PROBLEM OUTLINE

DANIELLE DEES, LAURA SLAYBAUGH

1. D

oes any string s in S* have a secondary structure?

Claim: If some string s in S* has a secondary structure, then some string s' in S* will have a restricted secondary structure.

L is defined as the language generated by the following context-free grammar.

$$
\begin{aligned}
W &\rightarrow tW|V_0 \\
V_0 &\rightarrow tV_1\bar{t}|tI_0\bar{t}|tM_0\bar{t} \\
V_1 &\rightarrow tV_2\bar{t}|H_1|tI_1\bar{t}|tM_1\bar{t} \\
V_2 &\rightarrow tV_3\bar{t}|H_2|tI_2\bar{t}|tM_2\bar{t} \\
V_3 &\rightarrow H_3|tI_3\bar{t}|tM_3\bar{t} \\
H_3 &\rightarrow \text{all hairpins closed with one pair with energy 3} \\
H_2 &\rightarrow \text{all hairpins closed with one pair with energy 2} \\
H_1 &\rightarrow \text{all hairpins closed with one pair with energy 1} \\
I_1 &\rightarrow tV_0|V_0t \\
I_2 &\rightarrow tV_1|V_1t|tI_1|I_1t \\
I_3 &\rightarrow tV_2|V_2t|tI_2|I_2t \\
M_0 &\rightarrow WM_0WM_0|WM_1WM_{-1}|WM_2WM_{-2}|WM_3WM_{-3} \\
M_1 &\rightarrow WM_{-2}WM_3|WM_{-1}WM_2|WM_0WM_1|WM_1WM_0|WM_2WM_{-1}|WM_3WM_{-2} \\
M_2 &\rightarrow WM_{-1}WM_3|WM_0WM_2|WM_1WM_1|WM_2WM_0|WM_3WM_{-1} \\
M_3 &\rightarrow WM_0WM_3|WM_1WM_2|WM_2WM_1|WM_3WM_0 \\
WM_{-3} &\rightarrow tWM_{-2}|V_0WM_{-3}|V_1WM_{-2}|V_2WM_{-1}|V_3WM_0 \\
WM_{-2} &\rightarrow tWM_{-1}|V_{-1}WM_{-3}|V_0WM_{-2}|V_1WM_{-1}|V_2WM_0|V_3WM_1 \\
WM_{-1} &\rightarrow tWM_0|V_{-2}WM_{-3}|V_{-1}WM_{-2}|V_0WM_{-1}|V_1WM_0|V_2WM_1|V_3WM_2 \\
WM_0 &\rightarrow tWM_1|V_{-3}WM_{-3}|V_{-2}WM_{-2}|V_{-1}WM_{-1}|V_0WM_0|V_1WM_1|V_2WM_2|V_3WM_3 \\
WM_1 &\rightarrow tWM_2|V_{-3}WM_{-2}|V_{-2}WM_{-1}|V_{-1}WM_0|V_0WM_1|V_1WM_2|V_2WM_3 \\
WM_2 &\rightarrow tWM_3... \\
WM_3 &\rightarrow
\end{aligned}
$$

This grammar should generate all strands which have a secondary structure with small loops, where the internal structures have energy between zero and a constant. Since this grammar is context-free, the language L is context free.

Let L' be the set of strands in S* that have secondary structure. L is a subset of L'. Is L' an empty set?

Claim: L' is empty if and only if L interected with S* is empty. Claim: L is context free because we have the grammar. Corrolary: L interected with S* is context free by formal language theory.

Theorem: The problem of deciding if a contex free language, C, is empty is decidable. If C is not empty then there is a short string in C where short depends on the size of the grammar for C. This short string could be tested to determine all the properties of C. Therefore, in this instance an algorithm must exist that will determine if L intersected with S* is empty. If this set is not empty, then we only have a set of short strings to deal with. However, this set of short strings can be rather large. In our particular instance of word design, these short strands are combinatorial permutations of words from an original set. We can test all the combinations of length “short” in polynomial time. Therefore, an efficient algorithm exists which will determine whether any string in S* will fold based on the fact that if no string in S^3 (or some other “small” number) folds, then there exists not string in S* which will have a secondary structure.