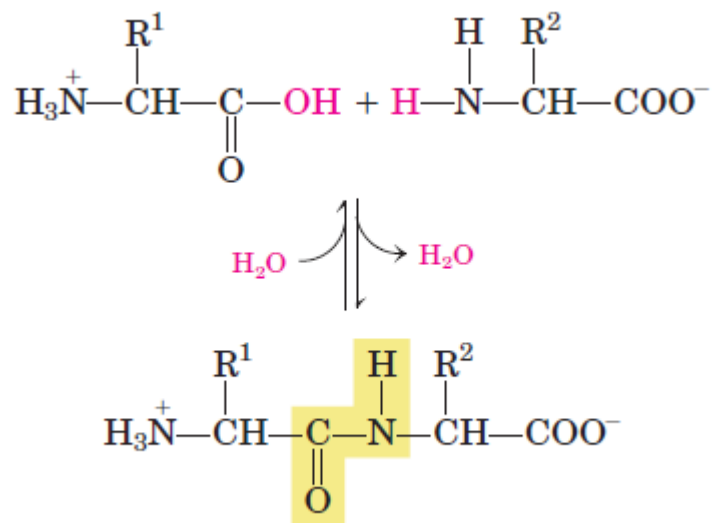
**Figura 4.** Estereoisomería en aminoácidos, son imágenes especulares no superponibles entre sí (enantiómeros). Los dos estereoisómeros de la alanina, L y D-alanina. Imagen tomada de (Nelson *et al.*, 2008).

Cuando los aminoácidos se forman por reacciones químicas ordinarias, se obtienen D-amino y L-amino indistintamente, sin embargo, los residuos encontrados en las proteínas en su forma natural, son exclusivamente L-amino, esto se debe a que las células únicamente pueden sintetizar los L-amino porque los sitios activos de las enzimas son asimétricos, lo que hace que las reacciones que catalizan sean estereoespecíficas. Se

pueden encontrar unos pocos residuos D-amino en unos pocos péptidos, por lo general pequeños, algunos péptidos presentes en las paredes celulares de algunas bacterias y en ciertos antibióticos peptídicos (Nelson *et al.*, 2008).

### 2.1.2. Cadenas de aminoácidos

Los polipéptidos biológicos varían en tamaño desde pequeños a muy grandes, y están conformados desde dos o tres, a miles de residuos de aminoácidos, unidos en forma de cadenas. Dos aminoácidos pueden unirse covalentemente mediante un enlace peptídico. Dicho enlace se forma por la deshidratación del grupo carboxilo de un aminoácido y el grupo amino del otro aminoácido (ver Figura 5). En condiciones fisiológicas del pH, el equilibrio de la reacción química favorece el mantenimiento de los aminoácidos en contra de la creación de los enlaces peptídicos. Para que la reacción sea más favorable a la formación del enlace peptídico, el grupo carboxilo debe ser modificado o activarse químicamente para que el grupo hidroxilo pueda eliminarse más fácilmente, función que realiza el ribosoma dentro de la célula (Nelson *et al.*, 2008).

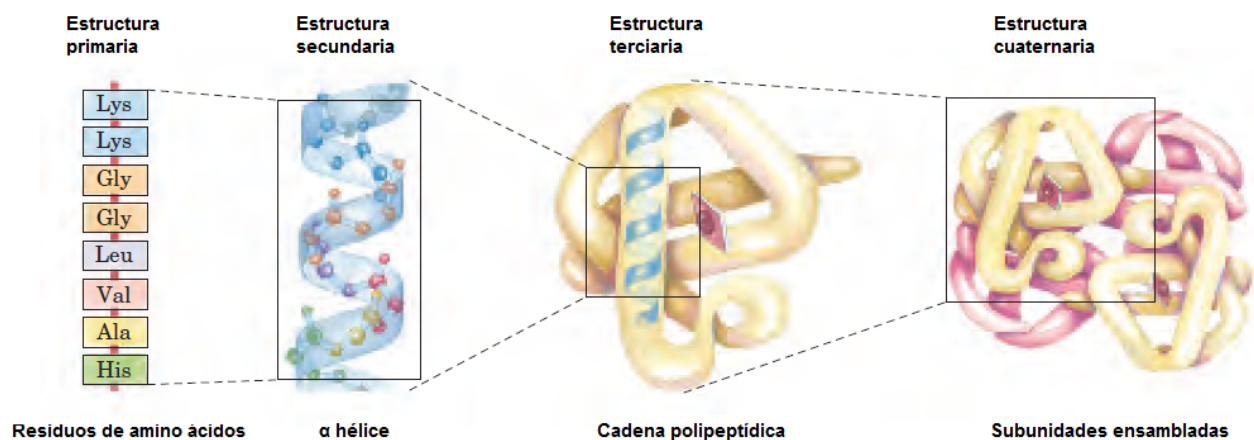


**Figura 5.** Formación de un enlace peptídico por condensación. El grupo amino de un aminoácido actúa como un nucleófilo para desplazar el grupo hidroxilo de otro aminoácido, formando un enlace peptídico. Imagen tomada de (Nelson *et al.*, 2008).

A la unión peptídica de dos aminoácidos se le conoce como dipéptido, al enlace peptídico de tres aminoácidos como tripéptido y así de manera similar para los tetrapéptidos, pentapéptidos, etc. Una cadena peptídica que está formada por solo unos pocos aminoácidos se le llama péptido, y a la unión de muchos aminoácidos como

polipéptido. Las proteínas pueden estar compuestas por miles de residuos de aminoácidos, por lo que el término proteína y polipéptido a veces se usan indistintamente. Las moléculas a las que se hace referencia como polipéptidos generalmente tienen pesos moleculares por debajo de 10000 g/mol, y las llamadas proteínas tienen pesos moleculares más altos. La representación de la cadena polipeptídica sin tener en cuenta su cadena lateral, es conocida como Columna Vertebral de la cadena peptídica. Algunas proteínas contienen dos o más cadenas polipeptídicas formando un único complejo, en estos casos los polipéptidos individuales no se consideran subunidades, sino que se les denominan cadenas de la proteína (Nelson *et al.*, 2008).

En la descripción y comprensión de la estructura de las proteínas, se usan cuatro niveles de complejidad, organizados por una jerarquía conceptual (ver Figura 6). La Estructura Primaria está descrita por la secuencia de residuos de aminoácidos. La Estructura Secundaria describe las disposiciones particularmente estables de la columna vertebral que dan lugar a patrones estructurales recurrentes. La Estructura Terciaria describe el plegamiento tridimensional de un polipéptido. La Estructura Cuaternaria se presenta cuando la proteína está formada por dos o más subunidades polipeptídicas (Nelson *et al.*, 2008).



**Figura 6.** Niveles estructurales en proteínas. La imagen mostrada corresponde a la hemoglobina. Imagen tomada de (Nelson *et al.*, 2008).

### 2.1.2.1. Estructura Primaria

Es una forma de representación de la proteína, solo muestra la secuencia ordenada de residuos de aminoácidos que componen a la estructura peptídica. Las diferencias en la estructura primaria son especialmente informativas. De cada proteína brinda número y secuencia de residuos de aminoácidos. La estructura primaria determina la



estructura tridimensional única y esto a su vez, determina la función de la proteína (Nelson *et al.*, 2008).

### 2.1.2.2. Estructura Secundaria

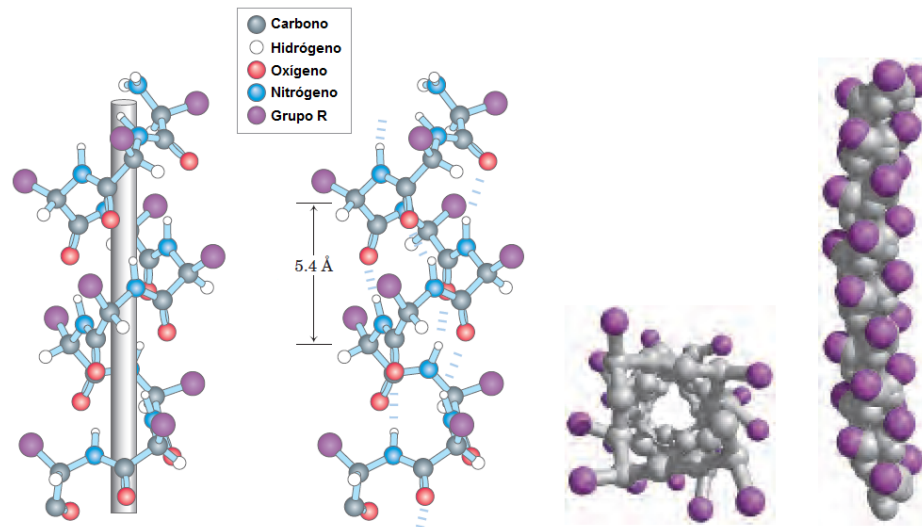
El término se refiere a la conformación local de alguna parte de un polipéptido, fue predicha utilizando principios químicos y observaciones experimentales en 1951. La estructura secundaria se enfoca en los patrones de plegamiento regulares comunes de la columna vertebral del polipéptido. Las estructuras secundarias más estables y que se encuentran ampliamente en las proteínas son las  $\alpha$  hélices y las  $\beta$  láminas (Nelson *et al.*, 2008).

**$\alpha$  hélice** es una estructura helicoidal y a su vez, la disposición más simple que la cadena polipeptídica podría adoptar con sus enlaces peptídicos rígidos. El esqueleto polipeptídico se enrolla estrechamente alrededor de un eje imaginario trazado longitudinalmente a través de la mitad de la hélice, con los grupos R de los residuos de aminoácidos, sobresaliendo hacia afuera desde el centro de la hélice (ver Figura 7). El giro helicoidal de la  $\alpha$  hélice está definida por la regla de la mano derecha<sup>3</sup>. Generalmente, una cuarta parte de los residuos de aminoácidos en los polipéptidos se encuentran en  $\alpha$  hélices; esto se debe a que la forma favorece los enlaces por puente de hidrógeno. Cada giro de la  $\alpha$  hélice se mantiene en giros adyacentes por tres o cuatro enlaces de hidrógeno. Los enlaces de hidrógeno dan a la estructura helicoidal su estabilidad.

**$\beta$  lámina** la columna vertebral de la cadena polipeptídica se extiende en forma de zigzag, ver Figura 8. Estos zigzag pueden disponerse lado a lado para formar una estructura que asemeja una serie de pliegues, llamados láminas u hojas betas. Esta disposición favorece la formación de enlaces de hidrógeno entre los segmentos adyacentes de la cadena polipeptídica. Los segmentos individuales que conforman la  $\beta$  lámina, generalmente se encuentran cerca en la cadena polipeptídica, aunque también pueden estar distante; incluso pueden ser segmentos en diferentes cadenas polipeptídicas. Los segmentos pueden formar  $\beta$  láminas

---

<sup>3</sup>Consiste en la semejanza con el giro de los dedos índice, medio, anular y meñique con respecto al pulgar, al cerrar en forma de puño la mano derecha. El dedo pulgar define la dirección de propagación de los residuos de aminoácidos.

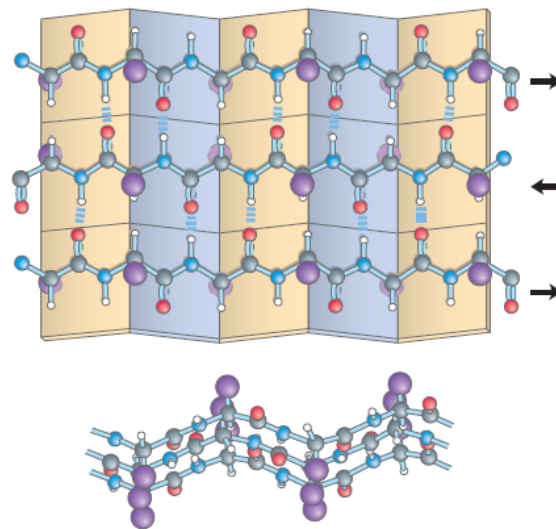


**Figura 7.** Cuatro modelos de la hélice alfa, mostrando diferentes aspectos de su estructura. Imagen tomada de (Nelson *et al.*, 2008).

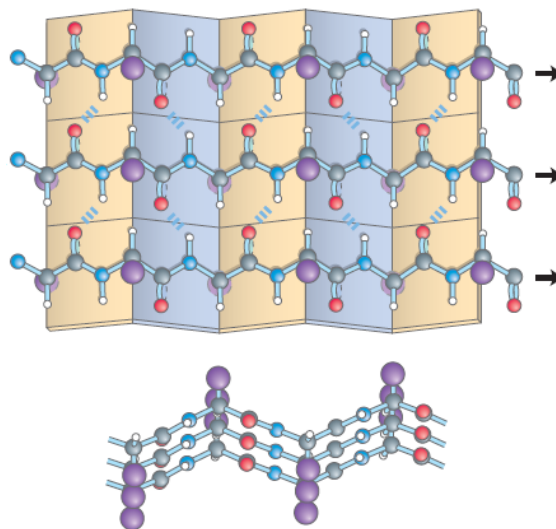
antiparalelas, cuando los aminoácidos de las cadenas de dos zigzags adyacentes, se encuentran en sentidos opuestos; pueden formar  $\beta$  láminas paralelas, cuando los aminoácidos de las cadenas de dos zigzags adyacentes, se encuentran en el mismo sentido; los patrones de enlace de hidrógeno son diferentes para ambos casos. Los grupos R de los aminoácidos adyacentes sobresalen de la estructura en zigzag en direcciones opuestas (Nelson *et al.*, 2008).

### 2.1.2.3. Estructura Terciaria

Describe la disposición tridimensional de todos los átomos que componen a la proteína. Muestra la interacción de los aminoácidos, que se encuentran separados en la estructura primaria y que se encuentran en diferentes estructuras secundarias, para lograr una estructura plegada en tres dimensiones. La ubicación de las estructuras secundarias y sus dobleces están determinadas por los aminoácidos que componen la cadena. Los segmentos de las cadenas peptídicas se mantienen en sus posiciones terciarias características mediante diferentes tipos de interacciones de enlaces débiles entre los segmentos (Nelson *et al.*, 2008).



(a) Anti paralela



(b) Paralela

**Figura 8.** La conformación beta de las cadenas polipeptídicas. Imagen tomada de (Nelson *et al.*, 2008).

#### 2.1.2.4. Estructura Cuaternaria

Se refiere a la disposición tridimensional de cadenas polipeptídicas o subunidades que conforman una proteína compuesta por más de una de dichas subunidades. Estas subunidades pueden ser idénticas o diferentes (Nelson *et al.*, 2008).

#### 2.1.3. Interacciones con otras moléculas

Las funciones de muchas de las proteínas implican la unión reversible con otras moléculas, estas moléculas son conocidas como ligandos. Un ligando puede ser cual-

quier tipo de molécula, incluso péptidos u otras proteínas. Esta propiedad transitoria de la interacción proteína-ligando es crítica para la vida, permitiendo que un organismo responda a las cambiantes circunstancias ambientales y metabólicas (Nelson *et al.*, 2008).

Un ligando se une a un sitio de la proteína llamado sitio de unión o sitio activo de la proteína; este sitio de unión es complementario al ligando en tamaño, forma, carga y carácter hidrófobo o hidrófilo. A esta interacción se le conoce como Acoplamiento Molecular, es altamente específica, la proteína es capaz de discriminar entre miles de moléculas en su entorno y unirse a una o pocas de ellas. Una proteína puede tener varios sitios activos separados para varios diferentes ligandos. Estas interacciones específicas son cruciales para mantener el orden y la vida de un organismo vivo. El acoplamiento de un ligando en el sitio activo de la proteína, a menudo, produce un cambio conformacional en la proteína, lo que hace que el sitio de unión sea más complementario al ligando. Esta adaptación se denomina ajuste inducido (Nelson *et al.*, 2008).

Las interacciones del ligando en el sitio activo pueden regularse haciendo uso de interacciones específicas con uno o más ligandos adicionales. Estos otros ligandos, pueden entorpecer la unión del primer ligando directamente en el sitio activo; o causar cambios conformacionales en la proteína que conlleven a un cambio estructural del sitio activo (Nelson *et al.*, 2008).

### **Llave-cerradura**

En 1894, Emil Fisher, propuso que las enzimas reconocen sus sustratos de la misma forma que una llave entra en la cerradura, existiendo solo una llave capaz de entrar en una cerradura. En esta similitud, la llave es el ligando y la cerradura es el sitio activo de la proteína. El ligando solamente se acoplaría al sitio activo que tenga complementariedad geométrica exacta con el sitio activo de la proteína objetivo; un ligando sin complementariedad geométrica con el sitio activo, no se acoplaría (Fischer, 1894).

## Ajuste inducido

Este modelo propuesto por Dan Koshland en 1958 es una actualización del modelo Llave-cerradura. Propone que la interacción de un ligando en el sitio activo de la proteína que se asemeja más al ajuste de una mano (ligando) introduciéndose a un guante (proteína). El modelo propuso que: se requiere la orientación precisa del ligando en el sitio activo; el ligando puede causar cambios posicionales de los residuos del sitio activo; y los cambios son inducidos por la unión de un ligando específico (Koshland, 1958).

## Selección de conformación

Este principio afirma que las proteínas plegadas existen como un grupo o conjunto de estados estructurales estrechamente relacionados y cada uno de estos estados tiene reactividades y/o afinidades de unión potencialmente diferentes. Esto sugiere que puede no ser necesario que el evento de unión induzca un cambio en la conformación en la proteína ya que las conformaciones similares al estado de unión con el ligando ya existe sin necesidad de estar presente el ligando (Beach *et al.*, 2005), por lo que puede ocurrir la unión del ligando directamente al estado estable.

## 2.2. Acoplamiento Molecular

Las interacciones entre el sitio activo y el ligando son de naturaleza mecánica cuántica, debido a la complejidad de los sistemas biológicos, la teoría cuántica no puede aplicarse directamente para describir dicha interacción. En consecuencia, la mayoría de los métodos utilizados en el acoplamiento molecular son de naturaleza empírica y generalmente carecen de generalidad. Los fenómenos mecánicos cuánticos, como la formación de un enlace covalente entre la proteína y el ligando tras el acoplamiento molecular y durante el estado de transición de la reacción, no se pueden predecir y/o evaluar utilizando estos métodos empíricos (Gu y Bourne, 2009). Al resultado de los métodos de acoplamiento molecular para predecir la interacción del ligando en el sitio activo se le llama pose de acoplamiento molecular.

El primer paso para realizar el acoplamiento molecular es identificar el sitio de unión del ligando en la proteína. Esto se hace para limitar el espacio de búsqueda so-

bre la superficie del receptor y así, minimizar los grados de libertad en la búsqueda. Generalmente el sitio activo se conoce a partir de las estructuras cristalinas del receptor unido a otros ligandos, aunque también es posible predecirse. Con frecuencia el sitio activo de una proteína es la cavidad más grande, pero no siempre ese es el caso positivo, por lo que se han desarrollado diferentes métodos de análisis y predicción del sitio activo (Gu y Bourne, 2009).

La siguiente etapa es identificar la función de puntuación, que distinguirá entre las poses generadas, la mejor solución que coincida estrechamente con la pose real del acoplamiento. Se considera al complejo proteína-ligando predicho más favorable, es el que es biológicamente relevante. A menudo las funciones de puntuación consideran el ligando y el sitio de unión como estructuras rígidas, esto no es realista, ya que las interacciones del ligando en el sitio activo son flexibles (Gu y Bourne, 2009).

### **2.2.1. Algoritmos de búsqueda**

Los algoritmos de acoplamiento generalmente contienen dos algoritmos de búsqueda, uno responsable de la colocación y exploración del ligando dentro del sitio de unión; el otro algoritmo busca los grados internos de libertad del ligando. Mientras que algunos programas de acoplamiento solo acoplan ligandos rígidos y evalúan conformaciones de ligandos predeterminadas, este enfoque permite generar soluciones cuando los recursos computacionales son limitados, sin embargo, no necesariamente se obtiene la mejor solución (Gu y Bourne, 2009).

Es posible buscar los seis grados de libertad de traslación y rotación del ligando en el sitio activo del receptor. Sin embargo, esta forma de búsqueda aumenta el tiempo de cómputo de forma exponencial, a medida que aumenta el número de grados de libertad del ligando. Los algoritmos de búsqueda intentan encontrar un equilibrio entre la búsqueda del panorama energético tan profundamente como sea posible, mientras encuentran una solución dentro de un tiempo razonable. Estos algoritmos se pueden dividir en tres categorías diferentes (Gu y Bourne, 2009):

**Búsqueda en cuadrícula:** Busca exhaustivamente el sitio de enlace mediante el movimiento del ligando rígido a través del grado de libertad disponible de forma

sistemática en el sitio activo. El tiempo de búsqueda y la precisión están determinados por el tamaño tridimensional de la cuadrícula y el tamaño de los pasos de búsqueda. Usa el método de la Transformada Rápida de Fourier para acelerar la búsqueda. La búsqueda está guiada por la complementariedad con la forma de la superficie de contacto y la complementariedad electrostática (Katchalski-Katzir *et al.*, 1992; Gabb *et al.*, 1997). Este enfoque no se usa en ligandos pequeños debido a que la mayoría son flexibles con un gran número de grados de libertad; sin embargo, se usa para el acoplamiento proteína-proteína, que generalmente son tratadas como moléculas rígidas.

**Descriptor del sitio activo:** Coloca el ligando en el sitio activo mediante la comparación de descriptores y es significativamente el más eficiente de los algoritmos de búsqueda (Ewing y Kuntz, 1997). Los descriptores para encontrar la solución generalmente son puntos con propiedades asociadas a ellos y colocados en el sitio activo del receptor. Los átomos del ligando se combinan con los descriptores del receptor teniendo alguna tolerancia. El algoritmo de coincidencias orienta al ligando en el sitio de unión que posteriormente se optimiza y se califica; de esta forma es posible explorar y optimizar un gran número de orientaciones. Para evitar un crecimiento exponencial asociado a la búsqueda sistemática y asegurar la eficiencia, se hace necesaria una poda cuidadosa del árbol de búsqueda (Ewing *et al.*, 2001; Moustakas *et al.*, 2006). Los descriptores que son asociados con propiedades físico-químicas y los átomos del ligando tienen que coincidir tanto geométrica como químicamente. Generalmente se hacen coincidir, como mínimo, un conjunto de tres a cuatro esferas en el receptor, con el mismo valor en los átomos del ligando. Se dice que se ha producido una coincidencia del ligando en el sitio de unión cuando las distancias entre las esferas son similares a las distancias interatómicas.

**Basado en energía:** Utilizan los campos de fuerza de la mecánica molecular para explorar con dinámica molecular o algoritmos de minimización de energía, la superficie de energía del ligando. Los mínimos del ligando deben ubicarse y evaluarse con respecto a su complementariedad con el receptor. Esta minimización de la energía es una búsqueda local por lo que solo es utilizada para optimizar las conformaciones generadas por otros algoritmos de búsqueda. En principio las búsquedas con dinámica molecular son búsquedas globales, pero tienden a es-

tancarse en mínimos locales, por lo que los tiempos de la simulación de búsqueda deben ser largos, para garantizar una cobertura completa del sitio de unión, lo que limita su aplicabilidad como una solución de alto rendimiento computacional en la identificación de poses proteína-ligando (David *et al.*, 2001). Algunas de las alternativas para acelerar el tiempo de búsqueda es la utilización de algoritmos genéticos (Morris *et al.*, 1998; Jones *et al.*, 1995) y las búsquedas tipo Monte-Carlo (Mcmartin y Bohacek, 1997) que son estocásticas y ayudan a que las simulaciones no se atasquen en mínimos locales cuando se buscan soluciones. Son estrategias de búsqueda que exploran todos los grados de libertad de forma simultánea, por lo que no se consideran como algoritmos de orientación.

Estos algoritmos de acoplamiento molecular presentan algunos puntos débiles para encontrar las soluciones óptimas. Con la precisión de los campos de fuerza actuales y el uso de una penalización constante por enlace rotativo, da lugar a una incertidumbre en el cálculo de la energía libre de entre 5 y 10 kcal/mol (Perola *et al.*, 2004). Por lo tanto es necesario una mayor precisión en los campos de fuerza y mejores funciones de penalización para la obtención de las energías libres de las conformaciones (Gu y Bourne, 2009).

### 2.2.2. Funciones de puntuación

En el acoplamiento molecular es de interés minimizar la energía libre de la unión entre el ligando y la proteína. Es una función de la energía entálpica y entrópica expresada por:

$$\Delta G_{bind} = \Delta H - T\Delta S, \quad (1)$$

donde  $\Delta G_{bind}$  es la energía libre de unión del ligando en el sitio activo de la proteína;  $\Delta H$  representa a la entalpía, que es el resultado de la eliminación del agua del sitio de unión y de la superficie del ligando, mientras es sustituido por las interacciones de van der Waals y coulómbicas;  $T$  representa la temperatura aplicada al sistema; y  $\Delta S$  representa los cambios en la entropía que son el resultado de restringir los grados internos de libertad del sitio de unión del receptor y el ligando, así como los cambios en los grados de libertad de traslación y rotación. Los cambios entrópicos como la flexibilidad, generalmente son desfavorables para la formación del complejo (Gu y



Bourne, 2009). El solvente va ganando entropía a medida que se libera del sitio de unión y de la superficie del ligando en la formación del complejo, también llamado efecto hidrofóbico (Kauzmann, 1959).

La energía libre de unión es un término complejo que puede calcularse directamente, pero debido a la naturaleza de la unión de energías libres, el costo computacional es considerable. La búsqueda de funciones de puntuación generalmente aplicables y altamente precisas que puedan evaluarse rápidamente aún continúa (Gu y Bourne, 2009).

La energía libre relativa puede aplicarse con mayor frecuencia. Los métodos más usados son la Integración Termodinámica y la Perturbación de la Energía Libre (Kollman, 1993). La energía libre de unión se obtiene transformando lentamente la pose de un ligando en otra pose. Estos métodos solo son factibles para pequeños cambios en la pose del ligando ya que el cálculo de energías libres de unión relativa aún tienen un costo considerable porque requieren mucho tiempo en los estados intermedios que deben muestrearse extensamente para pasar de una pose a otra (Gu y Bourne, 2009).

Los enfoques más prácticos para calcular las energías libres de unión, se pueden dividir en:

**Primeros principios:** Generalmente usan un campo de fuerza de Mecánica Molecular, aunque para ahorrar tiempo, a menudo no se calculan las fuerzas intramoleculares entre los átomos. Esta energía libre es solo una aproximación de la energía libre verdadera debido a que el campo de fuerza de la mecánica molecular solo tiene presente las fuerzas entálpicas entre la proteína y el ligando. Si el receptor es rígido, las energías aportadas por el receptor son precalculadas y almacenadas en una cuadrícula, esto acelera significativamente los cálculos y hace posible el uso de una función de puntuación de mecánica molecular en el acoplamiento molecular. Generalmente no se tienen en cuenta las interacciones con el solvente.

**Semi empíricos:** Debido a su naturaleza, los parámetros no son generalmente aplicables y deben derivarse para cada nuevo objetivo o serie química. Un ejemplo de método semi empírico es la Energía de Interacción Lineal. Sus cálculos requieren la disponibilidad de datos confiables del sitio activo. Aunque son simulados

solamente los estados iniciales y finales con Dinámica Molecular o con Monte-Carlo, el método es muy costoso en recursos computacionales y solo se usa para la optimización de soluciones potenciales y no para su identificación (Ljungberg *et al.*, 2001; Rizzo *et al.*, 2001).

**Empíricos:** Consiste en una serie de términos que desempeñan un papel en la conformación del complejo proteína-ligando. Una de las ventajas es que pueden evaluarse rápidamente y por lo tanto pueden ser usadas en el acoplamiento molecular. Los pesos de los términos se derivan ajustando la ecuación usando métodos de regresión a un conjunto de entrenamiento de complejos de proteína-ligando para los que se conoce tanto la información estructural como sus energías de interacción determinados experimentalmente. Aunque los términos utilizados representan principios físicos subyacentes a la formación del complejo, presenta varias debilidades: gran parte del detalle atómico se pierde debido al ajuste, el muestreo del espacio químico limitado por los complejos proteína-ligando da como resultado funciones empíricas más generales que específicas, y la mayoría de las funciones de puntuación empíricas consideran las interacciones como las distancias y los ángulos de los átomos involucrados.

**Basados en conocimiento:** Son funciones desarrolladas sobre la base del trabajo realizado en la comunidad de plegamiento de proteínas. Estas funciones se derivan utilizando solo complejos proteína-ligando, basándose en la hipótesis de que las interacciones que se observan con mayor frecuencia, son las más importantes en la estabilidad del complejo. Este enfoque tiene la ventaja que puede usar más complejos de proteína-ligando que el derivado de funciones empíricas, es más eficiente en consumo de recursos computacionales.

### 2.3. Métricas para las evaluaciones

En la siguiente sección se describirán las métricas que serán usadas para la evaluación de los resultados.

### 2.3.1. Desviación Cuadrática Media

La Desviación Cuadrática Media (Root Mean Standar Desviation, RMSD), es una medida que se usa para calcular qué tanto se asemejan las coordenadas atómicas de una estructura obtenida por métodos computacionales a las coordenadas atómicas de la misma estructura obtenida de forma experimental. La ecuación para la obtención del RMSD basado en las variaciones en las posiciones atómicas, se expresa por:

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n d_i^2}, \quad (2)$$

donde  $d_i$  es la distancia entre átomo  $i$ -ésimo de la estructura obtenida computacionalmente y el átomo  $i$ -ésimo obtenido experimentalmente en la estructura cristalina;  $n$  es el número total de átomos (Cohen y Sternberg, 1980). Generalmente se considera una buena predicción a una estructura obtenida computacionalmente con un  $RMSD \leq 2\text{\AA}$  (Gu y Bourne, 2009).

### 2.3.2. Coeficiente de Correlación

El Coeficiente de Correlación de las funciones de puntuación fue propuesta por Pearson en 1895 (Lee Rodgers y Nicewander, 1988). Es un índice sin dimensiones que es invariante a las transformaciones lineales de las variables. La ecuación para la obtención del coeficiente de correlación, se expresa por:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left[ \sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2}}, \quad (3)$$

donde  $r$  es el coeficiente de correlación,  $x_i$  es el elemento  $i$ -ésimo actual,  $y_i$  es el elemento  $i$ -ésimo predicho,  $\bar{x}$  es la media aritmética de los elementos actuales,  $\bar{y}$  es la media aritmética de los elementos predichos y  $n$  es el número total de elementos. El denominador ajusta las escalas de las variables para que los límites estén entre  $\pm 1$ , donde un valor muy cercano a  $+1$  representa una correlación positiva, un valor muy cercano a  $0$  ninguna correlación y un valor muy cercano a  $-1$  una correlación negativa.

### 2.3.3. Error Absoluto Medio

El Error Absoluto Medio (Mean Absolute Error, MAE) es una magnitud que expresa la media de los valores absolutos de un conjunto de errores. La ecuación para la obtención del MAE, se expresa por:

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i|, \quad (4)$$

donde  $e_i = y_i - x_i$ ,  $e_i$  es el error  $i$ -ésimo,  $y_i$  es el valor predicho y  $x_i$  es el valor real; y  $n$  es el número total de errores.

## 2.4. Huella de interacción

El enfoque de Huella Digital de Interacción (Interaction Fingerprint, IFP) representa una forma de codificar las interacciones de los complejos proteína-ligando. Un IFP codifica una presencia (1) o ausencia (0) de interacciones del ligando con aminoácidos específicos del sitio de unión de la proteína, formando una cadena binaria. Cada aminoácido del sitio de unión se describe mediante el mismo número de tipos de interacciones: hidrofóbico, aromático cara a cara, borde aromático cara a cara, enlace de hidrógeno donado por la proteína, enlace de hidrógeno donado por el ligando, enlace iónico con el catión de la proteína y el anión de la proteína, y el ion metálico. La longitud del IFP depende del tamaño del sitio de unión ligando-proteína (Chupakhin *et al.*, 2014).

### 2.4.1. Huella Estructural de la Interacción Proteína-ligando

Las huellas dactilares estructurales de la interacción proteína-ligando (Structural protein-ligand interaction fingerprints, SPLIF) cuantifican y comparan las interacciones ligando-proteína. Las estructuras tridimensionales del ligando y los fragmentos de la proteína que interactúan, se codifican explícitamente en la huella digital; todos los tipos de interacciones que pueden ocurrir entre los fragmentos están codificados implícitamente en SPLIF (Da y Kireev, 2014).

En la construcción de un SPLIF, por cada pose ligando-proteína obtenido, se deben

considerar que dos átomos están en contacto, si la distancia entre ellos está dentro de un umbral establecido. Por cada átomo del ligando y tomándolo como origen, se analizan los átomos de la proteína que se encuentran hasta cierta distancia de dicho átomo.

## **2.5. Descriptor molecular**

Un descriptor molecular es, según Todeschini y Consonni (2009), *el resultado final de un procedimiento lógico y matemático que transforma la información química codificada dentro de una representación simbólica de una molécula en un número útil o el resultado de un experimento estandarizado*. Se debe prestar atención al término útil, el cual puede significar que el descriptor puede dar más información sobre la interpretación de las propiedades moleculares y/o puede estar presente en un modelo de predicción de propiedades. Los descriptores moleculares son fácilmente interpretables y pueden o no dar información desde un punto de vista estructural o químico.

### **2.5.1. Descriptor Simple de Interacción Ligando-Receptor**

El descriptor simple de interacción ligando-receptor (Simple Ligand-Receptor Interaction Descriptor, SILIRID) es un descriptor molecular de tamaño fijo que caracteriza las interacciones ligando-proteína. Se puede obtener a partir de IFP o de SPLIF sumando los bits de los aminoácidos idénticos correspondientes. Esto da como resultado un vector de números enteros de tamaño fijo que corresponde al producto del número de aminoácidos de entrada y el IFP o el SPLIF calculado (Chupakhin *et al.*, 2014).

## **2.6. Relación Cuantitativa Estructura-Actividad**

La Relación Cuantitativa Estructura-Actividad (Quantitative Structure–Activity Relationships, QSAR) es el resultado del proceso que inicia con la descripción adecuada de las estructuras moleculares y termina con inferencias, hipótesis y predicciones sobre el comportamiento biológico de las moléculas en condiciones ambientales. Está basado en que la estructura molecular contiene características responsables de sus propiedades físicas, químicas y biológicas; y en su capacidad de capturar esas características en uno o varios descriptores moleculares. Mediante el modelado QSAR, la

reactividad de un compuesto químico puede ser inferida a partir de la estructura molecular de compuestos similares, cuyas propiedades ya han sido evaluadas (Todeschini y Consonni, 2009).

En un proceso QSAR es primordial el conocimiento exhaustivo de los compuestos a analizar, esto implica la obtención de datos experimentales confiables, los cuales pueden producirse experimentalmente o reportados por la literatura. La precisión debe evaluarse cuidadosamente, ya que la precisión de la propiedad estimada, no puede ser mayor al grado reportado por los datos de entrada. Posteriormente, es necesario definir un espacio químico confiable, seleccionar las características estructurales que se consideren sean las responsables de modelar el fenómeno en estudio. Es común el uso de numerosos descriptores moleculares, dentro de los cuales se espera estén las variables candidatas para el modelado y luego se aplica una técnica de reducción del número de variables (Todeschini y Consonni, 2009).

Por último es necesario la validación del modelo QSAR, generalmente se considera como el requisito más importante para un modelo QSAR aceptable. La capacidad predictiva del modelo se evalúa dividiendo los datos experimentales en dos conjuntos:

**Conjunto de entrenamiento:** Es el conjunto de elementos, mediante el cual se entrena el modelo predictivo.

**Conjunto de prueba:** Es el conjunto de elementos mediante el cual se evalúa la capacidad predictiva del modelo.

## Capítulo 3. Metodología

---

En el presente capítulo se describe la metodología diseñada para obtener los complejos moleculares para la realización de las pruebas, su preparación para el acoplamiento molecular, la obtención de las poses y los descriptores moleculares calculados de las poses.

### 3.1. Complejos químicos

En la presente sección se abordan los complejos moleculares con los que estaremos trabajando y analizando sus propiedades y composición, su selección y depuración.

#### 3.1.1. Banco de Datos de Proteínas

El Banco de Datos de Proteínas (Protein Data Bank, PDB) es uno de los recursos de datos experimentales para la investigación biomédica. Se estableció en 1971 luego de un simposio del Laboratorio de Cold Spring Harbor (Burley *et al.*, 2018b). Actualmente cuenta con 154 015<sup>1</sup> estructuras biomoleculares determinadas por cristalografía, espectroscopia de Resonancia Magnética Nuclear y por microscopía electrónica 3D. Está centrado en los modelos estructurales de proteínas, ácido desoxirribonucleico (ADN) y Ribonucleico (ARN) y sus complejos con metales y pequeñas moléculas (Burley *et al.*, 2018a). De los meta datos presentes para cada complejo, nos centraremos en los que nos brindarán información básica para la investigación:

- **PDB Id:** Es un identificador único de 4 caracteres alfanuméricos para cada complejo.
- **Secuencia(s):** Es una lista con los aminoácidos que componen la(s) secuencia(s) que conforman el complejo, identificador de la(s) secuencia(s) y la cantidad de aminoácidos de la(s) secuencia(s).
- **Estructura secundaria:** Son las uniones entre las columnas vertebrales de los aminoácidos de la(s) secuencia(s) para formar  $\alpha$  hélices y  $\beta$  láminas del complejo.

---

<sup>1</sup>Revisado el 22 de julio de 2019

- **Coordenadas 3D:** Son las coordenadas 3D de cada átomo que componen la(s) secuencia(s) del complejo.

En el Anexo A, Formato .1, podemos encontrar un ejemplo de fichero en formato PDB; correspondiente al complejo 1apm.

### **3.1.2. PeptiDB**

Base de datos de complejos proteína-péptido, propuesta por el Departamento de Microbiología y Genética Molecular de la Universidad Hebrea de Jerusalén, Israel (London *et al.*, 2010). PeptiDB contiene 103 estructuras de alta resolución y no redundantes. En los estudios computacionales realizados, comprobaron que la unión del péptido en el sitio activo de la proteína blanco, se realizó minimizando los cambios en la proteína y maximizando la entalpía de los enlaces de hidrógeno y del empaqueo de la cadena lateral. Por último comprobaron que el péptido se acopla en la cavidad más grande que tiene la proteína objetivo.

### **3.1.3. LEADS-PEP**

Base de datos de complejos proteína-péptido, propuesta por el Instituto Fraunhofer para la Biología Molecular y Aplicación Ecológica, Alemania (Hauser y Windshügel, 2016). LEADS-PEP contiene 53 complejos de proteína-péptido con alta resolución y longitudes de péptidos entre los 3 y 12 residuos; para lograr variedad en la selección se realizó agrupación por secuencias. En los estudios realizados, utilizaron la estructura flexible del péptido, comprobaron que para péptidos pequeños con longitud de entre 3 y 4 residuos, las herramientas computacionales de acoplamiento molecular se comportaron según lo previsto, ya con péptidos de longitud mayor, se requiere de un alto costo computacional y se obtiene un pobre desempeño.

### **3.1.4. PeptideInhibitor**

Base de datos de complejos proteína-péptido, obtenida el 13 de diciembre de 2018, procedente del PDB. El criterio de la consulta fue construido a partir de la Interfaz de búsqueda avanzada:

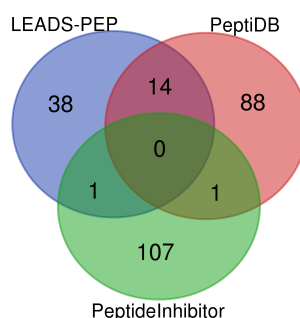


1. Fue seleccionada la opción “Structure Description”.
2. Se seleccionó la opción “Contains”.
3. En el cuadro de texto se introdujo la frase clave: “peptide inhibitor”.

La consulta arrojó 109 complejos que pueden ser redundantes en algunos casos ya que no se realizó otros procesos de filtrado.

### 3.1.5. DockingDB

Para la realización de la presente investigación se decidió unir, ver Figura 9, las bases de datos: PeptiDB, LEADS-PEP y PeptideInhibitor; en una única base de datos: DockingDB, ver Anexo E (Tabla 2).



**Figura 9.** Unión de PeptiDB, LEADS-PEP y PeptideInhibitor en DockingDB.

De la base de datos DockingDB, se realizó un filtrado de forma manual, usando AutoDock Tools, de las estructuras moleculares presentes en la misma. Los filtros se listan a continuación:

- Se eliminaron los heteroátomos que no forman parte de algún amino ácido del complejo biológico.
- Se seleccionó la cadena correspondiente al péptido (Cad. pept.) y se almacenó según los formatos de entrada de las herramientas de acoplamiento molecular (*pdb*, *pdbqt* y *mol2*). (En el Anexo A, formatos .1, .2, .3 y .4.)
- Se calcularon las rotaciones de los enlaces del péptido para un acoplamiento flexible (Pépt. rot.).

- Se comprobó la longitud y exactitud de los aminoácidos que componen el péptido con las respectivas coordenadas tridimensionales de los átomos que lo componen (Pépt. sec.).
- Se seleccionaron las cadenas que tengan algún amino ácido a no más de 8Å del péptido (Cad. prot.) y fueron almacenadas según los formatos de entrada de las herramientas de acoplamiento molecular.
- Al complejo 1xoc no fue posible realizarle las transformaciones necesarias por errores en el AutoDock Tools, (ver Anexo D, errores .1 y .2).

### 3.2. Acoplamiento molecular

Las herramientas computacionales para el acoplamiento molecular, usan funciones de puntuación basadas en estructuras las cuales buscan mitigar algunas de las deficiencias de los campos de fuerza tradicionales mediante el desarrollo de funciones de energía libre que reproducen las constantes de unión observadas. Estos enfoques usan una “ecuación maestra” expandida para modelar la energía libre de unión, añadiendo términos entrópicos a las ecuaciones de la mecánica molecular:

$$\Delta G = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{conform} + \Delta G_{tor} + \Delta G_{sol}, \quad (5)$$

donde  $\Delta G$  es la energía libre que se busca minimizar;  $\Delta G_{vdw}$  es la energía de los términos mecánicos moleculares típicos para la dispersión y la fricción;  $\Delta G_{hbond}$  es la energía que aportan los enlaces de hidrógeno;  $\Delta G_{elec}$  es la energía electrostática;  $\Delta G_{conform}$  las desviaciones de la geometría covalente;  $\Delta G_{tor}$  modela la restricción de las rotaciones internas de la conformación global; y  $\Delta G_{sol}$  representa los modelos que explican el efecto hidrofóbico de la estructura ya acoplada (Morris *et al.*, 1998).

Los métodos reportados para el acoplamiento molecular se dividen en dos amplias categorías: métodos de coincidencia y métodos de simulación de acoplamiento. Los métodos por coincidencias crean un modelo del sitio activo, que incluye sitios de enlaces de hidrógeno y sitios que son estéricamente accesibles, y luego intentan acoplar el ligando dado en el modelo como un cuerpo rígido al hacer coincidir su geometría con la del sitio activo, basados en el modelo de llave-cerradura. La segunda clase de

técnicas de acoplamiento, modela el acoplamiento de un ligando a un objetivo con mayor detalle: el ligando comienza aleatoriamente fuera del sitio activo de la proteína, y explora las traslaciones, orientaciones y conformaciones hasta que se encuentra un sitio de energía mínima. Estas técnicas son típicamente más lentas que las técnicas de emparejamiento, pero permiten modelar la flexibilidad dentro del ligando y pueden utilizar una mecánica molecular más detallada para calcular la energía del ligando dentro del sitio activo del blanco (Morris *et al.*, 1998).

En (Pagadala *et al.*, 2017) se realiza una revisión de las herramientas computacionales que implementan métodos para el acoplamiento molecular para pequeñas moléculas o ligandos aunque reconocen su uso en el diseño de péptidos inhibidores. En la últimas dos décadas se han desarrollado más de 60 herramientas las cuales se pueden encontrar bajo licencias académicas o comerciales. Entre estas herramientas podemos encontrar el: DOCK, AutoDock, FlexX, Surflex, GOLD, ICM, Glide, Cdocker, LigandFit, MCDock, FRED, MOE-Dock, LeDock, AutoDock Vina, rDock, UCSF Dock, entre muchos otros. Realizan una descripción de las dos estrategias de acoplamiento molecular, acoplamiento rígido y flexible, y la precisión alcanzada por las herramientas computacionales con cada estrategia. Llegando a la conclusión que aunque existen muchas herramientas de acoplamiento para la búsqueda y la predicción de poses, las funciones de puntuación no son precisas y deben mejorarse aún más.

En (Ciemny *et al.*, 2018) se realiza una revisión de las herramientas computacionales que implementan métodos para el acoplamiento molecular proteína-péptido. Realizan una descripción los métodos que implementan algunas de las herramientas como son el: GalaxyPepDock, PepComposer, Rosetta FlexPepDock, DynaDock, PepCrawler, HADDOCK peptide docking, PEP-FOLD 3, AutoDock Vina, DINC 2.0, Gold, Surflex-Dock, pepATTRACT, MDockPeP, CABS-dock, AnchorDock, ClusPro PeptiDock y PIPER-FlexPepDock; así como un análisis de los resultados de las herramientas para diferentes bases de datos de estructuras, entre ellas: LEADS-PEP y PeptiDB. Exponen la dificultad que representan los grados de libertad del péptido en el acoplamiento flexible, frente al acoplamiento rígido. Llegando a la conclusión que no siempre se puede obtener la pose nativa del complejo, ni saber qué tan precisas son las poses obtenidas, así como no se puede saber con exactitud cuántas poses son necesarias obtener para obtener la pose correcta.

### **3.2.1. Formas de acoplamiento molecular**

Se pueden realizar dos formas de acoplamiento molecular, según el sitio activo de la molécula objetivo: a ciegas y localizado; y dos formas según la rigidez de las estructuras involucradas: rígido y flexible.

Se realizó el acoplamiento a ciegas o global con el péptido y la proteína rígidos, donde no se conoce el sitio activo de la molécula blanco por lo que las poses del péptido pueden estar en toda la superficie de la molécula objetivo y la función de puntuación selecciona la cavidad donde se une el péptido con menor energía de acoplamiento.

Se realizó el acoplamiento dirigido con el péptido y la proteína rígidos, donde se conoce el sitio activo de la molécula blanco y se restringe las conformaciones del péptido mediante una caja y la función de puntuación selecciona la mejor conformación del péptido en la cavidad activa de la molécula objetivo.

Al igual que en el caso anterior, se realizó el acoplamiento dirigido con el péptido flexible y la proteína rígida, donde se conoce el sitio activo de la molécula blanco y se restringe las conformaciones del péptido mediante una caja y la función de puntuación selecciona la mejor conformación del péptido en la cavidad activa de la molécula objetivo.

En todos los casos se almacenaron los resultados de las energías libres y las coordenadas cartesianas de los átomos de cada pose calculada por las herramientas de acoplamiento molecular. Se calculó el RMSD para las coordenadas cartesianas de cada pose obtenida por las herramientas de acoplamiento molecular con respecto a las coordenadas de la pose obtenida del cristal de forma experimental.

A continuación se realizará una descripción de las herramientas de acoplamiento molecular que utilizamos en nuestros experimentos.

### **3.2.2. AutoDock4**

Un programa de acoplamiento molecular lanzado en 1990 por el Instituto de Investigación Scripps, Biología Molecular, La Jolla, CA (Goodsell y Olson, 1990). Utiliza un

método basado en una cuadrícula, se calcula la energía libre de interacción entre la molécula objetivo y el ligando, creando una red de energías la cual será usada como tabla de búsqueda durante el acoplamiento (Morris *et al.*, 2009). Incorpora un método de desolvatación basado en un conjunto típico de tipos de átomos y sus cargas. Permite el modelado flexible de porciones de las moléculas, lo que permite la rotación en torno a los grados de libertad de las mismas.

Para la función de energía libre se implementa una aproximación usando el ciclo termodinámico de Wesson y Eisenberg. A partir de AutoDock3 la función de energía libre se plantea como:

$$\begin{aligned} \Delta G = & \Delta G_{vdW} \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \\ & + \Delta G_{hbond} \sum_{i,j} E(t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{hbond} \right) \\ & + \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \\ & + \Delta G_{tor} N_{tor} \\ & + \Delta G_{sol} \sum_{i,cj} S_i V_j e^{(-r_{ij}^2/2\sigma^2)} \end{aligned} \quad (6)$$

donde los términos  $\Delta G_x$  en el lado derecho de la ecuación son coeficientes determinados empíricamente usando análisis de regresión lineal de un conjunto de complejos de proteína-ligando con constantes de unión conocidas. La  $E_{hbond}$ , es la energía promedio estimada del enlace de hidrógeno del agua con un átomo polar. Las sumas se realizan sobre todos los pares de átomos de ligando y átomos de proteína, además de todos los pares de átomos en el ligando que están separados por tres o más enlaces.

El método de búsqueda de las conformaciones, es un Algoritmo Genético Lamarckiano (Morris *et al.*, 1998). Los individuos de la población de conformaciones iniciales: mutan, intercambian sus parámetros conformacionales y compiten de una manera análoga a la evolución biológica en generaciones sucesivas; quedando al terminar, los individuos con conformaciones de energía libre mínima de acoplamiento. La modificación lamarckiana permite a las conformaciones individuales buscar mínimos locales para pasar esa información a generaciones posteriores.

El formato de los ficheros de entrada es *pdbqt* tanto para el ligando como para el receptor; en este fichero se encuentran las coordenadas 3D de los átomos y datos de torsión de la estructura. Necesita un fichero de configuración con formato *dprf*. El fichero de salida se encuentra en formato *dlog*; contiene las coordenadas 3D, las torsiones y la información energética de las mejores poses del ligando.

### 3.2.3. AutoDock Vina

Un nuevo programa de acoplamiento molecular propuesto en 2009 por el Instituto de Investigación Scripps, Biología Molecular, La Jolla, CA (Trott y Olson, 2009). Mejora la precisión de las predicciones de las conformaciones, calcula los mapas de cuadrículas y agrupa los resultados de forma transparente para el usuario, logrando una aceleración de hasta dos órdenes de magnitud en comparación con AutoDock4. Utiliza una variante de X-Score (Wang *et al.*, 2002), con un ajuste usando PDBbind (Wang *et al.*, 2005, 2004) y una función de puntuación que tiene en cuenta las contribuciones inter e intramoleculares.

La forma general de la función de puntuación para Vina es definida como:

$$c = \sum_{i < j} f_{t_i t_j}(r_{ij}), \quad (7)$$

donde la suma es sobre todos los pares de átomos que pueden moverse entre sí, excluyendo los átomos separados por hasta tres enlaces covalentes consecutivos,  $t_i$  es el tipo del átomo de índice  $i$ ,  $r_{ij}$  es la distancia interatómica y la función de interacción es definida como  $f_{t_i t_j}(r_{ij}) \equiv h_{t_i t_j}(d_{ij})$ , donde  $d_{ij} = r_{ij} - R_{t_i} - R_{t_j}$ ,  $R_t$  es el radio de van der Waals para el átomo de tipo  $t$  y es expresado en Å. Por último la función  $h_{t_i t_j}(d_{ij})$  es una suma ponderada de interacciones estéricas:

$$gauss_1(d) = e^{-(d/0.5)^2}, \quad (8)$$

$$gauss_2(d) = e^{-((d-3)/2)^2} \quad (9)$$

$$repulsion(d) = \begin{cases} d^2, & \text{si } d < 0 \\ 0, & \text{si } d \geq 0 \end{cases}, \quad (10)$$

$$hidrofobia(d) = \begin{cases} 1, & \text{si } d < 0.5 \\ -d + 1.5, & \text{si } 0.5 \leq d \leq 1.5, \\ 0, & \text{si } d > 1.5 \end{cases}, \quad (11)$$

y

$$enlaceHidrogeno(d) = \begin{cases} 1, & \text{si } d < -0.7 \\ -\frac{10}{7}d, & \text{si } -0.7 \leq d \leq 0. \\ 0, & \text{si } d > 0 \end{cases}. \quad (12)$$

En el caso de los metales son tratados formalmente como donadores de enlaces de hidrógeno. Todas las funciones de interacción  $f_{t_i t_j}$  se cortan en  $r_{ij} = 8\text{\AA}$ .

El método de búsqueda de las conformaciones lo realizan mediante el optimizador global de Búsqueda Local Iterativa (Trott y Olson, 2009), en su variante Broyden-Fletcher-Goldfarb-Shanno, que es un método eficiente, casi-Newton. Este tipo de método no solo usa el valor de la función de puntuación, sino también las derivadas de la misma respecto a sus argumentos, en este caso usaron como argumentos la posición y orientación de los ligandos, así como los valores de las torsiones de los residuos flexibles. Estas evaluaciones las realiza de forma simultánea aprovechando el paralelismo en múltiples CPUs con memoria compartida.

El formato de los ficheros de entrada es *pdbqt* tanto para el ligando como para el receptor; en este fichero se encuentran las coordenadas 3D de los átomos y datos de torsión de la estructura. Necesita un fichero de configuración con formato libre. El fichero de salida se encuentra en formato *pdbqt*; contiene las coordenadas 3D y las torsiones y la información energética de las mejores poses del ligando, ordenadas por sus puntuaciones energéticas y los RMSD de las conformaciones con respecto a la mejor conformación.

### 3.2.4. Dock6

Es un software de acoplamiento molecular desarrollado por el Departamento de Química Farmacéutica de la Universidad de California en San Francisco (Allen *et al.*, 2015). Utiliza un algoritmo de búsqueda de anclaje y crecimiento, un método de amplitud primero para el muestreo conformacional de moléculas pequeñas. Como función de energía utiliza una variante de energía repulsiva de van der Waals. Logró una aceleración sustancial al calcular esta energía interna al construir un conjunto de átomos que podrían formar una interacción de van der Waals interna al ligando, de esta forma se puede invocar en cualquier etapa del proceso de anclaje y crecimiento. Esto hace que mejore la capacidad de acoplamiento, así como el muestreo general. Incluye una nueva función de puntuación de huella, método eficaz para caracterizar la unión de los ligandos por residuos (Holden *et al.*, 2012); así como una nueva función RMSD basada en el algoritmo húngaro para moléculas simétricas (Allen y Rizzo, 2014).

La evaluación de energía busca minimizar la suma:

$$E_{total} = E_{interaction} + E_{internal} + E_{restraint}, \quad (13)$$

donde  $E_{interaction}$  es la energía de interacción entre el ligando y el receptor,  $E_{internal}$  es la energía interna del ligando y  $E_{restraint} = k(RMSD)^2$ . El valor por omisión de  $k$  es de  $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ .

El formato de los ficheros de entrada es *mol2* tanto para el ligando como para el receptor; en este fichero se encuentran las coordenadas 3D de los átomos. Necesita un fichero de configuración con formato *in*. El fichero de salida se encuentra en formato *mol2*; contiene las coordenadas 3D y la información energética de las mejores poses del ligando ordenadas por sus puntuaciones energéticas.

### 3.2.5. ClusPro

Es un servidor web para el acoplamiento molecular de proteína-proteína lanzado por el Departamento de Matemáticas Aplicadas y Estadísticas, Stony Brook University Nueva York en el año 2004 (Comeau *et al.*, 2004). El servidor realiza el muestreo de millones de conformaciones con las estructuras rígidas utilizando el programa de aco-



plamiento PIPER (Kozakov *et al.*, 2006), programa basado en el enfoque de correlación de la Transformada Rápida de Fourier (Katchalski-Katzir *et al.*, 1992); luego realiza el agrupamiento de las conformaciones basados en el RMSD de las 1000 estructuras con energía más bajas para encontrar los grupos que representarán los modelos más probables; y por último realiza un refinamiento de las estructuras seleccionadas minimizando la función de energía (Kozakov *et al.*, 2017).

El servidor genera cuatro conjuntos de conformaciones utilizando esquemas de puntuación: conformaciones balanceadas, donde no se favorece ninguna de las interacciones; conformaciones favoreciendo las interacciones electrostáticas; conformaciones favoreciendo las interacciones hidrofóbicas; y las conformaciones favoreciendo las interacciones de van der Waals y electrostática (Kozakov *et al.*, 2017).

La selección de los grupos la realiza buscando la conformación que tenga mayor número de vecinos a un radio IRMSD de 9Å; con esas conformaciones se crea un grupo y se eliminan del conjunto de selección y se realiza una nueva búsqueda. De esta manera genera hasta 30 agrupamientos (Kozakov *et al.*, 2017). Luego de la agrupación, minimizan la energía con 300 pasos de una red truncal fija usando solamente términos de van der Waals del potencial Charmm. En su configuración estándar, produce las estructuras de los centros de los 10 grupos más poblados.

El formato de los ficheros de entrada es *pdb* tanto para el ligando como para el receptor; en este fichero se encuentran las coordenadas 3D de los átomos. Los ficheros de salida se encuentran en formato *csv*, los grupos, los miembros del grupo y su peso energético; y *pdb*, contiene las coordenadas 3D de las mejores conformaciones del ligando centro del grupo al cual pertenece y las coordenadas 3D del receptor.

### **3.3. Manipulación de moléculas**

En esta sección se abordan algunas de las herramientas computacionales que permiten la manipulación de los complejos moleculares para que cumplan con las especificaciones de entrada y análisis de los resultados arrojados por cada una de las herramientas de acoplamiento molecular.

### **3.3.1. AutoDock Tools 4**

Es una herramienta computacional con interfaz gráfica efectiva para la preparación de coordenadas, diseño de experimentos y análisis de resultados. Facilita la confección de ficheros moleculares, mediante protonación, cálculo de cargas y la especificación de enlaces giratorios. Permite identificar el sitio activo, del objetivo, y el volumen y posición de la caja de acoplamiento molecular de forma visual (Morris *et al.*, 2009). Es el software por excelencia para la preparación de las moléculas y el análisis de los resultados con AutoDock4 y AutoDock Vina.

### **3.3.2. UCSF Chimera**

Es una herramienta computacional con interfaz gráfica para la preparación de las moléculas. Permite la visualización de múltiples moléculas a gran escala y la propiedad de colaboración interactiva de diferentes colaboradores remotos. Entre las funcionalidades que ofrece están: un visualizador de múltiples alineamientos; un visualizador de acoplamientos moleculares para la detección de ligandos acoplados; un visualizador de trayectorias de dinámica molecular; y un visualizador de volúmenes para el análisis de datos volumétricos (Pettersen *et al.*, 2004). Es la herramienta por excelencia para la preparación de las moléculas y el análisis de los resultados del Dock6.

## **3.4. Preparación del Acoplamiento Molecular**

En esta sección se muestran las herramientas computacionales necesarias para generar las estructuras requeridas en la entrada para cada una de las herramientas de acoplamiento molecular.

### **3.4.1. AutoGrid4**

Es una herramienta computacional para la generación de mapas de puntos a partir de las coordenadas 3D de la molécula blanco (Morris *et al.*, 1998). Estos mapas de puntos son parte de los parámetros de entrada la herramienta AutoDock4.

### **3.4.2. Grid**

Es una herramienta computacional para la generación de la malla de energías de atracción y repulsión a partir de las coordenadas 3D de la molécula blanco (Brozell *et al.*, 2012). La malla de energía es uno de los parámetros de entrada de la herramienta Dock6.

### **3.4.3. ShowBox**

Es una herramienta computacional para la generación de la caja que delimitará las conformaciones del ligando en la proteína blanco. La caja es uno de los parámetros de entrada de la herramienta Dock6.

### **3.4.4. SphGen**

Es una herramienta computacional que genera grupos de esferas superpuestas que describen la superficie de acceso al solvente a partir de las coordenadas 3D del sitio activo de la molécula blanco. Cada grupo representa un posible sitio de unión para los ligandos (Chen *et al.*, 1997). Estas esferas son parte de los parámetros de entrada de la herramienta Dock6.

## **3.5. Descriptores moleculares**

En la siguiente sección se describen las herramientas que serán usadas para la generación de los descriptores moleculares, así como para la interpretación de los mismos y la generación del modelo de aprendizaje de máquina.

### **3.5.1. PyPlif**

Es una herramienta para el cálculo de IFP desarrollada por el Centro de modelado Molecular de Yogyakarta, Indonesia (Radifar *et al.*, 2013). Fue desarrollada usando lenguaje Python y con dependencias de OpenBabel, por lo que es una herramienta libre y de código abierto.

Calcula la IFP al transformar las interacciones moleculares entre el ligando y la proteína en una matriz de bits según los residuos seleccionados y el tipo de interacción. Por cada residuo hay siete bits que representan los siete tipos de interacciones:

1. Apolar (van der Waals).
2. Cara a cara aromática.
3. Borde a cara aromática.
4. Enlace de hidrógeno (proteína como donante).
5. Enlace de hidrógeno (proteína como aceptor).
6. Interacción electrostática (proteína cargada positivamente).
7. Interacción electrostática (proteína cargada negativamente).

Con las matrices de bits de las poses del acoplamiento y de la pose de referencia, verifica la similitud usando el coeficiente de Tanimoto.

Requiere cuatro archivos de entrada: configuración con extensión *txt*, sitio activo de la proteína, el ligando de referencia, ambos en formato *mol2*, y características en formato *csv*. Al terminar la ejecución genera un fichero de salida en formato *csv*.

### **3.5.2. WEKA**

El Ambiente de Waikato para el Análisis del Conocimiento (Waikato Environment for Knowledge Analysis, WEKA), fue desarrollada por el Departamento de Ciencias de la Computación de la Universidad de Waikato, Nueva Zelanda (Hall *et al.*, 2009). Es un sistema desarrollado bajo licencia libre y de código abierto. Es posible implementar y adicionar nuevos algoritmos sin tener que preocuparse por la infraestructura de soporte para la manipulación de datos y la evaluación de esquemas.

Tiene como objetivo proporcionar una colección completa de algoritmos de aprendizaje automático y herramientas de preprocesamiento de datos. Permite probar rápidamente y comparar diferentes métodos de aprendizaje automático en nuevos conjuntos de datos. Su arquitectura modular y extensible permite que se ejecuten experimentos

de minería de datos a partir de su amplia colección de algoritmos de aprendizaje y herramientas incorporadas.

El entorno de trabajo incluye la exploración preliminar de los datos con facilidades de visualización de datos y herramientas de preprocesamiento; incluye también algoritmos de regresión, clasificación, agrupación, extracción de reglas por asociación y selección de atributos. Estas acciones se encuentran en la interfaz de usuario de exploración contenida en los paneles:

**Preproceso:** Son cargados los datos para su posterior transformación haciendo uso de los filtros. Los datos pueden provenir desde varias fuentes: archivos, URLs y bases de datos. Los formatos de archivos soportados son: *ARFF* de WEKA, *csv*, *LibSVM* y *C4.5*. Es posible generar datos de forma artificial y editarlos de forma manual.

**Clasificar:** Se encuentran los algoritmos de clasificación y regresión. Los datos de entrada para el conjunto de entrenamiento son obtenidos de los datos cargados en el Preproceso. Muestra información del poder predictivo del algoritmo de aprendizaje seleccionado sobre el conjunto de entrenamiento y de prueba suministrado; y una representación textual del modelo creado. Brinda acceso a representaciones gráficas de los modelos y visualiza errores de predicción mediante gráficos de dispersión y curvas de umbral. Los modelos pueden ser guardados y cargados en este panel.

**Cluster:** Permite ejecutar algoritmos de agrupación de los datos cargados en el panel de Preproceso. Proporciona estadísticas simples para la evaluación del rendimiento de la agrupación en clústeres. Visualiza la estructura del agrupamiento y los modelos se pueden almacenar de forma persistente de ser necesario.

**Asociar:** Contiene métodos y técnicas minería de asociación de los datos.

**Seleccionar atributos:** Brinda acceso a varios algoritmos y criterios de evaluación para identificar los atributos más importantes en el conjunto de datos cargados en el panel Preproceso. La solidez del conjunto de atributos seleccionados se puede validar a través de un enfoque basado en la validación cruzada. Hay que tener en cuenta que este panel está diseñado para el análisis de datos exploratorios.

**Visualizar:** Proporciona una matriz de gráficos de dispersión codificada por colores, así como la selección de profundizar en la selección de gráficos individuales y porciones de datos. También es posible obtener información de puntos de datos y su perturbación aleatoria para descubrir datos ocultos.

Para la descripción de los datos se tendrán en cuenta:

**Atributes:** muestra la cantidad total de atributos o columnas que contiene el conjunto.

**Instances:** muestra la cantidad total de instancias u objetos en el conjunto.

Para el filtrado de los datos, se tendrá en cuenta el filtro:

**Remove Useless:** es un filtro no supervisado de atributos. Elimina los atributos que no varían mucho, así como aquellos que varían por encima de un porcentaje de varianza máximo. Este filtro solo se aplica a atributos nominales.

Para la descripción de los atributos se tendrán en cuenta:

**Distinct:** muestra la cantidad de valores distintos que tienen los datos para ese atributo.

**Unique:** muestra la cantidad de instancias en los datos que tienen un valor para este atributo que ninguna otra instancia tiene.

Como algoritmos para la predicción, se tendrán en cuenta los siguientes algoritmos:

**Gaussian Processes:** es una función de clasificación que implementa procesos gaussianos para la regresión sin ajuste de hiperparámetros. Para la evaluación correcta del nivel de ruido, aplica la normalización o estandarización de todos los atributos, posteriormente los atributos nominales, son convertidos a binarios. Los valores ausentes se reemplazan por la media o por la moda global.

**Linear Regression:** es una función de clasificación que implementa la regresión lineal para la predicción. Utiliza el criterio de Akaike para la selección de los modelos, y es capaz de predecir instancias ponderadas.

**Multilayer Perceptron:** es una función de clasificación que implementa la propagación hacia atrás para entrenar un perceptrón de múltiples capas para clasificar instancias. La red puede ser construida mediante una heurística simple y sus parámetros pueden ser monitoreados y modificados durante el entrenamiento. Los nodos de salida pueden ser sigmoides o unidades lineales sin umbral.

**SimpleLinearRegression:** es una función de clasificación que implementa la regresión lineal simple. Selecciona el atributo que da como resultados el Error Cuadrático más bajo.

**SMOreg:** es una función de clasificación que implementa la máquina de soporte vectorial para la regresión. Los parámetros se pueden aprender utilizando varios algoritmos.

**IBk:** es un algoritmo de clasificación de los K-vecinos más cercanos. Puede seleccionar el mejor valor de K basado en la validación cruzada y se le puede hacer ponderación de distancias.

**KStar:** es un algoritmo de clasificación basado en instancias. La clase de una instancia de prueba es predicha a partir de la clase de una instancia similar a esta, según una función de similitud. Utiliza una función de ponderación basada en la entropía.

**Decision Table:** es un algoritmo basado en reglas para la clasificación. Construye un clasificador de mayoría a partir de una tabla de decisión simple.

**M5 Rules:** es un algoritmo basado en reglas para la clasificación. Genera una lista de decisiones usando "Divide y vencerás". En cada iteración, construye un árbol modelo usando M5 y convierte la "mejor" hoja en una regla.

**Zero R:** es un algoritmo basado en reglas para la clasificación. Predice, para clases numéricas, la media de las clases del entrenamiento.

**Decision Stump:** es un algoritmo basado en árboles. Hace regresión basado en el Error Cuadrático Medio.

**M5P:** es un algoritmo basado en árboles. Implementa el árbol para el modelo M5'.

**Random Forest:** es un algoritmo basado en árboles. Construye un bosque de árboles aleatorio.

**Random Tree:** es un algoritmo basado en árboles. Construye un árbol que considera K atributos elegidos aleatoriamente para cada nodo. No realiza podas. Realiza la estimación de las clases numéricas a partir de la media.

**REP Tree:** es un algoritmo basado en árboles. Construye un árbol de regresión usando la varianza de la información y lo poda una reducción de errores. Los valores de los atributos numéricos son ordenados una sola vez.

Para mayores detalles sobre Weka y su uso, se recomienda al lector revisar el manual de usuario, disponible en:

<https://www.cs.waikato.ac.nz/ml/weka/documentation.html>.



## Capítulo 4. Resultados

En el presente capítulo se exponen los resultados de los experimentos diseñados para la evaluación de las funciones de puntuación que implementan las diferentes herramientas computacionales de acoplamiento molecular.

### 4.1. Complejos moleculares

En la Tabla 1 mostramos 10 de los complejos presentes en DockinDB. Todos los complejos los podemos encontrar en el Anexo E, Tabla 2.

**Tabla 1.** Descripción general de 10 de las estructuras incluidas en el conjunto de datos de referencia DockingDB.

| PDB id | Cad. prot. | Cad. pépt. | Pépt. sec.                        | Pépt. rot. |
|--------|------------|------------|-----------------------------------|------------|
| 1apm   | E          | I          | TTYADFIASGRTGRRNAIHD              | 112        |
| 1eb1   | H          | A          | DYEPIPEEAF                        | 45         |
| 2aq9   | A          | X          | WMLDPIAGKWSR                      | 57         |
| 2cpk   | E          | I          | TTYADFIASGRTGRRNAIHD              | 106        |
| 2plx   | A          | B          | QCKVMCYAQRHSSPELLRRCLDNCEK        | 82         |
| 3qn7   | A          | B          | ACSRYEVDRCRGRGSAC                 | 83         |
| 5hhv   | AB         | I          | IHVTIPADLWDWIN                    | 64         |
| 5hi4   | AB         | I          | IHVTIPADLWDWIN                    | 64         |
| 5va9   | AB         | C          | YGHSHIRFGYSYHVSYCG** <sup>1</sup> | 41         |
| 5vb9   | A          | C          | CWVLEYDMFGALHCR                   | 34         |

### 4.2. Acoplamiento global

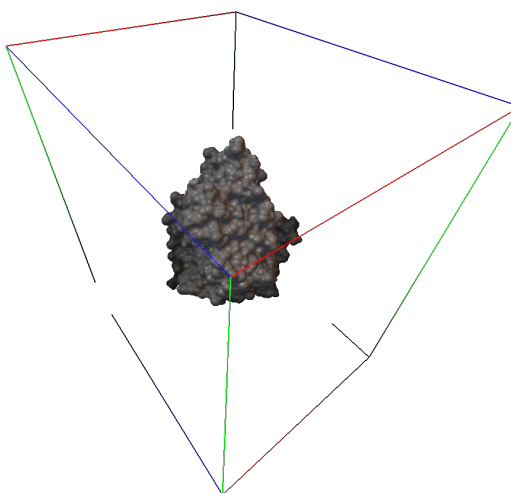
Para realizar el Acoplamiento Molecular global rígido con las herramientas AutoDock4, ClusPro, Dock6 y AutoDock Vina, se seleccionaron los complejos de la Tabla 1.

Haciendo uso de la herramienta AutoDock Tools fueron creados los ficheros de configuración. En el Anexo B, para el Autogrid4, ver Configuración .1; AutoDock4, ver

<sup>1</sup>El '\*\*' representa que no es un amino ácido natural.

Configuración .3; y el AutoDock Vina, ver Configuración .4. Se creó los ficheros de Configuración para el Sphgen, ver Configuración .6; el Showbox, ver Configuración .7; el Grid, ver Configuración .9; y el Dock6, ver Configuración .10. En cada caso un fichero de configuración fue generado para cada herramienta por complejo.

Se construyó una caja para cada complejo según las especificaciones de las herramientas que necesitan una caja para realizar el acoplamiento: AutoDock4, Dock6 y AutoDock Vina; el centro de la caja fue el centro geométrico de la proteína y los lados de la caja se calculó como el mínimo que contenga a la proteína más la máxima longitud de la cadena del péptido dado en Å (ver Figura 10). Haciendo uso de la herramienta UCSF Chimera se generó un fichero por complejo con el formato *dms* como entrada al Sphgen, ver el Anexo A, Formato .5. En cada caso fue generado un fichero de configuración para la herramienta por complejo. Los comandos de ejecución se encuentran en el Anexo E.



**Figura 10.** Representación de la caja para el acoplamiento molecular global del complejo 1apm.

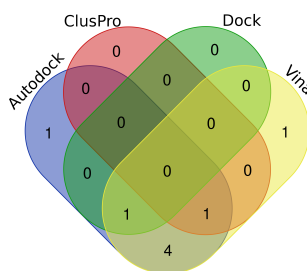
Una vez terminados los cálculos de las herramientas de acoplamiento molecular, se pueden observar los siguientes resultados:

1. Se obtuvieron hasta las 20 mejores poses por cada complejo, por cada herramienta.
2. El AutoDock Vina lanzó un error de memoria excedida en el complejo 5hi4, ver Anexo D, Error .6.

## Análisis según energía

Se ordenaron las poses en forma ascendente según su energía libre. En el Anexo E, Tabla 3 se puede observar la correlación que existe entre las diferentes funciones de energías y el RMSD de las poses calculadas.

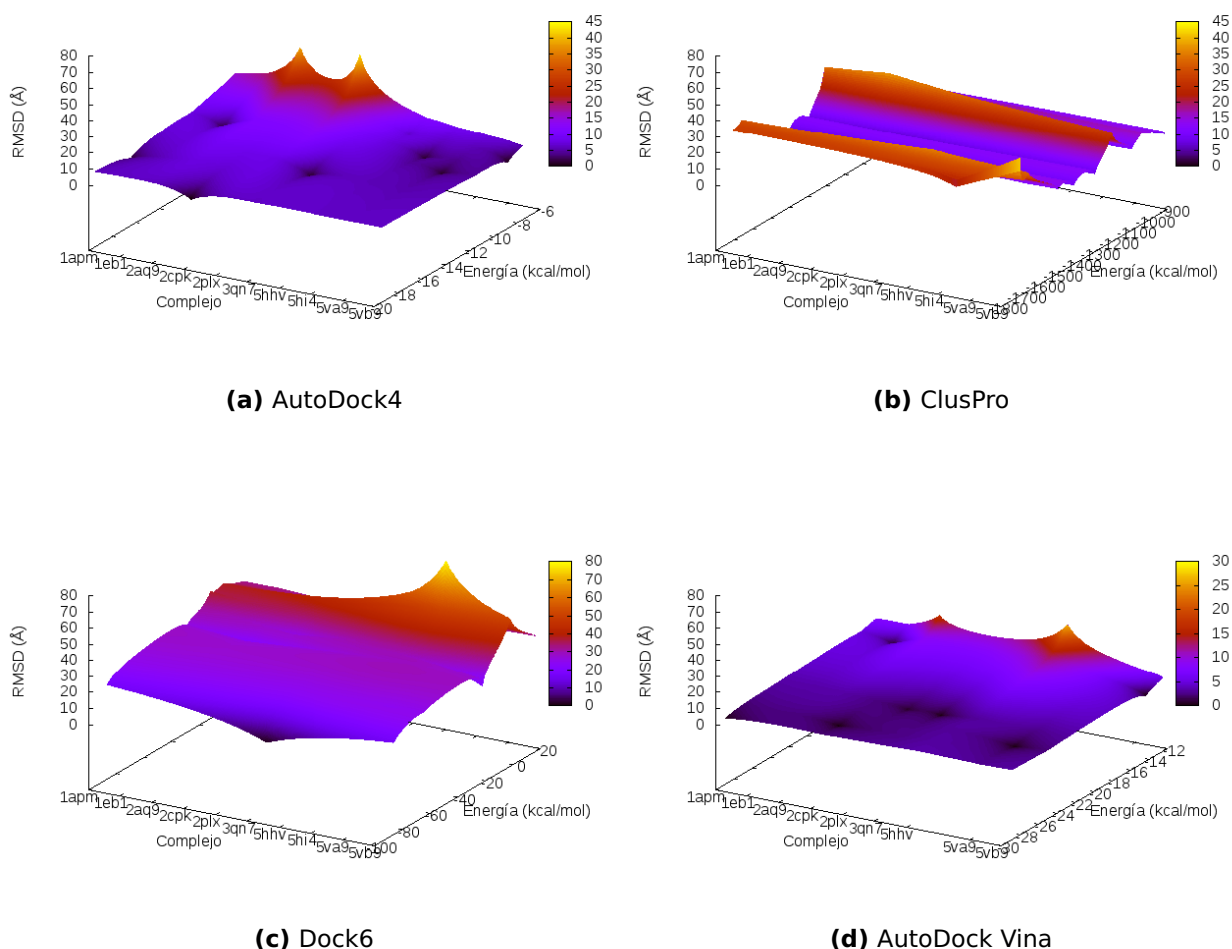
Al comparar los resultados de las cuatro herramientas, el AutoDock Vina tiene la mejor correlación lineal y positiva de 0.6349. El AutoDock4 logró el menor RMSD con la menor de sus energías en el 20% de los complejos, mientras que el AutoDock Vina logró el menor RMSD con la menor de sus energías en el 80% de los restantes complejos; el ClusPro y el Dock6 no lograron menor RMSD con su menor energía. El AutoDock4 logró un  $RMSD \leq 2\text{\AA}$  en el 70% de las poses de menor RMSD, el ClusPro y el Dock6 en el 10%, y el Vina en el 70% de las poses. En la Figura 11 se puede observar la distribución de las poses con  $RMSD \leq 2\text{\AA}$  por cada herramienta de acoplamiento molecular; como se observa hay dos complejos para los cuales no se calcularon poses por debajo de dicho umbral: 2aq9 y 5hhv. En la Figura 12 se puede observar cómo las poses de la menor energía reportan el menor RMSD para las herramientas AutoDock4 (a) y AutoDock Vina (d).



**Figura 11.** Unión de los conjuntos del acoplamiento molecular Global que representan las poses de  $RMSD \leq 2\text{\AA}$  por cada herramienta.

## Análisis según RMSD

Se ordenaron las poses en forma ascendente según su RMSD, se puede comprobar que de entre todas las herramientas, el AutoDock4 logró el menor RMSD en el 20% de las moléculas analizadas y el Autodoc Vina logró el menor RMSD en el 80%; el Dock6 y el ClusPro no lograron el menor RMSD en ninguno de los complejos analizados (ver Figura 13).

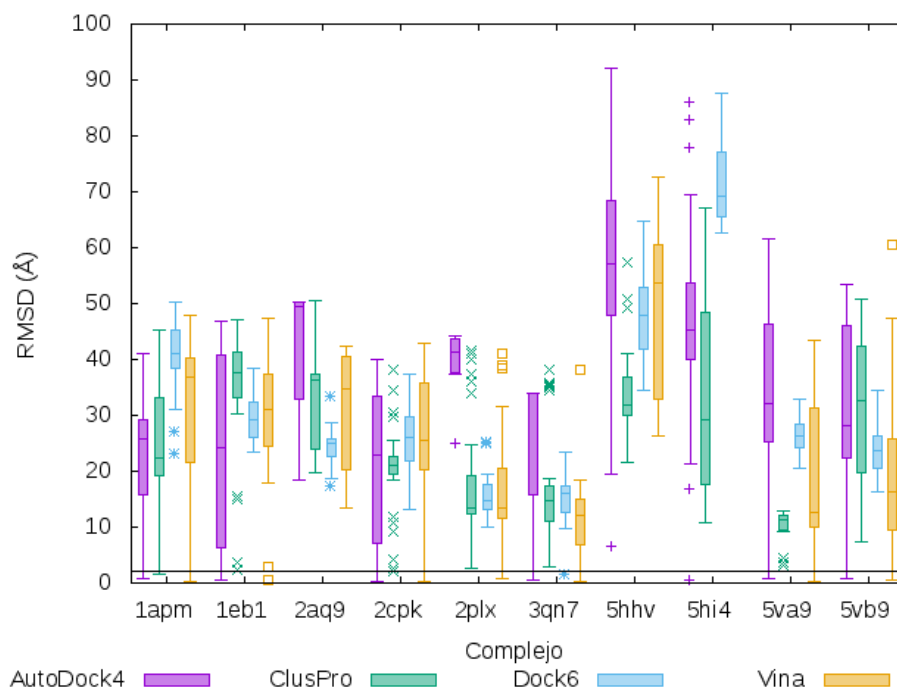


**Figura 12.** Relación entre el RMSD y la energía para el acoplamiento molecular global rígido.

### 4.3. Acoplamiento dirigido rígido

Con los complejos de la Tabla 1, se realizó el acoplamiento molecular dirigido rígido con las herramientas: AutoDock4, Dock6 y AutoDock Vina. ClusPro no fue utilizado por no existir una configuración donde se le especifique la localización del sitio activo de la proteína objetivo.

Al igual que en la sección 4.2 fue creado un fichero de configuración por cada herramienta por complejo. Los comandos de ejecución se encuentran en el Anexo E. Se construyó la caja para cada uno de los complejos según las especificaciones de las herramientas de acoplamiento molecular, el centro de la caja se encuentra ubicado en el centro geométrico de la estructura cristalográfica del péptido y los lados se



**Figura 13.** RMSD del acoplamiento molecular global con estructuras rígidas.

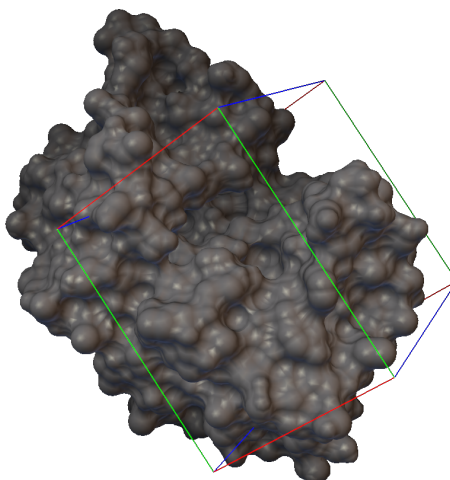
calcularon como la mínima longitud que contenga al péptido más la longitud de la cadena del péptido dado en Å(ver Figura 14).

Una vez terminados los cálculos de las herramientas de acoplamiento molecular, se pueden observar los siguientes resultados:

1. Se obtuvieron hasta las 10 mejores poses por cada complejo, por cada herramienta.
2. El Dock6 no logró poses dentro de la caja definida en el 50% de los complejos analizados.

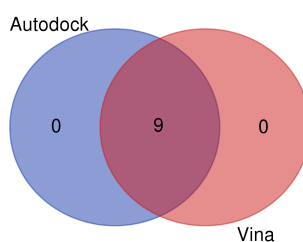
### Análisis según energía

Se ordenaron las poses en forma ascendente según su energía libre. En el Anexo E, Tabla 4 se puede observar la correlación que existe entre las diferentes funciones de energías y el RMSD de las poses calculadas. El Dock6 tiene la mejor correlación lineal y positiva de 0.8283. El AutoDock4 logró un  $RMSD \leq 2\text{Å}$  en el 90%, el Dock6 en el 0%, y el AutoDock Vina en el 90% de las poses. De entre todas las poses, el AutoDock4



**Figura 14.** Representación de la caja para el acoplamiento molecular local del complejo 1apm.

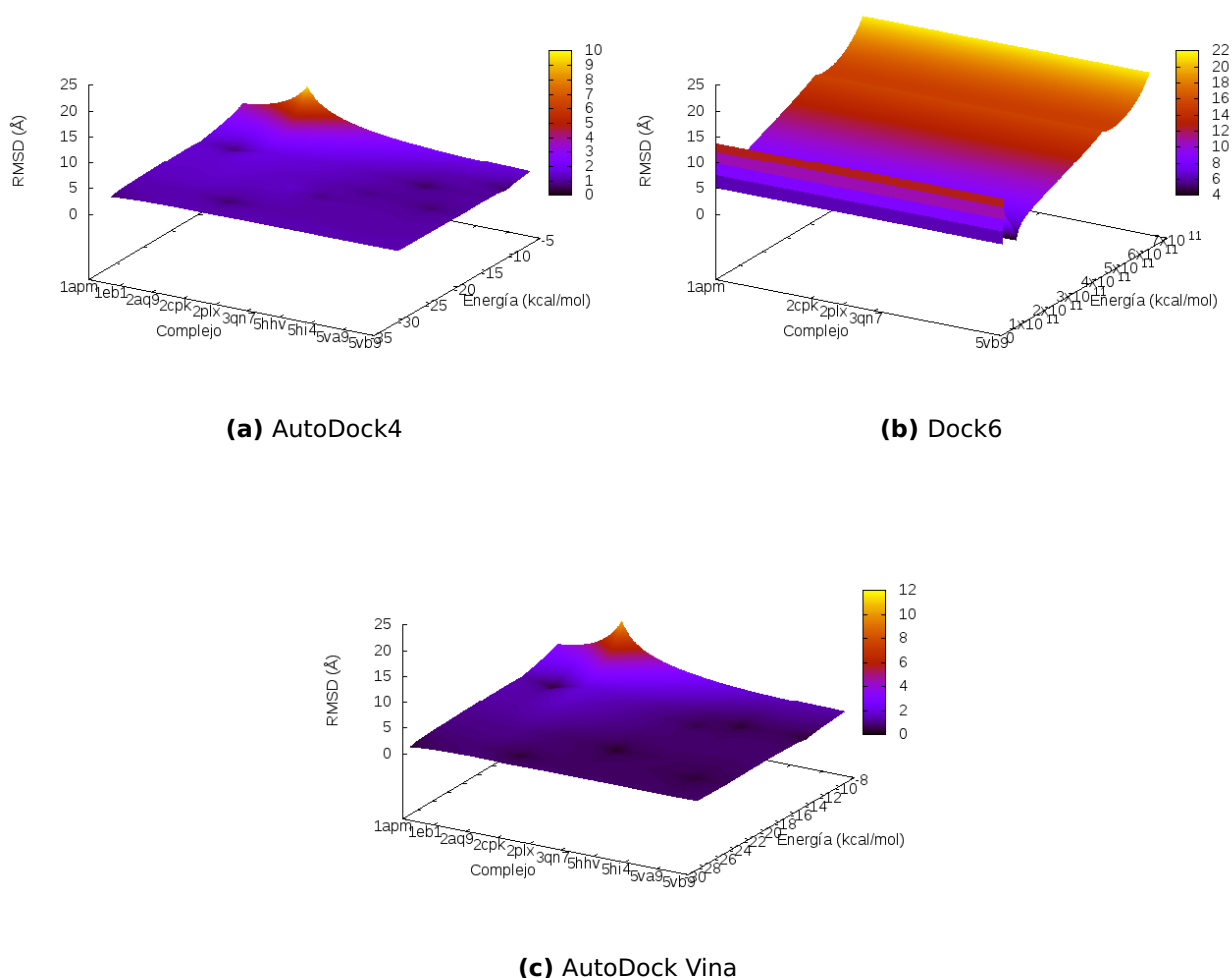
logró el menor RMSD en el 10%, mientras que el AutoDock Vina logró el menor RMSD en el 90%; el Dock6 no logró el menor RMSD con la menor energía. En la Figura 15 se puede observar la distribución de las poses con  $RMSD \leq 2\text{\AA}$  por cada herramienta de acoplamiento molecular; como se observa hay un complejo para los cuales no se calculó poses en dicho intervalo: 2aq9; el Dock6 no logró poses en dichos conjuntos. En la Figura 16 se puede observar cómo las poses de la menor energía reportan el menor RMSD para las herramientas AutoDock4 (a) y AutoDock Vina (c).



**Figura 15.** Unión de los conjuntos del acoplamiento molecular local que representan las poses de  $RMSD \leq 2\text{\AA}$  por cada herramienta.

### Análisis según RMSD

Se ordenaron las poses en forma ascendente según su RMSD, se puede comprobar que la herramienta AutoDock Vina logró el menor RMSD en el 100% de las moléculas analizadas (ver Figura 17).



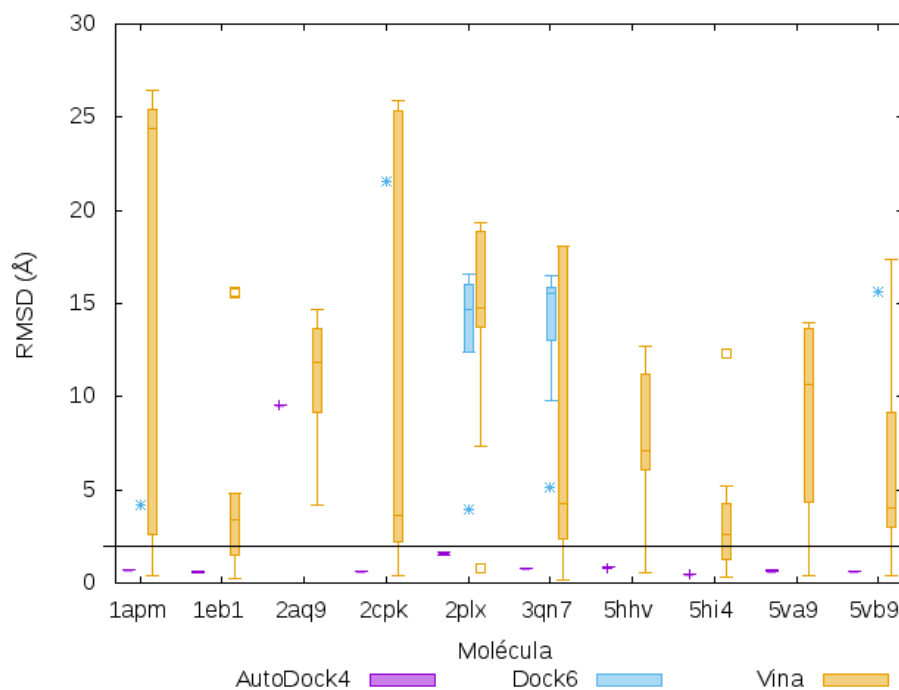
**Figura 16.** Relación entre el RMSD y la energía para el acoplamiento molecular dirigido rígido.

#### 4.4. Acoplamiento dirigido rígido de la DockingDB

Luego de los resultados anteriores se decidió realizar el acoplamiento dirigido rígido para la base de estructuras DockingDB con las herramientas: AutoDock4 y AutoDock Vina. Se decidió no realizar acoplamiento molecular con el Dock6 por sus discretos resultados en los experimentos anteriores.

Haciendo uso de la herramienta AutoDock Tools fueron creados los ficheros de configuración. En el Anexo B, para el Autogrid4, ver Configuración .2; AutoDock4, ver Configuración .3; y para el AutoDock Vina, ver Configuración .5. En cada caso fue generado un fichero de configuración para cada herramienta por complejo.

Se construyó una caja para cada uno de los complejos según las especificaciones



**Figura 17.** RMSD del acoplamiento molecular local con 10 estructuras rígidas.

de las herramientas de acoplamiento molecular, su centro fue localizado en el centro geométrico del péptido y con lado igual al tamaño mínimo que contenga al péptido más la longitud de la cadena del péptido en Å (ver Figura 14). Los comandos de ejecución se encuentran en el Anexo E.

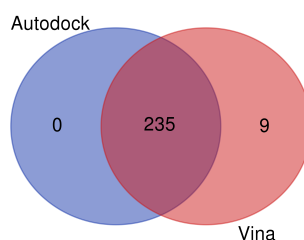
Una vez terminados los cálculos de las herramientas de acoplamiento molecular, se pueden observar los siguientes resultados:

1. Se obtuvieron hasta las 20 mejores poses por cada complejo, por cada herramienta.
2. El AutoDock4 y el AutoDock Vina lanzaron error de sintaxis atómica incorrecta para el átomo de Boro (B), para los tipos de átomos permitidos por AutoDock, en los complejos: 1ot5 y 1r64; ver Anexo D, errores .4 y .5.
3. El AutoGrid4 lanzó el error de receptor desconocido para el tipo (Se) para los complejos: 1n12, 3n55 y 4c2c; ver Anexo D, Error .3.



## Análisis según energía

Se ordenaron las poses en forma ascendente según su energía libre. En el Anexo E, Tabla 5 se puede observar la correlación que existe entre las diferentes funciones de energías y el RMSD de las poses calculadas. El AutoDock4 tiene la mejor correlación lineal y positiva de 0.2212. El AutoDock4 con la menor de sus energías logró un  $RMSD \leq 2\text{\AA}$  en el 94.4 % y el AutoDock Vina en el 96 % de las poses. De entre todas las poses, el AutoDock4 logró el menor RMSD en el 20.9 %, mientras que el AutoDock Vina logró el menor RMSD en el 77.9 % y el 1.2 % restante corresponde a los tres complejos para los cuales las herramientas reportaron errores. En la Figura 18 se puede observar la distribución de las poses con  $RMSD \leq 2\text{\AA}$  por cada herramienta de acoplamiento molecular; como se observa hay cinco complejos para los cuales no se calcularon poses en dicho intervalo: 2ot5, 1r64, 1tg1, 1xoc, 2aq9.



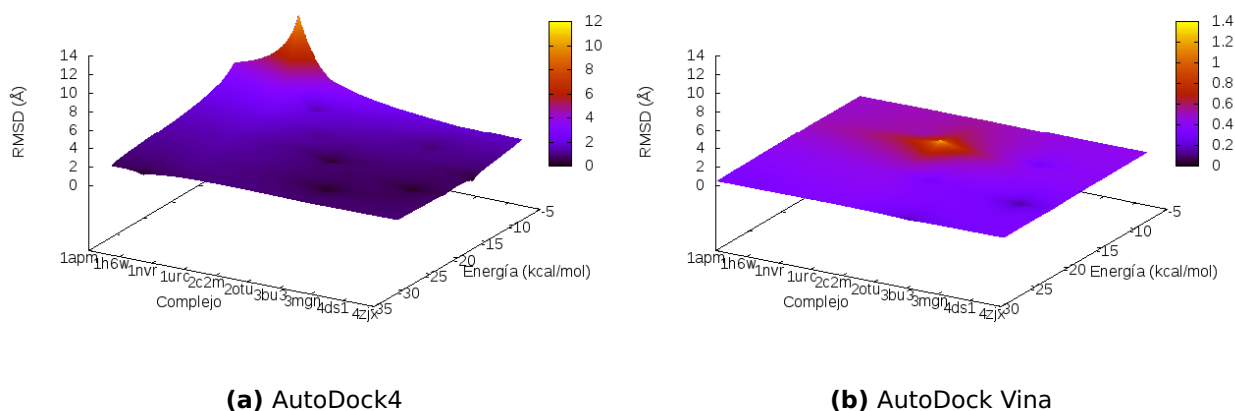
**Figura 18.** Unión de los conjuntos del acoplamiento molecular local que representan las poses de  $RMSD \leq 2\text{\AA}$  por cada herramienta en la DokingDB.

Según el orden mostrado en el Anexo E, Tabla 2, se seleccionó una muestra de diez complejos: 1apm, 1h6w, 1nvr, 1urc, 2c2m, 2otu, 3bu3, 3mgn, 4ds1 y 4zjx. En la Figura 19(b) se puede observar en la muestra seleccionada como las poses de la menor energía reportan el menor RMSD para la herramienta AutoDock Vina.

## Análisis según RMSD

Se ordenaron las poses en forma ascendente según su RMSD, se puede comprobar que el AutoDock4 logró el menor RMSD en el 19.7 %, el Autodoc Vina logró el menor RMSD en el 79.1 % y el 1.2 % fueron las moléculas analizadas que produjeron errores.

Según el orden mostrado en el Anexo E (Tabla 2), se seleccionó una muestra de 25 complejos: 1apm, 1d4t, 1g37, 1i8k, 1kl3, 1nvr, 1se0, 1u00, 1ymt, 2aoi, 2c2o, 2fmf, 2jam, 2p54, 2vj0, 3bu3, 3ds4, 3lnj, 3obq, 3s3j, 4ds1, 4ktx, 4whh, 5hhv y 5uml. En la



**Figura 19.** Relación entre el RMSD y la energía para el acoplamiento molecular dirigido rígido de una muestra de la DockingDB.

Figura 20 se puede observar en la muestra seleccionada como las poses de menor RMSD son para la herramienta AutoDock Vina.

Como se ha descrito con anterioridad el acoplamiento molecular rígido es utilizado principalmente en moléculas pequeñas, de pocos átomos, y como estamos realizando el estudio con péptidos, en moléculas de un gran número de átomos, se decide realizar el acoplamiento dirigido flexible.

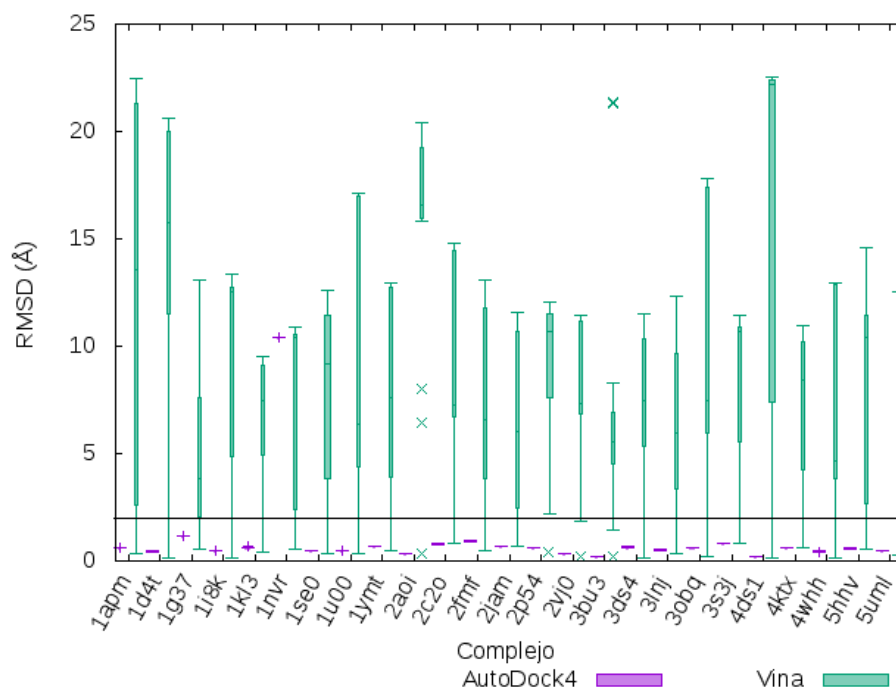
#### 4.5. Acoplamiento dirigido flexible de la DockingDB

Se decidió realizar el acoplamiento dirigido flexible para los complejos de la base de estructuras DockingDB con un número de rotaciones menores o iguales a 32, con las herramientas: AutoDock4 y AutoDock Vina. Se decidió no realizar acoplamiento molecular con el resto de los complejos porque no son permitidos por las herramientas de acoplamiento molecular.

Todos los parámetros de configuración y ejecución fueron los mismos que en la sección 4.4.

Una vez terminados los cálculos de las herramientas de acoplamiento molecular, al igual que en 4.4, se pueden observar los siguientes resultados:

1. Se obtuvieron hasta las 20 mejores poses por cada complejo, por cada herramien-



**Figura 20.** RMSD del acoplamiento molecular dirigido rígido de una muestra de la DockingDB.

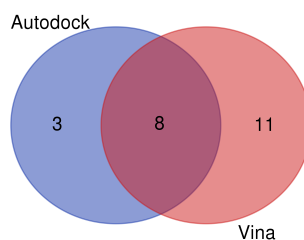
ta.

2. El AutoDock4 y el AutoDock Vina lanzaron error de sintaxis atómica incorrecta para el átomo de Boro (B), para los tipos de átomos permitidos por AutoDock, en los complejos: 1ot5; ver Anexo D, errores .4 y .5.
3. El AutoGrid4 lanzó el error de receptor desconocido para el tipo (Se) para los complejos: 4c2c; ver Anexo D, Error .3.

### Análisis según energía

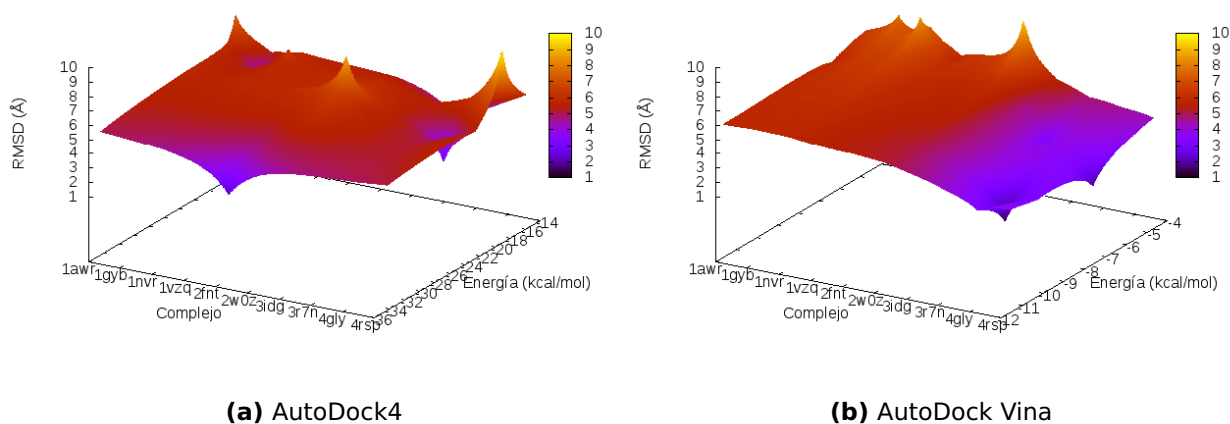
Se ordenaron las poses en forma ascendente según el valor de su energía libre. En el Anexo E, Tabla 6 se puede observar la correlación que existe entre las diferentes funciones de energías y el RMSD de las poses calculadas. El AutoDock Vina tiene la mejor correlación lineal y positiva de 0.3849. El AutoDock4 con la menor de sus energías logró un  $RMSD \leq 2\text{Å}$  en el 10.2% y el AutoDock Vina en el 17.6% de las poses. De entre todas las poses, el AutoDock4 logró el menor RMSD en el 41.7%, mientras que el AutoDock Vina logró el menor RMSD en el 54.6% y el 3.7% restante corresponde al complejo para el cual las herramientas reportaron errores. En la Figura 21 se puede

observar la distribución de las poses con  $RMSD \leq 2\text{\AA}$  por cada herramienta de acoplamiento molecular; como se observa hay 86 complejos para los cuales no se calcularon poses en dicho intervalo.



**Figura 21.** Unión de los conjuntos del acoplamiento molecular flexible local que representan las poses de  $RMSD \leq 2\text{\AA}$  por cada herramienta en la DockingDB.

Según el orden mostrado en el Anexo E, Tabla 2, se seleccionó una muestra de diez complejos: 1awr, 1gyb, 1nvr, 1vzq, 2fnt, 2w0z, 3idg, 3r7n, 4gly y 4rsp. En la Figura 22(b) se puede observar en la muestra seleccionada cómo las poses de la menor energía reportan el menor RMSD para la herramienta AutoDock Vina.



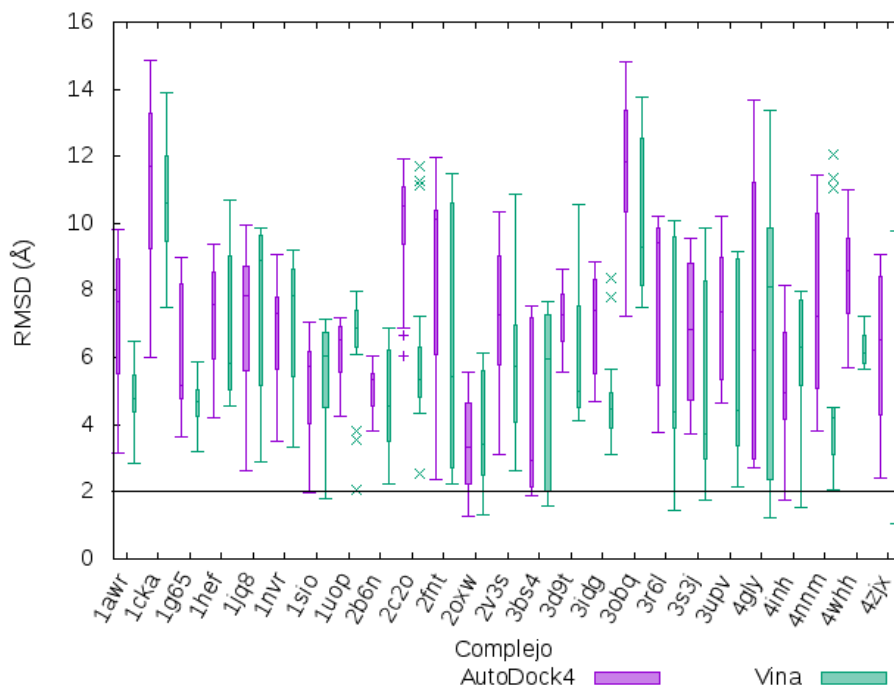
**Figura 22.** Relación entre el RMSD y la energía para el acoplamiento molecular dirigido rígido de una muestra de la DockingDB.

### Análisis según RMSD

Se ordenaron las poses en forma ascendente según su RMSD, se puede comprobar que el AutoDock4 logró el menor RMSD en el 24% y el AutoDock Vina logró el menor RMSD en el 75% y el 1% corresponde a errores de las moléculas analizadas.

Se seleccionó una muestra de 25 complejos: 1awr, 1cka, 1g65, 1hef, 1jq8, 1nvr, 1sio, 1uop, 2b6n, 2c2o, 2fnt, 2oxw, 2v3s, 3bs4, 3d9t, 3idg, 3obq, 3r6l, 3s3j, 3upv, 4gly,

4inh, 4nm, 4whh, 4zjx. En la Figura 23 se puede observar en la muestra seleccionada cómo las poses de menor RMSD son para la herramienta AutoDock Vina.



**Figura 23.** RMSD del acoplamiento molecular dirigido flexible de una muestra de los complejos con menos de 33 rotámetros de la DockingDB.

### Validación cruzada

Se tomaron las poses rígidas y flexibles de menor energía del AutoDock4 para cada complejo y se le calculó la energía con el AutoDock Vina. Se compararon los valores de energía de las poses y resultó en que las poses de menor energía del AutoDock Vina obtuvieron, en el 100% de los casos, una menor energía que las poses del AutoDock4 recalculadas con AutoDock Vina.

Se realizó el mismo experimento anterior, esta vez tomando las poses de menor energía del AutoDock Vina para cada complejo y calculando la energía con el AutoDock4. Se compararon los valores de energía de las poses y resultó que las poses rígidas de menor energía del AutoDock4 obtuvieron, en el 100% de los casos, una menor energía que las poses del AutoDock Vina recalculadas con AutoDock4; sin embargo, las poses flexibles del AutoDock Vina, recalculadas con AutoDock4, obtuvieron, en el 100% de los casos, una menor energía que las poses del AutoDock4.

#### 4.6. Cálculo de SILIRID

Para la obtención de los SILIRID para cada pose obtenida en las sub secciones 4.4 y 4.5, se realizó los cálculos de los SPLIF correspondientes a cada pose y posteriormente se realizó la suma de cada SPLIF en SILIRID. Los identificadores de las poses fueron nombrados de la siguiente forma: pdbID\_entry\_complejoID\_conf\_poseID; donde:

**pdbID:** es el identificador del PDB.

**entry:** palabra reservada.

**complejoID:** es un número entero mayor que 0 que representa el orden del complejo en la base de datos, ver Anexo E, Tabla 2.

**conf:** palabra reservada.

**poseID:** es un número entero mayor o igual a 0 que representa el orden de la pose. La pose con poseID igual a 0 es la pose obtenida del cristal. Las poses con poseID mayor que 0 y menor que 21 corresponden a las poses obtenidas con el AutoDock 4. Las poses con poseID mayor que 20 corresponden a las poses obtenidas con el AutoDock Vina.

##### 4.6.1. Obtención de los Descriptores Moleculares

Los SPLIF fueron generados haciendo uso de la herramienta PyPlif, para la misma, se construyó un archivo de configuración para cada uno de los complejos, ver Anexo B, Configuración .11, y el comando de ejecución se encuentra en el Anexo C. Para cada complejo se obtuvo un fichero que contiene los SPLIF de sus poses. Con los SPLIF de las poses de los complejos, se procedió a calcular los SILIRID de las poses. De los aminoácidos presentes en los sitios activos de los receptores, reportaron SILIRID los 20 aminoácidos más comunes en la naturaleza más los aminoácidos no comunes con los siguientes códigos del PDB: PTR, LLP, TPO, CME, MSE y CSO.

La base de datos de descriptores moleculares quedaría dispuesta por la columna Conf Name, contiene el nombre de las poses según el formato definido en la sección SILIRID; las columnas Estimated Free Energy Binding, Final Intermolecular Energy, vdW

Hbond desolv Energy, Electrostatic Energy, Final Total Internal Energy, Torsional Free Energy y Unbound Systems Energy, correspondientes a las energías obtenidas de AutoDock4; las columnas Affinity, gauss1, gauss2, repulsion, hydrophobic y Hydrogen, correspondientes a las energías obtenidas de Vina; las columnas AA\_# que contiene los SILIRID; por último se agregó la columna rmsd, donde se almacena el RMSD de cada pose. Las filas corresponden a cada pose arrojada por el cristal y las herramientas AutoDock4 y AutoDock Vina. Los datos quedan distribuidos en 197 atributos (columnas) y 8603 casos (filas) para las poses de acoplamiento molecular rígido; y 168 atributos y 4324 casos para las poses de acoplamiento molecular flexible.

Se decidió eliminar las columnas de los SILIRID que todos sus valores fueran 0. Al realizar un conteo de los SILIRID que sus valores son mayores que 0, tenemos que:

**Apolar (van der Waals)** aportó 23 columnas.

**Cara a cara aromática** aportó 3 columnas.

**Borde a cara aromática** aportó 3 columnas.

**Enlace de hidrógeno (proteína como donante)** aportó 20 columnas.

**Enlace de hidrógeno (proteína como aceptor)** aportó 22 columnas.

**Interacción electrostática (proteína cargada positivamente)** aportó 5 columnas.

**Interacción electrostática (proteína cargada negativamente)** aportó 7 columnas.

Por lo que se puede decir que los SILIRID que más información aportan son los: cara a cara aromática, borde a cara aromática; mientras que los más comunes son los: apolar (van der Waals), enlace de hidrógeno (proteína como aceptor).

Se decidió dividir la base de datos en dos conjuntos: entrenamiento y prueba. A continuación se describe el procedimiento seguido para la obtención de ambos conjuntos a partir de la base de datos de descriptores moleculares.

## Conjuntos de entrenamiento

A partir del orden descrito en el Anexo E, Tabla 2, se decidió tomar como conjunto de entrenamiento todas las poses que corresponden con  $a_k | (0 < k < 250) \wedge k \bmod 5 \neq 0$ , donde  $a$  corresponde a un complejo y  $k$  corresponde a la posición del complejo en la base de datos, para un subtotal de 200 complejos y el 80.32 % del total. Posteriormente se aplicó el filtro, del Weka, no supervisado para atributos: Remove Useless; con un porcentaje de varianza por encima del 99 %.

## Conjuntos de prueba

A partir del orden descrito en el Anexo E, Tabla 2, se decidió tomar como conjunto de prueba todas las poses que corresponden con  $a_k | (0 < k < 250) \wedge k \bmod 5 = 0$ , donde  $a$  corresponde a un complejo y  $k$  corresponde a la posición del complejo en la base de datos, para un subtotal de 49 complejos y el 19.68 % del total. No se le aplicaron filtros ya que los datos en este conjunto deben quedar en las mismas condiciones en las que se obtienen de las herramientas de acoplamiento molecular y de descriptores moleculares.

### 4.7. Obtención de los modelos computacionales para las poses rígidas

Por el origen de las poses, se diseñaron tres tipos de pruebas, para estudiar el poder de predicción de RMSD de cada una de las herramientas de acoplamiento molecular de poses rígidas.

#### 4.7.1. Poses del AutoDock4 y del Vina predicen poses del AutoDock4 y del Vina

En esta sección se decidió comprobar el poder de regresión que tienen las poses obtenidas del AutoDock4 y del AutoDock Vina para predecir qué RMSD tendría otras poses obtenidas del AutoDock4 y del AutoDock Vina.

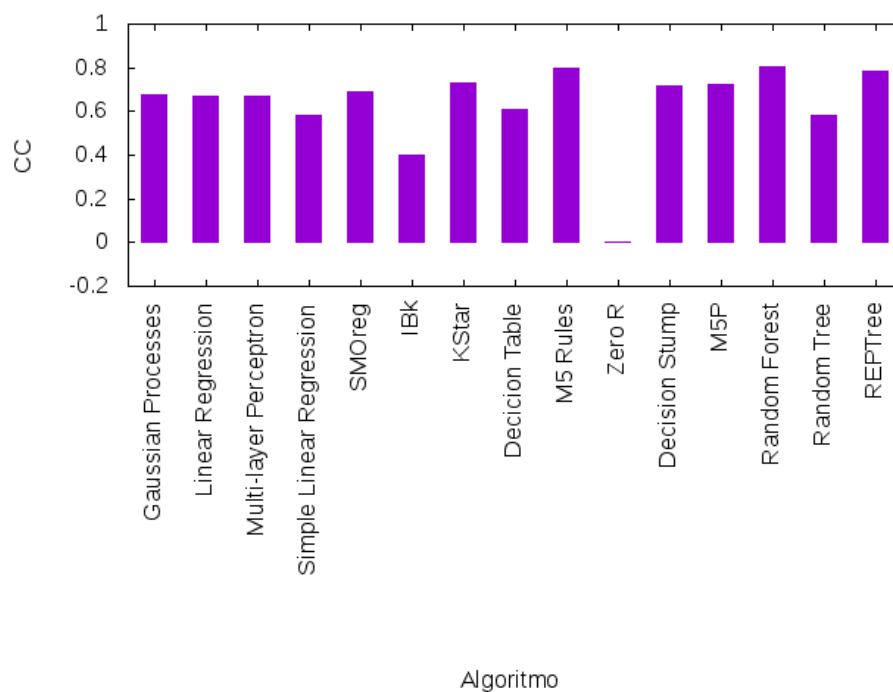
Se seleccionaron todas las poses en el conjunto de entrenamiento, con 89 atributos y 6900 casos. Los valores de RMSD siguen una distribución de 5725 valores distintos



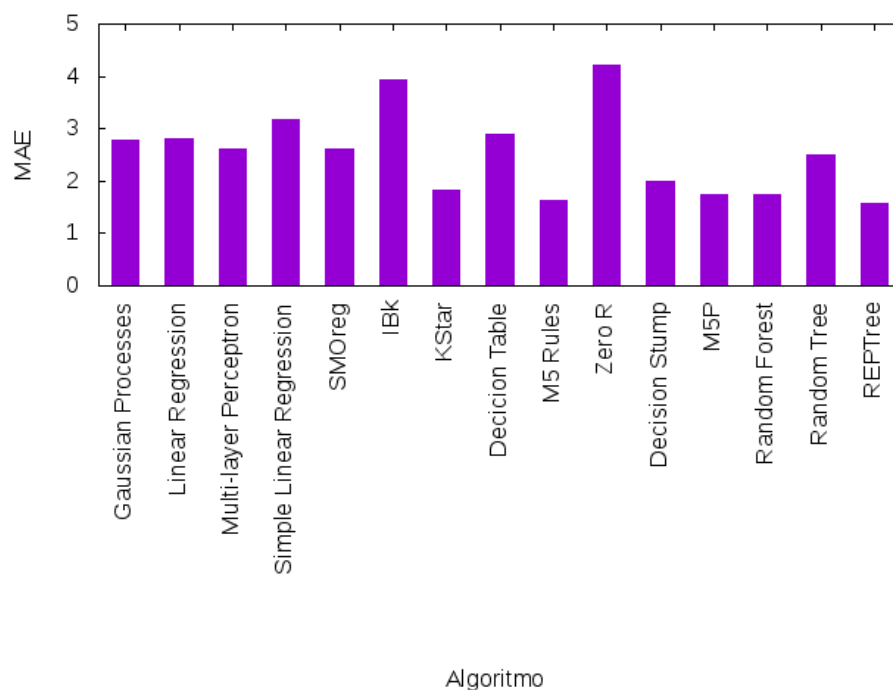
y 4952 de ellos únicos (72 %); un valor mínimo de 0Å, máximo de 28.154Å, una media de 4.018Å y una desviación estándar de 5.294Å.

Se seleccionaron todas las poses en el conjunto de prueba, con 197 atributos y 1703 casos. La diferencia en la cantidad de atributos entre entrenamiento y prueba, viene dada por el filtro de selección aplicado al conjunto de entrenamiento, Weka realiza el mapeo de los atributos según el nombre de estos, de esta forma son utilizados los atributos del conjunto de prueba que coincidan con los usados en el conjunto de entrenamiento. Los valores de RMSD siguen una distribución de 1547 valores distintos y 1446 de ellos únicos (85 %); un valor mínimo de 0Å, máximo de 58.931Å, una media de 3.792Å y una desviación estándar de 5.31Å.

Una vez realizados los cálculos de regresión, ver Anexo E, Tabla 7, resultó que el algoritmo que mejor coeficiente de correlación aportó fue el Random Forest con 0.802 y un error absoluto medio de 1.7367Å (ver Figura 24); mientras que el algoritmo de menor error absoluto medio fue el REPTree con 1.5665Å y un coeficiente de correlación de 0.785 (ver Figura 25).



**Figura 24.** Coeficiente de correlación de los algoritmos de predicción con las poses del AutoDock4 y del Vina prediciendo poses del AutoDock4 y del Vina.



**Figura 25.** Error absoluto medio de los algoritmos de predicción con las poses del AutoDock4 y del Vina prediciendo poses del AutoDock4 y del Vina.

#### 4.7.2. Poses del AutoDock4 predicen poses del AutoDock4

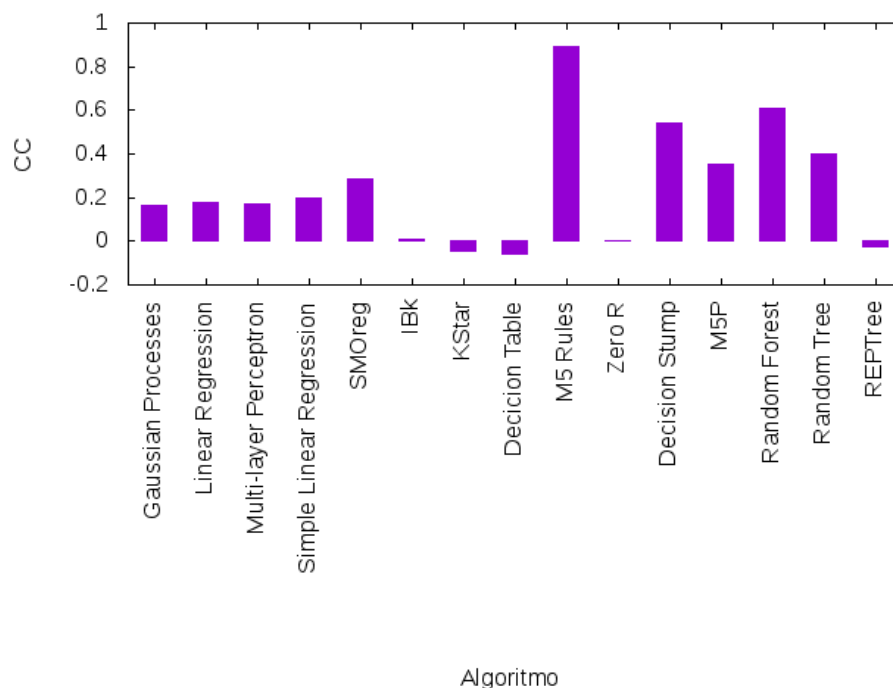
En esta sección se decidió comprobar el poder de regresión que tienen las poses obtenidas del AutoDock4 para predecir qué RMSD tendrían otras poses obtenidas del AutoDock4.

Del conjunto de entrenamiento, se seleccionaron las poses que fueron obtenidas a partir del AutoDock4, con 81 atributos y 3900 casos. Los valores de RMSD siguen una distribución de 3000 valores distintos y 2299 de ellos únicos (59%); un valor mínimo de 0.136Å, máximo de 18.539Å, una media de 1.035Å y una desviación estándar de 2.397Å.

Del conjunto de prueba, se seleccionaron las poses que fueron obtenidas a partir del AutoDock4, con 197 atributos y 960 casos. El conjunto de prueba contiene más atributos porque no fueron filtrados como el conjunto de entrenamiento, Weka realiza el mapeo de los atributos según el nombre de estos, de esta forma son utilizados los atributos del conjunto de prueba que coincidan con los usados en el conjunto de entrenamiento. Los valores de RMSD siguen una distribución de 859 valores distintos y 767 de ellos únicos (80%); un valor mínimo de 0.249Å, máximo de 6.943Å, una media

de 0.731Å y una desviación estándar de 0.952Å.

Una vez realizados los cálculos de regresión, ver Anexo E, Tabla 7, resultó que el algoritmo que mejor coeficiente de correlación aportó fue el M5 Rules con 0.8929 y un error absoluto medio de 0.422Å (ver Figura 26); mientras que el algoritmo de menor error absoluto medio fue el SMOreg con 0.3537Å y un coeficiente de correlación de 0.2883 (ver Figura 27).



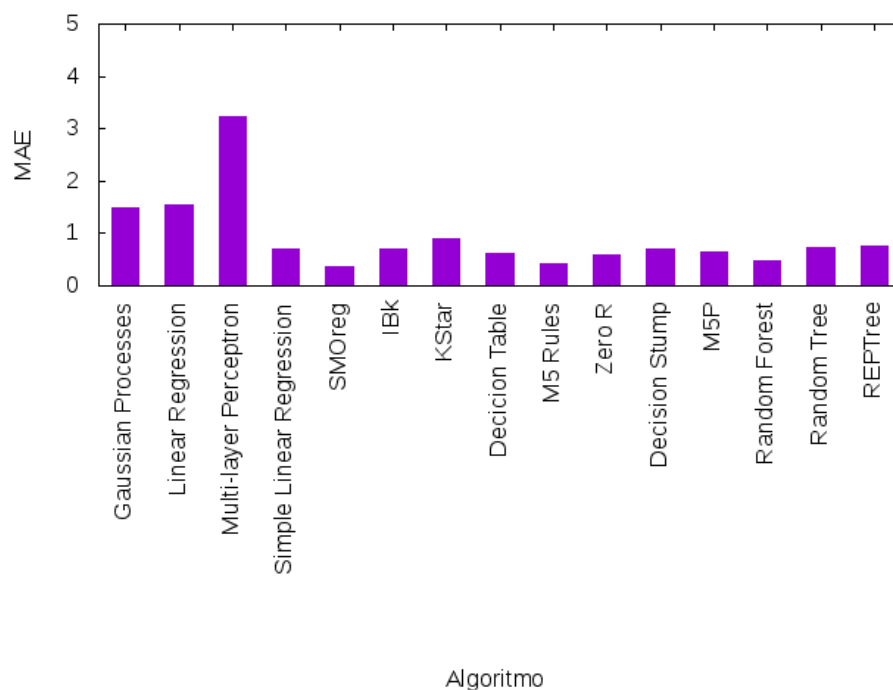
**Figura 26.** Coeficiente de correlación de los algoritmos de predicción con las poses del AutoDock4 prediciendo poses del AutoDock4.

#### 4.7.3. Poses del AutoDock Vina predicen poses del AutoDock Vina

En esta sección se decidió comprobar el poder de regresión que tienen las poses obtenidas del AutoDock Vina para predecir qué RMSD tendrían otras poses obtenidas del AutoDock Vina.

Del conjunto de entrenamiento, se seleccionaron las poses que fueron obtenidas a partir del AutoDock Vina, con 89 atributos y 2805 casos. Los valores de RMSD siguen una distribución de 2780 valores distintos y 2755 de ellos únicos (98%); un valor mínimo de 0.057Å, máximo de 28.154Å, una media de 8.444Å y una desviación estándar de 5.281Å.

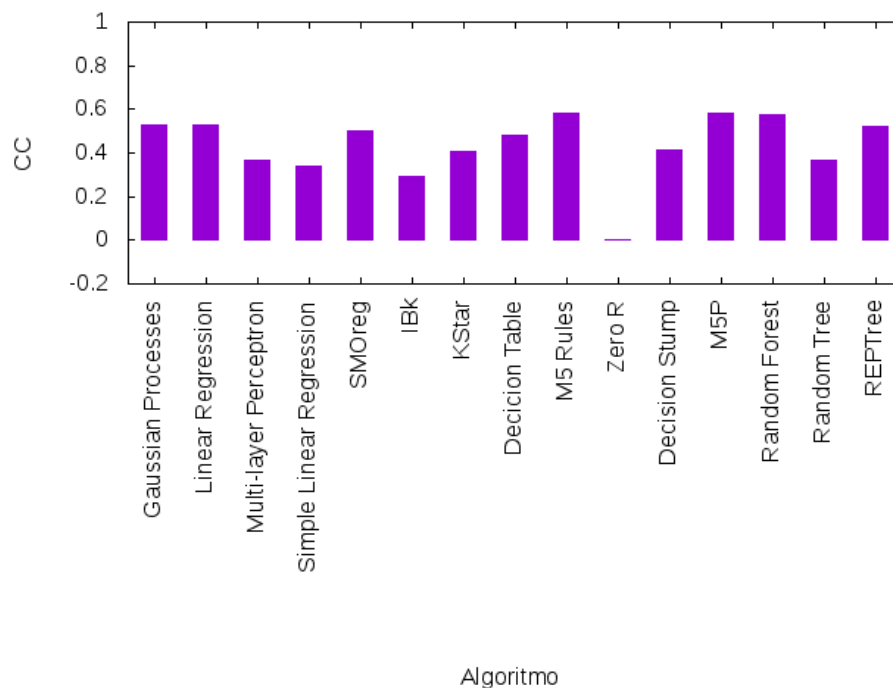
Del conjunto de prueba, se seleccionaron las poses que fueron obtenidas a partir



**Figura 27.** Error absoluto medio de los algoritmos de predicción con las poses del AutoDock4 prediciendo poses del AutoDock4.

del AutoDock Vina, con 197 atributos y 695 casos. El número de atributos varía del conjunto de entrenamiento al de prueba porque se le aplicaron filtros al conjunto de entrenamiento, disminuyendo de esta forma la cantidad de atributos, Weka realiza el mapeo de los atributos según el nombre de estos, de esta forma son utilizados los atributos del conjunto de prueba que coincidan con los usados en el conjunto de entrenamiento. Los valores de RMSD siguen una distribución de 692 valores distintos y 689 de ellos únicos (99%); un valor mínimo de  $0.073\text{\AA}$ , máximo de  $58.931\text{\AA}$ , una media de  $8.281\text{\AA}$  y una desviación estándar de  $5.81\text{\AA}$ .

Una vez realizados los cálculos de regresión, ver Anexo E, Tabla 7, resultó que el algoritmo que mejor coeficiente de correlación aportó fue el M5P con 0.5796 y un error absoluto medio de  $3.1674\text{\AA}$  (ver Figura 28); mientras que el algoritmo de menor error absoluto medio fue el Random Forest con  $3.1616\text{\AA}$  y un coeficiente de correlación de 0.5722 (ver Figura 29).



**Figura 28.** Coeficiente de correlación de los algoritmos de predicción con las poses del AutoDock Vina prediciendo poses del AutoDock Vina.

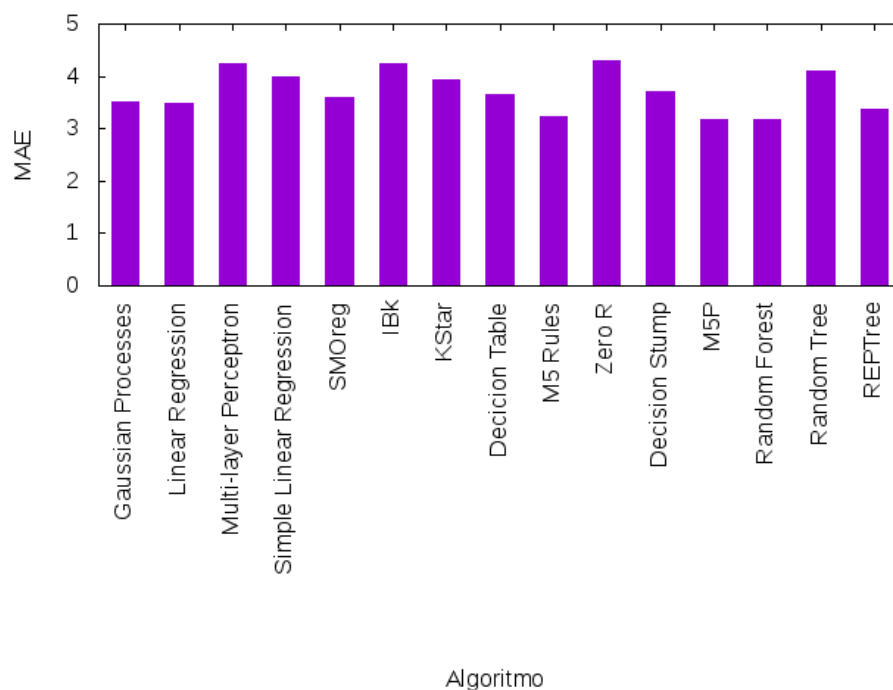
#### 4.8. Obtención de los modelos computacionales para las poses flexibles

Al igual que en la sección anterior, se decidió estudiar el poder de predicción de RMSD de cada una de las herramientas de acoplamiento molecular para poses flexibles.

##### 4.8.1. Poses del AutoDock4 y del Vina predicen poses del AutoDock4 y del Vina

Se seleccionaron todas las poses en el conjunto de entrenamiento, con 88 atributos y 3390 casos. Los valores de RMSD siguen una distribución de 3270 valores distintos y 3231 de ellos únicos (95%); un valor mínimo de 0Å, máximo de 15.618Å, una media de 6.707Å y una desviación estándar de 3.06Å.

Se seleccionaron todas las poses en el conjunto de prueba, con 169 atributos y 934 casos. La diferencia en la cantidad de atributos entre entrenamiento y prueba, viene dada por el filtro de selección aplicado al conjunto de entrenamiento, Weka realiza el mapeo de los atributos según el nombre de estos, de esta forma son utilizados



**Figura 29.** Error absoluto medio de los algoritmos de predicción con las poses del AutoDock Vina prediciendo poses del AutoDock Vina.

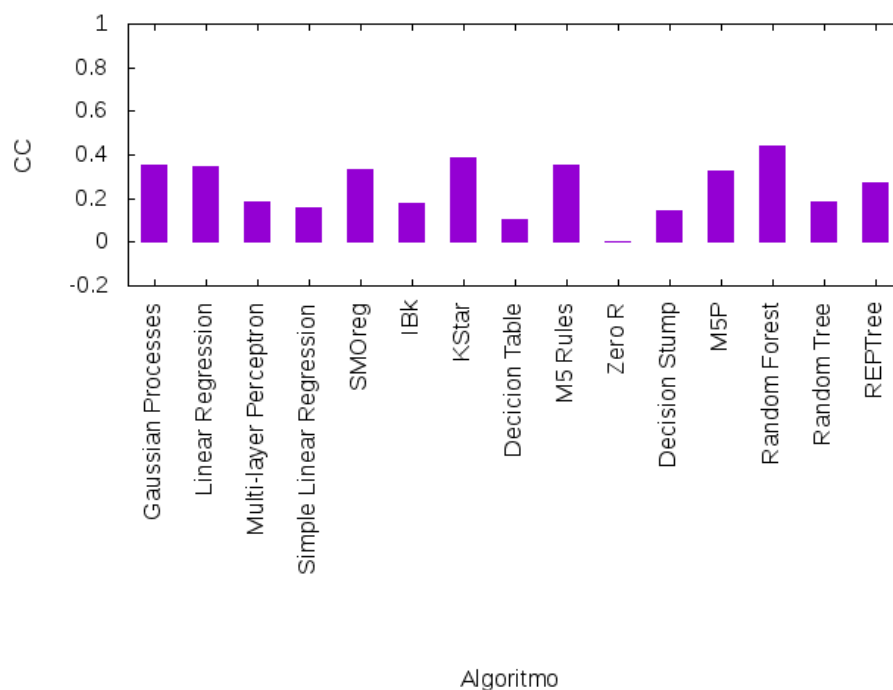
los atributos del conjunto de prueba que coincidan con los usados en el conjunto de entrenamiento. Los valores de RMSD siguen una distribución de 909 valores distintos y 905 de ellos únicos (97%); un valor mínimo de  $0\text{\AA}$ , máximo de  $13.736\text{\AA}$ , una media de  $6.155\text{\AA}$  y una desviación estándar de  $2.785\text{\AA}$ .

Una vez realizados los cálculos de regresión, ver Anexo E, Tabla 8, resultó que el algoritmo que mejor coeficiente de correlación aportó fue el Random Forest con 0.4415, coincidiendo con el menor error absoluto medio de  $2.0844\text{\AA}$  (ver figuras 30 y 31).

#### 4.8.2. Poses del AutoDock4 predicen poses del AutoDock4

Del conjunto de entrenamiento, se seleccionaron las poses que fueron obtenidas a partir del AutoDock4, con 88 atributos y 1660 casos. Los valores de RMSD siguen una distribución de 1654 valores distintos y 1648 de ellos únicos (99%); un valor mínimo de  $0.981\text{\AA}$ , máximo de  $15.458\text{\AA}$ , una media de  $7.194\text{\AA}$  y una desviación estándar de  $2.709\text{\AA}$ .

Del conjunto de prueba, se seleccionaron las poses que fueron obtenidas a partir del AutoDock4, con 169 atributos y 460 casos. El conjunto de prueba contiene más



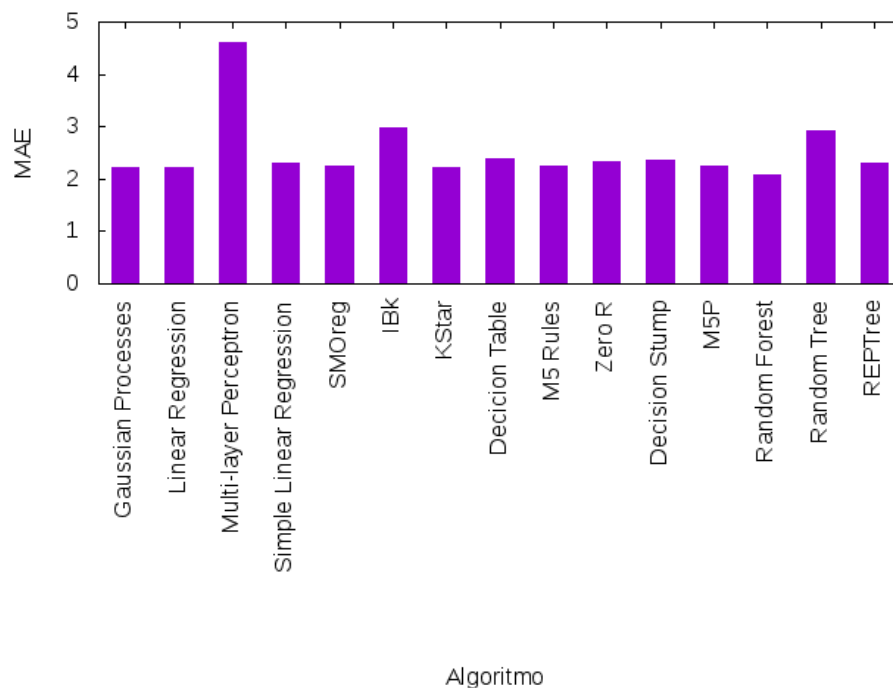
**Figura 30.** Coeficiente de correlación de los algoritmos de predicción con las poses flexibles del AutoDock4 y del Vina prediciendo poses del AutoDock4 y del Vina.

atributos porque no fueron filtrados como el conjunto de entrenamiento, Weka realiza el mapeo de los atributos según el nombre de estos, de esta forma son utilizados los atributos del conjunto de prueba que coincidan con los usados en el conjunto de entrenamiento. Los valores de RMSD siguen una distribución de 460 valores distintos y 460 de ellos únicos (100%); un valor mínimo de 1.254Å, máximo de 12.73Å, una media de 6.614Å y una desviación estándar de 2.517Å.

Una vez realizados los cálculos de regresión, ver Anexo E, Tabla 8, resultó que el algoritmo que mejor coeficiente de correlación aportó fue el Random Forest con 0.418 y un error absoluto medio de 1.9316Å (ver Figura 32); mientras que el algoritmo de menor error absoluto medio fue el M5 Rules con 1.8834Å y un coeficiente de correlación de 0.3684 (ver Figura 33).

#### 4.8.3. Poses del AutoDock Vina predicen poses del AutoDock Vina

Del conjunto de entrenamiento, se seleccionaron las poses que fueron obtenidas a partir del AutoDock Vina, con 85 atributos y 1647 casos. Los valores de RMSD siguen una distribución de 1635 valores distintos y 1623 de ellos únicos (99%); un valor míni-



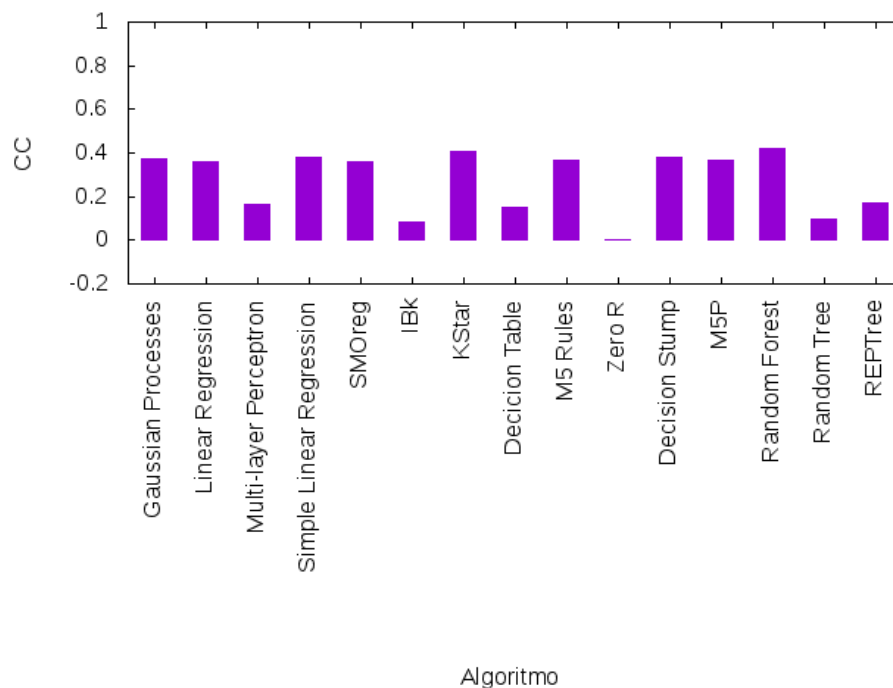
**Figura 31.** Error absoluto medio de los algoritmos de predicción con las poses flexibles del AutoDock4 y del Vina prediciendo poses del AutoDock4 y del Vina.

mo de 0.421Å, máximo de 15.618Å, una media de 6.553Å y una desviación estándar de 3.058Å.

Del conjunto de prueba, se seleccionaron las poses que fueron obtenidas a partir del AutoDock Vina, con 169 atributos y 451 casos. El número de atributos varía del conjunto de entrenamiento al de prueba porque se le aplicaron filtros al conjunto de entrenamiento, disminuyendo de esta forma la cantidad de atributos, Weka realiza el mapeo de los atributos según el nombre de estos, de esta forma son utilizados los atributos del conjunto de prueba que coincidan con los usados en el conjunto de entrenamiento. Los valores de RMSD siguen una distribución de 449 valores distintos y 447 de ellos únicos (99%); un valor mínimo de 1.019Å, máximo de 13.736Å, una media de 6Å y una desviación estándar de 2.728Å.

Una vez realizados los cálculos de regresión, ver Anexo E (Tabla 8), resulto que el algoritmo que mejor coeficiente de correlación aportó fue el KStar con 0.2899 y un error absoluto medio de 2.3116Å (ver Figura 34); mientras que el algoritmo de menor error absoluto medio fue el Random Forest con 2.2216Å y un coeficiente de correlación de 0.2702 (ver Figura 35).



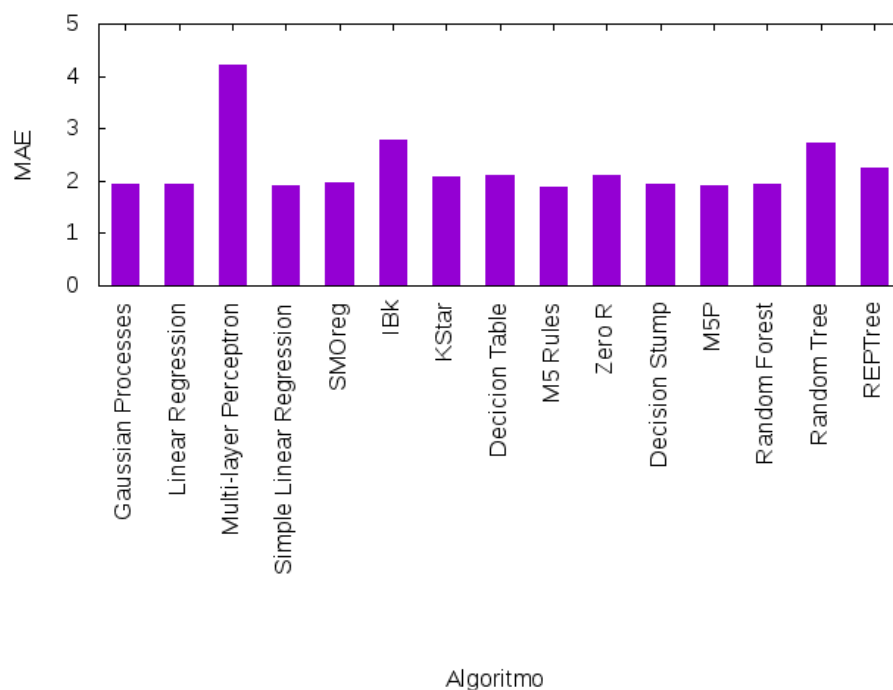


**Figura 32.** Coeficiente de correlación de los algoritmos de predicción con las poses flexibles del AutoDock4 prediciendo poses del AutoDock4.

#### 4.9. Análisis de los resultados obtenidos

En la subsección 4.4 encontramos que las herramientas de acoplamiento molecular lograron, con las poses con menor energía, un  $RMSD \leq 2\text{\AA}$  (Gu y Bourne, 2009) en el 94.4% de los casos para el AutoDock4 y en el 96% para el AutoDock Vina. Por otra parte, en la subsección 4.5 se observa que las herramientas bajaron su desempeño al 10.2% para el caso del AutoDock4 y al 17.6% para el AutoDock Vina. Estos resultados son semejantes a los obtenidos en Pagadala *et al.* (2017) y Ciemny *et al.* (2018), ya que para las poses obtenidas en modo rígido, los métodos de acoplamiento molecular son más exactos que para las poses obtenidas de forma flexible.

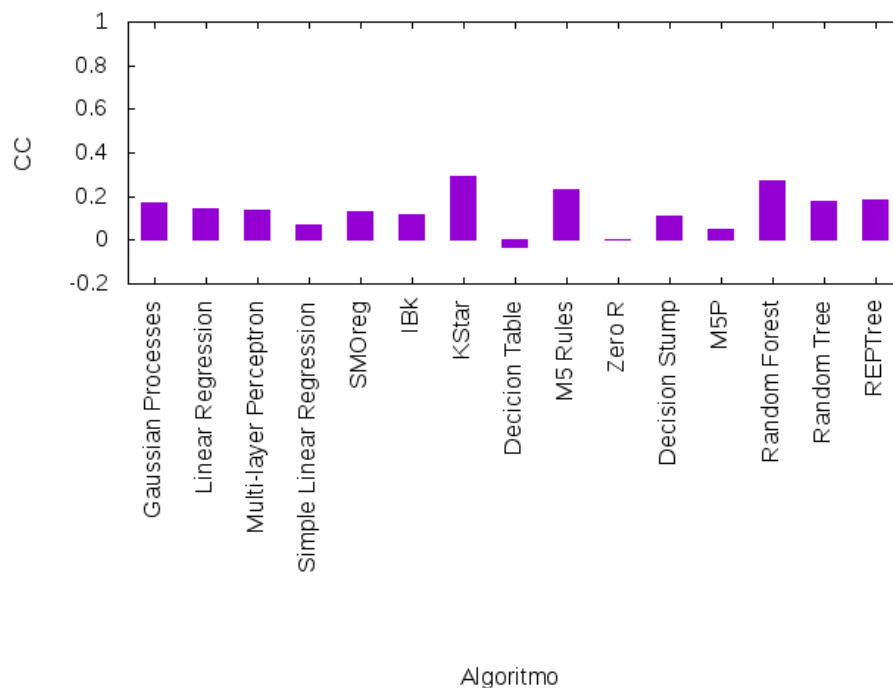
En la sección 4.7 encontramos que para el caso de las poses del AutoDock4 y del AutoDock Vina que predicen poses del AutoDock4 y del AutoDock Vina, el algoritmo con el menor error absoluto medio es el REPTree de Weka con  $\pm 1.5665\text{\AA}$ , siendo este un error relativamente alto ya que está muy cerca de los  $2\text{\AA}$  y un coeficiente de correlación de 0.785 que es un valor relativamente bueno. Para el caso de las poses del AutoDock Vina que predicen poses del AutoDock Vina, el algoritmo con el menor error absoluto medio es el Random Forest con  $\pm 3.1616\text{\AA}$ , siendo este un error muy



**Figura 33.** Error absoluto medio de los algoritmos de predicción con las poses flexibles del AutoDock4 prediciendo poses del AutoDock4.

alto ya que está por encima de los  $2\text{\AA}$  y un coeficiente de correlación de 0.5722 que no es un valor bueno. Sin embargo, en las poses del AutoDock4 que predicen poses del AutoDock4, se obtuvo que el algoritmo M5 Rules de Weka fue el que mejor coeficiente de correlación alcanzó con 0.8929 y un error absoluto medio de  $\pm 0.422\text{\AA}$ , siendo este un buen coeficiente de correlación y un error aceptable en la predicción por estar muy por debajo de los  $2\text{\AA}$ ; también se obtuvo que el algoritmo SMOreg de Weka alcanzó el menor error absoluto medio con  $\pm 0.3537\text{\AA}$  y un coeficiente de correlación de 0.2883 siendo este último relativamente bajo aunque el error sea aceptable por estar muy por debajo de los  $2\text{\AA}$ .

En la sección 4.8 encontramos que para el caso de las poses del AutoDock4 y del AutoDock Vina que predicen poses del AutoDock4 y del AutoDock Vina, el algoritmo con el menor error absoluto medio es el Random Forest con  $\pm 2.0844\text{\AA}$ , siendo este un error alto ya que está por encima de los  $2\text{\AA}$  y un coeficiente de correlación de 0.4415 que no es un valor bueno. Para el caso de las poses del AutoDock4 que predicen poses del AutoDock4, el algoritmo con el menor error absoluto medio es el M5 Rules con  $\pm 1.8834\text{\AA}$ , siendo este un error alto ya que está muy cerca de los  $2\text{\AA}$  y un coeficiente de correlación de 0.3684 que no es un valor bueno. Para el caso de poses del AutoDock

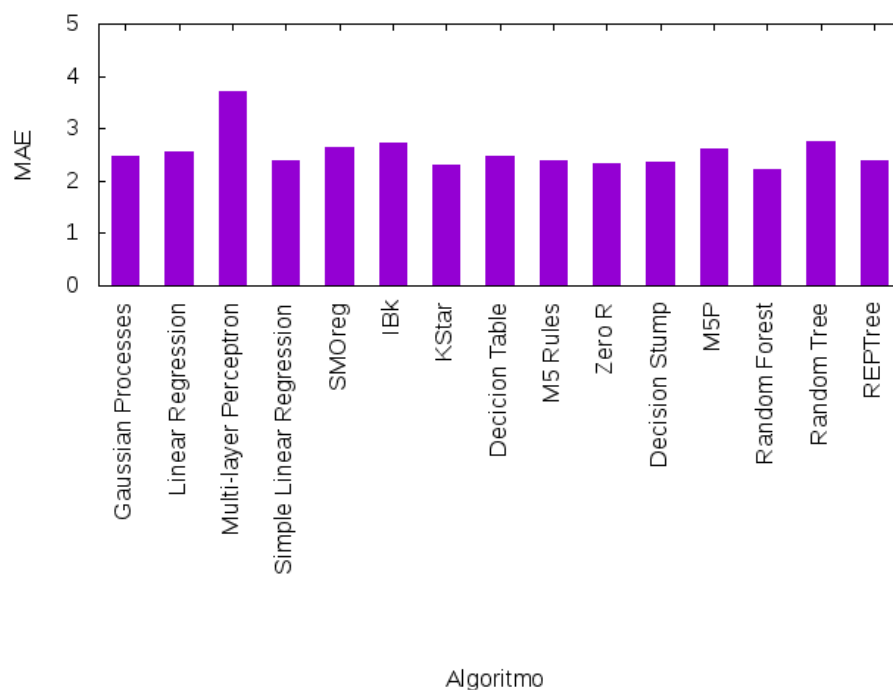


**Figura 34.** Coeficiente de correlación de los algoritmos de predicción con las poses flexibles del AutoDock Vina prediciendo poses del AutoDock Vina.

Vina que predicen poses del AutoDock Vina, el algoritmo con el menor error absoluto medio es el Random Forest con  $\pm 2.2216\text{\AA}$ , siendo este un error muy alto ya que está por encima de los  $2\text{\AA}$  y un coeficiente de correlación de 0.2702 que no es un valor bueno.

Se puede observar que las medidas para la evaluación de los mejores algoritmos de predicción de RMSD, el coeficiente de correlación y el error absoluto medio, no siempre coinciden en el mismo algoritmo de predicción. Un buen algoritmo de predicción debe tener un coeficiente de correlación positivo lo más cercano a 1 posible y al mismo tiempo su error absoluto medio debe ser positivo lo más cercano a 0 posible. Con el objetivo de tener una sola medida de evaluación para la calidad, se propone un cociente de predicción, el cual se maximice con la maximización del coeficiente de correlación y la minimización del error absoluto medio de los resultados. El cociente de predicción está dado por la ecuación:  $CP = \left| \frac{CC}{MAE} \right|$ ; donde CP, es el cociente de predicción, CC es el coeficiente de correlación y MAE es el error absoluto medio. El algoritmo de predicción que más alto cociente de predicción reporte, será el que obtenga la mejor relación coeficiente de correlación - error absoluto medio.

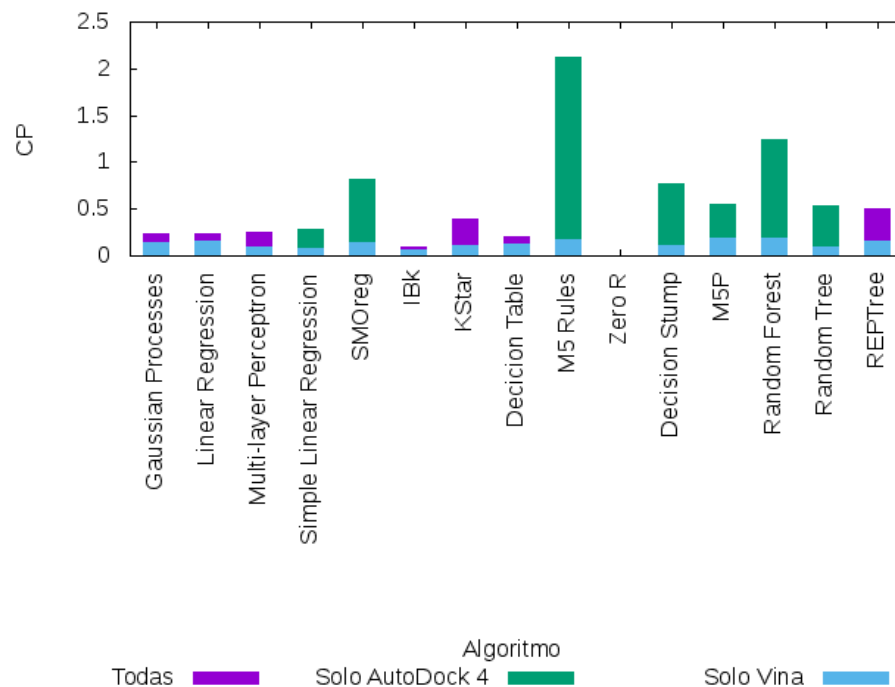
El modelo de predicción que mejor cociente de predicción reportó para las poses



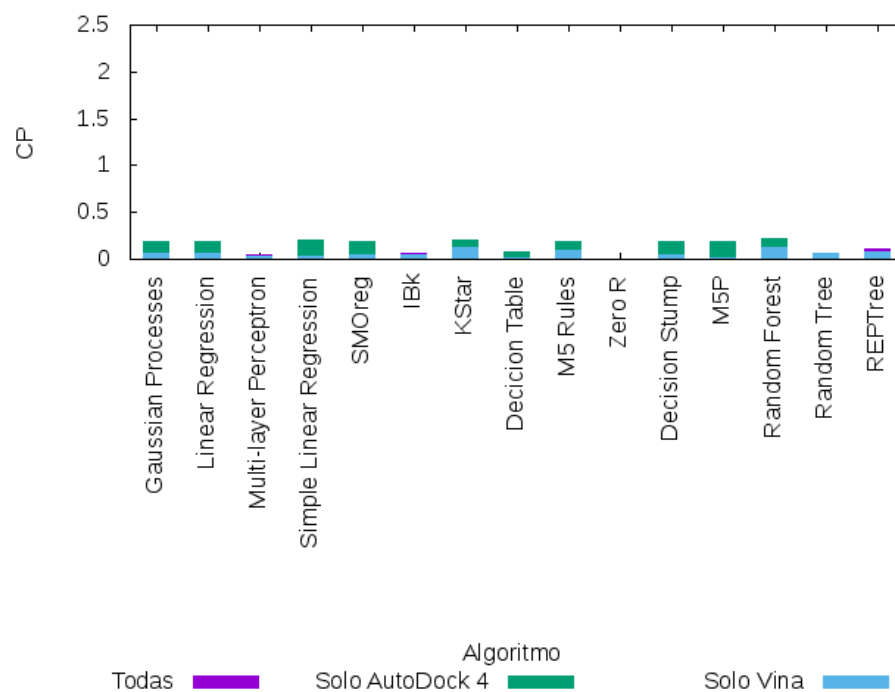
**Figura 35.** Error absoluto medio de los algoritmos de predicción con las poses flexibles del AutoDock Vina prediciendo poses del AutoDock Vina.

rígidas fue el M5 Rules de Weka para las poses del AutoDock4 cuando predicen poses del AutoDock4 con un cociente de predicción de 2.1159 (ver Figura 36), con un coeficiente de correlación de 0.8929 y un error absoluto medio de  $\pm 0.422\text{\AA}$ . Siendo este un resultado muy discreto ya que el AutoDock Vina, con un porcentaje del 96%, logra mejores predicciones de poses rígidias que el AutoDock4.

Para las poses flexibles el modelo que mejor cociente de predicción reportó fue el Random Forest para las poses de AutoDock4 que predicen poses de AutoDock4 con un cociente de predicción de 0.2164 (ver Figura 37), con un coeficiente de correlación de 0.418 y un error absoluto medio de  $\pm 1.9316\text{\AA}$ . Siendo este un mal resultado ya que el coeficiente de correlación es relativamente bajo y el error absoluto medio es relativamente alto; por otro lado, el AutoDock Vina, aunque con un porcentaje muy bajo, 17.6%, logra mejores predicciones de poses flexibles que el AutoDock4.



**Figura 36.** Cociente de predicción de los algoritmos de predicción para: Poses del AutoDock4 y del Vina predicen poses del AutoDock4 y del Vina (Todas), Poses del AutoDock4 predicen poses del AutoDock4 (Solo AutoDock4) y Poses del Vina predicen poses del Vina (Solo Vina); para las poses rígidas.



**Figura 37.** Cociente de predicción de los algoritmos de predicción para: Poses del AutoDock4 y del Vina predicen poses del AutoDock4 y del Vina (Todas), Poses del AutoDock4 predicen poses del AutoDock4 (Solo AutoDock4) y Poses del Vina predicen poses del Vina (Solo Vina); para las poses flexibles.

## Capítulo 5. Conclusiones

---

En el presente capítulo se presentan las conclusiones a las que se llegó con el desarrollo del presente trabajo de investigación, así como algunas perspectivas de investigación sobre el problema abordado.

### 5.1. Sumario

Se analizaron las diferentes funciones de puntuación de acoplamiento molecular. Se seleccionaron cuatro de las herramientas computacionales que implementan dichas funciones, quedando al final AutoDock4 y AutoDock Vina como las candidatas para obtener las poses. Se propusieron una serie de experimentos computacionales para estudiar el desempeño de las herramientas de acoplamiento molecular seleccionadas para la predicción de poses de acoplamiento en la interacción proteína-péptido. A continuación se realizó una evaluación de los descriptores moleculares en su capacidad de predecir las poses de acoplamiento. Los descriptores se dividen en dos grupos: funciones de energía y descripción de las interacciones atómicas proteína-péptido. Para los descriptores moleculares basados en la energía se usaron las energías arrojadas por las herramientas de acoplamiento molecular; y para los descriptores basados en las interacciones atómicas se usaron los descriptores basados en SPLIF.

A continuación se presentan las conclusiones a las que se llegó en base a los experimentos realizados.

### 5.2. Conclusiones

- La herramienta computacional AutoDock Vina, para el acoplamiento molecular rígido dirigido de proteína-péptido, tiene una exactitud del 96 % prediciendo poses con  $RMSD \leq 2\text{\AA}$  al seleccionar la pose de menor energía libre.
- La herramienta computacional AutoDock Vina, para el acoplamiento molecular flexible dirigido de proteína-péptido, tiene una exactitud del 17.6 % prediciendo poses con  $RMSD \leq 2\text{\AA}$  al seleccionar la pose de menor energía libre.

- No fue posible obtener un modelo de aprendizaje de máquina eficiente a partir de las poses obtenidas por las herramientas computacionales AutoDock4 y AutoDock Vina, usando como descriptores, las energías obtenidas por las herramientas computacionales y los SILIRIDs; y como algoritmos de predicción los aportados por Weka para la regresión en su versión estándar.
- El modelo de aprendizaje de máquina para predecir el RMSD de las poses rígidas obtenidas por las herramientas computacionales que mejor resultados reportó fue el M5 Rules de Weka con un error relativo medio de  $\pm 0.422\text{\AA}$  y un coeficiente de correlación de 0.8929.
- El modelo de aprendizaje de máquina para predecir el RMSD de las poses flexibles obtenidas por las herramientas computacionales que mejor resultados reportó fue el Random Forest de Weka con un error relativo medio de  $\pm 2.0423\text{\AA}$  y un coeficiente de correlación de 0.4709.

### 5.3. Propuesta de trabajo futuro

#### Descriptores moleculares

Para los descriptores moleculares se propone adicionar otros que permitan describir mejor las interacciones proteína-péptido. Por lo que se propone la utilización de descriptores moleculares basados en teoría de la información estructural, los cuales podrían describir la forma estructural que presenta el péptido en la cavidad y así predecir con más exactitud el RMSD.

#### Funciones de energía

Para las funciones de energía de acoplamiento molecular, se propone realizar experimentos que tengan en cuenta otros enfoques de energía libre del péptido flexible. Por lo que se propone la utilización de las energías del Rosetta FlexPepDock, DynaDock y PepCrawler (Ciemny *et al.*, 2018); en la predicción de los RMSD, logrando de esta forma encontrar cuáles de los métodos computacionales de acoplamiento molecular pueden describir mejor las poses y de esta forma lograr predecir con mejor

exactitud el RMSD.

### **Selección de características**

Se recomienda mejorar la selección de los descriptores moleculares antes del entrenamiento de las máquinas de aprendizaje. Por lo que se propone realizar un análisis de los métodos de selección de características brindados por Weka y realizar nuevos experimentos usando las elecciones aportadas por dichos métodos.

### **Predicción de poses de acoplamiento**

Predecir la conformación que tendría un péptido cuando este se acople en el sitio activo de su proteína blanco.



## Literatura citada

- Allen, W. J. y Rizzo, R. C. (2014). Implementation of the Hungarian Algorithm to Account for Ligand Symmetry and Similarity in Structure-Based Design. *Journal of Chemical Information and Modeling*, **54**(2): 518–529.
- Allen, W. J., Balius, T. E., Mukherjee, S., Brozell, S. R., Moustakas, D. T., Lang, P. T., Case, D. A., Kuntz, I. D., y Rizzo, R. C. (2015). DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry*, **36**(15): 1132–1156.
- Allouche, D., André, I., Barbe, S., Davies, J., de Givry, S., Katsirelos, G., O’Sullivan, B., Prestwich, S., Schiex, T., y Traoré, S. (2014). Computational protein design as an optimization problem. *Artificial Intelligence*, **212**: 59–79.
- Beach, H., Cole, R., Gill, M. L., y Loria, J. P. (2005). Conservation of  $\mu$ s- ms enzyme motions in the apo-and substrate-mimicked state. *Journal of the American Chemical Society*, **127**(25): 9167–9176.
- Brozell, S. R., Mukherjee, S., Balius, T. E., Roe, D. R., Case, D. A., y Rizzo, R. C. (2012). Evaluation of DOCK 6 as a pose generation and database enrichment tool. *Journal of Computer-Aided Molecular Design*, **26**(6): 749–773.
- Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Costanzo, L. D., Christie, C., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., Liang, Y., Lowe, R., Peisach, E., Periskova, I., Randle, C., Rose, A., Sekharan, M., Shao, C., Tao, Y.-P., Valasatava, Y., Voigt, M., Westbrook, J., Young, J., Zardecki, C., Zhuravleva, M., Kurisu, G., Nakamura, H., Kengaku, Y., Cho, H., Sato, J., Kim, J. Y., Ikegawa, Y., Nakagawa, A., Yamashita, R., Kudou, T., Bekker, G.-J., Suzuki, H., Iwata, T., Yokochi, M., Kobayashi, N., Fujiwara, T., Velankar, S., Kleywegt, G. J., Anyango, S., Armstrong, D. R., Berrisford, J. M., Conroy, M. J., Dana, J. M., Deshpande, M., Gane, P., Gáborová, R., Gupta, D., Gutmanas, A., Koča, J., Mak, L., Mir, S., Mukhopadhyay, A., Nadzirin, N., Nair, S., Patwardhan, A., Paysan-Lafosse, T., Pravda, L., Salih, O., Sehnal, D., Varadi, M., Vařeková, R., Markley, J. L., Hoch, J. C., Romero, P. R., Baskaran, K., Maziuk, D., Ulrich, E. L., Wedell, J. R., Yao, H., Livny, M., y Ioannidis, Y. E. (2018a). Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research*.
- Burley, S. K., Berman, H. M., Christie, C., Duarte, J. M., Feng, Z., Westbrook, J., Young, J., y Zardecki, C. (2018b). RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Science*, **27**(1): 316–330.
- Cerqueira, N. M., Gesto, D., Oliveira, E. F., Santos-Martins, D., Brás, N. F., Sousa, S. F., Fernandes, P. A., y Ramos, M. J. (2015). Receptor-based virtual screening protocol for drug discovery. *Archives of Biochemistry and Biophysics*, **582**: 56–67.
- Chen, Q., Shafer, R. H., y Kuntz, I. D. (1997). Structure-Based Discovery of Ligands Targeted to the RNA Double Helix <sup>†</sup>. *Biochemistry*, **36**(38): 11402–11407.
- Chupakhin, V., Marcou, G., Gaspar, H., y Varnek, A. (2014). Simple ligand–receptor interaction descriptor (silirid) for alignment-free binding site comparison. *Computational and structural biotechnology journal*, **10**(16): 33–37.

- Ciemny, M., Kurcinski, M., Kamel, K., Kolinski, A., Alam, N., Schueler-Furman, O., y Kmiecik, S. (2018). Protein–peptide docking: opportunities and challenges. *Drug discovery today*, **23**(8): 1530–1537.
- Cohen, F. E. y Sternberg, M. J. (1980). On the prediction of protein structure: The significance of the root-mean-square deviation. *Journal of Molecular Biology*, **138**(2): 321–333.
- Comeau, S. R., Gatchell, D. W., Vajda, S., y Camacho, C. J. (2004). ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Research*, **32**(Web Server): W96–W99.
- Craik, D. J., Fairlie, D. P., Liras, S., y Price, D. (2013). The future of peptide-based drugs. *Chemical biology & drug design*, **81**(1): 136–147.
- Da, C. y Kireev, D. (2014). Structural protein-ligand interaction fingerprints (splif) for structure-based virtual screening: method and benchmark study. *Journal of chemical information and modeling*, **54**(9): 2555–2561.
- David, L., Luo, R., y Gilson, M. K. (2001). Ligand-receptor docking with the mining minima optimizer. *Journal of computer-aided molecular design*, **15**(2): 157–171.
- Ewing, T. J., Makino, S., Skillman, A. G., y Kuntz, I. D. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-Aided Molecular Design*, **15**(5): 411–428.
- Ewing, T. J. A. y Kuntz, I. D. (1997). Critical evaluation of search algorithms for automated molecular docking and database screening. *Journal of Computational Chemistry*, **18**(9): 1175–1189.
- Fischer, E. (1894). Einfluss der configuration auf die wirkung der enzyme. *Berichte der deutschen chemischen Gesellschaft*, **27**(3): 2985–2993.
- Gabb, H. A., Jackson, R. M., y Sternberg, M. J. (1997). Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of molecular biology*, **272**(1): 106–120.
- Goodsell, D. S. y Olson, A. J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins: Structure, Function, and Genetics*, **8**(3): 195–202.
- Gu, J. y Bourne, P. E. (2009). *Structural bioinformatics*, Vol. 44. John Wiley & Sons.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., y Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, **11**(1): 10–18.
- Hauser, A. S. y Windshügel, B. (2016). LEADS-PEP: A Benchmark Data Set for Assessment of Peptide Docking Performance. *Journal of Chemical Information and Modeling*, **56**(1): 188–200.
- Holden, P. M., Kaur, H., Goyal, R., Gochin, M., y Rizzo, R. C. (2012). Footprint-based identification of viral entry inhibitors targeting HIVgp41. *Bioorganic & Medicinal Chemistry Letters*, **22**(8): 3011–3016.
- Huang, P.-S., Boyken, S. E., y Baker, D. (2016). The coming of age of de novo protein design. *Nature*, **537**(7620): 320–327.

- Jones, G., Willett, P., y Glen, R. C. (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of molecular biology*, **245**(1): 43–53.
- Jorgensen, W. L. (2004). The many roles of computation in drug discovery. *Science (New York, N.Y.)*, **303**(5665): 1813–8.
- Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., y Vakser, I. A. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proceedings of the National Academy of Sciences of the United States of America*, **89**(6): 2195–9.
- Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. En: *Advances in protein chemistry*, Vol. 14. Elsevier, pp. 1–63.
- Kollman, P. (1993). Free energy calculations: applications to chemical and biochemical phenomena. *Chemical reviews*, **93**(7): 2395–2417.
- Koshland, D. E. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences*, **44**(2): 98–104.
- Kozakov, D., Brenke, R., Comeau, S. R., y Vajda, S. (2006). PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Structure, Function, and Bioinformatics*, **65**(2): 392–406.
- Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., y Vajda, S. (2017). The ClusPro web server for protein-protein docking. *Nature protocols*, **12**(2): 255–278.
- Lee Rodgers, J. y Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, **42**(1): 59–66.
- Ljungberg, K. B., Marelius, J., Musil, D., Svensson, P., Norden, B., y Åqvist, J. (2001). Computational modelling of inhibitor binding to human thrombin. *European journal of pharmaceutical sciences*, **12**(4): 441–446.
- London, N., Movshovitz-Attias, D., y Schueler-Furman, O. (2010). The Structural Basis of Peptide-Protein Binding Strategies. *Structure*, **18**(2): 188–199.
- Mcmartin, C. y Bohacek, R. S. (1997). Qxp: powerful, rapid computer algorithms for structure-based drug design. *Journal of computer-aided molecular design*, **11**(4): 333–344.
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., y Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, **19**(14): 1639–1662.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., y Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, **30**(16): 2785–2791.
- Moustakas, D. T., Lang, P. T., Pegg, S., Pettersen, E., Kuntz, I. D., Brooijmans, N., y Rizzo, R. C. (2006). Development and validation of a modular, extensible docking program: DOCK 5. *Journal of Computer-Aided Molecular Design*, **20**(10-11): 601–619.

- Murgueitio, M. S., Bermudez, M., Mortier, J., y Wolber, G. (2012). In silico virtual screening approaches for anti-viral drug discovery. *Drug Discovery Today: Technologies*, **9**(3): e219–e225.
- Nelson, D. L., Lehninger, A. L., y Cox, M. M. (2008). *Lehninger principles of biochemistry*. Macmillan.
- Pagadala, N. S., Syed, K., y Tuszynski, J. (2017). Software for molecular docking: a review. *Biophysical reviews*, **9**(2): 91–102.
- Perola, E., Walters, W. P., y Charifson, P. S. (2004). A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins: Structure, Function, and Bioinformatics*, **56**(2): 235–249.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., y Ferrin, T. E. (2004). Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, **25**(13): 1605–1612.
- Radifar, M., Yuniarti, N., y Istyastono, E. P. (2013). Pyplif: Python-based protein-ligand interaction fingerprinting. *Bioinformatics*, **9**(6): 325.
- Rizzo, R. C., Tirado-Rives, J., y Jorgensen, W. L. (2001). Estimation of binding affinities for hept and nevirapine analogues with hiv-1 reverse transcriptase via monte carlo simulations. *Journal of medicinal chemistry*, **44**(2): 145–154.
- Todeschini, R. y Consonni, V. (2009). *Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references*, Vol. 41. John Wiley & Sons.
- Trott, O. y Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, **31**(2): NA–NA.
- Van Drie, J. H. (2007). Computer-aided drug design: the next 20 years. *Journal of Computer-Aided Molecular Design*, **21**(10-11): 591–601.
- Wang, R., Lai, L., y Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of computer-aided molecular design*, **16**(1): 11–26.
- Wang, R., Fang, X., Lu, Y., y Wang, S. (2004). The PDBbind Database: Collection of Binding Affinities for Protein–Ligand Complexes with Known Three-Dimensional Structures. *Journal of Medicinal Chemistry*, **47**(12): 2977–2980.
- Wang, R., Fang, X., Lu, Y., Yang, C.-Y., y Wang, S. (2005). The PDBbind Database: Methodologies and Updates. *Journal of Medicinal Chemistry*, **48**(12): 4111–4119.

## Anexo A - Formatos para el almacenamiento de los complejos moleculares

En el presente anexo se muestran extractos de los formatos de los ficheros para el almacenamiento de los complejos moleculares.

### Formato PDB

El formato PDB es un formato de archivo en forma de texto que describe las estructuras tridimensionales de las moléculas contenidas en el Banco de Datos de Proteínas. Proporciona una descripción y anotación de las estructuras de ácidos nucleicos y de proteínas, incluidas las coordenadas atómicas, las asignaciones de estructuras secundarias y la conectividad atómica, ver Formato .1.

**HEADER, TITLE y AUTHOR:** proporcionan información sobre el identificador, el título y los investigadores que definieron la estructura.

**REMARK:** contiene anotaciones de forma libre, ajustándose a información estandarizada.

**SEQRES:** proporciona las secuencias de las cadenas peptídicas.

**ATOM:** describe las coordenadas de los átomos que forman parte de la proteína.

**HETATM:** describe las coordenadas de los heteroátomos.

#### Formato .1. Extracto de un archivo en formato PDB del complejo 1apm.

```

HEADER  TRANSFERASE(PHOSPHOTRANSFERASE)          18-JAN-93   1APM
TITLE    2.0 ANGSTROM REFINED CRYSTAL STRUCTURE OF THE CATALYTIC
TITLE    2 SUBUNIT OF CAMP-DEPENDENT PROTEIN KINASE COMPLEXED WITH A
TITLE    3 PEPTIDE INHIBITOR AND DETERGENT
...
EXPDTA   X-RAY DIFFRACTION
AUTHOR   D.R.KNIGHTON,S.M.BELL,J.ZHENG,L.F.TENEYCK,N.-H.XUONG,
AUTHOR   2 S.S.TAYLOR,J.M.SOWADSKI
...
REMARK 350 BIOMOLECULE: 1
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: DIMERIC
REMARK 350 SOFTWARE DETERMINED QUATERNARY STRUCTURE: DIMERIC
REMARK 350 SOFTWARE USED: PISA
REMARK 350 TOTAL BURIED SURFACE AREA: 3000 ANGSTROM**2
REMARK 350 SURFACE AREA OF THE COMPLEX: 15620 ANGSTROM**2
REMARK 350 CHANGE IN SOLVENT FREE ENERGY: 8.0 KCAL/MOL
REMARK 350 APPLY THE FOLLOWING TO CHAINS: E, I
REMARK 350  BIOMT1  1  1.000000  0.000000  0.000000      0.00000
REMARK 350  BIOMT2  1  0.000000  1.000000  0.000000      0.00000

```

```

REMARK 350 BIOMT3 1 0.000000 0.000000 1.000000 0.000000
...
SEQRES 1 E 350 GLY ASN ALA ALA ALA LYS LYS GLY SEP GLU GLN GLU
...
SEQRES 27 E 350 ILE ASN GLU LYS CYS GLY LYS GLU PHE THR GLU PHE
SEQRES 1 I 20 THR THR TYR ALA ASP PHE ILE ALA SER GLY ARG THR GLY
SEQRES 2 I 20 ARG ARG ASN ALA ILE HIS ASP
...
ATOM 11 N GLU E 11 -7.219 -19.053 -12.525 1.00 65.59 N
ATOM 12 CA GLU E 11 -6.833 -17.952 -11.656 1.00 36.48 C
ATOM 13 C GLU E 11 -5.851 -17.061 -12.401 1.00 68.77 C
...
ATOM 2861 OXT ASP I 24 27.839 8.607 -5.342 1.00100.00 O
TER 2862 ASP I 24
HETATM 2863 C1 OCT E 354 -1.335 -8.264 -9.894 1.00 25.68 C
HETATM 2864 C2 OCT E 354 -2.610 -8.301 -9.016 1.00 25.36 C
HETATM 2865 C3 OCT E 354 -3.735 -9.205 -9.576 1.00 28.46 C
...
END

```

## Formato PDBQT

El formato PDBQT es un formato muy semejante al formato PDB incluyendo la carga parcial para cada átomo (Q) y el tipo de átomo definido por AutoDock (T). Para realizar el acoplamiento molecular rígido, ver Formato .2 y el acoplamiento molecular flexible, ver Formato .3.

### Formato .2. Extracto de un archivo en formato PDBQT del péptido 1apm rígido.

```

REMARK 0 active torsions:
REMARK status: ('A' for Active; 'I' for Inactive)
REMARK I between atoms: N_1 and CA_5
REMARK I between atoms: CA_5 and CB_8
REMARK I between atoms: CA_5 and C_6
...
REMARK I between atoms: CB_211 and CG_212
ROOT
ATOM 1 N THR I 5 26.695 -13.081 19.007 1.00 15.23 -0.064 N
ATOM 2 HN1 THR I 5 27.618 -13.049 18.574 1.00 0.00 0.275 HD
ATOM 3 HN2 THR I 5 26.687 -13.710 19.810 1.00 0.00 0.275 HD
...
ATOM 215 OXT ASP I 24 27.839 8.607 -5.342 1.00100.00 -0.646 OA
ENDROOT
TORSDOF 81

```

### Formato .3. Extracto de un archivo en formato PDBQT del péptido 1awr flexible.

```

REMARK 20 active torsions:
REMARK status: ('A' for Active; 'I' for Inactive)
REMARK 1 A between atoms: N_1 and CA_5
REMARK 2 A between atoms: CA_5 and CB_8
REMARK 3 A between atoms: CA_5 and C_6
...
REMARK 20 A between atoms: CA_45 and C_46
ROOT
ATOM 1 C GLY I 103 19.825 20.120 51.856 1.00 9.63 0.238 C
ATOM 2 O GLY I 103 19.714 20.437 50.647 1.00 6.12 -0.272 OA
ENDROOT

```

```

BRANCH 1 3
ATOM 3 CA GLY I 103 18.767 19.270 52.559 1.00 4.95 0.225 C
BRANCH 3 4
ATOM 4 N GLY I 103 17.457 19.206 51.928 1.00 6.14 -0.351 N
ATOM 5 HN GLY I 103 17.005 20.076 51.645 1.00 0.00 0.163 HD
...
ENDBRANCH 3 4
ENDBRANCH 1 3
...
TORSDOF 15

```

## Formato MOL2

El formato MOL2 almacena toda la información y las características del átomo, la posición y la conectividad, ver Formato .4.

**Formato .4.** Extracto de un archivo en formato MOL2 del péptido 1apm.

```

@<TRIPOS>MOLECULE
1apm
215 214 0 0 0
SMALL
GASTEIGER

@<TRIPOS>ATOM
1 N 26.6950 -13.0810 19.0070 N.4 5 THR5 -0.0640
2 HN1 27.6180 -13.0490 18.5740 H 5 THR5 0.2750
3 HN2 26.6870 -13.7100 19.8100 H 5 THR5 0.2750
...
215 OXT 27.8390 8.6070 -5.3420 O.co2 24 ASP24 -0.6460
@<TRIPOS>BOND
...
15 209 215 ar
16 202 199 2
17 202 204 1
...
214 1 3 1

```

## Formato DMS

El formato DMS almacena la información de las esferas de contacto que son generadas a partir de la información atómica del receptor, ver Formato .5.

**Formato .5.** Extracto de un archivo en formato dms del receptor 1apm.

```

GLU 11E N -7.219 -19.053 -12.525 A
GLU 11E N -5.824 -19.902 -12.376 SC0 0.911 0.851 -0.518 0.091
GLU 11E N -8.267 -19.477 -11.337 SC0 0.571 -0.639 -0.258 0.725
...
PHE 350E OXT 3.771 20.625 -12.952 SC0 0.568 0.000 -0.500 0.866

```

## Anexo B - Configuración de las herramientas de acoplamiento molecular

En el presente anexo se muestran los ficheros de configuración de las herramientas para realizar el acoplamiento molecular.

### Configuración de Autogrid4

Ficheros de configuración para la herramienta Autogrid4. Para el acoplamiento molecular global, ver Configuración .1 y para el acoplamiento molecular local, ver Configuración .2.

**Configuración .1.** Configuración de la herramienta Autogrid4 para el acoplamiento molecular global del complejo 1apm.

```
npts 87 123 109           # num.grid points in xyz
gridfld 1apmP.maps.fld   # grid_data_file
spacing 1.0              # spacing(A)
receptor_types A C HD N OA P SA # receptor atom types
ligand_types A C HD N NA OA SA # ligand atom types
receptor 1apmP.pdbqt     # macromolecule
gridcenter 9.242 1.701 -2.952 # xyz-coordinates or auto
smooth 0.5               # store minimum energy w/in rad(A)
map 1apmP.A.map          # atom-specific affinity map
map 1apmP.C.map          # atom-specific affinity map
map 1apmP.HD.map         # atom-specific affinity map
map 1apmP.N.map          # atom-specific affinity map
map 1apmP.NA.map         # atom-specific affinity map
map 1apmP.OA.map         # atom-specific affinity map
map 1apmP.SA.map         # atom-specific affinity map
elecmap 1apmP.e.map      # electrostatic potential map
dsolvmap 1apmP.d.map     # desolvation potential map
dielectric -0.1465       # <0, AD4 distance-dep.diel;>0, constant
```

**Configuración .2.** Configuración de la herramienta Autogrid4 para el acoplamiento molecular local del complejo 1apm.

```
npts 68 90 102           # num.grid points in xyz
gridfld 1apmP.maps.fld   # grid_data_file
spacing 0.375            # spacing(A)
receptor_types A C H HD N OA P SA # receptor atom types
ligand_types A C OA HD N # ligand atom types
receptor 1apmP.pdbqt     # macromolecule
gridcenter 20.497 -1.297 4.856 # xyz-coordinates or auto
smooth 0.5               # store minimum energy w/in rad(A)
map 1apmP.A.map          # atom-specific affinity map
map 1apmP.C.map          # atom-specific affinity map
map 1apmP.OA.map         # atom-specific affinity map
map 1apmP.HD.map         # atom-specific affinity map
map 1apmP.N.map          # atom-specific affinity map
elecmap 1apmP.e.map      # electrostatic potential map
dsolvmap 1apmP.d.map     # desolvation potential map
dielectric -0.1465       # <0, AD4 distance-dep.diel;>0, constant
```



## Configuración de AutoDock4

Fichero de configuración para la herramienta AutoDock4, ver Configuración .3

**Configuración .3.** Configuración de la herramienta AutoDock4 para el acoplamiento molecular del complejo 1apm.

```

autodock_parameter_version 4.2      # used by autodock to validate parameter set
outlev 1                             # diagnostic output level
intelec                              # calculate internal electrostatics
seed pid time                         # seeds for random generator
ligand_types A C OA HD N            # atoms types in ligand
fld 1apmP.maps.fld                  # grid_data_file
map 1apmP.A.map                     # atom-specific affinity map
map 1apmP.C.map                     # atom-specific affinity map
map 1apmP.OA.map                    # atom-specific affinity map
map 1apmP.HD.map                    # atom-specific affinity map
map 1apmP.N.map                     # atom-specific affinity map
elecmap 1apmP.e.map                 # electrostatics map
desolvmap 1apmP.d.map               # desolvation map
move 1apmL.pdbqt                    # small molecule
about 22.3434 -0.0427 8.7123        # small molecule center
tran0 random                         # initial coordinates/A or random
quaternion0 random                  # initial orientation
dihe0 random                         # initial dihedrals (relative) or random
torsdof 81                          # torsional degrees of freedom
rmstol 2.0                           # cluster_tolerance/A
extnrg 1000.0                        # external grid energy
e0max 0.0 10000                      # max initial energy; max number of retries
ga_pop_size 150                      # number of individuals in population
ga_num_evals 2500000                  # maximum number of energy evaluations
ga_num_generations 27000              # maximum number of generations
ga_elitism 1                          # number of top individuals to survive to next generation
ga_mutation_rate 0.02                # rate of gene mutation
ga_crossover_rate 0.8                 # rate of crossover
ga_window_size 10                     #
ga_cauchy_alpha 0.0                   # Alpha parameter of Cauchy distribution
ga_cauchy_beta 1.0                    # Beta parameter Cauchy distribution
set_ga                                # set the above parameters for GA or LGA
sw_max_its 300                        # iterations of Solis & Wets local search
sw_max_succ 4                         # consecutive successes before changing rho
sw_max_fail 4                         # consecutive failures before changing rho
sw_rho 1.0                            # size of local search space to sample
sw_lb_rho 0.01                        # lower bound on rho
ls_search_freq 0.06                   # probability of performing local search on individual
set_psw1                              # set the above pseudo-Solis & Wets parameters
unbound_model bound                   # state of unbound ligand
ga_run 20                             # do this many hybrid GA-LS runs
analysis                              # perform a ranked cluster analysis

```

## Configuración de AutoDock Vina

Ficheros de configuración para la herramienta AutoDock Vina. Para el acoplamiento molecular global, ver Configuración .4 y para el acoplamiento molecular local, ver Configuración .5.

**Configuración .4.** Configuración de la herramienta AutoDock Vina para el acoplamiento molecular global del complejo 1apm.

```
receptor = 1apmP.pdbqt
ligand = 1apmL.pdbqt
center_x = 9.242
center_y = 1.701
center_z = -2.952
size_x = 87
size_y = 123
size_z = 109
out = 1apmL.pdbqt
exhaustiveness = 120
num_modes = 20
energy_range = 100
```

**Configuración .5.** Configuración de la herramienta AutoDock Vina para el acoplamiento molecular local del complejo 1apm.

```
receptor = 1apmP.pdbqt
ligand = 1apmL.pdbqt
center_x = 20.497
center_y = -1.297
center_z = 4.856
size_x = 25.5
size_y = 33.75
size_z = 38.25
out = 1apmL.pdbqt
exhaustiveness = 120
num_modes = 10
energy_range = 100
```

## Configuración de Sphgen

Fichero de configuración para la herramienta Sphgen, ver Configuración .6

**Configuración .6.** Configuración de la herramienta Sphgen para el acoplamiento molecular global del complejo 1apm.

```
1apm.dms
R
X
0.0
4.0
1.4
rec.sph
```

## Configuración de Showbox

Ficheros de configuración para la herramienta Showbox. Para el acoplamiento molecular global, ver Configuración .7 y para el acoplamiento molecular local, ver Configuración .8.

**Configuración .7.** Configuración de la herramienta Showbox para el acoplamiento molecular global del complejo 1apm.

```
N
U
9.242
1.701
-2.952
87
123
109
recBox.pdb
```

**Configuración .8.** Configuración de la herramienta Showbox para el acoplamiento molecular local del complejo 1apm.

```
N
U
20.497
-1.297
4.856
25.5
33.75
38.25
recBox.pdb
```

## Configuración de Grid

Fichero de configuración para la herramienta Grid, ver Configuración .9

**Configuración .9.** Configuración de la herramienta Grid para el acoplamiento molecular del complejo 1apm.

```
compute_grids          yes
grid_spacing           0.3
output_molecule       no
contact_score          no
energy_score           yes
energy_cutoff_distance 9999
atom_model             a
attractive_exponent    6
repulsive_exponent     12
distance_dielectric    yes
dielectric_factor      4
bump_filter            yes
bump_overlap           0.75
receptor_file          1apmP.mol2
box_file               recBox.pdb
vdw_definition_file    vdw_AMBER_parm99.defn
score_grid_prefix      grid
```

## Configuración de Dock6

Fichero de configuración para la herramienta Dock6, ver Configuración .10

**Configuración .10.** Configuración de la herramienta Dock6 para el acoplamiento molecular global del complejo 1apm.

```

conformer_search_type          rigid
use_internal_energy            yes
internal_energy_rep_exp        12
internal_energy_cutoff         100.0
ligand_atom_file               1apmL.mol2
limit_max_ligands              no
skip_molecule                 no
read_mol_solvation             no
calculate_rmsd                 yes
use_rmsd_reference_mol         no
use_database_filter            no
orient_ligand                  yes
automated_matching             yes
receptor_site_file             rec.sph
max_orientations                1000
critical_points                 no
chemical_matching              no
use_ligand_spheres             no
bump_filter                     no
score_molecules                 yes
contact_score_primary          no
contact_score_secondary        no
grid_score_primary              yes
grid_score_secondary           no
grid_score_rep_rad_scale       1
grid_score_vdw_scale           1
grid_score_es_scale            1
grid_score_grid_prefix         grid
multigrid_score_secondary      no
dock3.5_score_secondary        no
continuous_score_secondary     no
footprint_similarity_score_secondary no
pharmacophore_score_secondary  no
descriptor_score_secondary     no
gbsa_zou_score_secondary       no
gbsa_hawkins_score_secondary   no
SASA_score_secondary           no
amber_score_secondary          no
minimize_ligand                yes
simplex_max_iterations           1000
simplex_tors_premin_iterations   0
simplex_max_cycles              1
simplex_score_converge          0.1
simplex_cycle_converge          1.0
simplex_trans_step              1.0
simplex_rot_step                 0.1
simplex_tors_step               10.0
simplex_random_seed             0
simplex_restraint_min           no
atom_model                      all
vdw_defn_file                   vdw_AMBER_parm99.defn
flex_defn_file                   flex.defn
flex_drive_file                  flex_drive.tbl
ligand_outfile_prefix           rigidL
write_orientations              no
num_scored_conformers          20
write_conformations             yes
cluster_conformations          no
rank_ligands                    no

```

## Configuración de PyPlif

Fichero de configuración para la herramienta PyPlif, ver Configuración .11

**Configuración .11.** Configuración de la herramienta PyPlif para el cálculo de los PLIF para las poses del complejo 1apm.

```
protein_reference 1apmP.mol2  
ligand_reference 1apmL.mol2
```

```
protein_ligand_folder 1apm  
residue_of_choice LYS47 THR48 LEU49 GLY50 THR51 GLY52 SER53 PHE54 GLY55  
...  
GLY287 ILE291 TRP296 PHE327 ASP328 ASP329 TYR330 TPO197  
output_file 1apm/1apm.csv
```

## Anexo C - Comandos de ejecución

En el presente anexo se muestran los comandos de ejecución de cada una de las herramientas utilizadas en la investigación. Todas las herramientas utilizadas fueron usadas en su versión para Sistema Operativo Linux, versión Red Hat Enterprise Linux Server, distribución 6.7 (Santiago), con 30 nodos, 24 núcleos y 120GB de RAM cada nodo.

### AutoGrid4

Para ejecutar AutoGrid4 se usó el comando `./autogrid4 -p conf.gpf -l grid.log` donde *autogrid4* es la llamada de la acción del comando, *conf.gpf*, es el archivo de configuración, ver Apéndice B, Configuración .1; y *grid.log* es el archivo donde se guardará el log de la llamada al comando *autogrid4*.

### AutoDock4

Para ejecutar AutoDock4 se usó el comando `./autodock4 -p conf.dpf -l 1apm.dlg` donde *autodock4* es la llamada de la acción del comando, *conf.dpf*, es el archivo de configuración, ver Apéndice B, Configuración .3; y *1apm.dlg* es el archivo donde se guardarán las poses del acoplamiento molecular realizados por AutoDock4, en este caso del complejo 1apm.

### Sphgen

Para ejecutar Sphgen se usó el comando `./sphgen` donde *sphgen* es la llamada de la acción del comando. El archivo de configuración, ver Apéndice B, Configuración .6, debe estar en el mismo directorio de ejecución del comando.

### Showbox

Para ejecutar Showbox se usó el comando `showbox < showbox.in` donde *showbox* es la llamada de la acción del comando, *showbox.in*, es el archivo de configuración, ver Apéndice B, Configuración .8.

## **Grid**

Para ejecutar Grid se usó el comando `./grid -i grid.in` donde *grid* es la llamada de la acción del comando, *grid.in*, es el archivo de configuración, ver Apéndice B, Configuración .9.

## **Dock6**

Para ejecutar Dock6 se usó el comando `./dock6 -i dock.in` donde *dock6* es la llamada de la acción del comando, *dock.in*, es el archivo de configuración, ver Apéndice B, Configuración .10.

## **AutoDock Vina**

Para ejecutar AutoDock Vina se usó el comando `./vina --config conf.txt` donde *vina* es la llamada de la acción del comando, *conf.txt*, es el archivo de configuración, ver Apéndice B, Configuración .5.

## **Pyplif**

Para ejecutar Pyplif se usó el comando `./pyplif -c 1apm_conf.txt` donde *pyplif* es la llamada de la acción del comando, *1apm\_conf.txt*, es el archivo de configuración, ver Apéndice B, Configuración .11; en este caso del complejo 1apm.

## Anexo D - Errores en tiempo de ejecución

En el presente anexo se muestran los errores arrojados por las herramientas durante la ejecución de los cálculos de acoplamiento molecular.

### AutoDock Tools

Durante la preparación de los complejos para el acoplamiento molecular, el AutoDock Tools arrojó el siguiente error al eliminar y posteriormente adicionar los átomos de hidrógeno en el complejo 1xoc, ver los errores .1 y .2.

**Error .1.** Error de la herramienta AutoDock Tools al intentar eliminar los átomos de hidrógeno del complejo 1xoc.

```
...
>>> ERROR *****
Traceback (most recent call last):
  File "/MGLToolsPckgs/ViewerFramework/VF.py", line 898, in tryto
    result = command( *args, **kw )
  File "/MGLToolsPckgs/Pmv/deleteCommands.py", line 485, in doit
    self.vf.dispatchEvent(event)
  File "/MGLToolsPckgs/ViewerFramework/VF.py", line 537, in dispatchEvent
    func(event)
  File "/MGLToolsPckgs/Pmv/dashboardCommands.py", line 577, in handleDeleteAtomsEvents
    if parents[0]==None: break
  File "/lib/python2.5/UserList.py", line 28, in __getitem__
    def __getitem__(self, i): return self.data[i]
IndexError: list index out of range
```

**Error .2.** Error de la herramienta AutoDock Tools al intentar adicionar los átomos de hidrógeno al complejo 1xoc.

```
...
ERROR *****
Traceback (most recent call last):
  File "/MGLToolsPckgs/ViewerFramework/VF.py", line 898, in tryto
    result = command( *args, **kw )
  File "/MGLToolsPckgs/Pmv/editCommands.py", line 1043, in doit
    hat = addh.addHydrogens(mol.allAtoms, method=method)
  File "/MGLToolsPckgs/PyBabel/addh.py", line 92, in addHydrogens
    Hatoms = self.place_hydrogens1(atoms, num_H_to_add)
  File "/MGLToolsPckgs/PyBabel/addh.py", line 115, in place_hydrogens1
    Hat = Hat + self.add_tertiary_hydrogen(a, SP3_C_H_DIST)
  File "/MGLToolsPckgs/PyBabel/addh.py", line 345, in add_tertiary_hydrogen
    s = [ m[0][0] + m[0][1] + m[0][2],
TypeError: 'NoneType' object is unsubscriptable
```

### AutoGrid4

Durante los cálculos de la malla de puntos, la herramienta AutoGrid4 arrojó el siguiente error que impidió la continuación de los cálculos para los complejos: 1n12, 3n55 y 4c2c; ver Error .3.



**Error .3.** Error de la herramienta AutoGrid4 al intentar calcular la malla de puntos a los complejos: 1n12, 3n55 y 4c2c.

```
...
./autogrid4: ERROR: Unknown receptor type: "Se"
— Add parameters for it to the parameter library first!
./autogrid4: Unsuccessful Completion.
```

## AutoDock4

Durante los cálculos de acoplamiento molecular la herramienta AutoDock4 arrojó los siguientes errores que impidieron la continuación de los cálculos para los complejos: 1ot5 y 1r64; ver Error .4.

**Error .4.** Error de la herramienta AutoDock4 al intentar realizar acoplamiento molecular a los complejos: 1ot5 y 1r64.

```
...
./autodock4: ERROR: Unknown ligand atom type "B";
add parameters for it to the parameter library first!
./autodock4: FATAL ERROR: ./autodock4: ERROR: Unknown ligand atom type "B";
add parameters for it to the parameter library first!
./autodock4: Unsuccessful Completion.
```

## AutoDock Vina

Durante los cálculos de acoplamiento molecular la herramienta AutoDock Vina arrojó los siguientes errores que impidieron la continuación de los cálculos para los complejos: 1ot5, 1r64, 5hi4; ver errores .5 y .6.

**Error .5.** Error de la herramienta AutoDock Vina al intentar realizar acoplamiento molecular a los complejos: 1ot5 y 1r64.

```
...
Parse error on line 50 in file "1ot5L.pdbqt": ATOM syntax incorrect: "B" is not a valid AutoDock type.
Note that AutoDock atom types are case-sensitive.
...
Parse error on line 86 in file "1r64L.pdbqt": ATOM syntax incorrect: "B" is not a valid AutoDock type.
Note that AutoDock atom types are case-sensitive.
```

**Error .6.** Error de la herramienta AutoDock Vina al intentar realizar acoplamiento molecular al complejo: 5hi4.

An error occurred: boost::thread\_resource\_error.

Please contact the author, Dr. Oleg Trott <ot14@columbia.edu>, so that this problem can be resolved. The reproducibility of the error may be vital, so please remember to include the following in your problem report:

- \* the EXACT error message,
- \* your version of the program,
- \* the type of computer system you are running it on,
- \* all command line options,

- \* configuration file (if used),
- \* ligand file as PDBQT,
- \* receptor file as PDBQT,
- \* flexible side chains file as PDBQT (if used),
- \* output file as PDBQT (if any),
- \* input (if possible),
- \* random seed the program used (this is printed when the program starts).

Thank you!

Error: insufficient memory!

Error: insufficient memory!

## Anexo E - Tablas

En el presente anexo se muestran las tablas que contienen la información sobre la base de datos utilizada en los experimentos y las correlaciones existentes entre la pose de menor energía y su correspondiente RMSD con la pose del cristal, obtenidas a partir de las herramientas de acoplamiento molecular y los resultados de las predicciones arrojadas por los algoritmos de Weka.

### DockingDB

Cuenta con 249 complejos, los cuales fueron filtrados para seleccionar la mínima cantidad de cadenas para realizar el acoplamiento molecular, ver Tabla 2. Como se puede observar, fue almacenado el identificador PDB del complejo (PDB id), el(los) identificador(es) de la(s) cadena(s) correspondiente(s) a la proteína (Cad. prot.), el identificador de la cadena del péptido (Cad. pépt.), la secuencia de aminoácidos de la cadena del péptido (Pépt. sec.) y el número de rotaciones de los enlaces del péptido (Pépt. rot.):

**PDB id:** Cadena de cuatro caracteres alfanuméricos que corresponden con el identificador del complejo en la PDB.

**Cad. prot.:** Identificador de un carácter alfabético para cada una de la(s) cadena(s) que componen a la proteína.

**Cad. pépt.:** Identificador de un carácter alfabético para la cadena que compone al péptido.

**Pépt. sec.:** Secuencia de caracteres alfabéticos correspondiente a los identificadores de una letra de los aminoácidos naturales. Los aminoácidos no naturales son representados por \*, ya que no tienen código de una letra.

**Pépt. rot.** Número natural mayor que 0 que representa la cantidad de rotaciones de los enlaces del péptido.

**Tabla 2.** Descripción general de las estructuras incluidas en el conjunto de datos de referencia DockingDB.

| PDB id | Cad. prot. | Cad. pépt. | Pépt. sec.           | Pépt. rot. |
|--------|------------|------------|----------------------|------------|
| 1apm   | E          | I          | TTYADFIASGRTGRRNAIHD | 112        |
| 1atp   | E          | I          | TTYADFIASGRTGRRNAIHD | 106        |

Continúa ...

| PDB id | Cad. prot. | Cad. pépt. | Pépt. sec.       | Pépt. rot. |
|--------|------------|------------|------------------|------------|
| 1awr   | C          | I          | HAGPIA           | 20         |
| 1b6j   | AB         | C          | NPIV**1          | 18         |
| 1b9j   | A          | B          | KLK              | 20         |
| 1ce1   | LH         | P          | GTSSPSAD         | 30         |
| 1cka   | A          | B          | PPPALPPKK        | 32         |
| 1cpi   | AB         | C          | NPIV**           | 18         |
| 1czq   | A          | D          | G*****           | 41         |
| 1czy   | C          | E          | PQQATDD          | 33         |
| 1d4t   | A          | B          | KSLTIYAQVQK      | 62         |
| 1d6w   | A          | I          | DFEEIPEEL*       | 49         |
| 1ddv   | A          | B          | TPPSPF           | 20         |
| 1dit   | H          | P          | DPGGGGGNGDFEEI** | 72         |
| 1dkx   | A          | B          | NRLLLT           | 34         |
| 1e8n   | A          | I          | GFGPFG*          | 22         |
| 1eb1   | H          | A          | DYEPIPEEAF       | 45         |
| 1eg4   | A          | P          | NMTPYRSPPPYVP    | 56         |
| 1elw   | A          | C          | GPTIEEVD         | 35         |
| 1er8   | E          | I          | PFHLLVY*         | 36         |
| 1g37   | A          | B          | FEAIPAEYL        | 40         |
| 1g65   | KL         | 3          | IT***            | 24         |
| 1gag   | A          | B          | PATGDFMNMSPVG*   | 70         |
| 1gvk   | B          | A          | NPI*             | 13         |
| 1gyb   | AB         | E          | GFSFG            | 20         |
| 1h6w   | A          | B          | SLNYIIKVKE       | 57         |
| 1h9l   | B          | A          | VEPI             | 16         |
| 1hc9   | A          | C          | WRYESSLLPYPD     | 66         |
| 1heF   | E          | I          | AAV**            | 25         |
| 1i73   | A          | B          | PL*              | 11         |
| 1i8k   | AB         | C          | KKGNVVTDH        | 53         |
| 1iak   | AB         | P          | STDYGLQINSRW     | 70         |
| 1iau   | A          | B          | IEP**            | 25         |
| 1ihj   | B          | C          | TEFCA            | 22         |
| 1jbu   | H          | X          | EEWEVLCWTWETCER  | 61         |
| 1jd5   | A          | B          | AIAYFIPD         | 33         |
| 1jq8   | AB         | P          | LAIYS            | 23         |
| 1jq9   | AB         | P          | FLSYK            | 28         |
| 1jwg   | A          | C          | DEDLLHI          | 35         |
| 1jyq   | A          | L          | N***             | 24         |
| 1kl3   | BC         | F          | HPQFEK           | 32         |
| 1klu   | AB         | C          | GELIGTLNAAKVPAD  | 65         |
| 1lvm   | A          | C          | ENLYFQ*          | 35         |
| 1mfg   | A          | B          | EYLGLDVPV        | 39         |
| 1mvu   | AB         | P          | VVQEALDKAREGR    | 72         |
| 1n12   | C          | D          | SDVAFRGNLLD      | 54         |
| 1n7f   | A          | C          | ATVRTYSC         | 40         |
| 1nln   | A          | B          | GVQSLKRRRCF      | 69         |
| 1nq7   | A          | B          | HKILHRLLE        | 59         |
| 1ntv   | A          | B          | NFDNPVYRKT       | 55         |
| 1nvr   | A          | B          | ASVSA            | 19         |
| 1nx1   | A          | C          | DAIDALSSDFT      | 50         |
| 1oai   | A          | B          | DSGFSFGSK        | 43         |
| 1ot5   | B          | D          | AK**             | 22         |
| 1ou8   | A          | C          | GRHGAANDENY      | 54         |
| 1pcg   | A          | E          | KILCRLLQ*        | 51         |
| 1pz5   | AB         | C          | MDWNMHAA         | 38         |
| 1qkz   | HL         | P          | ANGGASGQVK       | 44         |
| 1r64   | B          | D          | REK**            | 36         |
| 1rxz   | A          | B          | KSTQATLERWF      | 62         |

Continúa ...

<sup>1</sup>El '\*' representa que no es un amino ácido natural.

| PDB id            | Cad. prot. | Cad. pépt. | Pépt. sec.           | Pépt. rot. |
|-------------------|------------|------------|----------------------|------------|
| 1se0              | A          | B          | AIAYFIP              | 28         |
| 1sfi              | A          | I          | GRCKTSIPPICFPD       | 23         |
| 1sio              | A          | D          | IP**                 | 13         |
| 1ssh              | A          | B          | GPPPAMPARPT          | 38         |
| 1svz              | A          | C          | PQFSLW               | 28         |
| 1t4f              | M          | P          | RFMDYWEGL            | 49         |
| 1t7r              | A          | B          | SSRFESLFAG           | 50         |
| 1tg1              | A          | C          | LV*                  | 25         |
| 1tp5              | A          | B          | KKETWV               | 35         |
| 1tw6              | B          | D          | AVPIAQ               | 23         |
| 1u00              | A          | P          | ELPPVKIHC            | 40         |
| 1u8i              | AB         | C          | ELDKWAN              | 37         |
| 1u9l              | AB         | C          | NRPILSL              | 36         |
| 1uj0              | A          | B          | TPMVNRENK            | 51         |
| 1uop              | A          | B          | GFEP                 | 15         |
| 1urc              | B          | E          | RKLFG*               | 15         |
| 1vzq              | H          | I          | DFEEIPE              | 29         |
| 1w9e              | B          | S          | NEFYF                | 27         |
| 1x2r              | A          | B          | LDEETGEFL            | 45         |
| 1xoc <sup>2</sup> | A          | B          | VDSKNTSSW            | 0          |
| 1ymt              | A          | B          | RPTILYALLS           | 63         |
| 1yph              | DF         | B          | CGVPAIQPVLSG         | 48         |
| 1yuc              | A          | C          | ASRPAILYALLSSS       | 65         |
| 1ywo              | A          | P          | QPPVPPQRPM           | 42         |
| 1z9o              | D          | J          | EDEFYDALS            | 44         |
| 1zuk              | AB         | C          | RGPPPPPPHR           | 41         |
| 2a3i              | A          | B          | QKSLQLLQTE           | 71         |
| 2ak5              | B          | D          | RPPKPRPR             | 44         |
| 2aof              | AB         | C          | RPGNLQSRP**          | 100        |
| 2aoh              | AB         | C          | VSFNPQITA*           | 92         |
| 2aoi              | AB         | C          | RPGNLQS*             | 86         |
| 2aoj              | AB         | C          | VSFNPQITAA**         | 98         |
| 2aq9              | A          | X          | WMLDPIAGKWSR         | 57         |
| 2b1z              | B          | D          | KILHRLQLD            | 53         |
| 2b6n              | A          | B          | APT                  | 9          |
| 2b9h              | A          | C          | RRNLKGLNLLH          | 73         |
| 2bba              | A          | P          | NYLFSPNGPIARAW       | 64         |
| 2c1e              | AB         | C          | DEV**                | 28         |
| 2c2k              | AB         | C          | DEV**                | 28         |
| 2c2m              | AB         | C          | DEV**                | 29         |
| 2c2o              | AB         | C          | DEV**                | 29         |
| 2c2z              | AB         | C          | LET**                | 28         |
| 2c3i              | B          | A          | KRRRHPSG             | 52         |
| 2cch              | B          | E          | HTLKGRRLVFDN         | 69         |
| 2cpk              | E          | I          | TTYADFIASGRTGRRNAIHD | 106        |
| 2d0n              | AC         | B          | PSIDRSTKP            | 45         |
| 2d5w              | A          | C          | ASKPK                | 25         |
| 2ds8              | AB         | P          | ALRVVK               | 32         |
| 2dze              | AB         | X          | LARRLR               | 41         |
| 2fgr              | A          | B          | DNWQNGTS             | 40         |
| 2fmf              | A          | B          | QDQVDDLDSLGF         | 65         |
| 2fnt              | AB         | P          | RQVNFLG              | 25         |
| 2foj              | A          | B          | GARAHSS              | 33         |
| 2fvj              | A          | B          | HKLVLQTTT            | 53         |
| 2h9m              | A          | B          | ARTKQT               | 34         |
| 2ho2              | A          | B          | PPPPPPPPPL           | 22         |
| 2hpl              | A          | B          | DDLYG                | 23         |

Continúa ...

<sup>2</sup>No se pudo realizar la preparación del complejo para el acoplamiento molecular. Ver Anexo D, errores .1 y .2.

| PDB id | Cad. prot. | Cad. pépt. | Pépt. sec.                 | Pépt. rot. |
|--------|------------|------------|----------------------------|------------|
| 2ipu   | LH         | Q          | AEFRHDS                    | 38         |
| 2iv9   | AB         | P          | SFGDGFADF                  | 38         |
| 2j6f   | A          | C          | PPKPRRR                    | 44         |
| 2jam   | A          | D          | GVSKFA                     | 27         |
| 2l6e   | A          | B          | ITFDLLYGKKK**              | 62         |
| 2o02   | A          | P          | GHGQGLLDALDLAS             | 62         |
| 2o4j   | A          | C          | KNHPMLMNLKLD               | 56         |
| 2o9v   | A          | B          | VPPPVPPPS                  | 26         |
| 2otu   | AB         | P          | QQQQQQQQQQG                | 72         |
| 2oxw   | A          | X          | IAG                        | 10         |
| 2oy2   | A          | W          | IAG                        | 10         |
| 2p0w   | A          | P          | KGGKGLGKGGAKRHR            | 80         |
| 2p1t   | A          | B          | HKILHRLQD                  | 56         |
| 2p54   | A          | B          | ARHKILHRLQEQE              | 60         |
| 2plx   | A          | B          | QCKVMCYAQRHSSPELLRRCLDNCEK | 82         |
| 2puy   | B          | E          | ARTKQTARKS                 | 62         |
| 2pv2   | AB         | E          | NFTLKFWDIFRK               | 70         |
| 2qab   | A          | C          | KHKILHRLQD                 | 60         |
| 2qos   | C          | A          | LRYDSTAERLY                | 64         |
| 2r7g   | A          | B          | PPTLHELYDL                 | 45         |
| 2rd4   | AB         | C          | LVFFA                      | 20         |
| 2uae   | B          | E          | RLI**                      | 24         |
| 2v3s   | B          | C          | GRFQVT                     | 32         |
| 2vj0   | A          | P          | PKGWVTFE                   | 32         |
| 2w0z   | A          | B          | APPPRPPKP                  | 31         |
| 2w10   | A          | D          | PPPRPTAPKPLL               | 46         |
| 2xfx   | A          | C          | VGYPKVKEEML                | 57         |
| 2zjd   | A          | B          | GGDDDWTHLS                 | 45         |
| 3apr   | E          | I          | PFHVY*                     | 26         |
| 3bfq   | G          | F          | ADSTITIRGYVRDNR            | 87         |
| 3bfw   | C          | D          | ADSTITIRGYVRD              | 69         |
| 3brl   | A          | C          | ATSAKATQTD                 | 48         |
| 3bs4   | A          | B          | NIF                        | 15         |
| 3bu3   | A          | B          | YNYPEDYGDIEIG              | 65         |
| 3bwa   | A          | C          | FPTKDVALL                  | 33         |
| 3ch8   | A          | P          | PQPVDSWV                   | 33         |
| 3che   | A          | C          | D***                       | 21         |
| 3cvp   | A          | B          | NRASKL                     | 26         |
| 3d1e   | A          | P          | GQLGLF                     | 27         |
| 3d9t   | B          | D          | ATPFQE                     | 27         |
| 3ds0   | A          | T          | ITFEDLLDYYGP               | 57         |
| 3ds1   | A          | T          | ITFEDLLDYYG                | 55         |
| 3ds3   | A          | C          | ITFEDLLDYYGP               | 57         |
| 3ds4   | A          | T          | ITFEDLLDYYGP               | 57         |
| 3e2b   | A          | C          | ATSAKATQTD                 | 48         |
| 3eqs   | A          | B          | TSFAEYWNLIS                | 55         |
| 3eqy   | A          | C          | TSFAEYWNLISP               | 57         |
| 3gjq   | AB         | E          | WEHD*                      | 22         |
| 3gjt   | AB         | E          | IEPD*                      | 19         |
| 3gq1   | A          | C          | WLF                        | 14         |
| 3idg   | AB         | C          | ALDKWD                     | 30         |
| 3iwy   | AC         | B          | *****                      | 57         |
| 3jzo   | A          | P          | LTFEHWYWAQLTS              | 61         |
| 3lnj   | AC         | B          | *****                      | 63         |
| 3lny   | A          | B          | EQVSAV                     | 28         |
| 3lnz   | EG         | H          | TSFAEYWALLSP               | 54         |
| 3m17   | A          | I          | RFTGHFGYPC*****            | 30         |
| 3m1b   | C          | J          | FTGHFGYPC***               | 17         |
| 3mgn   | AB         | K          | G*****                     | 40         |
| 3mmg   | AB         | D          | ETVRFQSD                   | 45         |

Continúa ...

| PDB id | Cad. prot. | Cad. pépt. | Pépt. sec.        | Pépt. rot. |
|--------|------------|------------|-------------------|------------|
| 3n55   | A          | B          | PQIINRP           | 39         |
| 3nfk   | AB         | C          | SWESHKSGGETRL     | 70         |
| 3njg   | A          | B          | PQIINRP           | 35         |
| 3obq   | A          | B          | PTPSAPVPL         | 29         |
| 3q47   | B          | C          | NPISDVD           | 31         |
| 3qn7   | A          | B          | ACSRYEVDRCRGRGSAC | 83         |
| 3r5j   | AB         | E          | ADVA**            | 20         |
| 3r6g   | AB         | E          | VDVA**            | 21         |
| 3r6l   | AB         | E          | VDVA**            | 21         |
| 3r7b   | CD         | F          | DVA**             | 17         |
| 3r7n   | AB         | E          | DVA**             | 17         |
| 3rjm   | AB         | E          | VDV***            | 22         |
| 3rq7   | A          | B          | PLS****           | 33         |
| 3s3j   | A          | B          | VPL**             | 20         |
| 3s3p   | A          | B          | QPL**             | 26         |
| 3s3s   | A          | B          | VPL**             | 24         |
| 3t6r   | AB         | D          | ARTKQ             | 31         |
| 3tpx   | AC         | B          | *****             | 61         |
| 3upv   | A          | B          | PTVEEVD           | 31         |
| 3vqg   | A          | B          | VTLV              | 17         |
| 4btb   | A          | C          | PPPPPPPP          | 17         |
| 4c2c   | A          | C          | AVPA              | 11         |
| 4dgy   | HL         | A          | QLINTNGSWHIN      | 62         |
| 4ds1   | AC         | D          | NYAESGIQTDL       | 52         |
| 4eik   | A          | B          | SLARRPLPLP        | 48         |
| 4gly   | A          | B          | CCLGRGCENHRCL*    | 26         |
| 4gpl   | B          | A          | TPEP***           | 23         |
| 4imq   | A          | B          | LF**              | 19         |
| 4imz   | A          | B          | LF**              | 20         |
| 4inh   | A          | M          | AL**              | 18         |
| 4j44   | A          | B          | AIIV              | 14         |
| 4j8s   | A          | B          | RRLPIFNRSVS       | 69         |
| 4ks6   | A          | B          | G****             | 25         |
| 4ktx   | A          | B          | RGL***            | 25         |
| 4n7h   | A          | B          | EAPPSYAEV         | 36         |
| 4nnm   | B          | D          | YPTSII            | 27         |
| 4ofb   | A          | B          | TIDDEYRRKT****    | 75         |
| 4ore   | X          | A          | TFEYGDGI          | 32         |
| 4q6h   | A          | B          | VQDTRL            | 34         |
| 4qbr   | A          | P          | ARTKQTA           | 39         |
| 4rsp   | A          | B          | SVL**             | 25         |
| 4rxz   | A          | C          | TSFAEYWNLIS       | 55         |
| 4v3i   | A          | B          | DLTRP             | 25         |
| 4whh   | A          | B          | S****             | 32         |
| 4whk   | A          | B          | S****             | 31         |
| 4whl   | A          | B          | S****             | 32         |
| 4wlb   | AB         | D          | SLLKKLLD          | 45         |
| 4yec   | AB         | C          | VLT**             | 22         |
| 4zjx   | A          | B          | CRWTKCL**         | 22         |
| 4zvu   | CD         | E          | VEI**             | 21         |
| 5bjt   | E          | V          | RYFCTKWKHGWCCEEVG | 33         |
| 5eel   | BE         | N          | SFEGYDN**         | 15         |
| 5eeq   | B          | M          | SFEGYDNSFP**      | 17         |
| 5hhv   | AB         | I          | IHVTIPADLWDWIN    | 64         |
| 5hhx   | AB         | I          | IHVTIPADLWDWIN    | 64         |
| 5hi3   | AB         | I          | IHVTIPADLWDWIN    | 64         |
| 5hi4   | AB         | I          | IHVTIPADLWDWIN    | 64         |
| 5kgn   | A          | C          | DYPGDYCYLY***     | 32         |
| 5l20   | AB         | C          | VLT**             | 23         |
| 5lak   | AC         | I          | YYN**             | 27         |

Continúa ...

| PDB id | Cad. prot. | Cad. pépt. | Pépt. sec.           | Pépt. rot. |
|--------|------------|------------|----------------------|------------|
| 5lf0   | HI         | c          | IT***                | 23         |
| 5tyi   | A          | L          | KFEGDNE**            | 12         |
| 5u06   | A          | L          | KFEGDNE**            | 16         |
| 5uml   | AG         | C          | LTFLEYWAQLM          | 57         |
| 5umm   | A          | B          | LTFLEYWAQLMQ         | 64         |
| 5va9   | AB         | C          | YGHSHIRFGYSYHVSYCG** | 41         |
| 5vb9   | A          | C          | CWVLEYDMFGALHCR      | 34         |
| 5vk0   | AM         | B          | TSFAEYWLLSC***       | 45         |
| 5vk1   | AI         | B          | TSFEYWLLSP***        | 44         |
| 5yyf   | A          | B          | QTARK**              | 35         |
| 6m8w   | A          | B          | IA**                 | 14         |
| 6m8y   | A          | B          | IP**                 | 13         |

## Acoplamiento molecular rígido global

Resultado del RMSD de las poses de menor energía y su correlación lineal para el acoplamiento molecular rígido global, ver Tabla 3.

**Tabla 3.** Correlación de la pose de menor energía y su RMSD para el acoplamiento molecular rígido global.

| Complejo    | Auto Dock 4 |         | Clus Pro |         | Dock 6   |         | Vina           |         |
|-------------|-------------|---------|----------|---------|----------|---------|----------------|---------|
|             | Energía     | RMSD    | Energía  | RMSD    | Energía  | RMSD    | Energía        | RMSD    |
| 1apm        | -16.481     | 0.8443  | -1409.9  | 1.9704  | -11.3252 | 38.5539 | -28.9          | 0.2843  |
| 1eb1        | -11.311     | 0.6326  | -915.4   | 3.7144  | 10.9658  | 33.6400 | -14.2          | 0.4228  |
| 2aq9        | -7.518      | 39.5733 | -1205.6  | 37.4550 | -28.2690 | 18.6248 | -12.6          | 18.1073 |
| 2cpk        | -19.408     | 0.3153  | -1296.0  | 2.6027  | -18.4525 | 24.4442 | -27.0          | 0.2658  |
| 2plx        | -7.779      | 43.7088 | -1094.5  | 6.2341  | 16.0840  | 13.6141 | -23.0          | 0.6995  |
| 3qn7        | -14.650     | 0.6346  | -1043.5  | 18.1771 | -86.4666 | 1.6460  | -22.8          | 0.1875  |
| 5hhv        | -9.00       | 6.5946  | -1657.3  | 35.8352 | -6.1050  | 48.1382 | -12.5          | 26.8407 |
| 5hi4        | -10.337     | 0.5002  | -1701.3  | 17.3732 | -2.2693  | 80.7725 | * <sup>3</sup> | *       |
| 5va9        | -11.579     | 0.8303  | -1493.0  | 11.5580 | -65.6058 | 24.2678 | -25.1          | 0.3274  |
| 5vb9        | -8.791      | 0.7310  | -1702.4  | 44.7286 | -22.1822 | 19.5871 | -14.4          | 0.4306  |
| Correlación | 0.5701      |         | -0.4861  |         | 0.4694   |         | 0.6349         |         |

## Acoplamiento molecular rígido local

Resultado del RMSD de las poses de menor energía y su correlación lineal para el acoplamiento molecular rígido local para diez complejos, ver Tabla 4.

**Tabla 4.** Correlación de la pose de menor energía y su RMSD para el acoplamiento molecular rígido local.

| Complejo | Auto Dock 4 |        | Dock 6         |         | Vina    |        |
|----------|-------------|--------|----------------|---------|---------|--------|
|          | Energía     | RMSD   | Energía        | RMSD    | Energía | RMSD   |
| 1apm     | -30.786     | 0.6248 | 6.01E+10       | 4.2112  | -28.9   | 0.3225 |
| 1eb1     | -14.559     | 0.5704 | * <sup>4</sup> | *       | -14.2   | 0.2078 |
| 2aq9     | -7.022      | 8.3826 | *              | *       | -9.4    | 9.1561 |
| 2cpk     | -26.824     | 0.5417 | 6.23E+11       | 21.5241 | -27.0   | 0.3101 |
| 2plx     | -21.421     | 1.3869 | 5.93E+09       | 13.9279 | -22.9   | 0.7135 |
| 3qn7     | -21.647     | 0.6686 | 1.15E+09       | 5.1167  | -22.8   | 0.1374 |
| 5hhv     | -18.616     | 0.7782 | *              | *       | -16.1   | 0.4740 |
| 5hi4     | -14.263     | 0.4492 | *              | *       | -14.7   | 0.2573 |

Continúa ...

<sup>3</sup> Memoria excedida. Ver Anexo D, Error .6.

<sup>4</sup> Poses fuera de la caja de acoplamiento molecular.



| Complejo    | Auto Dock 4 |        | Dock 6   |         | Vina    |        |
|-------------|-------------|--------|----------|---------|---------|--------|
|             | Energía     | RMSD   | Energía  | RMSD    | Energía | RMSD   |
| 5va9        | -19.144     | 0.5890 | *        | *       | -25.0   | 0.3388 |
| 5vb9        | -11.275     | 0.5413 | 4.24E+11 | 15.6134 | -14.5   | 0.3328 |
| Correlación | 0.5497      |        | 0.8283   |         | 0.5379  |        |

Resultado del RMSD de las poses de menor energía y su correlación lineal para el acoplamiento molecular rígido local para los complejos de la DockinDB, ver Tabla 5.

**Tabla 5.** Correlación de la pose de menor energía y su RMSD para el acoplamiento molecular rígido local con toda la DockingDB.

| Complejo | Auto Dock 4 |         | Vina    |        |
|----------|-------------|---------|---------|--------|
|          | Energía     | RMSD    | Energía | RMSD   |
| 1apm     | -30.803     | 0.6208  | -28.9   | 0.3225 |
| 1atp     | -26.952     | 0.3922  | -25.8   | 0.2323 |
| 1awr     | -10.893     | 0.6494  | -11.6   | 0.3584 |
| 1b6j     | -17.995     | 0.4545  | -20.4   | 0.5185 |
| 1b9j     | -17.649     | 0.3643  | -15.8   | 0.1927 |
| 1ce1     | -12.95      | 0.3075  | -15.7   | 0.3119 |
| 1cka     | -17.859     | 0.4417  | -12.6   | 0.149  |
| 1cpi     | -17.8       | 0.3479  | -20.0   | 0.3585 |
| 1czq     | -7.334      | 13.4478 | -8.0    | 1.1879 |
| 1czy     | -11.502     | 0.3101  | -12.9   | 0.5841 |
| 1d4t     | -21.673     | 0.4579  | -22.7   | 0.1352 |
| 1d6w     | -14.856     | 0.7473  | -15.1   | 0.4511 |
| 1ddv     | -6.786      | 0.8129  | -8.9    | 0.182  |
| 1dit     | -17.857     | 1.2781  | -20.1   | 0.4228 |
| 1dkx     | -12.943     | 0.5318  | -14.3   | 0.4474 |
| 1e8n     | -14.014     | 0.3372  | -15.2   | 0.2527 |
| 1eb1     | -14.537     | 0.5896  | -14.1   | 0.3811 |
| 1eg4     | -11.967     | 0.6038  | -13.2   | 0.4725 |
| 1elw     | -14.555     | 0.2609  | -13.0   | 0.2426 |
| 1er8     | -19.121     | 0.3756  | -22.3   | 0.122  |
| 1g37     | -13.618     | 1.1794  | -13.4   | 0.5334 |
| 1g65     | -7.75       | 1.262   | -7.4    | 9.8384 |
| 1gag     | -21.076     | 0.3395  | -27.3   | 0.2818 |
| 1gvk     | -7.659      | 1.4694  | -5.4    | 0.9151 |
| 1gyb     | -9.537      | 0.779   | -9.4    | 0.4986 |
| 1h6w     | -30.894     | 0.412   | -28.8   | 0.3149 |
| 1h9l     | -7.525      | 0.9918  | -6.8    | 0.524  |
| 1hc9     | -18.552     | 0.447   | -19.6   | 0.1797 |
| 1hef     | -8.343      | 0.9132  | -6.6    | 1.3337 |
| 1i73     | -5.849      | 0.5183  | -8.9    | 0.7568 |
| 1i8k     | -20.493     | 0.4984  | -20.2   | 0.1539 |
| 1iak     | -28.673     | 0.3425  | -35.6   | 0.1902 |
| 1iau     | -16.426     | 1.2083  | -11.8   | 1.4301 |
| 1ihj     | -8.307      | 1.2025  | -8.2    | 1.6865 |
| 1jbu     | -29.729     | 0.3716  | -27.9   | 0.1756 |
| 1jd5     | -17.025     | 0.3477  | -13.2   | 0.2554 |
| 1jq8     | -11.045     | 0.3733  | -6.6    | 0.445  |
| 1jq9     | -8.269      | 0.5588  | -0.2    | 0.2093 |
| 1jwg     | -15.635     | 0.4774  | -14.4   | 0.2271 |
| 1jyq     | -21.358     | 0.2668  | -18.7   | 0.2615 |
| 1kl3     | -9.617      | 0.6269  | -14.3   | 0.431  |
| 1klu     | -27.009     | 0.3813  | -30.1   | 0.1248 |
| 1lvm     | -23.849     | 0.3048  | -24.7   | 0.104  |
| 1mfg     | -12.776     | 0.4361  | -14.9   | 0.2724 |
| 1mvu     | -9.957      | 0.4286  | -12.7   | 0.4327 |

Continúa ...

| Complejo | Auto Dock 4 |         | Vina    |         |
|----------|-------------|---------|---------|---------|
|          | Energía     | RMSD    | Energía | RMSD    |
| 1n12     | *5          | *       | -28.9   | 0.0825  |
| 1n7f     | -13.286     | 1.1063  | -13.0   | 0.2051  |
| 1nln     | -28.236     | 0.6196  | -21.7   | 0.461   |
| 1nq7     | -11.831     | 0.6133  | -10.9   | 0.3883  |
| 1ntv     | -20.04      | 0.4047  | -21.6   | 0.1783  |
| 1nvr     | -8.497      | 10.421  | -8.2    | 0.5326  |
| 1nx1     | -15.124     | 0.7029  | -17.1   | 0.4757  |
| 1oai     | -10.494     | 0.9298  | -12.2   | 0.8282  |
| 1ot5     | *6          | *       | *       | *       |
| 1ou8     | -16.045     | 0.3944  | -18.3   | 0.173   |
| 1pcg     | -13.759     | 0.778   | -9.9    | 0.4373  |
| 1pz5     | -12.449     | 0.3087  | -15.5   | 0.2952  |
| 1qkz     | -14.307     | 0.895   | -16.6   | 0.3673  |
| 1r64     | *7          | *       | *       | *       |
| 1rxz     | -17.192     | 0.5628  | -16.7   | 0.699   |
| 1se0     | -15.901     | 0.4698  | -13.4   | 0.3364  |
| 1sfi     | -18.678     | 0.5509  | -19.7   | 0.2157  |
| 1sio     | -8.751      | 0.989   | -7.9    | 1.375   |
| 1ssh     | -12.864     | 0.5254  | -12.0   | 0.0574  |
| 1svz     | -17.589     | 0.3738  | -19.8   | 0.1882  |
| 1t4f     | -12.0       | 0.3782  | -13.5   | 0.347   |
| 1t7r     | -15.183     | 0.4377  | -12.0   | 0.3014  |
| 1tg1     | -7.157      | 8.4989  | -7.0    | 11.8031 |
| 1tp5     | -14.726     | 0.622   | -12.2   | 0.4256  |
| 1tw6     | -12.523     | 0.5095  | -10.9   | 0.0735  |
| 1u00     | -15.438     | 0.4775  | -16.8   | 0.3111  |
| 1u8i     | -14.952     | 0.4664  | -13.7   | 0.5609  |
| 1u9l     | -17.125     | 0.5012  | -19.1   | 0.147   |
| 1uj0     | -14.122     | 0.9998  | -10.7   | 0.5422  |
| 1uop     | -10.274     | 0.3956  | -12.2   | 0.1236  |
| 1urc     | -10.806     | 1.0008  | -12.8   | 0.4981  |
| 1vzq     | -11.48      | 1.0251  | -11.4   | 0.4832  |
| 1w9e     | -13.133     | 0.3394  | -14.3   | 0.1869  |
| 1x2r     | -13.642     | 0.3702  | -14.9   | 0.1486  |
| 1xoc     | *8          | *       | *       | *       |
| 1ymt     | -14.602     | 0.6842  | -14.7   | 0.5118  |
| 1yph     | -11.279     | 0.9525  | -13.6   | 0.4683  |
| 1yuc     | -13.751     | 0.7601  | -15.0   | 0.4154  |
| 1ywo     | -11.005     | 0.5786  | -11.0   | 0.1573  |
| 1z9o     | -13.864     | 0.4318  | -13.4   | 0.3459  |
| 1zuk     | -21.825     | 0.5801  | -20.1   | 0.0856  |
| 2a3i     | -13.439     | 0.5977  | -10.4   | 0.5327  |
| 2ak5     | -10.442     | 16.4746 | -10.5   | 16.4588 |
| 2aof     | -20.558     | 18.5338 | -23.0   | 0.4113  |
| 2aoh     | -22.894     | 0.2525  | -27.0   | 0.2841  |
| 2aoi     | -19.49      | 0.3272  | -21.9   | 16.0236 |
| 2aoj     | -26.363     | 0.3515  | -28.0   | 0.5545  |
| 2aq9     | -7.018      | 8.3783  | -9.4    | 9.1457  |
| 2b1z     | -13.073     | 0.5887  | -10.2   | 0.2694  |
| 2b6n     | -4.862      | 6.9363  | -5.3    | 6.9938  |
| 2b9h     | -18.188     | 0.6652  | -16.1   | 0.3463  |
| 2bba     | -17.926     | 0.5454  | -20.1   | 0.139   |
| 2c1e     | -15.892     | 0.7959  | -14.3   | 0.6628  |

Continúa ...

<sup>5</sup>Error durante la construcción de la malla de puntos. Ver Anexo D, Error .3.

<sup>6</sup>Error en los cálculos de acoplamiento molecular. Ver anexo D, errores .4 y .5.

<sup>7</sup>Idem.

<sup>8</sup>No se pudo realizar la preparación del complejo para el acoplamiento molecular. Ver Anexo D, errores .1 y .2.

| Complejo | Auto Dock 4 |        | Vina    |        |
|----------|-------------|--------|---------|--------|
|          | Energía     | RMSD   | Energía | RMSD   |
| 2c2k     | -15.55      | 0.8783 | -14.2   | 1.0295 |
| 2c2m     | -15.818     | 1.1613 | -15.2   | 1.3137 |
| 2c2o     | -15.772     | 0.7875 | -14.4   | 0.848  |
| 2c2z     | -15.134     | 0.5806 | -13.5   | 0.3919 |
| 2c3i     | -19.55      | 0.4591 | -15.6   | 0.1829 |
| 2cch     | -19.913     | 0.5346 | -20.3   | 0.1273 |
| 2cpk     | -26.838     | 0.6583 | -27.0   | 0.3352 |
| 2d0n     | -17.205     | 0.4255 | -13.8   | 0.2724 |
| 2d5w     | -17.503     | 0.3201 | -17.4   | 0.2389 |
| 2ds8     | -14.065     | 0.2932 | -11.5   | 0.3235 |
| 2dze     | -20.702     | 0.364  | -18.0   | 0.2627 |
| 2fgr     | -9.224      | 1.6857 | -9.9    | 0.6146 |
| 2fmf     | -10.19      | 0.9603 | -9.6    | 0.5135 |
| 2fnt     | -24.959     | 0.4349 | -30.0   | 0.2131 |
| 2foj     | -7.953      | 0.7269 | -9.8    | 0.2732 |
| 2fvj     | -12.088     | 0.6573 | -11.0   | 0.2479 |
| 2h9m     | -11.221     | 0.7645 | -12.3   | 0.1649 |
| 2ho2     | -9.503      | 4.8879 | -9.1    | 4.8079 |
| 2hpl     | -12.173     | 0.3775 | -10.7   | 0.3144 |
| 2ipu     | -17.43      | 0.581  | -19.6   | 0.2508 |
| 2iv9     | -14.526     | 0.4192 | -20.8   | 0.2471 |
| 2j6f     | -16.111     | 0.5061 | -11.8   | 0.2707 |
| 2jam     | -9.255      | 0.6774 | -8.6    | 0.7119 |
| 2l6e     | -16.421     | 0.3579 | -17.6   | 0.2929 |
| 2o02     | -20.242     | 0.4757 | -23.6   | 0.2683 |
| 2o4j     | -11.659     | 0.4894 | -10.6   | 0.2455 |
| 2o9v     | -8.761      | 0.5803 | -10.7   | 0.274  |
| 2otu     | -19.601     | 0.2397 | -21.7   | 0.2909 |
| 2oxw     | -8.86       | 0.997  | -6.0    | 6.8043 |
| 2oy2     | -10.312     | 0.4225 | -8.7    | 0.1512 |
| 2p0w     | -24.027     | 0.3264 | -21.0   | 0.1089 |
| 2p1t     | -14.457     | 0.8334 | -13.8   | 0.3331 |
| 2p54     | -12.647     | 0.6006 | -12.5   | 0.3977 |
| 2plx     | -21.507     | 1.5151 | -23.0   | 0.6728 |
| 2puy     | -15.142     | 0.4881 | -14.7   | 0.1435 |
| 2pv2     | -22.296     | 0.5267 | -30.2   | 0.2434 |
| 2qab     | -12.389     | 0.9578 | -11.4   | 0.427  |
| 2qos     | -13.139     | 0.365  | -18.1   | 0.1662 |
| 2r7g     | -16.01      | 0.486  | -19.0   | 0.231  |
| 2rd4     | -11.317     | 0.6662 | -12.5   | 0.4346 |
| 2uue     | -9.099      | 1.741  | -10.1   | 0.4946 |
| 2v3s     | -13.958     | 0.3689 | -12.5   | 0.1239 |
| 2vj0     | -11.284     | 0.3348 | -12.3   | 0.1955 |
| 2w0z     | -14.191     | 0.4181 | -12.0   | 0.1594 |
| 2w10     | -18.861     | 0.5371 | -15.7   | 0.3799 |
| 2xfx     | -30.876     | 0.2384 | -30.7   | 0.1818 |
| 2zjd     | -17.783     | 0.5402 | -18.2   | 0.3116 |
| 3apr     | -21.016     | 0.3323 | -21.8   | 0.1559 |
| 3bfq     | -42.523     | 0.1423 | -37.7   | 0.2038 |
| 3bfw     | -39.522     | 0.1505 | -33.6   | 0.1115 |
| 3brl     | -18.017     | 0.312  | -17.4   | 0.2586 |
| 3bs4     | -14.328     | 0.3651 | -14.8   | 0.4669 |
| 3bu3     | -26.235     | 0.1915 | -29.7   | 0.1787 |
| 3bwa     | -26.006     | 0.2845 | -33.4   | 0.298  |
| 3ch8     | -20.892     | 0.3735 | -23.4   | 0.3285 |
| 3che     | -11.405     | 0.3323 | -11.6   | 0.6239 |
| 3cvp     | -13.817     | 0.3082 | -14.8   | 0.2229 |
| 3d1e     | -13.472     | 0.4808 | -12.5   | 0.37   |
| 3d9t     | -14.58      | 0.6478 | -12.0   | 0.0869 |

Continúa ...

| Complejo | Auto Dock 4     |        | Vina    |        |
|----------|-----------------|--------|---------|--------|
|          | Energía         | RMSD   | Energía | RMSD   |
| 3ds0     | -12.537         | 0.7394 | -17.4   | 0.3066 |
| 3ds1     | -14.438         | 0.6352 | -18.1   | 0.2735 |
| 3ds3     | -12.542         | 0.907  | -15.6   | 0.3029 |
| 3ds4     | -15.727         | 0.6507 | -19.6   | 0.1103 |
| 3e2b     | -17.101         | 0.4004 | -17.6   | 0.2936 |
| 3eqs     | -15.673         | 0.7167 | -18.2   | 0.1667 |
| 3eqy     | -14.863         | 0.5435 | -17.8   | 0.1827 |
| 3gjq     | -12.159         | 1.5158 | -11.6   | 0.8217 |
| 3gjt     | -9.22           | 0.6118 | -9.1    | 0.2977 |
| 3gq1     | -12.178         | 0.6004 | -10.4   | 0.2579 |
| 3idg     | -13.644         | 0.4022 | -12.9   | 0.4422 |
| 3iwy     | -19.406         | 0.4076 | -20.8   | 0.2835 |
| 3jzo     | -13.73          | 0.4752 | -15.5   | 0.3464 |
| 3lnj     | -18.442         | 0.538  | -18.8   | 0.3296 |
| 3lny     | -11.826         | 0.1905 | -12.8   | 0.3297 |
| 3lnz     | -20.603         | 0.4341 | -25.5   | 0.1508 |
| 3m17     | -10.915         | 0.8271 | -13.1   | 0.6602 |
| 3m1b     | -9.845          | 1.0238 | -11.5   | 1.0779 |
| 3mgn     | -13.021         | 0.8772 | -14.5   | 0.2734 |
| 3mmg     | -23.013         | 0.3186 | -25.4   | 0.1687 |
| 3n55     | * <sup>9</sup>  | *      | -20.4   | 0.3211 |
| 3nfk     | -22.882         | 0.2504 | -21.2   | 0.3367 |
| 3njg     | -23.082         | 0.3558 | -20.4   | 0.1918 |
| 3obq     | -12.015         | 0.6092 | -12.9   | 0.2296 |
| 3q47     | -17.243         | 0.473  | -14.4   | 0.1867 |
| 3qn7     | -21.505         | 0.6718 | -22.8   | 0.2526 |
| 3r5j     | -16.569         | 0.1764 | -12.3   | 0.2931 |
| 3r6g     | -16.061         | 0.6823 | -12.6   | 0.2259 |
| 3r6l     | -15.478         | 0.6351 | -10.8   | 0.2071 |
| 3r7b     | -15.588         | 0.2175 | -11.1   | 0.3008 |
| 3r7n     | -14.603         | 0.1869 | -10.2   | 0.3527 |
| 3rjm     | -14.587         | 0.7349 | -10.0   | 0.7506 |
| 3rq7     | -19.352         | 0.401  | -20.0   | 0.1524 |
| 3s3j     | -11.594         | 0.8195 | -10.7   | 0.8478 |
| 3s3p     | -11.214         | 0.9405 | -8.3    | 0.6294 |
| 3s3s     | -10.744         | 0.8896 | -9.8    | 0.6037 |
| 3t6r     | -16.703         | 0.3885 | -13.4   | 0.2755 |
| 3tpx     | -19.46          | 0.5313 | -23.1   | 0.2373 |
| 3upv     | -16.052         | 0.4577 | -14.1   | 0.2972 |
| 3vqg     | -10.444         | 0.2931 | -11.0   | 0.149  |
| 4btb     | -11.208         | 0.7129 | -12.4   | 0.279  |
| 4c2c     | * <sup>10</sup> | *      | -10.8   | 0.4894 |
| 4dgy     | -14.925         | 0.5975 | -18.6   | 0.4356 |
| 4ds1     | -22.356         | 0.1871 | -22.5   | 0.1264 |
| 4eik     | -16.91          | 0.3287 | -17.1   | 0.1234 |
| 4gly     | -22.137         | 0.3976 | -23.7   | 0.1979 |
| 4gpl     | -12.216         | 0.3577 | -13.0   | 0.4405 |
| 4imq     | -9.893          | 0.7881 | -6.6    | 0.88   |
| 4imz     | -14.182         | 0.7876 | -11.6   | 0.8192 |
| 4inh     | -11.363         | 0.8786 | -10.0   | 1.0438 |
| 4j44     | -10.852         | 0.4976 | -9.3    | 0.3226 |
| 4j8s     | -16.815         | 0.677  | -18.7   | 0.1756 |
| 4ks6     | -15.355         | 0.6415 | -14.6   | 0.4155 |
| 4ktx     | -13.84          | 0.6001 | -12.3   | 0.6425 |
| 4n7h     | -9.373          | 0.6156 | -11.6   | 0.4217 |
| 4nnm     | -12.901         | 0.4596 | -14.4   | 0.2596 |

Continúa ...

<sup>9</sup>Error durante la construcción de la malla de puntos. Ver Anexo D, Error .3.

<sup>10</sup>Idem.

| Complejo    | Auto Dock 4 |        | Vina    |        |
|-------------|-------------|--------|---------|--------|
|             | Energía     | RMSD   | Energía | RMSD   |
| 4ofb        | -18.092     | 0.4652 | -16.5   | 0.3583 |
| 4ore        | -12.88      | 0.4775 | -15.0   | 0.362  |
| 4q6h        | -12.567     | 0.3341 | -12.7   | 0.2511 |
| 4qbr        | -20.767     | 0.3697 | -14.3   | 0.2251 |
| 4rsp        | -15.778     | 0.6526 | -14.3   | 0.6494 |
| 4rxz        | -11.934     | 0.5088 | -14.6   | 0.1762 |
| 4v3i        | -6.41       | 0.6874 | -7.7    | 0.9967 |
| 4whh        | -18.819     | 0.441  | -19.6   | 0.1413 |
| 4whk        | -16.365     | 0.2724 | -17.0   | 0.1786 |
| 4whl        | -16.917     | 0.3775 | -18.2   | 0.2687 |
| 4wlb        | -14.041     | 0.7199 | -12.5   | 0.3402 |
| 4yec        | -12.036     | 1.1227 | -9.5    | 0.8932 |
| 4zjx        | -17.294     | 0.3685 | -14.9   | 0.3911 |
| 4zvu        | -11.282     | 0.5351 | -8.8    | 0.8899 |
| 5bjt        | -7.033      | 13.772 | -7.5    | 0.3417 |
| 5eel        | -17.316     | 0.5007 | -19.4   | 0.3195 |
| 5eeq        | -14.812     | 0.4912 | -15.4   | 0.2352 |
| 5hhv        | -18.392     | 0.5774 | -16.1   | 0.5211 |
| 5hhx        | -18.194     | 0.9071 | -17.4   | 0.3625 |
| 5hi3        | -21.862     | 0.4009 | -23.7   | 0.2151 |
| 5hi4        | -14.038     | 0.446  | -14.7   | 0.2573 |
| 5kgn        | -14.375     | 0.651  | -22.3   | 0.3595 |
| 5l20        | -11.063     | 1.1347 | -8.1    | 0.7933 |
| 5lak        | -13.982     | 0.6589 | -13.5   | 0.5569 |
| 5lf0        | -11.863     | 0.9622 | -9.1    | 1.3006 |
| 5tyi        | -14.856     | 0.5762 | -15.0   | 0.3089 |
| 5u06        | -17.891     | 0.4039 | -16.1   | 0.466  |
| 5uml        | -14.888     | 0.4765 | -17.7   | 0.3049 |
| 5umm        | -17.179     | 0.5853 | -19.4   | 0.3006 |
| 5va9        | -17.322     | 0.5359 | -23.7   | 0.3647 |
| 5vb9        | -11.014     | 0.4997 | -13.1   | 0.2535 |
| 5vk0        | -14.483     | 0.6282 | -17.8   | 0.3988 |
| 5vk1        | -13.594     | 0.5171 | -16.8   | 0.318  |
| 5yyf        | -13.774     | 0.4186 | -13.5   | 0.2327 |
| 6m8w        | -10.116     | 1.04   | -8.9    | 1.7018 |
| 6m8y        | -9.369      | 1.1496 | -6.2    | 1.3653 |
| Correlación | 0.2212      |        | 0.1910  |        |

## Acoplamiento molecular flexible local

Resultado del RMSD de las poses de menor energía y su correlación lineal para el acoplamiento molecular flexible local para los complejos flexibles de la DockinDB, ver Tabla 6.

**Tabla 6.** Correlación de la pose de menor energía y su RMSD para el acoplamiento molecular rígido local con los complejos flexibles.

| Complejo | Auto Dock 4 |        | Vina    |        |
|----------|-------------|--------|---------|--------|
|          | Energía     | RMSD   | Energía | RMSD   |
| 1awr     | -16.472     | 7.6452 | -6.4    | 5.6054 |
| 1b6j     | -19.592     | 3.1584 | -10.7   | 0.8328 |
| 1b9j     | -22.527     | 2.6928 | -8.4    | 1.1152 |
| 1ce1     | -22.009     | 5.6206 | -5.6    | 2.0131 |
| 1cka     | -24.58      | 9.2196 | -4.7    | 9.7505 |
| 1cpi     | -22.005     | 2.8503 | -10.6   | 0.7555 |
| 1ddv     | -20.465     | 5.9177 | -5.6    | 9.2705 |
| 1e8n     | -20.689     | 4.4591 | -8.4    | 5.3917 |

Continúa ...

| Complejo | Auto Dock 4 |         | Vina    |         |
|----------|-------------|---------|---------|---------|
|          | Energía     | RMSD    | Energía | RMSD    |
| 1g65     | -17.064     | 8.4282  | -6.6    | 4.2332  |
| 1gvk     | -14.082     | 3.1216  | -6.4    | 6.6292  |
| 1gyb     | -17.591     | 3.3495  | -4.7    | 7.9199  |
| 1h9l     | -16.8       | 3.2276  | -6.8    | 7.8483  |
| 1hef     | -21.327     | 9.3824  | -6.2    | 9.2272  |
| 1i73     | -11.222     | 5.7919  | -7.3    | 7.7055  |
| 1iau     | -21.065     | 3.5855  | -9.6    | 3.2595  |
| 1ihj     | -19.746     | 7.2848  | -4.3    | 2.4125  |
| 1jq8     | -20.374     | 4.3352  | -8      | 9.6378  |
| 1jq9     | -12.072     | 9.1521  | -7.4    | 10.3953 |
| 1jyq     | -19.321     | 7.7959  | -9.1    | 1.717   |
| 1kl3     | -23.897     | 6.3867  | -7.1    | 7.1882  |
| 1nvr     | -17.908     | 6.5874  | -5.2    | 8.6424  |
| 1ot5     | *11         | *       | *       | *       |
| 1se0     | -20.597     | 7.0211  | -5.5    | 11.0825 |
| 1sfi     | -29.821     | 1.1732  | -9.2    | 1.9785  |
| 1sio     | -13.726     | 6.2423  | -5.5    | 5.1893  |
| 1svz     | -20.707     | 5.3114  | -8.3    | 2.2393  |
| 1tg1     | -11.144     | 12.0318 | -7.1    | 12.0652 |
| 1tw6     | -16.733     | 4.1757  | -5.9    | 10.3912 |
| 1uop     | -18.318     | 6.4846  | -6.9    | 6.0906  |
| 1urc     | -14.622     | 6.4436  | -7.7    | 1.7923  |
| 1vzq     | -22.082     | 5.781   | -5      | 5.5735  |
| 1w9e     | -21.945     | 6.324   | -6.2    | 9.3097  |
| 2b6n     | -11.676     | 5.5742  | -4.1    | 6.8986  |
| 2c1e     | -18.068     | 10.3572 | -7.1    | 9.635   |
| 2c2k     | -17.175     | 6.424   | -6.9    | 5.1114  |
| 2c2m     | -19.783     | 11.0602 | -7.9    | 4.2128  |
| 2c2o     | -18.881     | 11.1821 | -7.6    | 11.1329 |
| 2c2z     | -21.224     | 9.5482  | -7.7    | 6.7203  |
| 2d5w     | -19.521     | 4.8217  | -9.1    | 2.9108  |
| 2ds8     | -21.664     | 9.9064  | -4.1    | 6.9488  |
| 2fnt     | -34.331     | 2.3533  | -11.6   | 5.4578  |
| 2ho2     | -20.937     | 8.8568  | -5.6    | 14.8211 |
| 2hpl     | -19.27      | 2.7437  | -5.5    | 7.1794  |
| 2jam     | -19.186     | 10.0284 | -4.5    | 10.1537 |
| 2oxw     | -12.238     | 1.7907  | -5.8    | 2.3747  |
| 2oy2     | -14.218     | 2.57    | -6.4    | 1.6835  |
| 2rd4     | -17.438     | 9.3189  | -8.4    | 8.5256  |
| 2uue     | -15.973     | 5.4533  | -5.8    | 3.3037  |
| 2v3s     | -19.797     | 7.106   | -5.3    | 3.4461  |
| 2vj0     | -21.286     | 8.5861  | -5.3    | 6.6704  |
| 2w0z     | -22.862     | 9.0209  | -4.8    | 9.3626  |
| 3apr     | -22.458     | 10.4675 | -9.2    | 10.7217 |
| 3bs4     | -16.821     | 1.8924  | -8.9    | 1.562   |
| 3che     | -14.165     | 6.263   | -7.1    | 6.6345  |
| 3cvp     | -18.903     | 9.5464  | -6.9    | 5.6471  |
| 3d1e     | -16.924     | 7.195   | -6      | 3.8053  |
| 3d9t     | -20.756     | 8.6428  | -6.6    | 4.5841  |
| 3gjq     | -17.736     | 3.2175  | -7      | 9.3226  |
| 3gjt     | -17.897     | 8.1842  | -7.6    | 8.2033  |
| 3gq1     | -15.639     | 3.2805  | -6.1    | 4.2796  |
| 3idg     | -23.816     | 5.5086  | -6.2    | 4.3903  |
| 3lny     | -20.285     | 8.7075  | -5.8    | 11.8209 |
| 3m17     | -25.679     | 4.9514  | -5.2    | 4.5981  |
| 3m1b     | -19.49      | 1.6662  | -5.9    | 1.6143  |
| 3obq     | -23.733     | 11.3537 | -6.3    | 13.198  |
| 3q47     | -23.844     | 7.3554  | -5.8    | 9.4086  |

Continúa ...

<sup>11</sup>Error en los cálculos de acoplamiento molecular. Ver anexo D, errores .4 y .5.

| Complejo    | Auto Dock 4 |         | Vina    |         |
|-------------|-------------|---------|---------|---------|
|             | Energía     | RMSD    | Energía | RMSD    |
| 3r5j        | -15.684     | 3.8429  | -7.4    | 2.045   |
| 3r6g        | -16.457     | 4.2663  | -7.5    | 2.0077  |
| 3r6l        | -15.676     | 8.4118  | -6.7    | 4.1773  |
| 3r7b        | -17.665     | 1.9531  | -7.4    | 2.3496  |
| 3r7n        | -15.962     | 1.9813  | -6.8    | 2.8435  |
| 3rjm        | -15.232     | 10.44   | -7      | 2.7133  |
| 3s3j        | -17.098     | 4.2105  | -6.6    | 3.1848  |
| 3s3p        | -17.029     | 7.7951  | -5.2    | 9.2678  |
| 3s3s        | -16.659     | 9.7011  | -5.8    | 9.0955  |
| 3t6r        | -21.988     | 7.1698  | -5.3    | 8.3389  |
| 3upv        | -20.434     | 10.1183 | -6.4    | 5.6704  |
| 3vqg        | -16.957     | 6.4196  | -5.8    | 7.2378  |
| 4btb        | -21.662     | 11.4722 | -6.7    | 4.4883  |
| 4c2c        | *12         | *       | -7.3    | 2.0543  |
| 4gly        | -22.64      | 2.7134  | -10.3   | 1.313   |
| 4gpl        | -18.939     | 7.3707  | -6.6    | 12.7198 |
| 4imq        | -15.668     | 5.2283  | -6.3    | 8.5177  |
| 4imz        | -17.761     | 6.663   | -7.6    | 2.3914  |
| 4inh        | -13.851     | 4.1476  | -6.8    | 1.5155  |
| 4j44        | -14.975     | 4.2699  | -5.4    | 1.7099  |
| 4ks6        | -20.323     | 8.936   | -7.8    | 9.0586  |
| 4ktx        | -19.158     | 7.1056  | -7.4    | 4.3336  |
| 4nnm        | -20.341     | 10.3101 | -6.7    | 3.1744  |
| 4ore        | -22.852     | 7.0279  | -7.1    | 6.4379  |
| 4rsp        | -19.081     | 9.9329  | -7.6    | 1.9443  |
| 4v3i        | -18.991     | 6.9167  | -4.8    | 7.3484  |
| 4whh        | -21.566     | 11.018  | -8.4    | 6.2136  |
| 4whk        | -20.972     | 10.1258 | -7.8    | 4.6585  |
| 4whl        | -20.164     | 5.457   | -8      | 1.885   |
| 4yec        | -17.45      | 9.2587  | -5.8    | 4.2184  |
| 4zjx        | -21.909     | 5.134   | -8.4    | 6.5224  |
| 4zvu        | -16.655     | 6.4319  | -6      | 6.8131  |
| 5eel        | -20.647     | 1.9874  | -11.2   | 0.4207  |
| 5eeq        | -21.122     | 1.3795  | -8.2    | 1.0189  |
| 5kgn        | -29.178     | 2.4343  | -8.8    | 2.6989  |
| 5l20        | -15.958     | 8.0592  | -5.6    | 3.9478  |
| 5lak        | -21.673     | 8.689   | -7.5    | 5.3349  |
| 5lf0        | -18.736     | 2.6612  | -6.1    | 4.2564  |
| 5tyi        | -19.833     | 1.0883  | -9.6    | 0.8739  |
| 5u06        | -23.266     | 1.6591  | -9      | 1.367   |
| 6m8w        | -13.878     | 2.5168  | -6.8    | 2.3234  |
| 6m8y        | -13.983     | 1.8695  | -6      | 1.9793  |
| Correlación | 0.0067      |         | 0.3849  |         |

## Predicción del RMSD del Acoplamiento Molecular Rígido Local

Resultados de las predicciones del RMSD a partir de los algoritmos de predicción del Weka. *Todas* significa a las poses del AutoDock4 y del AutoDock Vina que predicen las poses del AutoDock4 y del AutoDock Vina; *Solo AutoDock4* significa que las poses del AutoDock4 predicen las poses del AutoDock4; *Solo Vina* significa que las poses del AutoDock Vina predicen las poses del AutoDock Vina; *Todas AutoDock4* significa que las poses del AutoDock4 y del AutoDock Vina predicen las poses del AutoDock4; y *Todas Vina* significa que las poses del AutoDock4 y del AutoDock Vina predicen las poses del AutoDock Vina. *CC* representa el Coeficiente de Correlación y *MAE* representa el Error Relativo Medio. Ver Tabla 7.

<sup>12</sup>Error durante la construcción de la malla de puntos. Ver Anexo D, Error .3.

**Tabla 7.** Resultados de la predicción del RMSD de las poses para el acoplamiento molecular dirigido rígido.

| Algoritmo                | Todas  |        | Solo AutoDock4 |        | Solo Vina |        | Todas AutoDock4 |        | Todas Vina |        |
|--------------------------|--------|--------|----------------|--------|-----------|--------|-----------------|--------|------------|--------|
|                          | CC     | MAE    | CC             | MAE    | CC        | MAE    | CC              | MAE    | CC         | MAE    |
| Gaussian Processes       | 0.6758 | 2.7909 | 0.1671         | 1.4927 | 0.5248    | 3.5154 | 0.1694          | 2.2140 | 0.5318     | 3.5561 |
| Linear Regression        | 0.6723 | 2.8065 | 0.1796         | 1.5340 | 0.5288    | 3.4871 | 0.1636          | 2.2508 | 0.5314     | 3.5447 |
| Multi-layer Perceptron   | 0.6679 | 2.6119 | 0.1684         | 3.2228 | 0.3667    | 4.2382 | 0.1647          | 1.7626 | 0.4252     | 3.8124 |
| Simple Linear Regression | 0.5797 | 3.1743 | 0.1963         | 0.7072 | 0.3373    | 4.0023 | 0.4165          | 2.2944 | 0.2378     | 4.3573 |
| SMOreg                   | 0.6870 | 2.6022 | 0.2883         | 0.3537 | 0.4978    | 3.5873 | 0.1741          | 1.7856 | 0.5054     | 3.7380 |
| IBk                      | 0.4003 | 3.9234 | 0.0098         | 0.6931 | 0.2895    | 4.2285 | 0.1062          | 3.6816 | 0.2778     | 4.2556 |
| KStar                    | 0.7337 | 1.8330 | -0.0454        | 0.8923 | 0.4037    | 3.9465 | 0.5924          | 0.4138 | 0.4254     | 3.8598 |
| Decicion Table           | 0.6083 | 2.9010 | -0.0594        | 0.6054 | 0.4806    | 3.6428 | 0.3794          | 2.1355 | 0.3670     | 3.9579 |
| M5 Rules                 | 0.7985 | 1.6390 | 0.8929         | 0.4220 | 0.5791    | 3.2431 | 0.4519          | 0.5370 | 0.5821     | 3.1593 |
| Zero R                   | 0      | 4.2171 | 0              | 0.6008 | 0         | 4.3052 | 0               | 3.4079 | 0          | 5.3485 |
| Decision Stump           | 0.7176 | 1.9963 | 0.5388         | 0.7013 | 0.4115    | 3.7138 | 0.5388          | 0.7966 | 0.4107     | 3.6072 |
| M5P                      | 0.7262 | 1.7538 | 0.3536         | 0.6511 | 0.5796    | 3.1674 | 0.3286          | 0.7933 | 0.5657     | 3.1125 |
| Random Forest            | 0.8020 | 1.7367 | 0.6058         | 0.4888 | 0.5722    | 3.1616 | 0.5166          | 0.7117 | 0.5749     | 3.1511 |
| Random Tree              | 0.5840 | 2.4947 | 0.3987         | 0.7416 | 0.3640    | 4.1066 | 0.1293          | 1.3126 | 0.3973     | 4.1948 |
| REPTree                  | 0.7850 | 1.5665 | -0.0246        | 0.7600 | 0.5190    | 3.3799 | 0.6851          | 0.3351 | 0.5140     | 3.2889 |

## Predicción del RMSD del Acoplamiento Molecular Flexible Local

Resultados de las predicciones del RMSD a partir de los algoritmos de predicción del Weka. *Todas* significa a las poses del AutoDock4 y del AutoDock Vina que predicen las poses del AutoDock4 y del AutoDock Vina; *Solo AutoDock4* significa que las poses del AutoDock4 predicen las poses del AutoDock4; *Solo Vina* significa que las poses del AutoDock Vina predicen las poses del AutoDock Vina; *Todas AutoDock4* significa que las poses del AutoDock4 y del AutoDock Vina predicen las poses del AutoDock4; y *Todas Vina* significa que las poses del AutoDock4 y del AutoDock Vina predicen las poses del AutoDock Vina. CC representa el Coeficiente de Correlación y MAE representa el Error Relativo Medio. Ver Tabla 8.

**Tabla 8.** Resultados de la predicción del RMSD de las poses para el acoplamiento molecular dirigido flexible.

| Algoritmo                | Todas  |        | Solo AutoDock4 |        | Solo Vina |        | Todas AutoDock4 |        | Todas Vina |        |
|--------------------------|--------|--------|----------------|--------|-----------|--------|-----------------|--------|------------|--------|
|                          | CC     | MAE    | CC             | MAE    | CC        | MAE    | CC              | MAE    | CC         | MAE    |
| Gaussian Processes       | 0.3506 | 2.2109 | 0.3733         | 1.9292 | 0.1733    | 2.4636 | 0.4132          | 1.9153 | 0.2205     | 2.3897 |
| Linear Regression        | 0.3439 | 2.2184 | 0.3627         | 1.9495 | 0.1434    | 2.5697 | 0.4061          | 1.9171 | 0.2149     | 2.4008 |
| Multi-layer Perceptron   | 0.1873 | 4.5977 | 0.1611         | 4.2244 | 0.1395    | 3.7133 | 0.1887          | 3.8822 | 0.1053     | 5.1985 |
| Simple Linear Regression | 0.1556 | 2.3167 | 0.3778         | 1.9182 | 0.0714    | 2.3802 | 0.1413          | 2.0701 | 0.0714     | 2.3724 |
| SMOreg                   | 0.3344 | 2.2448 | 0.3592         | 1.9567 | 0.1316    | 2.6281 | 0.41            | 1.927  | 0.1918     | 2.4609 |
| IBk                      | 0.1786 | 2.972  | 0.0805         | 2.7735 | 0.1184    | 2.7321 | 0.1649          | 2.7875 | 0.1026     | 3.0666 |
| KStar                    | 0.3896 | 2.2198 | 0.4079         | 2.0648 | 0.2899    | 2.3116 | 0.3829          | 2.1049 | 0.3365     | 2.2613 |
| Decicion Table           | 0.1039 | 2.391  | 0.1529         | 2.1086 | -0.0347   | 2.4745 | 0.1458          | 2.1256 | 0.042      | 2.4534 |
| M5 Rules                 | 0.3497 | 2.2612 | 0.3684         | 1.8834 | 0.2332    | 2.378  | 0.3138          | 2.105  | 0.283      | 2.3349 |
| Zero R                   | 0      | 2.318  | 0              | 2.1018 | 0         | 2.326  | 0               | 2.0659 | 0          | 2.3513 |
| Decision Stump           | 0.1412 | 2.3595 | 0.3769         | 1.9482 | 0.1125    | 2.3479 | 0.1078          | 2.1288 | 0.0703     | 2.4028 |
| M5P                      | 0.3276 | 2.2389 | 0.3668         | 1.9021 | 0.0521    | 2.6226 | 0.3172          | 2.0417 | 0.2229     | 2.3512 |
| Random Forest            | 0.4415 | 2.0844 | 0.418          | 1.9316 | 0.2702    | 2.2216 | 0.4652          | 1.8682 | 0.3154     | 2.1702 |
| Random Tree              | 0.1864 | 2.9273 | 0.0944         | 2.7138 | 0.1801    | 2.7409 | 0.0179          | 2.9452 | 0.1734     | 2.9283 |
| REPTree                  | 0.2699 | 2.3134 | 0.1713         | 2.2378 | 0.1851    | 2.3963 | 0.2677          | 2.0919 | 0.1856     | 2.4224 |