

Computación Bio–inspirada

Tema 4: Modelos de computación molecular basados en ADN

David Orellana Martín
Mario de J. Pérez Jiménez

Grupo de Investigación en Computación Natural
Dpto. Ciencias de la Computación e Inteligencia Artificial
Universidad de Sevilla

dorellana@us.es (<http://www.cs.us.es/~dorellana/>)

marper@us.es (<http://www.cs.us.es/~marper/>)

Máster Universitario en Lógica, Computación e Inteligencia Artificial

Curso 2023-2024



Índice

- ♣ Preliminares.
- ♣ Modelos de computación molecular basados en ADN.
- ♣ Modelo **no restringido** de Adleman.
- ♣ Modelo **restringido** de Adleman.
- ♣ Modelo **débil** de Amos.
- ♣ Modelo **sticker** de Roweis.

Preliminares

Un **multiconjunto** \mathbf{m} sobre un **alfabeto** (**conjunto no vacío**) Σ es una aplicación de Σ en el conjunto \mathbf{N} de los números naturales.

- Para cada $x \in \Sigma$, $\mathbf{m}(x)$ es la **multiplicidad** del símbolo x en \mathbf{m} .

Soporte de un multiconjunto \mathbf{m} sobre Σ : $\{x \in \Sigma \mid \mathbf{m}(x) > 0\}$.

Multiconjunto finito (resp. vacío): soporte finito (resp. vacío): .

- Se puede formalizar un **tubo de ensayo** de un laboratorio como un conjunto finito de moléculas (repetidas o no): un **multiconjunto finito** de **moléculas**.

Cardinal de un multiconjunto finito sobre Σ : Suma de las multiplicidades de todos los símbolos de Σ .

Preliminares

Un **agregado** sobre un alfabeto Σ es un multiconjunto finito de elementos de Σ .

- Las moléculas que **no** tienen “orientación” pueden ser representadas por agregados sobre un alfabeto.
- Las moléculas que tienen “orientación” (ej. ADN) pueden ser representadas por cadenas sobre un alfabeto (ej. $\{\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}\}$).

Formalización de un **tubo de ensayo** de un laboratorio que contiene un conjunto finito de moléculas:

- Si las moléculas **no** están “orientadas”: **multiconjunto finito** de **agregados** sobre un alfabeto.
- Si las moléculas **sí** están “orientadas”: **multiconjunto finito** de **cadenas** sobre un alfabeto.

Unión de dos multiconjuntos m_1 y m_2 sobre Σ : Es el multiconjunto $m_1 \cup m_2$ sobre Σ definido por $(m_1 \cup m_2)(x) = m_1(x) + m_2(x)$, para cada $x \in \Sigma$.

- A veces, notaremos el multiconjunto unión por $m_1 + m_2$.

Modelos de computación molecular basados en ADN

Modelos de computación **orientados a programas**.

- * Estructura de datos (**tubos**).
 - Los tubos representan **multiconjuntos finitos** de moléculas.
- * Sintaxis:
 - **Operaciones básicas (moleculares)**.
 - **Programas moleculares** (sucesiones finitas de operaciones moleculares).
- * Semántica:
 - Función que trata de capturar, formalmente, la implementación real de las operaciones moleculares en el laboratorio.

Programas moleculares

Un programa molecular va a usar **tubos** (que formalizan los **tubos de ensayo**):

- ★ A través de unas operaciones moleculares (formalmente definidas).
- ★ Cada programa molecular posee un **tubo** de **entrada** y otro de **salida**.

Verificación formal de un programa molecular (diseñado para resolver un problema abstracto):

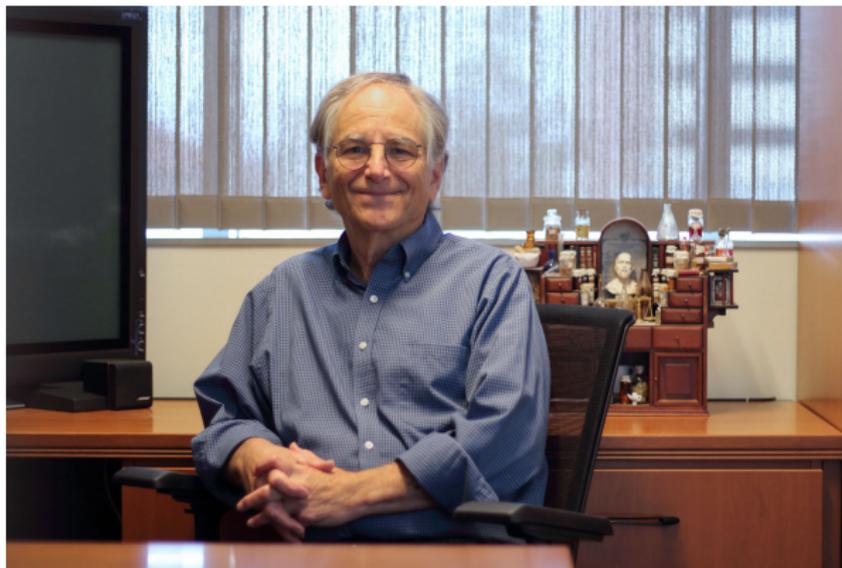
(a) **Corrección** (“*son todos los que están*”).

- Toda molécula del tubo de salida representa una solución correcta del problema.

(b) **Completitud** (“*están todos los que son*”).

- Toda molécula del tubo inicial que representa una solución correcta del problema, debe estar en el tubo de salida.

Modelos restringido y no restringido de Adleman



L.M. Adleman. On constructing a molecular computer. En R.J. Lipton, E.B. Baum, eds. *DNA based computers*, **American Mathematical Society**, 1996, pp. 1-22 (draft, 8 de enero de 1995).

Modelo no restringido de Adleman

Es un modelo de computación:

- * Sustrato computacional: **moléculas de ADN**.
- * Basado en procedimientos de **filtrado**.
- * **Sin memoria de acceso aleatorio**.

Alfabeto de trabajo: $\Sigma_{ADN} = \{\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}\}$.

- Toda cadena sobre el alfabeto Σ_{ADN} se puede identificar con una hebra simple de ADN (con su orientación natural $5' - 3'$).

Definición: *En el modelo no restringido de Adleman, un **tubo** es un multiconjunto de **cadena**s sobre Σ_{ADN} .*

- ★ Colección de moléculas de ADN, eventualmente repetidas.

Instrucciones moleculares básicas modelo no restringido de Adleman

♣ **Extraer** (T, γ) .

Entrada: un tubo T y una cadena γ sobre Σ_{ADN} .

Salida: dos tubos $+(T, \gamma) = \{\sigma \in T \mid \gamma \text{ es una subcadena de } \sigma\}$, y
 $-(T, \gamma) = \{\sigma \in T \mid \gamma \text{ no es una subcadena de } \sigma\}$.

♣ **Mezclar** (T_1, T_2) .

Entrada: dos tubos T_1 y T_2 .

Salida: un tubo $T_1 \cup T_2$ (unión de ambos, como multiconjuntos)

♣ **Amplificar** $(T, \{T_1, T_2\})$.

Entrada: un tubo T .

Salida: dos tubos T_1 y T_2 que son copias exactas de T .

♣ **Detectar** (T) .

Entrada: un tubo T .

Salida: **SÍ**, en el caso $T \neq \emptyset$, y **NO**, en caso contrario.

Es un **modelo** de computación universal (*D. Beaver, 1995*).

Modelo restringido de Adleman

Es un modelo de computación:

- * Sustrato computacional: **moléculas arbitrarias**.
- * Basado en procedimientos de **filtrado**.
- * **Sin memoria de acceso aleatorio**.

Definición: *En el modelo restringido de Adleman, un **tubo** sobre un alfabeto Σ es un multiconjunto finito de **agregados** sobre Σ .*

- * Colección de moléculas (no necesariamente “orientadas”) eventualmente repetidas.

Instrucciones moleculares básicas modelo restringido de Adleman

♣ **Extraer**(T, s).

Entrada: un tubo T y un un *símbolo* $s \in \Sigma$.

Salida: dos tubos $+(\mathbf{T}, \mathbf{s}) = \{\sigma \in T \mid s \in \sigma\}$; $-(\mathbf{T}, \mathbf{s}) = \{\sigma \in T \mid s \notin \sigma\}$.

♣ **Mezclar**(T_1, T_2).

Entrada: dos tubos T_1 y T_2 .

Salida: un tubo $T_1 \cup T_2$ (unión de ambos, como multiconjuntos)

♣ **Detectar**(T).

Entrada: un tubo T .

Salida: **SÍ**, en el caso $T \neq \emptyset$, y **NO**, en caso contrario.

Es un **modelo** de computación **universal** (*D. Beaver, 1995*).

Modelo débil de Amos

Martyn Amos (1971 - ...)



M. Amos. *DNA computation*, PhD thesis, The University of Warwick, 1997.

Modelo débil de Amos

Es un modelo de computación:

- * Sustrato computacional: **moléculas de ADN**.
- * Basado en procedimientos de **filtrado**.
- * **Sin memoria de acceso aleatorio**.

Definición: *En el modelo débil de Amos, un **tubo** es un multiconjunto finito de **cadena**s sobre Σ_{ADN} .*

- ★ Colección de moléculas de ADN, eventualmente repetidas.

Instrucciones moleculares básicas modelo débil de Amos

- ♣ **Quitar**($T, \{\gamma_1, \dots, \gamma_k\}$), con $k \geq 1$.

Entrada: un tubo T y unas cadenas $\gamma_1, \dots, \gamma_k$ sobre Σ_{ADN} .

Salida: un tubo T' obtenido de T eliminando aquellas moléculas que contenga alguna de esas cadenas.

- ♣ **Copiar**($T, \{T_1, \dots, T_k\}$), con $k \geq 1$.

Entrada: un tubo T .

Salida: k tubos T_1, \dots, T_k que son copias exactas de T .

- ♣ **Unión**($\{T_1, \dots, T_k\}, T$), con $k \geq 1$.

Entrada: k tubos T_1, \dots, T_k .

Salida: un tubo T que es la unión de T_1, \dots, T_k como multiconjuntos.

- ♣ **Selección**(T).

Entrada: un tubo T .

Salida: un elemento de T , seleccionado aleatoriamente, en el caso $T \neq \emptyset$, y **NO**, en caso contrario.

Es un **modelo** de computación **universal** (*D. Beaver, 1995*).

Modelo sticker de Roweis

Sam Roweis (abril de 1972- enero de 2010)



S. Roweis, E. Winfree, R. Burgoyne, N.V. Chelyapov., M.F. Goodman, P.W.K. Rothmund, L.M. Adleman. *A sticker-based model for DNA computation*. **Journal of Computational Biology**, 5, 4 (1998), 615-629.

Modelo sticker de Roweis

Es un modelo de computación:

- * Sustrato computacional: **moléculas de ADN**.
- * Basado en procedimientos de **filtrado**.
- * **Con memoria de acceso aleatorio**.

La diferencia con los modelos anteriores radica en la forma de representar la información.

Modelo sticker de Roweis

Representación de la información:

- * **Cadena de memoria** del tipo (n, p, m) , $n \geq p \cdot m$: hebra simple de ADN de longitud n que contiene p subcadenas (**regiones**) de longitud m .

Cadena del tipo (20,4,5)

ATCGG	TCATA	GCACT	AAAAA	
-------	-------	-------	-------	--

- * **Stickers** asociados a una cadena de memoria: cadena simple de longitud m complementaria con una región.

Stickers asociado a la cadena anterior

TAGCC	AGTAT	CGTGA	TTTTT
-------	-------	-------	-------

- * Región **activada**: está complementada por el sticker asociado.
- * Región **desactivada**: no está complementada por el sticker asociado.
- * **Complejo de memoria**, del tipo (n, p, m) $n \geq p \cdot m$: doble hebra formada por una cadena de memoria del tipo (n, p, m) complementada por algunos stickers.

Modelo sticker de Roweis

Codificación binaria de la información:

- * Dos complejos de memoria del tipo (30, 6, 5).



Conveniencia de disponer de una “**frontera**” natural entre las regiones de una cadena: evitar situaciones no deseadas.



Una forma de evitarlo:

- * Regiones impares: sólo **purinas** (A, G); regiones pares: sólo **pirimidinas** (C, T).

Otra forma:

- * La “frontera” está formada por una cadena “especial”.

Modelo sticker de Roweis

Definición: *En el modelo sticker de Roweis, un tubo es un multiconjunto finito de complejos de memoria del mismo tipo.*

- ★ Colección finita de complejos de memoria del mismo tipo, eventualmente repetidos.
- ★ Codifica sucesiones finitas de 0's y 1's.

Modelo sticker de Roweis

Comparación de los mecanismos de representación de la información en el paradigma Adleman y en el modelo sticker:

- * Ambos están basados en la direccionalidad y en la complementariedad de Watson–Crick.
- * En el paradigma Adleman se parte de cadenas simples y cortas que pueden formar doble hebras, con voladizos pero sin huecos.
- * En el modelo sticker se parte de cadenas largas (complejos) y cortas (stickers) formando dobles hebras con voladizos y posibles huecos.
- * Densidad de almacenamiento de información:
 - ★ En el paradigma Adleman: $\frac{1}{20}$.
 - ★ En el modelo sticker: $\frac{1}{m}$ (m es la longitud de las regiones).

Instrucciones moleculares básicas modelo sticker de Roweis

♣ **Extraer** (T, j) .

Entrada: un tubo T de complejos del tipo (n, p, m) y una región j , $1 \leq j \leq p$.

Salida: dos tubos $+(T, j)$ que contiene los complejos de T con la región j -ésima activada, y $-(T, j)$ que contiene los complejos restantes de T .

♣ **Activar** (T, j) .

Entrada: un tubo T de complejos del tipo (n, p, m) y una región j , $1 \leq j \leq p$.

Salida: un tubo T' obtenido de T activando, si procede, la región j -ésima de los complejos.

♣ **Desactivar** (T, j) .

Entrada: un tubo T de complejos del tipo (n, p, m) y una región j , $1 \leq j \leq p$.

Salida: un tubo T' obtenido de T desactivando, si procede, la región j -ésima de los complejos.

♣ **Mezclar** (T_1, T_2) .

Entrada: dos tubos T_1 y T_2 .

Salida: un tubo $T_1 \cup T_2$ (unión de ambos, como multiconjuntos)

♣ **Leer** (T) .

Entrada: un tubo T .

Salida: un elemento de T , seleccionado aleatoriamente, en el caso $T \neq \emptyset$, y **NO**, en caso contrario.

Es un **modelo** de computación universal (*D. Beaver, 1995*).

Modelo sticker de Roweis

Tubo de entrada del modelo sticker:

- * *Biblioteca* de orden (n, p, q, m) , $n \geq p \cdot m$; $1 \leq q \leq p$:
 - * Consta de **todos** los posibles complejos de memoria del tipo (n, p, m) con las últimas $p - q$ regiones desactivadas.
 - * Contiene 2^q complejos de memoria.

Hablaremos, simplemente, de una (p, q) -biblioteca:

- * Una (p, q) -biblioteca codifica **todos** los números binarios con p dígitos tales que son nulos los últimos $p - q$ dígitos.

