

# 1 RNA-Strukturen

Es gibt verschiedene Modelle auf denen RNA Sekundärstrukturvorhersage basieren kann. Diese Modelle wurden im Verlauf der Entwicklung komplexer um weitere, in den Vorläufermodellen nicht vorhandene, biologische Eigenschaften mit zu erfassen.

## 1.1 Nussinov Algorithmus(1978)

Das Ziel dieses Algorithmuses ist die Maximierung der Basenpaaranzahl. Der Algorithmus zieht einige biologische Gegebenheiten nicht in Betracht, zu diesen gehören:

- Pseudo-Knoten
- Interaktionen zwischen ungepaarten Regionen
- Stacking Interaktionen

## 1.2 Zuker & Stiegler Algorithmus(1981)

Der von Zuker & Stiegler entwickelte Algorithmus errechnet Sekundärstrukturen auf der Basis der "Minimum Free Energy". Dazu verwendet der Algorithmus experimentell ermittelte Energiewerte. Der beschriebene Algorithmus baut auf den Algorithmen von Nussinov (1978) und Smith & Waterman (1978) auf. In ihrem Paper sie definieren

- Hairpin loop
- Bulge loop
- Interior loop
- Bifurcation loop (wird heute als multiloop bezeichnet)
- stacking region

## 1.3 McCaskill's Algorithmus (1991)

# 2 Protein-Strukturen

Proteine besitzen mehrere Strukturebenen. Die basalste Ebene stellt die Primärstruktur dar. Sie listet die Sequenz der Aminosäuren (vom N- zum C-Terminus) auf. Die Sekundärstruktur beschreibt welche Bereiche der Primärstruktur sich in  $\beta$ -Faltblätter und  $\alpha$ -Helices falten. Die Tertiärstruktur gibt an wie die Sekundärstrukturelemente räumlich zu einander angeordnet sind. Wiederkehrende räumliche Anordnungen dieser Elemente werden als Motive bezeichnet. Die Quarternärstruktur beschreibt wie sich mehrere Polypeptidketten zu einem Multimer zusammen lagern. Proteine können entfaltet und zurück gefaltet (das funktioniert nur bedingt) werden. Es wird häufig angenommen, dass sich ein Protein in die thermodynamisch bevorzugte Konformation faltet.

Basierend auf dieser Annahme erwartete man, dass mathematische Modelle die Vorhersage der Proteinstruktur ermöglichen sollten. Daraus entwickelte sich die Idee des **Molecular Modelling**.

**Anfinsen's dogma** Die native Struktur eines Proteins wird einzig und allein von seiner Aminosäuresequenz bestimmt. Sie ist ein einzigartiges, stabiles und kinetisch zugängliches freies Energie Minimum.

**Levinthal's paradox** Die Anzahl der möglichen Konformationen die ein gegebenes Protein annehmen kann, ist astronomisch groß. Eine Polypeptidkette mit 100 Aminosäuren besitzt 99 Peptidbindungen und damit 198  $\Phi$ - und  $\Psi$ -Bindungswinkel. Wenn jeder dieser Bindungswinkel drei stabile Konformationen besitzt, existieren 3198 mögliche Konformationen.

## 2.1 Molecular Modelling

Um das Modellieren von Proteine dennoch zu ermöglichen werden vereinfachende Annahmen getroffen. So werden Atome als Punktladungen mit einer assoziierten Masse dargestellt. Die kollektive mathematische Beschreibung eines Proteins erfolgt durch die Potentialfunktion. Diese fasst die Energiebeiträge verschiedener physikalischer Kräfte und die Potentiale zwischen allen Paaren von Atomen ("pair potential"), diese bilden das Kraftfeld (force field). Molekulare Mechanik minimiert die statische Potentialenergie und Molekulare Dynamik (molecular dynamics) modelliert das Verhalten des Systems über die Zeit.

In ein Kraftfeld fließen alle Energiebeiträge ein die die Struktur eines Proteins beeinflussen. Diese sind:

- Energiebeiträge der chemischen Bindungen
  - Bindungslänge (durchschnittliche Entfernung zweier chemisch verbundener Atome im Molekül)
  - Bindungswinkel (Winkel zwischen zwei aufeinander folgenden Bindungen)
  - Torsionswinkel (Winkel zwischen den beiden Flächen (1. Fläche: Atome 1,2,3; 2. Fläche: Atome 2,3,4) die von vier Atomen in einer Kette gebildet werden)
- Nicht-Bindungsenergiebeiträge
  - van der Waals Energien
  - Elektrostatische Wechselwirkungen
  - conjugated system distortion
  - Wasserstoffbrücken
  - Lösungsmiteleinflüsse

Damit ergibt sich die Gesamtenergie zu:

$$E_{total} = E_{bond} + E_{angle} + E_{torsion} + E_{Waals} + E_{coulomb} \quad (1)$$

## 2.2 Statistische Potentiale

A statistical potential can be derived from known structures and probabilistic theory. The equation for the statistical potential is:

$$E_{total} = -kT \sum_s \ln \frac{P(s)}{P_R(s)} \quad (2)$$

## References

- [1] Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, et al. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biol 6(4): e96.