

Universidad de las Ciencias Informáticas

Facultad 6



Título: “Sistema de análisis de secuencia mRNA para el diseño de siRNA v2.0”

Trabajo de Diploma para optar por el título de
Ingeniero Informático

Autores: Roberto Duniel Rivero Soto.

Edecio Ramón Batista Rodríguez.

Tutores: Dr. Ricardo Bringas Pérez.

Ing. Vladir Antonio Parrado Cruz.

Junio, 2009

PENSAMIENTO

“El futuro de nuestra Patria tiene que ser, necesariamente, un futuro de hombres de ciencia”.

Fidel Castro Ruz.

DECLARACIÓN DE AUTORÍA

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

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

Roberto Duniel Rivero Soto

Firma del Autor

Edecio Ramón Batista Rodríguez

Firma del Autor

Vladir Antonio Parrado Cruz

Firma del Tutor

Ricardo Bringas Perez

Firma del Tutor

DATOS DE CONTACTO

Autor:

Roberto Duniel Rivero Soto.

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

Autor:

Edecio Ramón Batista Rodríguez.

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

Tutor:

Ing. Vladir Antonio Parrado Cruz.

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

Email: vparrado@uci.cu

Tutor:

Dr. Ricardo Bringas Perez.

Centro de Ingeniería Genética y Biotecnología, Ciudad de La Habana, Cuba.

Email: ricardo.bringas@cigb.edu.cu

AGRADECIMIENTOS

A todo el mundo.

DEDICATORIA

Para nuestras familias.

RESUMEN

El silenciamiento de genes por interferencia de RNA (iRNA) es una tecnología naciente que se está llevando a cabo por el alto impacto que esto tiene en las investigaciones biológicas, además esta técnica se perfila como una poderosa herramienta terapéutica para combatir enfermedades en diferentes organismos. Para los estudios con esta tecnología un componente fundamental lo constituyen las herramientas informáticas para el diseño de los RNA pequeños de interferencia (siRNA). A pesar de la existencia en Internet de aplicaciones que permiten realizar este diseño, el Centro de Ingeniería Genética y Biotecnología (CIGB) y la Universidad de las Ciencias Informáticas (UCI) han decidido desarrollar una herramienta propia con estos fines, que le permita desarrollar estas predicciones con mayor seguridad de los resultados obtenidos y agregar facilidades y funcionalidades a la aplicación a partir de recomendaciones y exigencias derivadas de la experiencia experimental de los investigadores de esa institución y así mejorar el ambiente de investigación de los especialistas en esta rama. Para ello se realizó una primera versión del sistema que cumplía con las funcionalidades requeridas por los especialistas en una primera etapa de definición del proyecto. Luego surgió la necesidad de extenderlo a una segunda versión, tomando como base la anterior, con el objetivo de hacerle mejoras al proceso de diseño e incluir nuevas funcionalidades.

PALABRAS CLAVE

Diseño de siRNA, miRNA, Silenciamiento de genes, siRNA, Sitios blancos.

ÍNDICE

AGRADECIMIENTOS	I
DEDICATORIA	II
RESUMEN	III
INTRODUCCIÓN	1
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA	4
Introducción	4
1.1 La informática y su relación con las ciencias biológicas.	4
1.2 Sistemas existentes para el diseño de siRNA y predicción de estructura secundaria.	5
1.2.1 siRNA Target Finder.	5
1.2.2 siRNA Design Software.....	5
1.2.3 Custom siRNA Design Tool.	5
1.2.4 DEQOR.....	6
1.2.5 IDT´Scitools RNAi Design.	6
1.2.6 siDirect.....	6
1.2.7 siRNA Wizard v3.0.	7
1.2.8 RNA Vienna.	7
1.2.9 alasSiRNA-Design 1.1.....	7
1.3 Limitaciones de los sistemas existentes.	7
1.4 Silenciamiento de genes.	8
1.4.1 Definición de siRNA.....	9
1.4.2 Definición de miRNA	10
1.4.3 miRNA y siRNA. Semejanzas y diferencias.....	11
1.4.4 mRNA	12
1.4.5 SNP.....	12
1.5 Tecnologías, metodologías y herramientas.	12
1.5.1 Tecnología J2EE.....	13
1.5.2 Lenguajes de desarrollo del software.....	13
1.5.3 Otras consideraciones generales para la selección de estas tecnologías.....	15
1.5.4 Herramientas de desarrollo utilizadas.....	15
1.5.5 Gestor de Base de Datos	17

1.5.6 Servidor de Aplicaciones	18
1.5.7 Proceso de desarrollo de software	20
1.5.8 Servicio Web	21
1.5.9 SOA (Service Oriented Architecture).....	22
1.6 Resumen de la tecnología, herramientas y metodologías a utilizar	22
1.7 Conclusiones.....	23
CAPÍTULO 2: CARACTERÍSTICAS DEL SISTEMA.....	24
Introducción	24
2.1 Objeto de estudio.....	24
2.1.1 Objetivos estratégicos de la organización.	24
2.1.2 Flujo actual de los procesos.	25
2.2 Procesos objeto de automatización.	25
2.2.1 Procesos que serán objeto de automatización.	25
2.3 Propuesta de sistema	27
2.3.1 Descripción general de la propuesta de sistema.....	27
2.3.2 Análisis comparativo de otras soluciones existentes con la propuesta.....	28
2.4 Modelo del Dominio	29
2.4.1 Reglas del negocio.....	30
2.5 Especificación de los requisitos de software	31
2.5.1 Requerimientos Funcionales	31
2.5.2 Requerimientos no funcionales	32
2.6 Definición de actores y casos de uso del sistema	35
2.7 Casos de Uso definidos	35
2.8 El diagrama de casos de uso para alasSiRNA Desing v2.0:	39
2.9 El diagrama de casos de uso para siRNA Web Services:	40
2.10 Especificación de los Casos de Uso	41
2.11 Conclusiones.....	53
CAPITULO 3: ANÁLISIS Y DISEÑO DEL SISTEMA	54
Introducción	54
3.1 Arquitectura de software orientado a servicios.	54
3.2 Objetivos del diseño	55
3.3 Vista Lógica del sistema.	55
3.4 Diseño de clases del sistema.....	56

3.4.1	Diseño de las clases de alasSiRNA - Design versión 2.0	56
3.4.2	Diseño de las clases del web services.	63
3.4.3	Diseño de la base de datos (clases persistentes).....	71
3.5	Descripción de las clases del web services.	72
3.6	Descripción de las Tablas de la Base de Datos.	82
3.7	Diagramas de Interacción (Secuencia, Colaboración)	86
3.8	Diagrama de Despliegue.	87
3.9	Tratamiento de errores	88
3.10	Seguridad	90
3.11	Conclusiones	90
CAPÍTULO 4:	IMPLEMENTACIÓN Y PRUEBA	91
	Introducción.	91
4.1	Diagrama de componentes	91
4.2	Diagramas de componentes del proyecto	91
4.2.1	Diagramas de componentes de la aplicación del alasSiRNA - Design v2.0.....	92
4.2.2	Diagramas de componentes del web services.	93
4.3	Obtención de algoritmos	94
4.3.1	Obtención de algoritmo para la búsqueda de la secuencia común en un gen determinado. .	95
4.3.2	Obtención de algoritmo para la búsqueda de la secuencia específica en un transcrito determinado.....	95
4.3.3	Obtención de algoritmo para calcular la energía libre entre el sitio blanco y el siRNA complementario a este.	96
4.4	Modelo de Prueba.	96
4.4.1	Casos de prueba de caja negra.....	96
4.5	Conclusiones.	106
CONCLUSIONES GENERALES.	108
RECOMENDACIONES	109
REFERENCIAS BIBLIOGRÁFICAS	110
BIBLIOGRAFÍA	113
ANEXOS	116
Anexo 1: Diagramas de Secuencia	116
Anexo 2: Pantallas del sistema.	122

GLOSARIO DE TÉRMINOS128

INTRODUCCIÓN

En 1998 Andrew Fire y Craig Mello descubrieron la existencia de un RNA de doble hebra de 20-21 nucleótidos perfectamente complementarias, denominados RNA pequeños de interferencia (siRNA por sus siglas en inglés), los cuales suprimen la expresión de diferentes genes mediante el corte del RNA mensajero (mRNA por sus siglas en inglés) complementario [1]; por tanto, el objetivo de los siRNAs es desactivar secuencias génicas específicas en el organismo, lo que sería de gran utilidad para el desarrollo de la biología molecular y así incorporar beneficios al resto del mundo.

La creación de manera artificial de siRNA es una poderosa herramienta terapéutica para combatir y buscar soluciones a diferentes problemas en plantas, animales y hombres. Además se ha demostrado que puede ser utilizado como una forma de terapia génica, es decir si cierta enfermedad es el efecto del aumento en la expresión de un gen específico y se conoce su identidad, se podría “apagar” su expresión por medio de inyecciones de siRNAs y curar así la enfermedad [2].

Esto ha causado una revolución científica, por lo que existen numerosos especialistas y científicos realizando estudios de esta índole en todo el mundo. Cuba no se queda atrás, el Centro de Ingeniería Genética y Biotecnología (CIGB) es una de las organizaciones pioneras en este tipo de investigación.

Actualmente en el mundo hay varias herramientas online publicadas en Internet, de las cuales se puede mencionar a: siRNA Design Tools [3], siRNA Design Software [4], Custom siRNA Design Tool [5], DEQOR [6], IDT Scitools RNAi Design [7], siDirect [8] y siRNA Wizard v3.0 [9].

La utilización de tales herramientas crean una serie de problemas como por ejemplo: que en la mayoría de las aplicaciones no se conocen los algoritmos que se han utilizado para la predicción de los siRNA y no son configurables para el uso de algoritmos propios. No dan la posibilidad de hacer un análisis con el por ciento de exactitud que se estime conveniente de acuerdo al estudio que se esté realizando, además no existe una unidad en las ideas expuestas en estos sitios, razón que provoca que exista diversidad sobre el mismo tema y no haya seguridad en los resultados obtenidos. Además a estas herramientas no se le pueden incluir facilidades, ni funcionalidades específicas que requiera un especialista [2].

Para cubrir con esta necesidad el CIGB y la Universidad de las Ciencias Informáticas (UCI) se han dado a la tarea de crear una aplicación capaz de dar solución a este problema. Para ello se hizo una primera versión de esta con las funcionalidades requeridas por los especialistas para aquella etapa, pero no con la capacidad de dar solución a todas las problemáticas planteadas en este tema de diseño de siRNA.

En el transcurso del proceso de desarrollo de la versión 1.1 del Sistema de análisis de secuencia mRNA para el diseño de siRNA han aparecido necesidades objetivas primero por parte de los especialistas del CIGB que requieren brindar las funcionalidades implementadas en esta versión desde su institución a otras instituciones incluso a nivel mundial sin mostrar el código fuente (por problemas de seguridad y autoría de los algoritmos) para fortalecer las relaciones y el avance de los estudios del tema de silenciamiento genético y de esta manera la comunidad que trabaja los siRNAs avance en su camino por encontrar soluciones a enfermedades muy dañinas. Por otra parte la versión implementada contempla solo el análisis de un transcrito de un gen y no se tienen en cuenta diseños específicos para transcritos o comunes a todos, tampoco permite visualizar la energía libre de los sitios blanco, lo cual es de suma importancia pues mientras menor es esta más funcional es el mismo. Criterios como estos permiten hacer diseños funcionales y más exactos de siRNAs.

Concretamente existen necesidades del proceso de diseño de siRNA que se deben tener en cuenta para realizar un buen diseño, por lo que el trabajo a realizar se orienta a ¿Cómo lograr un proceso amplio y funcional del diseño automático de siRNA?

Para la obtención de un siRNA funcional, a través de las investigaciones de los especialistas, se hace necesario estudiar el proceso de análisis de secuencia mRNA para el diseño de siRNA. Para esto, se incidirá específicamente en las aplicaciones informáticas Cliente-Servidor para el diseño de siRNA. Con la finalidad de dar solución al problema señalado se traza como **objetivo general** desarrollar la versión 2.0 del Sistema de análisis de secuencia mRNA para el diseño de siRNA.

Para dar cumplimiento al objetivo general del trabajo se trazaron los siguientes **objetivos específicos**:

- Realizar el levantamiento de nuevos requisitos.
- Diseñar las nuevas funcionalidades.
- Implementar las nuevas funcionalidades.
- Realizar pruebas al software.

Para cumplir con dichos objetivos, se trazaron las siguientes **tareas**:

- Revisión bibliográfica.
- Investigación de las tendencias y selección de las tecnologías para realizar el diseño de siRNA.
- Realización de las actividades del flujo de trabajo de requerimientos.
- Realización de las actividades del flujo de trabajo de diseño.
- Implementación de las nuevas funcionalidades.
- Realización de pruebas al sistema.

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

Introducción

En este capítulo se hace un análisis de la relación de la informática con las ciencias biológicas. Se presentan los fundamentos teóricos y prácticos que cimientan las soluciones seleccionadas. Se aborda la importancia del desarrollo de la aplicación. Además se mencionan y analizan una serie de aplicaciones ya desarrolladas vinculadas con el tema de diseño de siRNA. Se fundamentan las tendencias y tecnologías actuales utilizadas para el logro de una herramienta o producto con calidad. También se trata de cómo se organiza el trabajo, cuáles herramientas se utilizarán debido a sus características y la forma de uso.

1.1 La informática y su relación con las ciencias biológicas.

La informática en la actualidad es una solución a muchos problemas que hace algunos años era imposible responder. En la rama de las ciencias biológicas ha resaltado su papel, en especial en el área de Biología molecular. Esta interacción condicionó el surgimiento de una disciplina que creó vínculos indisolubles entre la Informática y las ciencias biológicas: la Bioinformática.

La Bioinformática ofrece la posibilidad de comparar y relacionar la información genética con fines deductivos. En los últimos años, se aprecia el crecimiento de una corriente de investigación y desarrollo de nuevas técnicas para la extracción del conocimiento, la minería de datos y la visualización, cuyo objetivo es acelerar los descubrimientos científicos, a partir de la reducción de los costos y el aumento del número de experimentos. Estas nuevas técnicas, útiles para la investigación de distintas enfermedades y el diagnóstico clínico, establecen el camino a seguir por bioinformáticos e investigadores en el futuro. Se puede afirmar que cualquier proceso en cualquier campo biomédico requiere un apoyo informático [10].

En la actualidad se han publicado varias herramientas en Internet relacionadas con la bioinformática. Las cuales han sido fuente utilizable para el presente trabajo, principalmente las publicadas en sitios online que diseñan sitios blancos, predicen la estructura secundaria del RNA y calculen la energía en la predicción de la estructura secundaria.

1.2 Sistemas existentes para el diseño de siRNA y predicción de estructura secundaria.

En la actualidad hay cierto número de herramientas online dedicadas al diseño de siRNA y varias aplicaciones de código abierto, publicados en Internet para la predicción de estructuras secundarias. De las que se destacan:

1.2.1 siRNA Target Finder.

Es una herramienta creada para encontrar los sitios blancos en una secuencia de mRNA de interés. Esta herramienta no da la posibilidad de consultar una base de datos para obtener la cadena mRNA, sino que solo da la opción de entrar la cadena ya conformada por el usuario para el análisis de esta secuencia. El resultado puede ser consultado con otra herramienta para más análisis lo que es muy trabajoso.

1.2.2 siRNA Design Software.

Toma una secuencia del mRNA y hace uso de las herramientas de diseño ya existentes para dar como resultado una lista de sitios blancos. No es conocido el algoritmo que implementa para el análisis de las secuencias de mRNA y cuenta con pocas reglas para obtener un resultado de calidad. También se tarda considerablemente al hacer el análisis de la secuencia solicitada y es muy trabajoso para ver la ayuda o guía.

1.2.3 Custom siRNA Design Tool.

Es una herramienta de diseño con notables ventajas dentro de las que se destaca una interfaz de usuario amigable y significativas mejoras en el proceso de identificación de siRNA. No obstante a pesar de estas ventajas, posee solo cuatro reglas para realizar la búsqueda de sitios blancos, es muy lento al seleccionar opciones ya que se refresca la pantalla y eso es molesto para el que visite esa aplicación. Además no es estable, en ocasiones está fuera de servicio. También pregunta el país que

está usando la herramienta lo cual no es muy confiable dado que para diseñar siRNAs no hace falta saber el país.

1.2.4 DEQOR.

Es una herramienta que opera mediante un uso excesivo de secciones que muestran las distintas soluciones según las reglas seleccionadas por el usuario. Evalúa la potencia inhibidora de las secuencias potenciales de siRNA e identifica las regiones del gen que tienen un alto poder de silenciamiento. No se tiene conocimiento del algoritmo que utiliza para el análisis de las secuencias. Brinda demasiadas reglas en cuanto a la configuración de vistas que complican el trabajo del usuario. El análisis de las secuencias se torna tedioso debido al tiempo que emplea en obtener y mostrar los resultados. Las vistas que ofrece no cuentan con la calidad que necesitan los especialistas puesto que muestra los resultados de forma poco intuitiva. Además se realiza análisis a archivos de tipo fasta y raw.

1.2.5 IDT´Scitools RNAi Design.

Esta aplicación a pesar de las ventajas que posee hace consultas a una base de datos externa (Centro Nacional para la Información de Biotecnología NCBI) que resulta confiable debido al prestigio del centro, pero no siempre los datos de esta pueden estar disponibles. No se conoce el algoritmo de diseño que emplea, ni es una interfaz amigable y es complicada de entender el funcionamiento de la misma.

1.2.6 siDirect.

Es una aplicación web online para diseñar los siRNA que acepta una secuencia como entrada y devuelve como resultado una lista de sitios blancos. Posee grandes ventajas en cuanto al diseño pero como ya se ha señalado en aplicaciones anteriores, no se conoce al algoritmo que emplea para realizar el análisis de las secuencias. Hace consultas a la base de datos del NCBI lo que constituye una limitante.

1.2.7 siRNA Wizard v3.0.

Esta herramienta no da la posibilidad de consultar una base de datos para obtener la cadena mRNA, sino que solo da la opción de entrar la cadena ya conformada por el usuario para el análisis de esta secuencia. Tiene una interfaz amigable que permite pegar la secuencia del RNA que se quiere analizar y además tiene la opción de buscar los miRNA (microRNA) que están en esa cadena, algo muy útil para el análisis de los siRNA.

1.2.8 RNA Vienna.

Consiste en un paquete para la predicción y la comparación de las estructuras secundarias del RNA. Predice la estructura secundaria a través de la minimización de la energía, que es la función más utilizada en el paquete. Usa el algoritmo de Nussinov para generar la mayor cantidad de cadenas complementarias.

Es mundialmente usado para predecir estructuras secundarias de RNA.

1.2.9 alasSiRNA-Design 1.1.

Este sistema ya liberado, contempla solo el análisis de un transcrito de un gen y no se tienen en cuenta diseños específicos para un transcrito o comunes a todos, tampoco permite visualizar la energía libre del sitio blanco, lo cual es de suma importancia pues mientras menor es esta más funcional es el mismo. Los especialistas del CIGB requieren brindar las funcionalidades implementadas en esta versión desde su institución a otras instituciones incluso a nivel mundial sin mostrar el código fuente (por problemas de seguridad y autoría de los algoritmos) para fortalecer las relaciones y el avance de los estudios del tema de silenciamiento genético, lo que en estos momentos es imposible.

1.3 Limitaciones de los sistemas existentes.

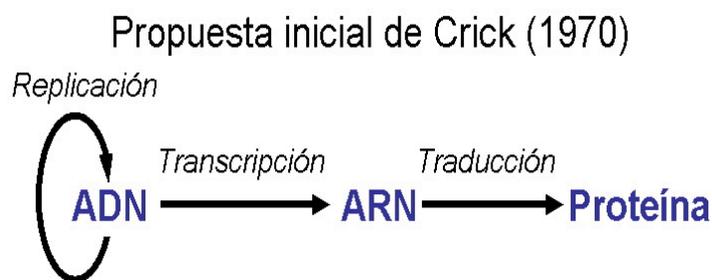
Las herramientas anteriormente presentadas, tienen su uso básico sobre Internet. Estas usualmente suelen sufrir cambios repentinos e ignorados por los usuarios. Se puede afirmar que los resultados obtenidos de los análisis, haciendo uso de estos sistemas, carecen de total confidencialidad. No se tienen en cuenta criterios de búsqueda como en el caso donde hay que considerar la existencia de los SNPs (polimorfismo de un solo nucleótido).

La mayoría de estas herramientas solo permiten hacer el análisis mediante una cadena mRNA entrada por el usuario. Se desconocen los algoritmos implementados para el análisis de las secuencias de mRNA y las reglas que estos algoritmos aplican para la predicción de los sitios blancos de siRNA [2]. A esto se le incluye que con excepción de siRNA Wizard v3.0 no se trata ningún tipo de información en relación a los miRNAs que está siendo analizado. La versión implementada como se pudo observar sigue estando limitada a los requerimientos actuales de los especialistas.

1.4 Silenciamiento de genes.

La información genética es el “programa” que guía la síntesis de proteínas específicas, responsables de las características estructurales y funcionales de un organismo. La molécula principal portadora de esta información es el ADN. En el ADN existen secciones llamadas “genes” que son la unidad de información hereditaria y que en muchos casos son responsables de una proteína específica [11].

Mecanismos bioquímicos en el núcleo celular seleccionan y transcriben parte de la información en un gen formando una copia de su mensaje en moléculas de RNA mensajero (transcripción). Las moléculas de RNA son utilizadas entonces para la síntesis de la proteína correspondiente en los ribosomas. El esquema del flujo de información genética se muestra en el diagrama siguiente [11]:



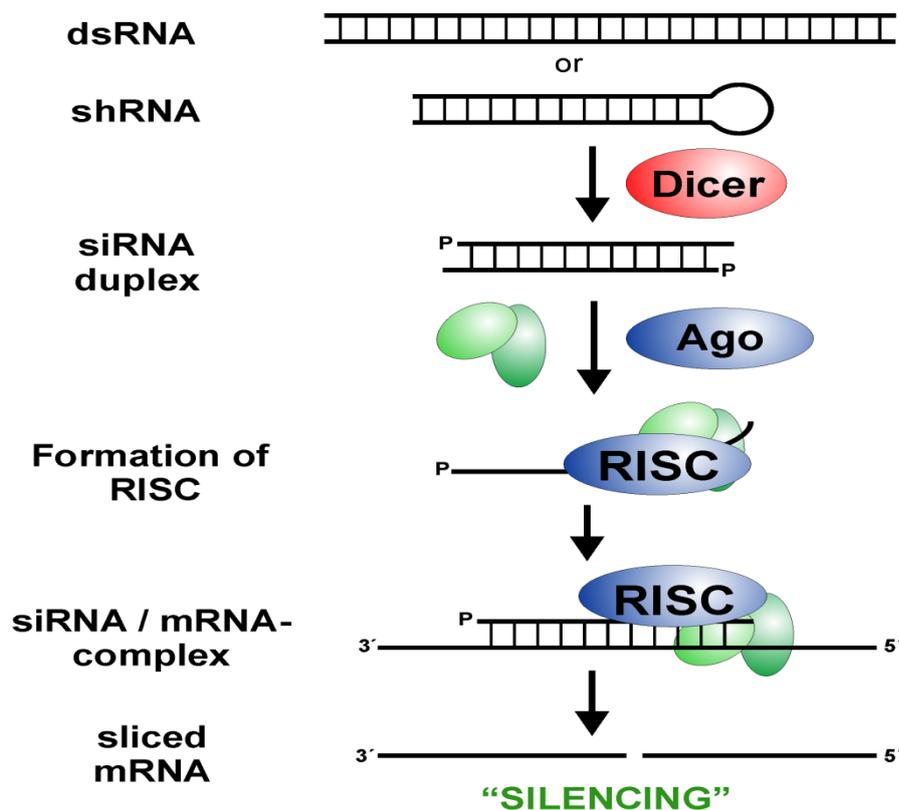
[11]

Fig. 1 Propuesta de Crick.

1.4.1 Definición de siRNA

Los siRNAs son moléculas de RNA doble hebra de 20-21 nucleótidos (nt) perfectamente complementarias, que se originan a partir de un RNA largo de doble hebra (dsRNA). Los dsRNAs pueden ser de origen endógeno, o de origen exógeno. La enzima responsable del procesamiento del dsRNA en moléculas de siRNAs es Dicer.

Los siRNAs suprimen la expresión de los genes diana mediante el corte del mRNA complementario en dos mitades, a través de la interacción de la hebra antisentido del siRNA con el complejo RISC (RNA-induced silencing complex). Las dos mitades del mRNA son posteriormente degradadas por la maquinaria celular, lo que conlleva la supresión de la expresión del gen [12].



[13]

Fig. 2 Proceso de siRNA.

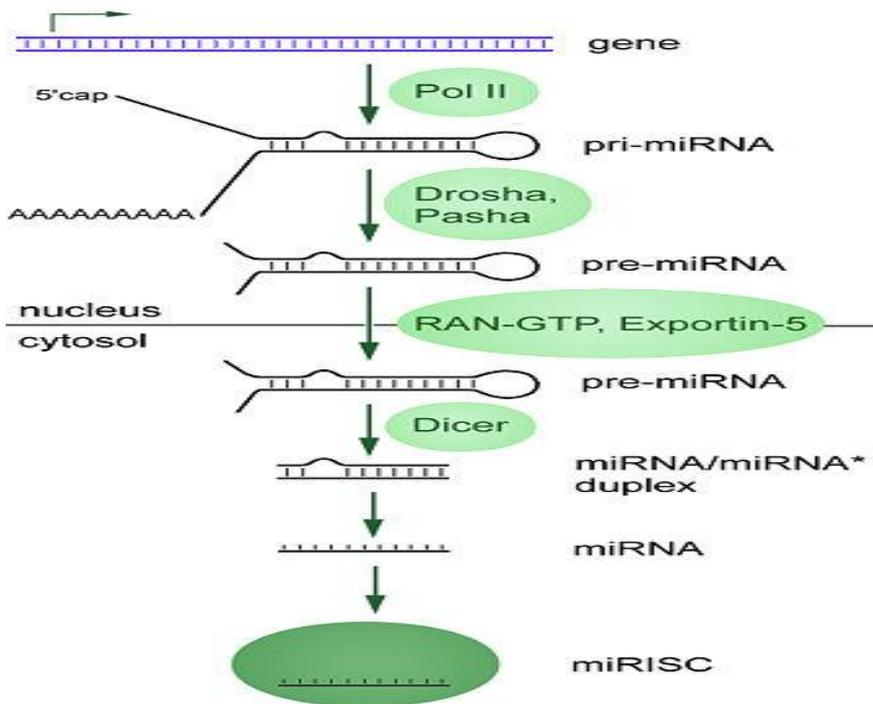
1.4.2 Definición de miRNA

Una vez descubiertos los siRNAs y la existencia en las células de proteínas que catalizan la degradación del mRNA, los investigadores se preguntaron si los siRNAs también estaban codificados en el genoma, y empezaron a purificar pequeños RNAs (19-25 nt) a partir de diferentes especies animales. Sin embargo, no encontraron siRNAs, sino los denominados microRNAs, que se habían identificado anteriormente de forma independiente.

Los miRNAs son moléculas de RNA transcritas a partir de genes de ADN, pero no son traducidos a proteínas. Algunos recuentos de miRNAs en humanos, identificaba que los miRNAs podrían representar como mínimo 3% de todos los genes humanos.

La función de los miRNAs está relacionada con la regulación de la expresión génica. De esta forma un miRNA es complementario de una parte de uno o más mRNA. Los miRNAs generalmente inhiben la traducción del mRNA.

Dado el número de dianas potenciales de los miRNAs estos podrían constituir otra capa del circuito regulatorio que existe en las células. Según esto, cualquier desregulación de los miRNAs podría conllevar grandes problemas regulatorios en la célula [14].



[15]

Fig. 3 Proceso de miRNA

1.4.3 miRNA y siRNA. Semejanzas y diferencias

El siRNA y el miRNA son RNAs de interferencia y su función por tanto es la misma.

El siRNA es creado de forma artificial por los especialistas en sus laboratorios una vez identificado el gen que se desea silenciar y los sitios blancos de su acción. La cadena de nucleótidos de un siRNA por tanto se diseña de acuerdo al objetivo específico del especialista teniendo en cuenta su investigación. Esto constituye una ventaja puesto que una vez creado el siRNA aumenta la probabilidad de que se silencie el gen deseado o se aplique la acción de la forma que se quiere.

A diferencia de los siRNA, el miRNA es creado naturalmente por el organismo. El apareamiento de los miRNA no es siempre exacto, por tal causa, en ocasiones, la expresión del gen no es inhibida, sino solo atenuada [2].

1.4.4 mRNA

El RNA mensajero es el ácido ribonucleico que contiene la información genética procedente del ADN para utilizarse en la síntesis de proteínas, es decir, determina el orden en que se unirán los aminoácidos. Sale del núcleo y se asocia a ribosomas, donde se construye la proteína. Cada tres nucleótidos (codón) corresponde un aminoácido distinto. Así, la secuencia de aminoácidos de la proteína está configurada a partir de la secuencia de los nucleótidos del mRNA [16].

1.4.5 SNP

SNP (polimorfismo de un solo nucleótido) es una variación en la secuencia de ADN que afecta a una sola base de una secuencia del genoma.

Los SNP forman hasta el 90% de todas las variaciones genómicas humanas, y aparecen cada 100 a 300 bases en promedio, a lo largo del genoma humano. Los SNP que se localicen dentro de una secuencia codificante pueden modificar o no la cadena de aminoácidos que producen, se llama SNP sinónimo a los primeros y SNP no-sinónimo a los segundos. Los SNP que se encuentren en regiones no codificantes pueden tener consecuencias en el proceso de traducción. Estos se consideran una forma de mutación puntual que ha sido lo suficientemente exitosa evolutivamente para fijarse en una parte significativa de la población de una especie [17].

1.5 Tecnologías, metodologías y herramientas.

Para el desarrollo del sistema, se realizó un estudio de las posibles herramientas a utilizar en su construcción, teniéndose en cuenta principalmente la tendencia actual y las novedades de cada una de estas. También se elaboraron estrategias o metodologías de desarrollo de software la cual cuenta con uso de estándares y tecnologías que utiliza el centro con el que se desarrolla el proyecto (CIGB) que promueven prácticas orientadas hacia la funcionalidad y la entrega del producto, así como de la implicación directa del cliente.

1.5.1 Tecnología J2EE.

La tecnología Java 2 Enterprise Edition (J2EE) proporciona una completa y potente plataforma orientada al desarrollo de aplicaciones corporativas distribuidas y a los servicios web.

Las aplicaciones desarrolladas sobre esta plataforma suelen tener una arquitectura en capas, que es lo que se pretende realizar en esta versión del sistema. Esto permite una mejor organización, reutilización y extensión del código en nuevas versiones.

Entre las características de este tipo de aplicaciones se encuentran las siguientes:

- Necesidad de alta productividad en el desarrollo de la aplicación.
- Integración con los sistemas existentes.
- Libertad de elección de plataformas de desarrollo y producción.
- Escalabilidad.
- Modelos flexibles de seguridad [18].

J2EE como plataforma de desarrollo aporta librerías, la mayoría de las cuales son gratuitas y de código abierto. Además esta tecnología es multiplataforma.

Debido a que la tecnología JSP es abierta y multiplataforma, los servidores web, plataformas y otros componentes pueden ser fácilmente actualizados o cambiados sin que afecte a las aplicaciones basadas en la tecnología JSP.

1.5.2 Lenguajes de desarrollo del software.

Para desarrollar aplicaciones web se podrían seleccionar numerosos lenguajes de programación usados en el mundo informático como son C#, Java y php, los cuales son comparados desde el punto de vista del desarrollador de una aplicación, en cuanto a la seguridad, a la velocidad y a la robustez entre otros factores.

Dada las características de robustez de java, de multiplataforma, además de ser software libre, es muy usado en el mundo del desarrollo de aplicaciones por lo que se estandarizó en el proyecto así como en el centro que trabaja conjunto con el equipo de trabajo al uso de java como lenguaje de programación.

Por la creciente importancia de exportación de funcionalidades a través de servicios web el cliente debe obtener la información necesaria para interactuar con dichos servicios. Para dar solución a este

problema se crea el lenguaje de descripción de servicios Web (WSDL, Web Service Description Language) el cual es utilizado por el sistema, dado su extensibilidad en la exportación de funcionalidades.

Hoy en día, UML está consolidado como el lenguaje estándar en el análisis y diseño de sistemas de cómputo en la programación orientada a objetos. En el proyecto se usa porque plasma un sistema de software previo al proceso de escribir el código.

1.5.2.1 Lenguaje JAVA.

Java ofrece todas las ventajas de un lenguaje potente y robusto, pues fue diseñado para crear software altamente fiable. Es un lenguaje de programación basado en clases y orientado a objetos. Sus características de memoria liberan a los programadores de responsabilidades y errores. Java es compilado en un código intermedio más abstracto que el código de máquina que es ejecutado por la máquina virtual de Java (J2EE).

Una de las características más significativas de Java es que posee una arquitectura neutral, es decir, el compilador Java compila su código a un fichero objeto de formato independiente de la arquitectura de la máquina en que se ejecutará. La independencia de la arquitectura representa solo una parte de su portabilidad. Java especifica los tamaños de sus tipos de datos básicos y el comportamiento de sus operadores aritméticos, de manera que los programas son iguales en todas las plataformas.

1.5.2.2 WSDL (Web Services Description Language).

WSDL (Web Services Description Language) es un dialecto que describe la interfaz pública a los servicios Web. Está basado en XML y describe la forma de comunicación, es decir, los requisitos del protocolo y los formatos de los mensajes necesarios para interactuar con los servicios listados en su catálogo. Las operaciones y mensajes que soporta se describen en abstracto y se ligan después al protocolo concreto de red y al formato del mensaje [19].

1.5.2.3 Lenguaje de Modelado (UML).

UML (Unified Modeling Language) es un lenguaje que permite modelar, construir y documentar los elementos que forman un sistema software orientado a objetos. Se ha convertido en el estándar de facto de la industria, debido a que ha sido impulsado por los autores de los tres métodos más usados de orientación a objetos: Grady Booch, Ivar Jacobson y Jim Rumbaugh. Estos autores fueron contratados por la empresa Rational Software Co. para crear una notación unificada en la que se pueda basar la construcción de sus herramientas CASE (Computer Aided Software Engineering) [20].

EL UML es una de las herramientas más gustada en el mundo para desarrollo de software, esta particularidad de su uso se debe a que permite a los desarrolladores de sistemas, generar diseños conforme a sus ideas de forma fácil y entendible.

1.5.3 Otras consideraciones generales para la selección de estas tecnologías

Se tuvo en cuenta otras tecnologías para la construcción del sistema, por lo que la selección se basó en una búsqueda de tecnologías para que la herramienta que se desarrolle cumpla con las condiciones establecidas por la sociedad de software libre. Además de cumplir los estándares establecidos tanto por el polo productivo de bioinformática como por centro que desarrolla conjunto con el proyecto (CIGB).

1.5.4 Herramientas de desarrollo utilizadas.

Existen varias herramientas para el desarrollo de un software, de ellas se hace uso de las herramientas CASE y las herramientas IDE.

1.5.4.1 Visual Paradigm.

Se puede definir a las Herramientas CASE como un conjunto de programas y ayudas que dan asistencia a los analistas, ingenieros de software y desarrolladores, durante todos los pasos y procesos del ciclo de vida de desarrollo de un proyecto [22].

Visual Paradigm para UML es una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue [22].

1.5.4.2 Rational Rose.

Es la herramienta CASE desarrollada por los creadores de UML que cubre todo el ciclo de vida de un proyecto: concepción y formalización del modelo, construcción de los componentes y certificación de las distintas fases. Permite una trazabilidad real entre modelo (análisis y diseño) y el código ejecutable.

Rational Rose es una herramienta con plataforma independiente que ayuda a la comunicación entre los miembros del equipo, a monitorear el tiempo de desarrollo y a entender el entorno de los sistemas. Una de las grandes ventajas es que utiliza la notación estándar en la arquitectura de software (UML), la cual permite a los arquitectos de software y desarrolladores visualizar el sistema completo utilizando un lenguaje común, además los diseñadores pueden modelar sus componentes e interfaces en forma individual y luego unirlos con otros componentes del proyecto.

1.5.4.3 Herramienta CASE utilizada en el desarrollo del sistema.

Se utiliza Visual Paradigm porque Rational Rose es una herramienta con plataforma independiente y el proyecto necesita una herramienta CASE que sea multiplataforma. Además Visual Paradigm es una herramienta que se caracteriza por tener gran capacidad de realizar ingeniería inversa.

1.5.4.4 NetBeans 6.5

NetBeans IDE es una herramienta para que los programadores puedan escribir, compilar, depurar y ejecutar programas. Está escrito en Java. Existe además un número importante de módulos para extender el NetBeans IDE. Es de código abierto y gratuito para uso tanto comercial como no comercial. El código fuente está disponible para su reutilización de acuerdo con la Common Development and Distribution License and the GNU General Public License [23].

1.5.4.5 Eclipse

Eclipse es una potente herramienta universal de entorno de desarrollo de software desarrollado en Java, usa java como lenguaje de programación aunque permite plugins para varios lenguajes más. La plataforma está construida en base a plugins, mecanismo que permite desarrollar, integrar y correr nuevos plugins. Otras características que aporta Eclipse son:

- Es una herramienta de código abierto.
- Eclipse es neutral y adaptable a cualquier tipo de lenguaje, por ejemplo C/C++, Cobol, C#, XML, etc. La característica clave de Eclipse es la extensibilidad.
- Soporta la programación orientada a objetos (POO) [24].

1.5.4.6 Herramienta IDE utilizada en el desarrollo del sistema.

Se utiliza Netbeans porque tiene un muy buen soporte de webapps (.war, jsp y servlets) del cual eclipse carece, los plugins de eclipse no son ni tan potentes, ni tan sencillos como el módulo que en Netbeans viene preinstalado. Lo cierto es que ambos IDE se van acercando en funcionalidades conforme maduran, pero netbeans es más intuitivo para realizar aplicaciones web con creación y utilización de webservices [25].

1.5.5 Gestor de Base de Datos

En cualquier proyecto informático para una selección de la base de datos a utilizar se requiere de un análisis profundo sobre cuál gestor utilizar. En la aplicación se analizan principalmente dos gestores.

1.5.5.1 Base de datos PostgreSQL

PostgreSQL es un servidor de base de datos relacional orientada a objetos de software libre. Está considerado como la base de datos de código abierto más avanzada del mundo [23]. PostgreSQL aproxima los datos a un modelo objeto-relacional, y es capaz de manejar complejas rutinas y reglas. Ejemplos de su avanzada funcionalidad son las consultas SQL declarativas, el control de concurrencia

multi-versión, el soporte de multi-usuario, transacciones, optimización de consultas, herencia, y arreglos.

1.5.5.2 Base de datos MySQL

MySQL es un sistema de gestión de bases de datos relacional, soporta una gran carga de forma muy eficiente, que mantiene el código fuente del servidor SQL, así como también el score.

Este gestor de bases de datos es, probablemente, el gestor más usado en el mundo del software libre, debido a su gran rapidez y facilidad de uso. Esta gran aceptación es debida, en parte, a que existen infinidad de librerías y otras herramientas que permiten su uso a través de gran cantidad de lenguajes de programación, además de su fácil instalación y configuración [27].

1.5.5.3 Gestor de base de datos utilizado en el desarrollo del sistema.

En este proyecto se usa PostgreSQL además de las ventajas que posee, porque la estructura de la base de la tecnología del cliente presenta dicho gestor de base de datos y se debe de cumplir con sus peticiones. Además de que es uno de los gestores que estableció en el polo de Bioinformática [28].

Es un proyecto open source, por lo que el desarrollo de PostgreSQL no es manejado por una sola compañía sino que es dirigido por una comunidad de desarrolladores y organizaciones comerciales las cuales trabajan en su desarrollo. Dicha comunidad es denominada el PGDG (PostgreSQL Global Development Group). Implementa el uso de rollback's, subconsultas y transacciones, haciendo su funcionamiento mucho más eficaz, y ofreciendo soluciones en campos en las que MySQL no podría.

1.5.6 Servidor de Aplicaciones

Las funciones de todos los servidores de aplicaciones son similares pero no todos los servidores de aplicaciones brindan sus servicios con la misma velocidad, seguridad y confiabilidad. Por tanto se realiza un análisis relacionado con esto.

1.5.6.1 Tomcat 6.0

Tomcat 6.0 es un contenedor de Servlets con un entorno JSP. Tomcat puede funcionar como servidor web por sí mismo. Es usado mundialmente como servidor web autónomo en entornos con alto nivel de tráfico y alta disponibilidad [29].

Es necesario señalar que el cliente de este proyecto tiene JBoss como servidor de aplicaciones y este utiliza internamente un servidor Tomcat. En este trabajo se decide usar este último como servidor de aplicaciones para gastar menos recursos de cómputo debido a que no se cuenta con toda la infraestructura necesaria para el correcto funcionamiento de JBoss. Además de que la versión de Tomcat que se utiliza es totalmente compatible con la versión de JBoss que utiliza el cliente y cumple con las políticas del polo al que pertenece este proyecto.

1.5.6.2 GlassFish

Glassfish es un servidor de aplicaciones que implementa la plataforma JavaEE5, por lo que soporta las últimas versiones de tecnologías como: JSP, JSF, Servlets, EJBs, Java API para Servicios Web (JAX-WS), Arquitectura Java para Enlaces XML (JAXB), Metadatos de Servicios Web para la Plataforma Java 1.0, y muchas otras tecnologías. Existen personas que contribuyen al desarrollo del mismo, agregándole más características importantes a Glassfish.

1.5.6.4 JBoss

JBoss es un servidor de aplicaciones J2EE de código abierto implementado en Java puro, puede ser utilizado en cualquier sistema operativo. Ha sido reconocido como el líder en el mercado de los servidores de aplicaciones Java. JBoss ha evolucionado en un completo y robusto servidor de aplicaciones [30].

1.5.6.5 Servidor de aplicaciones utilizado en el desarrollo del sistema.

El servidor usado es JBoss porque es Open Source, Escalable, de Alto desempeño, Arquitectura Modular y orientada a servicios. Además de ser el más usado en el mundo tiene dentro una versión de

Tomcat y trabaja con una especificación Servlet y con una especificación de JSP. Dado que esto es lo que necesita el proyecto, es utilizado como servidor de aplicaciones.

1.5.7 Proceso de desarrollo de software

Para desarrollar una aplicación con calidad se hace necesario apoyarse en una metodología de trabajo ya conocida y después de un análisis se puede determinar cuál se apega más al proyecto.

1.5.7.1 RUP (Rational Unified Process)

El proceso unificado de desarrollo (RUP), es en la actualidad uno de los más usados por las empresas de software y es validado continuamente para el perfeccionamiento de su uso. Está concebido para que desde el inicio del proceso, se establezca una definición acertada del proyecto, haciendo innecesarias las reconstrucciones parciales posteriores. Además está dirigido a la programación orientada a objetos que permite obtener sistemas escalables en el tiempo que no necesitarán grandes inversiones de recursos en sus modificaciones posteriores.

Las características principales de este proceso son:

- Guiado por los Casos de Uso
- Centrado en la Arquitectura
- Iterativo e incremental

A través de un proyecto aplicando RUP, los requerimientos funcionales son expresados en la forma de Casos de Uso, que guían la realización de una arquitectura ejecutable de la aplicación. Además el proceso focaliza el esfuerzo del equipo en construir los elementos críticos estructuralmente y del comportamiento (llamados Elementos Arquitecturales) antes de construir elementos menos importantes.

Este proceso se basa en el modelo en espiral que organiza las iteraciones por etapas y fases para obtener una estructura más sólida, clara y ajustable a las necesidades particulares de cada organización, facilitando además la administración del proyecto. Cada iteración se considera un

subproyecto que no solo genera documentación, sino también productos de software, permitiendo con esto que el usuario tenga puntos de verificación y control más rápidos, y que se realice un proceso continuo de pruebas y de integración desde las primeras iteraciones.

Además controla, a través de un esquema predefinido de escalabilidad y gestión de riesgos, el manejo de los mismos, con lo cual posibilita que se reconozcan previamente los posibles problemas, y fallos, para su temprana prevención y corrección.

Presenta la particularidad de que, en cada ciclo de iteración, se hace exigente el uso de artefactos, siendo por este motivo, una de las metodologías más importantes para alcanzar un grado de certificación en el desarrollo del software y como su enfoque esta basado en modelos utiliza un lenguaje bien definido para tal fin, el UML [31].

1.5.7.2 OpenUp

OpenUp es un marco de trabajo para procesos de desarrollo de software. Fue liberado por el Eclipse Process Framework (EPF). OpenUp se construyó sobre una donación realizada por IBM del Basic Unified Process. Fue entregada a Eclipse a fines de 2005 y renombrado como OpenUP en 2006. Preserva la esencia del Unified Process. Está pensado para proyectos pequeños. [32]

1.5.7.3 Metodología utilizada en el desarrollo del sistema.

Para el desarrollo correcto de la aplicación se utiliza la metodología OpenUP. Esta excluye la mayoría de las partes opcionales de RUP e incluye varios elementos nuevos. Es un proceso ágil y ligero, que promueve al desarrollo del software sobre las buenas prácticas, haciéndolo un proceso pequeño y extensible si es necesario. Es muy usado en proyectos pequeños con características similares a la aplicación. Además esta metodología fue creada por lo desarrolladores de java y por tanto es ideal para los proyectos que utilicen este lenguaje de programación.

1.5.8 Servicio Web

Un servicio Web (en inglés **Web service**) es un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones. Distintas aplicaciones de software desarrolladas en lenguajes de programación diferentes, y ejecutadas sobre cualquier plataforma, pueden utilizar los servicios web para intercambiar datos en redes de ordenadores como Internet. La interoperabilidad se consigue mediante la adopción de estándares abiertos. Las organizaciones OASIS y W3C son los comités responsables de la arquitectura y reglamentación de los servicios Web. Para mejorar la interoperabilidad entre distintas implementaciones de servicios Web se ha creado el organismo WS-I, encargado de desarrollar diversos perfiles para definir de manera más exhaustiva estos estándares [19].

1.5.9 SOA (Service Oriented Architecture)

La Arquitectura Orientada a Servicios (en inglés Service-Oriented Architecture o SOA), es un concepto de arquitectura de software que define la utilización de servicios para dar soporte a los requerimientos de software del usuario. SOA proporciona una metodología y un marco de trabajo para documentar las capacidades de negocio y puede dar soporte a las actividades de integración y consolidación [19]. La misma está formada por tres partes: un proveedor, un intermediario y un cliente que no presentan ningún acoplamiento entre ellos. El proveedor ofrece un servicio determinado y que el cliente no tiene por qué conocer directamente. El cliente aprende cómo utilizar el servicio a partir de la información que le ofrece el intermediario que normalmente simplifica el uso de dicho servicio. El cliente sólo sabe cómo utilizar el servicio, es decir, como enviar y recibir datos pero no conoce ningún detalle de su implementación interna.

1.6 Resumen de la tecnología, herramientas y metodologías a utilizar

Como resultado del estudio de las tecnologías y herramientas de desarrollo a utilizar se ha llegado a la conclusión de que el sistema se desarrollará sobre la plataforma J2EE usando como lenguaje de programación Java, y como gestor de bases de datos PostgreSQL 8.2. Se determinó además hacer uso del Lenguaje Unificado de Modelado (UML) por su estrecha integración con OpenUP, metodología adoptada. Como herramienta de modelado se utiliza Visual Paradigm. Como entorno de desarrollo integrado se decidió que debe ser NetBeans 6.5 y como servidor de aplicaciones se adoptó JBoos [2].

Se determinó realizar los web services con arquitectura orientada a servicios (SOA) en el lenguaje java.

1.7 Conclusiones

En el presente capítulo se analizaron las tecnologías, herramientas y metodologías para el desarrollo de la aplicación y se profundizó en algunos conceptos necesarios para la comprensión de la solución de este trabajo y la necesidad de la creación de una nueva aplicación dado el análisis de los sistemas actualmente utilizados para dar solución al problema del diseño de siRNA. Se detallaron limitantes de estos sistemas y por qué se realiza una nueva versión del sistema.

CAPÍTULO 2: CARACTERÍSTICAS DEL SISTEMA

Introducción

En el presente capítulo se describe el objeto de estudio, y se especifican en detalles los problemas existentes que dieron pie al desarrollo de la herramienta, se especifican los detalles de la construcción de la misma y la solución usando componentes del modelo del dominio de la metodología Open/Up, de este modelo se tendrán en cuenta la definición de las entidades y los conceptos principales. Se definen los requerimientos, tanto funcionales como no funcionales y se obtienen además, como resultado, el diagrama de casos de uso del sistema, las descripciones de los correspondientes casos de uso y el diagrama de objetos.

2.1 Objeto de estudio

De un análisis sencillo se puede definir que dentro de la biología molecular una de las esferas es la Genética. El sistema a desarrollar tiene como objeto de estudio las aplicaciones cliente-servidor para el diseño de siRNA con el objetivo de detallar su funcionamiento y ventajas. Esto proporcionará un mejor entendimiento en el funcionamiento de la aplicación y en su desarrollo.

El sistema que se propone solo tiene relación con la versión anterior al mismo, a la cual se le incluirá algunas mejoras y se le cambiará la arquitectura del mismo a petición del cliente.

2.1.1 Objetivos estratégicos de la organización.

En la actualidad, existe una gran cantidad de organizaciones que centran sus investigaciones al campo de la genética. Cuba se ha sumado al estudio de genes, el CIGB es una de las instituciones que ha alcanzado un alto nivel en las investigaciones dirigidas al desarrollo, producción y comercialización de las mismas, contribuyendo directamente al desarrollo económico y social. Su impacto está destinado al mejoramiento de la salud humana, las producciones agropecuarias, acuícolas, y al medio ambiente [2]. Debido a las numerosas investigaciones que se llevan a cabo ha

surgido la necesidad de crear un sistema capaz de predecir los sitios blancos para silenciar genes, pues esto es de gran utilidad en la curación de enfermedades tanto humanas como en animales y plantas.

2.1.2 Flujo actual de los procesos.

Actualmente los procesos fluyen de la siguiente manera:

El especialista a través de su navegador web accede a la versión 1.1, de esta forma puede obtener un siRNA, formando una secuencia a través de el id del gen, nombre del transcrito o nombre del gen o a través de una secuencia de mRNA entrada por dicho especialista. Luego que el sistema le muestra los sitios, el especialista diseña en el laboratorio los siRNA que actúan con estos sitios blancos.

2.2 Procesos objeto de automatización.

En el proceso de estudio e investigaciones sobre el análisis de secuencias de mRNA para el diseño de siRNA se desea automatizar todo lo referente a la información de los genes, cromosomas, así como las reglas que se necesitan para el análisis.

2.2.1 Procesos que serán objeto de automatización.

Las secuencias de mRNA se pueden analizar de varias formas. La segunda versión del sistema se centrará principalmente en la búsqueda de sitios blancos de la acción del siRNA tomando en cuenta los SNP. Incluirá el algoritmo desarrollado para esta búsqueda implementado en la versión anterior y añadirá nuevas funcionalidades relacionadas con seleccionar solo los sitios targets comunes a todos los transcritos de un gen determinado o buscar sitios targets específicos de un transcrito en particular, además de la incorporación del cálculo de energía libre en la unión del siRNA con su estructura secundaria.

En el mundo existe una gran tendencia a exportación de funcionalidades con servicios web y la aplicación tendrá todas sus funcionalidades con esta arquitectura orientada a servicios.

Análisis simple:

Primeramente se necesita una secuencia de mRNA que será objeto del análisis. Puede ser entrada por el usuario o resultar de consultas realizadas por el mismo a la base de datos. En este último caso, se necesitaría la información correspondiente al criterio de búsqueda seleccionado (nombre para el caso de un transcrito y símbolo o id de un gen). Cuando el especialista solicita la búsqueda de un gen, luego de obtener los datos correspondientes, se procede a buscar los transcritos asociados a este. Posteriormente se muestra de forma gráfica los datos de dichos transcritos para luego dar la opción al especialista de seleccionar uno de estos y realizar el análisis. Luego de haber mostrado los datos del transcrito solicitado se da la posibilidad de realizar el análisis de la secuencia de los nucleótidos del mismo.

También de estos transcritos que están asociados a un gen determinado se puede realizar análisis de gran especificidad como buscar la secuencia común a todos los transcritos de un gen determinado, así como buscar la secuencia específica de un transcrito especificado.

Después de realizada cualquiera de estas acciones, el especialista debe especificar las reglas necesarias para llevar a cabo el proceso de análisis para obtener los sitios blancos. Estas reglas son:

- La secuencia comienza con AA.
- Tiene T en la posición 10.
- Tiene una A en la posición 3.
- Tiene una A en la posición 19
- Contenido de CG de 30-50% (se da la posibilidad de escoger a conveniencia los porcentajes).
- Ausencia de GC en la posición 19.
- Ausencia de G en la posición 13.
- Tiene 3 o más A/T entre las bases 15 y 19.
- De las reglas seleccionadas se desea que se cumplan al menos un número específico seleccionado.

Como resultado del análisis se obtiene una representación gráfica de los sitios blancos que corresponden a la cadena de nucleótidos, incluyendo aquellos que contienen SNP si esa opción es especificada anteriormente por el especialista. También se da la posibilidad de visualizar de forma individual la información de cada sitio blanco [2].

Análisis teniendo en cuenta la opción incluir SNP:

Al hacer el análisis anteriormente descrito se tienen en cuenta los SNP. Estos, localizados en los exones, son la primera causa de variabilidad genética; producen cambios en el código que altera la cadena de aminoácidos [17]. Debido a que son porciones de código que implican una diferencia genética, no se pueden tener en cuenta para el diseño de los siRNA, pues estas variaciones en la secuencia del ADN pueden afectar la respuesta de los individuos a enfermedades, bacterias, virus, productos químicos, fármacos. En el resultado se muestran los sitios que contienen SNP, además de la secuencias con los SNPs incluidos [2].

Búsqueda de miRNA:

Se buscan y muestran gráficamente los miRNA que actúan en determinado gen señalando la posición donde actúan. Se visualizan además datos significativos del miRNA. También se brinda la opción de mostrar solamente los de un valor de significancia mínimo seleccionado en adelante.

2.3 Propuesta de sistema

2.3.1 Descripción general de la propuesta de sistema

Con el objetivo de mejorar la calidad de las investigaciones aplicadas y el desarrollo de los trabajos implementados en el CIGB, referentes al diseño de siRNA, se ha decidido desarrollar una aplicación que se enfoca en cómo llevar a cabo de manera rápida y eficaz el estudio de los genes, específicamente para lograr un diseño de siRNA más eficiente, estable y seguro que conlleve a significativos aportes en los descubrimientos científicos, además de que tiene que tener sus funcionalidades en servicios web para mayor utilización de las mismas. Estas funcionalidades para alcanzar este nivel son las siguientes:

- **Analizar secuencias mRNA:** Permitirá analizar una secuencia mRNA, introducida por el especialista o formada a partir de la información de genes y transcritos almacenada en la base de datos. Brindará la posibilidad de elegir las reglas deseadas para el análisis. Luego de esta operación se mostrará la localización de los sitios blancos, así como sus datos (representación gráfica, secuencia, posiciones inicial y final, la energía libre y el número de reglas que se cumplen) [2].
- **Mostrar Genes:** Se brindará la posibilidad de visualizar la información referente al gen buscado, con sus transcritos asociados y los datos más importantes de estos; además representa de forma gráfica los datos de los miRNA que actúan sobre los mismos, y en qué región lo hace.
- **Mostrar Transcritos:** Permitirá a los especialistas conocer la información referente a los transcritos. Mostrando además los datos de los exones asociados y entre estos, las regiones codificantes y las no traducidas.
- **Buscar Secuencia Específica de un Transcrito:** Permitirá a los especialistas conocer la información referente a los exones específicos de un transcrito determinado lo cual después se le puede realizar un análisis de secuencia para mayor especificidad en la investigación.
- **Buscar Secuencia Común de Transcritos de un Gen:** Permitirá a los especialistas conocer la información referente a los exones comunes de un gen determinado lo cual después se le puede realizar un análisis de secuencia para mayor especificidad en la investigación.

2.3.2 Análisis comparativo de otras soluciones existentes con la propuesta

Teniendo en cuenta otras soluciones existentes que se encuentran online en internet, se puede decir, que la propuesta que brinda esta investigación es la más acertada como solución al problema planteado por el especialista por las siguientes razones:

1. Mantiene la independencia de internet lograda con la aplicación anterior lo cual favorece el avance en las investigaciones.
2. Se brindan funcionalidades a partir de recomendaciones y exigencias derivadas de la experiencia de los investigadores.
3. Cuenta con algoritmos propios, bien conocidos por los usuarios finales.
4. Tiene una conexión a una base de datos segura, con gran confiabilidad y disponibilidad de la información.
5. Brinda servicios web para exportar las funcionalidades a cualquier investigador, tanto en el CIGB como en el resto del mundo.

Esta versión incluye funcionalidades en comparación con la solución anterior así como una arquitectura diferente (SOA) y más eficiente que brindará servicios, a través de web services, para mejorar los requerimientos de los especialistas. Se calcula la energía libre de los sitios blancos. Se adiciona la implementación de un algoritmo para formar secuencia que sea común a todos los transcritos y otro para formar secuencia de un transcrito determinado. El sistema quedará concebido como una herramienta potente, capaz de brindar servicios de calidad y eficiencia a los investigadores. Además de brindar dos servicios que no serán usados en la versión 2.0 de alasSiRNA - Design y que serán de gran ayuda para los investigadores y desarrolladores del CIGB. Estos servicios son los de mostrar SNPs que se encuentren en determinada región del cromosoma y mostrar SNPs que coinciden con posibles sitios targets del miRNA.

2.4 Modelo del Dominio

Un modelo del dominio captura los tipos más importantes de objetos en el contexto del sistema. Los objetos del dominio representan los eventos que suceden en el entorno en el que trabaja el sistema. La modelación del dominio tiene como objetivo fundamental la comprensión y descripción de las clases más importantes en el sistema.

Debido a que se hace difícil encontrar una estructura de los procesos de negocios que tienen que ver con el objeto de estudio, y haciéndose necesario poder entender el funcionamiento de la aplicación, se necesita describir mediante una serie de conceptos, entidades y sus relaciones, agrupados en un modelo del dominio para el logro de un fácil entendimiento de la aplicación.

2.4.1 Reglas del negocio.

Las reglas del negocio permiten definir los límites lógicos de las aplicaciones. En ellas las aplicaciones reflejan las restricciones que existen, evitando la realización de acciones no válidas. A continuación se enumeran las reglas definidas para el sistema:

1. Tener la información necesaria para el análisis de la secuencia de mRNA.
2. Seleccionar las reglas de cumplimiento para obtener el análisis de la secuencia de mRNA.

Definición de las entidades y los conceptos principales.

Con la aplicación de Modelo del dominio se detectaron las siguientes entidades y conceptos (objetos):

alasSiRNA - Design v1.1: Versión de la aplicación para el diseño de siRNA.

Diseño_siRNA: Es el resultado del análisis de secuencias mRNA.

Buscar_MiRNA: Es el resultado de la búsqueda de los miRNAs al buscar un gen

Buscar_Gen: Es la búsqueda de un gen a partir de un criterio dado.

Buscar_Transcrito: Es el resultado de los transcritos al buscar un gen y también se puede buscar individual.

Buscar_SNP: Busca todos los SNPs relacionados a un transcrito determinado.

Diagrama Modelo del Dominio.

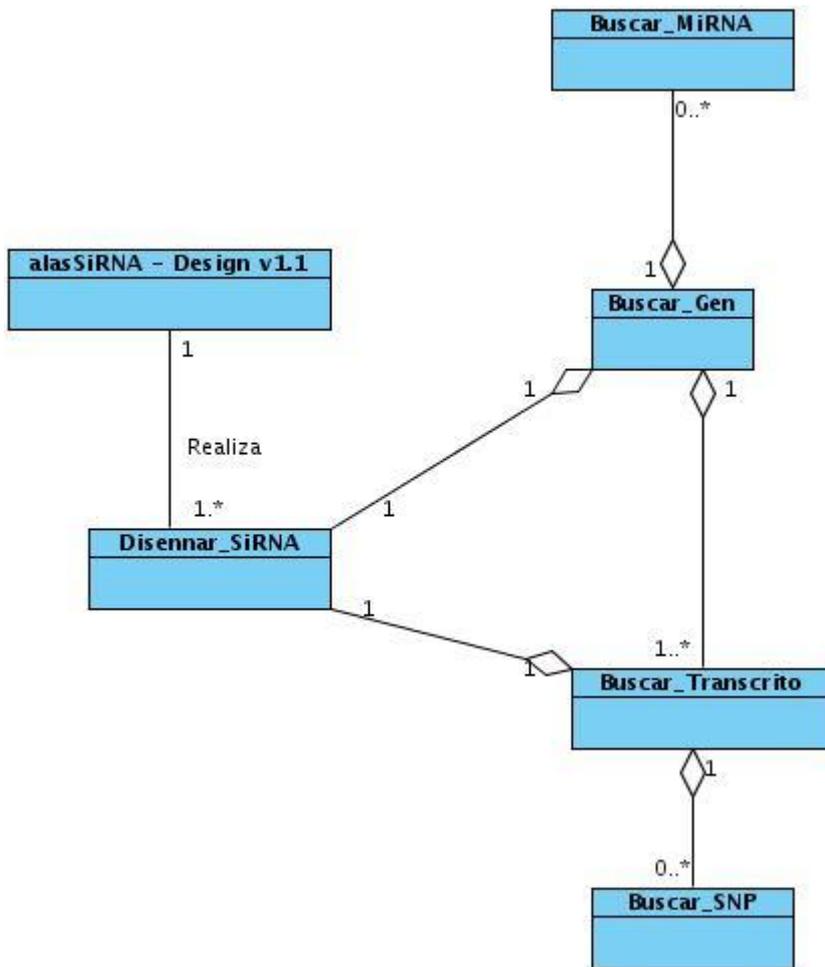


Fig. 4 Modelo del Dominio

2.5 Especificación de los requisitos de software

Para la especificación de los requisitos el cliente debe tener en cuenta además de las funcionalidades o requisitos funcionales del software los requisitos no funcionales del software.

2.5.1 Requerimientos Funcionales

Los requerimientos funcionales son capacidades o condiciones que el sistema debe cumplir. Permiten expresar una especificación más detallada de las responsabilidades del sistema en cuestión. Los requerimientos funcionales del software sombreados son nuevos en esta versión.

R 1.1: Buscar datos de un gen.

R 1.2: Visualizar datos de un gen.

R 1.3: Buscar datos de transcritos asociados al gen.

R 1.4: Visualizar datos de transcritos asociados al gen.

R 1.5: Buscar datos de los miRNA asociados al gen.

R 1.6: Visualizar datos de los miRNA asociados al gen.

R 2.1: Buscar los datos de un transcrito indicado.

R 2.2: Visualizar gráficamente los datos de un transcrito indicado.

R 2.3: Buscar los SNP asociados a un transcrito.

R 2.4: Visualizar gráficamente los SNP asociados a un transcrito

R 3.1: Buscar sitios blancos en la secuencia mRNA.

R 3.2: Mostrar gráfica de la secuencia de mRNA resaltando sitios blancos ordenados por reglas cumplidas.

R 3.3: Mostrar SNP asociados a la secuencia de mRNA.

R 3.4: Mostrar datos de un sitio blancos indicado.

R 3.5: Calcular energía libre al RNA complementario.

R4: Formar una secuencia que sea común a todos los transcritos

R5: Formar una secuencia única de un transcrito determinado

R6: Buscar SNPs que se encuentren en determinada región de un cromosoma.

R7: Buscar SNPs que coinciden con sitios target de miRNA.

2.5.2 Requerimientos no funcionales

Los requerimientos no funcionales son propiedades o cualidades que el producto debe tener. Representan las características del mismo.

Apariencia o interfaz externa:

La herramienta deberá tener un diseño de interfaz amigable y agradable, de forma tal que el usuario haga uso de la misma sin dificultad alguna, ajustándose a los estándares establecidos para el desarrollo de un buen diseño. Las diferentes interfaces tienen que estar redactadas en el idioma inglés ya que este es el idioma más idóneo para los estudios biológicos.

Usabilidad:

El sistema brindará la posibilidad de ser usado por cualquier persona cuyos conocimientos en cuanto al trabajo con los ordenadores sean básicos, solo es necesario contar con conocimientos especializados en biología para poder entender los resultados brindados por la aplicación.

Rendimiento:

El sistema está concebido para un ambiente cliente/servidor, que garantiza que la eficiencia del mismo en cuanto a la rapidez de respuesta ante las solicitudes de los usuarios sea la mejor, al igual que la velocidad de procesamiento de la información. Para ello se realiza la validación de los datos de entrada y la manipulación de eventos en el cliente y en el servidor, las que por cuestiones de seguridad o de acceso a los datos lo requieran. De esta forma se logra una mayor velocidad de procesamiento y un mayor aprovechamiento de los recursos.

SopORTE:

El sistema debe dar la posibilidad de mejoramiento y su ampliación a otras especificaciones que se le incorporen en un futuro. Además debe ser de fácil instalación, adaptable a numerosas plataformas y de fácil mantenimiento.

Portabilidad:

El sistema está concebido para que sea multiplataforma.

Confiabilidad:

El sistema debe brindar el máximo de confiabilidad y precisión de la información al usuario, para evitar errores.

Software:

Se debe disponer de Windows 95 (o versiones superiores) o Linux para la instalación de la aplicación. La aplicación se realizará en un ambiente Web, la base de datos es independiente de la aplicación. Se debe disponer de un navegador web cualquiera.

Hardware:

Para el desarrollo y puesta en práctica del proyecto se requieren máquinas servidoras con los siguientes requisitos:

Para el servidor de aplicaciones:

Procesador Pentium 3 o superior.

512 MB de RAM.

Para el servidor del Web Services:

Procesador Pentium 3 o superior.

512 MB de RAM.

Un mínimo de 4 GB de espacio disponible en disco duro.

Para el servidor de base de datos.

Procesador Pentium 3 o superior.

512 MB de RAM.

Un mínimo de 2 GB de espacio disponible en disco duro

Para las PC clientes se necesita:

Un procesador Pentium 3 o superior.

2.6 Definición de actores y casos de uso del sistema

Un actor es una entidad externa al sistema que interactúa con él. Siempre se beneficia de la realización de un caso de uso. Estimula el sistema con eventos de entrada y salida. Es un rol de un usuario que puede intercambiar información o puede ser un recipiente pasivo de información. Representa un ser humano, una máquina o un software que interactúa con el sistema.

Tabla.1 Actores del Sistema

Nombre del actor	Descripción
Especialista	Representa el usuario de la versión 2.0 de alasSiRNA - Design teniendo la posibilidad de interactuar con funcionalidades específicas del web services.
Aplicación_Cliente	Representa el sistema que accederá a los servicios del web services y que podrá obtener todas sus funcionalidades.

2.7 Casos de Uso definidos

Los casos de uso son “fragmentos” de funcionalidad que ofrece el sistema para aportar un resultado de valor para sus actores. Son una secuencia de acciones a llevar a cabo por el sistema como respuesta a los requerimientos funcionales. Se utilizan para obtener información de cómo debe trabajar el sistema. Describen el comportamiento del sistema mediante el intercambio acción y reacción.

En la elaboración de los casos de uso del sistema se tiene en cuenta por que parte se esta haciendo el caso de uso ya que no tiene la misma secuencia de pasos lógicos. En esta caso se realizan por la parte de la aplicación siRNA Web Services y por parte de alasSiRNA - Design v2.0.

Los casos de usos que se señalan con sombreado son los que incluye la nueva versión.

Los casos de uso definidos para alasSiRNA - Design v2.0 son los siguientes:

1. Realizar análisis de las secuencias mRNA.
2. Mostrar datos de un gen.
3. Mostrar datos de un transcrito.
4. Mostrar datos de miRNAs.
5. Mostrar la secuencia común a un gen
6. Mostrar la secuencia específica de un transcrito

Tabla.2 CU Realizar análisis de la secuencia mRNA

CU 1	Realizar análisis de la secuencia mRNA
Actor	Especialista
Descripción	Obtiene y muestra un diseño de siRNA a partir de un adecuado análisis de la secuencia mRNA que se haya seleccionado.
Referencia	R 3.1, R 3.2, R 3.3, R 3.4

Tabla.3 CU Mostrar datos de un gen

CU 2	Mostrar datos de un gen.
Actor	Especialista
Descripción	Obtiene y muestra información de un determinado gen.
Referencia	R 1.1, R 1.2, R 1.3, R 1.4.

Tabla.4 CU Mostrar datos de un transcrito.

CU 3	Mostrar datos de un transcrito.
-------------	---------------------------------

Actor	Especialista
Descripción	Obtiene y muestra información relevante de un determinado transcrito.
Referencia	R 2.1, R 2.2, R 2.3, R 2.4.

Tabla.5 CU Mostrar datos de miRNAs.

CU 4	Mostrar datos de miRNAs.
Actor	Especialista
Descripción	Obtiene y muestra información sobre los miRNAs de un determinado gen.
Referencia	R1.5.

Tabla.6 CU Mostrar la secuencia común de un gen.

CU 5	Mostrar la secuencia común de un gen
Actor	Especialista
Descripción	Obtiene y muestra la secuencia común a todos los transcritos de un gen determinado.
Referencia	R4

Tabla.7 CU Mostrar la secuencia específica de un transcrito determinado.

CU 6	Mostrar la secuencia específica de un transcrito
Actor	Especialista
Descripción	Obtiene y muestra la secuencia específica a un transcrito determinado.
Referencia	R 5

Los casos de uso definidos para siRNA Web Services son los siguientes:

1. Realizar análisis de las secuencias mRNA.
2. Buscar datos de un gen.
3. Buscar datos de un transcrito.
4. Calcular Energía Libre
5. Realizar secuencia común a un gen
6. Realizar secuencia específica de un transcrito

7. Listar SNPs que estén en determinada región del cromosoma.
8. Listar SNPs que coinciden con los miRNAs.

Tabla.8 CU Realizar análisis de la secuencia mRNA

CU 1	Realizar análisis de la secuencia mRNA
Actor	Aplicacion_Cliente
Descripción	Obtiene un diseño de siRNA a partir de un adecuado análisis de la secuencia mRNA que se haya escogido.
Referencia	R 3.1.

Tabla.9 CU Buscar datos de un gen

CU 2	Buscar datos de un gen.
Actor	Aplicacion_Cliente
Descripción	Obtiene información de un determinado gen.
Referencia	R 1.1, R 1.5.

Tabla.10 CU Buscar datos de un transcrito.

CU 3	Buscar datos de un transcrito.
Actor	Aplicacion_Cliente
Descripción	Obtiene información relevante de un determinado transcrito.
Referencia	R 2.1, R 2.2, R 2.3, R 2.4.

Tabla.11 CU Calcular Energía.

CU 4	Calcular Energía Libre
Actor	Aplicacion_Cliente
Descripción	Calcula la energía libre del sitio target al realizar el caso de uso: Realizar análisis de la secuencia mRNA
Referencia	R3.5.

Tabla.12 CU Realizar secuencia común de un gen.

CU 5	Realizar secuencia común de un gen
Actor	Aplicacion_Cliente
Descripción	Obtiene la secuencia común a todos los transcritos de un gen determinado.
Referencia	R4

Tabla.13 CU Realizar secuencia específica de un transcrito

CU 6	Realizar secuencia específica de un transcrito
Actor	Aplicacion_Cliente
Descripción	Obtiene la secuencia específica a un transcrito determinado.
Referencia	R 5

Tabla.14 CU Listar SNPs que estén en determinada región del cromosoma.

CU 7	Listar SNPs que estén en determinada región del cromosoma.
Actor	Aplicacion_Cliente
Descripción	Obtiene la lista de los SNPs en una determinada región del cromosoma.
Referencia	R 6

Tabla.15 CU Listar SNPs que coinciden con los miRNAs.

CU 8	Listar SNPs que coinciden con los miRNAs.
Actor	Aplicacion_Cliente
Descripción	Obtiene la lista de los SNPs en una determinada región del cromosoma que coinciden con algún miRNAs.
Referencia	R 7

2.8 El diagrama de casos de uso para alasSiRNA Desing v2.0:

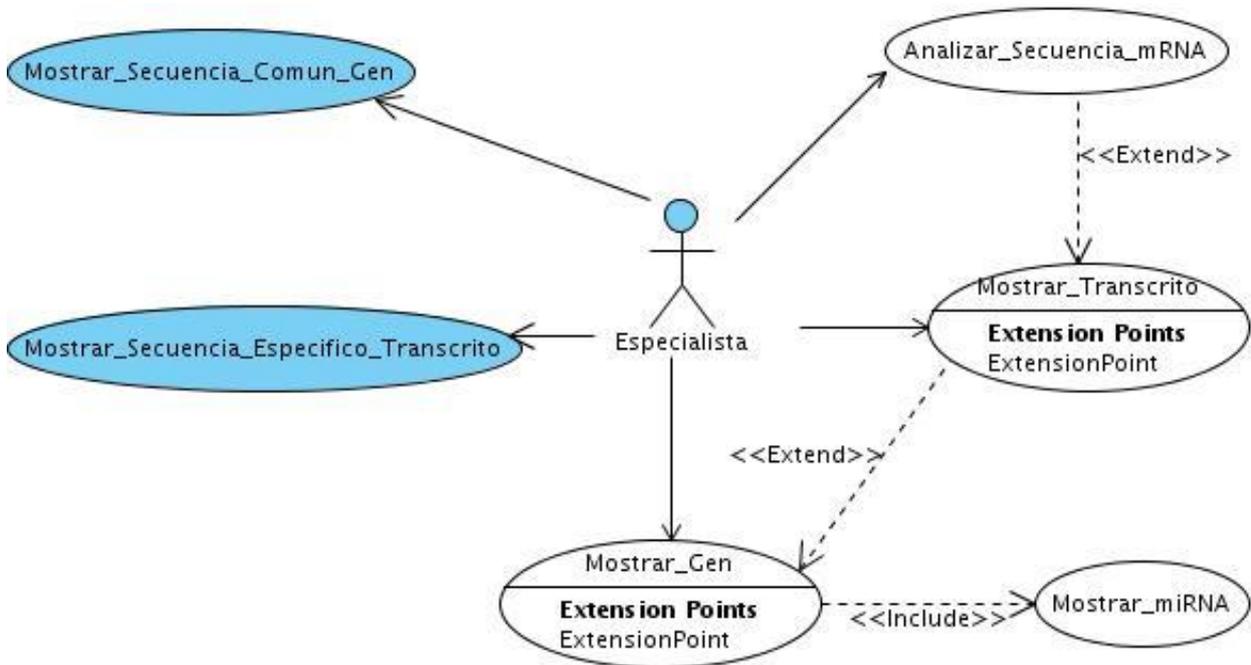


Fig. 5 Diagrama de casos de uso del sistema de alasSiRNA - Design v2.0

2.9 El diagrama de casos de uso para siRNA Web Services:

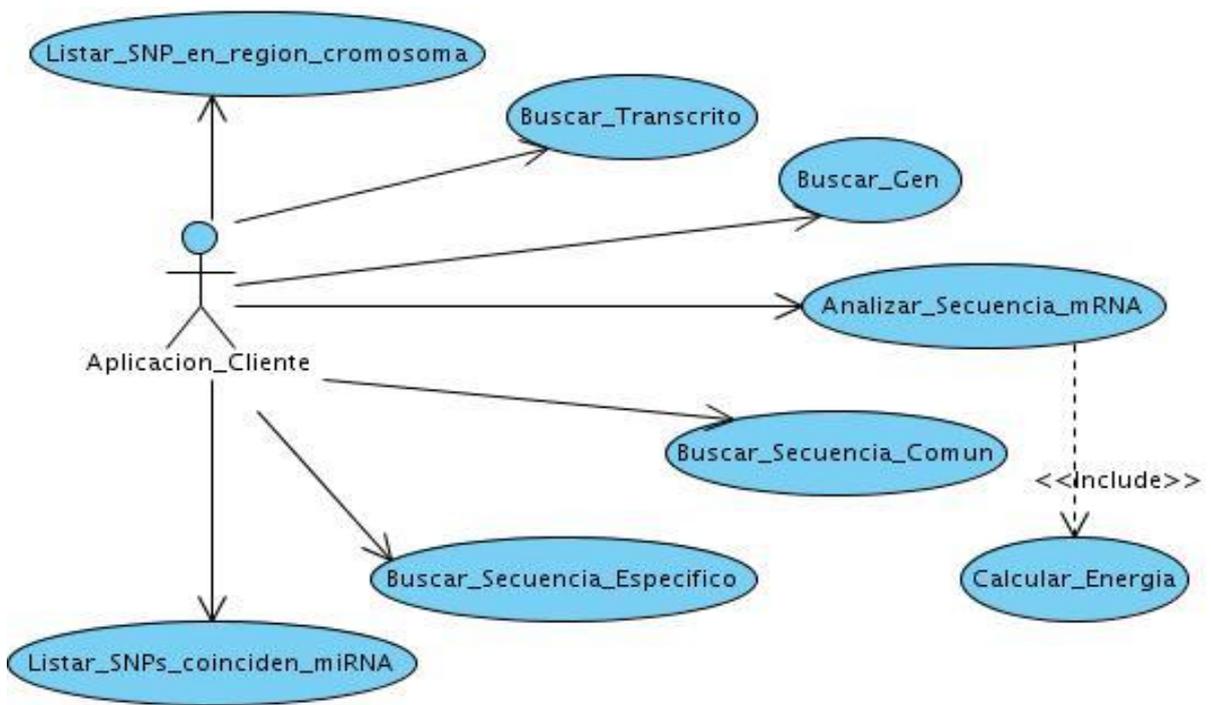


Fig. 6 Diagrama de casos de uso del sistema de siRNA Web Services.

2.10 Especificación de los Casos de Uso

En la especificación de casos de uso se tiene en cuenta por qué parte se está haciendo el caso de uso ya que no tiene la misma secuencia de pasos lógicos.

Caso de uso Analizar_Secuencia_mRNA. alasSiRNA - Design v2.0

Tabla.16 Descripción del CU Analizar_Secuencia_mRNA.

Caso de Uso:	Analizar_Secuencia_mRNA.	
Actores:	Especialista.	
Propósito:	Obtener un diseño de siRNA a partir de un adecuado análisis de la secuencia mRNA que se haya seleccionado.	
Resumen:	El caso de uso se inicia cuando luego de haber buscado y mostrado determinado gen o transcrito, el especialista decide analizar la secuencia encontrada o accede a la opción "Paste your own sequence", procede a analizarla seleccionando las reglas y el tipo de análisis deseado, para de esta manera obtener el diseño de siRNA esperado.	
Referencia:	R 3.1, R 3.2, R 3.3, R 3.4	
Precondiciones:	Haber realizado el CU Mostrar_Gen. Haber realizado el CU Mostrar_Transcrito	
Poscondiciones:		
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Sistema	
1) El especialista accede a la opción "Paste your own sequence".		
2) El especialista selecciona las reglas para realizar el análisis de la secuencia mRNA.		
3) El especialista pega la secuencia de mRNA o selecciona la opción "Sample"		

4) El especialista selecciona la opción de "Analyze".	5) Busca los sitios blancos y los SNP pertenecientes a la secuencia de mRNA en el web services. 6) Muestra la secuencia mRNA resaltando sitios blancos, sus datos y los SNP que coinciden con los sitios blancos. 7) Muestra una gráfica con los SNP y finaliza así la ejecución del caso de uso.
Flujos Alternos	
Acción del Actor	Respuesta del Sistema
	5.1) El web services analiza la secuencia y esta no presenta SNP. El sistema no muestra la gráfica con SNP.
Prioridad:	Crítico.
Escenario "Analizar secuencia sin SNP"	
Acción del Actor	Respuesta del Sistema
8) El especialista selecciona la opción de "Exclude SNPs region from the design (No For Paste Sequence)".	8.1) Muestra la secuencia mRNA resaltando sitios blancos, sus datos y excluye de la gráfica aquellos que contienen SNP.

Caso de uso Mostrar_Gen. alasSiRNA Desing v2.0

Tabla.17 Descripción del CU Mostrar_Gen.

Caso de Uso:	Mostrar_Gen.
Actores:	Especialista.
Propósito:	Mostrarle información relevante a los especialistas.
Resumen:	El caso de uso se inicia cuando el especialista selecciona una de las opciones de búsqueda de secuencias ('By gene id' o 'By gene simbol') y desea conocer información sobre los genes, sin necesidad de realizar un análisis previo. Entonces especifica un criterio de búsqueda, a partir del

	mismo se busca en el web services.
Referencia:	R 1.1, R 1.2, R 1.3, R 1.4, R 1.5, R 1.6.
Precondiciones:	
Poscondiciones:	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1) El especialista selecciona el criterio (By gene id o By gene simbol).	2) El sistema brinda la opción de introducir la información a partir del criterio de búsqueda seleccionado.
3) El especialista introduce el criterio y selecciona la opción "Search".	4) Sistema verifica si los datos están correctos 5) Busca datos del gen en el web services 5) Muestra datos de gen (Chromosome, Gene id, Symbol y los transcritos asociados). 7) Muestra datos de los transcritos asociados al gen. 8) Muestra los miRNA correspondientes al gen (start, end, nombre, Pvalue, Score) y finaliza así la ejecución del caso de uso.
Flujos Alternos	
Acción del Actor	Respuesta del Sistema
	4.1) Existen datos incorrectos. El sistema muestra el mensaje: "Gene which Id is 'a' doesn't exist. Check that all the charaters are numbers" y brinda la posibilidad de introducir nuevamente los datos.
	8.1) El sistema analiza la secuencia y esta no presenta miRNA. El sistema no muestra los miRNA
Prioridad:	Crítico.

Caso de uso Mostrar_Transcrito. alasSiRNA - Design v2.0

Tabla.18 Descripción del CU Mostrar_Transcrito.

Caso de Uso:	Mostrar_Transcrito.
Actores:	Especialista.
Propósito:	Mostrarle información relevante a los especialistas.
Resumen:	El caso de uso se inicia cuando luego de haber generado a partir de la especificación de un criterio de búsqueda los datos del gen requerido se muestran los transcritos asociados al mismo, el especialista selecciona un transcrito y del mismo se muestra su información.
Referencia:	R 2.1, R 2.2, R 2.3, R 2.4.
Precondiciones:	Haber realizado CU Mostrar_Gen.
Poscondiciones:	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1) El especialista escribe el nombre del transcrito.	2) El sistema pregunta al web services que el nombre del transcrito exista. 3) Busca los transcritos y exones asociados al transcrito en el web services. 4) Muestra Transcript Name y Gene id 5) Muestra la cantidad de exones. 6) Muestra una gráfica con datos de los exones asociados al transcrito. 7) Muestra una gráfica de los SNP presentes en el transcrito y finaliza la ejecución del caso de uso.
Flujos Alternos	
Acción del Actor	Respuesta del Sistema
	2.1) Existen datos incorrectos. El sistema muestra el mensaje de error:” Results: The transcripts which name is 'name' doesn't exist. Check that the letters are capitalized.” y brinda la posibilidad de introducir nuevamente los datos.
	7.1) El sistema no muestra SNP.

Prioridad:	Crítico.
-------------------	----------

Caso de uso Mostrar_miRNA. alasSiRNA - Design v2.0

Tabla.19 Descripción del CU Mostrar_miRNA.

Caso de Uso:	Mostrar_miRNA	
Actores:	Especialista.	
Propósito:	Mostrar los miRNA que silencian determinado gen identificado.	
Resumen:	El caso de uso se inicia cuando el especialista solicita al sistema buscar un gen determinado este muestra sus datos de los miRNA correspondientes.	
Referencia:	R 1.5, R 1.6.	
Precondiciones:	Haber realizado el CU Mostrar_Gen	
Poscondiciones:		
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Sistema	
3) El especialista elige un nuevo valor de p-value y accede a la opción "Show".	1) Muestra los datos de los miRNA que actúan sobre un gen determinado (id, pvalue, score) con un nivel de significancia menor que el predefinido. 2) Brinda la opción de cambiar el p-value. 4) El web services busca los datos de los miRNA que actúan sobre el gen con un nivel de significancia menor que el seleccionado por el especialista. 5) El sistema muestra el gen con los miRNAs de menor significancia.	
7) El especialista accede a la opción "Show all"	6) El sistema muestra la opción "Show all" 8) El sistema muestra todos los miRNA que actúan sobre el gen.	
Flujos Alternos		
Acción del Actor	Respuesta del Sistema	
	2.1) El sistema no encuentra miRNA para el gen especificado	

	por lo que no la muestra.
Prioridad:	Crítico.

Caso de uso Mostrar_Secuencia_Especifico_Transcrito. alasSiRNA - Design v2.0

Tabla.20 Descripción del CU Mostrar_Secuencia_Específico_Transcrito.

Caso de Uso:	Mostrar_Secuencia_Especifico_Transcrito	
Actores:	Especialista.	
Propósito:	Mostrar las secuencias de nucleótidos que son específicas en determinado Transcrito.	
Resumen:	El caso de uso se inicia cuando un especialista solicita a la aplicación buscar una secuencia específica a un transcrito determinado.	
Referencia:	R5.	
Precondiciones:		
Poscondiciones:		
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Sistema	
1) El especialista elige el transcrito a analizar y la opción "Specific sequence of a transcript". 2) El especialista introduce el criterio de búsqueda, (nombre del transcrito)	3) Busca la secuencia específica para los transcritos en el web services. 4) El sistema muestra la posibilidad de analizar dicha secuencia.	
Prioridad:	Crítico.	

Caso de uso Mostrar_Secuencia_Comun_Gen. alasSiRNA Desing v2.0

Tabla.21 Descripción del CU Mostrar_Secuencia_Comun_Gen.

Caso de Uso:	Mostrar_Secuencia_Comun_Gen
Actores:	Especialista.
Propósito:	Mostrar las secuencias de nucleótidos que son comunes en todos los

	Transcritos de ese gen.
Resumen:	El caso de uso se inicia cuando el especialista solicita al sistema buscar una secuencia común en un gen determinado.
Referencia:	R4.
Precondiciones:	
Poscondiciones:	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1) El especialista selecciona la opción "Common sequence of the transcripts". 2) El especialista introduce el criterio de búsqueda, (gen id o gen symbol)	3) El sistema busca la secuencia común para los transcritos de ese gen a través del web services. 4) El sistema muestra la posibilidad de analizar dicha secuencia.
Prioridad:	Crítico.

Caso de uso Analizar_Secuencia_mRNA. siRNA Web Services

Tabla.22 Descripción del CU Analizar_Secuencia_mRNA.

Caso de Uso:	Analizar_Secuencia_mRNA.
Actores:	Aplicacion_Cliente
Propósito:	Obtener un diseño de siRNA a partir de un adecuado análisis de la secuencia mRNA que se haya seleccionado.
Resumen:	El caso de uso se inicia cuando alguna aplicación solicita al web services analizar la secuencia de mRNA y procede a analizarla seleccionando las reglas y el tipo de análisis deseado, para de esta manera obtener el diseño de siRNA esperado.
Referencia:	R 3.1.
Precondiciones:	
Poscondiciones:	

Flujo Normal de Eventos	
Acción del Actor	Respuesta del Servicio
1) Aplicación cliente hace una petición al web services para analizar la secuencia de mRNA pasando como parámetro las reglas, la secuencia de mRNA y la secuencia de SNPs.	2) Busca sitios blancos en la secuencia mRNA. 3) Busca los SNP pertenecientes a la secuencia. 4) Devuelve la secuencia mRNA resaltando sitios blancos, sus datos y los SNP que coinciden con los sitios blancos.
Flujos Alternos	
Acción del Actor	Respuesta del Sistema
Prioridad:	Crítico.

Caso de uso Buscar_Gen. siRNA web Services

Tabla.23 Descripción del CU Buscar_Gen.

Caso de Uso:	Buscar_Gen.
Actores:	Aplicacion_Cliente.
Propósito:	Mostrarle información relevante a los especialistas.
Resumen:	El caso de uso se inicia cuando alguna aplicación accede a las funcionalidades del web services para conocer información sobre los genes. Entonces se busca en la base de datos la información necesaria para que sea devuelta para la aplicación que la accede.
Referencia:	R 1.1, 1.5.
Precondiciones:	
Poscondiciones:	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Servicio

1) Aplicación cliente hace una petición al web services para buscar el gen pasando como parámetro el id o el símbolo.	2) Busca datos de gen (Chromosome, Gene id, Symbol y los transcritos asociados).
	3) Busca transcritos asociados al gen.
	4) Busca los miRNA correspondientes al gen (start, end, nombre, Pvalue, Score).
	5) Devuelve el gen y valores asociados a este.
Prioridad:	Crítico.

Caso de uso Buscar_Transcrito. siRNA web services

Tabla.24 Descripción del CU Buscar_Transcrito.

Caso de Uso:	Buscar_Transcrito.	
Actores:	Aplicacion_Cliente.	
Propósito:	Mostrarle información relevante a los especialistas.	
Resumen:	El caso de uso se inicia cuando alguna aplicación accede a las funcionalidades del web services para conocer más información sobre los transcritos. Entonces se busca en la base de datos la información necesaria para que sea devuelta para la aplicación que la invoca.	
Referencia:	R 2.1, R 2.3.	
Precondiciones:		
Poscondiciones:		
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Servicio	
1) Aplicación cliente hace una petición al web services para buscar un transcrito pasando como parámetro el nombre del transcrito.	2) Busca transcrito. 3) Busca exones vinculados al transcrito. 4) Devuelve el transcrito con sus exones asociados.	
Prioridad:	Crítico.	

Caso de uso Calcular_Energia. siRNA web services

Tabla.25 Descripción del CU Calcular_Energia.

Caso de Uso:	Calcular_Energia	
Actores:	Aplicacion_Cliente.	
Propósito:	Mostrar la Energía del RNA de doble cadena en el silenciamiento de genes	
Resumen:	El caso de uso se inicia cuando una aplicación solicita al web services analizar secuencia y la energía se calcula en el sitio target y esta se devuelve a cada vez que se interactúa con analizar secuencia.	
Referencia:	R3.5.	
Precondiciones:	Haber realizado el CU Analizar_Secuencia_mRNA.	
Poscondiciones:		
Flujo Normal de Eventos		
Acción del Actor	Respuesta del servicio	
	Se realiza un análisis de sitios blancos. 2) Calcula la energía correspondiente al sitio target en cuestión. 3) Se adjunta este valor de la energía a cada sitio blanco.	
Prioridad:	Secundario.	

Caso de uso Buscar_Secuencia_Específico. siRNA web services

Tabla.26 Descripción del CU Buscar_Secuencia_Específico.

Caso de Uso:	Buscar_Secuencia_Específico
Actores:	Aplicacion_Cliente.
Propósito:	Mostrar las secuencias de nucleótidos que son específicas a determinado Transcrito.
Resumen:	El caso de uso se inicia cuando una aplicación solicita al web services buscar la secuencia específica de un transcrito determinado.

Referencia:	R5.
Precondiciones:	
Poscondiciones:	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Servicio
1) Aplicación cliente hace una petición al web services para buscar la secuencia específica de un transcrito pasando como parámetro el nombre.	2) El servicio busca el transcrito, gen y exones asociados al transcrito especificado. 3) El servicio busca la secuencia específica. 4) El servicio devuelve dicha secuencia.
Prioridad:	Crítico.

Caso de uso Buscar_Secuencia_Comun. siRNA web services

Tabla.27 Descripción del CU Buscar_Secuencia_Comun.

Caso de Uso:	Buscar_Secuencia_Comun
Actores:	Aplicacion_Cliente.
Propósito:	Mostrar las secuencias de nucleótidos que son comunes en todos los Transcritos de ese gen.
Resumen:	El caso de uso se inicia cuando una aplicación solicita al web services buscar la secuencia común a un gen determinado.
Referencia:	R4.
Precondiciones:	
Poscondiciones:	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Servicio
1) Aplicación cliente hace una petición al web services para buscar la secuencia común del gen pasándole como parámetro el símbolo o el id.	2) EL servicio busca el gen, transcrito y exones asociados al gen especificado. 3) El servicio busca la secuencia común para los transcritos de ese gen. 4) El servicio retorna la secuencia con los valores

	especificados
Prioridad:	Crítico.

Caso de uso Listar_SNP_en_region_cromosoma. siRNA web services

Tabla.28 Descripción del CU Listar_SNP_en_region_cromosoma.

Caso de Uso:	Listar_SNP_en_region_cromosoma
Actores:	Aplicacion_Cliente.
Propósito:	Listar las secuencias de SNPs para una región del cromosoma
Resumen:	El caso de uso se inicia cuando una aplicación solicita al web services listar los SNPs para una región del cromosoma
Referencia:	R4.
Precondiciones:	
Poscondiciones:	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Servicio
1) La aplicación hace una petición al web services para obtener los SNPs en una región del cromosoma, pasándole el cromosoma, y la posición inicial y final.	2) El servicio busca los SNPs 3) El servicio retorna los SNPs
Prioridad:	Secundario.

Caso de uso Listar_SNPs_coinciden_miRNA. siRNA web services

Tabla.29 Descripción del CU Listar_SNPs_coninciden_miRNA.

Caso de Uso:	Listar_SNPs_coinciden_miRNA
Actores:	Aplicacion_Cliente.

Propósito:	Listar los de SNPs para una región del cromosoma que coinciden con miRNAs
Resumen:	El caso de uso se inicia cuando una aplicación solicita al web services listar los SNPs para una región del cromosoma y que haga énfasis en los que coinciden con miRNAs
Referencia:	R4.
Precondiciones:	Haber realizado el CU Mostrar_Gen
Poscondiciones:	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Servicio
1) La aplicación hace una petición al web services para obtener los SNPs en una región del cromosoma y que coincidan con los miRNAs pasándole como parámetro un gen determinado.	2) El sistema busca los SNPs de ese gen. 3) El sistema realiza análisis con miRNAs asociados a ese gen 4) El servicio retorna los SNPs
Prioridad:	Secundario.

2.11 Conclusiones

En este capítulo se describió el objeto de estudio del presente trabajo y se plantearon los problemas existentes que dieron pie al desarrollo de la herramienta. Se ha brindado una definición de los requisitos que deberá cumplir el sistema. A través de estos se han detallado las especificaciones que han hecho los especialistas como parte del flujo de trabajo de levantamiento de requisitos. Se redefinió el diagrama de casos de uso del sistema planteado en la tesis anterior, agregando casos de usos en dependencia de las nuevas acciones que se le incluyeron a la aplicación, lo cual permitió que se representaran detalladamente las funcionalidades requeridas, lográndose una modelación más clara de la interacción del sistema y el especialista. Se especificaron estos casos de uso con la peculiaridad de realización del mismo caso de uso en la versión 2.0 de alasSiRNA - Design como en el web services.

CAPITULO 3: ANÁLISIS Y DISEÑO DEL SISTEMA

Introducción

Este capítulo ofrece una descripción de la arquitectura del sistema y de los principales objetivos del diseño. Se definen y diseñan diagramas de clases e interacción para el diseño de clases web, también se hace una breve comparación con la aplicación anterior. Se tratan los principios del diseño del sistema, se explica el tratamiento de errores. Se define el diseño de la base de datos a través de los modelos lógico y físico de la misma y se da una descripción de las tablas más importantes. Se muestra el despliegue de la aplicación a través de un diagrama de despliegue y se hace énfasis en la seguridad de la misma.

3.1 Arquitectura de software orientado a servicios.

Una arquitectura es un entramado de componentes funcionales que aprovechando diferentes estándares, convenciones, reglas y procesos, permite integrar una amplia gama de productos y servicios informáticos, de manera que pueden ser utilizados eficazmente dentro de la organización. La Arquitectura del Software es la organización fundamental de un sistema formado por sus componentes, las relaciones entre ellos y el contexto en el que se implantarán, y los principios que orientan su diseño y evolución. Para que un sistema sea extensible y reutilizable, como se pretende que sea el presente proyecto, su arquitectura tiene que estar diseñada en base a esto.

Por esta razón esta aplicación tiene una arquitectura orientada a servicios la cual ha surgido para afrontar problemas, como la mejor manera de hacer más con menos recursos. Promete hacer la reutilización y la integración mucho más fáciles, ayudando a reducir el tiempo de desarrollo y aumentando la agilidad organizacional. Además en la rama de la bioinformática se presenta una creciente evolución y un creciente cambio por lo que usando SOA con web services subyacentes proporciona mayor flexibilidad para afrontar dichos cambios tanto en el ambiente de negocios como en la infraestructura tecnológica.

3.2 Objetivos del diseño

El diseño tiene como principales objetivos comprender detalladamente los requisitos funcionales y no funcionales, sistemas operativos, tecnologías de distribución, restricciones relacionadas con el lenguaje de programación, tecnologías de interfaz de usuario. Además, el diseño crea un punto de partida para las actividades de implementación que siguen.

3.3 Vista Lógica del sistema.

La vista lógica describe el diseño más importante de las clases y su organización en paquetes y subsistemas, y la organización de éstos en capas. Muestra cómo la funcionalidad es diseñada en el interior del sistema, en términos de la estructura estática y comportamiento dinámico del sistema.

La siguiente representación tiene como objetivo principal desplegar las clases de la aplicación por lo que se pueden especificar tres subsistemas, el subsistema Sistemas de Diseño de siRNA se refiere a las aplicaciones que se conecten con el web services desde cualquier lugar del planeta, de esta representación abstracta hereda la versión 2.0 de alasSiRNA – Design v2.0 que será el primer cliente de siRNA Web Services el cual tiene diferentes servicios a utilizar por cualquiera de los usuarios en cualquier aplicación cliente.

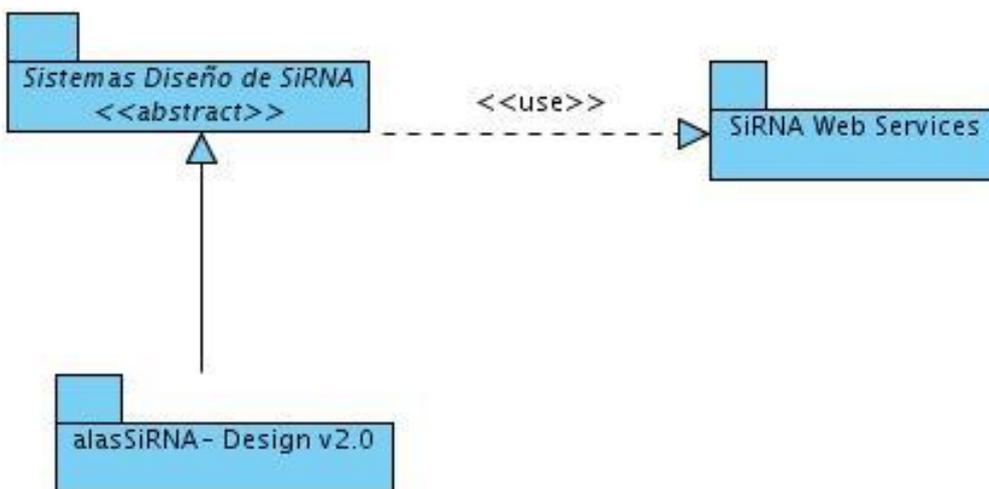


Fig. 7 Vista lógica del sistema.

3.4 Diseño de clases del sistema

Las clases que son diseñadas en esta aplicación tienen la particularidad que se ven desde 2 puntos de vista. Del lado del servicio web y del lado de la aplicación que se conecta al servicio web en este caso sería la versión 2.0 de alasSiRNA Desing, las cuales cuentan con características diferentes que serán detalladas en los siguientes epígrafes.

3.4.1 Diseño de las clases de alasSiRNA - Design versión 2.0

La versión 2.0 de alasSiRNA - Design escogió dentro de los varios patrones arquitectónicos que existen, el Controlador Frontal, que implementa el patrón Modelo Vista Controlador (MVC) con páginas Java, con la peculiaridad que accederá a los datos a través de un web services.

Como características fundamentales de este patrón están las siguientes capas:

Vista: Es la capa que contiene las clases dedicadas puramente a la creación de objetos de presentación. Para trabajar con las vistas se ha empleado el uso de plantillas.

Controlador: Es la capa que contiene todas las clases encargadas de garantizar la lógica de la aplicación. Según el patrón Controlador Frontal, se ha trabajado con estructuras denominadas acciones que no son más que clases encargadas de llevar el control del flujo en el sistema.

Modelo: Es la capa que contiene todas las clases que toman las decisiones importantes desde el punto de vista empresarial y del negocio a informatizar.

Para la representación de dichas clases se utilizó el lenguaje unificado de modelado el cual posee una extensión para el modelado de aplicaciones Web, propuesta por Conallen [22], dicha extensión es usada para el diseño de las clases. Los estereotipos que usa esta extensión son:



<<Server Page>> Representa la página Web que tiene código que se ejecuta en el servidor. Este código interactúa con recursos en el servidor. Las operaciones representan las funciones del código y los atributos las variables visibles dentro del alcance de la página.



<<Client Page>> Una instancia de Página Cliente es una página Web, con formato HTML; mezcla de datos, presentación y lógica. Son interpretadas por el browser. Cada página cliente solo puede ser construida por una página servidor.



<<Form>> Grupo de elementos de entrada que son parte de una página cliente. Se relaciona directamente con la etiqueta de igual nombre del HTML. Sus atributos son los elementos de entrada del formulario.

Existen otros componentes como funciones en java script para validación de campos y páginas en servlets para generar imágenes dinámicas.

El diseño de las clases en Open/Up es basado en casos de uso por lo que se diseñan clases por casos de uso con más significancia:

- Caso de uso Analizar_Secuencia_mRNA

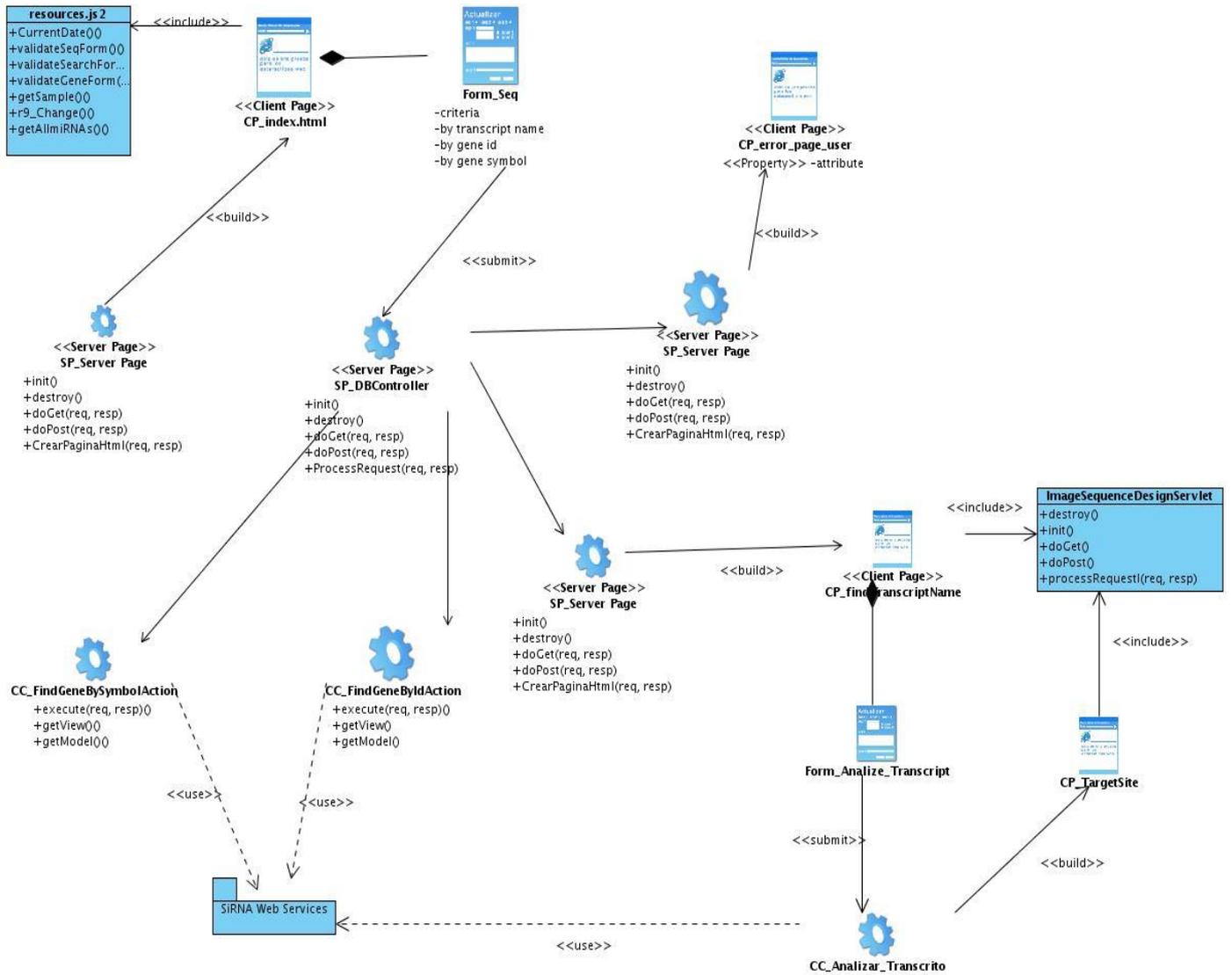


Fig. 8 Diagrama de clases del caso de uso Analizar_Secuencia_mRNA. alasSiRNA - Design v2.0

- Caso de uso **Mostrar_Gen**

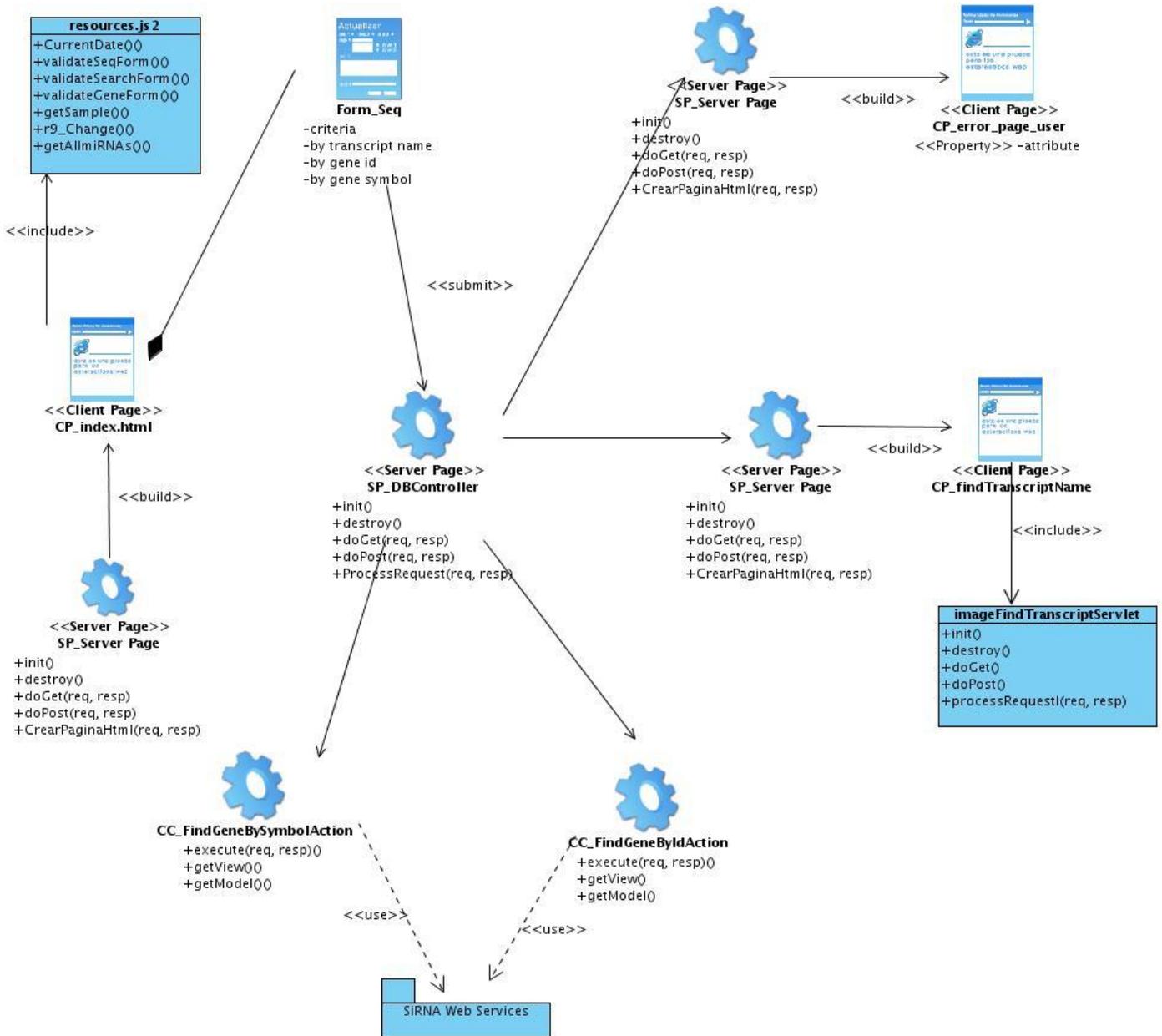


Fig. 9 Diagrama de clases del caso de uso **Mostrar_Gen**. alasSiRNA - Design v2.0

- Caso de uso **Mostrar_Transcrito**

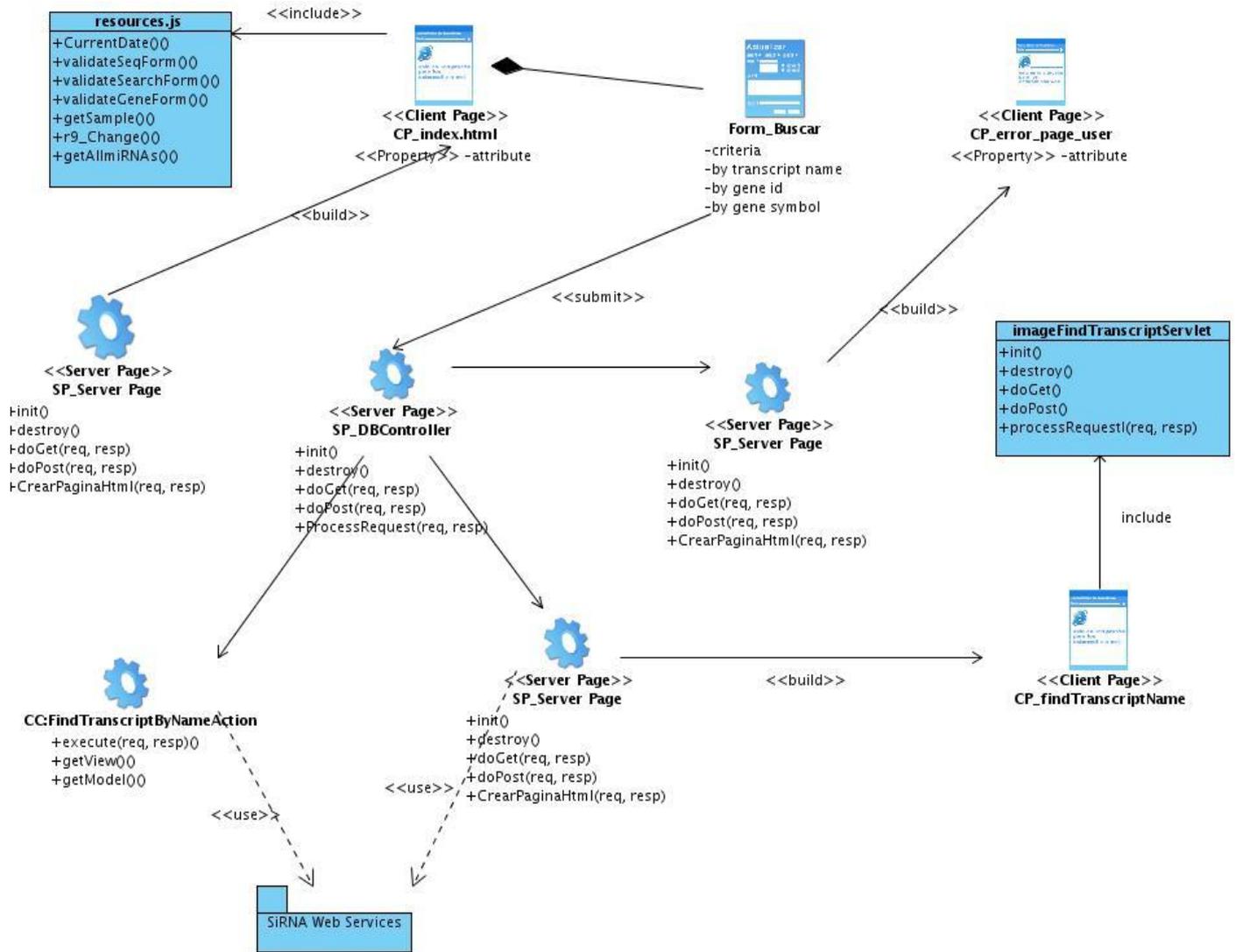


Fig. 10 Diagrama de clases del caso de uso **Mostrar_Transcrito**. alasSiRNA - Design v2.0

- Caso de uso **Mostrar_Secuencia_Especifico_Transcrito**

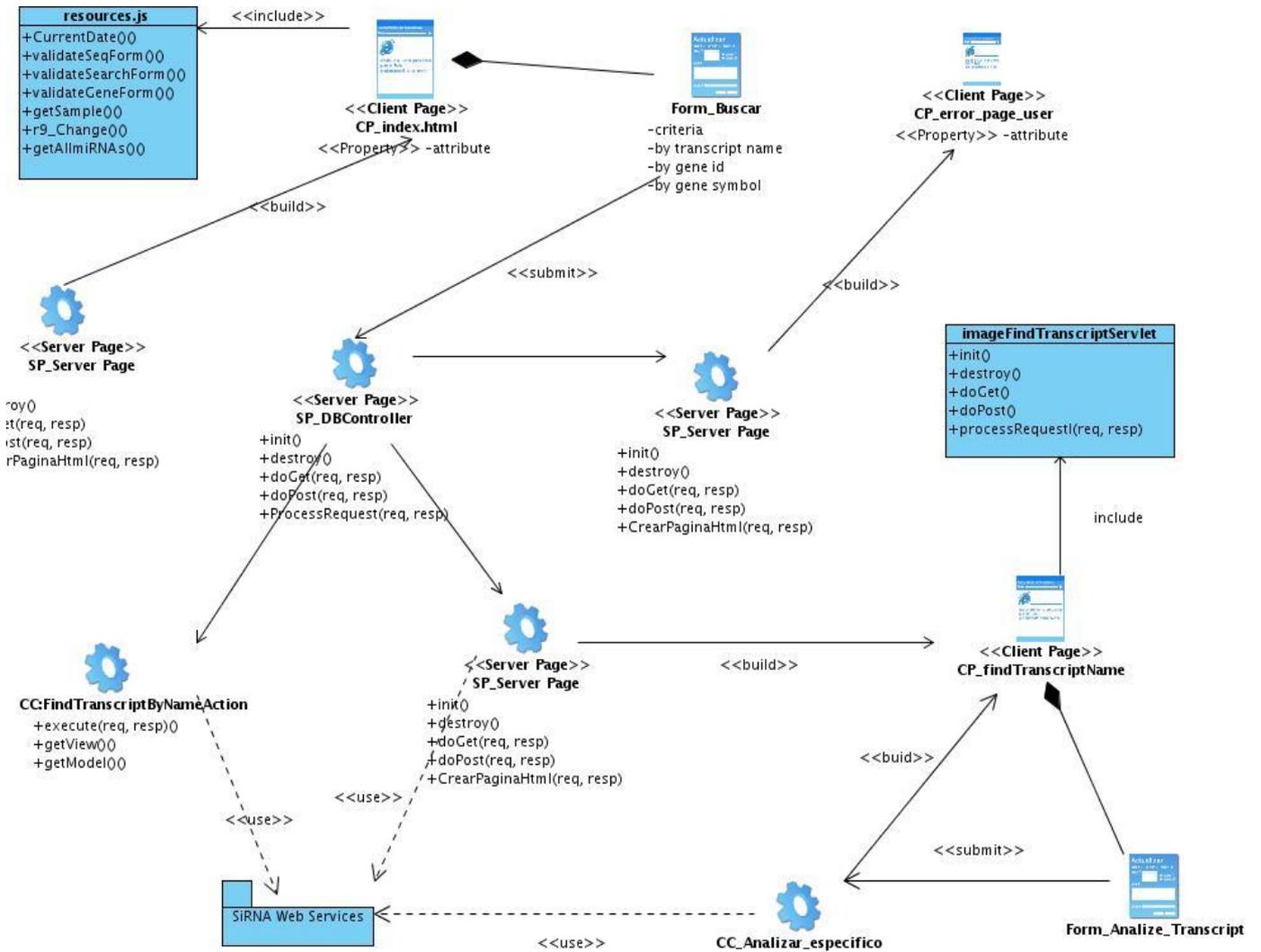


Fig. 11 Diagrama de clases del caso de uso **Mostrar_Secuencia_Especifico_Transcrito**. alasSiRNA - Design v2.0

- Caso de uso **Mostrar_Secuencia_Comun_Gen**

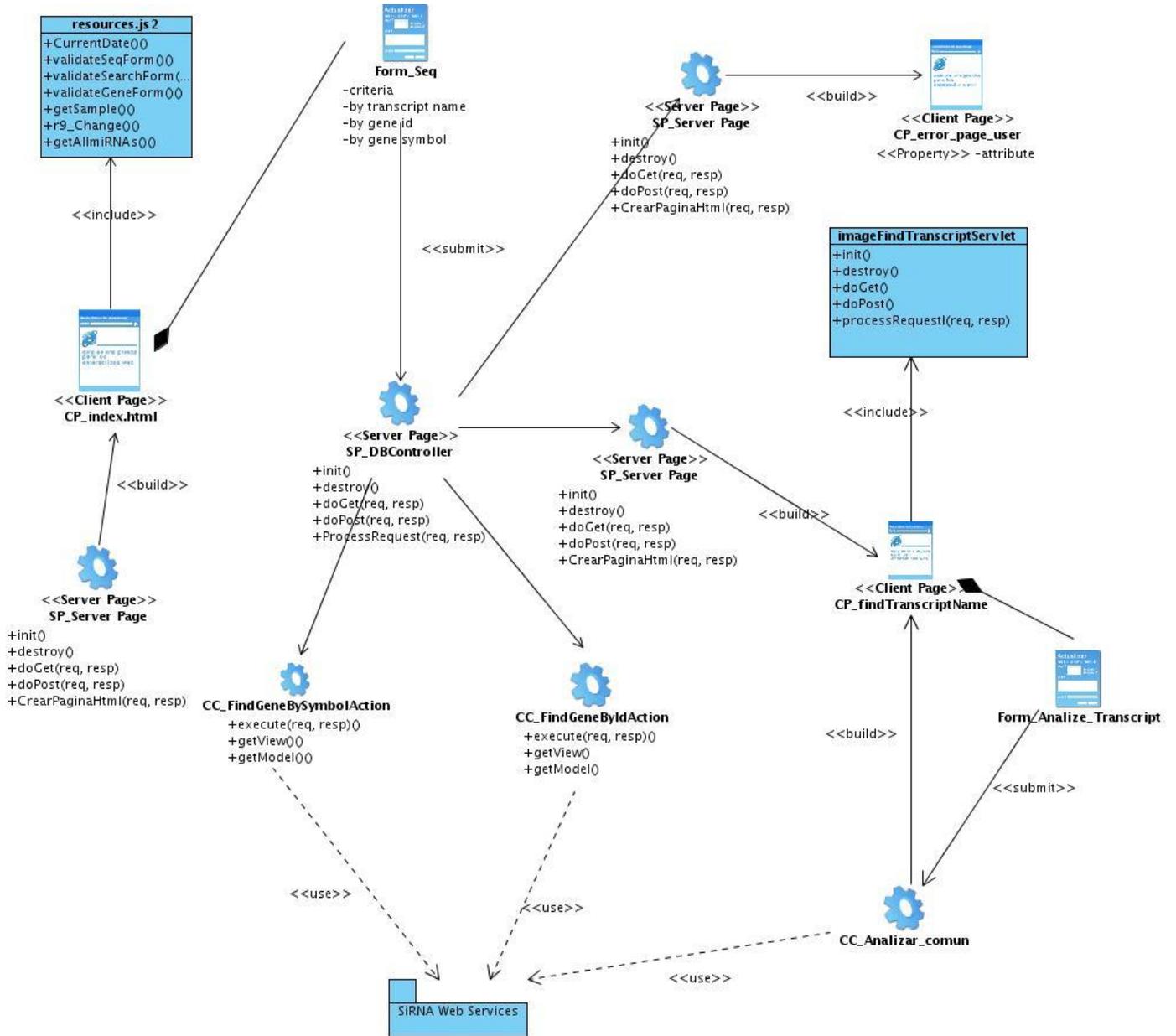


Fig. 12 Diagrama de clases del caso de uso **Mostrar_Secuencia_Comun_Gen**. alasSiRNA - Design v2.0

3.4.2 Diseño de las clases del web services.

Los servicios Web, no son aplicaciones con una interfaz gráfica con la que los usuarios puedan interactuar, sino que son software accesible a través de internet o en redes privadas que usen tecnologías internet por otras aplicaciones. De esta forma se puede desarrollar aplicaciones que hagan uso de otras aplicaciones que estén disponibles en internet interactuando con ellas.

Los siguientes diagramas de clases no siguen un patrón de arquitectura conocido pero cumplen con los requisitos de seguridad, confiabilidad y disponibilidad que debe tener todo web services.

Diagrama de clases del web services:

El diseño de las clases en Open/Up es basado en casos de uso por lo que se diseñan clases por casos de uso con más significancia:

- Caso de uso Analizar_Secuencia_mRNA

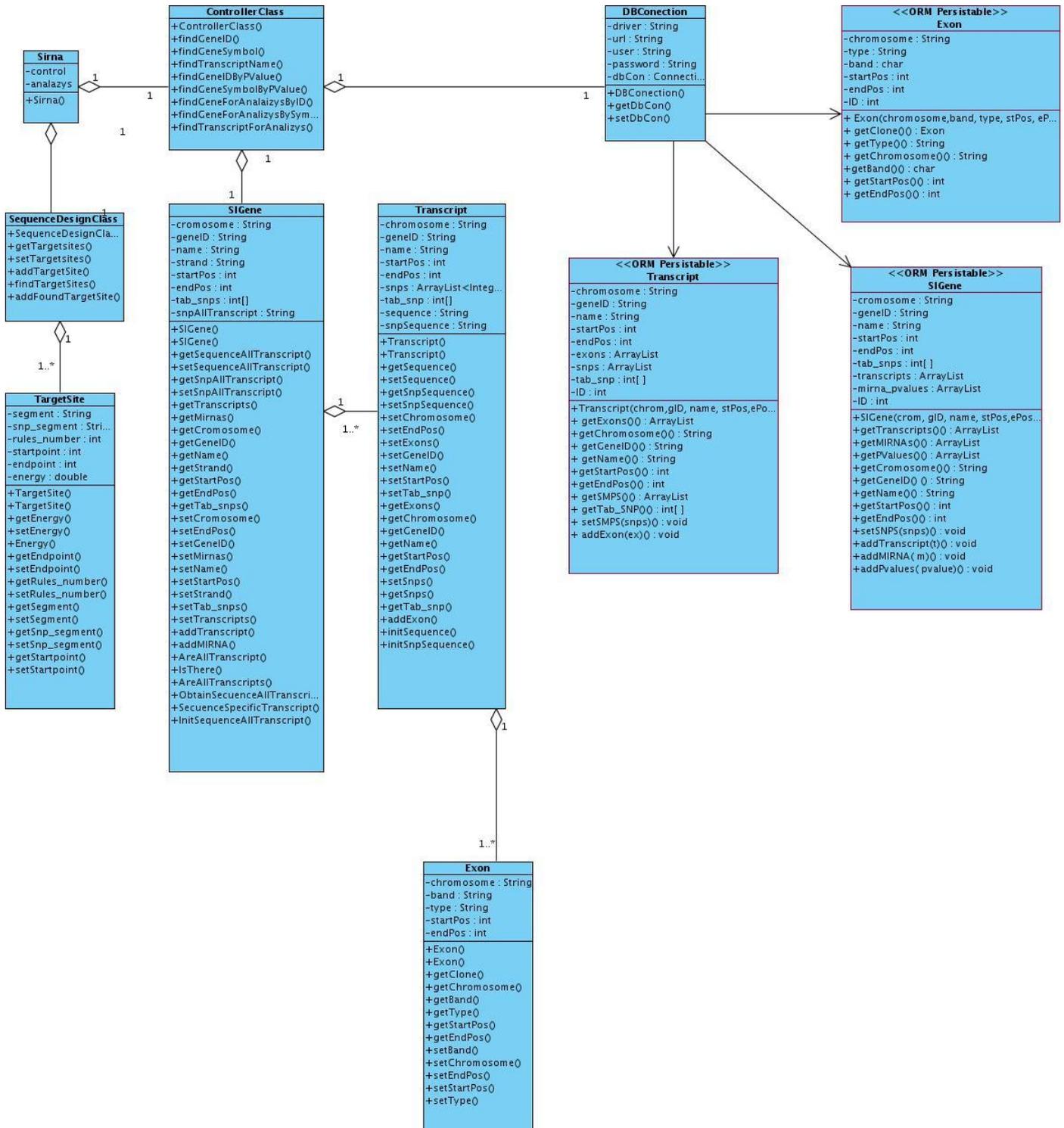


Fig. 13 Diagrama de clases del caso de uso Analizar_Secuencia_mRNA. siRNA Web Services.

• Caso de uso Buscar_Gen

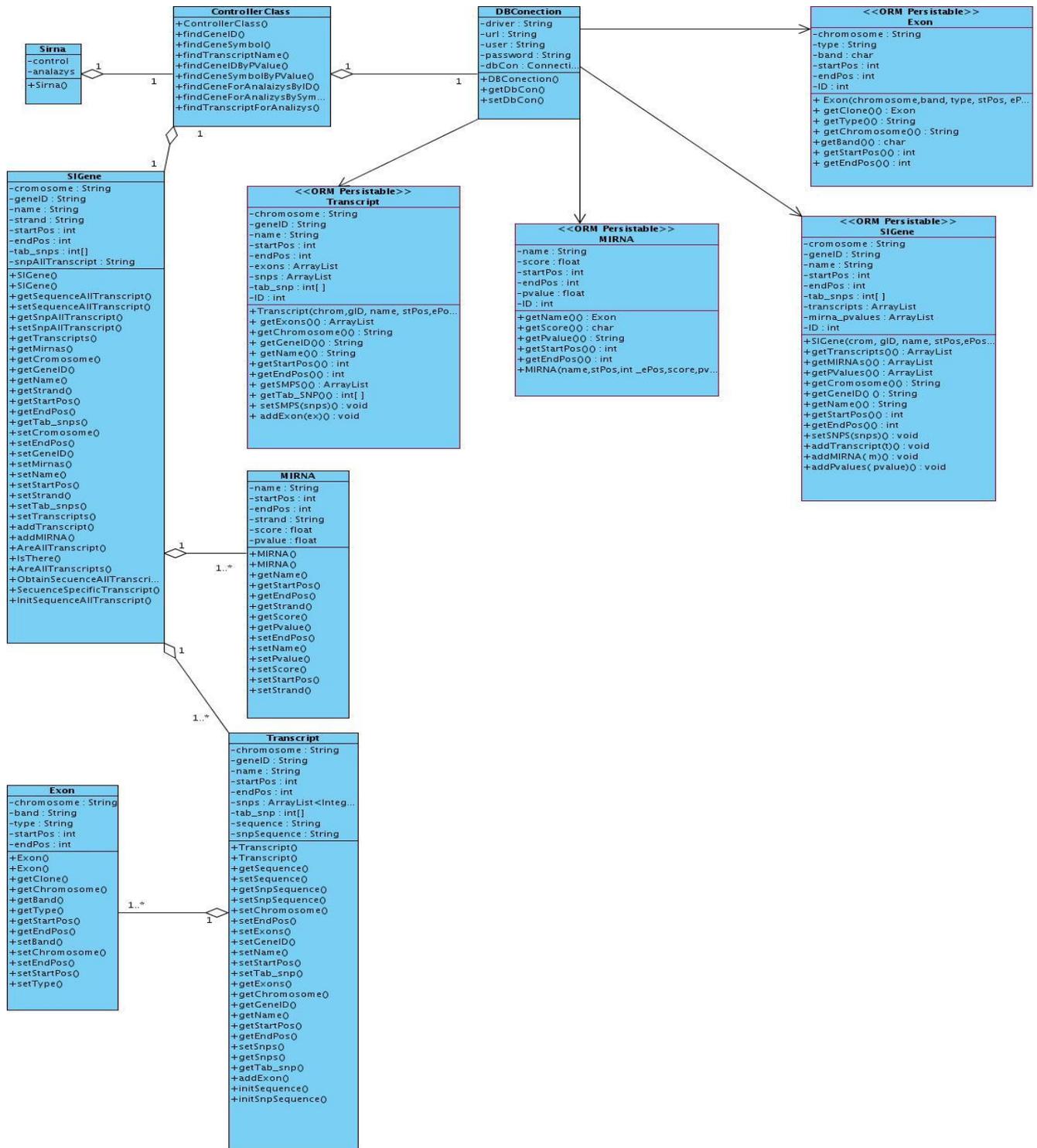


Fig. 14 Diagrama de clases del caso de uso Buscar_Gen. siRNA Web Services.

- Caso de uso Buscar_Transcrito

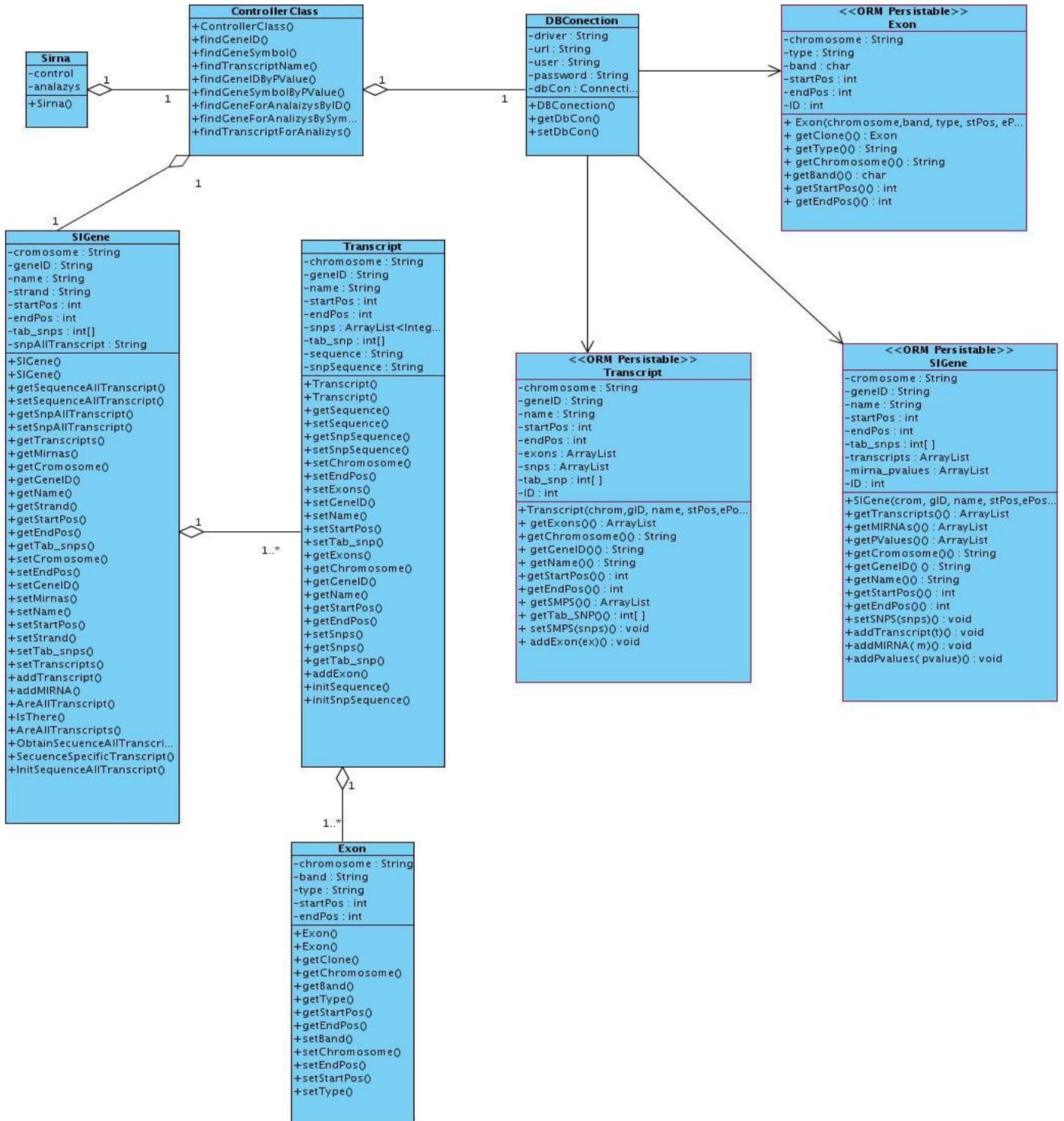


Fig. 15 Diagrama de clases del caso de uso Buscar_Transcrito. siRNA Web Services.

• Caso de uso Buscar_Secuencia_Comun

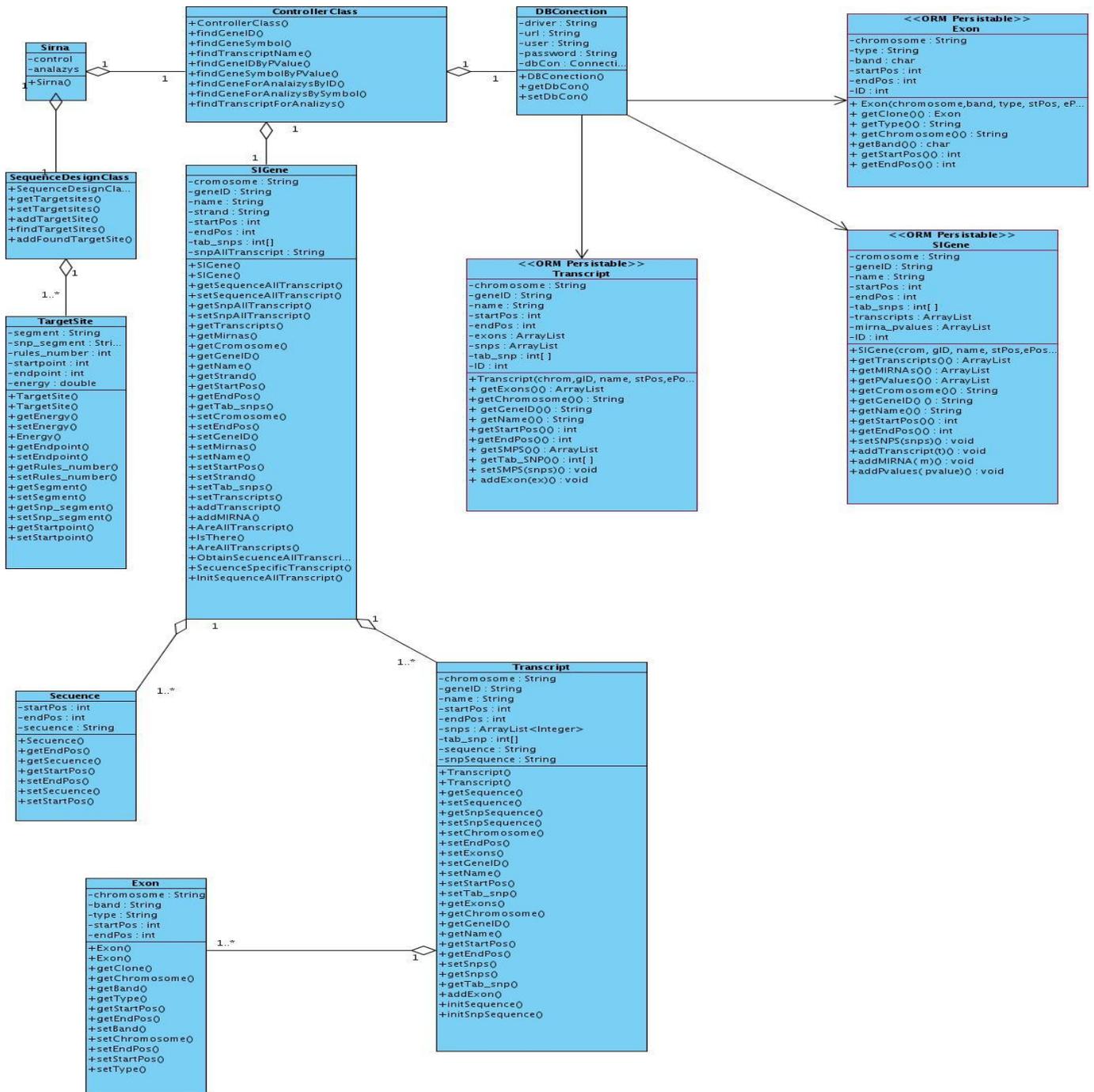


Fig. 17 Diagrama de clases del caso de uso Buscar_Secuencia_Comun. siRNA Web Services.

- Caso de uso Listar_SNP_en_region_cromosoma

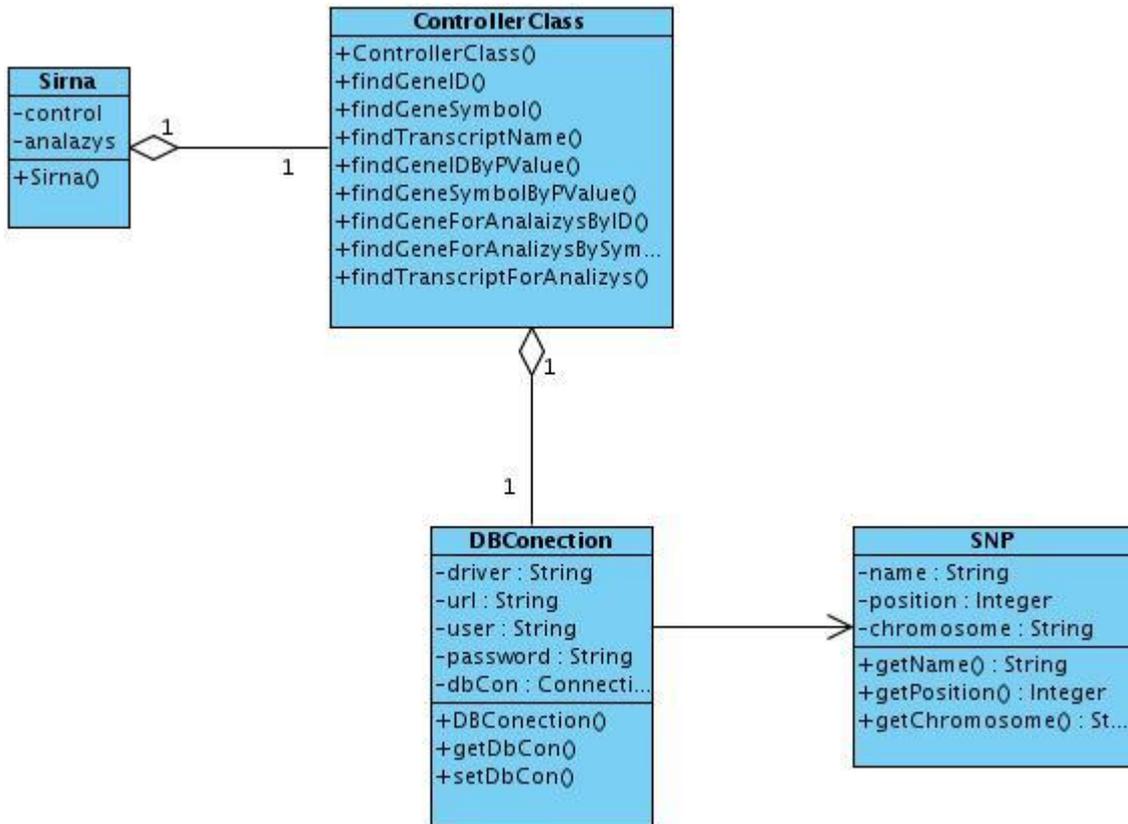


Fig. 18 Diagrama de clases del caso de uso Listar_SNP_en_region_cromosoma. siRNA Web Services.

- Caso de uso Listar_SNP_coinciden_miRNA

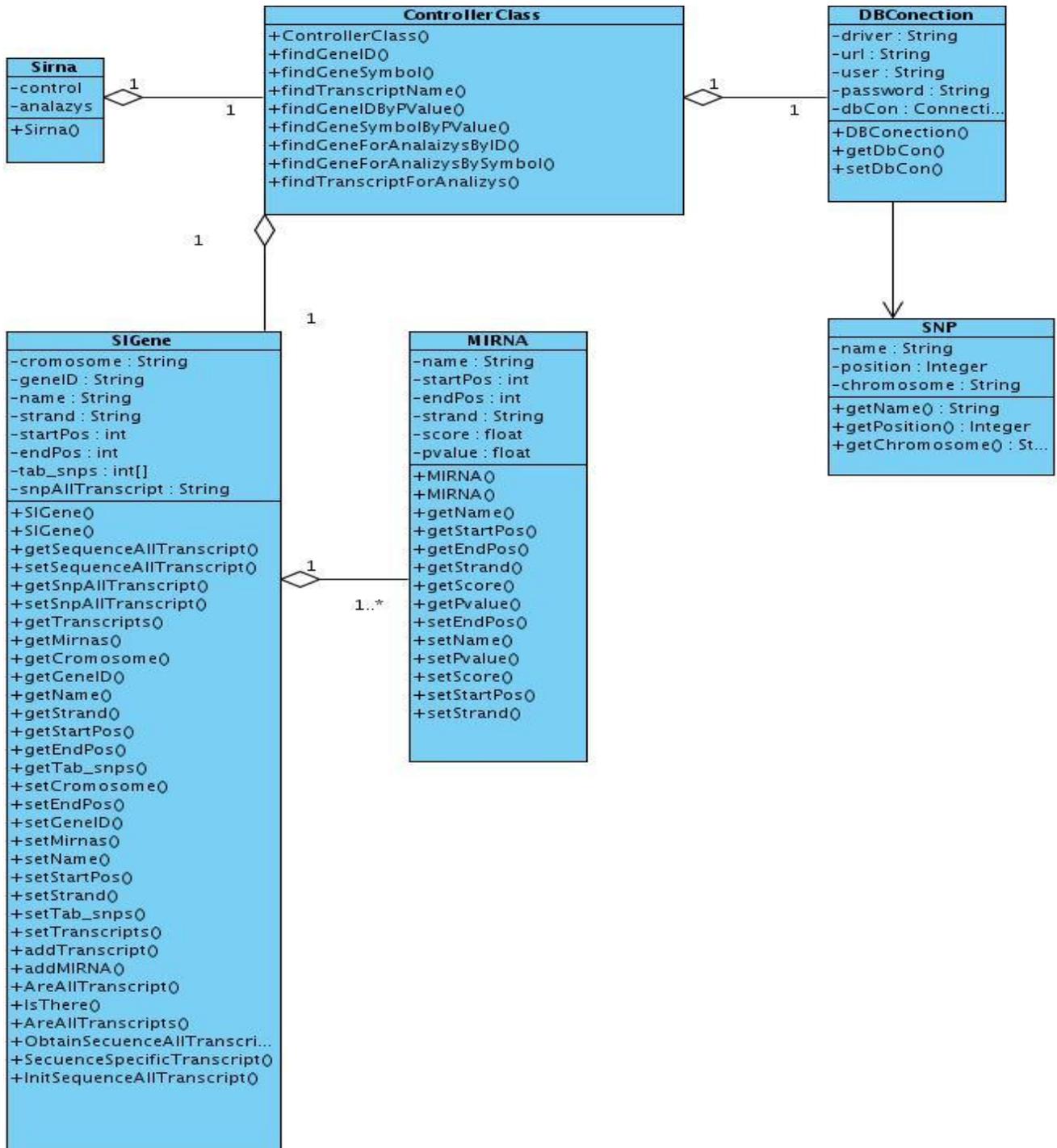


Fig. 19 Diagrama de clases del caso de uso Listar_SNP_coinciden_miRNA. siRNA Web Services.

3.4.3 Diseño de la base de datos (clases persistentes)

Para realizar el diseño de la base de datos de la aplicación se confeccionaron el diagrama de clases persistentes y el modelo de datos. Las clases persistentes son clases entidades que pueden mantener su valor en el espacio y el tiempo. Además las clases temporales son aquellas de las cuales el sistema se encarga de manejar y almacenar en tiempo de ejecución, razón por la que dejan de existir cuando termina la ejecución del programa.

A continuación se presenta el diagrama de clases persistentes que esta compuesto por dichas clases y las correspondientes relaciones entre ellas y el modelo físico de datos que es una representación de las tablas de la base de datos y sus relaciones.

Modelo lógico de datos.

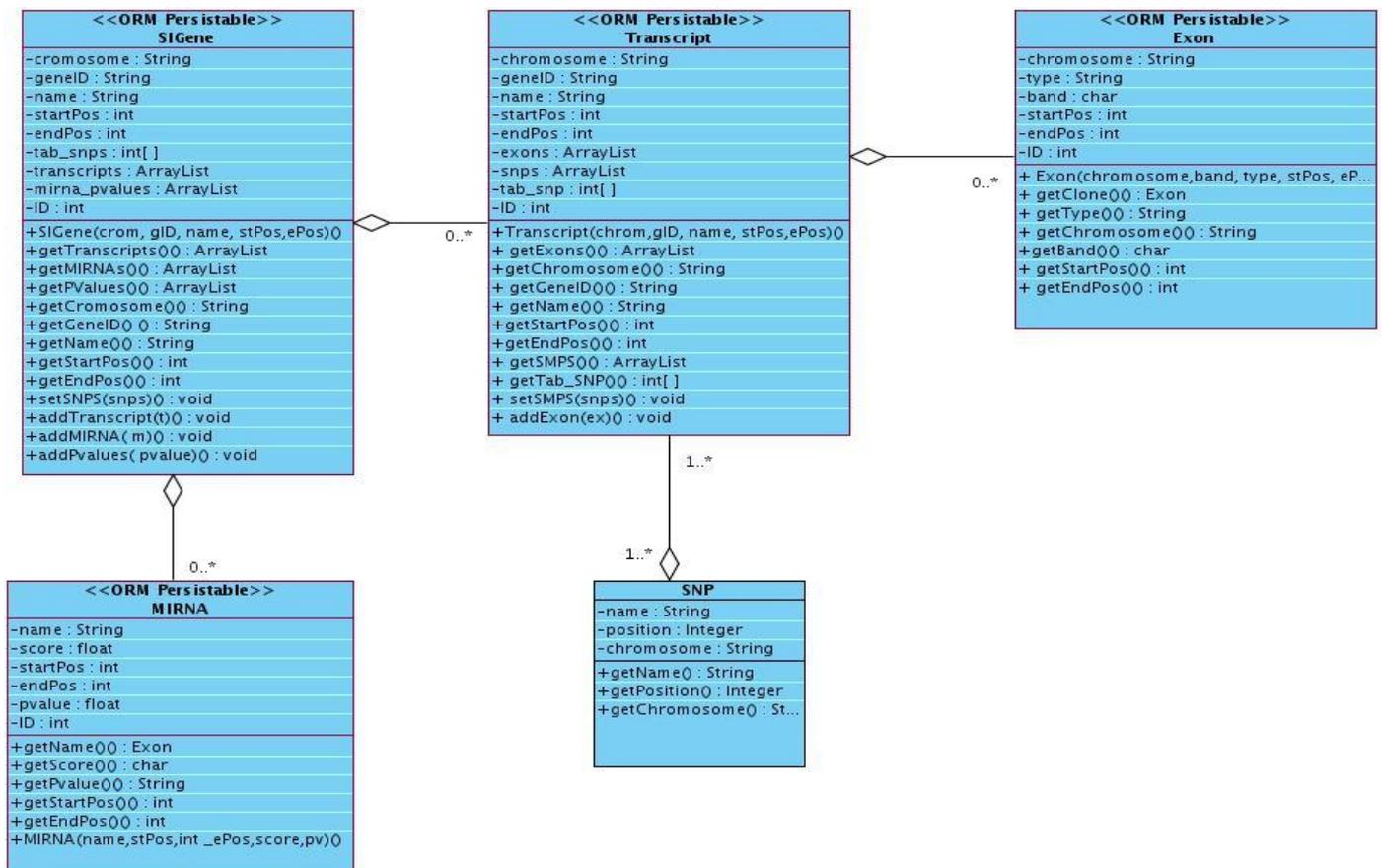


Fig. 20 Diagrama de clases persistentes.

Modelo físico de datos

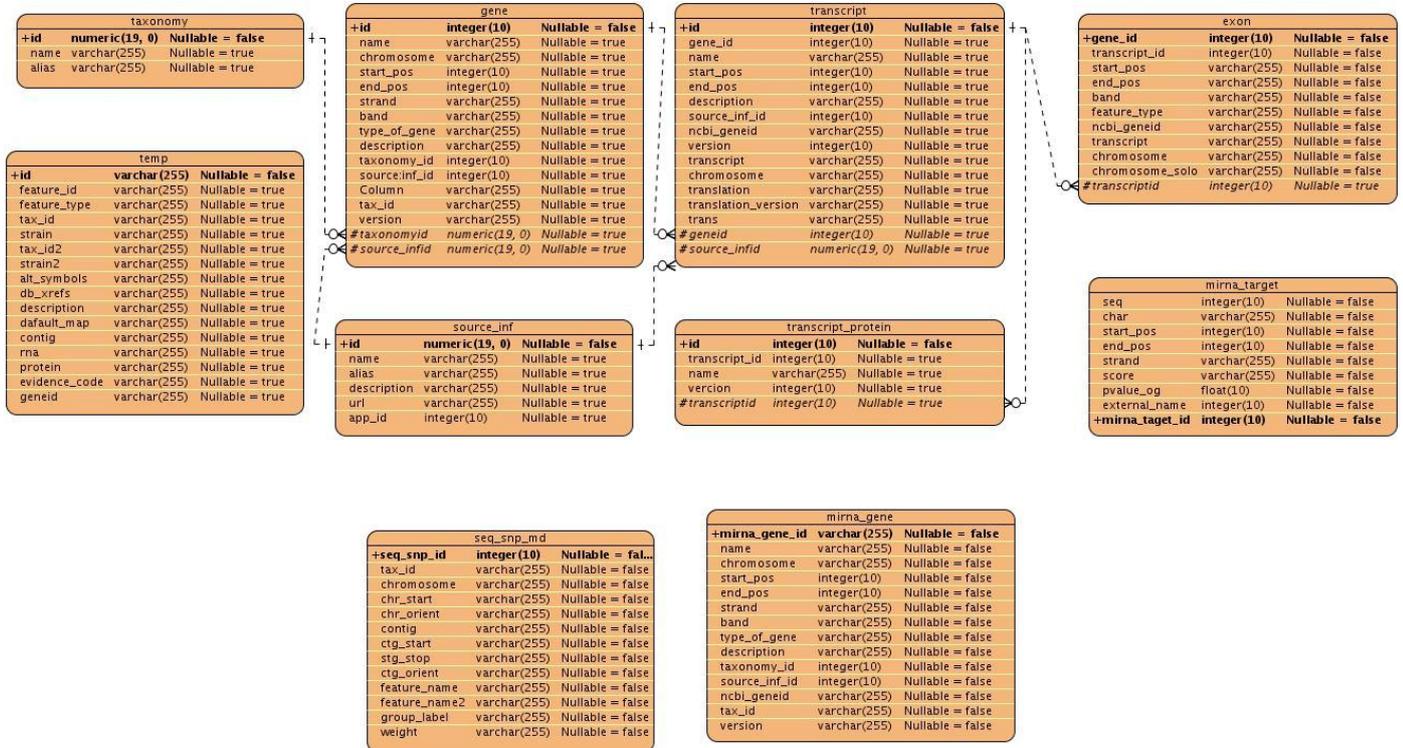


Fig. 21 Diagrama Entidad - Relación

3.5 Descripción de las clases del web services.

Tabla.30 Descripción de la clase siRNA.

Nombre: siRNA	
Atributo	Tipo
control	ControllerClass
analazys	SequenceDesignClass
Para cada responsabilidad:	
Nombre:	siRNA ()
Descripción:	Inicializa los valores de control y analazys.
Nombre:	FindGeneByID
Descripción:	Ejecuta el servicio web para encontrar gen por id.

Nombre:	FindTranscriptByName
Descripción:	Ejecuta el servicio web para encontrar transcrito por nombre
Nombre:	FindGeneBySymbol
Descripción:	Ejecuta el servicio web para encontrar gen por símbolo
Nombre:	FindGeneByIDPValue
Descripción:	Ejecuta el servicio web para encontrar un gen a través del id con miRNA con menor pvalue que el especificado
Nombre:	FindGeneSymbolByPValue
Descripción:	Ejecuta el servicio web para encontrar un gen a través del símbolo con miRNA con menor pvalue que el especificado
Nombre:	FindGeneForAnalizysAllTranscriptByID
Descripción:	Ejecuta el servicio web para buscar la secuencia común de todos los transcritos de ese gen a través de su id.
Nombre:	FindGeneForAnalizysAllTranscriptBySymbol
Descripción:	Ejecuta el servicio web para buscar la secuencia común de todos los transcritos de ese gen a través de su símbolo.
Nombre:	FindTranscriptForAnalizysSpecificSequence
Descripción:	Ejecuta el servicio web para buscar la secuencia específica a un transcrito especificado.

Tabla.31 Descripción de la clase ControllerClass.

Nombre: ControllerClass	
Atributo	Tipo
dbCon	Connection
Para cada responsabilidad:	
Nombre:	ControllerClass ()
Descripción:	Inicializa el valor de dbCon.
Nombre:	findGeneID(String)
Descripción:	Realiza la búsqueda para encontrar gen por id.
Nombre:	findGeneSymbol(String)
Descripción:	Busca un gen por el símbolo
Nombre:	findTranscriptName(criteria)

Descripción:	Realiza la búsqueda para encontrar transcrito por nombre
Nombre:	findGeneIDByPValue(String, String)
Descripción:	Realiza la búsqueda para encontrar un gen a través del id con miRNA con menor pvalue que el especificado
Nombre:	findGeneSymbolByPValue(String, String)
Descripción:	Realiza la búsqueda para encontrar un gen a través del símbolo con miRNA con menor pvalue que el especificado
Nombre:	findGeneForAnalizysByID(String)
Descripción:	Realiza la búsqueda para analizar el gen por su id.
Nombre:	findGeneForAnalizysBySymbol(String)
Descripción:	Realiza la búsqueda para analizar el gen por su símbolo.
Nombre:	findTranscriptForAnalizys(String)
Descripción:	Realiza el la búsqueda para analizar el transcrito por su nombre.

Tabla.32 Descripción de la clase DBConection.

Nombre: DBConection.	
Atributo	Tipo
driver	String
url	String
user	String
contraseña	String
dbcon	Connection
Para cada responsabilidad:	
Nombre:	DBConection ()
Descripción:	Cargar el driver en memoria y crear la conexión.
Nombre:	getDBCon()
Descripción:	El método simplemente retorna el valor de dbcon.
Nombre:	setDBCon(Connection)
Descripción:	El método cambia los valores de la conexión.

Tabla.33 Descripción de la clase SequenceDesignClass.

Nombre: SequenceDesignClass.

Atributo	Tipo
targetsites	ArrayList<TargetSite>
Para cada responsabilidad:	
Nombre:	SequenceDesignClass ()
Descripción:	Inicializa a targetsites.
Nombre:	getTargetsites ()
Descripción:	El método simplemente retorna el valor de targetsites.
Nombre:	setTargetsites (ArrayList<TargetSite>)
Descripción:	El método cambia los valores de la targetsites.
Nombre:	addTargetSite(TargetSite)
Descripción:	El método adiciona un valor de TargetSite a la lista.
Nombre:	addFoundTargetSite(boolean,String,int,TargetSite)
Descripción:	El método adiciona un valor de TargetSite encontrado a la lista.
Nombre:	findTargetSites(String _,boolean, boolean, boolean, boolean, boolean ,int ,int , Boolean, boolean, boolean, boolean, int, boolean, String)
Descripción:	El método encuentra los sitios blancos.

Tabla.34 Descripción de la clase TargetSite.

Nombre: TargetSite.	
Atributo	Tipo
segment	String
snp_segment	String
rules_number	int
startpoint	int
endpoint	int
energy	double
Para cada responsabilidad:	
Nombre:	TargetSite(String, int, int, int)
Descripción:	Inicializa los valores de los atributos.
Nombre:	TargetSite()
Descripción:	Construye los atributos con valores nulos.
Nombre:	getEnergy()

Descripción:	El retorna el valor de energy.
Nombre:	setEnergy(double)
Descripción:	El método cambia los valores de energy.
Nombre:	Energy(String)
Descripción:	El calcula los valores de la energía.
Nombre:	getEndpoint()
Descripción:	El retorna el valor de endpoint.
Nombre:	setEndpoint(int)
Descripción:	El método cambia los valores de la endpoint.
Nombre:	getRules_number()
Descripción:	El método retorna los valores de la rules_numbers.
Nombre:	setRules_number(int)
Descripción:	El método cambia los valores de la rules_numbers.
Nombre:	getSegment()
Descripción:	El método retorna los valores de la segment.
Nombre:	setSegment(String)
Descripción:	El método cambia los valores de la segment.
Nombre:	getSnp_segment()
Descripción:	El método retorna los valores de la snp_segment.
Nombre:	setSnp_segment(Strong)
Descripción:	El método cambia los valores de la snp_segment.
Nombre:	getStartpoint()
Descripción:	El método retorna los valores de startpoint.
Nombre:	setStartpoint(int)
Descripción:	El método cambia los valores de startpoint.

Tabla.35 Descripción de la clase SiGene.

Nombre: SiGene.	
Atributo	Tipo
geneID	String
chromosome	Srting
startPos	int

end_pos	int
strand	String
name	String
tab_snps	int[]
transcripts	ArrayList<Transcript>
mirnas	ArrayList<MIRNA>
sequenceAllTranscript	String
snpAllTranscript	String
Para cada responsabilidad:	
Nombre:	getGeneID ()
Descripción:	Devuelve el identificador del gen especificado.
Nombre:	getChromosome ()
Descripción:	Devuelve el nombre del chromosome.
Nombre:	getStartPos()
Descripción:	Devuelve la posición inicial de un gen en el cromosoma.
Nombre:	getEnd_Pos()
Descripción:	Devuelve la posición final de un gen en el cromosoma.
Nombre:	getStrand()
Descripción:	Devuelve la orientación del gen en el cromosoma, cadena (+) o cadena (-)
Nombre:	getName()
Descripción:	Devuelve el nombre del gen.
Nombre:	getTabs_Snps()
Descripción:	Devuelve los valores de tab_snps.
Nombre:	get Transcripts()
Descripción:	Devuelve los valores de transcripts.
Nombre:	getMirnas ()
Descripción:	Devuelve los valores de miRNAs.
Nombre:	getSequenceAllTranscript ()
Descripción:	Devuelve los valores de sequenceAllTranscript.
Nombre:	getSnpAllTranscript()
Descripción:	Devuelve los valores de snpAllTranscript.
Nombre:	SGene(String, String, String,String, int, int)

Descripción:	Inicializa los valores de los atributos.
Nombre:	setGeneID (String)
Descripción:	Cambia el identificador del gen especificado.
Nombre:	setChromosome (String)
Descripción:	Cambia el nombre del chromosome.
Nombre:	setStartPos(int)
Descripción:	Cambia la posición inicial de un gen en el cromosoma.
Nombre:	setEnd_Pos(int)
Descripción:	Cambia la posición final de un gen en el cromosoma.
Nombre:	setStrand(String)
Descripción:	Cambia la orientación del gen en el cromosoma, cadena (+) o cadena (-)
Nombre:	setName(String)
Descripción:	Cambia el nombre del gen.
Nombre:	setTabs_Snps(int[])
Descripción:	Cambia los valores de tab_snps.
Nombre:	set Transcripts(ArrayList<Transcript>)
Descripción:	Cambia los valores de transcripts.
Nombre:	setMirnas (ArrayList<int>)
Descripción:	Devuelve los valores de miRNAs.
Nombre:	addTranscript(Transcript)
Descripción:	Adiciona transcritos a la lista transcript
Nombre:	addMIRNA(MIRNA)
Descripción:	Adiciona miRNAs a la lista miRNA
Nombre:	ObtainSecuenceAllTranscript()
Descripción:	Obtiene la secuencia que tiene exones comunes en todos los transcritos.
Nombre:	SecuenceSpecificTranscript(int)
Descripción:	Obtiene la secuencia específica que tiene un transcrito especificado.

Tabla.36 Descripción de la clase Transcriptp.

Nombre: Transcript.	
Atributo	Tipo
geneID	String

name	String
startPos	int
end_pos	int
chromosome	String
exons	ArrayList<Exon>;
snps	ArrayList<Integer>;
tab_snp	int[]
sequence	String
snpSequence	String
Para cada responsabilidad:	
Nombre:	Transcript(String, String, String, int, int)
Descripción:	Inicializa los valores de los atributos
Nombre:	getGeneID ()
Descripción:	Devuelve el identificador del gen especificado.
Nombre:	getChromosome ()
Descripción:	Devuelve el nombre del chromosome.
Nombre:	getStartPos()
Descripción:	Devuelve la posición inicial del transcrito.
Nombre:	getEnd_Pos()
Descripción:	Devuelve la posición final del transcrito.
Nombre:	getName()
Descripción:	Devuelve el nombre del gen.
Nombre:	getTabs_Snps()
Descripción:	Devuelve los valores de tab_snps.
Nombre:	setGeneID (String)
Descripción:	Cambia el identificador del gen especificado.
Nombre:	setChromosome (String)
Descripción:	Cambia el nombre del cromosoma.
Nombre:	setStartPos(int)
Descripción:	Cambia la posición inicial de un gen en el cromosoma.
Nombre:	setEnd_Pos(int)
Descripción:	Cambia la posición final de un gen en el cromosoma.

Nombre:	setName(String)
Descripción:	Cambia el nombre del gen.
Nombre:	setTabs_Snps(int[])
Descripción:	Cambia los valores de tab_snps.
Nombre:	getSequence()
Descripción:	Retorna los valores de la secuencia.
Nombre:	setSequence(String)
Descripción:	Cambia los valores de la secuencia
Nombre:	getSnpSequence()
Descripción:	Retorna los valores de la snpsequence.
Nombre:	addExon(Exon)
Descripción:	Adiciona un exon a la lista de exones
Nombre:	setSnpSequence()
Descripción:	Cambia los valores de snpsequence
Nombre:	initSequence()
Descripción:	Inicializa los valores de la secuencia

Tabla.37 Descripción de la clase Section.

Nombre: Section.	
Atributo	Tipo
startPos	int
end_pos	int
counter	int
there	boolean
Para cada responsabilidad:	
Nombre:	Section (int, int, int, boolean)
Nombre:	getStartPos()
Descripción:	Devuelve la posición inicial.
Nombre:	getEnd_Pos()
Descripción:	Devuelve la posición final.
Nombre:	setStartPos(int)
Descripción:	Cambia la posición inicial.

Nombre:	setEnd_Pos(int)
Descripción:	Cambia la posición final.
Nombre:	getCounter ()
Descripción:	Retorna los valores de counter.
Nombre:	setCounter (int)
Descripción:	Cambia los valores de counter.
Nombre:	isThere ()
Descripción:	Retorna los valores si está.
Nombre:	setThere (boolean)
Descripción:	Cambia los valores de there.

Tabla.38 Descripción de la clase MIRNA.

Nombre: MIRNA	
Atributo	Tipo
StartPos	int
end_pos	int
strand	String
name	String
score	float
pvalue	float
Para cada responsabilidad:	
Nombre:	MIRNA(String, int, int, String, float, float)
Descripción:	Inicializa los valores de los atributos.
Nombre:	getStartPos()
Descripción:	Devuelve la posición inicial.
Nombre:	getEnd_Pos()
Descripción:	Devuelve la posición final.
Nombre:	getStrand()
Descripción:	Devuelve la orientación del gen en el cromosoma, cadena (+) o cadena (-)
Nombre:	getName()
Descripción:	Devuelve el nombre del gen.
Nombre:	setStartPos(int)

Descripción:	Cambia la posición inicial de un gen en el cromosoma.
Nombre:	setEnd_Pos(int)
Descripción:	Cambia la posición final de un gen en el cromosoma.
Nombre:	setStrand(String)
Descripción:	Cambia la orientación del gen en el cromosoma, cadena (+) o cadena (-)
Nombre:	setName(String)
Descripción:	Cambia el nombre del gen.
Nombre:	getScore()
Descripción:	Devuelve el score.
Nombre:	getPValue()
Descripción:	Devuelve el pvalue
Nombre:	setStartScore (float)
Descripción:	Cambia el score.
Nombre:	setPValue (float)
Descripción:	Cambia el pvalue.

En relación con la versión anterior la programación de los web services es totalmente nuevo por tanto estas clases no estaban estructuradas de esta manera en dicha versión anterior, aunque se utilizó el recurso de reutilizar código como buena práctica en la elaboración de estas clases.

3.6 Descripción de las Tablas de la Base de Datos.

Tabla.39 Descripción de la tabla gene.

Nombre: gene		
Descripción: En esta tabla se almacenan las informaciones referentes a los genes.		
Atributo	Tipo	Descripción
id	integer	Identificador autogenerado del gen.
name	varchar	Nombre del gen de de acuerdo al NCBI
chromosome	varchar	Número del cromosoma donde se localiza el gen
start_pos	int	Posición inicial en el cromosoma.
end_pos	int	Posición final en el cromosoma.

strand	char	Orientación del gen en el cromosoma, cadena (+) o cadena (-)
band	varchar	La banda del gen en la que se encuentra.
type_of_gene	varchar	Tipo del gen (codificador, RNA, etc.)
description	text	Descripción de la función del gen.
taxonomy_id	int	Id del organismo a que pertenece el gen de acuerdo a la base Taxonomy del NCBI
source_inf_id	int	Identificador de la fuente de información
ncbi_geneid	varchar	Identificador del gen de acuerdo al NCBI
tax_id		Identificador de la taxonomía a la que pertenece
version	varchar	Versión del gen.

Tabla.40 Descripción de la tabla transcript.

Nombre: transcript		
Descripción: En esta tabla se almacenan las informaciones de los transcritos.		
Atributo	Tipo	Descripción
id	varchar	Identificador autogenerado del transcrito
gene_id	int	Gene _ id del gene en la tabla Gene
name	varchar	Nombre del transcrito de de acuerdo al NCBI
start_pos	int	Posición inicial en el cromosoma
end_pos	int	Posición final en el cromosoma
description	text	Descripción del transcrito
source_inf_id	int	Identificador de la fuente de recursos de donde provino.
ncbi_geneid	varchar	Identificador del gen que produce el transcrito
version	int	Versión del transcrito
transcript	varchar	Identificador del transcrito de acuerdo al NCBI
chromosome	varchar	Cromosoma del gen que produce el transcrito
translation	varchar	Identificador de la secuencia de aminoácidos de acuerdo al NCBI.
translation_version	varchar	Versión de la secuencia aminoácidos.

Tabla.41 Descripción de la tabla exon.

Nombre: exon		
Descripción: En esta tabla se almacenan las informaciones de los exones.		

Atributo	Tipo	Descripción
id	varchar	Identificador autogenerado del exón
gene_id	int	Gene _ id del gene en la tabla Gene
transcript_id	int	transcript_ id del transcrito gene en la tabla Transcript
start_pos	varchar	Posición inicial en el cromosoma
end_pos	varchar	Posición final en el cromosoma
band	varchar	La banda del cromosoma donde se encuentra
feature_type	varchar	Tipo de exón (UTR o CDS)
ncbi_geneid	varchar	Identificador del gen a que pertenece
transcript	varchar	Identificador del transcrito a que pertenece
chromosome	varchar	Cromosoma al que pertenece
chromosome_solo	varchar	Cromosoma al que pertenece

Tabla.42 Descripción de la tabla chromosome.

Nombre: chromosome		
Descripción: En esta tabla se almacenan las informaciones referentes a los cromosomas.		
Atributo	Tipo	Descripción
id	varchar	Identificador autogenerado del cromosoma
name	varchar	Nombre del cromosoma (1,2,3 ... X, Y)
known_genes	varchar	Genes conocidos.
unknown_genes	varchar	Genes desconocidos.
SNP	int	cantidad de SNP
length	varchar	Longitud del cromosoma
taxonomy_id	varchar	Identificador del organismo.

Tabla.43 Descripción de la tabla source_Info.

Nombre: seq_snp_md		
Descripción: En esta tabla se almacenan todos los datos referentes a los SNPs.		
Atributo	Tipo	Descripción
seq_snp_id	integer	Identificador autogenerado de la fuente de información.
chromosome	varchar	Nombre del cromosoma asociado.
chr_start	varchar	Nombre del cromosoma en el que esta el SNP.

chr_srt	varchar	Posición donde está el SNP
---------	---------	----------------------------

Tabla.44 Descripción de la tabla taxonomy.

Nombre: taxonomy		
Descripción: En esta tabla se almacenan todo lo referente a la información de los organismos.		
Atributo	Tipo	Descripción
id	varchar	Identificador del organismo.
name	varchar	Nombre del organismo.
alias	varchar	Alias del organismo.

Tabla.45 Descripción de la tabla mirna_gene.

Nombre: mirna_gene		
Descripción: En esta tabla se almacenan las informaciones referentes a los genes.		
Atributo	Tipo	Descripción
mirna_gene_id	integer	Identificador autogenerado del gen.
name	varchar	Nombre del gen de de acuerdo al NCBI
chromosome	varchar	Número del cromosoma donde se localiza el gen
start_pos	int	Posición inicial en el cromosoma.
end_pos	int	Posición final en el cromosoma.
strand	char	Orientación del gen en el cromosoma, cadena (+) o cadena (-)
band	varchar	La banda del cromosoma donde se encuentra
type_of_gene	varchar	Tipo del gen (codificador, RNA, etc.)
description	text	Descripción de la función del gen.
taxonomy_id	int	Id del organismo a que pertenece el gen de acuerdo a la base Taxonomy del NCBI
source_inf_id	int	Identificador de la fuente de información
ncbi_geneid	varchar	Identificador del gen de acuerdo al NCBI
tax_id		Identificador de la taxonomía a la que pertenece
Versión	varchar	Versión del gen.

Tabla.46 Descripción de la tabla mirna_target.

Nombre: mirna_target		
Descripción: En esta tabla se almacenan las informaciones referentes a los genes.		
Atributo	Tipo	Descripción
mirna_target_id	integer	Identificador autogenerado del gen.
seq	varchar	Nombre del gen de de acuerdo al NCBI
chr	varchar	Número del cromosoma donde se localiza el gen
start_pos	int	Posición inicial en el cromosoma.
end_pos	int	Posición final en el cromosoma.
strand	char	Orientación del gen en el cromosoma, cadena + o cadena
band	varchar	La banda del cromosoma donde se encuentra
score	varchar	Valores propios de los SNP almacenados
pvalue_og	text	Valor de significancia del SNP
external_name	varchar	Gen al que silencia.

No se han incluido nuevas tablas a la base de datos en el sistema puesto que se ha decidido que la base de datos de la versión anterior tiene una forma rápida y efectiva de responder a las diferentes consultas que se le pueden hacer.

3.7 Diagramas de Interacción (Secuencia, Colaboración)

Los diagramas de interacción describen secuencias de intercambios de mensajes entre los roles que implementan el comportamiento de un sistema. Estos diagramas proporcionan una vista integral del comportamiento del sistema, muestran el flujo de control a través de muchos objetos.

Los diagramas de interacción se exhiben en dos diagramas centrados en distintos aspectos pero complementarios:

Los diagramas de secuencia para el diseño del sistema propuesto muestran la secuencia de mensajes entre objetos durante un escenario concreto (**anexo 1**).

Los diagramas de colaboración son útiles en la fase exploratoria para identificar objetos. La distribución de los objetos en el diagrama permite observar adecuadamente la interacción de un objeto con respecto a los demás.

3.8 Diagrama de Despliegue.

Los Diagramas de despliegue muestran la disposición física de los distintos **nodos** que componen un sistema y el reparto de los componentes sobre dichos nodos.

Un nodo es un elemento físico que existe en tiempo de ejecución y representa un recurso computacional, que generalmente tiene algo de memoria y, a menudo, capacidad de procesamiento. Estos se utilizan para modelar la topología del hardware sobre el que se ejecuta el sistema. Representa típicamente un procesador o un dispositivo sobre el que se pueden desplegar los componentes.

Los estereotipos permiten precisar la naturaleza del equipo:

Procesadores: Nodo con capacidad de procesamiento. Puede ejecutar un componente.

Dispositivos: Nodo sin capacidad de procesamiento. Representa cualquier otro dispositivo hardware.

Los nodos se relacionan mediante conexiones bidireccionales.

Las conexiones se modelan como asociaciones, con todas las características que implica. Pueden ser USB, HTTP, JDBC, ADO, HTTPS, entre otras.

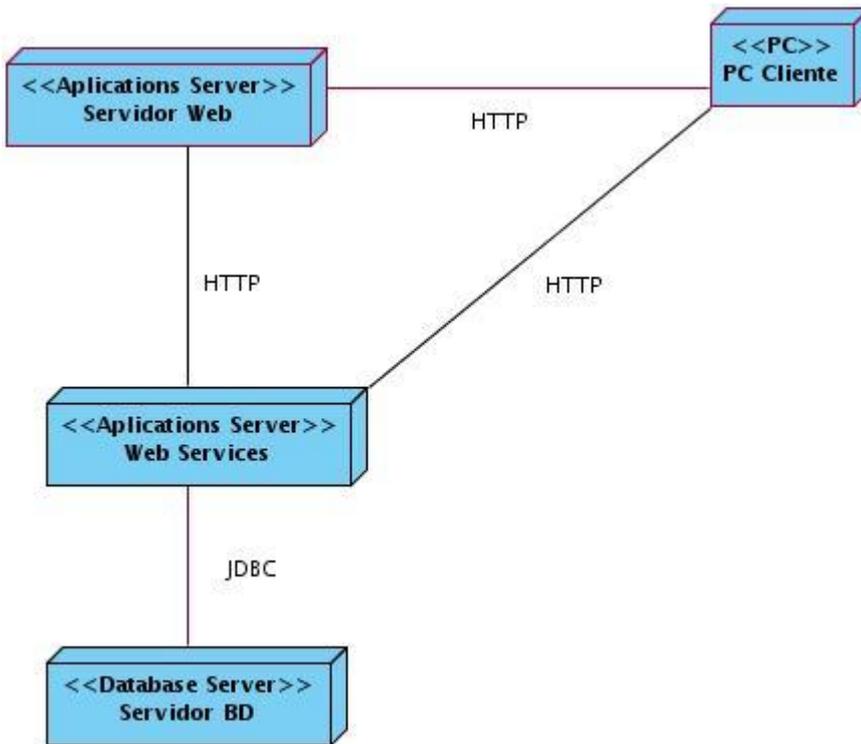


Fig. 22 Diagrama de despliegue

3.9 Tratamiento de errores

Para el tratamiento de errores se ve principalmente los que pueda cometer el usuario al interactuar con el sistema, por tanto se van a ver en la cliente del web services, por lo que siempre se verifican los datos de entrada antes de ser utilizados evitando incidentes y búsquedas innecesarias en la base de datos.

Esta aplicación hace uso de mensajes para señalarle al especialista que debe rectificar los datos introducidos antes de llevar a cabo las operaciones solicitadas por el mismo. Es de gran importancia verificar la integridad de la información introducida para de esta forma obtener un resultado de calidad. Un ejemplo de las validaciones que se realizan es cuando se selecciona un criterio de búsqueda pero no se introduce un valor [2].

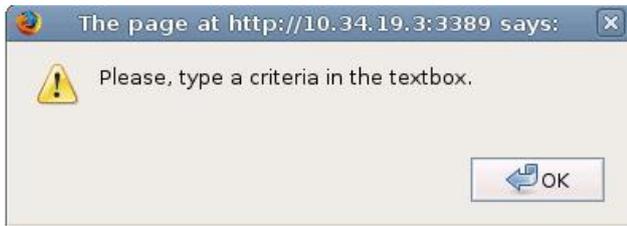


Fig. 23 Pantalla de error

En la aplicación también se tratan los errores a nivel de servidor en cuanto a las búsquedas que se realizan en la base de datos. Cuando un usuario busca una información que no está almacenada, el sistema a través de excepciones propias trata este tipo de error. El problema a resolver con esto es que cada vez que se realiza una búsqueda, esta desencadena dos o tres búsquedas más, si la primera existe. Una vez que se comprueba que no existe la primera, el sistema controla que las demás no se ejecuten a través de la excepciones que el implementa. Con esto no solo se evita búsquedas innecesarias sino también demora de tiempo en los procedimientos [2].

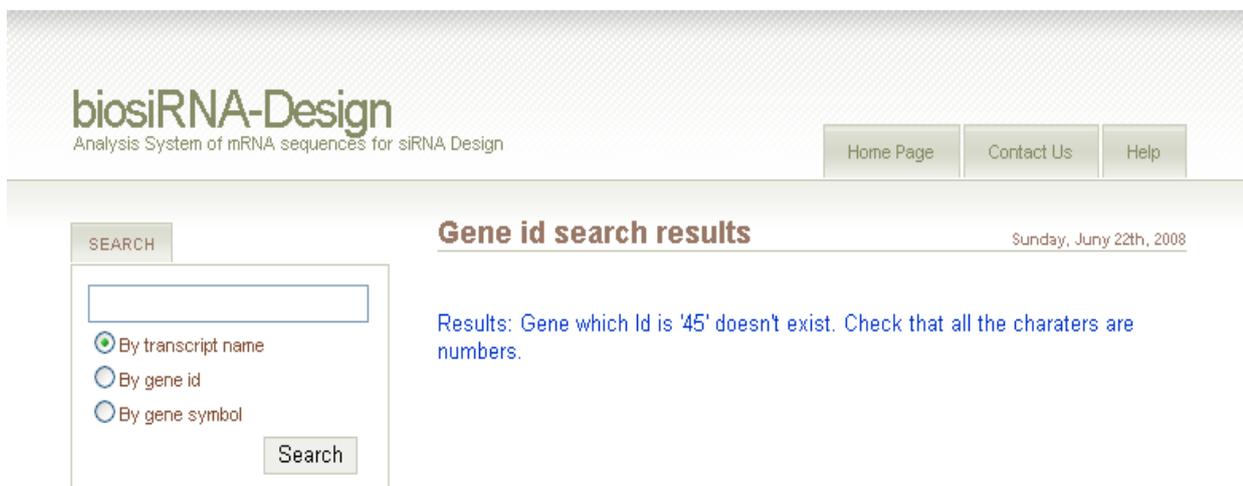


Fig. 24 Pantalla de error

Cuando en la aplicación ocurren errores inesperados, esta los trata de manera que no se pierde su integridad. Al ocurrir estos sucesos, el sistema los captura y redirecciona el usuario a una página de error mostrándole el mensaje de error.

3.10 Seguridad

La seguridad de la herramienta que se presenta se evidencia a través del servidor de base de datos PostgreSQL. A la información almacenada en este solo tendrá acceso el administrador del sistema, con otra aplicación en este caso de escritorio la cual actualizará las tablas de la base de datos de manera local en el centro cliente de la aplicación.

En la aplicación, se evidencia también la seguridad a través de la dirección web que se muestra en el navegador donde se está ejecutando el sistema. Todas las páginas se muestran bajo una misma dirección web, y las direcciones no coinciden con el nombre de la página mostrada. Con esto se evita que se acceda a las distintas páginas de la aplicación sin las precondiciones correspondientes, ya que se afectaría la integridad de toda la herramienta [2].

3.11 Conclusiones

En este capítulo se determina el diseño apropiado para el software en su conjunto, tanto para el cliente como para el web services, brindando una clara visión de los principios de diseño utilizados para la construcción del sistema propuesto, así como la definición de diseño aplicada. Para esto se expone primeramente la vista lógica del sistema. Luego se presentaron los diagramas de clases para la versión 2.0 de alasSiRNA - Design, la cual sigue el mismo patrón de arquitectura de la versión anterior, y los diagramas de clases para el web services. Además de los diagramas de interacción de dichas clases para una implementación más detallada y organizada de los procesos que llevará a cabo el sistema. Se realizó también el modelo de datos generado a partir del diagrama de clases persistentes y se describieron cada una de las tablas que conforman la base de datos y se obtuvo el modelo de despliegue con el cual quedó representada la distribución física del sistema y se enfatizó en el tratamiento de errores y en la seguridad del sistema. Se demostró la importancia de esta etapa y se ganó claridad en cuanto a la concesión del sistema a construir y se sentaron las bases para la fase de implementación y prueba del sistema.

CAPÍTULO 4: IMPLEMENTACIÓN Y PRUEBA

Introducción.

En el presente capítulo se realizará el modelo de implementación a partir de los resultados del flujo de trabajo anteriormente descrito (Diseño), Se utilizará los mismos algoritmos que tenía la versión anterior más la inclusión de 4 algoritmos que dan claridad a las investigaciones de los especialistas en el cual se detalla su funcionamiento.

Además se van a describir las pruebas realizadas al sistema, teniéndose en cuenta principalmente los casos de prueba de caja negra, que las mismas se realizan sobre la interfaz del software.

4.1 Diagrama de componentes

Los diagramas de componentes describen los elementos físicos del sistema y sus relaciones, muestran las opciones de realización incluyendo código fuente, binario y ejecutable.

Los componentes representan todos los tipos de elementos que entran en la fabricación de aplicaciones informáticas. Pueden ser simples archivos, paquetes, bibliotecas cargadas dinámicamente, cada uno de los cuales tiene un estereotipo diferente.

Son utilizados para modelar la vista estática de un sistema. Muestra la organización y las dependencias que existen entre un conjunto de componentes.

4.2 Diagramas de componentes del proyecto

El proyecto tiene como característica antes mencionada que tiene un web services que brinda respuesta a consultas y análisis realizados por cualquier aplicación cliente. Además tiene como interfaz

para conectarse a dicho web services la versión 2.0 de la aplicación alasSiRNA - Design. Por tanto se debe realizar sendos diagramas de componentes al proyecto.

4.2.1 Diagramas de componentes de la aplicación del alasSiRNA - Design v2.0.

- En el paquete **Vista** se encuentran todos los archivos que interactúan directamente con el usuario ya sean páginas html, o jsp y estos se encargan de visualizar la información resultante. Algunas de estas páginas tienen incluido código java script para validación de entrada de datos.
- En el paquete **Controlador** están incluidos los ficheros que hacen el procesamiento de los datos obtenidos desde la interfaz del siRNA web services para mandárselo a los clientes.
- En el paquete **Modelo** se agrupan los ficheros encargados de almacenar con una lógica definida la información manejada por la aplicación de acuerdo al tipo de procedimiento que se esté realizando.

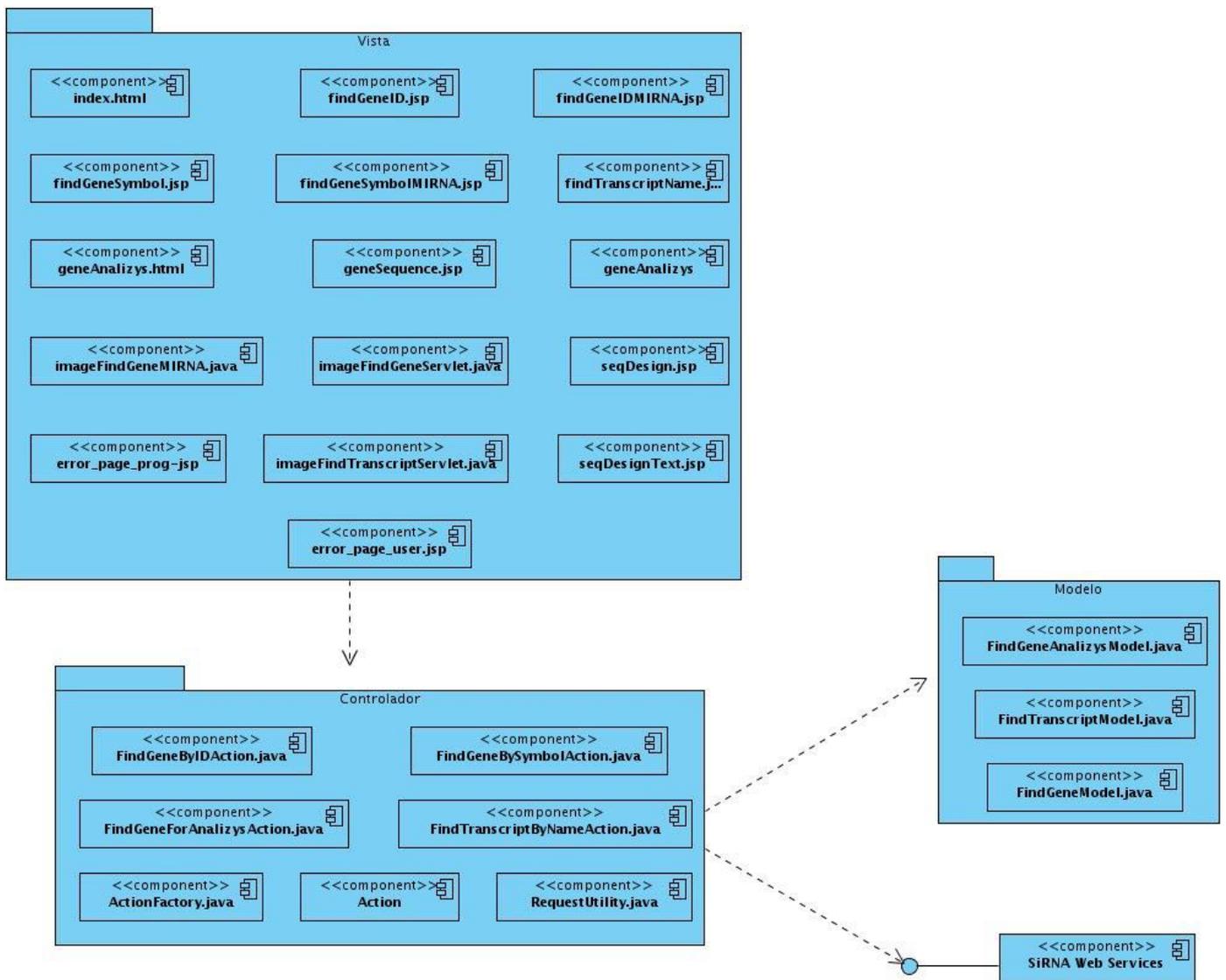


Fig. 25 Diagrama de componentes de alasSiRNA - Design v2.0.

4.2.2 Diagramas de componentes del web services.

Los web services no están ligados a una arquitectura de componentes en particular. Los protocolos facilitan a nivel base la comunicación entre las distintas infraestructuras de objetos distribuidos. Por este motivo, únicamente es necesario preocuparse del desarrollo y utilización de Servicios Web. Por tanto en este diagrama de componentes se agrupan las clases en dependencia de la utilidad de estas en la aplicación.

El paquete Controler está la clase ControlerClass que lo que realiza es controlar las clases y las distintas funcionalidades del web services.

El paquete DataAccess tiene los ficheros relacionados con el acceso a datos, ya sea a la base de datos como a los ficheros fasta con información de los cromosomas que se encuentra en el paquete Ficheros de Cromosomas.

En el paquete utilsclass están todos los archivos con clases que servirán de apoyo en la resolución de los problemas planteados.

El componente siRNA.java es el que devuelve todas las funcionalidades del web services.

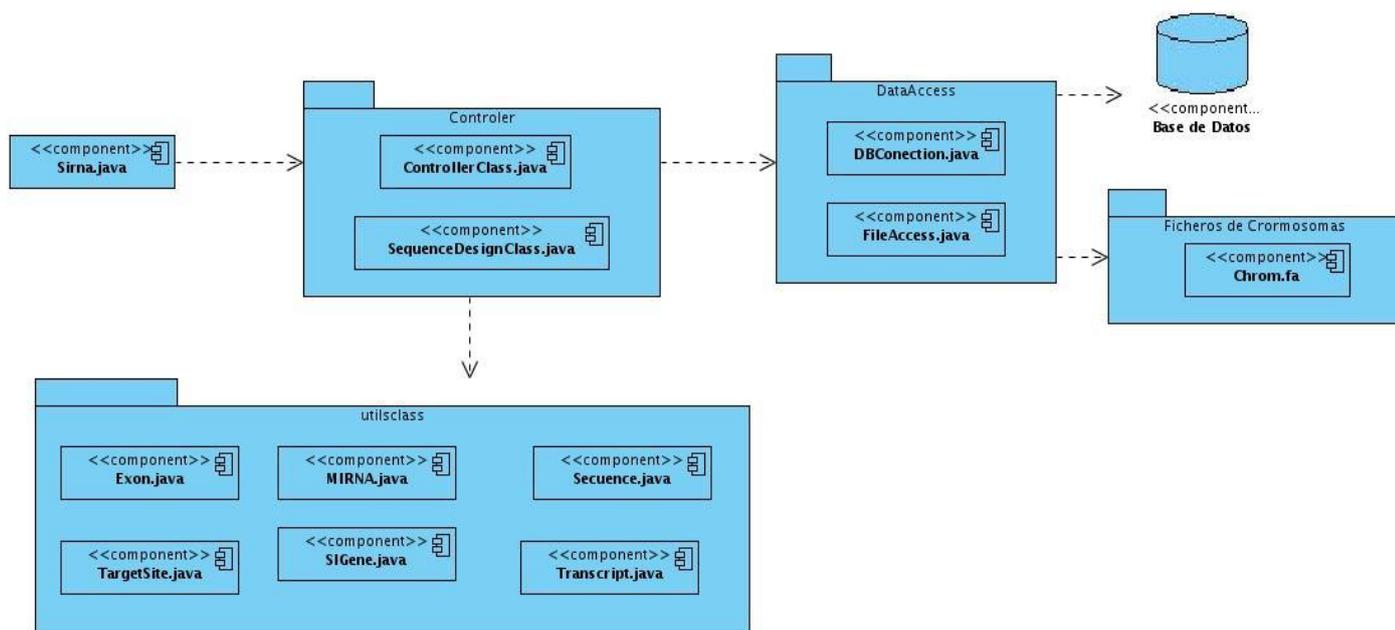


Fig. 26 Diagrama de componentes de siRNA Web Services.

4.3 Obtención de algoritmos

En la versión 2.0 de alasSiRNA - Design se incluyen nuevas funcionalidades las cuales serán descritas en este epígrafe.

4.3.1 Obtención de algoritmo para la búsqueda de la secuencia común en un gen determinado.

Para el análisis profundo de los mRNA se hace necesario realizar una búsqueda de un mRNA que sea óptimo a todo el gen para así al silenciar este RNA se silencie el gen completo el cual se describe de la siguiente manera:

- Una vez encontrado un determinado gen se procede con la búsqueda de los transcritos asociados a dicho gen.
- Luego de cada transcrito se buscan los exones y se realiza una intersección de estos.
- Después se concatenan las secuencias resultantes de la operación anterior.
- Luego se devuelve la cadena de RNA formada para que se le realice la búsqueda de sus sitios blancos.

4.3.2 Obtención de algoritmo para la búsqueda de la secuencia específica en un transcrito determinado.

Complementario al algoritmo anterior se añade la problemática de silenciar solamente la parte del gen que se desee a partir de los transcritos, algoritmo que se describe de la siguiente manera:

- Una vez encontrado un determinado transcrito se procede con la búsqueda del gen asociado a ese transcrito.
- Después se realiza la búsqueda de los transcritos que tiene este gen.
- Luego de cada transcrito se buscan los exones, se comparan y se desechan aquellos que coincidan con los del transcrito determinado.
- Después se concatenan las secuencias de los exones que son específicas del transcrito determinado inicialmente.
- Luego se devuelve la cadena de RNA formada para que se le realice la búsqueda de sus sitios blancos.

4.3.3 Obtención de algoritmo para calcular la energía libre entre el sitio blanco y el siRNA complementario a este.

Para una mayor especificación en los resultados obtenidos al buscar los siRNAs a determinada secuencia de mRNA se hace necesario calcular la energía libre ya que mientras menor sea esta más eficiente es el siRNA.

- Una vez encontrado un determinado el sitio target se recorre el mismo.
- En dicho recorrido se realiza una sumatoria con las siguientes especificaciones.

	A/U	C/G	G/C	U/A
A/U	-0.9	-1.8	-2.3	-1.1
C/G	-1.7	-2.9	-3.4	-2.3
G/C	-2.1	-2.0	-2.9	-1.8
U/A	-0.9	-1.7	-2.1	-0.9

Fig. 27 Tabla para la obtención de la energía libre.

- Luego se devuelve el resultado de la sumatoria que sería la energía asociada a cada uno de los sitios blancos.

4.4 Modelo de Prueba.

En cada una de las fases, en la construcción de un proyecto, hay que realizarle pruebas al software que se va desarrollando, centrándose principalmente en la fase de implementación en la cual se realiza el mayor esfuerzo de este flujo de trabajo. Las pruebas se hacen con el objetivo de medir la calidad del producto y de descubrir errores de cualquier tipo que pueda tener la aplicación.

4.4.1 Casos de prueba de caja negra

Las pruebas de caja negra se refiere a las pruebas que se llevan a cabo sobre la interfaz del software, por lo que los casos de prueba pretenden demostrar que las funciones del software son operativas,

que la entrada se acepta de forma adecuada y que se produce una salida correcta, así como que la integridad de la información externa se mantiene. Se examinan fundamentalmente algunos aspectos del modelo del sistema sin profundizar mucho en la estructura interna del software.

Caso de Prueba 1: Buscar datos de un gen.

Flujo Central.

1. El sistema brinda la opción de escoger un criterio de búsqueda para buscar y mostrar los datos de un gen (“By gene id”, “By gene symbol”).
2. El especialista:
 - Selecciona un criterio de búsqueda.
 - Introduce los datos requeridos (id o símbolo del gen)
 - Ordena la búsqueda.

El sistema:

3. Valida que se entre la información de un criterio de búsqueda.
4. Busca los datos del gen (número del cromosoma, id y símbolo).
5. Busca los datos de los transcritos asociados al gen (posición inicial y final, cantidad de exones, nombre y regiones codificantes y no traducidas de los exones asociados a los transcritos).
6. Busca los datos de los miRNA asociados al gen (nombre, Pvalue, SCORE, posición inicial y final).
7. Muestra los datos del gen.
8. Muestra gráficamente los datos de los transcritos asociados.
9. Muestra gráficamente los datos de los miRNA asociados al gen.
10. El sistema brinda la posibilidad de seleccionar un Pvalue (nivel de significancia) para visualizar los miRNA asociados al gen con un nivel de significancia que como máximo sea el especificado.
11. El sistema da la posibilidad de mostrar en detalle un transcrito y analizar su secuencia de nucleótidos.
12. Se realiza el caso de uso “Buscar datos de un transcrito” y se ejecuta el caso de uso “Realizar análisis de la secuencia mRNA” y así termina la ejecución del caso de uso.

Flujos Alternos:

2.2 El especialista introduce la información del criterio de búsqueda de un gen que no tiene miRNA.

2.2.1 El sistema no muestra los miRNA.

2.3 El especialista introduce la información del criterio de búsqueda de un gen que no tiene transcritos.

2.3.1 El sistema no muestra los transcritos.

13.1 Se culmina la ejecución del caso de uso.

10.1 El especialista selecciona un PValue y ordena que se muestren los miRNA.

10.2 El sistema muestra los miRNA que tengan un valor de significancia menor que el especificado.

3.1 Si el especialista no introduce los datos requeridos el sistema muestra un mensaje de error.

Iteraciones:

Tabla.47 Resultado del caso de prueba Buscar datos de un gen.

Clases válidas	Resultado esperado	Resultado de la prueba
<Se selecciona en Search la opción By gene id y se introduce el siguiente dato: 400>	<El sistema verifica que el dato entrado es correcto, brinda la opción de introducir la información a partir del criterio de búsqueda seleccionado, muestra datos de gen(id y símbolo, cromosoma correspondiente y cantidad de transcritos asociados), busca transcritos asociados al gen, muestra datos de los transcritos asociados al gen (posición inicial y final en la cadena, nombre, exones, regiones codificantes y regiones no traducidas) y muestra los miRNA correspondientes al gen y sus datos >	<Satisfactorio>
<Se selecciona en Search la opción By	<El sistema verifica que el dato entrado es válido, brinda la opción de introducir la información a partir del criterio de	<Satisfactorio>

gene symbol y se entra el siguiente dato: ARL1>	búsqueda seleccionado, muestra datos de gen(id y símbolo, cromosoma correspondiente y cantidad de transcritos asociados), busca transcritos asociados al gen, muestra datos de los transcritos asociados al gen (posición inicial y posición final en la cadena, nombre, exones, regiones codificantes y regiones no traducidas) y muestra los miRNA correspondientes al gen y sus datos>	
<Se selecciona en Search a opción By gene symbol y se entra el siguiente dato: GSR>	<El sistema verifica que el dato entrado es incorrecto y muestra el mensaje de error: "Gene which symbol is 'gsr' doesn't exist. Check that the letters are capitalized">	<Satisfactorio>
<Se selecciona la opción By gene id y se entra el siguiente dato: ADR2>	<El sistema verifica que el dato entrado es incorrecto y muestra el siguiente mensaje de error: "Gene which Id is 'adr2' doesn't exist. Check that all the characters are numbers">	<Satisfactorio>

Caso de Prueba 2: Mostrar un transcrito

Flujo Central:

1. El sistema brinda la opción de escoger un criterio de búsqueda para buscar y mostrar los datos de un transcrito.
2. El especialista selecciona el criterio de búsqueda "By transcript name", introduce la información requerida (nombre del transcrito) y ordena la búsqueda.
3. El sistema valida que se entre la información referente al criterio de búsqueda.
4. El sistema busca los datos del transcrito (posición inicial y final, cantidad de exones, nombre y regiones codificantes y no traducidas de los exones asociados al transcrito y regiones SNP).
5. El sistema busca la secuencia de nucleótidos del transcrito.
6. El sistema muestra gráficamente los datos del transcrito.
7. El sistema muestra la secuencia de nucleótidos.

8. El sistema da la posibilidad de analizar la secuencia de nucleótidos del transcrito visualizando las reglas para dicho análisis.
9. Se realiza el caso de uso "Realizar análisis de la secuencia mRNA" a partir de la acción 4 del flujo normal de eventos.

Flujos Alternos:

1.1 El sistema visualiza una interfaz donde muestra los transcritos asociados a un gen y sus datos y da la posibilidad de seleccionar uno de ellos para visualizar en detalle sus datos.

1.2 El especialista selecciona el nombre del transcrito y ordena la búsqueda.

1.3 El sistema busca los datos del transcrito (secuencia de nucleótidos y regiones SNP).

1.4 El sistema muestra gráficamente los datos del transcrito (posición inicial y final, cantidad de exones, nombre y regiones codificantes y no traducidas de los exones asociados al transcrito y regiones SNP).

1.5 Se realizan las acciones de la 6 a la 9 del flujo normal de eventos.

3.1 El sistema muestra un mensaje de error en caso de que el especialista no haya introducido la información solicitada.

9.1 Se sale del caso de uso.

Iteraciones:

Tabla.48 Resultado del caso de prueba Mostrar datos de un transcrito.

Clases válidas	Resultado esperado	Resultado de la prueba
<Se selecciona en la sección "Search" la opción By transcript name y se entra el siguiente dato: NM_001025091>	<El sistema verifica que el dato entrado es correcto, busca exones asociados al transcrito, muestra id y nombre del transcrito y cantidad de exones, muestra una gráfica con datos de los exones asociados al transcrito y muestra una gráfica de los SNP presentes en el transcrito y su cadena mRNA.>	<Satisfactorio>

<Se selecciona la opción By transcript name y se entra un transcrito que no tiene SNP>	<El sistema verifica que la entrada es válida, y no muestra SNP>	<Satisfactorio>
<Se selecciona la opción By transcript name y se entra el siguiente dato: NMOR>	<El sistema verifica que el dato entrado es incorrecto y muestra el mensaje de error: "The transcripts which name is 'nmor' doesn't exist. Check that the letters are capitalized.">	<Satisfactorio>

Caso de Prueba 3: Analizar una secuencia mRNA.

Flujo Central:

El especialista selecciona la opción "Paste your own sequence" para analizar la secuencia.

1. El sistema visualiza la interfaz para analizar la secuencia.
2. El especialista selecciona un criterio de búsqueda (id o símbolo del gen) y ordena la búsqueda.

El sistema:

3. El especialista selecciona las reglas para analizar la secuencia mRNA y ordena el análisis.
4. El sistema verifica que al menos se seleccione una regla y que la secuencia sea correcta.
5. El sistema verifica si el especialista ha seleccionado la regla de excluir los SNP de los resultados ("Exclude SNP region from the design (No For Pasted Sequence)").
6. El sistema busca sitios blancos en la secuencia y los SNP asociados a la misma.

7. El sistema muestra gráficamente los resultados de la búsqueda de los sitios blancos en la cadena mRNA y señala los sitios blancos con SNP ordenados por la cantidad de reglas cumplidas de las seleccionadas.
8. El sistema muestra una gráfica con los SNP de la secuencia.
9. El sistema brinda la posibilidad de visualizar en detalle la información de un sitio blanco.

Flujo Alternativo:

- 1.1 El usuario realiza la búsqueda de un gen determinado.
- 1.2 El sistema da la opción de analizar un transcrito determinado. De este paso sigue en la opción 3.

8.1 El especialista selecciona un sitio blanco para ver su información.

8.2 El sistema visualiza los datos del sitio blanco seleccionado (secuencia de nucleótidos, posiciones inicial y final de la misma, regiones SNP, energía libre y número de reglas).

7.1 El sistema busca los sitios blancos en la cadena excluyendo los que contienen SNP.

Iteraciones:

Tabla.49 Resultado del caso de prueba Analizar una secuencia mRNA.

Clases válidas	Resultado esperado	Resultado de la prueba
<Se selecciona en Paste your own sequence y se pega una secuencia de nucleótidos>	<El sistema muestra las reglas para el análisis y la posibilidad de pegar la secuencia, luego de presionar la opción analizar, busca sitios blancos en la secuencia, luego muestra la secuencia resaltando sitios blancos y sus datos, muestra una gráfica de la secuencia mRNA resaltando sitios blancos y brinda la opción de seleccionar de forma individual el segmento que contiene un sitio blanco específico luego muestra una nueva ventana con la información de los sitios blancos (Segmento de la cadena correspondiente,	<Satisfactorio>

	número de reglas que cumple el transcrito al que pertenece, posición inicial y final, número y la energía libre).>	
<Se selecciona en Paste your own sequence y no se pega nada en el cuadro de texto>	<El sistema muestra mensaje especificando que debe pegar una secuencia de mRNA">	<Satisfactorio>
<después de buscado un gen determinado se selecciona un transcrito y se da la opción de analizar >	<El sistema muestra las reglas para el análisis y la secuencia, luego busca sitios blancos en la secuencia, los SNP, luego de presionar la opción analizar, muestra la secuencia resaltando sitios blancos y sus datos, muestra una gráfica de la secuencia mRNA resaltando sitios blancos, muestra una gráfica con los SNP asociados al sitio blanco correspondiente y brinda la opción de seleccionar de forma individual el segmento que contiene un sitio blanco específico luego muestra una nueva ventana con la información de los sitios blancos (Segmento de la cadena correspondiente, número de reglas que cumple el transcrito al que pertenece, posición inicial y final, número y SNP en caso de tener).>	<Satisfactorio>

Caso de Prueba 4: Mostrar la secuencia comun a un gen especificado.

Flujo Central.

1. El especialista escoge la opción de analizar un gen específico.
2. El sistema brinda la opción de escoger un criterio de búsqueda para buscar y mostrar los datos de un gen ("By gene id", "By gene symbol").
3. El especialista:
 - Selecciona un criterio de búsqueda.
 - Introduce los datos requeridos (id o símbolo del gen)
 - Ordena la búsqueda.

El sistema:

4. Valida que se entre la información de un criterio de búsqueda.
5. Busca los datos del gen (número del cromosoma, id y símbolo).
6. Busca los datos de los transcritos asociados al gen (posición inicial y final, cantidad de exones, nombre y regiones codificantes y no traducidas de los exones asociados a los transcritos).
7. Analiza los exones comunes en todos los transcritos del gen determinado
8. Muestra la opción del caso de prueba 3
9. Muestra gráficamente los datos de los transcritos asociados.

Flujos Alternos:

- 4.1 Si el especialista no introduce los datos requeridos el sistema muestra un mensaje de error.

Iteraciones:

Tabla.50 Resultado del caso de prueba Mostrar la secuencia comun a un gen especificado.

Clases válidas	Resultado esperado	Resultado de la prueba
<Se selecciona en “Common sequence of the transcripts” en la opción By gene id y se introduce el siguiente dato: 400>	<El sistema verifica que el dato entrado es correcto brinda la opción de introducir la información a partir del criterio de búsqueda seleccionado, muestra los datos del caso de prueba 3>	<Satisfactorio>
<Se selecciona en “Common sequence of the transcripts” en la opción By gene symbol y se entra el siguiente dato: ARL1>	<El sistema verifica que el dato entrado es correcto brinda la opción de introducir la información a partir del criterio de búsqueda seleccionado, muestra los datos del caso de prueba 3>	<Satisfactorio>

<Se selecciona en en la opción "Common sequence of the transcripts" By gene symbol y se entra el siguiente dato: GSR>	<El sistema verifica que el dato entrado es incorrecto y muestra el mensaje de error: "Gene which symbol is 'gsr' doesn't exist. Check that the letters are capitalized">	<Satisfactorio>
<Se selecciona la opción By gene id y se entra el siguiente dato: ADR2>	<El sistema verifica que el dato entrado es incorrecto y muestra el siguiente mensaje de error: "Gene which Id is 'adr2' doesn't exist. Check that all the characters are numbers">	<Satisfactorio>

Caso de Prueba 5: Mostrar la secuencia específica a un transcrito especificado.

Flujo Central:

1. El especialista escoge la opción de analizar un transcrito específico.
2. El especialista selecciona el criterio de búsqueda "By transcript name", introduce el nombre del transcrito
3. El sistema busca dicho transcrito.
4. El sistema valida que se entre la información referente al criterio de búsqueda.
5. El sistema busca el gen asociado al transcrito determinado.
6. El sistema realiza el análisis de los exones que son específicos de ese transcrito.
7. El sistema procede al caso de prueba 3

Flujos Alternos:

2.1 El sistema muestra un mensaje de error en caso de que el especialista no haya introducido la información solicitada.

Iteraciones:

Tabla.51 Resultado del caso de prueba Mostrar la secuencia específica a un transcrito especificado.

Clases válidas	Resultado esperado	Resultado de la prueba
<p><Se selecciona en la sección “Specific sequence of a transcript” la opción By transcript name y se entra el siguiente dato: NM_001025091></p>	<p><El sistema verifica que el dato entrado es correcto, busca exones asociados al transcrito, y procede al caso de prueba 3.></p>	<p><Satisfactorio></p>
<p><Se selecciona en la sección “Specific sequence of a transcript” con la opción By transcript name y se entra el siguiente dato: NMOR></p>	<p><El sistema verifica que el dato entrado es incorrecto y muestra el mensaje de error: “The transcripts which name is ‘nmor’ doesn’t exist. Check that the letters are capitalized.”></p>	<p><Satisfactorio></p>

4.5 Conclusiones.

Se describieron detalladamente los diagramas de componentes que corresponden al modelo de implementación del sistema. Se analizaron los algoritmos que se le incluyen a esta nueva versión los cuales son descritos detalladamente. Se implementó con una arquitectura orientada a servicios a partir de los artefactos generados por los flujos de trabajos anteriores y se obtiene una aplicación con las siguientes pantallas del sistema (**anexo 2**).

Se realizaron las pruebas de caja negra donde se pudo comprobar las funcionalidades de la aplicación. Durante la etapa de prueba se comprobó la existencia de algunos errores que estaban en contra del

buen funcionamiento del software, los cuales fueron rectificadas a medida que se fueron realizando pruebas posteriores.

CONCLUSIONES GENERALES.

El presente trabajo realiza una propuesta de solución a los requisitos planteados por el especialista del CIGB que es nuestro cliente, referente a la utilización de una nueva arquitectura y la incorporación de algoritmos para aumentar la eficiencia en el diseño de los siRNA, cumpliendo los requisitos de confiabilidad y seguridad de toda la información que se maneja. Esta solución se puede usar en el Centro de Ingeniería Genética y Biotecnología de Cuba y en todas las demás instituciones nacionales e internacionales que necesiten hacer uso de dicha herramienta.

La herramienta adoptó una arquitectura orientada a servicios, la cual es más factible para los trabajos que requieren en la institución que utiliza dicha aplicación. Tiene un ambiente fácil de entender y usar por cualquier personal que esté calificado para la interacción con la misma. Se le agregaron funcionalidades como la búsqueda de una secuencia específica a un determinado transcrito, la búsqueda de la secuencia común a un determinado gen y el cálculo de la energía libre, las que aumentan el nivel de especificación en los análisis que se pueden realizar. El sistema cumple estándares de diseño y utiliza técnicas de programación orientada a objetos.

Al resultado del trabajo se le han realizado diferentes pruebas con resultados satisfactorios por lo que se concluye que los objetivos propuestos han sido cumplidos satisfactoriamente. Se incluyen además varias recomendaciones que deben tenerse en cuenta para un futuro seguimiento de este trabajo.

RECOMENDACIONES

Con el desarrollo del trabajo se han cumplido los objetivos del mismo, a pesar de ello nuevas ideas han ido surgiendo con el desarrollo del mismo, lo cual da la posibilidad de que se llegue a crear una aplicación más funcional y eficiente, que brinde toda clase de servicios sobre la base de las investigaciones en el campo de los estudios genéticos, por tanto recomendamos:

1. Continuar el desarrollo del sistema con el objetivo de implementar otras funcionalidades que se desean para próximas versiones.
2. Optimizar las búsquedas en ficheros y base de datos.
3. Realizar funcionalidades como servicios para la integración y utilización con otros sistemas que requieran dicha información.
4. Aplicar el sistema a todas las instituciones nacionales e internacionales donde pueda ser usado, permitiendo un buen uso del mismo.
5. Hacer uso de un framework en versiones superiores del sistema, con el objetivo de hacer más flexible y reutilizable la arquitectura lograda.

REFERENCIAS BIBLIOGRÁFICAS

- [1] ARN de interferencia, el detonante genético. [consultado el: 24 de febrero de 2009] Disponible en: http://www.diariomedico.com/edicion/diario_medico/mi_dm/bionoticias/genomica/es/desarrollo/694711_04.html
- [2] Tirado, N. and K. Roman, SISTEMA DE ANÁLISIS DE SECUENCIAS MRNA PARA EL DISEÑO DE SIRNA. V 1.1, in Facultad 6. 2009, UCI: Ciudad Habana. p. 108.
- [3] siRNA Target Finder. [Consultado el: 12 de febrero de 2009]. Disponible en: http://www.Ambion.com/techlib/misc/siRNA_tools.html
- [4] siRNA Design Software. [Consultado el: 16 de febrero de 2009]. Disponible en: <http://i.cs.hku.hk/~siRNA/software/siRNA.php>
- [5] Custom siRNA Design Tool. [Consultado el: 6 de febrero de 2009]. Disponible en: <http://www.dharmacon.com/sidesign/default.aspx>
- [6] DEQOR. [Consultado el 17 de febrero de 2009]. Disponible en: <http://cluster-1.mpi-cbg.de/Deqor/deqor.html>
- [7] IDT´Scitools RNAi Design.[Consultado el: 16 de febrero de 2009]. Disponible en: <http://www.idtdna.com/scitools/applications/rnai/rnai.aspx>
- [8] siDirect. [Consultado el 16 de febrero de 2009]. Disponible en: <http://genomics.jp/sidirect/index.php?type=fc>
- [9] siRNA Wizard v3.0. [Consultado el 16 de febrero de 2009]. <http://www.siRNAwizard.com/design.php>
- [10] Perezleo, L., et al. Impacto de la Bioinformática en las ciencias biomédicas.
- [11] Parrado, V., et al., Silenciamiento de Genes. Informática 2009, 2009.
- [12] Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. 2001 Jan 18;409(6818):363-6
- [13] Imagen procesor de siRNA [Consultado el 26 de febrero de 2009] Disponible en: <http://www.gene-quantification.de/siRNA-mechanism.png>
- [14] Glasser V. Tapping miRNA-Regulated Pathways. Genetic Eng Biotech News Mar 1 2008 (Vol. 28,

No. 5)

[15] Imagen miRNA [Consultado el: 11 de febrero de 2009]. Disponible en:

<http://es.wikipedia.org/wiki/MIRNA>

[16] Devlin, T. M. 2004. Bioquímica, 4ª edición. Reverté, Barcelona.

[17] SNP [Consultado el 26 de febrero de 2009] Disponible en:

http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml

[18] La tecnología J2EE. [Consultado el 8 de Febrero de 2009]. Disponible en:

<http://www.jtech.ua.es/j2ee/2003-2004/present.htm>

[19] Informatización, UCI. Arquitectura para los Sistemas que Conforman la Intranet Universitaria.in UCI. 2007. Habana: CUBA

[20] Desarrollo Orientado a Objetos con UML. [Consultado el : 18 de Febrero de 2009]. Disponible en:

<http://www.clikear.com/manuales/uml/introduccion.aspx>

[21] Herramientas CASE [Consultado el 8 de Febrero de 2009]. Disponible en:

<http://www.cyta.com.ar/biblioteca/bddoc/bdlibros/proyectoinformatico/libro/c5/c5.htm>

[22] Visual Paradigm. [Consultado el: 8 de Febrero de 2009]. Disponible en:

[http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_\(M%C3%8D\)_1472_0_p/](http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_(M%C3%8D)_1472_0_p/)

[23] NetBeans. [Consultado el: 8 de Febrero de 2009]. Disponible en:

http://www.netbeans.org/index_es.html

[24] Correa, Y. and Y. Diaz, Sistema de análisis de secuencias mRNA para el diseño de siRNA, in facultad 6. 2009, UCI: Habana. p. 120.

[25] Netbeans vs Eclipse [consultado el: 24 de marzo de 2009] disponible en:

<http://www.juanjonavarro.com/masquecodigo/2003/12/09/netbeans-frente-a-eclipse>

[26] PostgreSQL [Consultado el: 8 de Febrero de 2009]. Disponible en:

http://www.computerworld.com.au/article/62894/postgresql_affiliates_org_domain

[27] MySQL [consultado el: 25 de marzo de 2009]

http://www.netpecos.org/docs/mysql_postgres/x57.html

[28] Quiñones. E, A. Introducción a Postgres. Asociación Peruana de Software Libre

[29] Tomcat [Consultado el : 18 de Febrero de 2009]. Disponible en:

http://www.programacion.com/tutorial/tomcatintro/1/#1_intro

[30] JBOSS [consultado el: 24 de febrero de 2009] Disponible en: <http://www.jboss.org/>

[31] Booch, G., Rumbaugh, J., Jacobson, I. "El Lenguaje Unificado de Modelado". Addison-Wesley. 1999

[32] Garcilaso Jordana. Introducción Open UP [consultado el: 24 de febrero de 2009] Disponible en:

http://www.mug.org.ar/Descargas/Jornadas/Downloads_GetFile.aspx?id=3136

BIBLIOGRAFÍA

- Tirado, N. and K. Roman, SISTEMA DE ANÁLISIS DE SECUENCIAS MRNA PARA EL DISEÑO DE SIRNA. V 1.1, in Facultad 6. 2009, UCI: Ciudad Habana. p. 108.
- siRNA [Consultado el: 20 de febrero de 2009]. Disponible en: <http://es.wikipedia.org/wiki/siRNA>
- Ahmed, H., et al. Prediction of guide strand of microRNAs from its sequence and secondary structure, in BMC Bioinformatics 2009. Chandigarh, India. Disponible en: <http://www.biomedcentral.com/1471-2105/10/105>
- siRNA Target Finder. [Consultado el: 12 de febrero de 2009]. Disponible en: http://www.Ambion.com/techlib/misc/siRNA_tools.html
- siRNA Design Software. [Consultado el: 16 de febrero de 2009]. Disponible en: <http://i.cs.hku.hk/~siRNA/software/siRNA.php>
- Custom siRNA Design Tool. [Consultado el: 6 de febrero de 2009]. Disponible en: <http://www.dharmacon.com/sidesign/default.aspx>
- DEQOR. [Consultado el 17 de febrero de 2009]. Disponible en: <http://cluster-1.mpi-cbg.de/Degor/deqor.html>
- IDT Scitools RNAi Design.[Consultado el: 16 de febrero de 2009]. Disponible en: <http://www.idtdna.com/scitools/applications/rnai/rnai.aspx>
- siDirect. [Consultado el 16 de febrero de 2009]. Disponible en: <http://genomics.jp/sidirect/index.php?type=fc>
- siRNA Wizard v3.0. [Consultado el 16 de febrero de 2009]. <http://www.siRNAwizard.com/design.php>
- Correa, Y. and Y. Diaz, Sistema de análisis de secuencias mRNA para el diseño de siRNA, in facultad 6. 2009, UCI: Habana. p. 120.
- Perez, J. Web Services, in Universidad Politécnica de Valencia, Valencia: España. p. 15.
- Perezleo, L., et al. Impacto de la Bioinformática en las ciencias biomédicas.
- Parrado, V., et al., Silenciamiento de Genes. Informática 2009, 2009.

- MiRNA [Consultado el: 11 de febrero de 2009]. Disponible en: <http://es.wikipedia.org/wiki/MIRNA>
- ARN mensajero [Consultado el : 18 de febrero de 2009]. Disponible en: http://es.wikipedia.org/wiki/ARN_mensajero
- Single nucleotide polymorphism.[Consultado el 17 de febrero de 2009] Disponible en: http://en.wikipedia.org/wiki/Single_nucleotide_polymorphism
- Vienna RNA Package.[Consultado el 17 de febrero de 2009] Disponible en: <http://www.tbi.univie.ac.at/RNA/>
- La tecnología J2EE. [Consultado el 8 de Febrero de 2009]. Disponible en: <http://www.jtech.ua.es/j2ee/2003-2004/present.htm>
- Informatización, UCI. Arquitectura para los Sistemas que Conforman la Intranet Universitaria.in UCI. 2007. Habana: CUBA
- Herramientas CASE [Consultado el 8 de Febrero de 2009]. Disponible en: <http://www.cyta.com.ar/biblioteca/bddoc/bdlibros/proyectoinformatico/libro/c5/c5.htm>
- Visual Paradigm. [Consultado el: 8 de Febrero de 2009]. Disponible en: [http://www.freownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_\(M%C3%8D\)_14720_p/](http://www.freownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_(M%C3%8D)_14720_p/)
- NetBeans. [Consultado el: 8 de Febrero de 2009]. Disponible en: http://www.netbeans.org/index_es.html
- PostgreSQL [Consultado el: 8 de Febrero de 2009]. Disponible en: <http://es.wikipedia.org/wiki/PostgreSQL>
- Quiñones. E, A. Introducción a Postgres. Asociación Peruana de Software Libre
- Tomcat [Consultado el : 18 de Febrero de 2009]. Disponible en: http://www.programacion.com/tutorial/tomcatintro/1/#1_intro
- Desarrollo Orientado a Objetos con UML. [Consultado el : 18 de Febrero de 2009]. Disponible en: <http://www.cliekar.com/manuales/uml/introduccion.aspx>
- ARN de interferencia, el detonante genético. [consultado el: 24 de febrero de 2009] Disponible en: http://www.diariomedico.com/edicion/diario_medico/mi_dm/bionoticias/genomica/es/desarrollo/6

[94711_04.html](#)

- Netbeans vs Eclipse [consultado el: 24 de marzo de 2009] disponible en:
<http://www.juanjonavarro.com/masquecodigo/2003/12/09/netbeans-frente-a-eclipse>
- MySQL [consultado el: 25 de marzo de 2009]
http://www.netpecos.org/docs/mysql_postgres/x57.html
- Booch, G., Rumbaugh, J., Jacobson, I. "El Lenguaje Unificado de Modelado". Addison-Wesley. 1999
- Garcilaso Jordana. Introducción Open UP [consultado el: 24 de febrero de 2009] Disponible en:
http://www.mug.org.ar/Descargas/Jornadas/Downloads_GetFile.aspx?id=3136
- JBOSS [consultado el: 24 de febrero de 2009] Disponible en: <http://es.wikipedia.org/wiki/JBoss>
- SNP [Consultado el 26 de febrero de 2009] Disponible en:
http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml
- Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. 2001 Jan 18;409(6818):363-6
- Glasser V. Tapping miRNA-Regulated Pathways. Genetic Eng Biotech News Mar 1 2008 (Vol. 28, No. 5)
- Devlin, T. M. 2004. Bioquímica, 4ª edición. Reverté, Barcelona.
- PostgreSQL [Consultado el: 8 de Febrero de 2009]. Disponible en:
http://www.computerworld.com.au/article/62894/postgresql_affiliates_org_domain
- Ernesto Quiñones A. [Consultado el : 18 de Febrero de 2009]. Disponible en:
[http://www.apesol.org/ Asociación Peruana de Software Libre](http://www.apesol.org/)
- JBOSS [consultado el: 24 de febrero de 2009] Disponible en: <http://www.jboss.org/>

ANEXOS

Anexo 1: Diagramas de Secuencia

Diagrama de secuencia CU Mostrar_Secuencia_Comun_Gen. alasSiRNA - Design v2.0

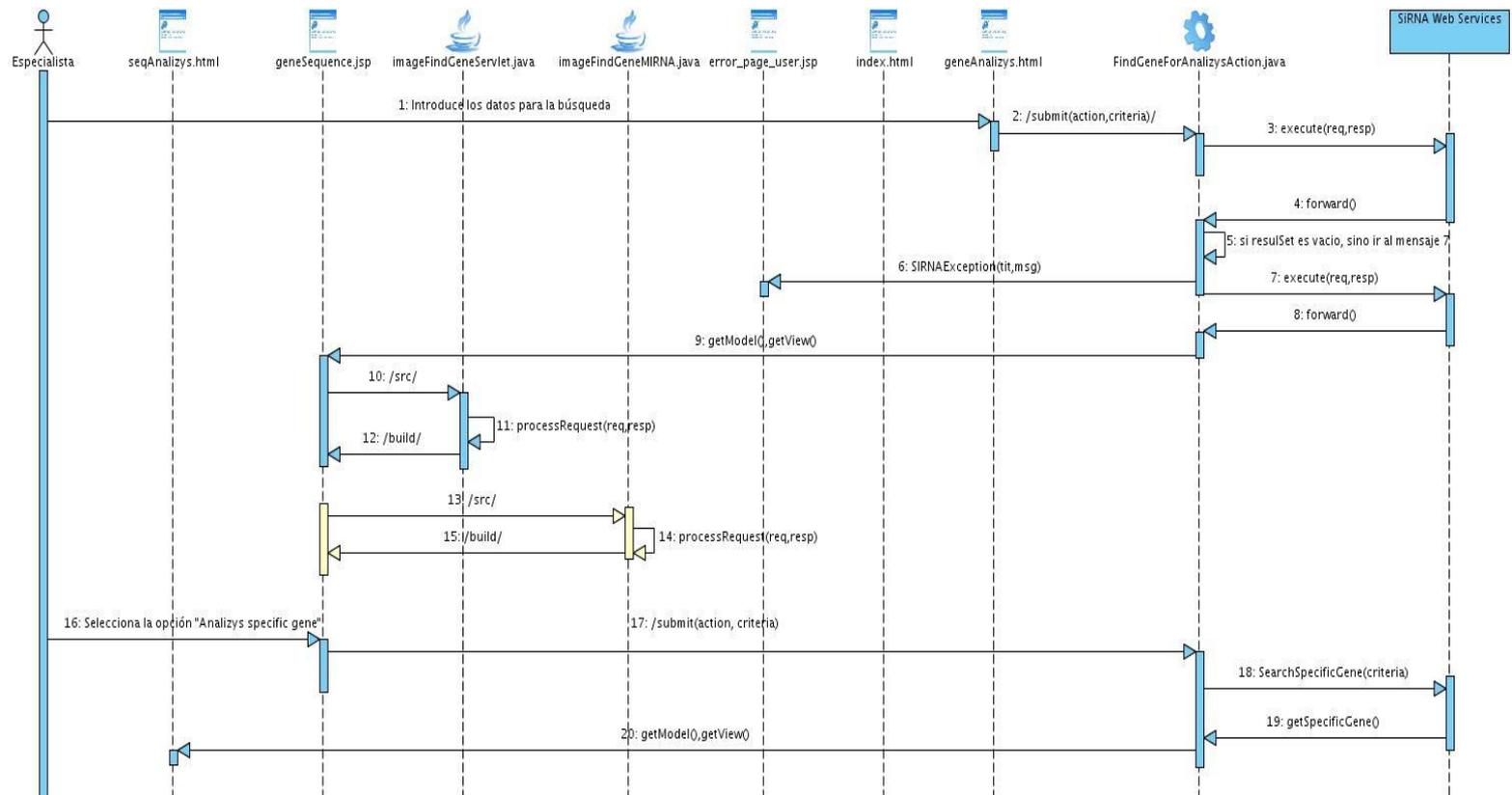


Diagrama de secuencia CU Analizar_Secuencia_mRNA. alasSiRNA - Design v2.0

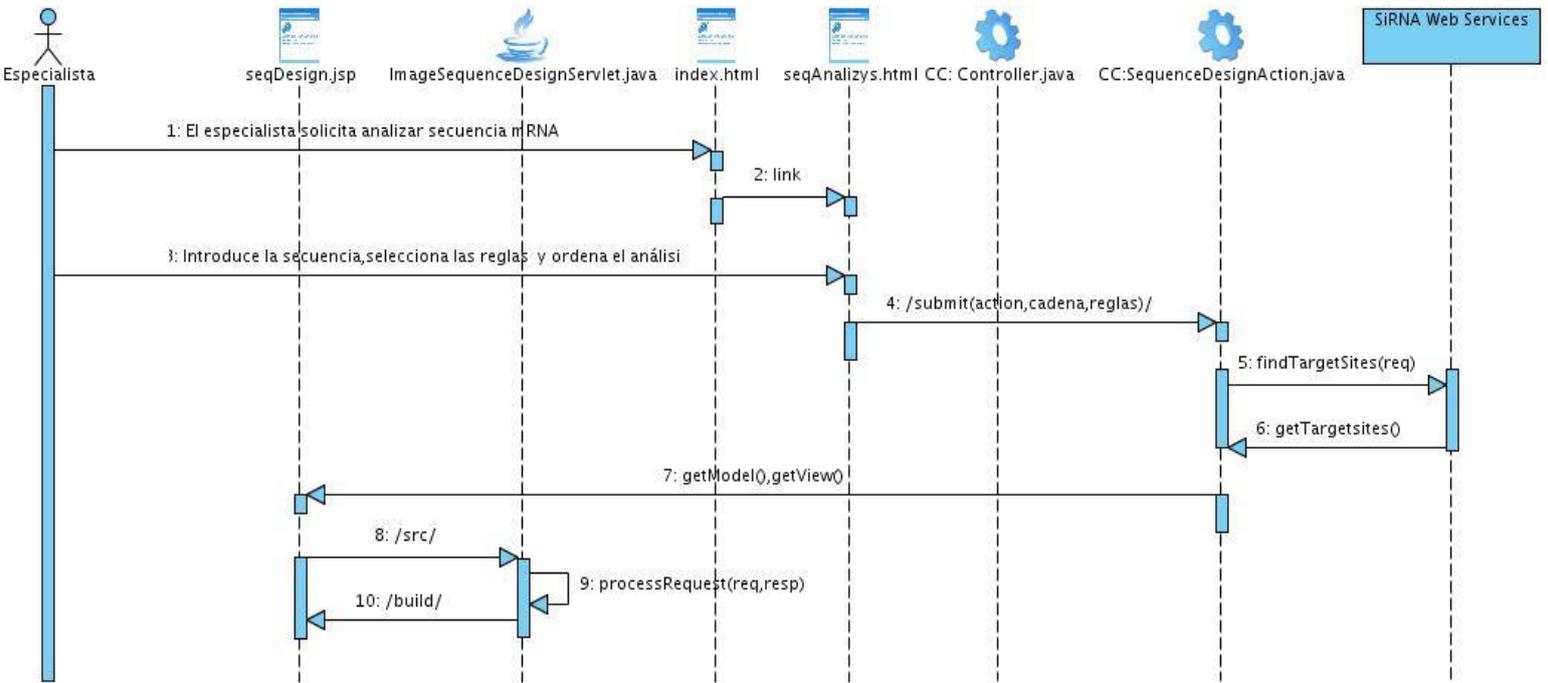


Diagrama de secuencia CU Mostrar_Secuencia_Especifico_Transcrito. alasSiRNA - Design v2.0

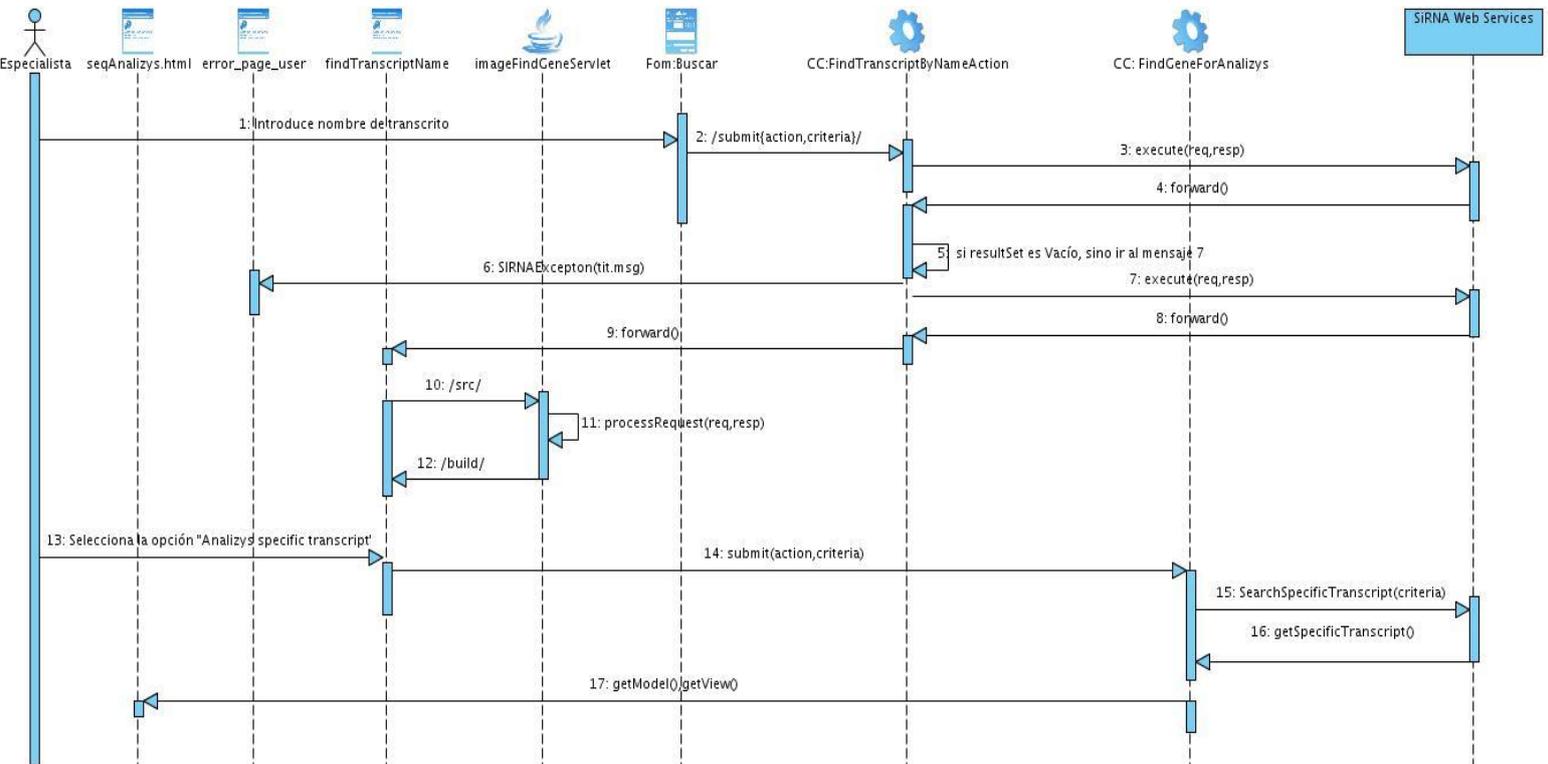


Diagrama de secuencia CU Mostrar_Transcrito. alasSiRNA - Design v2.0

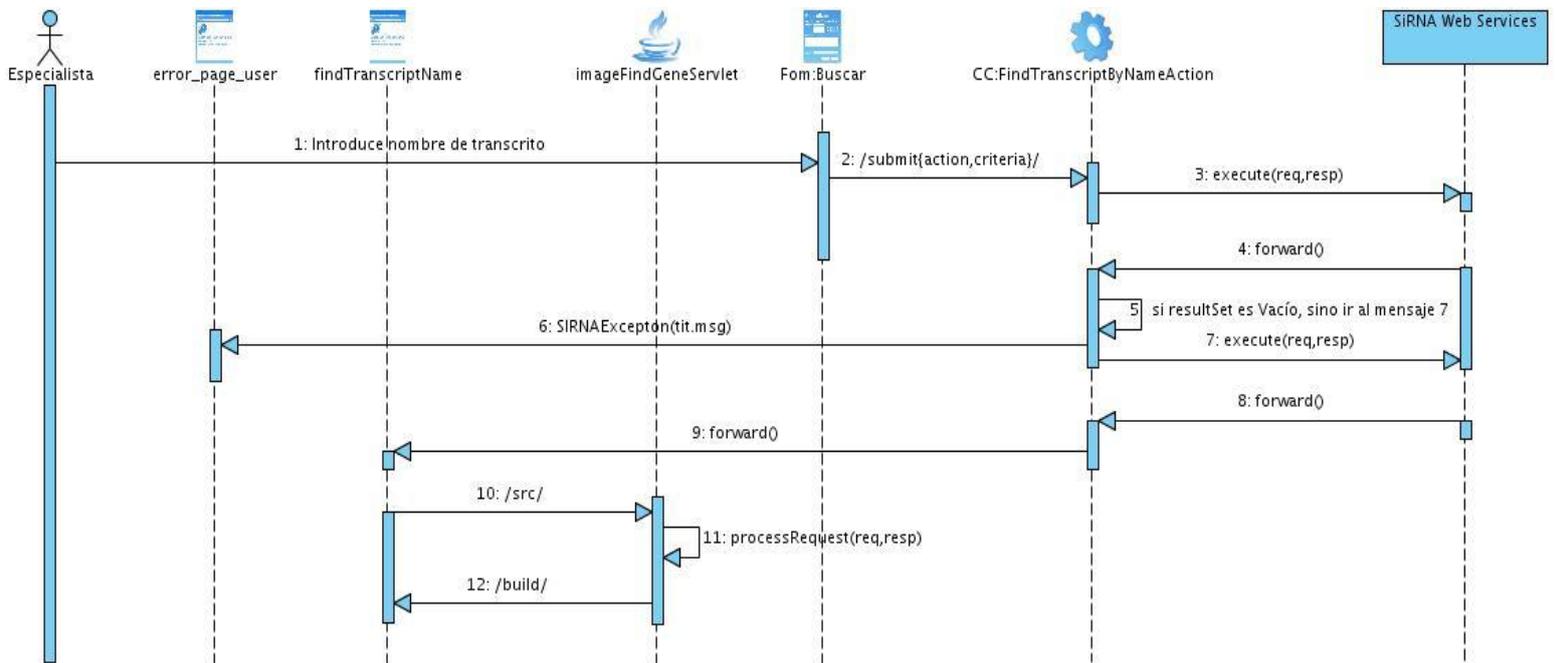


Diagrama de secuencia CU Mostrar_Gen. alasSiRNA - Design v2.0

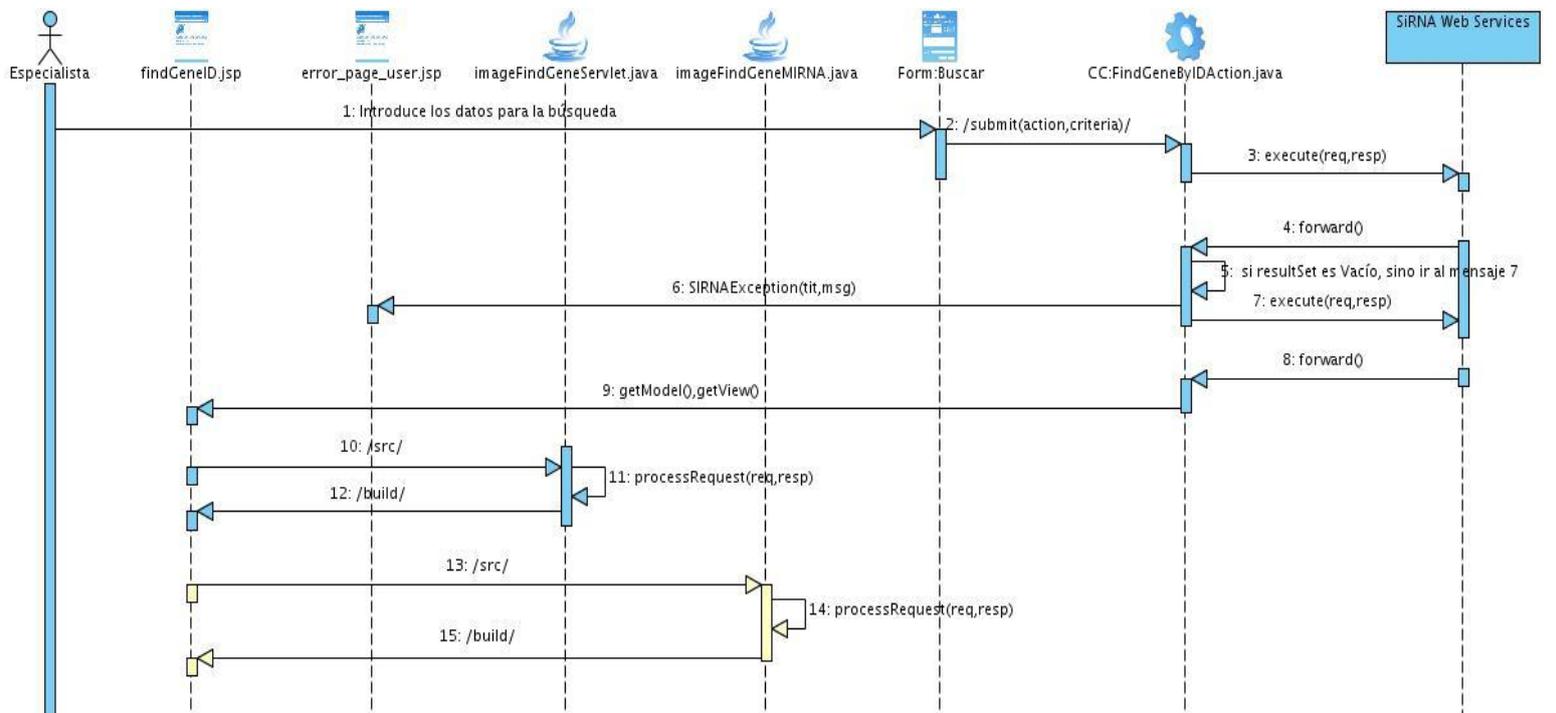


Diagrama de secuencia CU Listar_SNP_en_region_cromosoma. siRNA web Services

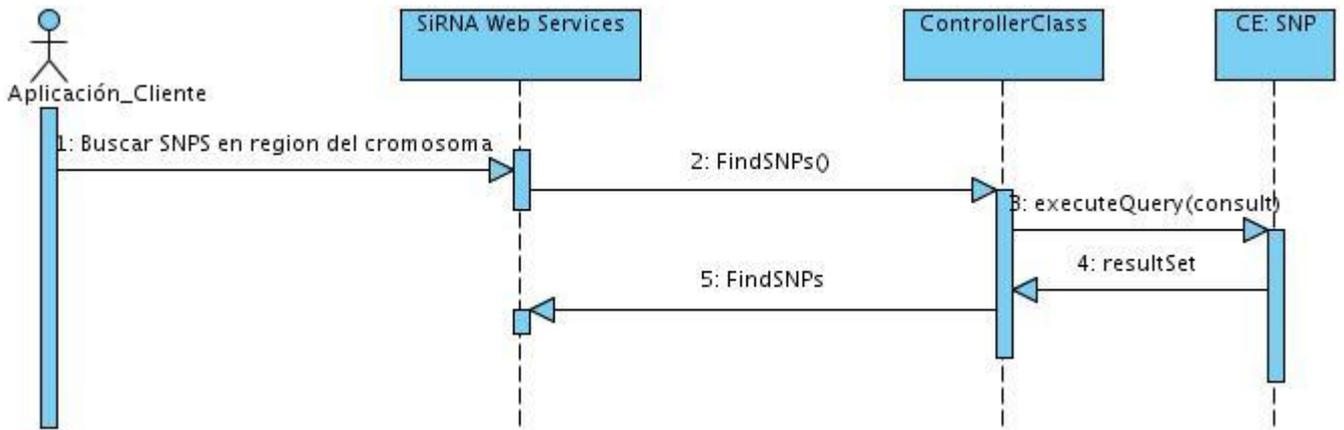


Diagrama de secuencia CU Buscar_Transcrito. siRNA web Services

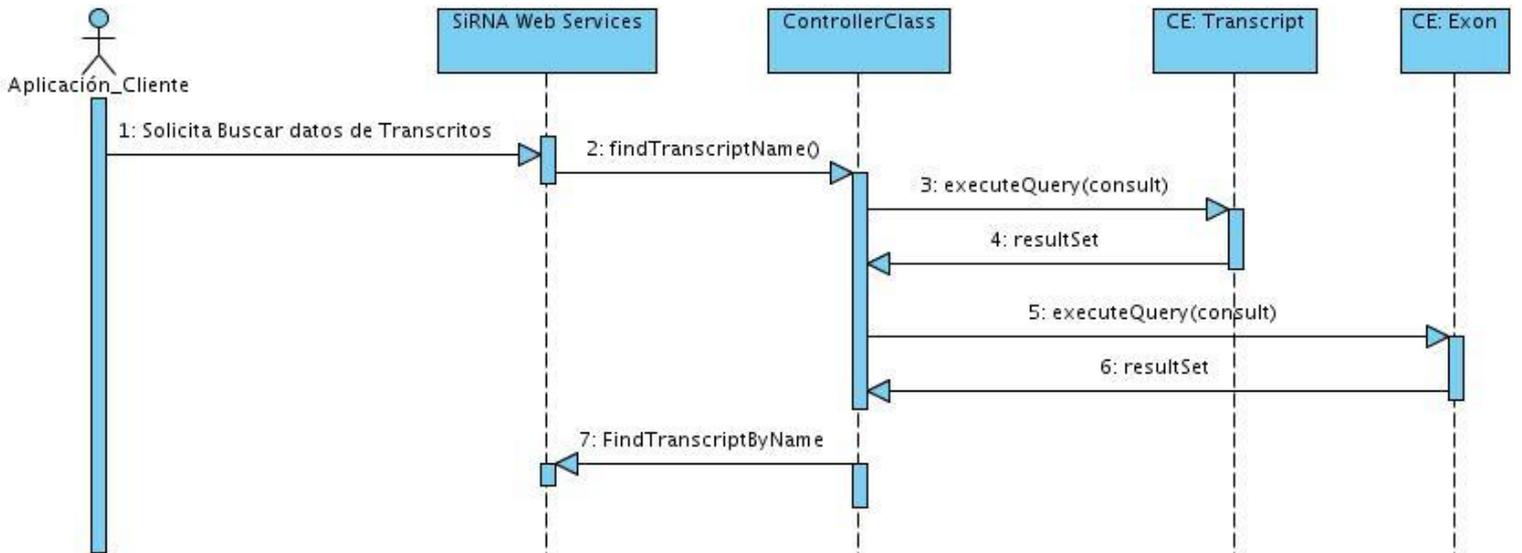


Diagrama de secuencia CU Buscar_Gen. siRNA web Services

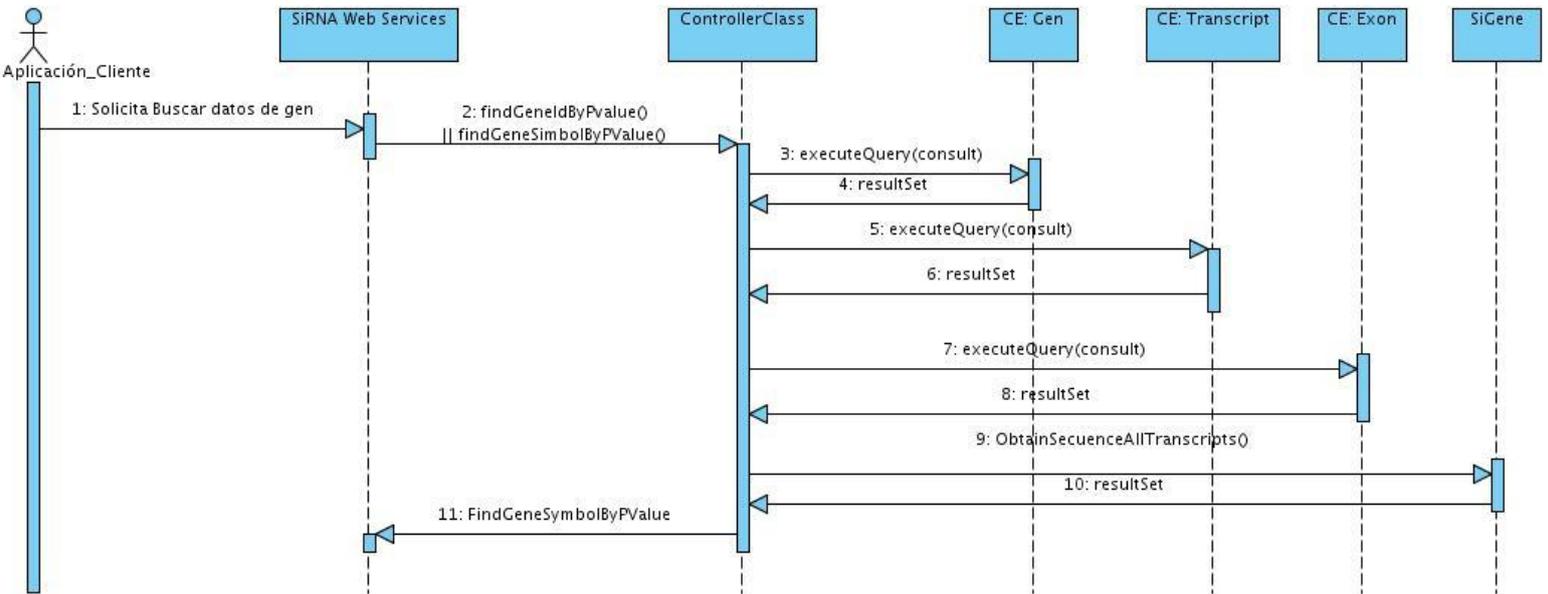


Diagrama de secuencia CU Analizar_Secuencia_mRNA. siRNA web Services

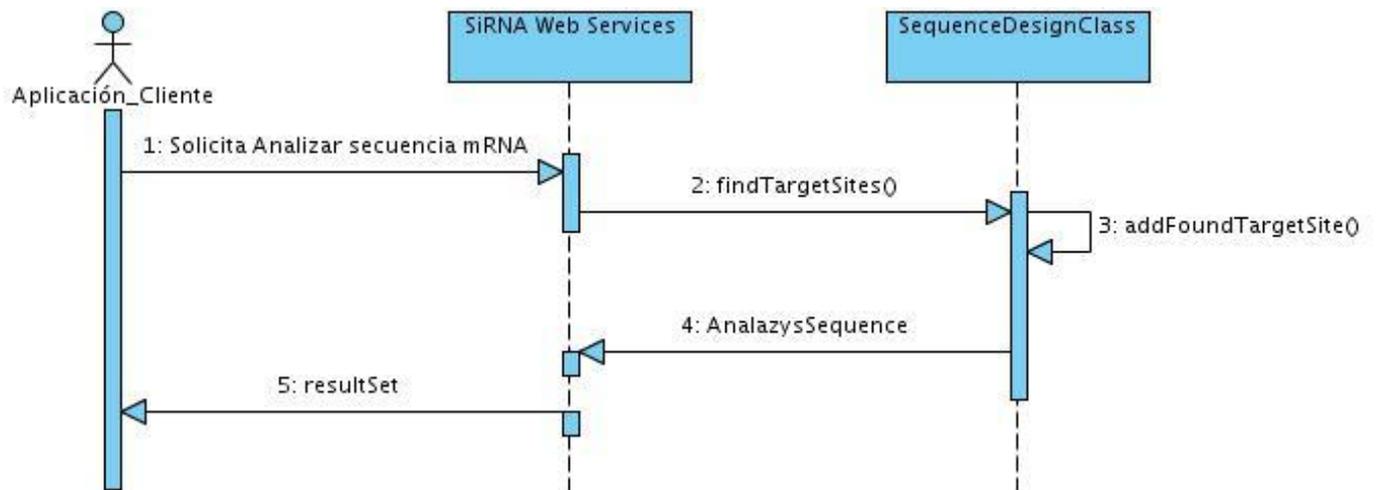


Diagrama de secuencia CU Buscar_Secuencia_Comun. siRNA web Services

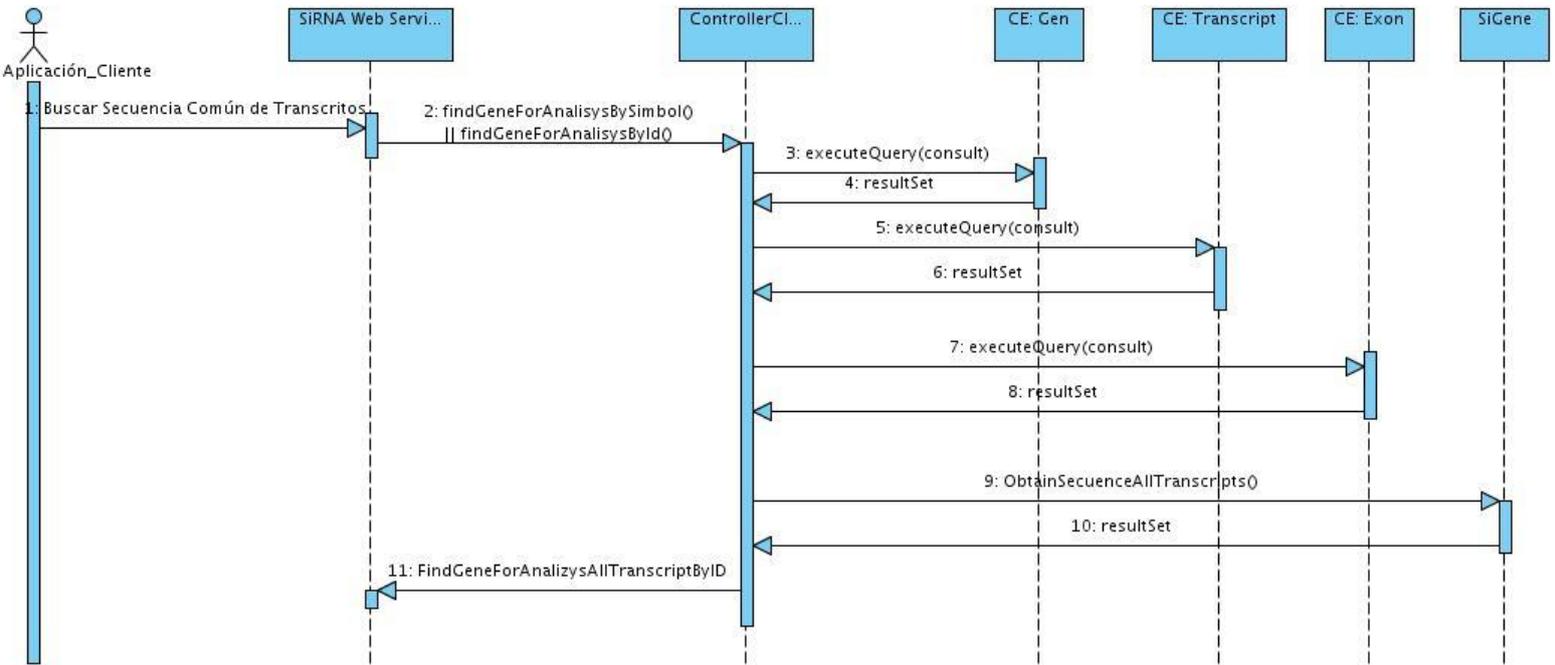


Diagrama de secuencia CU Buscar_Secuencia_Especifico. siRNA web Services

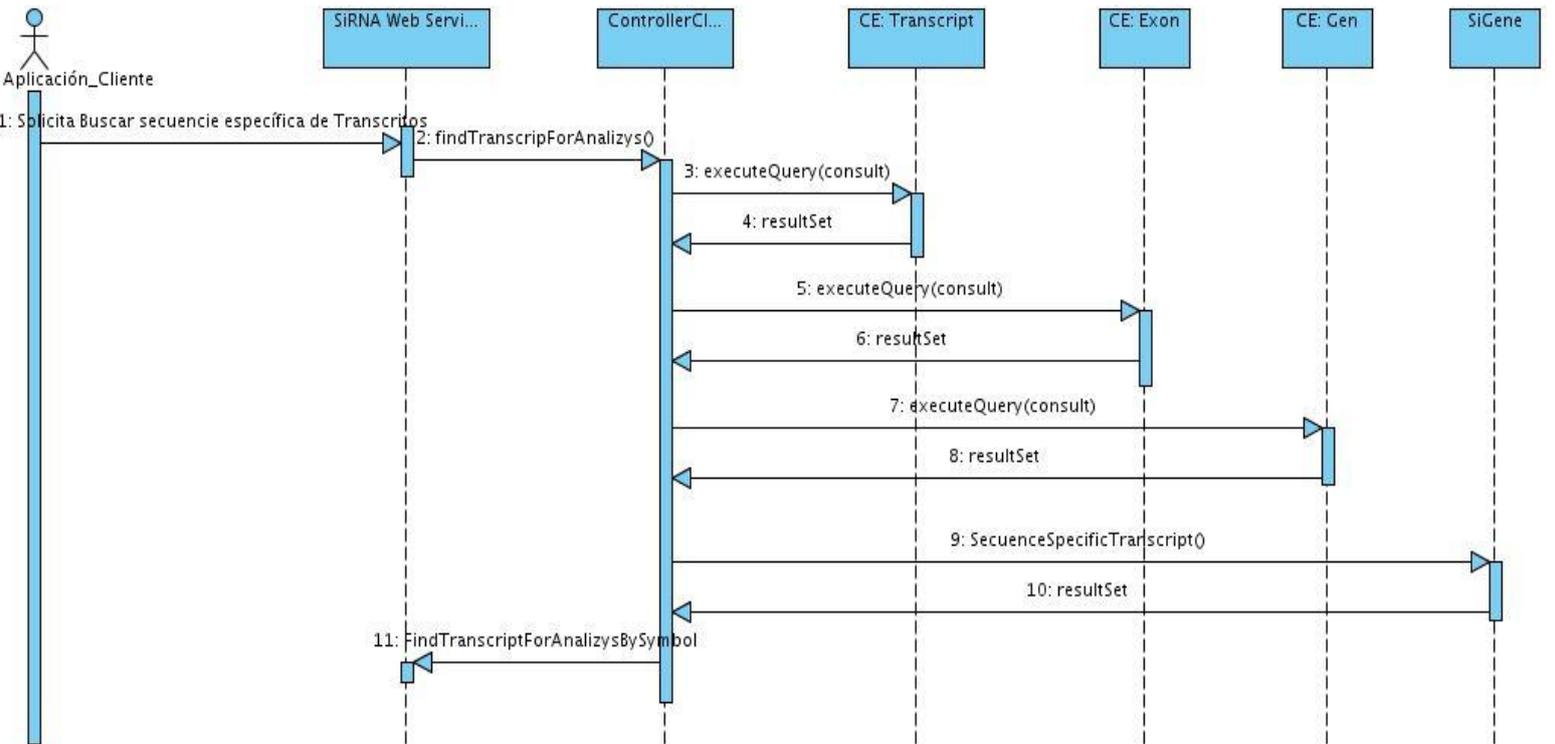
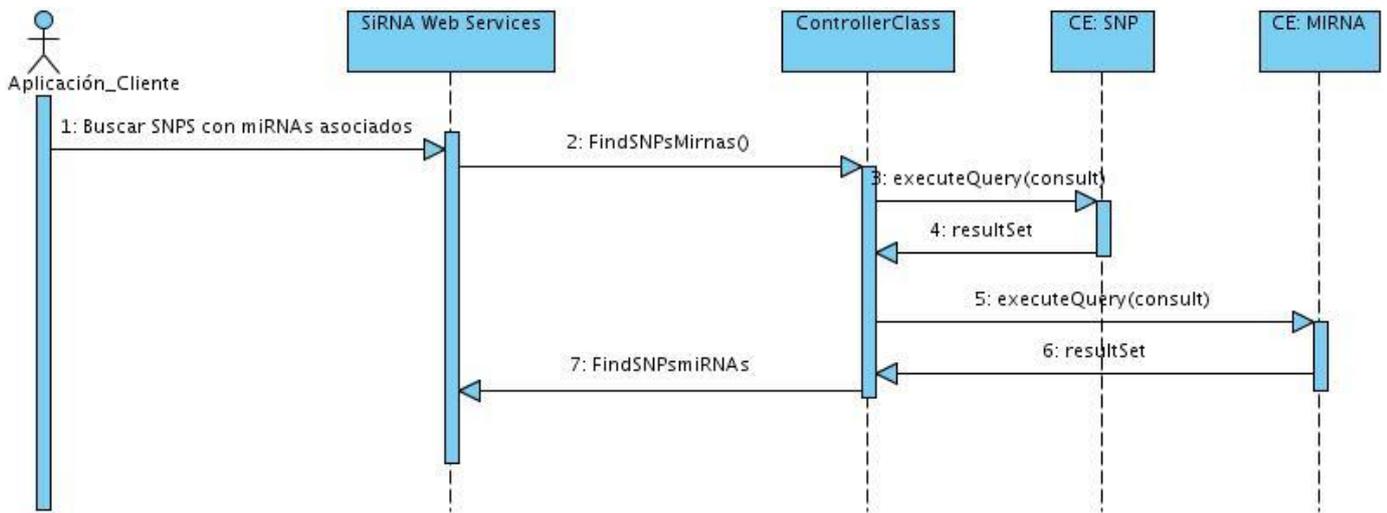


Diagrama de secuencia CU Listar_SNPs_coinciden_miRNA. siRNA web Services



Anexo 2: Pantallas del sistema.

Página principal

alassirna-Design
Analysis System of mRNA sequences for siRNA Design

Home Page | Contact Us | Help

SEARCH

By gene symbol
By gene id
By transcript name

Search

ANALYZE

Paste your own sequence
Common sequence of the transcripts
Specific sequence of a transcript

OTHER SITES

CIGB
NCBI
EBI
Heber Biotec S.A
Bussines Group
OMS
GAVI
EMBnet

Home Page Wednesday, July 3rd, 2009

What does alassirna-Design do?

alassirna-Design tool has been created to find target sites in a mRNA sequence taking into consideration SNPs. You can choose whether the analysis is going to include SNPs or not. You can't input the sequence that is going to be analyzed or select it from our database, and then you can select a set of rules for the design.

This tool shows up a graphic result of the analysis with the target sites, exons, coding and untranslated regions in the sequence.

Mostrar datos de un gen

SEARCH

By gene symbol
 By gene id
 By transcript name

Search

ANALYZE

Paste your own sequence
 Common sequence of the transcripts
 Specific sequence of a transcript

OTHER SITES

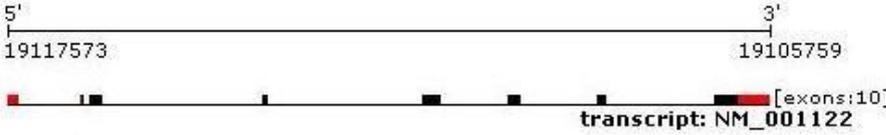
[CIGB](#)
[NCBI](#)
[EBI](#)
[Heber Biotec S.A](#)
[Bussines Group](#)
[OMS](#)
[GAVI](#)
[EMBnet](#)

Gene id search results Monday, May 11th, 2009

Chromosome: 9 Gene id: 123 Symbol: ADFP [-]

5' 3'

19117573 19105759



transcript: NM_001122 [exons:10]

■ - coding region ■ - untranslated region

Choose one transcript to analyze it

NM_001122 ▼

Analyze

miRNAs reported for gene 'ADFP'

Graphic of miRNAs

miRNAs Total: 82

5' 3'

19117573 19105759

Mostrar datos de los miRNAs de un gen

- [NCBI](#)
[EBI](#)
[Heber Biotec S.A](#)
[Bussines Group](#)
[OMS](#)
[GAVI](#)
[EMBnet](#)

miRNAs reported for gene 'ADFP'

Graphic of miRNAs

miRNAs Total: 82

miRNA	PValue	SCORE	start	end
mir-0021	1.8257E-4	16.1686	19106056	19106035
mir-00142	1.92431E-4	16.2501	19105879	19105859
mir-00142	1.92431E-4	16.7351	19105819	19105798
mir-00892	2.05256E-4	15.8452	19106161	19106140
mir-00487	2.43531E-4	15.953	19105909	19105885
mir-00487	2.43531E-4	15.0907	19105906	19105885
mir-00582	3.10643E-4	16.8566	19113511	19113488

Buscar datos de transcritos

SEARCH

By gene symbol
 By gene id
 By transcript name

Search

ANALYZE

Paste your own sequence
Common sequence of the transcripts
Specific sequence of a transcript

OTHER SITES

CIGB
NCBI
EBI
Heber Biotec S.A
Bussines Group
OMS
GAVI
EMBNet

Transcript name search results

Monday, May 11th, 2009

Transcript Name: NM_001085 Gene ID: 12

5' 3'
94148467 94160143

[exons:7]

■ - coding region ■ - untranslated region ■ - snp region

Analyze transcript sequence

```
ATTCATGAAAATCCACTACTCCAGACAGACGGCTTTGGAATCCA  
CCAGCTACATCCAGCTCCCTGAGGCAGAGTTGAGAATGGAGAG  
AATGTTACCTCTCCTGGCTCTGGGGCTCTTGGCGGCTGGGTTCT  
GCCCTGCTGTCCTCTGCCACCCTAACAGCCCACTTGACGAGGA  
GAATCTGACCCAGGAGAACCAAGACCGAGGGACACACGTGGA
```

The algorithm will apply these rules to find target sites:

- It begins with AA
- It contains T in 10th position
- It contains A in 3rd position
- It contains A in 19th position
- There is CG in 30 - 50 %

Buscar secuencia común a un gen determinado

SEARCH

Analyze common sequence

Monday, May 11th, 2009

By gene symbol
 By gene id
 By transcript name

Search gene to analyze common sequence

By gene symbol
 By gene id

ANALYZE

Paste your own sequence
Common sequence of the transcripts
Specific sequence of a transcript

OTHER SITES

[CIGB](#)
[NCBI](#)
[EBI](#)
[Heber Biotec S.A](#)
[Bussines Group](#)
[OMS](#)
[GAVI](#)
[EMBnet](#)

Analyze common sequence

Chromosome: 4 | Gene id: 118 | Symbol: ADD1

```

GCACCCAGGTCGGGCGGTGGGGGCGAGCGGAGGGGCTGAGG
GGCGGAGAGGCCTGGCGGGCCGCTGCTGCGGGCCAGGGGAC
GGGGGCGGAGCCGGAGCCGGAGCCGACGGGCGGTGGCCGCA
CTGGGACCCCGAATCCCGCGCGCTGCCACGATTGCTTCT
GAGGAACCTAGAAAGATTGTACAATGAATGGTATTCTCGTGCT
          
```

The algorithm will apply these rules to find target sites:

- It begins with AA
- It contains T in 10th position
- It contains A in 3rd position
- It contains A in 19th position
- There is CG in - %
- There isn't GC in 19th position

Buscar secuencia específica de un transcrito determinado

SEARCH

Analyze specific sequence

Monday, May 11th, 2009

By gene symbol
 By gene id
 By transcript name

Search transcript to analyze specific sequence

Transcript name:

ANALYZE

Paste your own sequence
Common sequence of the transcripts
Specific sequence of a transcript

OTHER SITES

[CIGB](#)
[NCBI](#)
[EBI](#)
[Heber Biotec S.A](#)
[Bussines Group](#)
[OMS](#)
[GAVI](#)
[EMBnet](#)

Analyze specific sequence

```

ATTCATGAAAATCCACTACTCCAGACAGACGGCTTTGGAATCCA
CCAGCTACATCCAGCTCCCTGAGGCAGAGTTGAGAATGGAGAG
AATGTTACCTCTCCTGGCTCTGGGGCTCTTGGCGGCTGGGTTCT
GCCCTGCTGCTCTGCCACCCTAACAGCCACTTGACGAGGA
GAATCTGACCCAGGAGAACCAAGACCGAGGGACACACGTGGA
          
```

The algorithm will apply these rules to find target sites:

- It begins with AA
- It contains T in 10th position
- It contains A in 3rd position
- It contains A in 19th position
- There is CG in - %
- There isn't GC in 19th position
- It not contains G in 13th position
- There is 3 or more A/T between the bases 15 and 19

Ver detalles del sitio blanco

The screenshot displays the 'alasSiRNA-Design' web application interface. The main page is titled 'Sequence Design' and shows details for Transcript Name: NM_014190 and Gene id: 118. The mRNA length is 3975. A 'Target Sites Total: 9' is indicated. A diagram shows the mRNA sequence with 9 target sites marked. A 'Target Site Information' dialog box is open, showing details for Target Site 7: Number of rules: 6 | Start In: 1715 | End In: 1736. The segment is AATTTACAGGACATTAAGACG with an energy of -26.6. The SNP area is A'A'TTTACAGGACATTAAGACG. The dialog box also shows a 'Close' button. The background page includes a search section with options 'By gene symbol', 'By gene id', and 'By transcript name', and a 'SEARCH' button. There are also 'Home Page', 'Contact Us', and 'Help' links at the top right. The date 'Monday, May 11th, 2009' is shown in the top right corner.

Target Site Information

Target Site Fields

Target Site 7:
Number of rules: 6 | Start In: 1715 | End In: 1736

Segment: AATTTACAGGACATTAAGACG
Energy: -26.6

SNP area: A'A'TTTACAGGACATTAAGACG

Close

GLOSARIO DE TÉRMINOS

Transcrito: Molécula de RNA transcrita a partir de una hebra complementaria de ADN.

Gen: Es la unidad de información hereditaria de los seres vivos y una secuencia lineal de nucleótidos en la molécula de ADN o RNA, que contiene la información necesaria para la síntesis de una macromolécula con función celular específica. El gen es considerado como la unidad de almacenamiento de información y unidad de herencia al transmitir esa información a la descendencia.

Cromosomas: Son las estructuras físicas de la célula eucariota que portan los genes. Estos cromosomas solo son visibles durante la división celular. Desde el punto de vista de su composición los cromosomas están formados de ADN y proteínas.

Exones: Son cada una de las regiones de un gen que contienen la información para producir la proteína codificada en el gen. Cada exón codifica una porción específica de la proteína completa.