



Abb. 2.1. Schematische Darstellung des durch die Begriffabfolge Struktur, Abstraktion, Analyse gegebenen bioinformatischen Arbeitsprinzips am Beispiel einer DNA-Sequenz. Mit dem oberen Bildelement ist hier das Makromolekül selbst mit seinen vielfältigen biochemischen Eigenschaften gemeint. Damit ist klar, welche enorme Abstraktion in dem Schritt zur Symbolsequenz liegt. Der Weg von der Sequenz zur zugewiesenen Funktion führt dann oft über spezielle Modelle

schaften analytisch herleiten, so dass man eine Information darüber erhält, welche Aspekte des realen Phänomens zwangsläufig aus den (oft wenigen) im Modell festgelegten Grundeigenschaften folgen. Während die beiden letztgenannten Punkte üblicherweise Gegenstand der *Theoretischen Biologie* sind, findet sich der erste Aspekt, die numerische Simulation, in einer Vielzahl bioinformatischer Untersuchungen. Die dabei verwendeten Modelle sind Wahrscheinlichkeitsmodelle. Ein Wahrscheinlichkeitsmodell erzeugt zum Beispiel Sequenzen. Die (in dem Modell möglichen) Sequenzen werden im Allgemeinen mit unterschiedlichen Wahrscheinlichkeiten hervorgebracht, die von den Parametern des Modells abhängen. Dieses elementare Szenario werden wir in den Kapiteln 2.2 und 2.3 diskutieren. Wir werden im Folgenden anhand einfacher Beispiele sehen, wie man Wahrscheinlichkeitsmodelle formuliert und vor allem, wie sich dann bei gegebener Modellstruktur die Modellparameter aus empirischen Befunden schätzen lassen. Die derzeit beste Darstellung, um von dieser pragmatischen und qualitativen Ebene zu einer mathematisch exakteren Ebene zu gelangen, ist das Buch von Durbin et al. (1998).

2.2 Wahrscheinlichkeitsmodelle

Einen Zugang zu dem abstrakten Begriff der Wahrscheinlichkeitsmodelle gewinnt man über Beispiele. Es ist klar, dass die für uns relevanten Objekte Sequenzen sind