Chapter 5

Concluding Remarks

In the first part of this thesis, I have attempted to evaluate the potential and the limitations of using structure information for the study of protein interactions. I have shown that protein domains known to be part of an interaction interface in a protein structure can be projected onto the protein interaction network. This reveals that while our current knowledge of interacting domain pairs is small, these domain pairs are significantly overrepresented in experimentally verified protein interactions in both eukaryotes as well as prokaryotes. There is also significant conservation of domain pairs between species, even though only approximately 5% of the protein interaction network is covered by the structural data. This presents a strong argument for solving the structures of more novel interacting domain pairs. A substantially higher coverage could for example provide enough information to identify the most likely binary pairs of interacting proteins in complexes identified using affinity-purification methods: those protein pairs with known interacting domain pairs can be assumed to be more likely to really interact.

In the following chapter, I demonstrated that the existing structural data can be employed successfully to investigate disease mutations on a molecular level. I described several genetic diseases which are the result of point mutations in a domain which is known to be involved in an interaction through a homologous structure. In the future,

binding kinetics experiments will hopefully confirm my predictions. My approach already exemplifies the power of structural homology based approaches applied to protein interactions. Within the possibilities of the incomplete datasets available, I estimated that 4% of all known disease mutations affect a protein interaction. Increased numbers of structural templates and more stringently defined domains, representing only a particular binding geometry or binding partner, could improve the sensitivity and specificity of my method further.

Interestingly, many of the mutations in interaction interfaces are inherited in a dominant fashion. In the last part of this thesis, I extended my analysis beyond structure-based domains to study the evolutionary pressures governing protein complexes in human. Specifically, I investigated the distribution of protein complexes with respect to large insertion and deletion polymorphisms often referred to as copy-number variations (CNVs). It is known that proteins vary regarding their duplicability and sensitivity to homozygous deletion. It has been argued that many dosage sensitive proteins are members of protein complexes. I observed in human that expression variation in members of protein complexes is significantly lower than in other selected proteins. Furthermore, I could show that members of protein complexes are rarely found inside CNVs. Combined, these two facts suggest that frequently, purifying selection acts against CNVs that contain genes encoding protein complexes, or genes in protein complexes have evolved to reside outside regions which are enriched for CNVs. It seems likely that such evolutionary pressures have been acting for some time, as the set of protein complex genes also has fewer paralogs on average than other genes. In congruence with the duplication/divergence theory of gene evolution, the studied genes of members of protein complexes are under stronger negative selection than the rest of the genome, as indicated by their low dN/dS rates.

An interesting alternative approach to the same question could be the analysis of known knock-out mice mutants. With the increasing availability of knock-out models

for various genes, it could be envisaged to differentiate between heterozygous as opposed to purely homozygous phenotypes, in a similar way as dominant and recessive mutations are defined in human disease. From my initial results presented in this thesis, I expect knock-outs of genes in protein complexes to be more often phenotypicaly active than other genes.

In summary, it can be said that the investigation of protein interactions has already brought about many exciting insights and fostered interconnections between previously unrelated fields. Combining structure information with protein interactions to explain genetic diseases is an example of such an integrative approach that will probably become more common in the coming years. Similarly, my analysis of large scale genomic variation in the context of protein interactions shows how network biology can provide insights into such fundamental questions as gene duplicability. However, as the field of protein interaction research is still in a comparatively early stage of development, many basic assertions still need to be made and many obstacles need to be overcome. Our understanding of the evolution of protein interactions is still incomplete. Being able to trace the processes that shaped the interaction networks of higher organisms would not only shed light on the origins of organismal complexity, but could also be of practical use: it is still unclear to what extent protein interactions are conserved between species. Moreover, it is also not yet fully understood what distinguishes a protein interaction interface from other surface regions. As a result of that, our ability to validate or even predict protein interactions is still limited. My findings point towards the possibility of reducing the complexity of protein interaction networks down to domain interaction networks as a more conserved unit of interaction evolution.