ISBN: 978-959-7213-02-4

SLD256 SERVICIO DE ACOPLAMIENTO MOLECULAR EN LA PLATAFORMA DE SERVICIOS BIOINFORMÁTICOS DE LA UCI

SLD256 MOLECULAR DOCKING SERVICE IN THE BIOINFORMATICS PLATFORM SERVICES OF UCI

Osvel Chávez Hernández[1], Longendri Aguilera Mendosa[2], Taymara Hernández Ortega[3]

1 Centro de Tecnolagías y Gestión de Datos, Universidad de las Ciencias Informáticas, Cuba, ochavez@uci.cu, Carretera a San Antonio de los Baños km 2 ½, Torrens, Boyeros, La Habana, Cuba, C.P.:

2 Universidad de las Ciencias Informáticas, Cuba, loge@uci.cu

3 Universidad de las Ciencias Informáticas, Cuba, tay@uci.cu

RESUMEN: *El proceso de identificación de un fármaco eficaz para el tratamiento de una determinada enfermedad tiene un alto costo en tiempo y capital humano, siendo necesario para ello un importante presupuesto.En la actualidad podemos contar con procedimientos computacionales que aminoran en gran medida estos altos costos, pero aún, en nuestro país, no son aprovechados por un gran número de especialistas debido a la limitada capacidad de cómputo que presentan algunos centros de investigación. El principal objetivo en el diseño de fármacos es encontrar aquellas moléculas candidatos que puedan interactuar de manera efectiva en un determinado blanco terapéutico. Para encontrar las conformaciones moleculares que tributen a la obtención de un complejo molecular estable se emplean los llamados métodos de Acoplamiento Molecular. En este trabajo se presenta un Servicio Web de Acoplamiento Molecular, el cual utiliza una Plataforma de Tareas Distribuidas desarrollada y desplegada en los laboratorios docentes de la Universidad de las Ciencias Informáticas (UCI), lográndose una disminución significativa del tiempo de respuesta de los métodos computacionales de Acoplamiento Molecular.*

Palabras Clave:acoplamiento molecular, sistema distribuido, servicios bioinformáticos.

ABSTRACT: *The identification of an effective drug for the treatment of a particular disease is costly in time and human capital, it being necessary for a major budget. At present we have procedures that reduce computational largely these high costs, but still, in our country, are not used by a large number of specialists due to limited computing power, for several research centers. The main objective in the design of drug candidates is to find those molecules that can interact effectively in a specific therapeutic target. To find the molecular conformations taxing to obtain a stable molecular complex called methods are employed molecular docking. This paper presents a Molecular Docking Web Service, which uses a Distributed Task Platform developed and deployed in the teaching laboratories of the University ofInformatics Sciences (UCI), achieving a significant reduction in response time computational methods Molecular Docking.*

KeyWords:molecular docking, distributed system, bioinformatics services.

1. INTRODUCCIÓN

El descubrimiento de un nuevo medicamento y desarrollo posterior del mismo son dos fases, que condicionan lograr un nuevo producto que sea útil en la terapéutica. La fase de descubrimiento se estima que dure aproximadamente unos 4 - 5 años y la fase de desarrollo se estima que dure aproximadamente unos 5 - 6 años más [1].

En la actualidad, gracias a los avances científicos y tecnológicos en el campo de la biología molecular, la ciencia ha podido explicar el funcionamiento de los medicamentos. Algunos han sido descubiertos al azar, como en el caso de la penicilina, o bien mediante el estudio de remedios tradicionales y productos naturales. Lo cierto es que numerosas macromoléculas han sido identificadas que desempeñan funciones vitales en el funcionamiento de la célula y cuya alteración está relacionada con ciertas enfermedades.

Puesto que las proteínas son los componentes macromoleculares más abundantes, no resulta sorprendente que éstas sean las dianas más frecuentes de los fármacos. Razón por lo cual, cuando se conoce la base biológica de una enfermedad o de un desarreglo metabólico y se pueden definir las proteínas que causan tal desajuste, se diseña un fármaco que es una pequeña molécula (llamada ligando) con las propiedades químicas y biológicas necesarias para unirse al sitio activo de esa proteína y moldear su funcionamiento a condiciones favorables. [**Figura 1**]

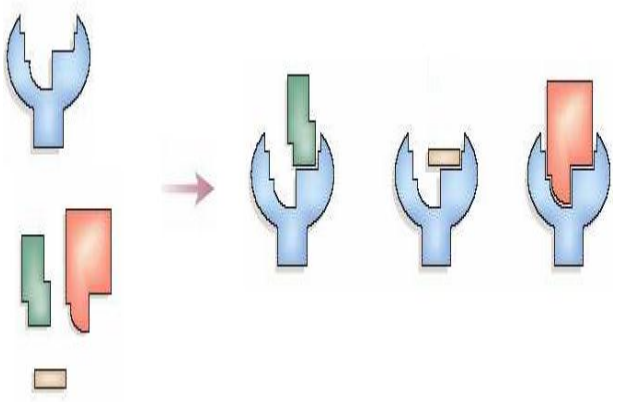

Figura 1Modelo de complementariedad de formas.

La importancia del diseño de fármaco basado en la estructura del receptores poder encontrar las mejores interacciones que ocurren entre los posibles candidatos a fármacos y el sitio de unión de la diana terapéutica. Lo que se traduce en un fármaco más efectivo. Con este propósito, se recurre a uno de los métodos computacionales más usados en este campo: el "acoplamiento molecular" (Molecular Docking en inglés) [2]. Este método consiste en calcular computacionalmente cuál es la posición más favorable que tendría una molécula en el blanco molecular, trabajando con sus representaciones tridimensionales. Es importante destacar el gran número de grados de libertad que pueden tener dos moléculas para unirse, acentuándose lo anterior cuando hablamos de macromoléculas biológicas como son las proteínas, por lo que este método demanda de gran potencia de cómputo.

En la industria farmacéutica es común trabajar con las llamadas "bibliotecas virtuales": colecciones de miles o millones de estructuras moleculares que se pueden sintetizar. Estas bibliotecas se usan en un proceso conocido como "Tamizaje Virtual" (Virtual Screening en inglés), a fin de seleccionar los mejores compuestos capaces de unirse a una macromolécula diana utilizando métodos computacionales. El tiempo en obtener la respuesta de este proceso se puede ver afectado por el tamaño de la biblioteca virtual y por los grandes cálculos que implica realizar simulaciones de acoplamiento molecular. Después de hacer los cálculos, las moléculas virtuales seleccionadas como posibles candidatos a fármacos son sintetizadas y evaluadas de forma experimental. De tal manera que las computadoras y los métodos computacionales ayudan a filtrar y reducir los costos experimentales.

En Internet podemos encontrar Servidores Profesionales que realizan el Acoplamiento Molecular y presentan el respaldo computacional para realizar los costosos cálculos que se derivan de este proceso, entre ellos podemos encontrar el DockingServer [3], el RosettaDock [4] y el Swissdock [5]. Si se realiza un análisis de los costos de las ejecuciones en estos servidores encontramos que aquellos servicios con mayor número de prestaciones tienen costos elevados, lo que constituye una limitante para los especialistas de países en vías de desarrollo. Por otro lado, estos servidores no le brindan a los especialistas la información referente a las políticas de protección de sus datos cuando consumen de sus servicios.

En Cuba existen instituciones que presentan recursos computacionales para realizar Acoplamiento Molecular, como son el Centro de Ingeniería Genética y Biotecnología (CIGB) y el Centro de Inmunología Molecular (CIM). Sin embargo no todos los especialistas cubanos cuentan con los recursos computacionales o algún servidor en la red nacional que brinde el servicio de Acoplamiento Molecular. A pesar de existir cooperaciones entre los especialistas de los diferentes centros, la demanda de recursos computacionales de altas prestaciones es superior a la existente en la red de investigación nacional.

La UCI cuenta con varios centros de desarrollo de software, donde se encuentra el Centro de Tecnología y Gestión de Datos (DATEC). En el departamento de Bioinformática de este centro se desarrolla una Plataforma de Servicios Bioinformáticos la cual tiene dentro de sus principales objetivos brindar un conjunto de servicios a los especialistas del país asociados al área de la bioinformática, siendo el servicio de Acoplamiento Molecular uno de los que se quiere implementar en esta plataforma.

En este trabajo se propone reducir el tiempo en obtener la respuesta de un proceso de Acoplamien-

to Molecular, que en una PC tradicional demora un tiempo excesivamente largo, y brindar el servicio a la comunidad de especialistas cubanos de esta rama la ciencia.

Por lo que proponemos un Servicio Web de Acoplamiento Molecular que aproveche los recursos computacionales presente en la institución; una Interfaz Web, la cual se muestra en forma de Portlet, que consume el Servicio Web y permite al especialista gestionar las ejecuciones sobre el Servicio Web así como administrar las respuestas dadas por el mismo.

2. CONTENIDO

En los siguientes sub-epígrafes se describen las herramientas que utilizamos para el desarrollo del Servicio Web de Acoplamiento Moleculary la implementación delaInterfaz Web en forma de Portlet; y los sistemas necesarios para brindar nuestro servicio. Así como, se discuten los resultados obtenidos en las pruebas realizadas al Servicio Web de Acoplamiento Molecular.

2.1 Materiales y Métodos

Para la búsqueda de los mejores acoplamientos entre un receptor y un ligando candidato, los métodos de Acoplamiento Molecular emplean funciones de puntuación basadas en distintos fundamentos teóricos. Estás funciones de puntuación se encuentran implementadas en varias herramientas que permiten el acoplamiento de una molécula receptora u objetivo con otra molécula que puede ser un ligando o una proteína [6]. Estas herramientas son utilizadas teniendo en cuenta las características del sistema de estudio así podemos encontrar entre las más utilizadas: Autodoc, Dock, Gold y FlexX.

Para esta versión del Servicio Web de Acoplamiento Molecular se utiliza Dock como herramienta de Acoplamiento Molecular, por tener precedentes de investigaciones anteriores con el mismo en el departamento [7].

Dock6.5 [8] es una herramienta que realiza el acoplamiento molecular, fue escrita en lenguaje C++ y es de código abierto. Utiliza un modelo 3D de la proteína objetivo y gira el ligando sobre el área elegida de dicha proteína para lograr una mayor conformación entre ella y el ligando [9]. Permite hacer acoplamiento molecular no solo proteína - ligando, sino también proteína - proteína, además de poder compilarse para cualquier arquitectura y ser de licencia libre.

Se realizó el Servicio Web de Acoplamiento Molecular sobre el middleware **Apache Axis2** [10] que permite la comunicación con otras aplicaciones y servicios sobre la Web de Apache. Su arquitectura se basa en desarrollar sobre un núcleo simple y extensible que proporciona las abstracciones básicas para el resto del sistema.

Para el despliegue del Servicio Web de Acoplamiento Molecular se utilizó el servidor de aplicaciones web **Apache Tomcat6.0** [11] por ser de código abierto, que une la tecnología Java Servlet y Java Server Pages. Es desarrollado en un entorno abierto y participativo y publicado bajo la licencia Apache2.0. Está destinado a ser una colaboración de los desarrolladores de los mejores de su clase de todo el mundo. En él se pueden montar aplicaciones de misión crítica a través de la Web [12]. Es un servidor HTTP y un contenedor de servlets (clases de Java), cargándolos y ejecutándolos de forma dinámica [13].

Para satisfacer la demanda de cómputo que requiere el Dock6.5 se utiliza **T-arenal2.0** [7], Plataforma de Cálculo Distribuido desarrollada en la UCI sobre la arquitectura cliente - servidor y desplegada en los laboratorios docentes de la misma universidad. Basado en la computación voluntaria [14], la cual plantea que: los clientes que son los encargados de realizar el procesamiento de cada uno de los sub problemas, son los que le piden al servidor la unidad de trabajo para su procesamiento. La forma en que el servidor debe generar las unidades de trabajo y la forma en que tiene que unir las respuestas dada por el sistema se especifican en la claseDataManager.java y la forma de ejecución del problema en la PC cliente se especifica en la clase Algorithm.java.

En el desarrollo del Servicio Web de Acoplamiento Molecular y dela Interfaz Weben forma de Portlet se usa el lenguaje de programación **Java** [15] por ser un lenguaje multiplataforma, orientado a objetos y de licencia libre.

Como IDE de desarrollo se utilizó el **Eclipse Helios3.6** [16], el cual permite trabajar con el lenguaje de programación propuesto, es fácil de integrar con Apache Axis2 y es un software de licencia libre.

El sistema gestor de base de datos que se utiliza es el **PosgreSQL8.4** [17], por ser un sistema de gestión de bases de datos objetos - relacional, de propósito general, multiplataforma, multiusuario, de código abierto y diseñado para ambientes de misión crítica y mantenimiento de grandes bases de datos. Soporta el almacenamiento de objetos binarios grandes. Contiene columnas auto-incrementales. Implementa la herencia de tablas y los disparadores comunes, por columna, condicionales. Permite las consultas recursivas.

Se desarrolló sobre el marco de trabajo **Spring2.6** [18] para la implementación de la Interfaz Web en forma de Portlet, el cual proporciona una amplia configuración de la programación y el modelo de las modernas aplicaciones empresariales, basadas en

Java. Un elemento clave de Spring es el apoyo de infraestructura a nivel de aplicación.

La conexión a la base de datos se realiza mediante la herramienta de mapeo objeto -relacional **Hibernate3.3.1** [19] por ser un conjunto de proyectos relacionados que permiten a los desarrolladores utilizar estilos de modelos de dominio en sus aplicaciones de manera que se extienden mucho más allá del mapeo objeto - relacional. Permite al desarrollador detallar cómo es su modelo de datos, qué relaciones existen y qué forma tienen. Genera las sentencias SQL y libera al desarrollador del manejo manual de los datos que resultan de la ejecución de dichas sentencias, manteniendo la portabilidad entre todos los motores de bases de datos con un ligero incremento en el tiempo de ejecución.

2.2 Resultados y Discusión

En el experimento fueron utilizados una base de datos de 500 ligandos del ZINC [20] y como proteína objetivo la proteína que presenta el código 1VRT del PDB [21], ambos provenientes de los DEMOS[1] del Dock6.5. Fueron utilizadas las máquinas de los laboratorios docentes de la UCI donde se encuentra montado el sistema T-arenal, todas con sistema operativo basado en GNU/Linux.

2.2.1 Servicio Web de Acoplamiento Molecular

La base de datos de 500 ligandos fue dividida en subconjuntos de 1, 50, 100, 200, 250 y 500 ligandos respectivamente. La **Tabla 1** muestra los tiempos de respuesta obtenidos como resultado del uso del Servicio Web de Acoplamiento Molecular.

Ligandos	1	50	100	200	250	500
Una PC	0h 0m 2s	0h 2m 56s	0h 6m 33s	0h 20m 43s	0h 28m 46s	1h 47m 2s
Servicio Acoplamiento Molecular	0h 1m 13s	0h 2m 12s	0h 3m 12s	0h 5m 11s	0h 7m 17s	0h 14m 33s

Tabla1Ejecuciones de Acoplamiento molecular en el Servicio Web de Acoplamiento Molecular.

Como se muestra en el **Gráfico 1**, el Servicio Web desarrollado ha dado solución al problema planteado en la introducción de este trabajo, ya que logra reducir hasta en un 75% el tiempo de respuesta de una gran simulación de Acoplamiento Molecular.

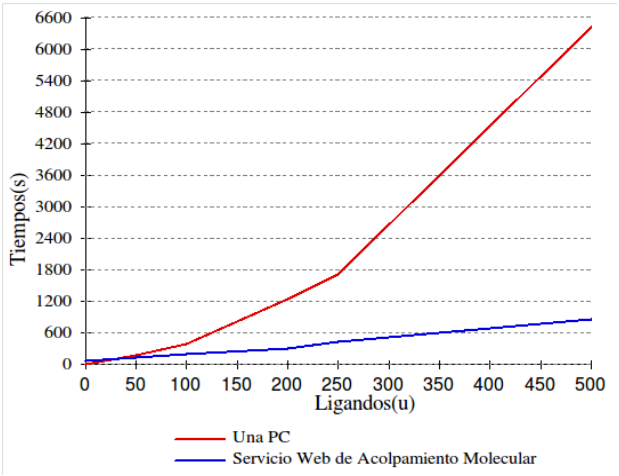

Gráfico1Tiempos en obtener respuesta de los subconjuntos de ligandos utilizados.

2.2.2 Speed Up

Para el cálculo del SpeedUp que es la relación que tienen las ejecuciones del proceso distribuido con la ejecución en una sola estación de trabajo, fueron utilizadas 1, 20, 40, 80 y 100 máquinas, dando como resultado: ver **Tabla 2.**

Cantidad de PCs	1	20	40	80	100
Tiempo ejecución	1h 55m 36s	0h 15m 29s	0h 7m 17s	0h 7m 20s	0h 9m 9s
SpeedUp	1	7,47	15,94	15,76	12,63

Tabla 2Ejecuciones de Acoplamiento molecular en el Servicio Web de Acoplamiento Molecular.

Como se muestra en el **Gráfico 2**, tenemos un incremento gradual de la velocidad de procesamiento a medida que se fueron incrementando la cantidad de PCs, hasta llegar a 40 máquinas a partir de la cual se logró una meseta y luego de esto a partir de las 80 PCs tiende a disminuir la velocidad de procesamiento. Es importante destacar que este comportamiento (aceleración - meseta - desaceleración) es característico de este tipo de proceso distribuido[7].

[1] Paquete de demostraciones para probar la aplicación.

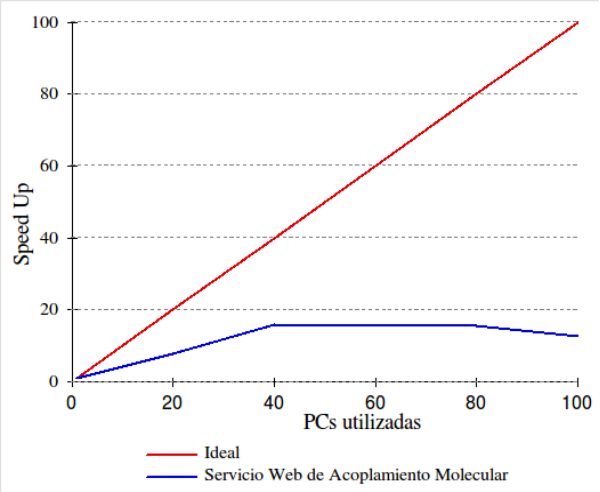

Gráfico 2SpeedUp del Servicio de Acoplamiento Molecular.

2.2.3 Interfaz Web

Para consumir el Servicio Web de Acoplamiento Molecular se creó unainterfaz Web en forma de Portlet, el cual permite a los especialistas gestionar las corridas[2] sobre el sistema y administrar los resultados dado por el Servicio Web de Acoplamiento Molecular.

Todas las acciones sobre los datos del especialista están a un clic de distancia.

Las corridas en el sistema se organizan según una cola con prioridad, tomando siempre la corrida con menor prioridad.La prioridad se calcula utilizando la siguiente fórmula:

$$\mathbf{P = T + F} \quad (1)$$

$$\mathbf{T = \sum_{i=1}^{n} Ar_i} \quad (2)$$

$$F = A * 10^{10} + M * 10^{8} + D * 10^{6} + h * 10^{4} + m * 10^{2} + s \quad (3)$$

donde:

P: prioridad de la corrida

T: tamañode todos los archivos que necesite la corrida del especialista

F: fecha en que se crea la corrida

Ar_i: archivo i-ésimo

A: año

M: mes

D: día

h: hora

m: minuto

s: segundo

Con esta estrategia de prioridad garantizamos que se ejecuten primero las corridas que alternen entre las contengan una menor cantidad de datos de entrada y las que lleven más tiempo de creadas en el sistema.

En la [**Figura 2**] se muestra la interfaz que le permite al especialista Crear, Modificar, Eliminar y Buscar sus corridas en el sistema.

Figura2 Interfaz Gestionar Corrida.

La **[Figura 3 y Figura 4]** muestran las interfaces de crear y modificar corrida, respectivamente, las cuales permiten subir los archivos necesarios para la creación o modificación de una corrida en el sistema.

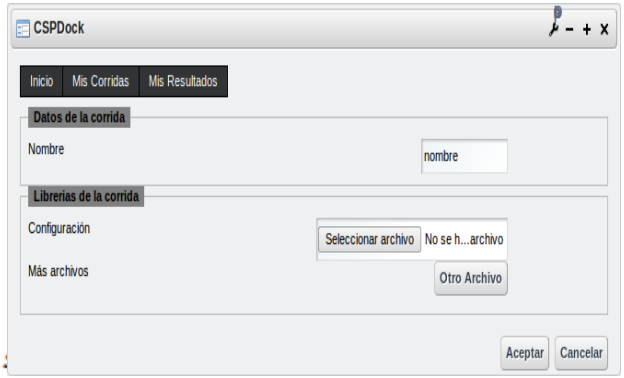

Figura 3 Interfaz Crear Corrida.

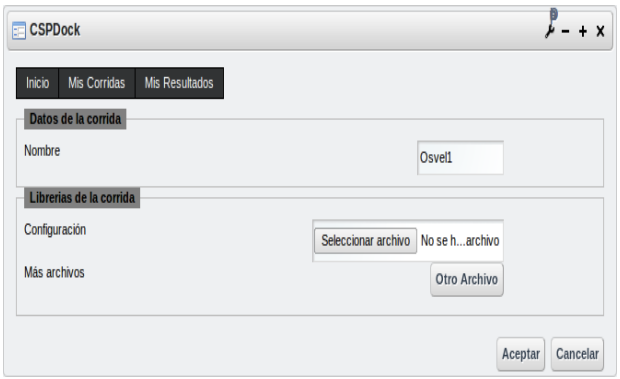

Figura 4 Interfaz Modificar Corrida.

En la **[Figura 5]** se muestra la interfaz que le permite al especialista Ver, Descargar, Eliminar y

[2]Todos los datos necesarios para realizar un Acoplamiento Molecular con Dock 6.5.

Buscar un resultado arrojado por el Servicio de Acoplamiento Molecular a una corrida ya terminada.

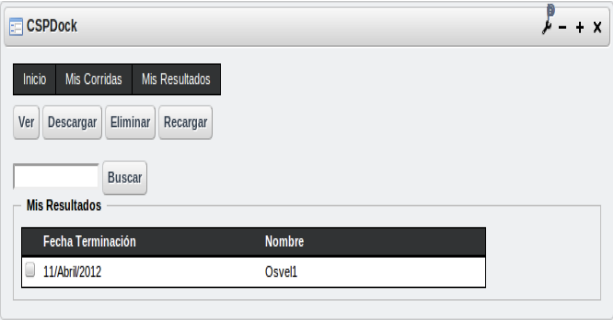

Figura 5 Interfaz Administrar Resultado.

La [**Figura 6**] muestra una interfaz correspondiente a la opción de Ver resultado.

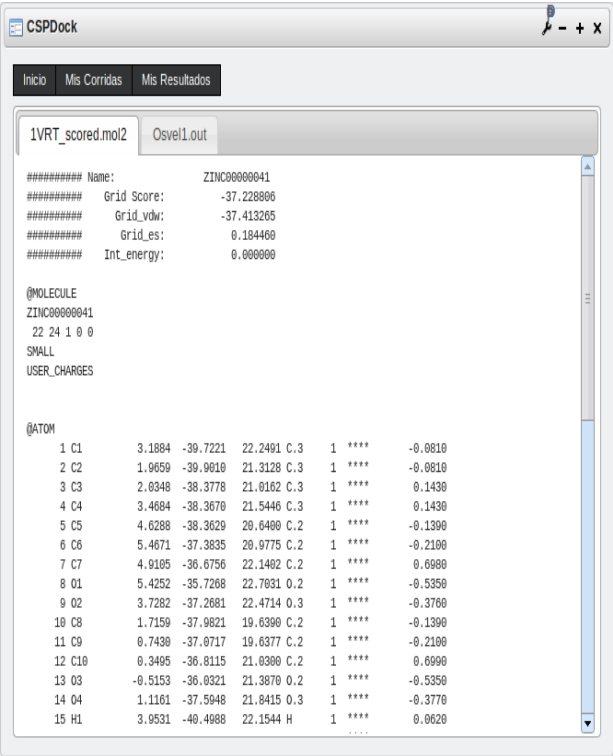

Figura 6 Interfaz Ver Resultado.

La [**Figura 7**] muestra la confirmación a la opción de Descarga.

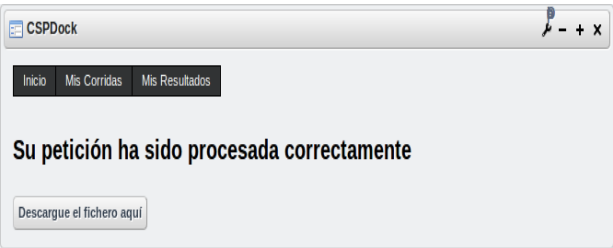

Figura 7 Interfaz Descargar Resultado.

Es importante destacar que tanto la Interfaz Web en forma de Portlet como el Servicio Web de Acoplamiento Molecular no cuentan con una Herramienta de Administración de archivos, esto le da al especialista garantía de que sus datos no serán administrados por otros usuarios que no sean ellos mismos. Así como que no existe una forma de recuperación de datos, una vez que el especialista haya modificado y/o eliminado un dato, este no puede volver a su condición inicial.

3. CONCLUSIONES

Se desarrolló un Servicio Web de Acoplamiento Molecular que puede ser consumido por otras aplicaciones y utiliza como soporte de cálculo la plataforma de tareas distribuidas desarrollada y desplegada en la UCI.

Se desarrolló una Interfaz Web en forma de Portlet el cual le permitirá a los especialistas gestionar las corridas que realicen sobre nuestro Servicio Web de Acoplamiento Molecular y administrar las respuestas dadas por el mismo.

Se logró reducir el tiempo en obtener la respuesta de una gran simulación de Acoplamiento Molecular, aprovechando los recursos computacionales existentes en la universidad.

4. REFERENCIAS BIBLIOGRÁFICAS

1. Abbas I.: "Integración de los modelos de simulación en el diseño de los ensayos clínicos." Tesis Doctorales en Red. 2003; Availablefrom: http://www.tesisenred.net/handle/10803/6506

2. Padilla Zúñiga A.J. y Rojo Domínguez A.:"Simulación del reconocimiento entre proteínas y moléculas orgánicas o Docking. Aplicación al diseño de fármacos."Mensaje Bioquímico, Vol26.Depto Bioquímica, Fac Medicina, Universidad Nacional Autónoma de México. Cd Universitaria, México, DF, MÉXICO. 2002; Aviablefrom: http://bq.unam.mx/wikidep/uploads/MensajeBioquimico/Mensaje_Bioq02v26p129_Rojo.pdf

3. E H.:"DockingServer: molecular docking calculations online."Acta Pharm Hung. 2009;Available from: http://www.ncbi.nlm.nih.gov/pubmed/19526678

4. Lyskov S.:"The RosettaDock server for local protein-protein docking."Oxford Journals.2008 abril;36(2).Available from: http://nar.oxfordjournals.org/content/36/suppl_2/W233.short

5. SitioOficial.: Swiss Intitute of Bioinformatics. Available from: http://www.swiss.vital-it.ch/

6. Kitchen D. B.:"Docking and scoring in virtual screening for drug discovery: methods and applications."Nature Reviews, Drug Discovery. 2004noviembre;3. Available from:http://www.nature.com/nrd/journal/v3/n11/pdf/n

rd1549.pdf

7. Aguilera Mendoza L., García Jacas C. R. y MirayesTaset D. M.: "Sistema de Cómputo Distribuido" Telem@tica2008; No. 16. Available from:http://www.fiec.espol.edu.ec/resources/download/revista/Telematica_A%C3%B1oVII_No16.pdf.

8. The Oficial UCSF DOCK Web-site; Available from:http://dock.compbio.ucsf.edu/DOCK_6/index.htm

9. Tsui B.:"PRIME 2011 Osaka University."2011;Available from: http://prime.ucsd.edu/progress_reports_2011/PRIME_Final_Final/BTsui_Final_2011.pdf

10. Perera S.:"Axis2, Middleware for Next Generation Web Services."Web Services, 2006 ICWS '06 InternationalConference on.2006 septiembre;p. 833 - 840.Available from:http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnum-ber=4032100&abstractAccess=no&userType=inst.

11. Apache Tomcat. Web Oficial Cite; Available from: http://tomcat.apache.org/

12. Mérida D.: "Adaptación de Contenidos Web Considerando las Características de los Dispositivos de Enlace Informático". 2006 Dic; 5(1): 9. Available from:http://atreides.udg.edu/bcds/images/bcds/papers/pdf/adaptaci_n_de_contenidos_web_considerando_las_caracter_sticas_de_los_dispositivos_de_acceso.pdf

13. "Introducción a Apache Tomcat 5.5." Universidad de Sevilla; Available-from:http://www.lsi.us.es/docencia/get.php?id=1923.

14. Informatica en salud 2009; Availablefrom: http://www.informatica2009.sld.cu/Members/dmmirayes/sistema-de-computo-distribuido/

15. "Lenguajes de Programación"; Availablefrom: http://www.lenguajes-de-programacion.com/lenguajes-de-programacion.shtml

16. Geer D.:"Eclipse becomes the dominant Java IDE."IEEEXplorer digital library. 2005 julio; 38(7): 16 - 18. Available from: http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=31455

17. Sitiooficial de PostgreSQL; Available from: http://www.postgresql.org/about/press/presskit91/es/

18. Spring Framework;Available from: http://www.springsource.org/spring-framework

19.JBoss Community;Available from: http://www.hibernate.org/

20.ZINC a free database of commercially-available compounds for virtual screening;Available from: http://zinc.docking.org/

21.PDB a Portal to Biological Macromolecular Structures;Available from: http://www.rcsb.org/pdb/home/home.do

5. SÍNTESISCURRICULARES DE LOS AUTORES

Osvel Chávez Hernández, Ingeniero en Ciencias Informáticas, graduado en la Universidad de las Ciencias Informáticas (UCI), graduado con título de Oro, promoción de 2012. Publicó como coautor en Uciencia 2012 y como autor en la Jornada Científico Estudiantil 2012 de la UCI; y en el Fórum de Ciencia y Técnica 2012 de la UCI.