Literatur – DNA Computing

T. Head, Formal language theory and DNA: An analysis of the generative capacity of specific recombinant behaviors. *Bull. Math. Biology* **49** (1987) 737–759.

L. M. Adleman, Molecular computation of solutions to combinatorial problems. *Science* **226** (1994) 1021–1024.

T. Head, Gh. Păun and D. Pixton, Language theory and molecular genetics. In: G. Rozenberg and A. Salomaa (eds.), *Handbook of Formal Languages*, Springer-Verlag, 1997, Vol. II, Chapter 7, 295–360.

Gh. Păun, G. Rozenberg and A.Salomaa, *DNA Computing - New Computing Paradigms*. Springer-Verlag, Berlin, 1998.

J. Dassow and Gy. Vaszil, Multiset splicing systems. *BioSystems* **74** (2004) 1–7.

Struktur der DNA I

Struktur der DNA II

Messen der Länge eines DNA Moleküls durch Gel-Elektrophorese

Operationen auf DNA Molekülen – Polymerase

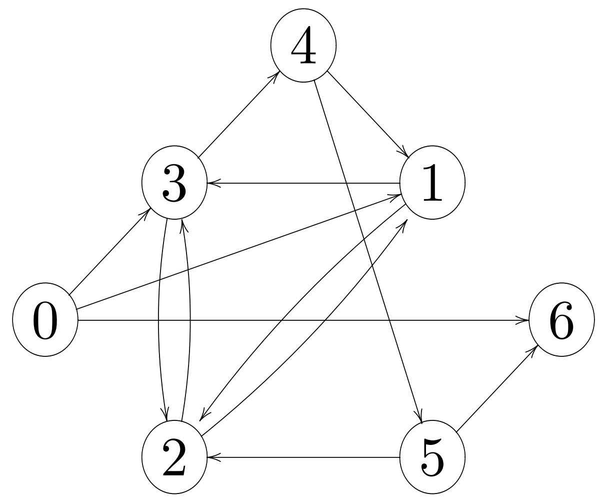
Operationen auf DNA Molekülen – Polymerase-Kettenreaktion

Operationen auf DNA Molekülen – Endonuklease

Operationen auf DNA Molekülen – Verbinden und Ligase

Operationen auf DNA Molekülen – Hybridisierung

Adlemans Experiment



V(2)	TATCGGATCGGTATATCCGA
E(2,3)	CATATAGGCTCGATAAGCTC
V(3)	GCTATTCGAGCTTAAAGCTA
E(3,4)	GAATTTCGATCCGATCCATG

Splicing-Schema und Splicing-Operation I

Definition:
Ein Splicing-Schema ist ein Paar (V, R), wobei
– V ein Alphabet ist und
– R eine Teilmenge von $V^*\#V^*\$V^*\#V^*$ ist.

Die Elemente von R heißen Splicing-Regeln.

Definition:
Wir sagen, dass $w \in V^*$ und $z \in V^*$ aus $u \in V^*$ und $v \in V^*$ durch die Splicing-Regel $r = r_1\#r_2\$r_3\#r_4$ entstehen, und schreiben dafür $(u, v) \vdash_r w$ und $(u, v) \vdash_r z$, wenn die folgenden Bedingungen erfüllt sind:
– $u = u_1r_1r_2u_2$ and $v = v_1r_3r_4v_2$,
– $w = u_1r_1r_4v_2$ and $z = v_1r_3r_2u_2$.

Splicing-Schema und Splicing-Operation II

Für eine Sprache L über V und ein Splicing-Schema (V,R) setzen wir

$$spl(L,R) = \{w \mid (u,v) \vdash_r w,\ u \in L,\ v \in L,\ r \in R\}.$$

Für zwei Sprachfamilien $\mathcal{L}_1$ and $\mathcal{L}_2$ setzen wir

$$\begin{aligned} spl(\mathcal{L}_1,\mathcal{L}_2) \quad = \quad & \{L \mid L = spl(L_1,L_2) \text{ für ein } L_1 \in \mathcal{L}_1 \\ & \qquad \text{und ein Splicing-Schema } (V,R) \text{ mit } R \in \mathcal{L}_2\}. \end{aligned}$$

Splicing-Operation – Beispiele

$L = \{a^n b^n \mid n \geq 0\}$ und $R = \{a\#b\$a\#b\}$
$spl(V, R) = \{a^n b^m \mid n \geq 1, m \geq 1\}$

$L \subset V^*$ beliebig, $L' \subset V^*$ beliebig, $(V \cup \{c\}), R)$, $R = \{\#xc\$c\# \mid x \in L'\}$
$spl(L\{c\}, R) = \{w \mid wz \in L$ für ein $z \in L'\}$

$\{a^n b^n\} \notin spl(\mathcal{L}(REG), \mathcal{L}(RE))$

Die Erzeugungskraft der Splicing-Operation

Satz:
Es gilt die folgende Tabelle, bei der im Schnittpunkt der Zeile X und der Spalte Y ein Z steht, falls $\mathcal{L}(Z) = spl(\mathcal{L}(X), \mathcal{L}(Y))$ gilt, und Z_1/Z_2 steht, falls $\mathcal{L}(Z_1) \subset spl(\mathcal{L}(X), \mathcal{L}(Y)) \subset \mathcal{L}(Z_2)$ ist.

	FIN	REG	CF	CS	RE
FIN	FIN	FIN	FIN	FIN	FIN
REG	REG	REG	REG/CF	REG/RE	REG/RE
CF	CF	CF	RE	RE	RE
CS	RE	RE	RE	RE	RE
RE	RE	RE	RE	RE	RE

Einige Lemmata I

Lemma:
Für Sprachfamilien $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}'_1, \mathcal{L}'_2$ mit $\mathcal{L}_1 \subseteq \mathcal{L}'_1$ und $\mathcal{L}_2 \subseteq \mathcal{L}'_2$ gilt $spl(\mathcal{L}_1, \mathcal{L}_2) \subseteq spl(\mathcal{L}'_1, \mathcal{L}'_2)$.

Lemma:
Falls $\mathcal{L}_1$ unter Konkatenation mit Symbolen abgeschlossen ist, so gilt $\mathcal{L}_1 \subseteq spl(\mathcal{L}_1, \mathcal{L}_2)$ für alle Sprachfamilien $\mathcal{L}_2$.

Lemma:
Falls $\mathcal{L}$ unter Konkatenation, Homomorphismen, inversen Homomorphismen und Durchschnitten mit regulären Mengen abgeschlossen ist, so gilt $spl(\mathcal{L}, \mathcal{L}(REG)) \subseteq \mathcal{L}$.

Einige Lemmata II

Lemma:
Falls $\mathcal{L}$ unter Konkatenation mit Symbolen abgeschlossen ist, so gilt für jede Sprache $L \in \mathcal{L}$, $L \subseteq V^*$ und jedes $c \notin V$ die Beziehung $\{c\}L \in spl(\mathcal{L}(REG), \mathcal{L})$.

Lemma:
Für jede rekursiv-aufzählbare Sprache L gibt es kontextfreie Sprachen L_1 und L_2 derart, dass $L = \{u \mid uv \in L_1 \text{ für ein } v \in L_2\}$ gilt.

Lemma:
Für jede rekursiv-aufzählbare Sprache $L \subset V^*$ gibt es eine kontextabhängige Sprache L' und Buchstaben c_1 und c_2, die nicht in V liegen, derart, dass $L' \subseteq L\{c_1\}\{c_2\}^*$ gilt und es für jedes $w \in L$ eine Zahl $i \geq 1$ derart gibt, dass $wc_1c_2^i \in L'$ ist.

Splicing-Systeme

Definition:
Ein Splicing-System ist ein Tripel $G = (V, R, A)$, wobei
– V ein Alphabet ist,
– R eine Teilmenge von $V^*\#V^*\$V^*\#V^*$ ist, und
– A eine Teilmenge von V^* ist.

Definition:
Die von einem Splicing-System G erzeugte Sprache $L(G)$ ist durch die folgenden Setzungen definiert:
– $spl^0(G) = A$ und $spl^{i+1}(G) = spl(spl^i(G), R)) \cup spl^i(G)$ für $i \geq 0$,
– $L(G) = \cup_{i \geq 0} spl^i(G)$.

Beispiel:
$G = (\{a, b\}, \{a\#b\$a\#b\}, \{(a^n b^n)^m \mid n \geq 1, m \geq 1\})$
$L(G) = \{a^{r_1} b^{s_1} a^{r_2} b^{s_2} \ldots a^{r_m} b^{s_m} \mid m \geq 1,\ r_i \geq 1,\ s_i \geq 1,\ 1 \leq i \leq m\}$

Erweiterte Splicing-Systeme

Definition:
i) Ein erweitertes Splicing-System ist ein Quadrupel $G = (V, T, R, A)$, wobei
– $H = (V, R, A)$ ein Splicing-System ist und
– T eine Teilmenge von V ist.

ii) Die von einem erweiterten Splicing-System G erzeugte Sprache ist durch $L(G) = L(H) \cap T^*$ definiert.

Beispiel:
$G = (\{a, b, c\}, \{a, b\}, \{\#c\$c\#a\}, \{c^m a^n b^n \mid n \geq 1\}$
$L(G) = \{a^n b^n \mid n \geq 1\}$

Definition: Für zwei Sprachfamilien $\mathcal{L}_1$ und $\mathcal{L}_2$ definieren wir $Spl(\mathcal{L}_1, \mathcal{L}_2)$ ($ESpl(\mathcal{L}_1, \mathcal{L}_2)$) als die Menge aller Sprachen $L(G)$, die durch Splicing-Systeme $G = (V, R, A)$ ($G = (V, T, R, A)$) mit $A \in \mathcal{L}_1$ und $R \in \mathcal{L}_2$ erzeugt werden.

Die Erzeugungskraft von Splicing-Systemen

Satz:
Es gilt die folgende Tabelle, wobei im Schnittpunkt der mit X markierten Zeile und der mit Y markierten Spalte ein Z steht, falls $\mathcal{L}(Z) = Spl(\mathcal{L}(X), \mathcal{L}(Y))$ gilt, und Z_1/Z_2 steht, falls $\mathcal{L}(Z_1) \subset Spl(\mathcal{L}(X), \mathcal{L}(Y)) \subset \mathcal{L}(Z_2)$ gilt.

	FIN	REG	CF	CS	RE
FIN	FIN/REG	FIN/RE	FIN/RE	FIN/RE	FIN/RE
REG	REG	REG/RE	REG/RE	REG/RE	REG/RE
CF	CF	CF/RE	CF/RE	CF/RE	CF/RE
CS	CS/RE	CS/RE	CS/RE	CS/RE	CS/RE
RE	RE	RE	RE	RE	RE

Die Erzeugungskraft von erweiterten Splicing-Systemen

Satz:
Es gilt die folgende Tabelle, wobei im Schnittpunkt der mit X markierten Zeile und der mit Y markierten Spalte ein Z steht, falls $\mathcal{L}(Z) = ESpl(\mathcal{L}(X), \mathcal{L}(Y))$ gilt.

	FIN	REG	CF	CS	RE
FIN	REG	RE	RE	RE	RE
REG	REG	RE	RE	RE	RE
CF	CF	RE	RE	RE	RE
CS	RE	RE	RE	RE	RE
RE	RE	RE	RE	RE	RE

Multimengen

Multimenge M over V – Abbildung von V^* in $\mathbf{N}$

$M(x)$ – Multiplizität von x

Die Multimenge M heißt genau dann endlich wenn es eine endliche Menge $U \subset V^*$ derart gibt, dass $M(x) = 0$ für alle $x \notin U$

Mächtigkeit – $\#(M) = \sum_{x \in V^*} M(x)$
Länge – $l(M) = \sum_{x \in V^*} M(x)|x|$

eine endliche Multimenge M ist als $[w_1, w_2, \ldots, w_n]$ darstellbar

$l(M) = |w_1 w_2 \ldots w_n|$
$\#_a(M) = \#_a(w_1 w_2 \ldots w_n)$ für $a \in V$

Multimengen-Splicing I

Für zwei Wörter w und v über V ist die Splicing-Regel $p = u_1\#u_2\$u_3\#u_4$ genau dann anwendbar auf w und v wenn $w = w_1u_1u_2w_2$ and $v = v_1u_3u_4v_2$ gelten, und als Resultat erhalten wir die Wörter $w' = w_1u_1u_4v_2$ und $v' = v_1u_3u_2w_2$. Wir schreiben dafür $[w, v] \Longrightarrow_p [w', v']$.

Für zwei Multimengen $M = [w_1, w_2, \ldots, w_n]$ und $M' = [v_1, v_2, \ldots, v_n]$ von Wörtern über V und eine Menge P von Splicing-Regeln über V definieren wir

- einen <u>sequentiellen Ableitungschritt</u> $M \Longrightarrow_s M'$ durch $[w_1, w_2] \Longrightarrow_p [v_1, v_2]$ und $w_i = v_i$ für $3 \leq j \leq n$ für ein $p \in P$ und eine passende Anordnung der Elemente in M und M',

Multimenge-Splicing II

- einen <u>maximal parallelen Ableitungsschritt</u> $M \Longrightarrow_{mp} M'$ durch $[w_{2i-1}, w_{2i}] \Longrightarrow_{p_i} [v_{2i-1}, v_{2i}]$ für $1 \leq i \leq k \leq \frac{n}{2}$ und $w_i = v_i$ für $2k+1 \leq j \leq n$ für gewisse $p_i \in P$ und eine passende Anordnung der Elemente in M und M', und durch die Forderung, dass es keine Multimenge $[w, w'] \subseteq [w_{2k+1}, w_{2k+2}, \ldots, w_n]$ gibt, auf die eine Splicing-Regel $p \in P$ (erfolgreich) angewendet werden kann,

- einen <u>streng maximal parallelen Ableitungsschritt</u> $M \Longrightarrow_{smp} M'$ durch $M \Longrightarrow_{mp} M'$ für ein k (wie im vorhergehenden Punkt) und es gibt kein M'' mit $M \Longrightarrow_{mp} M''$ für ein $k' > k$.

Multimengen-Splicing-Systeme

Ein Multimengen-Splicing-System ist ein Tripel $G = (V, P, M)$, wobei
– V ein Alphabet ist,
– P eine endliche Menge von Splicing-Regeln über V ist und
– M eine endliche Multimenge über V ist.

Wir definieren die von G erzeugte sequentielle, maximal parallele und streng maximal parallele Multimengen-Sprache $mL(G, s)$ bzw. $mL(G, mp)$ bzw $mL(G, smp)$ durch

$$\begin{aligned} mL(G, s) &= \{K \mid M \Longrightarrow_s^* K\}, \\ mL(G, mp) &= \{K \mid M \Longrightarrow_{mp}^* K\}, \\ mL(G, smp) &= \{K \mid M \Longrightarrow_{smp}^* K\}. \end{aligned}$$

Bezeichnung einiger Sprachfamilien

für $Y \in \{s, mp, smp\}$

$m\mathcal{L}(Y)$	–	Familie aller Sprachen $mL(G,Y)$, die durch ein Multimengen-Splicing-System G im Ableitungsmodus Y erzeugt werden können
$m\mathcal{L}_n(Y)$	–	Familie aller Sprachen $mL(G,Y)$, die durch ein Multimengen-Splicing-System $G = (V,P,M)$ mit $\#(M) = n$ im Ableitungsmodus Y erzeugt werden können

Einige Fakten I

Lemma:

Für jedes Multimengen-Splicing-System $G = (V, P, M)$, jedes $a \in V$, jeden Modus $Y \in \{s, mp, smp\}$ und jede Multimenge $K \in mL(G, Y)$ gilt

$$\#(K) = \#(M),\ l(K) = l(M) \text{ und } \#_a(K) = \#_a(M).$$

Satz:

Es seien n und m zwei verschiedene Zahlen und $Y \in \{s, mp, smp\}$ und $Y' \in \{s, mp, smp\}$ zwei Ableitungsmodi.
Dann sind $m\mathcal{L}_n(Y)$ und $m\mathcal{L}_m(Y')$ unvergleichbar.

Einige Fakten II

Satz:

i) Für $n \in \{1, 2, 3\}$ gilt $m\mathcal{L}_n(s) = m\mathcal{L}_n(mp) = m\mathcal{L}_n(smp)$.

ii) Für $n \geq 4$ sind $m\mathcal{L}_n(mp)$ und $m\mathcal{L}_n(smp)$ unvergleichbar mit $m\mathcal{L}_n(s)$.

iii) Für $n \geq 5$ ist $m\mathcal{L}_n(mp)$ nicht in $m\mathcal{L}_n(smp)$ enthalten.

iv) Für $n \geq 6$ sind die Familien $m\mathcal{L}_n(s)$, $m\mathcal{L}_n(mp)$ und $m\mathcal{L}_n(smp)$ paarweise unvergleichbar.