

Escuela de Ingeniería en computación

Arquitectura de computadores

II Semestre

Estudiantes:

Darío Monestel

Ricardo Dávila

Jeffrey Alvarado

Profesor:

Miguel Corea

Investigación

Traductor genético

Fecha de entrega:

...

2017

Índice

Introducción	2
Objetivo de la investigación	3
Flujo de la información genética	4
Síntesis de proteínas o traducción	5
¿Qué es la traducción?	5
ARN de transferencia	7
¿Qué son los aminoácidos?	9
¿Cuál es la finalidad de la síntesis de proteínas?	10
Flujo del programa	11
Conclusión	12
Referencias	13
Anexos	14

Introducción

En la siguiente investigación se expondrá el tema del ADN y su relación con los aminoácidos donde se implementará un traductor genético que tomará un archivo de texto con la cadena de ADN donde procederá a traducirlo primero a ARN y luego a proteínas que son las moléculas de los aminoácidos, luego desde otra implementación habrá otro traductor que hará lo mismo que el anterior pero de forma contraria es decir pasará de aminoácidos a AND.

Objetivo General

El objetivo de esta investigación es profundizar en el tema de biología mediante la programación implementando un programa que se encarga de traducir ADN y aminoácidos, con la finalidad de ser aplicado en el tema de la bioinformática para ofrecer soluciones al desarrollo en general.

Además para aplicar diferentes estructuras en el lenguaje de programación C, necesarias para el desarrollo de programas en otras áreas de la computación.

Objetivos específicos

- Parsear texto de un archivo específico
- Aplicación de un algoritmo de ordenamiento para manejar las cadenas de ARN.
- Aprender sobre el tema de la biología aplicado en la computación.
- Aplicar un diccionario de datos para como estrategia para el manejo de las estructuras de ADN.
- aplicar principios matemáticos para el manejo de las cadenas de ADN y ARN.
- Profundizar en el manejo del lenguaje C.
- Manejo de algoritmos mediante ciclos , condicionales , funciones entre otras.

Flujo de la información genética

La información genética en los seres vivos está contenida en las moléculas de ADN (ácido desoxirribonucleico). El ADN es una macromolécula formada por unidades denominadas nucleótidos, los nucleótidos que forman el ADN sólo pueden ser cuatro: A (adenina), T (timina), C (citosina) o G (guanina). Para que esta información pueda ser utilizada por las células debe transcribirse a una molécula de ARN (ácido ribonucleico). La molécula de ARN se copia fielmente a partir de la molécula de ADN en un proceso llamado transcripción. Existen diferencias químicas entre las moléculas que forman el ADN y el ARN, pero además el código difiere ya que la T del ADN es reemplazada por U (uracilo) en el ARN.

La información genética es traducida por la maquinaria celular para producir las proteínas usando el código genético, el cual determina la secuencia de aminoácidos codificada en el ADN y luego en el ARN. Durante la traducción la maquinaria celular utiliza la molécula de ARN como molde para sintetizar una cadena de aminoácidos codificada en la misma. Para ello interpreta el código leyendo de a 3 nucleótidos, esta unidad se denomina codón, cada codón codifica para un aminoácido.

Nótese que la información contenida en la molécula de ADN y de ARN es la misma, por lo cual se puede obtener la secuencia proteica codificada en una molécula de ADN a partir de su secuencia. Existen 64 codones posibles formados a partir de combinar los 4 nucleótidos del ADN, sin embargo existen sólo 20 aminoácidos. Algunos aminoácidos están codificados en más de un codón, además existen 3 codones denominados stop. Cuando la maquinaria celular lee alguno de estos 3 codones stop, la síntesis de la proteína codificada se detiene. Es decir, los codones stop determinan el final de la proteína y no codifican para ningún aminoácido.

El ADN se copia a ARN en un proceso que se llama "transcripción", podemos pensar en un escritor que transcribe un texto, es decir lo lee y lo copia, pero en este

caso, no se copia a otra molécula de ADN sino a una molécula de ARN. Estas moléculas, luego, van a ser "leídas" por una máquina que se llama ribosoma, en un proceso llamado "traducción". En este caso podemos pensar en un escritor que traduce el mismo texto a otro idioma. Este proceso convierte la información genética contenida en el ADN en información contenida en una secuencia de aminoácidos de una proteína. Para poder obtener la secuencia proteica codificada en una secuencia de ADN conocida debemos utilizar el código genético:

		Segunda base					
		U	C	A	G		
P r i m e r a b a s e	U	Phe UUU	Ser UCU	Tyr UAU	Cys UGU	U	T e r c e r a b a s e
		Phe UUC	Ser UCC	Tyr UAC	Cys UGC	C	
		Leu UUA	Ser UCA	Stop UAA	Stop UGA	A	
		Leu UUG	Ser UCG	Stop UAG	Trp UGG	G	
	C	Leu CUU	Pro CCU	His CAU	Arg CGU	U	
		Leu CUC	Pro CCC	His CAC	Arg CGC	C	
		Leu CUA	Pro CCA	Gln CAA	Arg CGA	A	
		Leu CUG	Pro CCG	Gln CAG	Arg CGG	G	
	A	Ile AUU	Thr ACU	Asn AAU	Ser AGU	U	
		Ile AUC	Thr ACC	Asn AAC	Ser AGC	C	
		Ile AUA	Thr ACA	Lys AAA	Arg AGA	A	
		Met AUG	Thr ACG	Lys AAG	Arg AGG	G	
	G	Val GUU	Ala GCU	Asp GAU	Gly GGU	U	
		Val GUC	Ala GCC	Asp GAC	Gly GGC	C	
		Val GUA	Ala GCA	Glu GAA	Gly GGA	A	
		Val GUG	Ala GCG	Glu GAG	Gly GGG	G	

Síntesis de proteínas o traducción

¿Qué es la traducción?

Qué es lo que se traduce? En realidad se traduce el lenguaje del ADN, que se lee de a bases, tanto en el ADN como en el ARN, a un nuevo lenguaje que es el de los aminoácidos que formarán un polipéptido. Por eso se dice que un gen codifica un polipéptido. Qué es lo que se traduce? En realidad se traduce el lenguaje del ADN, que se lee de a bases, tanto en el ADN como en el ARN, a un nuevo lenguaje que

es el de los aminoácidos que formarán un polipéptido. Por eso se dice que un gen codifica un polipéptido.

El ARNm tanto en procariotas como eucariotas lleva información para la síntesis de polipéptidos y esto se hace gracias a que los ribosomas y los ARNt van leyendo el ARN desde su extremo 5' al 3' donde desde el codón de inicio (AUG) hasta el de stop o paro (UAG,UAA y UGA) de a cada tres nucleótidos o codones, así se van incorporando aminoácidos que le corresponde a cada codón hasta llegar al stop donde finaliza la síntesis de la proteína.

Recordemos lo visto en la página de replicación y transcripción



Estructura general de un gen eucariota. La Región verde es la que se traduce a una proteína (desde el codón de inicio (UAG) hasta el de stop (UGA, UAA o UAG))

La traducción de un lenguaje a otro debe realizarla un verdadero traductor que comprenda ambos lenguajes. Este traductor es el ARN de transferencia (ARNt), ya que puede leer bases en el ARNm, a través de su anticodón que se unen por complementariedad al codón del ARNm, y puede asociarse a un aminoácido gracias a la unión de realizar la aminoacil sintetasa que lo carga con su respectivo aminoácido en su extremo 3'.

La síntesis proteica se desarrolla en el citoplasma celular donde se encuentran los ribosomas, que son partículas que en procariotas están formadas por 3 moléculas de rRNA asociados con alrededor de 52 moléculas proteicas. Estos proporcionan el sistema enzimático necesario para realizar la unión peptídica entre los aminoácidos que van a integrar la proteína, el sitio de unión para el mRNA, y el sitio de anclaje para los tRNA que portan los aminoácidos; estos sitios se denominan A y P. Los tRNA se asocian con las bases del mRNA mediante la interacción codón-anticodón.

Los primeros descubrimientos sobre el código genético mostraron que la información proveniente del ADN (transcrita a un mRNA) se lee de a tripletes de bases o nucleótidos. Cada triplete de bases se corresponde con un aminoácido. A esta información en el ARNm que lleva la información para la síntesis de proteínas estará dada por el Codón de inicio (AUG) y el codón de stop (UGA, UAA o UAG). Ese es el correcto marco de lectura de cada codón.

ARN de transferencia

Como sabemos, para que la síntesis proteica se lleve a cabo es necesario que esté presente una molécula de mRNA que proviene del núcleo, que lleva la información necesaria para decidir qué proteína se va a sintetizar. Este mRNA deberá asociarse a un ribosoma que le brindará el ambiente necesario para la traducción y finalmente será necesario que aparezca el primer aminoácido codificado por el primer codón del mRNA. Lo cierto es que el aminoácido no puede unirse solo sino que necesita una molécula adaptadora que lo porte. Esta molécula es el tRNA que posee en un extremo el aminoácido correcto y en otro extremo tiene un triplete de bases llamado ANTICODON que se aparea por complementariedad de bases con el codón presente en el mRNA.

Como cualquier RNA, está formado por una secuencia de bases (75 a 85 aprox.), que se representa en la clásica forma de hoja de trébol debido al apareamiento de bases que ocurre en cortas secuencias de su cadena (Estructura secundaria).

Cada ARNt está codificado en el ADN y hay uno para cada tipo de aminoácido, y la aminoacil sintetasa lo reconoce por esta estructura tridimensional en forma de L. O sea que hay ARNt de leucinas, de argininas, de serinas etc. La estructura tridimensional también le sirve para poder asociarse a los ribosomas en el proceso de la traducción de proteínas.

En este proceso entonces participan tres tipos de ARN, el de transferencia, el ribosomal que se asocia a proteínas formando los ribosomas y el ARNm que será leído para sintetizar un polipéptido.

La estructura de los ARN mensajeros ya ha sido descrita y sabemos que la traducción comienza en un triplete de bases (AUG) llamado el codón de inicio. Este codón determina entonces como se leerá el ARNm, o sea de a tripletes o codones que le siguen al AUG, por eso se dice que determina el marco de lectura del ARNm.

Este tipo de estructura terciaria le brinda a la molécula la posibilidad de, por un lado, poder aparear su anticodón con el codón del mRNA de una forma poco habitual, es decir que permite que se apareen G con U por ejemplo, además del clásico A-U y C-G; y esta configuración en forma de L le permite tener lo más alejado posible al anticodón del extremo aceptor, esto es necesario porque el anticodón se asocia con el codón sobre la subunidad pequeña del ribosoma y el aminoácido que está en el extremo aceptor se une con otro aminoácido sobre la subunidad grande del ribosoma.

El enlace entre un determinado aminoácido y su correspondiente tRNA es catalizado por la Aminoacil sintetasa, que es específica para cada aminoácido, y por lo tanto para todos los tRNA que lo puedan portar. A los tRNA diferentes que puedan portar el mismo aminoácido, se los llama iso-aceptores.

La enzima produce la reacción de carga del tRNA con su correspondiente aminoácido en dos pasos:

* Aminoácido + ATP ----- Aminoacil – AMP + Pirofosfato

* Aminoacil – AMP + tRNA ----- Aminoacil – tRNA + AMP

Un tRNA cargado con un aminoácido se denomina Aminoacil – tRNA y cuando porta una cadena polipeptídica en crecimiento se denomina Peptidil-tRNA.

Anteriormente se creía que la enzima reconocía al tRNA correspondiente a un determinado Aminoácido, por su anticodón ya que teóricamente es complementario del codón y por lo tanto de la información del DNA, pero luego se comprobó que la enzima reconoce al tRNA por su conformación exclusivamente y es tan específica que una mutación en una base del mismo haría que la enzima ya no pueda reconocerlo.

¿Qué son los aminoácidos?

Un aminoácido es una molécula orgánica con un grupo amino (-NH₂) y un grupo carboxilo (-COOH).¹ Los aminoácidos más frecuentes y de mayor interés son aquellos que forman parte de las proteínas. Dos aminoácidos se combinan en una reacción de condensación entre el grupo amino de uno y el carboxilo del otro, liberándose una molécula de agua y formando un enlace amida que se denomina enlace peptídico; estos dos "residuos" de aminoácido forman un dipéptido. Si se une un tercer aminoácido se forma un tripéptido y así, sucesivamente, hasta formar un polipéptido. Esta reacción tiene lugar de manera natural dentro de las células, en los ribosomas.

Todos los aminoácidos componentes de las proteínas son L-alfa-aminoácidos. Esto significa que el grupo amino está unido al carbono contiguo al grupo carboxilo

(carbono alfa) o, dicho de otro modo, que tanto el carboxilo como el amino están unidos al mismo carbono; además, a este carbono alfa se unen un hidrógeno y una cadena (habitualmente denominada cadena lateral o radical R) de estructura variable, que determina la identidad y las propiedades de cada uno de los diferentes aminoácidos. Existen cientos de radicales por lo que se conocen cientos de aminoácidos diferentes, pero sólo 22 (los dos últimos fueron descubiertos en los años 1986 -selenocisteína- y 2002 -pirrolisina-)² forman parte de las proteínas y tienen codones específicos en el código genético.

La unión de varios aminoácidos da lugar a cadenas llamadas péptidos o polipéptidos, que se denominan proteínas cuando la cadena polipeptídica supera una cierta longitud (entre 50 y 100 residuos aminoácidos, dependiendo de los autores) o la masa molecular total supera las 5000 uma y, especialmente, cuando tienen una estructura tridimensional estable definida.

¿Cuál es la finalidad de la síntesis de proteínas?

La finalidad de la síntesis de proteínas es permitir al organismo formar aquellas macromoléculas que necesita para llevar a cabo sus funciones. Y es que el cuerpo humano no es capaz de utilizar las proteínas ingeridas mediante la alimentación directamente, sino que necesita romper sus enlaces peptídicos y, a partir de los aminoácidos que contienen, crear nuevas estructuras.

Flujo del programa

Empezamos definiendo la estructura de los codón con su aminoácido respectivo luego usamos una función para compararlas para obtener el aminoácido que corresponde, antes de ello se debe validar la entrada para comprobar que sea efectivamente una cadena de ADN captada del archivo a indicar en la consola, finalmente realiza el proceso de traducción donde parsea el texto y toma los codón para traducirlos y retornarlos en consola. Lo anterior es referente al proceso de pasar de ADN a aminoácidos, con respecto al otro programa que pasa de aminoácidos a ADN tenemos una estructura similar al otro programa, donde se maneja un almacenamiento de datos similar a un diccionario, luego se realiza la función comparadora y se parsea el aminoácido donde antes se valida para que efectivamente la entrada sea correcta, no obstante se requiere de una implementación para permutar todas esas posibles cadenas de material genético mediante un par de funciones para dicho fin, donde estas se guardan en un archivo .txt generado por el programa para almacenar todas esas combinaciones de ARN generadas para luego finalmente obtener la salida con la cadena de ADN correspondiente.

Conclusión

En conclusión se aprendió aplicar diferentes funciones en el lenguaje C para parsear texto y aplicar estructuras de colas y pilas para el manejo de datos durante el proceso de traducción, además se abarcó parte del tema de la biología para tener

un dominio del problema a la hora de realizar la estructura de almacenamiento de las

tripletas(codón) que definen las cadenas de ADN para su posterior gestión a la hora de realizar la traducción respectiva.

En definitiva, todas las células del organismo necesitan de la síntesis de proteínas para llevar a cabo sus funciones estructurales y reguladoras. De hecho, si este proceso no tuviese lugar, serían incapaces de funcionar correctamente ni desarrollarse.

Así se espera que el lector comprenda mejor cuál es la finalidad de la síntesis de proteínas y cómo funcionan estas macromoléculas dentro de tu organismo. Sin duda, un proceso fundamental para la vida de todas las especies.

Referencias

Blog gen molecular. (2017). IT Glossary - Software as a Service. Obtenido de Blog gen molecular. [Mayo 20, 2017](#), de Gartner Sitio web:
<https://genmolecular.com/sintesis-de-proteinas-o-traduccion/>

Es.wikibooks.org.(2017).HSNblogDisponible en:
<https://www.hsnstore.com/blog/cual-es-la-finalidad-de-la-sintesis-de-proteinas/>
difusión 2017. Disponible en :
<http://difusion.df.uba.ar/ConectarIgualdad/Tutorial%20CodigoGenetico.pdf>

Biomodel 2017.Articulo encontrado en:
<http://biomodel.uah.es/model4/dna/codons.html>

Anexos

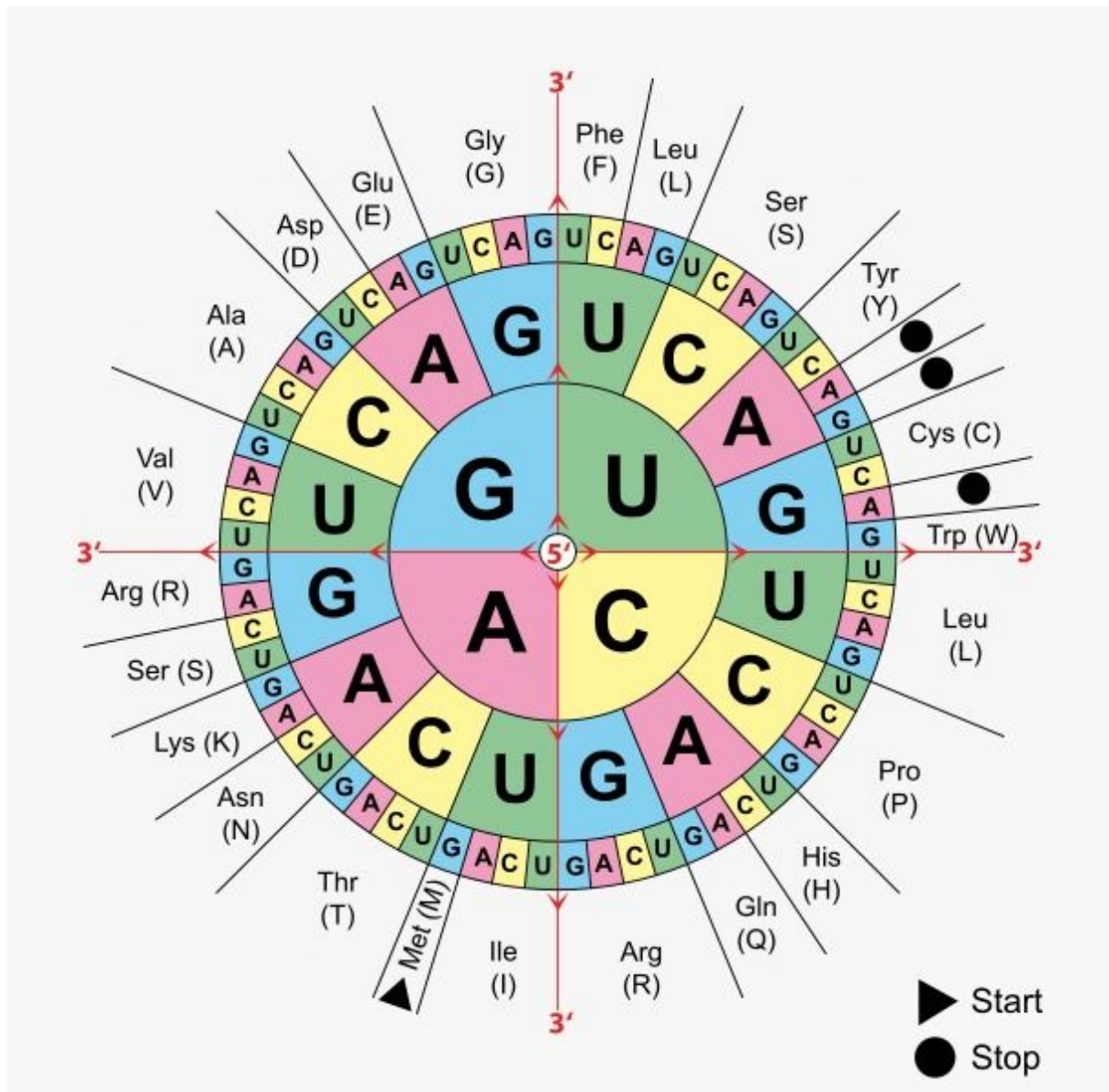


Tabla de aminoácidos:

Nombre	Nombre inglés	Código una letra	Código tres letras
Valina	Valine	V	Val
Leucina	Leucine	L	Leu
Treonina	Threonine	T	Thr
Lisina	Lysine	K	Lys
Triptófano	Tryptophan	W	Trp
Histidina	Histidine	H	His
Fenilalanina	Phenylalanine	F	Phe
Isoleucina	Isoleucine	I	Ile
Arginina	Arginine	R	Arg
Metionina	Methionine	M	Met
Alanina	Alanine	A	Ala
Prolina	Proline	P	Pro
Glicina	Glycine	G	Gly
Serina	Serine	S	Ser
Cisteína	Cysteine	C	Cys
Asparagina	Asparagine	N	Asn
Glutamina	Glutamine	Q	Gln
Tirosina	Tyrosine	Y	Tyr
Ácido Aspártico	Aspartic Acid	D	Asp
Ácido Glutámico	Glutamic Acid	E	Glu

ADN (DNA)	Hebra codificante (Codones)	5' > > > - - - - - T T C - - - - - > > > 3'
	Hebra molde (anticodones)	3' < < < - - - - - A A G - - - - - < < < 5'
ARNm (mRNA)	Mensajero (codones)	5' > > > - - - - - U U C - - - - - > > > 3'
ARNt (tRNA)	Transferente (anticodones)	3' < < < A A G < < < 5'
Proteína	Aminoácidos	amino > > > Fenilalanina > > > carboxilo