

## **Diplom- oder Projektarbeit: Nutzung struktureller Information zur Vorhersage von Transkriptionsfaktor-Bindestellen**

Kurze Signalsequenzen sind an der Steuerung vieler genomischer Prozesse beteiligt. Auf der Ebene der Transkriptionsregulation sind dies Transkriptionsfaktorbindestellen, an die Transkriptionsfaktoren binden und so die Transkription erleichtern oder hemmen. Die Vorhersage von Bindestellen bekannter Transkriptionsfaktoren wird häufig mithilfe statistischer Modelle bewerkstelligt. Eines der bekanntesten Modelle ist dabei die Gewichtsmatrix [4, 3] (PWM - position weight matrix), die von statistischer Unabhängigkeit zwischen den Positionen der Bindestelle auf der DNA ausgeht. Für viele Familien von Transkriptionsfaktoren, z.B. Zink-Finger, ist die dreidimensionale Struktur der Faktoren bekannt. Teilweise lässt sich aus dieser Struktur ableiten, an welchen Stellen es zu Interaktionen zwischen dem Faktor und der DNA kommen kann [1, 2].

Ziel der Arbeit ist es, die Struktur und Interaktionsmöglichkeiten für verschiedene Familien von Transkriptionsfaktoren zu recherchieren. Diese Strukturinformationen sollen dann für den Entwurf eines statistischen Modelles herangezogen werden, das spezifisch auf die Eigenschaften der Faktoren dieser Familie zugeschnitten ist. Das Modell soll dabei so implementiert werden, dass es sowohl generativ als auch diskriminativ im bestehenden Java-Framework ([www.jstacs.de](http://www.jstacs.de)) genutzt werden kann. Mithilfe von Jstacs soll das entworfene Modell dann mit bestehenden Modellen zur Vorhersage von Transkriptionsfaktor-Bindestellen verglichen werden.

### **Ansprechpartner:**

Jan Grau, Raum 4.09, 55-24768, [grau@informatik.uni-halle.de](mailto:grau@informatik.uni-halle.de)

### **Literatur**

- [1] Tommy Kaplan, Nir Friedman, and Hanah Margalit. Ab initio prediction of transcription factor targets using structural knowledge. *PLoS Comput Biol*, 1(1):e1, Jun 2005.

- [2] Re V. Morozov, James J. Havranek, David Baker, and Eric D. Siggia. Protein-dna binding specificity predictions with structural models. *Nucleic Acids Res*, 33:5781–5798, 2005.
- [3] R. Staden. Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Research*, 12:505–519, 1984.
- [4] G. D. Stormo, T. D. Schneider, L. M. Gold, and A. Ehrenfeucht. Use of the 'perceptron' algorithm to distinguish translational initiation sites. *NAR*, 10:2997–3010, 1982.