

1. ÜBUNG "BIOINFORMATIK", SS 17

Aufgabe 1: (1 Credits)

Titin is a muscle protein whose gene has the largest known coding sequence - 80,781 DNA bases. How many amino acids long is titin?

Aufgabe 2: (9 Credits)

Recall the *Nirenberg-Matthaei-Experiment*: We introduced the technique of transcribing synthetic mRNA in order to solve some of the genetic code. The synthetic mRNA was periodic in nature: $XXXX \dots$, $XXYXXY \dots$, $XYXXYY \dots$, etc. Derive all the information you can about the genetic code using only two letters A and C . Clearly define the synthetic mRNA and their protein products. Recall that only the presence of amino acids could be detected, not the sequence.

Aufgabe 3: (1+2+1 = 4 Credits)

Consider a gene as a subsequence of the DNA that encodes one protein and let S be a the protein (sequence of aminoacids) *CRICK*.

- (a) Which aminoacids are encoded?
- (b) How many different genes g can theoretically code for this sequence S ?
- (c) Write down one possible gene g encoding S .

Aufgabe 4: (6 Credits)

Let us consider a protein simply as a sequence of aminoacids. Consider the set R of DNA sequences of length $3n$ with $n \in \mathbb{N}$. Let $R' \subseteq R$ be the set of sequences $r \in R$ that can theoretically code for proteins. In particular, assume that each sequence $r = r_1 r_2 \dots r_{3n} \in R'$ *begins* with the startcodon coding for **Met**, *ends* with a stopcodon and none of the codons $r_i r_{i+1} r_{i+2}$ with $i \bmod 3 = 1$ and $3 < i < 3n - 2$ corresponds to a start- or stopcodon.

Determine the cardinality $|R'|$ for $n = 1$, $n = 2$ and $n > 3$.

Deadline: Tuesday - April 11, 2017 - 8.15pm