FINAL REPORT OF NK81950 GRANT

Molecular recognition via natural and artificially-evolved linear peptide motifs: structural, thermodynamic and kinetic principles

We have studied protein-protein interactions (PPIs) involving molecular recognition through linear motifs (LMs) to generate a comprehensive description of structural, thermodynamic and kinetic principles of these interactions. The model proteins included hub proteins interacting with motor proteins, serine proteinases, protein kinases and their scaffold proteins and amyloid formation.
The major results of the project, which have been published in 30 peer-reviewed research papers, are summarized in five chapters.

1. Protein-protein interactions of LC8 through linear motifs

LC8 (DYNLL) is a typical hub protein that binds to many partners containing LMs. Our primary goal was to thoroughly characterize the interaction network of this highly conserved small globular and dimeric protein, which is an important "dimerization engine" of all eukaryotic cells.

We compared the kinetic and thermodynamic parameters of the two mammalian isoform of LC8 binding to four LMs. Binding characteristics of the two isoforms *in vitro* are highly similar. Importantly, we found that bivalent ligands (dimeric peptides/proteins) show significant, three orders of magnitude increase in binding affinity due to an avidity effect decreasing the k_{off} of the complex. We described the mechanism of complex formation by a conformational selection mechanism. We have disproved a phosphorylation-based mechanism previously suggested to regulate the activity of LC8 (which was supposed to contribute to its tumor promoting activity in breast cancer cells). (*Radnai et al, JBC, 2010*)

The binding preference of the 8-residue long LM was investigated by directed evolution, phage-display. We were able to evolve a binding peptide with a 20-times higher affinity than the previously described strongest LC8 binding motif. The evolved peptide can serve as competitive inhibitor of LC8 functions. The structural basis of the stronger binding was explained by our atomic-resolution structures of LC8 in complex with the evolved peptide in monovalent and bivalent form. Based on the established binding preference we applied a bioinformatics approach to predict additional binding partners of the LC8 hub protein. (*Rapali et al, PLoS One, 2011*)

By extending our directed evolution studies we refined the surprisingly long list of predicted LC8-binding partners (>100 proteins). Validation of these predictions is an ongoing research (supported by other resources). (*Rapali, PhD Dissertation, 2013*)

We identified and characterized a novel LC8 binding protein ATMIN - an ATM kinase-interacting protein involved in DNA damage-induced nuclear foci formation - that interestingly displays 18 LC8-binding LMs. (*Rapali et al, BBRC, 2011*)

An extensive characterization of the LC8 LM of myosin 5a in free and LC8-bound form was performed using NMR spectroscopy, X-ray crystallography and molecular dynamics simulations. The atomic resolution structure of the complex explained how the non-canonical LM of myosin 5a fits into the common binding groove of LC8. Our results deepened the knowledge on the diverse partner recognition of the LC8 hub protein. (*Bodor et al, Biochemistry, 2014*)

Structural and functional aspects of the LC8 hub protein including its linear motif recognition and binding characteristics have been summarized in a review article. We also discussed the controversial role of LC8 as a cargo adapter protein of the dynein and myosin 5 motor proteins (*Rapali et al, FEBS J, 2011*)

The LC8 binding motif is very often localized in disordered regions close to coiled-coil domains that are stabilized upon LC8 interaction. We proposed a new role for coiled-coil structures in protein evolution: these structures easily evolve from or evolve to intrinsically disordered and/or charged single alpha-helix (CSAH) structural motifs (*Gáspári & Nyitray, BioMol Concepts, 2011*). CSAHs were previously recognized by us as novel protein structural elements that are relatively rare in proteomes, nevertheless are involved in many key biological processes. (*Gáspári et al, BBA, 2012*)

2. Linear motifs within the non-muscle myosin 2 (NM2) tail regions (the S100A4-NM2A interaction)

S100A4 is a vertebrate specific homodimeric Ca-binding protein that is overexpressed in metastatic tumors and is also involved in other pathologies. In the Ca-bound form it binds to several intra- and extracellular proteins including non-muscle myosin 2A (NM2A). We precisely delineated the LM of NM2A interacting with S100A4, quantitatively characterized the interaction and determined the atomic-resolution structure of the complex. It revealed an asymmetric mode of interaction unique in the S100 family: a single, predominantly α-helical myosin chain is wrapped around the Ca-bound S100A4 dimer occupying both symmetry related hydrophobic binding pockets. Based on these results, we proposed a model how S100A4 binding disrupts NM2A filaments leading to increased cell migration (a requirement for metastasis). Description of the complex will facilitate the design of specific inhibitors that interfere with the S100A4-NM2A interaction. (*Kiss et al, PNAS, 2012*)

We assigned a function to the C-terminal random coil region of S100A4. It appears to be involved in a conformational activation on Ca-binding. The activation of the entire C-terminal random coil may play a role in mediating the interactions with selected partner proteins of S100A4. These results provide an insight to the dynamic mechanism of C-terminal region as a mediator of S100A4-driven metastasis. (*Duelli et al, PLoS One, 2014*)

Despite having a high degree of sequence similarity the three NM2 paralogs (A, B and C) have isoform specific functions. We showed that S100A4 interacts not only with NM2A but also with NM2C (*Kiss et al, PNAS, 2011*). We carried out a combinatorial paralog scanning phage display experiment to determine the contribution of the α-helical S100A4 binding region of NM2 to isoform specificity. We found the unexpected result that selective binding is mostly determined by a single amino acid substitution (Ala in NM2A and NM2C vs. Asn in NM2B). The most diverse part of the binding region, the-non-helical tailpiece has less contribution to the binding affinity and affects more significantly the on rate constant of the interaction. We found that among the three isoforms NM2B has the most stable coiled-coil region. We suggest that weak S100A4 binding site combined by a stable coiled coil explains why S110A4 does not dissociate the NM2B filament (*Kiss et al 2014, in preparation*). (The S100 PPI project in our laboratory is further supported by another OTKA grant).

3. Linear motif preference of serine proteases

All serine proteinases have a primary specificity dictated by a substrate binding site in the immediate vicinity of the catalytic triad. This site forms a transient complex with an at least hexapeptide segment of the protein substrate that can be considered as a LM. As serine proteinases are highly potent enzymes many different types of inhibitors evolved to regulate them. The most diverse class of such inhibitors, reversible substrate-like inhibitors has 18 families and each family has a distinct 3D fold but possesses a common surface exposed loop that imitates a peptide substrate and binds to the target enzyme through an LM.

We evolved - using phage display - the canonical loop of several types of reversible inhibitors against various serine proteinases in order to assess whether there are general rules in this type molecular recognition. First we used the scaffold of the smallest natural trypsin inhibitor peptide, the 14 amino acid SFTI to evolve selective inhibitors against two enzymes of the lectin pathway of the complement system, MASP-1 and MASP-2. We evolved a MASP-2 specific and a dual specific inhibitor. The evolution provided a snapshot on the LM preference of these enzymes. With the combination of these two inhibitors we could prove that MASP-1, which was considered to be an auxiliary enzyme in lectin pathway activation, is much more important than formally believed. (*Kocsis et al., J. Immunol., 2010*)

We also evolved the analogous canonical loop of a larger, 35 amino acid inhibitor from the Pacifastin family against the same MASP enzymes. In spite of the highly analogous structures of the canonical loops of the two inhibitor families, the evolved motifs were characteristically different. This finding suggested that the optimal LM sequence is not dictated entirely by the proteinase. Based on high resolution X-ray structures we could rationalize the source of scaffold dependence. Moreover, on the Pacifastin scaffold we were able to evolve truly monospecific MASP-1 and MASP-2 inhibitors, respectively (*Héja et al, J. Biol. Chem. 2012*).

With the above inhibitors we could unequivocally identify MASP-1 as a master regulator of lectin pathway activation. We proved that MASP-1 is the only physiological activator of MASP-2, and therefore of the entire pathway (*Héja et al. PNAS, 2012*). Moreover, we could also provide a detailed kinetic analysis of lectin pathway activation (*Megyeri et al. JBC, 2013*).

We were invited to write a review paper on complement serine proteinase inhibitors (*Gál et al, Adv Exp Med Biol., 2013*).

Using a Pacifastin scaffold we determined the LM preference of a peculiar regulating enzyme, chymotrypsin C. We discovered that chymotrypsin C has a LM preference highly unique among serine proteinases as its major selectivity position is mapped to P4' (*Szabó A. et al J. Biol. Chem. 2011*).

The same Pacifastin scaffold was used to evolve an inhibitor that binds with extremely high affinity to the classic model enzyme, bovine trypsin. We could generate an ultra-high resolution (0.93Å) X-ray structure of the complex and could identify the protonation state of each catalytic triad residue and show that prior to the acylation step by Ser195, both Asp102 and His57 are in neutral state. This finding strictly contradicts the common model, in which Asp102 is permanently deprotonated throughout the entire catalytic cycle (*Wahlgren et al, J. Biol. Chem. 2011*).

We provided evidence for an extremely high resistance of a complex of crayfish trypsin (CFT) and bovine pancreatic trypsin inhibitor (BPTI) against heat- and chemical denaturation. To explore the structural features responsible for this extreme stability, we crystallized CFT in complex with BPTI, and identified extended contacts compared to the bovine trypsin-BPTI complex. (Molnár et el, FEBS J, 2013)

4. Molecular background of amiloid formation and the role of linear motifs

Amyloid formation is a complex multistep reaction, in which proteins often associate through linear motifs. The exact mechanism and the factors that can affect or inhibit the process *in vivo* are poorly understood. We studied the effect of two chaperon-like molecules (ERD14 plant chaperon and LC8 hub protein) on amyloid formation of β2-microglobulin (β2m) and Alzherimer's β-amyloid peptide. ERD14, a natively unfolded chaperon significantly slowed down the polymerization process. LC8 inhibited the fibril formation of β2m in the presence of low concentration of SDS under neutral conditions. The structure of β2m was proved to be native like in the presence of ERD14 or LC8 which can explain the decreased amyloidogenicity of the molecule (*manuscript is under submission*).

We developed a new method to study the effects of interacting molecules on the stability of amyloid fibrils. We found that β2m fibrils can dissociate at elevated temperatures and their thermal stability is a sensitive measure to characterize the interaction of additives and inhibitors with the amyloid form. Moreover this inexpensive simple technique enables high-throughput screening to find effective compounds that affect protein aggregation for both theoretical and pharmaceutical purposes. (*Kardos et al., Biochemistry, 2011*)

In collaboration with Yuji Goto's group we studied the thermodynamics of amyloid formation of β2m and other polypeptides and were capable of detecting the spontaneously nucleated process (*Ikenoue et al., PNAS, 2014*). We also investigated the effect of alcohol on the aggregation and structure of the amyloid fibrils of human insulin. We found that, depending

on the conditions, insulin can form several different types of amyloid fibrils with distinct morphologies and secondary structure compositions. *(Muta et al., J. Biol. Chem, 2014*)

Light chain-associated (AL) amyloidosis is characterized by dominant fibril deposition of the variable domain (VL) of IgG light chains, and thus its constant domain (CL) has been considered not to be amyloidogenic. We examined the *in vitro* fibril formation of the isolated CL in comparison with β2m. The results suggested that CL plays an important role in the development of AL-amyloidosis (*Yamamoto et al., FEBS Letters 2010*).

We have grown oligomers and amyloid fibrils of β-amyloid peptide of various morphologies, and also studied the *in vivo* effect of these aggregates by electrophysiology experiments on rats and proteomics studies on mouse brain samples. We found a strong correlation between the diverse *in vivo* effects and the aggregation time, aggregate size, and morphology. (*Orban et al., Brain Research, 2010, Szegő et al., Neuroendocrinology, 2011*)

We developed a new method for the improved, more accurate structure determination of protein and peptides by CD spectroscopy. The algorithm can distinguish parallel and antiparallel β- structures and can predict secondary structures of amyloid fibrils. The first manuscript is under publication. We presented our results at the Biophysical Society Meeting in Philadelphia (*Micsonai et al, manuscript submitted*)

5. Linear docking motifs in of MAPK pathway

Mitogen-activated protein kinases (MAPKs) have a docking groove that interacts with linear "docking" motifs in binding partners. Crystal structures of four complexes of MAPKs with LMs, representing JNK-specific, ERK-specific, or ERK- and p38-selective binding modes, revealed that the regions located between consensus positions in the docking motifs showed conformational diversity. Although the consensus positions in the docking motifs served as anchor points that bound to common MAPK surface features and mostly contributed to docking in a nondiscriminatory fashion, the conformation of the intervening region between the anchor points mostly determined specificity. These results suggest a coherent structural model for MAPK docking specificity that reveals how LMs binding to a common kinase docking groove can mediate diverse interaction patterns. (*Garai et al. Sci Signal, 2012*)

We determined the crystal structure of ERK5 in complex with an MKK5 construct comprised of the PB1 domain and the LM. We showed that ERK5 has distinct PPI surfaces compared with ERK2, which is the closest ERK5 paralog. The two MAPKs have characteristically different physiological functions and their distinct PPI surface topography enables them to bind different sets of activators and substrates. (*Glatz et al. JBC, 2013).*

LMs usually bind with only medium binding affinity ($K_d \sim 0.1\text{-}10\ \mu M$) to shallow protein-interaction surfaces on their binding partners. The crystallization of proteins in complex with LMs is often challenging because the energy gained upon crystal packing between symmetry mates in the crystal may be on a par with the binding energy of the protein-peptide complex. We presented a rational surface-engineering approach that involves mutating protein surface

residues that are distant from the peptide-binding ERK2 docking groove to alanines. The findings highlight the importance of negative selection in crystal engineering for weakly binding protein-peptide complexes. (*Gógl et al, Acta Crystallogr, 2013*)

In summary, we have made a significant progress in understanding the structural, thermodynamic and kinetic principles of protein-protein interactions involving linear motifs. Besides their theoretical importance our results also allow for selective interfering with linear motif binding, which opens new fields for potential therapeutic applications.