



UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS  
VERTEX, ENTORNOS INTERACTIVOS 3D, FACULTAD 5

# APLICACIÓN PARA EL CÁLCULO DE ÍNDICES TOPOGRÁFICOS DE PROTEÍNAS Y DESCRIPTORES DE MOLÉCULAS PEQUEÑAS

**Trabajo de diploma para optar por el título de Ingeniero en Ciencias Informáticas**

Autor: Alejandro Villegas La Rosa

Tutores: Dr. Ramón Carrasco Velar

Ing. Roberto Tellez Ibarra

**La Habana, 2016**

*A las estrellas no se sube por caminos llanos*  
*José Martí*

*A mis padres y mi familia completa*

---

## Agradecimientos

---

*Le agradezco a todos mis amigos y familiares...*

---

## Declaración de autoría

---

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales sobre esta, con carácter exclusivo.

Para que así conste firmamos la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

---

Alejandro Villegas La Rosa  
Autor

---

Dr. Ramón Carrasco Velar  
Tutor

---

Ing. Roberto Tellez Ibarra  
Tutor

En la Universidad de las Ciencias Informáticas (UCI), particularmente en el proyecto de Visualización y Minería de Grafos Ponderados para la Quimiinformática, se han desarrollado herramientas que permiten el cálculo de descriptores moleculares y de proteínas. Estas aplicaciones presentan problemas para la realización de estas operaciones, sobre todo en el cálculo de descriptores topográficos para aminoácidos, debido a deficiencias en el proceso de lectura escritura que incidían en el propio cálculo de los descriptores. Los datos utilizados sobre las estructuras tridimensionales de las proteínas, que son adquiridos del Protein Databank (PDB) (Banco de Datos de Proteínas), no se ajustan a los requerimientos técnicos del proyecto pues necesitan ser depurados primeramente. Debido a esto, es preciso almacenar gradualmente solo los datos que sean necesarios en una base de datos que pueda ser accedida por el equipo de investigación, siendo necesario el acceso remoto a la misma como facilidad adicional para el cálculo de los descriptores. Por otro lado, sería significativo realizar dichos cálculos a un set de Catepsinas B por la manera en la que interviene en el ciclo de vida de diferentes parásitos. En el presente trabajo se plantea como objetivo, desarrollar una herramienta para el cálculo de índices topográficos para aminoácidos y descriptores de moléculas pequeñas. La aplicación desarrollada permite el cálculo de los índices topográficos sobre aminoácidos, además de los descriptores moleculares AMR(refractividad molar), ALOGP(coeficiente de partición octanol - agua) y ALOGP2 según el modelo Ghose-Crippen. Además, proporciona el acceso remoto a una base de datos de Catepsinas B así como la visualización del efecto de perturbación de los aminoácidos en las proteínas. Los resultados encontrados en el caso de prueba del análisis perturbativo, sugieren la posibilidad de establecer un rango para las interacciones a larga distancia entre grupos de átomos como es el caso de los aminoácidos en las proteínas.

**Palabras clave:** Catepsinas B, descriptores moleculares, índices topográficos, proteína.

<b>Introducción</b>	<b>1</b>
<b>1 Fundamentación Teórica</b>	<b>4</b>
1.1 Introducción . . . . .	4
1.2 Descriptores moleculares o índices . . . . .	4
1.2.1 Índices topográficos . . . . .	5
1.3 Proteínas. Estructura química . . . . .	5
1.3.1 Proteínas . . . . .	5
1.3.2 Aminoácidos . . . . .	6
1.3.3 Catepsinas B . . . . .	7
1.4 Aplicaciones usadas para el cálculo de descriptores . . . . .	7
1.4.1 Dragon: . . . . .	7
1.4.2 Molconn-Z: . . . . .	7
1.4.3 Codessa(Comprehensive Descriptors for Structural and Statistical Analysis): . . . . .	8
1.5 Herramientas y tecnologías a utilizar en la aplicación . . . . .	8
1.6 Procedimiento para el cálculo de índices topográficos de aminoácidos . . . . .	9
1.6.1 Reducción del grafo de la proteína . . . . .	10
1.6.2 Cálculo del centro de masa . . . . .	10
1.6.3 Índice del estado electrotopográfico para aminoácidos en proteínas reducidas (S-state)	11
1.6.4 1.7.4. Índice del estado lipotopográfico para aminoácidos en proteínas reducidas . . . . .	12
1.6.5 Índice del estado refractotopográfico para aminoácidos en proteínas reducidas . . . . .	12
1.7 Coeficiente de partición ALogP . . . . .	12
1.8 Metodología de desarrollo de software . . . . .	13
1.8.1 Metodologías ágiles . . . . .	13
1.8.2 Selección de la metodología de desarrollo de software a utilizar . . . . .	14
1.9 Conclusiones parciales . . . . .	14
<b>2 Características y diseño del sistema</b>	<b>15</b>
2.1 Introducción . . . . .	15
2.2 Captura de requisitos . . . . .	15

2.2.1	Requisitos funcionales . . . . .	15
2.2.2	Requisitos no funcionales . . . . .	16
2.3	Historias de Usuario . . . . .	16
2.4	Planificación . . . . .	20
2.4.1	Estimación del esfuerzo por historia de usuario . . . . .	20
2.4.2	Plan de iteraciones . . . . .	21
2.5	Tarjetas CRC . . . . .	21
2.6	Diagrama de clases . . . . .	22
2.7	Descripción de la arquitectura . . . . .	23
2.8	Patrones de Diseño . . . . .	24
2.9	Conclusiones parciales . . . . .	24
<b>3</b>	<b>Implementación y pruebas del sistema</b>	<b>25</b>
3.1	Introducción . . . . .	25
3.2	Estilo de codificación . . . . .	25
3.3	Herramienta para el cálculo de índices topográficos de proteínas y descriptores de moléculas pequeñas . . . . .	26
3.3.1	Cálculo de los índices topográficos para aminoácidos . . . . .	27
3.3.2	Exportación de descriptores calculados . . . . .	27
3.3.3	Acceso remoto a la base de datos de Catepsinas B . . . . .	28
3.3.4	Visualización científica del efecto perturbativo en los aminoácidos . . . . .	28
3.4	Tareas de ingeniería . . . . .	30
3.5	Pruebas . . . . .	31
3.5.1	Pruebas de aceptación . . . . .	32
3.5.2	Análisis de las pruebas de aceptación . . . . .	34
3.5.3	Validación de los descriptores implementados en la aplicación . . . . .	34
3.6	Resultados obtenidos . . . . .	35
3.7	Conclusiones parciales . . . . .	38
	<b>Conclusiones</b>	<b>39</b>
	<b>Recomendaciones</b>	<b>40</b>
<b>4</b>	<b>Anexo</b>	<b>41</b>
4.1	Tablas de ingeniería . . . . .	41
	<b>Referencias bibliográficas</b>	<b>43</b>



---

## Índice de figuras

---

1.1	Estructura general de un aminoácido. . . . .	6
1.2	Estructura de una proteína primaria . . . . .	6
3.1	Cálculo de índices topográficos en aminoácidos de la proteína 1CPJ . . . . .	27
3.2	Opciones de conexión de la aplicación . . . . .	28
3.3	Ventana que permite la configuración para acceder a la base de datos remota . . . . .	28
3.4	Ventana para la selección del aminoácido que será analizado . . . . .	29
3.5	Análisis del efecto perturbativo sobre el aminoácido seleccionado . . . . .	29
3.6	Resultados de las pruebas de aceptación . . . . .	35
3.7	Efecto perturbativo de los aminoácidos restantes sobre el aminoácido Alanina 7 en la proteína 1CPJ. . . . .	36
3.8	Efecto perturbativo de los aminoácidos restantes sobre el aminoácido Arginina 49 de la cadena A en la proteína 1CPJ. . . . .	36
3.9	Efecto perturbativo de los aminoácidos restantes sobre el aminoácido Serina 4 de la cadena A en la proteína 1CPJ. . . . .	37
3.10	Efecto perturbativo de los aminoácidos restantes sobre el aminoácido Triptófano 215 de la cadena B en la proteína 1CPJ. . . . .	37

---

## Índice de tablas

---

2.1	Historia de usuario # 1 . . . . .	18
2.2	Historia de usuario # 2 . . . . .	18
2.3	Historia de usuario # 3 . . . . .	18
2.4	Historia de usuario # 4 . . . . .	19
2.5	Historia de usuario # 5 . . . . .	19
2.6	Historia de usuario # 6 . . . . .	19
2.7	Historia de usuario # 7 . . . . .	20
2.8	Estimación de esfuerzo por historia de usuario . . . . .	20
2.9	Plan de duración de las iteraciones . . . . .	21
2.10	Tarjeta CRC # 1 . . . . .	21
2.11	Tarjeta CRC # 2 . . . . .	22
2.12	Tarjeta CRC # 3 . . . . .	22
3.1	Tarea de ingeniería # 1 . . . . .	30
3.2	Tarea de ingeniería # 2 . . . . .	30
3.3	Tarea de ingeniería # 3 . . . . .	30
3.4	Tarea de ingeniería # 4 . . . . .	31
3.5	Tarea de ingeniería # 5 . . . . .	31
3.6	Tarea de ingeniería # 6 . . . . .	31
3.7	Prueba de aceptación # 1 . . . . .	32
3.8	Prueba de aceptación # 2 . . . . .	32
3.9	Prueba de aceptación # 3 . . . . .	32
3.10	Prueba de aceptación # 4 . . . . .	33
3.11	Prueba de aceptación # 5 . . . . .	33
3.12	Prueba de aceptación # 6 . . . . .	34
4.1	Tarjeta CRC # 4 . . . . .	41
4.2	Tarjeta CRC # 5 . . . . .	41
4.3	Tarjeta CRC # 6 . . . . .	42
4.4	Tarjeta CRC # 7 . . . . .	42

---

## Lista de algoritmos

---

---

## Lista de códigos fuentes

---

3.1	Ejemplo de código en Java . . . . .	25
-----	-------------------------------------	----

Durante el desarrollo de las investigaciones químicas el uso de las tecnologías de la información se ha convertido en un elemento sustancial para la descripción de moléculas y sus características estructurales. Frank Brown en el año 1998 introdujo el concepto de quimioinformática de la siguiente manera (Brown et al., 1998):

La quimioinformática es la combinación de aquellos recursos computacionales para transformar datos en información e información en conocimiento, con el propósito de hacer más rápidas y mejores decisiones en el ámbito de la identificación y optimización de fármacos.

Esta rama científica combina los campos de acción de la química y las ciencias computacionales en el área de la teoría de grafos y minería del espacio químico.

Durante los últimos años y con la aplicación de las nuevas técnicas de la información la industria farmacéutica ha desarrollado sus investigaciones hacia aquellos métodos que permitan describir de manera eficiente y de bajo costo computacional la estructura química de las moléculas candidatas a medicamentos, así como predecir la respuesta biológica de prácticamente cualquier compuesto orgánico, con altas probabilidades de acierto (Rodríguez-León; Ramón Carrasco-Velaz; Arencibia-Benítez; Gilberto-Rodríguez y Domínguez-Vega, 2010). Esto facilita que se puedan acortar los plazos de ejecución y los costos de la investigación.

El diseño de fármacos asistido por computadora ha sido establecido como un enfoque de investigación dirigido a descubrir moléculas usando herramientas computacionales. En esta temática se conjugan diversas disciplinas como la inteligencia artificial y la estadística, entre otras.

Por otro lado, las Catepsinas B, son enzimas endógenas que participan en la degradación de las proteínas y desempeñan diferentes roles en los organismos vivos sobre todo en el ciclo de vida de diferentes parásitos. Existen investigaciones que aseguran la importancia de esta proteína como candidato potencial para el desarrollo de fármacos contra la fasciola hepática lo que hace aún más necesario realizar una descripción estructural y de propiedad de las mismas. (Law; Smooker; Irving et al., 2003).

De igual forma, en nuestro país se han realizado estudios para el diseño racional de fármacos apoyados en información estructural a través de la modelación por homología de esta proteína (Naranjo; Díaz de Arce; Pérez et al., 2007).

Particularmente, en la Universidad de las Ciencias Informáticas (UCI), se han desarrollado herramientas para la quimioinformática. Entre ellas está la plataforma alasGRATO, donde existía un módulo especializado en el cálculo de descriptores moleculares y de proteínas. El módulo de proteínas presentaba problemas para la realización de estas operaciones. Actualmente el proyecto de Visualización y Minería de Grafos Ponderados

para la Quimioinformática se encuentra desarrollando la aplicación GraphMining, herramienta para el diseño de fármacos por minería de grafos, que permite el cálculo de descriptores híbridos, tanto para moléculas pequeñas como para aminoácidos en proteínas. Sin embargo este software presenta diferentes problemas para el cálculo de descriptores de aminoácidos. Esto se debe a deficiencias en el proceso de lectura escritura que incidían en el propio cálculo de los descriptores.

La fuente de datos estructurales de proteínas más amplia es el Protein Databank (PDB DB) un portal de información que contiene 119 823 estructuras tridimensionales de macromoléculas biológicas (Databank, 2016). Sin embargo, no resulta práctico, a los efectos de estudios concretos, utilizar todo el conjunto de los ficheros pdb de estructuras que contiene dicha base por lo que resulta aconsejable almacenar gradualmente solo los necesarios en una base de datos personalizada. Lo recomendable en estos casos, es que esta base se sitúe en un servidor para uso generalizado del equipo de investigación, dado que la misma crecerá continuamente, con lo cual se genera la necesidad del acceso remoto a dicha base de datos centralizada como facilidad adicional para el cálculo de cualquier tipo de ellas, provenientes del PDB Databank. Además, la estructura de la base de PDB DB, no se ajusta a los requerimientos técnicos del proyecto pues requiere incluso que esa información sea depurada.

Por todo lo anterior se presenta como **Objeto de estudio:** Cálculo de índices topográficos y descriptores moleculares.

El objeto de estudio delimita el **campo de acción:** Cálculo de índices topográficos para aminoácidos y descriptores de moléculas pequeñas.

**Objetivo general:**

Desarrollar una aplicación para el cálculo de índices topográficos para aminoácidos y descriptores de moléculas pequeñas.

**Tareas a realizar:**

Definición de los requisitos funcionales y no funcionales de la herramienta.

Implementación del cálculo de descriptores híbridos para grafos reducidos.

Implementación de descriptores para moléculas pequeñas.

Acceso remoto a la base de datos de Catepsinas B para realizar cálculos correspondientes.

Exportación de los descriptores calculados de los aminoácidos en las proteínas a documento Excel.

Visualización científica para analizar la perturbación de los aminoácidos en las proteínas.

Elaboración del informe que contemple el proceso de investigación y desarrollo definido.

**Posibles resultados:**

Aplicación para el cálculo de índices topográficos de proteínas y algunos descriptores de moléculas pequeñas.

Análisis del comportamiento de la perturbación entre los aminoácidos de las proteínas.

**Métodos Científicos:**

Métodos Teóricos:

Análítico-Sintético: para el análisis de la documentación existente relacionada con el tema, ex-

trayendo los elementos más importantes y necesarios para dar solución al problema existente, de manera que permita sintetizar todo lo obtenido relacionado con los estándares pertinentes y trabajos ya realizados anteriormente por expertos en el tema.

Métodos Empíricos:

Método de observación científica: Se emplea con el objetivo de observar el funcionamiento del módulo para el cálculo de índices topográficos de la herramienta GraphMining. Este método permite investigar el procedimiento empleado por dicha aplicación para efectuar el cálculo de estos descriptores.

El presente trabajo de diploma consta de un Resumen, Introducción, tres capítulos, Conclusiones Generales, Recomendaciones, Referencias Bibliográficas y Anexos.

Los capítulos que se presentan tiene la siguiente estructura:

**Capítulo 1. Fundamentación teórica:** En este capítulo se presenta el marco teórico estudiado y se detallan los conceptos fundamentales relacionados de la química para una mejor comprensión de la solución planteada. Se abordan las herramientas, tecnologías y la metodología empleada durante el desarrollo del trabajo.

**Capítulo 2. Características y diseño del sistema:** Se abordan las características de la aplicación desarrollada de acuerdo a las diferentes fases de la metodología empleada. Se describe los requisitos funcionales, la arquitectura utilizada y el diseño del sistema.

**Capítulo 3. Implementación y pruebas del sistema:** Describe la implementación de los algoritmos utilizados durante el desarrollo de la aplicación. Se presentan las tareas de ingeniería y todo lo relacionado con las pruebas realizadas al sistema. Se abordan los resultados obtenidos durante el cálculo de los descriptores mediante el análisis de la visualización de la perturbación de los aminoácidos.

## **1.1. Introducción**

En este capítulo se expone todo lo relacionado con los conceptos teóricos que sirven de base a este trabajo. También se explica la estrecha vinculación que existe entre la química y la informática, siendo esta última una eficaz herramienta para el desarrollo de diferentes estudios dentro de este campo. Así mismo, los programas usados actualmente para el cálculo de descriptores moleculares son expuestos así como sus elevados costos en la industria del software. Luego se detallan las herramientas, tecnologías y metodología empleada para el desarrollo de la solución propuesta. Además, se expone el procedimiento para la implementación de los algoritmos utilizados en la aplicación.

## **1.2. Descriptores moleculares o índices**

Actualmente existe una gran variedad de aplicaciones en el diseño de fármacos asistidos por computadora y en la quimioinformática que dependen de representaciones de moléculas mediante descriptores, los cuales a través de valores numéricos capturan sus características estructurales y propiedades.

Un descriptor molecular es 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 en el resultado de un experimento estandarizado (Todeschini y Consonni, 2000).

Dicho con otras palabras, los descriptores moleculares son valores numéricos que describen las propiedades de las moléculas y el comportamiento de compuestos orgánicos. Estos pueden variar en dependencia de la información codificada y el tiempo de cómputo. De igual manera la calidad de los mismos condiciona la calidad de los modelos matemáticos que describen los fenómenos químicos y biológicos. Asimismo juegan un papel fundamental en todos estos procesos convirtiéndose en la herramienta básica para transformar la información química a valores numéricos para posteriormente aplicar procedimientos informáticos.

En nuestros días se han reportado en la literatura una gran cantidad de los mismos, que van desde simples propiedades de volumen hasta elaborar formulaciones tridimensionales y huellas moleculares complejas, que



a veces consisten de miles de posiciones de bit. Una de las tareas más importantes en la investigación de la quimioinformática es la selección basada en el conocimiento de los descriptores que son adecuados para aplicaciones específicas.

Existen muchos criterios a la hora de clasificar estos índices, como también se les conoce entre los que se destacan los siguientes (Escalona-Arranz; Ramón Carrasco-Velaz y Padrón-García, 2008):

- Índices basados en propiedades químico-físicas: Son aquellos, como lo dice su nombre, que derivan de la determinación experimental de una determinada propiedad químico-física de la molécula.
- Índices mecánico-cuánticos: Son aquellos que se determinan por cálculos mecánico-cuánticos.
- Índices topológicos y topográficos: Son aquellos que se determinan a partir del grafo químico, o sea de la estructura química.

La solución informática que se presenta en este trabajo permite entre sus funcionalidades el cálculo de dos descriptores para moléculas pequeñas: la refractividad molar (de ahora en adelante AMR) de acuerdo al modelo de Ghose-Crippen y el coeficiente de partición octanol-agua de Ghose-Crippen-Viswanadhan (Viswanadhan; Ghose; Revankar y Robins, 1990) (de ahora en adelante ALOGP) . Ambos pertenecen a la primera de las clasificaciones mostradas. El descriptor ALOGP describe la hidrofobicidad de las moléculas, mientras que por otro lado la refractividad molar mide el volumen estérico de las mismas.

### 1.2.1. Índices topográficos

Concerniente a los índices topográficos, estos son descriptores que se calculan por lo general a partir de la matriz de adyacencia entre los elementos de un grafo, los cuales han sido ponderados por un valor numérico que contiene información tridimensional del grafo molecular (R Carrasco-Velaz, 2008). Dentro de estos índices híbridos se encuentran el Índice del Estado Electrotopográfico, el Índice del Estado Refractotopográfico y el Índice del Estado Lipotopográfico. Por su gran importancia, el desarrollo de nuevos descriptores y nuevas representaciones de proteínas capaces de codificar información relevante constituye un área emergente de investigación. En la solución informática que se presenta se realiza el cálculo de estos tres índices en distintas cadenas de aminoácidos que conforman la estructura de diversas proteínas.

## 1.3. Proteínas. Estructura química

Teniendo en cuenta que el cálculo de estos índices topográficos será efectuado sobre proteínas, específicamente Catepsinas B, es oportuno explicar qué son las proteínas y cómo están estructuradas, lo cual ayudará a entender mejor el procedimiento utilizado para implementar estos índices en el software desarrollado.

### 1.3.1. Proteínas

Las proteínas desempeñan un importante papel entre las moléculas constituyentes de los seres vivos (biomoléculas). Por lo general todos los procesos biológicos dependen de la presencia o la actividad de este

tipo de moléculas. Una proteína está compuesta por largos polímeros de aminoácidos todos los cuales tienen como elemento estructural común la estructura de un alfa-aminoácido.

### 1.3.2. Aminoácidos

Existen alrededor de 300 tipos de aminoácidos en la naturaleza, sólo 20 de los cuales sirven como componentes básicos de las proteínas. Otros funcionan como metabolitos, mensajeros y reguladores de los procesos biológicos (kessel2010). Como su nombre lo indica son compuestos orgánicos que contienen un grupo ácido carboxílico (-COOH) y un grupo amino (-NH<sub>2</sub>) en posición alfa, junto con una cadena lateral específica para cada uno. A continuación se presentan dos imágenes que representan lo antes expuesto: la primera es una representación general de un aminoácido mientras que la segunda describe la estructura de una proteína acompañada de estos.

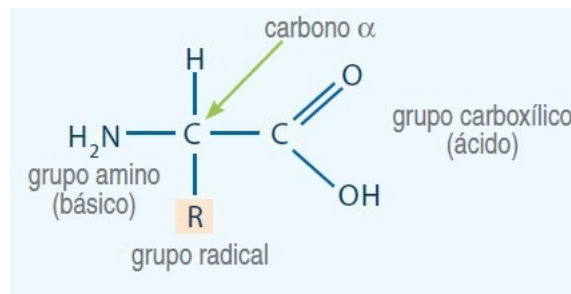


Figura 1.1. Estructura general de un aminoácido.

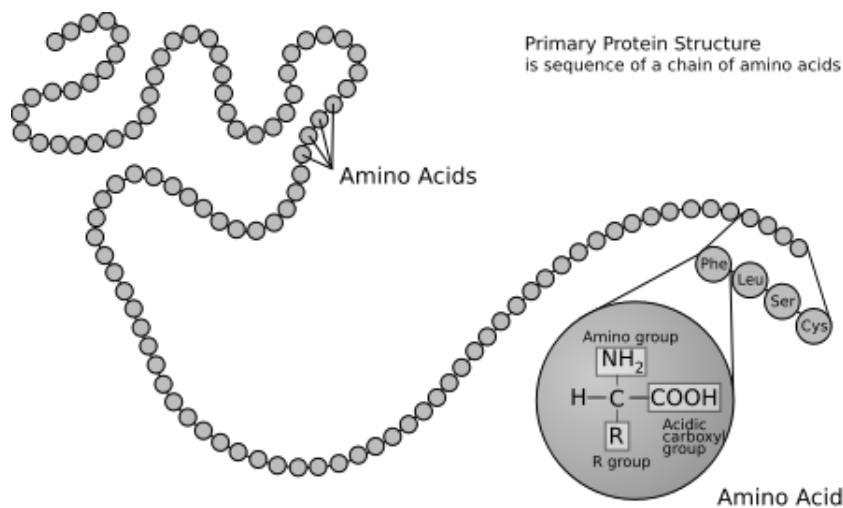


Figura 1.2. Estructura de una proteína primaria

### 1.3.3. Catepsinas B

Las catepsinas son enzimas <sup>1</sup> endógenas que participan en la degradación de las proteínas. En el caso particular de las Catepsinas tipo B, estas pertenecen a la superfamilia de la papaína, cisteino proteasa que desempeña diferentes roles en los organismos vivos y el ciclo de vida de numerosos parásitos, como por ejemplo, la fasciola y el plasmodium (malaria).

Valores elevados de algunas catepsinas han sido relacionados también con la existencia de diferentes enfermedades crónicas como el enfisema pulmonar, la arterioesclerosis, la artritis reumatoide y muchas otras enfermedades como el cáncer. Por esta razón y teniendo en cuenta la importancia de la Catepsinas B en el presente trabajo de diploma se realiza el cálculo de los índices topográficos sobre este tipo de compuestos, como caso de estudio.

## 1.4. Aplicaciones usadas para el cálculo de descriptores

Actualmente existen aplicaciones informáticas que permiten el cálculo de varios descriptores entre las que se destacan las siguientes (Michael J.S. Dewar y Karelson, s.f.):

### 1.4.1. Dragon:

Es una de las aplicaciones más populares para el cálculo de descriptores moleculares. Fue desarrollada por Milano Chemometrics y QSAR Research Group. Permite el cálculo de más de 5000 descriptores de de diferente naturaleza. Es multiplataforma y cuenta con varias versiones para Windows. Por otra parte en Linux existe una versión, DragonX, que sólo funciona en segundo plano por una línea de comandos. Aun así su diseño sólo permite realizar el cálculo de descriptores moleculares y no tiene la funcionalidad de realizar análisis QSAR. Para el primer año de uso la licencia tiene un costo de 2700 euros.

Como otras características técnicas de este software puede señalarse que el DRAGON 5.4 para Windows permite el cálculo en batch de hasta 50,000 moléculas con un máximo de 1,000 átomos por molécula, mientras que la versión Linux del DRAGON no tiene constreñimientos.

### 1.4.2. Molconn-Z:

Es de gran utilidad para realizar el cálculo de un amplio espectro de índices sobre todos topológicos entre los que se encuentran: índices de conectividad, Índice del Estado Electrotológico, índices de equivalencia topológica e índice topológico total y alguna información de los fragmentos del grafo, de los átomos, del tipo de enlace, etc. Este software permite el trabajo con bases de datos comerciales y los valores de los índices calculados son exportados a un fichero de salida estándar para poder ser utilizados en análisis estadísticos (Molconn-ZTM, 2006). El precio de la licencia para un usuario simple es de \$8 500.

---

<sup>1</sup>enzima: molécula de naturaleza proteínica y estructural que acelera las reacciones químicas.

### 1.4.3. Codessa(Comprehensive Descriptors for Structural and Statistical Analysis):

Este programa brinda al usuario la posibilidad de calcular más de 500 descriptores (topológicos, geométricos, constitucionales, termodinámicos, electrostáticos), un módulo para el establecimiento de modelos QSAR basado fundamentalmente en métodos estadísticos (análisis de regresión, heurísticos y análisis multivariados) y la integración completa con Gaussian 98/03. Es multiplataforma, pero es un programa propietario de elevado costo. Su licencia tiene un costo de \$ 4 200 para un usuario simple con un precio de soporte equivalente a \$ 800.

A pesar del gran uso que ofrecen estos programas no son gratuitos y el costo de adquisición de los mismos es muy elevado. Además no permiten calcular descriptores topograficos entre sus funcionalidades.

## 1.5. Herramientas y tecnologías a utilizar en la aplicación

- **Entorno de Desarrollo Integrado (IDE): Netbeans**

Para el desarrollo del presente módulo se seleccionó como IDE Netbeans en su versión 8.0.2 . Esta plataforma cuenta con diferentes módulos que permiten extenderlo y favorecen el trabajo modular. Además de ser el IDE oficial de Java 8 ofrece soporte para otros lenguajes como C/C++ y PHP. También provee herramientas y editores para sus frameworks y tecnologías. Puede ser instalado en todos los sistemas operativos que soporten Java, desde Windows y Linux hasta cualquier sistema MacOS X. Es un software gratuito y de código abierto que cuenta con una extensa comunidad de desarrolladores en todo el mundo (Netbeans, 2016).

- **Lenguaje de programación: Java**

Java es un lenguaje de programación de propósito general, concurrente, inflexible y orientado a objetos. Normalmente se compila para el conjunto de instrucciones bytecode y formato binario según la especificación de la máquina virtual de Java(JVM por sus siglas en inglés) (Java, 2016). Se utilizará este lenguaje debido a la posibilidad de ejecutar la aplicación en cualquier sistema operativo una vez que sea compilada. Además, es necesario el uso de diferentes tecnologías basadas en Java para el desarrollo de la herramienta que se presenta.

- **Biblioteca CDK**

The Chemistry Development Kit (CDK) es una biblioteca Java de código abierto para Quimioinformática y Bioinformática. Su arquitectura y capacidades, así como su desarrollo como proyecto de código abierto es descrito por un equipo de colaboradores internacionales de instituciones académicas e industriales. El CDK proporciona métodos para muchas tareas comunes en informática molecular, incluyendo estructuras químicas 2D y 3D, rutinas I/O, análisis y generación de SMILES, las búsquedas de anillo, verificación de isomorfismo, generación de estructura de diagrama, etc. Posee además escenarios de aplicación, así como información de acceso para los usuarios interesados y posibles contribuyentes (Christoph Steinbeck; Kuhn; Horlacher y Edgar Luttmann, 2003).

Debido a las características mencionadas anteriormente se utiliza dicha biblioteca para efectuar el

cálculo de los descriptores topográficos y moleculares puesto que facilita el proceso de implementación de los índices híbridos y descriptores moleculares.

- **Sistema Gestor de Base de Datos: PostgreSQL**

PostgreSQL es un sistema de base de datos relacional de gran alcance, de código abierto y libre. Cuenta con más de 15 años de desarrollo activo y una arquitectura probada que se ha ganado una sólida reputación de fiabilidad, integridad de datos y la corrección. Se ejecuta en todos los principales sistemas operativos, incluyendo Linux, Mac OS X y Windows. Es compatible con el almacenamiento de objetos binarios grandes, como imágenes, sonidos, o de vídeo. Tiene interfaces de programación de manera nativa para C/C++, Java, .Net, Perl, Python, Ruby entre otros. Es dirigido por una comunidad de desarrolladores que trabaja de manera desinteresada (PostgreSQL, 2016). Se utilizará para el acceso remoto a la base de datos de Catepsinas B de la aplicación.

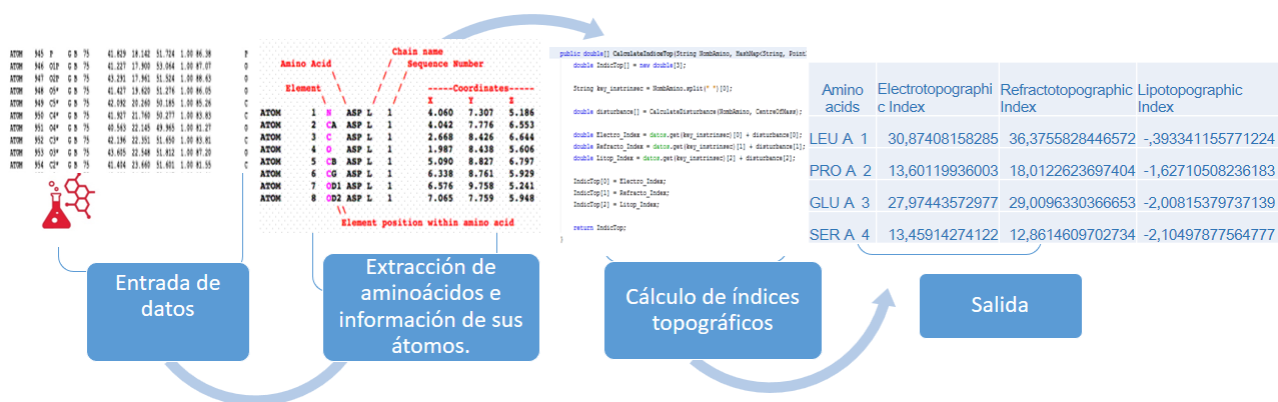
- **Herramienta CASE: Visual Paradigm**

Visual Paradigm es una potente herramienta para visualizar y diseñar elementos de software, posee características gráficas muy cómodas que facilitan la realización de los diagramas de modelación que sigue el estándar UML, proporciona a los desarrolladores una plataforma que les permita diseñar un producto con calidad de forma rápida, facilita la interoperabilidad con otras herramientas; tiene licencia gratuita, es multiplataforma y puede contar con plug-ins que le posibilitan la integración con algunas herramientas Java como son: Eclipse/IBM WebSphere, JBuilder, NetBeans, entre otras. La herramienta está diseñada para una amplia gama de usuarios, incluyendo ingenieros de software, analistas de sistemas, analistas de negocios y arquitectos de sistemas que estén interesados en la creación de grandes sistemas de software de manera confiable a través del paradigma Orientado a Objetos (Paradigm, 2016).

## 1.6. Procedimiento para el cálculo de índices topográficos de aminoácidos

Para efectuar el cálculo de los índices topográficos de los aminoácidos se procede de la siguiente forma:

- Lectura de ficheros de entrada.
- Reducción del grafo de la proteína.
- Cálculo del valor de la perturbación de cada aminoácido.
- Finalmente calcular los índices topográficos de cada aminoácido de la proteína.



Procedimiento para el cálculo de índices topográficos de aminoácidos

### 1.6.1. Reducción del grafo de la proteína

La reducción del grafo de la proteína cargada en la aplicación permite hacer cálculos mucho más veloces computacionalmente al reducir el espectro de las coordenadas atómicas de los aminoácidos que componen la proteína. Para lograr esto se procede de la siguiente manera:

1. Se ignoran las coordenadas de los heteroátomos que conforman el fichero .pdb
2. Se calcula el centro de masa de cada aminoácido de la proteína.
3. Se construye la matriz de coordenadas cartesianas de los centros de masas de los aminoácidos.

### 1.6.2. Cálculo del centro de masa

El centro de masa representa el punto en el que se concentra toda la masa del sistema para su estudio. Es el centro de simetría de distribución de un sistema de partículas. Se puede decir que el sistema formado por toda la masa concentrada en el centro de masas es un sistema equivalente al original (Ruiz, 2016).

La posición del centro de masa de una partícula viene dada por la expresión:

$$\vec{r}_{cm} = \frac{\sum_{i=0}^N m_i \vec{r}_i}{M} \quad (1.6.1)$$

$$M = \sum_{i=0}^N m_i \quad (1.6.2)$$

M : masa total del sistema de partículas.

$\vec{r}_i$  : vector de posición de la masa  $i$ -ésima respecto al sistema de referencia supuesto.

$m_i$  : masa de la partícula  $i$ -ésima.

El cálculo del centro de masa de los aminoácidos permite simplificar en gran medida las coordenadas de los átomos que lo conforman, para lograr así una única coordenada cartesiana. Esta coordenada representa

la concentración de toda la masa atómica en un único punto del aminoácido en el espacio.

### 1.6.3. Índice del estado electrotopográfico para aminoácidos en proteínas reducidas (S-state)

El índice del estado electrotopográfico fue definido por Carrasco y cols. como una variante para aminoácidos del índice del Estado Electrotopográfico para átomos (E-state) (R Carrasco-Velaz, 2008).

En este, cada átomo tiene asignado un valor de estado intrínseco  $I_i$  calculado de la siguiente manera:

$$I_i = [(2/N^2)\delta^v + 1]/\delta \quad (1.6.3)$$

$N$ : Principal número cuántico del átomo  $i$ .

$\delta^v$ : Número o valencia de electrones en el esqueleto molecular ( $Zv - h$ ).

$\delta$ : Número de electrones en la estructura (-h).

Luego el índice del estado electrotopográfico (S-state)  $S_i$  es el valor intrínseco modificado

$$S(A_i) = I_i + \Delta I_i \quad (1.6.4)$$

donde  $\Delta I_i$  es el efecto de perturbación

$$\Delta I_i = \sum_{j=1}^N (I_i - I_j)/r_{ij}^2 \quad (1.6.5)$$

$r_{ij}$  : Distancia topográfica entre los  $i$  y  $j$  átomos tomada de la matriz de distancias interatómicas calculadas por algún método químico-cuántico (AM1, PM3, etc).

Para el cálculo del Índice Electrotopográfico para Aminoácidos en una proteína, se calcula el aminoácido aislado y la suma de todos sus valores atómicos constituirá el valor intrínseco del aminoácido, el cual se tabula y se completan los valores intrínsecos de esa forma para los 20 aminoácidos esenciales.

El cálculo del centro de masa de los aminoácidos descrito con anterioridad redujo el número de vértices del grafo que representa a la proteína ya que se obtiene una coordenada cartesiana del centro de masa por cada aminoácido. Por este motivo cuando se calcula el índice del estado electrotopográfico para aminoácidos en proteínas reducidas la variable  $r_{ij}$  representaría en esta ocasión la distancia topográfica entre los centros de masa del cada par de aminoácidos  $i$  y  $j$ . Por lo demás el cálculo se efectúa de igual manera a lo explicado en el epígrafe precedente. Este nuevo procedimiento reduce en gran medida computacionalmente el cálculo de la perturbación  $I_i$

**1.6.4. 1.7.4. Índice del estado lipotopográfico para aminoácidos en proteínas reducidas**

La transformación de índices topológicos a topográficos es un procedimiento muy sencillo. Con la simple sustitución de la distancia topológica para el cálculo de la perturbación  $I_i$  por la distancia euclidiana de los átomos es suficiente [Ramón Carrasco Velar, «Empleo de grafos reducidos híbridos para estudios de clasificación de moléculas bioactivas», 2013.]. Para el cálculo de su predecesor, el índice lipotopográfico para átomos ( $\Lambda_{3D}$  Lstate3D) se emplea la expresión:

$$\Lambda_{3D} = I_i + \Delta I_i \quad (1.6.6)$$

donde  $I_i$  es el valor intrínseco del átomo  $i$  y  $\Delta I_i$  es el efecto de la perturbación.

$$\Delta I_i = \sum_{j=1}^N (I_i - I_j) / r_{ij}^2 \quad (1.6.7)$$

$r_{ij}$ : Distancia euclidiana entre los  $i$  y  $j$  átomos tomada de la matriz de distancias.

**1.6.5. Índice del estado refractotopográfico para aminoácidos en proteínas reducidas**

En cuanto al índice del estado refractotopográfico se calcula de la misma forma que el resto de los índices topográficos analizados con anterioridad.

$$\mathfrak{R}_{3D} = AR_i + \Delta AR_i \quad (1.6.8)$$

Siendo esta vez  $R_i$  el valor de la refractividad intrínseca del aminoácido  $i$  y  $\Delta R_i$  es un término de la perturbación definido por:

$$\Delta AR_i = \sum_{j=1}^N (AR_i - AR_j) / r_{ij}^2 \quad (1.6.9)$$

$r_{ij}$ : Distancia topográfica entre los centros de masa del aminoácido  $i$ .

**1.7. Coeficiente de partición ALogP**

La lipofilidad, denotada como  $P$  es la medida del grado en que una molécula dada prefiere un ambiente no polar (hidrofóbico) antes que un ambiente acuoso. (Marinich; Maguna; Okulik y Castro, s.f.).

El descriptor molecular más ampliamente utilizado de lipofilidad es el coeficiente de partición octanol-agua (conocido como  $\text{Log}P_{\text{octanol/agua}}$  o solo  $\text{Log}P$  cuando no se especifica). Este cociente se obtiene de la división de los valores de concentración de una sustancia disuelta en una fase acuosa o lipídica (grasa) cuando



ambas fases se ponen en contacto, se agitan y dejan reposar hasta el equilibrio. En el caso de sustancias solubles en agua (por ejemplo una sal), una pequeña cantidad pasará a la fase lipídica, en baja proporción. De ahí que el valor de

$$\log P = \frac{\text{concentracin de sal en la fase lipídica}}{\text{concentracin de sal en la fase}}$$

será menor que 1 y el logaritmo de P será negativo. Un análisis similar se puede realizar para sustancias lipídicas.

Por otra parte, el coeficiente de partición octanol-agua para átomos de Ghose-Crippen (ALogP) tal como lo plantean esos autores, está dado por,

$$ALogP = \sum_k a_k N_k \quad (1.7.1)$$

donde  $n_k$  es el número de átomos de tipo  $k$  y  $a_k$  es la contribución del tipo de átomo correspondiente (Viswanadhan; Ghose; Revankar y Robins, 1990). Para la solución informática propuesta se utiliza la biblioteca CDK que tiene implementado los procedimientos para el cálculo de este índice.

## 1.8. Metodología de desarrollo de software

Una metodología de desarrollo de software, en ingeniería de software, es un marco de trabajo usado para estructurar, planificar y controlar el proceso de desarrollo en sistemas de información, a través de un conjunto de procedimientos, técnicas, herramientas y un soporte documental que ayuda a los desarrolladores a realizar un nuevo software. Estas metodologías se clasifican en dos tipos: Metodologías tradicionales y ágiles.

### 1.8.1. Metodologías ágiles

Las metodologías ágiles proporcionan una serie de reglas encaminadas a satisfacer las necesidades del cliente y realizar la entrega del proyecto menos complicada y más satisfactoria logrando un producto de buena calidad en poco tiempo.

Las principales características que marcan la diferencia con respecto a las metodologías tradicionales son expresadas en el Manifiesto Ágil donde se valora (Christoph Steinbeck; Kuhn; Horlacher y Edgar Luttmann, 2003).

- Al individuo y las interacciones del equipo de desarrollo, sobre el proceso y las herramientas.
- Desarrollar software que funcione, más que conseguir una buena documentación.
- La colaboración con el cliente, más que la negociación de un contrato.
- Responder a los cambios, más que seguir estrictamente un plan.

### **1.8.2. Selección de la metodología de desarrollo de software a utilizar**

Después del análisis de las diferentes características, ventajas y desventajas de las metodologías de desarrollo de software y teniendo en cuenta las particularidades de la solución informática a desarrollar se elige la metodología ágil eXtreme Programming (XP). Esta metodología está definida para equipos pequeños de desarrollo, tiene prioridad de satisfacer al cliente mediante un desarrollo iterativo e incremental, abierto a los cambios y caracterizado por un código simple, al mismo tiempo que aplica un conjunto de prácticas que harán la entrega de la aplicación menos complicada y más satisfactoria para los clientes. En el caso particular de esta investigación, el cliente, que es también el especialista es parte del equipo de desarrollo, por lo que va a estar en condiciones de contestar rápida y correctamente a cualquier pregunta por parte del desarrollador.

## **1.9. Conclusiones parciales**

En el capítulo se expuso la vinculación de la informática como herramienta de apoyo al desarrollo de los estudios en diferentes disciplinas de la química como el diseño racional de fármacos asistido por computadora y la quimioinformática. Se explicaron los descriptores moleculares que fueron usados en la implementación de la herramienta así como algunos conceptos esenciales para lograr una mejor comprensión de la solución propuesta.

## 2.1. Introducción

Este capítulo describe las características del software desarrollado partiendo desde los requisitos funcionales hasta los artefactos empleados haciendo uso de la metodología XP. Se explica también el proceso de planificación de las entregas así como la estimación del esfuerzo según cada historia de usuario. Posteriormente se describe la arquitectura utilizada para desarrollar la aplicación y luego se presentan los patrones de diseño implementados.

## 2.2. Captura de requisitos

La captura de los requisitos funcionales y no funcionales es uno de los procedimientos más importantes para el desarrollo de cualquier software. En esta etapa el principal objetivo es identificar lo que el cliente realmente necesita, de manera que el equipo de desarrollo pueda trabajar en base a estos requerimientos y logre obtener un producto que funcione con la calidad deseada.

### 2.2.1. Requisitos funcionales

Los requisitos funcionales especifican acciones que debe poder realizar un sistema, sin tener en cuenta las restricciones físicas (Booch; Grady; Jacobson e Ivar y Rumbaugh, 2000). Además, definen su comportamiento de salida y entrada, y no alteran la funcionalidad del producto, esto quiere decir que los requerimientos funcionales se mantienen invariables sin importar con qué propiedades o cualidades se relacionan. A continuación se presentan los requerimientos funcionales tomados en cuenta durante el desarrollo del presente trabajo:

- RF-1.** Cargar archivos de proteínas y moléculas pequeñas.
- RF-2.** Conectar aplicación a la base de datos de Catepsinas.
- RF-3.** Calcular centro de masa de los aminoácidos

- RF-4.** Calcular índices topográficos para aminoácidos.
- RF-4.1** Calcular índice electrotopográfico para aminoácidos.
- RF-4.2.** Calcular índice refractotopográfico para aminoácidos.
- RF-4.3.** Calcular índice lipotopográfico para aminoácidos.
- RF-5.** Calcular descriptores moleculares para moléculas pequeñas.
- RF-5.1** Calcular coeficiente de partición (ALOGP) de moléculas pequeñas.
- RF-5.2.** Calcular descriptor ALOGP2 de moléculas pequeñas.
- RF-5.3.** Calcular refractividad molar(AMR) de moléculas pequeñas.
- RF-6.** Exportar descriptores calculados a documento Excel.
- RF-7.** Visualizar el efecto perturbativo de los aminoácidos en un gráfico de dispersión.

### 2.2.2. Requisitos no funcionales

Los requisitos no funcionales sólo describen atributos del sistema o atributos del entorno del sistema. Los requisitos no funcionales tomados en cuenta en el presente trabajo son los que se describen a continuación:

- **Seguridad**
  - RNFS-1.** Integridad: Los datos originales no deben sufrir pérdidas durante su representación.
- **Software**
  - RNFS-2.** En la estación de trabajo debe estar instalada la Máquina Virtual de Java(JVM) en su versión 7u51 o superior.
- **Hardware**
  - RNFH-3. Periféricos:** Mouse.
  - RNFH-4.** Procesador Intel Core i3 o superior.
  - RNFH-5.** 1 GB o más de memoria RAM.
  - RNFH-6.** CPU 3 GHZ o superior.
- **Restricciones en el Diseño e Implementación**
  - RNF-7.** Se empleará el lenguaje de programación Java en su versión 7 bajo el paradigma de Programación Orientada a Objetos.

## 2.3. Historias de Usuario

Las historias de usuario son la técnica que emplea la metodología XP para especificar los requisitos del software, basado en una breve descripción de su comportamiento utilizando una terminología del cliente sin lenguaje técnico. Estas historias se descomponen en tareas de programación y se asignan a los programadores para ser implementadas (LNCS, s.f.). Para estimar el esfuerzo asociado a la implementación de las historias de usuario los programadores establecen como medida el punto, que equivale a una semana ideal de programación.

A continuación se definen las historias de usuario que fueron abordadas en la solución propuesta

### 2.3. HISTORIAS DE USUARIO

---

Tabla 2.1. Historia de usuario # 1

Historia de usuario	
<b>Número:</b> 1	<b>Nombre:</b> Cargar archivos de proteínas y moléculas pequeñas.
<b>Usuario:</b> Especialista	
<b>Prioridad en negocio:</b> Alta	<b>Riesgo en desarrollo:</b> Bajo
<b>Puntos estimados:</b> 0.5	<b>Iteración asignada:</b> 1
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b> Permite al usuario cargar los archivos de proteínas y moléculas pequeñas para realizar los cálculos correspondientes.	
<b>Observaciones:</b> El usuario debe seleccionar a través del selector de archivos los ficheros con extensión .pdb o .mol para proteínas y moléculas respectivamente.	

Tabla 2.2. Historia de usuario # 2

Historia de usuario	
<b>Número:</b> 2	<b>Nombre:</b> Conectar aplicación a la base de datos de Catepsinas
<b>Usuario:</b> Especialista	
<b>Prioridad en negocio:</b> Alta	<b>Riesgo en desarrollo:</b> Alto
<b>Puntos estimados:</b> 1	<b>Iteración asignada:</b> 1
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b> Permite acceder a la base de datos de las Catepsinas de manera remota, para obtener las coordenadas del centro de masa de los aminoácidos y calcular los índices topográficos.	
<b>Observaciones:</b>	

Tabla 2.3. Historia de usuario # 3

Historia de usuario	
<b>Número:</b> 3	<b>Nombre:</b> Calcular centro de masas de los aminoácidos
<b>Usuario:</b> Especialista	
<b>Prioridad en negocio:</b> Alta	<b>Riesgo en desarrollo:</b> Alto
<b>Puntos estimados:</b> 0.5	<b>Iteración asignada:</b> 1
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b> Permite calcular el centro de masa de los aminoácidos de las proteínas cargadas.	
<b>Observaciones:</b> Sin observaciones.	

Tabla 2.4. Historia de usuario # 4

Historia de usuario	
<b>Número:</b> 4	<b>Nombre:</b> Calcular índices topográficos de los aminoácidos
<b>Usuario:</b> Especialista	
<b>Prioridad en negocio:</b> Alta	<b>Riesgo en desarrollo:</b> Alto
<b>Puntos estimados:</b> 3	<b>Iteración asignada:</b> 2
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b> Permite calcular el Índice de Estado Electrotopográfico, el Índice de Estado Refractotopográfico y el Índice de Estado Lipotopográfico para aminoácidos.	
<b>Observaciones:</b> Este índices topográficos se obtienen al reducir el grafo de la estructura química inicial utilizando el criterio de centro de masa. La salida mostrada aparece en el mismo orden del archivo de la proteína.	

Tabla 2.5. Historia de usuario # 5

Historia de usuario	
<b>Número:</b> 5	<b>Nombre:</b> Calcular descriptores moleculares para moléculas pequeñas
<b>Usuario:</b> Especialista	
<b>Prioridad en negocio:</b> Alta	<b>Riesgo en desarrollo:</b> Bajo
<b>Puntos estimados:</b> 1	<b>Iteración asignada:</b> 3
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b> Permite calcular el coeficiente de partición octanol-agua ALOGP de Ghose-Crippen , el descriptor ALOGP2 y el Índice de Refractividad Molar para Átomos (AMR).	
<b>Observaciones:</b> Sin observaciones.	

Tabla 2.6. Historia de usuario # 6

Historia de usuario	
<b>Número:</b> 6	<b>Nombre:</b> Exportar a Excel
<b>Usuario:</b> Especialista	
<b>Prioridad en negocio:</b> Alta	<b>Riesgo en desarrollo:</b> Bajo
<b>Puntos estimados:</b> 0.5	<b>Iteración asignada:</b> 3
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b> Permite exportar los índices topográficos de los aminoácidos y los descriptores moleculares calculados a una hoja electrónica Excel para realizar investigaciones con estos valores.	
<b>Observaciones:</b> Sin observaciones.	

Tabla 2.7. Historia de usuario # 7

Historia de usuario	
<b>Número:</b> 7	<b>Nombre:</b> Visualizar el efecto perturbativo de los aminoácidos en un gráfico de dispersión
<b>Usuario:</b> Especialista	
<b>Prioridad en negocio:</b> Alta	<b>Riesgo en desarrollo:</b> Alta
<b>Puntos estimados:</b> 1	<b>Iteración asignada:</b> 3
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b> Permite analizar la influencia de la perturbación de un aminoácido con los restantes de una misma proteína a través de un gráfico de dispersión.	
<b>Observaciones:</b>	

## 2.4. Planificación

### 2.4.1. Estimación del esfuerzo por historia de usuario

La estimación de esfuerzos por historia de usuario permite la identificación de cargas de trabajo del equipo de desarrollo y determina el plan de acción del proyecto. A partir de los puntos estimados para cada historia de usuario, se puede establecer el esfuerzo a realizar.

Un punto de estimación es el equivalente a una semana ideal de programación. Es necesario aclarar que 1 semana laboral equivale a los 5 días laborales de la misma.

Tabla 2.8. Estimación de esfuerzo por historia de usuario

Iteración	Historias de usuario	Puntos estimados (semanas)
1	1 Cargar Archivos de proteínas y moléculas pequeñas	0.5
	2 Conectar aplicación a la base de datos de Catepsinas	1
	3 Calcular centro de masa de los aminoácidos.	1
	4 Calcular índices topográficos para aminoácidos.	2
2	5 Calcular índices topográficos para aminoácidos.	2
	6 Calcular índices topográficos para aminoácidos.	2
3	7 Calcular descriptores moleculares para moléculas pequeñas.	1
	8 Exportar valores calculados a documento Excel	0.5
	9 Visualizar el efecto perturbativo de los aminoácidos en un gráfico de dispersión	1
<b>Total</b>		<b>11.0</b>



### 2.4.2. Plan de iteraciones

El ciclo de desarrollo de software guiado por XP se caracteriza por ser iterativo e incremental, por lo que se realizan varias iteraciones sobre el sistema antes de su fase de producción. (Ma Carmen Letelier, s.f.)

Cada iteración corresponde a diferentes historias de usuario las cuales son expresadas en tareas de programación. Cada una de ellas es asignada a un programador para su implementación.

La herramienta para el cálculo de índices topográficos de aminoácidos para proteínas y descriptores moleculares se realizará en 3 iteraciones, durante las cuales se implementarán de manera incremental todas las historias de usuario. En la primera iteración se priorizan las historias de usuario de las cuales dependen el resto. Todas están acompañadas de su tiempo de implementación estimado en semanas.

Tabla 2.9. Plan de duración de las iteraciones

Iteración	Historias de usuario	Duración (semanas)
1	1 Cargar Archivos de proteínas y moléculas pequeñas	4.5
	2 Conectar aplicación a la base de datos de Catepsinas	
	3 Calcular centro de masa de los aminoácidos.	
	4 Calcular índices topográficos para aminoácidos.	
2	5 Calcular índices topográficos para aminoácidos.	4
	6 Calcular índices topográficos para aminoácidos.	
3	7 Calcular descriptores moleculares para moléculas pequeñas.	2.5
	8 Exportar valores calculados a documento Excel	
	9 Visualizar el efecto perturbativo de los aminoácidos en un gráfico de dispersión	
<b>Total</b>		<b>11.0</b>

### 2.5. Tarjetas CRC

Las tarjetas CRC (Clase-Responsabilidad-Colaboración), son utilizadas por la metodología XP para la representación de las clases y guiar el proceso de desarrollo de la solución. Se dividen en tres secciones que contienen la información del nombre de la clase, sus responsabilidades y sus colaboradores. La estructura de estas tarjetas es la siguiente:

Tabla 2.10. Tarjeta CRC # 1

Tarjeta CRC	
Clase: Nombre de la clase que se esta modelando	
Responsabilidad	Colaboración

Continúa en la próxima página

Tabla 2.10. Continuación de la página anterior

Es una descripción de alto nivel del propósito de la clase.	Indica con cuáles clases se requiere la colaboración para cumplir con la responsabilidad.
---	---

Seguidamente se exponen las tarjetas CRC correspondientes a las clases principales de la aplicación. Para consultar las restantes puede dirigirse al anexo.

Tabla 2.11. Tarjeta CRC # 2

Tarjeta CRC	
<b>Clase:</b> CalcIndTop	
Responsabilidad	Colaboración
<ul style="list-style-type: none"> <li>• Calcular centro de masa de cada aminoácido para reducir el grafo químico.</li> <li>• Calcular valor de la perturbación de cada aminoácido.</li> <li>• Realizar el cálculo de los índices topográficos de cada aminoácido de la proteína cargada.</li> </ul>	Indice ExportExcel

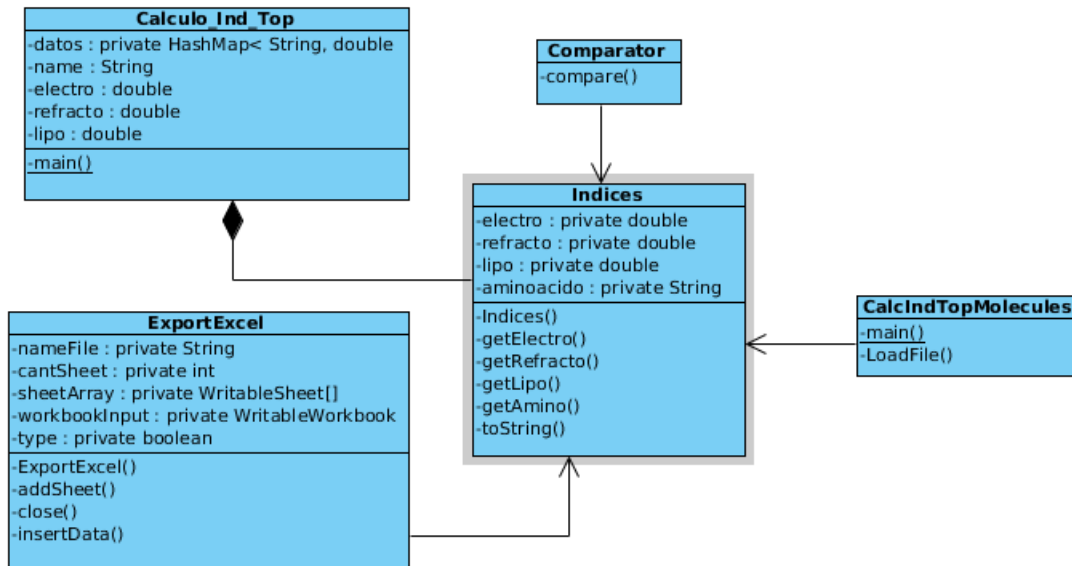
Tabla 2.12. Tarjeta CRC # 3

Tarjeta CRC	
<b>Clase:</b> CalcDescMolecules	
Responsabilidad	Colaboración
<ul style="list-style-type: none"> <li>• Calcular coeficiente de partición (ALOGP) de la molécula cargada.</li> <li>• Calcular la refractividad molar (AMR) de la molécula cargada.</li> <li>• Procesar datos del fichero de entrada usando las clases necesarias de la biblioteca CDK.</li> </ul>	ALOGPDescriptor FileManager

## 2.6. Diagrama de clases

A pesar que las tarjetas CRC son una notación alterna a los diagramas de clases en la metodología XP, de igual manera se considera hacer uso de este modelo puesto que permite mostrar la interacción de las clases

y sus colaboraciones con carácter más gráfico.



## 2.7. Descripción de la arquitectura

La definición oficial de Arquitectura de Software, conocida también como Arquitectura Lógica es la que brinda el documento de IEEE Std 1471-2000, adoptada también por Microsoft, la cual plantea que: La Arquitectura de Software es la organización fundamental de un sistema encarnada en sus componentes, las relaciones entre ellos, el ambiente y los principios que orientan su diseño y evolución (Hilliard, 2000.).

En el presente trabajo de diploma se propone como arquitectura base para la organización estructural del módulo, una arquitectura en capas. A continuación se detalla la estructura de la aplicación dividida según MVC:

- **Modelo:** Está compuesto por las clases que representan la información que será utilizada para el manejo de los datos. Estas clases describen las propiedades de los objetos y son accedidas directamente por la capa controladora.
- **Vista:** Es la interfaz visual con la cual interactúa el usuario. En la solución desarrollada esta capa está compuesta por componentes gráficos tales como ventanas, barras de herramientas, mensajes de diálogos y tablas con la salida de los datos.
- **Controlador:** Son las clases que reciben las peticiones del usuario, obtienen la información requerida de los modelos y luego la procesan para mostrar los resultados al usuario a través de las vistas.

## 2.8. Patrones de Diseño

Un patrón de diseño es una descripción de clases y objetos comunicándose entre sí, la cual es adaptada para resolver un problema de diseño general en un contexto particular. Su objetivo es agrupar la experiencia en diseño de programas orientados a objetos de forma tal que las personas puedan utilizarlos con efectividad (Martínez, s.f.).

Existen muchas otras definiciones asociadas al término que coinciden con las buenas prácticas asociadas al uso de los patrones de diseño. Seguidamente se exponen los patrones utilizados en la aplicación.

- Delegation(When Not Use Inheritance) / Delegación (Cuando no usa herencia)

Es una forma de extender y reutilizar las funcionalidades de una clase escribiendo una clase adicional con funcionalidad añadida utilizando instancias de la clase original. La clase ExportExcel hace uso de este patrón ya que reutiliza las funcionalidades de la API JExcel para la exportación de los índices topográficos.

- Singleton o Solitario

Se utiliza para garantizar que una clase solo tiene una instancia única por lo que provee un punto de acceso global a él. Solamente puede ser instanciado un objeto de esa clase en la aplicación. En la herramienta desarrollada la clase JDBCUtilConnection emplea este patrón, asegurando que se cree solo una instancia durante la ejecución de la herramienta. Además, esto permite no sobrecargar al servidor con demasiadas conexiones y ahorrar recursos del sistema.

- Patrones GRASP:

Patrón Experto: La clase CalcIndTop es la encargada de controlar todas las acciones del programa ya que es la que contiene todas las instancias e informaciones necesarias para llevar la lógica del programa en general.

Patrón Creador: La clase CalcIndTop es la encargada de crear la clase Índice puesto que contiene los índices topográficos que se pasarán a un objeto Índice cuando sea creado.

## 2.9. Conclusiones parciales

Se describió el diseño del sistema así como todo el proceso de desarrollo de software durante la implementación del mismo. Se caracterizó la arquitectura de software así como el patrón arquitectónico empleado. También se analizaron los patrones de diseño presentes en la aplicación y los requisitos tanto funcionales como no funcionales.

---

## Implementación y pruebas del sistema

---

### 3.1. Introducción

El presente capítulo describe el proceso de implementación de los algoritmos para el cálculo de los descriptores. Además, se detallan las tareas de ingeniería según plantea la metodología XP. También se presentan las pruebas de aceptación que aseguran la calidad de la herramienta desarrollada.

### 3.2. Estilo de codificación

La refactorización del código es una de las prácticas a aplicar durante el proceso de desarrollo de software según plantea la metodología XP, de manera tal que todo el personal con conocimiento del tema pueda corregir y extender cualquier parte del producto (Joskowicz, 2008). El uso de estándares durante el proceso de programación de la solución es fundamental, para así lograr que el código sea fácilmente entendible.

Por consiguiente en la solución abordada se utilizaron las siguientes convenciones:

- En todos los casos se utilizarán nombres descriptivos que ayuden a una mejor comprensión del código.
- Los nombres de los métodos y clases usadas estarán escritos en idioma inglés.
- Para la nomenclatura de las clases y los métodos se utilizará el estilo de capitalización CamelCase, con el cual se capitaliza la primera letra de cada palabra. Ejemplo `getAllCentreofMass`.
- La llave de apertura de cada estructura de código se localizará en la línea inferior a la declaración.
- No se utilizará una misma línea para definir más de una variable y siempre que sea posible éstas se inicializarán en su misma línea de declaración.
- Los nombres de las variables comenzarán siempre con minúsculas.

El siguiente fragmento muestra un ejemplo del código de la aplicación que evidencia lo antes expuesto.

Código fuente 3.1. Ejemplo de código en Java

```
1  for (Map.Entry<String , IAtomContainer> aminoacid : aminoacids.  
    entrySet()) {
```

```
2         String key = aminoacid.getKey();
3         IAtomContainer atomContainer = aminoacid.getValue()
4         ;
5
6         IsotopeFactory.getInstance(chemFile.getBuilder()).
7         configureAtoms(atomContainer);
8
9         Point3d threeDCentreOfMass = GeometryTools.
10        get3DCentreOfMass(atomContainer);
11
12        PDBAtom atomo = (PDBAtom) atomContainer.getAtom(0);
13
14        int serialNumb = atomo.getSerial();
15
16        String keyCentreofMass = String.format("%s %s",
17        key,
18        serialNumb
19        );
20
21        CentreOfMass.put(keyCentreofMass,
22        threeDCentreOfMass);
23    }
```

---

### 3.3. Herramienta para el cálculo de índices topográficos de proteínas y descriptores de moléculas pequeñas

La herramienta desarrollada permite el cálculo de los índices topográficos sobre aminoácidos en las proteínas. Además, implementa los descriptores moleculares AMR(refractividad molar) y ALogP (coeficiente de partición octanol-agua) de acuerdo al modelo de Ghose-Crippen. Asimismo, proporciona analizar el efecto perturbativo entre los aminoácidos de las proteínas a través del empleo de gráficas de dispersión que visualizan dicha información.

La aplicación está compuesta principalmente por tres funcionalidades: el cálculo de los índices para aminoácidos, el cálculo para descriptores moleculares, el acceso remoto a la base de datos de proteínas y por

último el análisis del efecto perturbativo entre los aminoácidos.

### 3.3.1. Cálculo de los índices topográficos para aminoácidos

Para el proceso de lectura de los archivos de moléculas pequeñas y proteínas se utilizó en la aplicación la biblioteca CDK la cual facilitó la carga y depuración de estos ficheros. Además se emplearon distintos métodos de esta biblioteca que facilitaron diversas tareas durante la implementación de los descriptores abordados en este trabajo.

A continuación se expone la interfaz que muestra el cálculo de los índices electrotopográfico, refractotopográfico y lipotopográfico de los aminoácidos de una proteína.

Amino Acids	Disturbance EL	Disturbance R	Disturbance L	Electrotopogr	Refractotopogr	Upotopographi
LEU A 1	3.3740815828...	2.7977828446...	0.3803588442...	30.874081582...	36.375582844...	-0.393341155...
PRO A 2	5.232193973...	-6.365237630...	-0.245405082...	13.601199360...	18.012262369...	-1.627105082...
GLU A 3	1.9744357297...	-0.434666963...	-0.368353797...	27.974435729...	29.009633036...	-2.008153797...
SER A 4	-5.3741309592...	-9.179539029...	-0.509479775...	13.459142741...	12.861460970...	-2.104978775...
PHE A 5	7.1162875480...	17.649932601...	1.8090688595...	37.116287548...	63.180832601...	2.1931688595...
ASP A 6	-0.903754800...	-5.143965763...	-0.231545044...	22.596245199...	21.388734236...	-1.583345044...
ALA A 7	-8.038441489...	-13.78688191...	-0.221231148...	10.294891843...	6.9230190883...	-1.544431148...
ARG A 8	10.234682684...	8.7159794717...	-1.260963404...	42.734682684...	46.015279471...	-3.545263404...
GLU A 9	1.7500244795...	-0.712940573...	-0.375952792...	27.750024478...	28.731359425...	-2.015752792...
GLN A 10	2.4347567382...	-1.909582139...	-0.910125573...	29.434756738...	27.756917860...	-2.944925573...
TRP A 11	13.508836936...	31.045217114...	2.3132772869...	48.842170269...	87.728417114...	3.0664772869...
SER A 12	-4.230149106...	-6.892055811...	-0.318246469...	14.603184226...	15.158944188...	-1.913744669...
ASN A 13	1.8720296165...	-2.318669468...	-0.517740289...	26.172029616...	24.436230531...	-2.264540289...
CYS A 14	-7.804977989...	-1.451529419...	0.7622908761...	9.028355434...	26.716470580...	0.1380908761...
PRO A 15	-4.761649922...	-5.329204641...	-0.140134121...	14.071683410...	19.048295358...	-1.521834121...
THR A 16	-0.323055142...	-3.531689852...	-0.804285517...	22.676944857...	23.140710147...	-2.638185517...
ILE A 17	4.3189328447...	11.315807171...	0.1702297471...	30.818932844...	48.809107171...	-0.951870252...
ALA A 18	-6.537825526...	-11.06208672...	-0.085436218...	11.765507807...	9.6478132719...	-1.408636218...
GLN A 19	4.5092681194...	0.1841023042...	-1.011703202...	31.509268119...	29.850602304...	-3.046503202...
ILE A 20	3.8965736203...	10.561618746...	0.1464846502...	30.386573620...	48.054918746...	-0.975615349...
ARG A 21	13.108312356...	11.281959482...	-1.561175763...	45.608312356...	48.581259482...	-3.845475763...
ASP A 22	0.8000028696...	-3.563358782...	-0.219621364...	24.300002869...	22.969341217...	-1.571421364...
GLN A 23	7.6926414158...	2.0417238982...	-1.428059643...	34.692641415...	31.708223898...	-3.462859643...

Figura 3.1. Cálculo de índices topográficos en aminoácidos de la proteína 1CPJ

### 3.3.2. Exportación de descriptores calculados

Puede resultar importante para el potencial usuario, tener la posibilidad de realizar diferentes análisis con los descriptores calculados por la aplicación. Estos análisis pueden deberse a un interés particular y diferente a las funcionalidades empleadas en la herramienta, determinado por el objetivo de la investigación que realice dicho usuario final. Por lo que se le creó la opción adicional de poder exportar la información resultante de los cálculos a una hoja electrónica en formato Excel, que permita la posterior manipulación de estos valores. De igual forma, la aplicación permite importar de regreso estos valores para su revisión desde la propia aplicación.

### 3.3.3. Acceso remoto a la base de datos de Catepsinas B

La aplicación desarrollada permite el acceso a una base de datos centralizada que contiene información estructural sobre un set de Catepsinas B sobre las cuales serán calculados los índices topográficos correspondientes a sus aminoácidos. La barra de herramientas del software desarrollado posee las siguientes tres opciones:

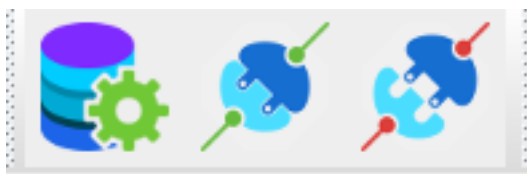



Figura 3.2. Opciones de conexión de la aplicación

 **Configuración de la conexión:** Permite configurar la conexión con el servidor de base de datos especificado. Al presionar el botón se muestra una ventana donde el usuario debe proporcionar los valores de los campos Server, Port, Name DB, User y Password necesarios para efectuar la conexión.

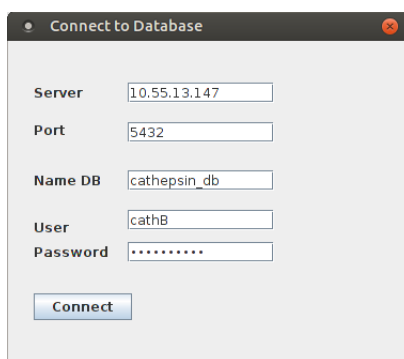



Figura 3.3. Ventana que permite la configuración para acceder a la base de datos remota

Una vez configurada la conexión la aplicación carga las proteínas localizadas en el servidor remoto.

 **Conectar servidor de base de datos:** Permite establecer la conexión al servidor de base de datos si ya fueron configurados los parámetros necesarios.

 **Desconectar servidor de base de datos:** Permite cerrar la conexión efectuada con anterioridad.

### 3.3.4. Visualización científica del efecto perturbativo en los aminoácidos

Esta funcionalidad permite analizar a través de un gráfico de dispersión la influencia de la perturbación que ejercen los demás aminoácidos de una proteína cargada en la aplicación sobre un aminoácido de dicha proteína especificado por el usuario. Al presionar el botón Analyze se visualiza una ventana con las proteínas cargadas en la herramienta y los aminoácidos correspondientes a la misma.



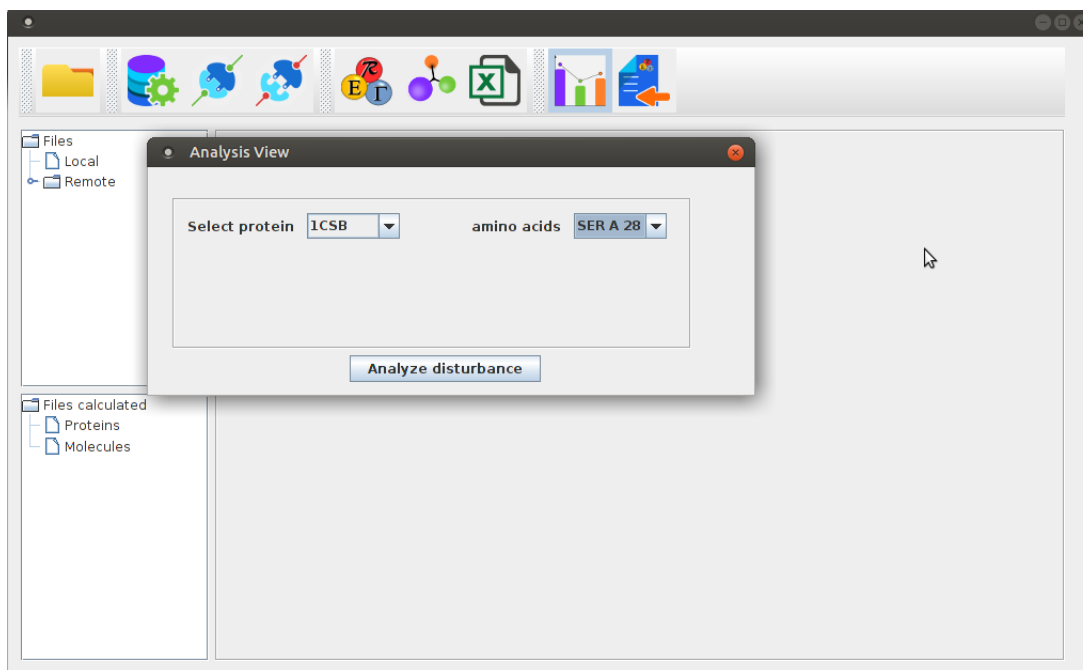


Figura 3.4. Ventana para la selección del aminoácido que será analizado

Cuando el usuario presiona el botón Analyze disturbance la aplicación visualiza la influencia de la perturbación del resto de los aminoácidos con respecto al aminoácido seleccionado a través de un gráfico de dispersión.

### 3.4. Tareas de ingeniería

Las historias de usuario son descompuestas en tareas de ingeniería para ser implementadas durante una iteración. A continuación se presentan las tareas de ingeniería desarrolladas.

- **Historia de Usuario : Calcular índices topográficos de los aminoácidos.**

Tabla 3.1. Tarea de ingeniería # 1

Tarea	
Número de tarea: 1	Número de Historia de usuario: 4
Nombre de la tarea: Calcular Índice de Estado Electrotopográfico de los aminoácidos	
Tipo de tarea: Desarrollo	Puntos estimados: 2
Fecha de inicio: 15 de febrero de 2016	Fecha de fin: 4 de marzo de 2016
Programador responsable: Alejandro Villegas La Rosa	

Continúa en la próxima página

Tabla 3.1. Continuación de la página anterior

<b>Descripción:</b> Se implementan varias funcionalidades auxiliares sobre las cuales dispone la función principal para el cálculo del índice electrotopográfico. Se calcula el valor de la perturbación total de cada aminoácido de la proteínas además de los valores intrínsecos correspondientes a cada aminoácido
--

Tabla 3.2. Tarea de ingeniería # 2

Tarea	
<b>Número de tarea:</b> 2	<b>Número de Historia de usuario:</b> 4
<b>Nombre de la tarea:</b> Calcular Índice de Estado Refractotopográfico de los aminoácidos	
<b>Tipo de tarea:</b> Desarrollo	<b>Puntos estimados:</b> 2
<b>Fecha de inicio:</b> 7 de marzo de 2016	<b>Fecha de fin:</b> 16 de marzo de 2016
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b> Se procede de igual manera que en el caso del índice electrotopográfico pero esta vez se emplean valores intrínsecos diferentes para los aminoácidos.	

Tabla 3.3. Tarea de ingeniería # 3

Tarea	
<b>Número de tarea:</b> 3	<b>Número de Historia de usuario:</b> 4
<b>Nombre de la tarea:</b> Calcular Índice de Estado Lipotopográfico de los aminoácidos	
<b>Tipo de tarea:</b> Desarrollo	<b>Puntos estimados:</b> 2
<b>Fecha de inicio:</b> 21 de marzo de 2016	<b>Fecha de fin:</b> 25 de marzo de 2016
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b> Se usa el mismo procedimiento que para el resto de los índices topográficos implementados a diferencia de los valores intrínsecos utilizados.	

- **Historia de Usuario : Calcular descriptores de moléculas pequeñas.**

Tabla 3.4. Tarea de ingeniería # 4

Tarea	
<b>Número de tarea:</b> 4	<b>Número de Historia de usuario:</b> 5
<b>Nombre de la tarea:</b> Calcular coeficiente de partición (ALOGP) de moléculas pequeñas	
<b>Tipo de tarea:</b> Desarrollo	<b>Puntos estimados:</b> 0.5
<b>Fecha de inicio:</b> 5 de abril de 2016	<b>Fecha de fin:</b> 8 de abril de 2016
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b> Se implementa el coeficiente de partición octanol-agua ALOGP de Ghose-Crippen. Para esto se hace uso del método ALOGPDescriptor de la biblioteca CDK.	

Tabla 3.5. Tarea de ingeniería # 5

Tarea	
<b>Número de tarea:</b> 5	<b>Número de Historia de usuario:</b> 5
<b>Nombre de la tarea:</b> Calcular descriptor ALOGP2 de moléculas pequeñas	
<b>Tipo de tarea:</b> Desarrollo	<b>Puntos estimados:</b> 0.5
<b>Fecha de inicio:</b> 11 de abril de 2016	<b>Fecha de fin:</b> 13 de abril de 2016
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b> Este descriptor calcula el cuadrado del coeficiente de partición octanol-agua (ALOGP). Se emplea la biblioteca CDK para el cálculo de este descriptor.	

Tabla 3.6. Tarea de ingeniería # 6

Tarea	
<b>Número de tarea:</b> 6	<b>Número de Historia de usuario:</b> 5
<b>Nombre de la tarea:</b> Calcular la refractividad molar (AMR) de moléculas pequeñas	
<b>Tipo de tarea:</b> Desarrollo	<b>Puntos estimados:</b> 0.5
<b>Fecha de inicio:</b> 19 de abril de 2016	<b>Fecha de fin:</b> 22 de abril de 2016
<b>Programador responsable:</b> Alejandro Villegas La Rosa	
<b>Descripción:</b>	

## 3.5. Pruebas

En la metodología XP la fase de producción es aquella en la cual se realizan las pruebas a la aplicación y las revisiones necesarias que aseguren la calidad del sistema antes de ser trasladado al entorno del cliente. Las pruebas son una actividad en la cual un sistema o componente es ejecutado bajo unas condiciones o requerimientos especificados, los resultados son observados y registrados, y es hecha una evaluación de algún aspecto determinado (Joskowicz, 2008).

### 3.5.1. Pruebas de aceptación

Las pruebas de aceptación tienen como objetivo verificar que el software esté listo y que puede ser usado por los usuarios finales para ejecutar aquellas funciones y tareas para las cuales fue construido. En la metodología XP las pruebas de aceptación son creadas según las historias de usuario. El cliente especifica uno o diversos escenarios para comprobar la correcta implementación de las historias de usuario. Este tipo de pruebas son consideradas como pruebas de caja negra. Seguidamente se presentan las pruebas de aceptación empleadas.

Tabla 3.7. Prueba de aceptación # 1

Caso de prueba de aceptación	
<b>Código:</b> HU1_P1	<b>Historia de usuario:</b> 2
<b>Nombre:</b> Cargar archivo de proteínas o moléculas pequeñas.	
<b>Descripción:</b> Prueba para la funcionalidad Cargar archivo de proteínas y moléculas pequeñas.	
<b>Condiciones de ejecución:</b>	
<b>Pasos de ejecución:</b> El usuario intenta cargar archivos de proteínas o moléculas pequeñas.	
<b>Resultados esperados:</b> Los archivos de proteínas o moléculas pequeñas se cargan correctamente en la aplicación.	

Tabla 3.8. Prueba de aceptación # 2

Caso de prueba de aceptación	
<b>Código:</b> HU2_P1	<b>Historia de usuario:</b> 7
<b>Nombre:</b> Conectar aplicación a la base de datos de Catepsinas.	
<b>Descripción:</b> Prueba para la funcionalidad Conectar la aplicación a la base de datos de Catepsinas.	
<b>Condiciones de ejecución:</b> Los campo para la configuración de la conexión remota deben ser correctos.	
<b>Pasos de ejecución:</b> El usuario intenta conectarse a la base de datos de Catepsinas.	
<b>Resultados esperados:</b> Se produce la conexión de la aplicación a la base de datos de Catepsinas correctamente.	

Tabla 3.9. Prueba de aceptación # 3

Caso de prueba de aceptación	
<b>Código:</b> HU2_P1	<b>Historia de usuario:</b> 7
<b>Nombre:</b> Calcular índices topográficos de los aminoácidos.	
<b>Descripción:</b> Prueba para la funcionalidad calcular índices topográficos de los aminoácidos.	
<b>Condiciones de ejecución:</b> El usuario carga un fichero de proteínas a la aplicación o se conecta de manera remota a la base de datos de Catepsinas.	

Continúa en la próxima página

Tabla 3.9. Continuación de la página anterior

<p><b>Pasos de ejecución:</b>                  El usuario presiona el botón para el cálculo de índices topográficos de la aplicación.                  El usuario selecciona las proteínas sobre las cuales desea calcular los índices de sus aminoácidos.</p>
<p><b>Resultados esperados:</b> La aplicación visualiza una tabla con los índices topográficos para aminoácidos de la proteína seleccionada.</p>

Tabla 3.10. Prueba de aceptación # 4

Caso de prueba de aceptación	
<b>Código:</b> HU5_P1	<b>Historia de usuario:</b> 5
<b>Nombre:</b> Calcular descriptores moleculares de moléculas pequeñas.	
<b>Descripción:</b> Prueba para la funcionalidad Calcular descriptores moleculares de moléculas pequeñas.	
<b>Condiciones de ejecución:</b> El usuario debe cargar previamente archivos de moléculas pequeñas.	
<b>Pasos de ejecución:</b> El usuario presiona el botón para el cálculo de los descriptores de moléculas pequeñas.	
<b>Resultados esperados:</b> Se visualiza una ventana que contiene el valor de los descriptores calculados de la molécula seleccionada.	

Tabla 3.11. Prueba de aceptación # 5

Caso de prueba de aceptación	
<b>Código:</b> HU6_P1	<b>Historia de usuario:</b> 7
<b>Nombre:</b> Exportar índices topográficos a formato Excel.	
<b>Descripción:</b> Prueba para la funcionalidad Exportar descriptores a Excel.	
<b>Condiciones de ejecución:</b> El usuario debe previamente calcular los índices topográficos para aminoácidos o los descriptores moleculares. El usuario debe seleccionar un directorio donde salvar el documento con los valores calculados.	
<b>Pasos de ejecución:</b> El usuario selecciona los descriptores que desea exportar de la lista de ficheros con su información calculada. El usuario presiona el botón Export Excel y selecciona el directorio para exportar los índices topográficos de los aminoácidos.	

Continúa en la próxima página

Tabla 3.11. Continuación de la página anterior

<b>Resultados esperados:</b> Se exportan los índices topográficos a un documento Excel en el directorio especificado por el usuario.
--

Tabla 3.12. Prueba de aceptación # 6

Caso de prueba de aceptación	
<b>Código:</b> HU7_P1	<b>Historia de usuario:</b> 7
<b>Nombre:</b> Visualización del efecto de perturbación sobre los aminoácidos de las proteínas.	
<b>Descripción:</b> Prueba para la funcionalidad Visualizar el efecto perturbativo de los aminoácidos en un gráfico de dispersión.	
<b>Condiciones de ejecución:</b> Deben existir proteínas cargadas en la aplicación proveniente de archivos locales como de la conexión remota con el servidor de base de datos.	
<b>Pasos de ejecución:</b> El usuario selecciona el botón destinado a la visualización de la información calculada. Luego el usuario selecciona el aminoácido que desea analizar de la proteínas cargadas en la aplicación	
<b>Resultados esperados:</b> La aplicación visualiza un gráfico de dispersión que representa el efecto de perturbación de los aminoácidos restantes sobre el que fue seleccionado.	

### 3.5.2. Análisis de las pruebas de aceptación

Durante cada una de las iteraciones definidas se desarrollaron 15 casos de pruebas de aceptación. A continuación se muestra en un gráfico de pastel los resultados de las pruebas realizadas.

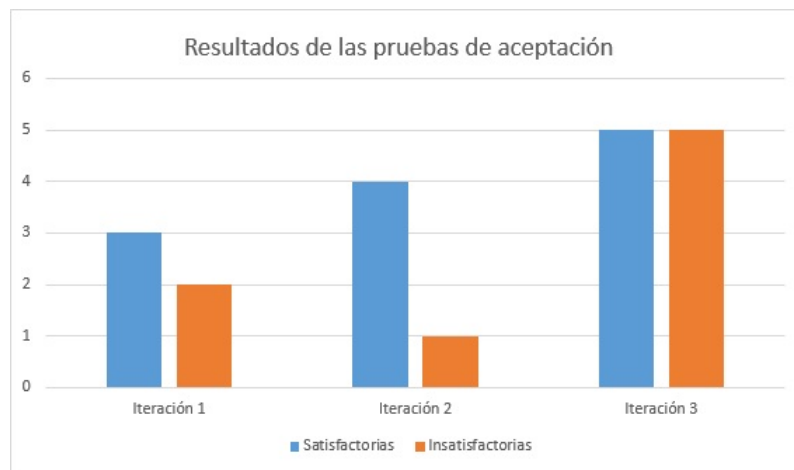


Figura 3.5. Resultados de las pruebas de aceptación

En la primera iteración se realizaron 5 pruebas, donde 2 de ellas permitieron detectar problemas durante

el proceso de conexión a la base de datos, puesto que la aplicación no validaba correctamente los campos vacíos. Esta vulnerabilidad fue corregida posteriormente. En la segunda iteración fueron detectados errores durante la visualización del gráfico de dispersión que fueron solucionados más adelante. En la tercera iteración todas las pruebas efectuadas fueron satisfactorias.

### 3.5.3. Validación de los descriptores implementados en la aplicación

Los descriptores moleculares que se implementaron en la aplicación fueron calculados utilizando los métodos provistos por la biblioteca CDK. Como se explicó en el primer capítulo, esta biblioteca proporciona un conjunto de funcionalidades para la informática molecular que son frecuentemente utilizadas por la comunidad científica, lo que acredita aún más el adecuado funcionamiento de estos algoritmos.

En el caso de los índices topográficos para aminoácidos han sido utilizados en trabajos anteriores y en publicaciones científicas a nivel internacional por Carrasco y cols. (R Carrasco-Velaz, 2008) ,los cuales fueron validados a través de diversos modelos para la evaluación de los resultado. En tal sentido se puede afirmar que estos valores calculados en la aplicación son apropiados.

## 3.6. Resultados obtenidos

Un hallazgo interesante al realizar el cálculo de proteínas seleccionadas se puede observar al analizar las gráficas de dispersión realizadas. Estas reflejan el comportamiento de la perturbación sobre un determinado aminoácido ejercida por los restantes. Las diferentes gráficas visualizadas por la aplicación poseen todas un comportamiento similar, tipo oscilador anarmónico, determinado por el término de perturbación de la expresión de cálculo. Los efectos dados por la distancia entre los aminoácidos decrece a mayor lejanía, sin embargo, no es el mismo comportamiento para cada uno de ellos.

Existen diferencias visibles en las gráficas en las que se puede afirmar que la disminución del efecto que sobre él ejercen los restantes comienza a disminuir en unos casos a 12 y en otros a 16-18 angstrom. Esto constituye un resultado interesante pues según se conoce, los efectos a larga distancia, como es el caso, desaparecen prácticamente a los 10-12 angstrom.

A modo de ejemplo se muestra el efecto perturbativo visualizado por la aplicación desarrollada sobre los aminoácidos Alanina 7, Arginina 49, Serina 4, todos de la cadena A de la proteína y del aminoácido triptófano 215 de la cadena B ambas cadenas pertenecientes a la proteína 1CPJ.

Es posible observar el desplazamiento de los valores de perturbación a rangos negativos en aminoácidos pequeños como la Alanina (no polar) y serina (polar), y a valores positivos como en la Arginina (positivo) y el Triptófano (voluminoso) .

Esto brinda una posibilidad de análisis para trabajos posteriores con proteínas como serían los estudios de homología de proteínas o de docking. En este último caso, será necesario el desarrollo de nuevas funciones de scoring para las minimizaciones. En el caso de los estudios de homología habrá que acudir a otras técnicas de minería como la minería de grafos

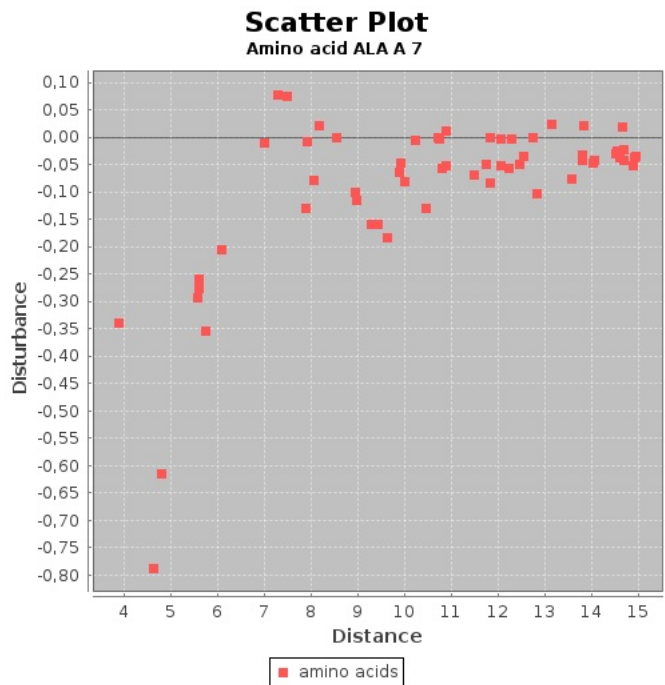


Figura 3.6. Efecto perturbativo de los aminoácidos restantes sobre el aminoácido Alanina 7 en la proteína 1CPJ.

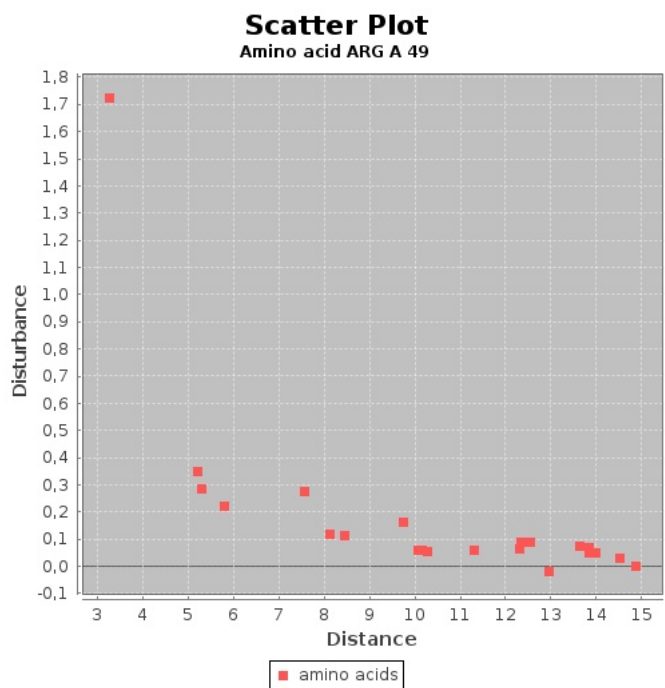


Figura 3.7. Efecto perturbativo de los aminoácidos restantes sobre el aminoácido Arginina 49 de la cadena A en la proteína 1CPJ.



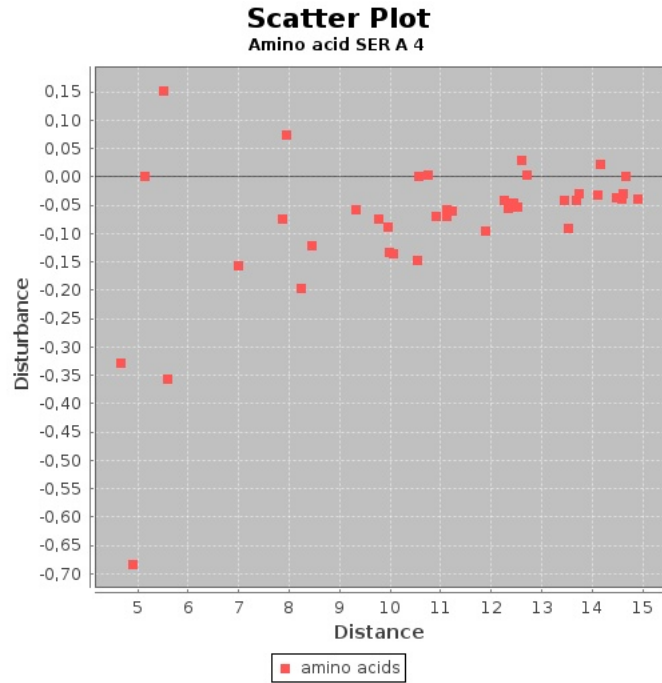


Figura 3.8. Efecto perturbativo de los aminoácidos restantes sobre el aminoácido Serina 4 de la cadena A en la proteína 1CPJ.

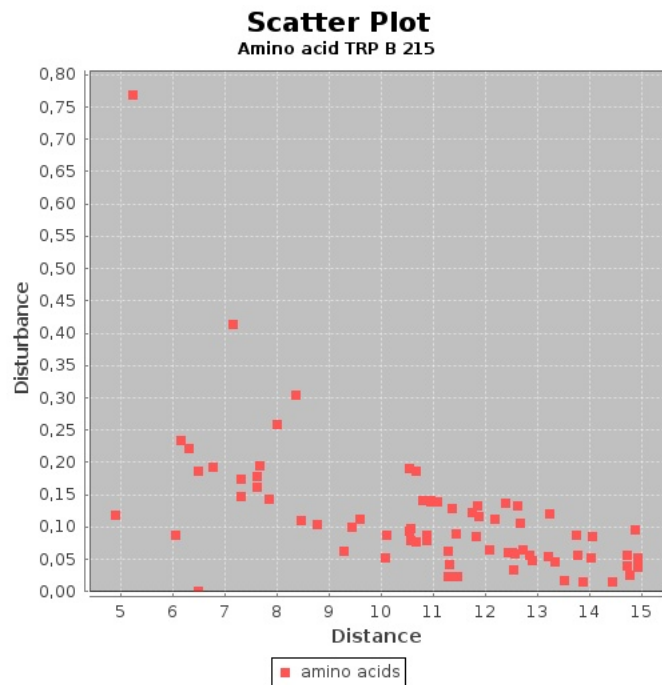


Figura 3.9. Efecto perturbativo de los aminoácidos restantes sobre el aminoácido Triptófano 215 de la cadena B en la proteína 1CPJ.

### **3.7. Conclusiones parciales**

Durante la etapa de implementación el uso de los estándares de codificación hizo posible desarrollar un código reutilizable y limpio capaz de ser comprendido por cualquier equipo de desarrollo que pretenda realizarle mantenimiento. La representación de la interfaz de prueba permitió visualizar la estructura y funcionamiento de la aplicación. Esto hizo posible realizar pruebas de aceptación a la aplicación que validan la calidad del producto creado. El gráfico de dispersión visualizado por la aplicación permitió realizar un análisis que evidencia el comportamiento de los aminoácidos a través de una relación distancia-perturbación.

---

## Conclusiones

---

Como resultado final se puede concluir que:

- La herramienta desarrollada permite el cálculo de índices topográficos para aminoácidos y de dos descriptores de moléculas pequeñas. Además, es posible el acceso remoto a una base de datos de Catepsinas B para facilitar la realización de diversos estudios sobre los valores calculados de estas proteínas.
- Los resultados encontrados en el caso de prueba, sugieren la posibilidad de establecer un rango para las interacciones a larga distancia entre grupos de átomos como es el caso de los aminoácidos en las proteínas.

---

## Recomendaciones

---

Se recomienda extender las funcionalidades de la aplicación a través de la implementación de nuevos descriptores. Es preciso profundizar en las nuevas investigaciones relacionadas con la Catapsina B debido a la manera en que interviene en diferentes enfermedades que tienen amplia repercusión en la actualidad como el ébola, por solo citar un ejemplo.

## 4.1. Tablas de ingeniería

Tabla 4.1. Tarjeta CRC # 4

Tarjeta CRC	
<b>Clase:</b> Indice	
Responsabilidad	Colaboración
<ul style="list-style-type: none"> <li>Almacenar nombre de los aminoácidos que conforman la proteína.</li> <li>Almacenar el valor de los índices topográficos calculados para estos aminoácidos</li> </ul>	ExportExcel CalcIndTop FileManager

Tabla 4.2. Tarjeta CRC # 5

Tarjeta CRC	
<b>Clase:</b> ExportExcel	
Responsabilidad	Colaboración
<ul style="list-style-type: none"> <li>Exportar los datos de los nuevos valores calculados.</li> <li>Manipular información contenida dentro de las hojas electrónicas con la salida de los datos de la aplicación.</li> </ul>	CalcIndTop

Tabla 4.3. Tarjeta CRC # 6

Tarjeta CRC	
<b>Clase:</b> FileManager	
<b>Responsabilidad</b>	<b>Colaboración</b>
<ul style="list-style-type: none"><li>• Leer archivo con los datos de la estructura química.</li><li>• Procesar datos del fichero de entrada usando las clases necesarias de la biblioteca CDK.</li></ul>	CalcIndTop CalcDescMolecules

Tabla 4.4. Tarjeta CRC # 7

Tarjeta CRC	
<b>Clase:</b> CalcDescMolecules	
<b>Responsabilidad</b>	<b>Colaboración</b>
<ul style="list-style-type: none"><li>• Calcular coeficiente de partición (ALOGP) de la molécula cargada.</li><li>• Calcular la refractividad molar (AMR) de la molécula cargada.</li><li>• Procesar datos del fichero de entrada usando las clases necesarias de la biblioteca CDK.</li></ul>	ALOGPDescriptor FileManager

---

## Referencias bibliográficas

---

- BOOCH, James; GRADY; JACOBSON e IVAR Y RUMBAUGH, J, 2000. *El Proceso Unificado de Desarrollo de Software* (vid. pág. 15).
- BROWN, Frank K et al., 1998. Chemoinformatics: what is it and how does it impact drug discovery. *Annual reports in medicinal chemistry*, vol. 33, págs. 375-384 (vid. pág. 1).
- CARRASCO-VELAR, R, 2008. Nuevos descriptores atómicos y moleculares para estudios de estructura-actividad: Aplicaciones [PhD. Thesis]. *La Habana: Universidad de La Habana* (vid. págs. 5, 11, 35).
- CHRISTOPH STEINBECK, and Yongquan Han; KUHN, Stefan; HORLACHER, Oliver y EDGAR LUTTMANN, and Egon Willighagen, 2003. The Chemistry Development Kit (CDK):An Open-Source Java Library for Chemo- and Bioinformatics. *Journal of Chemical Information and Computer Sciences*, vol. 43, n.º 2, págs. 493-500. PMID: 12653513. Url: <http://pubs.acs.org/doi/abs/10.1021/ci025584y> (vid. págs. 8, 13).
- DATABANK, Protein, 2016. *Protein Databank*. [Online; accessed 13-Mar-2016]. Url: <http://www.rcsb.org/pdb/home/home.do> (vid. pág. 2).
- ESCALONA-ARRANZ, Julio César; CARRASCO-VELAR, Ramón y PADRÓN-GARCÍA, Juan Alexander, 2008. *Diseño racional de fármacos* (vid. pág. 5).
- HILLIARD, Rich, 2000. IEEE-Std-1471-2000 Recommended Practice for Architectural Description of Software-Intensive Systems. (Vid. pág. 23).
- JAVA, 2016. *java*. [Online; accessed 15-Feb-2016]. Url: <https://www.java.com> (vid. pág. 8).
- JOSKOWICZ, José, 2008. *Reglas y Prácticas en eXtreme Programming*. Dirección: ( ) (vid. págs. 25, 31).
- LAW, Ruby HP; SMOOKER, Peter M; IRVING, James A et al., 2003. Cloning and expression of the major secreted cathepsin B-like protein from juvenile *Fasciola hepatica* and analysis of immunogenicity following liver fluke infection. *Infection and immunity*, vol. 71, n.º 12, págs. 6921-6932 (vid. pág. 1).
- LNCS. *Ingeniería del software: Metodologías y ciclos de vida* (vid. pág. 16).
- MA CARMEN LETELIER, Patricio y Penadés. *Métodologías ágiles para el desarrollo de software: eXtreme Programming (XP)*. (Vid. pág. 21).
- MARINICH, Jorge A; MAGUNA, Fabiana P; OKULIK, Nora B y CASTRO, Eduardo A. Caracterización óptima de la estructura mediante varios índices de complejidad y conectividad molecular. (Vid. pág. 12).

- MARTÍNEZ, Francisco Javier. *Guía de construcción de software en Java con patrones de diseño* (vid. pág. 24).
- MICHAEL J.S. DEWAR, A Katritzky y KARELSON, M. *Providers of Solutions for Computational Chemistry*. Dirección: <> (vid. pág. 7).
- MOLCONN-ZTM, 2006. Price List for North America and Europe. Url: <<http://www.edusoft-lc.com/molconn/mconpric.html>> (vid. pág. 7).
- NARANJO, D; DÍAZ DE ARCE, Heydi; PÉREZ, E et al., 2007. Modelación por homología de la catepsina b de Fasciola hepatica. *Revista de Salud Animal*, vol. 29, n.º 1, págs. 58-64 (vid. pág. 1).
- NETBEANS, 2016. *Netbeans*. [Online; accessed 15-Feb-2016]. Url: <<https://www.netbeans.org>> (vid. pág. 8).
- PARADIGM, Visual, 2016. *Visual Paradigm*. [Online; accessed 15-Feb-2016]. Url: <<http://www.visualparadigm.com>> (vid. pág. 9).
- POSTGRESQL, 2016. *PostgreSQL, Official Web*. [Online; accessed 15-Feb-2016]. Url: <<http://www.postgresql.org/about/>> (vid. pág. 9).
- RODRÍGUEZ-LEÓN, Alexis René; CARRASCO-VELAR, Ramón; ARENCIBIA-BENÍTEZ, Orlando; GILBERTO-RODRÍGUEZ, José y DOMÍNGUEZ-VEGA, Dairon, 2010. Herramienta para el cálculo de descriptores atómicos y moleculares. (Vid. pág. 1).
- RUIZ, José Javier Sandonís, 2016. *Teoría centro de masa*. [Online; accessed 20-Feb-2016]. Url: <[personales.unican.es/junqueraj/JavierJunquera\\_files/.../Teoria\\_Centros\\_de\\_Masa.pdf](http://personales.unican.es/junqueraj/JavierJunquera_files/.../Teoria_Centros_de_Masa.pdf)> (vid. pág. 10).
- TODESCHINI, Roberto y CONSONNI, Viviana, 2000. *Handbook of Molecular Descriptors* (vid. pág. 4).
- VISWANADHAN, Vellarkad N; GHOSE, Arup K; REVANKAR, Ganapathi R y ROBINS, Roland K, 1990. An estimation of the atomic contribution to octanol-water partition coefficient and molar refractivity from fundamental atomic and structural properties: Its uses in computer aided drug design. *Mathematical and Computer Modelling*, vol. 14, págs. 505-510 (vid. págs. 5, 13).