

Bioinformatik von RNA- und Proteinstrukturen

Inhaltsverzeichnis

1	Formale Sprachen	1
1.1	formale Grammatik G	1
1.2	Klassifikation von formalen Sprachen	2
1.3	Hidden Markov Model	2
2	Einleitung	3
3	Strukturvorhersage	4
3.1	Nussinov	4
3.2	Turner-Modell (Nearest-Neighbor-Modell)	4
3.3	Zuker-Algorithmus	4
3.3.1	suboptimales Falten	4
3.4	Wuchty-Algorithmus	4
3.4.1	Wuchty-Backtracking	4
3.5	McCaskill	4
3.6	stochastisches Backtracking	4
3.7	Strukturvorhersagen verbessern	4
3.8	Konsensusstrukturvorhersagen	4
3.9	Wie kann RNA evolvieren?	4
3.9.1	Neutrale Netzwerke	4
3.9.2	SHAPE-Abstraktion	4
3.9.3	Energielandschaften	4
3.9.4	Faltungskinetik	4
3.9.5	Barriers Trees	4
3.9.6	Flooding-Algorithmus	4
3.9.7	Co-transcriptional folding	4

1 Formale Sprachen

Formale Sprache¹ L über Alphabet Σ

$L \subseteq \Sigma^*$

mit Σ^* = Kleensche Hülle² von Σ

$$\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n$$

$\Sigma^0 = \{\varepsilon\}, \Sigma^1 = \Sigma, \Sigma^2 = \Sigma \times \Sigma$

$\varepsilon \rightarrow$ leeres Wort (leere Menge)

Beispiel: $\Sigma = \{a\}, \Sigma^* = \{\varepsilon, a, aa, aaa, \dots\}, L = \{a, aa, aaaa, \dots\}$

1.1 formale Grammatik G

$G = (N, \Sigma, P, S)$ mit

- N = Nichtterminale
- Σ = Alphabet
- P = Produktionsregeln
- S = Startsymbol ($\in N$)

$P \subseteq (N \cup \Sigma)^* / N(N \cup \Sigma)^* \rightarrow (N \cup \Sigma)^*$

Beispiel:

$G = (\{S\}, \{a\}, \{S \rightarrow aaS, S \rightarrow a\}, S)$

führt zu: $S \rightarrow aaS \rightarrow aaa$

¹https://de.wikipedia.org/wiki/Formale_Sprache

²https://de.wikipedia.org/wiki/Kleenesche_und_positive_H%C3%BClle

1.2 Klassifikation von formalen Sprachen

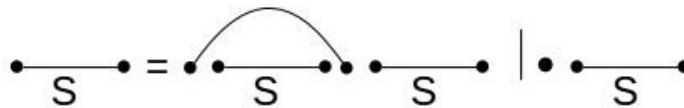
durch die Comsky-Hierarchie³:

- Typ 0 = rekursiv auszählbar ($\alpha N \beta \rightarrow \gamma$)
- Typ 1 = kontext-sensitiv ($\alpha N \beta \rightarrow \alpha \gamma \beta$)
- Typ 2 = kontext-frei, $N \rightarrow (N \cup \Sigma)^* \rightarrow$ stochistisch kontextfreie Grammatik (SCFG) \rightarrow Dynamics Programming
- Typ 3 = regular ($N \rightarrow \Sigma | \Sigma N$) \rightarrow dann immer Hidden Markov Model (HMM) modellierbar

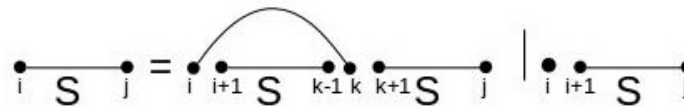
bei Alignments: $\boxed{S} \rightarrow \boxed{S} \begin{smallmatrix} \vdots \\ \vdots \end{smallmatrix} \mid \boxed{S} \begin{smallmatrix} \cdot \\ \cdot \end{smallmatrix} \mid \boxed{S} \begin{smallmatrix} \bullet \\ \bullet \end{smallmatrix} \mid \varepsilon$

Erweiterung mit Wahrscheinlichkeit: $G=(N, \Sigma, P, S, \Omega)$
mit Ω = Wahrscheinlichkeit für Produktionsregeln

jetzt auf RNA-Vorhersagen:



scoring scheme: Bewertung von $\sigma (\frown) = 1$, $(\sigma (\dashrightarrow))$, $\sigma (\bullet) = 0$
scoring function: max Basepairs: + (Summe), Anzahl der Strukturen: \cdot (Multiplikation) choice function: max Basepairs: max, Anzahl der Strukturen: + (Summe)



1.3 Hidden Markov Model

M: Match, I: Insertion, D: Deletion

Grammatik:

³<https://de.wikipedia.org/wiki/Chomsky-Hierarchie>

- $M \rightarrow M_{A_A} | \dots | I | D$
- $I \rightarrow I_{A_} | \dots | D | M$
- $D \rightarrow D_{_A} | \dots | M | I$

Beispiel:

Faltungsgrammatik

$$S \rightarrow (S)S|.S|\varepsilon$$

Nichtterminale = S, Alphabet = $\{ (,), . \}$

Beispiel in Baumdarstellung:

weiteres Beispiel: Sankoff, Kombination von zwei Grammatiken (Alignment und Faltung)

Alignmentgrammatik

$$S \rightarrow .S|_S|\varepsilon$$

$$G = (N = \{S\}, \Sigma = \{., _\}, P = \{S \rightarrow .S|_S|\varepsilon\}, S)$$

$$\text{Alignment: } G^2 = G \times G = (N \times N, \Sigma \times \Sigma, P^2, (S, S))$$

$$P^2 = P \times P = \begin{pmatrix} S \\ S \end{pmatrix}$$

2 Einleitung

Struktur: Form \rightarrow Funktion

Funktion folgt Form, Form folgt Sequenz

Proteine, RNA, DNA: Sequenzen

4 Strukturlevels:

- primäre Struktur (Sequenz): 1 Dimension
- sekundäre Struktur (grobe Annäherung an Struktur): 2 Dimensionen
- tertiäre Struktur (räumliche Struktur): 3 Dimensionen
- quatiäre Struktur (räumliche Anordnung von interagierenden Strukturen): 4 Dimensionen

Behandlung hauptsächlich 2D

3 Strukturvorhersage

3.1 Nussinov

3.2 Turner-Modell (Nearest-Neighbor-Modell)

3.3 Zuker-Algorithmus

3.3.1 suboptimales Falten

3.4 Wuchty-Algorithmus

3.4.1 Wuchty-Backtracking

3.5 McCaskill

3.6 stochastisches Backtracking

3.7 Strukturvorhersagen verbessern

3.8 Konsensusstrukturvorhersagen

3.9 Wie kann RNA evolvieren?

3.9.1 Neutrale Netzwerke

3.9.2 SHAPE-Abstraktion

3.9.3 Energielandschaften

3.9.4 Faltungskinetik

3.9.5 Barriers Trees

3.9.6 Flooding-Algorithmus

3.9.7 Co-transcriptional folding