# Life Sciences Informatics

Die DNA ist wie ein Programm, was sich selbst über Millionen von Jahren geschrieben hat. Sie besteht aus 4 unterschiedlichen Bausteinen, als A, C, T und G abgekürzt, die in einer Doppelhelix hängen. Weiters gibt es auch die RNA, welche an einer einzelnen Helix hängt und eine andere Basis hat (U statt G). Sie ist instabil, damit sie schnell auf- und abgebaut werden kann. RNA-Sequenzierung ist heute ein beliebtes Forschungsfeld geworden, dadurch, dass Sequenzierapparate günstiger geworden sind, wurden in den letzten Jahren sehr viele Genome bekannt. Die Informationsdichte von DNA ist extrem hoch, in jeder menschlichen Zelle steckt ein aufgewickelter 2 Meter langer DNA-"Faden".

Relevant zur Informatik ist die Sequenzierung, da sie hohe Anforderungen an Rehcenleistung hat. Heutzutage wird die Sequenzierung parallelisiert, indem das Genom in kleine Stücke zerteilt wird und die entstehenden Sequenzen danach wieder zusammengesetzt werden. Die Zusammensetzung ist algorithmisch interessant. Weiters sind durch die Datenmenge, die bei der Sequenzierung von Genomen entsteht, auch Speichermedien interessant. Beim Vergleich von Samples von Genen gegen eine Sequenzdatenbank fallen auch extrem hohe Rechenlasten an.

Gene kodeieren Proteinsequenzen, welche aus 20 Bausteinen (Aminosäuren) bestehen. Unklar ist die Frage, wie man von einer Sequenz zu ihrer dreidimensionalen Struktur kommt (die Zuordnung von einer Sequenz zu einer Struktur ist surjektiv). Ein interessanter Fakt ist, dass nur von 150.000 Proteinen die 3D-Struktur bekannt ist, obwohl die Kodierung 146 Millionen mögliche Sequenzen hat.

Da die Basen der Sequenzen unterschiedliche Eigenschaften wie Hydrophobie oder Hydrophilie haben, formt sich aus der Kette irgendwann ein stabiles Konstrukt, wenn man sie in Wasser gibt. Dies ist der Faltungsvorgang von Proteinen, welcher oft im Computer simuliert wird. Hierzu muessen alle Wechselwirkungen zwischen den einzelnen Bausteinen berücksichtigt werden. Die zur Simulation benötigten Rechenressourcen sind für größere Proteine heute noch nicht verfügbar.

# Feedback

* "Visualisierung" von Zahlen mit Vergleichen: 22000 DVDs, Milchstraße mit Genketten der Einwohner von London durchqueren
* interessante Vortragsweise
* manchmal interessant klingende Themen übersprungen