Summary

Almost all cellular processes depend on finely tuned protein-protein-interactions. A substantial number of these interactions are mediated by a relatively small number of sequentially and structurally conserved protein-interaction domains using conserved mechanisms. Some of these interactions represent also interesting targets for new pharmacological intervention strategies.

In this work, qualitative and quantitative interaction models were built on the sequence and structure level to understand the mechanisms of binding as well as specificity and to predict new potential interactions for protein-interaction-domains (WW and PDZ domains). These statistical models were trained using experimental interaction data obtained by screening 42 WW domains and 3 PDZ domains with peptide libraries.

For WW domains, a classification system was developed which for the first time allowed the assignment of WW domains as well as their proline-rich ligands to the following 6 different specificity groups based solely on their sequences: the Y-specificity group characterized by the xPPx**Y**-recognition motif, the R_a-group with the (P/F/I/V)P(P/G)PPP**R**-motif, the R_b-group with the PPP**R**GPPP-motif, the L-group with the PP**L**PP-motif, the phosphoserin(*po*S)/phosphothreonine(*po*T)-group with the **(*po*T/*po*S)P**-motif and the poly-P-group with the (**P**/I/V/L)**PPPPP**-motif. As the specificity determining residues in the domains could also be identified for 5 of the 6 specificity groups, 66% of all 482 known WW domains (SMART-database, as of 11/15/2001) could be assigned to one of the specificity groups. Subsequently, the predicted tyrosine-specificity for 3 WW domains with before unknown specificity was experimentally confirmed.

For statistical interaction models based on domain or complex structures, the 3D-structures of all 42 experimentally analyzed WW domains were built by homology modeling. Furthermore, models of the domain/ligand-complex structures for each of the 6 specificity groups were devised based on systematic analyses of the recognition motifs (substitutional

analyses). Thus, the first consistent hypothesis for the recognition of arginine-containing ligands by WW Domains of the R_a- and R_b-specificity groups could be suggested.

Based on the modeled structures of the 42 WW domains, a quantitative structure-activity-model of tyrosine-specificity was built using the *Comparative-Molecular-Field-Analysis* (CoMFA). This CoMFA-model allowed the prediction of affinity towards tyrosine-containing ligands based on the structure of the domains.

Based on the modeled WW-domain/xPPx**Y**-ligand complexes, an enhanced quantitative structure-activity-model of tyrosine-specificity was built using the *Comparative-Binding-Energy-Analysis* (COMBINE). The explicit representation of ligands in this COMBINE-model allowed particularly the design of a ligand with higher affinity towards the first WW domain of the yeast NEDD4-like Ubiquitin-ligase *y*RSP5. For the threonine-to-aspartate mutation in position −5 relative to the tyrosine of the peptide G**T$_{-5}$**PPPPYTVG an almost 3-fold increase in affinity was predicted. This significant increase was confirmed experimentally. Interestingly, an aspartate is found in the same position within the recognition motif in the known *y*RSP5 interaction partner YM95_YEAST, emphasizing the *in vivo* relevance of this prediction.

Finally, the *Quantitative-Specificity-Profile* (QSP) method was developed. For the first time it allows the quantification of the specificity of a protein-interaction domain in terms of a ligand sequence-dependent affinity function defined for the complete ligand sequence space. By using QSP, not only is the specificity visualized intuitively (term scheme) but the affinities are predicted for all potential ligands. QSP-models were built for the PDZ domains of *h*AF6, *h*ERBIN and *m*SNA1 (http://www.fmp-berlin.de/nmr/pdz). They were used to predict the optimal peptidic ligands (superbinder) for each of the three domains. Subsequently, these were confirmed experimentally. In addition QSP-models also permit the quantification of selectivity for competing domains. For the three PDZ domains, significant differences in selectivity as well as a surprisingly large overlap of the ligand sequence spaces were found.