

**DESARROLLO DE SOFTWARE PARA IDENTIFICACIÓN DE ANOMALÍAS NO
COMUNES EN LA PROTEÍNA P53 Y CASOS DE HEPATOCARCINOMA
CELULAR**

INVESTIGADORES

ING. LUIS CARLOS TOVAR GARRIDO

ASESORES

MIGUEL ALCALDE ALVITES

CO-INVESTIGADORES

DANIEL ANDRÉS OROZCO MÉNDEZ

JAVIER DAVID CASTILLO BELTRÁN



**UNIVERSIDAD DE CARTAGENA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
CARTAGENA DE INDIAS**

**DESARROLLO DE SOFTWARE PARA IDENTIFICACIÓN DE ANOMALÍAS
NO COMUNES EN LA PROTEÍNA P53 Y CASOS DE HEPATOCARCINOMA
CELULAR**

GRUPO DE INVESTIGACIÓN

GIMÁTICA

LÍNEA DE INVESTIGACIÓN

BIOINFORMÁTICA

INVESTIGADORES

ING. LUIS CARLOS TOVAR GARRIDO

ASESORES

MIGUEL ALCALDE ALVITES

CO-INVESTIGADORES

DANIEL ANDRÉS OROZCO MÉNDEZ

JAVIER DAVID CASTILLO BELTRÁN



**UNIVERSIDAD DE CARTAGENA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
CARTAGENA DE INDIAS**

CONTENIDO

1	RESUMEN	5
2	ABSTRACT	6
3	INTRODUCCIÓN	7
4	PLANTEAMIENTO DEL PROBLEMA	9
4.1	DESCRIPCIÓN DEL PROBLEMA.....	9
4.2	FORMULACIÓN DEL PROBLEMA.....	10
5	MARCO DE REFERENCIA.....	11
5.1	MARCO TEÓRICO.....	11
5.1.1	HEPATOCARCINOMA CELULAR	11
5.1.2	EL GEN Y LA PROTEÍNA P53	12
5.1.3	P53 Y LOS TIPOS DE CÁNCER	13
5.1.4	ACCIÓN MOLECULAR DE P53	14
5.1.5	P53 FUNCIÓN Y DISFUNCIÓN	15
5.1.6	ALTERACIONES DE P53 Y LA HEPATOCARCINOGENESIS.	17
5.1.7	P53 ESPECTRO DE MUTACIONES EN HCC.....	17
5.1.8	SECUENCIACIÓN DE P53.....	19
5.1.9	ESTRUCTURA DE LAS PROTEÍNAS.....	22
5.1.10	SECUENCIACIÓN DE AMINOÁCIDOS.....	28
5.1.11	DETERMINACIÓN DE LA ESTRUCTURA PROTEICA.....	29
5.1.12	ALGORITMO DE ALINEAMIENTO UNIÓN DE VECINOS (NEIGHBOR-JOINING).....	31
5.1.13	HERRAMIENTAS PARA EL MODELADO 3D.....	31
5.1.14	ENTORNO INTEGRADO DE DESARROLLO (IDE, INTEGRATED DEVELOPMENT ENVIRONMENT).....	33
5.2	ESTADO DEL ARTE	34
5.2.1	RELACIÓN ENTRE LA PROTEÍNA P53 Y CASOS DE HEPATOCARCINOMA CELULAR	34
5.2.2	MÚLTIPLE ALINEAMIENTO DE SECUENCIAS DE AMINOÁCIDOS USANDO CLUSTAL	35
6	JUSTIFICACIÓN.....	38
7	OBJETIVOS	40
7.1	OBJETIVO GENERAL.....	40

7.2	OBJETIVOS ESPECÍFICOS.....	40
8	ALCANCE.....	41
9	METODOLOGÍA.....	42
9.1	TIPO DE INVESTIGACIÓN	42
9.2	FASES DEL DISEÑO Y DESARROLLO	43
9.2.1	FASE INICIAL	43
9.2.2	FASE DE ELABORACIÓN	44
9.2.3	FASE DE CONSTRUCCIÓN	44
9.2.4	FASE DE TRANSICIÓN.....	46
10	RESULTADOS Y DISCUSIÓN.....	47
10.1	FASE INICIAL	47
10.2	FASE DE ELABORACIÓN	47
10.2.1	ESPECIFICACIÓN DE REQUISITOS DEL SISTEMA	50
10.3	FASE DE CONSTRUCCION	53
10.3.1	DESARROLLO DEL SOFTWARE	53
10.4	FASE DE TRANSICIÓN.....	56
10.4.1	PRUEBA DE MÓDULOS EN TIEMPO REAL.....	56
10.4.2	PRUEBAS DE REQUISITOS FUNCIONALES.....	58
11	CONCLUSIONES	62
12	RECOMENDACIONES	64
13	REFERENCIAS.....	65

1 RESUMEN

P53 es una proteína supresora de tumores, que cumple un papel importante en la apoptosis (muerte celular programada por el organismo) y el control del ciclo celular. Si p53 no funciona de la forma correcta podría permitir que las células anormales proliferen, dando como resultado cáncer. El mal funcionamiento de p53 puede ser provocado por una mutación de él mismo.

Estudios basados en la proteína p53 demuestran que diferentes tipos de mutaciones podrían estar ligados a específicos tipos de cáncer. Estos estudios intentan analizar específicamente el comportamiento de la secuencia aminoacídica de p53 con base a los exones que pertenece cada codón, estableciendo así las regiones que son consideradas mutables o donde puede existir una mutación y definiendo como afecta esta mutación al desarrollo de un tumor cancerígeno específico.

El objetivo principal de esta investigación es facilitar el estudio de la secuencia aminoacídica de p53. De tal forma que se pueda realizar comparaciones entre la secuencia de aminoácidos de p53 en su forma común con ejemplos de otras p53 que tendrían mutaciones en su secuencia de aminoácidos, comparar los resultados obtenidos con otras investigaciones que relacionan el tipo de mutación con un tipo de hepatocarcinoma celular (HCC) específico y posteriormente graficar en un modelo 3D la estructura de la proteína mutada.

Al finalizar esta investigación se puede concluir que, con base a los estudios de p53 y casos de HCC por parte de otros científicos, se puede determinar uno o varios tipos de HCC específico cuando existe una mutación en exón determinado de p53 y estudiar a nivel molecular su estructura, y por ende, se considera que los resultados de la investigación son positivos y con posibilidad de mejora. Esto implica que se debe incentivar la investigación en este campo específico de la bioinformática para mejorar los resultados del análisis de p53 y su influencia en casos de HCC específicos.

2 ABSTRACT

P53 is a tumor suppressor protein, which plays an important role in apoptosis (cell death programmed by the organism) and cell cycle control. If p53 does not work the right way it could allow abnormal cells to proliferate, resulting in cancer. Malfunction of p53 can be caused by a mutation of itself.

Studies based on the p53 protein show that different types of mutations may be linked to specific types of cancer. These studies try to specifically analyze the behavior of the p53 amino acid sequence based on the exons belonging to each codon, thus establishing the regions that are considered mutable or where a mutation may exist and defining how this mutation affects the development of a specific cancerous tumor.

The principal objective of this research is to facilitate the study of the amino acid sequence of p53. In such a way that comparisons can be made between the amino acid sequence of p53 in its common form with cases of other p53 that would have mutations in its amino acid sequence, compare the results obtained with other investigations that relate the type of mutation to a type of Specific cellular hepatocarcinoma (HCC) and subsequently plot the structure of the mutated protein in a 3D model.

At the end of this investigation it can be concluded that, based on p53 studies and cases of HCC by other scientists, a type of specific HCC can be determined when there is a mutation in exon determined of p53 and to study at molecular level its structure, And therefore, it is considered that the results of the research are positive and with the possibility of improvement. This implies that research in this specific area of bioinformatics should be encouraged to improve the results of p53 analysis and its influence on specific HCC cases.

3 INTRODUCCIÓN

Los laboratorios de genética y bioquímica requieren maquinarias e instrumentos muy costosos que pocas industrias los fabrican, en especial en países como Perú y Colombia. La elaboración de software para el ámbito del área biológica y/o médica está siendo muy utilizada a nivel mundial como medio alternativo al uso de técnicas de alto costo, de tal forma que permita simular un proceso químico realizado por un espectrómetro de masas, por ejemplo, desde un computador.

El enfoque de los últimos años en el campo de la genética y estudio de la tumorigénesis se dirige al estudio de enfermedades que generan tumores cancerígenos y los posibles marcadores que permiten identificarlos dentro del genoma humano, como es el caso de la proteína p53 y la proteína p21. La proteína p53 detalla precisamente el tipo de Hepatocarcinoma Celular que puede presentar un paciente dependiendo de la mutación que esta haya tenido, dependiendo de atributos de la proteína (región en que se encuentre y cambio de la base nitrogenada del aminoácido mutado).

Estudiar la proteína p53 implica llevar a cabo los procedimientos de secuenciación de la proteína y de múltiple alineamiento de las secuencias proteómicas, procedimientos químicos con pasos muy meticulosos los cuales deben ser realizados por expertos en el tema y con máquinas e instrumentos necesarios en los laboratorios donde se lleve a cabo el estudio. En el estado del arte se puede encontrar métodos químicos utilizados por investigadores para realizar estudio de las mutaciones en la proteína, explicando detalladamente el proceso y evidenciando la necesidad que este proceso tiene de ser automatizado. Además se detalla en el estado del arte los antecedentes de investigaciones con el mismo propósito, especificando los software que actualmente cumplen por separado los objetivos específicos de esta investigación.

Los co-investigadores Daniel Andrés Orozco Méndez¹ y Javier David Castillo Beltrán¹, del grupo de investigación GIMATICA, en conjunto de los investigadores Luis Carlos Tovar Garrido² y Miguel Alcalde Alvites³, desarrollaron un software que permite analizar y visualizar las distintas mutaciones puntuales de p53 en su estructura terciaria (modelo 3D) y a lo largo de este documento se describirá el proceso de desarrollo. De esta forma se espera incentivar los estudios y procesos investigativos del Hepatocarcinoma Celular y los procesos de alineamientos de secuencias proteómicas y ácido nucleicas, e incluso investigaciones con enfoques distintos como el diseño de fármacos en Colombia y Perú. Igualmente esta investigación es un gran aporte en los temas de investigación relacionados con las técnicas de inteligencia artificial aplicadas en bioinformática las cuales está adscritas a la línea de investigación de inteligencia computacional suscrita al grupo de investigación GIMATICA.

¹ Estudiante de Ingeniería de Sistemas de la Universidad de Cartagena.

² Profesor de la Universidad de Cartagena. MsC en Ciencias Computacionales.

³ Estudiante de genética y bioinformática de la Universidad Nacional Mayor de San Marcos.

4 PLANTEAMIENTO DEL PROBLEMA

4.1 DESCRIPCIÓN DEL PROBLEMA

El enfoque mundial acerca de buscar blancos⁴ para enfrentar enfermedades que causan tumores o actividades tumorigénicas ha ido variando con el tiempo convirtiéndose, al pasar de los años, en un tema de gran impacto a nivel investigativo en la comunidad bióloga y médica. Actualmente, en Perú y Colombia, se buscan marcadores moleculares que estén implicados en el proceso de tumorigénesis, es decir, el proceso en el cual las células normales se transforman en células cancerígenas dando paso a los tumores relacionados con el cáncer. Uno de ellos es la proteína p53.

Esta proteína en su estado normal es un supresor de tumores, pero cuando adquiere alguna mutación puntual (cambia un solo nucleótido) o una acumulación de ellas, pierde esta función y ocurre la aparición de tumores. Así lo afirman León & Bosques (2005) “Normalmente la proteína p53 activada es requerida para detener el ciclo celular en la fase G1 como resultado de la estimulación directa a p21cip1⁵. Una ausencia o inhibición de esta propiedad conduce a inestabilidad genómica.¹² Alternativamente, cuando los daños en el ADN son más severos o cuando hay una gran replicación viral, p53 puede activar la vía apoptótica”

Los estudios se enfocan en determinar las posiciones de las mutaciones que aparecen en pacientes como los que presentan Hepatocarcinoma Celular (HCC), siendo este uno de los principales enfoques de esta investigación. Determinar la posición mutada y a qué otra base nitrogenada mutó la secuencia génica es de interés, como también lo es el análisis de las proteínas. Esto quiere decir, enfocarnos también en la secuencia de aminoácidos que se

⁴ Un blanco en biología se refiere a un gen que sirve como objeto de estudio por mostrar un patrón de mutación asociados a ciertas enfermedades.

⁵ El gen p21 es el principal blanco transcripcional del gen de supresión tumoral p53

traduce y la estructura terciaria (aminoácido) que está conforma, ya que es la estructura la que determina la función correcta o incorrecta que tiene la proteína.

Hoy en día estos estudios e investigaciones se están haciendo de forma manual, el proceso de identificación de la mutación y su cadena de aminoácidos está retrasando la toma de decisiones las cuales podría detectar y combatir algún tipo de tumor cancerígeno o actividad tumorigénica. A nivel mundial existen alternativas a este tipo de problemas de identificación, en su mayoría existentes en papel, basados en estudios realizados previamente en pacientes los cuales presentaron una anomalía común en la proteína p53 y fueron diagnosticados posteriormente con un tipo de cáncer en específico.

Para poder realizar estos estudios se necesita un espectrómetro de masas que permite realizar el proceso con mucha precisión y certeza como dice Rosas, N. y Torres, E. (1999), debido a que se especializa en el estudio de secuencias a nivel microscópico independientemente del entorno en que se encuentre. El acceso a esta máquina es limitado, puesto que sólo los mejores laboratorios de Perú y Colombia son quienes la poseen. Al obtener la secuencia de aminoácidos es necesario realizar una serie de procesos químicos en laboratorios de genética y bioquímica para obtener la múltiple alineación de secuencias en la que se establece la región en la que se presenta la mutación (codón o exón) y el nucleótido que cambió.

4.2 FORMULACIÓN DEL PROBLEMA

¿Cómo se podría automatizar el múltiple alineamiento de las secuencias de la proteína p53 y sus mutaciones a fin de detectar la presencia de Hepatocarcinoma Celular (HCC) incluyendo un modelado 3D de la secuencia de esta proteína?

5 MARCO DE REFERENCIA

5.1 MARCO TEÓRICO

5.1.1 HEPATOCARCINOMA CELULAR

El Hepatocarcinoma Celular (HCC) es el tumor maligno primario más conocido en el hígado. Relacionada con la cirrosis hepática, virus de la hepatitis b y c (VHC y VHB). Siendo el más agresivo el VHC, comparándolo con el VHB. Además existe una diferencia de incidencia geográfica, con un mayor porcentaje de individuos afectados en Asia y África. La sintomatología característica es sensibilidad o dolor abdominal, particularmente en el cuadrante superior derecho, agrandamiento del abdomen (ascitis), tendencia al sangrado o a la formación de hematomas y la ictericia (coloración amarillenta de la piel y los ojos).

Plantea Soussi, T. (1994) que los estudios genéticos moleculares han revelado que las alteraciones genéticas de los proto-oncogenes y genes supresores de tumores, por ejemplo p53 entre estos últimos, son de gran importancia en la carcinogénesis humana. De hecho, más de 20 genes implicados en al menos cuatro vías cancerígenas están implicados en el desarrollo de HCC.

La activación de oncogenes de la familia ras y otros ha sido detectada durante HCC inducida químicamente en roedores, pero hay poca evidencia de tal activación en tumores humanos; las alteraciones en oncogenes se han detectado en sólo una pequeña proporción de casos de HCC. Por el contrario, existen pruebas de que los genes supresores de tumores tales como los que codifican p53, la pRb y p16INKa se alteran en diferentes etapas de hepatocarcinogénesis y que podrían causar directa o indirectamente la inestabilidad cromosómica, y promover esta proliferación celular y la neovascularización. Es así que la frecuente pérdida de un alelo del gen supresor de tumores p53, localizado en el cromosoma

17p13.1, y las mutaciones en el alelo restantes reportan que ocurre en diversos tipos de cáncer humano incluyendo HCC.

5.1.2 EL GEN Y LA PROTEÍNA P53

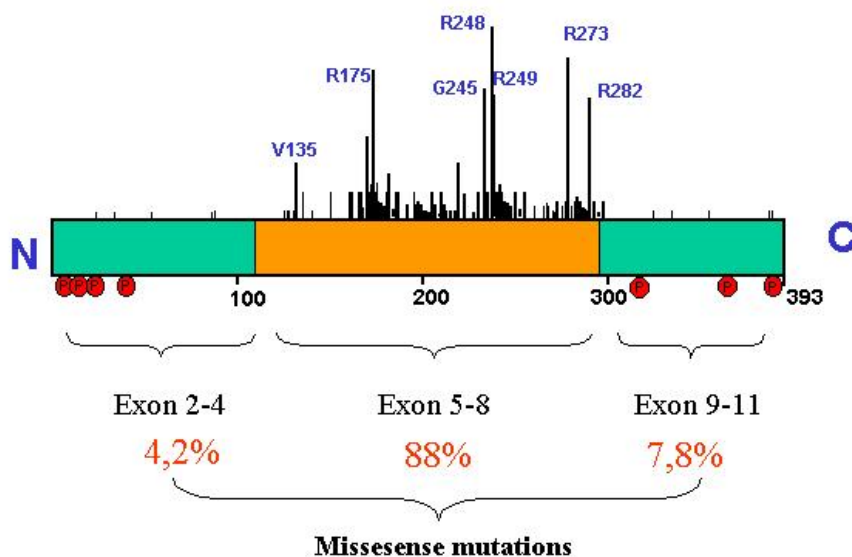
El gen p53 codifica una fosfoproteína nuclear de 393 aminoácidos que se une a secuencias específicas de ADN en el genoma humano. La proteína p53 se ha demostrado que funcionan como una proteína reguladora de la transcripción. Activa la transcripción de varios genes con papeles en el control del ciclo celular, incluyendo GADD45 y los que codifican p21WAF1 / CIP1, MDM2 (un regulador negativo de p53) y 14-3-3 σ (un regulador de la progresión en G2-M). También activa diversos genes que probablemente funcionan en la apoptosis, incluyendo el gen para Bax (un pro-apoptótico, proteína relacionada con Bcl-2) y varios genes que codifican proteínas implicadas en la generación de especies reactivas de oxígeno (ROS).

El gen p53 ha sido implicado en el control del ciclo celular (la detención del ciclo en G1 y G2 para evaluar la integridad genómica), la reparación del ADN y la síntesis, la diferenciación celular, la represión de la transcripción, plasticidad genómica y la apoptosis. Desempeña un papel clave en el reconocimiento y la respuesta al daño del ADN. Cuando el ADN está dañado, la célula expresa un mayor nivel de p53. Esta proteína entonces bloquea la fase G1, lo que permite a la célula para activar los sistemas de reparación del ADN. Cuando el daño es demasiado grande para ser reparado, la sobreexpresión de p53 se activa la apoptosis, la destrucción de la célula.

5.1.3 P53 Y LOS TIPOS DE CÁNCER

La secuenciación del gen p53 en los mamíferos, anfibios, aves y peces ha revelado cinco dominios altamente conservados, cuatro de los cuales caen dentro de los exones 5 a 8: el dominio II (codones 117-142), el dominio III (codones 171-181), dominio IV (codones 234-258), y el dominio V (codones 270-286).

Tal como dice Harris, C. C. (1991), el noventa y ocho por ciento de las 280 mutaciones de sustitución de base cae dentro de una región de 600 pares de base del producto del gen p53, de los codones 110 al 307. Esta secuencia abarca los exones 5 a 8, donde la mayoría de los aminoácidos conservados evolutivamente se concentran.



Mutaciones en p53

Figura 1. Porcentaje de mutaciones por codones agrupadas en tres grupos de exones.

(Castaños, E. 2015).

Las transiciones G: C a A: T constituyen la mayoría de las mutaciones de tumores de colon (79%), y la mayoría de éstas ocurren en dinucleótidos CpG. En los cánceres de mama esporádicos, se encontraron sólo cuatro tumores con transiciones gen p53 en los sitios CpG

(13%). Mutaciones G:C a T:A son la sustitución más frecuente en el cáncer de pulmón de células no pequeñas (nonSCLC). Las transversiones son también excepcionalmente frecuentes entre las mutaciones de p53 en tumores de esófago en comparación con patrones de sustitución de base de la mayoría de otros tipos de cáncer.

5.1.4 ACCIÓN MOLECULAR DE P53

La proteína p53 es capaz de formar tetrámeros que permiten actuar como un dominante negativo. Los alelos mutantes de p53 que producen heterodímeros con p53 wild type, producen un cambio conformacional que impide la unión a los elementos regulados por p53. Por lo tanto, p53 del tipo mutante suprime la actividad de p53 wild type. Y de hecho, ciertos mutantes de p53 missense (mutación sin sentido) pueden obtener "actividad oncogénica". Unos diez años más tarde, a principios de 1990, la proteína p53 fue reconocido como un gen supresor de tumores y el gen más frecuentemente mutado en el cáncer humano con una tasa de mutación de más del 50% en los casos de cáncer humano. La función de p53 en la tumorigénesis después se demostró que p53 también puede actuar como un factor de transcripción implicado en la regulación del ciclo celular y la apoptosis. Esto fue seguido por el descubrimiento de sus múltiples funciones en el desarrollo, la diferenciación, la amplificación de genes, recombinación de ADN, la segregación cromosómica, y la senescencia celular. Una variedad de estudios en los últimos años proporcionó evidencia de que el gen supresor de tumores p53 desempeña un papel importante en hepatocarcinogenesis independientemente de la etiología. Sin embargo, la frecuencia de mutaciones de p53 y su espectro de mutación con 75% mutaciones sin sentido son excepcionalmente diversas en su posición y la naturaleza, que afecta a más de 200 codones dispersos principalmente a lo largo de la porción central del gen.

5.1.5 P53 FUNCIÓN Y DISFUNCIÓN

p53 contiene un dominio de ácido cerca de su extremo N-terminal que es similar a los observados previamente en factores de transcripción bien caracterizados, cuando este dominio de ácido se fusiona a la región de unión a ADN de GAL4, la proteína quimérica resultante podría activar la transcripción de un operón GAL4. Aunque muchas proteínas contienen tales regiones ácidas, la fuerza de la activación y la localización nuclear de p53 sugieren que p53 está implicado en el control transcripcional, ya sea directamente o a través de un complejo con otras proteínas que se unen a genes específicos.

Si p53 se une específicamente al ADN y contiene un dominio ácido de activación en su extremo N-terminal, sería de esperar que p53 pudiera activar la expresión de genes adyacentes a un sitio de unión de p53.

P53 puede unirse formando unidades tetraméricas, por la simetría y las cuatro copias del sitio central 5 de ATGPyPyPyB. Además p53 puede activar directamente la transcripción de sitios de unión de p53. El nivel de activación de la transcripción se correlaciona con precisión con la fuerza de la unión a sitios de unión a p53 in vitro. La transactivación p53 sitio de unión mediada de genes informadores por p53 se puede observar en levaduras así como en células de mamífero. La activación transcripcional a través de sitios de unión a p53 es una característica importante y bioquímicamente evaluables de la proteína p53 normal.

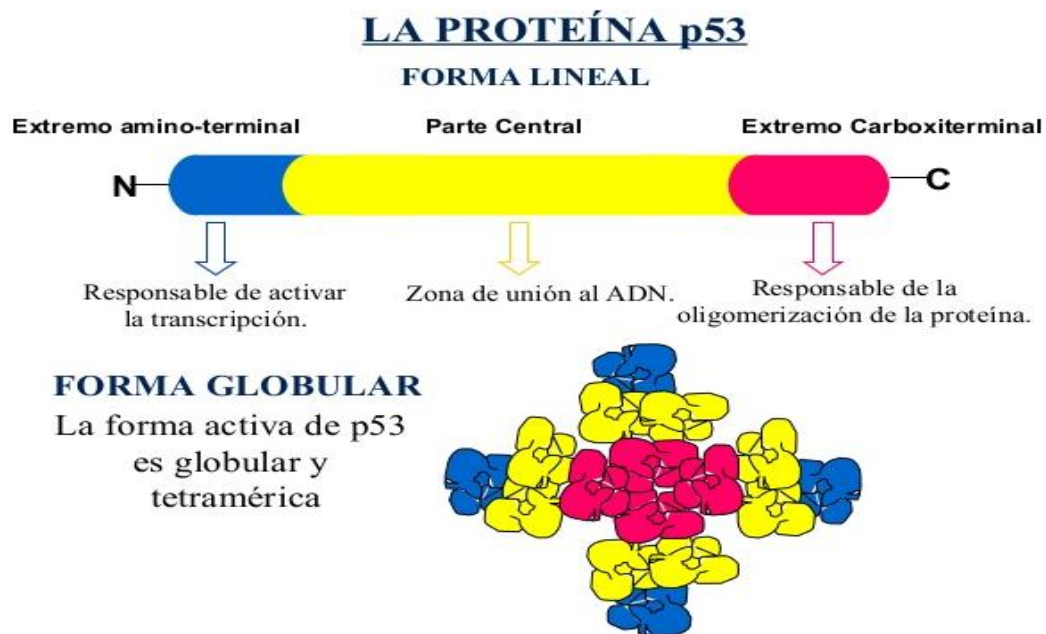


Figura 2. Representación gráfica de la proteína p53 y sus partes. (Asturnatura, 2011).

Un gran número de mutantes de p53 humanos se han descrito, la mayoría se producen como cambios de sentido erróneo en uno de los cuatro hotspots, entre los aminoácidos 129 a 146, 171 a 179, 234 a 260, 270 a 277. Todos los mutantes pierden la capacidad de unirse a sitios de unión de p53 y en consecuencia no pueden activar la expresión de genes adyacentes. Además de perder de manera uniforme la capacidad de unirse a sitios de unión para p53, algunos mutantes parecen cambiar la conformación global de la p53. Algunos mutantes de p53 permiten la unión a proteínas de choque térmico (HSPs), y otros alterar la activación del dominio ácido de manera que no puede funcionar cuando se fusiona al dominio de unión al ADN de GAL4.

Además de mutaciones intragénicas de p53, las alteraciones de otros genes aparentemente pueden llevar a las mismas consecuencias fisiológicas. Los primeros ejemplos de esto fueron proporcionados por oncogenes virales tumorales de ADN, tales como el gen de SV40, el gen E1B de adenovirus, y el gen E8 del virus del papiloma humano. Cada uno de estos genes codifican proteínas que se unen a p53, y en el caso de E8, estos resultados producen una degradación de p53.

5.1.6 ALTERACIONES DE P53 Y LA HEPATOCARCINOGENESIS.

Aproximadamente el 50% de todos los cánceres implican un gen p53 defectuoso, normalmente inactivado por una mutación puntual o delección del gen. Estas alteraciones se cree que previene la oligomerización y formación de los complejos tetraméricos de p53 que se unen a secuencias específicas de ADN, alterando así la función fisiológica de la proteína de tipo salvaje. Las mutaciones genéticas de p53 y la integración del ADN del virus de la hepatitis B (HBV) en el genoma del huésped son los cambios genéticos más frecuentes conocidos en HCC humano. Las alteraciones del gen p53 están presentes en 30-60% de los pacientes con HCC, y un punto de acceso mutación en p53 se ha descrito en pacientes con HCC en áreas de alta exposición a la aflatoxina. Los metabolitos de la aflatoxina B1 promueven sitios purínicos y mutaciones de G a T en el ADN cromosómico, y en el 50% de los pacientes con HCC de las áreas de exposición alta de aflatoxinas se encontraron una transversión (en la tercera posición) en el codón 249 de G a T en p53 (exón 7); por lo tanto, la aflatoxina B1 que contamina los alimentos en las zonas endémicas tiene un papel claro en hepatocarcinogenesis. El wild type P53 es polimórfico en el residuo 72, donde un cambio de una sola base de causa una sustitución de prolina (Pro) para la arginina (Arg) (CCC → CGC) en el dominio de transactivación. También hubo un caso de transversión de G a C en el codón 249.

5.1.7 P53 ESPECTRO DE MUTACIONES EN HCC

Se han observado para 169 muestras de tejido resecado principalmente de pacientes japoneses utilizando análisis del polimorfismo de conformación de cadena sencilla y secuenciación directa. Cuarenta y nueve tumores (29%) mostraron una mutación p53 (39 mutaciones puntuales y 10 frameshifts). Las mutaciones puntuales componen 18 transiciones, sólo 4 de los cuales ocurrieron en los sitios CpG, y 21 transversiones. Dos

dominios conservadas evolutivamente, IV y V, contenían 65% de todas las mutaciones y el codón 249 fue el sitio de la mutación más frecuente.

La pérdida alélica frecuentes sobre específicos brazos cromosómicos (1, 4q, 5q, 13q, 16p, 16q y 17p) indica que la disfunción de diversos genes supresores de tumores localizados en estos brazos cromosómicos está involucrado en el desarrollo de HCC. Entre éstos, el gen p53, localizado en el cromosoma 17p13, ha sido bien analizado y frecuente pérdida de un alelo y las mutaciones en el otro alelo se han informado de que ocurra en diversos tipos de cáncer humano, incluyendo en el HCC.

En el carcinoma de colon y de tipo Burkitt linfoma maligno, 50-65% de las mutaciones están representados por las transiciones de C a T en los sitios CpG. Por otro lado, en el cáncer de pulmón de células no pequeñas, el 57% de las mutaciones puntuales son G a T transversiones, lo que sugiere una forma específica de lesión directa del ADN debido a la presencia de benzo (α) pireno en el tabaco. Un tipo específico de mutación de p53 también se ha demostrado en los HCC de sujetos expuestos a alimentos contaminados con aflatoxina B1, en los que la mayoría de las mutaciones son G a T transversiones que ocurre en el codón 249.

Se ha informado de que la transversión G a T en la tercera base del codón 249 es una característica específica de los HCC endémicas relacionados con aflatoxinas. Se ha encontrado siete mutaciones en el codón 249, en diferentes sitios (dos en la primera posición, tres en la segunda posición y dos en la tercera posición), de diferentes tipos (cinco G a T y dos de A a T), y en diferentes tipos de la infección por el virus de la hepatitis (tres del VHB, tres del VHC, y el último no clasificado).

La inactivación de p53 requiere alteraciones de ambos alelos, lo que indica que un papel recesivo es predominante en la progresión del Hepatocarcinoma Celular en lugar de un papel dominante negativo. Es concebible que las mutaciones en ambos alelos pueden inactivar el gen p53; sin embargo, nunca encontramos estos casos. El gen p53 parece ser inactivado en su mayoría por una combinación de pérdida de un alelo y la mutación en el alelo restante.

5.1.8 SECUENCIACIÓN DE P53

Existen muchos métodos de secuenciación desde que se inició con Sanger hasta las más actuales que son las tecnologías NGS (Next Generation Sequencing) o conocidas como las de alto rendimiento. El concepto detrás de esta tecnología es que unos pequeños fragmentos de ADN son secuencialmente identificados a partir de sus señales emitidas, ya que cada fragmento se resintetiza a partir de una hebra molde de ADN; entonces las NGS extienden este proceso a través de millones de reacciones de una manera masiva en paralelo, en lugar de limitarse a una sola o unos pocos fragmentos de ADN. Este avance permite una rápida secuencia de las grandes extensiones de pares de bases de ADN que abarcan todo el genoma, con esto son capaces de secuenciar cientos de gigabases de datos en una sola corrida.

Entre los métodos de tecnologías NGS más conocidas están:

- La amplificación clonal in vitro, básicamente su uso se debe a que los métodos de detección molecular frecuentemente no son lo suficientemente sensibles para la secuenciación de una sola molécula, y este si obtiene muchas copias de cada molécula individualmente. Uno de los métodos es la PCR de emulsión, en la que se aíslan las moléculas individuales de ADN junto con microesferas recubiertas con cebadores en burbujas acuosas dentro de una fase oleosa. Posteriormente una PCR recubre cada microesfera con copias clonales de la biblioteca de moléculas aisladas y seguidamente se inmovilizan para ser más tarde secuenciadas.
- El método Illumina, para esta técnica hay que considerar que la muestra de ADN genómico (ADNg) se fragmenta en una biblioteca de pequeños segmentos que deben ser uniforme y preciso por secuenciación de millones de reacciones paralelas. Las bases nitrogenadas recién identificados, llamadas lecturas, son luego montadas usando un genoma de referencia conocido como un scaffold (resecuenciación), o en ausencia de un genoma de referencia (secuenciación de novo). El conjunto completo de lecturas alineadas revela la secuencia completa de cada cromosoma en la muestra gDNA.

- La secuenciación por ligación, es otro método enzimático de secuenciación que emplea una ADN ligasa en lugar de una polimerasa para identificar la secuencia objetivo. Este tipo de secuenciación utiliza un reservorio de todos los oligonucleótidos posibles de una longitud específica, marcados de acuerdo con la posición en la que se secuencia. Los oligonucleótidos se hibridan y ligan; el ligamiento preferente de las ADN ligasas por su secuencia específica produce una señal correspondiente con la secuencia complementaria en esa posición específica.
- El pirosecuenciamiento es un método de secuenciación de ADN en tiempo real basado en la liberación de los pirofosfatos (PPi) que tiene lugar en la reacción de polimerización del ADN a partir de sus dNTPs. Frente a otras técnicas de secuenciación, esta variante no requiere correr en un gel los fragmentos de ADN generados en la reacción de polimerización, ni marcadores fluorescentes o ddNTPs (dideoxinucleótidos). Este método requiere de la preparación de una molécula monocatenaria de ADN a la cual se híbrida un pequeño cebador. Igualmente, la pirosecuenciación requiere de los 4 dNTPs, la polimerasa de ADN, así como tres enzimas: sulfurilasa (y el sustrato adenosina 5'- fosfosulfato o APS), luciferasa (y el sustrato luciferina) y una apirasa. A medida que la reacción transcurre, se irá sintetizando la cadena complementaria e iremos obteniendo una serie de picos de señal en el pirograma que nos permitirán determinar la secuencia.

Todos estos métodos son en equipos de alto costo como el 454 GS FLX+ (Roche), HiSeq 2000/2500 (Illumina), 5500xl SOLiD (Life Technologies), PacBio RS (Pacific Biosciences), 454 GS Junior (Roche), Ion Personal Genome Machine (Life Technologies), Ion Proton (Life Technologies) y MiSeq (Illumina).

Para el proceso de secuenciamiento de p53 generalmente se realiza previamente una reacción de secuenciación, como en casi todas los tipos de secuenciamiento, en este caso puede ser realizado por un kit de secuenciamiento BigDye Terminator v1.1 Cycle (Applied Biosystems) realizando un protocolo como el siguiente:

Mix

- 7 μ l de producto de PCR purificado
- 1.25 μ l de Buffer
- 0.5 μ l de primers 10 μ M
- 1.5 μ l de Big Dye

Programación durante 30 ciclos

96 °C 10 sec

50 °C 5 sec

60 °C 4 min

Luego se realiza una purificación de la reacción de secuenciación antes del análisis, para esto la se lleva a cabo por el servicio de secuenciación con placas de filtración multiscreen de 96 pocillos. Finalmente los productos se analizaron mediante un secuenciador automático capilar (ABI PRISM® 3100 Genetic Analyzer, Applied Biosystems), basado en el método de Sanger.

5.1.9 ESTRUCTURA DE LAS PROTEÍNAS

Las proteínas presentan niveles de estructuración, estos son 4:

5.1.9.1 ESTRUCTURA PRIMARIA

La estructura primaria de las proteínas se refiere a la secuencia de aminoácidos, es decir, la combinación lineal de los aminoácidos mediante un tipo de enlace covalente, el enlace peptídico. Los aminoácidos están unidos por enlaces peptídicos siendo una de sus características más importantes la coplanaridad de los radicales constituyentes del enlace.

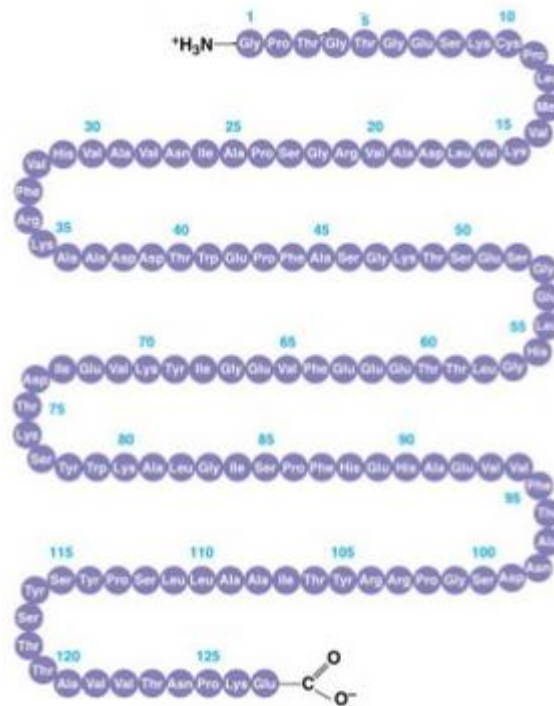


Figura 3. Estructura primaria de una proteína. Muestra la cadena de aminoácidos de manera lineal. (Asturnatura, 2011).

La estructura lineal del péptido definirá en gran medida las propiedades de niveles de organización superiores de la proteína. Este orden es consecuencia de la información del material genético: Cuando se produce la traducción del RNA se obtiene el orden de aminoácidos que van a dar lugar a la proteína. Se puede decir, por tanto, que la estructura primaria de las proteínas no es más que el orden de aminoácidos que la conforman. Los aminoácidos son 20: leucina, prolina, alanina(Ala), valina(Val), metionina(Met), triptófano(Trp), fenilalanina(Phe), isoleucina(Ile), glicina(Gly), serina(Ser), asparagina(Asn), glutamina(Gln), treonina(Thr), tirosina(Tyr), cisteina(Cys), aspartato(Asp), glutamato(Glu), histidina(His), lisina(Lys) y arginina(Arg).

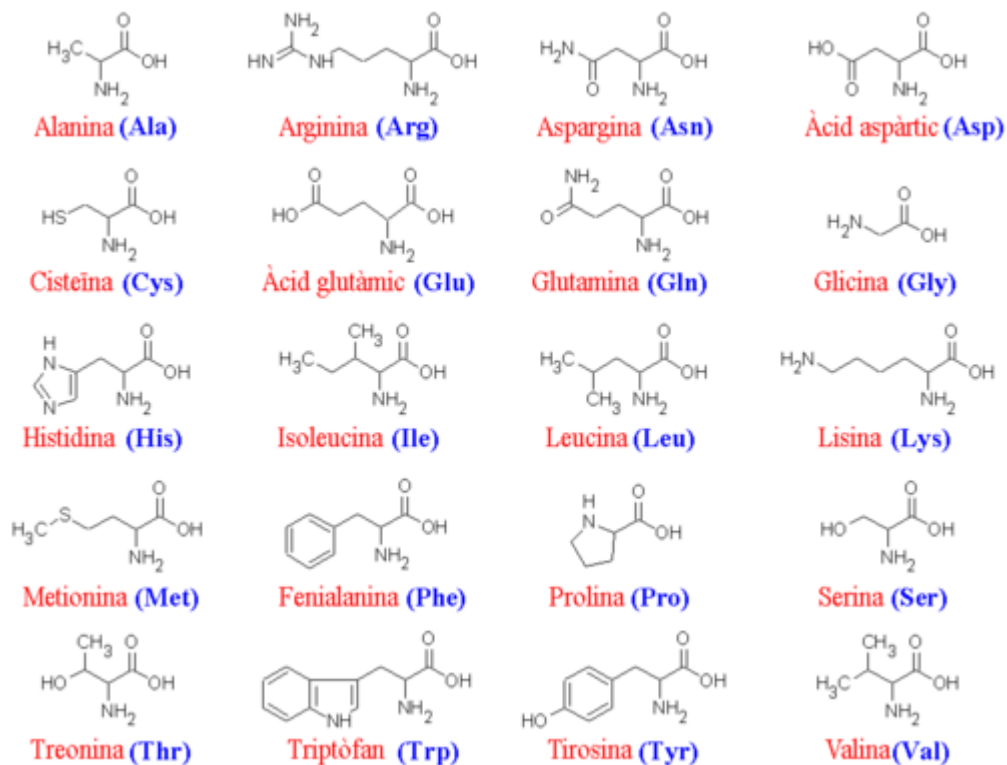


Figura 4. Estructura y composición de los 20 Aminoácidos. (Colegio Glenn Doman, 2011).

5.1.9.2 ESTRUCTURA SECUNDARIA

La estructura secundaria de las proteínas es la disposición espacial local del esqueleto proteico, gracias a la formación de puentes de hidrógeno entre los átomos que forman el enlace peptídico, es decir, un tipo de enlace no covalente, sin hacer referencia a la cadena lateral. Existen diferentes tipos de estructura secundaria: la estructura secundaria ordenada (de forma repetitiva donde se encuentran hélices alfa y cadenas beta, y no repetitivos donde se encuentran los giros beta y comba beta), la estructura secundaria no ordenada y la estructura secundaria desordenada. Los motivos más comunes son la hélice alfa y la lámina beta (Hoja plegada beta).

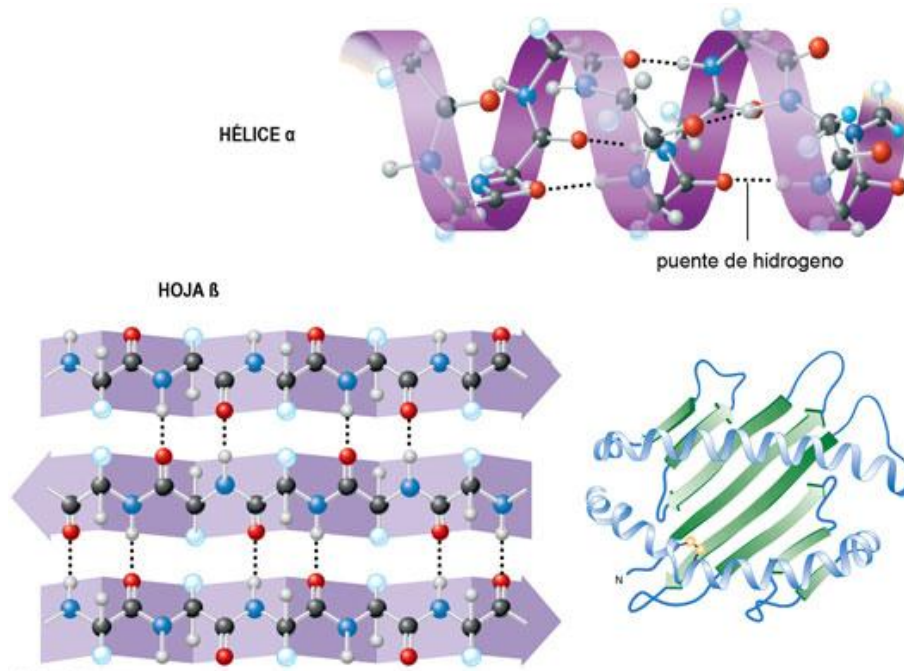


Figura 5. Estructura secundaria de una proteína. Se muestra la hélice α , hoja plegada β y una proteína mostrando los dos tipos de estructura en su cadena. (Asturnatura, 2011).

5.1.9.2.1 HÉLICE ALFA

Los aminoácidos en una α -hélice están dispuestos en una estructura helicoidal dextrógira, con unos 3.6 aminoácidos por vuelta. Cada aminoácido supone un giro de unos 100° en la hélice, y los carbonos α de dos aminoácidos contiguos están separados por 1.5\AA (Armstrongs). La hélice está estrechamente empaquetada, de forma que no hay casi espacio libre dentro de la hélice. Todas las cadenas laterales de los aminoácidos están dispuestas hacia el exterior de la hélice.⁶

El grupo amino del aminoácido (n) puede establecer un enlace de hidrógeno con el grupo carbonilo del aminoácido (n+4). De esta forma, cada aminoácido (n) de la hélice forma dos puentes de hidrógeno con su enlace peptídico y el enlace peptídico del aminoácido en (n+4) y en (n-4). En total son 7 enlaces de hidrógeno por vuelta. Esto estabiliza enormemente la hélice como indican Sippl, M. J., & Melo, F. (2013).

5.1.9.2.2 LÁMINA BETA

La lámina- β se forma por el posicionamiento paralelo de dos cadenas de aminoácidos dentro de la misma proteína, en el que los grupos amino de una de las cadenas forman enlaces de hidrógeno con los grupos carboxilo de la opuesta. Es una estructura muy estable que puede llegar a resultar de una ruptura de los enlaces de hidrógeno durante la formación de la hélice alfa. Las cadenas laterales de esta estructura están posicionadas sobre y bajo el plano de las láminas. Dichos sustituyentes no deben ser muy grandes, ni crear un impedimento estérico, ya que se vería afectada la estructura de la lámina.

5.1.9.3 ESTRUCTURA TERCIARIA

Es el modo en que la cadena polipeptídica se pliega en el espacio, es decir, cómo se enrolla una determinada proteína, ya sea globular o fibrosa. Es la disposición de los dominios en el espacio.

La estructura terciaria se realiza de manera que los aminoácidos apolares se sitúan hacia el interior y los polares hacia el exterior en medios acuosos. Esto provoca una estabilización por interacciones hidrofóbicas, de fuerzas de van der Waals y de puentes disulfuro (covalentes, entre aminoácidos de cisteína convenientemente orientados) y mediante enlaces iónicos.

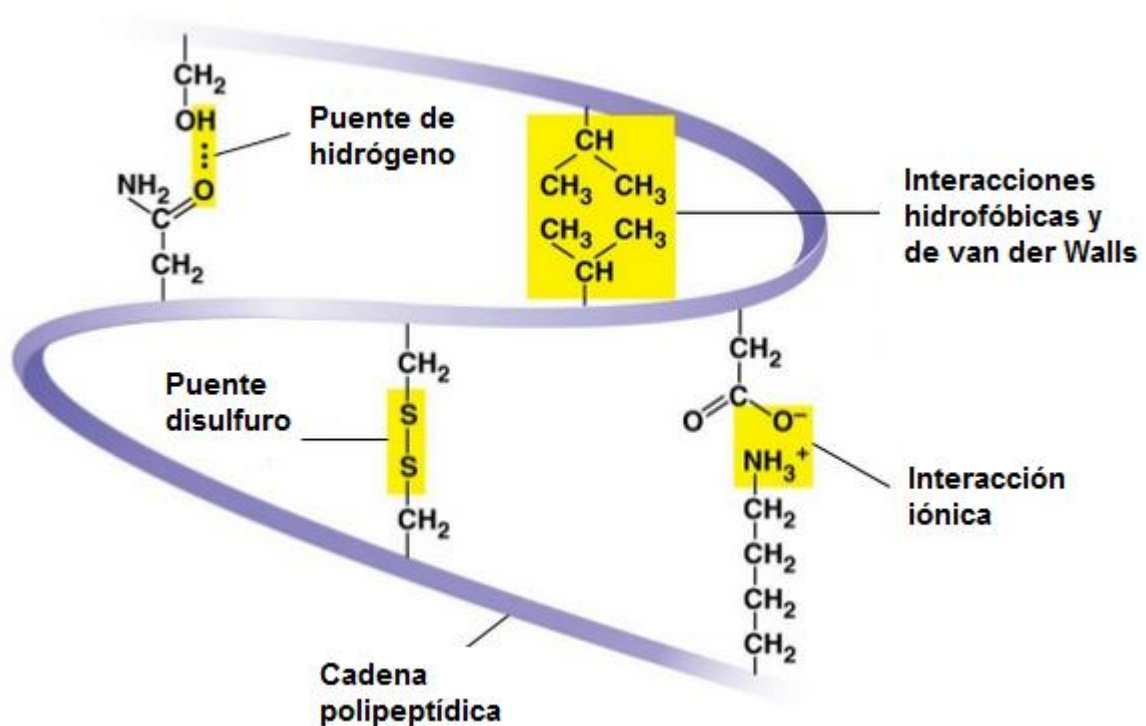


Figura 6. Estructura terciaria de una proteína. Se muestra la cadena polipeptídica, sus interacciones hidrofóbicas y de van der Waals, interacción iónica y sus puentes de hidrógeno y disulfuro. (Asturnatura, 2011).

5.1.9.4 ESTRUCTURA CUATERNARIA

La estructura cuaternaria deriva de la unión de varias cadenas peptídicas que, asociadas, conforman un ente, un multímero, que posee propiedades distintas a la de sus monómeros componentes. Dichas subunidades se asocian entre sí mediante interacciones no covalentes, como pueden ser puentes de hidrógeno, interacciones hidrofóbicas o puentes salinos. Para el caso de una proteína constituida por dos monómeros, un dímero, éste puede ser un homodímero, si los monómeros constituyentes son iguales, o un heterodímero, si no lo son.

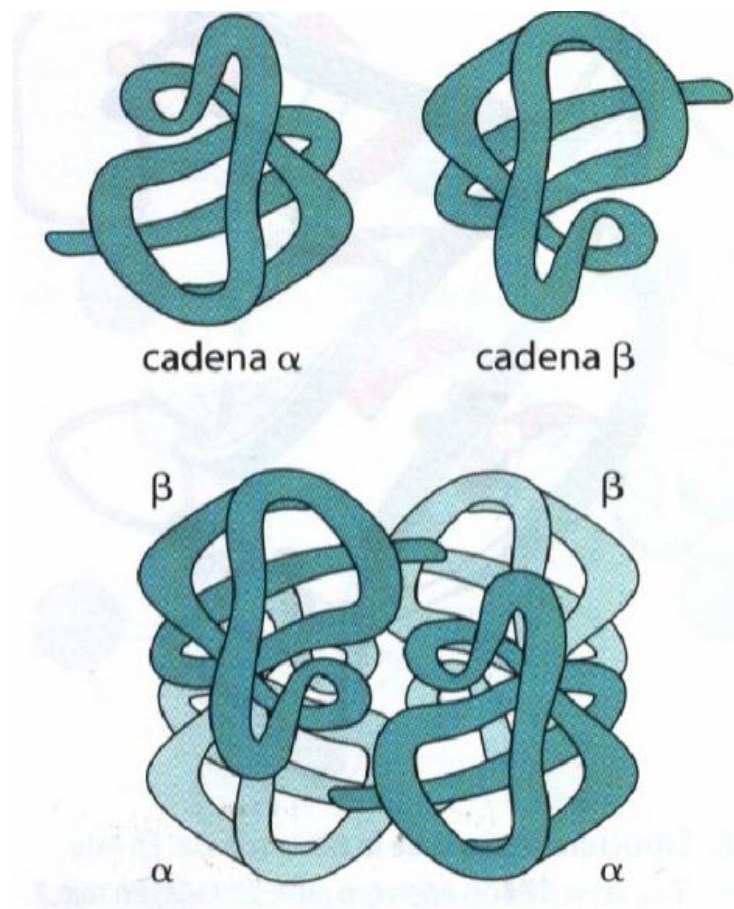


Figura 7. Estructura cuaternaria de una proteína. Esta proteína consta de dos cadenas polipeptídicas α y dos cadenas polipeptídicas β . (Anónimo, 2015).

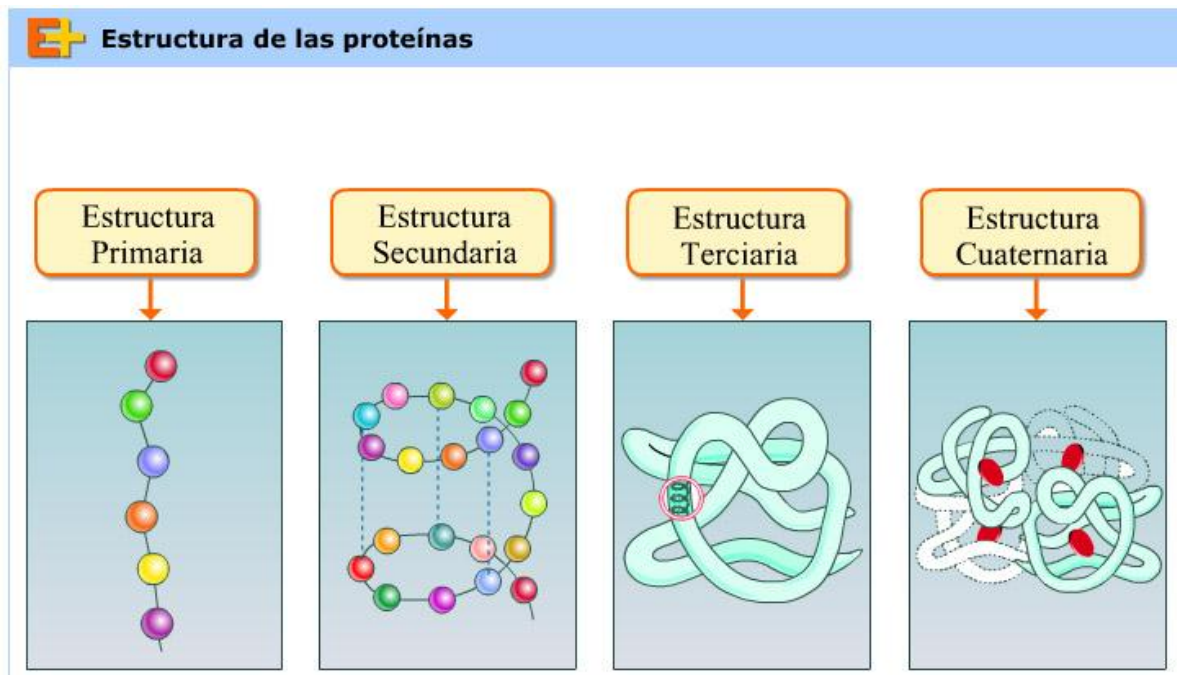


Figura 8. Resumen de estructura de una proteína. (Castaños, E. 2015)

5.1.10 SECUENCIACIÓN DE AMINOÁCIDOS

Uno de los métodos es la degradación de Edman, desarrollada por Pehr Edman, es un método de secuenciación de aminoácidos en un péptido. En este método, el residuo amino terminal se etiqueta y se separa del péptido sin afectar a los enlaces peptídicos entre los otros residuos.

El procedimiento es el siguiente: Marcaje del aminoácido N-terminal con Fenilisotiocianato (PITC) en medio alcalino, posteriormente ocurre una hidrólisis entre los aminoácidos 1 y 2 en medio ácido, el resto péptido se encontrará intacto, para luego identificar del aminoácido N-terminal por HPLC, y finalmente repetición de este proceso hasta culminar con toda la secuencia peptídica.

5.1.11 DETERMINACIÓN DE LA ESTRUCTURA PROTEICA

Existen ciertos métodos para determinar la estructura terciaria de las proteínas:

5.1.11.1 CRISTALOGRAFÍA DE RAYOS X

La cristalografía de rayos X es una técnica experimental para el estudio y análisis de materiales, basada en el fenómeno de difracción de los rayos X por sólidos en estado cristalino.

Los rayos X son difractados por los electrones que rodean los átomos por ser su longitud de onda del mismo orden de magnitud que el radio atómico. El haz de rayos X emergente tras esta interacción contiene información sobre la posición y tipo de átomos encontrados en su camino. Los cristales, gracias a su estructura periódica, dispersan elásticamente los haces de rayos X en ciertas direcciones y los amplifican por interferencia constructiva, originando un patrón de difracción. Existen varios tipos de detectores especiales para observar y medir la intensidad y posición de los rayos X difractados, y su análisis posterior por medios matemáticos permite obtener una representación a escala atómica de los átomos y moléculas del material estudiado

5.1.11.2 RESONANCIA MAGNÉTICA NUCLEAR

La resonancia magnética nuclear (RMN) es un fenómeno físico basado en las propiedades mecánico-cuánticas de los núcleos atómicos. RMN también se refiere a la familia de métodos científicos que explotan este fenómeno para estudiar moléculas (espectroscopia de RMN), macromoléculas (RMN biomolecular), así como tejidos y organismos completos, mostrando imagen por resonancia magnética.

La espectroscopia de RMN es una de las principales técnicas empleadas para obtener información física, química, electrónica y estructural sobre moléculas. Es una poderosa serie de metodologías que proveen información sobre la topología, dinámica y estructura tridimensional de moléculas en solución y en estado sólido.

5.1.11.3 ESPECTROMETRÍA DE MASAS

La espectrometría de masas (ES) es una técnica de análisis que permite determinar la distribución de las moléculas de una sustancia en función de su masa. El espectrómetro de masas es un dispositivo que permite analizar con gran precisión la composición de diferentes elementos químicos e isótopos atómicos, separando los núcleos atómicos en función de su relación carga-masa (z/m). Puede utilizarse para identificar los diferentes elementos químicos que forman un compuesto, o para determinar el contenido isotópico de diferentes elementos en un mismo compuesto. Con frecuencia se encuentra como detector de un cromatógrafo de gases, en una técnica híbrida conocida por sus iniciales en inglés, GC-MS.

El espectrómetro de masas mide razones carga/masa de iones, calentando un haz de material del compuesto a analizar hasta vaporizar e ionizar los diferentes átomos. El haz de iones produce un patrón específico en el detector, que permite analizar el compuesto.

5.1.12 ALGORITMO DE ALINEAMIENTO UNIÓN DE VECINOS (NEIGHBOR-JOINING)

Naruya, S. y Masatoshi, N (1987) fueron los creadores del algoritmo de unión de vecinos (Neighbor-joining) el cual está enfocado en la reconstrucción de árboles filogenéticos desde datos evolucionarios. El principio de este método es encontrar pares unidades taxonómicamente operacionales (vecinos) que minimicen el número total de ramas de cada área de alineamiento. El tamaño de la rama, al igual que la topología de un árbol parsimonioso, pueden ser obtenidos rápidamente a través de este algoritmo. Naruya y Masatoshi usaron una simulación de computador para estudiar la eficiencia de este método en la obtención del correcto árbol sin raíz en comparación con otros cinco métodos de construcción de árboles: the unweighted pair group method of analysis, Farris's method, Sattath and Tversky's method, Li's method, y Tatenó et al's modified Farris method. El algoritmo de Neighbor-joining al igual que el algoritmo de Sattath and Tversky's fueron generalmente mejores que los otros algoritmos.

5.1.13 HERRAMIENTAS PARA EL MODELADO 3D

El modelado 3D es la representación esquemática de un mundo conceptual de tres dimensiones, visible a través de un conjunto de objetos, elementos y propiedades que, una vez procesados, se convertirán en una imagen y animación en 3D. Este conjunto de características suele estar formado por objetos poligonales, tonalidades, texturas, sombras, reflejos, transparencias, translucidez, refacciones, iluminación (directa, indirecta y global), etc. Algunas de las herramientas de modelado disponibles son:

Anim8or: Es un programa de modelado en 3D gratuito desarrollado por R. Steven Glanville, donde se puede crear y editar objetos, figuras y escenas tridimensionales, partiendo de las formas más básicas. El programa tiene además soporte para fuentes TrueType, realiza operaciones en tiempo real basadas en OpenGL y es capaz de importar ficheros 3DS (3D

Studio), LWO (Lightwave) y OBJ(Wavefront), crear animaciones y escenas 3D exportando a vídeo AVI e imagen JPG o BMP, trabajar con texturas, sombras, focos de luz y mucho más (Glanville, 2011).

3ds Max: Software de modelado que proporciona una solución completa de modelado, animación, simulación y renderización a los creadores de juegos, cine y gráficos de movimiento.

Blender: Es un programa informático multiplataforma, dedicado especialmente al modelado, animación y creación de gráficos tridimensionales (Foundation, 2013). Blender es una herramienta que realiza los procesos de modelado 3D, texturizado, animaciones y render de una manera óptima. Además, Blender es un software libre, por tanto es gratis y permite una facilidad de soporte y comunidad de desarrolladores.

Unity: Es un motor de videojuegos que permite crear entornos interactivos, ya sean videojuegos o aplicaciones con diferentes fines (Unity Technologies, 2016). Unity permite agregar funcionalidad por medio del lenguaje de programación C, C++ o C# a modelos 3D creados anteriormente.

Jmol: Visor de Java código abierto para estructuras químicas en tres dimensiones. Jmol es una herramienta muy útil para el modelado de secuencias y contiene complementos muy importantes para la estructura de una proteína.

Biojava: Es un proyecto de código abierto dedicado a proveer un framework Java para el proceso de información biológica. Biojava contiene módulos para la instalación de modelos de proteína, permite cargar estructuras y modificarlas, permite hacer alineamiento de estructuras y secuencias y visualiza las estructuras en un entorno 3D.

5.1.14 ENTORNO INTEGRADO DE DESARROLLO (IDE, INTEGRATED DEVELOPMENT ENVIRONMENT)

Es un entorno de programación que integra un conjunto de herramientas que facilita el trabajo del desarrollador de software, incorporando sólidamente la edición orientada al lenguaje, la compilación o interpretación, la depuración, las medidas de rendimiento, la incorporación de los fuentes a un sistema de control de fuentes, etc., normalmente de forma modular.

Eclipse (Murphy, Kersten, & Findlater, 2006): Entorno de desarrollo integrado de código abierto multiplataforma para desarrollar lo que el proyecto llama "Aplicaciones de Cliente Enriquecido" y se integra fácilmente con los SDK de AR a utilizar.

NetBeans: es un entorno de desarrollo integrado libre, hecho principalmente para el lenguaje de programación Java. Existe además un número importante de módulos para extenderlo. NetBeans IDE1 es un producto libre y gratuito sin restricciones de uso.

Se usó NetBeans debido a su versatilidad en el desarrollo de aplicaciones de escritorio, facilidad de soporte, comunidad de desarrolladores para ayudas técnicas y por su condición de software libre.

5.2 ESTADO DEL ARTE

5.2.1 RELACIÓN ENTRE LA PROTEÍNA P53 Y CASOS DE HEPATOCARCINOMA CELULAR

Torres E y Rosas N (1999c) realizaron un estudio en Lima-Perú llamado: Asociación entre la proteína p53 mutada, grado de infiltración y tamaño del tumor en cáncer colorrectal; se pretendía determinar una mutación en la proteína p53 de pacientes con cáncer colorrectal y relacionarla con el tamaño del tumor y su grado de infiltración. Este estudio fue realizado con 40 casos diagnosticados como carcinoma colorrectal de los registros del Hospital Central FAP en los años 1987 – 1996. Se empleó el método inmunohistoquímico estreptavidina-biotina-peroxidasa con el anticuerpo do-7 y con pretratamiento con microondas para determinar la mutación de la proteína p53.

La máxima incidencia del carcinoma colorrectal se produce entre los 60 y 70 años, y menos del 20% de los casos se presentan en personas menores de 50 años de edad. En los tumores rectales, la proporción varón/mujer es de 2:1. En tumores más proximales no existen diferencia entre sexo. Los países con mayor incidencia de este caso de carcinoma son EE.UU., Canadá, Australia, Suecia y otros países desarrollados. Su frecuencia es significativamente inferior en Japón, América del sur y África.

En el 70-80% de los carcinomas de colon se han encontrado deleciones en la región que contiene el gen supresor p53, el cual está localizado en 17p 13.1. El gen p53 es el objetivo más frecuente de las alteraciones genéticas en el cáncer humano según Chang (1993) citado por Torres (1999a). La proteína p53 tiene el papel de suprimir el crecimiento de tumores, evitando la propagación de células con daños genéticos, convirtiéndola en un blanco fácil para los carcinógenos durante la transformación neoplásica.

En los 40 casos seleccionados se hizo una evaluación de la expresión de la proteína p53 mutada, se hizo un corte histológico de 4µm y se recogió en una lámina portaobjetos con

adherente (poli-lisina) para evitar el desprendimiento del tejido en el proceso de inmunotinción. Luego de desparafinar e hidratar el corte histológico, se procedió a la recuperación antigénica en un horno microondas por 10 minutos a una temperatura estandarizada para la visualización final de la inmunotinción. Luego se lleva a cabo el proceso de inmunotinción agregando al corte histológico un anticuerpo primario comercial, un segundo anticuerpo marcado con moléculas de biotina para reconocer el primer anticuerpo como antígeno, y moléculas de Streptavidin marcadas con enzima peroxidasa quienes reconocen a las moléculas de biotina como moléculas afines. Por último se agrega un sustrato cromógeno para obtener una visualización de color marrón dorado en la mutación de la proteína.

Los resultados de esta investigación indican que el tamaño del tumor en los casos de cáncer de colon en la población estudiada se relaciona directamente con la proteína p53, luego de comparar los grados de color del corte histológico con el tamaño del tumor y el porcentaje de mutación de la proteína. Además los resultados indican un bajo grado de diferenciación en la proteína p53 mutada, de tal manera que no se presentó relación entre estas, lo cual es similar al trabajo realizado por Poller (1999b).

5.2.2 MÚLTIPLE ALINEAMIENTO DE SECUENCIAS DE AMINOÁCIDOS USANDO CLUSTAL

Chenna R, Sugawaral H, Koike1 T, Lopez R, Gibson T, Higgins D y. Thompson J (2003) exponen el uso de series de programas Clustal para el alineamiento de aminoácidos entre secuencias ácidos nucleicas y proteínas. El enfoque de esta investigación es exponer la importancia del método Clustal y su popularidad en la práctica de estudios bioquímicos relacionados con secuencias de aminoácidos, al igual que describir el funcionamiento de este método.

Uno de los pilares de la informática moderna es el alineamiento de secuencias de proteínas, el cual habilita a biólogos a estudiar el alineamiento de secuencias conservadas a

través de la evolución y la relación ancestral entre diferentes organismos. El múltiple alineamiento se construye progresivamente por una serie de parejas alineadas, siguiendo el orden en su árbol de guía donde la comparación inicial usa un rápido algoritmo de alineamiento basado en palabras y a partir de este se construye un árbol guía.

Los programas más usados a nivel mundial para el múltiple alineamiento de secuencias provienen de series de programas Clustal, donde Des Higgins (1988) escribió y diseñó el primer programa Clustal el cual contiene un algoritmo de programación dinámico-eficiente y usa estrategias de alineamiento progresivo desarrollado por Feng y Doolittle (1987). ClustalV, segunda generación de Clustal, nace en 1992 cuando se incorpora perfiles de alineamiento (alineamientos de los alineamientos existentes) y la facilidad para generar árboles de múltiples alineamientos usando el algoritmo llamado neighbor joining. Dos años después nace la tercera generación de Clustal, Llamada ClustalW, quien implementó una serie de mejoras al algoritmo, como lo es el tamaño de las secuencias y la posición específica de las mutaciones. Y la última generación es ClustalX quien mejoró la experiencia del usuario al usar el programa, agregando ventanas con scrollbars y todos los parámetros son visibles a través del menú.



Figura 9. Múltiple alineamiento de cuatro oxidoreductas unido a una secuencia proteínica dominante en ClustalW. Los cambios de colores representan el rango de residuo. Colegio Glenn Doman (2011).

Las series de Clustal fueron desarrolladas para ofrecer robustez y portabilidad de programas que biológicamente precisen alineamientos dentro de un razonable límite de tiempo. ClustalW y ClustalX son actualizados periódicamente para permanecer con las características principales por el cual fueron desarrollados. La razón de su portabilidad es atribuida a la facilidad de guardar el resultado de múltiples alineamientos en formato FASTA, el cual es un formato universal de secuencias de aminoácidos representados por una cadena lineal de caracteres.

Se puede concluir de esta investigación que la serie de programas Clustal posee algoritmos de alineamientos muy óptimos para los trabajos realizados por biólogos, sin embargo permanece en constante mejora, lo cual nos dice que es posible mejorar la experiencia visual de los resultados obtenidos en los formatos FASTA llevándolo a un modelo 3D.

Es necesario resaltar que no se han encontrado en Colombia, evidencias de estudios realizados para la automatización de secuencias de aminoácidos ni existe registro de un software desarrollado que permita realizar este proceso.

6 JUSTIFICACIÓN

La detección de Hepatocarcinoma Celular (HCC) implica una mutación de la secuencia de la proteína p53, siendo esta un indicio de presencia tumorigénica en el cuerpo de quien se realiza el análisis de la proteína. Es de interés saber qué tipo de mutación presenta una persona, para así hallar un patrón y prevenir posibles futuros casos de tumores cancerígenos. Un estudio de la proteína p53 que permita encontrar los diferentes cambios que se presentan en esta misma, es de utilidad para diagnosticar diferentes tipos de cáncer. Evaluar casos que presenten similitudes tanto en mutaciones como en diagnósticos hace pensar que existe un patrón identificable para las personas que son propensas a padecer un caso específico de cáncer, y lo hay. Pero para llegar a estas conclusiones se necesita evaluar los distintos casos presentados anteriormente lo cual conlleva un extenso y cuidadoso análisis investigativo.

El análisis investigativo hablado anteriormente se realiza por medio de una serie de procesos químicos en laboratorios de genética y bioquímica. Esto implica analizar la proteína con un espectrómetro de masas para hallar la secuencia proteómica y luego se procede a identificar las mutaciones de cada aminoácido por medio de alineamiento de secuencias. Este procedimiento requiere laboratorios con instrumentos muy costosos – el costo de un espectrómetro de masas puede llegar hasta unos \$250.000 USD MedWOW (2014) - y no muy comunes en países como Colombia o Perú, dificultando así el estudio de este tipo de investigaciones en los países mencionados.

Estudios realizados anteriormente, como indica Naruya, S. y Masatoshi, N (1987), hablan de la posibilidad de software que permiten realizar el proceso de múltiple alineamiento de secuencias entre proteínas. Así como explica Waterhouse, A.M. (2009) este método implementa la automatización del proceso de alineamiento, de tal forma que se disminuyen los costos de las investigaciones de manera significativa (debido a la definición de software libre con fines investigativos) y es viable económicamente porque se desarrolló el software a partir de los conocimientos de los coinvestigadores y librerías con

licenciamiento libre, no hubieron costos adicionales durante la investigación y el desarrollo del software. Además, esta investigación incentivó el estudio de identificación de casos de Hepatocarcinoma Celular tanto para biólogos como para las ciencias de la computación en países como Colombia y Perú.

El modelado 3D de la proteína p53 es de gran ayuda para la toma de decisiones en el diagnóstico de casos de Hepatocarcinoma Celular, debido a que permite observar gráficamente el estado en que se encuentra cada aminoácido de la proteína y muestra de manera rápida y fácil cómo se vería una mutación en la proteína.

Se desarrolló un software que puede realizar un alineamiento de secuencias de la proteína p53 y su mutación, que además señala los puntos importantes de la mutación y el comportamiento de los aminoácidos que componen la proteína en un modelo 3D. De esta manera se espera facilitar futuros estudios del Hepatocarcinoma Celular desde el ámbito investigativo hasta la parte económica, teniendo en cuenta que también nos permite ahorrar un recurso tan importante como lo es el tiempo, ya que un sistema que integre alineamiento de secuencias y modelado 3D ahorraría el tiempo invertido en ambas fases, agilizando la investigación y los diagnósticos encontrados.

Se espera que al haber elaborado este software haya un aumento del estudio en áreas como la farmacogenómica y proteómica, poniéndole interés a la automatización de procesos de análisis por medio de software.

7 OBJETIVOS

7.1 OBJETIVO GENERAL

Desarrollar un software que permita analizar mutaciones de la proteína p53 e identifique casos de Hepatocarcinoma Celular (HCC), mediante comparaciones entre la secuencia de la proteína p53 y casos evidenciados de Hepatocarcinoma, a fin de automatizar el proceso en el estudio de tumorigénesis.

7.2 OBJETIVOS ESPECÍFICOS

- Desarrollar un software que identifique las mutaciones a partir de la información obtenida en el análisis de la secuencia de la proteína p53.
- Clasificar la mutación hallada con respecto a la región a la cual pertenece y a la secuencia de bases nitrogenadas que la compongan.
- Predecir la estructura 3D de la proteína p53 mutada.
- Reconocer el posible tipo de Hepatocarcinoma Celular que puede estar relacionado con la mutación hallada en la proteína p53.

8 ALCANCE

El alcance de este proyecto se delimitó por los siguientes aspectos:

- El horizonte temporal marcado por el estudio y la investigación corresponde a lo establecido en el diagrama de actividades extendiéndose aproximadamente seis (6) meses.
- El alcance espacial del estudio se limita a investigaciones realizadas en los países de Perú y Colombia. En Perú, específicamente a la Universidad Nacional Mayor de San Marcos, en su facultad de ciencias y biología. En lo que a Colombia respecta, específicamente a la Universidad de Cartagena.
- Por tratarse de una investigación de tipo bibliográfica, aplicada y experimental. Basada en investigaciones previas y con información sólida y eficaz, se espera obtener resultados que contribuyan a la comunidad científica a seguir descubriendo nuevas formas experimentales de continuar con este proyecto.
- El objetivo principal del proyecto es desarrollar un software que identifique tipos de hepatocarcinoma celular (HCC) ya identificados a fin de automatizar el proceso en el estudio de tumorigenesis. El software hará modelado 3D de la proteína estudiada. El proyecto no identificara nuevos hepatocarcinomas celulares, solo reporta los casos detectados de HCC, sin embargo también reporta los casos como “caso desconocido” cuando la mutación presentada no tiene coincidencia en los registros de la base de datos. El proyecto no diagnosticara a personas de enfermedades diferentes al hepatocarcinoma celular. el software no editará modelos 3D después de creados, solo creara modelos 3D a partir de una secuencia dada.
- el código fuente del software a entregar se hará de código abierto para ayudar a la comunidad científica en futuras investigaciones referentes al tema.

9 METODOLOGÍA

9.1 TIPO DE INVESTIGACIÓN

Con el fin de alcanzar el completo desarrollo de los objetivos planteados en este proyecto, el proyecto se dividió en distintas clasificaciones para entender el tipo de naturaleza de su investigación y definir la metodología que permitió alcanzar dicho propósito, el proyecto se enmarca en las siguientes clasificaciones:

- Investigación bibliográfica: Debido a la importancia de otras investigaciones relacionadas con el estudio de la proteína p53 en la identificación de Hepatocarcinoma Celular.
- Investigación aplicada: Debido al perfil indagatorio que lleva a este proyecto a la estructuración de una herramienta para la detección de Hepatocarcinoma Celular con base a la proteína p53.
- Investigación experimental: A razón de que es necesario investigar la estructura gráfica de cada base nitrogenada que conforma la proteína para su modelado 3D.

Para obtener la información necesaria del estudio de la proteína p53 y el proceso de alineamiento de las mutaciones se realizaron veinte (20) reuniones con profesores e investigadores biotecnológicos que se encuentran vinculados a la Universidad de Cartagena y a la Universidad Nacional Mayor de San Marcos. Cinco (5) reuniones presenciales y cuatro (4) reuniones virtuales a través de la herramienta Skype con el investigador Miguel Alcalde Alvites estudiante de genética y biotecnología de la Universidad Nacional Mayor de San Marcos y once (11) reuniones presenciales con el ingeniero de sistemas Luis Carlos Tovar Garrido profesor de la Universidad de Cartagena, además de dar un respaldo y un aval del proceso investigativo proteínico, genético y bioquímico.

Para alcanzar los objetivos planteados se determinó utilizar la metodología de desarrollo RUP (Rational Unified Process) cumpliendo con las fases de dicha metodología y haciendo el uso de UML (Lenguaje Unificado de Modelado)

9.2 FASES DEL DISEÑO Y DESARROLLO

Se utilizó RUP como metodología de desarrollo. Su avance tuvo un proceso iterativo por fases e internamente por objetivos, de tal manera que pudimos controlar a nivel general y específicamente toda la evolución del software y la investigación.

9.2.1 FASE INICIAL

El objetivo principal de esta fase fue establecer los límites del sistema de análisis y modelado 3D de la proteína p53. Inicialmente se realizaron reuniones para establecer el flujo y comportamiento del sistema, donde se planteó la estructura de los archivos FASTA que se ingresan al programa, cómo esta secuencia iba a ser mostrada al usuario final y cómo debería ser visualizado el modelo 3D para optimizar la experiencia del usuario con el sistema. También se definió un modelo de cómo se vería la interfaz de usuario y sus interacciones con el sistema. Por último se concretó la estructura del modelo 3D general y como se resaltaría la presencia de una mutación específica en p53. Con la información recogida en dichas reuniones se diseñó los artefactos tales como: modelo de requisitos, modelo de dominio, diagrama de casos de uso, descripción de los casos de uso y diagrama de actividades que se encuentran más adelante en los resultados de esta investigación. En segunda instancia, en la fase de prueba, se diseñarán simples pruebas técnicas que podrán ser realizadas por uno de nuestros usuarios finales para conocer su grado de satisfacción al usar el sistema.

9.2.2 FASE DE ELABORACIÓN

Desde este punto el enfoque se rige a través del desarrollo de la arquitectura del sistema basada en la obtención y gestión de los requerimientos, tanto funcionales como no funcionales. A través de la gestión de los requerimientos se hizo la recolección de información, detección de inconsistencias, validación, los cambios y vinculación de estos dentro del sistema de información (GUILLERM, DEMMOU, & SADOU, 2008).

Desde esta fase y hasta la fase de construcción se constroyeron los artefactos tales como diagrama de casos de uso, diagramas de actividades y diagrama de clases que se encuentran más adelante en los resultados de esta investigación. Para la descripción de tal arquitectura se tuvo en cuenta los diferentes modelos o vistas (estática, dinámica, funcional) en las cuales de incluyeron los diferentes artefactos de la Ingeniería de Software.

9.2.3 FASE DE CONSTRUCCIÓN

En esta fase se comenzó a desarrollar el sistema para el análisis y modelado 3D de la proteína p53. El proceso de desarrollo estuvo ligado a los objetivos específicos de esta investigación, considerando cada objetivo como una etapa clave durante el proceso y evaluando y desarrollando cada objetivo en tiempos diferentes. El proceso de desarrollo de cada objetivo fue así:

Desarrollar un software que identifique las mutaciones a partir de la información obtenida en el análisis de la secuencia de la proteína p53.

Se desarrolló un software en lenguaje Java con código abierto (para futuras investigaciones relacionadas) donde el usuario puede primeramente ver la secuencia de bases nitrogenadas de p53 y luego ingresar al software un archivo de texto simple en formato FASTA el cual contiene una fracción de la secuencia de una p53 mutada y se es almacenada en el software. Posteriormente el sistema busca las diferencias más notorias de bases

nitrogenadas entre ambas secuencias para encontrar las posibilidades de cambio o mutación en la proteína para su posterior estudio.

Clasificar la mutación hallada con respecto a la región a la cual pertenece y a la secuencia de bases nitrogenadas que la compongan.

En este punto, y con la información recibida por el software en la actividad anterior, el sistema procede a realizar un alineamiento de la mutación puntual gracias a un algoritmo de ordenamiento de secuencias que se implementó. Este alineamiento se realiza con base a los exones y codones de la proteína p53 y así se puede validar la mutación, la secuencia de bases nitrogenadas que la compone y la posición en la que se encuentra el aminoácido mutado.

Predecir la estructura 3D de la proteína p53 mutada

Durante el proceso de desarrollo se hizo un análisis de los requerimientos funcionales y no funcionales del sistema, y se planteó en principio realizar el modelado 3D con la herramienta Blender y posteriormente realizar el módulo 3D del software con Unity. Al seguir avanzando con el desarrollo, se determinó que Jmol era una herramienta que cumplía los principales objetivos de esta investigación, además que fue hecho en java permite hacer una integración fácil y cómoda con el software. Tras realizar el análisis de la secuencia de aminoácidos, el software permite la opción de exportar la secuencia mutada y posteriormente ejecutar un visualizador del modelo 3D de la proteína.

Reconocer el posible tipo de Hepatocarcinoma Celular que puede estar relacionado con la mutación hallada en la proteína p53

Por medio de estudios de los casos de Hepatocarcinoma Celular relacionados directamente con la proteína p53 hechos por otros investigadores, se realizó un estimado del posible tipo de Hepatocarcinoma Celular presentado en el gen ingresado anteriormente. Para esto existe una base de datos global donde investigadores relacionan mutaciones con tipos de canceres desarrollados, llamada UMDTP53 en la versión actualizada en 2012. Se utilizó la base de datos UMDTP53 como base para el desarrollo de este objetivo, como sustentan Hainaut, P., Soussi, T., Hollstein, M., Greenblatt, M., Hovig, E., Harris, C., Montesano, R. (1997)

9.2.4 FASE DE TRANSICIÓN

En esta fase se realizaron pruebas del software en conjunto con participantes de esta investigación. De esta manera se llevó a cabo el proceso de retroalimentación, donde se busca añadirle mejoras o corregir problemas con el software, dependiendo del resultado de las pruebas. Al culminar esta fase se realizó la documentación y el software terminó su fase final en el desarrollo.

10 RESULTADOS Y DISCUSIÓN

10.1 FASE INICIAL

Inicialmente se hizo necesario dos reuniones presenciales, en un salón de la Universidad de Cartagena, entre los co-investigadores Daniel Andrés Orozco Méndez y Javier David Castillo Beltrán, del grupo de investigación GIMATICA, en conjunto de los investigadores Luis Carlos Tovar Garrido y Miguel Alcalde Alvites, donde se propuso el inicio de este proyecto. En esta reunión se explicó el objetivo, alcance, y el resultado esperado y así mismo se establecieron los puntos clave del software, requerimientos funcionales y no funcionales, y las limitaciones que tiene el software durante el proceso del análisis y modelado 3D.

10.2 FASE DE ELABORACIÓN

Atendiendo a las indicaciones de la segunda fase del modelo RUP, se inició con la elaboración de los modelos para identificar los artefactos de ingeniería de software que permiten el análisis y de p53 y el modelado 3D. Además de la búsqueda de la base de datos con información relevante a casos de hepatocarcinoma celular y sus mutaciones en p53 y la identificación de los factores importantes de estos estudios para el resultado esperado en el software elaborado.

Con base a los estudios de la información mostrada en la base de datos de los casos de hepatocarcinoma celular y sus mutaciones en p53 realizados por Hainaut, P., Soussi, T., Hollstein, M., Greenblatt, M., Hovig, E., Harris, C., Montesano, R. (1997) podemos definir como factores importantes los siguientes factores:

A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	143	GTG to GCG	428	T to C	Cx3	Colon	1				Val->Ala		
2	175	CGC to CAC	524	G to A	Cx1	Colon	1			yes	Arg->His		
3	132	AAG to CAG	394	A to C	BT 20	Breast	2		L		Lys->Gln		
4	249	AGG to AGC	747	G to C	BT 549	Breast	2		L		Arg->Ser		
5	280	AGA to AAA	839	G to A	MDA231	Breast	2		L		Arg->Lys		
6	285	GAG to AAG	853	G to A	BT 474	Breast	2		L		Glu->Lys		
7	157	GTC to TTC	469	G to T	OZ1	HCC	3				Val->Phe		
8	249	AGG to AGT	747	G to T	OZ2	HCC	3				Arg->Ser		
9	249	AGG to AGT	747	G to T	OZ3	HCC	3				Arg->Ser		
10	249	AGG to AGT	747	G to T	OZ4	HCC	3				Arg->Ser		

Figura 10. Sección de la base de datos de Excel como ejemplo. Hainaut, P (1997)

- **Posición de la mutación:** Existen 394 aminoácidos donde cada uno tiene 3 bases nitrogenadas en la cadena original de aminoácidos de la proteína p53. Este factor determina la posición, entre todas las bases nitrogenadas de la cadena de aminoácidos, donde se presenta una mutación.
- **Exón:** La proteína p53 está conformada por 11 exones (del 1 al 11) los cuales son la forma en que se distribuyen los 394 aminoácidos que componen a la proteína p53. En otras palabras, este factor determina la región en la que se encuentra el aminoácido donde se presenta una mutación.
- **Codón:** El codón es la posición del aminoácido, en un rango de 394 posibilidades, donde se presenta la mutación.
- **Codón normal:** Este factor es la cadena de tres bases nitrogenadas que representa el codón o aminoácido donde se encuentra una mutación.
- **Codón mutado:** Dependiendo de la mutación, este factor contiene la cadena de tres bases nitrogenadas resultantes después de la mutación en el codón o aminoácido presentado anteriormente.

- **Aminoácido normal:** Este factor es el nombre del codón o aminoácido donde se encuentra una mutación.
- **Aminoácido mutado:** Este factor es el nombre del codón o aminoácido nuevo resultante de la mutación presentada anteriormente.
- **Origen:** El origen es el tipo de tumor o alteración del cuerpo humano relacionado con la mutación de p53. Este factor determina cual fue el origen en caso de haberse presentado una mutación en las personas que aparecen en la base de datos.
- **Cáncer:** Este factor es el tipo de cáncer asociado a la mutación presentada en la base de datos
- **Autores:** Este factor contiene los autores de los artículos científicos o investigaciones que respaldan la información presentada en cada fila de la base de datos. Estos autores son quienes hicieron la relación entre una mutación de la proteína p53 de un paciente con cáncer y el tipo de cáncer y origen que este presenta.

Los criterios de aceptación en el análisis de una mutación se basan en que una mutación es presentada cuando sólo una base nitrogenada, de la cadena natural de aminoácidos de p53, presenta un cambio. Teniendo en cuenta que las bases nitrogenadas son: A (adenina), C (citocina), G (guanina), T (timina); una mutación es entonces el cambio de una de ellas a cualquier otra de las tres restantes. Este cambio a su vez produce un cambio en la cadena de aminoácidos de p53 y dependiendo del cambio puede tener o no coincidencia con los estudios de mutaciones en p53 y su relación con hepatocarcinomas celulares realizados por los científicos respectivos.

Una mutación puede ser presentada en archivo con formato FASTA; el cual es un archivo de texto plano que contiene información relevante a la cadena misma o una porción de la cadena de p53 donde hubo una mutación. Cuando la mutación es encontrada en una

porción de la cadena de aminoácidos de p53 entonces esta porción debe ser una secuencia continua de aminoácidos, uno tras otro, que se alinean de inicio a fin en la cadena normal de aminoácidos de p53.

10.2.1 ESPECIFICACIÓN DE REQUISITOS DEL SISTEMA

Teniendo en cuenta la información recolectada tras las reuniones presenciales y con base a los procesos referentes al análisis de mutaciones en p53 y el modelado 3D de su estructura, se establecieron los requisitos del software.

Los requisitos se dividieron en funcionales y no funcionales. Los requisitos funcionales describen las transformaciones que el sistema realiza sobre las entradas para producir salidas. Los requisitos no funcionales son características que de una u otra forma puedan limitar al sistema, por ejemplo seguridad, confiabilidad, etc.

10.2.1.1 REQUISITOS FUNCIONALES

La lista de requisitos funcionales para el desarrollo del software de Identificación de anomalías no comunes en la proteína p53 y casos de hepatocarcinoma celular fue diseñada basado en los objetivos del proyecto, reuniones presenciales entre investigadores y co-investigadores y los requerimientos. A continuación se presentan los requisitos funcionales

ID. Requisitos	Nombre del Requisito	Usuario	Descripción
RF-001	Agregar archivos al software	Investigador	El investigador debe agregar archivos formato FASTA por medio de un gestor de archivos al software.
RF-002	Analizar secuencia de aminoácidos	Software	El software debe analizar la cadena de aminoácidos del archivo FASTA.
RF-003	Alinear secuencia de aminoácidos	Software	El software debe mostrar la subcadena de aminoácidos del archivo FASTA alineada con base a la cadena de aminoácidos de p53.
RF-004	Visualizar modelo 3D	Software	El software debe graficar en un modelo 3D la proteína p53, resaltando la zona donde hubo una mutación
RF-005	Visualizar casos de hepatocarcinoma celular	Software	El software debe mostrar la lista de casos de hepatocarcinoma celulares que coincidan con la mutación presentada anteriormente basado en la base de datos de estudios de científicos.

Tabla 1. Requisitos funcionales

10.2.1.2 REQUISITOS NO FUNCIONALES

La lista de requisitos no funcionales para el desarrollo del software de Identificación de anomalías no comunes en la proteína p53 y casos de hepatocarcinoma celular fue diseñada basado en los objetivos del proyecto, reuniones presenciales entre investigadores y co-investigadores y los requerimientos, teniendo en cuenta que la interfaz gráfica de usuario debería ser agradable y el software fácil de usar para los usuarios. A continuación se presentan los requisitos no funcionales:

ID. Requisitos	Nombre del Requisito	Descripción
RF-001	Entorno visual	El interfaz gráfico debe ser agradable para los usuarios del software.
RF-002	Usabilidad	El software es desarrollado para que los usuarios investigadores puedan usarlo de manera fácil.
RF-003	Seguridad	El software debe garantizar veracidad e integridad en los datos entregados.
RF-004	Rendimiento	El software maneja varias ventanas abiertas en simultáneo, debe ser capaz de manejar varias tareas.

Tabla 2. Requisitos no funcionales

10.3 FASE DE CONSTRUCCION

10.3.1 DESARROLLO DEL SOFTWARE

Para un desarrollo correcto y apropiado de la fase de construcción de la metodología de desarrollo RUP, se realizó una investigación acerca de las tecnologías que permitieran desarrollar la mejor solución informática a los requisitos planteados, esto para evitar posibles riesgos que se presenten durante el desarrollo. Una de las tecnologías que se decidió implementar fue la librería de Biojava ya que esta tecnología posee un amplio conjunto de herramientas que facilitan el desarrollo de software orientados a la bioquímica y son de gran ayuda para la realización de los requisitos funcionales planteados. También se definió la interfaz como una interfaz fácil y rápida de usar para que la experiencia con el usuario sea más amena.

Se decidió utilizar la tecnología JAVA como entorno de trabajo para el desarrollo del software por ser una tecnología de código abierto (open source) y por su facilidad para la integración con Biojava, el cual también fue desarrollado en JAVA. Así mismo, la interfaz de usuario se desarrolló bajo las herramientas nativas que incluye java para la creación y modelo de vistas. La integración con la base de datos que contiene los estudios previos de los de hepatocarcinoma celular, y su relación con las mutaciones en la proteína p53, se realizó con una conexión entre JAVA y una base de datos en Excel, debido a la facilidad de esta para realizar consultas a la base de datos cuando se trabaja de manera local y aprovechando la naturaleza de este proyecto al ser un software para uso local.

Para el análisis de la cadena de aminoácidos de la proteína p53 mutada, la cual se ingresa al software a través de un archivo en formato FASTA, fue desarrollado bajo los principios del algoritmo de alineamiento de unión de vecinos (neighbor joining) y con complementos que fueron utilizados por Biojava para realizar un alineamiento más preciso. El modelado 3D de la estructura de la proteína p53, tanto mutada como en su estado principal, se desarrolló con ayuda de las herramientas que provee la librería Jmol, la cual es una tecnología

especializada en el modelado 3D de estructuras moleculares y contiene las características necesarias para desarrollar este requerimiento.

El modelo de casos de uso propuesto muestra los procesos relacionados con el análisis de una mutación en la proteína p53 en formato FASTA y el modelado 3D

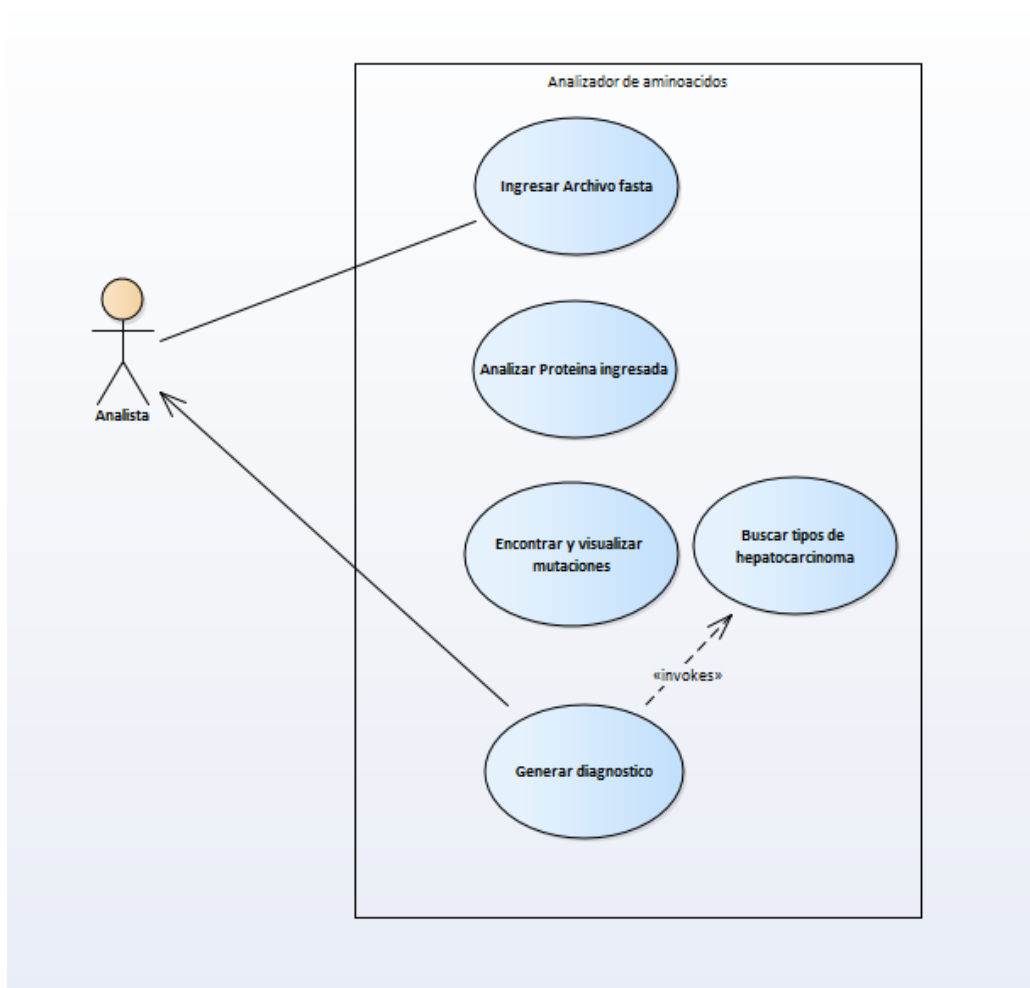


Figura 10. Modelo de casos de uso

Para analizar el modelo de negocio se emplearon el modelo de dominio y parte de la vista lógica de la arquitectura del software y los requisitos funcionales del mismo. En el modelo de dominio (Ver figura 11 Modelo de Dominio), se delimita el software, definiendo claramente sus principales entidades y las relaciones entre ellas, tal como ocurre en el mundo real.

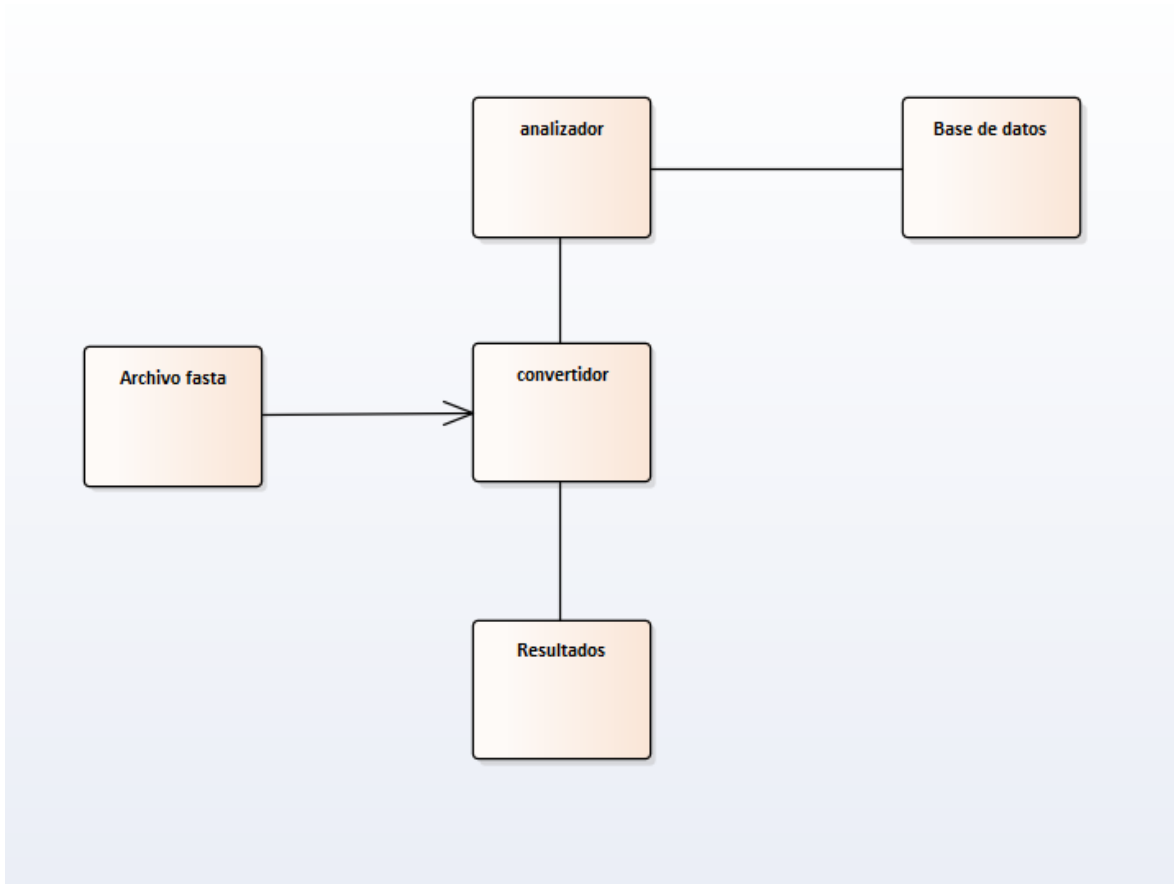


Figura 11. Modelo del dominio

10.4 FASE DE TRANSICIÓN

10.4.1 PRUEBA DE MÓDULOS EN TIEMPO REAL

Para comprobar el correcto funcionamiento del sistema, se hicieron pruebas en tiempo real. En esta prueba se tomó como base los procesos referentes al análisis de la cadena de aminoácidos de la proteína p53 mutada, la integración con la base de datos que contiene los estudios previos de los de hepatocarcinoma celular, y su relación con las mutaciones en la proteína p53 y el modelado 3D de la estructura de la proteína p53, en la cual participó el investigador y estudiante de genética y bioinformática de la Universidad Nacional Mayor de San Marcos, Miguel Alcalde Alvites en representación del usuario analista.

Para el usuario analista se realizaron las siguientes actividades:

- **Ingresar mutación en formato FASTA:** El usuario analista utiliza archivos en formato FASTA con secuencias de aminoácidos de la proteína p53, con una mutación, para ingresarlo al software a través de un gestor de archivos que este mismo provee.
- **Visualizar resultado del análisis:** Una vez el software recibe una mutación en archivo de formato FASTA, este procede a analizarlo, mediante el algoritmo de alineamiento, y muestra en pantalla los resultados encontrados al detectar una mutación en la secuencia ingresada, los cuales son: codón de la mutación, posición del codón y aminoácido al que mutó.
- **Visualizar relación de la mutación encontrada en casos de HCC:** Con los datos obtenidos del análisis de la secuencia mutada se realiza una consulta a la base de datos para encontrar una relación entre la mutación y uno o muchos casos de hepatocarcinoma celular que presentaron también esta mutación, incluyendo datos importantes de estos casos de hepatocarcinoma celular como el científico encargado

de la investigación, tipo de carcinoma y forma en que se presenta en el cuerpo humano. Posteriormente el software muestra en pantalla los resultados encontrados.

- **Visualizar modelo 3D de la proteína p53:** Para finalizar el estudio por el cual se hace uso de este software, se muestra en pantalla un modelo 3D de la proteína p53 resaltando la posición molecular donde se presentó la mutación. El usuario tiene la posibilidad de interactuar con el modelo, permitiendo acercarlo o alejarlo, cambiar las características del modelo y modificar la estructura en la que se muestra.

El usuario analista realizó las actividades correspondientes, para lo cual el software funcionó adecuadamente, cumpliendo a satisfacción todos los requerimientos establecidos en la fase inicial del proyecto, logrando así satisfacer al usuario final.

10.4.2 PRUEBAS DE REQUISITOS FUNCIONALES

- **Agregar archivos al software:** El investigador debe agregar archivos formato FASTA por medio de un gestor de archivos al software.

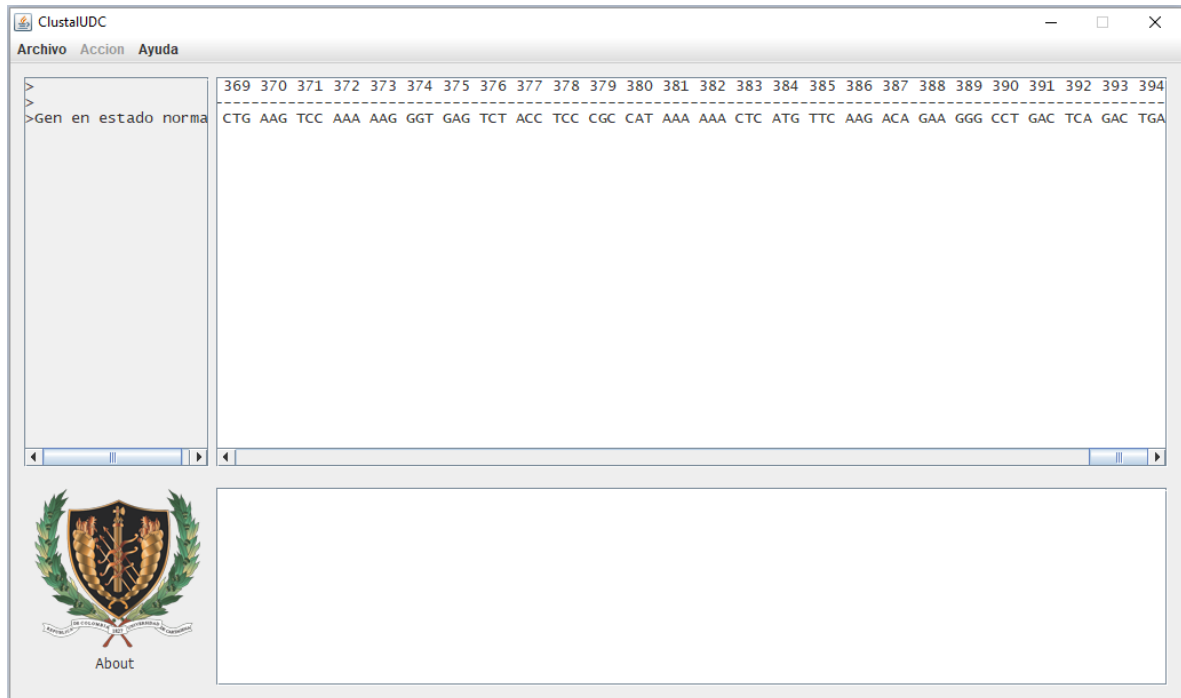


Figura 12. Ventana principal del software.

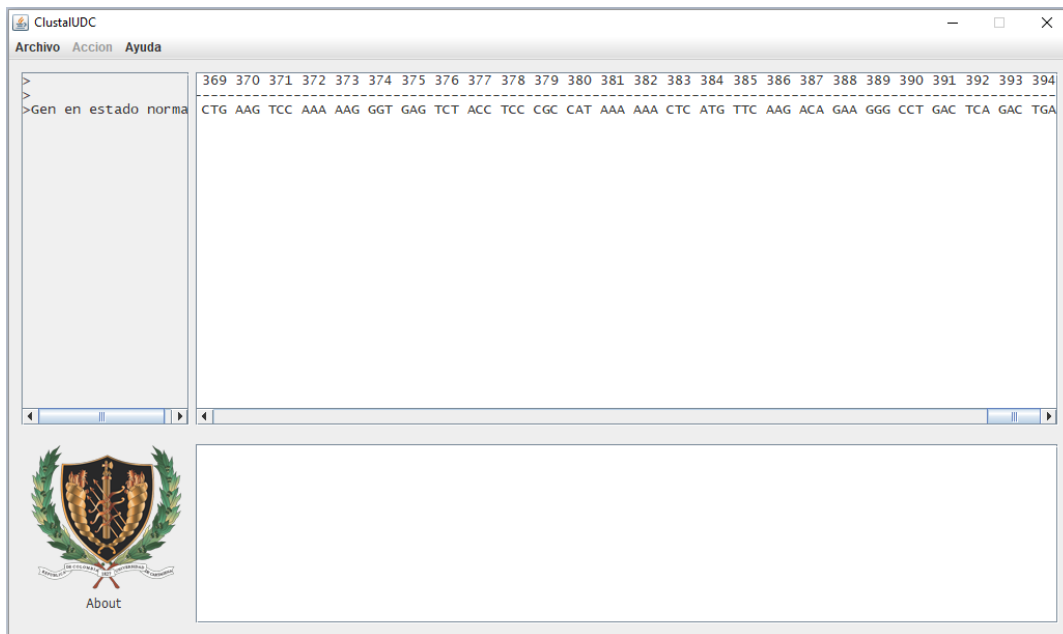


Figura 13. Gestor de archivos para ingresar archivos FASTA.

- **Analizar secuencia de aminoácidos:** El software debe analizar la cadena de aminoácidos del archivo FASTA.

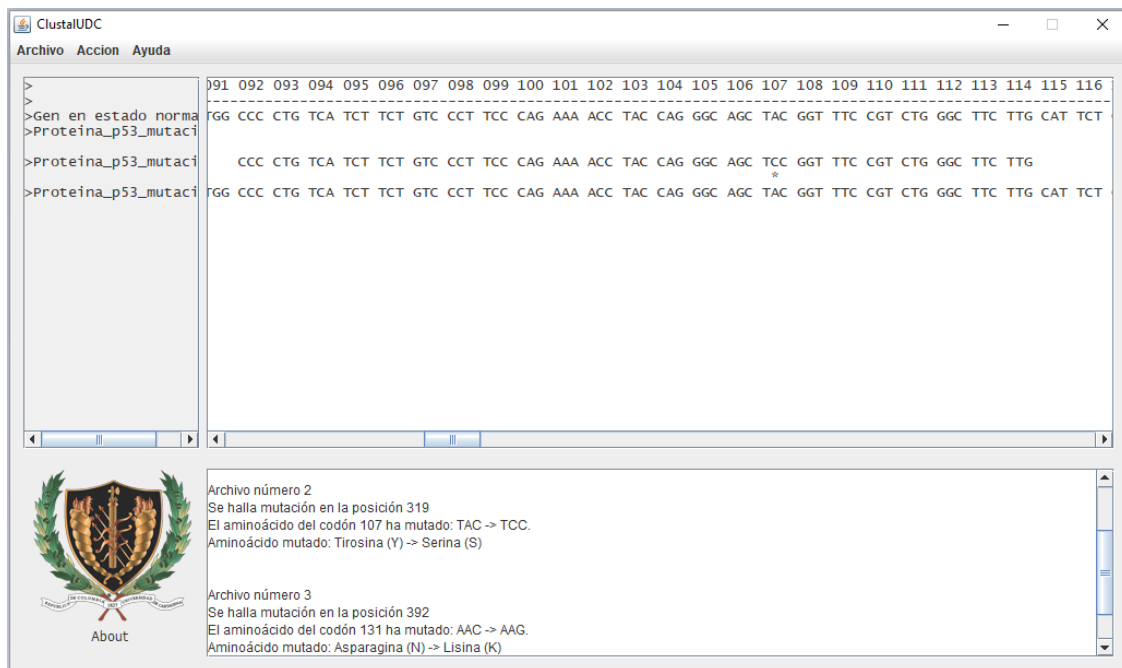


Figura 14. Ventana de resultados al analizar y alinear una secuencia mutada.

- **Alinear secuencia de aminoácidos:** El software debe mostrar la subcadena de aminoácidos del archivo FASTA alineada con base a la cadena de aminoácidos de p53. (Ver figura 14)
- **Visualizar modelo 3D:** El software debe graficar en un modelo 3D la proteína p53, resaltando la zona donde hubo una mutación.

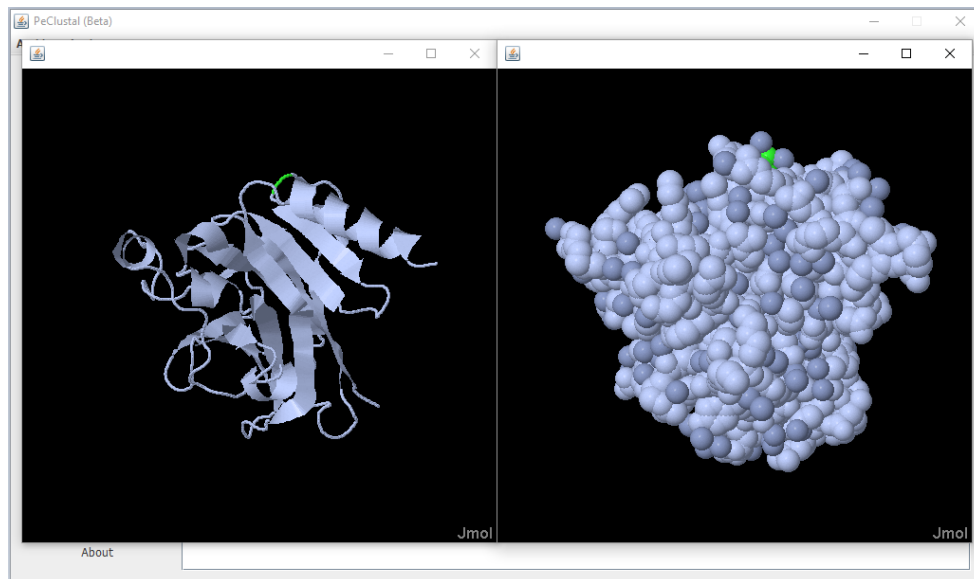


Figura 15. Ventana del modelo 3D.

- **Visualizar casos de hepatocarcinoma celular:** El software debe mostrar la lista de casos de hepatocarcinoma celulares que coincidan con la mutación presentada anteriormente basado en la base de datos de estudios de científicos.

No.	Mutatio...	Exon	Codon	WT Co...	Mutant ...	WT AA	Mutant ...	Origin	Cancer	Authors
1	393.0	5.0	131.0	AAC	AAG	Asn	Lys	Tumour	Bladder carcinoma	Spruck CH, Rideo...
2	393.0	5.0	131.0	AAC	AAG	Asn	Lys	Tumour	Esophageal ADC	Soslow RA, Altorki...
3	393.0	5.0	131.0	AAC	AAG	Asn	Lys	Tumour	Esophageal SCC	Goan YG, Chang ...

Figura 16. Ventana de resultados al realizar la consulta a la base de datos.

11 CONCLUSIONES

Este proyecto tiene como resultado un software para identificar anomalías no comunes en la proteína p53 y casos de hepatocarcinoma celular, alinear secuencias de aminoácidos y graficar en modelos 3D mutaciones de la proteína p53. Este sistema es utilizado para facilitar el estudio de la proteína p53 por parte de investigadores de ciencias relacionadas con la bioquímica y bioinformática.

Para la realización de este proyecto se propuso un objetivo general y objetivos específicos, de tal forma que al cumplirlos satisfactoriamente se obtienen las siguientes conclusiones:

- La proteína p53 es el principal supresor de tumores cancerígenos del cuerpo humano, el análisis de su estructura puede determinar mutaciones en casos de hepatocarcinoma celular o diferentes tipos de cáncer.
- Cuando una persona posee un tipo de hepatocarcinoma celular, o tumores cancerígenos, le es realizado un estudio enfocado a la proteína p53. Estudiar esta proteína tiene como resultado avances en la prevención y tratamiento del cáncer y los registros de su estudio son almacenados en bases de datos como la usada para realizar este proyecto. Hainaut, P. (1997).
- Es importante contar con la ayuda de un experto en el campo en el cual se realiza el proyecto, en este caso el de la bioinformática, para realizar un estudio más preciso y certero. La comunicación con los demás investigadores fue posible a través de las iteraciones que recomienda la metodología RUP durante el desarrollo del proyecto.
- Destacar los requerimientos funcionales y no funcionales es una de las labores más importantes en la planeación de un producto software. Gracias a los tiempos propuestos por la metodología RUP y a las reuniones con los investigadores y co-investigadores se realizó un estudio adecuado a los requerimientos del software.
- Parte del desarrollo del proyecto estuvo enfocado a librerías de código abierto en lenguaje JAVA llamadas BioJava y Jmol. Este tipo de librerías contienen material

suficiente para que investigadores puedan tomar el trabajo de otros expertos en sistemas de información y aplicarlo en sus propias investigaciones. Estas librerías son de uso gratuito y contienen términos y condiciones establecidas por sus dueños y dependiendo de su uso. Andreas Prlic (2012)

- Al comparar los resultados provenientes de la base de datos, se evidencia que existen casos donde mutaciones en la proteína P53 también se relacionan con tipos de cánceres diferentes al hepatocarcinoma celular. Esta conclusión puede generar nuevas investigaciones relacionadas.

El desarrollo de este proyecto responde la pregunta inicial de la formulación del problema planteada “¿Cómo se podría automatizar el múltiple alineamiento de las secuencias de la proteína p53 y sus mutaciones a fin de detectar la presencia de Hepatocarcinoma Celular (HCC) incluyendo un modelado 3D de la secuencia de esta proteína?” entregando como resultado un software que cumple los objetivos generales y específicos, los requerimientos planteados y es respaldado por el aval del investigador y experto bioinformático del proyecto.

Al culminar este proyecto, es necesario resaltar como resultado inesperado los casos en los que una sola mutación coincide con muchos casos de HCC basado en la base de datos utilizada. Este resultado permite concluir que existen mutaciones que son presentadas en múltiples casos de HCC y otras mutaciones que tal vez pueden presentarse solo en un caso de HCC, una interesante discusión que puede ayudar a científicos e investigadores en futuros proyectos.

Además, esta investigación es pionera en el estudio de la bioinformática aplicada en el programa de Ingeniería de Sistemas de la Universidad de Cartagena, abriendo paso así a un grupo de posibilidades para investigaciones futuras en el programa en convenio con programas de esta u otras universidades interesadas en estudios relacionados.

12 RECOMENDACIONES

- Es necesario contar con un experto en el tema, quien pueda resolver las dudas cuando se hace el análisis de requerimiento, especialmente en proyectos de investigación.
- Como ayuda para programadores, existen proyectos de código abierto y tecnologías que pueden ser usadas de forma gratuita en investigaciones, teniendo en cuenta los términos y condiciones de los creadores de estos, en especial para proyectos con alto nivel de dificultad en el campo de la programación.
- El software desarrollado con este proyecto puede ser mejorado por futuros investigadores, ya sea mejorando el algoritmo de alineamiento que se ha utilizado, actualizando la interfaz de usuario o adaptándolo a un requerimiento similar que alguien pueda necesitar.
- El estudio realizado en este proyecto puede ser tomado como referencia para futuros proyectos que busquen investigar y analizar esta u otra proteína relacionada con otras enfermedades.

13 REFERENCIAS

Cariello, N. F., Cui, L., Beroud, C., & Soussi, T. (1994). Database and software for the analysis of mutations in the human p53 gene. *Cancer research*, 54(16), 4454-4460.

Teufel, A., Staib, F., Kanzler, S., Weinmann, A., Schulze-Bergkamen, H., & Galle, P. R. (2007). Genetics of hepatocellular carcinoma. *World Journal of Gastroenterology*, 13(16), 2271.

Hollstein, M., Sidransky, D., Vogelstein, B., & Harris, C. C. (1991). p53 mutations in human cancers. *Science*, 253(5015), 49-53.

Hsu, I. C., Metcalf, R. A., Sun, T., Welsh, J. A., Wang, N. J., & Harris, C. C. (1991). Mutational hot spot in the p53 gene in human hepatocellular carcinomas.

Slater, A. W., Castellanos, J. I., Sippl, M. J., & Melo, F. (2013). Towards the development of standardized methods for comparison, ranking and evaluation of structure alignments. *Bioinformatics*, 29(1), 47-53.

León Cruz, G., & Bosques Diego, O. D. J. (2005). Infección por el virus del papiloma humano y factores relacionados con la actividad sexual en la génesis del cáncer de cuello uterino. *Revista Cubana de Obstetricia y Ginecología*, 31(1), 0-0.

Rosas, N., Torres, E. (1999a). Asociación entre la proteína p53 mutada, grado de infiltración y tamaño del tumor en cáncer colorrectal. *Sección de anatomía patológica*,

Hospital central FAP. Citando a Chang F. The p53 tumor suppressor gene as a common cellular target in human carcinogenesis. *Am J Gastroenterol* 1993; 88: 174-86.

Rosas, N., Torres, E. (1999b). Asociación entre la proteína p53 mutada, grado de infiltración y tamaño del tumor en cáncer colorrectal. Sección de anatomía patológica, Hospital central FAP. Citando a Poller D, Baxter K, Shepherd N. P53 and RBi protein expression: Are they prognostically useful in colorectal cancer? *Br J Cancer* 1997; 75(1), 87-93.

Rosas, N., Torres, E. (1999c). Asociación entre la proteína p53 mutada, grado de infiltración y tamaño del tumor en cáncer colorrectal. Sección de anatomía patológica, Hospital central FAP. Universidad Nacional Mayor de San Marcos. 60(2), 85-29.

Pinto, Y., Ibañez, M., Rangel, N., Ramirez, S., Sanchez, W., Vanegas, D., (2007) Polimorfismos del gen p53 en cáncer mamario familiar en una población Colombiana. *Rev Colomb cir.* 22(1).

Chenna R, Sugawara H, Koike T, Lopez R, Gibson T, Higgins D, Thompson J. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research.* 31(13), 3497-3500.

Higgins, D.G. Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, 73, 237-244.

Feng, D.F., Doolittle, R.F. (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.*, 25, 351-360.

Saitou,N., Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4, 406–425.

Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25, 4876–4882.

Silva, A., Gutiérrez, A., Ariaz, C., Lazaro, I (2002). Estructura, regulación y funciones del gen supresor de tumores p53. Velazquez. CIB, CSIC. 144, 28006 Madrid.

Lastra, E. (2013) Genética y cáncer. Sección de oncología médica. Hospital general Yagüe. Recuperado el 2 de mayo de 2016, de Slideshare: <http://es.slideshare.net/jaaldi/gentica-y-cncer-16103448>

Murphy, G., Kersten, M., & Findlater, L. (2006). How are Java software developers using the Eclipse IDE? *IEEE*, 76,83.

Foundation, B. (2013). Blender. Recuperado el 10 de abril de 2016, de Blender: <http://www.blender.org>

Glanville, S. (2011). Anim8or. Recuperado el 10 de abril de 2016, de Anim8or: <http://www.anim8or.com/main/index.html>

Asturnatura (2011). Estructura secundaria de proteínas. Recuperado el 14 de abril de 2016, de Asturnatura: <http://www.asturnatura.com/articulos/proteinas/estructura-secundaria.php>

Naruya, S., Masatoshi, N. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. The University of Chicago. Mol. Bios. Evol. 4(4), 406-425.

MedWOW (2014). Información comercial de un espectrómetro de masas. Recuperado el 5 de mayo de 2016, de MedWOW: <http://es.medwow.com/used-mass-spectrometer-equipment/48.med>

Anónimo (2015). Estructura cuaternaria de de proteínas. Recuperado el 16 de abril de 2016, de: <http://m.exam-10.com/himiya/13762/index.html>

Castaños, E. (2015). Niveles de organización de las proteínas. Recuperado el 16 de abril de 2016, de LidiaConLaQuímica: <https://lidiakonlaquimica.wordpress.com/2015/07/04/niveles-de-organizacion-de-las-proteinas/>

Colegio Glenn Doman (2011). Introducción a la bioquímica. Recuperado el 17 de abril de 2016 de: http://www.colegioglennoman.edu.co/2011_aula_bioquimica%20clase%202.htm

Unity Technologies (2016) Manual de usuario Unity. Recuperado el 3 de mayo de 2017 de: <https://docs.unity3d.com/es/current/Manual/index.html>

Jmol (2016) Un visor Java de código abierto para estructuras químicas en tres dimensiones. Recuperado el 4 de mayo de 2017 de: <http://www.jmol.org/>

Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009) "Jalview Version 2 - a multiple sequence alignment editor and analysis workbench"

Bioinformatics 25 (9) 1189-1191 doi: 10.1093/bioinformatics/btp033

BioJava: an open-source framework for bioinformatics in 2012

Andreas Prlic; Andrew Yates; Spencer E. Bliven; Peter W. Rose; Julius Jacobsen; Peter V. Troshin; Mark Chapman; Jianjiong Gao; Chuan Hock Koh; Sylvain Foisy; Richard Holland; Gediminas Rimsa; Michael L. Heuer; H. Brandstatter-Muller; Philip E. Bourne; Scooter Willis

Bioinformatics (2012) 28 (20): 2693-2695.

Hainaut, P., Soussi, T., Hollstein, M., Greenblatt, M., Hovig, E., Harris, C., Montesano, R. (1997) Database of p53 gene somatic mutations in human tumors and cell lines: updated compilation and future prospects. Oxford University Press – Nucleic Acids Research 1997; 25(1): 151 - 157
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC146396/pdf/250151.pdf>