PD-124261 Final Report

INTRODUCTION

The complement system (CS) is an essential part of the innate immune system. It can be triggered by various danger signals through three routes, the classical (CP), the lectin (LP) and the alternative (AP) pathway. The CS is responsible for the elimination of invading microorganisms and dangerously altered self-structures. Improperly regulated CS is involved in pathological processes and the recognized number of such pathologies is steadily growing. Initiator complexes of the CP and the LP consist of a large non-enzymatic multimeric protein that serves as pattern recognition molecule (PRM), which is associated with serine protease (SP) zymogens that are activated upon the PRM binds to its target. For example, the PRM of the CP is C1q, which is associated with 2 C1s and 2 C1r SPs yielding the C1 complex. Complexes of the LP are structurally similar to C1 of the CP, and these two pathways lead to the formation of the same C3 convertase, C4b2a, which cleaves C3 to yield C3b and C3a. The AP has no PRM, and its activator SPs are factor B (FB) and factor D (FD). The AP can spontaneously activated on any surfaces not protected by specific negative regulatory proteins, and it generates a different, C3bBb type C3 convertase. As the AP uses C3b to form a C3 convertase, it serves as an amplification loop for all three pathways. The lectin pathway is composed of a heterogeneous mixture of complexes, each containing one PRM and one or two dimers of associated SPs, MASP-1, MASP-2 or MASP-3, or related non-enzymatic proteins MAp19 or MAp44. At least five different lectin pathway PRMs have been described: MBL, three ficolins, and a heteromer (termed CL-LK) of collectin L1 and collectin K1 (also called collectin 11). While it was soon recognized that MASP-1 and MASP-2 contribute to LP activation, their specific functions were a debated subject, and no function was detected for MASP-3. Our research group developed the first selective inhibitors against each MASP enzyme. This way we deciphered the exact roles of MASP-1 and MASP-2 in LP activation, discovered, that MASP-1 is the exclusive in vivo activator of MASP-2, and also revealed a major function for MASP-3. We demonstrated that MASP-3 is the exclusive FD activator in resting blood, and therefore MASP-3 represents a hitherto hidden link between two CS pathways.

The 3MC (Malpuech-Michels-Mingarelli-Carnevale) syndrome is a rare autosomal congenital disorder characterized - among other features - by facial deformations and mental disability. Mutations of the MASP1 and the COLEC11 gene have been shown to be associated with this syndrome. As explained, the products of both genes are proteins of the lectin pathway. The COLEC11 gene encodes the collectin kidney 1 protein (also called as collectin K1, CL-K1, collectin 11, CL-11) while the MASP1 gene has three splice products, MASP-1, MASP-3 and MAP44. Interestingly, 3MC causing mutations in the MASP1 gene mapped to the SP coding region of MASP-3 and were associated with diminished MASP-3 protease activity. This brought up the possibility, that MASP-3 might have a role in embryogenesis.

The past decade has seen an explosion in zebrafish research. Advantages of this vertebrate model system (small size, external fertilization, optical transparency, great regeneration

capabilities) have made it a prime subject of developmental studies, but recently it has been used with remarkable success to study human pathogenesis too. This has been made possible by the fact that zebrafish is the ideal model organism for high throughput screening of chemical libraries (often seen as the antechamber of successful clinical trials), and by the availability of a wide range of transgenic and mutant strains. Highly efficient transgenesis techniques have been developed, and the revolution in novel genome editing and targeted gene regulation techniques, based on the Clustered Regularly Interspaced Palindromic Repeats (CRISPR) system, was promptly implemented in zebrafish, too. Zebrafish is also an established model of the 3MC syndrome. Both the MASP1 gene (codes only for MASP-3 in zebrafish as it lacks MASP-1 exons) as well as the COLEC11 gene have been silenced with synthetic antisense morpholino oligonucleotides in zebrafish (*Danio rerio*) and in both cases craniofacial abnormalities were detected. These results suggest that zebrafish is a suitable vertebrate model for the 3MC syndrome and could be used for the exploratory study of the mechanism of MASP-3 action in a whole animal.

RESULTS

1. Detecting complement components in zebrafish

1.1. Spatiotemporal expression of complement genes

To investigate the spatiotemporal expression of MASPs and their pattern recognition molecules CL-K1 and MBL2 in zebrafish embryos in situ hybridization (ISH) assays were carried out. Using RNA probes of 800 – 1000 bases in length, high MASP-3 expression level could be detected in the cranial region in 2-day post fertilization (dpf) embryos (Fig. 1). This expression pattern traces out presumably the choroid plexus, which has been implicated as a potential site for interactions between the immune- and the central nervous systems. The cranial expression is almost completely attenuated in 5 dpf larvae, only liver-specific expression can be detected at this stage. Interestingly, CL-K1 expression is restricted to the liver, no cranial expression was detected. MASP-2 and MBL2 expression could not be observed in the investigated embryonal stages. These results suggest that expression site of MASP-3 and its presumed major pattern recognition molecule, CL-K1 do not overlap. Moreover, it seems, that MASP-2 and MBL2 do not play any roles at this period of embryonic development. Most probably these proteins play only immunological roles that are not needed at this embryonic stage.

1.2. Detection of active MASP-3 *in vivo*

A MASP-3 inhibitor, which binds only to activated, as opposed to zymogen MASP-3, could be equipped with a recombinant epitope tag or labeled directly. Such a construct could be used much like a primary monoclonal antibody in various immunoassays. However, there are two challenges. No such inhibitor is known to exist, and no one has ever isolated or expressed any zebrafish MASP enzymes. I decided to clone the genes of zebrafish MASP-2 and MASP-3 catalytic fragments (MASP-cf), express, purify and activate the proteins, and finally, carry out phage display type directed protein evolution experiments using various canonical inhibitor libraries against both enzymes, although with a focus on MASP-3.

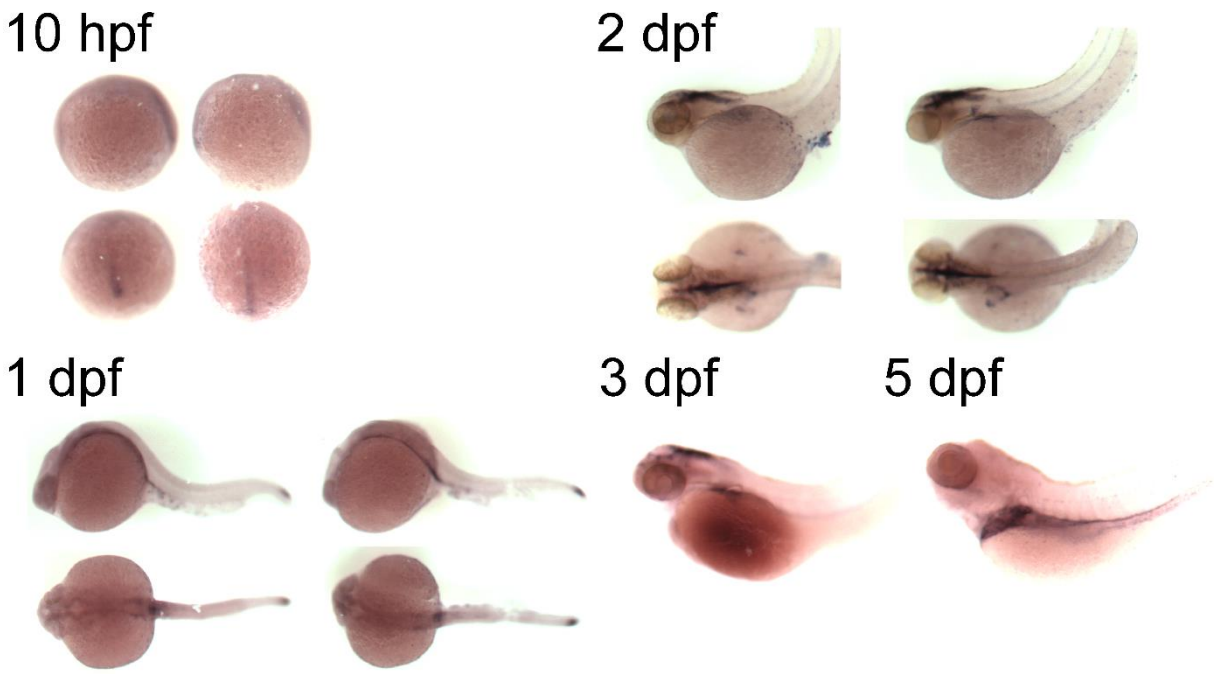

Figure 1. *MASP1* expression in zebrasih emryos and larvae

1.2.1. Producing zebrafish MASP catalytic fragments

Firstly, I worked out and optimized a refolding and purification processes and achieved over 1 mg highly purified protein per 1 L culture. This yield was enough to provide proteins for all the subsequent experiments. Similarly to the human orthologue, zebrafish MASP-2cf auto-activated, i.e. the low activity zymogens activated each other, and the activated forms quantitatively processed all zymogens. On the other hand, zebrafish MASP-3cf, again, just like its human orthologue, did not auto-activate. Thus, the activation, i.e. the specific cleavage of the zymogen had to be worked out. For human MASP-3cf activation, human MASP-1 is used. However, a similar treatment of the zebrafish MASP-3 zymogen resulted in additional cleavage products. Therefore, a different activation procedure had to be established. Based on the conserved double basic Lys-Arg sequence at the activation site of MASP-3, where cleavage after the Arg should occur, I used a proprotein convertase enzyme, KEX2, which is known to be specific for such sites. In fact, by using KEX2, I managed to cleave zymogen zebrafish MASP-3 only at the desired cleavage site (Fig. 2A).

1.2.2. Developing a selective MASP-3 inhibitor

The inhibitors previously developed against human MASPs proved not to be potent enough against the zebrafish enzymes. I conducted phage display experiments to evolve variants that are specific for zebrafish MASP-2 and MASP-3, respectively. For this purpose, I started with protease inhibitor scaffolds SGCI and TFPI as these were good scaffolds for the previous directed evolutions of human MASP-1, MASP-2 and MASP-3. Nevertheless, based on structural considerations I decided to include a third unrelated scaffold as well, eglin C, an

elastase inhibitor of medicinal leech. Display of this inhibitor required a more sophisticated approach as the C-terminus of the inhibitor has a key function and therefore had to remain freely available on the phage-displayed protein. This was achieved by a heterodimer coiled-coil protein fusion system. Since neither the SGCI nor the TFPI library enriched on MASP-3, I selected the eglin C library on MASP-3 resulting in a satisfying enrichment. Based on the sequence logo (Fig. 2B), 4 eglin C variants (eglin C^{PD1-4}) were cloned into a prokaryotic expression vector along with the wild-type protein and a variant bearing a Leu-to-Arg mutation at the P1 site. Note, that each evolved variant carries an Arg at position P1. The inhibitors were produced with high efficiency as recombinant proteins (typical yield was 30 mg/L). According to inhibitory constant (KI) measurements all variants (with the exception of wild-type eglin C) bind to zebrafish MASP-3 with a relatively high affinity (4 – 20 nM). Importantly, these variants do not inhibit zebrafish MASP-2 at all (Fig. 2C-D).

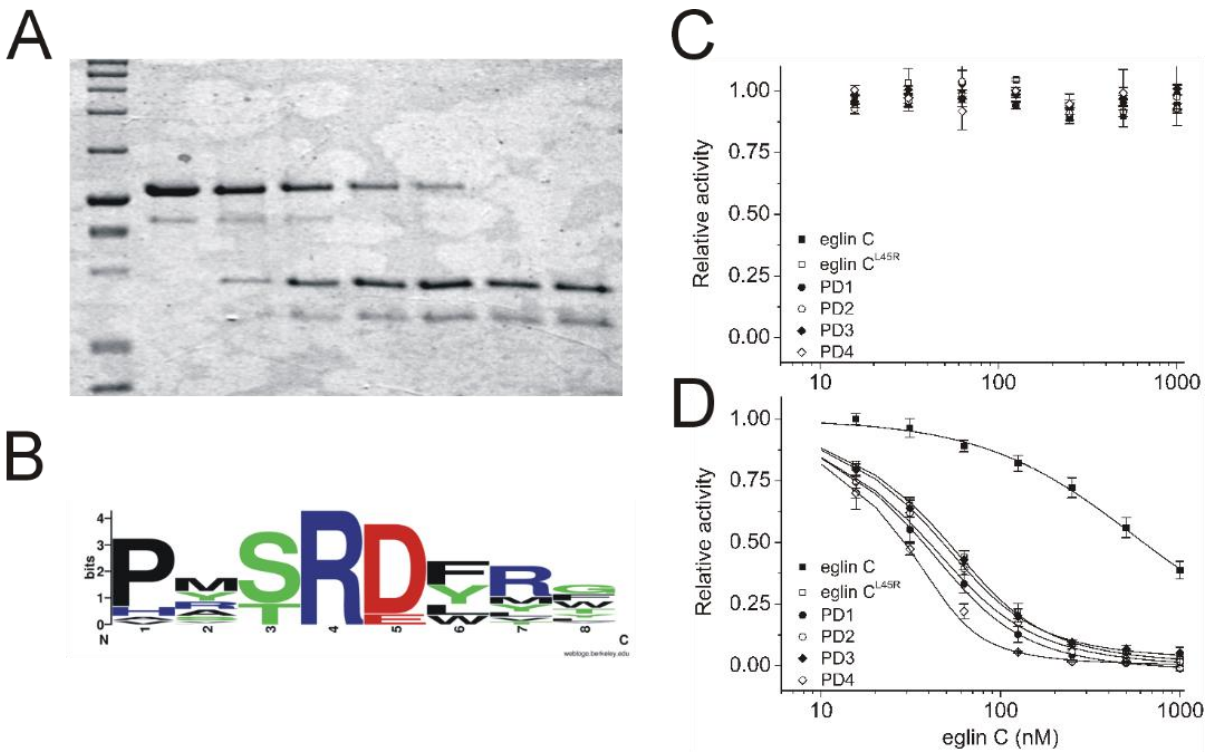

Figure 2. Development of potent zebrafish MASP3 inhibitors (A) Activation of MASP-3cf; (B) the sequence LOGO of the selected eglin C clones; (C-D) Inhibiton of zebrafish MASP-2cf and MASP-3cf by the evolved eglin C variants

1.2.3. Fluorescent detection of active MASP-3 *in vivo*

To detect the spatiotemporal localization of active MASP-3 in zebrafish embryos we used the method of immunofluorescence staining with Máté Varga at Department of Genetics (ELTE). For that purpose I produced eglin C inhibitors bearing an HA epitope tag at their N-terminus. The inhibitors were microinjected in zebrafish embryos at various developmental stages, then their localization (or "pharmacokinetics") were followed by immunofluorescence using a DyLight488-conjugated anti-HA antibody. According to our results, the inhibitors, that were microinjected in the yolk sac of 1-cell stage embryos, were detected in embryonic cells 4 hours post fertilization (hpf). However, at 1 day post fertilization (dpf) stage the HA-tagged inhibitors could not be detected. Note that we demonstrated by *in situ* hybridization (ISH) experiments that MASP-3 expression is most pronounced at 2-4 dpf stage, thus an alternative way of administration had to be developed. Next we microinjected embryos at 1 dpf stage (into the yolk sac, next to the heart tube), then the embryos were fixed at 2 dpf stage and stained. To determine whether the staining is specific for the given eglin C variant developed against zebrafish MASP-3 control experiments were carried out using HA-tagged non-binding inhibitor variants. For this purpose we used TFMI-3-HA (the inhibitor evolved previously against human MASP-3) as a negative control, since TFMI-3 does not inhibits zfMASP-3 *in vitro*. The localization of the signal developed using HA-eglin C^{PD4} and DyLight488-conjugated anti-HA antibody was highly reminiscent of the MASP-3 ISH staining, however, it seemed not to be specific, as TFMI-3-HA showed very similar localization. To overcome this problem, I developed an alternative, direct mode of detection as follows. Eglin C, as a member of potato type I serine protease inhibitor family, has a unique property among canonical serine protease inhibitors that it does not contain stabilizing disulfides. Thus, the introduction of a well-positioned unpaired cysteine residue in eglin C probably does not interfere with protein folding. I produced Ser5Cys variant of wild-type and zfMASP-3-specific eglin C^{PD4} in *E. coli*, purified and selectively labelled them at Cys5 with a molar excess of 5-iodoacetamido-fluorescein. After separating the conjugates from the non-reacted dye, labelled inhibitors were injected into 1 dpf zebrafish embryos. Interestingly, inhibitor localization differed from that detected using immunofluorescence staining. The cranial region was either not stained or weakly stained by both eglin C variants, while high fluorescence was detected in the pancreas. This latter phenomenon can be explained by the fact that the pancreas is abundant in digestive proteases with low selectivity, which bind both $eglinC^{wt}$ and eglin C^{PD4} with Met and Arg residues at the P1 position, respectively. Importantly, none of the eglin C variants (HA-tagged or fluorescently labelled) induced phenotypical changes in the zebrafish embryos.

2. Knock-out of MASP1 and COLEC11

To perform MASP-3 serine protease domain knock-out, several guide RNAs were synthesized in-house and microinjected in complex with Cas9 endonuclease into 1-cell stage embryos. We successfully carried out DNA cleavage in the 5' region of the *MASP1* gene exon 12 coding for the MASP-3 serine protease domain. The embryo pool, injected with the effective guide RNA / Cas9 mixture, were grown until reaching reproductive maturity. To investigate whether the fish carry the mutation in germ line and to characterize the *MASP1* mutations they were crossbred with wild-type animals. Genotyping the resulting embryos revealed that an 8 bp deletion occurred with the highest probability. This frameshift mutation leads to an early STOP

codon within 80 bp (Fig. 3A) resulting in a *MASP1* transcript lacking an intact serine protease domain (i.e. 85 % of the SP is absent) (Fig. 3B). Note that if this truncated protein is not eliminated by quality control mechanisms, it is capable to bind to PRMs such as CL-K1 through the first three non-catalytic MASP-3 domains (CL-K1 mutations are also associated with 3MC syndrome.) The animals (all heterozygotes), which carried this 8 bp deletion were crossed, progenies were grown until reaching reproductive maturity. *Strikingly, the homozygote mutant animals was phenotypically normal!*

This finding is clearly unexpected, for two reasons. One is that in zebrafish knockdown experiments using antisense morpholinos that aimed to prevent MASP-3 mRNA translation resulted in severely impaired embryogenesis. These morpholinos were directed against the *MASP1* initiation site and the exon 3 – intron 3 splice site and were predicted to entirely prevent any kind of MASP-3 production, i.e. not even a truncated form should have been produced. If so, the most plausible interpretation would be that the first five non-enzymatic domains of MASP-3 are fully responsible for the developmental role of MASP-3. On the other hand, it would entirely contradict the human results: in unrelated families having 3MC syndrome the patients were shown to carry homozygous mutations rendering their MASP-3 serine protease domain inactive. In this case the full-length protein is produced, but it is catalytically inactive. To make the case even more complicated, very recently knockout mice were produced, in which either the MASP-1 specific exon or the MASP-3 specific exon was selectively deleted. *The surprising result was that none of these mice had 3MC-type severe developmental disorders!*

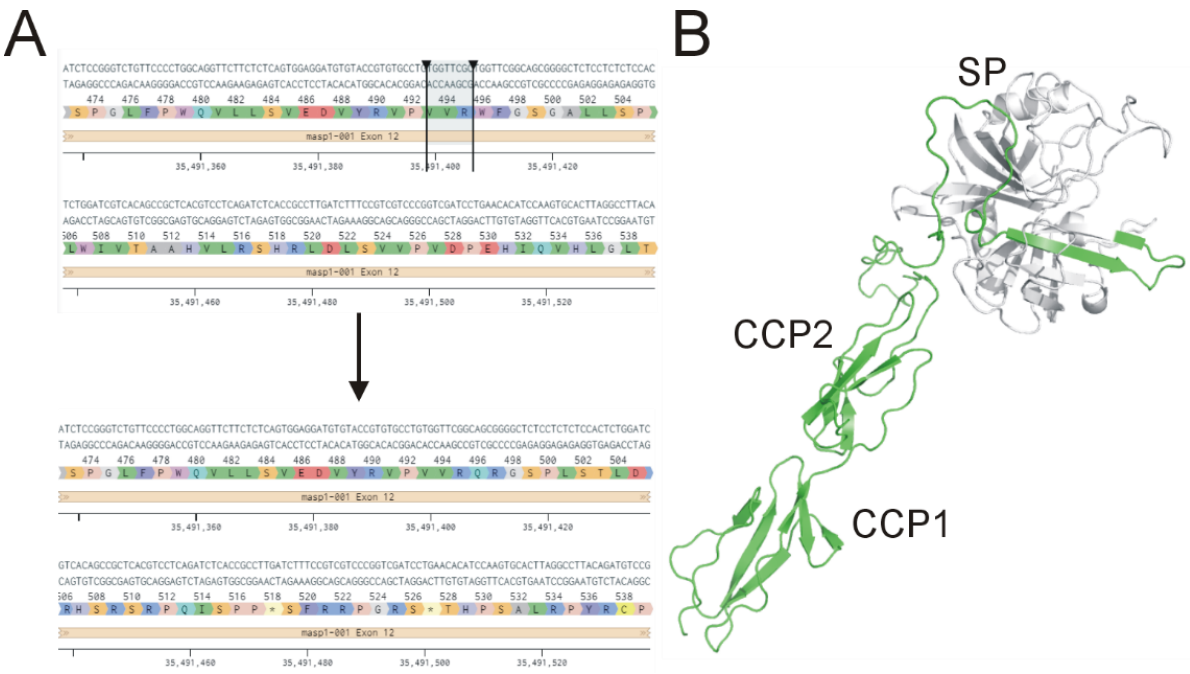

Figure 3. Knock-out of MASP-3 SP domain. (A) Cas9 cleavage occurring in the upstream region of the SP domain coding exon 12 results in an 8 bp deletion, which leads to premature termination. (B) The untranslated region of MASP-3 fragment is shown in grey in the 3D structural model of zebrafish MASP-3cf (green).

While at first our negative results were somewhat discouraging, now we consider them as important discovery. These results provide and independent second proof that in spite of previous suggestions, MASP-3 has no evolutionary conserved function in vertebrate embryonic development.

The inevitable question is why the previous morpholino-based zebrafish experiments have led to an opposite result. One possibility is, that in spite of various controls, in those morpholino experiments the treatment was not entirely specific and some kind of off-target effect caused the observed phenotypes. The morpholino based approach is known to be prone to such off-target artifacts.

After getting negative results in two vertebrate species with highly specific MASP-3 KO, even the human results should be thoroughly revisited. One should be aware that only a few families have been diagnosed with 3MC syndrome. This raises the possibility that the observed MASP-3 and COLEC-11 mutations are necessary but not sufficient for 3MC syndrome and in fact are linked to some yet unidentified additional genetic abnormalities.

FUTURE PLANS

To better understand the apparent discrepancy observed in the zebrafish model, we decided to achieve complete MASP1 and COLEC11 gene knockouts. We already started the experiments and successfully disrupted both genes. The selection of the mutants is in progress.

If knocking out the entire *MASP-1* gene still does not lead to any phenotype, it would suggest that the morpholino approach was compromised by some artifacts and the 3MC syndrome is either not directly linked to the lack of the MASP-3 protein, or this lack is needed to be combined with some other genetic defect.

However, if it recapitulated the findings of the morpholino experiments, it would suggest that at least in zebrafish, some of the non-enzymatic domains would carry the developmental function of MASP-3. This, however, would suggest that the human and the zebrafish system would be more related and the mouse system would be an outlier, a rather improbably outcome.

In all, it seems that more investigations are needed in the case of all three models to reveal the causative genetic background of the 3MC syndrome in humans and the exact (if any) developmental roles of MASP-3 in the various vertebrate species.