Preface

Gene regulatory networks are fixed determinants of cellular control and the abundance of differentially expressed regulatory proteins, called transcription factors, their driving signal. In concert with specific epigenetic marks, transcription factors define the active subset of the network to govern distinct cellular states in time and space.
Eukaryotic gene expression is generally controlled through molecular logic circuits combining regulatory signals of several transcription factors. Recently, it has been shown that complexation of regulatory proteins is a prevailing and highly conserved mechanism of signal integration within critical regulatory pathways, like body part formation or differentiation. A knowledge of potential assembly candidates could provide the basic information that is needed to infer possible target genes as well as the exerted mechanism of influence. There already exists a plethora of approaches to predict protein complexes from protein-protein interaction data. However, those are generally designed to detect large self-contained functional complexes and lack the ability to reveal dynamic and highly modular combinatorial complex assemblies, a property of crucial importance for the signal integration exerted by transcription factor complexes.

The method proposed in this thesis combines protein-protein interaction networks and domain-domain interaction networks with the well-known cluster-quality metric cohesiveness. A novel growth algorithm is described that locally maximizes the metric on the holistic level of protein interactions while sophisticated connectivity constraints are preserved. Assuming that each domain can only support one interaction, the domain topology can be utilized to account for the exclusive and thus combinatorial nature of physical interactions between proteins. During the growth process, the complex candidate is thought to be backed by a spanning tree of simultaneously possible domain interactions which restrict further expansion possibilities. Consequently, every addition of a protein requires the choice of an applicable domain interaction which again influences later steps. Often many options have to be taken into account by branching of the algorithm, which naturally allows for the justified prediction of a manifold of transcription factor complexes from a common start.

The proposed approach outperformed popular complex prediction methods by far for the prediction of transcription factor complexes in yeast. The evaluation was based on established benchmarks assessing accordance with several reference complex datasets as well as measures of biological relevance. Additionally, many of the predictions of the proposed method could be associated with target genes and a potential regulatory effect. Furthermore, predicted candidates could be mapped to distinct functions during a defined cellular state and condition by analyzing the expression coherence among their regulated genes for cell cycle expression data. Many findings were backed up by literature evidence.

The results encourage an application to higher eukaryotes where the combinatorial interplay between transcription factors is more pronounced. The knowledge of putative transcription factor complexes - DNA-binding members and recruited potentially regulatory active proteins - offers novel capabilities in the automatized modeling of gene regulatory networks which may assist to surpass nowadays models.

A condensed summary of the novel concept, the main method and the results for yeast was previously published in [1] prior to the production of this book.

http://www.springer.com/978-3-658-08268-0

Predicting Transcription Factor Complexes
A Novel Approach to Data Integration in Systems Biology
Will, T.
2015, XIX, 142 p. 29 illus., Softcover
ISBN: 978-3-658-08268-0