

RNAemo

Especificación de Requerimientos de Software

Franco Gaspar Riberi

12 de abril de 2012

Índice

1. Introducción	3
1.1. Propósito	3
1.2. Convenciones del Documento	3
1.3. Audiencia Esperada	3
1.4. Alcance	4
1.5. Descripción general del documento	5
2. Descripción General	6
2.1. Perspectiva del Producto	6
2.1.1. Interfaces del Sistema	7
2.1.2. Interfaces de Usuario	7
2.1.3. Interfaces de Hardware	7
2.1.4. Interfaces de Software	7
2.1.5. Interfaces de Comunicaciones	8
2.1.6. Restricciones de Memoria	8
2.1.7. Operaciones	8
2.1.8. Requerimientos de Instalación	9
2.2. Funciones del Producto	9
2.3. Características de Usuario	9
2.4. Restricciones	10
2.5. Trabajo Futuro	10
3. Requerimientos	10
3.1. Funciones del Sistema	10
3.1.1. Interfaces Externas	10
3.1.2. Requerimientos Funcionales	11
3.2. Restricciones de Rendimiento	13
3.3. Base de Datos	13
3.4. Restricciones de Diseño	14
3.5. Atributos del Software	14
Apendices	15
A. Definiciones, Acrónicos y Abreviaturas	16
B. Manejo de inputs	19
C. Referencias	20

1. Introducción

En esta sección se describe un panorama completo del SRS.

1.1. Propósito

El propósito de este documento es la especificación de requerimientos de software en el marco de la tesis de grado de la carrera Licenciatura en Ciencias de la Computación de la UNRC, “**Estudio de la relación entre divergencia en el uso de codones sinónimos entre virus y huésped y presencia de microRNA**”. Internamente este proyecto será denominado “**RNAemo**”. Los requerimientos del software son provistos por integrantes de **FuDePAN** en su carácter de autores intelectuales de la solución que se pretende implementar y colaboradores de dicha tesis.

Además, este documento establece la primera etapa de dicha tesis y será utilizado como parte de la validación final del proyecto.

1.2. Convenciones del Documento

Las palabras clave DEBE, NO DEBE, REQUERIDO, DEBERÁ, NO DEBERÁ, DEBERÍA, NO DEBERÍA, RECOMENDADO, PUEDE y OPCIONAL en este documento son interpretadas como esta descrito en el documento RFC 2119 [8].

1.3. Audiencia Esperada

A continuación se enumeran las personas involucradas en el desarrollo de la tesis y que por lo tanto, representan la principal audiencia del presente documento.

- *Dr. Roberto Daniel Rabinovich*: Miembro del INBIRS, anteriormente CNRS. Profesor Titular de Virología (Departamento Ciencias Biológicas, CAECE). Colaborador de **FuDePAN**.
- *Lic. Lucía Fazzi*: Licenciada en Genética.
- *Maria Pilar Adamo*: Colaborador de tesis, **FuDePAN**.
- *Daniel Gutson*: Colaborador de tesis, **FuDePAN**.
- *Lic. Guillermo Biset*: Colaborador de tesis, **FuDePAN**.
- *Lic. Laura Tardivo*: Directora de tesis, UNRC.

- *Ac. Franco Gaspar Riberi*: Tesista, UNRC.

1.4. Alcance

El producto que se especifica en este documento se denomina **“Estudio de la relación entre divergencia en el uso de codones sinónimos entre virus y huésped y presencia de microRNA”**, y su principal objetivo es contrastar formalmente una idea encomendada y postulada por el Dr. Roberto Daniel Rabinovich que involucra principalmente la molécula de RNA.

Para la comprensión de la hipótesis se deben tener en cuenta algunos hechos referidos al código genético y los microRNA tales como:

- El código genético está organizado en tripletes o codones.
- El código genético es degenerado¹: existen más tripletes o codones que aminoácidos, de forma que un determinado aminoácido puede estar codificado por más de un triplete. Esos tripletes son conocidos como codones sinónimos.
- En cada especie, se ha seleccionado una proporción de uso de esos codones que guarda relación con la proporción de RNA de transferencia correspondiente de manera de optimizar la síntesis proteica.
- Para algunos patógenos intracelulares como los virus, existe una divergencia entre el uso de codones sinónimos [20] utilizado por el virus y el huésped correspondiente. El origen de esa divergencia no está suficientemente esclarecido.
- Un microARN [18][19] ($_{mi}$ ARN o $_{mi}$ RNA por sus siglas en inglés) es un ARN monocatenario, de una longitud de entre 21 y 25 nucleótidos, y que tiene la capacidad de regular la expresión de otros genes mediante diversos procesos, utilizando para ello la ruta de ribointerferencia². Se encuentran codificados en el genoma y juegan un papel importante en la regulación de la expresión proteica, en la embriogénesis³, procesos cancerosos e infecciones virales. La generación de los microRNA puede variar según el órgano o temporalmente.

¹Implica que al menos uno de los tres nucleótidos (en general el último) puede ser distinto y sin embargo codificar para el mismo aminoácido.

²También conocido como RNA_i por el acrónimo del inglés RNA interference. Corresponde a un mecanismo de silenciamiento post-transcripcional de genes específicos, ejercido por moléculas de ARN que, siendo complementarias a un ARN mensajero, conducen habitualmente a la degradación de éste.

³Proceso que se inicia tras la fertilización de los gametos para dar lugar al embrión.

En este trabajo se estudiará si la divergencia en el uso de codones sinónimos entre virus y huésped contribuye a disminuir la interferencia de los $miRNA$ en la expresión de los RNA_m de origen viral. De esa manera se contribuirá a comprender mejor la relación virus-huésped y la evolución viral.

Dado un conjunto de small- RNA_s ⁴ y una colección de RNA_m de un determinado virus, dado que hay un RNA_m por cada gen, determinar si existen small- RNA_s que se van a hibridar al RNA_m . Luego contabilizar la cantidad de small- RNA_s que se hibridarían tanto a la secuencia original como a la secuencia complementaria. De manera similar, realizar el mismo estudio con el genoma humanizado⁵ contabilizando la cantidad de small- RNA_s que se hibridan.

Que se hibride o no un small- RNA_s a un determinado RNA_m involucra distintas reglas, siendo la más importante la complementariedad de bases. Pero además existen otras, como la presencia de determinadas bases, proporción de uniones GC y motivos específicos.

El proyecto involucrará un determinado virus aún no definido. Para tal fin se aplicarán criterios biológicos.

El sistema a desarrollar comprenderá las siguientes características:

- Abarcar en su totalidad los requerimientos del problema.
- Construir un sistema que puede ser extendido en otros proyectos, brindando un diseño flexible.
- Que proponga un buen uso de las prácticas de diseño para su mejor desempeño.
- Que posea abundante documentación clara y precisa.
- Lograr un código fuente bien escrito y estructurado respetando las buenas prácticas de programación.

1.5. Descripción general del documento

La estructura de este documento sigue las recomendaciones de la “Guía para la especificación de requerimientos de la IEEE” (IEEE Std 830-1998) [1]. El documento está organizado en las siguientes secciones generales:

En la *sección 1*, Introducción, se presenta una primera aproximación al proyecto.

⁴Moléculas de RNA muy pequeñas, dentro de las cuales se encuentra la $miRNA$, $siRNA$, entre otras.

⁵La humanización refiere al reemplazo de nucleótidos en los tripletes que forman la secuencia de RNA. Se acerca a la proporción de uso de codones utilizado en el humano.

En la *sección 2*, Descripción General, se presenta una descripción general de **RNAemo**, sus principales funcionalidades, interfaces y perfiles de usuarios.

En la *sección 3*, Requerimientos, se detallan los requerimientos funcionales específicos de **RNAemo** y los principales atributos que debe cumplir el software.

Por último, se detalla un pequeño apéndice.

2. Descripción General

Esta sección describe los requisitos del producto de modo general. Los requisitos específicos se describen en la sección 3.

2.1. Perspectiva del Producto

La bioinformática es una disciplina dedicada al análisis de elementos biológicos utilizando a la informática como herramienta principal para generar simulaciones, probar teorías, o realizar cálculos complejos entre otros aspectos. En particular, el producto a desarrollar apunta al cálculo complejo de ciertos datos, los cuales son de gran interés para la biología. El mismo, no será un componente de un sistema de mayor envergadura, sino que por el contrario, será totalmente autónomo e independiente. Además, se espera que sea modular permitiendo:

- Obtener resultados intermedios y numéricos.
- Combinable con otros componentes existentes o a ser desarrollados en el futuro, para obtener nuevos resultados.

El sistema se compondrá de los siguiente módulos:

- **Módulo generador de secuencias variando coeficiente de humanización:** dada una secuencia, genera secuencias a distintos coeficientes de humanización.
- **Módulo de matching:** dada una secuencia “larga”, dada una secuencia small-RNA_s, y dado un porcentaje de matching, devuelve todos los sitios donde ocurre el matching con ese porcentaje como mínimo, tanto cadena más como menos. Este módulo podrá considerar la estructura secundaria en una siguiente versión.
- **Módulo maestro (“Master Of Puppets”):** utilizando los módulos anteriores, el presente contabiliza y genera los gráficos leyendo una base de datos de small-RNAs.

2.1.1. Interfaces del Sistema

El producto será capaz de correr al menos en sistemas GNU/Linux, por lo cual sólo se utilizarán librerías compatibles con el mismo.

2.1.2. Interfaces de Usuario

En primera instancia, el usuario interactuará con el sistema mediante una CLI, no se proveerá de interface gráfica de usuario (GUI).

2.1.3. Interfaces de Hardware

El producto de software no requerirá hardware específico alguno para su correcto funcionamiento; cabe destacar que debido al uso de paralelismo, su performance mejorará en arquitecturas multi-core.

2.1.4. Interfaces de Software

El sistema necesitará de la siguiente información inicial, la cual será tomada como parámetros:

- Colección de RNA_m .
- Cantidad de random por RNA_m .
- Discretización del eje X.
- Argumento de % de matching (min,max,step).

Las librerías requeridas para el funcionamiento del sistema son las siguientes:

1. **BioPP:** Librería de C++ para el manejo de estructuras biológicas, código genético, entre otras funciones.
<http://code.google.com/p/biopp/><http://code.google.com/p/biopp/>
2. **FuD:** Framework para la implementación de aplicaciones distribuidas.
<http://code.google.com/p/fud/><http://code.google.com/p/fud/>
3. **MiLi:** Colección de pequeñas bibliotecas C++, compuesta unicamente por headers. Sin necesidad de instalación, sin un makefile, sin complicaciones. Soluciones simples para problemas sencillos.
<http://code.google.com/p/mili/><http://code.google.com/p/mili/>

4. **Rnafolding:** Provee, entre otras, las funcionalidades necesarias para obtener la energía libre de una secuencia de nucleótidos. Casi la totalidad su código proviene del proyecto **VAC-O**.

vac-o.googlecode.com.

<http://code.google.com/p/vac-o/><http://code.google.com/p/vac-o/>

2.1.5. Interfaces de Comunicaciones

No hay requerimientos especificados.

2.1.6. Restricciones de Memoria

Este proyecto no presenta restricciones en cuanto a la cantidad de memoria mínima necesaria para operar. El sistema DEBERÁ manejar la memoria en forma correcta y sin que ocurran memory leaks.

Las dependencias externas que serán utilizadas en el producto no deberán ser tenidas en cuenta en el chequeo de memory leaks u otros problemas.

2.1.7. Operaciones

El modo de operación del sistema será:

1. Invocación por consola por parte del usuario. Esta invocación DEBERÁ tener asociada los siguiente parámetros:
 - Colección de RNA_m .
 - Cantidad de random por RNA_m .
 - Discretización del eje X.
 - Argumento de % de matching (min,max,step).
2. Realización de cálculos internos.
3. En caso de que no se produzcan situaciones erróneas:
 - Exhibición de tablas comparativas de posible hibridación entre microRNA presente en el huésped humano y el RNA presente en la naturaleza y el genoma viral humanizado.
 - Exhibición de un gráfico con resumen estadístico a través del cual se podrá inferir, según la tendencia de las curvas, si la divergencia en el uso de codones entre virus-huésped y la presencia de microRNA pueden estar relacionados.

De lo contrario, el sistema DEBERÁ informar de tal suceso.

2.1.8. Requerimientos de Instalación

Como build system e instalador de código se usará *fudepan-build* [15].

2.2. Funciones del Producto

- DEBERÁ ser capaz que tomar como entrada diferentes parámetros.
- DEBERÁ controlar la validez de los parámetros de entrada.
- DEBERÁ calcular la secuencia complementaria de la secuencia original de RNA_m .
- DEBERÁ humanizar la secuencia original de RNA_m .
- DEBERÁ realizar permutaciones sobre la secuencia de RNA_m conservando la cantidad de cada nucleótido y su tamaño.
- DEBERÁ determinar si existen small-RNA_s que se van a hibridar al RNA_m .
- DEBERÁ contabilizar los small-RNA_s que se hibridan a la secuencia de RNA_m original.
- DEBERÁ contabilizar los small-RNA_s que se hibridan a la secuencia complementaria de RNA_m .
- DEBERÁ contabilizar los small-RNA_s que se hibridan a la secuencia humanizada de RNA_m .
- DEBERÁ realizar cálculos internos.
- DEBERÁ mostrar los resultados al usuario mediante un gráfico.

2.3. Características de Usuario

Se identifican tres tipos de usuarios para el sistema:

1. *Usuario final*: este tipo de usuario refiere a aquellas personas profesionales que utilizarán el producto. Sólo deberán interactuar gráficamente (mediante CLI), cargando los datos de entrada necesarios y ejecutando el programa para luego obtener el resultado.
2. *Usuario de extensión*: refiere a profesionales con conocimientos biológicos capaces de utilizar los posibles resultados para nuevas investigaciones o trabajos.

3. *Usuario desarrollador*: refiere a usuarios con ciertos conocimientos específicos en programación, que serán los encargados de ampliar o extender las funcionalidades del sistema, como así también posibles mejoras algorítmicas.

2.4. Restricciones

El producto DEBERÁ ser desarrollado utilizando el lenguaje de programación C++ [3] y bajo la licencia de software GPLv3 [9]. Además, se DEBERÁ respetar los lineamientos generales impuestos por **FuDePAN** (Thesis Guideline y Coding Guideline).

2.5. Trabajo Futuro

CONSULTAR, no estoy seguro

- A partir del resultado obtenido por el presente sistema, en futuras iteraciones se puede utilizar el mismo para definir moléculas de RNA interferentes. Esto involucra dos situaciones:
 1. Dada una cierta heurística aún no definida, encontrar un target a interferir (pequeña secuencia de nucleótidos), el cual será atacado por el RNA interferente.
 2. Encontrar la secuencia óptima para interferir ese target elegido. Se debe tener en cuenta, que el proceso de folding está relacionado al factor temperatura del huésped. Es decir, diseñar RNA que expresen secuencias sensibles al modificar el factor temperatura.
- Tener en cuenta la estructura secundaria.

3. Requerimientos

En esta sección se detallan específicamente los requerimientos del producto.

3.1. Funciones del Sistema

3.1.1. Interfaces Externas

No hay requerimientos especificados.

3.1.2. Requerimientos Funcionales

- **Nombre del requerimiento:** Cargar una colección de RNA_m. (RF1)
Propósito: Obtener los RNA_m que codifiquen para un mismo virus. Contrala los cuales se determinará si ciertos small-RNA_s se hibridan.
Input: Colección de RNA_m.
Procesamiento:
Output:

- **Nombre del requerimiento:** Cargar una cantidad de random por RNA_m. (RF2)
Propósito: refiere a cuantas secuencias random a generar por cada mensajero para realizar controles. Hacer random, significa realizar permutaciones al azar sobre la secuencia.
Input: int.
Procesamiento:
Output:

- **Nombre del requerimiento:** Cargar el número de discretización del eje X. (RF3)
Propósito: determinar la cantidad de “particiones” a realizar entre 0 y 1, donde:
0: todos los tripletes son los no humanizados (puede haber varias secuencias con 0).
1: todos los tripletes son los preferibles por el genoma humano.
Input: int.
Procesamiento:
Output:

- **Nombre del requerimiento:** Cargar el porcentaje de matching. (RF4)
Propósito: determinar el matching entre small-RNA_s y RNA_m.
Input: %min: int, %max:int, step:int
Procesamiento: Dado un % min, un % max de matching y un step, identificar todos los matching que ocurren por cada step.
Output:

- **Nombre del requerimiento:** Obtener una secuencia complementaria. (RF5)

Propósito: dada una secuencia, llamamemosla cadena original, obtener su cadena complementaria a partir del principio de complementariedad de bases ($A=T, G=C$).

Input: secuencia original de RNA_m .

Procesamiento: Recorrer la cadena original, y reemplazar cada $A \rightarrow T$, $G \rightarrow C$, $T \rightarrow A$ y $C \rightarrow G$.

Output: secuencia complementaria.

- **Nombre del requerimiento:** Obtener una secuencia humanizada (RF6)

Propósito: Obtener una secuencia humanizada para contabilizar los small- RNA_s que se hibridan.

Input: secuencia original de RNA_m .

Procesamiento: Tomar una secuencia de RNA_m , y reemplazar nucleótidos en los tripletes conservando la expresión del aminoácido. Se acerca a la proporción de uso de codones utilizado en el humano.

Output: secuencia humanizada.

- **Nombre del requerimiento:** Determinar si existen small- RNA_s que se hibridan a un RNA_m (RF7)

Propósito: Determinar si existen small- RNA_s que matcheen a cierto % de matching con un RNA_m .

Input: % de matching, RNA_m , small- RNA_s .

Procesamiento:

Output: bool.

- **Nombre del requerimiento:** Contar small- RNA_s en un RNA_m (RF8)

Propósito: Contabilizar los small- RNA_s que se hibridan a la secuencia original de RNA_m .

Input: % de matching, RNA_m , small- RNA_s .

Procesamiento:

Output: int.

- **Nombre del requerimiento:** Contar small- RNA_s en una secuencia complementaria de RNA_m (RF9)

Propósito: Contabilizar los small- RNA_s que se hibridan a la secuencia complementaria de RNA_m .

Input: % de matching, secuencia complementaria de RNA_m , small- RNA_s .

Procesamiento:

Output: int.

- **Nombre del requerimiento:** Contar small- RNA_s en una secuencia humanizada de RNA_m (RF10)

Propósito: Contabilizar los small- RNA_s que se hibridan a la secuencia humanizada de RNA_m .

Input: % de matching, secuencia humanizada de RNA_m , small- RNA_s

Procesamiento:

Output: int.

- **Nombre del requerimiento:** Realizar permutaciones al azar (RF11)

Propósito: permutar una secuencia de RNA_m al azar conservando la misma cantidad de cada nucleótidos y el mismo tamaño de secuencia.

Input: secuencia de RNA_m .

Procesamiento: intercambiar el orden de cada nucleótido de la secuencia de RNA_m .

Output: conjunto de permutaciones de la secuencia de entrada.

- **Nombre del requerimiento:** Exhibición de gráficos con resúmenes estadísticos (RF12)

Propósito: Exhibir el resultado de la ejecución del producto para ciertos parámetros de entrada.

Input: # small- RNA_s que hibridaron en el RNA_m , en el complemento de un RNA_m , y en la humanización de un RNA_m .

Procesamiento:

Output: gráfico.

3.2. Restricciones de Rendimiento

No hay requerimientos especificados.

3.3. Base de Datos

El sistema requerirá de una base de datos de small- RNA_s . En caso alternativo, se permitirá el uso de BLAST para generar secuencias.

3.4. Restricciones de Diseño

El producto DEBERÁ cumplir con los siguientes principios de diseño de la programación orientada a objetos. Los 5 primeros, son también conocidos por el acrónimo “**SOLID**” [2].

- Single responsibility principle (SRP)
- Open/closed principle (OCP)
- Liskov substitution principle (LSP)
- Interface segregation principle (ISP)
- Dependency inversion principle (DIP)
- Law of Demeter (LoD)

3.5. Atributos del Software

El código del producto DEBERÁ:

- Compilar sin advertencias, o las advertencias aceptadas DEBERÁN estar documentadas.
- Cumplir con el estándar ANSI C++ y el “*coding style*” definido por **FuDePAN**.

El software DEBERÁ:

- Funcionar sin memory leaks.
- Tener al menos un 85 % de cobertura con pruebas automatizadas.

Appendices

A. Definiciones, Acrónimos y Abreviaturas

- **RNAemo:** Nombre que recibe el presente producto.
- **UNRC:** Universidad Nacional de Río Cuarto.
- **FuDePAN:** Fundación para el Desarrollo de la Programación en Ácidos Nucleicos [17].
- **INBIRS:** Instituto Biomédico en Retrovirus y SIDA.
- **CNRS:** Centro Nacional de Referencia para el SIDA.
- **SIDA:** acrónimo de síndrome de inmunodeficiencia adquirida. También abreviada como VIH-sida o VIH/sida.
- **CAECE:** Centro de Altos Estudios en Ciencias Exactas.
- **IEEE:** Institute of Electrical and Electronics Engineers.
- **SOLID:** acrónimo nemotécnico introducido por Robert C. Martin en la década de 2000, que representa cinco principios básicos de programación y diseño orientado a objetos
- **GPL:** *General Public License*.
- **SRS:** Especificación de requerimientos.
- **FuD:** FuDePAN Ubiquitous Distribution [14]. Framework para el desarrollo de aplicaciones distribuidas a través de disposiciones heterogéneas y dinámicas de nodos de procesamiento.
- **CLI:** Interfaz de Línea de Comandos, por su acrónimo en inglés de Command Line Interface. Permite dar instrucciones a algún programa informático por medio de una línea de texto.
- **Nucleótido:** molécula orgánica formada por la unión covalente de un monosacárido de cinco carbonos (pentosa), una base nitrogenada y un grupo fosfato.
- **Tripletes:** conjunto de tres nucleótidos que determinan un aminoácido concreto, también conocido como codón.
- **Aminoácido:** molécula orgánica que conforma la proteína.

- **RNA:** Ácido ribonucleico. Es un tipo de ácido nucleico compuesta por nucleótidos esencial para la vida.
- **DNA:** Ácido desoxirribonucleico. Es un tipo de ácido nucleico, forma parte de todas las células.
- **RNA_m:** RNA mensajero. Se encuentra tanto en el núcleo como en el citoplasma celular. Su función es portar el código genético para las proteínas, es decir, transportan las instrucciones de codificación de las proteínas desde el DNA.
- **siRNA:** short interfering RNA. Corresponde a una clase de RNA de cadena doble presente en células eucariotas.
- **miRNA o microRNA:** Corresponde a una clase de RNA de cadena simple presente en células eucariotas.
- **small-RNA_s:** Moléculas muy pequeñas de RNA. Dentro de la clasificación de RNA, aparecen como RNA no codificante.
- **Proteína:** macromolécula formada por cadenas lineales de aminoácidos. Se considera proteína a aquellas cadenas de aminoácidos enlazados cuyo peso molecular es superior 6000 Daltons.
- **Virus:** Entidad biológica que para reproducirse necesita de una célula huésped.
- **$\Delta(G)$ o energía libre:** Es el potencial químico que se minimiza cuando un sistema alcanza el equilibrio a presión y temperatura constante. [13]
- **RNA no codificante:** es aquel RNA que no genera proteínas. Se encuentra el RNA transcripcional y small-RNA_s.
- **Humanización-Deshumanización:** refiere a mutar de forma silente una secuencia. Esto significa, mutar nucleótidos de un triplete conservando la expresión del aminoácido. La diferencia entre humanización y deshumanización radica en que si los tripletes por los que se muta son o no los preferenciales.
- **Mutación silente:** Las mutaciones silentes ocurren cuando se produce un cambio de un sólo nucleótido de DNA dentro de una porción de un gen codificador para una proteína que no afecta la secuencia de aminoácidos que componen la proteína para el gen. Un cambio en un

nucleótido, sin embargo, no siempre cambia el significado de un triplete. El triplete mutado puede aún representar el mismo aminoácido. Y cuando los aminoácidos de una proteína siguen siendo los mismos, esta mantiene su estructura y función.

- **BLAST:** *Basic Local Alignment Search Tool* [16]. Es un programa de alineamiento de secuencias, ya se de DNA, RNA o proteínas. Es capaz de comparar una secuencia problema (denominada query) contra una gran cantidad de secuencias almacenadas en una base de datos. Encuentra las secuencias de la base de datos que tienen mayor parecido a la secuencia query. BLAST es desarrollado por los Institutos Nacionales de Salud del gobierno de Estados Unidos.
- **Codones sinónimos:** término más conocido como “*codon usage bias*”.
COMPLETE
- **Distancia de hamming:** se denomina distancia de Hamming a la efectividad de los códigos de bloque y depende de la diferencia entre una palabra de código válida y otra. Cuanto mayor sea esta diferencia, menor es la posibilidad de que un código válido se transforme en otro código válido por una serie de errores. A esta diferencia se le llama distancia de Hamming, y se define como el número de bits que tienen que cambiarse para transformar una palabra de código válida en otra palabra de código válida.

B. Manejo de inputs

Para la manipulación de los datos se usaran cadenas de caracteres que representan tanto cadenas de DNA como cadenas de RNA para representar genes como nucleóticos.

- **nuc_arn** \rightarrow a | u | c | g | -
- **gen_arn** \rightarrow (**nuc_arn**)⁺
- **nuc_adn** \rightarrow a | t | c | g | -
- **gen_adn** \rightarrow (**nuc_adn**)⁺

Para formar cadenas más complejas, tales como aminoácido y proteínas, se usará:

- **aminoacido** \rightarrow Ala | Arg | Asn | ...
- **proteina** \rightarrow **aminoacido**(**aminoacido**)⁺

C. Referencias

- [1] IEEE Recommended Practice for Software Requirements Specifications. Copyright © 1998 by the Institute of Electrical and Electronics Engineers, Inc. All rights reserved. Published 1998. Printed in the United States of America. ISBN 0-7381-0332-2.
- [2] SOLID: “Design Principles and Design Patterns”, Robert C. Martin. http://www.objectmentor.com/resources/articles/Principles_and_Patterns.pdf
- [3] C++: Lenguaje de programación. <http://www.cplusplus.com>
- [4] G. Biset, D. Gutson, and M. Arroyo, “A framework for small distributed projects and a protein clusterer application”, 2009.
- [5] G. Biset, D. Gutson, and M. Arroyo, “Fud: Design and implementation of a framework for small distributed applications”, 2009.
- [6] B. Meyer, “Object-Oriented Software Construction”, Second Edition, Santa Barbara: Prentice Hall Professional Technical Reference, 1997.
- [7] G. Booch, J. Rumbaugh, and I. Jacobson, “Unified Modeling Language User Guide”, Second Edition, 2005.
- [8] RFC 2119. <http://tools.ietf.org/html/rfc2119>
- [9] GNU General Public License. <http://www.gnu.org/licenses/>
- [10] H. Curtis, N. Sue Barnes, A. Schnek and G. Flores, “Biología”, Editorial Médica Panamericana S.A, 2006, ISBN: 950-06-0423-X.
- [11] B. Pierce, “Genética. Un enfoque conceptual”, Tercera Edición, Editorial médica panamericana S.A, ISBN: 978-84-9835-216-0.
- [12] A. Blanco, “Química Biológica”, Séptima Edición, Editorial El Ateneo.
- [13] $\Delta(G)$: http://en.wikipedia.org/wiki/Gibbs_free_energy
- [14] FuD : <http://code.google.com/p/fud/>

- [15] fudepan-build: <http://fudepan-build.googlecode.com>
- [16] BLAST: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- [17] FuDePAN: <http://www.fudepan.org.ar/>
- [18] Vinay S. Mahajan, Adam Drake and Jianzhu Chen, “Virus-specific host miRNAs: antiviral defenses or promoters of persistent infection?”.
- [19] Man Lung YEUNG, Yamina BENNASSER, Shu Yun LE and Kuan Teh JEANG, “siRNA, miRNA and HIV: promises and challenges”.
- [20] Gareth M. Jenkins and Edward C. Holmes, “The extent of codon usage bias in human RNA viruses and its evolutionary origin”, 2003.
- [21] Comeron JM and Aguadé M. “An evaluation of measures of synonymous codon usage bias”, 1998.
- [22] Haruhiko Siomi and Mikiko C. Siomi, “On the road to reading the RNA-interference code”.