

HERRAMIENTA SOFTWARE PARA EL ANÁLISIS E IDENTIFICACIÓN DE PROTEÍNAS MEDIANTE IMÁGENES DE ELECTROFORESIS EN 2D.

CRISTIAN CAMILO PEÑA ALVAREZ

LUZBY STIWENSON LOPEZ AMADO



UNIVERSIDAD INDUSTRIAL DE SANTANDER

ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

FACULTAD DE INGENIERÍAS FÍSICO MECÁNICAS

BUCARAMANGA

2009

**HERRAMIENTA SOFTWARE PARA EL ANÁLISIS E IDENTIFICACIÓN DE
PROTEÍNAS MEDIANTE IMÁGENES DE ELECTROFORESIS EN 2D.**

CRISTIAN CAMILO PEÑA ALVAREZ

LUZBY STIWENSON LOPEZ AMADO

Este proyecto es presentado como requisito para optar
al título de Ingeniero de Sistemas

Director

Henry Arguello Fuentes

Magíster en Potencia Eléctrica

Codirector

Dr. Rodrigo Torres Sáez

Director Grupo de Investigación en Bioquímica e Ingeniería de Proteínas

UNIVERSIDAD INDUSTRIAL DE SANTANDER

ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

FACULTAD DE INGENIERÍAS FÍSICO MECÁNICAS

BUCARAMANGA

2009

A mi querida familia que a pesar de pasar por diversas situaciones me han ofrecido su constante apoyo, colaboración en el transcurso de mi vida, por su amor, cariño brindado y formación que me ayudan a estar más cerca de mis metas profesionales.

A mi madre Marlen Alvarez que siempre ha confiado en mí y que con sus palabras me han enseñado a reflexionar y mejorar en mi vida personal. Por tolerarme mis actitudes y guiarme en el camino correcto.

A mi padre Jose Peña que ha sido mi ídolo desde que tengo uso de razón, siempre lo he admirado y me he esforzado por llegar a ser como él, y porque decidió estar en el mejor hogar que es nuestra familia.

A Elsa Jimenez de Joya que me abrió las puertas de su casa y de su corazón y me trató como un miembro más de su grandiosa familia. Que Dios la tenga en su gloria.

A mi hermosa novia Diana Cristina Joya por ser la luz de mi vida y ser mi fuente de inspiración, por acompañarme incondicionalmente a cumplir con este sueño, por llenar mi vida de amor, cariño, alegría y ternura. Por darme el honor de tenerla junto a mí y conocer a un ser tan maravilloso como lo eres tú.

A mis hermanas que siempre me han sacado sonrisas y arrugas en mi frente, pero me hace feliz tenerlas conmigo para seguir compartiendo buenos momentos con ellas.

A mis tías María Clemencia Peña, Marlen Peña, Trinidad Peña que en todo momento que necesite su auxilio estuvieron dispuestos a colaborarme y ofrecer su “cuotica” para tener los recursos necesarios para efectuar este proyecto.

A mis abuelos Víctor Peña, Aura Amaya y Graciela Cruz que siempre me han consentido en toda mi vida y ofrecido su amor hacia mí.

A todos mis amigos de infancia y compañeros de la universidad que me han acompañado en este camino.

Cristian Camilo

A mis padres, en quienes he encontrado el apoyo y la consideración necesarios para llevar a cabo mi propia formación personal y profesional.

A mi mami, Zoraida, por su dedicación a cultivar los valores morales y enseñarnos el valor del esfuerzo y la honradez a toda nuestra familia, y su disposición de ayudarme y aconsejarme en todas las circunstancias a las que me he enfrentado en la vida.

A mi papá, Juan Antonio, por estar siempre pendiente de las necesidades de todos los miembros de nuestra familia y me ha enseñado la importancia de no desfallecer a pesar de las dificultades.

A mis hermanos, Hernando y Laura, de quienes siempre espero lo mejor, porque han tenido la paciencia para soportar mis actitudes, a veces salidas de tono, a causa del cansancio y la falta de tiempo para cultivar el amor fraternal.

A mi gran amigo, Jorge Valero, porque siempre me apoyó y me mostró que la perseverancia siempre permite lograr las metas y mostrarme que mi futuro sí estaba en la UIS.

A toda mi familia, por su fe en mis capacidades y ser los cimientos sobre los que se sostiene mi ser.

A todos los que han permanecido a mi lado en este camino, largo y pedregoso, que llaman la vida y del cual siempre se sale bien librado si se ha compartido con sinceridad y dando lo mejor de sí para el beneficio de todos.

Luzby Stíwenson

AGRADECIMIENTOS

Los autores del proyecto desean expresar sus más sinceros agradecimientos:

A **Henry Arguello Fuentes**, Director de este proyecto, quien a pesar de sus múltiples ocupaciones siempre estuvo dispuesto a colaborar con sus conocimientos, además de su apoyo incondicional en el transcurrir de esta labor.

Al **Dr. Rodrigo Torres Sáez**, Codirector del proyecto, quien fue nuestro guía y punto de referencia en el proceso de generación de conocimiento y nos ha ofrecido desde el primer día su total soporte a este proceso.

A **Christine Hoogland** y **Khaled Mostaguir** del Instituto Suizo de Bioinformática, por su apoyo desde la distancia, y su innegable interés por el fortalecimiento de la investigación y la academia.

A todas las personas que directa o indirectamente se han involucrado con este trabajo continuo de crecimiento personal y profesional.

A la Universidad Industrial de Santander por permitirnos la oportunidad de desarrollar este proyecto, y proporcionarnos los recursos necesarios para este propósito.

CONTENIDO

INTRODUCCIÓN	1
1. OBJETIVOS	4
1.1 OBJETIVO GENERAL	4
1.2 OBJETIVOS ESPECÍFICOS	4
2. PLANTEAMIENTO DEL PROBLEMA	6
2.1. ANTECEDENTES	6
2.2. JUSTIFICACIÓN	7
3. ELECTROFORESIS	9
3.1. ELECTROFORESIS DE PROTEÍNAS EN GELES DE POLIACRILAMIDA. -	9
3.2. SDS-PAGE	10
3.3. ELECTROFORESIS BIDIMENSIONAL	11
3.3.1. PUNTO ISOELÉCTRICO	12
3.3.2. MASA MOLECULAR	13
3.4 ELECTROFORESIS Y RAZE-2D	14
4. IMÁGENES DE ELECTROFORESIS	16
4.1. MASA MOLECULAR Y SU ESCALA EN LAS IMÁGENES	16
4.2. PUNTO ISOELÉCTRICO EN LAS IMÁGENES	16
4.3. INTERPRETACIÓN DE IMÁGENES CON RAZE-2D	17

<u>5. BASES DE DATOS BIOLÓGICAS PARA EL ALMACENAMIENTO DE INFORMACIÓN DE PROTEÍNAS</u>	<u>20</u>
5.1 SERVIDOR EXPASY	20
5.2 SWISS 2D PAGE	20
5.3 MAKE 2-DB II	21
5.4 BASE DE DATOS LOCAL Y RAZE-2D	26
<u>6. RAZE-2D</u>	<u>28</u>
6.1 IMPLEMENTACIÓN CON JAVA	28
6.2 POSTGRESQL	29
6.3 NETBEANS	31
6.4 ALGORITMO DE TRADUCCIÓN DE LA ESCALA DE MASA MOLECULAR	32
6.5 ALGORITMO PARA EL CÁLCULO DEL PORCENTAJE DE SIMILITUD	33
6.6 ALGORITMO PARA EL CÁLCULO DEL PUNTO ISOELÉCTRICO Y LA MASA MOLECULAR DE UN 'SPOT' EN BASE A SU POSICIÓN EN LA IMAGEN.	38
6.7 MANUAL DE USO DE LA HERRAMIENTA	41
<u>7. METODOLOGÍA DE PROTOTIPADO EVOLUTIVO</u>	<u>58</u>
7.1 ESTUDIO INICIAL	59
7.2 GENERACIÓN DE LA BASE DE DATOS LOCAL	59
7.3 ELABORACIÓN DEL PRIMER PROTOTIPO	61
7.4 ELABORACIÓN DEL PROTOTIPO FINAL	62
7.5 PRUEBAS Y RESULTADOS	63
7.5.1 PRUEBAS A LA HERRAMIENTA Y LA BASE DE DATOS	63
7.5.2 PRUEBAS COMPARATIVAS CON OTRA HERRAMIENTA	65
7.5.3 ANALISIS DE LOS ERRORES Y CONCLUSIONES	65

8. CONCLUSIONES	68
9. RECOMENDACIONES	71
BIBLIOGRAFIA	72

LISTA DE FIGURAS

FIGURA 1. Efecto del SDS sobre la estructura de la proteína.	11
FIGURA 2. Imagen resultante de un proceso de electroforesis 2D. (Arabidopsis Thaliana).....	13
FIGURA 3. MAKE 2D-DB II	22
FIGURA 4. Archivos de texto plano como fuente para la creación de la base de datos.	24
FIGURA 5. Imagen de gel de referencia (Arabidopsis Thaliana) y su respectivo spreadsheet.	25
FIGURA 6. Reporte XML de Melanie Viewer.....	26
FIGURA 7. Tunel de comunicación Raze2D-Base de Datos	31
FIGURA 8. Distribución de la masa molecular y el punto isoeléctrico en un gel de electroforesis 2D.	32
FIGURA 9. Conversión de la escala de la masa molecular a una escala lineal.	33
FIGURA 10. Montaje universal de las proteínas presentes en la SWISS 2D-PAGE.....	34
FIGURA 11. Establecimiento del área de búsqueda a partir del punto base y los rangos.....	34
FIGURA 12. Puntos encontrados en el área de búsqueda.	35
FIGURA 13. Dpunto y Dmax a partir de las coordenadas del punto base P0 y de Pmax y P1.	36

FIGURA 14. Deducción de Dmax, Dpunto y el porcentaje de similitud.	37
FIGURA 15. Cálculo de la recta de interpolación por medio de mínimos cuadrados.....	39
FIGURA 16. Conversión de la escala de imagen real a imagen digital hecho por RAZE-2D.....	40
FIGURA 17. Visualización de la Herramienta Raze2D.....	41
FIGURA 18. Ingresando los datos iniciales del gel a analizar.....	44
FIGURA 19. Ingresando Marcador al Gel	45
FIGURA 20. Imagen que corresponde a un mapa (Arabidopsis) con algunos marcadores ya insertados aplicando la opción Negativo al gel	46
FIGURA 21. Ingresando los spots	47
FIGURA 22. Imagen de las proteínas encontradas según el criterio de búsqueda establecido.....	49
FIGURA 23. Archivo de texto el cual contiene la lista de los spots indicados por el usuario	51
FIGURA 24. Archivo de texto con los datos encontrados de las proteínas similares al spot señalado	52
FIGURA 25. Guardar información	53
FIGURA 26. Imagen del gel con sus respectivos marcadores y spots señalados por el usuario.....	54
FIGURA 27. Consultando en la Base de Datos Swiss2DPage	56
FIGURA 28. Consultando en la base de datos UniProt a través de la herramienta.....	57

FIGURA 29. Diagrama de casos de uso del Administrador	89
FIGURA 30. Diagrama de caso de uso consulta en la base de datos Usuario Investigador	90
FIGURA 31. Diagrama de caso de uso análisis del gel 2D de Usuario Investigador	91
FIGURA 32. Diagrama de Entidad-Relación de las tablas utilizadas de la Base de Datos Swiss2DPage.....	98
FIGURA 33. Instalando MAKE2D-DB II.....	104
FIGURA 34. Imagen del instalador de Java	107
FIGURA 35. Opciones de la instalación	108
FIGURA 36. Finalizando la instalación.....	109
FIGURA 37. Evaluación por parte de un experto	111
FIGURA 38. Evaluación de satisfacción de usuario final (PAGINA 1)	112
FIGURA 39. Evaluación de satisfacción de usuario final (PAGINA 2)	113
FIGURA 40. Evaluación de satisfacción de usuario final (PAGINA 3)	114

LISTA DE TABLAS

Tabla 1. Resultados de las pruebas de error relativo y absoluto aplicados a la herramienta RAZE-2D. 64

Tabla 2. Error encontrado entre Raze2D y PDQUEST Advanced 65

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE

Nombre del mapa analizado: ARABIDOPSIS THALIANA

Fecha : 22/12/2008

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	P37702	5,26	5,26	0	0
Masa Molecular(Mw[KDa])		62,957	68,57	8,915609067	5,613
Punto Isoeléctrico (pI)	O50008	6,15	6,22	1,138211382	0,07
Masa Molecular(Mw[KDa])		76,088	80,85	6,25854274	4,762
Punto Isoeléctrico (pI)	O03042	5,64	5,64	0	0
Masa Molecular(Mw[KDa])		51,074	51,12	0,090065395	0,046
Punto Isoeléctrico (pI)	P99055	5,13	5,13	0	0
Masa Molecular(Mw[KDa])		43,852	44,29	0,998814193	0,438
Punto Isoeléctrico (pI)	P10896	5,01	5,01	0	0
Masa Molecular(Mw[KDa])		47,034	47,21	0,374197389	0,176
Punto Isoeléctrico (pI)	P27140	5,71	5,71	0	0
Masa Molecular(Mw[KDa])		28,62	28,61	0,034940601	0,01

PROMEDIO

1,484198397

Ea [Pi]

0,011666667

Tabla 3. Error para el mapa ARABIDOPSIS THALIANA	75
Tabla 4. Error para el mapa DICTYSLUG	76
Tabla 5. Error para el mapa ECOLI	77
Tabla 6. Error para el mapa BAT-MOUSE	78
Tabla 7. Error para el mapa YEAST	79
Tabla 8. Error para el mapa HUMAN-KIDNEY	80
Tabla 9. Error para el mapa STAPHYLOCOCCUS	81
Tabla 10. Error para el mapa MUSCLE-MOUSE	82
Tabla 11. Error para el mapa LYMPHOCYTE-HUMAN	83
Tabla 12. Error para el mapa HUMAN-LIVER	84
Tabla 13. Error para el mapa EPIMASTIGOTES con PDQUEST ADVANCED	85
Tabla 14. Error para el mapa EPIMASTIGOTES 338 con PDQUEST ADVANCED	

ANEXO B. CASOS DE USO	86
Tabla 15. Spot de la Base de Datos Swiss2DPage .	95
Tabla 16. Spotentry de la Base de Datos Swiss2DPage .	96
Tabla 17. Entry de la Base de Datos Swiss2DPage	96
Tabla 18. Gel de la Base de Datos Swiss2DPage	97

LISTA DE ANEXOS

ANEXO A. TABLAS DE PRUEBAS DE ERROR DE RAZE-2D.	75
ANEXO B. CASOS DE USO	87
ANEXO C. GENERACIÓN DE LA BASE DE DATOS SWISS2DPAGE	92
ANEXO D. MANUAL DE INSTALACIÓN DE LA BASE DE DATOS CON MAKE 2D-DB II	99
ANEXO E. MANUAL DE INSTALACIÓN DE LA HERRAMIENTA RAZE-2D	106
ANEXO F. EVALUACIÓN DE LA SATISFACCIÓN DEL USUARIO	111

RESUMEN

Título: HERRAMIENTA SOFTWARE PARA EL ANÁLISIS E IDENTIFICACIÓN DE PROTEÍNAS MEDIANTE IMÁGENES DE ELECTROFORESIS EN 2D.*

Autores: Cristian Camilo Peña Álvarez, Luzby Stíwenson López Amado.**

Palabras Clave: Bases De Datos Biológicas, Bioinformática, Imágenes De Electroforesis 2D, Proteómica.

Descripción: La electroforesis en 2D, es un proceso experimental importante para la identificación de proteínas. Raze2D es una herramienta software capaz de realizar una interpretación de imágenes de electroforesis 2D, y hacer un análisis predictivo de las proteínas presentes en dicha imagen. Esto se hace a través de la generación de valores cuantificados de cercanía en los valores de punto isoeléctrico y masa molecular entre los 'spots' presentes en el gel, y comparados con los valores almacenados en una base de datos internacional (SWISS 2D-PAGE), que ha sido replicada dentro de las instalaciones de la Universidad Industrial de Santander. Esta herramienta, diseñada y generada en base a las necesidades particulares del Grupo de Investigación en Bioquímica e Ingeniería de Proteínas de la Universidad Industrial de Santander, se constituye como una herramienta de carácter educativo y libre; que permite apoyar el proceso de identificación de nuevas proteínas. El proyecto ha sido dividido en 3 fases: la primera fase, encaminada a la replicación de la base de datos SWISS 2D-PAGE; y siguiendo las especificaciones metodológicas planteadas bajo el modelo de prototipado evolutivo, las fases dos y tres corresponden al desarrollo de los prototipos de la aplicación, teniendo en cuenta los requerimientos generados con la experiencia de los expertos al usar la herramienta.

* Proyecto de Grado.

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática.
Director: MPE. Henry Argüello Fuentes. Codirector: Dr. Rodrigo Torres Sáez.

ABSTRACT

Title: SOFTWARE TOOL FOR ANALYSIS AND IDENTIFICATION OF PROTEIN FROM 2D ELECTROPHORESIS IMAGES

Authors: Cristian Camilo Peña Álvarez, Luzby Stíwenson López Amado.

Keywords: biological databases, Bioinformatics, 2D electrophoresis images, Proteomics.

Description: The 2D electrophoresis is an experimental process for the identification of proteins. Raze2D is a software tool capable of performing an interpretation of 2D electrophoresis images, and a predictive analysis of proteins present in the image. This is done through the generation of quantified values of closeness in the values of isoelectric point and molecular mass between the 'spots' in the gel, and compared with the values stored in an international database (SWISS 2D-PAGE), which has been replicated within the premises of the Industrial University of Santander. This tool, designed and produced based on the particular needs of the Grupo de Investigacion en Bioquímica e Ingeniería de Proteínas of the Universidad Industrial de Santander, was established as an educational and free-software tool, which support the process of identification of new proteins. The project has been divided into 3 phases: the first aimed at the replication of the database SWISS 2D-PAGE, and following the methodological set under the evolutionary prototyping model; the phases two and three relate to the prototypes application development, taking into account the requirements generated by the experience of experts when using the tool.

* Proyecto de Grado.

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática.
Director: MPE. Henry Argüello Fuentes. Codirector: Dr. Rodrigo Torres Sáez.

GLOSARIO

2D-PAGE: (PAGE: PolyAcrilamide Gel Electrophoresis) Electroforesis en gel de poliacrilamida en dos dimensiones.

BSD: Son las iniciales de Berkeley Software Distribution (en español, Distribución de Software Berkeley) y se utiliza para identificar un sistema operativo derivado del sistema Unix nacido a partir de los aportes realizados a ese sistema por la Universidad de California en Berkeley.

Bytecode: El bytecode es un código intermedio más abstracto que el código máquina. Los programas en bytecode son interpretados por la máquina virtual dado que es análogo a un ordenador. Su ventaja es su portabilidad.

GNU GPL: Licencia pública general de GNU. Es una licencia creada por la Free Software Foundation, orientada principalmente a proteger la libre distribución, modificación y uso de software.

Licencia BSD: Es la licencia de software otorgada principalmente para los sistemas BSD (Berkeley Software Distribution). Pertenece al grupo de licencias de software libre. Esta licencia tiene menos restricciones en comparación con otras como la GPL estando muy cercana al dominio público. La licencia BSD al contrario que la GPL permite el uso del código fuente en software no libre.

PH: Potencial de Hidrógeno. Es una medida de acidez o basicidad de una solución. Es la concentración de aniones y cationes de Hidrogeno presente en una sustancia.

Proteómica: Se define como el uso de técnicas para obtener medidas cuantitativas del nivel de proteínas para valorar la expresión génica a nivel celular.

Punto Isoeléctrico: Es el PH al que una sustancia tiene carga neta cero.

UNIX: Sistema operativo portable, multitarea y multiusuario.

INTRODUCCIÓN

Debido a la necesidad de organizar, analizar y distribuir información biológica, los biólogos y bioquímicos son los primeros en hacer un acercamiento a la tecnología computacional (bioinformática y biocomputación) como elemento fundamental para su trabajo diario.

La tecnología proporciona un elemento teórico y las herramientas prácticas, para que los científicos puedan explorar por ejemplo las proteínas y el DNA¹. Las proteínas son moléculas compuestas por numerosos residuos denominados aminoácidos. Durante los últimos 20 años se ha determinado que muchas proteínas de diverso origen con una función similar, también tienen secuencias similares de aminoácidos. Así, existen las secuencias correspondientes del DNA que son similares aunque la proteína bajo análisis ocurre en diversas especies tales como ratones y seres humanos. Así mismo, se han buscado diferencias y semejanzas en el nivel del DNA entre un ratón y un ser humano para muchas secuencias similares.

Desde el principio de los años 90, un gran número de 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 grandes de datos los cuales se recogen y se almacenan en bases de datos biológicas, la mayoría de las cuales son publicadas y accesibles como la SWISS 2D-PAGE².

¹ Ácido desoxirribonucleico, contiene la información genética usada en el desarrollo y el funcionamiento de los organismos vivos conocidos y de algunos virus, siendo el responsable de su transmisión hereditaria.

² SWISS 2D-Page contiene datos de proteínas identificadas en mapas de electroforesis en geles de poliacrilamida de 1D y 2D. <http://us.expasy.org/ch2d/>

Además de recopilar todos estos datos, es necesario comparar los hallazgos y determinar las semejanzas y las diferencias en cada individuo. Las proteínas, como cadenas de aminoácidos, son importantes ya que cumplen una labor en los seres vivos y son las biomoléculas más versátiles y más diversas. Realizan una enorme cantidad de funciones diferentes, entre las que destacan la estructural, la reguladora, transportadora, defensiva, enzimática o contráctil. Los Investigadores y estudiantes de bioquímica son conscientes de estos hechos y reconocen la importancia del estudio y la identificación de nuevas proteínas, a través de sus procesos experimentales (Electroforesis 2D, Espectrometría de masas, etc.).

El presente proyecto presenta una herramienta de cómputo educativa para la comunidad de la escuela de Química, capaz de realizar el análisis e identificación de proteínas generadas por la técnica de separación denominada Electroforesis 2D a través de las imágenes producidas en el laboratorio, y la búsqueda de semejanzas entre las proteínas generadas en estos procedimientos y las almacenadas en Bases de datos biológicas mundialmente reconocidas, a través de valores cuantificados.

Se presenta al lector el planteamiento del problema, junto con los objetivos del proyecto (Capítulos 1 y 2), para proseguir con la fundamentación teórica de la electroforesis (Capítulo 3), mostrado en paralelo con el desarrollo de la herramienta.

En el capítulo 4 se introduce el uso de las imágenes de electroforesis 2D y la interpretación de dichas imágenes con la herramienta Raze2D. De igual manera, se relaciona la herramienta con las bases de datos biológicas (Capítulo 5), exponiendo las más representativas para la comunidad científica internacional, y en los anexos C y D, se profundiza en la generación de la base de datos y en la instalación de la herramienta utilizada para tal fin.

El capítulo 6 se centra en las cualidades de la herramienta, así como en la presentación de los algoritmos implementados como pilar del funcionamiento de Raze2D y el manual de uso de la herramienta. El anexo E hace relación con el manual de instalación de la herramienta.

El capítulo 7 hace mención a la metodología utilizada en el desarrollo de la herramienta, las pruebas realizadas a la herramienta y la interpretación de los resultados obtenidos de dichas pruebas, que se han de presentar más explícitamente en el anexo A. El anexo B presenta los diagramas y la presentación de los casos de uso tenidos en cuenta en el desarrollo de Raze2D. De igual manera el anexo F muestra la evaluación de la satisfacción del usuario final de la herramienta, como muestra de la finalización del proceso de diseño y elaboración de la herramienta.

Por último, el capítulo 8 presenta las conclusiones generadas con el desarrollo del proyecto aquí presentado, y el capítulo 9 menciona las recomendaciones que se presentan al proyecto, destinadas a la continuidad del mismo y a futuras mejoras e implementaciones.

1. OBJETIVOS

1.1 OBJETIVO GENERAL

Diseñar y construir una herramienta computacional que apoye el proceso de identificación de proteínas mediante la interpretación y el análisis de imágenes de electroforesis 2D, y que responda a las necesidades particulares de la Escuela de Química de la Universidad Industrial de Santander.

1.2 OBJETIVOS ESPECÍFICOS

- Generar una base de datos que contenga la información de las proteínas según el estándar generado por el SIB (base de datos SWISS 2D-PAGE³), codificada en archivos de texto plano.
- Diseñar un algoritmo basado en imágenes de electroforesis 2D que permita establecer el porcentaje de similitud de las proteínas existentes en la base de datos (SWISS 2D-PAGE) con las encontradas en dichas imágenes.
- Diseñar y desarrollar un software de carácter educativo, libre y portable, capaz de interpretar imágenes de electroforesis 2D proporcionadas por el usuario, y generar un valor que permita diferenciar y cuantificar la cercanía de los valores estimados a los valores encontrados de las

³ Swiss 2D-Page contiene datos de proteínas identificadas en mapas de electroforesis en geles de poliacrilamida de 1D y 2D. <http://us.expasy.org/ch2d/>

proteínas existentes en la base de datos (SWISS 2D-PAGE) a través de un algoritmo y de esta manera constituirse como una herramienta efectiva para la identificación de proteínas nuevas.

- Verificar el funcionamiento de la herramienta desarrollada mediante la comprobación de la precisión en los datos encontrados, utilizando como fuente los datos existentes en la base de datos (SWISS 2D-PAGE) y con la supervisión de expertos de la Escuela de Química Pura de la Universidad Industrial De Santander.

2. PLANTEAMIENTO DEL PROBLEMA

2.1. ANTECEDENTES

El Instituto Suizo de Bioinformática (SIB, por sus siglas en inglés), es una fundación académica y sin ánimo de lucro, establecida el 30 de Marzo de 1998, cuya misión es promover la investigación, el desarrollo de bancos de datos y tecnologías computacionales, y estar involucrada con actividades de enseñanza y servicios en el área de la bioinformática en Suiza, con colaboraciones internacionales [www.wikipedia.org].

Este instituto mantiene numerosas bases de datos biológicas⁴, entre las que se encuentran la SWISS-PROT, el servidor Expasy que está dedicado al estudio de secuencias y estructuras de proteínas, y la SWISS 2D-PAGE, utilizadas ampliamente alrededor del mundo por investigadores del área de bioquímica.

El grupo de Investigación en Ingeniería Biomédica – GIIB de la UIS, fundado en 1992, ha presentado una larga trayectoria en la generación de conocimiento mediante investigaciones relacionadas con la Ingeniería Biomédica, aunque ha incrementado sus esfuerzos en el desarrollo de investigaciones en otras áreas de conocimiento, como son el tratamiento digital de imágenes, señales, redes computacionales, supercomputación, y desarrollo de herramientas para dispositivos móviles.

⁴ Swiss-Prot, Prosite, EPD, Swiss-2Dpage, Human Chromosome 21, TrEST, TrGen, AGBD, Hits, Swiss Model Repository, GermOnline

Fruto de este esfuerzo, aparecen trabajos interdisciplinarios entre este grupo de investigación y otros grupos reconocidos, tales como el grupo de Investigación en Bioquímica e Ingeniería de Proteínas, dirigido por el Dr. Rodrigo Torres Sáez, reconocido investigador en este campo, quien ha manifestado su interés en el estudio de enfermedades tropicales, como la leishmaniasis, y el estudio de proteínas relacionadas con dichas enfermedades.

Para la identificación de proteínas, se han desarrollado en el mundo algunas herramientas computacionales capaces de hacer interpretaciones de imágenes de electroforesis, que se limitan únicamente a las imágenes sin establecer relaciones entre los datos hallados y los almacenados en las bases de datos biológicas, además, las licencias de uso de este tipo de software son costosas.

En la actualidad no se conoce de ninguna herramienta capaz de hacer el proceso de interpretación de las imágenes, integrado con la consulta a las bases de datos para identificar proteínas nuevas, y mucho menos la generación de valores cuantificados para medir la cercanía de las proteínas encontradas en el análisis con las proteínas identificadas en dichas imágenes.

2.2. JUSTIFICACIÓN

El SIB (Swiss Institute of Bioinformatics) a través de sus nacientes estándares establece los mecanismos para el almacenamiento de la información de las proteínas en una base de datos a partir de archivos de texto plano (SWISS 2D-PAGE). La identificación de nuevas proteínas se hace en primera instancia a través del estudio de las imágenes generadas en los procesos de electroforesis, y las herramientas computacionales que permiten hacer este tipo de procesos son de carácter privativo y tienen unos costos elevados por sus licencias de uso.

Se reconoce la necesidad de una herramienta de carácter educativo, libre y personalizado, que permita a los estudiantes e investigadores de la escuela de Química de la Universidad Industrial de Santander el análisis de las imágenes producidas en los laboratorios, y así identificar las proteínas encontradas en dichos procesos, al igual que el análisis de los resultados obtenidos teniendo como punto de referencia la información de la base de datos(SWISS 2D-PAGE) para así generar un porcentaje de similitud, y poder identificar si corresponden a proteínas nuevas.

3. ELECTROFORESIS

3.1. ELECTROFORESIS DE PROTEÍNAS EN GELES DE POLIACRILAMIDA.

La **electroforesis** es una técnica para la separación de mezclas complejas (proteínas o ácidos nucleicos) según la movilidad de estas en un campo eléctrico a través de una matriz porosa, la cual finalmente las separa por tamaños moleculares y carga eléctrica, dependiendo de la técnica que se use. Los ácidos nucleicos ya disponen de una carga eléctrica negativa, que los dirigirá al polo positivo, mientras que las proteínas se cargan con sustancias como el SDS (detergente) que incorpora cargas negativas de una manera dependiente del peso molecular. Para la separación se usa un gel de agarosa o poliacrilamida (fibras cruzadas, como una malla). Al poner la mezcla de moléculas y aplicar un campo eléctrico, éstas se moverán y deberán ir pasando por la malla, por la que las pequeñas se moverán mejor, más rápidamente. Así, las más pequeñas avanzarán más y las más grandes quedarán cerca del lugar de partida

La electroforesis de proteínas en geles de poliacrilamida, comúnmente denominada electroforesis en poliacrilamida (PAGE, 'polyacrilamide gel electrophoresis') es sin duda alguna una de las técnicas más ampliamente usadas para caracterizar mezclas complejas de proteínas. Es un método conveniente, rápido y económico a nivel de muestra pues se requieren sólo cantidades del orden de microgramos de proteína. La separación de cadenas de aminoácidos es una de las muchas aplicaciones de la electroforesis.

Existen dos tipos de electroforesis: SDS-PAGE y electroforesis 2D (Bidimensional).

3.2. SDS-PAGE

SDS es un detergente de acción desnaturalizante⁵ (ver FIGURA 1), que se une a las cadenas de aminoácidos que forman la proteína desnaturalizadas con una relación de 1.4 g de SDS por gramo de proteína, uniéndose aproximadamente una molécula de SDS por cada dos aminoácidos de la cadena. Esta unión masiva de moléculas de SDS bloquea la carga propia de la molécula de proteína y le confiere al complejo una carga neta negativa proporcional a su masa, haciendo que todas las proteínas acomplejadas con SDS viajen hacia el ánodo. La separación de los complejos SDS-proteína es proporcional sólo a la masa de la proteína pues todas tienen la misma carga por unidad de masa. Se puede entonces determinar el peso molecular aparente de cualquier proteína por comparación con un patrón de proteínas de pesos moleculares conocidos. Las movibilidades de las proteínas en los geles de SDS-PAGE son funciones lineales del logaritmo de su peso molecular.

SDS-PAGE es una técnica molecular de biología que se usa para separar las proteínas según su masa molecular. Las proteínas que se analizarán primero se mezclan con un SDS (Detergente) que separa la estructura secundaria de la proteína y aplica una carga negativa a cada proteína en proporción a su masa molecular. Sin el detergente (SDS), diversas proteínas con los pesos moleculares similares no emigrarían igual. La adición del SDS soluciona este problema y revela la proteína, dando una carga negativa uniforme cercana a lo largo de la longitud del polipéptido⁶.

⁵ Desnaturalizante: que separa la estructura terciaria de la proteína resultando en cadenas de aminoácidos (estructura secundaria).

⁶ Polipéptido son un tipo de moléculas formadas por cadenas lineales de aminoácidos.

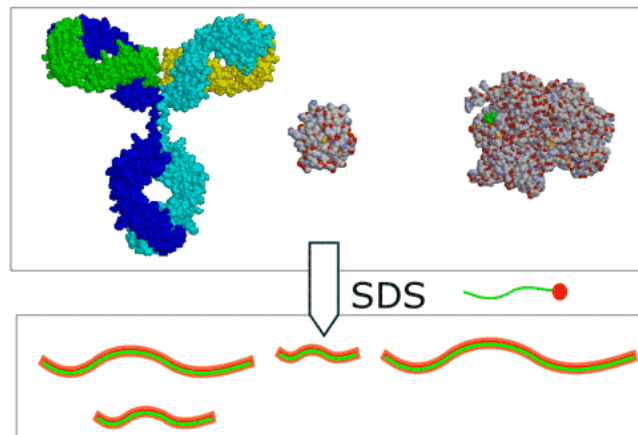


FIGURA 1. Efecto del SDS sobre la estructura de la proteína.

3.3. ELECTROFORESIS BIDIMENSIONAL

La electroforesis bidimensional (2D) se basa en separar las proteínas de una mezcla según dos propiedades moleculares, una en cada dimensión. El procedimiento más usado se basa en la separación en una primera dimensión mediante isoelectroenfoque (según su punto isoeléctrico) y una segunda dimensión según el peso molecular mediante electroforesis en poliacrilamida-SDS.

Este procedimiento se lleva a cabo fase por fase en geles porosos especiales con características muy particulares, que permiten la movilidad de las sustancias a través del gel, mediante la aplicación del SDS o el campo eléctrico según sea el caso.

3.3.1. PUNTO ISOELÉCTRICO

El punto isoeléctrico es el pH al que una sustancia anfótera⁷ (en este caso, proteínas) tiene carga neta cero. A este valor de pH la solubilidad de la sustancia es casi nula.

Las moléculas complejas, tales como las proteínas, se combinan con los iones hidrógeno y con otros iones presentes en la disolución, dando lugar a la carga neta de la molécula. A la concentración de iones hidrógeno, o al pH, para el cual la concentración del ion híbrido de una proteína es máxima y el movimiento neto de las moléculas de soluto en un campo eléctrico es prácticamente nulo, se le denomina punto isoeléctrico.

El isoelectroenfoque es una técnica que se basa en el desplazamiento de las moléculas en un gradiente de pH. Las moléculas como los aminoácidos, se separan en un medio en el que existe una diferencia de potencial y un gradiente de pH. La región del ánodo (+) es ácida y la del cátodo (-) es alcalina. Entre ambos se establece un gradiente de pH tal que las moléculas que se han de separar tengan su punto isoeléctrico (pI)⁸ dentro del rango. Las sustancias que se encuentran inicialmente en regiones de pH inferior a su punto isoeléctrico estarán cargadas positivamente y migrarán hacia el cátodo, mientras que aquellas que se encuentren en medios con pH más bajos que su pI tendrán carga negativa y migrarán hacia el ánodo. La migración les conducirá a una región donde el pH coincidirá con su pI, tendrán una carga

⁷ El anfoterismo es el término general que describe la habilidad de una sustancia para reaccionar como ácido o como base, según la sustancia con que reaccione.

⁸ El punto isoeléctrico, es el valor de pH en el cual las moléculas no presentan migraciones en un campo eléctrico y se determina experimentalmente por electroforesis.

neta nula y se detendrán. De esta forma las moléculas se sitúan en estrechas bandas donde coincide su pI con el pH.

El valor del punto isoeléctrico es uno de los puntos fundamentales en el estudio de proteínas porque cada proteína tiene un punto isoeléctrico característico, que es calculable por medio de algoritmos que trabajan en base a su secuencia de aminoácidos o experimentalmente considerando su posición en una imagen de electroforesis. Debido a esto, y para el caso de RAZE-2D, es uno de los pilares de la interpretación de las imágenes de electroforesis.

3.3.2. MASA MOLECULAR

El SDS-PAGE es el proceso que se utiliza para la separación de mezclas complejas (proteínas) en un gel. La separación se hace precisamente a través de la masa molecular de cada una de sus proteínas.

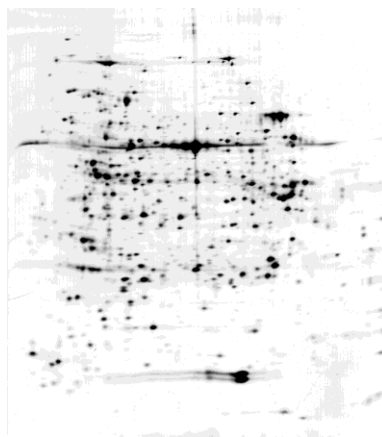


FIGURA 2. Imagen resultante de un proceso de electroforesis 2D. (Arabidopsis Thaliana)

La masa molecular, identificable por la posición en una imagen de electroforesis (ver FIGURA 2), o a través de la espectrometría de masas. Cada proteína tiene un valor específico de masa molecular y es uno de los datos más relevantes en el estudio de las proteínas, que se encuentra almacenado en muchas bases de datos biológicas. SWISS 2D PAGE es la base de datos que utiliza la herramienta para la comparación de los 'spots' (manchas) encontradas en las imágenes con las proteínas almacenadas en dicha base de datos.

3.4 ELECTROFORESIS Y RAZE-2D

Al tener las proteínas en el gel ya separadas en 2D, se obtiene una serie de manchas ('spots') en la imagen, que corresponden a las proteínas presentes en la mezcla y se prosigue con la determinación del punto isoelectrico y la masa molecular respectiva de cada molécula.

El cálculo de estos valores se puede hacer manualmente, pero las herramientas informáticas facilitan en gran medida este trabajo. Existen productos software de carácter comercial que son capaces del cálculo del punto isoelectrico y la masa molecular de un compuesto representado por un spot en el gel, pero desafortunadamente son herramientas especializadas en sólo este tipo de tareas⁹, sus licencias de uso son costosas y restringen el formato de las imágenes a utilizar. Uno de los mayores problemas de la 2D PAGE es el análisis y comparación de mezclas muy complejas de proteínas.

⁹ PDQUEST Advanced. Más información en: http://www.genmall.com.tw/products/info/Bulletin_3121%20PDQuest%202D%20Analysis%20Software%20Brochure%20Rev%20B.pdf

RAZE-2D es una herramienta diseñada y creada para apoyar la investigación y la formación de estudiantes de bioquímica en el estudio de proteínas y permite hacer el cálculo de los valores de punto isoelectrico y masa molecular de los 'spots' presentes en las imágenes suministradas, necesarios en el análisis cuantitativo, a través de algoritmos que traducen las coordenadas y calculan la distribución del pH y la masa molecular en el gel. Este tipo de herramientas contribuyen con el desarrollo de actividades investigativas y de formación académica para la generación de nuevos profesionales y la posibilidad de producir nuevos conocimientos en esta área de estudio.

4. IMÁGENES DE ELECTROFORESIS

4.1. MASA MOLECULAR Y SU ESCALA EN LAS IMÁGENES

Las imágenes de electroforesis, contienen información acerca de las proteínas presentes en el gel. Uno de los datos más importantes es la masa molecular. Sin embargo, el cálculo de éste a partir de las imágenes de electroforesis depende de la distribución en el gradiente de masas moleculares específico para el gel. La movilidad electroforética (distancia movida por la proteína en el gel durante el proceso) está relacionada linealmente con el logaritmo de la masa molecular, y en este sentido es posible interpolar¹⁰ estos valores a partir de una serie de valores conocidos, para determinar experimentalmente el valor de la masa molecular de una proteína en particular. En el caso de las imágenes digitales, el fenómeno es el mismo, pero teniendo en cuenta que la movilidad electroforética está dada por pixeles y no por las unidades de longitud conocidas.

4.2. PUNTO ISOELÉCTRICO EN LAS IMÁGENES

El punto isoeléctrico de una proteína se puede calcular en base a la ubicación de su 'spot' en el gel. Es importante saber que el efecto del campo eléctrico en los geles de poliacrilamida, o incluso la misma disposición del gel hace que la escala lineal que se maneja comúnmente, pueda perderse. RAZE-2D utiliza un algoritmo para solucionar estos inconvenientes.

¹⁰ se denomina interpolación a la construcción de nuevos puntos partiendo del conocimiento de un conjunto discreto de puntos.

SWISS 2D PAGE, la base de datos relacional utilizada por la herramienta, tiene como datos principales en sus entradas de registros el punto isoeléctrico y la masa molecular, que en últimas se traducen en la identidad propia de la proteína, junto con otros datos que hacen relación con el nombre del gel al que pertenecen y el número de acceso como identificador principal en la base de datos.

4.3. INTERPRETACIÓN DE IMÁGENES CON RAZE-2D

Las imágenes de electroforesis 2D (ver FIGURA 2), tienen características muy particulares, que permiten la separación de mezclas complejas. Estas características van desde su eje x que representa la distribución lineal del pH, el eje y que corresponde a la distribución exponencial del eje de masa molecular; hasta el hecho de permitir la movilidad de las muestras generándose unas manchas (spots) de las que se puede obtener su valor correspondiente de punto isoeléctrico y masa molecular simplemente con determinar su ubicación en el gel.

RAZE-2D posee gran versatilidad en cuanto al formato de las imágenes a trabajar, recibiendo imágenes en formato .bmp, .gif, .jpg, .png, entre otros. Esto permite que las imágenes suministradas por los usuarios provengan de diferentes fuentes sin que ello resulte en una disminución de la precisión en los datos calculados. Las imágenes digitales de los geles de electroforesis presentan las mismas cualidades que las imágenes reales del proceso, pero con la diferencia en que las distancias no están dadas en centímetros o milímetros, sino en los pixeles que componen dicha imagen. En este sentido, RAZE-2D se encarga de la conversión de las coordenadas para hacer los

cálculos de los valores de punto isoeléctrico y masa molecular de una manera intuitiva y amigable al usuario.

El usuario ingresa a la herramienta la imagen del gel en el formato que desee, junto con los datos iniciales del gel, como el punto isoeléctrico inicial y final, la masa molecular inicial y final y la cantidad de marcadores¹¹ a usar. Después de esto, se procede a definir los marcadores y con estos datos la herramienta está en capacidad de calcular el punto isoeléctrico y la masa molecular en cualquier ubicación del gel que el usuario escoja mediante un clic del ratón.

Para una mayor facilidad en la identificación de los spots, la herramienta posee funcionalidades para el filtrado de las imágenes. Entre los filtros que posee Raze2D están filtro paso-bajo, filtro paso-alto, filtro mediana y negativo. Esto favorece al usuario la identificación de los spots ya que hace posible el suavizado de la imagen, la eliminación de ruido, e incluso convertir al negativo de la imagen.

La herramienta guarda la posición seleccionada y los datos calculados del punto isoeléctrico y masa molecular, ofreciendo la opción de hacer el análisis predictivo de las proteínas que más se acerquen a las características calculadas (pI y Mw). Esto se hace a través de la selección de unos rangos que el usuario a libre voluntad puede elegir incluyendo el mapa (gel de electroforesis) con el cual comparar el 'spot' seleccionado.

A continuación la herramienta genera una lista de las proteínas que cumplen con el criterio de búsqueda y genera unos valores cuantificados que permiten

¹¹ Los marcadores son proteínas específicas que tienen un punto isoeléctrico y masa molecular conocidos, y que se añaden a la muestra para facilitar los cálculos del punto isoeléctrico y la masa molecular de las demás proteínas presentes en el gel.

calificar la cercanía de las proteínas encontradas en la base de datos, o en el mapa seleccionado, a la proteína (spot) escogida en la imagen.

Junto a esta lista, y si el usuario desea obtener más información sobre la proteína, con un doble clic se dirige al sitio web de la base de datos SWISS 2D PAGE donde aparece información más detallada de la proteína escogida.

La herramienta muestra como dato adicional en su interfaz gráfica información más completa, como la proteína más probable, su punto isoelectrico, masa molecular y porcentaje de similitud. Todo esto luego de haber sido analizado el spot y hecha la consulta en la base de datos.

5. BASES DE DATOS BIOLÓGICAS PARA EL ALMACENAMIENTO DE INFORMACIÓN DE PROTEÍNAS

5.1 SERVIDOR EXPASY

El servidor Expasy (**Expert Protein Analysis System**), es un servidor de Proteómica del Instituto Suizo de Bioinformática, que se encarga del análisis de secuencias y estructuras de proteínas y 2D PAGE, y funciona desde el 1° de Agosto de 1993. En su sitio web: www.expasy.org se encuentran los links hacia los diferentes sitios web de las bases de datos biológicas que mantiene, entre ellas SWISS 2D PAGE, y ofrece numerosas alternativas de consulta a la base de datos, y se ha convertido en una gran herramienta para bioquímicos y estudiantes en todo el mundo desde entonces. Además de la SWISS 2D PAGE, y en asociación con otras entidades, mantiene la base de conocimiento sobre proteínas y estructuras de las mismas: UniProt (de *Universal Protein*) que es el recurso de proteínas universal, un repositorio central de datos sobre proteínas. Esto lo ha convertido en el recurso líder mundial almacenando información sobre proteínas.

5.2 SWISS 2D PAGE

La base de datos SWISS-2DPAGE, creada en 1993, es mantenida en colaboración por la Central de Laboratorio de Química Clínica del Hospital Universitario de Ginebra y el Instituto Suizo de Bioinformática (SIB). La versión actual contiene alrededor de 4300 entradas en 35 mapas (geles) de referencia a partir de muestras humanas, muestras de ratón, *Arabidopsis thaliana*, *Dictyostelium discoideum*, *Escherichia coli*, *Saccharomyces cerevisiae* y

*Staphylococcus aureus*¹². A través del servidor ExPASy, la base de datos SWISS-2DPAGE puede ser consultada por palabras clave (descripción de proteínas, nombre de proteínas, nombre del gel, especies, autor, etc.) o gráficamente, haciendo clic en un punto. También es posible calcular la estimación de la ubicación de un punto a partir de la secuencia de la proteína de la SWISS-PROT.

La base de datos SWISS 2D_PAGE es la utilizada por la herramienta ya que cuenta con datos de proteínas identificadas y extraídas de procesos de electroforesis 2D, identificando en ellas el punto isoeléctrico y la masa molecular, vitales para el análisis cuantitativo de las imágenes.

5.3 MAKE 2-DB II

Make2D-DB II es una herramienta para crear, convertir, publicar, interconectar y mantener actualizadas las bases de datos de electroforesis 2D. Convierte los datos de varios formatos en un formato relacional. La herramienta ofrece la posibilidad de actualizar automáticamente los datos relacionados con las numerosas fuentes de datos externos en una forma muy coherente. También es posible con esta herramienta interconectar dinámicamente varias bases de datos remotas o proyectos para formar una base de datos virtual mundial accesible desde un único punto de entrada.

¹² Swiss 2dPage ten years later:

<http://www3.interscience.wiley.com/journal/109083909/abstract?CRETRY=1&SRETRY=0>

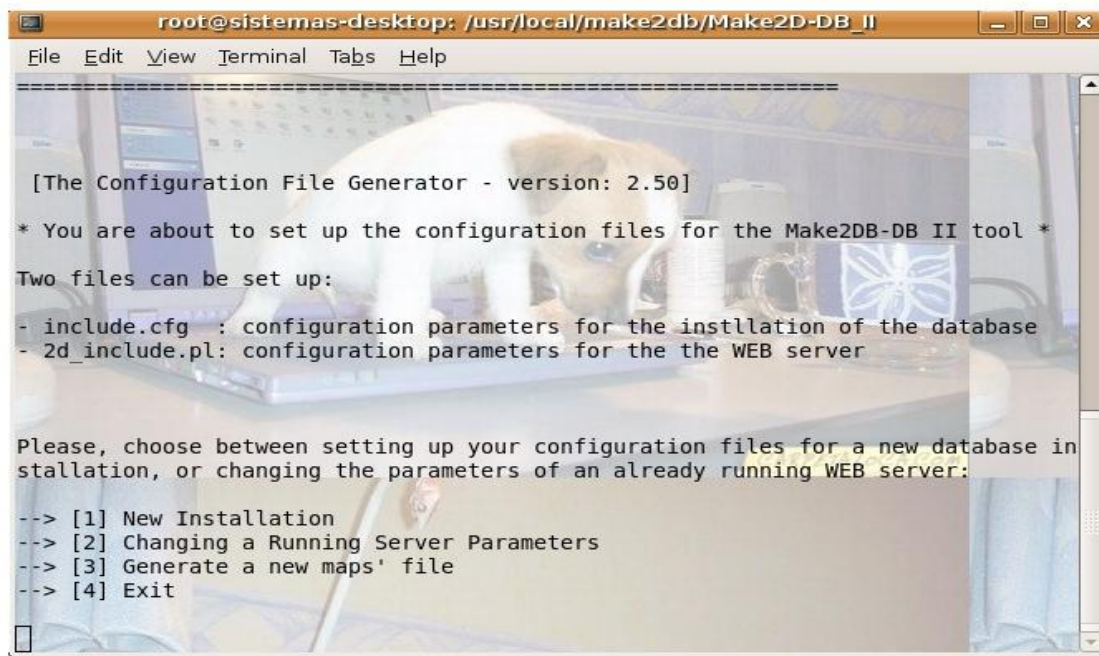


FIGURA 3. MAKE 2D-DB II

El manejo del MAKE 2D-DB II (ver FIGURA 3) es muy intuitivo, gracias a su interfaz donde todas las opciones y procedimientos que la herramienta puede realizar están bien especificados; se ejecuta desde una consola de terminal de Linux.

La herramienta es open-source (de código abierto) y totalmente libre. Maneja reportes XML hechos por el software Melanie/ImageMaster 2D Platinum¹³, hojas de cálculo (por ejemplo Excel / CSV informes), así como listas de texto simple. También es posible crear nuevas bases de datos relacionales a partir

¹³ Software privativo que identifica proteínas de geles de 2D capaz de exportar los resultados en archivos XML. <http://www.expasy.org/melanie/>

de cero. Se ejecuta en la mayoría de sistemas operativos basados en UNIX (Linux, Solaris / SunOS, IRIX).

Make2D-DB II está diseñado para garantizar un alto nivel de coherencia de los datos. Permite la interconexión dinámica entre cualquier número similar de bases de datos remotas y ofrece muchas otras características, incluidas las actualizaciones automáticas de datos referenciados a bases de datos externas, motor de búsqueda y visualización de datos combinado con las exportaciones en diversos formatos.

Gracias a esta herramienta se puede generar localmente una copia de la base de datos SWISS2D-PAGE la cual contiene información de proteínas existentes e identificadas de diferentes especies con la que se realizará el análisis respectivo con los procesos de electroforesis desarrollados por la escuela de química para así identificar las proteínas y determinar su similitud.

MAKE 2D-DB II presenta 3 opciones para la generación de las bases de datos:

- A partir de los archivos de texto plano (ver FIGURA 4), codificados con etiquetas en cada línea donde se especifican uno a uno los registros que han de ir almacenados en la base de datos, y las imágenes de los geles donde se encuentren las proteínas citadas en dichos registros.

```

1 SWISS-2DPAGE Release 18.0
2 September 2008
3
4 ID 1433S_HUMAN; STANDARD; 2DG.
5 AC P31947;
6 DT 15-JAN-1999, integrated into SWISS-2DPAGE (release 9).
7 DT 15-MAY-2003, 2D annotation version 1.
8 DT 26-SEP-2006, general annotation version 7.
9 DE 14-3-3 protein sigma (Stratifin) (Epithelial cell marker protein 1).
10 GN Name=SFN; Synonyms=HME1;
11 OS Homo sapiens (Human).
12 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
13 OC Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
14 OC Catarrhini; Hominidae; Homo.
15 OX NCBI_TaxID=9606;
16 MT DLD1_HUMAN.
17 IM DLD1_HUMAN.
18 RN [1]
19 RP MAPPING ON GEL.
20 RA Demalte-Annessi I., Sanchez J.-C., Hoogland C., Rouge V., Binz P.-A.,
21 RA Appel R.D., Hochstrasser D.F.;
22 RL Submitted (JAN-1999) to the SWISS-2DPAGE database.
23 ZD -!- MASTER: DLD1_HUMAN;
24 ZD -!- PI/Mw: SPOT 2D-001EP6=4.60/26759;
25 ZD -!- PEPTIDE MASSES: SPOT 2D-001EP6: 1008.798; 1070.551; 1205.655;
26 ZD 1546.668; 2189.874; 2833.435; TRYPSIN.
27 ZD -!- MAPPING: Peptide mass fingerprinting [1].
28 CC -----
29 CC This SWISS-2DPAGE entry is copyright the Swiss Institute of Bioinformatics.
30 CC There are no restrictions on its use by non-profit institutions as long as
31 CC its content is in no way modified and this statement is not removed.
32 CC Usage by and for commercial entities requires a license agreement (See
33 CC http://www.expasy.org/ch2d/license.html or send email to license@isb-sib.ch).
34 CC -----
35 DR Cornea-2DPAGE; P31947; 1433S_HUMAN.
36 DR OGP-www; P31947.
37 DR UniProtKB/Swiss-Prot; P31947; 1433S_HUMAN.
38 //
39 ID 3HIDH_HUMAN; PRELIMINARY; 2DG.
40 AC P31937;
41 DT 01-AUG-1993, integrated into SWISS-2DPAGE (release 0).

```

FIGURA 4. Archivos de texto plano como fuente para la creación de la base de datos.

- A través de spreadsheets (archivos de texto tabulados, reportes de Excel, ver FIGURA 5), y las imágenes relacionadas con las proteínas citadas. Es importante que cada imagen se encuentre con su respectivo spreadsheet donde haga la relación entre la ubicación de cada spot en dicha imagen y sus datos experimentales a almacenar.

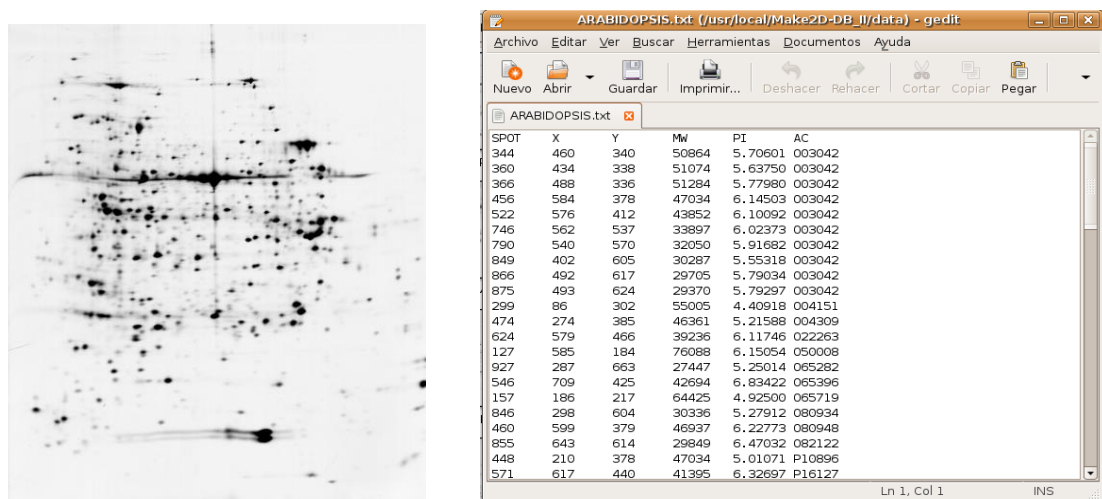


FIGURA 5. Imagen de gel de referencia (*Arabidopsis Thaliana*) y su respectivo spreadsheet.

- Como tercera opción, están los reportes XML generados por el Melanie Viewer o el Imagemaster 2D Platinum (Herramienta privativa del Instituto Suizo de Bioinformática, ver FIGURA 6).

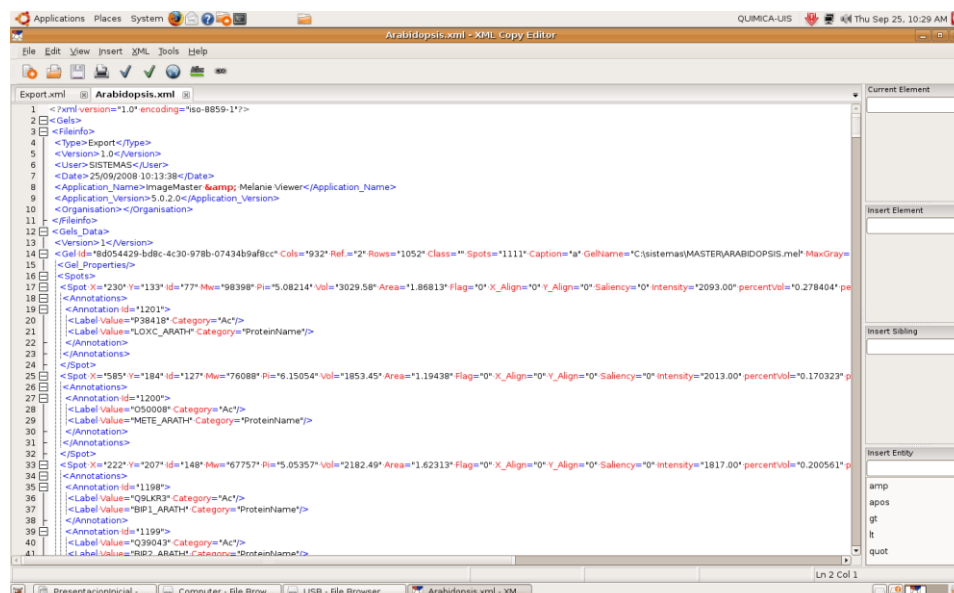


FIGURA 6. Reporte XML de Melanie Viewer.

Para el montaje de la base de datos local se utilizaron spreadsheets (archivos de texto tabulados) obtenidos como exportaciones a partir de la herramienta Melanie Viewer 5.02, y los archivos .mel, fuentes de la base de datos original SWISS 2D-PAGE, de donde se tomaron también las imágenes de los mapas de referencia.

5.4 BASE DE DATOS LOCAL Y RAZE-2D

Hasta este momento, se ha hecho énfasis en el uso de las bases de datos biológicas para el estudio de proteínas existentes y nuevas. Para RAZE-2D, no es la excepción, y por tanto se hace necesaria la selección de una, o varias bases de datos como referencia para la consulta y el análisis de las imágenes. La pregunta ahora es ¿cuál es la base de datos a utilizar? Uniprot es la base

de datos mundial con la información de proteínas utilizada ampliamente por los investigadores y bioquímicos en la comunidad científica y la academia. Sin embargo, no es la base de datos escogida para el análisis de las imágenes, debido a que en muchas de sus entradas, no se tiene un valor exacto identificado de punto isoeléctrico para las proteínas, a diferencia de la SWISS 2D PAGE, que por ser basada en los resultados de procesos de electroforesis 2D sí los tiene.

Gracias al uso de herramientas open-source como MAKE 2-DB II, es posible la generación de bases de datos relacionales a partir de archivos de texto plano, e imágenes de electroforesis; y tomando como punto de partida el hecho de que SWISS 2D PAGE es una base de datos libre, se hizo la replicación de esta base de datos de una forma consistente, y con esto mejorar los tiempos de respuesta en las consultas necesarias a dicha base de datos.

Para la Escuela de Química de la Universidad Industrial de Santander, quienes son los usuarios finales a los que está dirigida la herramienta, es importante el almacenamiento de la información recopilada a través del estudio de proteínas mediante electroforesis 2D, es muy importante la generación de esta base de datos local, dada su expreso interés en el estudio de enfermedades tropicales, como la leishmaniasis, y el efecto de fármacos en el tratamiento de las mismas.

6. RAZE-2D

6.1 IMPLEMENTACIÓN CON JAVA

Java es un lenguaje de programación orientado a objetos que nació alrededor de los años 90 y fue creado por la empresa Sun Microsystems¹⁴. El lenguaje mismo se inspira en la sintaxis de C++, incorpora sincronización y manejo de tareas en el lenguaje mismo (similar a Ada¹⁵) e incorpora interfaces como un mecanismo alternativo a la herencia múltiple de C++ pero elimina herramientas de bajo nivel, que suelen inducir a muchos errores, como la manipulación directa de punteros a memoria.

Una de las razones primordiales para la selección de este lenguaje en particular para el desarrollo de la herramienta, es su capacidad de permitir ejecutar las aplicaciones en diferentes plataformas existentes en el mercado. Las aplicaciones java están compiladas en bytecode que son interpretadas por la máquina virtual o JRE, gracias a esto permite la portabilidad en ejecución. Es una ventaja significativa para los desarrolladores de software, pues antes tenían que hacer un programa para cada sistema operativo, por ejemplo Windows, Linux, Mac, etc. Esto lo consigue porque se ha creado una Máquina Virtual de Java para cada sistema que hace de puente entre el sistema operativo y el programa de Java y posibilita que este último se entienda perfectamente.

¹⁴ Historia del lenguaje Java:
http://www.cad.com.mx/historia_del_lenguaje_java.htm

¹⁵ Ada es un lenguaje multipropósito, orientado a objetos y concurrente, pudiendo llegar desde la facilidad de Pascal hasta la flexibilidad de C++.

Alrededor del 2007 Sun Microsystems liberó la mayor parte de las tecnologías java bajo la licencia GNU GPL de tal forma que todo el software de Sun es ahora software libre, lo que posibilita el desarrollo de software sin recurrir al pago de licencias de uso por el trabajo en este lenguaje de programación, facilitando el mismo desarrollo del proyecto.

La tercera razón que justifica la escogencia de este lenguaje, es la posibilidad del uso de librerías especializadas para el tratamiento de imágenes, lo que posibilitó la inclusión de funcionalidades a la herramienta, tales como el filtrado en las imágenes, la conversión a escala de grises y la conversión a negativo. Todo esto para facilitar el trabajo del usuario final en la identificación de los spots y de esta manera, hacer de la herramienta un producto aún más amigable, porque la calidad de las imágenes puede variar dependiendo del dispositivo de adquisición de dichas imágenes, que puede ser un escáner de oficina o un instrumento más especializado.

6.2 POSTGRESQL

Es un sistema de gestión de base de datos relacional orientado a objetos. Producido y licenciado como software libre, es la mejor opción para el desarrollo de la herramienta, dadas sus características:

Alta Ocurrencia: Permite que mientras un proceso acceda a la tabla, otro pueda hacer operaciones sobre dicha tabla sin ningún tipo de bloqueo.

Alta variedad de tipos nativos: PostgreSQL provee nativamente soporte para números de precisión arbitraria, texto de largo ilimitado, figuras geométricas (con una variedad de funciones asociadas), direcciones IP (IPv4 e IPv6), bloques de direcciones estilo CIDR, direcciones MAC y arrays.

Claves ajenas denominadas también claves foráneas.

Vistas.

Herencia de tablas.

Integridad transaccional.

Tipos de datos y operaciones geométricas.

Disipadores (Triggers).

Instalación ilimitada: Con PostgreSQL, no hay costo asociado a la licencia del software, permitiendo la instalación en un número ilimitado de servidores.

Extendible: El código fuente está disponible para todos sin costo, esto es complementado por la comunidad de profesionales y entusiastas de PostgreSQL alrededor del mundo que también extienden PostgreSQL todos los días.

Multiplataforma: PostgreSQL está disponible en casi cualquier Unix (34 plataformas en la última versión estable) y plataformas Windows.

Todas estas cualidades se traducen en la implantación de una base de datos robusta, con una estructura relacional muy coherente, montada en un servidor Linux, con políticas de seguridad definidas. Sin embargo, fruto de las mismas políticas de seguridad del servidor en el que fue colocada la base de datos (Servidor de ciencias de la UIS), se restringe el acceso al servidor a través de la mayoría de puertos. Lo que hizo necesario la utilización de un “túnel” (ver FIGURA 7) mediante una página php que se encargara de hacer las consultas a la base de datos y de esta manera transmitir las consultas por medio del puerto http (puerto 80), a través de la navegación en internet en donde se envía la consulta a realizar utilizando de un mecanismo de encriptación para garantizar la seguridad, y la pagina php la desencripta y hace la consulta.

Cuando se tenga el resultado, éste se encripta nuevamente y se envía al programa para que este se los muestre al usuario.

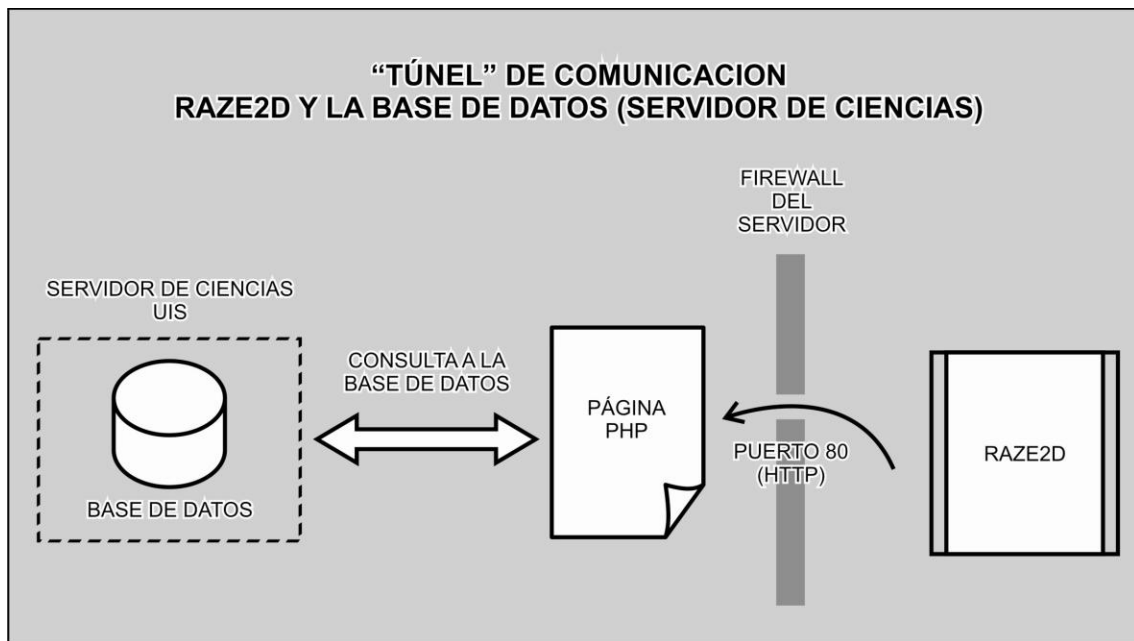


FIGURA 7. Tunel de comunicación Raze2D-Base de Datos

6.3 NETBEANS

Netbeans es un IDE (Entorno desarrollo integrado). Una herramienta pensada para programadores para escribir, depurar, compilar y ejecutar programas. Además permite desarrollar en distintos lenguajes tales como Ruby, C++ y lógicamente java. Netbeans fue la herramienta escogida porque fue desarrollada como software libre y tiene un potencial muy reconocido por los desarrolladores de software en el mundo entero. Permite además manejar de una manera muy amigable y fácil la inclusión de librerías externas, como en el caso de la librería JAI de Java para el tratamiento de imágenes. También es muy notable el editor de formularios, que es mucho más completo que el de

otros entornos de desarrollo, además es muy reconocido por las ventajas para los estudiantes y las personas que recién están empezando a desarrollar aplicaciones en Java.

6.4 ALGORITMO DE TRADUCCIÓN DE LA ESCALA DE MASA MOLECULAR

Los geles de poliacrilamida utilizados para los procesos de electroforesis 2D presentan en su eje vertical la escala para la masa molecular, que es exponencial (ver FIGURA 8), pero el logaritmo de la masa molecular presenta una escala lineal (si M_w es exponencial, entonces $\text{Log}(M_w)$ es lineal). De esta manera se puede convertir una escala exponencial en una escala lineal (FIGURA 9) para poder cuantificar las distancias entre un punto y otro.

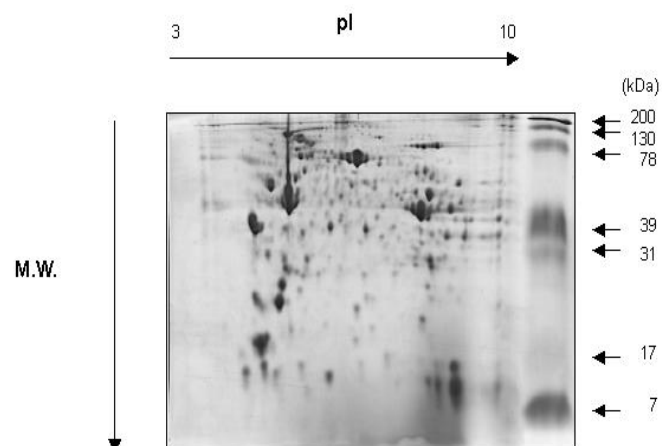


FIGURA 8. Distribución de la masa molecular y el punto isoeléctrico en un gel de electroforesis 2D.

El algoritmo para el cálculo de las distancias entre spots se basa en la formulación matemática para calcular la distancia entre 2 puntos en un plano cartesiano, en el cual ambos ejes presentan una relación lineal.

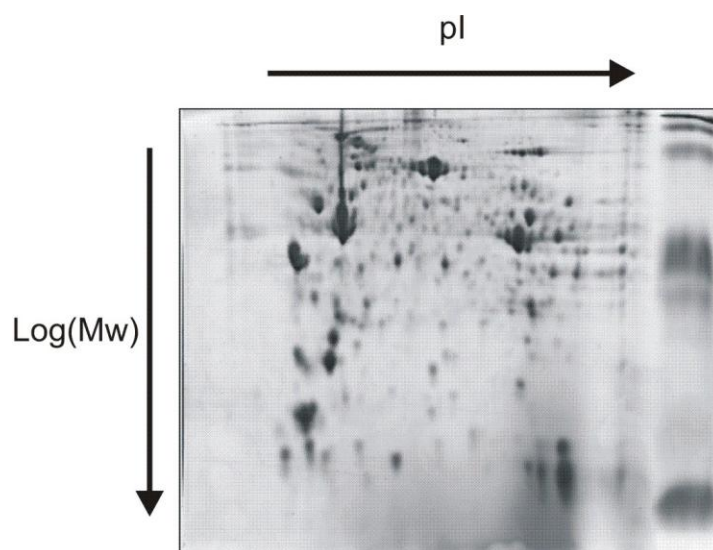


FIGURA 9. Conversión de la escala de la masa molecular a una escala lineal.

6.5 ALGORITMO PARA EL CÁLCULO DEL PORCENTAJE DE SIMILITUD

Como primera medida, se recurre a la suposición de que es posible realizar un montaje universal (en el laboratorio) en el que todas las proteínas encontradas en la base de datos se encuentran presentes en un solo gel (ver FIGURA 10).



FIGURA 10. Montaje universal de las proteínas presentes en la SWISS 2D-PAGE.

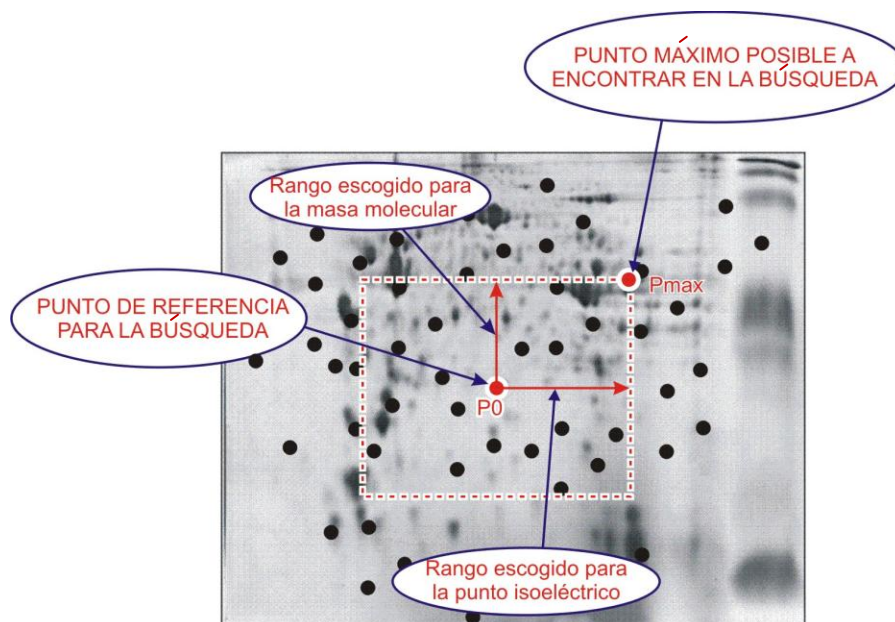


FIGURA 11. Establecimiento del área de búsqueda a partir del punto base y los rangos.

La búsqueda en la base de datos se realiza con los valores tomados de la imagen, y unos rangos establecidos por el usuario (ver FIGURA 11 y FIGURA 12). De la cual se obtiene una lista de las proteínas que cumplen con el criterio

de búsqueda ya establecido (los valores de punto isoeléctrico y masa molecular que calcula la herramienta al seleccionar un spot).

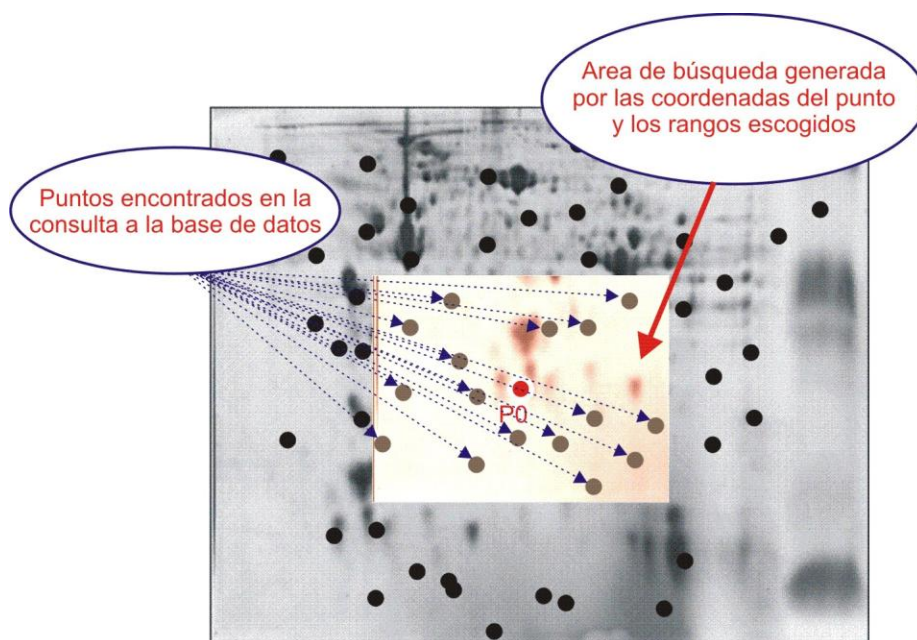


FIGURA 12. Puntos encontrados en el área de búsqueda.

Con la traducción de las coordenadas de masa molecular y con la escala de pH se puede obtener un “sistema de coordenadas” para proteínas que nos permite establecer las distancias entre dos puntos. Pero para cuantificar la cercanía de los puntos encontrados frente al punto tomado como base para establecer el criterio de búsqueda, es necesario establecer una unidad de referencia y en base a dicha referencia poder definir un porcentaje de similitud. Para este fin, se utilizan los rangos utilizados en la construcción del criterio de búsqueda para definir así la distancia máxima posible a encontrar en la consulta a la base de datos (D_{max}), y a esta distancia encontrada, se le asigna el valor de similitud 0%. Con base en este valor, se compara la distancia entre cada punto

encontrado en la base de datos con el punto base (D_{punto}), y según los rangos establecidos. Este valor se ha de conocer como el porcentaje de similitud de cada punto encontrado.

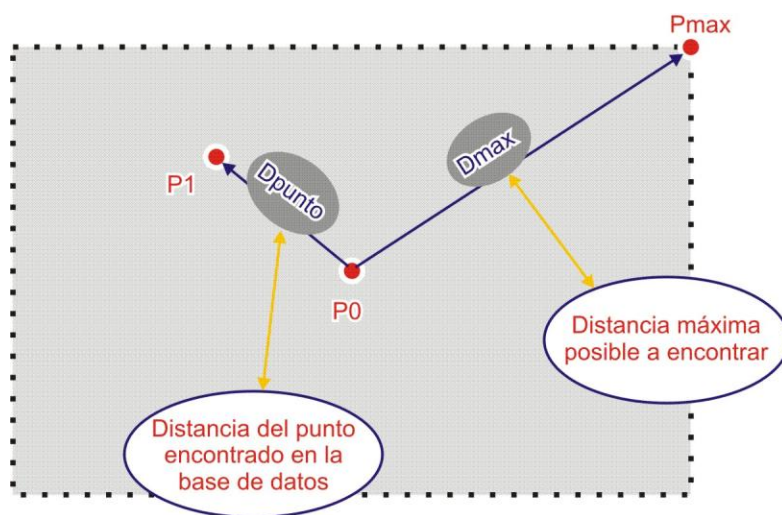


FIGURA 13. D_{punto} y D_{max} a partir de las coordenadas del punto base P_0 y de P_{max} y P_1 .

La distancia a calcular entre dos spots se basa en el cálculo de la distancia (ver FIGURA 13) entre dos coordenadas en un plano cartesiano, pero teniendo en cuenta que la distancia entre dos puntos en el eje y (el de la masa molecular) tiene que ser traducido a un eje de lineal de logaritmo de masa molecular.

Los puntos a los que se le ha de calcular la distancia son: el punto base para la consulta a la base de datos (P_0), y cada punto encontrado en dicha búsqueda (P_1).

P_0 tiene coordenadas (P_0 , M_0) que corresponden a sus valores de punto isoeléctrico y masa molecular respectivamente. Y P_1 (P_1 , M_1) son los valores

de cada punto encontrado en la base de datos. Las coordenadas de Pmax corresponden a los valores de P0 sumados con los valores de rango seleccionados para la consulta en la base de datos. Pmax (P₀+R₁, M₀+R₂), donde R₁ es el valor de rango escogido para la masa molecular y R₂ es el valor de rango escogido para el punto isoelectrico.

$$D_{\max} = \sqrt{\left[\text{Log} \left(\frac{M_0 + R_2}{M_0} \right) \right]^2 + (R_1)^2} \quad (1)$$

$$D_{\text{punto}} = \sqrt{\left[\text{Log} \left(\frac{M_1}{M_0} \right) \right]^2 + (P_1 - P_0)^2} \quad (2)$$

$$\text{Similitud (\%)} = \left(1 - \frac{D_{\text{punto}}}{D_{\max}} \right) \times 100 \quad (3)$$

FIGURA 14. Deducción de Dmax, Dpunto y el porcentaje de similitud.

Para el algoritmo del cálculo del porcentaje de similitud (ver FIGURA 14) hay que tener en cuenta que la movilidad electroforética se presenta en una relación lineal con el logaritmo de la masa molecular. Por esta razón, en el cálculo de las distancias no se toma como parámetro la masa molecular, sino el logaritmo de la masa molecular.

6.6 ALGORITMO PARA EL CÁLCULO DEL PUNTO ISOELÉCTRICO Y LA MASA MOLECULAR DE UN ‘SPOT’ EN BASE A SU POSICIÓN EN LA IMAGEN.

En los geles de poliacrilamida utilizados para la electroforesis 2D, cada posición en su escala vertical u horizontal corresponde a un único valor de punto isoeléctrico y masa molecular. Basándose en esto es posible interpretar una imagen digital (imagen escaneada) del gel; y cada uno de sus píxeles¹⁶ corresponde de la misma manera a un único valor de punto isoeléctrico y masa molecular.

Sin embargo, debido a la acción del campo eléctrico al que se ve sometido el gel durante una de las fases de la electroforesis (isoelectroenfoque), es posible que esta distribución pierda su consistencia, y por ello los bioquímicos utilizan los marcadores, que ellos mismos pueden identificar a simple vista dentro de la imagen.

Gracias al uso de los marcadores, es posible identificar la nueva distribución de los valores de punto isoeléctrico y masa molecular presentes en cada posición de la imagen, a través de la interpolación de sus valores (teniendo en cuenta la escala original que poseen). La masa molecular, como es sabido, no tiene una escala lineal, por lo tanto, la interpolación se hace en relación con el logaritmo de la masa molecular frente a la movilidad electroforética (Distancia que se ha

¹⁶ Píxel o pixel: (acrónimo del inglés *picture element*, "elemento de imagen") es la menor unidad homogénea en color que forma parte de una imagen digital, ya sea esta una fotografía, un fotograma de vídeo o un gráfico. [Tomado de www.wikipedia.com]

movido desde el vértice superior del gel), y que en últimas correspondería a la posición relativa frente a otros marcadores.

PARA EL MÉTODO DE REGRESIÓN LINEAL
POR MÍNIMOS CUADRADOS

Ecuación general de la recta
 $y = mx + n$
 con:

$$S_x = \sum x_i \quad S_{xx} = \sum x_i^2$$

$$S_y = \sum y_i \quad S_{xy} = \sum x_i y_i$$

N = cantidad de puntos a utilizar para la regresión.

$$m = \frac{N \cdot S_x - S_x \cdot S_y}{N \cdot S_{xx} - S_x \cdot S_x}$$

$$n = \frac{S_{xx} \cdot S_y - S_x \cdot S_{xy}}{N \cdot S_{xx} - S_x \cdot S_x}$$

FIGURA 15. Cálculo de la recta de interpolación por medio de mínimos cuadrados.

En los geles de poliacrilamida, la masa molecular presenta una tendencia lineal entre la movilidad electroforética y el logaritmo de la masa molecular, de esta manera se puede obtener la ecuación de la línea recta¹⁷ que define el

¹⁷ El método utilizado para hacer el cálculo de la ecuación de la línea fue el de mínimos cuadrados.

comportamiento de esta tendencia (FIGURA 15), y aplicarla al rango que corresponde dentro de los marcadores utilizados para tal fin.

Para la generación de los valores de punto isoeléctrico y masa molecular, el software escoge por cada coordenada (punto isoeléctrico: x ó masa molecular: y) los marcadores que se encuentran más cerca, uno hacia adelante y uno hacia atrás, y entre ellos genera la ecuación que determina la variación, y luego es posible asignar el valor al punto.

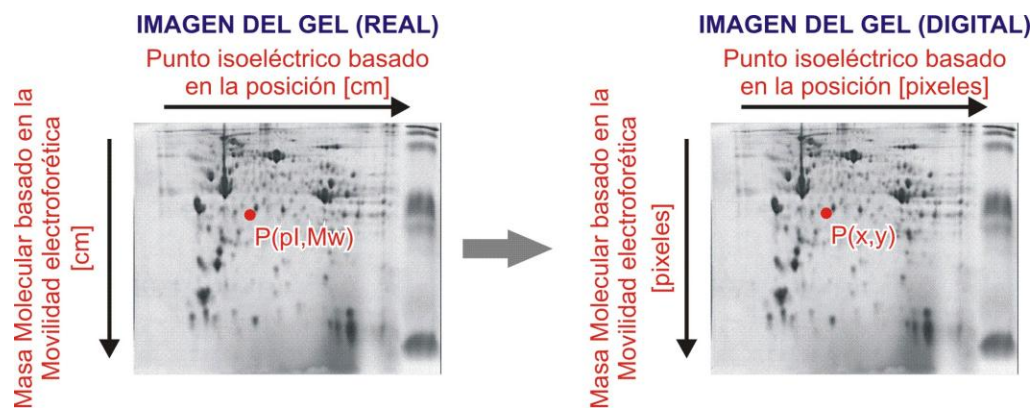


FIGURA 16. Conversión de la escala de imagen real a imagen digital hecho por RAZE-2D.

6.7 MANUAL DE USO DE LA HERRAMIENTA

RAZE-2D es una herramienta libre y educativa capaz de analizar imágenes de electroforesis en 2D, para determinar semejanzas con otras proteínas identificadas según el SIB. La herramienta consta de varios menús agrupando y facilitando el uso de la misma.

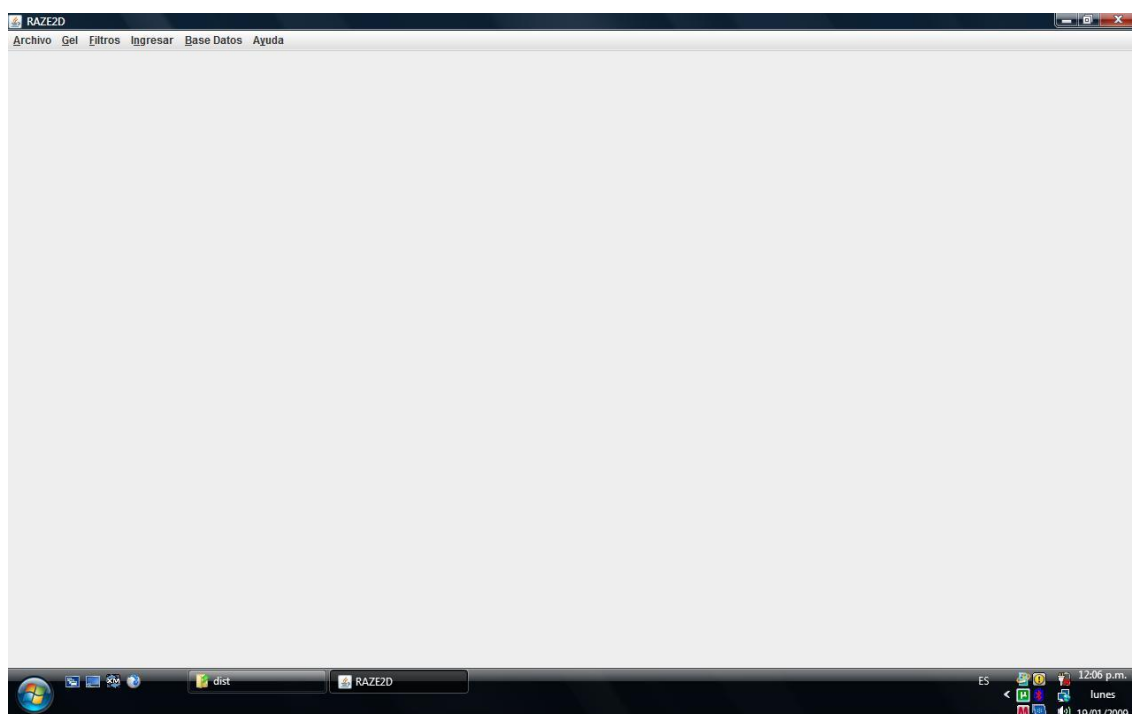


FIGURA 17. Visualización de la Herramienta Raze2D

1. **Archivo:** En el menú Archivo se puede encontrar los siguientes ítems:

- * Abrir: Ítem encargado de abrir la imagen resultante del proceso de electroforesis 2D.
- * Guardar: Ítem encargado de almacenar en cualquier parte del disco que desee el usuario la imagen analizada por la herramienta.
- * Salir: Ítem encargado de abandonar el programa.

2. **Gel:** En el menú Gel se puede encontrar los siguientes ítems:

- * Conversión a Grises: Ítem encargado de convertir la imagen a escala de grises.
- * Imagen Completa: Ítem encargado de ampliar la imagen a su tamaño original.
- * Negativo: Ítem encargado de convertir la imagen a negativo para una mejor visualización.
- * Restaurar Imagen: Ítem encargado de restaurar la imagen a su estado original.

3. **Filtros:** En el menú filtros se utiliza para limpiar las imágenes, eliminar el ruido existente en el gel. Se encuentran los siguientes ítems:

- * Paso Bajo: Filtro utilizado para permitir el paso de frecuencias más bajas y atenuar las frecuencias más altas. Los filtros utilizados son Media 5x5, Media 3x3, Gauss.
- * Paso Alto: Filtro utilizado para atenuar los componentes de baja frecuencia pero no las de alta frecuencia. Los filtros utilizados son Sharpen y Laplaciano.

4. **Ingresar:** En el menú Ingresar se puede encontrar los siguientes ítems:

- * Marcadores: Ítem utilizado para ingresar los marcadores correspondientes a cada imagen a analizar.
- * Spots: Ítem utilizado para ingresar los 'spots' o proteínas a analizar en el gel correspondiente.
- * Deshacer Marcador: Ítem encargado de eliminar el último marcador añadido.

5. **Base de Datos:** En el menú Base de Datos se utiliza para consultar a las dos bases de datos disponibles. Se encuentran los siguientes ítems:

- * Swiss2DPage: Ítem utilizado para consultar los mapas existentes en la base de datos Swiss2DPage.
- * Uniprot: Ítem utilizado para consultar información sobre cualquier proteína en la base de datos Uniprot a través de internet.

PASOS A SEGUIR PARA ANALIZAR UNA IMAGEN DE ELECTROFORESIS 2D CON RAZE2D

1. Abrir Imagen

Se accede al menú archivo y se hace clic en el ítem Abrir. A continuación aparecerá una imagen como la FIGURA 18, en la cual se ingresa los datos iniciales del gel que se va a analizar, entre ellos el punto isoeléctrico inicial y final, masa molecular inicial y final y por último el número de marcadores correspondiente del gel.

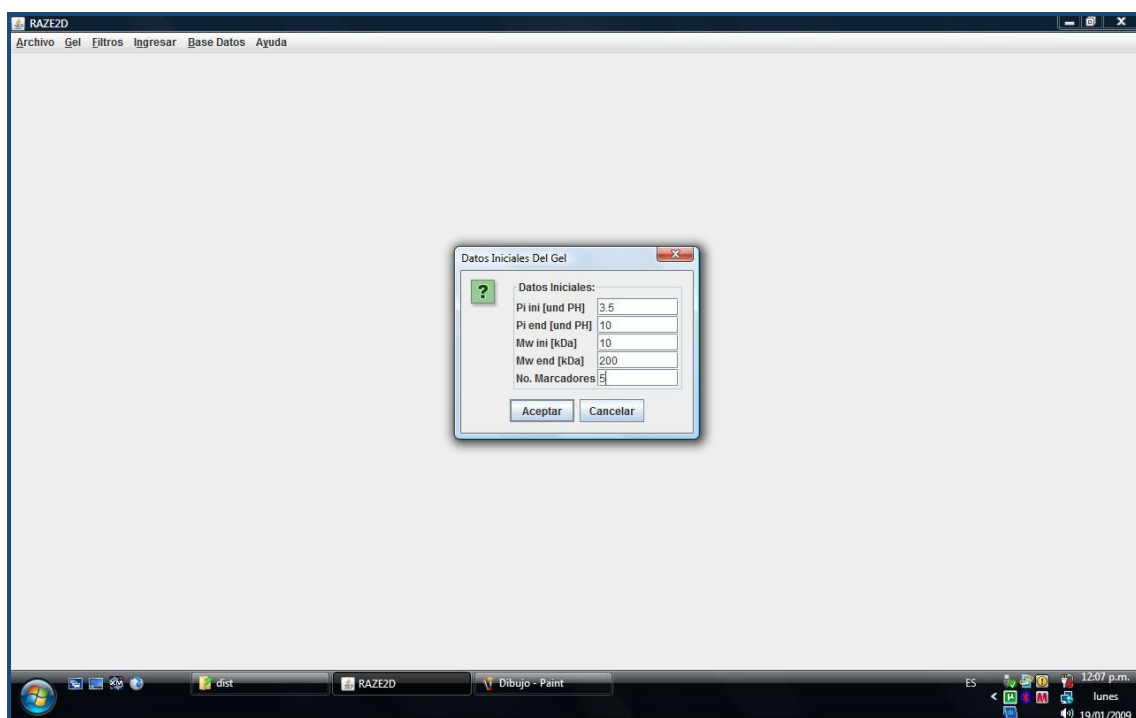


FIGURA 18. Ingresando los datos iniciales del gel a analizar

La imagen del gel se visualizará en la parte izquierda con sus respectivas coordenadas iniciales de punto isoeléctrico y masa molecular.

2. Ingresar Marcadores

Después de abrir la imagen se procede a insertar los marcadores correspondientes del gel. Se dirige al menú Insertar->Marcadores y se proporciona clic a donde compete cada marcador en el gel y se digita los valores de su punto isoeléctrico y masa molecular.

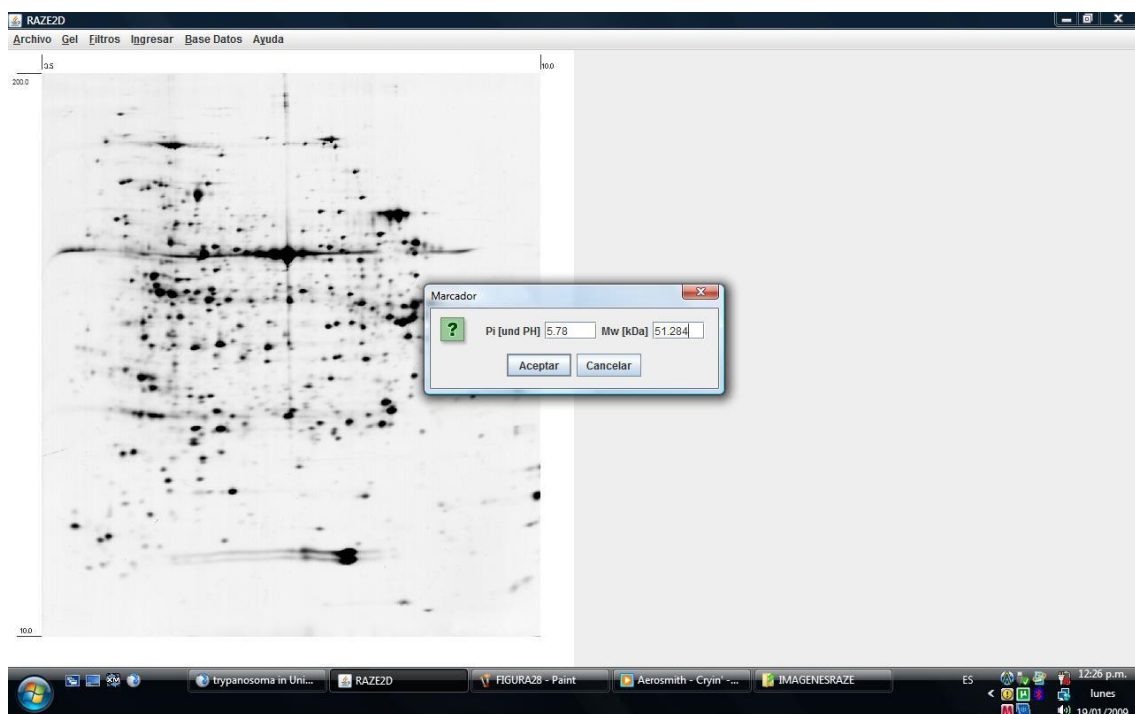


FIGURA 19. Ingresando Marcador al Gel

La herramienta automáticamente informará a través de un mensaje cuando se haya insertado el último marcador según los datos iniciales. Para una mejor

visualización del gel ya que en algunos casos no se pueden visualizar bien el punto deseado a ingresar, la herramienta posee varias opciones:

- Negativo o Conversión a Grises: Se traslada al menú Gel y se aplica la opción que más se adecúe al gel.
- Filtro paso alto: Se dirige al menú filtros y se aplica cualquier filtro que mejore la claridad de la imagen.

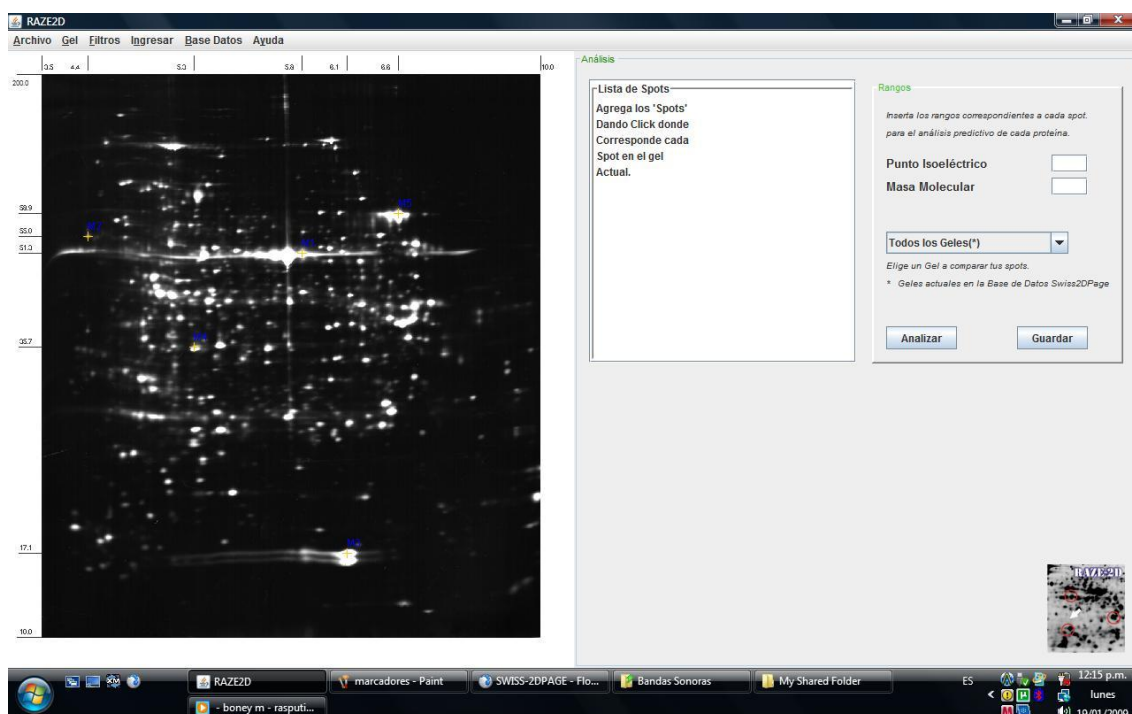


FIGURA 20. Imagen que corresponde a un mapa (Arabidopsis) con algunos marcadores ya insertados aplicando la opción Negativo al gel

3. Ingresar los Spots

Luego de tener la totalidad de marcadores en el gel, aparecerá un mensaje indicando que se debe proceder a ingresar los spots. Así que se dirige al menú Insertar->Spots, y a continuación se hace clic en cualquier parte del gel donde se desee conocer su punto isoeléctrico y su masa molecular.

Al ingresar cada spot en la parte derecha de la herramienta se tendrá una lista donde irán apareciendo los spots que se esté deseando conocer sus valores de punto isoeléctrico y masa molecular con su respectiva numeración para facilitar su búsqueda e identificación (FIGURA 21).

Al realizar esta acción se obtendrá una imagen como la siguiente

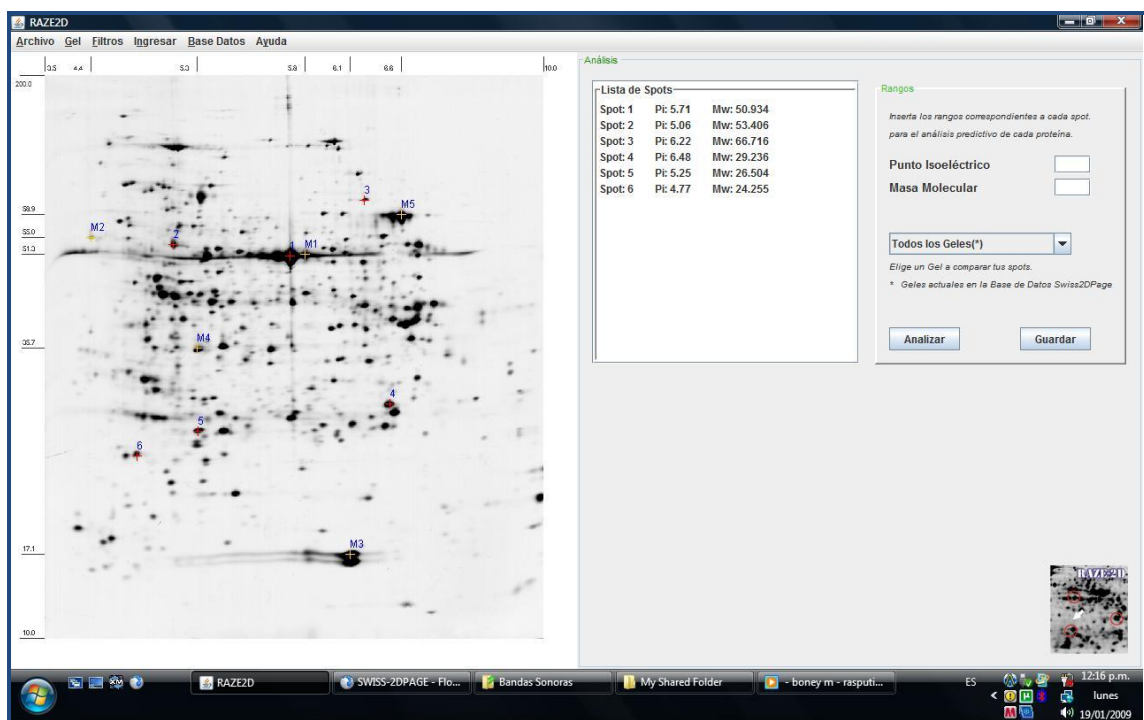


FIGURA 21. Ingresando los spots

4. Encontrando Similitudes de los spots.

Ahora se procederá a señalar el spot de la lista disponible, especificar los rangos de búsqueda para el punto isoeléctrico, masa molecular y elegir del combo algún mapa o todos los mapas en los cuales se desee encontrar algún porcentaje de similitud de alguna proteína identificada en la base de datos Swiss2DPage.

Al realizar este procedimiento perfectamente se obtendrá un resultado como la FIGURA 22, en la cual se ilustra una tabla con las 20 proteínas más cercanas (si existen) resultantes según su criterio de búsqueda con su respectivo punto isoeléctrico, masa molecular, mapa al que pertenece y lógicamente el porcentaje de similitud que tiene con el spot señalado a analizar. Luego se podrá ver un tooltip en el gel con los datos de la proteína más cercana en el spot que se eligió para realizar la búsqueda.

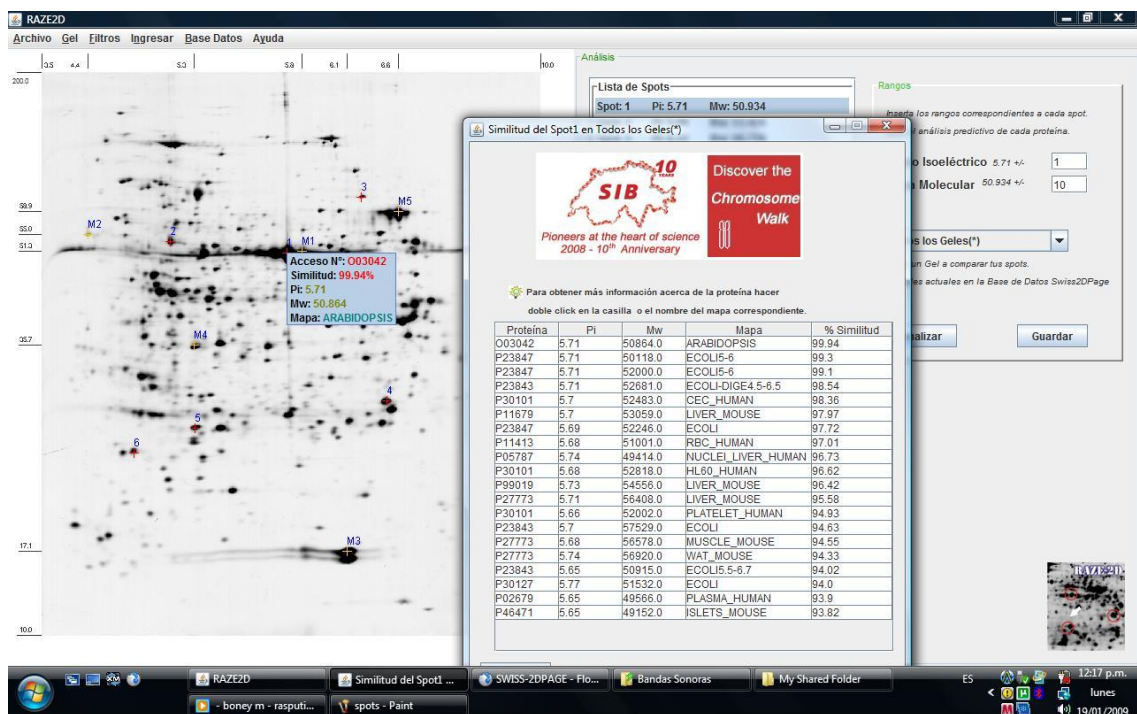


FIGURA 22. Imagen de las proteínas encontradas según el criterio de búsqueda establecido

En la tabla mostrada con las proteínas encontradas se puede hacer doble clic sobre cada número de acceso de la proteína (AC) para obtener más información sobre dicha proteína y también tener la opción de saber la posición de la proteína en el gel que pertenece dando doble clic en la columna Mapa.

5. Guardando Información

RAZE-2D permite al usuario guardar la información generada por la herramienta en archivos de texto refiriéndose a la tabla de proteínas encontradas o a la lista de spots generadas por el usuario. (Ver FIGURA 23 y FIGURA 24). Como se puede observar claramente en la FIGURA 21 o FIGURA 22, cada parte tiene un botón denominado “Guardar” el cual ayudará al usuario a ubicar el archivo de texto en cualquier parte del disco de su máquina (Ver FIGURA 25). Para una mejor visualización del archivo de texto plano, es recomendable al usuario de Windows abrirlo con el programa WORDPAD.

Además de poder almacenar la información anterior, se puede almacenar la imagen del gel con sus respectivos marcadores y números de spots señalados por el usuario en una imagen de formato JPEG en cualquier parte del disco duro. (Ver FIGURA 26)

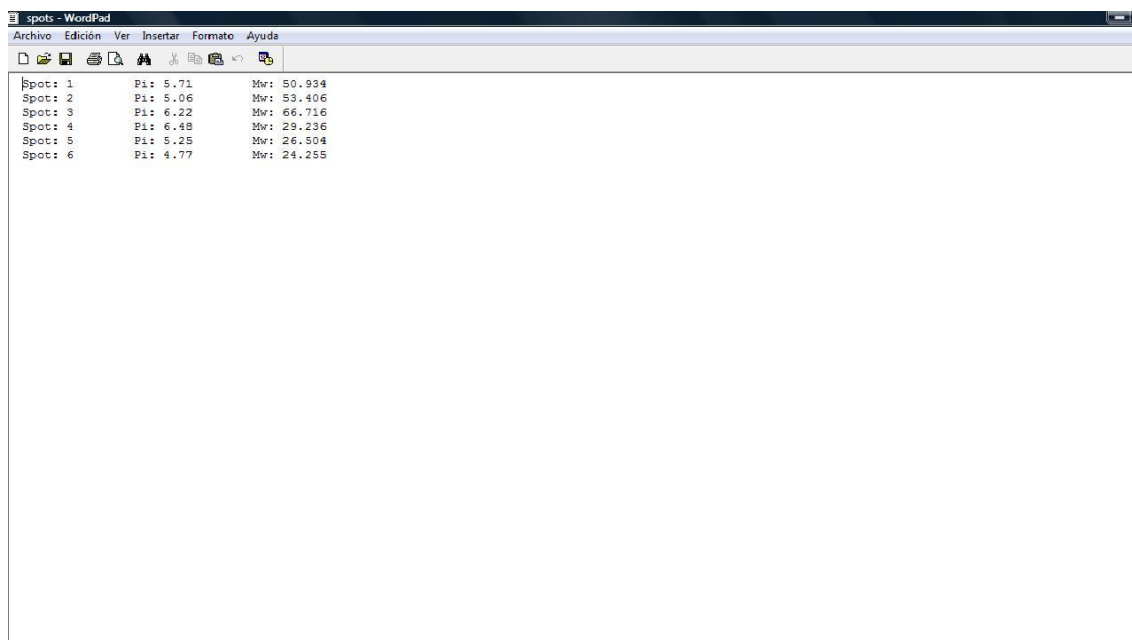


FIGURA 23. Archivo de texto el cual contiene la lista de los spots indicados por el usuario

proteinas - WordPad				
Archivo Edición Ver Insertar Formato Ayuda				
Proteína	PI	MW	%	Nombre Gel
O03042	5.71	50864.0	99.94	ARABIDOPSIS
P23847	5.71	50118.0	99.3	ECOLIS-6
P23847	5.71	52000.0	99.1	ECOLIS-6
P23843	5.71	52681.0	98.54	ECOLI-DIGE4.5-6.5
P30101	5.7	52483.0	98.36	CEC_HUMAN
P11679	5.7	53059.0	97.97	LIVER_MOUSE
P23847	5.69	52246.0	97.72	ECOLI
P11413	5.68	51001.0	97.01	RBC_HUMAN
P05787	5.74	49414.0	96.73	NUCLEI_LIVER_HUMAN
P30101	5.68	52813.0	96.62	HL60_HUMAN
P99019	5.73	54556.0	96.42	LIVER_MOUSE
P27773	5.71	56408.0	95.58	LIVER_MOUSE
P30101	5.66	52002.0	94.93	PLATELET_HUMAN
P23843	5.7	57529.0	94.63	ECOLI
P27773	5.68	56578.0	94.55	MUSCLE_MOUSE
P27773	5.74	56920.0	94.33	WAT_MOUSE
P23843	5.65	50915.0	94.02	ECOLIS.5-6.7
P30127	5.77	51532.0	94.0	ECOLI
P02679	5.65	49566.0	93.9	PLASMA_HUMAN
P46471	5.65	49152.0	93.82	ISLETS_MOUSE

FIGURA 24. Archivo de texto con los datos encontrados de las proteínas similares al spot señalado

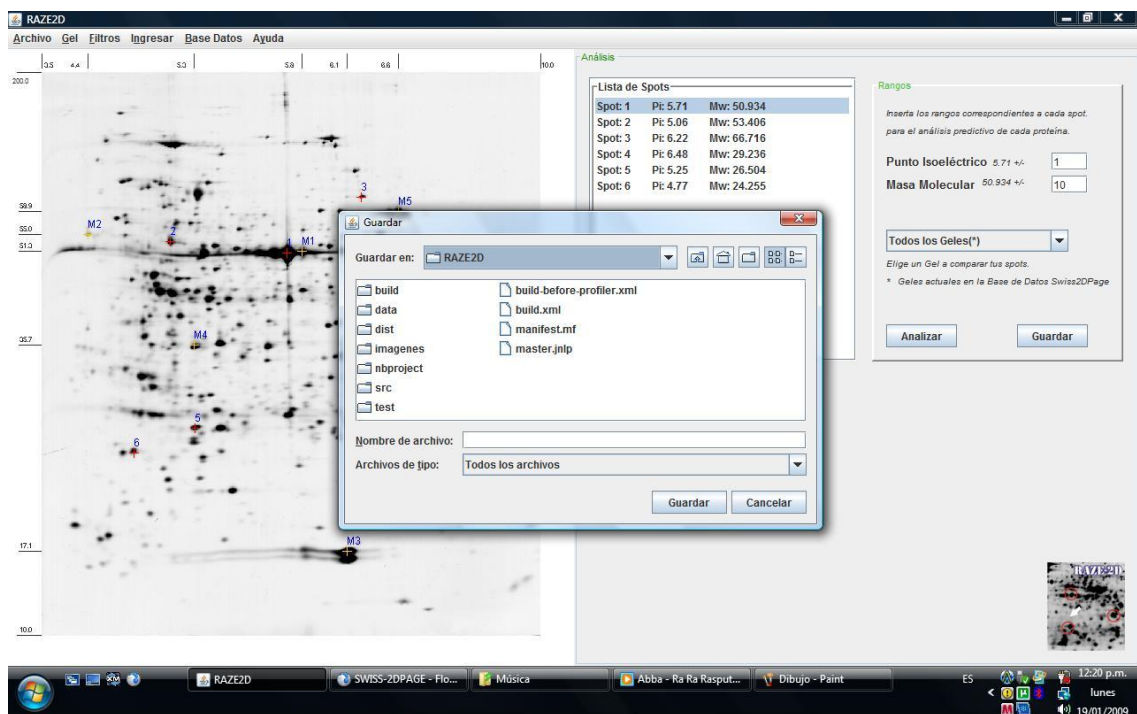


FIGURA 25. Guardar información

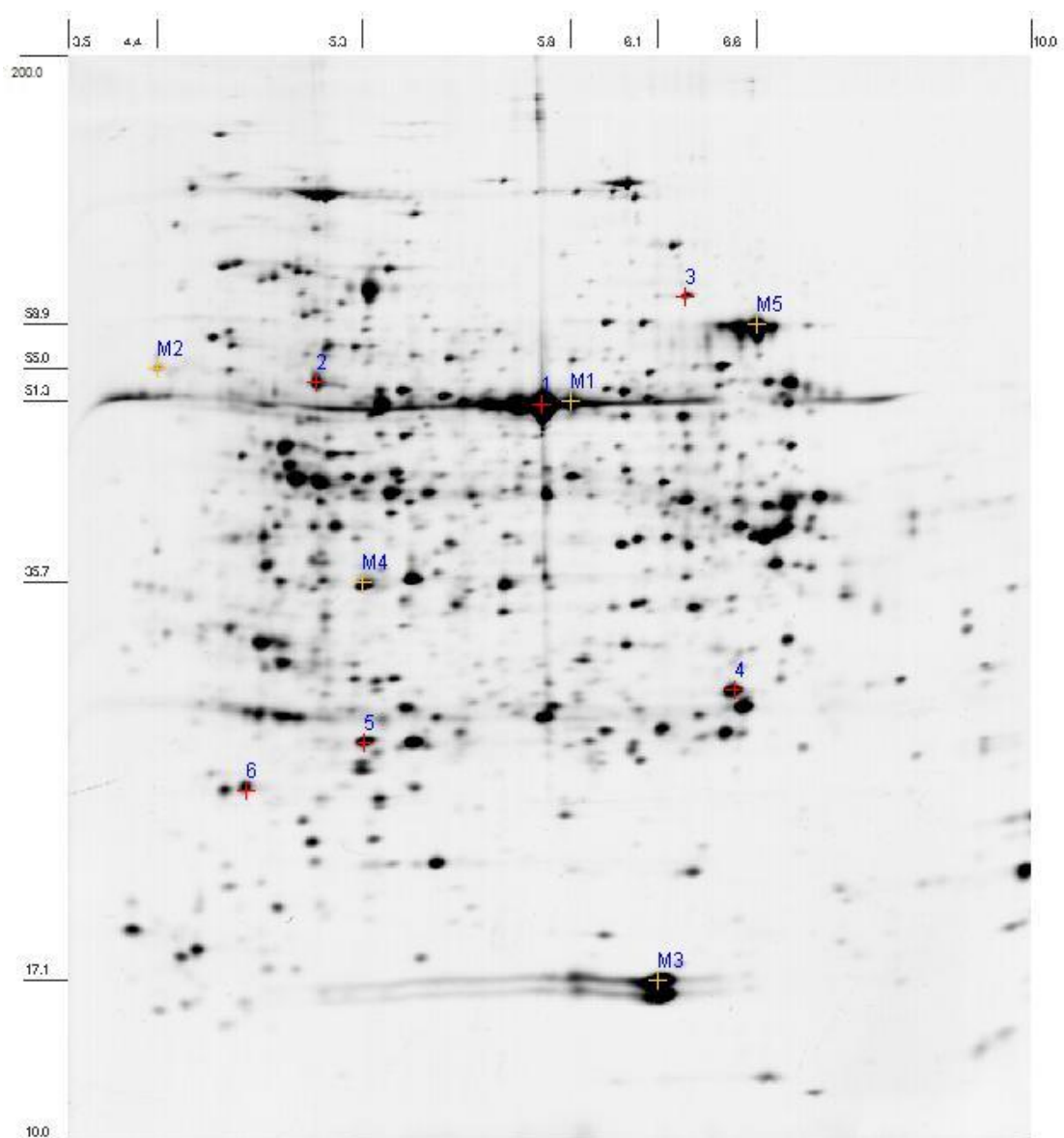


FIGURA 26. Imagen del gel con sus respectivos marcadores y spots señalados por el usuario

6. Consultando en las Base de Datos

Raze2D permite al usuario consultar sobre cada mapa existente en la base de datos SWISS2DPAGE donde se desglosará los datos pertenecientes de cada proteína en dicho mapa. (Ver FIGURA 27). Para acceder a esta opción se dirige al menú Base Datos->Swiss2DPage.

Si se desea hacer una consulta más avanzada RAZE-2D posee la opción de consultar en la base de datos Uniprot la cual contiene variedades de información relacionadas con proteínas. (Ver FIGURA 28), con las mismas opciones de búsquedas que se encuentran en su página web, (<http://www.uniprot.org>) para así facilitar al usuario dirigirse directamente a la página. Para esta consulta la herramienta utiliza el navegador por defecto para poder realizarla, solamente se digita la información a consultar y se proporciona clic en el botón buscar. Para acceder a esta opción se dirige al menú Base Datos->Uniprot.

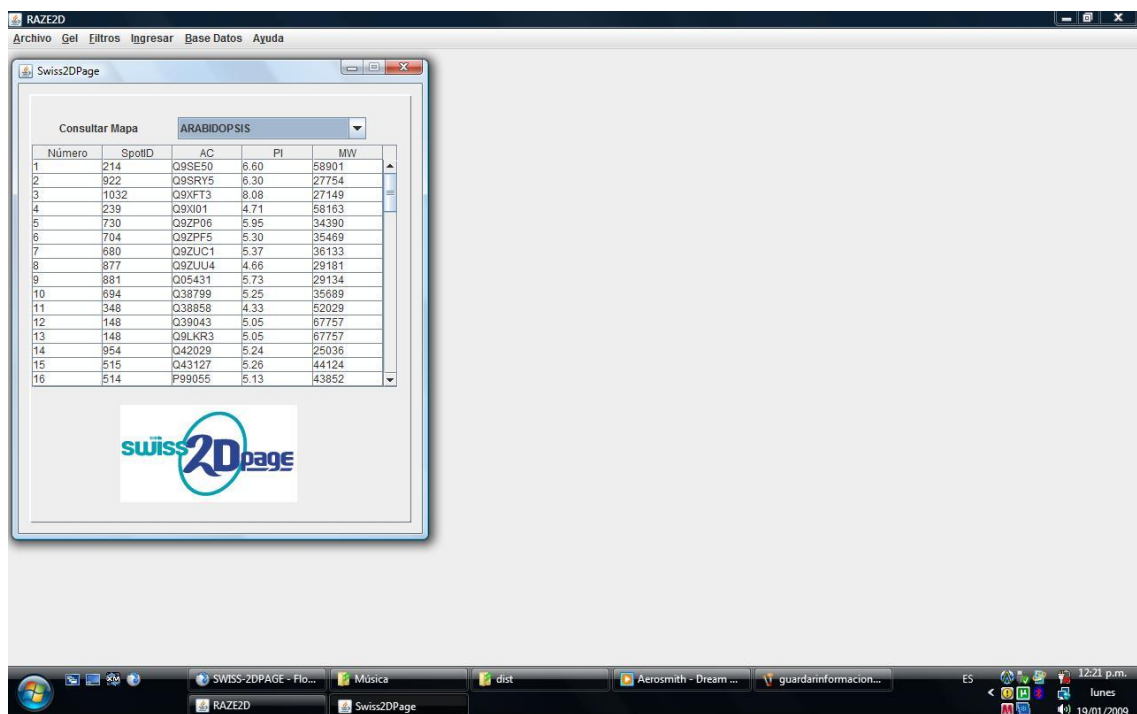


FIGURA 27. Consultando en la Base de Datos Swiss2DPage

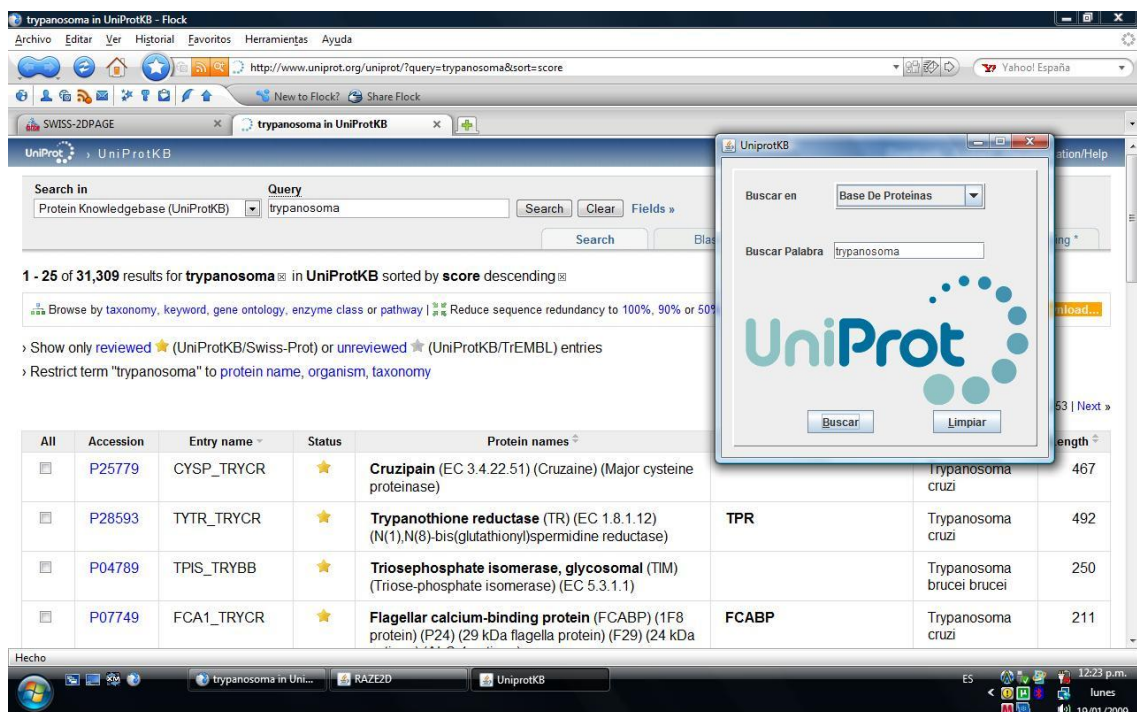


FIGURA 28. Consultando en la base de datos UniProt a través de la herramienta

7. METODOLOGÍA DE PROTOTIPADO EVOLUTIVO

La metodología utilizada para el desarrollo de la herramienta computacional (RAZE-2D), es el prototipado de sistemas, que permite la implementación de los diseños realizados, de tal manera que se concibe de una manera evolutiva e iterativa la implementación de la aplicación de la siguiente manera:

Identificación de requerimientos conocidos: Permite la generación de los prototipos, basados en los requerimientos preliminares expresados por el usuario.

Desarrollo de un modelo de trabajo: En este paso se pueden establecer las condiciones necesarias para el buen desarrollo de cada uno de los prototipos, y estas son:

- El lenguaje para el diálogo o conversación entre el usuario y el sistema.
- Pantallas y formatos para la entrada de datos.
- Módulos esenciales de procesamiento.
- Salida del sistema.

Utilización del prototipo: El prototipo es puesto a prueba bajo condiciones reales por el usuario, para poder así identificar los puntos claves a mejorar o los nuevos requerimientos del sistema.

Revisión del prototipo: Se obtiene la información acerca de los componentes del sistema que han de ser corregidos, mejorados, rediseñados, etc. Y se planea el desarrollo del siguiente prototipo.

7.1 ESTUDIO INICIAL

En esta primera etapa se realizó un estudio previo del problema, la identificación de los procesos involucrados en la identificación de proteínas tales como la electroforesis, el estándar de bases de datos del SIB (Swiss Institute of Bioinformatics). De igual manera se establecieron los recursos necesarios para llevar a cabo el proyecto.

Acercamiento al problema e identificación de necesidades: Estudio de las características del proceso de electroforesis, propiedades de los geles de poliacrilamida utilizados en dichos procesos, el procedimiento de identificación de proteínas utilizando la base de datos SWISS 2D-PAGE.

Recopilación de bibliografía y conceptualización: Interpretación de los diferentes estudios hechos sobre el tema, evaluación del estado del arte en cuanto a la identificación de proteínas por parte de la comunidad científica internacional.

Determinación de recursos: Se hizo un análisis de los recursos necesarios para el montaje de la base de datos, así como las herramientas computacionales a utilizar para el desarrollo del software.

7.2 GENERACIÓN DE LA BASE DE DATOS LOCAL

En esta fase se hizo la generación de la base de datos local en la cual se almacenó la información de proteínas de la base de datos SWISS 2D-PAGE.

Para esto se utilizó una herramienta computacional facilitada por el SIB (MAKE 2D-DB II¹⁸), y se obtuvo los elementos necesarios para su implementación.

Reconocimiento de la estructura relacional de la Base de Datos SWISS

2D-PAGE: Gracias a la ayuda proporcionada por la Dra. Christine Hoogland y el Dr. Khaled Mostaguir del Instituto Suizo de Bioinformática (SIB) se tuvo acceso a la información concerniente al diagrama relacional de la base de datos SWISS 2D-PAGE.

Asimilación de la herramienta MAKE 2D-DB II: MAKE 2D-DB II es la herramienta utilizada para la generación de la base de datos local. Con base en la información publicada en el servidor Expasy sobre MAKE 2D-DB II se instalaron los componentes necesarios para el correcto funcionamiento de esta herramienta, que se ejecuta en un sistema operativo basado en UNIX(GNU-Linux, por ejemplo).

Adquisición de los archivos fuente de la base de datos: La base de datos SWISS 2D-PAGE es una base de datos con copyright que exime de algún costo a entidades educativas y entidades sin ánimo de lucro, por lo que fue posible acceder a los archivos fuente de la base de datos, en formato de la herramienta Melanie Viewer¹⁹(archivos .mel), de donde se obtuvieron las imágenes de los geles utilizados en el montaje y las anotaciones sobre las proteínas encontradas.

¹⁸ Herramienta utilizada por el Instituto Suizo de Bioinformática para la generación de bases de datos relacionales con información experimental sobre proteínas encontradas por medio de electroforesis 2D. La versión utilizada por los autores del proyecto fue la 2.50.

¹⁹ Herramienta privativa para el estudio de imágenes de electroforesis 2D utilizada por el Instituto Suizo de Bioinformática. La versión utilizada por los autores del proyecto fue la 5.02.

Generación de archivos de texto plano a partir de los archivos fuente: La herramienta Melanie Viewer en su versión 5.02 permite la exportación de las anotaciones de proteínas encontradas en los diferentes geles mediante archivos de texto plano. De esta manera el conjunto de imágenes de los geles y las anotaciones relacionadas con cada gel permitió la generación de la base de datos local con la información perteneciente a la base de datos SWISS 2D-PAGE.

Implementación de la Base de datos local: MAKE 2D-BD II genera automáticamente la base de datos relacional al a partir de diferentes tipos de datos fuente²⁰(en nuestro caso, .png y .txt).

7.3 ELABORACIÓN DEL PRIMER PROTOTIPO

En esta etapa se desarrolló el primer prototipo de la aplicación, que incluía las rutinas necesarias para la consulta en la base de datos, y su presentación a los usuarios para su uso y revisión. De esta manera se establecieron nuevos requerimientos que fueron implementados en el siguiente prototipo junto con los módulos siguientes.

Identificación de requerimientos conocidos: A través de entrevistas con los usuarios a los que va destinada la herramienta (estudiantes y profesionales del grupo de investigación en bioquímica e ingeniería de proteínas de la UIS), se definieron los primeros requerimientos del software desarrollado.

²⁰ MAKE 2D-DB II permite la utilización de diferentes tipos de archivos fuente, entre los que se encuentran exportaciones xml, archivos de texto plano .txt para anotaciones sobre imágenes, archivos de texto plano .dat. En cuanto a imágenes admite formatos .png y .gif.

Conceptualización suplementaria: En esta etapa se inició la adquisición del entorno de desarrollo usado (Netbeans IDE), al igual que la información relacionada con el lenguaje de programación utilizado (JAVA), el establecimiento del mecanismo de conexión a la base de datos y la generación de la página php para este fin.

Desarrollo del modelo de trabajo: En esta fase se realizó el desarrollo del primer prototipo de la herramienta, siguiendo los requerimientos obtenidos y utilizando los componentes antes descritos.

Presentación del prototipo para uso y revisión por parte de los usuarios: El usuario recibió el prototipo desarrollado e hizo las recomendaciones para el mejoramiento de la herramienta, que fueron tenidas en cuenta para la elaboración del prototipo final.

7.4 ELABORACIÓN DEL PROTOTIPO FINAL

En esta segunda fase de elaboración de prototipo, se implementaron las correcciones y los nuevos requerimientos expresados por los usuarios en la revisión del primer prototipo. Además se incluyeron las rutinas para la interpretación de la imagen y el cálculo del porcentaje de similitud, establecido en el algoritmo diseñado.

Identificación de nuevos requerimientos: Recopilación de nuevos requerimientos expresados por los usuarios tras la experiencia del uso del primer prototipo.

Conceptualización suplementaria: Se hizo la instrucción en el uso de las librerías de tratamiento de imágenes necesarias (Java Advanced Imaging), y el diseño del algoritmo de cálculo del porcentaje de similitud.

Desarrollo del modelo de trabajo: Implementación de rutinas de interpretación de imágenes y del algoritmo para el cálculo del porcentaje de similitud.

Presentación del prototipo final.

7.5 PRUEBAS Y RESULTADOS

7.5.1 PRUEBAS A LA HERRAMIENTA Y LA BASE DE DATOS

El cálculo del error es muy importante para determinar el nivel de precisión que la herramienta maneja en la generación de los valores de punto isoelectrico y masa molecular, ya que define la exactitud misma del algoritmo planteado.

En la imagen presentada a continuación se muestra la tabla de resumen del error relativo y absoluto encontrado al calcular el valor de punto isoelectrico y masa molecular para algunas proteínas escogidas al azar en 10 mapas (geles) de referencia de la base de datos SWISS 2D-PAGE, para así obtener un valor de error relativo promedio total (el ANEXO A contiene las tablas de las pruebas por cada mapa). Para el desarrollo de esta prueba se escogió una proteína al azar en uno de los mapas de referencia de la SWISS 2D-PAGE y se compararon los valores almacenados en dicha base de datos con los valores que presentaba la herramienta al seleccionar la ubicación del mismo spot en el gel.

PRUEBA DE ERROR RELATIVO Y ABSOLUTO

MAPA	ERROR RELATIVO PROMEDIO	ERROR ABSOLUTO PROMEDIO [Pi]	ERROR ABSOLUTO PROMEDIO [Mw]
ARABIDOPSIS THALIANA	1,180885536	0,01125	1,42275
DICTYSLUG	0,350109923	0,01625	0,09025
ECOLI	1,437643816	0,04375	1,053625
BAT-MOUSE	1,282075151	0,0275	0,83125
YEAST	0,372565968	0,025	0,129625
HUMAN-KIDNEY	0,739323974	0,02375	0,252
STAPHYLOCOCCUS	0,604941082	0,04	0,14
MUSCLE-MOUSE	0,870729773	0,00875	0,3845
LYMPHOCYTE	0,121891786	0,005	0,04575
HUMAN-LIVER	0,907371668	0,0275	0,725125
PROMEDIO	0,786753868	0,023	0,5078875

Tabla 1. Resultados de las pruebas de error relativo y absoluto aplicados a la herramienta RAZE-2D.

El error se calculó de la siguiente manera:

Error Absoluto:

$$Ea = |\text{Valor medido} - \text{Valor exacto}| \quad (4)$$

Error Relativo:

$$Er = \left| \frac{Ea}{\text{Valor exacto}} \right| \quad (5)$$

Para los errores calculados, comparando el desempeño de la herramienta frente a datos tomados de la base de datos (Swiss2DPage) y frente a los datos generados por otra herramienta (ver Sección 7.5.2), se realizó un análisis y se generaron conclusiones que son expuestas en la sección 7.5.3.

7.5.2 PRUEBAS COMPARATIVAS CON OTRA HERRAMIENTA

El desempeño de la Raze2D es comparado con el desempeño de otra herramienta (de carácter comercial, PDQUEST Advanced), mediante la comparación de los valores generados de punto isoeléctrico y masa molecular en geles de muestra proporcionados el Grupo de Investigación en Bioquímica e Ingeniería de Proteínas de la Universidad Industrial de Santander.

PRUEBA DE ERROR RELATIVO Y ABSOLUTO

HERRAMIENTA UTILIZADA: PDQUEST ADVANCED

MAPA	ERROR RELATIVO PROMEDIO	ERROR ABSOLUTO PROMEDIO [Pi]	ERROR ABSOLUTO PROMEDIO [Mw]
EPIMASTIGOTES	0,576092838	0,028333333	0,7185
EPIMASTIGOTES 338	0,574323856	0,038333333	0,365833333
PROMEDIO	0,575208347	0,033333333	0,542166667

Tabla 2. Error encontrado entre Raze2D y PDQUEST Advanced

7.5.3 ANALISIS DE LOS ERRORES Y CONCLUSIONES

El error encontrado en los geles, se produce por la sensibilidad del mouse y el manejo que el usuario le da al software, ya que el cálculo del punto isoeléctrico y la masa molecular de un spot se basa en la ubicación que el mismo usuario le

dé al spot, que es señalado mediante un clic en la imagen, y cualquier variación en la posición del puntero del mouse, significa una variación en la posición en la imagen y por consiguiente, una variación en la posición en los píxeles en el eje horizontal y vertical, que son los utilizados por el algoritmo para el cálculo del punto isoeléctrico y la masa molecular.

El procedimiento realizado fue el siguiente:

Para el cálculo del error frente a la base de datos, se utilizaron las imágenes de geles contenidas la base de datos Swiss2DPage. Se analizaron en la herramienta, colocando siete marcadores proporcionados por el Grupo de Investigación en Bioquímica e Ingeniería de Proteínas, y basados en la ubicación proporcionada por el sitio web de la base de datos Swiss2DPage de Suiza: <http://www.expasy.org/ch2d/> seleccionamos los spots correspondientes a las proteínas a las cuales calcular el punto isoeléctrico y la masa molecular.

Para el cálculo del error frente a otra herramienta, se utilizó la herramienta que normalmente emplea el Grupo de Investigación en Bioquímica e Ingeniería de Proteínas, PDQUEST Advanced. Se seleccionaron 2 imágenes de geles al azar y con los marcadores utilizados por la herramienta, se interpretaron las imágenes en Raze2D, escogiendo los spots identificados por PDQUEST y se tabularon los resultados de dichos análisis (ver ANEXO A).

Los errores absoluto y relativo generados en los cálculos realizados por la herramienta, se consideran aceptables por los usuarios, y según la evaluación de la satisfacción del usuario (Ver ANEXO F) realizada por un experto, Dr. Rodrigo Torres Sáez, el cálculo de los parámetros generado por la herramienta es útil.

Se realizó una evaluación por parte del usuario final, donde se presentaron sus apreciaciones acerca de la herramienta, mediante valores cuantitativos, de 1 a 5 (con 5 como máxima nota), y fue llevada a cabo por Rocío del Pilar Niño,

estudiante de Maestría en la escuela de Química de la Universidad Industrial de Santander (Ver ANEXO F).

8. CONCLUSIONES

Este trabajo es la muestra irrefutable de la capacidad de la ingeniería de sistemas de extender su campo de acción hacia la investigación y la generación de nuevo conocimiento a través del trabajo interdisciplinario desde el análisis de los procesos y la generación de algoritmos que faciliten la labor investigativa en la universidad, así como el apoyo informático.

La utilización de nuevos y más potentes lenguajes de programación y de entornos de desarrollo, como en este caso JAVA y Netbeans, hace posible dar continuidad a los procesos de desarrollo de software en cualquiera de sus áreas de influencia, desde la educación hasta la investigación.

El prototipado evolutivo se convierte en una herramienta metodológica importante para el desarrollo de software del cual es difícil establecer con claridad los requerimientos del sistema, y permite que el desarrollo de la herramienta se continúe con mejoras y nuevos requerimientos proporcionados por los usuarios, para darle continuidad al proyecto aquí presentado.

Con el desarrollo de este trabajo se logró la generación de la herramienta computacional RAZE-2D, al igual que una base de datos local con una réplica de la información contenida en la base de datos SWISS-2DPAGE del Instituto Suizo de Bioinformática, y brindando a los usuarios una verdadera opción de análisis cuantitativo de imágenes de electroforesis en 2D y la interpretación de la posibilidad de ser encontradas proteínas nuevas.

Se brinda un enfoque a la interpretación de los 'spots' encontrados en las imágenes de electroforesis en 2D, a través de la incorporación de un "sistema de coordenadas" capaz de cuantificar la cercanía de las proteínas encontradas en una consulta a la base de datos con los datos interpretados por la

herramienta computacional y los rangos de consulta seleccionados por el usuario.

La fundamentación matemática presente en el estudio de la ingeniería de sistemas UIS, es un gran soporte para la generación de algoritmos que faciliten la labor interpretativa de los investigadores, con la posibilidad de establecer métodos cada vez más rápidos y con más precisión en la información generada.

La adquisición de herramientas computacionales de carácter libre, como es el caso de MAKE 2D-DB II, que se rigen por los nacies estándares para el almacenamiento de la información relacionada con las proteínas identificadas experimentalmente en procesos de electroforesis 2D a través de archivos de texto plano, permite la incorporación de nuevos estudios y la posibilidad de su presentación al mundo y a la comunidad científica internacional, de una manera comprensible.

Es un gran orgullo presentar como conclusión el hecho de existir una posibilidad de que Colombia sea el primero en Latinoamérica en generar resultados de análisis de proteínas y estudios sobre proteínas nuevas identificadas y publicadas en la base mundial de proteínas UniProt, y en la base de datos SWISS 2D-PAGE, a través del uso de la herramienta MAKE 2D-DB II.

El trabajo realizado con el desarrollo de este proyecto establece el punto de partida para el trabajo conjunto entre el Grupo de investigación en Ingeniería Biomédica-GIIB y el Grupo de Investigación en Bioquímica e Ingeniería de Proteínas, en el desarrollo de herramientas computacionales que colaboren en las labores investigativas dentro de la Universidad Industrial de Santander.

Es importante reconocer la importancia de darle apoyo computacional a los procedimientos y técnicas y de esta manera mejorar su aplicabilidad en la academia y la investigación.

9. RECOMENDACIONES

- Siguiendo el modelo metodológico escogido, se debe continuar en la generación de nuevos prototipos, recogiendo los nuevos requerimientos del sistema basados en la experiencia del usuario al trabajar con la herramienta RAZE-2D. Estas mejoras incluyen la generación de rutinas para el cálculo del punto isoeléctrico a partir de la secuencia de aminoácidos de la proteína, posibilitando así el uso de la base de datos mundial UniProt; incluir la interpretación de imágenes de electroforesis en una sola dimensión, y el desarrollo de mecanismos para el análisis y comparación de imágenes de varios geles de electroforesis gráficamente.
- Es importante continuar con el seguimiento al desarrollo de nuevos prototipos del software y comenzar con la tarea de generar imágenes de electroforesis propias, basadas en estudios hechos en los laboratorios de bioquímica de la UIS y almacenar toda esa información en la base de datos local.
- Como una segunda fase del proyecto, y continuando con el uso de la herramienta MAKE 2D-DB II, es posible la generación de un sitio web para el acceso a la base de datos local, y que éste acceso no se haga solamente desde el software desarrollado hasta ahora.
- Apoyar las labores conjuntas entre los diferentes grupos de investigación de la universidad, e incluso con entidades externas, como una forma de cumplir con la misión de la universidad refiriéndonos al trabajo interdisciplinario y las relaciones con el exterior.

BIBLIOGRAFIA

- [1] Aristizábal Ramírez Diego Luis (2006), *Conceptos Básicos del Procesamiento Digital de Imágenes Usando OrquideaJAI*, Escuela de Física Universidad Nacional de Colombia.
- [2] Barrett, G ad Elmore, D. Cambridge University Press, (1998), *Amino acids and peptides*.
- [3] Calvin Austin; Pawlan Pawlan, (2007) *Advanced Programming For The Java TM 2 Platform*
- [4] Ceballos Sierra Javier, (2006) *Java 2: Interfaces Gráficas y Aplicaciones para Internet*. Número Clasificación: 006.76/C387j, Biblioteca de la Universidad Industrial de Santander.
- [5] Keogh Jim, (2003), *J2EE: Manual de Referencia*, Numero Clasificación: 005.2768/K37j. Biblioteca Universidad Industrial de Santander.
- [6] Niño Jackson David, (2005), *Manejo avanzado de imágenes en java*. Facultad de Ing. Electrónica. Universidad el Bosque.
- [7] Schildt Herbert, (2006), *Fundamentos de Java 3° Edición*. Translate from the third English edition of Java a Beginner's Guide, McGraw-Hill.
- [8] Palos Juan Antonio, (2005), *Gráficos con Java 2D₁*, Formato PDF.
- [9] Palos Juan Antonio, (2004), *Swing y JFC (Java Foundation Classes*, Formato PDF.
- [10] Sillero Antonio, (2004), *Isoelectric point determination of proteins and other macromolecules: Oscillating method*, Facultad de Medicina,

Departamento de Bioquímica, Instituto de Investigaciones Biomédicas Alberto Sols, CSIC/UAM, Madrid, España.

[11] Suarez Leo, (2003), *El paquete java.io. Manejo de las I/O*. Formato PDF.

[12] Tabb David L, (2002), An algorithm for isoelectric point estimation. Formato PDF. Disponible en: <http://fields.scripps.edu/DTASelect/20010710-pl-Algorithm.pdf>

[13] Java Swing, (2004), “Manual de Java Swing” Disponible en: <http://jungla.dit.upm.es/~santiago/docencia/apuntes/Swing/>

[14] Patrick Keegan's Blog. (2008), “A Personal Data Storage Application With Embedded Java DB” Disponible en: <http://weblogs.java.net/blog/pkeegan/>

[15] wiki de guía de tutoriales de algunos lenguajes de programación, (2006), “Chuwiki”. Disponible en:

<http://www.chuidiang.com/chuwiki/index.php?title=Portada>

[16] Curso Interactivo Tratamiento digital de la imagen, (2003), Instalación de JAI. Disponible en: http://campusvirtual.uma.es/tdi/www_netscape/instrucciones.php

[17] Foro Code Makers, (2007), “Graficos y Java 2d”. Disponible en: http://foro.code-makers.es/graficos_y_java_2d-t1374.0.html

[18] MAKE 2D-DB II, Documentación oficial, (2001), Disponible en: <http://www.expasy.org/ch2d/make2ddb/>

[19] Practical PostgreSQL Chapter 12. JDBC, (2005), “Using the PostgreSQL Driver” Disponible en: <http://www.postgresql.org/docs/manuals/>

[20] liNpOx, Programación y algo más en Linux, (2007), “JDBC: Conectar Java con PostgreSQL” Disponible en: <http://linpox.wordpress.com/2008/05/23/jdbc-conectar-java-con-postgresql/>

[21] Métodos en Biología celular, (2006), “ELECTROFORESIS EN GELES DE POLIACRILAMIDA” Disponible en: <http://www.ub.es/biocel/wbc/tecnicas/page.htm>

[22] Tutorial de electroforesis, (2004), “2D GEL ELECTROPHORESIS FOR PROTEOMICS TUTORIAL” Disponible en: http://www.aber.ac.uk/~mpgwww/Proteome/Tut_2D.html

ANEXO A. TABLAS DE PRUEBAS DE ERROR DE RAZE-2D.

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE

Nombre del mapa
analizado:

ARABIDOPSIS THALIANA

Fecha:

22/12/2008

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	P37702	5,26	5,26	0	0
Masa Molecular(Mw[KDa])		62,957	68,57	8,915609067	5,613
Punto Isoeléctrico (pI)	O50008	6,15	6,22	1,138211382	0,07
Masa Molecular(Mw[KDa])		76,088	80,85	6,25854274	4,762
Punto Isoeléctrico (pI)	O03042	5,64	5,64	0	0
Masa Molecular(Mw[KDa])		51,074	51,12	0,090065395	0,046
Punto Isoeléctrico (pI)	P99055	5,13	5,13	0	0
Masa Molecular(Mw[KDa])		43,852	44,29	0,998814193	0,438
Punto Isoeléctrico (pI)	P10896	5,01	5,01	0	0
Masa Molecular(Mw[KDa])		47,034	47,21	0,374197389	0,176
Punto Isoeléctrico (pI)	P27140	5,71	5,71	0	0
Masa Molecular(Mw[KDa])		28,62	28,61	0,034940601	0,01

PROMEDIO

1,484198397

Ea [Pi] 0,011666667

Ea [Mw] 1,840833333

Tabla 3. Error para el mapa ARABIDOPSIS THALIANA

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE

Nombre del mapa
analizado:

DICTYSLUG

Fecha:

22/12/2008

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	P46800	6,88	6,87	0,145348837	0,01
Masa Molecular(Mw[KDa])		30,813	31,024	0,684775906	0,211
Punto Isoeléctrico (pI)	P29447	6,4	6,45	0,78125	0,05
Masa Molecular(Mw[KDa])		10,202	10,272	0,686139973	0,07
Punto Isoeléctrico (pI)	P02887	6,26	6,29	0,479233227	0,03
Masa Molecular(Mw[KDa])		27,201	27,201	0	0
Punto Isoeléctrico (pI)	P16642	4,88	4,88	0	0
Masa Molecular(Mw[KDa])		10,777	10,768	0,083511181	0,009
Punto Isoeléctrico (pI)	P34121	5,06	5,06	0	0
Masa Molecular(Mw[KDa])		14,864	14,797	0,450753498	0,067
Punto Isoeléctrico (pI)	P02887	6,54	6,53	0,152905199	0,01
Masa Molecular(Mw[KDa])		27,139	27,201	0,228453517	0,062

PROMEDIO

0,307697612

Ea [Pi] 0,016666667

Ea [Mw] 0,069833333

Tabla 4. Error para el mapa DICTYSLUG

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE

Nombre del mapa
analizado:

ECOLI

Fecha:

22/12/2008

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	P0A7F3	6,84	6,9	0,877192982	0,06
Masa Molecular(Mw[KDa])		16,989	15,532	8,576137501	1,457
Punto Isoeléctrico (pI)	P0A8M0	5,64	5,66	0,354609929	0,02
Masa Molecular(Mw[KDa])		92,858	96,047	3,434275991	3,189
Punto Isoeléctrico (pI)	P63284	5,3	5,3	0	0
Masa Molecular(Mw[KDa])		88,421	91,393	3,361192477	2,972
Punto Isoeléctrico (pI)	P0A786	6,13	6,22	1,468189233	0,09
Masa Molecular(Mw[KDa])		35,391	35,256	0,381452912	0,135
Punto Isoeléctrico (pI)	P0A6P9	5,34	5,34	0	0
Masa Molecular(Mw[KDa])		46,601	46,511	0,193128903	0,09
Punto Isoeléctrico (pI)	P0A6P1	5,16	5,15	0,19379845	0,01
Masa Molecular(Mw[KDa])		33,686	33,415	0,804488512	0,271

PROMEDIO

1,637038907

Ea [Pi]

0,03

Ea [Mw]

1,352333333

Tabla 5. Error para el mapa ECOLI

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE

Nombre del mapa
analizado:

BAT-MOUSE

Fecha:

22/12/2008

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	P14211	4,29	4,31	0,466200466	0,02
Masa Molecular(Mw[KDa])		54,136	54,295	0,293704744	0,159
Punto Isoeléctrico (pI)	P07724	5,58	5,58	0	0
Masa Molecular(Mw[KDa])		62,343	62,842	0,800410632	0,499
Punto Isoeléctrico (pI)	P04117	8,66	8,67	0,115473441	0,01
Masa Molecular(Mw[KDa])		12,742	12,72	0,172657354	0,022
Punto Isoeléctrico (pI)	P02088	7,61	7,59	0,262812089	0,02
Masa Molecular(Mw[KDa])		12,093	12,076	0,140577193	0,017
Punto Isoeléctrico (pI)	P12787	5,08	5,05	0,590551181	0,03
Masa Molecular(Mw[KDa])		12,037	12,01	0,224308382	0,027
Punto Isoeléctrico (pI)	P63017	4,99	4,98	0,200400802	0,01
Masa Molecular(Mw[KDa])		38,586	39,359	2,003317265	0,773

PROMEDIO

0,439201129

Ea [Pi] 0,015

Ea [Mw] 0,2495

Tabla 6. Error para el mapa BAT-MOUSE

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE

Nombre del mapa
analizado:

YEAST

Fecha:

22/12/2008

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	P00330	6,6	6,57	0,454545455	0,03
Masa Molecular(Mw[KDa])		41,097	41,215	0,287125581	0,118
Punto Isoeléctrico (pI)	P00358	6,65	6,63	0,30075188	0,02
Masa Molecular(Mw[KDa])		32,035	32,158	0,383955049	0,123
Punto Isoeléctrico (pI)	P06169	5,9	5,9	0	0
Masa Molecular(Mw[KDa])		58,619	58,655	0,061413535	0,036
Punto Isoeléctrico (pI)	P00950	8,73	8,82	1,030927835	0,09
Masa Molecular(Mw[KDa])		25,384	25,175	0,823353293	0,209
Punto Isoeléctrico (pI)	P00330	6,69	6,68	0,149476831	0,01
Masa Molecular(Mw[KDa])		40,924	41,058	0,327436223	0,134
Punto Isoeléctrico (pI)	P14540	5,75	5,74	0,173913043	0,01
Masa Molecular(Mw[KDa])		36,899	37,034	0,365863574	0,135

PROMEDIO

0,363230192

Ea [Pi] 0,026666667

Ea [Mw] 0,125833333

Tabla 7. Error para el mapa YEAST

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE

Nombre del mapa
analizado:

KIDNEY

Fecha:

22 DE DICIEMBRE DE 2008

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	p06576	4,99	4,97	0,400801603	0,02
Masa Molecular(Mw[KDa])		48,24	48,22	0,04145937	0,02
Punto Isoeléctrico (pI)	P60709	5,26	5,18	1,520912548	0,08
Masa Molecular(Mw[KDa])		41,60	42,35	1,802884615	0,75
Punto Isoeléctrico (pI)	P30101	5,37	5,39	0,372439479	0,02
Masa Molecular(Mw[KDa])		52,20	52,32	0,226044979	0,12
Punto Isoeléctrico (pI)	P04179	6,86	6,87	0,145772595	0,01
Masa Molecular(Mw[KDa])		22,28	21,82	2,064631957	0,46
Punto Isoeléctrico (pI)	P00367	6,81	6,81	0	0,00
Masa Molecular(Mw[KDa])		48,50	48,45	0,103092784	0,05
Punto Isoeléctrico (pI)	P22352	5,36	5,31	0,932835821	0,05
Masa Molecular(Mw[KDa])		22,79	22,51	1,211270078	0,28
PROMEDIO				0,551384114	
				Ea [Pi]	0,03
				Ea [Mw]	0,279

Tabla 8. Error para el mapa HUMAN-KIDNEY

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE

Nombre del mapa
analizado:

STAPHYLOCOCCUS

Fecha:

22 DE DICIEMBRE DE 2008

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO (%)	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	P99157	4,990	4,990	0,000	0,000
Masa Molecular(Mw[KDa])		19,702	19,719	0,086	0,017
Punto Isoeléctrico (pI)	P99128	5,760	5,760	0,000	0,000
Masa Molecular(Mw[KDa])		61,744	62,059	0,510	0,315
Punto Isoeléctrico (pI)	P99083	4,780	4,500	5,858	0,280
Masa Molecular(Mw[KDa])		67,907	68,110	0,299	0,203
Punto Isoeléctrico (pI)	P99115	6,120	6,140	0,327	0,020
Masa Molecular(Mw[KDa])		57,996	57,637	0,619	0,359
Punto Isoeléctrico (pI)	P99098	5,010	5,010	0,000	0,000
Masa Molecular(Mw[KDa])		22,611	22,594	0,075	0,017
Punto Isoeléctrico (pI)	P99070	5,460	5,460	0,000	0,000
Masa Molecular(Mw[KDa])		32,616	32,560	0,172	0,056
PROMEDIO				0,497	
				Ea [Pi]	0,05
				Ea [Mw]	0,16

Tabla 9. Error para el mapa STAPHYLOCOCCUS

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE

Nombre del mapa
analizado:

MUSCLE-MOUSE

Fecha:

22 DE DICIEMBRE DE 2008

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	P70402	5,760	5,760	0,000	0,000
Masa Molecular(Mw[KDa])		63,521	64,594	1,689	1,073
Punto Isoeléctrico (pI)	P99041	5,270	5,240	0,569	0,030
Masa Molecular(Mw[KDa])		42,937	43,279	0,797	0,342
Punto Isoeléctrico (pI)	P02088	7,200	7,190	0,139	0,010
Masa Molecular(Mw[KDa])		11,731	11,875	1,228	0,144
Punto Isoeléctrico (pI)	P05977	5,040	5,040	0,000	0,000
Masa Molecular(Mw[KDa])		23,161	23,015	0,630	0,146
Punto Isoeléctrico (pI)	P58771	4,810	4,790	0,416	0,020
Masa Molecular(Mw[KDa])		33,858	33,803	0,162	0,055
Punto Isoeléctrico (pI)	P32848	4,840	4,840	0,000	0,000
Masa Molecular(Mw[KDa])		12,935	12,999	0,495	0,064
PROMEDIO				0,383	
				Ea [Pi]	0,01
				Ea [Mw]	0,304

Tabla 10. Error para el mapa MUSCLE-MOUSE

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE

Nombre del mapa
analizado:

LYMPHOCYTE

Fecha:

22 DE DICIEMBRE DE 2008

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	P52566	5,030	5,030	0,000	0,000
Masa Molecular(Mw[KDa])		23,952	24,013	0,255	0,061
Punto Isoeléctrico (pI)	P62937	7,820	7,830	0,128	0,010
Masa Molecular(Mw[KDa])		17,900	17,901	0,006	0,001
Punto Isoeléctrico (pI)	P07237	4,650	4,660	0,215	0,010
Masa Molecular(Mw[KDa])		43,614	43,657	0,099	0,043
Punto Isoeléctrico (pI)	P04075	7,890	7,890	0,000	0,000
Masa Molecular(Mw[KDa])		37,732	37,724	0,021	0,008
Punto Isoeléctrico (pI)	P06733	7,370	7,360	0,136	0,010
Masa Molecular(Mw[KDa])		42,232	42,249	0,040	0,017
Punto Isoeléctrico (pI)	P22626	7,860	7,860	0,000	0,000
Masa Molecular(Mw[KDa])		34,149	34,075	0,217	0,074
PROMEDIO				0,070	
				Ea [Pi]	0,005
				Ea [Mw]	0,034

Tabla 11. Error para el mapa LYMPHOCYTE-HUMAN

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE

Nombre del mapa
analizado:

LIVER

Fecha:

22 DE DICIEMBRE DE 2008

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	P31929	6,190	6,180	0,162	0,010
Masa Molecular(Mw[KDa])		145,434	145,186	0,171	0,248
Punto Isoeléctrico (pI)	p02768	5,800	5,790	0,172	0,010
Masa Molecular(Mw[KDa])		66,471	70,979	6,782	4,508
Punto Isoeléctrico (pI)	P30049	4,500	4,510	0,222	0,010
Masa Molecular(Mw[KDa])		13,244	13,216	0,211	0,028
Punto Isoeléctrico (pI)	P02766	5,380	5,430	0,929	0,050
Masa Molecular(Mw[KDa])		13,732	13,708	0,175	0,024
Punto Isoeléctrico (pI)	P68871	6,850	6,950	1,460	0,100
Masa Molecular(Mw[KDa])		12,558	12,464	0,749	0,094
Punto Isoeléctrico (pI)	P00738	6,300	6,330	0,476	0,030
Masa Molecular(Mw[KDa])		15,953	15,891	0,389	0,062
PROMEDIO				0,744	
				Ea [Pi]	0,035
				Ea [Mw]	2,822666667

Tabla 12. Error para el mapa HUMAN-LIVER

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE (PDQUEST ADVANCED)

Nombre del mapa
analizado:

EPIMASTIGOTES

Fecha:

05 DE ENERO DE 2009

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO(%)	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	1	5,900	5,950	0,847	0,050
Masa Molecular(Mw[KDa])		76,200	77,815	2,119	1,615
Punto Isoeléctrico (pI)	2	5,700	5,690	0,175	0,010
Masa Molecular(Mw[KDa])		72,100	72,298	0,275	0,198
Punto Isoeléctrico (pI)	3	5,500	5,520	0,364	0,020
Masa Molecular(Mw[KDa])		67,700	68,464	1,129	0,764
Punto Isoeléctrico (pI)	4	5,500	5,500	0,000	0,000
Masa Molecular(Mw[KDa])		59,900	60,291	0,653	0,391
Punto Isoeléctrico (pI)	5	5,500	5,480	0,364	0,020
Masa Molecular(Mw[KDa])		63,600	63,667	0,105	0,067
Punto Isoeléctrico (pI)	6	5,400	5,470	1,296	0,070
Masa Molecular(Mw[KDa])		67,500	68,776	1,890	1,276
PROMEDIO				0,576	
				Ea [Pi]	0,028333333
				Ea [Mw]	0,7185

Tabla 13. Error para el mapa EPIMASTIGOTES con PDQUEST ADVANCED

FORMATO DE EVALUACION DE LOS PARAMETROS 2D-PAGE (PDQUEST ADVANCED)

Nombre del mapa
analizado:

EPIMASTIGOTES 338

Fecha:

05 DE ENERO DE 2009

NOMBRE DEL PARAMETRO	PROTEINA ANALIZADA	BASE DE DATOS SWISS 2D-PAGE	RAZE-2D	ERROR RELATIVO(%)	ERROR ABSOLUTO
Punto Isoeléctrico (pI)	1	5,400	5,450	0,926	0,050
Masa Molecular(Mw[KDa])		45,000	44,890	0,244	0,110
Punto Isoeléctrico (pI)	2	5,400	5,390	0,185	0,010
Masa Molecular(Mw[KDa])		46,100	45,450	1,410	0,650
Punto Isoeléctrico (pI)	3	4,900	4,970	1,429	0,070
Masa Molecular(Mw[KDa])		50,100	49,560	1,078	0,540
Punto Isoeléctrico (pI)	4	5,000	5,000	0,000	0,000
Masa Molecular(Mw[KDa])		50,000	49,560	0,880	0,440
Punto Isoeléctrico (pI)	5	5,000	5,030	0,600	0,030
Masa Molecular(Mw[KDa])		43,400	43,250	0,346	0,150
Punto Isoeléctrico (pI)	6	5,000	5,070	1,400	0,070
Masa Molecular(Mw[KDa])		44,100	43,795	0,692	0,305
PROMEDIO				0,574	
				Ea [Pi]	0,038333333
				Ea [Mw]	0,365833333

Tabla 14. Error para el mapa EPIMASTIGOTES 338 con PDQUEST
ADVANCED

ANEXO B. CASOS DE USO

DESCRIPCION DE LOS CASOS DE USO

ADMINISTRADOR

- **Preparar Datos:** El administrador prepara los datos de acuerdo a una de las tres formas existentes en preparar los datos según el SIB.
- **Configurar Datos:** El administrador configura los datos de acuerdo a la herramienta Make2D-DB II.
- **Ingresar al sistema:** El administrador puede ingresar al servidor de base de datos por medio de un usuario y contraseña.
- **Agregar usuario:** El administrador agrega un usuario nuevo para que pueda realizar la consulta a la base de datos.
- **Actualización registros:** El administrador restaura los registros de la base de datos basados en los nuevos registros proporcionados por el SIB (Instituto Suizo De Bioinformática).

USUARIO INVESTIGADOR

CASO DE USO: ANÁLISIS DEL GEL 2D.

- **Ingresar a la herramienta:** Cualquier usuario puede ingresar a la herramienta ya que es de carácter libre y es el único rol que necesita la herramienta sin necesidad de contraseña alguna.
- **Abrir Gel:** El usuario investigador abre una imagen resultante del proceso de electroforesis para su posterior análisis.
- **Ingresar Datos Iniciales:** El usuario ingresa los datos iniciales correspondientes al gel que se desea analizar y el respectivo número de marcadores.

- **Ingresar Marcadores:** El usuario ingresa los respectivos valores de punto isoeléctrico y masa molecular de las proteínas conocidas (marcadores).
- **Generar escalas:** Genera la escala bidimensional (punto isoeléctrico y masa molecular) correspondiente al gel.
- **Cambiar Visualización Gel:** El usuario cambia la visualización de la imagen del Gel en las opciones disponibles por la herramienta (Escala de Grises, Negativo).
- **Filtración de la Imagen:** El usuario filtra la imagen para eliminar ruido presente en la imagen.
- **Ingresar Spots:** El usuario hace clic en el gel donde se desee conocer su respectivo valor de punto isoeléctrico y masa molecular.
- **Analizar Spot:** El usuario encuentra el porcentaje de similitud que tiene el spot señalado con alguna proteína de la base de datos Swiss2DPage.
- **Guardar Información Spot:** El usuario guarda en un archivo de texto los valores de punto isoeléctrico y masa molecular generados para cada spot.
- **Consultar Proteína Encontrada:** El usuario consulta sobre la proteína más cercana al spot señalado toda la información relacionada a dicha proteína.
- **Guardar Información Proteína:** El usuario guarda la información de las proteínas más cercanas correspondiente al spot analizado.
- **Guardar Gel:** El usuario guarda la imagen del gel con cada uno de sus spots señalados en dicho gel.

CASO DE USO: CONSULTAS EN LAS BASE DE DATOS.

- **Consultar Base Datos Swiss2DPage:** El usuario consulta las proteínas existentes en los 36 mapas (geles) de Swiss2DPage.

- **Consultar Base Datos Uniprot:** El usuario consulta la información relacionada a cualquier proteína que desee conocer de la base de datos Uniprot.
-

DIAGRAMAS DE CASOS DE USO

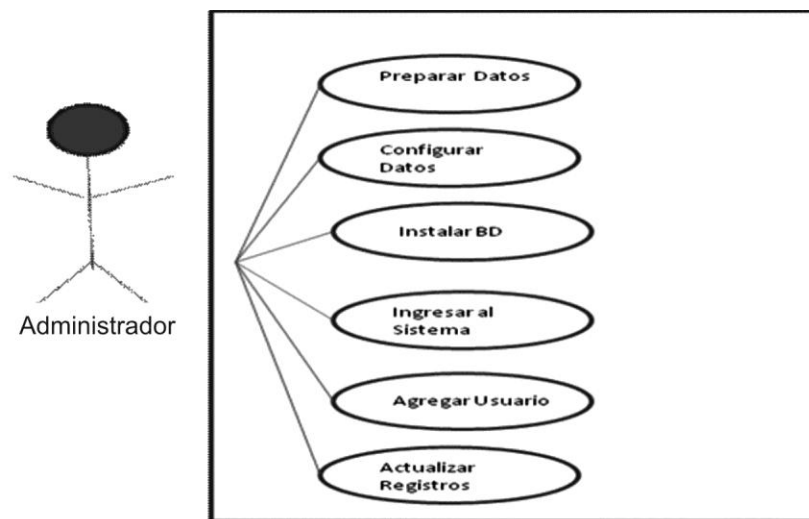


FIGURA 29. Diagrama de casos de uso del Administrador

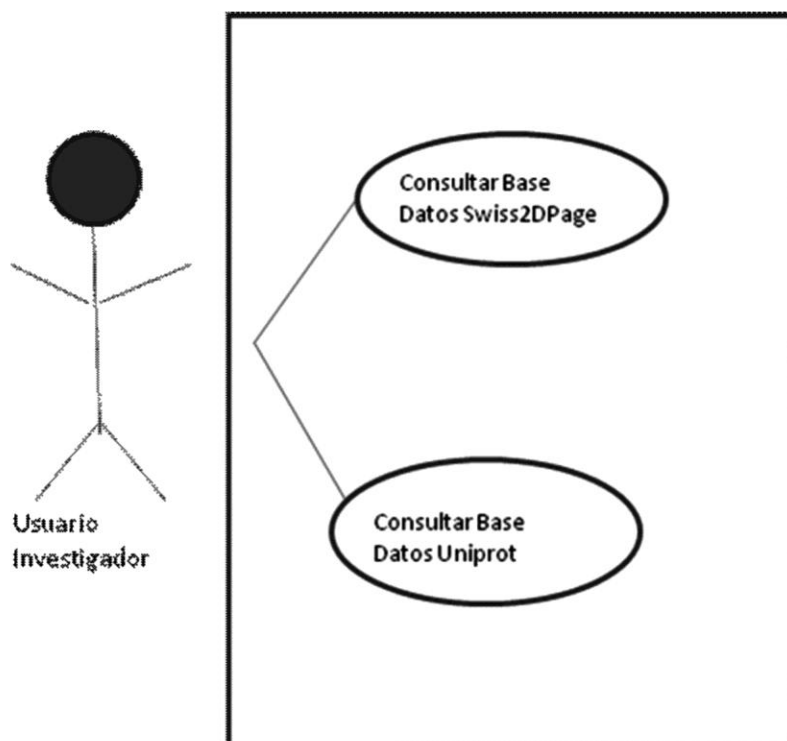


FIGURA 30. Diagrama de caso de uso consulta en la base de datos Usuario Investigador

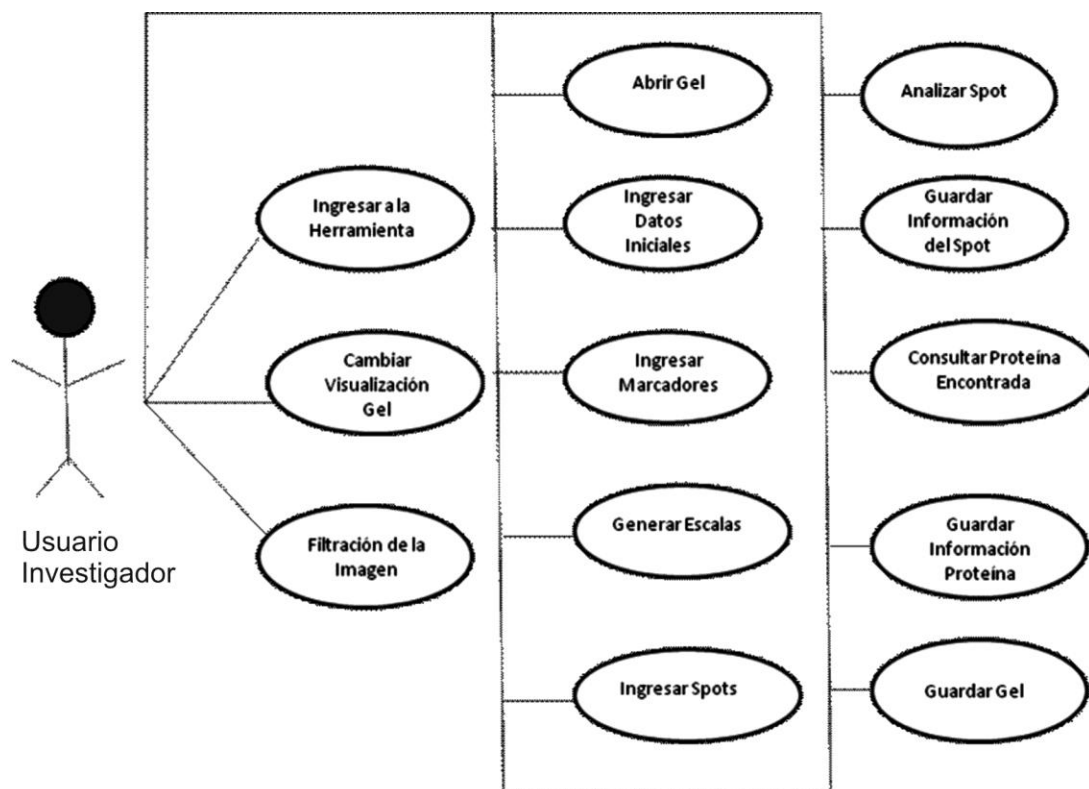


FIGURA 31. Diagrama de caso de uso análisis del gel 2D de Usuario Investigador

ANEXO C. GENERACIÓN DE LA BASE DE DATOS SWISS2DPAGE

Fase de Construcción

La última versión de la herramienta Make2D-DBII, modificada por Khaled Mostaguir para aspirar al grado de Doctor de Ciencias en la Universidad de Génova²¹ está hecha en el lenguaje perl junto con los datos proteómicos suministrados ayuda a generar un servidor web y una base de datos relacional con dichos datos.

Gracias a la herramienta Make2D-DBII se genera la base de datos Swiss2DPAGE en el sistema gestor de base de datos PostgreSQL él cual consta de cuatro esquemas en el que constituye la información total de la base de datos. Los cuatro esquemas son:

- **Core:** Este es el esquema central que contiene todos los datos. Posee todas las relaciones y refleja el modelo de datos casi en su totalidad. Este esquema es autosuficiente ya que posee todos los datos y las operaciones requeridas para el comportamiento de la base de datos. Contiene datos privados que solo el administrador y los usuarios privilegiados pueden modificar.
- **Public:** La estructura del esquema público es un reflejo del esquema Core (todos los índices y relaciones), pero sin las operaciones, los procedimientos, las asociaciones y las limitaciones del esquema Core. Un usuario normal sólo puede acceder a los datos del esquema public.
- **Common:** Este esquema contiene unos procedimientos comunes para los anteriores esquemas. El esquema Core y Common constituyen en

²¹ Documento PhD de Khaled Mostaguir. <http://mordor.isb-sib.ch/make2ddb/readme/Thesis%20-%20Make2D-DB-II-%20Khaled%20Mostaguir.pdf>

conjunto la totalidad del modelo de datos y por lo tanto son autosuficientes. El esquema *Common* tiene una sola tabla, con una única tupla, que se relaciona con la base de datos de identidad y estado actual.

- **Log:** La función de este esquema es la de registrar cualquier modificación de datos aplicado al esquema *Core*. Cualquier tupla (registro) borrado o modificado se reporta aquí. El esquema *Log* contiene una copia de todas las relaciones del esquema *Core*. Cada tabla tiene tres adicionales atributos: *modificationDate* enunciando cuando ha ocurrido la modificación, *updateOrDeleted* indicando cuando los datos han sido actualizados o eliminados y *userModifiedName* indicando el nombre del usuario que hizo la modificación. Como el esquema *public* no tiene ni asociaciones y operaciones.

Dirección URL para acceder a los distintos esquemas implementados

❖ El esquema core:

http://world-2dpage.expasy.org/make2ddb/database_schema/core_schema.html

❖ El esquema public:

http://world-2dpage.expasy.org/make2ddb/database_schema/public_schema.html

❖ El esquema common:

http://world-2dpage.expasy.org/make2ddb/database_schema/common_schema.html

❖ **El esquema log:**

http://world-2dpage.expasy.org/make2ddb/database_schema/log_schema.html

❖ **Todos los esquemas:**

http://world-2dpage.expasy.org/make2ddb/database_schema/all.html

Tablas y campos de la base de datos

Como ya se ha explicado anteriormente, el esquema *core* es el que contiene la totalidad de la información. El esquema posee 27 tablas, algunas relacionadas con la información proteómica y otras tablas relacionadas con las rutas de imágenes, autores, e información relacionada con el servidor web. Para las consultas que realiza la herramienta fueron necesarias las siguientes tablas:

spot

spot		
Columna	Tipo de Dato	No Nulo
spotid	character varying (16)	NOT NULL
gelid	integer	NOT NULL
mw	integer	NOT NULL
pi	numeric(4,2)	
xcoordinate	integer	
ycoordinate	integer	
odrelative	real	
volumerealtive	real	
userstamp	character varying (128)	
update	timestamp without time zone	

Tabla 15. Spot de la Base de Datos Swiss2DPAGE.

spotentry

spotentry		
Columna	Tipo de Dato	No Nulo
spotid	character varying (16)	
gelid	integer	
ac	character varying (32)	NOT NULL
fragment	boolean	
showflag	boolean	NOT NULL
showflagswitch	boolean	NOT NULL
userstamp	character varying (128)	
update	timestamp without time zone	

Tabla 16. Spotentry de la Base de Datos Swiss2DPage.

entry

entry		
Columna	Tipo de Dato	No Nulo
ac	character varying (32)	NOT NULL
id	character varying (32)	NOT NULL
entryclass	character varying (32)	
identmethod	character varying (16)	
description	text	
releasecreation	integer	
subreleasecreation	integer	
genenames	character varying (1024)	
keywords	character varying (1024)	
organismid	integer	
organismstrain	character varying (256)	
entrychecksum	Character (16)	
showflag	boolean	
showflagswitch	boolean	
userstamp	character varying (128)	
update	timestamp without time zone	

Tabla 17. Entry de la Base de Datos Swiss2DPage

gel

gel		
Columna	Tipo de Dato	No Nulo
gelid	integer	NOT NULL
melaniegeneratedid	character varying (64)	
shortname	character varying (32)	NOT NULL
fullname	character varying (256)	
dimension	smallint	
startpi	numeric(4,2)	
endpi	numeric(4,2)	
startmw	bigint	
endmw	bigint	
organismid	integer	
organismstrain	character varying (256)	
gelparentid	integer	
gelpreparationid	integer	
gelinformaticsid	integer	
bisourceinformaticsid	integer	
showflag	boolean	
showflagswitch	boolean	
userstamp	character varying (128)	
update	timestamp without time zone	

Tabla 18. Gel de la Base de Datos Swiss2DPAGE

Diagrama de Entidad-Relación

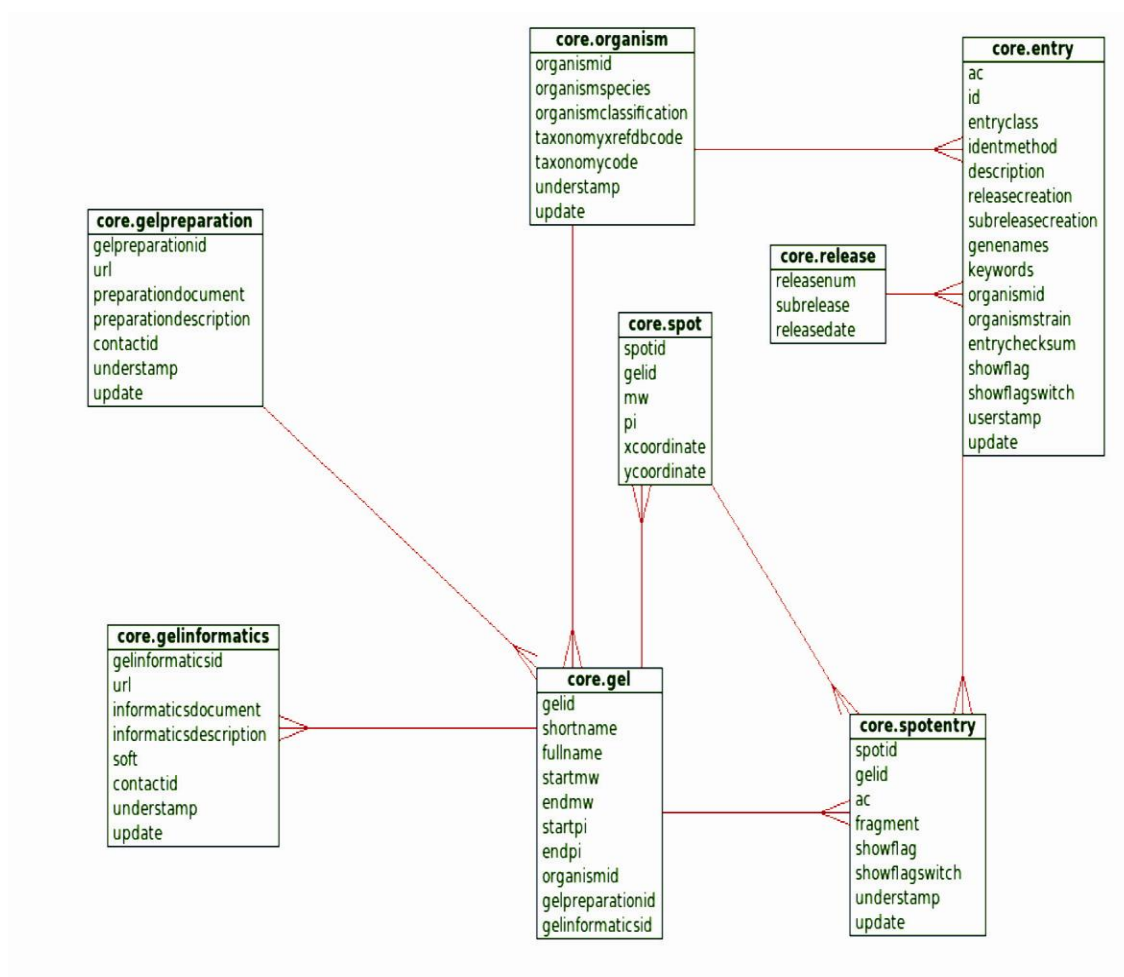


FIGURA 32. Diagrama de Entidad-Relación de las tablas utilizadas de la Base de Datos Swiss2DPAGE

ANEXO D. MANUAL DE INSTALACIÓN DE LA BASE DE DATOS CON MAKE 2D-DB II

Texto basado en la documentación oficial de la herramienta en el sitio web del servidor Expasy:

http://www.expasy.org/ch2d/make2ddb/1.Readme_main.html.

La herramienta MAKE 2D-DB II está diseñada para funcionar en sistemas operativos basados en UNIX (GNU-Linux) para la instalación y montaje de bases de datos, y para su correcta utilización es necesario instalar el lenguaje PERL y unos módulos importantes para la interacción del lenguaje con las bases de datos y otros componentes. Es muy importante que antes de comenzar el proceso, se tenga en cuenta que la instalación de muchos de los paquetes y módulos se deben hacer desde la cuenta de administrador del sistema.

Antes de instalar los módulos de Perl es necesario instalar el POSTGRESQL, esto se puede hacer a través de los gestores de paquetes (Synaptic, aptitude, apt-get, yum, etc.), o visitando su sitio oficial de descargas en <http://www.postgresql.org/download/>.

Con la instalación del POSTGRESQL es necesario generar unas variables de entorno (que se pueden colocar en el archivo /etc/environment) que apunten a la ubicación de las librerías de postgresql y la carpeta donde se ha de crear la base de datos, así:

```
POSTGRES_LIB="/usr/lib/postgresql/8.3/lib"
```

PGLIB="/usr/lib/postgresql/8.3/lib"

PGDATA="/usr/local/pgsql/data"*

*Nota: es necesario crear la carpeta pgsql/data *(se elige como ruta para la instalación de la base de datos) y cambiar el dueño de la carpeta al usuario de la base de datos, a través del comando chown.*

Agregar al PATH=/usr/lib/postgresql/8.3/bin

La instalación de PERL se puede hacer mediante los mismos gestores de paquetes en cualquier distribución de Linux (y por lo general viene incluido en la instalación básica).

Desde el mismo gestor de paquetes se pueden instalar los siguientes componentes:

libwww-perl

libxml-perl

libncurses5

libncurses5-dev

libreadline5

libreadline5-dev

libreadline5-dbg

tcl

readline-5.2

Después de la instalación de PERL se puede utilizar la consola para la descarga e instalación de los módulos, ejecutando la siguiente línea en una terminal:

```
perl -MCPAN -e shell
```

El prompt del terminal cambia a una sesión en PERL. El primer módulo a instalar es el de la actualización de los componentes existentes para el manejo de los paquetes en perl mediante CPAN, así:

```
install Bundle::CPAN
```

Como todas las instalaciones en Linux, es importante contar con una conexión a internet para la descarga de los paquetes desde los repositorios, en este caso los de CPAN. Por lo general la herramienta hace algunas preguntas acerca de instalación de otros paquetes faltantes, y es importante permitirle a la herramienta hacer estas descargas e instalaciones.

Después de la actualización de los paquetes “básicos” de CPAN, se puede proceder a la instalación de los módulos necesarios para el funcionamiento de MAKE 2D-DB II.

```
install DBI
```

Si por alguna circunstancia no se ha podido instalar el módulo libwww-perl y el libxml-perl desde el gestor de paquetes, se pueden instalar desde la consola de PERL así:

```
install XML::Parser::PerlSAX
```

```
install XML::Twig
```

Si las variables de entorno quedaron bien definidas, se puede proceder a instalar el módulo para la generación de la base de datos:

```
install DBD::Pg
```

```
install Compress::Zlib
```

Ahora se debe inicializar el clúster de la base de datos (si se coloca bien las variables de entorno al inicio). Es importante tener presente que la cuenta de postgres que inicialice el clúster de la base de datos se considerará la cuenta dueña de la base de datos a generar, y las operaciones siguientes se harán desde la misma cuenta.

Si la cuenta que ha de ser la dueña de la base de datos no es *postgres*, se debe crear una nueva cuenta desde esta sesión de usuario, mediante la siguiente sentencia SQL:

```
CREATE USER nombre_del_usuario WITH PASSWORD 'password_usuario'  
CREATEDB CREATEUSER
```

Esta operación no se debe realizar como root (Administrador), hay que hacerla como un usuario sin privilegios especiales, por ejemplo: postgres, que es la cuenta que se genera por defecto en la instalación del PostgreSQL.

```
/usr/lib/postgresql/8.3/bin/initdb
```

En la carpeta donde se ha de generar la base de datos (la que aparece en el PGDATA) se encuentra un archivo llamado `pg_hba.conf` el cual hay que modificar en las líneas des-comentadas para que quede de la siguiente forma:

```
local all all trust
host all all 127.0.0.1/32 trust
```

Seguido de esto se procede a arrancar el servidor de bases de datos:

```
nohup /usr/lib/postgresql/bin/postmaster -i
      /usr/lib/postgresql/8.3/bin/createlang plpgsql template1
```

A continuación se puede proceder al uso de la herramienta, y se consigue desde el sitio de descarga oficial en <http://www.expasy.org/ch2d/make2ddb.html>.

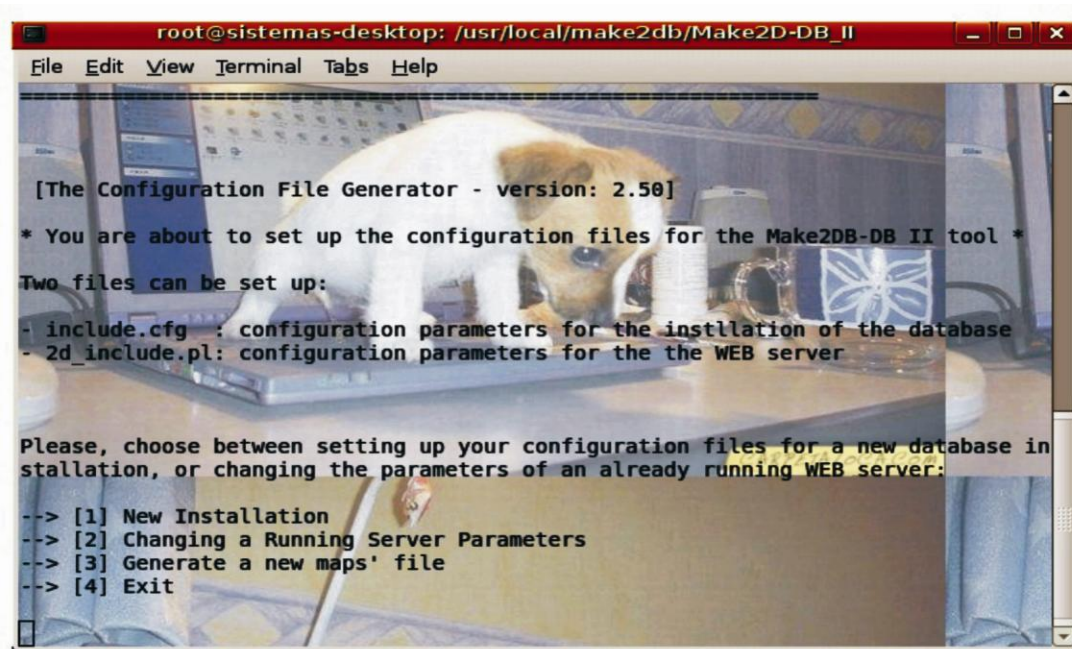


FIGURA 33. Instalando MAKE2D-DB II

Luego de adquirir la herramienta y descomprimirla en el lugar deseado, se procede a su ejecución, desde la consola de terminal:

```
perl make2db.pl -m config
```

Aparece entonces una interfaz como la mostrada en la FIGURA 33, la herramienta es muy intuitiva, pero requiere que el usuario esté muy consciente de las características de la base de datos a generar, y solicita datos sobre el tipo de fuente a utilizar, ya sea un archivo de texto plano, spreadsheets o reportes XML; y las imágenes de los geles. De la misma manera la herramienta solicita datos como la cuenta de postgres que ha de ser dueña de la base de datos, el formato de los archivos de imagen, etc.

Cuando se le han indicados todos los parámetros necesarios a la herramienta, se procede a la revisión de los archivos utilizados como fuente para la generación de la base de datos:

```
perl make2db.pl -m check
```

La herramienta hace la revisión de la consistencia de la información suministrada por el usuario y si no encuentra ningún problema, se puede proceder ya a la generación de la base de datos, así:

```
perl make2db.pl -m transform
```

De esta manera se procede a la generación de una base de datos orientada a objetos con la herramienta MAKE 2D-DB II.

ANEXO E. MANUAL DE INSTALACIÓN DE LA HERRAMIENTA RAZE-2D

RAZE-2D es una herramienta desarrollada en JAVA, por tanto es indispensable que se tenga correctamente instalada y configurada la máquina virtual de JAVA.

El sitio de descarga oficial de Sun Microsystems es:

<http://www.java.com/es/download/manual.jsp>

El proceso de instalación puede variar de un sistema operativo a otro, sin embargo en el sitio oficial existe mucha información al respecto, y en diferentes idiomas en:

<http://www.java.com/es/download/help/5000010500.xml#selfextracting>

En las diferentes distribuciones de Linux, existen varias formas de hacer la instalación del JRE (Java Runtime Environment), como son, adquirir los archivos autoextraíbles (RPM) o los binarios autoextraíbles (.bin). Hay que tener en cuenta que la tarea de instalación es recomendable realizarla desde la cuenta de administrador del sistema (root), y hacerla en una ruta que sea accesible por todos los usuarios para que todas las otras cuentas tengan acceso y puedan usar el JRE.

En cualquiera de los dos casos el archivo descargado tiene una extensión .bin. Es necesario cambiarle los permisos al archivo descargado y añadirle permisos de ejecución, con el comando:

`chmod a+x jre-6u<versión>-linux-i586.bin (autoextraíble)`

`chmod a+x jre-6u<versión>-linux-i586-rpm.bin (autoextraíble RPM)`

Este comando se ejecuta desde una consola de terminal, y se recomienda que sea realizado por el usuario administrador (root).

Ahora que se tienen permisos de ejecución para los archivos descargados, se puede proceder en su instalación, desde la consola de terminal:

```
./jre-6u<versión>-linux-i586.bin
```

```
./jre-6u<versión>-linux-i586-rpm.bin
```

En el caso de los rpm, es necesario un paso adicional para la instalación:

```
rpm -iv jre-6u<versión>-linux-i586.rpm
```

Para Windows el proceso es muy distinto, aunque existe un instalador para descargar, y es muy intuitivo, no requiere ninguna experiencia ni conocimiento en particular sobre la instalación de programas.



FIGURA 34. Imagen del instalador de Java

Al ejecutar el instalador de Java en Windows aparece un acuerdo de licencia (ver FIGURA 34) que el usuario debe aceptar para poder continuar con el proceso.



FIGURA 35. Opciones de la instalación

Las opciones de instalación, presentadas en la FIGURA 35, hacen referencia a otros productos que la empresa Sun Microsystems ofrece junto con la instalación de Java, y que son complementos para el buscador Google que se pueden añadir como barra de herramienta o como aplicación en la barra de tareas de Windows. Al terminar la instalación aparece una ventana de agradecimiento por la adquisición del producto por parte de Sun Microsystems (FIGURA 36).



FIGURA 36. Finalizando la instalación

Java está disponible para varios sistemas operativos: Windows, Mac, GNU-Linux y Solaris. Gracias a ello se ha convertido en la mejor opción a tomar por los desarrolladores de software, ya que permite que los productos generados sean muy portables.

Después de instalar JAVA, es necesario instalar las librerías de manejo de imágenes que requiere RAZE-2D para funcionar (JAI: Java Advanced Imaging). Estas librerías tienen sus propios instaladores y se distribuyen con la herramienta. Sin embargo, en el caso de Linux, el proceso es un poco más largo, y se explica a continuación.

Los binarios para la instalación están disponibles para descargar en

<https://jai.dev.java.net/binary-builds.html>

Nota: el procedimiento mencionado a continuación es recomendable realizarlo como usuario administrador del sistema (root).

Cuando se tiene instalado el JRE de java, los archivos quedan guardados en una ruta específica, a menos que la instalación se haya hecho a mano y el usuario haya escogido alguna otra ruta, esta ruta es:

`/usr/lib/jvm/java-6-sun/jre`

Ahora hay que ubicarse en la ruta mencionada, y desde ahí ejecutar los siguientes comandos:

`chmod u + x $downloaddir/jai-1_1_3-lib-linux- $ARCH-jre.bin`

Esto significa que hay que colocarle permisos de ejecución al archivo .bin que se encuentra en la carpeta descargada en la dirección \$downloaddir y el archivo \$ARCH es el que corresponde al archivo de la versión de JAI que se está instalando (la descargada).

Después hay que ejecutar el archivo .bin desde la ubicación especificada anteriormente.

`$downloaddir/jai-1_1_3-lib-linux-% $ $ARCH-jre.bin`

ANEXO F. EVALUACIÓN DE LA SATISFACCIÓN DEL USUARIO

FORMATO DE EVALUACIÓN PARA LOS PARÁMETROS 2D-PAGE

Experto Evaluador: Dr. Rodrigo Torres Sáez

Parámetro 2D-PAGE	Considera que el cálculo del parámetro es útil	
	SI	NO
Punto isoeléctrico calculado	X	
Masa molecular calculada	X	
Lista de coincidencias en la base de datos por mapa	X	
Lista de coincidencias en la base de datos en todos los mapas	X	
Porcentaje de similitud	X	

Firma: 

FIGURA 37. Evaluación por parte de un experto

Para la evaluación por parte del usuario final, el formato presentaba una casilla para efectuar una valoración cuantitativa (con 5 como nota máxima).

FORMATO DE EVALUACION DE SATISFACCION DEL USUARIO FINAL

Usuario: Rosío del Pilar Niño Moyano.

REVISION DE LA PRESENTACION DE LA INTERFAZ		
ACTIVIDAD	VALORACION	NOTAS ADICIONALES
1. ¿La interfaz muestra el título correcto?	5	
2. ¿Las etiquetas de los campos corresponden a los datos solicitados o suministrados?	5	
3. ¿El tamaño de los campos es el adecuado?	5	
4. ¿Están los controles habilitados correctamente?	5	
5. ¿Los colores presentados en la interfaz son agradables a la vista?	5	
6. ¿La interfaz tiene la dimensión adecuada?	5	
7. ¿Los campos están distribuidos de forma correcta?	5	
REVISION DE FUNCIONALIDAD		
8. ¿La interfaz lleva a cabo la tarea que se		

FIGURA 38. Evaluación de satisfacción de usuario final (PAGINA 1)

necesita?	5	
9. ¿El usuario recibe la información suficiente de las acciones que realiza en la herramienta software?	5	
10. ¿Los datos mostrados en la interfaz cambian en forma sincronizada?	5	
11. ¿Es rápido y fácil el manejo de la interfaz?	4	
12. ¿Los campos están dispuestos en el orden correcto?	5	
13. ¿Las imágenes son manejadas con facilidad?	5	
REVISION DE LA HERRAMIENTA EN GENERAL		
14. ¿La herramienta software es amigable?	5	
15. ¿La herramienta software no necesita capacitación para poder utilizarla (es intuitiva)?	4	
16. ¿Utilizaría la herramienta en su labor cotidiana?	5	Con seguridad.
17. ¿Recomendaría el uso de esta herramienta?	5	11

FIGURA 39. Evaluación de satisfacción de usuario final (PAGINA 2)

18. ¿Se siente satisfecho con el producto entregado?	5	
COMENTARIOS ADICIONALES: Es un producto necesario para los trabajos que se desarrollan en el GIBIP.		

FIGURA 40. Evaluación de satisfacción de usuario final (PAGINA 3)