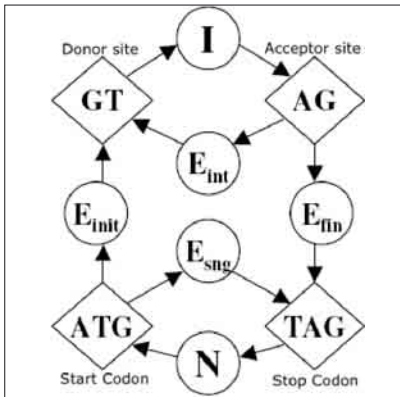




Markus Boos  
Cassian Strässle

## Prediction of Genes in Eukaryotic DNA

Diplomanden	Markus Boos, Cassian Strässle
Examinator	Prof. Dr. Guido Schuster
Experte	Gabriel Sidler, Eivycom GmbH, Zürich
Themengebiet	Digitale Signalverarbeitung
Projektpartner	Institut für elektrotechnik og databehandling, Universitetet i Stavanger (Norwegen)



Generalisiertes Hidden Markov Modell

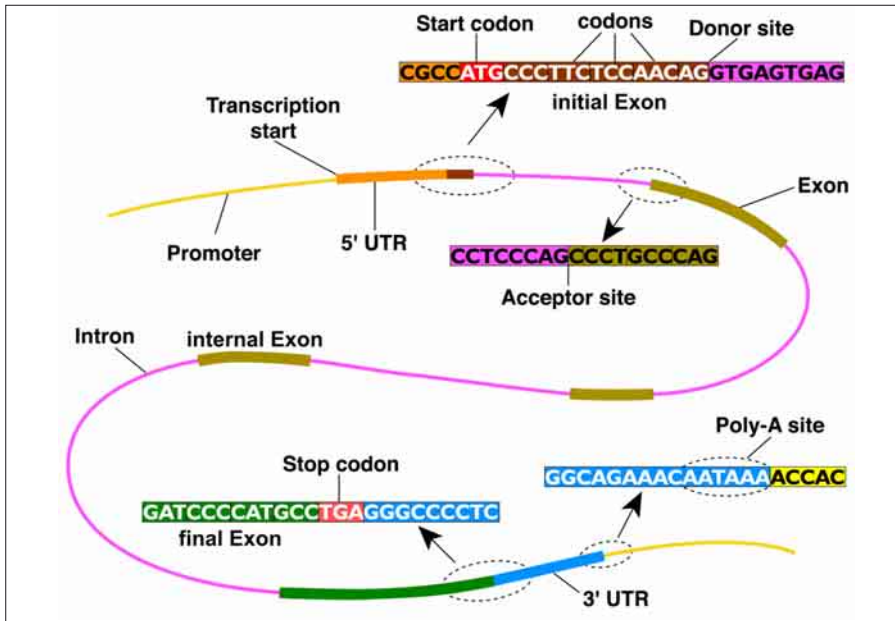
**Aufgabenstellung:** Desoxyribonukleinsäure (DNA) besteht aus vier Nukleotiden (Adenin, Cytosin, Guanin, Thymin), welche als Tripletts ein Codon bilden. Ein Codon codiert eine Aminosäure, wobei verschiedene Codons die gleiche Aminosäure erzeugen können. Aus den Aminosäuren entstehen Proteine, welche die Grundbausteine aller Zellen sind.

Der Aufbau der DNA-Sequenz folgt einer bestimmten Syntax (Übersicht DNA Syntax). Ribosomen wandeln anhand dieser Syntax die DNA in entsprechende Proteine um (Proteinbiosynthese). Die DNA von Eukaryoten (Lebewesen mit Zellkern)

besteht aus mehreren Millionen Nukleotiden. Bis anhin musste die DNA experimentell analysiert werden, um herauszufinden, welche Proteine aus ihr erzeugt werden.

Mit Hilfe der digitalen Signalverarbeitung soll es in Zukunft möglich sein, die DNA schneller und billiger zu analysieren. Unsere Aufgabe ist es, entsprechende Methoden zu suchen und eine davon zu implementieren.

**Ziel der Arbeit:** Wir erstellen eine Übersicht über die heute angewendeten Methoden der digitalen Signalverarbeitung, welche für die DNA-Analy-



Übersicht DNA-Syntax

se verwendet werden. Diese Übersicht dient der Universität i Stavanger als Einstieg in die Bioinformatik. Zudem implementieren wir eine solche Methode in Matlab. Die Implementation soll als Grundlage für weitere Projekte dienen.

**Lösung:** Die Übersicht wird mit Informationen aus Fachbüchern und Nachforschungen im Internet erstellt.

Als Analyseverfahren implementieren wir ein generalisiertes Hidden Markov Modell, welches die Eigenschaften von bestimmten Bereichen und ihre Abfolge modelliert (Generalisiertes Hidden Markov Modell). Die Eigenschaften der Bereiche werden aus vorhandenen Gendaten ermittelt und in geeigneten statistischen Modellen erfasst. Hidden Markov Modelle wurden bis anhin in automatischen Spracherkennungssystemen verwendet.