

# FASTA Datenstrukturen

- FASTA ist ein Datenformat in der Bio-Informatik
- Dient zur Speicherung der Struktur von Nukleinsäuren und Proteinen, bsp. RNA-Strukturen
- Bedient sich eines einfachen Textformats

# FASTA Datenstrukturen

```
> Beispiel FASTAFormat | Typ 1  
; Dies ist ein Kommentar  
; Es sind nur bestimmte Buchstaben und  
; Buchstabenketten erlaubt  
CATGCTCAACCCGATTTGAAC  
GAAGTAAAGTCTCTGAGCATT  
TACTCAAACGGTATGAATAA  
AACGATACTTTGCTCAGAGTA  
AGTTCGAATTC
```

# String-Alignment bei FASTA

- Die einfache Textstruktur ermöglicht String-Alignment-Algorithmen den Vergleich von FASTA Daten
- Bsp. Vergleich von im FASTA-Format gespeicherten RNA-Strukturen

# Needleman-Wunsch-Algorithmus (NWA)

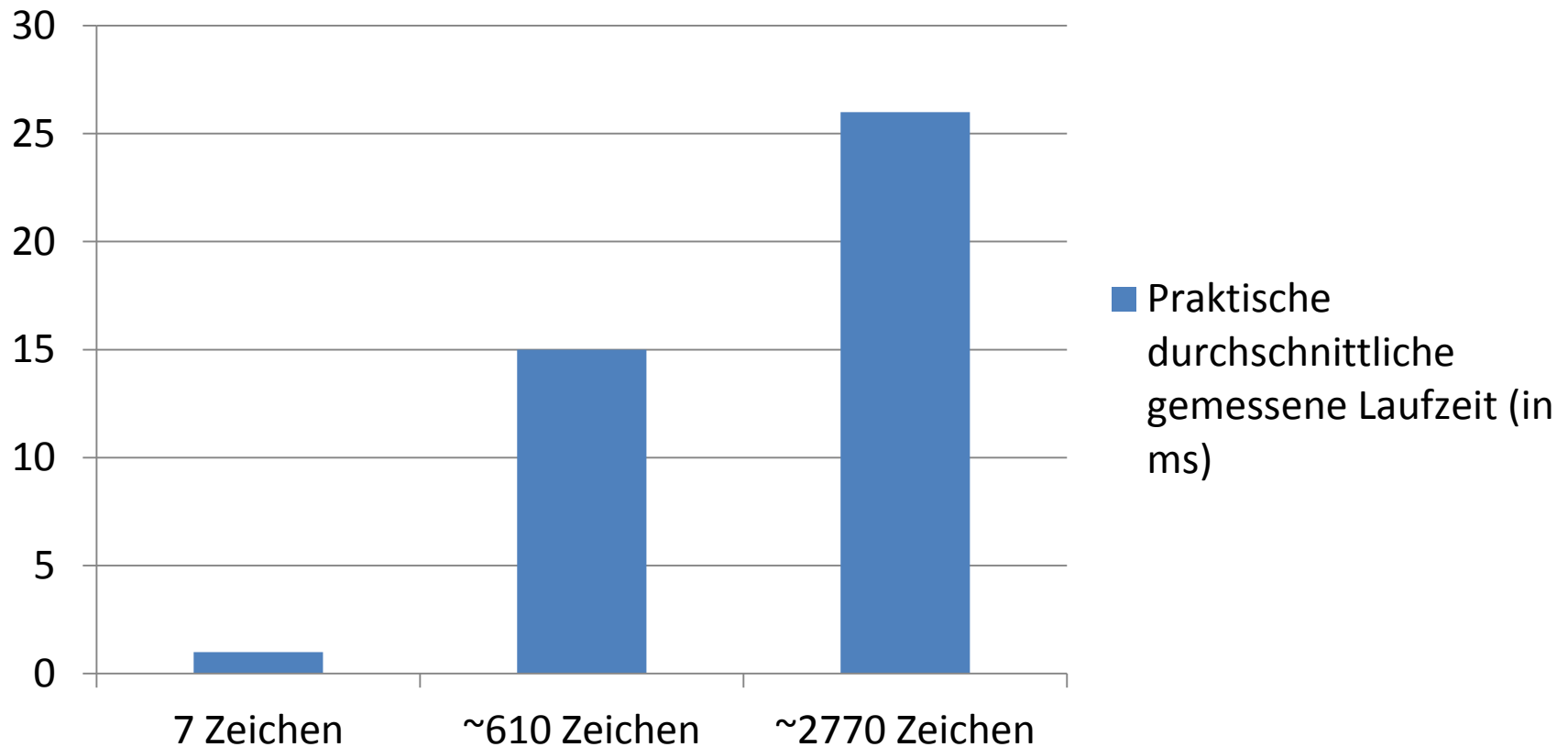
- Dient zum String-Alignment von Zeichenketten
- Erzeugt eine Zahlenmatrix anhand der Größen der zu vergleichenden Zeichenketten
- Buchstaben werden schrittweise verglichen
- Backtracking Algorithmus mithilfe der Matrix erstellt die verglichenen Zeichenketten des String-Alignments

# Laufzeit des NWA

- Eine Matrix mit der Größe  $n \cdot m$  muss mehrmals vollständig bearbeitet werden
- Theoretische Laufzeit beträgt  $\Theta(nm)$

# Laufzeit des NWA

Praktische durchschnittliche gemessene Laufzeit (in ms)



# Laufzeit des NWA

Theoretische Laufzeit nach ~630 Zeichen (in ms)

