

Universidad de las Ciencias Informáticas

Facultad 6



Título: “Caracterización del Índice del Estado Refractotopográfico para Aminoácidos”

Trabajo de Diploma para optar por el Título de
Ingeniero en Ciencias Informáticas.

Autores: Elezky Martínez Falcón
Maidelín Prieto Soto

Tutor: Ing. Alexis René Rodríguez León

Co-Tutor: Dr. Ramón Carrasco Velar

Ciudad de La Habana

Cuba

Junio 2009

Usted sabe que ha alcanzado la perfección en el diseño, no cuando usted no tiene nada que añadir, sino cuando no tiene nada más que quitar.

Antoine de Saint Exupery.



DECLARACIÓN DE AUTORÍA

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

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Elezky Martínez Falcón

Firma del Autor

Maidelín Prieto Soto

Firma del Autor

Ing. Alexis René Rodríguez León

Firma del Tutor

DATOS DE CONTACTO

Tutor:

Ing. Alexis René Rodríguez León

Universidad de las Ciencias Informáticas, Ciudad de La Habana, Cuba.

Email: arodriguezl@uci.cu

Co-Tutor:

Dr. Ramón Carrasco Velar.

Universidad de las Ciencias Informáticas, Ciudad de La Habana, Cuba.

Email: rcarrasco@uci.cu

AGRADECIMIENTOS

Una sola persona no puede ser responsable de un logro como éste. Nuestro más sincero agradecimiento:

A nuestro tutor Alexis René, por ser nuestro guía en el desarrollo de la tesis, por brindarnos su ayuda y por soportarnos.

A nuestro tutor Carrasco por sus consejos y dedicación.

Agradecemos a la Universidad de las Ciencias Informáticas, por darnos la oportunidad de formarnos en ella durante los últimos cinco años y hacer de nuestros sueños una realidad.

A todos muchas gracias ya que sin ustedes estas páginas estarían en blanco.

Maidelín:

Quiero agradecer el resultado de toda mi carrera universitaria a mis padres, por su amor y comprensión, por darme aliento para llegar siempre más allá, gracias por su ejemplo.

A mi abuela Ofelia por ser mi otra mamá, por todo su cariño y dedicación.

A mi Abuelo Domingo donde quiera que esté, por todos los consejos que me dio, para que fuera alguien en la vida, sé que ahora se sentiría muy orgulloso de mí.

A mi hermana por ser mi mejor amiga y creer siempre en mí.

A Omar por estar siempre cuando lo necesité, porque nunca me falló en estos cinco años, por ser ante todo mi amigo.

A mi compañera de tesis Elezky por su apoyo en los buenos y los malos momentos que pasamos, por ser mi amiga inseparable.

A mis amigas Ingrid e Imara por estar siempre conmigo, por ser incondicionales y hacer que la vida aquí fuese más fácil.

A toda mi familia por su preocupación.

A Maritza y la gente del apartamento, por acogernos con tanto amor en nuestro último año.

Por último, pero no menos importantes a todos mis compañeros y profesores, que me apoyaron durante todo este tiempo y pusieron su granito de arena para la culminación de este trabajo.

Elezky:

A mis padres Blanca y Jorge por cuidarme siempre, por ser mis guías, mi horizonte, mi límite ante los excesos, mis amigos y mis consejeros, porque sus fuerzas y amor me han dirigido por la vida y sobre todo, por darme las alas que necesitaba para volar.

A mi abuela María por su cariño y sus enseñanzas.

A Enma y Ester por acogerme como una hija.

A Wilber, Dunia, Zeus, Ivón, Carlos Iván por su apoyo y ayuda.

A mi compañera de tesis y amiga Maidelín por tu confianza, amistad, por hacer esta tesis conmigo y por estar ahí.

A Lianet, a Ingrid, a Imara por estos años de amistad.

A mis amigos sin tener en cuenta el tiempo que llevamos conociéndonos, por estar cuando los he necesitado.

A todos los profesores, estudiantes, compañeros, que de una forma u otra han contribuido a mi formación.

A todos gracias por dejarme formar parte de su historia, estoy orgullosa y agradecida por que formen parte de la mía.

DEDICATORIA

“A nuestros padres”

RESUMEN

Para obtener los principios activos responsables de los efectos curativos de un medicamento es importante el uso de descriptores o índices que permitan establecer relaciones entre la estructura químico-física de los compuestos y su actividad biológica con el fin de describir cuantitativamente una propiedad químico-física a partir de un fenómeno biológico dado y así agilizar el descubrimiento de nuevos fármacos. Existen descriptores atómicos, moleculares y se han definido algunos para proteínas. En el presente trabajo, que forma parte del proyecto denominado: "Plataforma para la Predicción de Actividad Biológica en Compuestos Orgánicos" se hace la caracterización del Índice del Estado Refractotopográfico para Aminoácidos, los cuales forman las proteínas, macromoléculas importantes en la naturaleza. Este trabajo constituye un estudio novedoso pues el índice nunca ha sido aplicado a estas estructuras. Para agilizar la caracterización se desarrolló una herramienta con el fin de acelerar los cálculos.

Palabras Clave:

Bioinformática, Descriptor, Diseño de Fármacos.

ÍNDICE

RESUMEN	IV
INTRODUCCIÓN	1
REVISIÓN BIBLIOGRÁFICA	5
1.1 Análisis bioinformático y modelado molecular	5
1.2 Estudios QSAR	6
1.2.1 Técnicas de los estudios QSAR	7
1.3 Descriptores	9
1.4 Índice del Estado Refractotopológico para Átomos	10
1.4.1 Características del Índice del Estado Refractotopológico para Átomos	11
1.4.2 Aportes del índice a los estudios QSAR.....	11
1.5 Conclusiones	11
PROGRAMAS Y METODOLOGÍAS	12
2.1 Lenguaje de Programación: Java	12
2.2 Entorno de desarrollo: Eclipse	13
2.3 Librerías Utilizadas	14
2.3.1 BioJava	14
2.3.2 JFreeChart	15
2.3.3 JCommon.....	16
2.4 Jmol	16
2.5 Estructuras de datos utilizadas	17

2.5.1	Lista.....	17
2.5.2	Tablas Hash	17
2.6	Programa Estadístico.....	18
2.7	Swiss PDB Viewer (SPDBV).....	18
2.8	Procedimientos Utilizados	19
2.8.1	Centro de masa.....	19
2.8.2	Redefinición del Descriptor Atómico Índice del Estado Refractotopológico (R-State, \mathcal{R}) para aplicarlo a proteínas	20
2.8.3	Gráficas	22
2.8.4	Coeficiente de Correlación.....	23
2.9	Conclusiones.....	23
	RESULTADOS Y DISCUSIÓN.....	24
3.1	Herramienta para realizar el cálculo del descriptor	24
3.1.1	Descripción de la Aplicación.....	24
3.2	Principales Cálculos Implementados	28
3.2.1	Cálculo del Centro de Masa.....	29
3.2.2	Cálculo del Índice Refractotopográfico	29
3.3	Fichero de Salida.....	30
3.4	Resultados Experimentales	31
3.4.1	Experimento 1	32
3.4.2	Experimento 2	36
3.4.3	Experimento 3	40
3.5	Conclusiones	47
	CONCLUSIONES GENERALES.....	48
	RECOMENDACIONES.....	49

REFERENCIAS BIBLIOGRÁFICAS.....	50
BIBLIOGRAFÍA.....	51
ANEXOS.....	53
GLOSARIO DE TÉRMINOS.....	66

INTRODUCCIÓN

Originalmente, el descubrimiento de nuevos medicamentos se apoyaba en la modificación de moléculas simples de origen natural, esto se realizaba al azar o en el mejor de los casos los científicos estaban guiados por algunos estudios cuantitativos del efecto que diferentes modificaciones introducidas en la estructura de un fármaco conocido ejercían sobre una actividad farmacológica determinada. Los estudios y descubrimientos se realizaban lentamente a causa del desconocimiento de la estructura de muchos compuestos. Estos procedimientos basados en técnicas de ensayo y error consumen mucho tiempo y requieren enormes recursos económicos, por lo que fue necesario encontrar nuevas vías para la solución de estos problemas.

En los últimos años se han buscado alternativas para acelerar el proceso de descubrimiento de nuevos medicamentos, la integración de varias disciplinas científicas, Biología, Química, Matemáticas, Estadística e Informática se conoce actualmente como Bioinformática, que ha dado respuesta a numerosos problemas en el análisis y modelado de datos biológicos ya que se realiza un estudio más profundo y cuantitativo de los compuestos basado en encontrar correlaciones entre la estructura química y la actividad biológica de éstos.

La tendencia en los estudios bioinformáticos actuales se enfoca hacia las interacciones de los medicamentos con las biomoléculas. Este tipo de estudio constituye un paso de avance en el descubrimiento de nuevos fármacos. Las proteínas son importantes biomoléculas de los seres vivos cuyas propiedades son tan útiles que se emplean como agentes terapéuticos y catalizadores. Conocer la relación estructura-propiedad de las proteínas es de gran utilidad en el campo de la biotecnología y la industria farmacéutica, su comprensión podría acelerar el descubrimiento de fármacos más eficaces para el tratamiento de muchas enfermedades.

Los métodos para el descubrimiento de nuevos medicamentos dependen en gran medida de la selección de los descriptores o índices utilizados para caracterizar la estructura química de los compuestos, por lo que es necesaria la definición de

nuevos índices capaces de reflejar de una manera más acertada las propiedades de los compuestos.

La redefinición de descriptores empleados tiempo atrás en la descripción del comportamiento de átomos y moléculas, para ser aplicados en proteínas, es una nueva alternativa en el diseño de nuevos fármacos pues se ha demostrado que resulta factible de realizar. Los nuevos índices redefinidos son capaces de reflejar un comportamiento análogo observado a una escala menor cuando se aplican a estas estructuras. En el país, algunos investigadores han realizado trabajos significativos siguiendo este enfoque.

El proceso de redefinición de un descriptor, que anteriormente fue utilizado con éxito en los estudios cuantitativos estructura-propiedad de un determinado tipo de compuesto, para aplicarlo a otro nivel, como son las macromoléculas, es de interés para las investigaciones realizadas en La Universidad de las Ciencias Informáticas (UCI) debido al gran potencial que representa para algunos de los proyectos productivos desarrollados en la misma.

La Facultad de Bioinformática de la UCI está desarrollando un proyecto de investigación para el tamizaje virtual de compuestos orgánicos basado en técnicas de teoría de grafo e inteligencia artificial, denominado: "Plataforma para la Predicción de Actividad Biológica en Compuestos Orgánicos". En el mismo se determinó que el comportamiento físico-químico en átomos o moléculas, podría ser análogo en macromoléculas, específicamente en proteínas. Se realizó la redefinición teórica del Índice del Estado Refractotopológico para Átomos [1], ahora aplicado a las proteínas, definiendo el nuevo descriptor como Índice del Estado Refractotopográfico para Aminoácidos. Se hace necesaria la caracterización del mismo para determinar si es viable aplicarlo a las proteínas.

Ante tal situación se plantea el siguiente **problema científico**: Es necesario identificar el comportamiento del valor del Índice del Estado Refractotopográfico para Aminoácidos.

El problema planteado se enmarca en el **objeto de estudio**: Estudio computacional químico-físico y topográfico de proteínas.

En aras de solucionar el problema planteado se define como **objetivo general**: Caracterizar el comportamiento del valor del Índice del Estado Refractotopográfico para Aminoácidos, en diferentes proteínas.

Para cumplir este objetivo se fijan los siguientes **objetivos específicos**:

- ✓ Realizar cálculos matemáticos para determinar el valor del Índice del Estado Refractotopográfico para Aminoácidos
- ✓ Implementar una herramienta para agilizar los cálculos necesarios
- ✓ Graficar y analizar los resultados obtenidos
- ✓ Realizar análisis estadístico.

Para dar cumplimiento a los objetivos trazados se desarrollarán las siguientes tareas:

- ✓ Implementación de un algoritmo para calcular el centro de masa de cada aminoácido en una proteína
- ✓ Obtención de las distancias y las perturbaciones entre los aminoácidos de una proteína
- ✓ Implementación del algoritmo para la obtención de los valores del Índice para cada aminoácido de la proteína
- ✓ Visualización y análisis de la ubicación espacial de los aminoácidos de una proteína, teniendo en cuenta el valor del índice
- ✓ Visualización del comportamiento distancia-perturbación de cada aminoácido respecto al resto a través de gráficas
- ✓ Visualización del comportamiento distancia-perturbación de cada aminoácido respecto a los aminoácidos polares
- ✓ Visualización del comportamiento distancia-perturbación de cada aminoácido respecto a los aminoácidos apolares
- ✓ Determinación de una distancia de corte a partir de los resultados observados en las gráficas

- ✓ Repetición de los cálculos para determinar el Índice teniendo en cuenta la distancia de corte
- ✓ Análisis del coeficiente de correlación.

La estructura del presente trabajo está conformada por tres capítulos, los cuales se mencionan a continuación:

Capítulo I. Fundamentación Teórica.

En este capítulo se ofrece una explicación de los conceptos y fundamentos teóricos necesarios para la comprensión de esta tesis.

Capítulo II. Programas y Metodologías.

Se fundamentan las tendencias, técnicas y tecnologías de software usadas para la solución del problema planteado y los procedimientos empleados para el desarrollo del presente trabajo.

Capítulo III. Resultado y Discusión.

Se exponen los experimentos realizados y el análisis de los resultados obtenidos durante la investigación.

CAPÍTULO 1

REVISIÓN BIBLIOGRÁFICA

En este capítulo se ofrece una explicación de los conceptos y fundamentos teóricos necesarios para la comprensión de esta tesis. Percibir la necesidad actual del descubrimiento de nuevos medicamentos es importante en el tratamiento de varias enfermedades, esto requiere del estudio estructural y cuantitativo de las biomoléculas como una de las principales fuentes en el desarrollo de los mismos.

1.1 Análisis bioinformático y modelado molecular

En las últimas décadas, avances en la Biología Molecular y en las tecnologías disponibles para el desarrollo de investigaciones han dado lugar a nuevos descubrimientos en esta rama de la Biología. Esto ha aportado gran cantidad de información biológica, haciéndose difícil el procesamiento de los datos. Se requiere del uso de técnicas innovadoras para manipular este creciente flujo de información en cuanto a su visualización, modelación, interpretación y análisis. La manipulación de esta enorme cantidad de datos resultantes de la investigación biológica actual necesita del uso de nuevas técnicas, métodos y herramientas adecuadas para el modelado de procesos biológicos. Se adoptó por guardar esta información en bases de datos de tipo biológicas, éstas se han convertido en un elemento importante para los científicos, pues les facilitan la comprensión de una serie de fenómenos. Constituyen una biblioteca de información sobre ciencias de la vida que ha sido acumulada a partir del análisis computacional, literatura publicada y experimentos científicos. La información contenida en bases de datos biológicas incluye funciones,

estructura y localización (tanto celular como cromosómica) de genes, efectos clínicos de mutaciones, así como similitudes de secuencias y estructuras biológicas. De éstas se obtiene la mayor parte de la información utilizada para los análisis bioinformáticos.

En el término genérico modelado o modelización molecular se incluyen todas aquellas técnicas computacionales que tienen como finalidad el estudio tridimensional estructural de las moléculas. Los programas de modelado molecular permiten simular su estructura y calcular su estabilidad (absoluta o relativa) de los distintos conformeros posibles para las mismas. Además, dependiendo del tipo de programa utilizado, se pueden calcular diversas propiedades físico-químicas, entre ellas el momento dipolar global de las moléculas, el calor de formación, el pKa (fuerza que tienen las moléculas de disociarse), el coeficiente de reparto, así como propiedades directamente relacionadas con la estructura tridimensional, como la densidad electrónica en diversos puntos del espacio.[2]

El modelado molecular es uno de los conceptos más extendidos en química y biología computacional, se emplea para identificar nuevos candidatos a fármacos a través de la creación de modelos moleculares en tres dimensiones.

Como resultado del modelado molecular se pueden calcular diferentes propiedades físico-químicas que describen la estructura de los compuestos, las cuales son importantes en el descubrimiento de nuevos medicamentos puesto que son empleadas en los estudios QSAR (Quantitative Structure Activity Relationship) esenciales para agilizar el proceso de hallazgo de un fármaco novedoso.

1.2 Estudios QSAR

Los estudios QSAR establecen una relación cuantitativa objetiva entre la actividad biológica y estructura química de determinados compuestos. Requieren disponer de datos biológicos de calidad, definir descriptores químicos relevantes y elegir un modelo adecuado para investigar las relaciones.

Los datos biológicos y físico-químicos pueden analizarse estadísticamente de dos formas:

- ✓ En forma conjunta o separada con métodos clasificatorios de reconocimiento de tendencias
- ✓ Usando técnicas de regresión para establecer correlaciones cuantitativas.

Los datos estructurales pueden tomarse a distintos niveles, desde propiedades intrínsecas de la molécula no relacionadas a las interacciones hasta descripciones estéricas-electrónicas como el potencial electrostático mapeado sobre la superficie de Van der Waals.

Las propiedades utilizadas pueden ser experimentales, calculadas o una combinación de ambas. Se ha demostrado que la actividad biológica de un conjunto de compuestos que actúan con el mismo mecanismo de acción puede modelarse matemáticamente a partir de parámetros estructurales simples. La aplicación de estudios QSAR reduce tiempo y costos del proceso de diseño de fármacos.

Etapas de QSAR

- ✓ Planeamiento de los objetivos
- ✓ Determinación de la actividad biológica
- ✓ Descripción de los compuestos
- ✓ Análisis de los objetivos
- ✓ Interpretación de la relación obtenida
- ✓ Predicción de las actividades.

El método QSAR utiliza técnicas novedosas, ya que se vale de programas computacionales para su ejecución, entre estas técnicas se encuentra el QSAR tradicional, QSAR por redes neuronales y 3D-QSAR.

1.2.1 Técnicas de los estudios QSAR

En la actualidad existen diferentes técnicas por las que se puede desarrollar un estudio QSAR y se pueden clasificar en:

QSAR tradicional

El método tradicional incluye el tratamiento estadístico de los datos por métodos multivariados, que incluyen el análisis de regresión, análisis clúster y análisis por componentes principales, entre otras técnicas estadísticas. En ellos se valora la actividad biológica como variable dependiente del conjunto de descriptores moleculares que constituyen las variables independientes. Como resultado de esto se obtienen modelos que describen la actividad biológica como una determinada función matemática de los descriptores moleculares, bien sean estos estructurales, químico-físicos, o ambos.[4]

QSAR por redes de neuronas

Los avances en neurofisiología, y las nuevas técnicas experimentales como la electroencefalografía, la tomografía axial computarizada, el tratamiento de imágenes por RMN, la tomografía de positrones o la de fotones, han incrementado la comprensión de la anatomía del cerebro y de los procesos químico-físicos que ocurren en él. A partir de este conocimiento, se diseñaron modelos matemáticos de adquisición de conocimientos del cerebro humano llamados Redes de Neuronas Artificiales (RNA), o simplemente, Redes de Neuronas (RN). Los modelos de redes neuronales se basan en la interacción colectiva y en paralelo de una gran cantidad de procesadores simples que simulan la conducta de las neuronas biológicas.[4]

QSAR tridimensional (3D-QSAR)

El QSAR tridimensional es una técnica moderna que consiste en el ordenamiento tridimensional de varios elementos que otorgan características determinadas a la molécula de manera tal que se pueden relacionar objetivamente con su actividad biológica.

Ventajas de QSAR-3D

- ✓ El mismo protocolo puede ser aplicado a distintos problemas
- ✓ El método puede manejar grupos de compuestos estructuralmente distintos y predecir nuevos compuestos con características que no existían en el conjunto de calibración

- ✓ Los mapas de contorno tienen relación directa con las interacciones importantes droga-receptor
- ✓ Si se conoce la estructura del receptor y las estructuras se alinean ubicándolas en el sitio activo, los mapas de contorno resultantes coincidirán con los aminoácidos del receptor importantes para la unión y proveerán información sobre cuáles son las interacciones principales.

Las relaciones estructura química - actividad biológica de los compuestos pueden ser cuantificadas en los estudios QSAR a partir del cálculo de los descriptores o índices matemáticos.

1.3 Descriptores

Se conocen como descriptores, los cuantificadores matemáticos que relacionan la estructura molecular de los compuestos a partir de parámetros estructurales simples, lo que posibilita interpretar las propiedades moleculares y describir el comportamiento de compuestos orgánicos, la calidad de los mismos condiciona la calidad de los modelos matemáticos que describan los fenómenos biológicos.

En el campo de los descriptores se conjugan diferentes disciplinas como el álgebra, la teoría de gráficas o de grafos, la teoría de la información, la química computacional, las teorías de la reactividad química y la química-física. Además, la programación, el sofisticado software y hardware son importantes herramientas.[4]

Dentro de un amplio grupo de descriptores se encuentran los constitucionales, éstos son derivados simplemente de la fórmula química, por ejemplo el peso molecular o el número de átomos de nitrógeno de una estructura; los geométricos que tienen en cuenta el análisis de superficie, cálculo de ángulos y distancia entre grupos, los lipofílicos que miden la tendencia de un compuesto X determinado, a formar enlaces hidrofóbicos; los electrónicos que tienen en cuenta la carga eléctrica; los estéricos se obtienen a partir de la forma del sustituyente y el volumen molar; los topológicos que indican una caracterización matemática de una molécula, donde los sitios ocupados por átomos son reemplazados por los "vértices" y las conexiones entre ellos por "aristas" conformando el grafo químico, en este tipo de descriptores existe cierta pérdida de información pues se representa un objeto tridimensional por un número

simple y los topográficos que son semejantes a los topológicos pero tienen en cuenta la estructura en tres dimensiones del compuesto; los híbridos en los que se combinan aspectos estructurales de los compuestos con propiedades físico-químicas o biológicas. Estos descriptores se dividen en dos grandes clasificaciones, atómicos y moleculares en dependencia de la estructura a la cual son aplicados.

1.4 Índice del Estado Refractotopológico para Átomos

El Índice del Estado Refractotopológico para Átomos (R-State) es un descriptor atómico, topológico e híbrido definido por Carrasco y Padrón en el año 2004. Este índice depende de un valor intrínseco del átomo que es una función de la distancia topológica a los átomos de su entorno.

Se desarrolla a partir de la teoría de grafos y la Partición de la Refractividad Molar definida por Ghose y Crippen en 1986. El índice se basa en la influencia de las fuerzas de dispersión de cada átomo sobre cada uno de los restantes en la molécula, modificado por la topología molecular.

El **R-State** (\mathfrak{R}_i), para un átomo i se define por la ecuación 1.1:

$$\mathfrak{R}_i = AR_i + \Delta AR_i$$

Donde AR_i es el valor de refractividad intrínseco del átomo i y ΔAR_i es un término de perturbación (aquellas fuerzas presentes en los átomos que modifican el estado de los demás átomos de la molécula), definida por la ecuación 1.2:

$$\Delta AR_i = \sum \frac{(AR_i - AR_j)}{R_{ij}^2}$$

Donde se suman todos los vértices j adyacentes en el grafo, AR_i y AR_j son los valores intrínsecos de la refractividad de los átomos i y j , respectivamente, y R_{ij}^2 es el número de átomos del camino más corto entre los átomos i y j , incluyendo tanto a i como a j .

1.4.1 Características del Índice del Estado Refractotopológico para Átomos

A diferencia de otros índices topológicos, los cuales no consideran la influencia de los átomos de hidrógeno, este índice sí incluye sus contribuciones al valor intrínseco de los átomos pesados. De acuerdo con la naturaleza del fenómeno estudiado, esta inclusión refleja la capacidad potencial de interacción total del grupo con una supuesta proteína ligando y esto es importante no solamente para el valor intrínseco del átomo al cual está directamente enlazado sino para los otros grupos presentes en la molécula. [1]

El índice presenta una variación de su valor en la medida que el átomo cambia de posición. Se considera que en la medida que el átomo se encuentre embebido en la molécula tendrá menos posibilidades de interactuar con un agente externo.

1.4.2 Aportes del índice a los estudios QSAR

Teniendo en cuenta la naturaleza híbrida del índice, es de considerar que su principal aplicabilidad esté en el campo de las relaciones entre la estructura química y la actividad biológica. [1]

1.5 Conclusiones

En este capítulo se expuso de forma general los métodos que relacionan la estructura química con la actividad biológica, QSAR; se abordó el tema de los descriptores haciendo énfasis en el Índice del Estado Refractotopológico para Átomos que constituye el punto de partida para la confección de este trabajo.

CAPÍTULO 2

PROGRAMAS Y METODOLOGÍAS

En este capítulo se fundamentan las tendencias, técnicas y tecnologías de software usadas para la solución del problema planteado. También se presentan los procedimientos empleados para el desarrollo de este trabajo.

Para la selección de las tecnologías a utilizar, se tuvo en cuenta la tendencia actual y las novedades de cada una de ellas. Esta selección se realizó de acuerdo a las necesidades y a los recursos que se dispone.

2.1 Lenguaje de Programación: Java

Para seleccionar un lenguaje de programación adecuado se debe tener en cuenta varios aspectos, dentro de ellos, conocer que un lenguaje de programación es aquel elemento dentro de la informática que permite crear programas mediante un conjunto de instrucciones, operadores y reglas de sintaxis; que pone a disposición del programador para que éste pueda comunicarse con los dispositivos hardware y software existentes.

Es la notación para la descripción precisa de algoritmos o programas informáticos. Lo constituye el conjunto de instrucciones que permite al programador pensar claramente sobre la complejidad del problema a resolver, de manera que pueda ordenarlas convenientemente para la creación de un programa ejecutable por la computadora. Se dividen en lenguajes de alto y bajo nivel según se acerquen más o menos a las formas de comunicación humana, respectivamente.

Para el desarrollo de este trabajo se seleccionó Java, lenguaje de programación que según sus características se ha convertido en uno de los más utilizados en el mundo. Java es un lenguaje de programación orientado a objetos con una sintaxis fácilmente accesible y cómoda de desarrollar, elaborado a partir de los lenguajes C y C++, de donde hereda sus características principales, por ejemplo, sus tipos de datos primitivos. En este lenguaje se trata de reducir los errores más comunes que se cometen en la programación. Elimina la aritmética de punteros y referencias, ya no es necesario liberar la memoria usada. Posee un entorno robusto, verifica su código al mismo tiempo que lo escribe, y una vez más antes de ejecutarse; esto trae como consecuencia que se detecte un mayor número de errores en tiempo de compilación. Tiene una gestión avanzada de memoria llamada recolección de basura para los objetos ya inservibles sin una instrucción específica, y un manejo de excepciones orientado a objetos integrados.

El paquete de utilidades de Java viene con un conjunto completo de estructuras de datos complejas y sus métodos asociados, que serán de inestimable ayuda para implementar applets y otras aplicaciones más complejas. Se dispone también de estructuras de datos habituales, como pilas y tablas hash, como clases ya implementadas. [7]

En Java no se permiten los accesos ilegales a memoria. Se puede determinar si el código no produce desbordamientos en la pila, si el acceso a los campos de un objeto es legal dependiendo de la visibilidad. Java fue diseñado para crear software fiable y para ejecutarse independientemente de la plataforma.

Por las características señaladas se seleccionó como el más adecuado para el desarrollo de este trabajo.

2.2 Entorno de desarrollo: Eclipse

De la lista de entornos de desarrollo integrados que se basan en Java se determinó que Eclipse es el que reúne las características necesarias para desarrollar este trabajo, pues es una plataforma de desarrollo de código abierto extensible que permite integrar una amplia gama de herramientas y librerías. Es un desarrollo de IBM cuyo código fuente fue puesto a disposición de los usuarios. [8]

La arquitectura de plug-in permite integrar diversos lenguajes de desarrollo sobre el mismo IDE. Posee un conjunto de servicios para el desarrollo de Java, C, C++, que se conectan a partir de estos componentes antes mencionados e introducir otras aplicaciones.

Conserva el registro de las versiones, genera y mantiene la documentación de cada etapa del proyecto, permite que distintos desarrolladores colaboren en la creación de un proyecto. Permite la reutilización de código y la prueba de errores eficientemente.

2.3 Librerías Utilizadas

Los análisis bioinformáticos involucran una amplia gama de de tareas y procesos. Se han escrito diversos programas para varias aplicaciones bioinformáticas usando cada lenguaje disponible. A causa del tamaño de los conjuntos de datos en Bioinformática, el tiempo computacional no es trivial y es deseable la eficiencia en la velocidad computacional. Se han creado comunidades que crean librerías programadas en Java para facilitar el trabajo con esas estructuras, teniendo BioJava un sitio entre los primeros.

2.3.1 BioJava

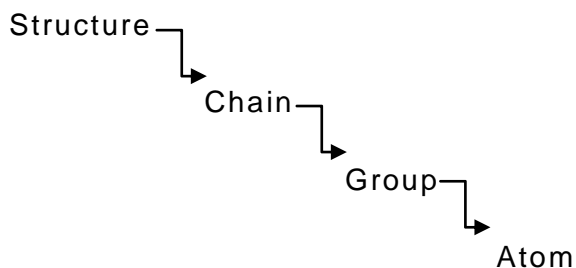
BioJava es un proyecto que se ha utilizado para escribir código que recupere información de las bases de datos biológicas comunes, es una colección de 40 paquetes organizados en 9 categorías principales, cada categoría contiene clases e interfaces creadas para desarrollar tareas específicas como alineamiento de secuencias y manejo de secuencias. Es una librería de código abierto (licencia GNU), es actualizada y mejorada continuamente. Está provista de una colección que incluye objetos, clases y métodos para manipular secuencias biológicas. Es útil para la automatización de proyectos en Bioinformática que se basan en software libre.

BioJava contiene potentes rutinas de análisis estadístico y herramientas para analizar los formatos de archivos estándares, por ejemplo, los archivos **.pdb**¹.

¹ **.pdb**: extensión del fichero PDB, *Protein Data Bank*. Base de Datos de estructuras de proteínas creada en 1971 por Walter Hamilton. Posee cerca de 55000 estructuras (continúa creciendo), las cuales cuentan con la ubicación espacial, determinadas experimentalmente por cristalografía de rayos X y resonancia magnética nuclear (RMN).

Clases de BioJava Utilizadas:

- ✓ **PDBFileReader**: Se utiliza para leer los datos del fichero **.pdb** y los devuelve en una estructura (*Structure*)
- ✓ **Structure**: Permite el acceso a la información de la proteína contenida en el archivo **.pdb**. La clase *Structure* provee acceso a los datos siguiendo la siguiente jerarquía:



- ✓ **Aminoacid**: Contiene la información de los aminoácidos de la proteína.
- ✓ **Atom**: Es la clase interfaz que posee la información de los átomos de cada aminoácido.
- ✓ **AtomImpl**: Es la implementación de la clase Atom.
- ✓ **Calc**: Esta clase posee herramientas para realizar operaciones sobre átomos, aminoácidos, etc.
- ✓ **Chain**: Es una clase interfaz que corresponde a cada cadena en el fichero **.pdb**, una cadena consiste en una lista de grupos (*Group*)
- ✓ **Group**: Es una estructura para almacenar cada grupo de átomos dentro de una secuencia de la proteína. Existen tres tipos de grupos: aminoácidos, heteroátomos y nucleótidos.

2.3.2 JFreeChart

JFreeChart es una biblioteca de clases de Java para la generación de gráficas que se distribuye bajo los términos de la licencia GNU/LGPL (Lesser General Public License) que permite su uso en software propietario. Se incluye soporte para: gráficas de pie,

gráficas de barra, gráficas lineales, gráficas de intervalos de tiempo, planillas métricas, entre otras. Se utiliza en aplicaciones, programas internos de servidores, Java Server Page (JSP) o programas internos. Se puede mostrar salidas en diversas formas, ejemplo archivos de imágenes PNG, JPEG, o los formatos de archivo (incluyendo PDF, EPS y SVG).

Clases Utilizadas:

- ✓ **JFreeChart:** Clase encargada de coordinar todo el proceso de dibujo de las gráficas. Puede trabajar con grandes cantidades de datos creando gráficos diferentes
- ✓ **ChartFactory:** Provee varios métodos que son convenientes a la hora de crear los gráficos
- ✓ **PlotOrientation:** Se encarga de dar la orientación del gráfico ya sea horizontal o vertical
- ✓ **XYDataset:** Interfaz que define una colección de datos de la forma (x, y) teniendo como característica que cada serie de datos es independiente a la otra
- ✓ **XYSeries:** Estructura de datos temporal que asocia los valores numéricos
- ✓ **XYSeriesCollection:** Colección de objetos *XYSeries* que implementa la interfaz *XYDataset*.

2.3.3 JCommon

JCommon es una colección de clases de gran utilidad, usadas por **JFreeChart**. Esta librería incorpora una gran cantidad de clases de apoyo para la creación de gráficas que pueden ser muy ventajosas a la hora de realizar los trabajos con las mismas. **JCommon** se distribuye bajo los términos de la licencia LGPL.

2.4 Jmol

Jmol es un visualizador de Java de código abierto para estructuras químicas en tres dimensiones. Es una miniaplicación (*applet*), componente para la integración en

sistemas, programa autónomo que se ejecuta localmente. Es multiplataforma, compatible con sistemas operativos Windows, Mac OS X y Linux/Unix, así como con programas de visualización molecular como Rasmol y Chime. Realiza una representación gráfica tridimensional de alto rendimiento sin requisitos de hardware, sólo requiere de la instalación de la Máquina Virtual de Java.

Se destaca por ofrecer numerosas funcionalidades nuevas en la representación y análisis de estructuras, reconocer numerosos formatos moleculares. Puede leer de archivos con formato CIF/mmCIF², MOPAC³, PDB, entre otros. Se pueden realizar animaciones, formas esquemáticas para las estructuras secundarias de biomoléculas, también se pueden ejecutar mediciones en cuanto a distancia, ángulo y ángulo de torsión. Exporta los resultados procesados a **.jpg**, **.png**, **.ppm**, **.pdf** y **PovRay**. Puede ser utilizado como librería para incluirlo en otras aplicaciones.

2.5 Estructuras de datos utilizadas

2.5.1 Lista

Estructura de datos similar a los arreglos. La diferencia es que el orden de los elementos enlazados puede ser diferente al orden de almacenamiento en la memoria o el disco, permitiendo que el orden de recorrido de la lista sea diferente al de almacenamiento. Existen diferentes tipos de listas, Lista Simplemente Enlazada (SE), Listas Doblemente Enlazadas (DE), Listas Circulares (LC) y Listas Circulares Doblemente Enlazadas (LCDE). Las listas enlazadas contienen en cada elemento una referencia a otro elemento del mismo tipo, de ahí que resulte fácil el proceso de adición, inserción, y eliminación de uno o varios elementos.

2.5.2 Tablas Hash

Una tabla hash es una estructura de datos que generaliza lo que se conoce por Arreglo, pero en este caso se usan en tareas de asociación de claves para realizar la búsqueda utilizando estos mismos datos que se mantienen almacenados.

² **CIF/mmCIF**: (Crystallographic Information File/Macromolecular Crystallographic Information File) formato estándar de la Unión Internacional de Cristalografía.

³ **MOPAC**: (Molecular Orbital PACkage) es la salida de un programa de Química Cuántica, desarrollado en el grupo de J.J.P. Stewart.

Las tablas hash se suelen implementar con arreglos de una dimensión, aunque se pueden hacer implementaciones multidimensionales basadas en varias claves. Las tablas hash tienen como ventaja que sin importar el número de elementos que se tenga almacenado en la tabla, la complejidad temporal de la búsqueda es $O(1)$ llegando a $O(n)$ en casos extremos.

2.6 Programa Estadístico

En términos generales, se puede decir que la estadística se ocupa de describir, presentar y analizar datos con el objetivo de ayudar a comprenderlos.

SPSS

Es un programa que está diseñado para facilitar el análisis estadístico de datos. Se puede usar en múltiples plataformas Linux, Mac OS y Windows. Entre sus mayores virtudes está la de interactuar con bases de datos relativamente grandes, graficar y realizar un análisis estadístico a partir de los datos de salida que muestra. Tiene la capacidad de representar de forma efectiva los resultados en tablas y gráficas de alta variedad y calidad, por ejemplo los gráficos de barra y de dispersión, útiles para comparar resultados obtenidos.

2.7 Swiss PDB Viewer (SPDBV)

Fue creado por Nicolas Guex en el año 1994. Es una aplicación que proporciona una interfaz de fácil uso en la que se puede señalar el o los aminoácidos que se desean, exponer sus datos, mostrar el esqueleto carbonado de la proteína, entre otras funcionalidades. Permite analizar varias proteínas al mismo tiempo, pueden ser superpuestas, a fin de deducir alineaciones estructurales y comparar sus sitios activos. Mutaciones de aminoácidos, ángulos y distancias entre los átomos son fáciles de obtener gracias a la intuitiva interfaz gráfica y las funcionalidades del programa. Además posee diversas herramientas de modelación e integra archivos de comandos.

2.8 Procedimientos Utilizados

Para calcular el Índice del Estado Refractotopográfico para Aminoácidos, se decidió primeramente ver su ubicación espacial debido a que son estructuras tridimensionales; como las proteínas están compuestas por aminoácidos y éstos a su vez por varios átomos los cuales presentan distintas ubicaciones espaciales fue necesario concentrar la masa de todos los átomos contenidos de cada aminoácido en un punto, para definirlo así se aplica la teoría del centro de masa.

2.8.1 Centro de masa

Dado un sistema constituido por un conjunto de partículas puntuales de masas $m_1, m_2, m_3, \dots, m_N$, el centro de masa (CM) del sistema es un punto abstracto cuyo vector de posición se calcula por medio de la ecuación 2.1:

$$\vec{R} \equiv \frac{1}{M} \sum_{i=0}^N m_i \vec{r}_i,$$

Donde 2.2:

$$M = \sum_{i=0}^N m_i$$

M es la masa total del sistema.

Desde el punto de vista matemático es evidente que el vector \vec{R} representa un promedio pesado de las posiciones ocupadas por todas las partículas que forman al sistema. Con respecto a la definición del centro de masa hay dos observaciones que deberían resultar obvias, en primer lugar, que la posición del CM no tiene que coincidir con la posición de alguna de las partículas del sistema y en segundo, que debido a que encontrar \vec{R} involucra el cálculo de un promedio ponderado en el que los pesos de los objetos a promediar (las posiciones de cada partícula del sistema) están dados por las masas de las partículas, el CM tiende a acercarse a aquellas regiones en que se concentre más masa. [9]

Aquí cada partícula es un átomo del aminoácido en cuestión, de los cuales se buscaron sus coordenadas del PDB. La masa atómica de cada uno de los átomos que se encuentran en las proteínas se extrajo de la tabla periódica de los elementos. [Tabla 2.1]

Tabla 2.1 Masa Atómica de los elementos que componen los aminoácidos.

No.	Elemento	Masa Atómica
1	O	15,999
2	C	12,011
3	N	14,006
4	S	32,064
5	F	18,998
6	Cl	35,453
7	Br	79,909
8	I	126,904
9	P	30,973

El vector \vec{R} que se obtiene da como resultado las coordenadas del aminoácido que se desea conocer su posición.

Esta teoría se ha aplicado pues se conocen los datos suficientes (las coordenadas de posición extraídas del fichero **.pdb**) para ver a la proteína en cuestión como un sólo punto con coordenadas (x, y, z) y la masa atómica de cada uno de los elementos que componen el aminoácido que se está analizando.

2.8.2 Redefinición del Descriptor Atómico Índice del Estado Refractotopológico (R-State, \mathcal{R}) para aplicarlo a proteínas

Se toma como base el Índice del Estado Refractotopológico para Átomos antes explicado [epígrafe 1.4] pero esta vez para aplicarlo a proteínas se aplica la fórmula 1.1 teniendo como datos las coordenadas del centro de masa [epígrafe 2.8.1] y utilizándolo para cada aminoácido.

El Índice del Estado Refractotopográfico para un aminoácido i se define por la ecuación 2.3:

$$\mathcal{R}_i = AR_i + \Delta AR_i$$

Donde AR_i es el valor de refractividad intrínseco del aminoácido i , estos valores fueron calculados utilizando el plug-in Cálculo de Descriptores de la “Plataforma para la Predicción de Actividad Biológica en Compuestos Orgánicos”, quedando en la tabla 2.2 y ΔAR_i es un término de perturbación definido por la ecuación 2.4:

$$\Delta AR_i = \sum_{i=1}^n \frac{AR_i - AR_j}{R_{ij}}$$

Los términos AR_i y AR_j son los valores intrínsecos de la refractividad de los aminoácidos i y j , respectivamente, y R_{ij} es la distancia topográfica entre i y j . Dada por la fórmula 2.5:

$$R_{ij} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Donde x_1, y_1, z_1 son las coordenadas del centro de masa de cada aminoácido i y x_2, y_2, z_2 son las coordenadas del centro de masa de cada aminoácido j .

Tabla 2.2 Valor de refractividad intrínseco de cada aminoácido.

No.	Aminoácido	Valor Intrínseco
1	Alanina(ALA)	21.5695
2	Glicina (GLY)	16.9227
3	Asparagina (ASN)	28.0889
4	Glutamina (GLN)	35.0261
5	Arginina (ARG)	45.9192
6	Ácido aspártico (ASP)	28.0889
7	Aspartato (ASX)	55.0021
8	Cisteína (CYS)	28.8885
9	Ácido Glutámico (GLU)	32.7357
10	Glutamato (GLX)	64.2957

11	Histidina (HIS)	41.0948
12	Isoleucina (ILE)	35.5099
13	Leucina (LEU)	35.5099
14	Lisina (LYS)	39.4354
15	Metionina (MET)	38.1821
16	Fenilalanina (PHE)	48.1983
17	Prolina (PRO)	29.0321
18	Serina (SER)	23.2046
19	Taurina. (THE)	27.8514
20	Triptófano (TRP)	59.7794
21	Tirosina (TYR)	49.8334
22	Valina (VAL)	30.8631

2.8.3 Gráficas

Las gráficas son elementos empleados para visualizar el comportamiento de valores numéricos que facilitan la comprensión del significado de éstos.

Gráficas de Dispersión

Muestran una serie como un conjunto de puntos, los valores se representan mediante la posición de los puntos en el gráfico, en el eje de las x se representa una variable independiente y en el eje de las y una dependiente; así, si las variables están correlacionadas, el gráfico muestra algún tipo de tendencia. Este tipo de gráficos suele usarse para comparar datos y son ideales para controlar la distribución de los valores, es el tipo de gráfico más recomendable si se tiene gran cantidad de puntos. Se usan normalmente para mostrar y comparar valores numéricos, como datos científicos, estadísticos y de ingeniería.

Gráficas de líneas

Muestran una serie como un conjunto de puntos conectados entre sí mediante una línea, manifiestan la relación existente entre dos variables. Cuando no se tienen gran cantidad de datos ofrecen más claridad que las gráficas de dispersión.

2.8.4 Coeficiente de Correlación

El coeficiente de correlación es la medida descriptiva de la intensidad de la asociación lineal entre dos variables x y y . Los valores de coeficiente de correlación siempre están entre -1 y +1. Un valor de +1 indica que las dos variables x y y , tienen una relación lineal positiva perfecta. Esto es, todos los puntos están dados en una línea recta con pendiente positiva. Un valor de -1 indica que x y y tienen una relación lineal negativa perfecta, y que todos los puntos de datos están en una recta con pendiente negativa. Los valores cercanos a cero indican que x y y no tienen relación lineal.[10]

2.9 Conclusiones

En este capítulo se mencionaron las herramientas y los métodos utilizados: Java como lenguaje de programación y Eclipse como entorno de desarrollo, las librerías utilizadas tales como BioJava, JFreeChart, JCommon. Se expuso de forma general, la teoría utilizada para la realización de este trabajo. Se describieron los algoritmos basados en el cálculo de descriptores atómicos que se le han aplicado a proteínas.

CAPÍTULO 3

RESULTADOS Y DISCUSIÓN

En este capítulo se exponen los experimentos realizados y el análisis de los resultados obtenidos como parte de la investigación realizada.

3.1 Herramienta para realizar el cálculo del descriptor

Debido a la gran cantidad de datos que contienen los ficheros de cada proteína y con el objetivo de agilizar eficientemente el tiempo de cálculo, se implementó el algoritmo del Índice del Estado Refractotopográfico para Aminoácidos. Esta aplicación permite graficar los resultados, realizar análisis estadístico, recalculer el índice teniendo en cuenta una cota de distancia definida por el usuario y guardar en un fichero información de interés para un posterior estudio.

3.1.1 Descripción de la Aplicación

Luego de ejecutar la herramienta, se accede al menú **Archivo** opción **Abrir** y se selecciona el fichero **.pdb** de la proteína a analizar, la herramienta visualiza la proteína empleando *Jmol* y muestra algunos datos de la misma, [Figura 3.1].

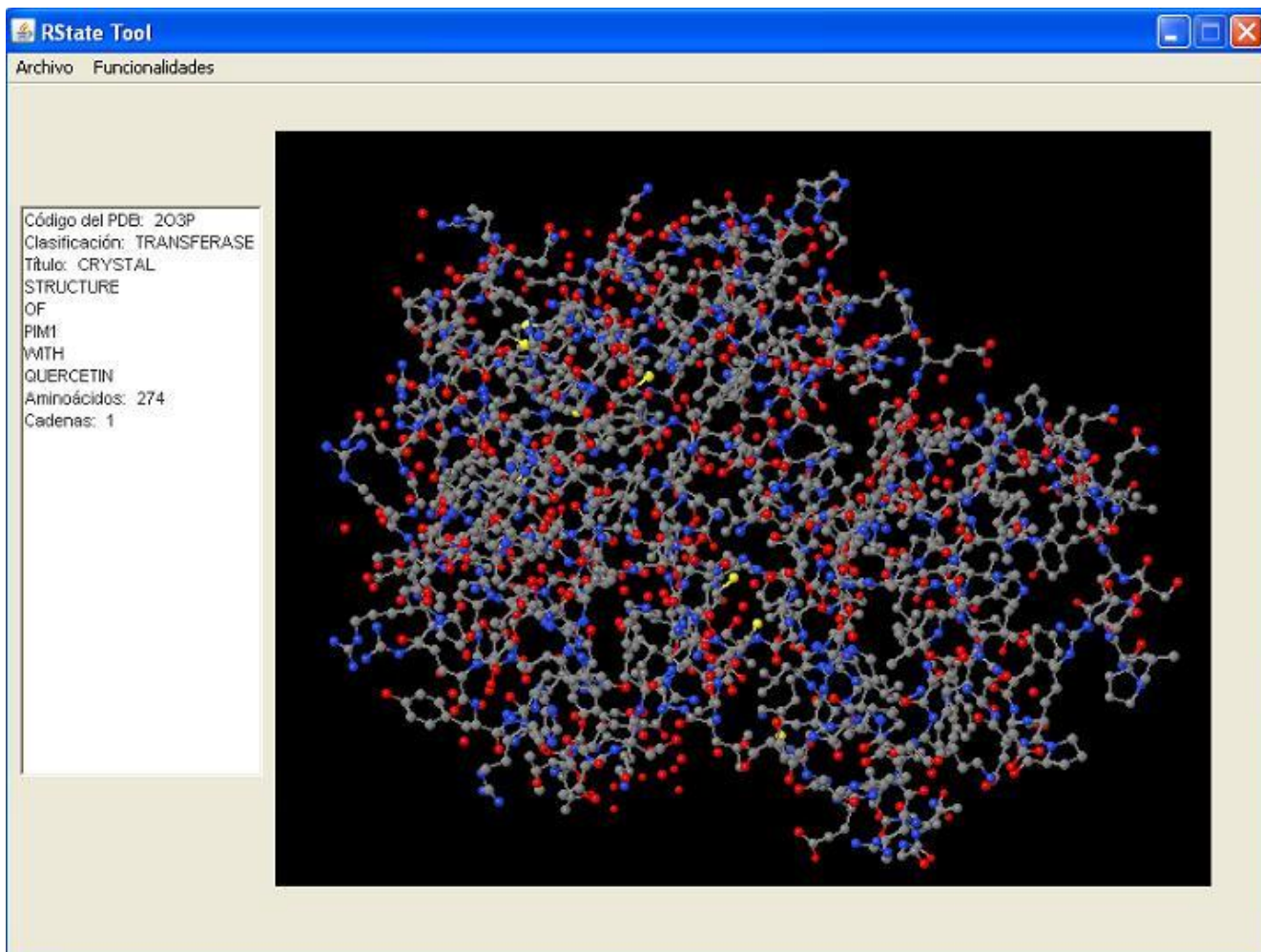


Figura 3.1 Ventana Principal de la Herramienta.

Al acceder al menú **Funcionalidades** en la opción **Analizar Proteína**, se muestra una ventana compuesta por tres pestañas: **Cálculos**, **Comportamiento** y **Acotar RState**. En la pestaña **Cálculos** se muestra el Índice del Estado Refractotopográfico para Aminoácidos calculado, [Figura 3.2]. Consta de un botón que permite presentar el centro de masa para cada aminoácido de la proteína.

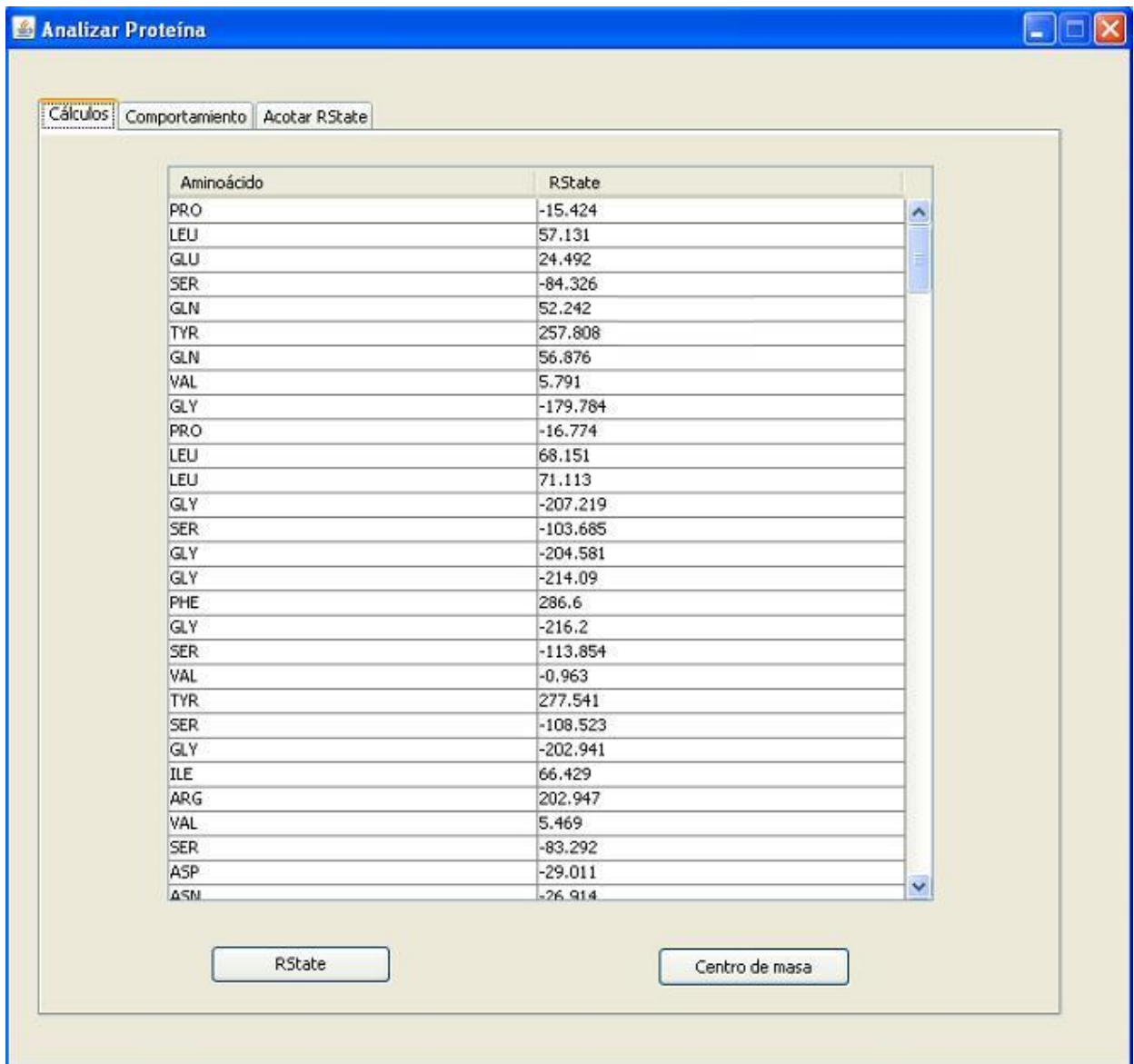


Figura 3.2 Ventana que muestra el Índice del Estado Refractotopográfico para Aminoácidos.

En la pestaña **Comportamiento** se brinda la opción de escoger un aminoácido y exhibir, mediante el botón **Ver**, la relación distancia-perturbación del aminoácido seleccionado con el resto [Figura 3.3]; también es posible graficar cómo se comporta dicha relación respecto a: el resto de los aminoácidos, los aminoácidos apolares, polares y un gráfico de dispersión.

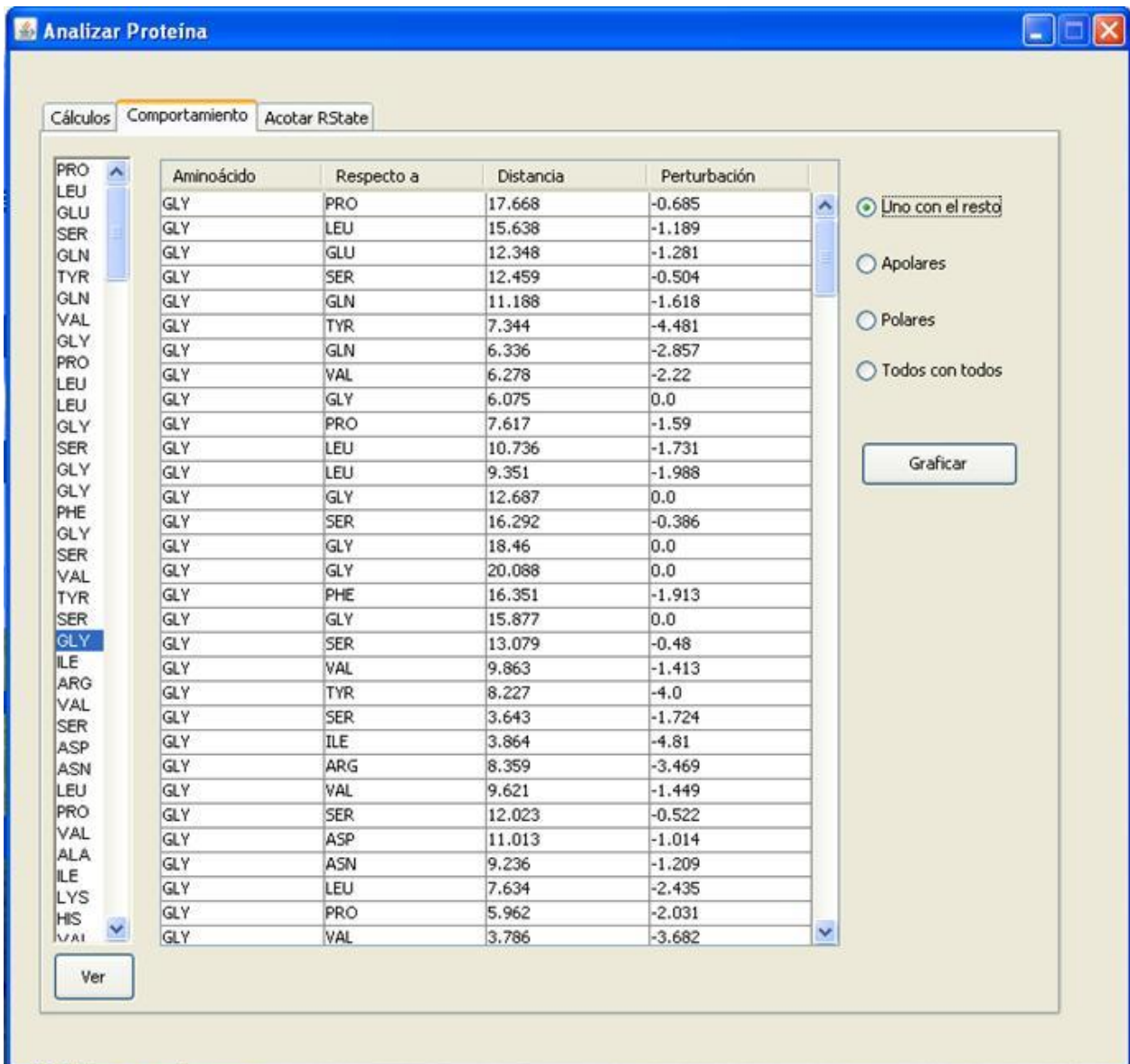


Figura 3.3 Comportamiento distancia-perturbación entre los aminoácidos.

La perturbación ejercida por el resto de los aminoácidos sobre uno en específico disminuye en la medida en que aumenta la distancia, o sea, que los aminoácidos más alejados no ejercen un efecto significativo para dicho aminoácido. De aquí se intuye que sería posible despreciar en los cálculos aquellos aminoácidos más distantes.

En la pestaña **Acotar RState**, la herramienta permite acotar el índice de acuerdo a una distancia establecida por el usuario. El botón **Calcular** recalcula el índice para la nueva distancia establecida y muestra el Coeficiente de Correlación entre los valores

obtenidos en esta variante y el índice calculado sin despreciar ningún aminoácido. [Figura 3.4]

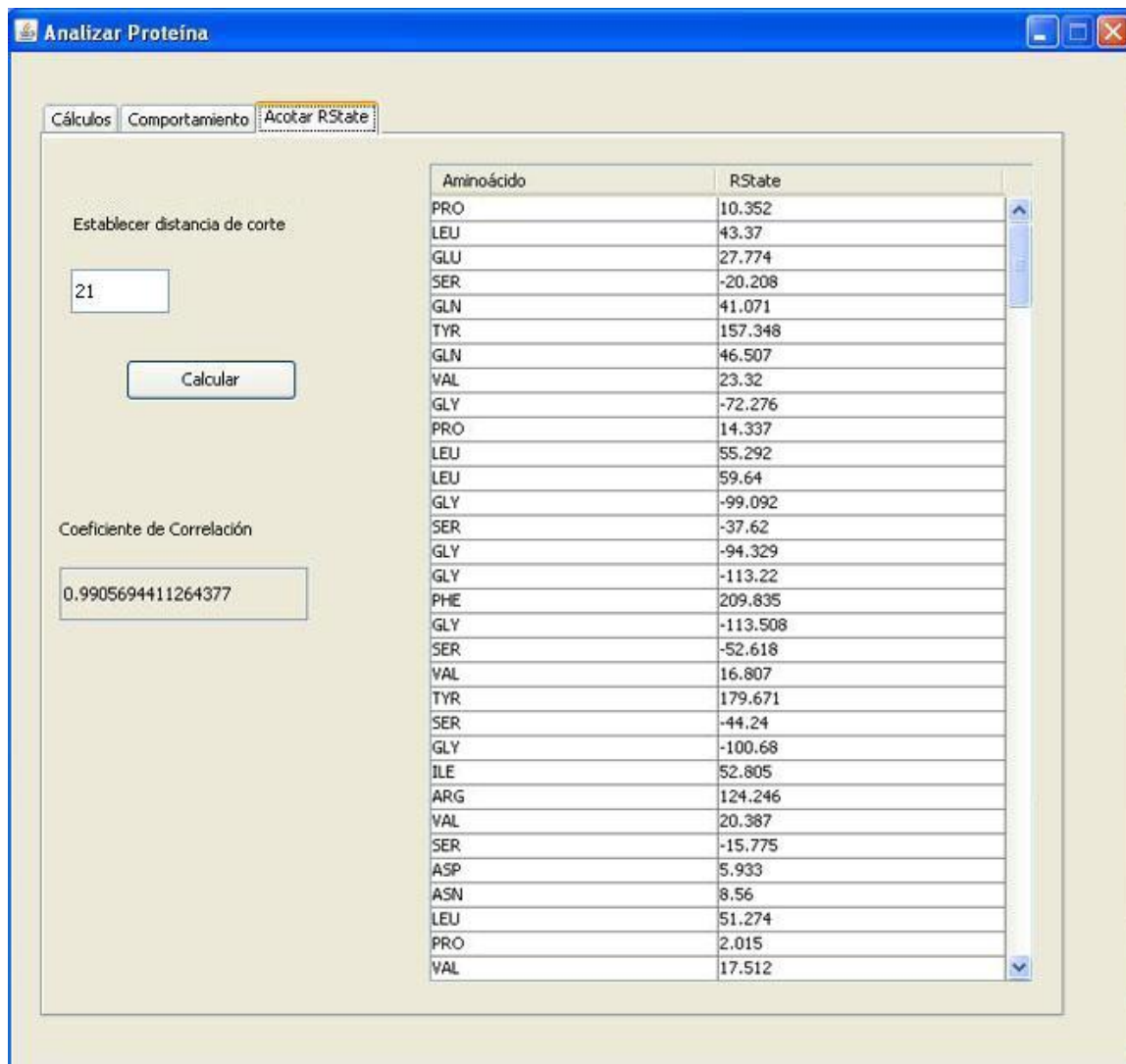


Figura 3.4 Índice del Estado Refractotopográfico para Aminoácidos con distancia establecida por el usuario.

3.2 Principales Cálculos Implementados

De la Base de Datos de Proteínas se escogieron los archivos en formato **.pdb**, estos archivos contienen información tales como: las cadenas por la cuales está formada, las coordenadas de los átomos que la componen en tres dimensiones, la secuencia de los aminoácidos, entre otras. La lectura de estos ficheros se realizó para obtener

los datos requeridos por el descriptor y estructurarlos en una Lista. Esta lectura se realizó a través de la clase *PDBFileReader* de *BioJava*.

3.2.1 Cálculo del Centro de Masa

Teniendo en cuenta la definición de Centro de Masa expuesta en el capítulo anterior, se tomaron de la Lista las coordenadas de cada átomo en todos los aminoácidos y con esto se calculó el centro de masa para cada aminoácido en la proteína.

¿Cómo se implementó?

Para implementar el algoritmo del cálculo del centro de masa, de cada aminoácido se obtuvo su Lista de átomos y de ésta las coordenadas de cada átomo. Se suma la multiplicación del valor de cada coordenada por la masa del átomo correspondiente, finalmente cada resultado se divide entre la sumatoria de la masa de todos los átomos. Esto retorna como resultado un vector con las coordenadas x, y, z del aminoácido, o sea, su centro de masa.

Descripción general de las funciones utilizadas

- ✓ **length:** Devuelve el tamaño del arreglo.
- ✓ **size:** Devuelve la longitud de la lista.
- ✓ **get:** Recibe una posición y devuelve el elemento que está en esa posición.
- ✓ **add:** Recibe un valor y lo adiciona al final de una lista.
- ✓ **getAtoms:** Dado un aminoácido devuelve el arreglo de átomos.
- ✓ **getPDBName:** Devuelve el nombre del aminoácido
- ✓ **getPDBCode:** Devuelve el número que identifica al aminoácido en la proteína.
- ✓ **getCentroide:** Devuelve el vector que contiene las coordenadas del centro de masa de un aminoácido.

3.2.2 Cálculo del Índice Refractotopográfico

¿Cómo se implementó?

Una vez computado el centro de masa de cada aminoácido, se calculó la distancia de cada uno respecto al resto. A cada distancia obtenida se le sumó el efecto perturbativo [ecuación 2.4 y 2.5] que generan los demás aminoácidos sobre el aminoácido en cuestión.

Todo este proceso se realiza para cada aminoácido de la proteína. El resultado final del algoritmo es una colección de valores que corresponden con el número de aminoácidos que contiene la proteína. Así, si se tienen 198 aminoácidos en la proteína, entonces se obtienen 198 valores del índice.

Descripción general de las funciones utilizadas

- ✓ **size**: Devuelve la longitud de la lista
- ✓ **get**: Recibe una posición y devuelve el elemento que está en esa posición
- ✓ **add**: Recibe un valor y lo adiciona al final de una lista
- ✓ **getAminoType**: Devuelve el nombre del aminoácido en un carácter
- ✓ **getDistance**: Calcula dadas dos coordenadas calcula la distancia
- ✓ **getPDBName**: Devuelve el identificador del aminoácido.

3.3 Fichero de Salida

La información obtenida luego de procesar los datos extraídos del fichero **.pdb** por la aplicación se almacenan en un fichero con extensión **.rsp** nombre derivado de Índice del Estado Refractotopográfico para Aminoácidos en la proteína.

Características del nuevo fichero:

PROTEIN_CODE: Código que identifica la proteína.

PROTEIN_NAME: Nombre de la proteína.

CENTER_MASS: Centro de masa para cada aminoácido dispuesto de la siguiente forma; nombre del aminoácido, número que lo identifica, coordenadas del centro de masa x, y, z, respectivamente.

END: Fin de la sección del centro de masa.

RELATIONSHIP_DISTANCE_PERTURBATION: Los valores correspondientes a la distancia y la perturbación de cada aminoácido respecto al resto dispuestos de la siguiente forma; nombre del aminoácido objetivo, número identificador, nombre del aminoácido que ejerce fuerza perturbativa sobre al aminoácido objetivo, valor de la distancia entre ellos, valor de la perturbación comentada anteriormente.

END: Fin de la sección de la relación distancia perturbación.

RSTATE: Valores del Índice del Estado Refractotopográfico para cada aminoácido, dispuestos de la siguiente forma; nombre del aminoácido, número y valor del índice.

END: Fin de la sección del Índice del Estado Refractotopográfico para Aminoácidos.

DELIMIT_RSTATE: Índice del Estado Refractotopográfico para Aminoácidos delimitado por una distancia de corte.

END: Fin de la sección del Índice del Estado Refractotopográfico para Aminoácidos acotado.

Este fichero puede resultar de gran utilidad para los estudios posteriores del Índice Refractotopográfico pues se pueden escoger los datos y realizar otros análisis que resulten de interés futuro.

3.4 Resultados Experimentales

Para identificar el comportamiento del índice, se tomó una muestra heterogénea de 5 proteínas, clasificadas en globulares y filamentosas:

Globulares:

PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE: Oncogén que ha sido implicado en el desarrollo de leucemias, linfomas y cáncer de próstata, es el blanco de programas de desarrollo de fármacos. Su código en el PDB es 2O3P, posee 274 aminoácidos.

HIV-1 PROTEASE: Proteasa aspártica esencial para el ciclo de vida del VIH, el retrovirus que causa el SIDA. Su código en el PDB es 1RPI, posee 2 cadenas simétricas de 99 aminoácidos cada una.

ACTIN ALPHA 1: Forma los microfilamentos, uno de los tres componentes fundamentales del citoesqueleto de las células de los organismos eucariotas. Su código en el PDB es 1J6Z, posee 368 aminoácidos.

Filamentosas:

COLLAGEN X: Cadena corta de Colágeno expresado específicamente por condrocitos hipertróficos del cartílago de la placa de crecimiento durante la formación del hueso endocondrial. Su código en el PDB es 1GR3, posee 132 aminoácidos.

FIBROBLAST GROWTH FACTOR 7/1 CHIMERA: Derivado de células estromales FGF-7 (del inglés Fibroblast Growth Factor) que se une y activa solamente el receptor FGFR2IIIb en células epiteliales. Su código en el PDB es 1QQL, posee 131 aminoácidos.

Se realizaron 3 experimentos para realizar una caracterización inicial del Índice del Estado Refractotopográfico para Proteínas.

3.4.1 Experimento 1

En la definición del Índice del Estado Refractotopológico [epígrafe 1.4] se describe el efecto que tiene la posición del átomo en el valor del índice, pues dada la naturaleza de las fuerzas de dispersión es lógico suponer que un átomo interactuará menos con la proteína ligando en la misma medida en que esté más sumergido en su esqueleto o tenga un menor volumen estérico. [1]

Este experimento se realizó para comprobar la analogía que pudiera existir en el comportamiento del índice, aplicado ahora a proteínas. De aquí fue necesario observar cómo se comporta el valor del índice de cada aminoácido respecto a la posición que éste ocupa en la proteína. Para ello se calculó el Índice del Estado Refractotopográfico para Proteína a las 5 proteínas seleccionadas utilizando la herramienta implementada en esta tesis.

Del fichero resultante se tomaron los valores del índice y se organizaron ascendentemente [tablas 3.1 y 3.2]. Utilizando el SPDBV se representó cada proteína y se localizaron los aminoácidos con valores más pequeños (utilizando el color rojo) y más grandes (empleando el verde) [figuras 3.5 - 3.10].

Tabla 3.1 Valores del Índice del Estado Refractotopográfico para Proteínas de los 20 aminoácidos que muestran un menor valor del mismo en la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE

No.	Aminoácido	Índice
229	GLY	-269,528
102	GLY	-253,844
190	GLY	-247,103
220	GLY	-243,465
188	GLY	-243,393
203	GLY	-239,024
180	GLY	-230,348
238	GLY	-224,054
162	GLY	-222,507
50	GLY	-216,200
48	GLY	-214,090
251	GLY	-210,751
137	GLY	-209,684
45	GLY	-207,219
47	GLY	-204,581
55	GLY	-202,941
78	GLY	-201,138
99	GLY	-196,653
224	ALA	-180,444
41	GLY	-179,784

Tabla 3.2 Valores del Índice del Estado Refractotopográfico de los 20 aminoácidos que muestran un mayor valor del mismo en la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE

No.	Aminoácido	Índice
281	PHE	262,645
132	PHE	262,737
130	PHE	267,643
100	PHE	268,714
198	TYR	275,118
53	TYR	277,541
218	TYR	279,446
148	PHE	281,038
147	PHE	282,901
49	PHE	286,600
233	TYR	289,769
187	PHE	300,148
207	TYR	314,190
77	TRP	399,907
109	TRP	406,451
269	TRP	407,960
212	TRP	411,454
289	TRP	424,049
149	TRP	441,683
226	TRP	473,317

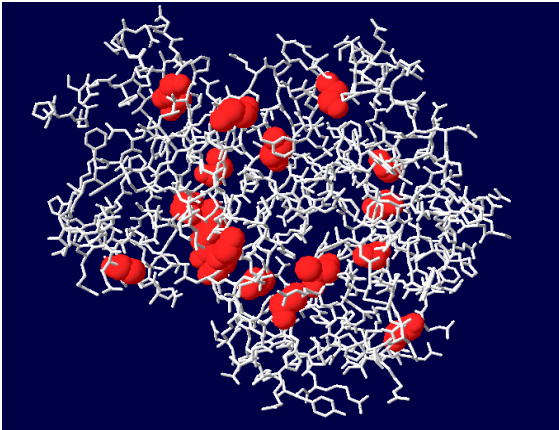


Figura 3.5 Los 20 valores menores en la proteína *PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE*.

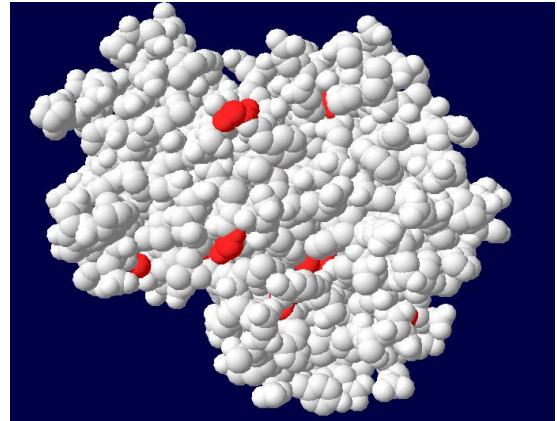


Figura 3.6 Los 20 valores menores en la proteína *PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE* representados sobre la superficie de Van der Waals.

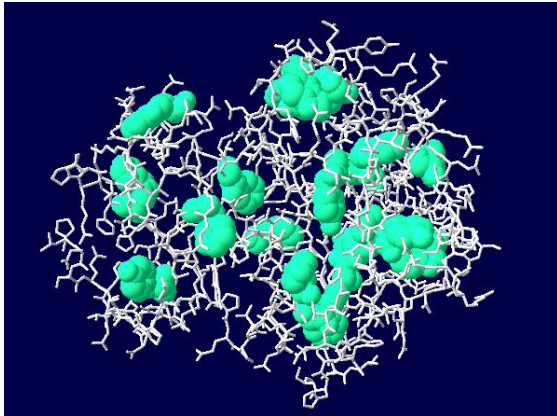


Figura 3.7 Los 20 valores mayores en la proteína *PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE*.

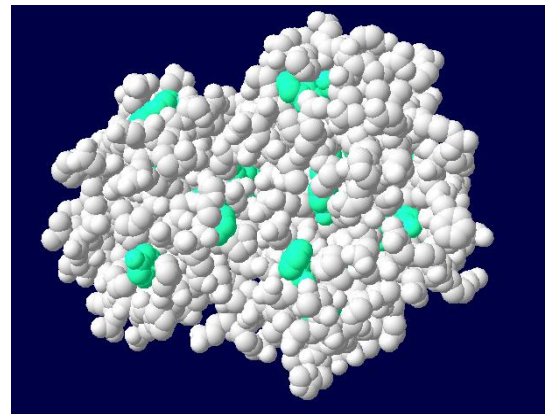


Figura 3.8 Los 20 valores mayores en la proteína *PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE* representados sobre la superficie de Van der Waals.

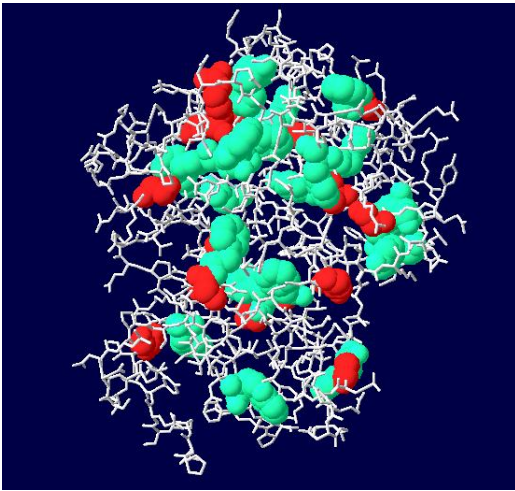


Figura 3.9 Los 20 valores menores y los 20 mayores en la proteína *PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE*

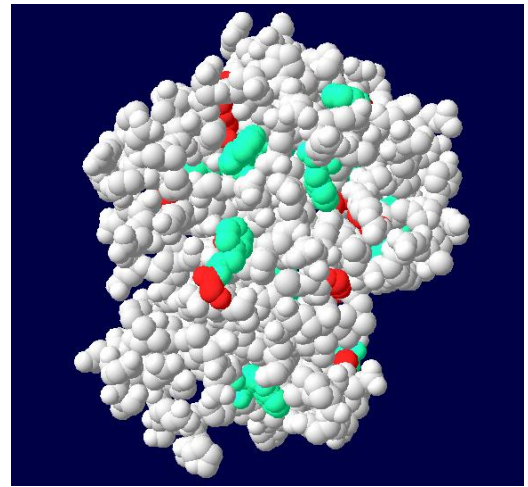


Figura 3.10 Los 20 valores menores y los 20 mayores en la proteína *PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE* representados sobre la superficie de Van der Waals.

Un primer resultado exhibe que, la mayoría de los 20 aminoácidos de menor valor, correspondieron a la Glicina, para los 20 aminoácidos de mayor valor, la mayoría correspondió al Triptófano y Fenilalanina, en determinadas proteínas de las analizadas, algunas poseían pocos aminoácidos de este tipo, entonces aparece entre los valores mayores el aminoácido Tirosina. Este resultado puede estar condicionado por el valor intrínseco de estos aminoácidos, factor que influye en el cálculo del índice [Tabla 2.2]. De estos tres aminoácidos el de mayor volumen presenta por tener dos anillos aromáticos condensados, es el Triptófano. Esto le confiere un mayor valor intrínseco pero a su vez una mayor polarizabilidad, que es una de las propiedades químico-físicas que describe el índice. Un análisis similar, pero a la inversa puede realizarse para la Glicina. La Fenilalanina, con solamente un anillo aromático, posee menor volumen y, aunque también polarizable por la nube aromática del anillo bencénico, su valor es menor que el del Triptófano pero mucho mayor que el de la Glicina. En la Tirosina, la presencia del grupo hidroxilo en el anillo aromático le confiere a este aminoácido una mayor polaridad en detrimento de su polarizabilidad. Por esa razón, en los casos estudiados, una disminución de la cantidad de Triptófano en la secuencia, incrementa la posibilidad de alcanzar mayores valores a este aminoácido.

Se observó también que no todos los aminoácidos seleccionados con valores pequeños se encontraban sumergidos en la estructura ni todos los aminoácidos con valores más grandes se encontraban en la superficie. En el caso de estructuras pequeñas, cuando se analizó la capacidad del Índice del Estado Refractotopológico para Átomos de reflejar la magnitud en que el átomo que se analizaba se encontraba sumergido en la estructura 3D de la molécula, era evidente que valores mayores del índice significaban una mayor capacidad para interactuar con agentes externos o lo que se planteaba como equivalente, que se encontraban más hacia la superficie de la molécula. El comportamiento del índice en el caso de las proteínas tiene una significación equivalente pero no idéntica. En el caso de moléculas pequeñas, el concepto de un átomo sumergido o estéricamente impedido son prácticamente sinónimos, lo cual no es así para macromoléculas. En general, un menor valor del índice representa la disminución de la capacidad de la entidad química, de ser polarizada, es decir de la polarizabilidad, y viceversa. Por supuesto, cuando el aminoácido está por debajo de la superficie de la proteína, se hace claro que no solamente esta estéricamente impedido, sino que resulta difícil que adquiera una polaridad por efecto de agentes externos. Una situación semejante se puede presentar si el aminoácido analizado se encuentra en la superficie pero rodeado por otros que perturban su estructura disminuyendo su polarizabilidad y por ende, su capacidad para ser polarizado por una especie polar externa.

En este resultado pudiera influir la naturaleza de cada tipo de aminoácido o su valor intrínseco, lo cual sugirió un segundo experimento donde se considere el comportamiento de aminoácidos de un mismo tipo.

3.4.2 Experimento 2

Para cada una de las 5 proteínas se agruparon los aminoácidos de un mismo tipo, teniendo en cuenta los valores del índice se organizaron ascendientemente. Se tomó el menor y el mayor valor del índice para cada tipo de aminoácido [tablas 3.3 – 3.5], los aminoácidos con estos valores se representaron utilizando el SPDBV [figuras 3.11 – 3.13].

Tabla 3.3 Menor y Mayor valor del Índice del Estado Refractotopográfico para el aminoácido Alanina en la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

No.	Aminoácido	Índice	Menor Valor	Mayor Valor
224	ALA	-180,444	X	-
154	ALA	-177,033	-	-
144	ALA	-175,005	-	-
223	ALA	-174,762	-	-
191	ALA	-162,17	-	-
65	ALA	-156,816	-	-
272	ALA	-156,698	-	-
138	ALA	-143,087	-	-
300	ALA	-137,575	-	X

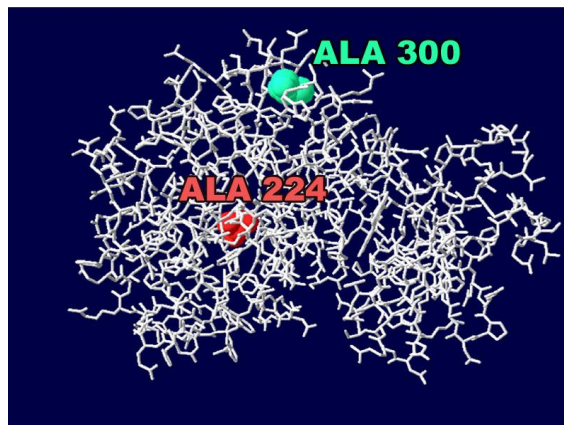


Figura 3.11 Visualización de la Alanina de menor valor del índice y la Alanina de mayor valor del índice para la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

Tabla 3.4 Menor y Mayor valor del Índice del Estado Refractotopográfico para el aminoácido Asparagina en la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

No.	Aminoácido	Índice	Menor Valor	Mayor Valor
172	ASN	-63,653	X	-
286	ASN	-45,966	-	-
160	ASN	-44,266	-	-
178	ASN	-38,321	-	-
61	ASN	-26,914	-	-
82	ASN	-25,320	-	X

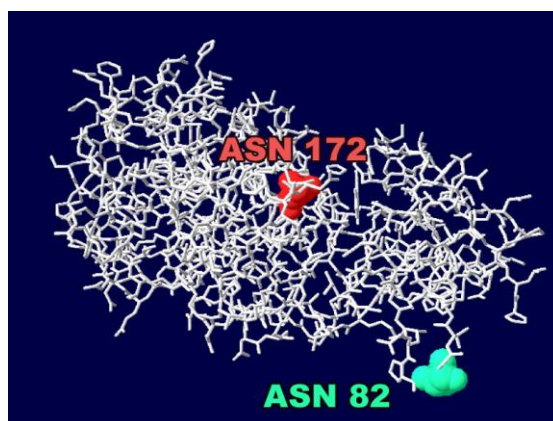


Figura 3.12 Visualización de la Asparagina de menor valor del índice y la Asparagina de mayor valor del índice para la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

Tabla 3.5 Menor y Mayor valor del Índice del Estado Refractotopográfico para el aminoácido Glicina en la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

No.	Aminoácido	Índice	Menor Valor	Mayor Valor
229	GLY	-269,528	X	-
102	GLY	-253,844	-	-
190	GLY	-247,103	-	-
220	GLY	-243,465	-	-
188	GLY	-243,393	-	-
203	GLY	-239,024	-	-
180	GLY	-230,348	-	-
238	GLY	-224,054	-	-
162	GLY	-222,507	-	-
50	GLY	-216,200	-	-
48	GLY	-214,090	-	-
251	GLY	-210,751	-	-
137	GLY	-209,684	-	-
45	GLY	-207,219	-	-
47	GLY	-204,581	-	-
55	GLY	-202,941	-	-
78	GLY	-201,138	-	-
99	GLY	-196,653	-	-
41	GLY	-179,784	-	-
83	GLY	-156,406	-	X

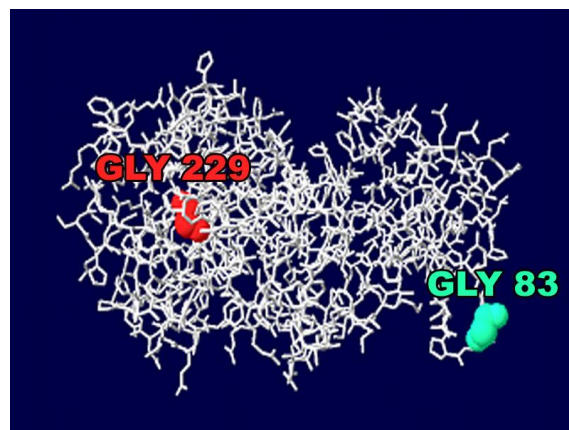


Figura 3.13 Visualización de la Glicina de menor valor del índice y la Glicina de mayor valor del índice para la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

Como resultado se obtuvo que, para cada aminoácido de la proteína, el mayor valor del índice correspondió al aminoácido menos bloqueado por sus vecinos más cercanos, generalmente localizado hacia la superficie y el de menor valor se encontraba más bloqueado por sus vecinos más cercanos [Anexo 1], generalmente localizado hacia el interior de la proteína.

Es válido destacar que esta conducta se manifiesta en la generalidad de los aminoácidos, pero existen ciertos casos como por ejemplo la Arginina, que describe un comportamiento anómalo [Tabla 3.6 y Figura 3.14].

Es un caso particular que requiere de un estudio más detallado que escapa del alcance de esta tesis. El valor del índice se comporta de forma contraria a lo

esperado para los aminoácidos. La base de cálculo del índice es a partir de los aminoácidos como molécula neutra y al grupo guanidino es altamente polar y susceptible de ser polarizado fuertemente por los otros de su entorno, con lo cual cambiaría el valor intrínseco del mismo por la posible formación de estructuras cargadas.

Tabla 3.6 Menor y Mayor valor del Índice del Estado Refractotopográfico para el aminoácido Glicina en la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

No.	Aminoácido	Índice	Menor Valor	Mayor Valor
214	ARG	177,011	X	-
250	ARG	177,865	-	-
258	ARG	181,924	-	-
217	ARG	184,040	-	-
274	ARG	184,227	-	-
85	ARG	189,471	-	-
112	ARG	191,287	-	-
256	ARG	193,132	-	-
268	ARG	200,335	-	-
57	ARG	202,947	-	-
73	ARG	206,144	-	-
136	ARG	206,186	-	-
205	ARG	210,033	-	-
179	ARG	211,320	-	-
145	ARG	216,475	-	-
156	ARG	221,567	-	-
221	ARG	222,385	-	-
278	ARG	223,302	-	-
122	ARG	230,461	-	-
105	ARG	232,208	-	-
166	ARG	248,244	-	X

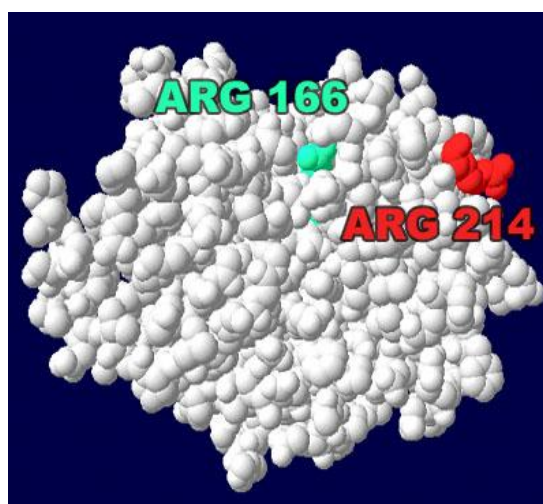


Figura 3.14 Visualización de la Arginina de menor valor del índice y la Arginina de mayor valor del índice para la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

Ya con este experimento se describe un comportamiento del índice el cual sugiere que valores mayores del mismo corresponden con residuos menos impedidos estéricamente y que valores mayores corresponden a residuos que se encuentran sumergidos o estéricamente impedidos, con lo cual se establece una analogía con respecto al Índice del Estado Refractotopológico para Átomos [epígrafe 1.4] que constituye un primer paso en la caracterización del Índice del Estado Refractotopográfico para Aminoácidos.

3.4.3 Experimento 3

Como se describió en el experimento anterior, el valor del índice de un aminoácido se ve más influenciado por sus vecinos más cercanos. Intuitivamente, este comportamiento da lugar a suponer que aquellos aminoácidos que se encuentren más lejanos, ofrecerán muy poca o ninguna influencia (perturbación) sobre él.

Se concibió este experimento para corroborar esta suposición y así poder acotar las distancias cuando se calcula el índice. Esta modificación agilizaría los cálculos y concentraría más el análisis en los aminoácidos más cercanos.

Tomándose los datos de las distancias y perturbaciones inherentes a cada aminoácido [Anexo 2] respecto al resto para cada proteína analizada, se implementó en la herramienta desarrollada, una funcionalidad que construyó una gráfica de dispersión [Figura 3.15] donde se observó el comportamiento Distancia vs. Perturbación de todos los aminoácidos: a partir de una distancia determinada, los valores perturbativos se comportaban asintóticamente, tendiendo a 0 (cero).

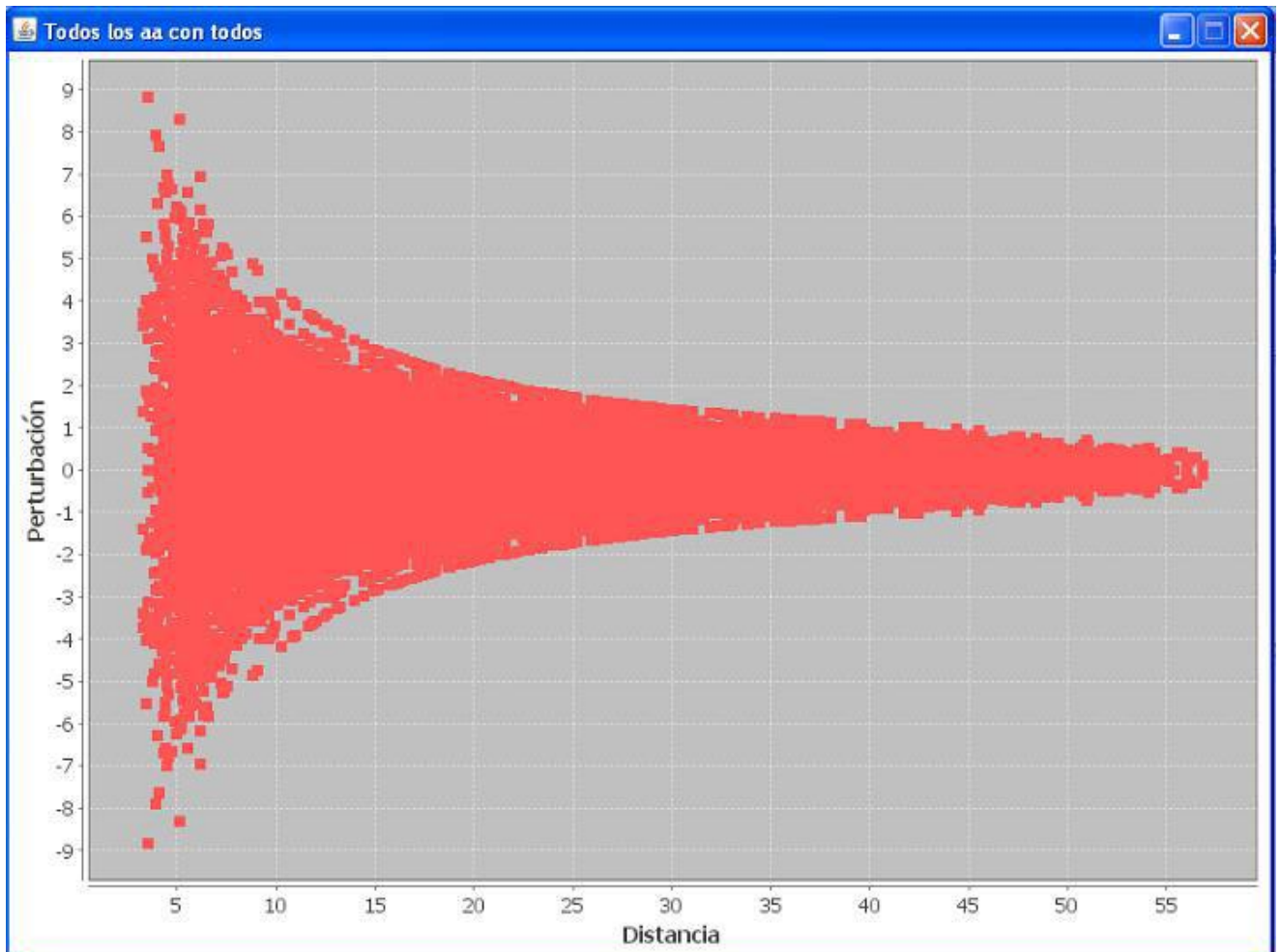


Figura 3.15 Gráfico de dispersión del comportamiento distancia-perturbación de todos los aminoácidos de la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

No obstante el resultado anterior, cabría interrogarse si este comportamiento se conserva para el análisis específico en cada aminoácido respecto al resto. Entonces se decidió realizar este análisis para cada aminoácido particularmente y comprobar si en algún caso específico no se cumple lo planteado anteriormente.

Se graficó, utilizando la herramienta desarrollada, el comportamiento Distancia vs. Perturbación cada aminoácido con el resto en las proteína seleccionadas [figuras 3.16 y 3.17] y se corroboró que mientras aumentaba la distancia, el valor perturbativo disminuía tendiendo a 0 (cero).

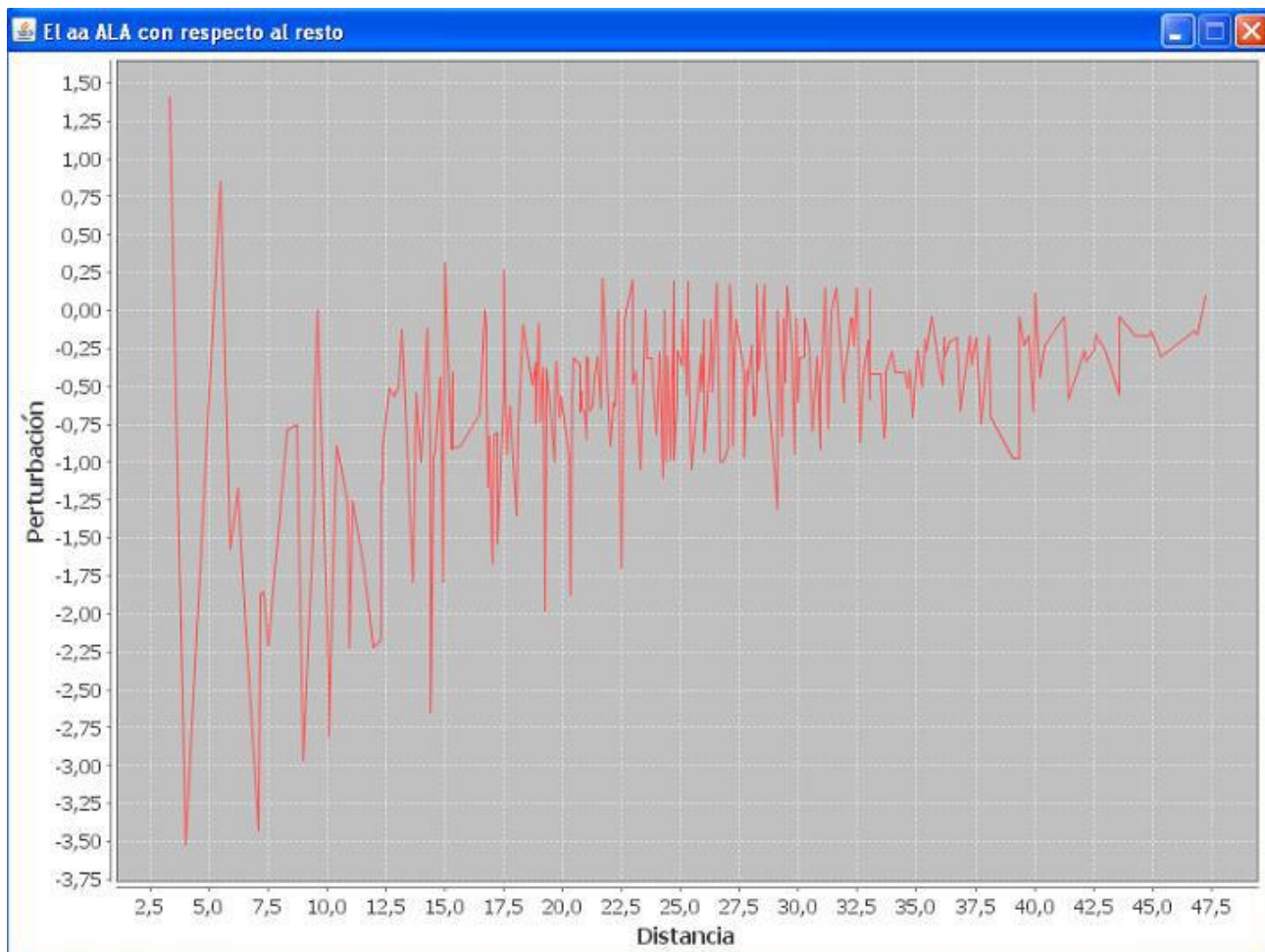


Figura 3.16 Comportamiento distancia-perturbación del aminoácido Alanina respecto al resto de los aminoácidos de la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

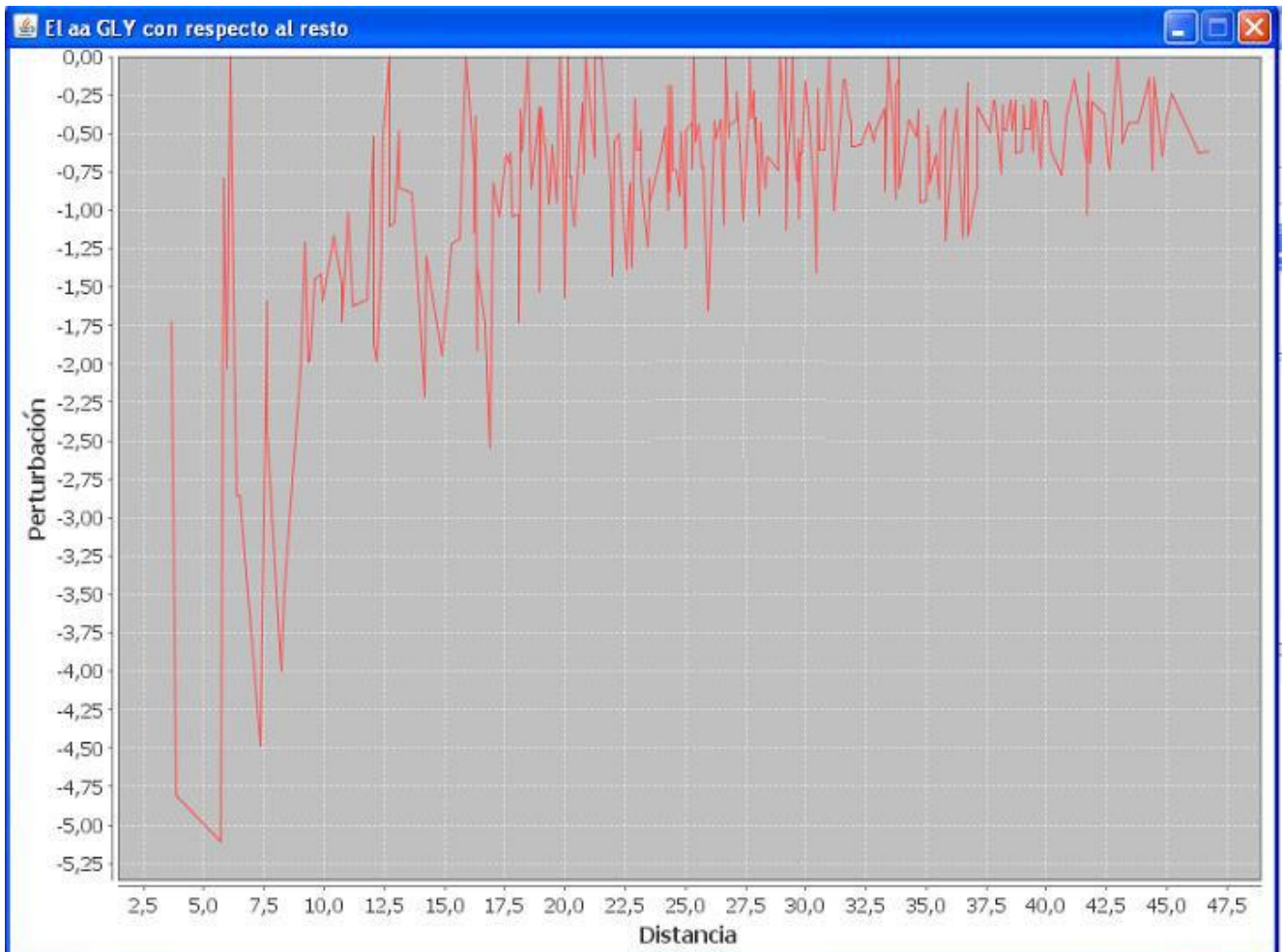


Figura 3.17 Comportamiento distancia-perturbación del aminoácido Glicina respecto al resto de los aminoácidos de la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

Otro aspecto a tener en cuenta en este caso sería el criterio de la polaridad de los aminoácidos, pues un cambio en esta propiedad influye en las interacciones de campos eléctricos e interacciones dipolo-dipolo que están presentes entre los aminoácidos y pudiera entonces afectar el comportamiento que se está analizando.

Teniendo en cuenta que se pueden clasificar los aminoácidos en dos grandes grupos: polares y apolares [Tabla 3.7], se tomaron todos los aminoácidos de cada proteína y se graficaron los valores del índice de cada uno de ellos respecto a los valores del índice de los aminoácidos polares y apolares [figuras 3.18 y 3.19].

Tabla 3.7 Clasificación de los aminoácidos según su polaridad.

No.	Aminoácido	Polaridad	No.	Aminoácido	Polaridad
1	Alanina(ALA)	polar	12	Isoleucina (ILE)	polar
2	Glicina (GLY)	polar	13	Leucina (LEU)	polar
3	Asparagina (ASN)	apolar	14	Lisina (LYS)	apolar
4	Glutamina (GLN)	apolar	15	Metionina (MET)	polar
5	Arginina (ARG)	apolar	16	Fenilalanina (PHE)	apolar
6	Ácido aspártico (ASP)	apolar	17	Prolina (PRO)	apolar
7	Aspartato (ASX)	apolar	18	Serina (SER)	apolar
8	Cisteína (CYS)	apolar	19	Taurina. (THE)	apolar
9	Ácido Glutámico (GLU)	apolar	20	Triptófano (TRP)	apolar
10	Glutamato (GLX)	apolar	21	Tirosina (TYR)	apolar
11	Histidina (HIS)	apolar	22	Valina (VAL)	polar

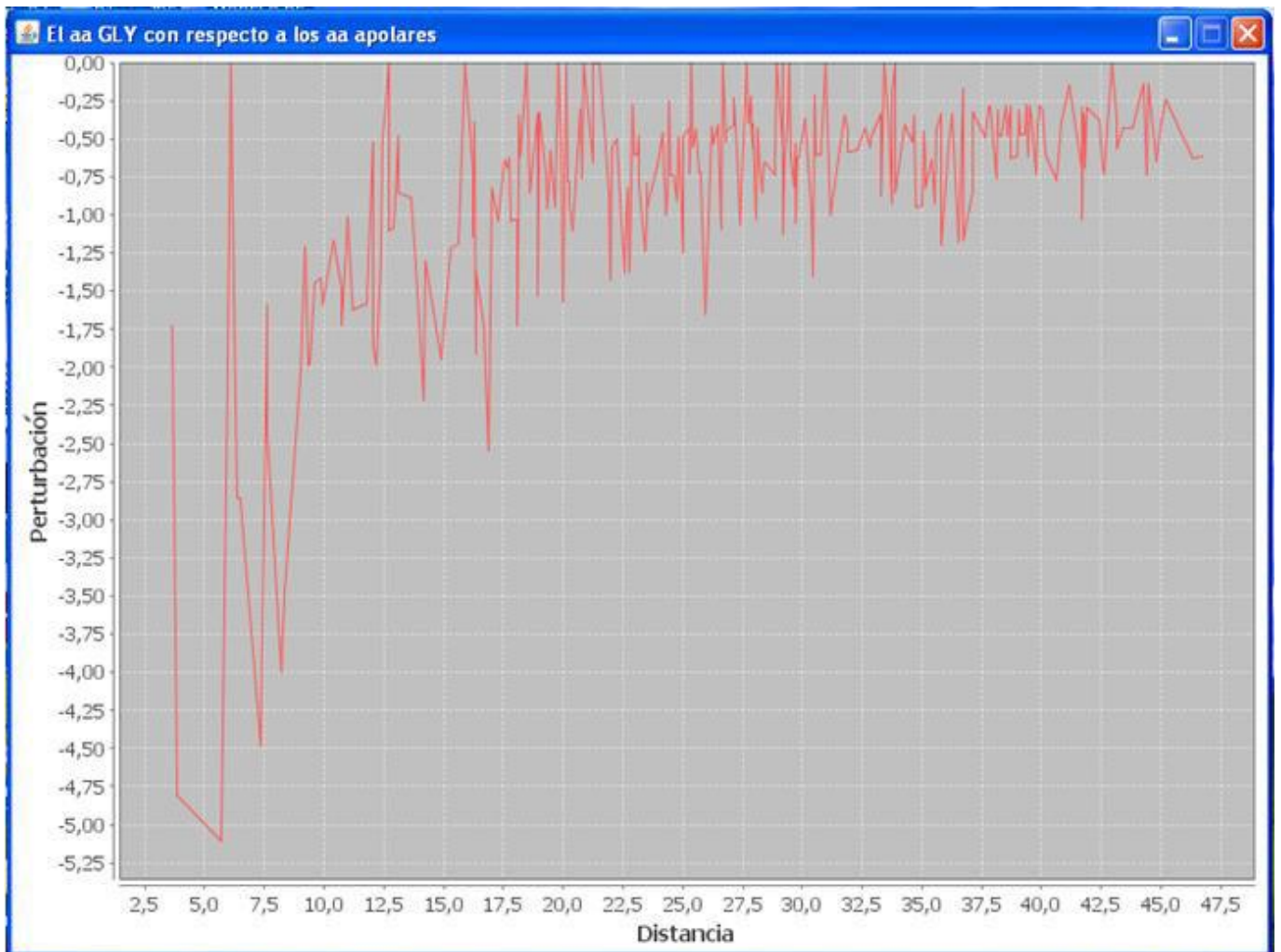


Figura 3.18 Comportamiento distancia-perturbación del aminoácido Glicina respecto a los aminoácidos apolares de la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

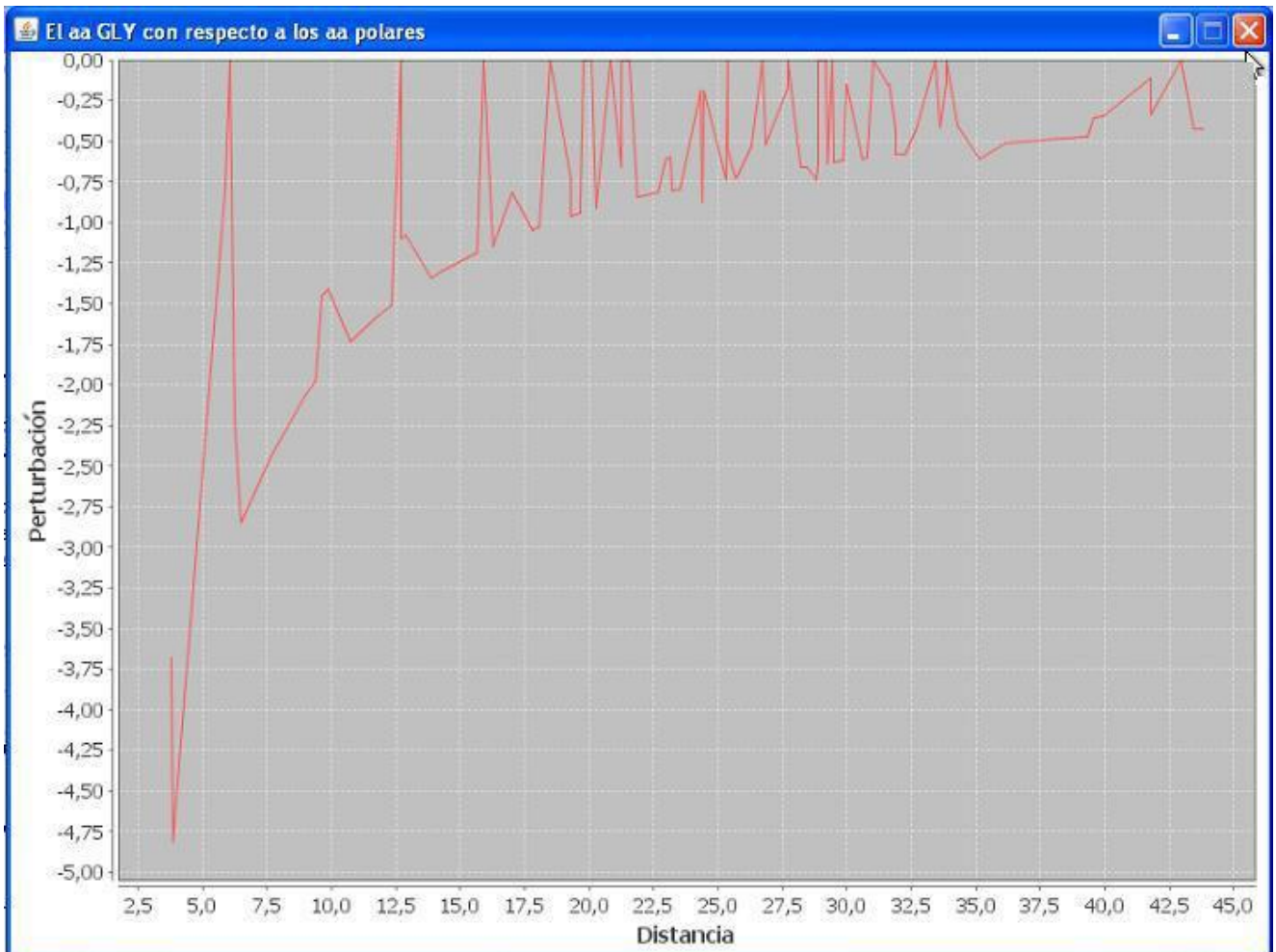


Figura 3.19 Comportamiento distancia-perturbación del aminoácido Glicina respecto a los aminoácidos polares de la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE.

Ambas gráficas muestran que, a medida que aumenta la distancia, la perturbación disminuye tendiendo a 0 (cero). Este comportamiento se mantuvo en todos los aminoácidos de cada una de las proteínas analizadas. Este resultado permite entonces sugerir una cota o rango de distancia de corte fuera de la cual los valores de la perturbación no son significativos y así minimizar el tiempo de cálculo del índice.

Se implementó en la herramienta una funcionalidad que permite recalcular el Índice del Estado Refractotopográfico para Aminoácidos a partir del establecimiento de una cota de distancia, dentro de la cual los valores de perturbación de cada aminoácido serían tomados en cuenta. Se observó que la cota de distancia varía de una proteína

a otra. El factor que aparentemente más influye en esta variación es la cantidad de aminoácidos que posea una proteína, pues mientras más aminoácidos tiene, se encuentra una mayor influencia del resto sobre el aminoácido analizado.

Además, se calculó el coeficiente de correlación entre el índice para esta modificación y el índice sin tener en cuenta la distancia de corte. Se considera válida una cota de distancia cuando la correlación es mayor que 0.99 [Tabla 3.8].

Tabla 3.8 Resultados para las proteínas analizadas después de aplicar la distancia de corte.

Proteína	Número de aminoácidos	Distancia de corte	Coefficiente de Correlación
ACTIN ALPHA 1	368	25	0,990
HIV-1 PROTEASE	198	24	0,991
PROTO-ONCOGENE SERINE/THREONINE- PROTEIN KINASE	274	21	0,991
COLLAGEN X	132	12	0,990
FIBROBLAST GROWTH FACTOR 7/1 CHIMERA	131	12	0,991

Los valores resultantes del cálculo del coeficiente de correlación fueron validados con el uso del software SPSS, arrojando un resultado idéntico.

3.5 Conclusiones

En este capítulo se realizaron tres experimentos, en 5 proteínas diferentes, para observar el comportamiento en las mismas y una posible analogía del descriptor redefinido y su antecedente. Se evidenció cierta similitud, aunque en casos específicos se observó un comportamiento anómalo. Los resultados de los cálculos realizados se recogen en un fichero que la herramienta implementada proporciona como salida, para un análisis posterior.

CONCLUSIONES GENERALES

- ✓ Se caracterizó mediante diferentes procedimientos el Índice del Estado Refractotopográfico para Aminoácidos, encontrándose que los mayores valores del mismo corresponden a aminoácidos más externos o más accesibles y que los menores valores corresponden a los que están, o sumergidos o menos accesibles
- ✓ Se implementó una herramienta visual en esta etapa de la investigación para la caracterización del Índice del Estado Refractotopográfico para Aminoácidos
- ✓ Se sugiere una distancia de corte para minimizar el tiempo de los cálculos.

RECOMENDACIONES

- ✓ Caracterizar el Índice del Estado Refractotopográfico para Aminoácidos, de acuerdo a otras propiedades físico-químicas
- ✓ Aplicar los experimentos a un conjunto más amplio de proteínas
- ✓ Realizar un estudio para el comportamiento anómalo identificado (aminoácido Arginina)
- ✓ Realizar un estudio teniendo en cuenta el entorno en que se encuentra cada aminoácido dentro de una proteína ya que esto puede influir en el valor del índice.

REFERENCIAS BIBLIOGRÁFICAS

1. Carrasco, R., *Nuevos descriptores atómicos y moleculares para estudios de estructura-actividad: Aplicaciones*. 2008: Ciudad de La Habana. p. 141.
2. López, M.d.C.A., *Introducción a la química terapéutica*. 1992: Editorial McGraw Hill. 850.
3. Escalona, J. and R. Carrasco. *Introducción al diseño racional de fármacos*. 2008.
4. Todeschini, R. and V. Consonni, *Handbook of molecular descriptors*. 2000: p. 667.
5. Giraldo, J.C.R. (2006) *Estado del arte de la Bioinformática*.
6. Andrade, M.G., et al. *Mensaje Bioquímico: La Estructura y la visualización Molecular de Proteínas*. 2005. México DF.: Edgar Vázquez Contreras.
7. León, J.d.C.y., *Guía de Iniciación al Lenguaje JAVA*. 1999.
8. Gutiérrez, J. *El entorno de desarrollo Eclipse* 2001 [Disponible en: www.uv.es/~jgutierr/MySQL_Java/TutorialEclipse.pdf].
9. Caicedo, M.I. *Sistemas de Partículas I Centro de Masa y Teorema del Momentum*. 2006 [Disponible en: www.fis.usb.ve/~mcaicedo/education/fisica2/centermass.pdf].
10. Anderson, D.R., D.J. Sweeney, and T.A. Williams, *Estadística para administración y economía*. 2004, Publicado por Cengage Learning Editores.

BIBLIOGRAFÍA

1. *Protein Data Bank Atomic Coordinate and Bibliographic Entry Format Description.* 1992
2. *Protein Data Bank Atomic Coordinate and Bibliographic Entry Format Description.* 1996
3. (2008) *Protein Data Bank Changes Guide.*
4. *Protein Data Bank Atomic Coordinate and Bibliographic Entry Format Description.* 2008
5. Badenas, F.G. *Métodos computacionales de modelado molecular y diseño de fármacos.*
6. Björkholm, P. *Method for recognizing local descriptors of protein structures using Hidden Markov Models.*
7. Escalona, J. and R. Carrasco. *Introducción al diseño racional de fármacos.* 2008
8. Font, M. *Diseño de Fármacos Racional Asistido por Ordenador.*
9. García, V.E.S., *Selección de nuevos antibacterianos por Topología Molecular.* 2004.
10. Hall, S.R. and B. McMahon. *International Tables for Crystallography Volume G: Definition and exchange of crystallographic data.*
11. Muñoz, F., J. Frau, and M. Coll. *Modelización Molecular. Aplicación a biomoléculas.*
12. Pastor, M., *Development and Applications of 3D Molecular Descriptors.* 2004.
13. Sarioz, D., T.Y. Kong, and G.T. Herman. *History Trees as Descriptors of Macromolecular Structures.* 2004 [cited.
14. Todeschini, R. and V. Consonni, *Handbook of molecular descriptors.* 2000: p. 667.
15. Giraldo, J.C.R. (2006) *Estado del arte de la Bioinformática.*
16. Benítez, O.A. and J.G.R. Hernández, *Algoritmos para el Cálculo de Descriptores Atómicos y Moleculares.* 2008, UCI.
17. López, M.d.C.A., *Introducción a la química terapéutica.* 1992: Editorial McGraw Hill. 850.
18. Anderson, D.R., D.J. Sweeney, and T.A. Williams, *Estadística para administración y economía.* 2004, Publicado por Cengage Learning Editores.
19. Andrade, M.G., et al., *Mensaje Bioquímico: La Estructura y la visualización Molecular de Proteínas.* Vol. XXIX. 2005, México DF.: Edgar Vázquez Contreras.

20. León, J.d.C.y., *Guía de Iniciación al Lenguaje JAVA*. 1999.
21. Carrasco, R., *Nuevos descriptores atómicos y moleculares para estudios de estructura-actividad: Aplicaciones*. 2008: Ciudad de La Habana. p. 141.
22. Garit, J.A.C. and Y. Marrero, *Linear indices of the "macromolecular graph's nucleotides adjacency matrix" as a novel approach in bioinformatics studies. 1. Prediction of paromomycin's affinity constant with HIV-1 Ψ -RNA packaging region*. 2005.
23. Garit, J.A.C. and Y.M. Ponce, *Linear indices of the "macromolecular graph's nucleotides adjacency matrix" as a novel approach in bioinformatics studies. 1. Prediction of paromomycin's affinity constant with HIV-1 Ψ -RNA packaging region*. 2005.
24. Marrero, Y., *Protein Quadratic Indices of the "Macromolecular Pseudograph's α -Carbon Atom Adjacency Matrix". 1. Prediction of Arc Repressor Alanine-mutant's Stability*. 2004.
25. Marrero, Y., *TOMOCOMD-CAMPS and Protein Bilinear Indices: Novel Bio-Macromolecular Descriptors for Protein Research*. 2008.
26. Marrero, Y., et al., *Nucleic Acid Quadratic Indices of the "Macromolecular Graph's Nucleotides Adjacency Matrix". Modeling of Footprints after the Interaction of Paromomycin with the HIV-1 Ψ -RNA Packaging Region*. 2004.
27. Yovanovic, A., *La Conminución como una Operación de transferencia Macromolecular de masa*. 2004.
28. Castañón, H.G.d.T., *Modelización molecular de los receptores de adenosina y sus ligandos en el marco de diseño de fármacos asistido por ordenador*, in *Departamento de Ciencias Experimentales de la Salud*. 2004, Universidad Pompeu Fabra: Barcelona. p. 132.
29. Gutiérrez, J. *El entorno de desarrollo Eclipse 2001* [Disponible en: http://www.uv.es/~jgutierrez/MySQL_Java/TutorialEclipse.pdf].
30. Caicedo, M.I. *Sistemas de Partículas I Centro de Masa y Teorema del Momentum*. 2006 [Disponible en: www.fis.usb.ve/~mcaicedo/education/fisica2/centermass.pdf]
31. Spegazzini, N., *Diseño de un Modelo de Interacción de Benzodiazepinas y Flavonoides con acción Ansiolítica derivado de su Análisis QSAR*, in *Departamento de Investigación*. 2007, Universidad de Belgrano: Buenos Aires. p. 56.

Anexo 1 Tablas Resumen de la visualización de los menores y mayores valores de cada tipo de aminoácido de las 5 proteínas

Tabla 1 Resultados para la proteína *PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE*

Proteína	Aminoácido	Menor Valor	Mayor Valor	Comportamiento
<i>PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE</i>	ALA	-180,444	-137,575	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	ARG	177,011	248,244	Ambos aminoácidos hacia la superficie. El menor valor del índice es el aminoácido menos bloqueado.
	ASN	-63,653	-25,320	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	ASP	-64,551	-29,011	“ “
	CYS	-53,886	-34,612	“ “
	HIS	148,951	128,240	“ “
	GLU	11,044	27,280	“ “
	GLY	-269,528	-156,406	“ “
	ILE	49,686	72,074	“ “
	LEU	52,312	71,299	“ “
	LYS	111,883	140,435	“ “
	MET	92,326	111,935	“ “
	PHE	209,625	300,148	Los dos aminoácidos hacia la superficie.
	PRO	-48,702	-15,424	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	SER	-159,107	-83,292	“ “
	THR	-61,183	-34,055	“ “
	TRP	399,907	473,317	Los dos aminoácidos hacia la superficie.
	TYR	223,984	314,190	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	VAL	-19,256	5,791	“ “

Tabla 2 Resultados para la proteína *HIV-1 PROTEASE*

Proteína	Aminoácido	Menor Valor	Mayor Valor	Comportamiento
<i>HIV-1 PROTEASE</i>	ALA	-106,862	-92,130	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	ARG	149,944	227,160	Ambos aminoácidos hacia la superficie. El menor valor del índice es el aminoácido menos

bloqueado.			
ASN	-16,997	-5,038	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
ASP	-13,879	-12,364	" "
CYS	-8,462	-3,185	" "
GLN	63,051	68,990	" "
GLU	41,390	47,555	" "
GLY	-181,552	-108,150	" "
HIS	137,606	137,855	" "
ILE	73,512	90,080	" "
LEU	70,780	88,973	" "
LYS	101,542	122,922	" "
MET	122,449	122,516	Los dos aminoácidos hacia la superficie.
PHE	214,961	232,579	" "
PRO	-6,171	6,714	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
THR	-21,705	-9,713	" "
TYR	233,995	234,386	Los dos aminoácidos hacia la superficie.
VAL	17,907	25,164	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado

Tabla 3 Resultados para la proteína ACTIN ALPHA 1

Proteína	Aminoácido	Menor Valor	Mayor Valor	Comportamiento
ACTIN ALPHA 1	ALA	-190,238	-102,843	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	ARG	225,447	317,410	Ambos aminoácidos hacia la superficie. El menor valor del índice es el aminoácido menos bloqueado.
	ASN	-55,780	-27,506	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	ASP	-54,774	-23,799	" "
	CYS	-41,535	-28,394	" "
	GLN	65,616	100,911	" "
	GLU	29,063	45,753	" "
	GLY	-295,726	-164,820	" "
	HIS	144,403	224,855	" "
	ILE	75,045	105,050	" "
	LEU	78,614	109,199	" "
	LYS	135,701	190,075	" "
	MET	115,929	161,850	" "

PHE	275,088	353,891	Los dos aminoácidos hacia la superficie.
PRO	-37,748	-14,742	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
SER	-155,578	-76,386	“ “
THR	-62,888	-24,734	“ “
TRP	511,707	607,644	Los dos aminoácidos hacia la superficie.
TYR	294,117	388,949	“ “
VAL	-1,709	16,490	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado

Tabla 4 Resultados para la proteína COLLAGEN X

Proteína	Aminoácido	Menor Valor	Mayor Valor	Comportamiento
COLLAGEN X	ALA	-88,820	-64,394	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	ARG	134,160	144,327	Ambos aminoácidos hacia la superficie. El menor valor del índice es el aminoácido menos bloqueado.
	ASN	-23,481	-8,811	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	ASP	-20,865	-11,590	“ “
	CYS	-10,130		Un sólo aminoácido de este tipo en la proteína
	GLN	46,240	55,004	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	GLU	27,074	34,323	“ “
	GLY	-153,680	-92,631	“ “
	HIS	107,934	119,897	“ “
	ILE	50,974	63,401	“ “
	LEU	50,489	62,702	“ “
	LYS	77,176	95,382	“ “
	MET	66,640	82,737	“ “
	PHE	179,749	206,801	Los dos aminoácidos hacia la superficie.
	PRO	-11,012	1,870	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	SER	-82,033	-60,742	“ “
	THR	-26,827	-10,226	“ “
	TRP	294,298	324,621	Los dos aminoácidos hacia la superficie.
	TYR	157,458	226,001	“ “
	VAL	13,000	11,805	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado

Tabla 5 Resultados para la proteína FIBROBLAST GROWTH FACTOR 7/1 CHIMERA

Proteína	Aminoácido	Menor Valor	Mayor Valor	Comportamiento
FIBROBLAST GROWTH FACTOR 7/1 CHIMERA	ALA	-112,174	-74,904	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	ARG	121,734	157,614	“ “
	ASN	-37,080	-19,491	“ “
	ASP	-27,226	-16,831	Los dos aminoácidos hacia la superficie.
	CYS	-31,897	-21,128	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	GLN	35,493	41,038	“ “
	GLU	16,169	22,291	“ “
	GLY	-168,590	-124,201	“ “
	HIS	87,947	99,944	“ “
	ILE	45,134	53,132	“ “
	LEU	39,333	51,190	“ “
	LYS	77,011	93,058	“ “
	MET	64,698	79,727	“ “
	PHE	170,587	195,454	Los dos aminoácidos hacia la superficie.
	PRO	-25,892	-11,976	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado
	SER	-89,486	-40,853	“ “
	THR	-42,602	-20,157	“ “
	TRP	294,891	297,602	Los dos aminoácidos hacia la superficie.
	TYR	169,714	211,929	Los dos aminoácidos hacia la superficie.
	VAL	-1,222	2,915	Menor valor : aminoácido más bloqueado Mayor valor: aminoácido menos bloqueado

Anexo 2 Tabla ejemplo del comportamiento distancia-perturbación

Tabla 6 Comportamiento distancia-perturbación del aminoácido Glicina en la proteína PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE

Aminoácido	No. Respecto a	No. Distancia	Perturbación
GLY	23	GLU 4	15,064 -1,050

GLY	23	THR	5	12,145	-0,900
GLY	23	THR	6	9,163	-1,193
GLY	23	ALA	7	8,285	-0,561
GLY	23	LEU	8	10,319	-1,801
GLY	23	VAL	9	9,283	-1,502
GLY	23	CYS	10	13,984	-0,856
GLY	23	ASP	11	14,763	-0,756
GLY	23	ASN	12	18,244	-0,612
GLY	23	GLY	13	20,320	0,000
GLY	23	SER	14	23,694	-0,265
GLY	23	GLY	15	22,808	0,000
GLY	23	LEU	16	19,725	-0,942
GLY	23	VAL	17	17,306	-0,806
GLY	23	LYS	18	14,460	-1,557
GLY	23	ALA	19	11,519	-0,403
GLY	23	GLY	20	8,584	0,000
GLY	23	PHE	21	7,203	-4,342
GLY	23	ALA	22	3,846	-1,208
GLY	23	ASP	24	3,551	-3,144
GLY	23	ASP	25	5,890	-1,896
GLY	23	ALA	26	7,658	-0,607
GLY	23	PRO	27	9,878	-1,226
GLY	23	ARG	28	9,743	-2,976
GLY	23	ALA	29	13,555	-0,343
GLY	23	VAL	30	16,325	-0,854
GLY	23	PHE	31	19,297	-1,621
GLY	23	PRO	32	22,809	-0,531
GLY	23	SER	33	23,363	-0,269
GLY	23	ILE	34	26,764	-0,694
GLY	23	VAL	35	28,898	-0,482
GLY	23	GLY	36	31,719	0,000
GLY	23	ARG	37	33,925	-0,855
GLY	23	PRO	38	37,720	-0,321
GLY	23	ARG	39	41,069	-0,706
GLY	23	HIS	40	44,766	-0,540
GLY	23	GLN	41	43,284	-0,418
GLY	23	GLY	42	44,344	0,000
GLY	23	VAL	43	41,942	-0,332
GLY	23	MET	44	39,319	-0,541
GLY	23	VAL	45	41,810	-0,333
GLY	23	GLY	46	41,540	0,000

GLY	23	MET	47	38,144	-0,557
GLY	23	GLY	48	37,280	0,000
GLY	23	GLN	49	38,707	-0,468
GLY	23	LYS	50	33,185	-0,678
GLY	23	ASP	51	34,106	-0,327
GLY	23	SER	52	31,038	-0,202
GLY	23	TYR	53	31,253	-1,053
GLY	23	VAL	54	25,457	-0,548
GLY	23	GLY	55	25,567	0,000
GLY	23	ASP	56	25,492	-0,438
GLY	23	GLU	57	27,708	-0,571
GLY	23	ALA	58	30,022	-0,155
GLY	23	GLN	59	29,211	-0,620
GLY	23	SER	60	32,556	-0,193
GLY	23	LYS	61	34,402	-0,654
GLY	23	ARG	62	36,249	-0,800
GLY	23	GLY	63	39,398	0,000
GLY	23	ILE	64	39,265	-0,473
GLY	23	LEU	65	35,847	-0,519
GLY	23	THR	66	36,421	-0,300
GLY	23	LEU	67	33,134	-0,561
GLY	23	LYS	68	31,471	-0,715
GLY	23	TYR	69	29,181	-1,128
GLY	23	PRO	70	27,357	-0,443
GLY	23	ILE	71	24,795	-0,750
GLY	23	GLU	72	29,872	-0,529
GLY	23	GLY	74	24,771	0,000
GLY	23	ILE	75	27,069	-0,687
GLY	23	ILE	76	25,883	-0,718
GLY	23	THR	77	30,358	-0,360
GLY	23	ASN	78	30,481	-0,366
GLY	23	TRP	79	28,162	-1,522
GLY	23	ASP	80	30,908	-0,361
GLY	23	ASP	81	28,981	-0,385
GLY	23	MET	82	23,874	-0,890
GLY	23	GLU	83	26,175	-0,604
GLY	23	LYS	84	27,114	-0,830
GLY	23	ILE	85	22,307	-0,833
GLY	23	TRP	86	20,302	-2,111
GLY	23	HIS	87	23,073	-1,048
GLY	23	HIS	88	21,302	-1,135

GLY	23	THR	89	17,211	-0,635
GLY	23	PHE	90	17,161	-1,822
GLY	23	TYR	91	20,873	-1,577
GLY	23	ASN	92	21,266	-0,525
GLY	23	GLU	93	17,497	-0,904
GLY	23	LEU	94	13,786	-1,348
GLY	23	ARG	95	18,377	-1,578
GLY	23	VAL	96	15,916	-0,876
GLY	23	ALA	97	17,753	-0,262
GLY	23	PRO	98	17,928	-0,675
GLY	23	GLU	99	20,069	-0,788
GLY	23	GLU	100	16,942	-0,933
GLY	23	HIS	101	13,370	-1,808
GLY	23	PRO	102	13,031	-0,929
GLY	23	THR	103	13,961	-0,783
GLY	23	LEU	104	12,289	-1,512
GLY	23	LEU	105	16,092	-1,155
GLY	23	THR	106	16,412	-0,666
GLY	23	GLU	107	20,913	-0,756
GLY	23	ALA	108	22,808	-0,204
GLY	23	PRO	109	24,563	-0,493
GLY	23	LEU	110	28,306	-0,657
GLY	23	ASN	111	25,873	-0,432
GLY	23	PRO	112	28,944	-0,418
GLY	23	LYS	113	29,362	-0,767
GLY	23	ALA	114	29,923	-0,155
GLY	23	ASN	115	26,879	-0,415
GLY	23	ARG	116	23,460	-1,236
GLY	23	GLU	117	28,237	-0,560
GLY	23	LYS	118	27,892	-0,807
GLY	23	MET	119	22,608	-0,940
GLY	23	THR	120	23,277	-0,470
GLY	23	GLN	121	27,757	-0,652
GLY	23	ILE	122	24,741	-0,751
GLY	23	MET	123	20,399	-1,042
GLY	23	PHE	124	22,873	-1,367
GLY	23	GLU	125	27,838	-0,568
GLY	23	THR	126	25,567	-0,427
GLY	23	PHE	127	21,933	-1,426
GLY	23	ASN	128	22,922	-0,487
GLY	23	VAL	129	19,342	-0,721

GLY	23	PRO	130	18,592	-0,651
GLY	23	ALA	131	17,295	-0,269
GLY	23	MET	132	18,656	-1,140
GLY	23	TYR	133	16,321	-2,016
GLY	23	VAL	134	18,379	-0,759
GLY	23	ALA	135	16,836	-0,276
GLY	23	ILE	136	19,509	-0,953
GLY	23	GLN	137	18,158	-0,997
GLY	23	ALA	138	19,127	-0,243
GLY	23	VAL	139	17,962	-0,776
GLY	23	LEU	140	13,931	-1,334
GLY	23	SER	141	15,786	-0,398
GLY	23	LEU	142	18,158	-1,024
GLY	23	TYR	143	15,077	-2,183
GLY	23	ALA	144	12,917	-0,360
GLY	23	SER	145	16,525	-0,380
GLY	23	GLY	146	16,757	0,000
GLY	23	ARG	147	19,775	-1,466
GLY	23	THR	148	20,733	-0,527
GLY	23	THR	149	24,002	-0,455
GLY	23	GLY	150	24,418	0,000
GLY	23	ILE	151	26,119	-0,712
GLY	23	VAL	152	22,862	-0,610
GLY	23	LEU	153	24,854	-0,748
GLY	23	ASP	154	22,053	-0,506
GLY	23	SER	155	24,044	-0,261
GLY	23	GLY	156	23,608	0,000
GLY	23	ASP	157	25,287	-0,442
GLY	23	GLY	158	27,524	0,000
GLY	23	VAL	159	26,539	-0,525
GLY	23	THR	160	27,453	-0,398
GLY	23	HIS	161	26,243	-0,921
GLY	23	ASN	162	28,382	-0,393
GLY	23	VAL	163	24,971	-0,558
GLY	23	PRO	164	27,227	-0,445
GLY	23	ILE	165	23,859	-0,779
GLY	23	TYR	166	27,748	-1,186
GLY	23	GLU	167	25,496	-0,620
GLY	23	GLY	168	21,844	0,000
GLY	23	TYR	169	26,626	-1,236
GLY	23	ALA	170	25,473	-0,182

GLY	23	LEU	171	29,038	-0,640
GLY	23	PRO	172	30,138	-0,402
GLY	23	HIS	173	33,599	-0,719
GLY	23	ALA	174	31,782	-0,146
GLY	23	ILE	175	29,461	-0,631
GLY	23	MET	176	32,066	-0,663
GLY	23	ARG	177	30,943	-0,937
GLY	23	LEU	178	31,677	-0,587
GLY	23	ASP	179	32,162	-0,347
GLY	23	LEU	180	31,677	-0,587
GLY	23	ALA	181	28,667	-0,162
GLY	23	GLY	182	27,929	0,000
GLY	23	ARG	183	28,981	-1,001
GLY	23	ASP	184	32,392	-0,345
GLY	23	LEU	185	31,551	-0,589
GLY	23	THR	186	31,587	-0,346
GLY	23	ASP	187	34,719	-0,322
GLY	23	TYR	188	37,749	-0,872
GLY	23	LEU	189	35,658	-0,521
GLY	23	MET	190	36,948	-0,575
GLY	23	LYS	191	41,935	-0,537
GLY	23	ILE	192	40,486	-0,459
GLY	23	LEU	193	40,261	-0,462
GLY	23	THR	194	43,523	-0,251
GLY	23	GLU	195	45,238	-0,350
GLY	23	ARG	196	44,822	-0,647
GLY	23	GLY	197	46,940	0,000
GLY	23	TYR	198	43,587	-0,755
GLY	23	SER	199	43,158	-0,146
GLY	23	PHE	200	39,147	-0,799
GLY	23	VAL	201	40,510	-0,344
GLY	23	THR	202	38,348	-0,285
GLY	23	THR	203	34,834	-0,314
GLY	23	ALA	204	34,735	-0,134
GLY	23	GLU	205	37,552	-0,421
GLY	23	ARG	206	32,081	-0,904
GLY	23	GLU	207	29,770	-0,531
GLY	23	ILE	208	33,116	-0,561
GLY	23	VAL	209	32,992	-0,423
GLY	23	ARG	210	27,124	-1,069
GLY	23	ASP	211	28,554	-0,391

GLY	23	ILE	212	32,127	-0,579
GLY	23	LYS	213	28,936	-0,778
GLY	23	GLU	214	25,348	-0,624
GLY	23	LYS	215	28,142	-0,800
GLY	23	LEU	216	31,899	-0,583
GLY	23	CYS	217	31,203	-0,383
GLY	23	TYR	218	33,679	-0,977
GLY	23	VAL	219	32,138	-0,434
GLY	23	ALA	220	35,662	-0,130
GLY	23	LEU	221	36,875	-0,504
GLY	23	ASP	222	40,855	-0,273
GLY	23	PHE	223	41,168	-0,760
GLY	23	GLU	224	44,185	-0,358
GLY	23	ASN	225	44,509	-0,251
GLY	23	GLU	226	40,645	-0,389
GLY	23	MET	227	44,010	-0,483
GLY	23	ALA	228	46,999	-0,099
GLY	23	THR	229	44,845	-0,244
GLY	23	ALA	230	45,020	-0,103
GLY	23	ALA	231	48,247	-0,096
GLY	23	SER	232	47,931	-0,131
GLY	23	SER	233	44,035	-0,143
GLY	23	SER	234	43,639	-0,144
GLY	23	SER	235	39,603	-0,159
GLY	23	LEU	236	39,832	-0,467
GLY	23	GLU	237	42,076	-0,376
GLY	23	LYS	238	36,958	-0,609
GLY	23	SER	239	38,383	-0,164
GLY	23	TYR	240	34,771	-0,947
GLY	23	GLU	241	36,917	-0,428
GLY	23	LEU	242	37,855	-0,491
GLY	23	PRO	243	36,655	-0,330
GLY	23	ASP	244	41,127	-0,272
GLY	23	GLY	245	41,122	0,000
GLY	23	GLN	246	42,960	-0,421
GLY	23	VAL	247	41,484	-0,336
GLY	23	ILE	248	40,082	-0,464
GLY	23	THR	249	41,765	-0,262
GLY	23	ILE	250	38,195	-0,487
GLY	23	GLY	251	39,701	0,000
GLY	23	ASN	252	41,251	-0,271

GLY	23	GLU	253	39,212	-0,403
GLY	23	ARG	254	35,079	-0,827
GLY	23	PHE	255	37,624	-0,831
GLY	23	ARG	256	39,829	-0,728
GLY	23	CYS	257	34,163	-0,350
GLY	23	PRO	258	32,646	-0,371
GLY	23	GLU	259	36,492	-0,433
GLY	23	THR	260	35,320	-0,309
GLY	23	LEU	261	31,807	-0,584
GLY	23	PHE	262	34,243	-0,913
GLY	23	GLN	263	39,249	-0,461
GLY	23	PRO	264	37,759	-0,321
GLY	23	SER	265	41,814	-0,150
GLY	23	PHE	266	40,353	-0,775
GLY	23	ILE	267	38,696	-0,480
GLY	23	GLY	268	41,699	0,000
GLY	23	MET	269	38,522	-0,552
GLY	23	GLU	270	42,534	-0,372
GLY	23	SER	271	38,288	-0,164
GLY	23	ALA	272	37,020	-0,126
GLY	23	GLY	273	35,178	0,000
GLY	23	ILE	274	31,311	-0,594
GLY	23	HIS	275	35,089	-0,689
GLY	23	GLU	276	37,540	-0,421
GLY	23	THR	277	33,130	-0,330
GLY	23	THR	278	31,401	-0,348
GLY	23	TYR	279	37,090	-0,887
GLY	23	ASN	280	36,519	-0,306
GLY	23	SER	281	32,350	-0,194
GLY	23	ILE	282	33,145	-0,561
GLY	23	MET	283	37,746	-0,563
GLY	23	LYS	284	36,209	-0,622
GLY	23	CYS	285	33,843	-0,354
GLY	23	ASP	286	36,132	-0,309
GLY	23	ILE	287	38,594	-0,482
GLY	23	ASP	288	36,995	-0,302
GLY	23	ILE	289	32,156	-0,578
GLY	23	ARG	290	35,635	-0,814
GLY	23	LYS	291	33,523	-0,672
GLY	23	ASP	292	29,629	-0,377
GLY	23	LEU	293	29,032	-0,640

GLY	23	TYR	294	32,911	-1,000
GLY	23	ALA	295	28,403	-0,164
GLY	23	ASN	296	25,262	-0,442
GLY	23	ASN	297	26,577	-0,420
GLY	23	VAL	298	21,689	-0,643
GLY	23	MET	299	22,887	-0,929
GLY	23	SER	300	19,848	-0,317
GLY	23	GLY	301	19,707	0,000
GLY	23	GLY	302	21,344	0,000
GLY	23	THR	303	24,824	-0,440
GLY	23	THR	304	23,031	-0,475
GLY	23	MET	305	21,687	-0,980
GLY	23	TYR	306	27,285	-1,206
GLY	23	PRO	307	28,110	-0,431
GLY	23	GLY	308	29,907	0,000
GLY	23	ILE	309	28,103	-0,661
GLY	23	ALA	310	28,317	-0,164
GLY	23	ASP	311	31,447	-0,355
GLY	23	ARG	312	34,393	-0,843
GLY	23	MET	313	30,836	-0,689
GLY	23	GLN	314	32,504	-0,557
GLY	23	LYS	315	37,211	-0,605
GLY	23	GLU	316	36,685	-0,431
GLY	23	ILE	317	34,463	-0,539
GLY	23	THR	318	37,441	-0,292
GLY	23	ALA	319	40,182	-0,116
GLY	23	LEU	320	39,192	-0,474
GLY	23	ALA	321	39,106	-0,119
GLY	23	PRO	322	40,357	-0,300
GLY	23	SER	323	42,455	-0,148
GLY	23	THR	324	40,653	-0,269
GLY	23	MET	325	36,618	-0,581
GLY	23	LYS	326	32,871	-0,685
GLY	23	ILE	327	32,546	-0,571
GLY	23	LYS	328	27,233	-0,827
GLY	23	ILE	329	26,856	-0,692
GLY	23	ILE	330	22,500	-0,826
GLY	23	ALA	331	21,336	-0,218
GLY	23	PRO	332	17,966	-0,674
GLY	23	PRO	333	17,295	-0,700
GLY	23	GLU	334	13,040	-1,213

GLY	23	ARG	335	18,809	-1,542
GLY	23	LYS	336	16,950	-1,328
GLY	23	TYR	337	13,172	-2,498
GLY	23	SER	338	13,427	-0,468
GLY	23	VAL	339	13,529	-1,030
GLY	23	TRP	340	8,371	-5,120
GLY	23	ILE	341	8,471	-2,194
GLY	23	GLY	342	9,830	0,000
GLY	23	GLY	343	8,876	0,000
GLY	23	SER	344	5,270	-1,192
GLY	23	ILE	345	7,559	-2,459
GLY	23	LEU	346	10,679	-1,741
GLY	23	ALA	347	8,272	-0,562
GLY	23	SER	348	6,230	-1,008
GLY	23	LEU	349	10,410	-1,786
GLY	23	SER	350	12,177	-0,516
GLY	23	THR	351	14,841	-0,736
GLY	23	PHE	352	15,675	-1,995
GLY	23	GLN	353	15,923	-1,137
GLY	23	GLN	354	19,840	-0,912
GLY	23	MET	355	18,242	-1,165
GLY	23	TRP	356	14,789	-2,898
GLY	23	ILE	357	20,307	-0,915
GLY	23	THR	358	21,784	-0,502
GLY	23	LYS	359	24,689	-0,912
GLY	23	GLN	360	26,908	-0,673
GLY	23	GLU	361	24,834	-0,637
GLY	23	TYR	362	26,630	-1,236
GLY	23	ASP	363	29,212	-0,382
GLY	23	GLU	364	30,830	-0,513
GLY	23	ALA	365	30,392	-0,153
GLY	23	GLY	366	29,586	0,000
GLY	23	PRO	367	29,585	-0,409
GLY	23	SER	368	29,549	-0,213
GLY	23	ILE	369	25,950	-0,716
GLY	23	VAL	370	23,680	-0,589
GLY	23	HIS	371	26,498	-0,912
GLY	23	ARG	372	26,094	-1,111

GLOSARIO DE TÉRMINOS

Actividad Biológica: Capacidad inherente de una sustancia para alterar una o más funciones químicas o fisiológicas de una célula.

Aminoácido: Principal constituyente de las proteínas, su estructura incluye un grupo amino, un grupo carboxilo y una cadena lateral de composición variable.

Angstrom: Unidad de longitud equivalente a 0,00000001cm. Usualmente, en física se escribe de la siguiente manera: 1×10^{-8} cm.

Centro de masa: Es un punto que representa el promedio pesado de todas las partículas que forman el sistema.

Coefficiente de correlación: Provee una medida de como dos variables aleatorias están asociadas en una muestra.

Confómeros: Compuestos químicos.

Descriptor: Número que describe la estructura química o una propiedad de la molécula o fragmento de ésta.

Grafo molecular: Representación pictórica de la topología molecular.

Índice topográfico: Número que se calcula generalmente a partir de la matriz de adyacencia o de distancias entre los elementos de un grafo que han sido ponderados por un valor numérico que contiene información tridimensional del grafo molecular.

Proteína: Macromolécula compuesta por una o más cadenas polipeptídicas. Cada una con una secuencia particular de aminoácidos conectadas por enlace peptídico. Las proteínas realizan numerosas funciones vitales dentro y fuera de la célula.

Superficie de Van der Waals: Define una superficie molecular y un volumen encerrado por dicha superficie.

Tamizaje: Son pruebas sistemáticas de laboratorio que permiten identificar características comunes en una muestra determinada.

Topografía molecular: Es la información que puede obtenerse de la distribución espacial de los átomos de una molécula.

Topología molecular: Es toda la información (y la única) que puede obtenerse de la conectividad mutua entre todos los pares de átomos en una molécula.

Valor intrínseco: Valor propio o característico de una entidad por sí misma y no por causas exteriores.