# Research interests and project proposal

My research focuses on two main areas: **protein interaction maps of whole genomes** and on **protein domain interactions** and their specificity.

My philosophy is that we should study model organisms or physiological processes in systematic, collaborative and coordinated efforts to which I can contribute my experience in protein interaction technology.

Previous (and still ongoing) projects

In my previous work I used high-throughput methods and bioinformatics to study protein interaction networks in yeast. We not only found a large number of hitherto unknown interactions in yeast (Uetz et al. 2000) but also determined the size and topology of a cellular proteome network for the first time (Schwikowski et al. 2000).
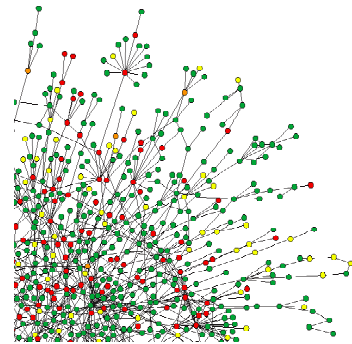
I am particularly interested in **protein-protein interaction networks** of whole organisms and their analysis by comparative approaches. The comparative bioinformatics of such networks is not well understood because there is still *too little data from different model organisms*. We have been generating such data for several model organisms and comparing the results. This analysis addresses overall network topology as well as questions about individual proteins. For example, how much can protein sequences diverge until their interactions get lost or change specificity? How do different protein families and domains compare in this respect? Which amino acids do we have to change to change binding (and hence biological) specificity? Can we understand these properties in structural terms (based on NMR or X-ray data)?

At the end of my postdoctoral work, I started to study **protein domains** and their protein-protein interactions. One important aim of these studies is to identify protein ligands for little known domains or domains of completely unknown function. Surprisingly, there are still dozens of protein domains in both prokaryotes and eukaryotes for which no ligand (and thus no function) is known. Many of them are supposed to interact with other domains. Once we have identified their interaction partners, I would like to learn how their binding specificity is determined. Recently, we have done this kind of analysis for **PX, DEP, and FF domains of yeast**. The ligands that we identified for these domains demarcate breakthroughs in the understanding of these domains. For example, we have convincingly demonstrated for the first time that the PX domain is a *bona fide* protein interaction domain (Vollert & Uetz, 2004). Previously, it was thought that the PX domain is only a lipid-binding domain. In collaboration with a local NMR group we are now studying the exact binding interaction modes of the PX domains with its targets. Functional studies are under way as well.

Protein interaction maps of viruses

In the course of several collaborative projects we just finished the two-hybrid interaction maps of Kaposi-Sarcoma Associated Herpesvirus (**KSHV**) and *Sulfolobus spindle-shaped virus* 1 (**SSV1**). These are the most extensive interaction studies in viruses carried out to date. In collaboration with Jürgen Haas (Munich) we are currently working out the interaction maps of Varicella Zoster Virus (VZV) and Cytomegalovirus (CMV). KSHV was the first herpesvirus studied this way and with VZV and CMV we will have studied all 3 major herpesviral subfamilies in a comprehensive fashion.

Analysis of single proteins and protein domains

We are using **peptide mapping** in order to analyze the binding specificity of protein domains to peptides. Peptides can be used to derive **consensus sequences for binding.** Such consensus sequences will eventually allow the computational prediction of interacting proteins from other genome sequences.
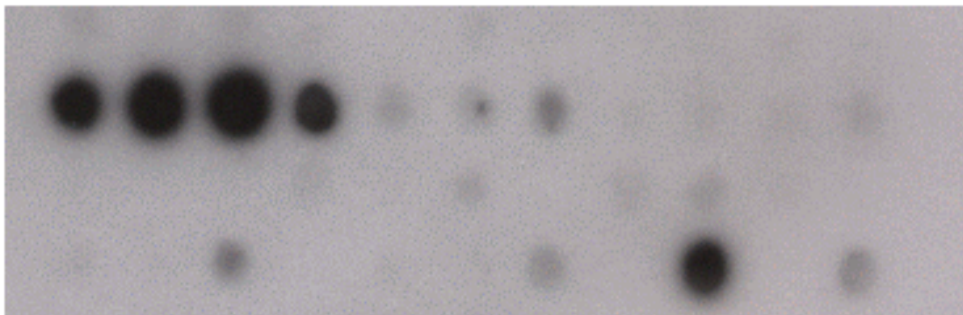
Membrane with overlapping peptides of the yeast Yip1p protein, synthesized on the membrane using SPOT synthesis. The membrane-bound peptides were then incubated with a GST-PX fusion protein and detected by anti-GST antibodies and ECL. The 4 spots in the left half indicate the binding site of the PX domain while the single spot on the bottom right is a positive control (Vollert & Uetz, unpublished).

Obviously, I am also very interested in extending my studies through **collaborations** with other labs that study the atomic details and *in vivo* relevance of such interactions (e.g. using NMR, X-Ray crystallography, and/or mass spectrometry etc.).

Future Projects

Protein interaction maps and systems biology of bacteria

Surprisingly, there is not a single organism for which a complete protein interaction map has been accomplished. All published reports are only partial analyses. I suggest to analyze two small bacterial proteomes both experimentally and computationally by comparing their interaction and signaling patterns, as well as integrate these data with metabolomic, expression, phenotypic, and structural data.

Comparative interactomics of *Treponema pallidum* and *Borrelia burgdorferi*

We just started a project aiming at the protein interaction map of *Treponema pallidum*, the causative agent of Syphilis (in collaboration with Tim Palzkill, Baylor College of Medicine). This will be the main focus of the lab for the near future. *Treponema* was chosen because it has an extremely reduced genome of only ~1,000 open reading frames (ORFs) which we have already cloned into two-hybrid vectors. So far, we have carried out screens with all cell division and motility-related proteins. All 1000 screens will be done by spring 2005. This analysis should give us an idea of the connectivity of *Treponema* proteins and their function (of which nearly 50% are unknown!). In particular, it will tell us how metabolic pathways, signaling pathways and structural units in the cell are connected.

We plan to supplement the two-hybrid data with an interaction map generated by the use of protein arrays (which can be easily made using the cloned ORFs!). The latter will also be used to identify phosphorylation sites and other modifications on a genome-wide level, e.g. by screening it with kinases and radio-labelled ATP. The goal of this project is to identify signaling pathways and regulatory interactions on a genome-wide level.
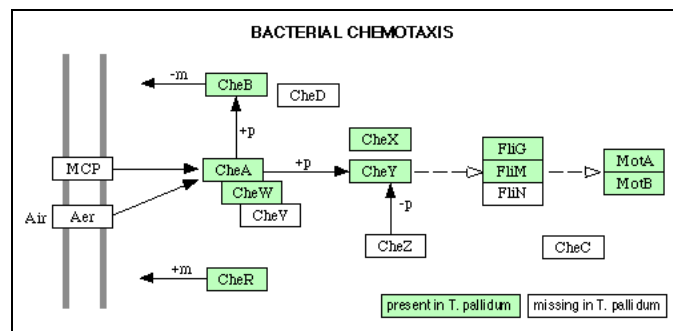
Comparative interactomics in bacteria: an example.

Subsequently, *Borrelia burgdorferi*, the spirochete that causes Lyme disease, will be analyzed in a similar fashion. A comparison of both species will not only cross-validate the results obtained from each other but will also answer many **biological questions**, e.g. (1) Which protein interactions (in addition to proteins!) are conserved (i.e. "essential") in *Borrelia* and *Treponema*? (2) Which (conserved) domains and/or motifs mediate these interactions? (3) How do different processes in a cell interact (i.e. communicate) both in terms of protein-protein interactions as well as via small molecules? (4) Can signaling (and metabolic) pathways explain the very different lifestyles of *Treponema* (vertebrates only) and *Borrelia* (vertebrates *and* ticks)? (5) How do phage interact with their hosts? (both whole genome arrays can be screened with proteins of species-specific phage!). (6) What is the function of the numerous proteins of yet unknown function? ? I am especially interested in conserved proteins of unknown function as their homologues can also be studied in *E. coli* or *B. subtilis*. This group of ~100 proteins will be subjected to more detailed analysis.

Undergraduate and graduate teaching interests

Given the complexity of molecular networks, I consider it as of foremost priority to make the daunting amount of information accessible to students. Students need to understand both the **principles** of biology but also the importance of molecular **detail**. I try to reflect this conflict also in my research where I apply *global* approaches to *single* proteins. Since I am interested in comprehending the whole universe of complexity from protein (domains) to cells to species to biodiversity, I would like to communicate this also in my teaching – essentially the unity of biology across all those levels.

More specifically, I will be able to teach **molecular biology** and **genetics**, **genomics** and **proteomics**, **biochemistry**, as I have done in the past:

Classes currently taught (as adjunct faculty at the University of Karlsruhe):
- A lecture series in **molecular biology** (undergraduate and graduate level)
- A practical course in **yeast molecular biology**, including an introduction to **bioinformatics.**
- A seminar on genome and and proteome research.

Graduate students also need to learn how to reconcile the need for detail and the "big picture". That is why I have most students do a genome-wide screen of protein interactions during which they learn to appreciate the "genomic view" of complexity. They are then allowed and in fact encouraged to follow up their pet proteins and interactions in order to work out the mechanistic and biological details. I encourage all PhD students in the lab to give **journal clubs** that cover these different levels of complexity.

Finally, I also participate regularly in an **outreach** program of the FZK that offers lectures to lay people as well as high school teachers and students.