

# Algorithmen der Sequenzanalyse: verrauschte Motive

AlgSeq – 28/10/2024

Prof. M. Sammeth

# AlgSeq - Übungen (4)

## (1) Inexakte DnaA Boxen

Wie in den vorigen Übungen gesehen, erreicht das GC-Ungleichgewichtsdiagramm von *E. coli* ein Minimum an Position **3.923.620**, mit einem Algorithmus für das “Häufige Wörter mit RC” Problem konnten jedoch keine relevanten Klumpen gefunden werden.

Bestimmen Sie mit einem Algorithmus für das **Häufige Wörter mit RC und Unmittelbaren Nachbarn Problem** das/die häufigsten  $k$ -mere in einem Fenster von 500 nt, das an der Position **3.923.620** beginnt und vergleichen Sie Ihr Ergebnis zu der experimentell bestimmten DnaA Box in *E. coli*: **TTATCCACA**.

**Datei:** `genom_Escherichia_coli.fasta` (s. Übungen zu Algo03)

## (2) Implantierte Motive

Versuchen Sie auf irgendeine Weise das Subtiles Motiv Problem zu lösen, die entsprechenden 10 Sequenzen für Dna sind in der beiliegenden Datei vorgegeben.

**Datei:** `Algo04_subtiles_motiv.txt`