OTKA K108642 Final report.

There is a complex and partially unraveled link between chronic inflammation and cancer. These pathological phenomena arise from disturbances in certain cellular signaling networks. The main goal of our project is to reveal and describe the underlying protein-protein interactions in these networks. Most enzymes, particularly kinases and proteases, have limited specificity, but their complex interaction network results in highly specific signaling systems. Protein-protein interactions allow information transmission over the network by direct and allosteric mechanisms. The major obstacle to developing specific inhibitors of signaling pathways is the lack of detailed knowledge on these interactions.

In connection with a previous phosphoglycerate-kinase-angiogenesis project, our attention turned to another tumor suppressor protein, Rassf1A, which belongs to the Ras protein family. Although Rassf1A is not an enzyme, it influences various cellular processes: apoptosis, cell cycle and microtubular stability among others. There are several proteins that interact with Rassf1A, e.g. Aurora kinase A. This kinase is important in various cancerous processes and is able to phosphorylate Rassf1A, as evidenced by *in vivo* experiments. Rassf1A (Ras-association domain family 1 isoform A), is frequently silenced in a wide range of cancers. However, the mechanism by which Rassf1A exerts its tumor suppressor effects has not been clarified. Rassf1A is involved in three important cellular processes: microtubule stability, mitosis and induction of apoptosis. Thus, loss of function of Rassf1A leads to accelerated cell cycle progression and resistance to apoptotic signals, resulting in increased cell proliferation. Therefore, development of targeted drugs to restore Rassf1A function is of great significance. In order to achieve this, the phosphorylation status of Rassf1A, sensitively influencing its various actions should be controlled. In this OTKA project, our approach was the characterization of the molecular interactions of Rassf1A with its phosphorylating partner, Aurora A, using isolated recombinantly expressed molecules *in vitro*.

Another major target in our OTKA project was Rho-associated protein kinase (ROCK2), a large (1388 amino acids), multidomain, multifunctional protein involved in tumorigenesis, inflammation processes and neurodegenerative disorders. Our goal was to describe its activation mechanism based on structural information and develop allosteric inhibitors to inhibit its action on specific targets including amyloid precursor protein (APP) and beta secretase (BACE1), involved in neurodegenerative diseases

Protein expression and purification. Although our laboratory possesses significant experience in the field of protein-protein interactions and allosteric signal transduction, the present project was an entirely new line of research. The first year – as it was outlined in the work plan - was devoted to the technical establishment of the experimental work with proteins, not produced and studied in the laboratory before. The proposed work in the project is built upon detecting protein-protein interactions by various biochemical and biophysical methods. The protein demand of these experiments is very high regarding both quality and quantity, therefore we had to set up efficient methods to express and purify the protein kinases and their interacting partners. The target proteins are multidomain intracellular proteins, with their size exceeding 1350 residues, as a consequence they are extremely difficult to express.

For optimal production of various kinases, and their interacting partners, we designed a novel, versatile expression platform utilizing four different expression systems, bacterial E.coli, yeast - Pichia pastoris, and two insect cell systems (transiently transfected and stably transfected, using baculovirus, Sf9 cell lines). All expression vectors contain a cleavable N-terminal maltose binding protein (MBP) tag (for better expression and affinity purification), a compatible multicloning site and a C-terminal poly-His tag (to facilitate affinity purification of non-degraded full-length proteins). The expression platform is suitable for expression of all the proteins used in this project. One of the target proteins (ROCK2 kinase) was chosen to test the system. We believe this work could be published in a high-impact biotechnology journal. The manuscript is in preparation. Another line of the expression system optimization study was the development of a bacterial secretion system, where we analyzed the mechanism of the flagellar secretion system. This work was published in BBA Molecular Cell Research.

The following, commercially not available proteins were produced in mg quantities at high purity: Aurora kinases A,B,C (full length and the kinase domains); Rassf1A,C and their truncated variants (both MBP and GST fusion constructs); ROCK1,2 (full length and the kinase domains); LIM Kinase1; cofilin-1; Protein Kinase A; RhoA GTPase; intracellular domains of Alzheimer Precursor Protein (APP) and Beta Secretase. Several proteins were also produced in fluorescently labelled form. To our knowledge, full length recombinants ROCK1/2 were never previously obtained in purified form. These proteins are members of two signaling pathways with a suspected link (LIM Kinase). Homogeneous proteins were produced; the yields were several mgs in E. coli and several hundreds of micrograms in Sf9 insect cells.

Methodological advances. Several new experimental and computational methods were developed to assist the project. These have broader uses and can be utilized in other projects as well.

Novel method for conformational entropy calculation. The proteins we study in this project are multidomain proteins with significant flexibility, and functional changes like phosphorylation or binding involve flexibility changes. To quantitatively describe these changes, we developed a novel method to calculate entropy from conformational ensembles, and demonstrated its applicability to describe changes in interdomain flexibility.

We developed, adapted and applied the following functional assays to check the quality of the purified kinase samples: Phosphorylation of synthetic peptides was monitored by Z'lyte assay; ADP consumption was monitored using ADP-Glo and tested using a coupled enzyme assay containing molar excess of PK, LDH, PEP and NADH. Phosphorylation of natural substrates was monitored by P32-ATP incorporation. Development, selection and adaptation of physical binding assays were completed. The following assays were applied in our protein-protein binding experiments: Pull-down assay (affinity chromatography and Western blot); Microscale Thermophoresis (MST); Surface Plasmon Resonance (SPR); isothermal titration microcalorimetry, differential scanning fluorimetry (DSF); fluorescence polarization (FP);

fluorescence energy transfer (FRET); small angle X-ray scattering (SAXS); analytical gel filtration, native gel electrophoresis.

Detection of pathophysiologically relevant interactions between Rassf1 and Aurora kinase A. It is known that Rassf1A (tumor suppressor) can be phosphorylated by Aurora A, and that Rassf1A activates Aurora A. It was published that Rassf1C (oncoprotein), while a transcript variant of Rassf1A with a different N-terminus, is unable to cooperate with Aurora A. We expressed and purified the full length Rassf1A/C and a truncated mutant named Rassf1ΔN (a variant lacking the N-terminal 119 residues) which contains only their common C-terminal domains. Using radioactive ATP incorporation assay, we demonstrated that not only Rassf1A but also Rassf1C and Rassf1ΔN can be phosphorylated by Aurora A. This is a novel observation; contradicting the previous, generally accepted view therefore further studies should be devoted to the physiological function of this phosphorylation.

We expressed and purified 3 truncated forms of the recombinant human Rassf1A (Fig.1).

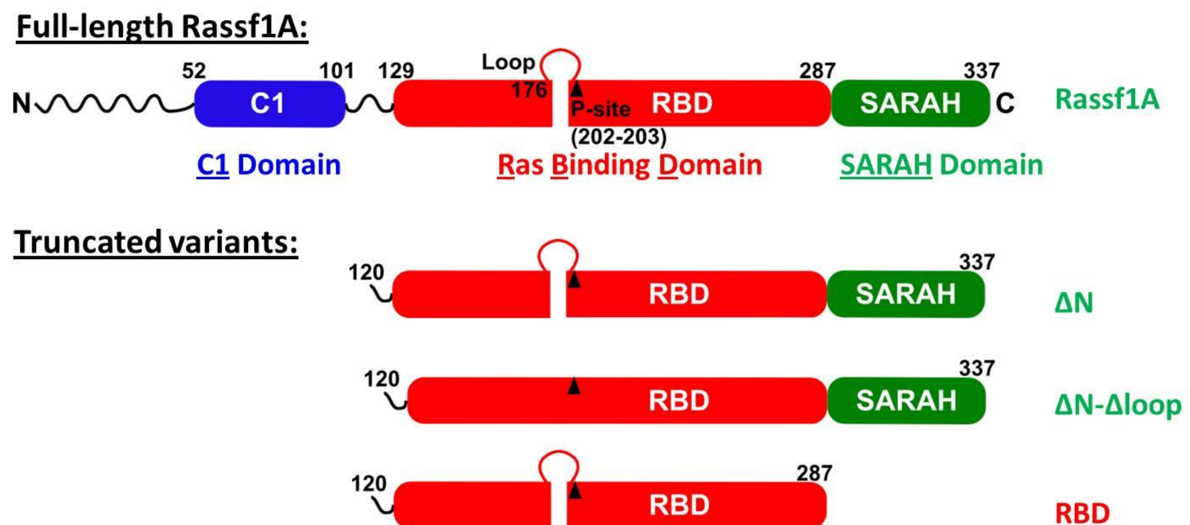

Figure.1. Full length and truncated variants of the Ras-association domain family 1 isoform A (Rassf1A) protein and its truncated variants expressed and studied in the present OTKA project.

The full-length Rassf1A had a strong tendency to aggregate while the truncated forms of Rassf1A, ΔN, and the RBD (RAS-binding domain) that in addition to the N-terminal 119 residues lacks the C-terminal 53 residues (the helical SARAH-domain). This variant had a reduced tendency to aggregate. This supports the prediction based on the sequence that the N-terminal part of Rassf1A is highly disordered under native conditions. We also designed and expressed a loop deletion mutant to test the significance of the disordered loop preceding the phosphorylation site of Rassf1A in its interaction with Aurora A.

To map the different functions into the domain structure of Rassf1A, we measured the kinetics of phosphorylation of the various constructs (Fi.1.) by the kinase domain of AuroraA. The following observations and conclusions were made: 1.The k_{cat} values of phosphorylation of the forms ΔN and RBD are the same within the experimental error. This implies that **the SARAH domain do not poses apparent catalytic role.** 2. The K_m value of the

phosphorylation, however, is significantly increased in the RBD form, lacking the SARAH domain. Thus, **SARAH domain takes part in the binding to Aurora A**. 3.Deletion of the loop near the phosphorylation site (i.e. the Δloop-mutant) largely lowers the k_{cat} value, without significantly affecting K_m. In addition, a synthetic peptide with the same sequence as that of the loop failed to alter the functional interaction between Rassf1A and Aurora A. This suggests that **the loop is important in the catalysis**, but doesn't take part in the physical binding to the enzyme. 4. Based on size-exclusion chromatography experiments, the ΔN variant consists of an equilibrium mixture of monomeric and dimeric states, while the RBD cannot form dimers. This is consistent with the literature, since Rassf1A is expected to homo-dimerize via its SARAH domain, which is absent from the RBD variant. 5. Kinetic analyses of phosphorylation of the ΔN form in a highly extended concentration range by Aurora A exhibits biphasic character, as revealed by an Eadie-Hofstee plot. The higher-concentration range (where ΔN is presumed to exist mostly in dimeric form) yields kinetic parameters very similar to that of the RBD form. Thus, involvement of the SARAH domain in an interaction with another Rassf1A molecule has a similar effect on the phosphorylation as the deletion of this domain. These data suggest that **only the monomeric form of Rassf1A** (due to the specific interaction with its SARAH-domain) **is the fully competent substrate for the enzyme Aurora A.**

Molecular modeling: Since experimental structural information is not yet available, we constructed structural models for the full-length Rassf1A and Rassf1C molecules using protein structure prediction methods to design and interpret our experimental work. Although there is no known experimental structure for the full length Rassf1 variants, but parts of the structure such as the Ras-association domain, the C-terminal SARAH domain and an N-terminal region have homologs of known structure. Since, we have no information on the missing regions and the shape of the full-length structure, we used multiple protein structure prediction schemes to obtain full-length models, including the state-of-the art methods I-TASSER and RBO Aleph. Although considerable uncertainties remain, conclusions can be made based on multiple modeling runs. Models suggest an interaction between the N-terminal part and the Ras-associating domain, forming a relatively compact structure. This interaction does not completely block the site that is phosphorylated by Aurora A but may modify the shape of the surrounding region, thereby influencing binding. Judging by the available complex structure between the Ras-binding domain of RASSF5 and Ras, the binding of Ras proteins occurs by one edge of the main beta sheet of the domain. From the full-length models and the predicted Ras-binding interface, we can narrow down the dimerization interface to a smaller region on the protein surface.

Network modeling. Since the general aim of our research was to identify potential points of intervention into erroneous signaling processes. To identify such points of intervention at systems level, we built a logical signaling network model by combining data from curated databases containing regulatory interactions, posttranslational modifications, 3D structures of protein complexes, and tissue-specific expression data. The logical rules for the model were inferred from the collected data. Performing simulations with this model allows us to

mechanistically study the effect of the removal of Rassf1A or inhibiting its interactions. The logical model was used to reveal possible cross-talks of the Rassf1A pathway with other signaling pathways and adjusting the scope of our research.

Small-angle X-ray scattering and structural model to reconciliate the membrane-distal and membrane-proximal functions of Rho-associated coiled-coil containing protein kinase. To understand the mechanism of ROCK2 activation structural information is necessary, however the full-length structure is unknown. There are two contradicting structural models for the full-length ROCK kinase in the literature. To resolve this contradiction, we performed small-angle X-ray scattering (SAXS) measurements at the EMBL synchrotron facility in Hamburg. Based on the available literature and our preliminary in vitro SAXS data, we constructed a new model for the regulatory mechanism of ROCK2, which is in line with both sets of experimental data. Our experiments indicate that the ROCK2 protein forms neither a stable auto-inhibited loop-like structure nor a fully extended one, but instead it populates a conformational ensemble between those extremes. For the main in vivo membrane distal function of ROCK2, i.e. the phosphorylation of myosin light chain phosphatase (MLCP), the extended conformation is required. The binding of Rho protein may induce a population shift towards the extended state by rigidifying the coiled coil region. SAXS intensities were obtained for full-length dimeric ROCK2 molecule with and without two RhoA subunits bound. The following procedure was used to find conformations compatible with SAXS data.

Coiled-coil segments in the ROCK2 sequence were predicted with MARCOIL, and their structures were built with the CCFold program. The Modeller program was used to construct the structure of the full-length dimer (2x1792 residues), with and without bound RhoA. The geometric simulation program FRODAN was used to obtain a broad sampling of the conformational space of ROCK2. The CRYSOL program was used to generate theoretical SAXS curves.

The models best fitting the experimental SAXS data were selected. In the model for ROCK2 without bound RhoA, the chain is folded in half and an N-terminal domain is in contact with the C-terminal domain. In the model for RhoA-bound ROCK2, the molecule is less compact, and the termini are not in proximity. Analysis of the generated structural ensembles supports the conclusion that RhoA binding causes a population shift towards states where the termini are not in contact.

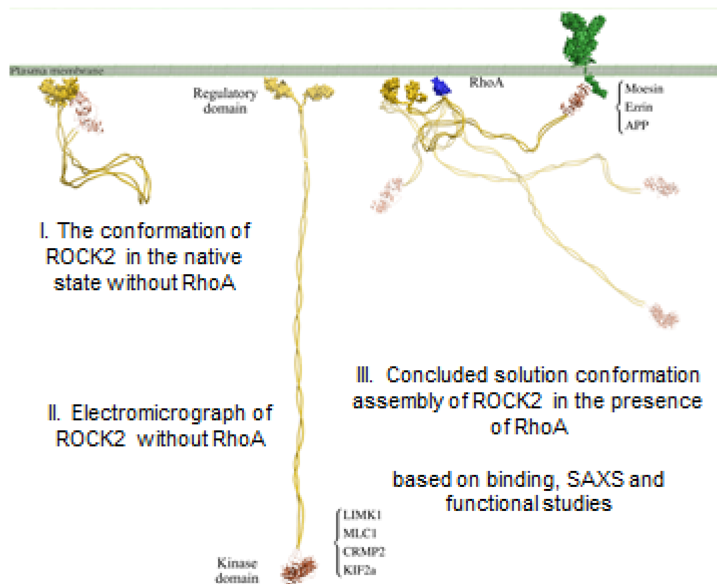

Figure 2. ROCK2 is a large flexible protein. Based on our structural and functional experiments a model was constructed that explains both membrane- distant and proximal functions.

Based on structural and kinetic data we constructed a model that can explain the membrane-proximal and -distal activities of ROCK2 as well. The autoinhibitory structure is not a frozen low energy structure, only a local energy minimum, where an interacting partner can easily exchange the role of the C-terminal. Our results using the full ROCK2 and its kinase domain separately showed clearly that the size of the substrate can make a significant difference in the behavior of the full ROCK2 enzyme. On a small peptide substrate, the full-length ROCK2 had only 1% of the activity of the isolated kinase domain, but on the natural substrate of ROCK2, LIM kinase 1, the activities of the two forms were nearly equal. This unexpected result can be explained as the consequence of a conformational equilibrium governed by flexibility: a big substrate with a large interaction surface can competitively displace the C-terminal domains, but a small peptide cannot. Another interesting result was the inability of RhoA to enhance ROCK2 activity on these substrates, showing that the role of RhoA is not to directly influence the conformational equilibrium between ROCK2 states, but to make possible substrate-driven conformation changes.

Substrate specificities of ROCK1/2. We also characterized the interactions between the proteins of the Rho/ROCK/LIMK/cofilin signaling pathway. We showed that LIMK1 is phosphorylated on Thr-508 only by the ROCK2 kinase but not ROCK1, while LIMK2 is a substrate of only ROCK1. From phylogenetic analysis and structure comparisons of the ROCK1 and ROCK2 kinases, we predicted a site that probably determines substrate specificity, and designed a pair of mirror image mutations to swap the substrate specificities of the kinases towards LIMK1 and LIMK2. The designed mutants were expressed, purified and tested; the experimental findings support the role of this residue in determining the substrate specificity of ROCK1/2.

Interactions between ROCK2 and APP/BACE1. The study of the interaction network of ROCK2 was further broadened in the direction of neurodegenerative disorders, in particular Alzheimer's disease. ROCK2 was shown to phosphorylate Amyloid Precursor Protein (APP) and β-secretase (BACE1) on their intracellular flanking regions. Bioinformatic and physical methods were used to map the interactions between ROCK2 and its target proteins. We found

that BACE1 interacts with ROCK2 at the substrate binding groove (K_d=36 μM) only weakly. In contrast, the APP intracellular domain forms a stable complex with ROCK2 (K_d=25nM), while neither the APP646-664 nor the APP665-695 fragments do. The findings suggest that APP646-664 binds in the substrate binding groove, while APP665-695 binds to an allosteric site, which is mainly hydrophobic as indicated by molecular modeling. We also showed that APP binds to ROCK2 only in the presence of BACE1, indicating that binding of BACE1 induces a conformation change on ROCK2 and the newly formed structure is capable of interacting with APP. This result implies that **ternary interaction is required for the proper phosphorylation of APP which is required for the trafficking of APP to the early endosome.**

To locate the allosteric binding site on ROCK2, we designed and expressed two mutants, V178D and I257D. The results were surprising: the mutations did not affect the binding of APP to ROCK2, but the kinase activity substantially decreased, to 18% (V178D) and 0.01% (I257D) of the wild-type. The mutations affected the ATPase activities to a much smaller extent (2.0 and 0.1-fold). Our results show that allosteric communication between the ATP and substrate binding sites does exist.

A ROCK2 inhibitor could have an anti-Alzheimer's disease potential, but as ROCK2 has several other downstream targets as well, a protein-protein interaction (PPI) inhibitor is needed to block amyloid formation. We tested 300 small-molecule putative PPI inhibitors both for ROCK2 binding using DSF and for PPI inhibition using a fluorescence polarization assay. We identified 11 potential pathway-specific allosteric drug candidates. This finding has attracted the attention of a major pharmaceutical company, and a product-oriented applied research program was launched with their support. Publication of these results is therefore delayed due to the process of securing intellectual property rights.

Protein-protein interactions in the alternative pathway of the complement system. We did continue our systematic studies on regulatory protein-protein interactions in the complement network. The complement system in interaction with other networks plays an important role in inflammation and in the elimination of dangerously altered self-cells (apoptotic cells, cancer cells). An intact complement system is indispensable for maintaining the immune homeostasis; however, uncontrolled, pathological activation of the complement system can contribute to the development of serious disease conditions. The complement system is a proteolytic cascade system which can be activated via three different routes: the classical, the lectin and the alternative pathways. Our aim was to clarify the exact mechanism of complement activation and to find the relevant targets for developing drugs against complement related diseases.

Mannose-binding-lectin (MBL)-associated serine proteases (MASPs) are the enzymatic constituents of the lectin pathway of the complement system. They are complexed with large pattern recognition molecules such as MBL, other collectins, and ficolins. We have made a significant contribution to clarifying the main function of 2 out of the 3 MASPs. We reviled that MASP-1 auto activates first, then activates MASP-2, and finally both participate in the

formation of the C4b2a convertase. This discovery led to the revision of the accepted mechanism of activation of the complement lectin pathway.

There is a third MBL-associated serine protease (MASP-3) whose function has been unknown. In the recent years we revealed the physiological function of MASP-3. It was suggested in the literature that MASP-3 may participate in the activation of the alternative pathway, however the mechanism was unclear. We developed a unique, sensitive assay system to monitor the activation of pro-Factor D (pro-FD) into Factor D (FD) in normal human plasma. FD is the initiation protease of the alternative pathway that circulates in a cleaved form in the resting blood. We recombinantly expressed pro-FD, labeled it with fluorescent dye and followed its conversion by SDS PAGE and cation exchange chromatography. We used our selective and high-affinity inhibitors against MASP-1 (SGMI-1), MASP-2 (SGMI-2) and MASP-3 (TFMI-3) to identify the protease responsible for the conversion. We found that inhibition of MASP-3 completely and permanently blocked the conversion of pro-FD into FD in the normal human plasma. Inhibition of MASP-1 and MASP-2 had no measurable effect. In this way **we proved that MASP-3 is the exclusive activator of FD in the human blood, and showed that the lectin and the alternative pathways of the complement system are deeply connected.** Since only activated MASP-3 is able to cleave pro-FD, we determined the activation status of MASP-3 in the normal human plasma. We found that almost 80% of the MASP-3 molecules are present as cleaved, activated form in the blood. In the course of this work we discovered another important connection between the lectin and the alternative pathways. Inhibition of MASP-1 prevented the activation of the alternative pathway on bacterial lipopolysaccharide (LPS)-coated surface but not on zymosan-coated surface. Since LPS is the main component of the outer cell membrane of Gram-negative bacteria, we also tested the alternative pathway activation on the surface of E. coli cells. We found that inhibition of MASP-1 prevented C3b-deposition on the surface of E. coli. **We can conclude that MASP-1 is important for defense against Gram-negative bacterial infections**. This mechanism may also be important in the case of other surfaces (e.g. biomedical materials, liposomal drugs) provoking adverse, harmful complement response. In case of complement-related diseases we can target MASP-3 to achieve a complete and long-term inhibition of the alternative pathway. If we target MASP-1, we can block the lectin pathway activation generally and the alternative pathway on certain surfaces.

Summary. There is a complex and partially unraveled link between chronic inflammation and cancer. These pathological phenomena arise from disturbances in certain cellular signaling pathways which are integrated into complex networks. For technical reasons only some, well confinable parts can be studied with the hope of success. We picked three related ones: Rassf1 regulation, Rho/ROCK signaling pathway and the complement cascade. Protein-protein interactions allow information transmission over the network by direct and allosteric mechanisms. The major obstacle to developing specific inhibitors of signaling pathways is the lack of detailed knowledge on these interactions. Most enzymes, particularly kinases and proteases, have limited specificity, but their complex interaction network results in highly specific signaling systems. The main goal of our project was to reveal and describe the

underlying protein-protein interactions in these networks, with special emphasize on the allosteric ones.

The following results were achieved and published:

We expressed and purified the proteins required for this study (35 entities) including the challenging full length ROCK2 – 2x1388 amino acids).

Combining reaction kinetics, structural experiments and computer modelling we developed a structure based functional model to reconcile the membrane-distant and membrane-proximal functions of ROCK2. An allosteric binding site was localized and mapped with the aim of inhibiting the formation of the ROCK2-BACE-APP complex, connected with Alzheimer disease.

The AuroraA binding site on Rassf1A was localized and characterized. We have revealed the role of the free SARAH domain (monomeric form of Rassf1A) in the recognition of AuroraA. We have shown that the loop at the RAS binding site is not involved in the binding of AuroraA, but its role is significant in the catalytic step.

We proved that Mannan Associated Serine Protease-3 (MASP-3) is the exclusive activator of Factor D in human blood, revealing that the lectin and alternative pathways of the complement system are deeply connected. We found that inhibition of MASP-1 prevents C3b-deposition on the surface of E. coli, and concluded that MASP-1 is important for defense against Gram-negative bacterial infections.

The proteins we studied in this project are multidomain proteins with significant flexibility, and their functions: phosphorylation or binding involve flexibility changes. To quantitatively describe these changes, we developed a novel method to calculate entropy from conformational ensembles, and demonstrated its applicability to describe changes in interdomain flexibility.

During our basic research we never ignored the potential industrial applicability. Our long lasting collaboration with pharmaceutical and biotech companies helped us in this. We continuously participate in the process of developing recombinant protein expression technologies and analytical methods at the biotechnology division of Richter Gedeon pharmaceutical company. The discovery and mapping of an allosteric site on ROCK2 is a potential target for inhibition of the ROCK2-BACE-APP complex connected with Alzheimer-disease. Together with ELTE Biochemistry Department one startup company "Evolveritas" was launched with the investment of venture capital, for the medical exploitation of our patented highly specific protease inhibitors, and our knowledge on the mechanism of complement activation. The listed results were presented in twenty-seven research papers (IF: 101.822), and one book. Patenting processes have delayed publication some of our results regarding the specific inhibition of certain, medically important protein-protein interactions. One of our discoveries was quoted in a major immunology textbook (Janeway's Immunobiology, 9th edition, 37-76; 2017).