

**ESTUDIO E IMPLEMENTACIÓN DE ALGUNAS HERRAMIENTAS
BIOINFORMATICAS PARA SU POSIBLE INTEGRACIÓN EN UN SISTEMA QUE
PERMITA LA CARACTERIZACIÓN DEL CITOCROMO C de *Kluyveromyces
fragilis***

**LAURA VICTORIA AMAYA COTE
MIGUEL JOSÉ MOSCOTE GÓMEZ**

**LINEA DE INVESTIGACION SISTEMAS DE INFORMACIÓN E INGENIERÍA DEL
SOFTWARE - BIOTECNOLOGÍA
FACULTAD DE INGENIERÍA DE SISTEMAS
UNIVERSIDAD AUTONOMA DE BUCARAMANGA
BUCARAMANGA
2004**

**ESTUDIO E IMPLEMENTACIÓN DE ALGUNAS HERRAMIENTAS
BIOINFORMATICAS PARA SU POSIBLE INTEGRACIÓN EN UN SISTEMA QUE
PERMITA LA CARACTERIZACIÓN DEL CITOCROMO C de *Kluyveromyces
fragilis***

**LAURA VICTORIA AMAYA COTE
MIGUEL JOSÉ MOSCOTE GÓMEZ**

Tesis de Grado para optar el título de
Ingeniera(o) de Sistemas

Director

Dra. GRACIELA CHALELA A. MSc. Dr. rer. nat.

**LINEA DE INVESTIGACION SISTEMAS DE INFORMACIÓN E INGENIERÍA DEL
SOFTWARE – BIOINFORMÁTICA
FACULTAD DE INGENIERÍA DE SISTEMAS - LABORATORIO DE
INVESTIGACIÓN Y MEDIO AMBIENTE
UNIVERSIDAD AUTONOMA DE BUCARAMANGA
BUCARAMANGA**

2004

Nota de aceptación

Firma del Presidente del Jurado

Firma de Jurado

Firma de Jurado

Bucaramanga, 22 de Octubre de 2004

A Dios por ser el primero en mi vida, por darme la fortaleza y permitirme los medios para culminar esta etapa profesional.

A mi Padre Pedro Pablo y a mi Madre Luz Marina por el amor, comprensión y apoyo brindado, gracias por hacer de nuestro hogar ese valuarte y un lugar de descanso.

A mis Hermanos Lilitiana Patricia y Pedro Samuel, a Tomasito y demás seres queridos que durante este proceso me han acompañado.

Laura Victoria Amaya Cote

*A Dios , ser supremo que me iluminó y me dio fuerzas para no desfallecer en
ningún momento.*

*A mí admirada Madre, Ena, que con su amor incondicional, dedicación, apoyo,
fortaleza en momentos difíciles y su grandísimo esfuerzo, logré alcanzar este
nuevo triunfo en mi vida.*

*A mi Padre que con sus palabras sabias me ayudó a salir adelante durante todo
este proceso.*

*A mis queridas Tías quienes me apoyaron en todo momento para que alcanzara
este sueño.*

*A mis hermanos quienes estuvieron siempre conmigo brindándome su apoyo
incondicional.*

*Y a mi adorada sobrina Maria Valentina quien me dio fuerzas para seguir
adelante y luchar para que este sueño se hiciera realidad.*

Miguel José Moscote Gómez

AGRADECIMIENTOS

A la Doctora Graciela Chalela Álvarez, nuestra directora del proyecto por su valiosa orientación y apoyo a lo largo de este proyecto, quien no solo nos guió durante este proceso académico sino que también nos infundó valores personales y profesionales.

Al Doctor José de Jesús Pérez, quien con sus conocimientos fue una pieza importante para el desarrollo de este proyecto.

A los Ingenieros Daniel Arenas Seleey y Freddy Méndez Ortiz quienes con su asesoría nos orientaron hacia el buen desarrollo de nuestro proyecto de grado.

A Maria de Pilar Velasco quien con su asesoría durante las prácticas de laboratorio fue clave para alcanzar los logros en el área de biología.

A nuestros compañeros y a todas aquellas personas que nos acompañaron durante todo este proceso.

Y a la Universidad Autónoma de Bucaramanga, en especial a la Facultad de Ingeniería de Sistemas y al Laboratorio de Biotecnología y Medio Ambiente, por proporcionarnos las herramientas y los recursos necesarios para la realización de nuestro proyecto.

CONTENIDO

GLOSARIO	14
RESUMEN	21
INTRODUCCIÓN	24
1. PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACION	27
2. ANTECEDENTES	29
3. OBJETIVOS	33
3.1 OBJETIVO GENERAL	33
3.2 OBJETIVOS ESPECIFICOS	33
4. ESTADO DEL ARTE	34
5. MARCO TEORICO	36
5.1 CONCEPTOS BIOLÓGICOS	36
5.1.1 Levaduras	36
5.1.2 Transporte de Electrones y Fosforilación Oxidativa	37
5.1.3 Sistema Citocromo	40
5.1.4 Mitocondrias	40
5.1.5 Ribosomas	44
5.1.6 Proteínas	44
5.1.7 Métodos y determinación de crecimiento de los microorganismos	52
5.1.8 Métodos de secuenciación	54
5.2 CONCEPTOS INFORMÁTICOS	57
5.2.1 Sistema	57
5.2.2 Sistema de Información	57
5.2.3 Lenguaje UML	58
5.2.4 Plataforma de Desarrollo Windows	61
5.2.5 Lenguajes e Interfaces de Desarrollo	62
5.2.6 Microsoft Access.	64
5.3 HERRAMIENTAS BIOINFORMATICAS	66
5.3.1 Rasmol	66
5.3.2 Treeview	67
5.3.3 Fasta	69
5.3.4 Clustal X	70
5.3.5 Kegg	72
5.3.6 Yeast Microarray Global Viewer	72
5.3.7 SRS (Embl/Ebi	74
5.3.8 Entrez	75

5.3.9 NCBI Blast Server	76
6. DISEÑO METODOLÓGICO	78
6.1 ETAPA 1: METODOLOGÍA PARA EL ESTUDIO BIOLÓGICO DEL <i>KLUYVEROMYCES FRAGILIS</i> .	79
6.1.1 Fases	79
6.2 ETAPA 2: METODOLOGÍA PARA EL DESARROLLO DEL SOFTWARE	80
6.2.1 Fases	82
7. RESULTADOS	84
7.1 MODELAMIENTO DE LAS HERRAMIENTAS BIOINFORMATICAS LOCALES EXISTENTES	84
7.1.1 Rasmol	84
7.1.2 Clustal X	86
7.1.3 Fasta	87
7.1.4 Treeview	88
7.2 SISTEMA INTEGRADO DE INFORMACION BIOINFORMÁTICA	89
7.3 MODELAMIENTO DEL SISTEMA DE INFORMACION BIOINFORMATICO	90
7.3.1 Diagramas de casos de usos	90
7.3.2 Diagramas de secuencias.	95
7.3.3 Diagrama de clases	103
7.4 CARACTERIZACIÓN DEL CITOCROMO C DE <i>KLUYVEROMYCES FRAGILIS</i>	106
7.4.1 Secuencia genética del Citocromo C de <i>Kluyveromyces fragilis</i>	106
7.4.2 Secuencia de aminoácidos del Citocromo C de <i>Kluyveromyces fragilis</i>	107
7.5 ANÁLISIS DE RESULTADOS EN EL SISTEMA DE INFORMACION BIOINFORMÁTICO	108
7.5.1 Análisis en Fasta	108
7.5.2 Análisis en Clustal X	109
7.5.3 Análisis en Rasmol	115
7.5.4 Análisis en TreeView	117
7.5.5 Análisis en KEGG	117
7.5.6 Análisis en Blast	118
8. CONCLUSIONES	120
9. RECOMENDACIONES	122
BIBLIOGRAFIA	123
ANEXO A.	128

LISTA DE TABLAS

Tabla 1. Representación del código genético junto con los aminoácidos que lo componen 48

LISTA DE FIGURAS

Figura 1. Fotografía electrónica de <i>Kluyveromyces fragilis</i>	36
Figura 2. Representación de la cadena respiratoria: matriz de la mitocondria (Tomado de Luenga, 2003)	39
Figura 3. Ultraestructura mitocondrial y proceso energético.(Tomado de Luenga, 2.003)	41
Figura 4. Representación gráfica del ciclo de Krebs y sus interconexiones metabólicas. (Tomado de Biopsicología).	43
Figura 5. Representación de la alanina; aminoácido mas pequeño después de la glicina	46
Figura 6. Representación de la secuenciación por el N-terminal.	55
Figura 7. Representación de la ruptura de una proteína por medio de Bromuro de cianógeno.	56
Figura 8. Representación de una molécula 2MCG en modelo Wireframe	66
Figura 9. Representación de una molécula 2MCG en modelo Spacefill	67
Figura 10. Representación del árbol de la filogenia de primates en Unrooted	68
Figura 11. Representación del árbol de la filogenia de primates en Rectangular cladogram	68
Figura 12. Interfaz modo gráfico donde solicita archivo de Fasta	69
Figura 13. Archivo generado como resultado del software.	70
Figura 14. Interfaz de Clustal X, mostrando resultado en alineamiento múltiple	71
Figura 15. Archivo de resultado para impresión	71
Figura 16. Interfaz en línea de KEGG	72
Figura 17. Interfaz en línea de YMGV	73

Figura 18. Resultados arrojados por el software.	74
Figura 19. Interfaz de la bases de datos de SRS	75
Figura 20. Interfaz para visualización de resultado	75
Figura 21. Interfaz de la bases de datos de Entrez	76
Figura 22. Interfaz de la bases de datos de Blast	77
Figura 23. Diagrama general del proceso Bioinformático (Realizado por los autores)	78
Figura 24. Representación del ciclo de vida del RUP [2].	81
Figura 25. Diagrama de casos de uso Herramienta Rasmol	85
Figura 26. Diagrama de casos de uso Herramienta Clustal X	87
Figura 27. Diagrama de casos de uso Herramienta Fasta	87
Figura 28. Diagrama de casos de uso Herramienta Fasta	89
Figura 29. Diagrama de casos de uso, nivel 0	90
Figura 30. Diagrama de casos de uso, nivel 1	91
Figura 31. Diagrama de secuencia – Crear archivo	96
Figura 32. Diagrama de secuencia – Determinar extensión y herramientas para visualizar archivo	98
Figura 33. Diagrama de secuencia – Ejecutar herramienta	99
Figura 34. Diagrama de secuencia – Crear enlaces Web	100
Figura 35. Diagrama de secuencia – Modificar enlace Web	101
Figura 36. Diagrama de secuencia – Eliminar enlace Web	102
Figura 37. Diagrama de clases	104
Figura 38. Representación tridimensional de la secuencia de aminoácidos de <i>Kluyveromyces fragilis</i> .	116
Figura 39. Representación de <i>Kluyveromyces fragilis</i> en Rasmol vista 'cintas'	116
Figura 40. Filogenia del citocromo c de <i>Kluyveromyces fragilis</i>	117
Figura 41. Genes de enzimas de <i>Kluyveromyces fragilis</i> que participan en la fosforilación oxidativa.	118
Figura 42. Ventana principal del sistema Bioinformático	127

Figura 43. Administrar Herramientas	128
Figura 44. Consulta Herramientas relacionadas con Extensiones.	129
Figura 45. Abrir archivo.	130
Figura 46. Herramientas Bioinformáticas compatibles.	131
Figura 47. Ejecutar Herramienta.	131

LISTA DE ANEXOS

ANEXO A. PANTALLAS DEL SISTEMA DE INFORMACIÓN BIOINFORMÁTICO	127
---	-----

GLOSARIO

ACIDO CÍTRICO: Para efecto de la presente investigación, el ácido cítrico como ácido tricarboxílico hace referencia al Ciclo de Krebs.

ÁCIDO DESOXIRRIBONUCLEICO (ADN): El componente químico dentro del núcleo de una célula que lleva las instrucciones para elaborar los organismos vivientes.

ÁCIDO RIBONUCLEICO (ARN): El componente químico que resulta de la transcripción del ADN. En el ARN, la letra U, que corresponde al uracilo, substituye a la Timina del ADN. En las células se encuentran tres tipos de ARN. El ARN mensajero, el ARN de transferencia y ARN ribosomal.

ADENOSÍN TRIFOSFATO: Es el ATP (adenosín trifosfato) la molécula que interviene en todas las transacciones de energía que se llevan a cabo en las células; por ella se la califica como "moneda universal de energía". El ATP está formado por adenina, ribosa y tres grupos fosfatos, contiene enlaces de alta energía entre los grupos fosfato; al romperse dichos enlaces se libera la energía almacenada.

ALGORITMO: Es un conjunto finito de instrucciones o pasos que sirven para ejecutar una tarea o resolver un problema.

AMINOÁCIDOS: Los aminoácidos son compuestos orgánicos que al combinarse forman las proteínas. Los aminoácidos son el resultado de la digestión de las proteínas. Están clasificados en aminoácidos "esenciales" (deben ser consumidos) y "no esenciales" (los puede producir el organismo a partir de los aminoácidos esenciales).

BIOINFORMÁTICA: Una disciplina científica que se interesa por todos los aspectos relacionados con la adquisición, almacenamiento, procesamiento, distribución, análisis e interpretación de información biológica, mediante la aplicación de técnicas y herramientas de las matemáticas, de la biología y de la informática, con el propósito de comprender el significado biológico de una gran variedad de datos.

BIOLOGÍA MOLECULAR: Es una ciencia cuyo objetivo fundamental es la comprensión de todos aquellos procesos celulares que contribuyen a que la información genética se transmita eficientemente de unos seres a otros **y** se exprese en los nuevos individuos.

VÍAS METABÓLICAS: Son los diferentes pasos bioquímicas que realiza la célula de cualquier organismo para llevar a cabo reacciones de asimilación y desasimilación.

VÍAS REGULADORAS: Son aquellas que permiten influenciar los procesos de supresión (o no inducción) de un gen para construir una determinada proteína

CITOCROMO: Proteína coloreada que contiene hierro y participa en los procesos de respiración celular.

CITOPLASMA: Región celular situada entre la membrana plasmática y el núcleo, conteniendo los orgánulos celulares.

CODÓN: La información genética se escribe con cuatro letras, pero que están en grupos de tres. Cada grupo de tres se llama codón y lo que hace es codificar un aminoácido o un símbolo de puntuación (Comenzar, Stop).

COENZIMAS: Son grupos químicos que ayudan a las enzimas a llevar a cabo sus funciones anabólicas y catabólicas.

ENZIMAS: Las enzimas son proteínas que catalizan básicamente todas las reacciones químicas en las células de todos los organismos vivos. Las enzimas metabolizan un sustrato para convertirlo en un producto sin que la enzima sufra ninguna modificación estructural o metabólica, las enzimas se clasifican según el grupo químico que catalizan.

ESPECTROMETRÍA: Proceso químico en donde se utiliza como sistema de medición el espectro fotómetro, fundamentalmente se utiliza para describir cualitativamente y cuantitativamente la composición de una sustancia orgánica o no.

EUCARIOTA: Son células con núcleo diferenciado y con compartimentos intracelulares separados entre si por membranas independientes

EXONES: Componentes importantes del ácido ribonucleico recién transcrito. Combina secuencias de expresión con secuencias de no expresión. Las secuencias de expresión se denominan ex Los aminoácidos son compuestos orgánicos que al combinarse forman las proteínas. Los aminoácidos son el resultado de la digestión de las proteínas.

FILOGENIA: Parte de la Biología que se ocupa de las relaciones de parentesco entre los distintos grupos de seres vivos. Origen y desarrollo evolutivo de las especies, y en general, de las estirpes de seres vivos.

GENOMA: Es el conjunto de genes que especifican todos los caracteres que pueden ser expresados en un organismo. Un genoma es todo el material genético de un ser vivo. Es el juego completo de instrucciones hereditarias para la

construcción y mantenimiento de un organismo, y pasar la vida a la siguiente generación. En la mayoría de los seres vivos, el genoma está hecho por un químico llamado A.D.N. El genoma contiene genes, empacados en cromosomas y afectan características específicas del organismo.

GLUCÓLISIS: Proceso que consiste en una serie de diez reacciones, cada una catalizada por una enzima determinada, que permite transformar una molécula de glucosa en dos moléculas de un compuesto de tres carbonos, el ácido pirúvico.

GLUCOSA: Molécula que se sintetiza en la fotosíntesis de las plantas, y esta contenida como molécula compleja en los nutrientes de todos los seres vivos y de cuya degradación química obtiene energía todos los seres vivos. Se compone de seis átomos de carbono, doce de hidrógeno y seis radicales -OH.

HERRAMIENTAS BIOINFORMÁTICAS: Es software que permite el análisis, representación, visualización, comparación de microorganismos biológicos

HIDROFOBICO: La propiedad que tiene una sustancia de repelar el agua.

HONGOS: Son organismos con núcleo que se reproducen por esporas, carecen de clorofila (por lo tanto no son fotosintéticos), se reproducen sexual o asexualmente y pueden tener estructuras filamentosas o no.

IN SITU: Se refiere a los procesos o las reacciones que se realizan en un determinado lugar dentro de la célula.

ITERACIÓN: Una iteración es un ciclo completo de desarrollo que produce una versión (interna o externa) de un producto ejecutable, que constituye un subconjunto del producto final en desarrollo, que luego se ira incrementado de iteración en iteración hasta convertirse en el sistema final. Cada iteración pasa a

través de varios flujos de trabajo del proceso, aunque con un énfasis diferente en cada uno de ellos, dependiendo de la fase en que se encuentre.

LEVADURA: Nombre genérico de ciertos hongos unicelulares, de forma ovoidea, que se reproducen por gemación o división. Suelen estar unidos entre sí en forma de cadena, y producen enzimas capaces de descomponer diversos cuerpos orgánicos, principalmente los azúcares, en otros más sencillos.

METABOLISMO: Conjunto de reacciones químicas que efectúan constantemente las células de los seres vivos con el fin de sintetizar sustancias complejas a partir de otras más simples, o degradar aquellas para obtener estas.

METABOLISMO GLUCIDICO: Es la liberación de energía de un aceptor a un receptor y se refiere a la degradación de la glucosa para la síntesis de la ATP .

METABOLISMO LIPÍDICO: Este es el metabolismo de los lípidos o de las grasas.

METABOLISMO PROTEICO: Es el proceso de degradación de las proteínas para formar compuestos simples como aminoácidos, las enzimas que intervienen e el procesos de llaman peptidazas.

MITOCONDRIA: Mitocondria del griego: *mitochondrias*, thread + *chondros*, gránulo es el sitio donde se lleva a cabo la respiración celular, el metabolismo aerobio, en la mayoría de los organismos eucariontes (con núcleo verdadero).

MOLÉCULAS: Es una de las diferentes formas en las que se pueden agrupar los átomos de uno o varios elementos químicos, y se caracterizan por estar formadas por un número determinado de átomos. Si en una molécula cambian el número de átomos, o se altera la distribución de los mismos, nos encontraremos con una molécula distinta correspondiente a un compuesto distinto.

NUCLEÓTIDOS: Los nucleótidos son los bloques de construcción a partir de los cuales se construyen los ácidos nucleicos. Los nucleótidos están unidos en una cadena polinucleotídica con un esqueleto que consiste de series alternadas de residuos de azúcar y fosfato.

ORGÁNULOS: Componentes de las células eucariotas con morfología y funciones bien diferenciadas. Los principales orgánulos son: Núcleo: Síntesis y reparación del DNA, Nucleólo: Síntesis de los ribosomas y proceso del RNA. Mitochondrias: Respiración celular.

OXIDACIÓN: Es el proceso electroquímico por el cual un ión o átomo pierde uno o varios electrones. Cuando un ión o átomo se oxida: Pierde electrones, Actúa como agente reductor, Es oxidado por un agente oxidante, Aumenta su estado o número de oxidación.

PÉPTIDOS: Son un tipo de moléculas formadas por la unión de varios aminoácidos mediante enlaces peptídicos. La unión de un bajo número de aminoácidos da lugar a un péptido:

- **Oligopéptido:** Número de aminoácidos < 10.
- **Polipéptido:** Número de aminoácidos > 10.
- **Proteína:** número de aminoácidos > 50.

POLÍMEROS: La materia esta formada por moléculas que pueden ser de tamaño normal o moléculas gigantes llamadas polímeros. Los polímeros se producen por la unión de cientos de miles de moléculas pequeñas denominadas monómeros que forman enormes cadenas de las formas más diversas.

PROBIÓTICO: Son los microorganismos y/o las sustancias producidas por estos que sirven para estimular el crecimiento o para efectuar reacciones de asimilación

de nutrientes. Los probióticos corresponden a una preparación de un producto que contiene microorganismos viables en suficiente número, los cuales alteran la microbiota (por implantación o colonización) en un compartimiento del huésped provocando efectos beneficiosos sobre la salud del mismo.

PROTEASA: Es una enzima que degrada otras proteínas. Hay muchos tipos de proteasas y cada una degrada un tipo específico de proteínas. Las proteasas son importantes en la regulación celular, porque constantemente es necesario cortar y reparar moléculas.

PROTEÓMICA: La proteómica estudia el proteoma, el conjunto de proteínas que se expresan a partir de un genoma. La enorme cantidad de información generada por los proyectos de secuenciación de genomas, y la necesidad de descifrar esta información están desplazando el foco de atención hacia el estudio directo de las proteínas, su estructura, su función, sus interacciones y sus modificaciones.

SISTEMA: Es el conjunto de cosas que ordenadamente relacionadas entre sí, contribuyen a un determinado objetivo.

SISTEMA DE INFORMACIÓN: Es un conjunto formal de procesos que, operando sobre una colección de datos estructuradas recopilan, elaboran y distribuyen la información necesaria para las operaciones de una empresa.

SISTEMA REDOX: Proceso de oxidorreducción entre el reductor y el oxidante, en donde existe una transferencia de electrones que genera una corriente eléctrica, marcada por una diferencia de potencial entre ambos.

UBIQUINONA: Corresponde a una proteína que naturalmente está presente en el interior de las células y cumple la función de activar la respiración celular.

RESUMEN

La Bioinformática como nueva área de interés en nuestra sociedad y como uno de los nuevos acontecimientos científicos que comprende la investigación y el desarrollo de herramientas bioinformáticas permiten entender el flujo de información biológica, partiendo desde los genes hasta las estructuras moleculares; por esta razón, hoy en día la Bioinformática se ha convertido en un instrumento necesario por medio del cual los investigadores logran profundizar en sus estudios científicos.

Actualmente algunas instituciones interesadas en la investigación biológica e informática como la UNAB, requieren de herramientas que permitan un mejor estudio y entendimiento de la composición estructural y funcional de las proteínas en particular, así como el diseño de modelos que los expliquen. Para lograr realizar este tipo de investigaciones se necesita tener claro conocimiento sobre aspectos biológicos e informáticos, lo cual se logra tomando como punto de partida que la Bioinformática. Dado su carácter multidisciplinar, permite mayor comprensión sobre el significado biológico de una gran cantidad de datos que sin la ayuda de los sistemas de información no tendrían el desarrollo científico que han alcanzado.

Por lo anterior esta investigación como proyecto de grado ha realizado un estudio biológico que busca comprender una serie de procesos biológicos básicos en las células y un estudio de diferentes herramientas bioinformáticas como son: Rasmol, Treeview, Fasta, Clustal X, SRS, YMGV, BLAST SERVER, KEGG, Entrez; las cuales de manera gráfica permiten visualizar, modificar y comprender la composición estructural de las proteínas.

Para la realización de este trabajo se plantearon elementos metodológicos procedentes tanto del campo biológico como de las tecnologías informáticas, que articulados de manera armónica dieron lugar a un proceso metodológico complejo que permitió de una parte, el estudio biológico del Citocromo C de **Kluyveromyces fragilis** y por otra, el análisis de las herramientas Bioinformáticas y el desarrollo del Sistema de Información Bioinformático. Dicho proceso se organizó a manera de fases que permiten cumplir los objetivos específicos del proyecto indicando qué y cómo se lograrán cada uno de ellos, así como las técnicas y recursos que lo operacionalizan.

El estudio biológico se abordó desde el crecimiento y desarrollo de **Kluyveromyces fragilis** hasta la obtención de la secuencia de aminoácidos de la proteína del Citocromo C, mediante las siguientes fases: Cultivo de la levadura **Kluyveromyces fragilis** sobre medio modificado de Malta dextrosa agar y medio de Ogy; Obtención de biomasa mediante cultivo en medios líquidos modificados, como el caldo de Malta en Biorreactores de 5 litros; Proceso de Secuenciación.

La metodología para la elaboración del análisis de las herramientas Bioinformáticas y el desarrollo del Sistema de Información Bioinformático se fundamentó sobre el Proceso Unificado de Racional (RUP), a través de la siguiente secuencia:

- **Iniciación.** Recolección de información y determinan de los recursos necesarios, además de un planteamiento de fases iterativas donde se planifica el proceso.
- **Elaboración.** Análisis de toda la documentación y de las diferentes herramientas bioinformáticas dando como resultado la determinación de la viabilidad de integración de estas en un único sistema.
- **Construcción.** De ser posible la integración de estas herramientas bioinformáticas en un único sistema, en esta fase se desarrollará de forma

iterativa e incremental el producto software requerido para el Centro de Investigación en Biotecnología y Medioambiente de la UNAB, usando el software de desarrollo apropiado para su construcción.

- **Transición.** Evaluación del funcionamiento de Sistema de Información Bioinformático
- **Conclusiones.** Muestra los resultados obtenidos y si estos cumplieron con los objetivos propuestos al inicio de la investigación.

El estudio buscaba evaluar la posibilidad de integración de estas herramientas Bioinformáticas en un único sistema que le facilite a los investigadores del Laboratorio de Investigación y Medio Ambiente de la UNAB, el estudio de las proteínas microbianas en especial la del Citocromo C de *Kluyveromyces fragilis*.

Como resultado del proceso, se logró: la integración de las diferentes herramientas bioinformáticas que al comienzo de la investigación se plantearon. Además se realizó la caracterización del Citocromo C de **Kluyveromyces fragilis**, cumpliendo así con los objetivos de esta investigación.

INTRODUCCIÓN

Debido a que la Bioinformática desde hace algún tiempo atrás surge como nueva disciplina científica, ha merecido gran interés mundial especialmente por organismos pioneros dedicados a la investigación y el avance científico. Este tema actualmente no es ajeno a la Universidad Autónoma de Bucaramanga, la cual lo ha considerado como un gran desafío debido en parte a los desarrollos y aplicaciones derivadas de las investigaciones sobre biología molecular que plantea el presente siglo.

Este proyecto titulado **ESTUDIO E IMPLEMENTACIÓN DE ALGUNAS HERRAMIENTAS BIOINFORMATICAS PARA SU POSIBLE INTEGRACIÓN EN UN SISTEMA QUE PERMITA LA CARACTERIZACIÓN DEL CITOCROMO C de *Kluyveromyces fragilis*** se realiza con el fin de iniciar el proceso Bioinformático dentro de la UNAB, es por ello que se parte de una estructura interdisciplinar entre el Laboratorio de Computo Especializado y el Laboratorio de Investigación en Biotecnología y Medio Ambiente, teniendo como base la importancia del trabajo en equipo y la aplicación de la ciencia biológica y de la tecnología para afrontar este reto.

En el presente documento y en los diferentes capítulos que lo componen se podrá apreciar como los objetivos propuestos al inicio de la investigación se fueron desarrollando para así obtener los resultados esperados.

Este documento se encuentra estructurado por capítulos, en el capítulo primero titulado Planteamiento del problema y justificación cual contiene los aspectos que conllevaron a realizar esta investigación. En este se muestra cual es la situación actual y como a través de este proyecto de grado se pretende dar inicio a la era de la Bioinformática en la UNAB.

El capítulo segundo, Antecedentes muestra al contexto en que se concibió y se desarrolló el presente proyecto, destacando la importancia de la iniciación de esta nueva línea de investigación en la UNAB, debido a los avances científico – tecnológicos y la importancia de la unión entre las herramientas tecnológicas como una herramienta de apoyo a la investigación Biológica.

En los Objetivos que constituyen el capítulo tercero, se especifica de manera concreta el alcance del proyecto y los presupuestos científicos para su desarrollo

En el capítulo cuarto, Estado del Arte se expone una recopilación del desarrollo histórico de la Bioinformática y también enfatiza cómo ésta ha ido evolucionando dentro de la UNAB, poniendo de manifiesto la importancia de este aporte científico – tecnológico. Mostrando la importancia de la inclusión de esta línea dentro del proyecto investigativo de la UNAB.

En el capítulo quinto, Marco Teórico se muestran los fundamentos conceptuales, tecnológicos y metodológicos tanto del área biológica como informática que orientan la elaboración de este proyecto, la relación que existe entre ellos en el contexto del proyecto, e igualmente la metodología propuesta para obtener los resultados. Teniendo en cuenta las diferentes etapas necesarias para el logro de los objetivos propuestos. A su vez también se recopila la información necesaria sobre las herramientas Bioinformáticas tomadas en cuenta para el desarrollo de este proyecto de grado.

El capítulo sexto titulado Desarrollo Metodológico muestra una descripción de las fases planteadas para el desarrollo de este proyecto, teniendo en cuenta que este es multidisciplinar y por lo tanto comprende un área biológica y un área informática.

En el capítulo séptimo, Resultados se describe todo el proceso desarrollado para cumplir con los objetivos propuestos de este proyecto, cumpliendo paso a paso con la metodología de este proyecto. Iniciando por el análisis de las Herramientas Bioinformáticas locales con su respectivo modelado y el por qué de la determinación de su posible integración en un único sistema; la obtención de la secuencia de aminoácidos del Citocromo C de **Kluyveromyces fragilis**, y la caracterización de esta levadura en el Sistema de información Bioinformático.

En el capítulo octavo, se exponen las Conclusiones en donde se enfatiza el proceso final y los logros obtenidos en la investigación y se analizan cada una de las herramientas bioinformáticas integradas que permitirá la aplicación de tales herramientas en el proceso biológico concluyendo la eficiencia de la integración y su aplicación.

En los anexos se ilustra mediante las pantallas del sistema las guías de utilización del Sistema de Información Bioinformático.

1. PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACION

La Bioinformática como nueva área de interés en nuestra sociedad y como uno de los nuevos acontecimientos de la ciencia, se puede entender como la intersección entre la biología y la tecnología de la información y proporciona las herramientas y recursos necesarios para favorecer la investigación biológica. Este campo interdisciplinario comprende la investigación y desarrollo de herramientas útiles para llegar a entender el flujo de información desde los genes a las estructuras moleculares, su función bioquímica, su conducta biológica y, finalmente, la influencia en las enfermedades.

Una de las definiciones generalmente aceptada de Bioinformática es: "Una disciplina científica que se interesa por todos los aspectos relacionados con la adquisición, almacenamiento, procesamiento, distribución, análisis e interpretación de información biológica, mediante la aplicación de técnicas y herramientas de las matemáticas, de la biología y de la informática, con el propósito de comprender el significado biológico de una gran variedad de datos". [1]

Dentro del contexto científico-tecnológico en el Laboratorio de Investigaciones en Biotecnología y Medioambiente (LINBYA) de la UNAB, se requieren herramientas bioinformáticas que permitan un mejor estudio y entendimiento de la composición estructural y funcional de las proteínas, en especial las derivadas de los procesos metabólicos realizados en la mitocondria de los microorganismos, como en las levaduras.

El reto en la construcción de software para el área bioinformática, es el establecimiento de una arquitectura que permita la realización de búsquedas inteligentes, comunicación con bases de datos y la unión con herramientas de análisis que permitan dar respuesta a problemas biológicos concretos.

Por esta razón se ha decidido por medio de este proyecto, estudiar algunas herramientas bioinformáticas actuales, buscando su posible integración en un solo sistema que le facilite a los investigadores del Laboratorio de Investigación en Biotecnología y Medioambiente (LINBYA) de la UNAB, el estudio de la proteína del Citocromo C de **Kluyveromyces fragilis** y que se pueda usar en estudios posteriores con diferentes microorganismos.

2. ANTECEDENTES

No se puede mirar la historia de la bioinformática sin describir inicialmente la historia de la biología. En realidad son los biólogos y los bioquímicos quienes hacen su primer acercamiento a la tecnología computacional como elemento fundamental para su trabajo diario. [2]

La Bioinformática comprende tres subespecialidades:

- **Bioinformática en sentido estricto.** La investigación y desarrollo de la infraestructura y sistemas de información y comunicaciones que requiere la biología moderna. (Redes y bases de datos para el genoma, estaciones de trabajo para procesamiento de imágenes).
- **Biología Molecular Computacional.** La computación que se aplica al entendimiento de cuestiones biológicas básicas, mediante la modelización y simulación. (Sistemas de Vida Artificial, algoritmos genéticos, redes de neuronas artificiales).
- **Biocomputación.** El desarrollo y utilización de sistemas computacionales basados en modelos y materiales biológicos. (Biochips, biosensores, computación basada en DNA). Los computadores basados en DNA se están empleando para la secuenciación masiva y el screening de diversas enfermedades, explotando su característica de procesamiento paralelo implícito. [1]

Los sistemas informáticos que se emplean en el área bioinformática cumplen las siguientes funciones:

- Adquisición de datos.
- Software para visualización.
- Generación y ensamblaje de secuencias.
- Análisis de datos.
- Programas para análisis de secuencias.
- Predicción de estructura de proteínas.
- Paquetes de integración y ensamblaje de mapas genéticos.
- Software para clasificación y comparación.
- Técnicas de inteligencia artificial.
- Gestión de datos.
- Bases de datos locales o accesibles mediante redes de comunicaciones.
- Distribución de datos.
- Redes de comunicaciones.

Los principales campos de aplicación de la bioinformática, se pueden resumir como:

- Gestión de datos en el laboratorio.
- Automatización de experimentos.
- Ensamblaje de secuencias contiguas.
- Predicción de dominios funcionales en secuencias génicas.
- Alineación de secuencias.
- Búsquedas en las bases de datos de estructuras.
- Determinación de la estructura de macromoléculas.
- Predicción de la estructura de macromoléculas.
- Evolución molecular.
- Árboles filogenéticos.

La nueva generación de informática introduce el concepto de Bioinformática de Segunda Generación caracterizada porque en los últimos años, se ha trabajado con muchas bases de datos que almacenaban información biológica a medida que iban apareciendo. Sin embargo esto no sólo ha tenido efectos positivos ya que muchos científicos se quejan de la creciente complejidad que representa encontrar información útil en este "laberinto de datos".

Para mejorar esta situación, se desarrollan técnicas que integran la información dispersa, gestionan bases de datos distribuidas, las seleccionan automáticamente, evalúan su calidad, y facilitan su accesibilidad para los investigadores. Se habla de Bioinformática Integradora, que cada vez, con más énfasis, reside en Internet y no en bases de datos locales. [1]

En unos años, la comunidad científica tendrá a su disposición la secuencia de bases (3.000 millones) que componen el Genoma Humano; sin embargo, esta información (estructural) es insuficiente para un entendimiento completo de su función, regulación y variación (aproximadamente 80.000 genes). Los procesos celulares son gobernados por el repertorio de genes expresados y su patrón de actividad temporal. Se necesitan herramientas para gestionar información genética en paralelo. Por ello, se emplean nuevas tecnologías para extracción de conocimiento, minería de datos y visualización. [3,4]

La bioinformática, en este sentido, ofrece la capacidad de comparar y relacionar la información genética con una finalidad justificada, siendo capaz de ofrecer unas respuestas que no parecen obvias a la vista de los resultados de los experimentos.

La potencia de estos sistemas trae consigo la obtención, en tiempos muy breves, de grandes volúmenes de información, (secuencias, mutaciones, datos de expresión génica, determinaciones analíticas de interés clínico, screening de

fármacos) que necesitan ser gestionados con técnicas bioinformáticas para extraer conocimiento de utilidad en la investigación biomédica.

Este proyecto se considera como una investigación pionera en el campo de la Bioinformática ya que no existen antecedentes de trabajos similares dentro de la UNAB, contando solo con la propuesta de un Prototipo de sistema para almacenamiento de información de secuencias biológicas en estructuras de árboles de posiciones, realizado por Julie Alexandra Quiroga Ríos estudiante de la facultad de Ingeniería de Sistemas en el año 1.998.

Con este trabajo de investigación se pretende iniciar la era de la Bioinformática en la UNAB y es por ello que se hace una sinergia entre el Laboratorio de Computo Especializado y el Laboratorio de investigación en biotecnología y Medio Ambiente (LINBYA) de la UNAB.

3. OBJETIVOS

3.1 OBJETIVO GENERAL

Conocer el funcionamiento de algunas herramientas Bioinformáticas aplicadas a la caracterización de la proteína del Citocromo C de **Kluyveromyces fragilis**, buscando su posible integración en un único sistema para apoyo de futuros estudios en el área de en la UNAB.

3.2 OBJETIVOS ESPECIFICOS

Determinar y caracterizar el ciclo de vida de **Kluyveromyces fragilis**.

Conocer los procesos biológicos básicos relacionados con la caracterización de la proteína del Citocromo C de la levadura **Kluyveromyces fragilis**.

Profundizar sobre algunas de las herramientas bioinformáticas disponibles para el estudio de la proteína del Citocromo C de **Kluyveromyces fragilis** como son: Fasta, Clustal X, TreeView, Rasmol, Kegg y Yeast Microarray Global Viewer.

Analizar el funcionamiento de las diferentes herramientas bioinformáticas descritas anteriormente, usando una metodología apropiada con el fin de comprender sus características básicas, siguiendo la metodología apropiada correspondiente a cada programa de Bioinformática.

Evaluar la posibilidad de integración de las herramientas bioinformáticas en un único sistema, que permita la caracterización del Citocromo C de **Kluyveromyces fragilis**.

4. ESTADO DEL ARTE

Desde el principio de los años 90, muchos laboratorios han estado analizando el genoma completo de varias especies tales como bacterias, levaduras, ratones y seres humanos. Durante estos esfuerzos de colaboración, se han generado cantidades enormes de datos los cuales se recogen y se almacenan en grandes bases de datos, la mayoría de las cuales son publicadas y accesibles.

Debido al incremento en complejidad y capacidad tanto de las computadoras como de las técnicas de investigación, se necesitan "puentes" humanos que puedan entender ambas disciplinas y sean capaces de comunicarse con los expertos de los dos campos. [1, 2, 4]

Actualmente hay cerca de 80 genomas que están accesibles públicamente por Internet y unas tres centenas de otros en bases de datos privadas. Genbank, la mayor base de datos de secuencias genéticas del mundo contiene actualmente cerca de 10 mil millones de pares de bases equivalentes a cerca de 12 millones de registros (genes). [1, 2, 4]

El auge que experimenta la Bioinformática es visible con el surgimiento en el mundo de compañías orientadas a prestar esta clase de servicios relacionados con el manejo de la información biológica. Los datos en sí mismos no son comercializables, pero la información implícita de ellos si lo es.

En la Revista Colombiana: Ciencia y Tecnología se hace un importante recuento de la capacidad de la Bioinformática para ayudar a resolver problemas y procesos biológicos, señalando que las Universidades, la Academia de Ciencias Exactas, Físicas y Naturales y entidades particulares, han dado comienzo muy recientemente a la aplicación de estas herramientas, entrando así a este nuevo

concepto que sin duda ayudará a realizar la tarea biológica con mayor precisión y a mas bajo costo. Competir exitosamente exige estar adelantados en el futuro, estar a la vanguardia en los desarrollos de la ciencia y la tecnología. Por esto es que la ciencia aplicada es urgente, para que Colombia pueda entrar a competir en un mundo globalizado, para que sus productos tengan un mayor valor agregado a través de la innovación. Por otro lado, porque es urgente modificar los actuales sistemas de producción que ya son obsoletos, frente a los nuevos que ofrecen los acelerados avances de la ciencia y la tecnología de hoy. [5]

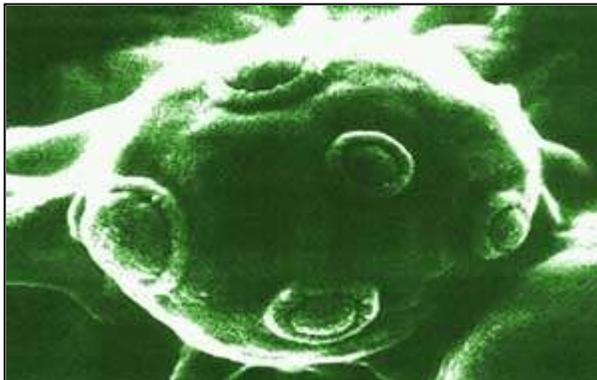
La Universidad Autónoma de Bucaramanga no es ajena a estos grandes desafíos debido en parte a los desarrollos y aplicaciones derivadas de las investigaciones sobre biología molecular que plantea el presente siglo, por eso desarrolló una estructura interdisciplinar entre el Laboratorio de Computo Especializado y el Laboratorio de Investigación en Biotecnología y Ambiente teniendo como base la claridad de la importancia del trabajo en equipo y la aplicación de la ciencia biológica y de la tecnología. Debido a los grandes avances del genoma humano, esta multidisciplina se viene transformando y dando origen a otras tres áreas emergentes: Genómica: mapeo, secuenciación y caracterización de genes, Proteómica: estudio de la interacción de las proteínas, mecanismos de control y expresión génica, entre otras y Bioinformática: basadas en supercomputadores con la capacidad de almacenar grandes bases de datos sobre secuencias genómicas de hombre y otros organismos vivos, desarrollo de algoritmos que permitan comparar bases e interacción de datos de librerías genómicas, librerías químicas, proteínas, química combinatoria, entre otras las cuales generarán grandes cambios en los mercados bursátiles del mundo entero.

5. MARCO TEORICO

5.1 CONCEPTOS BIOLÓGICOS

5.1.1 Levaduras. Las levaduras son hongos no filamentosos, es decir, que carecen de micelio. Su estructura reproductiva asexual es la Blastospora, que es una especie de yema que se desprende de manera irregular de la célula madre y da lugar a una célula exacta. También se reproducen sexualmente formando cuerpos fructíferos desnudos, denominados Ascos con unidades celulares que son las Ascosporas. (Figura 1)

Figura 1. Fotografía electrónica de *Kluyveromyces fragilis*



Este grupo perteneciente al Dominio Eukarya y al subdominio Fungi, agrupa a especies y géneros de importancia médica e industrial y están muy ligadas a los alimentos. Desde el punto de vista científico las levaduras se han convertido en modelos de estudio desde hace más de 100 años y se caracterizan porque poseen un ciclo de vida que permite realizar los ensayos que requieren tanto la biología molecular como la genética actual; tienen un genoma pequeño, solo unas cuantas veces mayor que el de las bacterias y 200 veces menor que el genoma de las células de los mamíferos.

Kluyveromyces fragilis es una levadura muy importante como probiótico y se está estudiando en el Laboratorio de Investigación en Biotecnología y Ambiente de la UNAB, su papel en los procesos de asimilación de nutrientes y la capacidad de servir de apoyo al aparato digestivo de los rumiantes, cualidad que será muy importante en la conversión de leche y carne de mejor calidad y cantidad en lugares de poca disponibilidad de buenos pastos para el pastoreo. La calidad nutricional se mide por la capacidad de biotransformación del alimento, proceso en el que juega papel importante la proteína del Citocromo C mitocondrial.

Uno de los paradigmas más importantes en la biología, es el que establece que el DNA se transcribe a RNA, el cual se traduce, generándose todas las proteínas que están presentes en los diferentes momentos del ciclo celular. Son las proteínas las responsables de mantener el funcionamiento de la célula y, es por ello que el estudio simultáneo de todo el juego de proteínas presentes en los diferentes momentos de la célula ha resultado de gran interés. Las proteínas que son sintetizadas por las levaduras participan en un gran número de procesos metabólicos y energéticos como el transporte de electrones y la fosforilación oxidativa. [6]

5.1.2 Transporte de Electrones y Fosforilación Oxidativa. La mayor parte de la energía libre que se produce durante la oxidación de la glucosa a CO₂ es retenida en las coenzimas reducidas FADH₂ y NADH generadas en la glucólisis y el ciclo del ácido cítrico, éstas durante la respiración liberan electrones, que irán pasando por una serie de transportadores en la membrana interna mitocondrial, formando 4 complejos:

- 1.-Complejo I.- De la NADH deshidrogenasa a la ubiquinona
- 2.-Complejo II .-De la Succinato deshidrogenasa a la ubiquinona.
- 3.-Complejo III.-Citocromo reductasa. De la ubiquinona al Citocromo C .

4.-complejo IV.-Citocromo C oxidasa, que acepta electrones del Citocromo C y los cede al oxígeno.

La citocromo oxidasa, dímero (hierro y cobre) cataliza de forma eficiente la reducción del oxígeno. La célula puede utilizar el oxígeno para la respiración solo porque la citocromo oxidasa sitúa al oxígeno en un centro bimetálico especial, donde se mantiene unida entre un átomo de hierro ligado a un grupo hemo y a un átomo de cobre, hasta que se capte un total de 4 electrones; solo entonces los 2 átomos de la molécula de oxígeno serán liberados como 2 moléculas de agua sin ningún peligro. La energía resultante se usa para el bombeo de los protones hacia fuera de la membrana interna mitocondria y provoca un gradiente electroquímico que es positivo - ácido y negativo- alcalino en la matriz mitocondrial. [7]

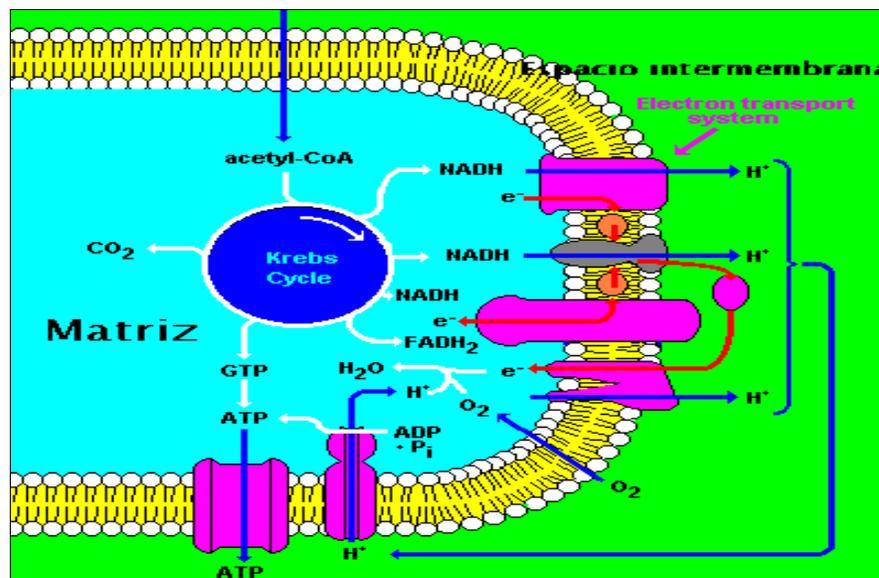
La ATP sintasa es la enzima que cataliza la fosforilación, posee una zona esférica llamada factor I dirigida hacia la matriz mitocondrial, y por otra parte el factor Fo integrado a la membrana. Dado que la membrana es básicamente impermeable a los protones, por lo tanto el gradiente no se desarma por una constante reentrada de los mismos, teniendo en cuenta que la ATP sintasa (conocido como complejo F1, ATP asa mitocondrial) contiene el único canal para la entrada de protones, por lo tanto a medida que los protones pasan por el canal, se produce la siguiente reacción $ADP + P_i = ATP$. [8, 9]

El acoplamiento de los dos procesos se realiza a través del gradiente de potencial electroquímico de protones que es generado por la cadena respiratoria durante la transferencia de electrones desde los sustratos hasta el oxígeno. Una de las proteínas esenciales para el transporte de electrones es el Citocromo C, que se caracteriza porque recibe los electrones del complejo ubiquinona Citocromo C₁ y transfiere al complejo Citocromo Oxidasa. Esta proteína ha sido extraída de la membrana interna mitocondrial y ha sido purificada y cristalizada; la proteína es

esférica de aproximadamente 104 a 113 residuos de aminoácidos y un grupo hemo unido covalentemente, el carácter hidrófóbico del entorno del grupo hemo hace que el potencial redox del Citocromo C sea más positivo. [7, 8]

La secuencia de más de 80 especies eucariotas ha sido determinada por Emil Smith y Col., el hallazgo más importante es que 26 de los 104 aminoácidos han permanecido inalterables durante más de 1500 millones de años. Algunos ligandos como Metionina e Histidina son invariantes, en algunas regiones la secuencia es casi invariable en todas las moléculas de Citocromo C, el arrollamiento compacto de la proteína requiere la presencia de Glicina en sitios determinados, varios residuos invariantes de Lisina y Arginina están localizados en los centros cargados positivamente en la superficie de la molécula. [10] (Figura 2)

Figura 2. Representación de la cadena respiratoria: matriz de la mitocondria (Tomado de Luenga, 2003)



Por todas las características mencionadas y siendo el Citocromo C una proteína esencial para la cadena respiratoria, en este trabajo se utilizarán diferentes herramientas de la bioinformática para estudiar algunas características

estructurales del Citocromo C de la levadura y su relación con los Citocromos C de otros organismos, así mismo su filogenia, su similitud o diferencia; también se determinará la importancia de su participación en el transporte de electrones y en la fosforilación oxidativa y por otra parte se buscará en la base de datos la secuencia de nucleótidos que corresponde al gen que codifica esta proteína.

5.1.3 Sistema Citocromo. El Citocromo C es una proteína que está presente en todos los organismos que contienen una cadena respiratoria mitocondrial: plantas, animales y microorganismos. Este transportador evolucionó hace más de mil millones de años.

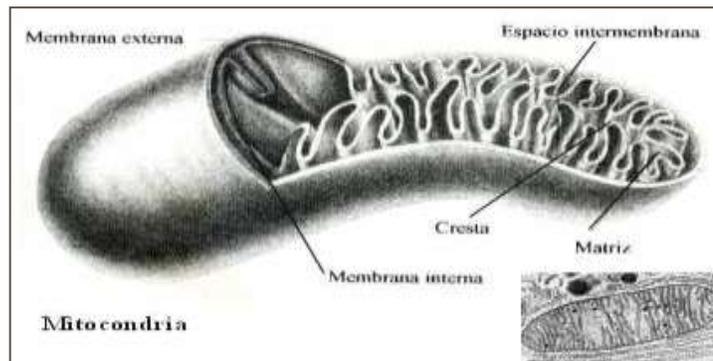
Es un componente de la cadena respiratoria mitocondrial de la mayoría de los Eucariotas; se encuentra situado en la membrana mitocondrial interna expuesto hacia el lado intermembrana. Está codificado por el DNA nuclear y se sintetiza como un precursor que posee una presecuencia N-terminal de 61 aminoácidos. Dicha presencia está formada por una región N-terminal muy básica de 35 aminoácidos, característica de un dominio que dirige a la matriz, una región central de 19 aminoácidos sin carga y una región C-terminal acídica de 7 aminoácidos. Parte de esta presecuencia es separada por una proteasa de la matriz cuando el polipéptido se inserta en la membrana interna. A continuación, el grupo *hemo* se une a la proteína y provoca un cambio de conformación para que pueda actuar como una segunda proteasa del espacio intermembrana que elimina el resto de la presecuencia. De esta manera, el Citocromo C queda insertado con la orientación adecuada de la membrana. [11, 12, 13]

5.1.4 Mitocondrias. Las Mitocondrias son diminutas estructuras celulares de doble membrana responsables de la conversión de nutrientes en el compuesto rico en energía trifosfato de adenosina (ATP), que actúa como combustible celular.

Por esta función que desempeñan, llamada respiración, se dice que las mitocondrias son el motor de la célula.

Se encuentran Mitocondrias en las células Eucarióticas (células con el núcleo delimitado por membrana). El número de Mitocondrias de una célula depende de la función de ésta. Las células con demandas de energía particularmente elevadas, como las musculares, tienen muchas más mitocondrias que otras. Por su parecido con las bacterias aeróbicas (es decir, que necesitan oxígeno), los científicos creen que las mitocondrias han evolucionado a partir de una relación simbiótica o de cooperación entre una bacteria aeróbica y una célula eucariótica ancestral, (Teoría Endosimbiótica. Las mitocondrias son los orgánulos celulares encargados de suministrar la mayor parte de la energía necesaria para la actividad celular, actúan por tanto, como centrales energéticas de la célula y sintetizan ATP a expensas de los carburantes metabólicos (glucosa, ácidos grasos y aminoácidos). (Figura 3)

Figura 3. Ultraestructura mitocondrial. (Tomado de Luenga, 2.003)



La ultraestructura mitocondrial está en relación con las funciones que desempeña: en la matriz se localizan las enzimas responsables de la oxidación de los ácidos grasos, los aminoácidos, el ácido pirúvico y el Ciclo de Krebs. [14, 15, 16]

En la membrana interna están los sistemas dedicados al transporte de los electrones que se desprenden en las oxidaciones anteriores y un conjunto de

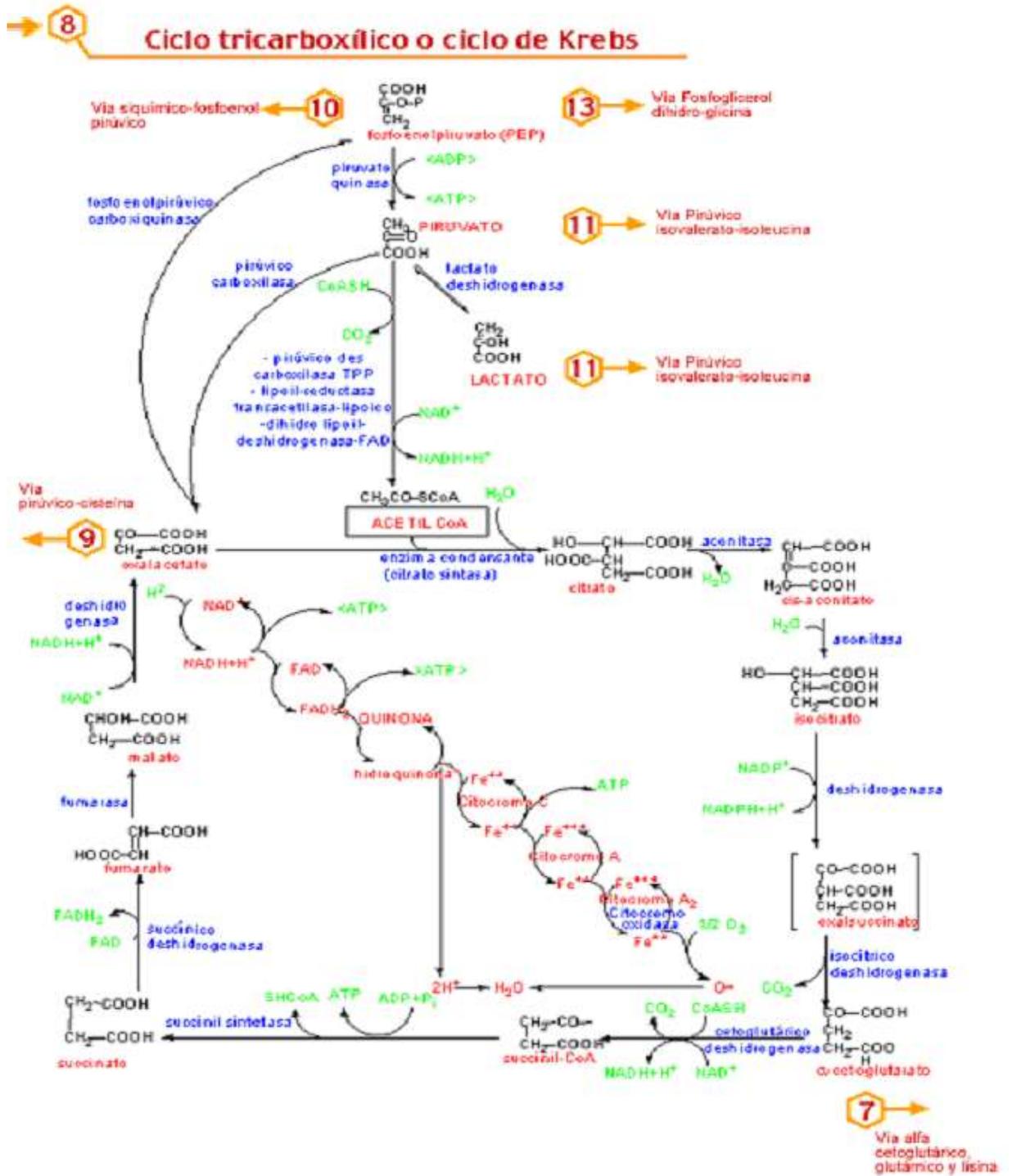
proteínas encargadas de acoplar la energía liberada del transporte electrónico con la síntesis de ATP; estas proteínas le dan un aspecto granuloso a la cara interna de la membrana mitocondrial.

También se encuentran dispersas por la matriz una molécula de ADN circular y unos pequeños ribosomas implicados en la síntesis de un pequeño número de proteínas mitocondriales.

El producto más importante de la degradación de los carburantes metabólicos es el acetil-CoA, (ácido acético activado con la coenzima A), que continúa su proceso de oxidación hasta convertirse en CO₂ y H₂O, mediante un conjunto de reacciones que constituyen el ciclo de Krebs punto central donde confluyen todas las rutas catabólicas de la respiración aerobia. Este ciclo se realiza en la matriz de la mitocondria.

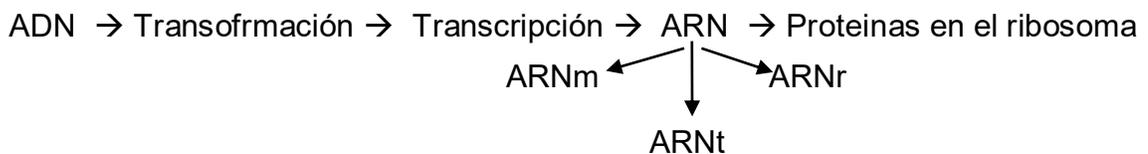
En este ciclo se consigue la oxidación total de los dos átomos de carbono del resto acetilo, que se eliminan en forma de CO₂; los electrones de alta energía obtenidos en las sucesivas oxidaciones se utilizan para formar NADH Y FADH₂, que luego entrarán en la cadena respiratoria, en donde los Citocromos juegan un papel importante. [7, 8, 16, 17]

Figura 4. Representación gráfica del ciclo de Krebs y sus interconexiones metabólicas. (Tomado de Biopsicología).



La vía metabólica 8 (Figura 4), es la ruta central del metabolismo intermediario, además de ser la responsable de la obtención de la energía biológica por parte de las células vivas desde la utilización de las estructuras mitocondriales. El papel de ruta central viene especialmente determinado por el cruce del metabolismo glucídico, con la figura del ácido pirúvico, con el metabolismo lipídico desde la acetil CoA y con el metabolismo protéico con las conexiones de los aminoácidos desde los ácidos oxalacético y alfa-cetoglutarico principalmente. [18]

5.1.5 Ribosomas. Son orgánulos membranosos, visibles solo al microscopio electrónico debido a su reducido tamaño en las células procariotas 29nm y 32nm en las células eucariotas. Su función es ensamblar proteínas.



5.1.6 Proteínas. Las proteínas, del griego Proteion, son moléculas de peso molecular elevado, formados por aminoácidos formados por enlaces peptídicos. La unión de un bajo número de aminoácidos da lugar a un péptido: oligopéptido: aminoácidos <10; polipéptido: aminoácidos > 10; proteinas: aminoácidos > 50.

Son biomoléculas formados básicamente por carbono, hidrógeno, oxígeno y nitrógeno, y pueden contener Hierro (Fe), Fósforo (P), Azufre (S), Magnesio (Mg) y Cobre (Cu). Son las más abundantes de las Biomoléculas, constituyen más del 50% de las células.

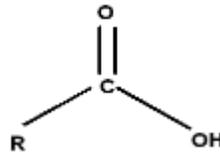
Las proteínas naturales están formadas por aminoácidos de la serie Levógira. Los aminoácidos se clasifican en: alifáticos neutros de cadena apolar (glicina, alanina, valina, leucina e isoleucina); alifáticos neutros de cadena no ionizable (serina y

treonina); aromáticos neutros (fenilalanina, tirosina y triptófano); con azufre (metionina y cisteina); ácidos di carboxílicos (ácido glutámico y aspártico); básicos (lisina, arginina e histidina). [1, 3, 18]

Las proteínas son polímeros lineales de α -aminoácidos con amplia variabilidad estructural y funciones biológicas muy diversas. La variedad de proteínas es elevadísima, y para su clasificación se suele recurrir a: criterios físicos, químicos, estructurales y funcionales.

Los aminoácidos se unen entre sí mediante enlaces peptídicos, entre el carboxilo de un aminoácido con el grupo amina del otro, con pérdida de agua. Se denomina péptido al polímero de aminoácidos cuya masa es menor de 6 Kda.

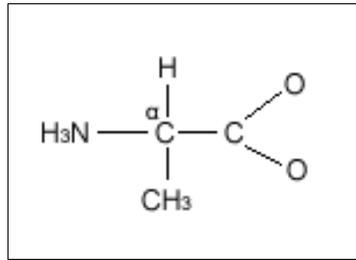
Un aminoácido es una molécula que contiene un grupo carboxilo – COOH y un grupo amino NH_2 – libres. Pueden expresarse como $\text{NH}_2\text{-CHR-COOH}$.



R= grupo orgánico o un OH. Característico para cada ácido.

Los aminoácidos son 20: alanina, arginina, asparagina, aspártico, cisteína, fenilalanina, glicina, glutamato, glutamina, histidina, isoleucina, leucina, lisina, metionina, prolina, serina, tirosina, treonina, triptófano, valina.

Figura 5. Representación de la alanina; aminoácido más pequeño después de la glicina



La estructura primaria: es el ordenamiento o secuencia de aminoácidos en una cadena polipeptídica. La secuencia de los aminoácidos esta determinada por los genes que controlan la síntesis proteica.

Se sintetizan tanto en el citoplasma de la célula como en las mitocondrias, como es el caso de la proteína del Citocromo C. La síntesis se lleva a cabo mediante conversión del DNA a RNA y con ayuda de los Ribosomas, de los RNA: mitocondrial, ribosomal y de transferencia y de moléculas de aminoácidos, mediante la etapa denominada de traducción.

En la síntesis de las proteínas un concepto importante para recordar, es que la base fundamental del mecanismo transcripcional del DNA es que de éste resulta una copia en forma de RNA, en un proceso que podríamos definir como "uno a uno", pues cada nucleótido en la cadena de DNA se copia en la de RNA (con la variante Timina que se copia en Uracilo). Analizar este proceso desde el punto de vista bioinformático no representa un gran problema de modelamiento, lo que sí constituye un problema un poco mayor es la traducción de dicho RNA en proteína pues se encuentran involucrados otros procesos y variables que complejizan la simulación, tales como las ediciones del RNAm y la utilización de un código específico de traducción (código genético), entre otros.

Como se sabe, existen 64 codones (que resultan de la combinación de los 4 nucleótidos, tomados de a tres a la vez), de esta manera un aminoácido (de ahora

en adelante "aa"), puede ser codificado por entre 1 y 6 codones diferentes. Por otra parte el hecho de que los nucleótidos de DNA (traducidos ya en RNA) se lean de a 3 para especificar una proteína, implica que este DNA cuenta con 3 marcos de lectura dependiendo del lugar desde el cual se empieza a leer, es decir desde el nucleótido 1, 2 o 3. Como las dos cadenas de DNA pueden ser copiadas en RNA, entonces tenemos 6 marcos de lectura posibles para una misma proteína, 3 por cada cadena del DNA, si a esto le sumamos la especificidad de uso codónico de algunos organismos, la complejidad del modelamiento por medios informáticos se eleva. Esta es una de las razones principales por las cuales la predicción de proteínas a partir de DNA es un poco más difícil que la simple transcripción de DNA a RNA.

La información genética está en el DNA y esa información se copia en RNA. El Ribosoma lee el RNA y ensambla la proteína. Toda proteína es una cadena de aminoácidos. El Ribosoma lee tres letras del código genético que ha transportado en el RNAt (tres letras codifican un aminoácido y se llama codon) y lo complementa con el anticodon del RNAm (complementario).

Así sucesivamente va leyendo las siguientes letras y ensambla el aminoácido correspondiente al lado del anterior y así hasta que llegue un codon que trae la señal de parada. (Stop: UAA, UAG). El AUG codifica la metionina y sirve como sitio de iniciación, el primer AUG en un RNAm es la región que codifica el sitio donde la translación de proteínas se inicia. (Tabla 1)

Tabla 1. Representación del código genético junto con los aminoácidos que lo componen

		Base 2			
		U	C	A	G
Base 1	U	UUU Fenilalanina	UUC Serina	UAA Tirosina	UGU Cisteína
		UUC Fenilalanina	UCC Serina	UAC Tirosina	UGC Cisteína
		UUA Leucina	UCA Serina	UAA Stop	UGA Stop
		UUG Leucina	UCG Serina	UAG Stop	UGG Triptófano
	C	CUU Leucina	CCU Prolina	CAU Histidina	CGU Arginina
		CUU Leucina	CCC Prolina	CAC Histidina	CGC Arginina
		CUA Leucina	CCA Prolina	CAA Glutamina	CGA Arginina
		CUG Leucina	CCG Prolina	CAG Glutamina	CGG Arginina
	A	AUU Isoleucina	ACU Treonina	AAU Asparagina	AGU Serina
		AUC Isoleucina	ACC Treonina	AAC Lisina	AGC Serina
		AUA Isoleucina	ACA Treonina	AAA Lisina	AGA Arginina
		AUG Metionina	ACG Treonina	AAG Lisina	AGG Arginina
	G	GUU Valina	GCU Alanina	GAU Ácido glutámico	GGU Glicina
		GUC Valina	GCC Alanina	GAC Ácido glutámico	GGC Glicina
		GUA Valina	GCA Alanina	GAA Ácido glutámico	GGA Glicina
		GUG Valina	GCC Alanina	GAG Ácido glutámico	GGG Glicina

La "columna vertebral" del análisis de secuencias en bioinformática lo constituyen los **alineamientos** de dichas secuencias. Comparaciones que se realizan entre dos o más secuencias a la vez con el fin de encontrar entre estas ciertas regularidades o patrones. Como fue mencionado anteriormente, a pesar de que en

teoría es posible inferir fácilmente las propiedades de una proteína dada su estructura primaria, en la práctica resulta un hecho muy complicado. Principalmente esta dificultad surge del hecho, resaltado anteriormente, de que entre el paso de DNA → RNA → PROTEÍNA, surgen algunos procesos que complican en gran medida la inferencia de la secuencia polipeptídica a partir del DNA. Entre estos procesos los más relevantes son:

Al confrontar una secuencia de proteína con el DNA es necesario encontrar en que lugar de dicho DNA se inicia la codificación de ésta, así como el lugar en el cual se detiene. Este problema es particularmente difícil en el humano, pues contiene mucho más DNA del requerido para codificar el número de proteínas existentes en su organismo, así que una porción de DNA al azar tiene muchas probabilidades de no codificar nada.

La presencia de exones en el DNA. Algunas veces este problema se puede minimizar mediante la utilización de cDNA (via RNA), en vez del DNA en sí mismo, pues este cDNA contiene mucho menos ruido (material extraño) . Sin embargo son muchos los casos en los cuales el RNA no se puede analizar y es necesario acudir al DNA original.

- No todo el RNA codifica la proteína. Pueden existir regiones de RNA no codificantes, las cuales pueden llegar a ser más largas que la región codificadora en sí misma.
- No todos los RNAs codifican proteínas (RNAr, RNAt y RNAsn por ejemplo).

De esta manera, no existe una solución global y completa para determinar una secuencia de proteína codificada a partir de una de DNA. Sin embargo, se puede combinar una serie de aproximaciones informáticas (sumadas a procedimientos en laboratorio húmedo), y obtener resultados aceptables.

En este punto es importante realizar un pequeño paréntesis y resaltar que hemos estado hablando de dos aspectos diferentes del análisis de secuencias, uno de ellos es la deducción de la secuencia de aminoácidos (su estructura primaria), a partir de la secuencia de DNA, y por otra parte la inferencia de función (estructura terciaria o cuaternaria) de la proteína a partir de la estructura primaria así el Citocromo C es una proteína conjugada, globular y monomérica. Podemos afirmar que estas dos aproximaciones han sido el principal interés en el desarrollo de la bioinformática en el mundo. [19]

Hablando más concretamente de la inferencia de función a partir de la estructura primaria, generalmente se utilizan dos aproximaciones:

- 1 Homología, compara la estructura primaria de la proteína analizada con la de otras proteínas cuya función ya se conoce.
- 2 “Ab initio”, que busca determinar la estructura que minimice la energía libre. Esta aproximación se realiza generalmente mediante métodos Monte-Carlo o redes neurales.

Sin embargo, incluso después de haber determinado la estructura terciaria de la proteína, aún no se han desarrollado técnicas que permitan inferir las propiedades funcionales de ésta proteína con base en su estructura, siempre se acude a la comparación con proteínas ya conocidas.

Aunque la visión dada hasta el momento pueda parecer bastante pesimista, el análisis de secuencias es, sin embargo, de gran importancia primero porque en sí mismo es un intento hacia la consecución de métodos exitosos en el análisis de secuencias, y su potencial es vasto. En segundo lugar el análisis de secuencias actual permite realizar una serie de estudios que de otra manera serían casi imposibles o por lo menos muy poco efectivos, a saber.

Identificación de la estructura primaria de una proteína a partir de DNA. Es de este punto del cual se ha estado hablando la mayor parte del tiempo. Por ejemplo, si se está tratando de encontrar el lugar del DNA desde el cual se codifica para la proteína, es útil conocer cuales péptidos son codificados por los seis marcos de lectura, así una sección corta que contenga muchos codones de parada, es un candidato muy malo a región codificadora. Obviamente este análisis no dice de donde a donde se codifica la proteína, pero por lo menos permite suponer donde están ocurriendo las cosas. [3]

Búsqueda de secuencias similares. Después de haber determinado una secuencia, ya sea mediante técnicas de laboratorio húmedo o *in silico*, generalmente surge una pregunta más o menos obvia: "*¿existe una secuencia parecida a esta?*", en otras palabras: "*¿ya ha visto alguien una secuencia como esta?*". Para esto se realizan búsquedas de secuencias similares a la nuestra en bases de datos destinadas para estos propósitos: GenBank, EMBL, DDBJ, SwissProt, PIR etc. [1, 2]

Inferencias evolutivas. Se pueden realizar inferencias interesantes comparando las secuencias de proteínas equivalentes (homólogos) entre diferentes organismos, que permitan realizar testimonios al respecto de la evolución de dichas especies a partir de sus ancestros comunes.

Similitud, Identidad y homología. Cuando se habla de Identidad se hace referencia a la ocurrencia de exactamente los mismos residuos en la misma posición de las secuencias alineadas.

El término **Similitud** hace referencia a las concordancias aproximadas entre las secuencias alineadas, estas concordancias tienen significado únicamente cuando se evalúan de acuerdo a una matriz de sustitución. La similitud permite asignar información de una secuencia en otra, nada más. Cuando se tiene una similitud

entre secuencias > 25% a lo largo de la longitud de la secuencia, se puede hablar de alta similitud y generalmente son casi homólogas. Una similitud alta en regiones cortas puede ser del 60% en 20 residuos. Por otra parte generalmente valores $E < 0.02$ (estadísticamente se puede contar con valores < 0.05 , lo que significa que se tiene un 95% de confianza en que el puntaje del alineamiento no se realizó por azar) son significantes e implican homología, mientras valores de $E > 1$ no son significantes. Cuando se dice que A tiene una alta **Homología** con B, decimos que A y B son muy parecidas pero que sus ancestros también, lo primero siempre se sabe pero lo segundo no. En realidad un porcentaje de homología indica "identidad y/o similitud" pero no necesariamente una relación evolutiva. Cuando las secuencias son homólogas comparten un plegamiento y todos los aminoácidos en este plegamiento deben ser conservados. [4, 10, 19]

5.1.7 Métodos y determinación de crecimiento de los microorganismos

El crecimiento de un microorganismo depende de varios factores como: El microorganismo en sí, de acuerdo a sus características biológicas; la composición del medio de cultivo, teniendo en cuenta que las células crecerán más lentamente en un medio mínimo que en un medio rico con todos los nutrientes; las condiciones de incubación: la temperatura de incubación, la concentración de CO₂ en la atmósfera de cultivo, la agitación, etc; la procedencia y características del inóculo: la curva de crecimiento variará dependiendo de que se parta de un cultivo estacionario o de uno en crecimiento exponencial, que se pase un cultivo crecido en MM+Glc a MM+Gal o de MR a MM+Glc o de un medio de malta o de malta modificado.

Por otro lado para lograr medir el crecimiento de una población se pueden utilizar los siguientes métodos:

- **Métodos directos.** Son aquellos en los que se cuenta el nº total de células en un volumen conocido y así se estima el nº total en la población. Entre ellos se encuentran: Recuento directo del número de células al microscopio usando cámaras de recuento; Contadores electrónicos, al suspender las células en un fluido conductor por el que pasa una corriente eléctrica cada una de estas genera un impulso; Recuento de células viables, haciendo diluciones se puede obtener el crecimiento de colonias aisladas procedentes de una única célula; Métodos indirectos, estos se caracterizan por que utilizan parámetros distintos del nº de células, pero que pueden relacionarse de forma directa con éste como la masa celular, actividad metabólica, etc.[22, 23, 24, 25]
- **Métodos indirectos.** Determinación del peso seco, se trata de medir la masa celular presente en un cultivo y estimar el número de células proporcional a ésta; Determinación del nitrógeno celular midiendo el contenido proteico del cultivo; Determinación del DNA; Determinación de una actividad metabólica según el tipo de microorganismo, por ejemplo el consumo de oxígeno (en aerobios), liberación de CO₂ u otro producto de fermentación, etc; Medida de la turbidez del cultivo, basándose en la capacidad de las células y partículas en general de absorber y/o dispersar la luz que incide sobre ellas. [22]

La ecuación que permite calcular el tiempo requerido para que una célula microbiana se divida dando lugar a dos células, este se denomina tiempo de generación G y es pues el tiempo que necesita una población para duplicar el número de células. Si debe considerar un tiempo inicial, T_o, y otro final, T_f, se tiene que la expresión de G es: $G = (T_f - T_o) / n$, donde: n es el número de generaciones, y viene dado por el número de ciclos de división ocurridos en la población en un tiempo dado. [22, 26,27]

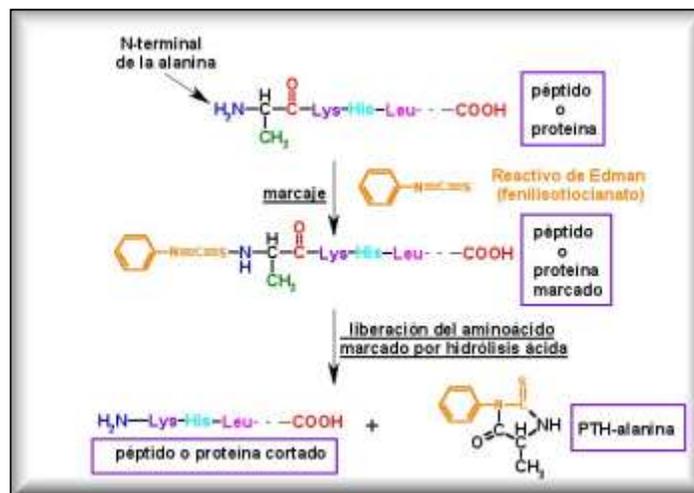
5.1.8 Métodos de secuenciación. Para realizar la secuenciación de proteínas e igualmente determinar la cantidad de cada aminoácido en éstas, es preciso empezar por someterla a hidrólisis, con ácidos, álcalis o enzimas, según el aminoácido que vaya a determinarse. Existen algunos métodos para la identificación de aminoácidos, entre ellos podemos encontrar: [23, 25, 28, 29, 30, 31]

- **Métodos químicos.** El primer método, y desde el punto de vista cuantitativo el método mas sencillo, consiste en aislamiento directo de los aminoácidos o de sus derivados, a base de diferencias en sus solubilidades.
- **Métodos microbiológicos.** Este método puede llegar a demostrar una relación directa entre la cantidad de cada uno de aminoácidos en un medio y la capacidad de crecimiento del microorganismo. Usando una curva de crecimiento *patrón* puede mostrar esta relación de acuerdo a cantidades específicas para el aminoácido.
- **Método de disolución isotópico.** Este método puede ser extraordinariamente exacto, pero requiere de aminoácidos marcados isotópicamente, es muy laborioso y a veces, como en el caso de isótopos estables, requiere del uso de equipo de ensayo costoso (espectrómetro de masas).
- **Análisis cromatográfico.** Con mucha diferencia, el método mas útil y mas preciso para determinar la cantidad de cada aminoácido en un hidrolizado de proteína es la separación de los aminoácidos en una columna de adsorbente, seguida por la elusión consecutiva de cada ácido.
- **Método de Sanger.** Las proteínas pueden ser caracterizadas por un análisis ulterior de los residuos N y C-terminal. La mayoría de las proteínas tienen dos

extremos: un amino (N-terminal) y un carboxilo (C-terminal). Este es históricamente el más importante; sin embargo, no es el más adecuado para el objeto de estudio de este proyecto de grado porque el reactivo reacciona lentamente y los aminoácidos son difíciles de detectar.

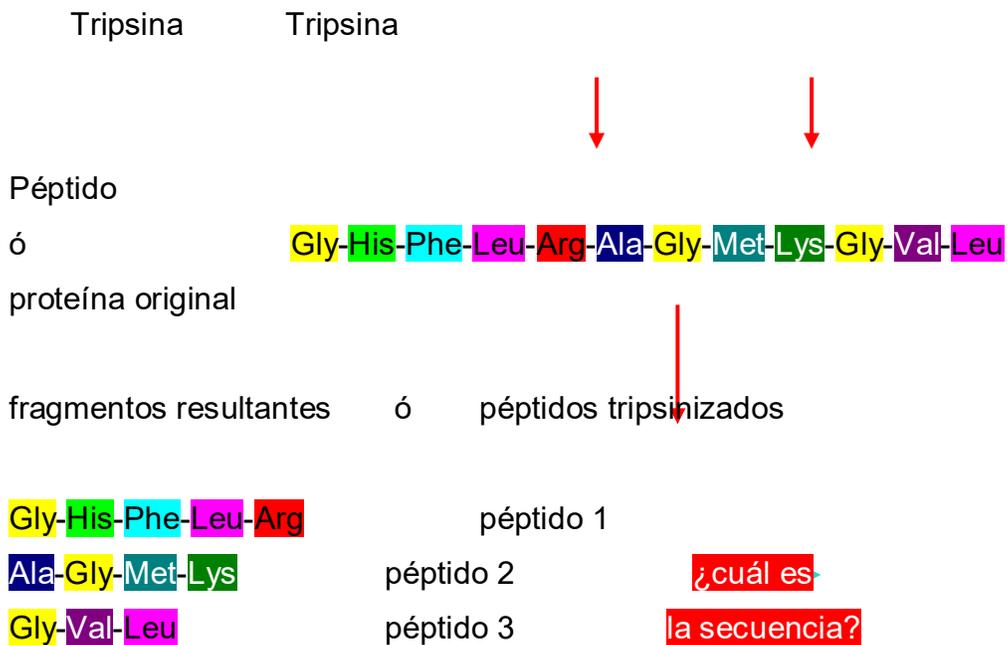
- Método de Edman.** Este método es el elegido para la realización de esta investigación. En este procedimiento se utiliza el reactivo fenilisotiocianato, que marca y libera tan solo el residuo N-terminal sin destruir el resto de la proteína. El aminoácido del N-terminal liberado se puede aislar e identificar; este se conoce como una feniltiohidantoina. Esto deja al descubierto un nuevo residuo N-terminal de la proteína; el producto proteico puede reciclarse varias veces a través de la reacción de Edman, y en cada paso pierde un aminoácido del N-terminal. En la actualidad se utiliza un analizador automático o secuenciador, que puede programarse para controlar la mezcla adecuada de reactivos, la separación de los productos y la identificación de los derivados feniltiohidantoinicos de los aminoácidos N-terminales.[32, 33, 34, 35] (Figura 6)

Figura 6. Representación de la secuenciación por el N-terminal.



Por ejemplo la tripsina que es una enzima digestiva que se produce por el páncreas e *in vitro* es utilizada comúnmente para romper específicamente a los

polipéptidos de gran tamaño, en fragmentos susceptibles de ser secuenciados. La tripsina rompe a los enlaces peptídicos que se encuentran en el lado carboxilo de una lisina o arginina. Los fragmentos resultantes se denominan péptidos tripsinizados: [36, 37, 38] (Figura 7)



El tratamiento de una proteína con bromuro de cianógeno se utiliza también comúnmente para romper de manera específica a las proteínas. El bromuro de cianógeno rompe la cadena en el lado carbonilo de un residuo de metionina. Este agente como la tripsina es muy específico y usualmente lleva a la producción de un número limitado de fragmentos: [39, 40]

Figura 7. Representación de la ruptura de una proteína por medio de Bromuro de cianógeno.

Bromuro de cianógeno



proteína original

fragmentos resultantes



Gly-His-Phe-Leu-Arg-Ala-Gly-Met

Lys-Gly-Val-Leu

péptido B

péptido A

¿cuál es

la

secuencia?

5.2 CONCEPTOS INFORMÁTICOS

5.2.1 Sistema. De manera general, el diccionario de la Real Lengua Española define un sistema como *“el conjunto de cosas que ordenadamente relacionadas entre sí, contribuyen a un determinado objetivo”*. A partir de dicha definición se puede concluir las partes principales de un sistema: los componentes del sistema, las relaciones entre ellos, que determinan la estructura del sistema y el objetivo del sistema. Además se destacan otros componentes como el entorno del sistema (que es todo aquello que lo rodea) y los límites del sistema (que es la frontera entre lo que es el sistema y lo que constituye el entorno). Para las técnicas del análisis de los sistemas se sigue la definición del enfoque sistémico o teoría general de sistemas. [20, 21]

5.2.2 Sistema de Información. Un sistema de información esta constituido, por las bases de datos, los programas de aplicación, los procedimientos manuales y automatizados, e incluye los sistemas computacionales que realizan procesamiento. [20, 21]

Uno de los conceptos mas complejos define un sistema de información como un conjunto formal de procesos que, operando sobre una colección de datos estructuradas según las necesidades de la empresa recopilan, elaboran y

distribuyen la información (o parte de ella) necesaria para las operaciones de dicha empresa y para las actividades de dirección y control correspondiente (decisiones) para desempeñar su actividad de acuerdo a su estrategia de negocios. [20, 21]

- **Elementos de un sistema de información.** Los elementos que componen un sistema de información son: [20]
 - **Los procedimientos y las prácticas habituales de trabajo.** Se siguen al ejecutar toda clase de actividades necesarias para el buen funcionamiento de la empresa.
 - **La información.** Es el elemento fundamental del sistema, su razón de ser.
 - **Las personas o usuarios.** Son los individuos o unidades de la organización que introducen, manejan o usan la información.
 - **El equipo de soporte.** Para la comunicación, el procesamiento y almacenamiento de la información.

5.2.3 Lenguaje UML. El Lenguaje Unificado de Modelado (Unified Modeling Language, UML) es un lenguaje estándar, sencillo y muy expresivo que permite cubrir una visualización general del software orientado a objetos a desarrollar. UML (Lenguaje Unificado de Modelado) puede utilizarse para visualizar, construir, especificar y documentar las características generales de un sistema. Su utilización es independiente del lenguaje de programación y de las características de los proyectos, ya que UML (Lenguaje Unificado de Modelado) ha sido diseñado para modelar cualquier tipo de proyectos. [41]

El lenguaje UML nos provee de diferentes diagramas los cuales hacen más fácil la comprensión del sistema a crear. Estos diagramas son:

- **Diagramas de casos de uso:** los diagramas de casos de uso se emplean para modelar la vista de casos de uso de un sistema. Estos diagramas son importantes para visualizar, especificar y documentar el comportamiento de un elemento.

Los diagramas de casos de uso se emplean para visualizar el comportamiento de un sistema (los servicios visibles externamente que proporciona el sistema en el contexto de su entorno), de forma que los usuarios puedan comprender como utilizar ese elemento y que los desarrolladores puedan implementarlo. Normalmente un diagrama de casos de uso contiene:

- **Casos de usos:** Un caso de uso es una descripción de las acciones de un sistema desde el punto de vista del usuario. Representados en una elipse. Cada caso de uso contiene un nombre que indica su funcionalidad.
 - **Actores:** los actores son los usuarios que interactúan con el sistema. Se representan con un muñeco.
 - **Relaciones de dependencia, generalización y asociación:** son las interacciones entre los actores y los casos de uso.
- **Diagrama de clases:** Los diagramas de clases se utilizan para modelar la vista de diseño estática de un sistema. Esto incluye modelar el vocabulario sistema, modelar las colaboraciones o modelar esquemas.
Los diagramas de clases dan una vista que soporta principalmente los requisitos funcionales de un sistema, los servicios que el sistema debe proporcionar a sus usuarios finales.

Los diagramas de clases contienen normalmente los siguientes elementos:

- **Clases:** una clase es una descripción de un conjunto de objetos que comparten los mismos atributos, operaciones, relaciones y semántica. Una clase esta representada por un rectángulo que dispone de tres apartados, el primero para indicar el nombre, el segundo para los atributos y el tercero para los métodos.
 - **Cada clase debe tener un nombre que la distinga de las otras clases.** Un atributo representa alguna propiedad del elemento que se esta modelando que es compartida por todos los objetos de esa clase. un método u operación es la implementación de un servicio que puede ser requerido a cualquier objeto de la clase para que muestre un comportamiento.
 - **Interfaces:** son las vistas con las que el usuario interactúa con el sistema.
 - **Relaciones de dependencia, generalización y asociación: existen tres tipos de relaciones especialmente importantes:** dependencia, que representan relaciones de uso entre clases; generalizaciones, que conectan clases generales con sus especializaciones, y asociaciones, que representan relaciones estructurales entre objetos.
- **Diagramas de secuencias:** Es un diagrama que destaca la ordenación temporal de los mensajes, estos diagramas se utilizan para modelar los aspectos dinámicos de un sistema. Los diagramas de secuencias pueden utilizarse para visualizar, especificar, construir y documentar la dinámica de una sociedad particular de objetos, o se pueden utilizar para modelar un flujo de control particular de un caso de uso. Estos diagramas son útiles debido a que permiten observar la vida de los objetos en el sistema, identificar llamados o posibles errores que se puedan presentar en el modelado estático, que imposibiliten el flujo de información o de llamadas entre los componentes del sistema.

Normalmente los diagramas de secuencias contienen:

- **Línea de vida de un objeto:** representa la vida del objeto durante la interacción. En un diagrama de secuencias, un objeto se representa como una línea vertical punteada con un rectángulo de encabezado y con rectángulos a través de la línea principal que denotan la ejecución de métodos.
- **Activación:** muestra el periodo de tiempo en el cual el objeto se encuentra desarrollando alguna operación, bien sea por si mismo o por medio de alguno de sus atributos.
- **Mensaje:** el envío de mensajes entre objetos se denota mediante una línea sólida dirigida, desde el objeto que emite el mensaje hacia el objeto que lo ejecuta.

5.2.4 Plataforma de Desarrollo Windows. Windows es un sistema operativo que fue creado y lanzado al mercado por la empresa Microsoft el 20 de Noviembre de 1985. La característica particular que siempre ha mantenido el sistema operativo Windows hasta el momento es el manejo de una interfaz gráfica fácil de usar, la cual tuvo un gran cambio al incluirse los menús desplegables.

Con el paso del tiempo, el sistema operativo de Microsoft ha evolucionado, actualmente existen varias versiones de acuerdo a las necesidades del usuario, entre ellas podemos encontrar: Windows 95, Windows 98, Windows Me, Windows 2000, Windows XP entre otros. [42]

5.2.5 Lenguajes e Interfaces de Desarrollo

- **Java.** Java se ha desarrollado a partir del concepto de un código ejecutable independiente de la plataforma. Los investigadores de Sun Microsystems lo crearon para que fuera un poderoso sistema de programación y transferencia de información utilizable en la Web. Java es un lenguaje de programación diseñado para suministrar contenidos ejecutables a las redes informáticas.

Java surgió en 1991 cuando un grupo de ingenieros de Sun Microsystems trataron de diseñar un nuevo lenguaje de programación destinado a electrodomésticos [6]. La reducida potencia del cálculo y memoria de los electrodomésticos llevo a desarrollar un lenguaje sencillo capaz de generar código de tamaño muy reducido.

Debido a la existencia de distintos tipos de CPUs y a los continuos cambios, era importante conseguir una herramienta independiente del tipo de CPU utilizada. Desarrollaron un código “neutro” que no dependía del tipo de electrodoméstico, el cual se ejecutaba sobre una “maquina virtual” denominada **Java virtual Machine (JVM)**. Era la JVM quien interpretaba el código neutro convirtiéndolo a código particular de la CPU utilizada. Esto permitía lo que luego se ha convertido en el principal lema del lenguaje: “Write Once, Run Everywhere”. A pesar de los esfuerzos realizados por sus creadores, ninguna empresa de electrodomésticos se interesó por el nuevo lenguaje.

Como lenguaje de programación para computadores, Java se introdujo a finales de 1.995. La clave fue la incorporación de un intérprete de Java en la versión 2.0 del programa Netscape Navigator, produciendo una verdadera revolución en Internet. Java 1.1 apareció a principios de 1997, mejorando sustancialmente la primera versión del lenguaje. Java 1.2, mas tarde rebautizado como Java 2, nacio a finales de 1998.

Java es el lenguaje ideal para aprender la informática moderna, porque incorpora todos estos conceptos de un modo estándar y mucho más sencillo. El principal objetivo del lenguaje Java es llegar a ser el “nexo universal” que conecte a los usuarios con la información, esta está situada en el ordenador local, en un servidor Web, en una base de datos o en cualquier otro lugar.

Java permite fácilmente el desarrollo tanto de arquitecturas cliente-servidor como de aplicaciones distribuidas, consistentes en crear aplicaciones capaces de conectarse a otros ordenadores y ejecutar tareas en varios ordenadores simultáneamente, repartiendo por lo tanto el trabajo.

Java hace realidad la animación, la iteración (ciclo completo de desarrollo que produce una versión (interna o externa) de un producto ejecutable, que constituye un subconjunto del producto final en desarrollo), el procesamiento, las aplicaciones distribuidas y nuevas formas de comunicación, además posee una gran flexibilidad para ser implementado sobre diferentes plataformas. Java es una herramienta no comercial y de fácil acceso, cuya principal característica es el desarrollo orientado a objetos. [43]

- **Netbeans – Jcreator.** Netbeans IDE es un gran entorno de desarrollo para Java, hecho en Java y open-source el cual soporta una variedad de lenguajes como Java, C, C++, entre otros. Netbeans es un software libre que facilita la creación de interfaces de manera rápida y eficiente permitiendo agilizar en factor de tiempo el desarrollo de la aplicación. Netbeans es un entorno de desarrollo, una herramienta para programadores para escribir, compilar, corregir errores y para ejecutar programas.[44]

Jcreator es un ambiente de desarrollo integrado para Java. Este programa permite editar, compilar, ejecutar y hacer “debug”, es decir sacar “bugs” o bichos, que son los errores en un programa.[45]

5.2.6 Microsoft Access. Microsoft Access es un sistema gestor de bases de datos relacionales (SGBD). Este es un programa para la gestión de información que permite diseñar las estructuras necesarias para almacenar la información y los medios para su introducción y explotación. Una base de datos suele definirse como un conjunto de información organizada sistemáticamente. [46]

En la terminología propia de las bases de datos hay tres conceptos claves dentro de las tablas, los cuales son:

- Un **campo** es cada uno de los tipos de datos que se van a usar. Se hace referencia a los campos por su nombre.
- Un **registro** está formado por el conjunto de información en particular.
- Un **dato** es la intersección entre un campo y un registro.

ELEMENTOS DE ACCESS [46]

- **Tablas.** Las tablas con el componente básico o elemental de las bases de datos. O lo que es lo mismo, una base de datos está principalmente compuesta por varias tablas relacionadas. Las tablas contienen datos sobre algo o alguien, proveedores, clientes, libros en una biblioteca, compras, ventas, etc.
- **Consultas.** Las consultas son preguntas que un usuario hace a la base de datos. Con ellas puede obtener información de varias tablas y con la estructura que más le interese. Además, las consultas pueden archivarse de forma que la próxima vez que se quiera hacer la misma pregunta no tendrá que volver a plantearla, será suficiente con llamar a la consulta

previamente creada. La importancia de las consultas es enorme, de hecho es la potencia de esta herramienta la que permite que los gestores de base de datos sean casi imprescindibles en nuestro trabajo diario.

- **Formularios.** Los formularios son un mecanismo que facilita enormemente la operatoria general con tablas, principalmente a la hora de mostrar, introducir y modificar datos. Un uso adecuado de éstos redundaría bastante en el nivel de manejabilidad de una aplicación o de un sistema de información desarrollado con Access.
- **Informes.** Los informes permiten presentar la información con una apariencia altamente profesional a la hora de imprimir nuestros datos.
- **Páginas de acceso a datos.** Una página de acceso a datos es una página Web que se puede utilizar para agregar, modificar, ver o manipular datos actuales en una base de datos de Microsoft Access o de SQL Server. Se pueden crear páginas que se utilizarán para especificar y modificar datos, de manera similar a los formularios de Access. También se pueden crear páginas que muestren registros agrupados jerárquicamente, de manera similar a los informes de Access.
- **Macros.** Las macros son un mecanismo de automatización de Microsoft Access. Utilizando éstas es posible automatizar tareas repetitivas eliminando la posibilidad de introducir errores de operación y liberando tiempo para emplearlo en otras actividades. Podemos decir que una macro no es más que una lista de tareas que queremos que Access lleve a cabo automáticamente.
- **Módulos.** Los módulos son objetos donde se almacena código escrito en lenguaje de programación denominado Access Basic.

5.3 HERRAMIENTAS BIOINFORMATICAS

5.3.1 Rasmol. Rasmol es un programa de gráficos moleculares que permite la visualización de cualquier tipo de estructura molecular definida. Además, es un programa de gran versatilidad que permite no sólo obtener diferentes modelos de representación de una molécula, sino que también permite colorear, resaltar y seleccionar átomos y/o regiones particulares, por lo que facilita el aprendizaje de los fenómenos estructurales y de su relación con la actividad biológica.

Igualmente Rasmol permite construir animaciones sobre una molécula, para presentar diferentes características de la molécula como la estructura secundaria, la conformación de dominios, la localización de sitios activos, los cambios estructurales provocados por mutaciones de la macromolécula, entre otros. [47] (Figura 8, 9)

Figura 8. Representación de una molécula 2MCG en modelo Wireframe

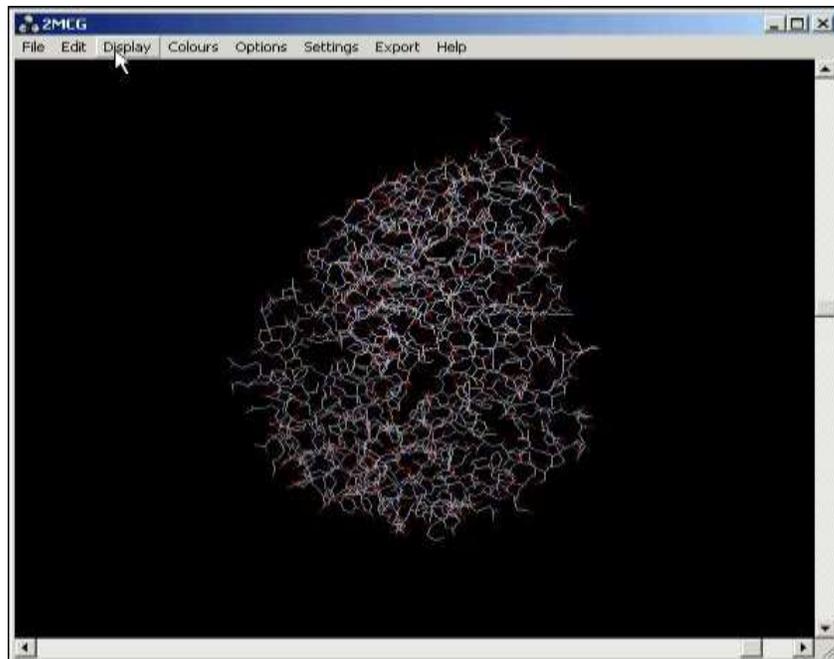
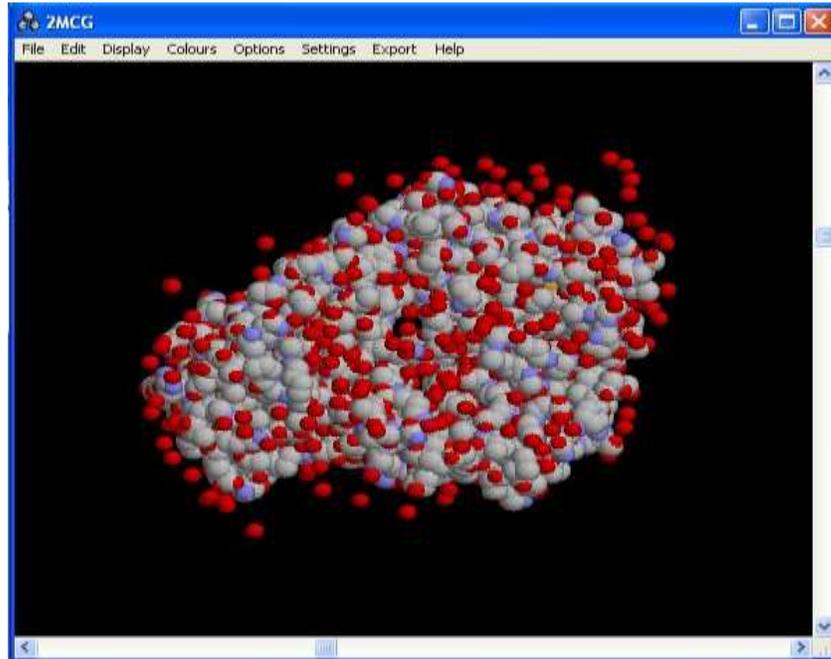


Figura 9. Representación de una molécula 2MCG en modelo Spacefill



5.3.2 Treeview. TreeView es un programa simple que exhibe los árboles del origen de la filogenia en un ambiente gráfico, realizando procesos de estructuración y análisis de datos de varias maneras, permitiendo un análisis experimental sobre el origen de la filogenia del objeto de estudio.

TreeView proporciona una visión del árbol del origen de la filogenia , contenido en archivos con formatos de otros programas como: NEXO, de PHYLIP, de Clustal W o de Clustal X, entre otros. [48] (Figura 10, 11)

Figura 10. Representación del árbol de la filogenia de primates en Unrooted

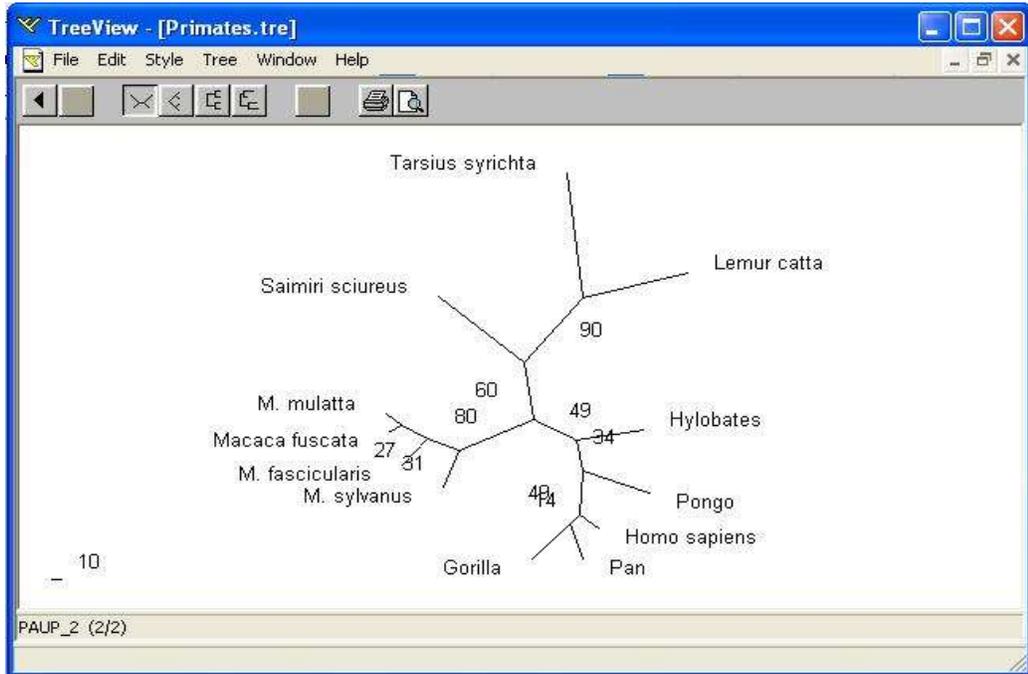
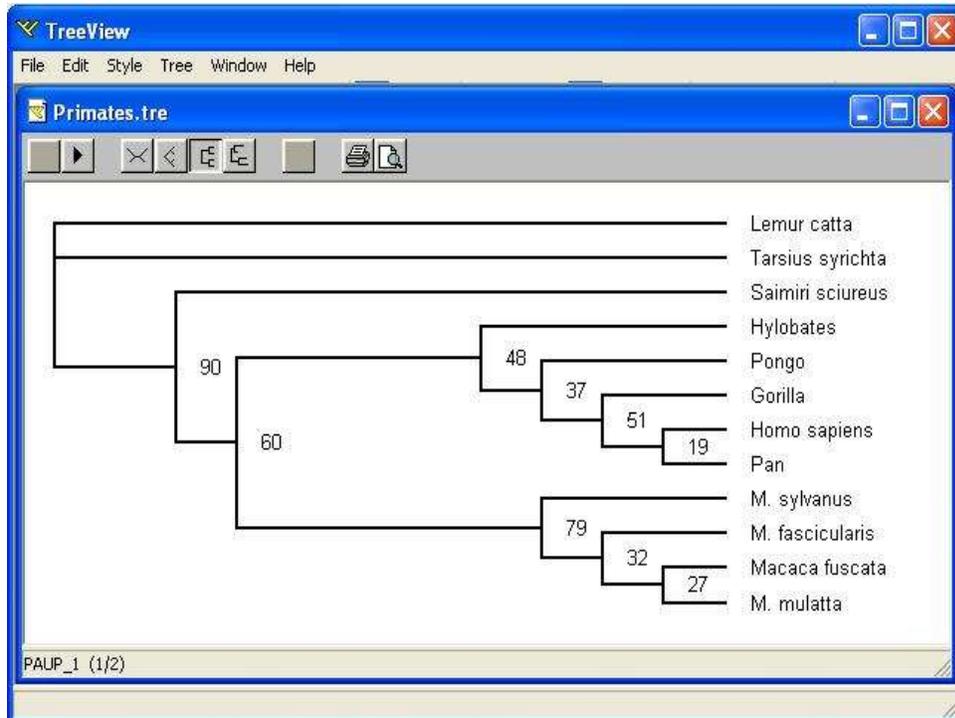


Figura 11. Representación del árbol de la filogenia de primates en Rectangular cladogram



5.3.3 Fasta. Fasta permite comparar un par de proteínas o secuencias de ADN, como también la comparación de una sola proteína o secuencia de ADN a una secuencia que se encuentra almacenada en una base de datos o biblioteca.

Fasta puede identificar secuencias largas que sean poco semejantes o secuencias altamente divergentes durante el proceso de comparación, con el fin de determinar de esta manera la existencia de semejanzas y homologías entre las secuencias completas de proteínas o de ADN almacenadas en las bases de datos. [49] (Figura 12, 13)

Figura 12. Interfaz modo gráfico donde solicita secuencias en Fasta

The image shows a web interface titled "Compare Two Sequences with FASTA". At the top, it displays the IP address "200.21.227.66" and the URL "fasta.bioch.virginia.edu". Below this, there is a section for choosing the program and sequences to compare. The "Program" dropdown menu is set to "FASTA: protein.protein DNA.DNA". There are radio buttons for "Protein", "DNA (both-strands)", "DNA (forward only)", and "DNA (rev-comp only)", with "Protein" selected. A checkbox for "Query post-trans modifications" is present and unchecked. The "Enter first (query) sequence" section has a dropdown menu set to "FASTA format" and a "Subset range" input field. Below this is a large text area for entering the query sequence. To the right of this area are two links: "Entrez protein sequence browser" and "Entrez DNA sequence browser". The "Enter second (library) sequence" section has a dropdown menu set to "FASTA format" and another large text area for entering the library sequence. At the bottom right, there is a "Compare Sequences" button.

Figura 13. Archivo generado como resultado del software.

```
# "C:\Documents and Settings\Usuario 1\Escritorio\fasta\fasta34.exe"
FASTA searches a protein or DNA sequence data bank
version 3.4t21 May 14, 2003
Please cite:|
W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

fasta.tre, 379 aa
vs fasta1.tre library

379 residues in 1 sequences
Altschul/Gish params: n0: 379 Lambda: 0.158 K: 0.019 H: 0.100

FASTA (3.45an0 Mar 2002) function [optimized, BL50 matrix (15:-5)] ktup: 2
join: 37, opt: 25, open/ext: -10/-2, width: 16
The best scores are:
gi|532319|pir|TVFV2E|TVFV2E envelope protein (379) 2442 563.0 4.6e-165

>>gi|532319|pir|TVFV2E|TVFV2E envelope protein (379 aa)
initn: 2442 init1: 2442 opt: 2442 Z-score: 2995.8 bits: 563.0 E(): 4.6e-165
Smith-Waterman score: 2442; 100.000% identity (100.000% ungapped) in 379 aa o

10 20 30 40 50 60
gi|532 ELRLRYCAPAGFALLKCNDADYDGFKTNCNSVSVVHCTNLMNTT VTTG LLLNGSYSENRT
:
:
:
gi|532 ELRLRYCAPAGFALLKCNDADYDGFKTNCNSVSVVHCTNLMNTT VTTG LLLNGSYSENRT
10 20 30 40 50 60
```

5.3.4 Clustal X. Clustal X proporciona un ambiente integrado para realizar alineaciones múltiples de la secuencia y del perfil para analizar resultados. Además le permite al usuario seleccionar las opciones requeridas para la alineación múltiple tradicional de la secuencia y del perfil. También permite modificar el orden de la alineación, seleccionar un subconjunto de secuencias para ser alineadas y puede seleccionar una gama secundaria de la alineación para ser realineado y para ser insertado nuevamente dentro de la alineación original. [50] (Figura 14, 15)

Figura 14. Interfaz de Clustal X, mostrando resultado en alineamiento múltiple

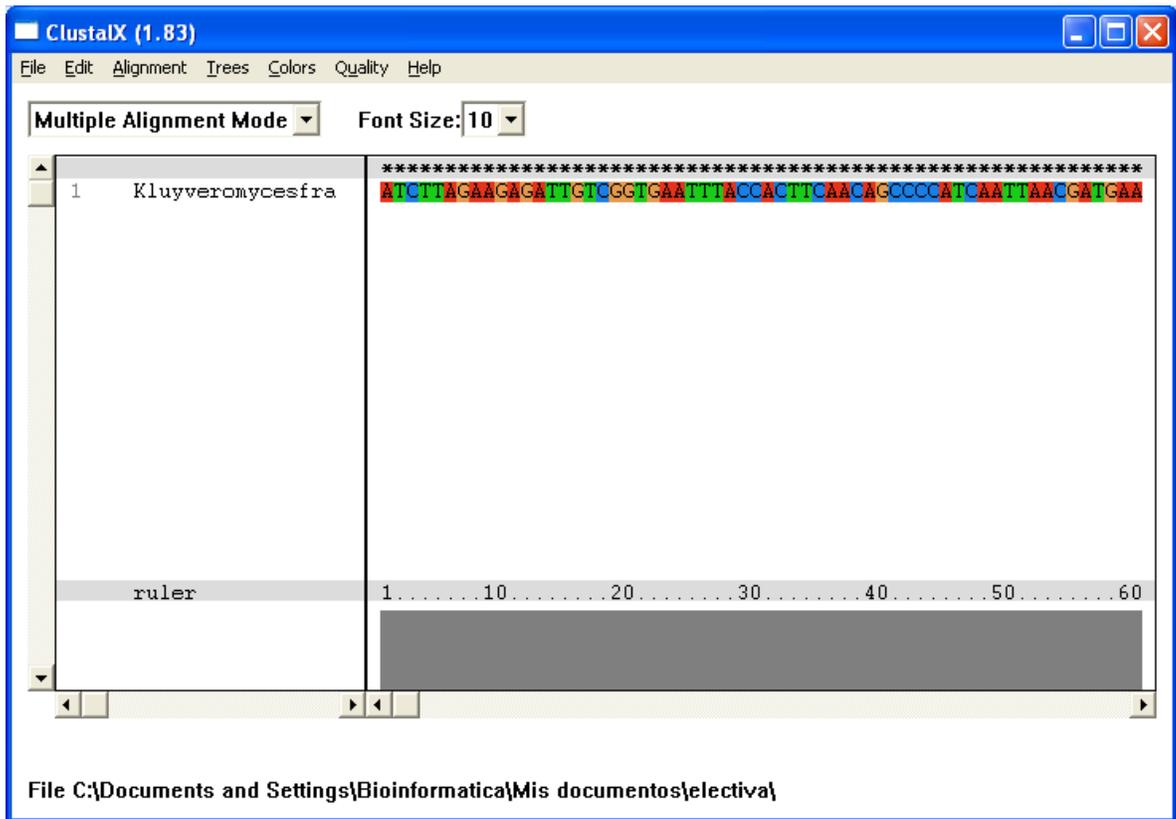
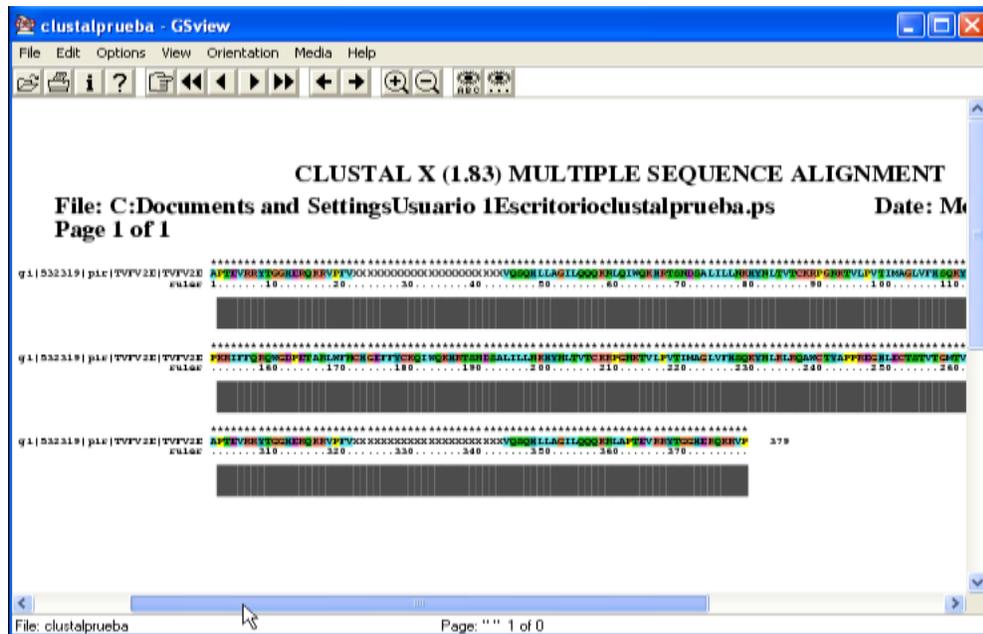
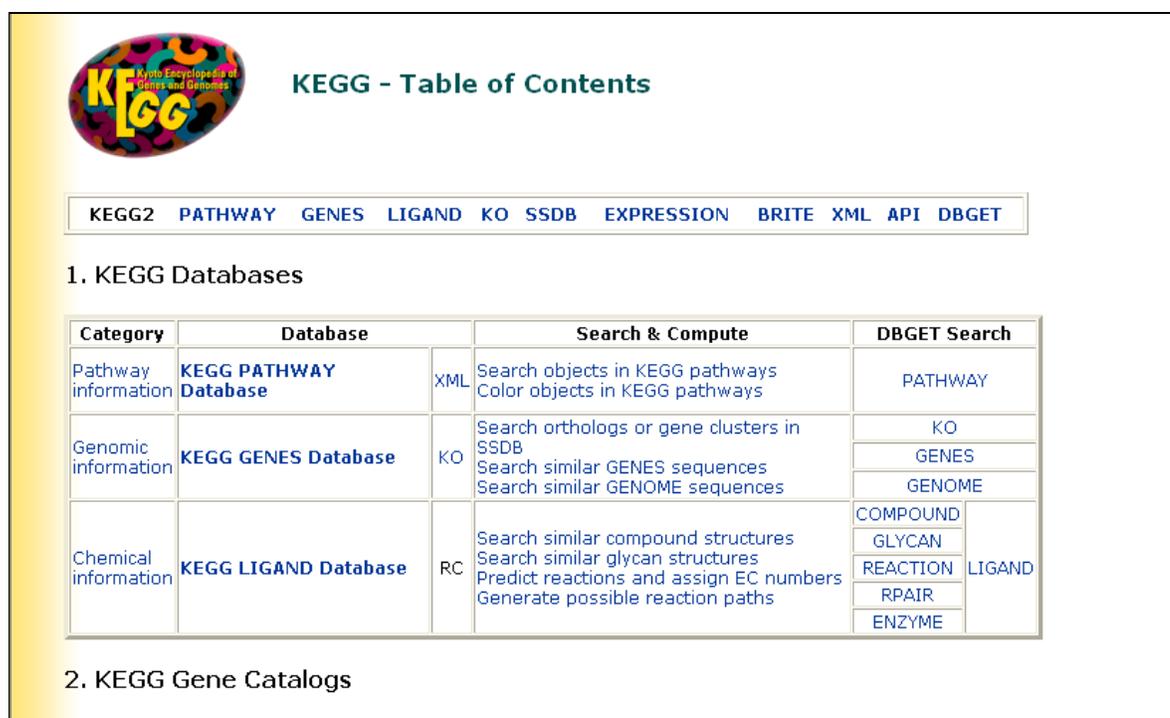


Figura 15. Archivo de resultado para impresión



5.3.5 Kegg. En mayo 1995 se inicio un proyecto llamado Kegg (Enciclopedia de Kyoto de genes y del genoma). El objetivo primario de Kegg es automatizar el conocimiento actual de interacciones moleculares, conocer caminos metabólicos, caminos reguladores y secuencias moleculares. Kegg mantiene los catálogos genéticos para todos los organismos que se han estudiado ligando cada producto de los genes a un componente en el pathway. Kegg también organiza una base de datos de todos los compuestos químicos en células vivas y liga cada uno de estos compuestos a otro componente. Los datos en Kegg son representados por diagramas gráficos (los mapas del camino y los mapas del genoma) o los textos jerárquicos (los catálogos de genes y los catálogos moleculares). [51] (Figura 16)

Figura 16. Interfaz en línea de KEGG



The screenshot shows the KEGG website interface. At the top left is the KEGG logo, which is a colorful, abstract shape with the letters 'KEGG' and the text 'Kyoto Encyclopedia of Genes and Genomes'. To the right of the logo is the title 'KEGG - Table of Contents'. Below the title is a navigation menu with the following items: KEGG2, PATHWAY, GENES, LIGAND, KO, SSDB, EXPRESSION, BRITE, XML, API, and DBGET. Below the navigation menu is the section '1. KEGG Databases'. Under this section is a table with the following structure:

Category	Database		Search & Compute	DBGET Search
Pathway information	KEGG PATHWAY Database	XML	Search objects in KEGG pathways Color objects in KEGG pathways	PATHWAY
Genomic information	KEGG GENES Database	KO	Search orthologs or gene clusters in SSDB Search similar GENES sequences Search similar GENOME sequences	KO GENES GENOME
Chemical information	KEGG LIGAND Database	RC	Search similar compound structures Search similar glycan structures Predict reactions and assign EC numbers Generate possible reaction paths	COMPOUND GLYCAN REACTION RPAIR ENZYME LIGAND

Below the table is the section '2. KEGG Gene Catalogs'.

5.3.6 Yeast Microarray Global Viewer. Yeast Microarray Global Viewer (YMGV) es una base de datos que proporciona una vista sintética de los perfiles genéticos de levaduras. El YMGV exhibe una representación gráfica de las variaciones en la

expresión genética para cada uno de los experimentos genómicos realizados, permitiendo la recuperación rápida de las condiciones experimentales que tienen un efecto sobre la expresión de estos genes. Además, proporciona herramientas a grupos aislantes de genes que comparten perfiles similares de la transcripción dentro un subconjunto definido de experimentos, a través de herramientas estadísticas que permiten mostrar un gravamen crítico de los datos publicados. [52] (Figura 17, 18)

Figura 17. Interfaz en línea de YMGV

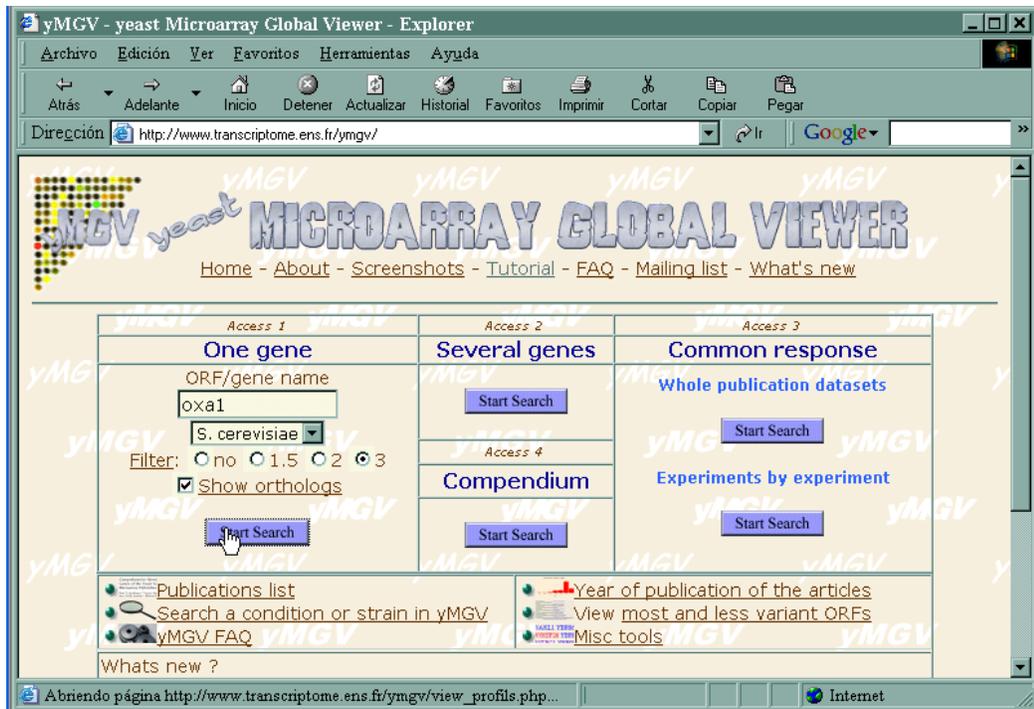
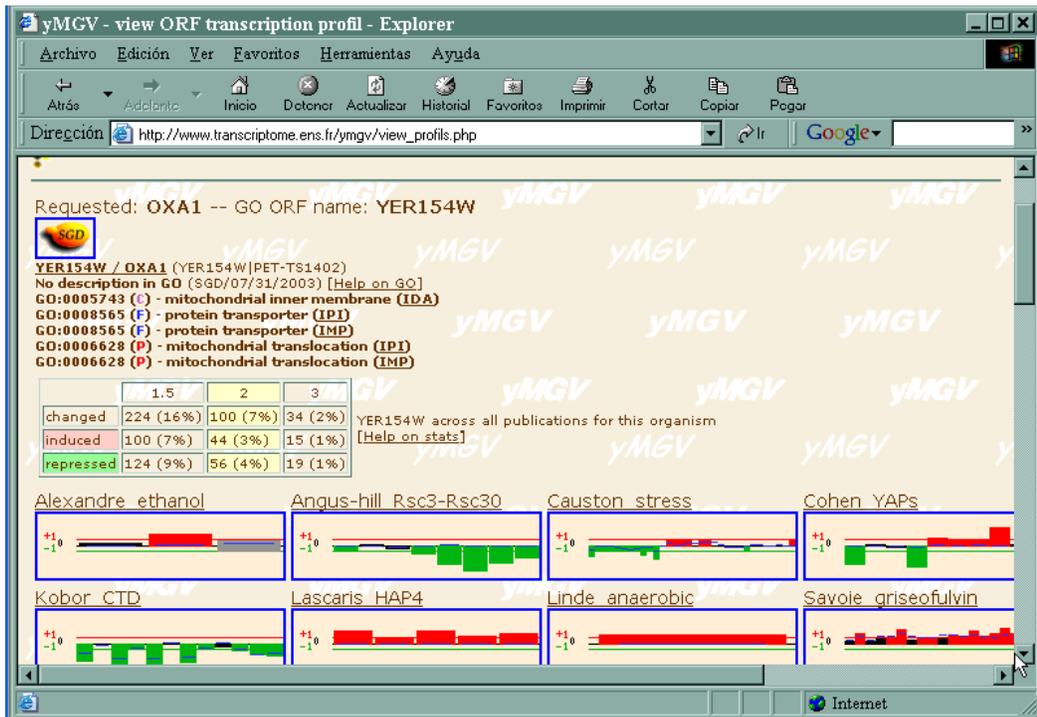


Figura 18. Resultados arrojados por el software.



5.3.7 SRS (Embl/Ebi). El SRS (Sistema de recuperación de la secuencia), es el sistema de aplicaciones más utilizado en el mundo para acceder y recuperar información de las bases de datos de secuencias biológicas, genómicas y demás bases de datos relacionadas. [53]

SRS está instalado en más de 30 servidores públicos a través del mundo y da actualmente el acceso a más de 300 bases de datos. SRS utiliza bancos de datos moleculares biológicos que se encuentran disponibles en formato ASCII (estándar para el intercambio de datos). Además, casi todos los usos en biología molecular (tal como Blast , PHRAG, TACG, entre otros.), tienen como producto de salida un archivo de texto. [53] (Figura 19, 20)

Figura 19. Interfaz de la bases de datos de SRS

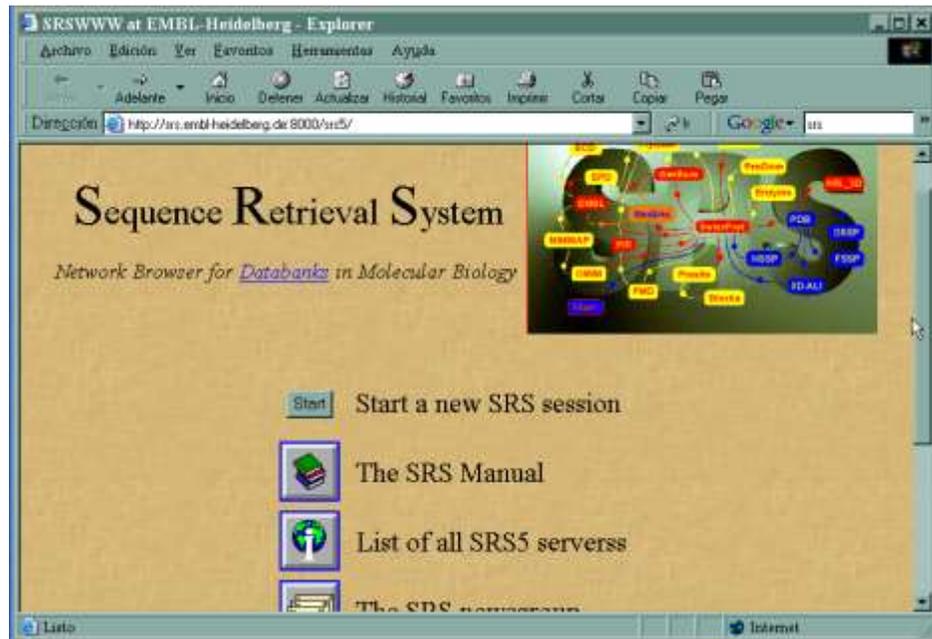
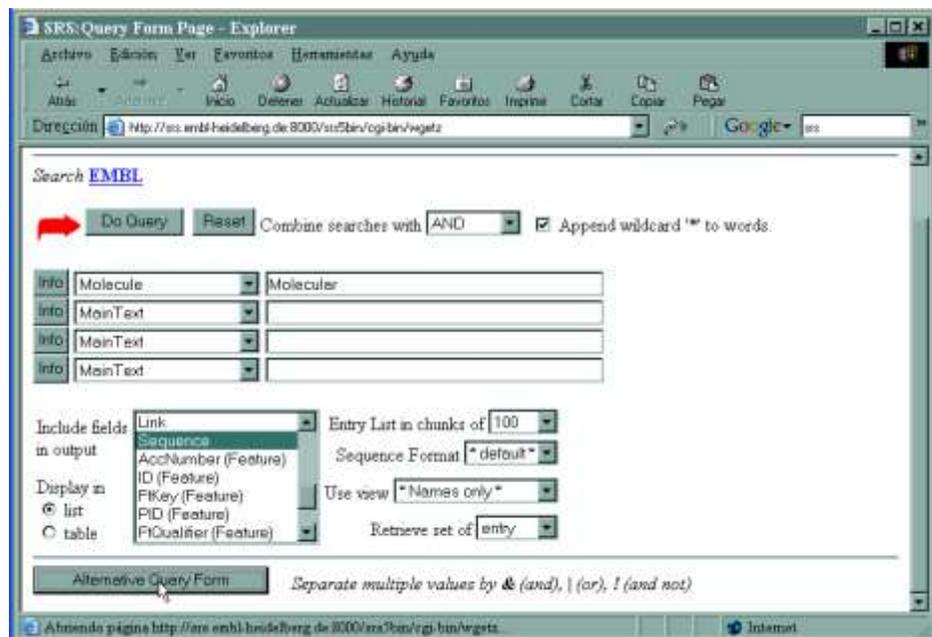


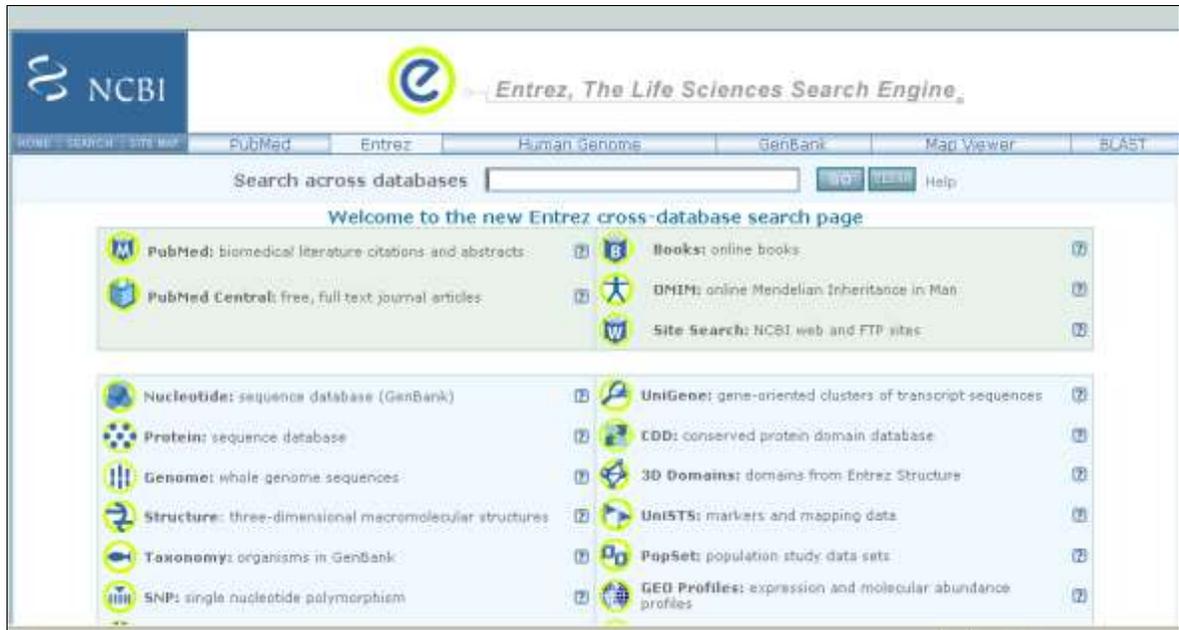
Figura 20. Interfaz para visualización de resultado



5.3.8 Entrez. Entrez fue desarrollado por el Centro Nacional para la Información de la Biotecnología (NCBI). Entrez es un sistema de búsqueda que integra

información de bases de datos en NCBI. Estas bases de datos incluyen secuencias de nucleótidos y proteínas, estructuras macromoleculares y cromosomas completos a través de PubMed.[54] (Figura 21)

Figura 21. Interfaz de la bases de datos de Entrez.



5.2.9 NCBI Blast Server. Blast (herramienta básica de la búsqueda de la alineación) utiliza un algoritmo heurístico en la búsqueda de la alineación y es empleado por los programas: blastp, blastn, blastx, tblastn, y tblastx.

Los programas de Blast fueron adaptados para buscar las semejanzas dentro de una secuencia. Los cinco programas de Blast descritos aquí realizan las tareas siguientes: el blastp compara la secuencia de un aminoácido contra una base de datos de la secuencia de la proteína; el blastn compara una secuencia del nucleotido contra una base de datos de la secuencia del nucleotido; el blastx compara los productos conceptuales de la traducción de una secuencia del nucleotido contra una base de datos de la secuencia de la proteína; el tblastn

compara una secuencia de la proteína contra una base de datos de la secuencia del nucleotido traducida dinámicamente en seis marcos de lectura; el tblastx compara las traducciones de una secuencia del núcleo contra una base de datos de la secuencia del nucleotido. [55] (Figura 22)

Figura 22. Interfaz de la bases de datos de Blast.

The image shows the NCBI BLAST website interface. At the top left is the NCBI logo. To the right, the word "BLAST" is displayed in large blue letters. Below the logo is a navigation bar with links for PubMed, Entrez, BLAST, OMIM, Taxonomy, and Structure. A yellow banner across the top of the main content area reads: "NEW 12 May 2004 BLAST 2.2.9 has been released. Read more...". The main content is organized into a grid of categories:

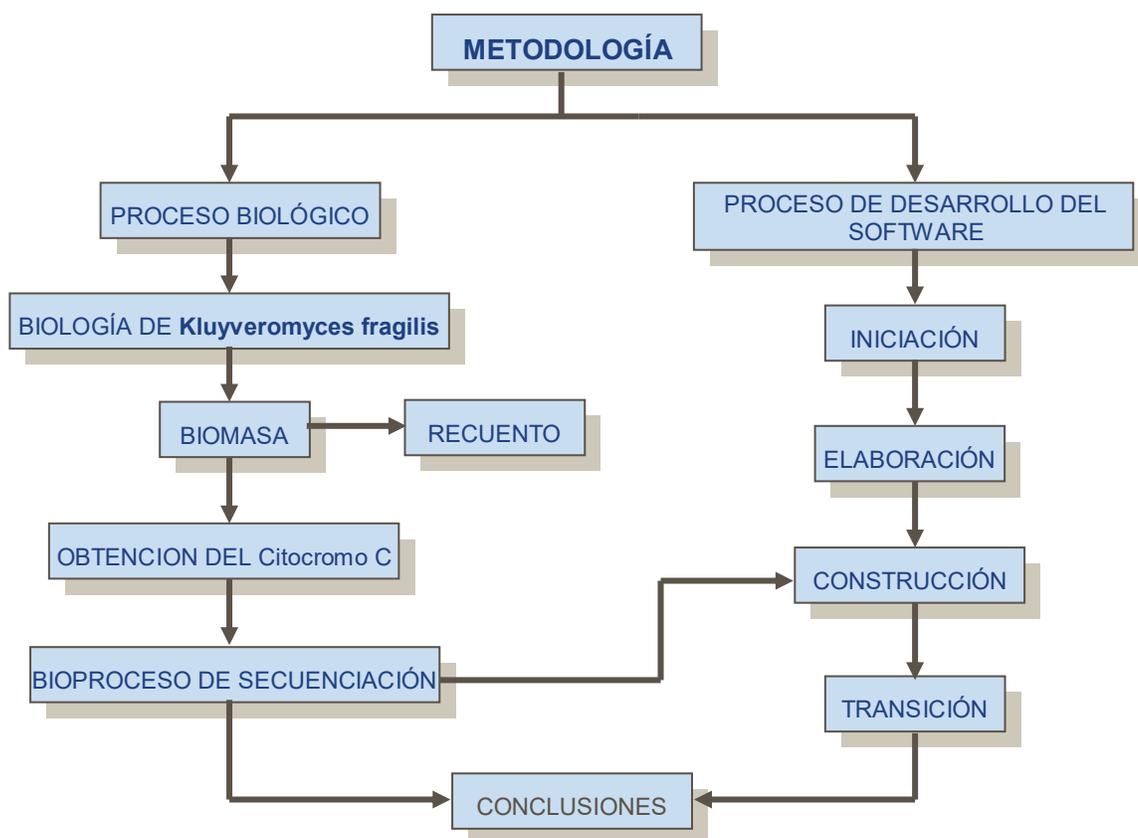
- Nucleotide**
 - Discontiguous megablast
 - Megablast
 - Nucleotide-nucleotide BLAST (blastn)
 - Search for short, nearly exact matches
 - Search trace archives with megablast or discontiguous megablast
- Protein**
 - Protein-protein BLAST (blastp)
 - PHI- and PSI-BLAST
 - Search for short, nearly exact matches
 - Search the conserved domain database (rpsblast)
 - Search by domain architecture (cdart)
- Translated**
 - Translated query vs. protein database (blastx)
 - Protein query vs. translated database (tblastn)
 - Translated query vs. translated database (tblastx)
- Genomes**
 - Chicken, cow, pig, dog, sheep, cat **NEW**
 - Environmental samples
 - Human, mouse, rat
 - Fugu rubripes, zebrafish
 - Insects, nematodes, plants, fungi, malaria
 - Microbial genomes, other eukaryotic genomes
- Special**
 - Search for gene expression data (GEO BLAST)
 - Align two sequences (bl2seq)
- Meta**
 - Retrieve results by RID
 - Get this page with javascript-free links

A left sidebar contains links for Info, Education, Download, and Support.

6. DISEÑO METODOLÓGICO

Debido a que la bioinformática comprende dos grandes ciencias: biología e informática, es necesario para cumplir los objetivos propuestos dividir la metodología de desarrollo en dos etapas. Una etapa comprende la metodología para el estudio biológico del **Kluyveromyces fragilis**, y la otra comprende la metodología a aplicar en el proceso de desarrollo del software.

Figura 23. Diagrama general del proceso Bioinformático (Realizado por los autores)



6.1 ETAPA 1: Metodología para el estudio biológico del *Kluyveromyces fragilis*.

El estudio biológico comienza desde el crecimiento y desarrollo de *Kluyveromyces fragilis* hasta la obtención de la secuencia de aminoácidos de la proteína del Citocromo C, proceso con las siguientes fases:

6.1.1 Fases

- Fase 1. Cultivo de la levadura *Kluyveromyces fragilis* sobre medio modificado de Malta dextrosa agar y medio de Ogy.
- Fase 2. Obtención de biomasa mediante cultivo en medios líquidos modificados, como el caldo de Malta en Biorreactores de 5 litros.
- Fase 3. Proceso de Secuenciación: Los péptidos se pueden secuenciar (establecer el orden en que se encadenan los aminoácidos en la secuencia) automáticamente utilizando el reactivo de EDMAN que separa y derivatiza el aminoácido N-terminal. Se realizan ciclos automáticos de separación e identificación cromatográfica. Más allá del aminoácido 20 o 30 el resultado ya no es fiable. Los péptidos se pueden secuenciar también por espectrometría de masas. Los péptidos son automáticamente fragmentados y de la masa de los fragmentos se deduce la secuencia. [56, 57, 58].

Las proteínas se pueden secuenciar tras romperlas en péptidos de tamaño adecuado y separarlos por cromatografía. Para la rotura se emplean enzimas o reactivos que cortan en sitios específicos. Es necesario realizar dos roturas con reactivos diferentes para obtener dos conjuntos de péptidos, todos los cuales son secuenciados. Los péptidos de un conjunto tienen secuencias que solapan las de

los péptidos del otro grupo y por comparación se puede establecer el orden en que se encadenan en la proteína entera. Es mucho más fácil y rápido determinar la secuencia de una proteína por secuenciación de su gen. [56, 57, 58]

El protocolo para la purificación del Citocromo C de **Kluyveromyces fragilis** comprende después de la separación por el método de Edman un proceso de filtración para eliminar el precipitado de hidróxido de aluminio, el cual se unió a la membrana de la Mitocondria debido a la carga negativa de los fosfolípidos y dejó libre a la proteína en solución. Para este proceso se usó cromatografía de intercambio iónico, la cual separa a las proteínas en función de sus cargas. El Citocromo C posee numerosos grupos cargados positivamente, la columna usada fue del tipo Amberlita CG-50.

El siguiente paso fue remover el exceso de sales empleado según el protocolo por filtración cromatográfica en gel, la cual separa las proteínas basado en su tamaño, como en el proceso se obtienen proteínas el método usado para su determinación fue el método de Bradford y la concentración del Citocromo C se midió por espectrofotometría.

Partiendo de una concentración inicial de 15 gramos de Biomasa de **Kluyveromyces fragilis** obtenida del Birreactor se llevaron a cabo los distintos procedimientos para la identificación en cada una de las fracciones que contenía el Citocromo C.

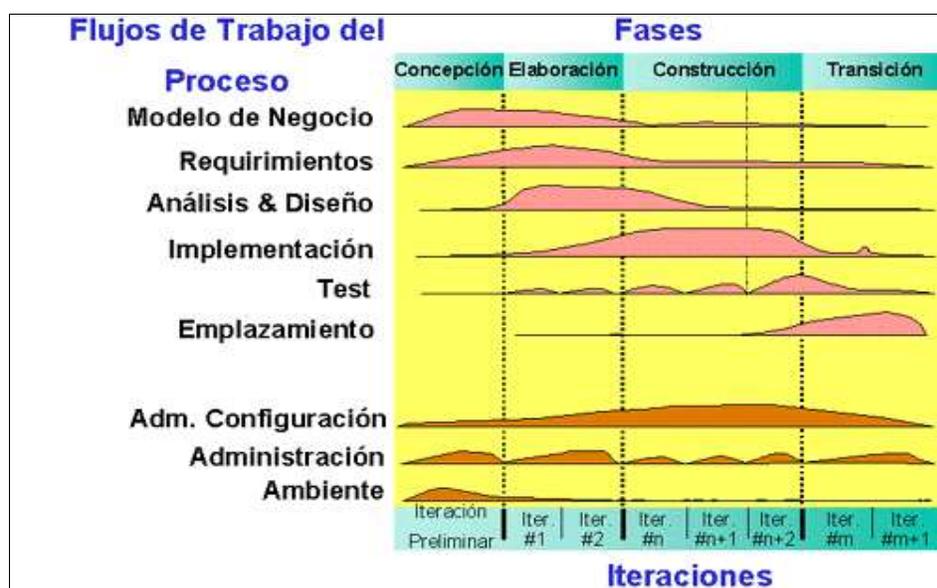
6.2 ETAPA 2: Metodología para el desarrollo del software

El diseño metodológico a seguir para realizar este proyecto es El Proceso Unificado de Rational (RUP) ya que se adapta especialmente a UML (Lenguaje Unificado de Modelado). Este proceso de desarrollo permite obtener un software

de mayor calidad satisfaciendo las necesidades de los usuarios finales dentro de la planificación establecida.

Las características más sobresalientes del Proceso Unificado de Rational son: *centrado en la arquitectura* dando desde el inicio del proceso una estructura robusta que facilite el desarrollo del software y su mantenimiento posterior. También este proceso estará *dirigido por casos de uso* que permiten tener un conocimiento sobre las actividades y escenarios utilizados para guiar el flujo de procesos desde la captura de requisitos hasta las pruebas. Este proceso soporta *técnicas orientadas a objetos* basándose en los conceptos: objeto y clase y las relaciones entre ellos, usando UML (Lenguaje Unificado de Modelado) como la notación común. Además, el Proceso Unificado de Rational se basa en un proceso *adaptable y configurable* cubriendo las necesidades de cada proyecto para el cual se necesite usar. Los aspectos calidad y riesgo, van contenidas durante el proceso de esta metodología, de esta manera podemos decir que el Proceso Unificado de Rational impulsa un *control de calidad y una gestión del riesgo continuo* para garantizar el éxito del proyecto en cada fase de desarrollo. [2]

Figura 24. Representación del ciclo de vida del RUP [41].



6.2.1 Fases

Esta metodología contiene fases de desarrollo, las cuales permiten cumplir los objetivos específicos de los proyectos indicando qué y cómo se lograrán cada uno de ellos, también las técnicas y recursos a utilizar y los resultados o productos que se esperan alcanzar.

- **Iniciación.** Durante esta fase se recolecta la información adecuada para comprender los aspectos necesarios y fundamentales que permiten desarrollar esta investigación. Además se utilizaran diferentes alternativas: medios informáticos como Internet, también lecturas en material impreso como lo son artículos, revistas, libros, tesis, entre otros, y además la asesoría por parte de la directora y del asesor, todo esto con el fin de complementar la información necesaria para el desarrollo del proyecto.

También en esta fase se determinan los recursos necesarios para llevar a cabo el proyecto y un planteamiento de fases iterativas donde se planifica el proceso.

- **Elaboración.** Una vez recolectada la información bioinformática, es necesario realizar un análisis de toda la documentación y de las diferentes herramientas bioinformáticas usadas para este proyecto, identificando las características y los riesgos que se puedan presentar durante el avance de este.

Para la comprensión de la composición estructural y la caracterización del Citocromo C de *Kluyveromyces fragilis*, se llevarán a cabo prácticas de laboratorio, tutorías sobre el tema, entre otros; las cuales permitirán tener un conocimiento claro y suficiente sobre esta temática, con el fin de obtener una documentación completa.

El resultado del análisis de las herramientas bioinformáticas será un documento que refleje las características principales de estas, determinando su estado del arte, a través del Lenguaje de Modelado Unificado UML, determinando así su posibilidad de integración en un único sistema. De ser posible la integración en un sistema se procederá a evaluar las herramientas existentes y plataformas para lograr desarrollar este proyecto.

- **Construcción.** De ser posible la integración de estas herramientas bioinformáticas en un único sistema, en esta fase se desarrollará de forma iterativa e incremental el producto software requerido para el Centro de Investigación en Biotecnología y Medioambiente de la UNAB, usando el software de desarrollo apropiado para su construcción.
- **Transición.** En esta fase de desarrollo, los usuarios del sistema evaluarán su funcionamiento, determinando si cumple con los requerimientos mínimos y necesarios para el desarrollo de futuras investigación en el área de Bioinformática.
- **Conclusiones.** Como parte final de este proyecto se emitirán unas conclusiones en donde se mostrará cuales fueron los resultados obtenidos y si estos cumplieron con los objetivos propuestos al inicio de la investigación.

7. RESULTADOS

7.1 MODELAMIENTO DE LAS HERRAMIENTAS BIOINFORMATICAS LOCALES EXISTENTES

En el modelado de las diferentes herramientas bioinformáticas locales, se mostraran los diagramas de casos de usos que representaran claramente las opciones con las que interactúa el usuario.

7.1.1 Rasmol. Rasmol es una herramienta que permite el análisis de moléculas biológicas a través de representaciones gráficas. En Rasmol se puede visualizar cualquier tipo de estructura molecular definida porque está ideado para hacer posible la visualización, la enseñanza y la producción de imágenes con calidad de publicación.

A su vez Rasmol se compone de una interfaz gráfica y una interfaz modo comando. Todo lo que se realiza dentro de la interfaz gráfica se ve reflejado en la interfaz modo comando y viceversa.

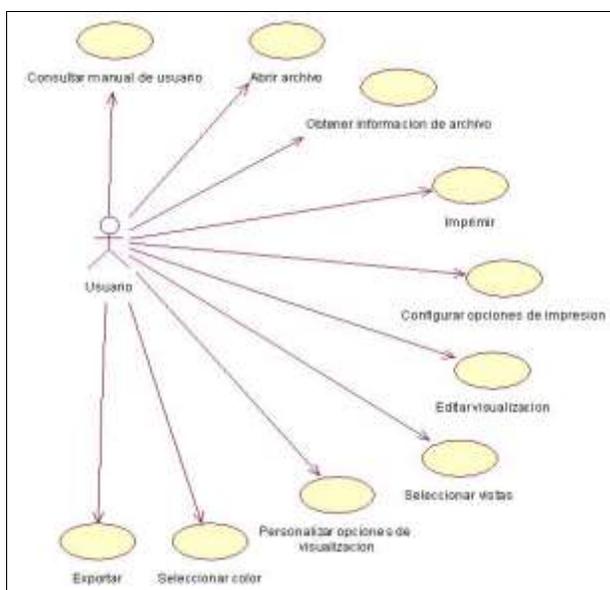
Las funciones más representativas de Rasmol comprende la representación de la molécula cargada en las formas de estructuras de alambre, enlaces en cilindros de Dreiding, esferas de espacio relleno (CPK), bolas y barras, cintas macromoleculares (lisas, sólidas y en filamentos) y superficie de puntos. La molécula exhibida puede ser girada, desplazada, ampliada (zoom) y/o cortada en rebanadas interactivamente usando, el ratón, las barras de desplazamiento de windows, o en línea de comandos. RasMol puede leer una sucesión de comandos previamente preparada en un archivo de script (guión) (o vía comunicación

interactiva) para permitir cargar una imagen dado o un punto de vista concreto, de forma rápida. RasMol también puede crear un archivo script conteniendo los comandos requeridos para regenerar una imagen en uso. (Figura 25)

Finalmente, la imagen generada podría exportarse en una variedad de formatos incluyendo GIF, PPM, BMP, PICT, ficheros de salida Sun o como un script MolScript o Kinemage [47]. Además esta herramienta nos permite trabajar con archivos de diferentes tipos como lo son:

- Brookhaven Databank: *.pdb; *.ent
- Alchemy File Format: *.alc; *.mol
- Sybyl MOL2 Format: *.syb; *.mol
- MDL Mol File Format: *.mdl; *.mol
- MSC (XMol) XYZ Format: *.xyz
- CHARMn File FormT: *.CHM
- MOPAC File Format: *.mop

Figura 25. Diagrama de casos de uso Herramienta Rasmol

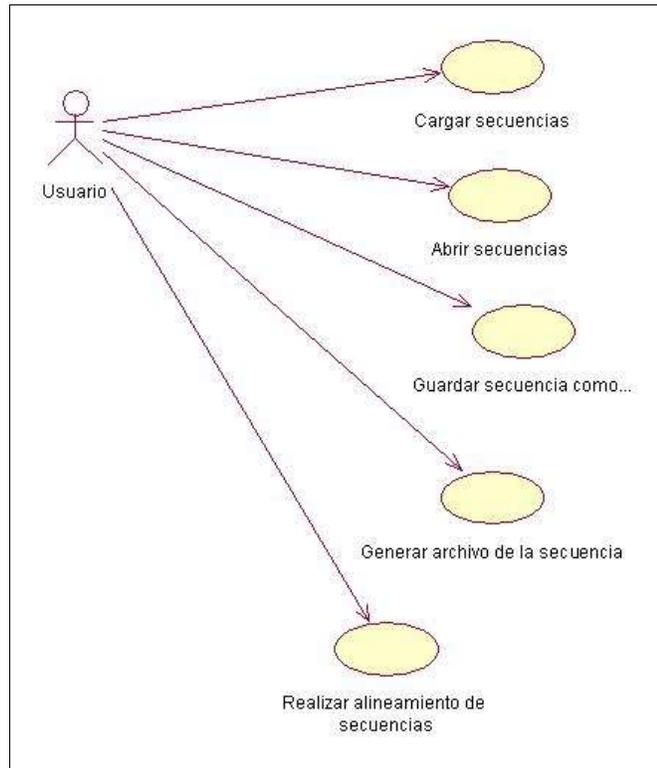


7.1.2 Clustal X. Clustal X trabaja bajo una interfaz gráfica y se caracteriza por ser un programa de propósito general para el alineamiento múltiple de secuencias de ADN o de proteínas.

La interfaz de Clustal X puede variar de acuerdo a las necesidades del usuario ya que tiene dos opciones: Multiple Alignment Mode y Profile Alignment Mode, con las cuales se puede estudiar una o dos secuencias. Clustal reconoce automáticamente diversos tipos de formato de archivo. Los diferentes formatos son:

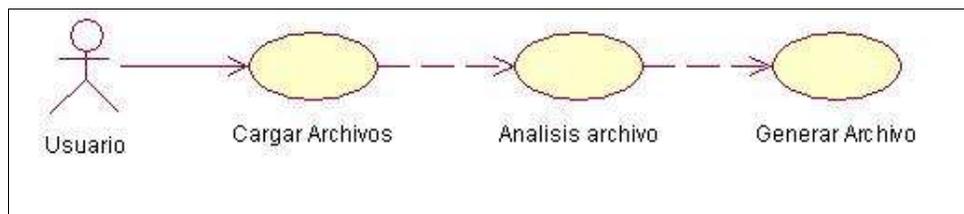
Los formatos de FASTA y de NBRF/PIR son reconocidos teniendo ">" como el primer carácter en el archivo. Los formatos de EMBL/Swiss Prot son reconocidos por las letras "identificación" en el comienzo del archivo (el símbolo para el campo conocido de la entrada). El formato de CLUSTAL es reconocido por la palabra CLUSTAL al principio del archivo. El formato de GCG/MSF es reconocido por uno del siguiente: - la palabra PileUp en el comienzo del archivo. - la palabra !! AA_MULTIPLE_ALIGNMENT o !! NA_MULTIPLE_ALIGNMENT en el comienzo del archivo. - la palabra MSF en la primera línea del archivo, y los caracteres. en el extremo de esta línea. El formato de GCG/RSF es reconocido por la palabra !! RICH_SEQUENCE al principio del archivo. (Figura 26)

Figura 26. Diagrama de casos de uso Herramienta Clustal X



7.1.3 Fasta. Esta herramienta permite cargar de manera manual dos archivos de secuencias biológicas para que sean comparados y arroja un archivo donde se muestran los resultados de la comparación. Igual puede comparar con una base de datos. La interfaz grafica mostrada por esta herramienta es de modo comando. (Figura 27)

Figura 27. Diagrama de casos de uso Herramienta Fasta

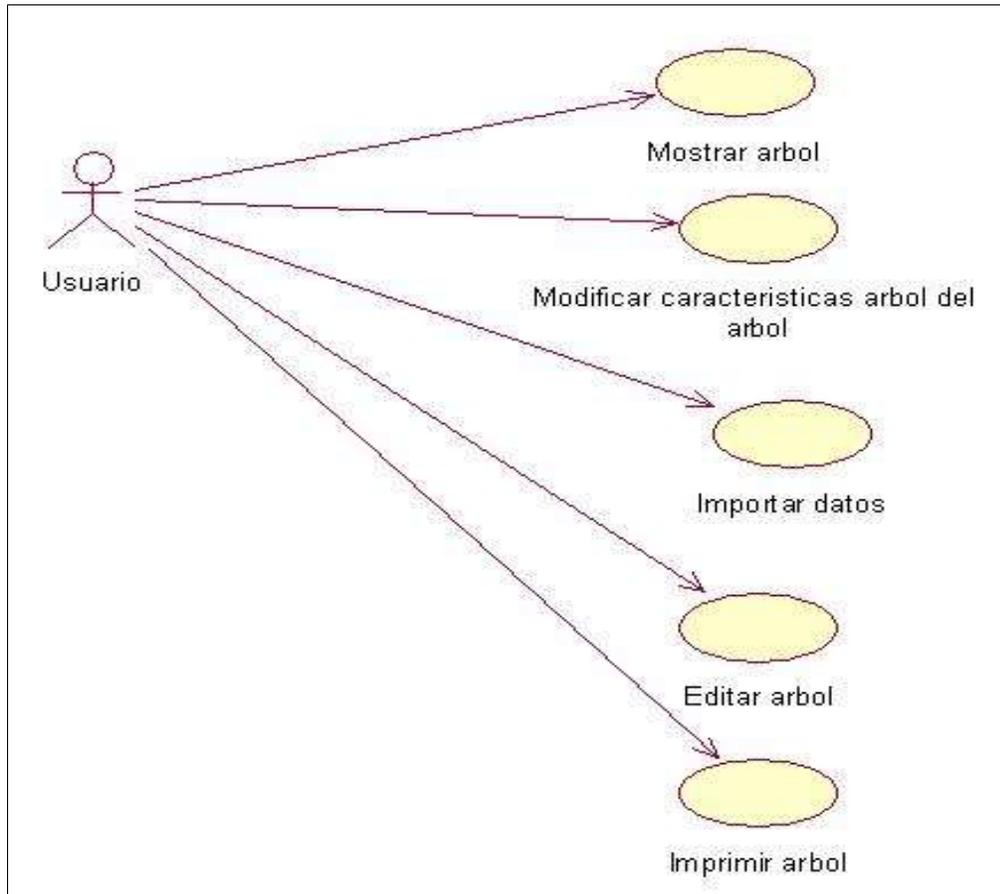


7.1.4 Treeview. Esta herramienta permite mostrar el árbol de la filogenia del objeto a estudiar, igualmente permite que el resultado arrojado sea guardado como grafica, importar una lista de nombres taxonómicos. Además permite modificar el árbol representado y que este sea mostrado con las preferencias requeridas por el usuario. Esta herramienta permite modificar tanto el tipo de letra, tamaño y forma de las letras mostradas en los resultados arrojados. Tiene una opción en la que se puede cambiar el tipo de árbol mostrado, se puede imprimir el árbol y en general un grupo de funciones que permiten modificar la estructura del árbol. Otra opción reflejada en esta herramienta es la de mostrar las ventanas de representación de los resultados. (Figura 28).

Además esta herramienta nos permite trabajar con archivos de diferentes tipos como lo son:

- Nexus (*.tre;*.trees)
- Phylip (*.phy)
- Clustal W(*.dnd;*.ph;*.phb)
- Henning86(*.hng)

Figura 28. Diagrama de casos de uso Herramienta Fasta



7.2 SISTEMA INTEGRADO DE INFORMACION BIOINFORMÁTICA

Una vez realizado el proceso de análisis a las herramientas Bioinformáticas locales se concluyó que si es viable su integración en un único sistema que permita no solo la caracterización del Citocromo C de **Kluyveromyces fragilis** sino que además le provee al usuario una herramienta para el análisis molecular de cualquier tipo de microorganismo.

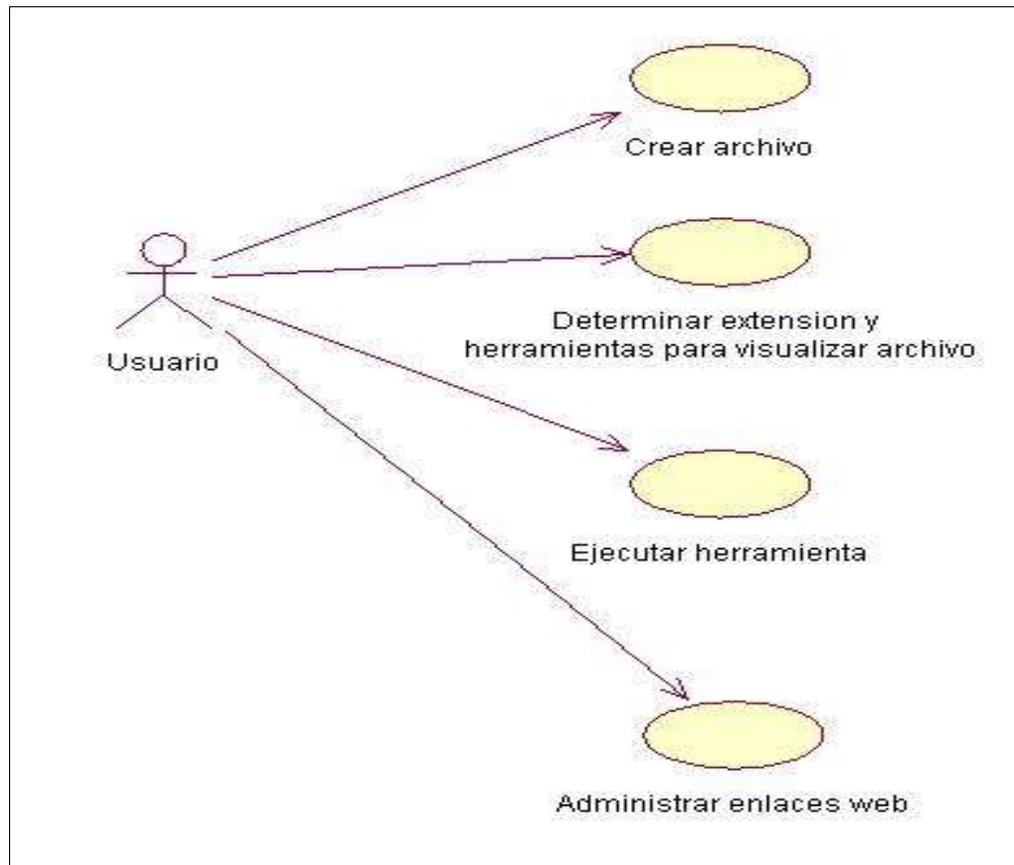
Después de comprobar que si es viable la construcción de este sistema Bioinformático se procedió al desarrollo de este.

7.3 MODELAMIENTO DEL SISTEMA DE INFORMACION BIOINFORMATICO

7.3.1 Diagramas de casos de usos

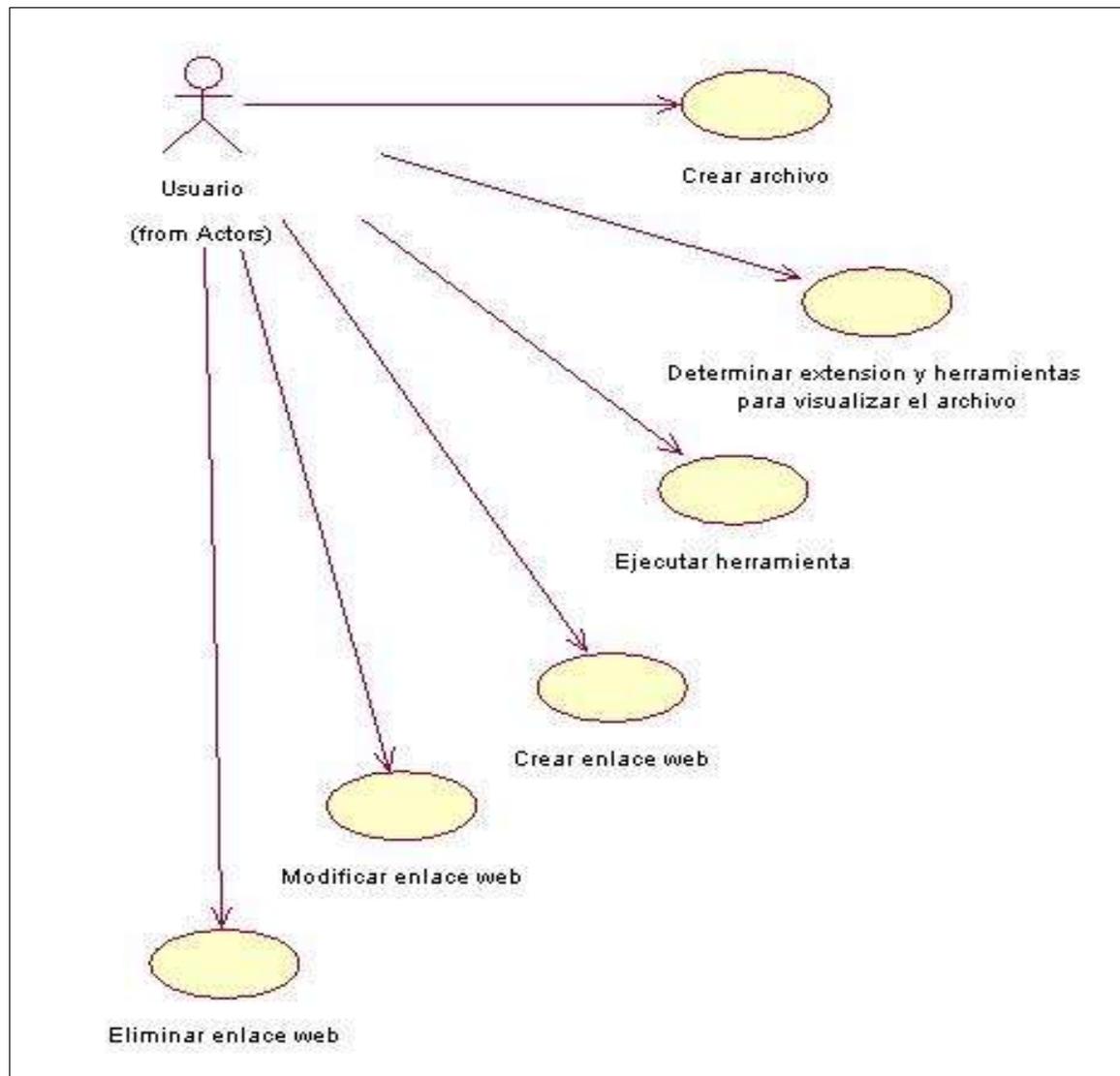
- **Casos de uso nivel 0.** (Figura 29). En este diagrama se representa de manera general el comportamiento del sistema desde el punto de vista del usuario, además de la forma, tipo y orden como los elementos generales interactúan.

Figura 29. Diagrama de casos de uso, nivel 0



- **Casos de uso nivel 1.** (Figura 30) En este diagrama se define de manera clara y precisa como el usuario interactuará con el sistema a desarrollarse.

Figura 30. Diagrama de casos de uso, nivel 1



A continuación se describe la función que realiza cada uno de los casos de uso que tiene el sistema a desarrollar:

Crear Archivo

- Descripción: El usuario desea crear un archivo y debe seleccionar la herramienta bioinformática a la cual le desea crear un archivo.

- PRE – condición: Se parte del hecho de que el usuario necesita crear un archivo para visualizar en una herramienta bioinformática.
- Flujo de eventos:
 - Path Básico:
 1. Seleccionar la opción crear archivo.
 2. Seleccionar la herramienta con la cual desea visualizar archivo.
 3. Ingresar datos que conforman el archivo.
 4. ingresar el nombre al archivo.
 5. Guardar el archivo en la ubicación predeterminada.
 - Path Alternativo:
 1. El usuario cancela el proceso.
 2. Devuelve al usuario a la pantalla principal.
- Post – Condición: Se ha creado un archivo de acuerdo con las especificaciones del usuario.

Determinar Extensión y Herramientas para visualizar el archivo

- Descripción: El usuario carga un archivo en el sistema, y este realiza un análisis de la extensión del archivo para comprobar que herramientas bioinformáticas puede utilizar.
- PRE – condición: Se parte del hecho de que el usuario necesita el análisis de un archivo biológico.
- Flujo de eventos:
 - Path Básico:
 1. Ingresar el archivo.
 2. El sistema realiza el análisis de la extensión del archivo.
 3. Se muestran las herramientas bioinformáticas que pueden utilizarse.
 - Path Alternativo:

1. Se envía un mensaje de error por que la extensión del archivo no es compatible con ninguna de las herramientas bioinformáticas.
 2. Se solicita al usuario ingresar nuevamente un archivo compatible con las herramientas bioinformáticas.
 3. Se repiten pasos 2 y 3 del Path básico.
- Post – Condiciones: Se muestran las herramientas bioinformáticas que utilizan este archivo.

Ejecutar Herramienta

- Descripción: El usuario desea visualizar un archivo en alguna de las herramientas bioinformáticas existentes en el sistema.
- PRE – condición: Se parte del hecho de que el usuario necesita visualizar un archivo biológico.
- Flujo de eventos:
 - Path Básico:
 1. Seleccionar la opción visualizar archivo.
 2. Seleccionar la herramienta con la cual desea visualizar archivo.
 3. Se ejecuta la herramienta bioinformática escogida.
 - Path Alternativo:
 1. El usuario cancela el proceso.
 2. Devuelve al usuario a la pantalla principal
- Post – Condiciones: El usuario puede visualizar el archivo en la herramienta bioinformática seleccionada, utilizando todas las características de la misma.

Crear Enlaces Web

- Descripción: El usuario desea agregar en el sistema un nuevo enlace Web.

- PRE – condición: Se parte del hecho de que el usuario necesita un nuevo enlace Web dentro del sistema.
- Flujo de eventos:
 - Path Básico:
 1. Seleccionar la opción Administración.
 2. Seleccionar la opción crear nuevo enlace Web
 3. Se ingresan los atributos del nuevo enlace Web.
 4. Se guardan los cambios en el sistema.
 - Path Alternativo:
 1. El usuario cancela el proceso.
 2. Devuelve al usuario a la pantalla principal de administración.
- Post – Condiciones: En el sistema queda creado un nuevo enlace Web.

Modificar Enlaces Web

- Descripción: El usuario puede modificar los atributos dentro del sistema de los enlaces Web existentes, si estas en algún momento llegan a cambiar.
- PRE – condición: Cuando se requieren cambios en las aplicaciones Web.
- Flujo de eventos:
 - Path Básico:
 1. Seleccionar la opción Administración.
 2. Seleccionar la opción modificar enlace Web
 3. Se muestran los enlaces Web existente.
 4. Seleccionar el enlace Web que desea modificar.
 5. Se realizan las modificaciones al enlace Web.
 6. Se guardan los cambios en el sistema.
 - Path Alternativo:
 1. El usuario cancela el proceso.
 2. Devuelve al usuario a la pantalla principal de administración.
- Post – Condiciones: La información del enlace Web queda actualizada.

Eliminar Enlaces Web

- Descripción: El usuario puede eliminar un enlace Web existente.
- PRE – condición: Cuando el enlace Web a dejado de funcionar en Internet o ya no satisface las necesidades del usuario.
- Flujo de eventos:
 - Path Básico:
 1. Seleccionar la opción Administración.
 2. Seleccionar la opción eliminar enlace Web
 3. Se muestran todas las aplicaciones Web.
 4. Selecciona la aplicación Web a eliminar.
 5. Se elimina la aplicación Web.
 6. Se guardan los cambios en el sistema.
 - Path Alternativo:
 1. El usuario cancela el proceso.
 2. Devuelve al usuario a la pantalla principal de administración.
- Post – Condiciones: La aplicación Web queda eliminada del sistema.

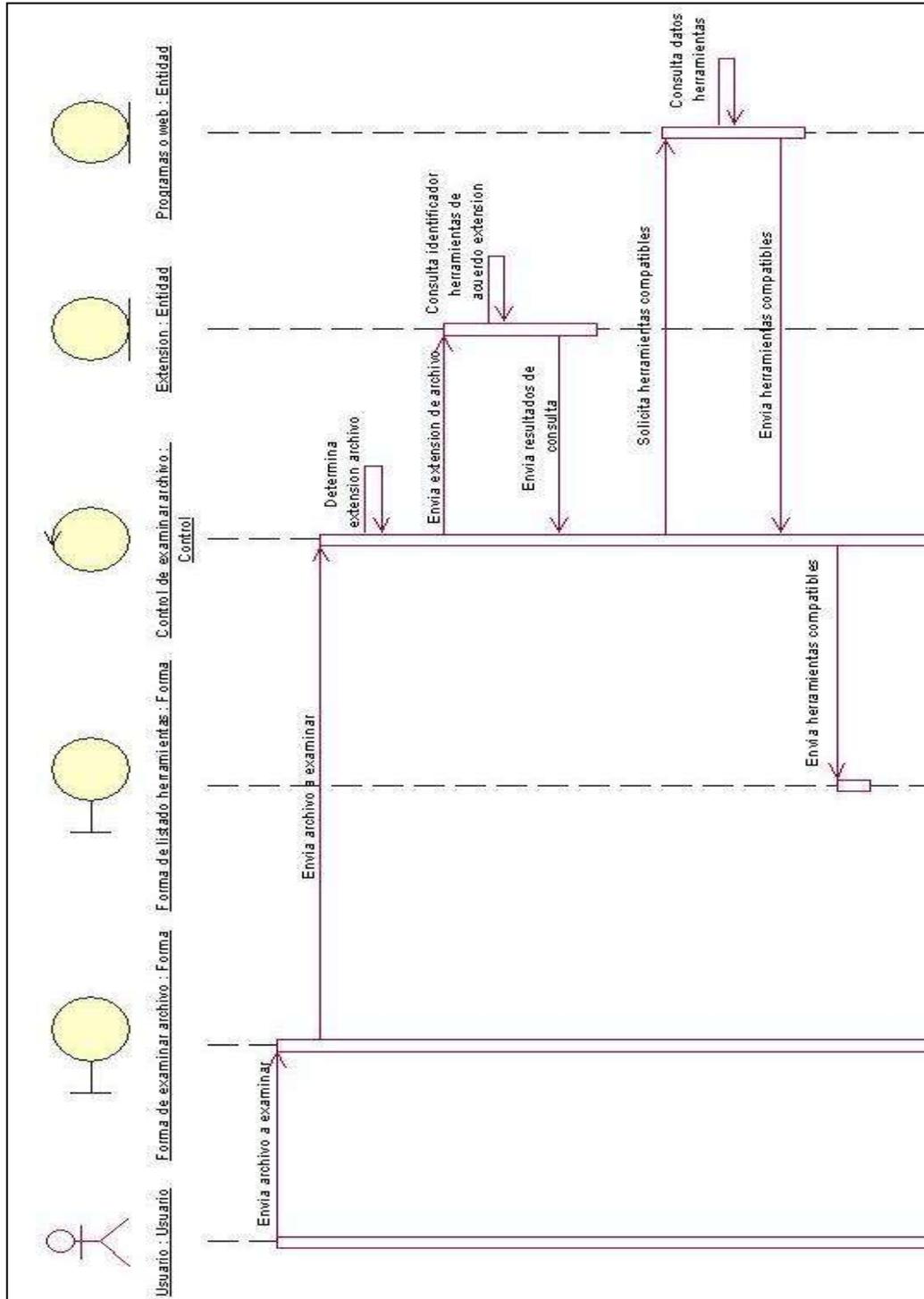
7.3.2 Diagramas de secuencias. En el análisis realizado se crearon de los diferentes casos de uso de nivel 1 los correspondientes diagramas de secuencia: en estos diagramas se muestra la interacción de un conjunto de objetos en una aplicación a través del tiempo. Esta descripción es importante porque puede dar detalle a los casos de uso, aclarándolos al nivel de mensajes de los objetos existentes, como también muestra el uso de los mensajes de las clases diseñadas en el contexto de una operación. A continuación mostraremos los diferentes diagramas de secuencias generados a partir de los diferentes casos de usos:

Este diagrama representa la secuencia que realiza el sistema (Figura 31), la cual es:

1. Al entrar al sistema, este consulta las herramientas para solicitar sus extensiones.
2. El usuario selecciona la herramienta a crear.
3. El usuario ingresa los datos que contiene el archivo a crear.
4. Cuando el usuario decide guardar el archivo, el sistema le muestra con que extensión debe guardar el archivo.
5. El archivo ha sido guardado con la extensión deseada.

Determinar Extensión Y Herramientas Para Visualizar El Archivo

Figura 32. Diagrama de secuencia – Determinar extensión y herramientas para visualizar archivo

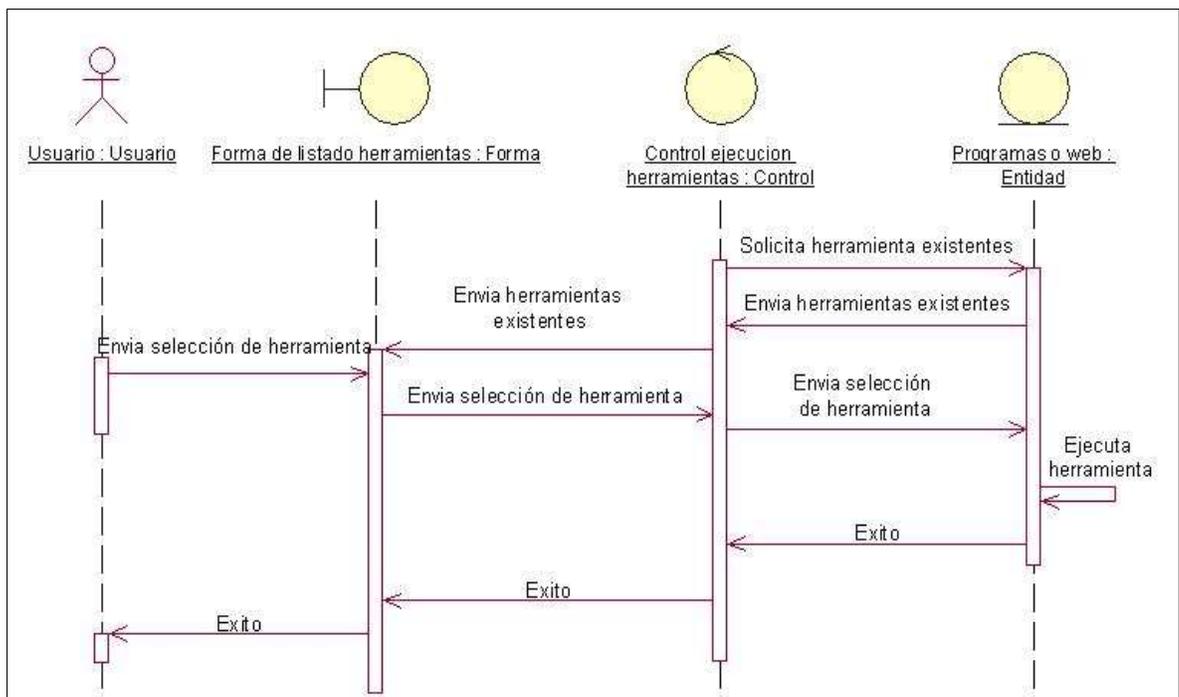


La secuencia general de este diagrama (Figura 32), es la siguiente:

1. El usuario ingresa el archivo a examinar, para consultar las herramientas disponibles para el.
2. El sistema determina la extensión del archivo.
3. El sistema consulta el identificador de la extensión.
4. El sistema consulta las herramientas compatibles con el archivo.
5. El sistema muestra herramientas compatibles.

Ejecutar Herramienta

Figura 33. Diagrama de secuencia – Ejecutar herramienta



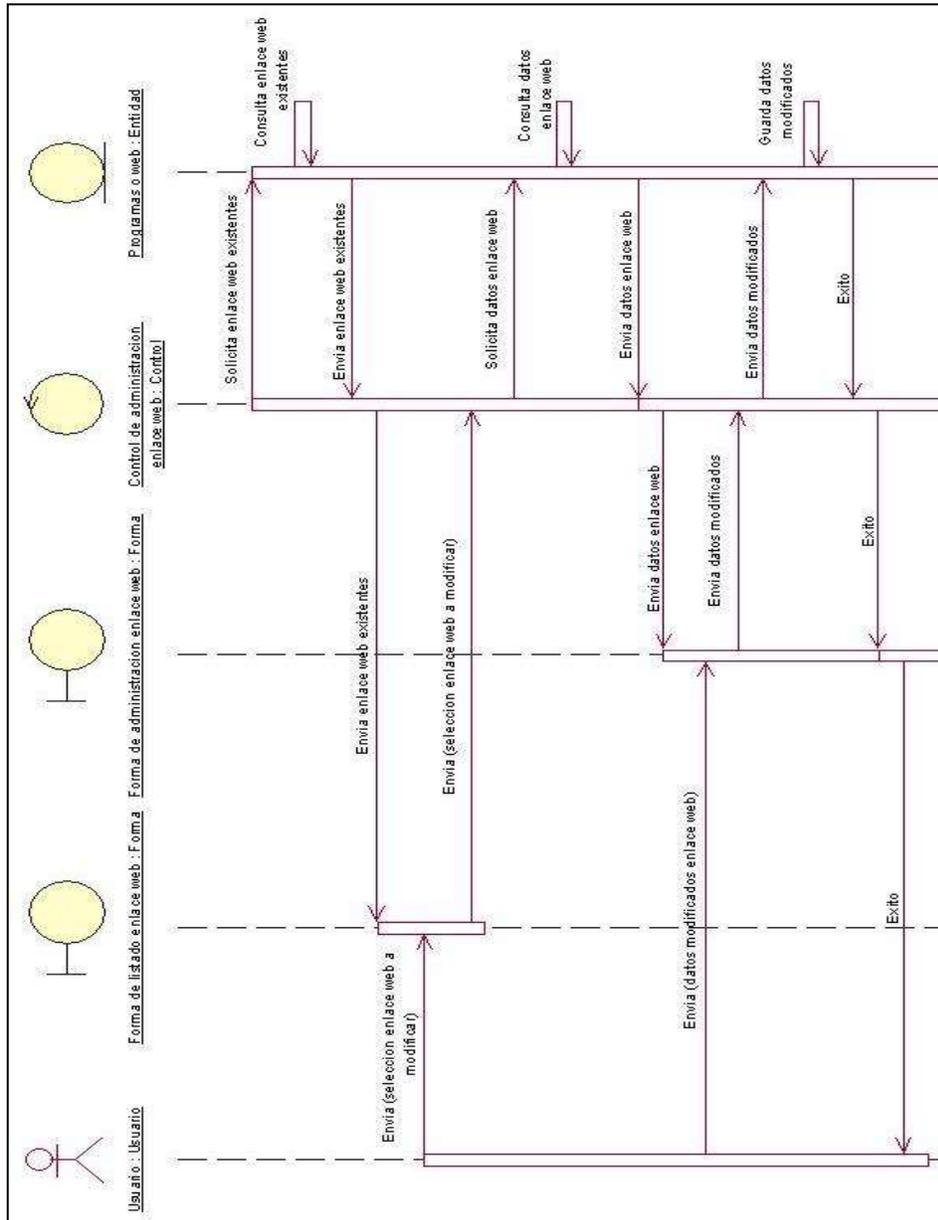
La secuencia general de este diagrama (Figura 33), es la siguiente:

1. El sistema consulta las herramientas existentes.
2. El sistema muestra las herramientas existentes.
3. El usuario selecciona la herramienta a ejecutar.
4. El sistema ejecuta la herramienta seleccionada.

3. El sistema verifica datos.
4. El sistema crea el nuevo enlace Web.

Modificar Enlaces Web

Figura 35. Diagrama de secuencia – Modificar enlace Web

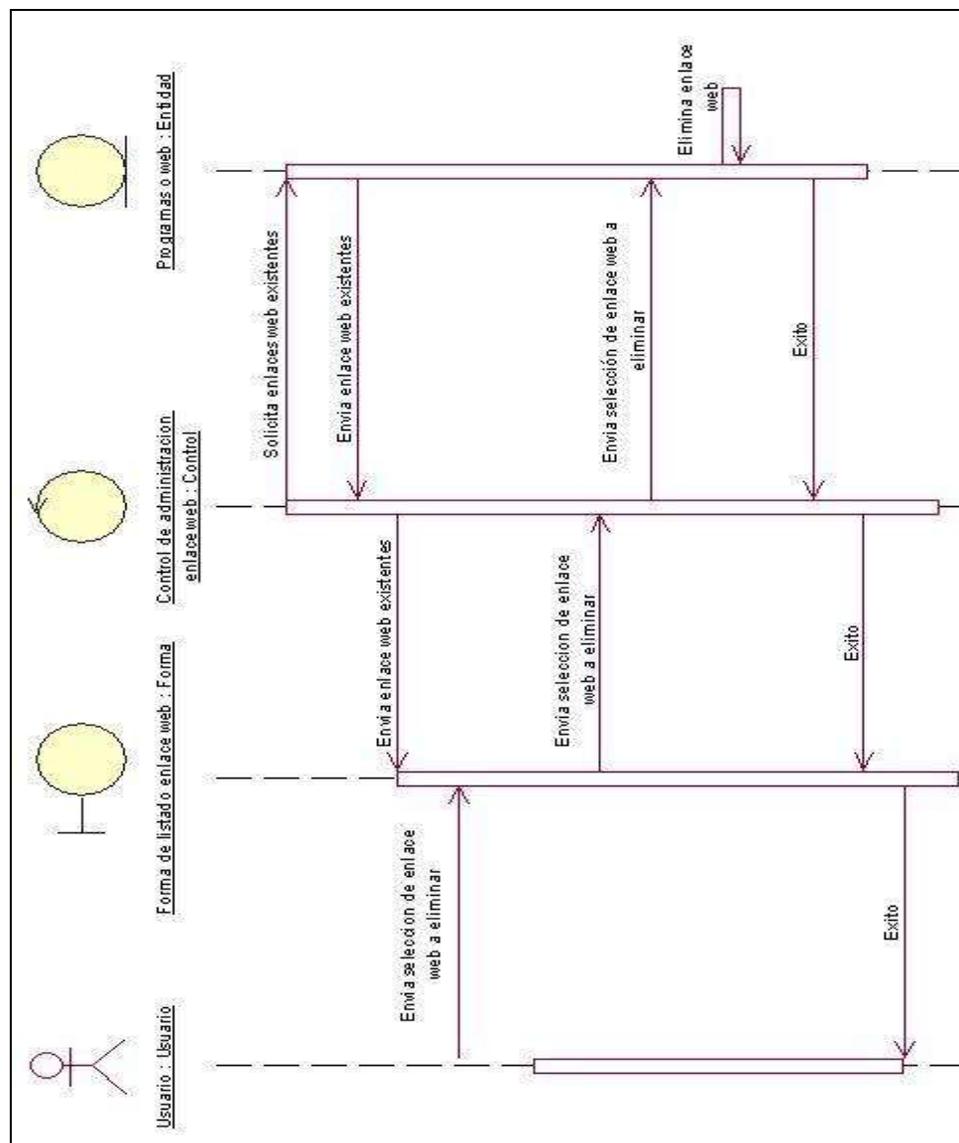


La secuencia general de este diagrama (Figura 35), es la siguiente:

1. El sistema consulta los enlaces Web existentes.
2. El usuario selecciona el enlace Web a modificar.
3. El sistema consulta los datos del enlace Web.
4. El usuario modifica los datos del enlace Web.
5. El sistema guarda los datos modificados.

Eliminar Enlaces Web

Figura 36. Diagrama de secuencia – Eliminar enlace Web

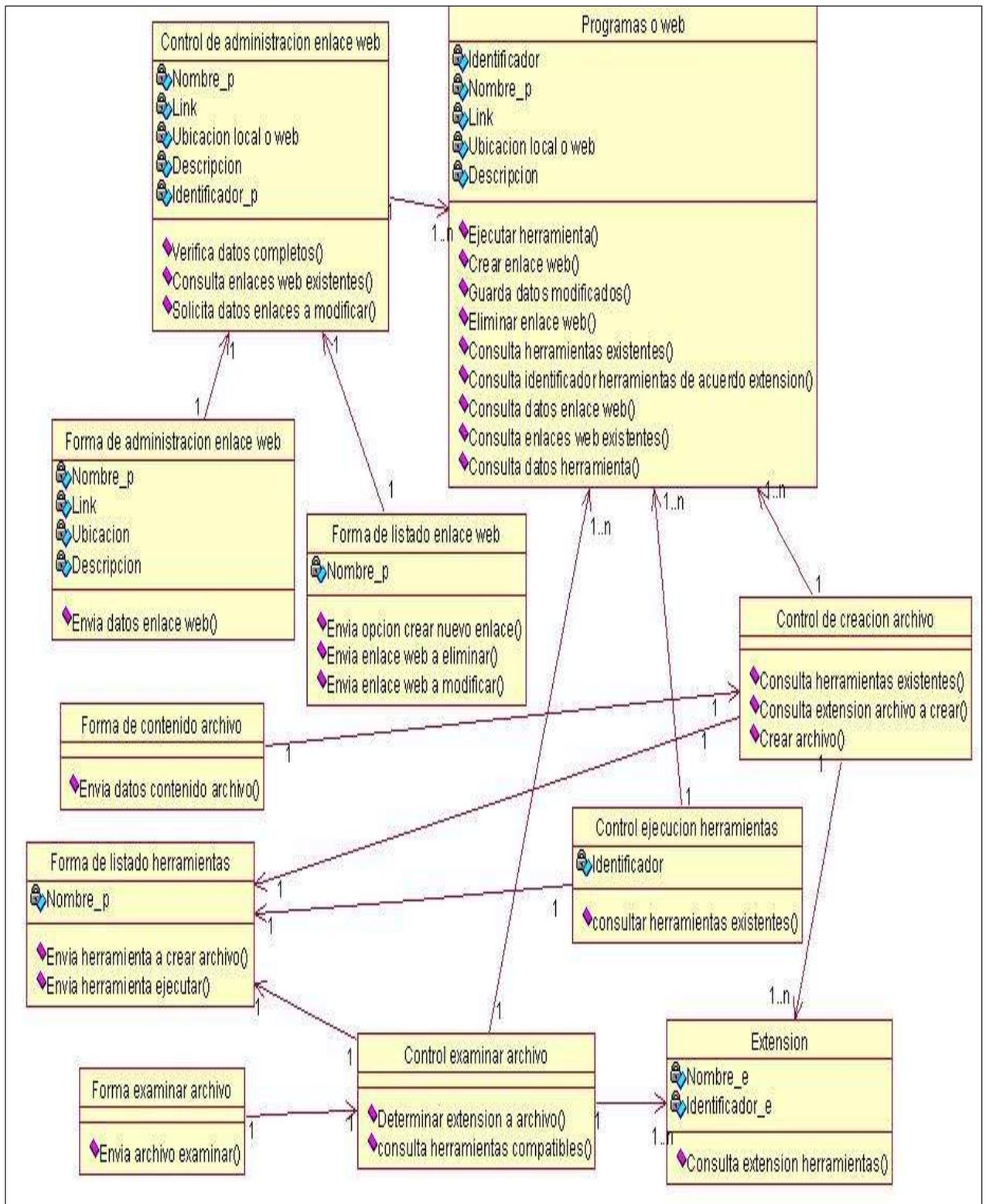


La secuencia general de este diagrama (Figura 36), es la siguiente:

1. El sistema consulta los enlaces Web existentes.
2. El usuario selecciona el enlace Web a eliminar.
3. El sistema elimina el enlace Web.

7.3.3 Diagrama de clases. (Figura 37). El diagrama de clases se considera principal para el análisis y diseño de un sistema porque en él se representan además de las clases sus relaciones estructurales y operaciones.

Figura 37. Diagrama de clases



A continuación se muestra la descripción de todos los componentes de este diagrama:

- **Programas o Web:** Esta entidad contiene toda la información de las diferentes herramientas que se encuentran en el sistema, es en ella donde está definida toda la información que a estas se refiere, como lo es el identificador, nombre, link, ubicación local o Web y descripción.
- **Extensión:** En esta entidad se encuentran almacenadas todas las extensiones con las que pueden trabajar las diferentes herramientas bioinformáticas contenidas en el sistema, se encuentra compuesta por un identificador y por el nombre de la extensión.
- **Control de administración de enlace Web:** este control es el encargado de comunicar todo lo referente administrar los enlaces Web, que en su defecto corresponde a la parte de crear, modificar y eliminar enlaces Web.
- **Control de creación de archivo:** Este control realiza los métodos de consultar herramientas, crear archivo y consultar extensiones, los cuales son los encargados de crear un archivo para la herramienta deseada por el usuario.
- **Control ejecución herramienta:** este control es el encargado de consultar la herramienta que el usuario desea ejecutar para trabajar, este control realiza la consulta de las herramientas existentes.
- **Control examinar archivo:** es el encargado de consultar las herramientas disponibles para el archivo que el usuario ha ingresado.

- Forma de administración de enlace Web: Esta es la interfaz que interactúa con el usuario, es en ella donde el usuario puede crear, eliminar, modificar y consultar los enlaces Web existentes en el sistema.
- Forma de listado de enlace Web: esta es la interfaz que le muestra al usuario los enlaces Web existentes en el sistema.
- Forma de listado de herramientas: esta es la interfaz que le muestra al usuario todas las herramientas Web existentes en el sistema.
- Forma de contenido archivo: Esta es la interfaz en donde el usuario ingresa los datos que contiene el archivo a crear.
- Forma de examinar archivo: Esta es la interfaz donde el usuario ingresa el archivo el cual desea analizar, para saber que herramientas tiene el sistema disponible para trabajar con el.

Después de analizar el sistema propuesto se procedió al desarrollo de este, teniendo como base todos los requerimientos y requisitos arrojados por este análisis. (Anexo 1)

7.4 CARACTERIZACIÓN DEL CITOCROMO C DE *Kluyveromyces fragilis*

7.4.1 Secuencia genética del Citocromo C de *Kluyveromyces fragilis*.

De acuerdo al desarrollo de esta investigación y al largo proceso de estudio de la proteína del Citocromo C de ***Kluyveromyces fragilis*** se logró la secuenciación de sus aminoácidos, utilizando diversos métodos para alcanzar este logro; dando como resultado la siguiente secuencia:

ATCTTAGAAG AGATTGTCGG TGAATTTACC ACTTCAACAG CCCCATCAAT TAACGATGAA GTTATCCAC
 AATCAGACGG TTCTCTTATC ATTGAGGGCT CCGCCAATTT ACGTGATTTG AATAAATTAT TTGACTGGAA
 TCTCGATACC GAAGATGCAC GTACCTTCAA CGGCTTAATT TTAGAGCATT TAGAAGAAAT TCCAGAAGAA
 GGAACGGTAT GTGAAATTA TGGGCTACAA ATCACGATTC TAGAAGTCAA TTACAACATG ATTAACAAG
 CCAAAGTCAT TAACTTTAA TTCAACATGT GGCTAAGCGA TGTCAATCAAG ACATCGCTTT TTTCTCCGT
 ACATGAATGT TTGATCCAAC ACAACATTTA TTCAACATTG GATAAATAAT CATCCTAAAT CGCACGAAT
 TCTTATTTAC CCCGTTTTTG GCTTTTGCTA GAATCTTGCA ATTGAAATTA ATTCTCAATA CCGTATAATG
 TTCAACCTTA TTTTGCGATA CAAATTAAG GGTATTAAA ATGCCACACT TATTCCACTT CTTACAACAA
 TATATTGATT ACAATTATTC TAGGCTTACT TGCCCTTCATG AGCTTTATTA TGGGTTGGCT TGTGATTGCA
 CGCTTTCTCT CTTAAGTCGC GTCAACGTGG CATCTTATGA AAGCATGCAT GAATTAGACA TTGACTTACA
 ACGCCACCTC ACAGCTATCT CTACAATCGG TTCTAATGCA CCTTATGTAG GTTTGCTTGG TACCGTCATT
 GGTATTCTCT TAACTTTCTA TGAATTAGGT CACTCTGGTG GCGATATTGA TGCGGCGGT ATTATGGTGC
 ACTTATCATT AGCCTTAAAA GCCACAGCAG TAGGTATTTT AGTCGCCATT CCTGCAATGG TTTGTTACAA
 CGGTTTAGGA CGTAAAGTCG AAGTTAATCG TTTGAAATGG TTTGCCTTAA ATGAGAAAAA AGCCAAACAA
 CAAGCATAGG GAGCCTCATG AAAAAGTTG ATGAAATCAA CATTATCCCT TTTATTGACA TCATGTTGGT
 ACTATTGGCT ATCGTTCTGA TTACAGCCTC TTTTATTTCA CAAGGTAAAA TCCAAGTGAA TGTACCAAAA
 GCAAGTTCAA CAGTTGCGTT TCGTTTCTGAT GATTTAGCCA AATTGCTGAC TATTACGGAA AGTGGTGAAA
 TTTTTTATCA CGATAAACCG ATTACTATAG TAGCATTGGA ACAAGAAATC AGTAATTGGG AAAAAGATCA
 AAAAGTCACC TTGAAGGTAG ATGCAAAATC CAGTTTCCAA GATTTCTGTT CTATCACTGA TATTATGGCT
 AAAAATGAAA TTAATAATGT CGCTATCGTG ACGGTTAAAG AAAAGGCACC TCAATGATAG ATAAAAGTCG
 TTCTTGATC GGGTTTGCAA TTTCATTGCT TTTTACGCA AGTTTTGTCT CTTTCTGTG TGGATTGTA
 CAAAAGACG ATGACAGCGC GAATGGATTT GCTGCCGATA TCATCTCAAC TCATATTTCC ATGGAAATGC
 TGGCGGCTAC CGTTTTAGAA GAACCAGAGC CGGAACCAGA GCCGGCGCCT CCGGTAGTAG AACCTGAAC
 GCCAAAAGAA GAAGTCGCAG ATCCGACGGT AAAACCTGAG CCACAAAAG AACCCGAAAA ACCAAAAGAG
 CCTGAAAAGC AAAAGAGAA ACCGAAGGAA AAACAAAAG AAAAGCCGAA AAAACCGAAG AAAGAACAAC
 GTGATTTACC AAAGTCAGAT CGCCAAATTG ATTCTAATTC ATCGATCAAT CAACAAGCGA CCACAACAGG
 TAACATCACA ACCAATAATC CGAATCTGGT CGGTAAAGGT AATAGCACAG ATGAAGTCAA TGCTTATCGC
 TCGGCTTTAC GCAGAGAAAT TGAAAAACAT AAACGCTATC CAAACCGTGC ACGCATGATG CGCAAACAAG
 GTGTGGTAAC AATCACGTTT CATCTTAATA ATGCCGGCGT AATTAGTAAT GCGCGAATCA GCAAATCTTC
 TGGCTCAGAA GAATTAGATA ACGCTGCACT GGTAGCTGTC AATAATGCC GTCCAATTGG TCCATTGCCT
 GCTGGTATGC CAAATGAAGT GAGCGTTTCT GTCAGTTTCA GAATCACAAA TTAATAAAGT GCGGTAATTT
 TTACCGCACT TTTTCTCTC TATTAGAATT CCTCTCATT GTGCATTTAT CTAGGACATC TTTTATAAAA
 ACTGTGTCAT TTTTATCTT CCTATTAGGA TATCTAACGA TTATCTTCTG CTCATCAATA AGGTAAATAA
 AAATGACTAA AAAACCTAT TTTCCGCA

7.4.2 Secuencia de aminoácidos del Citocromo C de *Kluyveromyces fragilis*.

Tomando como base la secuencia genética obtenida, se transformó a una secuencia de aminoácidos, dando como resultado lo siguiente:

1. ILE LEU GLU GLU ILE VAL GLY GLU PHE THR THR SER THR ALA PRO SER ILE ASN ASP GLU VAL ILE PRO
2. GLN SER ASP GLY SER PHE ILE ILE GLU GLY SER ALA ASN LEU ARG ASP LEU ASN LYS LEU PHE ASP TRP ASN
3. LEU ASP THR GLU ASP ALA ARG THR PHE ASN GLY LEU ILE LEU GLU HIS LEU GLU GLU ILE PRO GLU GLU
4. GLY THR VAL CYS GLU ILE ASN GLY LEU GLN ILE LEU GLU VAL ASN TYR ASN MET ILE LYS GLN
5. ALA LYS VAL ILE LYS LEU STOP PHE ASN ILE TRP LEU SER ASP VAL ILE LYS THR SER LEU PHE TYR SER VAL
6. HIS GLU CYS LEU ILE GLN HIS ASN ILE TRP LEU SER THR LEU ASP LYS STOP SER SER STOP ILE ALA ARG ILE
7. SER TYR LEU PRO ARG PHE TRP LEU LEU LEU GLU SER CYS ASN STOP ASN STOP PHE SER ILE PRO TYR ASN
8. VAL GLN PRO TYR PHE ALA ILE GLN ILE LYS GLY LEU LEU LYS CYS HIS THR TYR SER THR SER TYR ASN ASN
9. ILE LEU ILE THR ILE ILE LEU GLY LEU LEU ALA PHE MET SER PHE ILE MET GLY TRP LEU VAL ILE ALA
10. ARG PHE LEU SER STOP VAL ALA SER THR TRP HIS LEU MET LYS ALA TYR MET ASN STOP THR LEU THR TYR
11. ASN ALA THR SER GLN LEU SER LEU GLN SER VAL LEU MET HIS LEU MET STOP VAL CYS LEU VAL PRO SER
12. LEU VAL PHE SER STOP LEU SER MET ASN STOP VAL THR LEU VAL ALA ILE LEU MET ARG ARG LEU LEU TRP CYS
13. THR TYR HIS STOP PRO STOP LYS PRO GLN GLN STOP VAL PHE STOP SER PRO PHE LEU GLN TRP CYS VAL THR THR
14. VAL STOP ASP VAL LYS SER LYS LEU ILE VAL STOP ASN GLY LEU PRO STOP MET ARG LYS LYS PRO ASN ASN

15. LYS HIS ARG GLU PRO HIS GLU LYS VAL STOP STOP ASN GLN HIS TYR PRO PHE TYR STOP HIS HIS VAL GLY
 16. THR ILE GLY TYR ARG SER ASP TYR SER LEU PHE TYR PHE THR ARG STOP ASN PRO SER GLU CYS THR LYS
 17. SER LYS PHE ASN SER CYS VAL SER PHE ARG STOP PHE SER GLN ILE ALA ASP TYR TYR GLY LYS TRP STOP ASN
 18. PHE LEU SER ARG STOP THR ASP TYR TYR SER SER ILE GLY THR ARG ASN GLN STOP LEU GLY LYS ARG SER
 19. LYS SER HIS LEU GLU GLU ARG CYS LYS ILE GLN PHE PRO ARG PHE ARG PHE TYR HIS STOP TYR TYR GLY
 20. STOP LYS STOP ASN STOP LYS CYS ARG TYR ARG ASP GLY STOP ARG LYS GLY THR SER MET ILE ASP LYS SER ARG
 21. SER CYS ILE GLY PHE ALA ILE SER LEU LEU PHE HIS ALA SER PHE VAL SER PHE LEU TYR TRP ILE VAL
 22. GLN LYS ASP ASP ASP SER ALA GLY PHE ALA ALA ALA ILE ILE SER THR HIS ILE SER MET GLU MET
 23. LEU ALA ALA THR VAL LEU STOP STOP PRO GLU PRO GLU PRO GLU PRO ALO PRO PRO VAL VAL GLU PRO GLU LEU
 24. PRO LYS GLU GLU VAL ALA ASP PRO THR VAL LYS PRO GLU PRO PRO LYS GLU PRO GLU LYS PRO LYS GLU
 25. LEU GLU LYS PRO LYS GLU LYS PRO GLU GLU LYS PRO LYS GLU LYS PRO LYS LYS PRO LYS LYS GLU GLN
 26. ARG ASP LEU PRO LYS SER ASP ARG GLN ILE ASP SER ASN SER SER ILE ASN GLN GLN ALA THR THR THR GLY
 27. ASN ILE THR THR ASN ASN PRO ASN LEU VAL GLY LYS GLY ASN SER THR ASP GLU VAL ASN ALA TYR ARG
 28. SER ALA LEU ARG ARG GLU ILE GLU LYS HIS LYS ARG TYR PRO ASN ARG ALA ARG MET MET ARG LYS GLN
 29. GLY VAL VAL THR ILE THR PHE HIS LEU ASN ASN ALA GLY VAL ILE SER ASN ALA ARG ILE SER LYS SER
 30. SER GLY SER GLU GLU LEU ASP ASN ALA ALA LEU VAL ALA VAL ASN ASN ALA ARG PRO ILE SER LYS SER
 31. ALA GLY MET PRO ASN GLU VAL SER VAL PRO VAL SER PHE ARG ILE THR ASN STOP LYS SER ALA VAL ASN
 32. PHE THR ALA LEU PHE SER LEU TYR STOP ASN SER SER HIS LEU CYS THR TYR LEU GLY HIS LEU SER STOP LYS
 33. LEU VAL HIS PHE LEU SER SER TYR STOP ASP ILE STOP ARG LEU SER SER ALA HIS GLN STOP GLY LYS STOP
 34. LYS STOP LEU LYS ASN PRO ILE PHE ALA

7.5 ANÁLISIS DE RESULTADOS EN EL SISTEMA DE INFORMACION BIOINFORMÁTICO

Teniendo el sistema de información Bioinformático fue posible realizar la caracterización de la proteína del Citocromo C de *Kluyveromyces fragilis*, dando como resultado según la herramienta Bioinformática, lo siguiente:

7.5.1 Análisis en Fasta. Para realizar el análisis comparativo de *Kluyveromyces fragilis* en esta herramienta bioinformática se utilizó un microorganismos llamado *Kluyveromyces marxianus*

```

# /seqprg/slib/bin/fasta34 -s BL50 -w 80 -m 5 -m 6 -f -10 -g -2 -q @
/www/htdocs/tmp/tmp_6686.lib 2
FASTA searches a protein or DNA sequence data bank
  version 3.4t24 July 21, 2004
Please cite:
  W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

Query library @ vs /www/htdocs/tmp/tmp_6686.lib library searching
/www/htdocs/tmp/tmp_6686.lib library
  1>>>gi|51893954|ref|YP_076645.1| putative 5-formyltetrahydrofolate cyclo-ligase
[Kluyveromyces fragilis] - 246 aa
  
```

vs /www/htdocs/tmp/tmp_6686.lib library

197 residues in 1 sequences
Altschul/Gish params: n0: 246 Lambda: 0.158 K: 0.019 H: 0.100

FASTA (3.47 Mar 2004) function [optimized, BL50 matrix (15:-5)] ktup: 2
join: 36, opt: 24, open/ext: -10/-2, width: 16
Scan time: 0.000

The best scores are: opt bits E(1)
gi|51892330|ref|YP_075021.1| 5-formyltetrahydrofolate cyclo-ligase [S (197) 52 18.7
0.11 [align](#)

>>gi|51893954|ref|YP_076645.1|, 246 aa vs /www/htdocs/tmp/tmp_6686.lib library

>>gi|51892330|ref|YP_075021.1| 5-formyltetrahydrofolate cyclo-ligase [Kluyveromyces marxianus (197 aa)
initn: 48 initl: 48 opt: 52 Z-score: 62.2 bits: 18.7 E(): 0.11
Smith-Waterman score: 101; 26.396% identity (46.193% similar) in 197 aa overlap (28-216:29-197)

[Entrez lookup](#) [Re-search database](#) [General re-search](#)

>gi|518 28- 216: -----:
 10 20 30 40 50 60 70
gi|518 MEKGQIRQVWVDELTRRGQAAPKPVHGRIPNFVGSAAAAARLRELEVYRQARVVVKVGPDSPLHP IRAMVLRDGGKLYM
 : : : : : : :
gi|518 MDLSKQELRRRMIAARQALDPAERARLSGRAQRAVLAAPPEARARTVLLY-----IPVRGEVDTAALAAAGRQAGKRLLL
 10 20 30 40 50 60 70

 80 90 100 110 120 130 140 150
gi|518 P-TPRLTAGFVV--LEGVEPGREREATSLTNFRFRFGREVRLEEIDP--IDLVVAGTVAVSQSTGVIRIGKAGAGYDMEFAL
 : : : : : : : :
gi|518 PRVERAERGLRLHRWDGTPEQLVVRGAYGIPPEPRP-----DLPQEDPRAVDLVVVPGVAFDRR-GRRLGYGGGYDRLLPQ
 80 90 100 110 120 130 140

 160 170 180 190 200 210 220 230
gi|518 LCRLGKLRPDPVVTAVHDLQVVDAPDHPGLVLPWDPHDISVDYIVTP---TRTIARRAHPQPTIDWSLLEPERLRA
 : : : : : : : :
gi|518 LAR-----AAAIGLGYGFQLVE-----RLPAEAHDVPLDALATDLGLTRFPA--SADPSP
 150 160 170 180 190

 240
gi|518 LQPLLELRRMQGQGNP

246 residues in 1 query sequences
197 residues in 1 library sequences
Scomplib [34t24]
start: Tue Oct 19 18:54:59 2004 done: Tue Oct 19 18:55:00 2004
Total Scan time: 0.000 Total Display time: 0.016

Function used was FASTA [version 3.4t24 July 21, 2004]

7.5.2 Análisis en Clustal X. De igual manera, para realizar el análisis comparativo de *Kluyveromyces fragilis* en esta herramienta bioinformática se utilizó el microorganismos *Kluyveromyces marxianus*.

CLUSTAL X (1.83) multiple sequence alignment

```

Kluyveromycesmarxianus      ATTGAATAATGGGGCTGCAAAAATTAAGGGTGCTAAAATATATACCAGTA
Kluyveromycesfragilis      -----

Kluyveromycesmarxianus      CTACCTAAAGCAGTATGCATTCCCATTATTACTTGGAGGAAGGGCCTTCG
Kluyveromycesfragilis      -----

Kluyveromycesmarxianus      AAAGCGTTTGAAACAGGAGTAGGAGTAGGAGTATTACGGAGAAATGGGCA
Kluyveromycesfragilis      -----

Kluyveromycesmarxianus      TTGAATAGAGGTAGCAAAGGAAAAGCCGTGACAATGCGGCAATGTACCCA
Kluyveromycesfragilis      -----

Kluyveromycesmarxianus      ACCTATGAGTCTGTGTGTGTAAAGTGTGTGTCTTTTTTTTTTTAATTT
Kluyveromycesfragilis      -----

Kluyveromycesmarxianus      TAATTTTTTTTTTTAATTATTGAAAAACCAGGACTACGTCAGGTTAGGAG
Kluyveromycesfragilis      -----

Kluyveromycesmarxianus      AATGGATATTGAAAATATTAATGTACCAGTAGTTTTGCATGTATAATTAT
Kluyveromycesfragilis      -----ATCTTAGAAGAGAT----TGTCGGTGAATTTAC-CACTTCAACAG
                * * * * * * * * * * * * * * * * * * * *

Kluyveromycesmarxianus      TATTACCAATAGACGAAGTGGAGGTATTAGAGCGATTTGGGGGAATTGTG
Kluyveromycesfragilis      CCCCATCAATTAACGA--TGAAGTTATCCCA-CAATCAGACGGTTCCTCTT
                * * * * * * * * * * * * * * * * * * * *

Kluyveromycesmarxianus      ATTGTATATATAGTGTGTATGTTTTACGGTACTGAAAGATCGTGTGTGTA
Kluyveromycesfragilis      ATC--ATTGAGGGCTCCGCCAATTTACGTGATTTGAATAAATTATT-TGA
                ** * * * * * * * * * * * * * * * * * * * *

Kluyveromycesmarxianus      GGAAAAATAAGGCACAGGAAAAATTCAGATTTTTTAAAGGTTTATTACTA
Kluyveromycesfragilis      CTGGAATCTCGATACCGAAGATGCACGTACCTTCAACGGCTTAATT--TT
                ** * * * * * * * * * * * * * * * * * * * *

Kluyveromycesmarxianus      GGAGCAGTAGTAGTAGTAGAGTAGTGGTAGCAAAAGCCGAGATCTAGGTA
Kluyveromycesfragilis      AGAGCATT-----TAGAAGA--AATTCAGAAGAAGGAACGGTATGTGAA
                ***** * * * * * * * * * * * * * * * * *

Kluyveromycesmarxianus      GAGACAGATCTGTAATACATAGCTAGTTGAAGTTGAAGAAGTGCAAACGT
Kluyveromycesfragilis      ATTAATGGGCTACAAATCACGATT--CTAGAAGTGAATTACAACATG-AT
                * * * * * * * * * * * * * * * * * * * *

Kluyveromycesmarxianus      TAAGAAAAAAATGTCTAAATTTGATGTGAACAGTTATTGAGTGAATTG
Kluyveromycesfragilis      TAAACAAGCCAAAGTC---ATTAACCTTAA---TTCAACATGTGG-CTA
                *** * * * * * * * * * * * * * * * * * * *

Kluyveromycesmarxianus      AACCAGGATGAAAAGATTTCTTACTTTCTGCAGTTGATTTCTGGCATA
Kluyveromycesfragilis      AGCGATGTCATCAAGACATCGCTTTTTTCTTCCGTACATGAATGTTTGAT
                * * * * * * * * * * * * * * * * * * * *

Kluyveromycesmarxianus      TAAGAAGATTGAACGGTTGGGAATCCAGCGGTGAGGGTTTCTGATGGTC
Kluyveromycesfragilis      CCAACAC----AACATTTA----TTCAACATT--GGATAAATAATCATC
                * * * * * * * * * * * * * * * * * * * *
    
```


Kluyveromycesmarxianus Kluyveromycesfragilis	TGGAGAATGGCCAGAATCTACTTCAAACAACACCAAGGAAACCTCGGAC AAGTGAATGTACCAAAGCAAGTT----CAACAGTTGCGTTTCGTTTCAGA *
Kluyveromycesmarxianus Kluyveromycesfragilis	CTGTTGAGAGAGATTGCTGCTGACTCTATTGTTTTATGAAGAACAAA- TGATTTAGCCAAATTGCTGACTATTACGGAAAGTGGTGAAATTTTTTATC *
Kluyveromycesmarxianus Kluyveromycesfragilis	ACAATTATCTTACCTCTAAAGAAAGAAGACAATATCATGTCATTGGCCCA ACGATAAACCGATTACTATAGTAGCATTGGAACAAGAAATCAGTAATTGG *
Kluyveromycesmarxianus Kluyveromycesfragilis	AATGCTAAAGCAAAGACTAGTTCGGTGGTGGTT-CAGCATCTATGAACT GAAAA-AGATCAAAAAGTCACCTTGAAGGTAGATGCAAAATCCA---GTT *
Kluyveromycesmarxianus Kluyveromycesfragilis	CCTACTATGTTGTTTCTCCGTATGAAGGTATCGTCAATAAGCTGGGCAAA TCCAAGATTTGTTTCTATCACTGATATATGGCTAAAAATGAAATTTAAA *
Kluyveromycesmarxianus Kluyveromycesfragilis	GAGGTCGATTACACCGTAGGGCCCTATTACACAAATCGATTGGAGGTTT AATGTCGCT---ATCGTGACGGTTAA---AGAAAAGGCACCTCAA----T *
Kluyveromycesmarxianus Kluyveromycesfragilis	GGCAGAGAGTAGTTTGATCGATGCTGCAAAACCAGCAGATGCTGAAAAATG GATAGATAAAAG-----TCGTTCTTGC-----ATCGGGTTTGCAATTC *
Kluyveromycesmarxianus Kluyveromycesfragilis	CTGGATTAATTGC-CAAGTTTTACTCCAATCCAGTAGAAGAGAGATCTGA ATTGCTTTTTACGCAAGTTTTGTCTCTTCTGTATTGGA-----TTGT *
Kluyveromycesmarxianus Kluyveromycesfragilis	AGATGAAGAACCATTCCACGTTACCAAAGTCAATAGATCCAATGTTTAC- ACAAAAAGA-----CGATGACAGCGGAATGGATTGCTGCCGATA *
Kluyveromycesmarxianus Kluyveromycesfragilis	TTATTTGATTTCAAACATGAGAAAGTGGATCCAAAGAACCCTTACTTTTT TCATCTCAACTC--ATATTTCCATGGAATGCTGGCGGTACCGTTTTAG *
Kluyveromycesmarxianus Kluyveromycesfragilis	TGTAACCTTAACCGGA--CAGTAC--GTGCCTCAAGAAGATGGTGATTAT AAGAACCAGAGCCGGAACCAGAGCCGGCGCCTCCGGTAG-TAGAACCTGA *
Kluyveromycesmarxianus Kluyveromycesfragilis	ATCTTCAGTCTTCAAGTTTATGGTTCTGGTTTGTCTACTTAAACGATGA ACTGCCAAAAGAAGAAGTCGCAGATCCGACG-GTAAACCTGAGCCACCA *
Kluyveromycesmarxianus Kluyveromycesfragilis	GTTGATTATTGACCAAAAGCACAACCAAGAAAGGGGTAGTTTCTGCTTTG AAAGAACCAGAAAACCAAAAGAGCCTGAAAAGCCAAAAGAGAAAC--CG *
Kluyveromycesmarxianus Kluyveromycesfragilis	GAGCTGGTACCAAAGAAAGAACCAAAAAGTTGACTTTGAAGAAGGGCCAA AAGGAAAACCAAAAAGAAAGCCGAAAAAAC-----CGAAGAAAGAACAA *
Kluyveromycesmarxianus Kluyveromycesfragilis	GTTTATAATGTAAGAGTTGAGTACGGTTCTGGCCAACTTCAAGTTTGGT CGTGATTTACCAAAGTCAGATCGCCAAATTGATTCATAATTCA---TCGAT *

Kluyveromycesmarxianus Kluyveromycesfragilis	TGGGGAATTCGGTGCAGGTGGATTCCAAGCTGGTGTCTTAAGGCGATCG CAATCAA-----CAAGCGA--CCACAACAGGTAACATCA---CAACCA * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	ATGATGACGAGGAGATTAGAAACGCAGCAGAATTAGCAGCTAAGCATGAT ATAATCCGAATCTGGT-----CGGTAAAGGTAATAGCACAG-ATGA- * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	AAGGCTGTGTTGATAATTGGATTAATGGTGAATGGGAAACCGAAGGTTA -AGTCAATGCTTATCGCTCGGCTTTACGCAGA----GAAATTGAAAAACA * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	TGACAGAGAAAACATGGATTTGCCAAAAAGAACAATGAATTAGTTCGTG T--AAACGCTATCCAAACCGTGCACGCATGATGCGCAAAACAAGGTGTGG * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	CTGTTTTGAAAGCAAATCCAAATACTGTTATCGTTAACCAATCAGGTACC TAACAATCACGTTCCATCTTAATAATGCCGGCGT-AATAGTAATGCCGG * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	--CCAGTTGAGTTCCCTTGGTTAGAAGAGGCAAATGCGCTAGTTCAAGCT AATCAGC-AAATCTTCTGGCTCAGAAGAATTAGATAA----- * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	TGGTACGGTGGTAATGAATTGGGTAATGCTATCGCAGATGTCTTGTACGG CGCTGCACTGGTAGCTG--TCAATAATGCCCGTCCAATTGGTCCATTGCC * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	TGACGTGGTTCCAAATGGTAAGTTATCGCTCTCTGGCCATTTAAGTTGC TGCTGGTATGCCAAATG--AAGTGAGCGTTCCTGTGAGTTTCAGAATCAC * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	AAGATAATCCAGCCTTTTTAAACTTCAAGACCGAGTTCGGAAGAGTTGTT AAATTAATAAAAGTGCAGTAAATTTT----ACCGCACTTT-----TTTCT * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	TACGGTGAGGATATCTTTGTTGGTTATAGGTAAGTACGAAAAGCTTCAAAG CTCTATTAGAATTCCTCTCATTTGTGCACTTATCTAGGACATCTTTTATA * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	AAAGGTAGCCTTCCCCTTCGGATATGGTCTATCGTATACAACATTCGAAC AAAACCTGTGCATTTTTTATCTTCCCTAT-TAGGATATCTAACGATT--AT * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	TAGATATTTCTGACTTCAAGGTAAGTATGATAAGATAGATATTTTCAGTT CTTCTGCTCATCAAT--AAGGTAATAAAAAATGAC-TAAAAAACCCCTATT * * * * * * * * * * * * * * *
Kluyveromycesmarxianus Kluyveromycesfragilis	GATGTGAAGAATACTGGTGATAAATTTGCTGGCTCCGAGGTGGTGCAAGT TTCGCCA----- * *
Kluyveromycesmarxianus Kluyveromycesfragilis	CTACTTCAGCGCTCTAAACTCTAAGGTCTCGAGACCGGTTAAGGAGTTGA -----
Kluyveromycesmarxianus Kluyveromycesfragilis	AGGGATTCGAAAAAGTTCATTTGGAACCAGGTGAGAAGAAGACAGTTAAT -----

Kluyveromycesmarxianus Kluyveromycesfragilis	ATTGAACTAGAATTGAAAGATGCAATTCCTACTTTAACGAAGAGCTCGG -----
Kluyveromycesmarxianus Kluyveromycesfragilis	TAAATGGCACGTTGAAGCAGGTGAATACTTGGTTTCAGTTGGTACTTCTT -----
Kluyveromycesmarxianus Kluyveromycesfragilis	CTGATGATATACTTTCCGTCAAAGAGTTTAAAGTAGAAAAAGACTTGAC -----
Kluyveromycesmarxianus Kluyveromycesfragilis	TGGAAAGGTTTGTGAAGAATGCTAAAATGGTTTAGTGTTC CAATCCAGG -----
Kluyveromycesmarxianus Kluyveromycesfragilis	TGCAAGTTCATTGTACAGTTATAATTATATATATGTGTAACGTGCAATG -----
Kluyveromycesmarxianus Kluyveromycesfragilis	ACCCATCATAACAGAGAGTTATTTCGCTATTAAACACAAAAACAAACAA -----
Kluyveromycesmarxianus Kluyveromycesfragilis	CCAGTAACTACATGAAATGAATAGGTATTAAAGTCTTGAATTTCCCATGA -----
Kluyveromycesmarxianus Kluyveromycesfragilis	AATACGAACCTTTACAGTTTGAACTTAACAATATGGCCTTTTTAAGCCAT -----
Kluyveromycesmarxianus Kluyveromycesfragilis	ATCAACCTCATGAAATTACGGGAAGGAAACGATGAAAGGTCAAAGCCTA -----
Kluyveromycesmarxianus Kluyveromycesfragilis	TTAAGCATAGTACTGCATCTAAGGGAGAGTGGTACCCTTTACAAGGTTTT -----
Kluyveromycesmarxianus Kluyveromycesfragilis	TTGTTTTCCAGAGTAGTTTGC GAATACTACAAATACGTTGAATTTTTGAA -----
Kluyveromycesmarxianus Kluyveromycesfragilis	GTTACATTTTCATTACGTAACATTTAAACTAATTAGATAGTAAATAATAAA -----
Kluyveromycesmarxianus Kluyveromycesfragilis	CATCGCAATACACATTAATCATTGAATTAACCTATACAGCTTAGATTTCG -----
Kluyveromycesmarxianus Kluyveromycesfragilis	CAGAATATATTCTAACAGTAACTGTTAGAATAATCCATTAAGAGTCTAAA -----
Kluyveromycesmarxianus Kluyveromycesfragilis	GCCTGTGGCTTTTTAATGATGAATCCACAAGACTTTTTGCTGCAATTA -----

Kluyveromycesmarxianus Kluyveromycesfragilis	GGAGAAGATCAAGCAGAATAAAAAACAAATTATGAAGTACGGAAACTTCT -----
Kluyveromycesmarxianus Kluyveromycesfragilis	TGCACCTAACAAAATATATTGAAAAGATGGCTTTAACAGATTCTGCCTC -----
Kluyveromycesmarxianus Kluyveromycesfragilis	TGAAAGCTTTTCGACATGATCAGCATCGCTCTTTAGAGGCTCTTGCTCTT -----
Kluyveromycesmarxianus Kluyveromycesfragilis	TCAAATTTTGAGCATTGCAACTCTAACGTCATTTCGTTGGACCAAAGTT -----
Kluyveromycesmarxianus Kluyveromycesfragilis	GCCCTGACTGAGCCAAGAATGCTTGATCAACGATCCTTTCTTGGGTTTGG -----
Kluyveromycesmarxianus Kluyveromycesfragilis	AGCTTCAAAGACAACCTTCTAATTCTTCTAAGCTTCTACCCTTAGTTTCAA -----
Kluyveromycesmarxianus Kluyveromycesfragilis	CGAAGAAGAAGTAGATAACAATAAATTCGAAAATATCGAAGAAAACGTAG -----
Kluyveromycesmarxianus Kluyveromycesfragilis	AACACATAGAACCAATATTTGATATTCTTCATGCCTTTGGAGTAGCAAAT -----
Kluyveromycesmarxianus Kluyveromycesfragilis	TGATTAACAAATTGGGCAACACCAGAAACCAACATGTTGAGGAGTTGGGC -----
Kluyveromycesmarxianus Kluyveromycesfragilis	CTTAGATCT -----

7.5.3 Análisis en Rasmol. Rasmol permite realizar una representación grafica del Citocromo C de Kluyveromyces fragilis, a continuación se puede ver la representación de esta molécula en dos vistas diferentes: (Figura 38, 39)

Figura 38. Representación tridimensional de la secuencia de aminoácidos de *Kluyveromyces fragilis*.

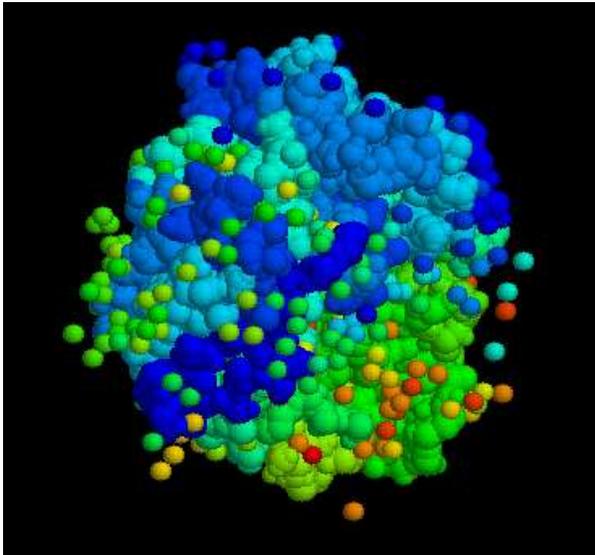
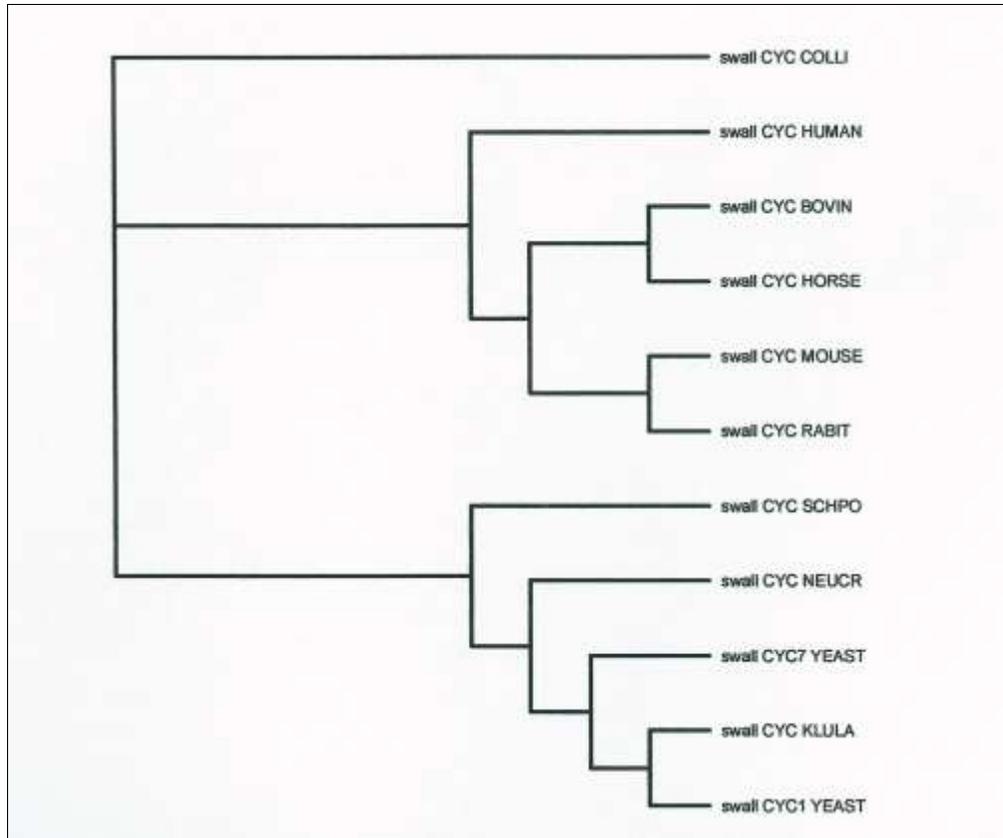


Figura 39. Representación de *Kluyveromyces fragilis* en Rasmol vista 'cintas'.



7.5.4 Análisis en TreeView. El análisis de *Kluyveromyces fragilis* en esta herramienta Bioinformática permitió conocer el árbol del origen de la filogenia. (Figura 40)

Figura 40. Filogenia del citocromo c de *Kluyveromyces fragilis*



7.5.5 Análisis en KEGG. El análisis de los genes de enzimas de *Kluyveromyces fragilis* que participan en la fosforilación oxidativa. (Figura 41)

Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 1
Number of Hits to DB: 164
Number of extensions: 61
Number of successful extensions: 1
Number of sequences better than 10.0: 1
Number of HSP's better than 10.0 without gapping: 1
Number of HSP's gapped: 1
Number of HSP's successfully gapped: 1
Number of extra gapped extensions for HSPs above 10.0: 0
Length of query: 105
Length of database: 711,361,694
Length adjustment: 80
Effective length of query: 25
Effective length of database: 711,361,614
Effective search space: 17784040350
Effective search space used: 17784040350
Neighboring words threshold: 9
Window for multiple hits: 0
X1: 16 (7.3 bits)
X2: 129 (49.7 bits)
X3: 129 (49.7 bits)
S1: 42 (21.9 bits)
S2: 68 (30.8 bits)

8. CONCLUSIONES

Se logró una profundización en un área científica que desde hace muy poco tiempo se ha explorado inclusive a nivel mundial. A nivel local, la UNAB hace su primer acercamiento por medio de este trabajo de grado investigativo sobre la Bioinformática.

El desarrollo de esta investigación ha permitido tener claridad en conceptos biológicos que hasta el momento eran ajenos. Conocer sobre las levaduras tipo el **Kluyveromyces fragilis**, las mitocondrias, los ribosomas, el ciclo de Krebs y en general una terminología biológica que de pronto parecía muy compleja, pero que con el tiempo se hace más amigable para el desarrollo de la investigación.

De la fase experimental preliminar se puede concluir que el microorganismo **Kluyveromyces fragilis** crece muy bien sobre el sustrato y la cosecha de biomasa es buena. Igualmente al realizar un estudio previo de secuenciación de la biomasa recolectada después de 10 días de cultivo, se obtuvo 1 gramo; este se trató con el reactivo de Edman y se envió para la lectura en un secuenciador Shimadzu a la Universidad de Texas at Houston, departamento de Biología molecular, logrando obtener la secuencia genética de este microorganismo.

El análisis realizado a las diferentes herramientas bioinformáticas permitió comprender el funcionamiento de cada una de ellas y sobre todo conocer los requerimientos. Al utilizar las diferentes herramientas bioinformáticas se ha observado como se pueden representar un sin número de datos desde diferentes vistas, los cuales hasta hace poco eran únicamente entendibles para los investigadores en Biología y áreas afines.

El modelado con la herramienta UML permitió tener un claro conocimiento de los requerimientos del sistema, los cuales a través de los diferentes diagramas que provee esta herramienta permitió modelar la funcionalidad del sistema, estos diagramas son: diagrama de casos de uso, nivel 0 y 1; diagrama de secuencias; diagrama de clases. UML permitió con la elaboración de estos diagramas obtener claridad y un punto de inicio para la implementación del sistema a desarrollar.

Se logró realizar la integración de las herramientas bioinformáticas descritas en los objetivos mediante un Sistema de Información Bioinformático que permite administrar herramientas Bioinformáticas, crear nuevos archivos de contenido biológico, verificar compatibilidad de archivos con herramientas para ser analizados y por ultimo ejecutar estas herramientas.

A partir del Sistema de Información Biológico se pudo realizar la caracterización del Citocromo C de **Kluyveromyces fragilis** tomando como secuencia de comparación la de **Kluyveromyces marxianus**, obteniendo datos de gran importancia para el desarrollo de futuras investigaciones con esta levadura y concluir que existen grandes semejanzas las cuales denotan una línea troncal común entre las dos.

El análisis de los resultados demuestra el cumplimiento tanto de los objetivos propuestos como de la metodología. Los aportes demostrados en los resultados permiten por primera vez mostrar las características moleculares de una levadura con un importante uso como prebiótico.

9. RECOMENDACIONES

El análisis de las conclusiones permiten la utilización de estas herramientas integradas como una muy valiosa alternativa para la determinación estructural de las proteínas que componen no solo el Citocromo sino de la molécula completa de microorganismos.

De esta forma la contribución de la presente investigación permite iniciar la línea de investigación en Bioinformática con aportes viables y a la vez confiables, toda vez que cualquier dato que se introduce dentro de los parámetros señalados en la investigación dan como resultado la comprensión molecular.

BIBLIOGRAFIA

- [1] Pagina Web. VSNS BioComputing Division [Citado 13 de Abril 2004].
<<http://www.techfak.uni-bielefeld.de/bcd/ForAll/welcome.html>>
- [2] C Giulivi, JJ Poderoso, A Boveris Production of nitric oxide by mitochondria. J. Biol. Chem. 1998; 273: 11038-11043.
- [3] GIBAS, C.; Jambeck P.2001. Developing Bioinformatics computer skills. O'Really & Associates, Inc., United Status of America.
- [4] CHALELA A, Graciela. 2000. Introducción a la bioinformática. Cuadernos de Ciencias. UIS.
- [5] Revista Colombia: Ciencia y Tecnología, vol. 20, num. 3, pág 36, Julio del 2.002
- [6] CHALELA A, Graciela. 2003. Funciones Probióticas de *Kluyveromyces fragilis*. Revista Colombiana de Biotecnología. UNAL (En prensa)
- [7] LEHNINGER L, Albert. 1990. Bioquímica, Segunda edición. Ediciones Omega. Barcelona.
- [8] PEÑA DIAZ, Antonio et all. 1990. Bioquímica, Segunda edición. Editorial Limusa. Mexico.
- [9] Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson J: The mitochondrion. In Molecular Biology of the Cell, 2nd edition (Robertson M. Ed) Garland Publishing, New York 1989; 342-356
- [10] Casey R: Membrane reconstitution of energy-conserving anzymes of oxidate phosphorylation. Biochim Biophys Acta 1984; 768: 319-347.
- [11] TZAGOLOFF A: Evolution of mitochondrial studies. I Mitochondria (Siekevitz P, Ed.) Plenum Press, New York 1982; 1 – 14.
- [12] Smoly J, Kuylenstierna B, Ernster L: Topological and functional organization of the mitochondrion. Proc Natl Acad Sci USA 1970; 66: 125-131.
- [13] Klingenberg M: Enzyme profile in mitochondria. Meth Enzymol 1967; 10: 3-7.

- [14] Ernster L, Kuylenstierna B: Outer membrane of mitochondria, in *Membranes of Mitochondria and Chloroplasts* (Racker E. Ed) Van Nostrand Reinhold, New York 1970; 172-212.
- [15] Tzagoloff A: Oxidate pathways in mitochondria, in *Mitochondria* (Siekevitz P, Ed) Plenum Press, New York 1982; 39-60.
- [16] Williamson J, Copper R: Regulation of the citric acid cycle in mammalian systems, *FEBS Lett.* 1980; suppl 117: 73-85.
- [17] Ernster L, Lee C: Biological oxidoreductions. *Annu Rev Biochem* 1964; 33: 729-788.
- [18] RUIZ A. M. 1999. *Bioquímica estructural: Conceptos fundamentales y 383 tests con respuesta razonada.* Editorial Tébar flores, Alfa Omega grupo editor. Mexico.
- [19] DAVIS, Leonard et al. 2001. *Basic Methods in Molecular Biology.* Second edition Appleton And Laug (Ed).
- [20] JAMES Senn. 1991. *Analisis y Diseño Sistema Información.* Edición Paperback.
- [21] Roger S. Presuman. 2002. *Ingenieria del Software, Un Enfoque Practico 5* Edición Paperback.
- [22] Gaskell SJ. Electrospray: Principles and practice. *J Mass Spectr* (1997); 32: 677-688.
- [23] Gygi SP, Rist B, Aebersold R. Measuring gene expression by quantitative proteome analysis. *Current Opinion in Biotechnology* (2000); 11: 396-401.
- [24] Giometti CS. Et al. A two-dimensional electrophoresis data base of human breast epithelial cell proteins. *Electrophoresis* (1997); 18: 573-581.
- [25] Hannash SM et al. Two-dimensional gel electrophoresis of cell proteins in childhood leukemia, with silver staining: a preliminary report. *Clin Chem* (1982); 28:1026-1030.
- [26] Niimi M. et al *Candida albicans* pathogenicity: a proteomic perspective. *Electrophoresis* (1999);11:2299-308.
- [27] Sanchez JC et al. Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients. *Electrophoresis.* (1997);18:324-7.

- [28] Kyrpides N. Genomes Online Database (GOLD): A monitor of complete and ongoing genome projects world wide. *Bioinformatics* (1999); 15:773-774.
- [29] Lawson SR et al. Quantitative protein changes in metastatic versus primary epithelial ovarian carcinoma. *Gynecol Oncol* (1991); 41:22-27.
- [30] Lin X et al. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* (1999); 402:761-8.
- [31] Mann M, Hojrup P and Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom* (1993); 22:338-45
- [32] Mann M, Talbo G. Developments in matrix-assisted laser desorption/ionization peptide mass spectrometry. *Curr Opin Biotechnol.* (1996);7:11-19. Review
- [33] Merchant M and Weinberger RS. Recent advancements in surface-enhanced laser desorption ionisation-time of flight mass spectrometry. *Electrophoresis* (2000); 21:1164-77.
- [34] A Boveris, L Costa, E Cadenas, JJ Poderoso Regulation of mitochondrial respiration by adenosine diphosphate, oxygen, and nitric oxide. *Meth. Enzymol.* 1999;301: 188-198.
- [35] Okuzawa K et al. Characterization of gene expression in clinical lung cancer materials by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* (1994);15:382-390.
- [36] Shevchenko A et al. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc Natl Acad Sci U S A* (1996);93:14440-5.
- [37] Traini M et al. Towards an automated approach for protein identification in proteome projects. *Electrophoresis* (1998);19:1941-9.
- [38] Wilkins MR et al.. High-throughput mass spectrometric discovery of protein post-translational modifications. *J Mol Biol* (1999); 289:645-57.
- [39] Velculescu VE et al. Characterization of the yeast transcriptome. *Cell* (1997); 88:243-251.
- [40] Wilkins MR et al. Progress in proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* (1995); 13:19-50.

[41] BOOCH, Grady. RUMBAUGH, James. JACOBSON, Ivar. El lenguaje unificado de modelado UML. España: Addison Wesley, 1.999.

[42] WINDOWS
<http://www.microsoft.com/colombia>

[43] ESCUDERO, Fernando. Al descubierto Java. España: Prentice Hall, 1.997, 1ra. edición.

[44] BOUDREAU Tim. NetBeans: The Definitive Guide

[45] Página Web. Xinox Software. JCREATOR. [Citado el 22 Junio de 2.004]. <<http://www.jcreator.com>>

[46] Página Web. Microsoft Access, Bases de datos. [Citado el 22 Junio de 2.004]. <<http://club.telepolis.com/ortihuela/access.htm>>

[47] Página Web. RasMol and OpenRasMol. RASMOL. [Citado el 20 Marzo de 2.004]. <<http://www.openrasmol.org/>>

[48] Página Web. Taxonomy, Systematics, and Bioinformatics at the University of Glasgow. TREEVIEW. [Citado el 15 Mayo de 2.004]. <<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>>

[49] Página Web. The Marine Biological Laboratory. FASTA. [Citado el 6 Mayo de 2.004]. <<http://workshop.molecularevolution.org/software/fasta/>>

[50] Página Web. Institut de Génétique et de Biologie Moléculaire et Cellulaire. CLUSTAL X. [Citado el 10 Marzo de 2.004]. <<http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html>>

[51] Pagina Web. Kyoto Encyclopedia of Genes and Genomes. [Citado el 20 Agosto de 2.004] <<http://www.genome.jp/kegg>>

[52] Página Web. Yeast microarray global viewer. [Citado el 14 de Mayo de 2.004]. <<http://transcriptome.ens.fr/yimgv/who.php>>

[53] Página Web. Sequence Retrieval System. Network Browser for Databanks in Molecular Biology. [Citado el 5 Junio de 2.004]. <<http://srs.embl-heidelberg.de:8000/srs5>>

[54] Página Web. National Center for Biotechnology Information. ENTREZ. [Citado el 22 Junio de 2.004]. <<http://www.ncbi.nih.gov/Entrez>>

[55] Página Web. Nacional Center for Biotechnology Information. NCBI BLAST SERVER. [Citado el 20 Junio de 2.004]. <<http://www.ncbi.nlm.nih.gov/BLAST>>

[56] GLICK R. Bernard et all. 2002. Molecular Biotechnology. ASM PRESS. Washington.

[57] Hannash SM et al. Data base analysis of protein expression patterns during T-cell ontogeny and activation. Proc Natl Acad Sci USA (1993); 90:3314-3318.

[58] CHALELA A, Graciela. 2001. Caracterización de levaduras. Producciones UIS.

ANEXO A.

PANTALLAS DEL SISTEMA DE INFORMACIÓN BIOINFORMÁTICO

El sistema de información Bioinformático esta orientado a todas aquellas personas investigadoras en el campo de la Bioinformática, pues este es de gran ayuda al integrar como ya se ha mencionado en este documento diversas herramientas para el análisis biológico.

Cuando el usuario ingresa al sistema podrá visualizar la siguiente ventana (Figura 42), en donde encontrará las diferentes funciones o módulos que tiene este sistema.

Figura 42. Ventana principal del sistema Bioinformático



Una vez que el usuario ha elegido la opción Administrar Herramientas, visualizará la siguiente ventana: (Figura 43)

Figura 43. Administrar Herramientas

Aplicación de herramientas Bioinformáticas

Buscar Agregar Actualizar Limpiar Eliminar Consultar Ayuda

Id:
Nombre:
Link:
Ubicacion:
Descripcion:
Extension1:
Extension2:
Extension3:
Extension4:
Extension5:

Conexión exitosa

En esta ventana el usuario podrá administrar las herramientas bioinformáticas existentes, es decir las que están incluidas en la base de datos o inclusive las que desee ingresar. En el menú superior el usuario encontrará dichas opciones, éstas son: Buscar, Agregar, Actualizar, Limpiar, Eliminar, Consultar y Ayuda.

Si el usuario desea realizar una búsqueda, el deberá ingresar el nombre de la herramienta Bioinformática en el campo 'Nombre' y luego dar clic sobre el botón Buscar.

El usuario podrá limpiar los campos del formulario dando clic en el botón Limpiar.

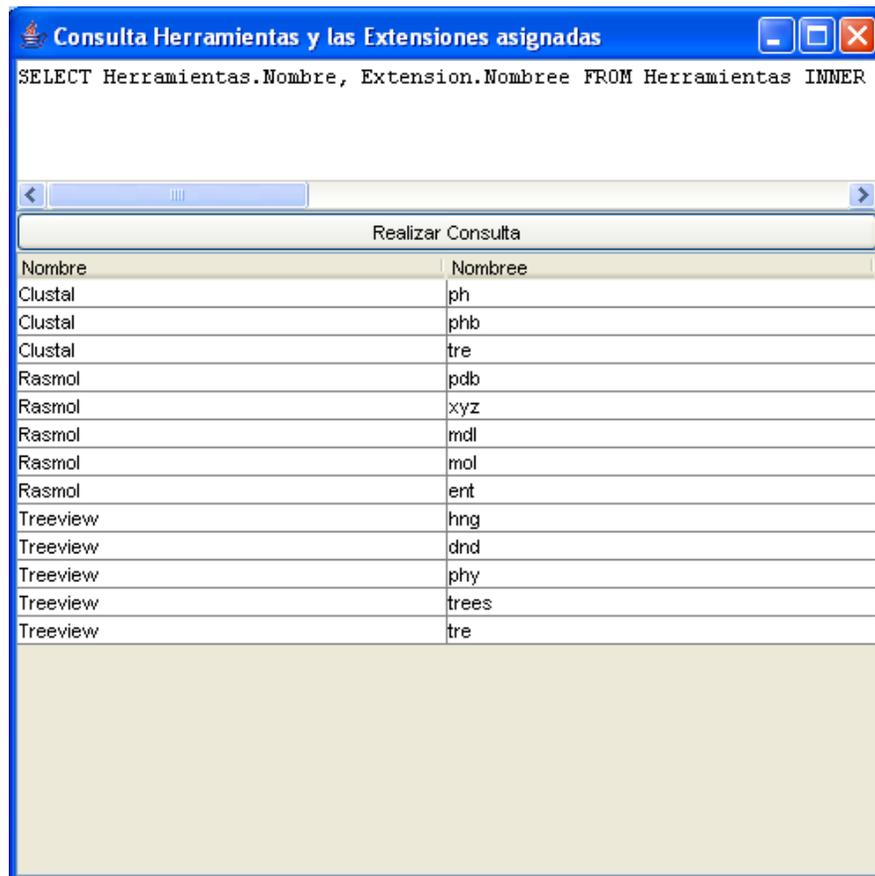
Si el usuario desea ingresar una nueva herramienta Bioinformática en línea lo podrá hacer diligenciando los campos activos y dando clic en el botón Agregar.

Si el usuario desea modificar los datos de alguna herramienta Bioinformática deberá primero consultarla y luego modificar el campo que desee, esta operación se hará dando clic en el botón Actualizar.

Para eliminar una herramienta Bioinformática, el usuario deberá primero consultarla y luego dar clic en el botón eliminar, la herramienta se eliminará por completo del sistema.

Si el usuario desea hacer una consulta general sobre las herramientas Bioinformáticas y la relación existente con las Extensiones dentro de la Base de Datos podrá dar clic en el Botón Consultar, el usuario podrá visualizar la siguiente ventana que se desplegará ha realizar esta operación: (Figura 44)

Figura 44. Consulta Herramientas relacionadas con Extensiones.



The screenshot shows a window with a blue title bar containing the text "Consulta Herramientas y las Extensiones asignadas". Below the title bar, a text area contains the SQL query: "SELECT Herramientas.Nombre, Extension.Nombre FROM Herramientas INNER". Below the query is a horizontal scrollbar. Underneath the scrollbar is a button labeled "Realizar Consulta". Below the button is a table with two columns: "Nombre" and "Nombree". The table contains 15 rows of data.

Nombre	Nombree
Clustal	ph
Clustal	phb
Clustal	tre
Rasmol	pdb
Rasmol	xyz
Rasmol	mdl
Rasmol	mol
Rasmol	ent
Treeview	hng
Treeview	dnd
Treeview	phy
Treeview	trees
Treeview	tre

Y finalmente, el usuario podrá consultar una ayuda que contempla las funciones principales de este modulo de Administración de herramientas de este sistema Bioinformático.

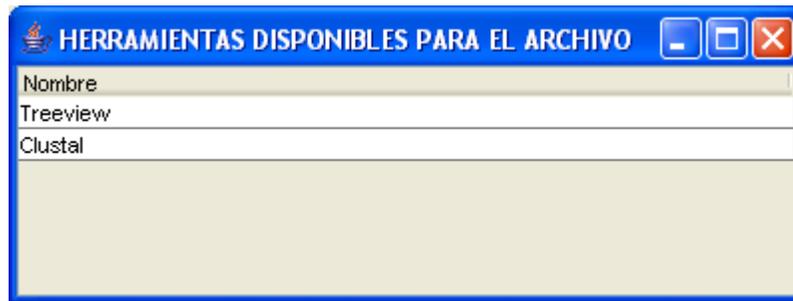
Cuando el usuario ha seleccionado del menú principal la opción de Examinar Archivo podrá conocer las herramientas Bioinformáticas con las que podrá analizar el archivo, por ello inicialmente el usuario visualizará la siguiente ventana en donde podrá seleccionar el archivo al que le desea conocer la extensión: (Figura 45)

Figura 45. Abrir archivo.



Después de haber hecho esta selección se desplegará una ventana en donde se mostrarán las herramientas Bioinformáticas que se pueden usar para el análisis biológico de la información contenida en el archivo anteriormente seleccionado. (Figura 46)

Figura 46. Herramientas Bioinformáticas compatibles.



Una vez que el usuario ha seleccionado la opción de Ejecutar Herramienta. El sistema automáticamente desplegará una ventana listando el nombre de la herramienta Bioinformática con su respectivo número de identificación dentro del sistema. (Figura 47)

El usuario podrá escribir en un cuadro de texto la herramienta seleccionada y por ultimo dar clic en el botón Ejecutar.

Una vez hecho esto, la herramienta seleccionada se desplegará y el usuario podrá interactuar directamente con ella.

Figura 47. Ejecutar Herramienta.

