From The Department of Biosciences and Nutrition
Karolinska Institutet, Stockholm, Sweden

A CRYSTALLOGRAPHIC VIEW OF EGG-SPERM INTERACTION

Hamed Sadat

Karolinska Institutet

Stockholm 2016

Published by Karolinska Institutet.
Printed by E-Print AB 2016

ISBN 978-91-7676-372-8

A Crystallographic View of Egg-Sperm Interaction

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Hamed Sadat

Principal Supervisor:
Professor Luca Jovine
Karolinska Institutet
Department of Biosciences and Nutrition

Co-supervisor:
Professor Henrik Garoff
Karolinska Institutet
Department of Biosciences and Nutrition

Opponent:
Professor Graham Wallis
University of Otago, New Zealand
Department of Zoology

Examination Board:
Professor Adnane Achour
Karolinska Institutet
Department of Medicine, Solna

Professor Rudolf Ladenstein
Karolinska Institutet
Department of Biosciences and Nutrition

Dr David Drew
Stockholm University
Department of Biochemistry and Biophysics

I dedicate this thesis to my mother and father

ABSTRACT

Animal reproduction is a heavily studied biological event, however our knowledge regarding the molecular basis of the recognition between gametes - egg and sperm - still remains very incomplete. In particular, it is currently unclear whether protein epitopes important for fertilization have been conserved among different species during evolution, and why gamete interaction proteins generally are subjected to strong positive Darwinian selection. Furthermore, there has not been a reported X-ray crystal structure of an egg-sperm protein complex, in any organism, prior to sperm penetration.

The aim of this thesis is to shed light on these important biological questions, by applying structural biology techniques to a selection of key reproductive proteins ranging from invertebrate to vertebrate.

In PAPER I, it was hypothesized that, sperm binding region in yeast, invertebrate and vertebrate adopts the same three dimensional fold as egg zona pellucida (ZP) protein ZP3, which has been suggested to be involved in mediating binding between gametes in humans. This also led us to the hypothesis that the N-terminal region of ZP2 might contain three ZP-N domains. Based on this theory, despite the lack of sequence similarity, we suggest that the 22 tandem repeats observed in the invertebrate egg coat protein vitelline envelope receptor for lysin (VERL), as well as a protein domain in yeast, essential for mating, also adopts the same ZP-N fold. This suggests that, even though being separated by 0.6-1 billion years of evolution, these reproductive proteins all share a common structural fold important for egg-sperm recognition.

In PAPER II, we facilitate protein expression by using an approach to obtain milligram recombinant protein transiently expressed in mammalian cells, either as a rigid or cleavable fusion with a mammalianized version of bacterial maltose-binding protein (mMBP). We show that mMBP can increase protein expression up to 200-fold and also assist crystallization of target proteins, which did not crystallize by themselves. Furthermore, we show that mMBP can also be used during structure determination as a search model for molecular replacement, highlighting the diversity of mMBP, not only for increasing protein expressing but also guiding crystallography.

In PAPER III, we show that invertebrate egg coat protein VERL, consisting of 22 tandem repeats, previously suggested to be ZP-N domains (paper I), shares a similar 3D structure as mammalian egg coat protein ZP2. By solving the crystal structures of VERL repeats and mammalian egg coat protein ZP2 ZP-N1, we show that these reproductive proteins share a common 3D fold, proving that this specific immunoglobulin-like fold has been conserved during evolution. Furthermore, we identified the binding region between VERL and lysin, while at the same time suggesting why VERL repeats are under positive selection. We also solved the crystal structure of the VERL/lysin complex. This structure is not only the first structure of an egg-sperm protein complex on the egg coat, but also further elucidates the role of positive Darwinian selection on these reproductive proteins. Finally, our structures

also suggest a very simple mechanism of how lysin-VERL binding creates a hole by splaying apart the VERL filaments allowing the sperm to enter the egg and fuse with the plasma membrane.

In PAPER IV, the crystal structure of the mammalian egg protein Juno, known previously as folate receptor (FR) 4, was solved. Previous data showed that Juno is crucial for fertilization, since female mice lacking Juno are infertile. Although our structure shows a similar fold as FRα and FRß, Juno is not able to bind vitamin B_9/folic acid. One of the reasons why Juno is not able to bind vitamin B_9 might be due to high flexibility within the folate binding region. Even though Juno has lost the ability to bind vitamin B_9 it has gained the ability to bind sperm, via sperm protein Izumo1.

LIST OF SCIENTIFIC PAPERS

I. Willie J. Swanson, Jan E. Aagaard, Victor D. Vacquier, Magnus Monné, **Hamed Sadat Al Hosseini**, Luca Jovine. *The molecular basis of sex: linking yeast to human*. Mol Biol Evol. 2011 Jul;28(7):1963-6. doi: 10.1093/molbev/msr026. Epub 2011 Jan 31.

II. Marcel Bokhove[*], **Hamed Sadat Al Hosseini**[*], Takako Saito[*], Elisa Dioguardi, Katharina Gegenschatz-Schmid, Kaoru Nishimura, Isha Raj, Daniele de Sanctis, Ling Han, Luca Jovine. *Easy mammalian expression and crystallography of maltose-binding protein-fused human proteins*. J Struct Biol. 2016 Apr;194(1):1-7. doi: 10.1016/j.jsb.2016.01.016. Epub 2016 Feb 3.

III. **Hamed Sadat Al Hosseini**[*], Isha Raj[*], Elisa Dioguardi, Kaoru Nishimura, Ling Han, Daniele de Sanctis, Luca Jovine.
Structural basis of egg-sperm recognition at fertilization.
Manuscript

IV. Ling Han[*], Kaoru Nishimura[*], **Hamed Sadat Al Hosseini**, Enrica Bianchi, Gavin J. Wright, Luca Jovine. *Divergent evolution of vitamin B_9 binding underlies Juno-mediated adhesion of mammalian gametes.* Curr Biol. 2016 Feb 8;26(3):R100-1. doi: 10.1016/j.cub.2015.12.034.

* Co-first authors

CONTENTS

1 INTRODUCTION 1
1.1 Introduction to fertilization 1
1.1.1 Evolution of reproductive proteins 2
1.1.2 Why abalone as a model organism? 3
1.2 Chicken or the egg? 4
1.2.1 Sperm acrosome reaction 5
1.2.2 Sperm protein lysin 6
1.2.3 ZP domain 7
1.2.4 Egg coat protein VERL 9
1.3 VERL - lysin binding and dissolution of the VE 9
1.3.1 Fusion to the plasma membrane 11
2 AIM AND SIGNIFICANCE 14
3 METHODOLOGY 15
3.1 Protein expression 15
3.2 Protein purification 16
3.3 Protein crystallization 18
4 RESULTS AND CONCLUSIONS 19
4.1 Paper I 19
4.2 Paper II 20
4.3 Paper III 20
4.4 Paper IV 21
5 FINAL CONCLUSIONS AND FUTURE PROSPECTIVE 23
5.1 MBP in protein expression and crystallography 24
5.2 Bridge between invertebrate and vertebrate fertilliztion 24
6 ACKNOWLEDGEMENTS 27
7 REFERENCES 29

LIST OF ABBREVIATIONS

VERL	Vitelline envelope receptor for lysin
MBP	Maltose-binding protein
mMBP	Mammalian maltose-binding protein
SEC	Size exclusion chromatography
IMAC	Immobilized Metal Affinity Chromatography
ZP	Zona Pellucida
ZP1	Zona Pellucida protein 1
ZP2	Zona Pellucida protein 2
ZP3	Zona Pellucida protein 3
ZP4	Zona Pellucida protein 4
VE	Vitelline envelope
MR	Molecular replacement
PDB	Protein data bank
AR	Acrosome reaction
SP	Signal peptide
AVEXIS	Avidity-based extra cellular interaction screen
VEZP14	Vitelline Envelope Zona Pellucida protein 14
C	Cysteine
PEI	Polyethylenimine
FR	Folate receptor
NTR	N-terminal region

1 INTRODUCTION

1.1 Introduction to fertilization

The ability to reproduce is vital for the survival of a species. All sexually reproducing organisms depend on the process of fertilization, which is the union of egg and sperm, resulting in a zygote. There are two main classifications of fertilization: internal fertilization and external fertilization. In vertebrates, such as mammals, internal fertilization is achieved by copulation, where sperm and egg must interact in a species-specific manner (Edward C. Metz, 1994; Wassarman et al., 2001). Even though this is a crucial stage in life and a popular topic for scientific studies, our knowledge regarding this process still remains unclear at the molecular level. For example, we still have not been able to identify a key sperm protein being directly involved in binding to the extracellular matrix that covers the mammalian egg, called the zona pellucida (ZP). Prior to egg interaction the sperm must undergo a process called capacitation (Austin, 1952), which involves several biochemical modifications that allow the sperm to bind to the ZP (Bailey, 2010). During the maturation of sperm it must also undergo acrosome reaction (AR), a process where sperm releases the content in the sperm head to be able to penetrate the ZP (Florman and First, 1988). AR occurs after the capacitation process (Bailey, 2010). Once AR has occurred sperm comes in contact with the ZP. In mouse the ZP is made up of three glycol proteins, ZP1, ZP2 and ZP3 (**Figure 1**). While ZP1 has shown to be involved in supporting the structural integrity of the ZP (Rankin et al., 1999; Huang et al., 2014), ZP2 and ZP3 have been purposed to act a sperm receptors (Bleil and Wassarman, 1980; Wassarman and Litscher, 2001; Wassarman, 2008). Furthermore, ZP2 has also been suggested to be involved in block to polyspermy. However, it is still unclear which of ZP2 and ZP3 is the most crucial protein for the initial binding process.

The fertilization process occurs very differently in marine invertebrate species that use external fertilisation to reproduce. This event occurs by spawning the gametes into open sea water where fertilisation occurs. In the case of the marine invertebrate abalone several species reproduce during the same time and at the same location, yet cross fertilisation is very rare (Vacquier et al., 1990; Vacquier and Lee, 1993; Shaw et al., 1995). Abalone sperm also undergoes the process of AR, once the sperm has been released and comes in contact with the vitelline envelope (VE), which corresponds to the mammalian ZP. The egg sperm recognition event in both vertebrates and invertebrates is believed to be species specific. It is believed that one reason why reproductive proteins are amongst the fastest evolving proteins is that they have to interact in a species specific way (Swanson and Vacquier, 2002). One force which is believed to act upon these proteins is suggested to be adaptive evolution (Wolfner, 1997; Hellberg and Vacquier, 1999; Yang and Bielawski, 2000). Since reproductive proteins are fast evolving, and the fact that there is roughly

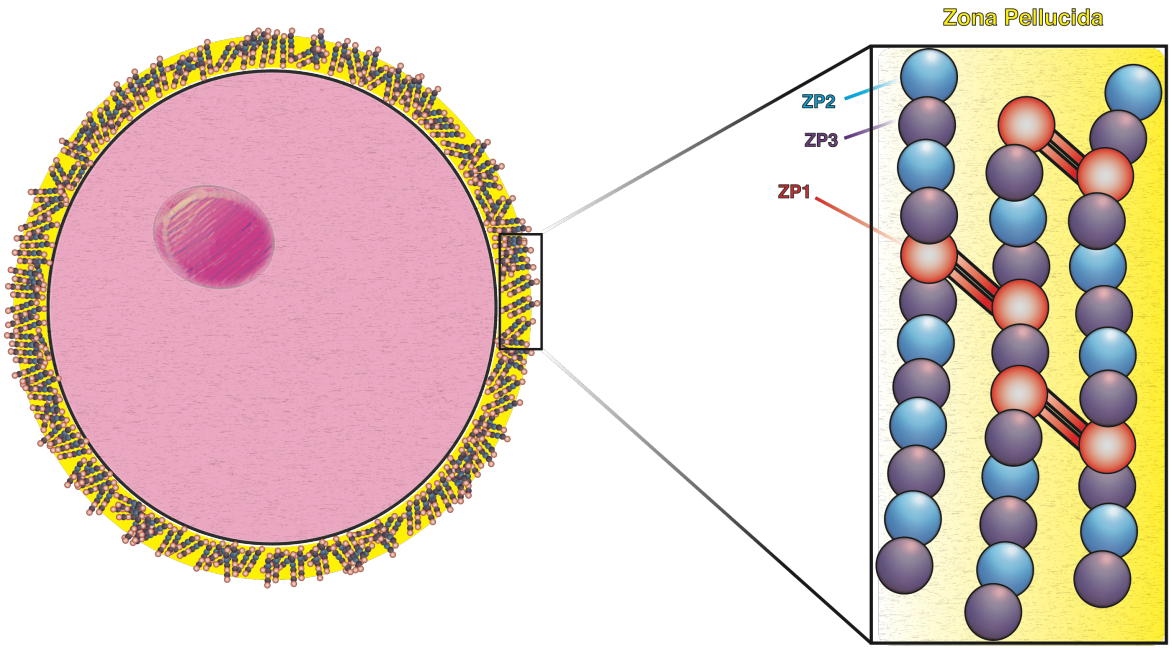

Figure 1. Arrangement of the ZP glycoproteins that form the extracellular matrix covering the egg. The mammalian egg is covered by either 3 or 4 glycoproteins (ZP1-3/ZP1-4), depending on the species, which forms the ZP. This figure is a represents mouse ZP, which is composed of ZP1, ZP2 and ZP3, while mouse ZP4 gene has shown to be a pseudo gene (Goudet et al., 2008). However, human ZP contains ZP1-3 as well as ZP4 (Bauskin et al., 1999; Wassarman, 2008). After penetrating the ZP, sperm fuses with the plasma membrane through Izumo1 (named after a Japanese marriage shrine), which binds to egg protein Juno (named after the roman goddess of fertility) (Bianchi et al., 2014; Wassarman, 2014; Han et al., 2016).

600 million years of evolution between invertebrates and vertebrates, it is difficult to predict if these reproductive systems share any related features during fertilisation. We wanted to focus on well characterised key reproductive proteins involved in abalone fertilisation to be able to better understand some of these essential biological questions, and to determine if there is a bridge which would connect invertebrate and vertebrate fertilisation.

1.1.1 Evolution of reproductive proteins

Why are some genes evolving faster than others? What are the selective forces behind this phenomenon? These are some interesting questions which are not only of interest for evolution biologist but for the entire filed of biology. It has been suggested that there are several forces involved in the rapid evolution of reproductive proteins such as sexual selection, sperm competition and sexual conflict, either as an individual event or in combination (Swanson and Vacquier, 2002). Rapid evolution is not only limited to genes involved in reproduction, but also genes that have a role in the immune system (Swanson and Vacquier, 2002). For instance, the genes that encode for human leukocyte antigens are among the fastest evolving genes in the human

genome (Hertz et al., 2011). The authors suggest that one reason behind this might be because these genes constantly must adapt to rapidly evolving pathogens (Hertz et al., 2011).

One of the forces that is involved in driving rapid evolution of reproductive proteins is adaptive evolution, which is a change that occurs in the gene, allowing the organism to better adapt to the existing environment (Hawks et al., 2007). In *Drosophila* it has been shown that adaptive evolution is the driving force contributing to pushing reproductive tissues to diverge twice as fast as non-reproductive tissues (Swanson and Vacquier, 2002). It has been shown, by analyzing 750 protein-protein interactions, that interface regions of proteins involved in different biological functions are mostly conserved in comparison to exposed areas (Von Eichborn et al., 2010). Since reproductive proteins are driven by positive selection (Swanson and Vacquier, 2002), a selective force that is beneficial for the fitness of the organism (Morgan et al., 2010),it makes them very interesting targets for studying evolution.

1.1.2 Why abalone as a model organism?

The main reason why abalone has been used as a model organism for studying the process of fertilization is that abalones produce large amount of gametes, which are easy to harvest in the laboratory by using hydrogen peroxide to induce the release of the gametes (Morse et al., 1977). Understandably, the study of mammalian fertilization is much more attractive, but there are much more obstacles for researchers when using a mammalian model organism, such as cost, availability of gametes and ethical restraints. Since we use recombinant proteins in our studies we would not have the above mentioned complications. Even though we also studied proteins in the mammalian system (performed by PhD student Elisa Dioguardi) for our structural studies there was some uncertainties regarding egg-sperm protein complex in the mammalian system. Therefore, we choose to study the proteins involved in egg-sperm recognition in abalone too, which will also reveal if there is an evolutionary link between invertebrate and vertebrate reproductive protein.

At the start of this project, there were several mammalian sperm protein candidates displaying egg coat binding capabilities during fertilization (Tanphaichitr et al., 2007). One of these proteins was sp56 (Cheng et al., 1994; Foster et al., 1997; Buffone et al., 2008), which was initially suggested to bind to the egg coat protein ZP3 (Bookbinder et al., 1995; Buffone et al., 2008). Although extensive binding trials, between sp56 and ZP3, were performed in our lab, no evidence of a protein complex between these proteins was detected in vitro (Elisa Dioguardi personal communication). Furthermore, in 2012 it was shown that sp56 knockout mice still were able to fertilize the egg, suggesting that the role of sp56 is not essential (Muro et al., 2012). This is just one example of a sperm protein candidate that was thought to be directly involved in binding to the ZP glycoproteins located on the egg, but was proven not to be crucial during this event. This highlights the complexity and difficulty to

identify a putative binding partner during this step of fertilization in vertebrates. One of the very few systems where such a complex, at this stage of fertilization, has been well characterized is the abalone (genus *Haliotis*) system, which is the main reason that we chose to use abalone as a model organism for structure determination of a gamete recognition complex. Upon reaching the egg, the abalone sperm undergoes AR, a process which will be described more in detail later, which causes sperm to release protein lysin, one of the two major acrosomal proteins in abalone sperm (Lewis et al., 1982). Lysin is responsible for dissolving the VE, the extracellular matrix that covers the egg. Abalone VE (Lewis et al., 1982) is composed of roughly 30 different subunits (Aagaard et al., 2010). One of these subunits have been identified as vitelline envelope receptor for lysin (VERL), a glycoprotein with a molecular weight of two million Dalton (Shaw et al., 1993).

1.2 Chicken or the egg?

Chicken or the egg dilemma questions what actually came first, the chicken or the egg? Aristotle stated the following: “If there has been a first man he must have been born without father or mother – which is a repugnant to nature. For there could not have been a first egg to give a beginning to birds, or there would have been a first bird which gave a beginning to eggs; for a bird comes from an egg” (Fénelon, 1825). So what does this more than 2000-year-old question have to do in relationship with abalone reproductive proteins VERL and lysin?

Sequence alignment of lysin between 25 different abalone species revealed a hypervariable region at the N-terminus that is significantly different among the various species. Furthermore, it was also determined that lysin contains 23 residues that are under positive selection (Yang et al., 2000). It has been suggested that the N-terminal region (NTR) (residues 1-12) as well as the positively selected residues play a crucial role during the species-specific binding to VERL (Lee et al., 1995; Kresge et al., 2001a; Mintseris and Weng, 2005). To better understand the involvement of lysin regions during VE dissolution, so called dissolution assay experiments were performed using lysin from different abalone species (Lyon and Vacquier, 1999). It was shown that by adding purified recombinant lysin from pink abalone to isolated egg VE from red abalone the dissolution of the red VE is not completed, compared to the control, which was red lysin. Furthermore, the dissolution of red VE was decreased by half when replacing the NTR of red lysin to pink lysin. This data suggested that the hypervariable region of lysin has some implication during the VE dissolution process (Lyon and Vacquier, 1999).

It is not only lysin that is under positive selection, also VERL has been suggested to be under positive selection (Swanson and Vacquier, 1998; Aagaard et al., 2010). The identification of positive selection in VERL was more difficult to detect since there are not as many sequences of VERL from different abalone species available, in comparison to lysin. By using maximum likelihood analysis it was shown that VERL

repeats 1 and 2 are under positive selection, while repeats 3-22 were suggested to be under concerted evolution (Swanson and Vacquier, 1998; Galindo et al., 2003). It has been suggested that VERL repeat 1 and 2 are under positive selection since they are involved in the species specific binding to lysin, while repeat 3-22 are involved in the non-specific interaction (Galindo et al., 2002).
If we would link VERL and lysin to the chicken and the egg question, the question here would be, which of these reproductive proteins are adapting to each other? Is sperm protein lysin adapting to the egg protein VERL or is VERL adapting to lysin? Or are they both adapting to each other at the same time? There are some data suggesting that in fact these reproductive proteins co-evolve, meaning that when one molecule changes the other one also changes at a comparable rate (Clark et al., 2009).

1.2.1 Sperm acrosome reaction

A vital process during sperm maturation is the AR, which takes place during gamete interaction, which allows sperm to penetrate the egg coat and fuse to the plasma membrane (Bleil and Wassarman, 1983; Wassarman, 1990; Yanagimachi, 1994) (**Figure 2**).

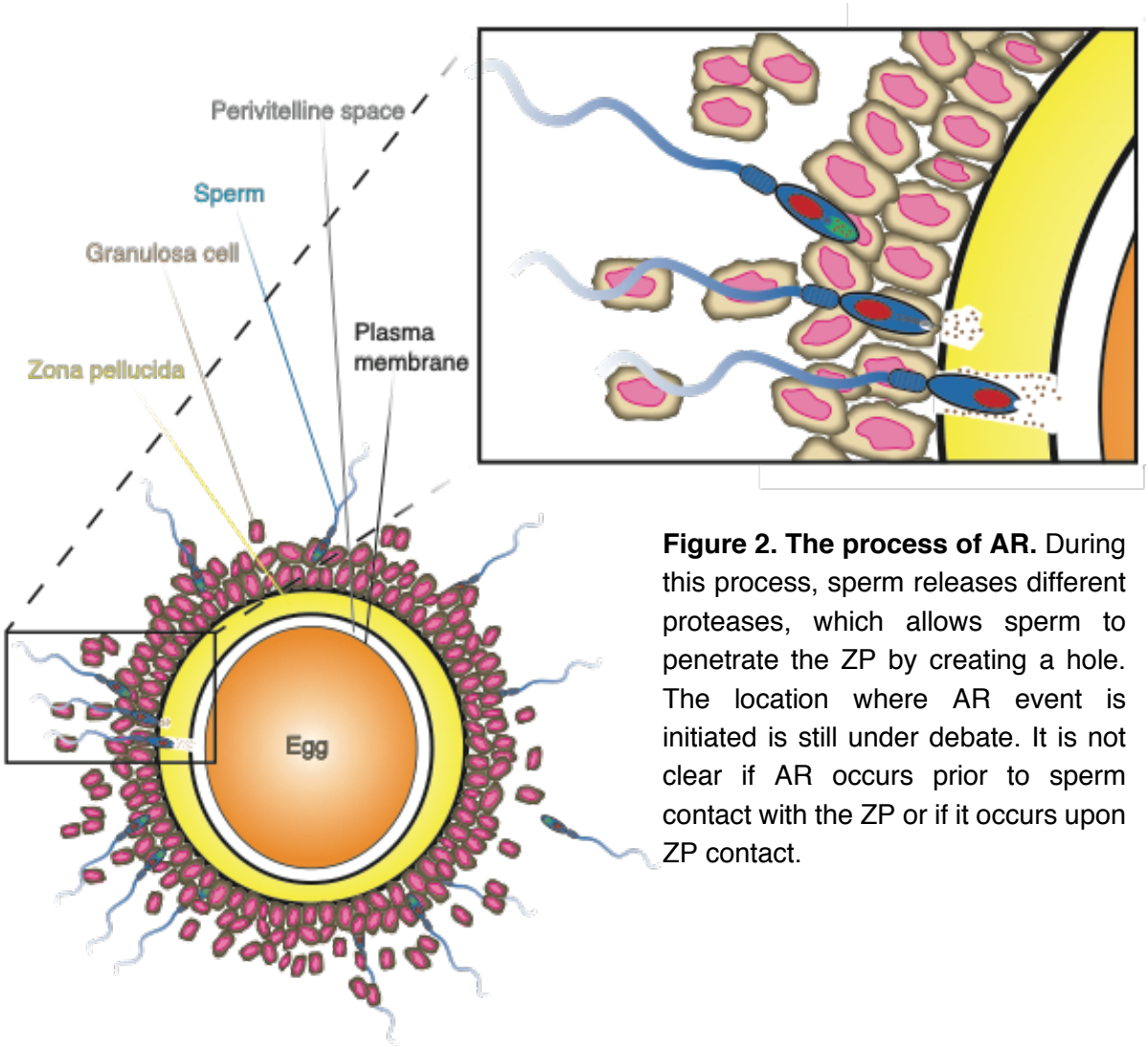

Figure 2. The process of AR. During this process, sperm releases different proteases, which allows sperm to penetrate the ZP by creating a hole. The location where AR event is initiated is still under debate. It is not clear if AR occurs prior to sperm contact with the ZP or if it occurs upon ZP contact.

In mouse, it was shown that egg coat protein ZP3 is responsible for initiation of this event (Bleil and Wassarman, 1983; Beebe et al., 1992), which is Ca^{2+}-dependent (Brucker and Lipford, 1995; Yanagimachi, 2011), and only acrosome reacted sperm is able to penetrate the ZP. During AR the inner acrosomal membrane is exposed and the inner content of the acrosome in the sperm head is released. In humans one of the major hydrolytic enzymes that is released during AR is acrosin. While some studies suggest that acrosin has a key role in the process of sperm penetration of the ZP (Jones and Brown, 1987; Tranter et al., 2000; Adham et al., 1997), others suggest that acrosin acts as a molecule that accelerates the distribution of acrosomal proteins during AR (Yamagata et al., 1998). Contrary to the above mentioned studies, it was shown that knock out mouse-sperm lacking ß-acrosin, which is the active form of acrosin, is still able to penetrate the ZP, although with a delay (Baba et al., 1994). This suggests that the role of acrosin is not crucial during egg penetration. In abalones, when sperm comes in contact with the VE two main acrosomal proteins are released, lysin and sp18 (Lewis et al., 1982; Shaw et al., 1993; Kresge et al., 2001b). It has been suggested that lysin and sp18 have different functions during fertilization.

1.2.2 Sperm protein lysin

Since gamete interaction in abalones is achieved externally they have developed a way to ensure a high rate of fertilization, which is by releasing a large amount of gametes during spawning, roughly 10^9 sperm ml^{-1} (Riffell et al., 2002). In 1982 Lewis et al identified a 13 kDa, later determined to 16 kDa, sperm protein named lysin (Lewis et al., 1982) to be one of the two major proteins in the acrosome. When abalone sperm comes in contact with the VE (Lewis et al., 1982; Sakai YT, 1982), AR takes place and lysin is released (Sakai YT, 1982), resulting in the dissolution of the VE. By isolating VE fragments in seawater and then adding the isolated VE to a cuvette containing lysin, turbidometric analysis was performed. This showed that lysin is active for roughly 15-30 seconds and then the activity stops. To establish if the activity of lysin in dissolving the VE is enzymatic or not, different enzyme inhibitors was added during VE dissolution to inhibit lysins effect. This showed that the enzyme inhibitors had no effect on lysin ability to dissolve the VE, which was was completely dissolved 4 to 6 minutes after adding lysin, suggesting that lysin acts in a non-enzymatic approach to dissolve the VE (Lewis et al., 1982).

It was also shown that by boiling isolated lysin prior to VE dissolution, lysin was no longer able to dissolve the VE, suggesting that its ability to dissolve the VE is dependent on the native structure (Lewis et al., 1982). When the complete amino acid sequence of red abalone lysin was reported it revealed a very positively charged protein, with many hydrophobic residues (Fridberger et al., 1985). Roughly more than 2 decades ago, when the X-ray crystal structure of red lysin was solved, it was the first structure of a reproductive protein (Shaw et al., 1993). The monomer structure revealed some specific features, such as a five α-helical bundle, two positively

charged tracks running along the sides of the molecule as well as a hydrophobic patch (Shaw et al., 1993; Kresge et al., 2001a) (**Figure 3**). Even though the first crystal structure of lysin was solved in a monomeric state, a different crystal structure revealed that it may also exist as a dimer (Diller et al., 1994; Shaw et al., 1995; Kresge et al., 2000b). This was further confirmed by light-scattering, which revealed a molecular weight of 38,2 kDa (Diller et al., 1994). Lysins monomeric and dimeric state was also determined by using fluorescence microscopy, which showed a loss of energy transfer when probed lysin dimer was exposed to VE, suggesting that lysin dimer dissociates when lysin is exposed to the VE (Shaw et al., 1995). All the above mentioned data supports the idea that the initial state of lysin is dimeric but upon contact with the VE the lysin dimer becomes monomer, which is the active form of the protein (Shaw et al., 1995).

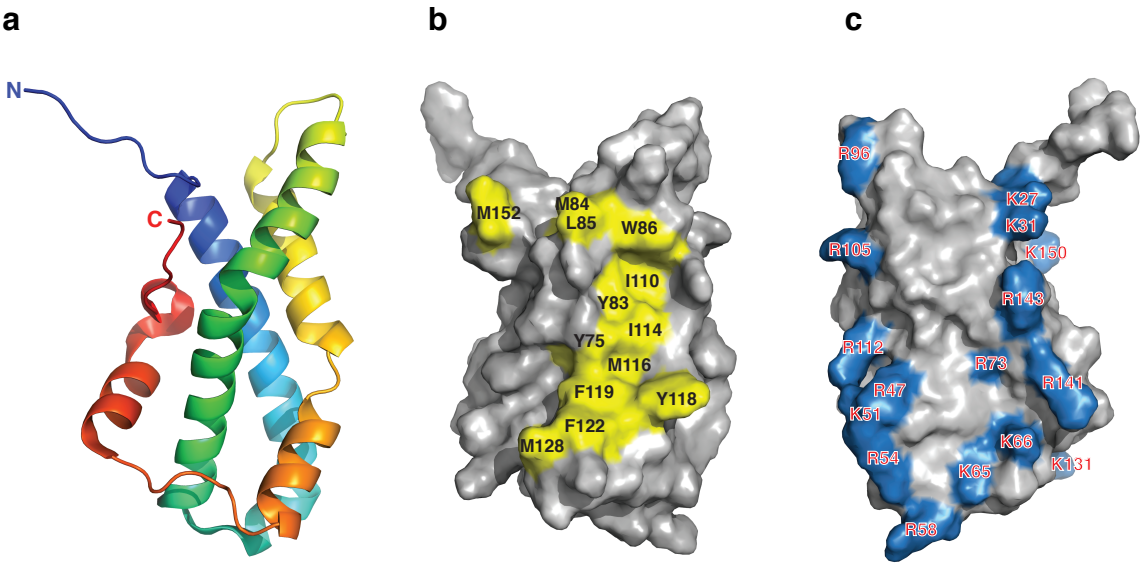

Figure 3. Crystal structure of red abalone lysin. (a) Cartoon representation showing the overall structure of red lysin. (b) Surface representation displaying the hydrophobic patch in yellow. (c) Surface representation highlighting the basic track region in blue running along the side of lysin.

1.2.3 ZP domain

The ZP domain was first identified in 1992 by sequence analysis of proteins ZP2, ZP3, uromodulin, transforming growth factor receptor III and zymogen granule membrane protein 2 (Bork and Sander, 1992). These proteins are involved in egg-sperm interaction, kidney function and TGF-β signaling (Bokhove et al., 2016a; Bork and Sander, 1992), highlighting the diverse role and function of the ZP domain. Since the first rapport of such a domain many more ZP domain containing proteins have been identified in different tissues and organisms (Jovine et al., 2005; Jovine et al., 2002). The proteins involved in egg coat formation all have a ZP domain composed of a

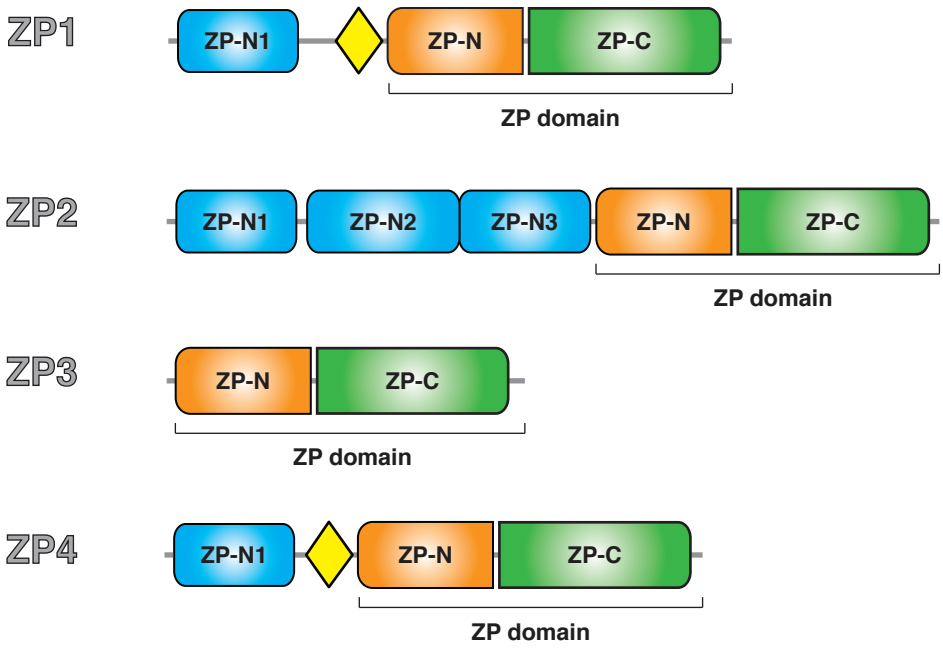

Figure 4. ZP domain architecture of ZP1-4. All ZP proteins share a common ZP domain composed of ZP-N (blue) and ZP-C (green). ZP1 and ZP4 also have a trefoil domain (yellow). In addition to the ZP domain, ZP1 contains an additional ZP-N repeat believed to be involved in filament formation (Okumura et al., 2004; Okumura et al., 2015), while ZP2 contains 3 additional ZP-N repeats. The importance of the first ZP-N1 repeat of ZP2 was reported by using humanized ZP in transgenic mice, where human sperm binds to ZP-N1 (Baibakov et al., 2012). Although the role of the additional ZP-N repeat of ZP4 is not clear, it has been suggested that human ZP4 is involved in binding to capacitated sperm (Gupta, 2015).

ZP-N and a ZP-C domain (**Figure 4**). One role of the ZP domain on the egg coat is to form filaments, by polymerization, that covers the egg coat (Galindo et al., 2002; Jovine et al., 2006).

In 2008 our group solved the crystal structure of the ZP-N domain of ZP3, the subunit that is believed to be involved in mediating the initial binding between gametes in mammals (Monne et al., 2008). The structure revealed that it belongs to a distinct subtype of the immunoglobulin (Ig) superfamily of proteins. Two years later our group also solved the crystal structure of full-length ZP3 (Monne et al., 2008; Han et al., 2010). While it has been suggested that the ZP3 is the primary sperm receptor, the NTR of ZP2 has been suggested to act as a secondary binding site for acrosome reacted sperm (Tsubamoto et al., 1999; Gahlay et al., 2010), making these reproductive proteins very attractive for structural studies. Analysis of the crystal structure of ZP3 showed some specific structural features which characterizes a ZP-N domain, such as a β-sandwich fold, two disulfide bonds with invariant 1-4, 2-3 connectivity, a conserved tyrosine and a unique E´strand (Monne et al., 2008; Han et al., 2010). The ZP domain is not only conserved in vertebrates but it has also been identified in invertebrates, such as previously mentioned VERL (Jovine et al., 2005; Monné et al., 2006).

1.2.4 Egg coat protein VERL

VERL is one of the major proteins found in the VE of abalone. It is a glycoprotein with a molecular mass of two million dalton, with half of the mass being carbohydrates (Swanson and Vacquier, 1997). The identification of VERL was done by coupling lysin to beads and passing dissolved purified VE over the lysin-coupled beads. The sample was further investigated by gel electrophoresis, which identified VERL. Isolated VERL was shown to be able to inhibit lysin's ability to dissolve the VE (Swanson and Vacquier, 1997), it was also proposed that recombinant VERL can inhibit lysin's ability to dissolve the VE (Galindo et al., 2002). Furthermore, VERL binding to lysin was detected by sucrose density gradient ultracentrifugation, which showed that lysin co-sediment with VERL. Electron microscopy images of VERL shows that it is a long rod shaped molecule with a size of roughly 13 nm in diameter (Swanson and Vacquier, 1997). When the full-length sequence of VERL was identified, it revealed that the long rod shaped molecule is composed of 22 tandem repeats, with each repeat being connected by a linker (Galindo et al., 2002). It was also revealed that the C-terminal region of VERL contains a ZP domain, (Galindo et al., 2002), composed of a ZP-N and a ZP-C domain (Jovine et al., 2005; Monné et al., 2006; Wassarman and Litscher, 2008) (**Figure 5**).

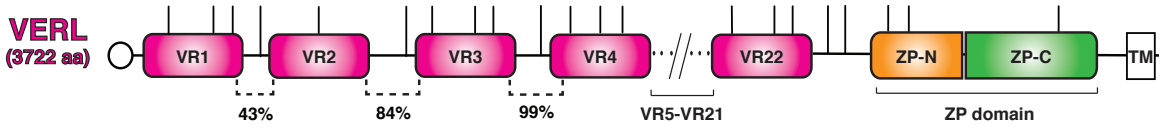

Figure 5. Domain organization of VERL. VERL repeats 1-22 (VR1-VR22) (highlighted in magenta). Open circle and lines correspond to signal peptide and N-linked glycosylation sites, respectively. Sequence identity between VR1-VR2 and VR2-VR3 is 43% and 84%, respectively. Repeats 3-22 are ~ 99% identical.

1.3 VERL-lysin binding and dissolution of the VE

Even though there are many different models suggesting how the binding between lysin and VERL contributes to the dissolution of the VE, there are no data showing a clear mechanism behind this process. Also, it is still not clear, regarding the stoichiometry between lysin and VERL, how many lysin molecules is binding to each VERL repeat (Swanson and Vacquier, 1998; Kresge et al., 2001a; Galindo et al., 2003). A brief summary regarding the stoichiometry and the different hypothesis, suggesting how lysin binding to VERL contributes to the dissolution of the VE, is highlighted below as well as a general model for the opening of the VERL filaments (**Figure 6**):

1) The discovery of the hydrophobic regions in lysin suggested that these regions may be involved in binding to the hydrophobic region of glycoproteins on the VE. It

was also suggested that since VE filaments are polymers the insertion of lysin into these polymers would break the polymer configuration (Lewis et al., 1982).

2) When solving the X-ray crystal structure of lysin, the structure gave raise to the idea that the positively charged basic tracks would assist the binding to the VE, which then would allow the unfolding of the VE proteins via the hydrophobic patch. As an alternative mechanisms, it was also suggested that the hydrogen bonding between VE glycoproteins would disrupt when lysin binds to the VE or that lysin binding induced a conformational change in the VE proteins (Shaw et al., 1993; Messier and Stewart, 1994).

3) By using fluorescently labeled lysin, stoichiometric binding assay was performed suggesting that for each VERL molecule there are 7 lysin molecules bound (Swanson and Vacquier, 1997).

4) As previously mentioned, the hypervariable region of lysin have been shown not only to be important for species specific binding but also for dissolving the VE. Furthermore, it was also suggested that roughly 60 lysin molecules bind to each VERL molecule (Swanson and Vacquier, 1998; Lyon and Vacquier, 1999), suggesting that roughly 3 molecules of lysin bind to each VERL repeat.

5) The dimer structure of red lysin suggested an alternative mode of action for species specific binding between lysin and VERL (Kresge et al., 2000a). It has been purposed that the dimer formation of lysin creates a cleft, which might have an important role in species specificity. This idea was further supported when the crystal structure of green abalone lysin was solved (Kresge et al., 2000b), which showed that this specific region differs between the crystal structures between red and green lysin, further suggesting that this region is important in species specific recognition.

6) It has been suggested that the 23 positively selected residues in lysin is important during species specific recognition to the egg coat (Yang et al., 2000).

7) When the full length sequence of VERL revealed that repeat 1 and 2 are under positive selection it was suggested that these regions are involved in the species specific recognition to lysin (Galindo et al., 2002).

8) In 2013 Aagaard et al suggested that a ZP-N motif alone is sufficient to bind lysin, such as VERL repeat 1 (Aagaard et al., 2013). Furthermore, they also suggested that 2 lysin molecules binds to each VERL repeat (Galindo et al., 2003; Aagaard et al., 2013).

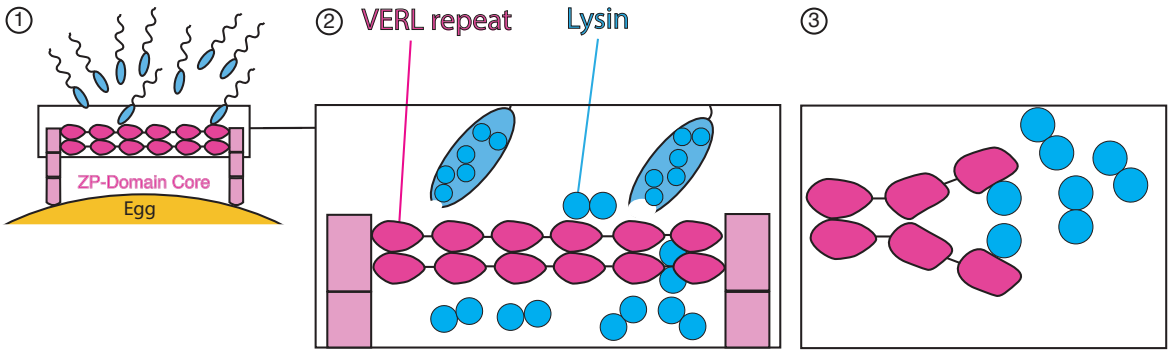

Figure 6. Lysin binding triggers filament dissociation. A general model Illustrating how lysin binding to VERL opens the VERL filaments.1) Sperm is released and comes in contact with the egg VE. 2) AR reaction takes place and lysin is released. 3) Upon contact with VERL, lysin dimers dissociate, allowing the monomer to bind VERL repeats. This forces the VERL filaments to open and allow the sperm to enter and interact with the egg plasma membrane.

There are many different hypotheses on where the interacting region on lysin and VERL are located, how many lysin molecules are bound per VERL repeat and how the binding contributes to dissolving the VE. All these suggestions have been proposed based on the sequence of lysin and VERL, as well as the crystal structure of red and green abalone lysin. However, we wanted to better understand this interaction and the mechanism, therefore we solved crystal structure of VERL as well as the VERL-lysin complex.

1.3.1 Fusion to the plasma membrane

Once the initial binding event between gametes has taken place on the egg coat, sperm is able to penetrate and fuse to the egg plasma membrane. Although no convincing mammalian egg-sperm recognition complex has been identified on the egg coat, the identification of egg protein Izumo1 and sperm protein Juno reveals the first egg-sperm protein complex in mammals at the plasma membrane, crucial for fertilization. Mouse sperm protein Izumo1 was first identified in 2005 by Inoue et al where they showed that Izumo$^{-/-}$ mice, even though they were healthy, were infertile. They also observed that even though sperm was able to bind to the ZP and penetrate, sperm had lost the capability to fuse with the egg (Inoue et al., 2005). It was not until roughly a decade later that the binding partner of Izumo1 was identified. In 2014 Bianchi et al identified mouse egg protein folate receptor 4 (Folr4), which they renamed to Juno, as the binding partner for Izumo1 (**Figure 7**) (Bianchi et al., 2014). By using avidity-based extracellular interaction screen (AVEXIS) Bianchi et al could detect low affinity binding (K_d ~ 12 μM) between mouse Izumo1 and Juno. The importance of Juno during fertilization was further investigated by blocking the binding between Juno and Izumo1 *in vitro*, resulting in inhibition of fertilization. Moreover, the role of Juno was highlighted

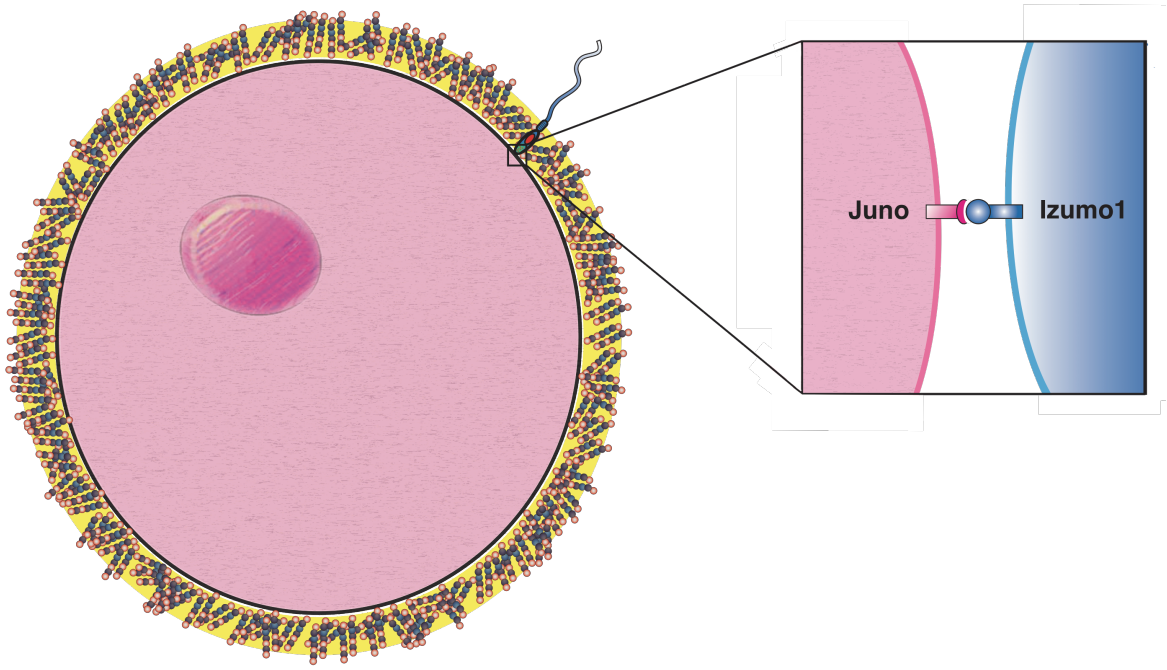

Figure 7. Gamete interaction at the plasma membrane. Once sperm has penetrated the egg ZP (shown in yellow) it binds to the underlying membrane, where egg protein Juno, the receptor for sperm protein Izumo1, is located. It has been shown that approximately 40 minutes, post fertilization, Juno is almost completely lost and by using electron microscopy, Juno has been detected in extracellular vesicles (Bianchi et al., 2014).

by generating Juno-deficient mice. Even though female Juno$^{-/-}$ mice exhibited normal mating behavior they did not produce any pups. This is the first egg-sperm protein complex that have been identified in mammals, which is crucial for triggering gamete membrane fusion. One important note is that Juno is only conserved among mammals (Grayson, 2015).

Roughly 1 year after the reported Juno-Izumo1 complex, our group solved the crystal structure of mouse Juno (paper IV). The crystal structure revealed a similar fold as folate receptors (FR), FRα and FRß (Han et al., 2016). The structure also showed a highly flexible region at the binding site, where vitamin B_9/folic acid is binding to the folate receptors. We conclude that even though Juno has lost the ability to bind vitamin B_9 it obtained the ability to interact with Izumo1. And as previously mentioned, Juno is only conserved in mammals, while FR homologues are discovered in all vertebrates, suggesting that they must have evolved form an ancestral FR. To better understand the basic mechanism behind the interaction between Juno and Izumo1, two independent groups solved the X-ray crystal structure of human Juno and Izumo1 as well as the complex (Ohto et al., 2016; Aydin et al., 2016). The structure of unbound human Izumo1 revealed an elongated structure that is assembled by two separate domains, the Izumo1 domain and the Ig-like-domain, which is connected via a ß-hairpin region that also acts as the binding region for Juno (Aydin et al., 2016; Ohto et al., 2016). At the same time, our group solved the crystal structure of mouse Izumo1 (Nishimura et al., 2016). By comparing the structures, we can see that mouse and

human Izumo1 are almost structurally identical, with a difference in the hook region of mouse Izumo1 (Aydin et al., 2016; Ohto et al., 2016; Nishimura et al., 2016). Nishimura et al also discovered that Izumo1 is structurally similar to *Plasmodium parasites* proteins SPECT1 and TRAP, which are involved in host cell traversal and hepatocyte attack as well as gliding motility and penetration, respectively (Nishimura et al., 2016). The complex structure of human Juno and Izumo1 do not show any structural differences between bound and unbound state, although in one article the authors suggest that there is a conformational change in Izumo1 upon binding to Juno (Aydin et al., 2016).

The counterpart of Izumo1 in abalone might be sperm protein Sp18. As previously mentioned, the two major acrosomal proteins in abalone, lysin and sp18 (Lewis et al., 1982), are believed to have two separate roles during fertilization (Kresge et al., 2001b). It has been shown that sp18 must have a different role in comparison to lysin during fertilization, due to its inability to dissolve the VE (Kresge et al., 2001b). By solving the crystal structure of green abalone sp18 (Kresge et al., 2001b) it was clear that sp18 and green abalone lysin share very similar structural features, suggesting that the two proteins arose by gene duplication (Kresge et al., 2000b; Kresge et al., 2001b; Swanson and Vacquier, 2002). This suggest that the dissolution of the VE by lysin is not dependent on the structural conformation of lysin, rather the surface properties of the protein. Sp18 has suggested to be involved in membrane fusion rather than the initial recognition event (Swanson and Vacquier, 1995).

2 AIM AND SIGNIFICANCE

The majority of the studies presented in this thesis has aimed at further understanding the process of fertilization using crystallography as a tool. We sought to better understand if there is structural fold that has been conserved between invertebrate and vertebrate recognition proteins during evolution.

Finally, by isolating a recombinant egg-sperm protein complex *in vitro* we also wanted to determine the X-ray crystal structure of such a complex. This will help us to better understand the mechanism behind sperm penetration during fertilization and increase our knowledge regarding the role of positive selection in egg-sperm recognition proteins.

Specific aims:

1. Determine, by bioinformatics, if evolutionary distant reproductice proteins share a ZP-N fold, such as ZP3.

2. Structurally confirming our hypothesis by solving the crystal structure of VERL and ZP2 ZP-N1.

3. Reconstitute egg-sperm protein complex *in vitro* by using recombinant VERL and lysin.

4. Determine the crystal struture of VERL-lysin complex.

5. To better understand the role of positive selection in VERL and lysin.

6. Obtain a more specific model on how VERL-lysin complex formation contributes to the opening of the egg coat during sperm entry.

3 METHODOLOGY

3.1 Protein expression

Recombinant protein can be used in many aspects of science such as structural determination and biological function characterization. This information can later be used in the industry for drug development and therapeutic purposes. But the process of obtaining properly folded purified protein is time consuming and challenging, which requires careful planning, even before beginning the purification process. Prior to protein expression trials there are many different factors to be considered, such as choosing a suitable expression vector, designing the genes of interest for either expressing the full-length protein or just a domain of interest, selection of an appropriate detection tag as well as the location of the tag. Once the gene of interest is designed one needs to select the expression system of choice. The main expression systems are yeast, bacteria, insect cells and mammalian cells, all with their own advantages and disadvantages. Also during this step there are many different factors to be considered such as, time, handling and cost, before choosing the desired expression system .The most classical expression system is the bacterial expression system due to easy handling, time saving, low cost and high yields of protein expression (Makrides, 1996; Baneyx, 1999). However, this system does have some disadvantages such as the inability to perform several post-translational modifications in comparison with eukaryotic systems (Makrides, 1996), as well as difficulty obtaining properly folded protein. Another widely used expression system is insect cells, such as *Spodoptera frugiperda* Sf9 cells. This was the expression system of choice at the initial stages of the VERL/lysin project. Due to several reasons we decided to choose this expression system: the ability of developing baculovirus strains for large scale protein expression, correct posttranslational modifications and the formation of disulfide bonds (Unger and Peleg, 2012). Even though crystallization trials were performed using protein produced from this system, we were unable to obtain well diffracting crystals. We therefore decided to move to a mammalian expression system by using human embryonic kidney (HEK) 293 cells (Graham et al., 1977), a system which was more well established in our lab. Despite being very time consuming the use of mammalian stable cell lines is still common, mainly for large scale protein production in pharmaceutical companies, due to the ability of mammalian cells to express properly folded protein that is post-translational modified (Wurm, 2004; Jenkins et al., 2008). Furthermore, even though stable cell line development is still used for structural studies (Chaudhary et al., 2012) the screening process and time consumption is not efficient enough. In some cases, one need to screen several (sometimes hundreds) different constructs, to be able to obtain a well behaving protein sample that is able to produce good diffracting crystals.

To perform large scale transient transfection in mammalian cells using commercial transfection reagents would be extremely expensive. One way of reducing cost is to use polyethyleneimine (PEI) as a transfection reagent, which was first described roughly 20 years ago (Boussif et al., 1995). Since then, large scale transient transfections, using PEI as a transfection reagent, have become more widely used, mainly due to low cost (Schlaeger and Christensen, 1999; Geisse and Henke, 2005; Aricescu et al., 2006).

By taking advantage of the protocol described by Aricescu et al, combined with our mMBP fusion system we are able to produce large amount of correctly folded, post-translational modified recombinant protein, in a fast, low cost and effective approach for our structural and functional studies (Aricescu et al., 2006; Bokhove et al., 2016b).

3.2 Protein purification

Depending on the requirements, protein purification may be achieved in one step, for a semi pure sample, or in several steps to obtain as pure material as possible. The degree of sample purity depends on what purpose the protein is prepared for. In general, a rule of thumb for X-ray crystallographic studies is that it is better to have less quantity of protein which is very pure than to have higher quantity of semi pure material. The purification process can be achieved in different ways by using the characteristic features of the target protein such as, charge, size, and specific binding regions. One of the most common purification techniques, which is also applied in our laboratory, is immobilized metal affinity chromatography (IMAC). By adding a stretch of poly-histidine, at either the N- or C-terminus of the protein, one can use them as a purification tag, due to their ability to bind metal ions such as nickel (E. Hochuli, 1988). This will separate the histidine-labeled protein from the non-labeled proteins. At this stage the protein purity is not sufficient enough for crystallization trials, due to unspecific binding of other proteins to the nickel resin. As previously mentioned one can take advantage of the protein size as a purification approach by applying it to a method termed size-exclusion chromatography (SEC). This is a technique that was first used 60 years ago (Ruthven, 1956) when starch columns was used for separation of molecules based on their size (Ruthven, 1956). The principle of this chromatographic technique is that the concentrated protein sample is inject on to a column containing a resin (stationary phase), while the large molecules will pass through the column fast the smaller molecules gets trapped in the pores of the resin resulting in a slower separation. This allows an efficient separation between large and small molecules and provides a purer sample (**Figure 8**). We routinely take advantage of the above mentioned techniques to obtain pure samples for crystallisation trials.

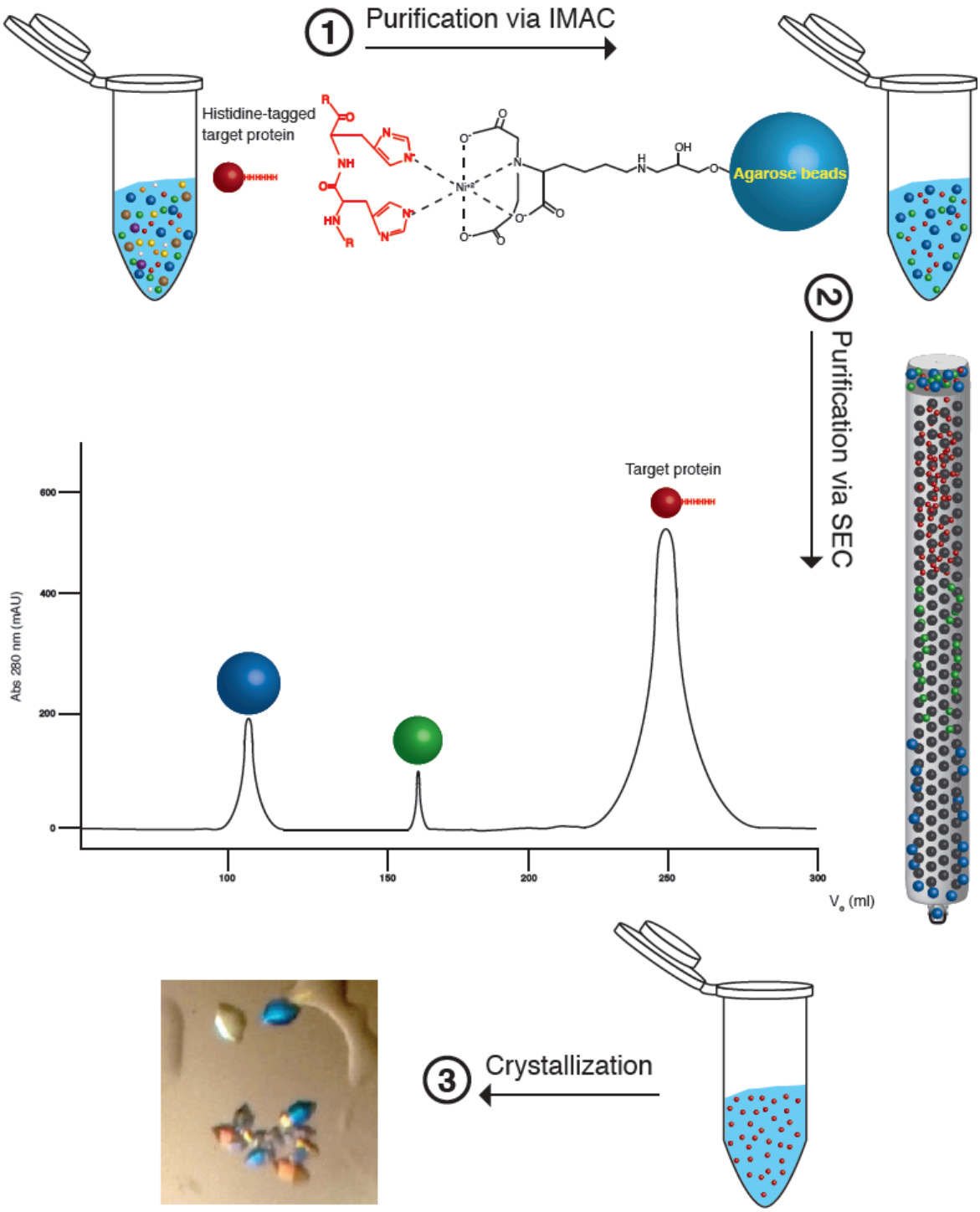

Figure 8. Protein purification procedure. Once the histidine-tagged target protein has been expressed, the first step (shown as number **1**) is to separate it from impurities by binding to nickel beads. **2**) The second purification step is to further separate the target protein, from unspecific bound proteins to the nickel beads, by using SEC, which separates proteins based on molecular weight and shape. **3**) Finally, if the purity of the target protein is suitable, crystallization trials can be performed.

3.3 Protein crystallization

Protein crystallization has been known for roughly 180 years when Friedrich Ludwig Hünefeld accidentally crystallized hemoglobin from earthworm blood (Hünefeld, 1840). The process of obtaining protein crystals is a trial and error process, even though we have gained more knowledge on how to place our sample in the most optimal conditions for crystallization (Chayen, 1998; Luft et al., 2007; Thakur et al., 2008), this does not mean that we will obtain crystals every time we put our sample in that particular condition. The process of finding the precise condition for crystallization of different proteins is still impossible to predict. Some important factors that can have a determining role in crystal formation are, protein purity, protein solubility, protein folding, temperature and crystal nuclei formation (Delucas et al., 2005; Galkin and Vekilov, 2000). Even if the conditions are favorable for crystal formation, the crystals still might not be good enough in terms of diffraction, size or form, which can hinder structure determination. One way to tackle this problem is to optimize the crystallization conditions to overcome the above mentioned problems. Optimization can be achieved by changing different factors such as protein concentration, precipitant or precipitant concentration, temperature and pH (Messick and Marmorstein, 2004; Gosavi et al., 2008). X-ray protein crystallography is a very powerful technique for solving the structure of macro-molecules to better understand their biological function as well as developing drugs for pharmaceutical purposes. In recent years the process of protein crystallization screening has become more and more automated (Zhu et al., 2014) with the use of machines such as the mosquito crystallization robot, which we routinely use in our lab (Bokhove et al., 2016b). Even though major advances have been done during recent years in the field of X-ray protein crystallography we still do not know why some proteins crystallize very easily and why some proteins do not. I truly believe that protein crystallography is as much of an art as it is science.

4 RESULTS AND CONCLUSIONS

4.1 Paper I

Rapid adaptive evolution, driven by positive selection, occurs when there is an advantageous mutation in a gene, resulting in a change in the encoded protein sequence. If there are repeated rounds of positive selection, this will eventually lead to the rapid evolution of the amino-acid sequence (Vacquier and Swanson, 2011; Lou et al., 2014). This phenomenon is acting in reproductive proteins in both invertebrates and vertebrates (Swanson et al., 2001; Swanson and Vacquier, 2002). There are several hypotheses that have been presented on why reproductive proteins are some of the fastest evolving proteins. One reason, which has been suggested to be a driving force behind adaptive evolution is co-evolution, where change in one species affects another species. Also sperm competition and sexual conflict are believed to play a role in evolution of reproductive proteins (Parker and Partridge, 1998; Swanson and Vacquier, 2002). Due to the rapid change in sequence in these proteins it is difficult to find structural similarities between proteins from different organisms.

In this project we took advantage of the specific structural hallmarks of the ZP-N domain of ZP3, which we used as a template to find structural similarities amongst remote evolutionary proteins involved in reproduction.

By applying bioinformatics analysis, we detected the sequence of vitelline envelope zona pellucida protein 14 (VEZP14), a cell adhesion protein in yeast *Saccharomyces cerevisiae* Sag1p and VERL repeat 10 as potential ZP-N domains, despite having an overall sequence that is very different. Based on these findings, a homology model of Sag1p and VERL repeat 10 was created. In the case of VERL, this suggests that VERL repeat 10, as well as other VERL repeats, might adopt a ZP-N like fold. In addition to the cysteine connectivity of C1-C4 and C2-C3 that is seen in VERL, Sag1p and ZP-N of ZP3, VERL repeat 2 has an additional disulfide bond (C201-C294). We hypothesized that the additional disulfide bond may be involved in intermolecular disulfides, resulting in homo-dimerization of VERL (Swanson and Vacquier, 1997). To confirm our hypothesis, we designed expression constructs of the first 4 repeats of VERL and a mutated version where the two additional cysteines were mutated. The two constructs were expressed in insect cells and the proteins were analyzed by immunoblotting in reducing and non-reducing conditions. As suggested, the wild type protein forms a disulfide-linked homodimer in non-reducing condition whereas the double cysteine mutant mostly forms a monomer.

In conclusion, this paper reveals a common structural domain, important for gamete recognition, is shared among these reproductive proteins suggesting that the ZP-N fold has been conserved during evolution, despite 0.6-1 billion years of evolution.

4.2 Paper II

Overexpression of target proteins may in many cases be a rate limiting factor for crystallographic studies, due to degradation, aggregation, size of the protein of interest and low amount of target protein. One way to solve some of these problems is by expressing a fragment of the protein which is of interest, even then, gaining large amount of properly folded protein may be difficult. Also it is important to choose the correct expression system for protein production because of post-translational modifications, which are required for the correct function of the protein.

The malE gene in *Escherichia coli* (*E. coli*) encodes for MBP, which is involved in the uptake and transfer of maltose and maltodextrins (Duplay et al., 1984; Nikaido, 1994). MBP can also be used as a fusion partner for target proteins expressed in *E. coli* to increase protein solubility and prevent aggregation. In this study we took advantage of the above mentioned features of MBP, which we applied in a mammalian expression system using HEK293 cells. Since many proteins of interest are of human origin, this system allows the target proteins to acquire the correct post-translational modifications, which would not have been possible using a bacterial expression system. The fusion to MBP also significantly increased the protein expression of our target proteins. By introducing specific mutations in MBP, which benefits affinity purification as well as crystallization of MBP fused target proteins (Center et al., 1998; Walker et al., 2010), we designed a mammalianized version of MBP (mMBP). We fused several target proteins to mMBP, either with signal peptide (SP) for secretion or without SP for intracellular expression. We also designed a vector which includes a TEV protease cleavage site to be able to remove mMBP after protein expression. By doing so, the unfused target protein may be used for functional assays, where mMBP would not interfere. Furthermore, mMBP can also be used as an additional purification tag by binding to amylose beads. Finally, the major advantage of a mMBP fused target proteins for structural determination is the use of mMBP for phasing during MR.

In this paper we have taken all the advantages of MBP and applied them in a mammalian expression system, which is the system of choice when expressing human proteins, which are dependent on different post-translational modifications. We were able to solve the crystal structure of several target proteins that did not crystallize when unfused.

4.3 Paper III

There is no clear structural evidence showing if the repetitive domains of VERL are ZP-N domains, which would be a direct link between invertebrate and vertebrate fertilization. We also have not yet been able to determine a crystal structure of a single gamete recognition complex taking place on the egg coat. Finally, we still lack a detailed picture of how a sperm is able to unravel the extracellular matrix surrounding the egg. We wanted to experimentally demonstrate if there is a link between

invertebrate and vertebrate reproduction by solving the crystal structure of VERL repeats and mammalian ZP2 ZP-N1. We also wanted to better understand the mechanism of how sperm binds to the egg coat and splays apart the VERL filaments to create a hole. This was achieved by solving the crystal structure of a complex between VERL and lysin.

In this paper we solved the crystal structure of mouse ZP2 ZP-N1, mMBP fused single domain of VERL, as well as the complex between VERL and lysin. For mMBP fused proteins we used mMBP as a search model for structure determination (paper II). The structures revealed that these reproductive proteins in fact share a common ZP-N fold, with some structural differences. The complex structure determination between VERL and lysin was achieved by co-expressing mMBP fused VERL together with lysin. Despite extensive crystallization trials this protein complex did not yield well diffracting crystals, we decided to remove mMBP by limited proteolysis using elastase, which yielded well diffracting crystals. Also by co-expressing unfused VERL with lysin we were able to obtain crystals in different conditions and space group. The complex crystal structure reveals a very accurate model of how lysin is breaking apart VERL filaments on the egg coat, creating a hole for sperm to enter and interact with the egg plasma membrane.

4.4 Paper IV

Until just 2 years ago no convincing egg-sperm recognition pair had been identified in mammals during gamete fusion. Earlier in vivo studies show that both sperm protein Izumo1 and egg protein CD9 have an important role during gamete fusion (Chen et al., 1999; Miyado et al., 2000; Kaji et al., 2000; Kaji et al., 2002). It was shown that Izumo1 deficient mice were completely infertile while $CD9^{-/-}$ females showed severe reduction of fertility. It was also later shown that recombinant Izumo1 still binds to CD9-/- deficient eggs equally well as to the wild type eggs. This does not diminish the significance of CD9 but it suggests that the binding partner of Izumo1 is a different protein (Inoue et al., 2013; Bianchi et al., 2014). In 2014 this was confirmed when egg protein Juno was identified as the binding partner to Izumo1, and the importance of Juno was highlighted when Juno deficient female mice were shown to be infertile (Bianchi et al., 2014).

In this paper we solved the crystal structure of the binding partner of Izumo1, Juno, previously known as folate receptor 4 (Bianchi et al., 2014). The crystal structure reveals a similar fold as FRα and FRß, with major flexibility within the area involved in binding folate in these two receptors. Furthermore, while the loop regions of FRs are ordered, the corresponding loop area in Juno is either disordered (loop 1 and 3) or adopts a different confirmation (loop 2), suggesting that this might be one reason why Juno is incapable of binding vitamin B_9/folic acid. Even though, the interaction between Juno-Izumo1 is conserved in mammals, yet mouse Juno does not bind human

Izumo1, despite sharing a sequence identity of 70%, indicating some degree of species specificity. When looking at the sequence, corresponding to loop 1, 2 and 3, it is clear that these regions are relatively different between human and mouse, suggesting that these might be regions that are important for determining species specificity. To test our observations, we designed several mouse Juno expression constructs where the loop region 1-3 were replaced with the corresponding human sequence. By using AVEXIS it was shown that when loop 1 of mouse was replaced with human no more binding was observed. Also by changing mouse Juno loop 3 to human we could detect binding to human Izumo1, suggesting that these areas are involved in binding.

In conclusion this paper suggests that Juno and FR evolved from an ancestral gene and that Juno is no longer able to bind to Vitamin B_9 but instead is able to interact with the sperm via Izumo1.

5 FINAL CONCLUSIONS AND FUTURE PROSPECTIVE

We have yet to identify a gamete recognition protein complex on the mammalian egg coat, which is essential for fertilization. By identifying and solving the structure of such a complex it would help us to better understand the initial egg-sperm binding at the molecular level. To be able to confront this important biological question we chose a model organism where egg-sperm recognition proteins have been well identified. This not only helped us to better understand if there is a structural link, which has been conserved during evolution, between invertebrate and vertebrate reproductive proteins, it also allowed us to study the process of gamete interaction at the molecular level.

In 2014 the binding partner for mammalian sperm protein Izumo1 was identified as egg protein Juno. Juno-Izumo1 complex has shown to be essential for fertilization in mammals. Roughly 2 years after the discovery of Juno the individual crystal structures of Izumo1 and Juno was solved, both from human and mouse, as well as the human complex structure.
However, Juno-Izumo1 would never interact, if the initial recognition event did not take place on the mammalian egg coat. The importance of identifying an egg-sperm protein complex, taking place on egg coat, would be of great value. This would not only help us to better understand the mechanism behind the initial egg-sperm interaction, but would also be a major step forward towards developing a target specific contraceptive.

In **paper I** we hypothesized that there might be a common domain important for gamete recognition during fertilization ranging from yeast to human. This hypothesis was possible due to the crystal structure of ZP3 ZP-N, which our group previously solved. Based on this structure we could perform bioinformatics analysis suggesting that ZP2, VERL, VEZP14 and Sag1p all contain ZP-N domains. To investigate 3D structure of some of these proteins, we used an expression system based on MBP as described in **paper II**. As previously mentioned, one of the rate limiting steps in protein crystallography is the amount of properly folded protein needed for structural studies. Also another difficulty during structure determination is the phase problem. In this paper we show that not only we can increase protein expression but we can also use MBP as a purification tag. Moreover, we could solve the crystal structure of mMBP fused target proteins, which did not crystallize as unfused. And finally we can solve the phase problem by using mMBP for MR during structural determination. In **paper III** we took advantage of our mMBP developed protein expression system to solve the crystal structure of mMBP fused VERL domains, which revealed that they adopt a ZP-N fold. To further confirm the link between vertebrate and invertebrate fertilization Ling Han and Elisa Dioguardi solved the crystal structure of ZP-N1 of ZP2, which also showed a similar fold. Furthermore, we solved the crystal structure of VERL/lysin complex. The complex structure helped us to predict a more precise model for the mechanism of how lysin contributes to unravelling the VERL filaments on the VE, allowing sperm entry. In **paper IV** Han et al solved the crystal structure of egg plasma membrane protein Juno. The structure revealed a FR fold similar to that of FRα and

FRβ. The structure also revealed that the binding region, corresponding to the region involved in binding vitamin B_9 in FRα and FRβ, was either not ordered due to high flexibility or adopted a different arrangement in the structure. This allowed us to hypothesize that this might be one reason why Juno is not able to bind vitamin B_9 but instead has evolved to bind sperm protein Izumo1.

5.1 MBP in protein expression and crystallography

In protein crystallography it is not always easy to obtain sufficient quantity of well-behaved protein that can be purified and used for crystallization studies. Although MBP has been used as a fusion partner since the beginning of the 90's (Szmelcman et al., 1990; Blondel and Bedouelle, 1990), it had previously not been used in a mammalian expression system as a fusion partner for structural studies. We show the efficiency of MBP by fusing it to several human target proteins in a mammalian expression system. By using this established expression system, we can easily grow several liters of mammalian expressed mMBP fused target proteins for either crystallization trials or functional assays. This system does also have its limitations, such as the difficulty to use mMBP for MR when the target protein is much bigger than mMBP. Or as in the case of mMBP fused VERL-lysin complex crystallization trials, were we did not get any complex crystals, suggesting that mMBP was interfering with crystal formation.
This system will prove to be an even more powerful tool in the future if the fusion of mMBP to target molecules can be implemented in other expression systems for crystallographic studies. Furthermore, one can also implement the mMBP fusion system for structural studies using other techniques different from crystallography.

5.2 Bridge between invertebrate and vertebrate fertilization

Over 95% of all animal species are invertebrates (Van Wettere and Lewbart, 2007), however until recently it was not known whether there were specific protein epitopes that had remained conserved during evolution between invertebrates and vertebrates.

In this thesis we show a clear structural connection between invertebrate and vertebrate fertilization by solving the crystal structure of ZP-N1 of ZP2 as well as single VERL domains. Furthermore, by solving the crystal structure of VERL/lysin complex, we managed to show detailed information at the molecular level between gamete recognition proteins. This clearly show that although invertebrates, such as abalone, and vertebrate's, such as humans, are separated by roughly 600 million years of evolution, they still share a common structural feature, which is important during

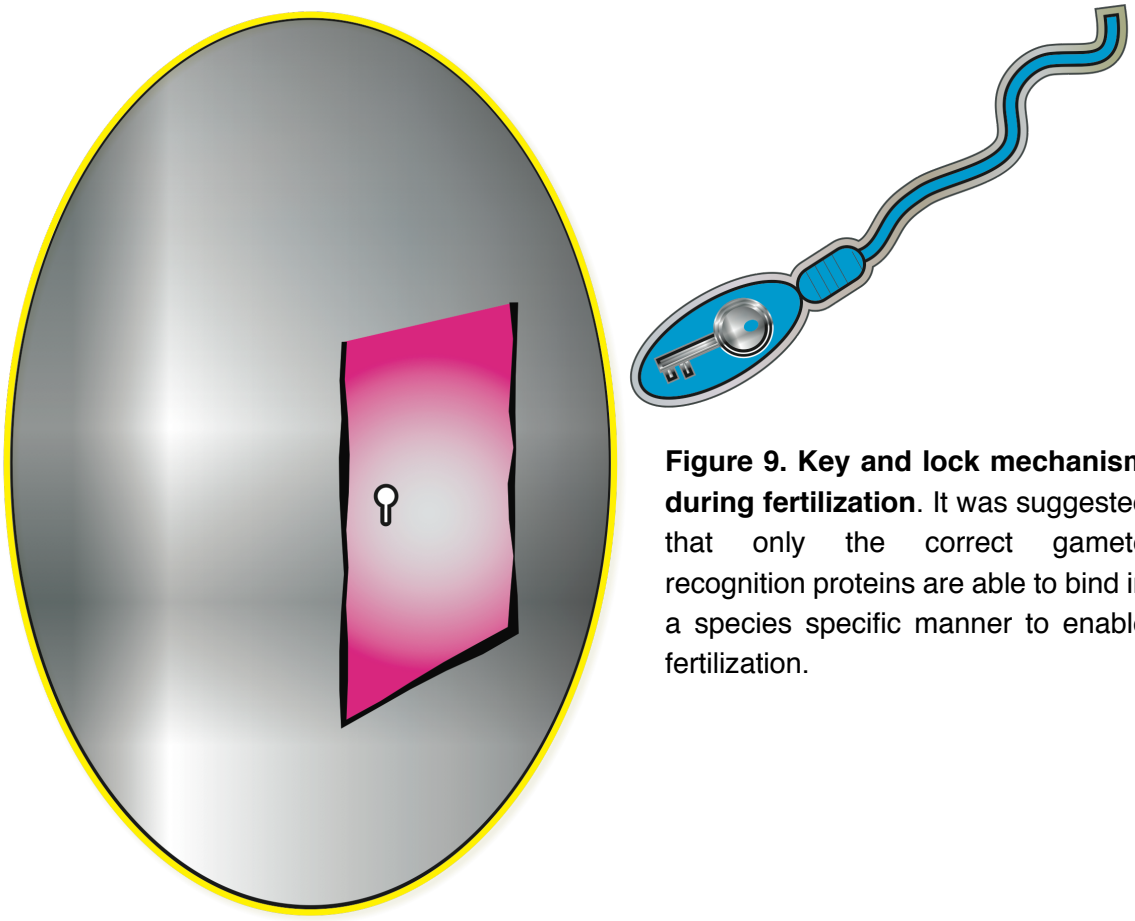

Figure 9. Key and lock mechanism during fertilization. It was suggested that only the correct gamete recognition proteins are able to bind in a species specific manner to enable fertilization.

fertilization. The mechanism behind this process is a very interesting scientific event, yet we still have not been able to obtain clear picture at the molecular level in mammals.

In abalone, the mechanism of dissolving the VE is dependent on lysin and we have highlighted this mechanism at the molecular level by solving the above mentioned crystal structures. However, the equivalent molecule, corresponding to lysin, does not seem to exist in vertebrates, or has not yet been identified. Even though the general idea in mammals is that sperm undergoes AR and binds to the ZP (Florman and Storey, 1982; Bleil and Wassarman, 1983), perhaps this event is different than previously proposed, suggesting that there might be a triggering event that initiates sperm penetration rather than a direct binding to the ZP. If we identified a sperm protein which is involved in complex formation with a ZP glycoprotein, the crystal structure of such a complex would be of interest for the whole biological field. It would not only help us to better understand the mechanism behind egg-sperm interaction on the egg coat but it would also open new possibilities for developing new contraceptives.

The study of sex in a scientific way started already in the Renaissance (Chianchi, 1998). It was not until 1914 that Frank R. Lillie suggested that the process of egg-sperm interaction during fertilization is dependent on a key and lock mechanism (**Figure 9**) (Lillie, 1914). This metaphor was addressed in this thesis by applying structural biology on central reproductive proteins during abalone fertilization. Although VERL/lysin complex structure does not completely answer this hypothesis,

it does shed some light in the key and lock metaphor suggested more than a century ago. Most importantly, we have gained a very first detailed picture of the initial egg and sperm interaction taking place on the egg coat.

6 ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to everyone who has contributed to all the different aspects of these studies and making this dissertation possible.

My first word of gratitude goes to my main supervisor **Luca**. This work would not have been possible without your help and guidance. We have been through a lot during my time as a PhD student. Your passion and enthusiasm for science is admirable and something that is very contagious. From day one you have pushed, challenged and tested me, which helped me to develop into an independent scientist, I hope. Your door has always been open and you have always been available regardless of day and time. Your knowledge in science is impressive and I am grateful that you shared your knowledge with me, both in science and in life, thank you.

Ling my friend, I have been fortunate to have worked with you. Your hard work and your drive for perfectionism is remarkable, even when making a 4-liter buffer you add the additional 100 μl NaCl to get the concentration just right ☺. We have shared many good (and some less good) moments together. I am thankful for your support, guidance and friendship during my time as a PhD student, I will surely miss you.

Elisa my partner in crime. I can't imagine going through this process without you. Thank you for tolerating all my nagging and whining about, well about everything. I still remember the day you expressed sp56 and you saw the protein by western blot, you were jumping up and down in the lab (just for seeing a protein being expressed), we have come a long way since then. I wish you all the best in life.

The past and present members of the LJO lab, **Marcel** your knowledge in science is impressive and I am thankful for all the help and patience you had when teaching me about crystallography. **Kaoru** thank you for the advice and for all the scientific discussions we have, mostly me talking and you listening, your Swedish is improving. **Amy**, **Takako**, **Isha**, **Romina** thank you all for bearing with me, I wish you all the best. A special thanks to **Isha**, who helped pushing our project forward.

Amir one second you talk about cars, next second you talk about the latest motor cycle helmet. Although I have difficulties following your conversations we still get along very well. Thank you for always being there for me no matter what it is. We started off as colleagues and we ended up as close friends. I wish you all the best.

My brother **Homan**, I know that whenever I need anything I can count on you, thank you for always having my back.

My beautiful and intelligent wife **Sara**, what would I do without you! I don't know how many evenings I have discussed VERL-lysin binding, and what the latest pull-down indicated or whatever residue 87 was doing in the interface. Thank you for you always listening and believing in me. Your patience has been amazing, specially these last

couple of months. I am looking forward to share an amazing life with you.
Together we can achieve anything.

Finally, my parents. Words can't express how much you mean to me. You always supported me in all my decisions in life, meaning if I studied you supported me and if I didn't, well I guess we will never find out ☺.
My father **Ali**, you always thought me the value of hard work, your advice is priceless to me regarding all aspects of life. You are a through source of inspiration.
My mother **Fahimeh**, your kindness and understanding always makes me feel secure. Every time we speak you ask me how things are going and what's happening with the projects, thank you for always showing an interest and always being there for me.

7 REFERENCES

Aagaard, J. E., Springer, S. A., Soelberg, S. D., and Swanson, W. J. (2013). Duplicate abalone egg coat proteins bind sperm lysin similarly, but evolve oppositely, consistent with molecular mimicry at fertilization. PLoS Genet *9*, e1003287.

Aagaard, J. E., Vacquier, V. D., MacCoss, M. J., and Swanson, W. J. (2010). ZP domain proteins in the abalone egg coat include a paralog of VERL under positive selection that binds lysin and 18-kDa sperm proteins. Mol Biol Evol *27*, 193-203.

Adham, I. M., Nayernia, K., and Engel, W. (1997). Spermatozoa lacking acrosin protein show delayed fertilization. Mol Reprod Dev *46*, 370-376.

Aricescu, A. R., Lu, W., and Jones, E. Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr D Biol Crystallogr *62*, 1243-1250.

Austin, C. R. (1952). The capacitation of the mammalian sperm. Nature *170*, 326.

Aydin, H., Sultana, A., Li, S., Thavalingam, A., and Lee, J. E. (2016). Molecular architecture of the human sperm IZUMO1 and egg JUNO fertilization complex. Nature

Baba, T., Azuma, S., Kashiwabara, S., and Toyoda, Y. (1994). Sperm from mice carrying a targeted mutation of the acrosin gene can penetrate the oocyte zona pellucida and effect fertilization. J Biol Chem *269*, 31845-31849.

Baibakov, B., Boggs, N. A., Yauger, B., Baibakov, G., and Dean, J. (2012). Human sperm bind to the N-terminal domain of ZP2 in humanized zonae pellucidae in transgenic mice. J Cell Biol *197*, 897-905.

Bailey, J. L. (2010). Factors regulating sperm capacitation. Syst Biol Reprod Med *56*, 334-348.

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. Curr Opin Biotechnol *10*, 411-421.

Bauskin, A. R., Franken, D. R., Eberspaecher, U., and Donner, P. (1999). Characterization of human zona pellucida glycoproteins. Mol Hum Reprod *5*, 534-540.

Beebe, S. J., Leyton, L., Burks, D., Ishikawa, M., Fuerst, T., Dean, J., and Saling, P. (1992). Recombinant mouse ZP3 inhibits sperm binding and induces the acrosome reaction. Dev Biol *151*, 48-54.

Bianchi, E., Doe, B., Goulding, D., and Wright, G. J. (2014). Juno is the egg Izumo receptor and is essential for mammalian fertilization. Nature *508*, 483-487.

Bleil, J. D., and Wassarman, P. M. (1980). Mammalian sperm-egg interaction: identification of a glycoprotein in mouse egg zonae pellucidae possessing receptor activity for sperm. Cell *20*, 873-882.

Bleil, J. D., and Wassarman, P. M. (1983). Sperm-egg interactions in the mouse: sequence of events and induction of the acrosome reaction by a zona pellucida glycoprotein. Dev Biol *95*, 317-324.

Blondel, A., and Bedouelle, H. (1990). Export and purification of a cytoplasmic dimeric protein by fusion to the maltose-binding protein of *Escherichia coli*. Eur J Biochem *193*, 325-330.

Bokhove, M., Nishimura, K., Brunati, M., Han, L., de Sanctis, D., Rampoldi, L., and Jovine, L. (2016a). A structured interdomain linker directs self-polymerization of human uromodulin. Proc Natl Acad Sci U S A *113*, 1552-1557.

Bokhove, M., Sadat Al Hosseini, H., Saito, T., Dioguardi, E., Gegenschatz-Schmid, K., Nishimura, K., Raj, I., de Sanctis, D., Han, L., and Jovine, L. (2016b). Easy mammalian expression and crystallography of maltose-binding protein-fused human proteins. J Struct Biol *194*, 1-7.

Bookbinder, L. H., Cheng, A., and Bleil, J. D. (1995). Tissue- and species-specific expression of sp56, a mouse sperm fertilization protein. Science *269*, 86-89.

Bork, P., and Sander, C. (1992). A large domain common to sperm receptors (Zp2 and Zp3) and TGF-ß type III receptor. FEBS Lett *300*, 237-240.

Boussif, O., Lezoualc'h, F., Zanta, M. A., Mergny, M. D., Scherman, D., Demeneix, B., and Behr, J. P. (1995). A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. Proc Natl Acad Sci U S A *92*, 7297-7301.

Brucker, C., and Lipford, G. B. (1995). The human sperm acrosome reaction: physiology and regulatory mechanisms. An update. Hum Reprod Update *1*, 51-62.

Buffone, M. G., Zhuang, T., Ord, T. S., Hui, L., Moss, S. B., and Gerton, G. L. (2008). Recombinant mouse sperm ZP3-binding protein (ZP3R/sp56) forms a high order oligomer that binds eggs and inhibits mouse fertilization in vitro. J Biol Chem *283*, 12438-12445.

Center, R. J., Kobe, B., Wilson, K. A., Teh, T., Howlett, G. J., Kemp, B. E., and Poumbourios, P. (1998). Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein. Protein Sci *7*, 1612-1619.

Chaudhary, S., Pak, J. E., Gruswitz, F., Sharma, V., and Stroud, R. M. (2012). Overexpressing human membrane proteins in stably transfected and clonal human embryonic kidney 293S cells. Nat Protoc *7*, 453-466.

Chayen, N. E. (1998). Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. Acta Crystallogr D Biol Crystallogr *54*, 8-15.

Chen, M. S., Tung, K. S., Coonrod, S. A., Takahashi, Y., Bigler, D., Chang, A., Yamashita, Y., Kincade, P. W., Herr, J. C., and White, J. M. (1999). Role of the integrin-associated protein CD9 in binding between sperm ADAM 2 and the egg integrin α6ß1: implications for murine fertilization. Proc Natl Acad Sci U S A *96*, 11830-11835.

Cheng, A., Le, T., Palacios, M., Bookbinder, L. H., Wassarman, P. M., Suzuki, F., and Bleil, J. D. (1994). Sperm-egg recognition in the mouse: characterization of sp56, a sperm protein having specific affinity for ZP3. J Cell Biol *125*, 867-878.

Chianchi, M. (1998). Leonardo, the Anatomy (Florence: Giunti Editore S.p.A.).

Clark, N. L., Gasper, J., Sekino, M., Springer, S. A., Aquadro, C. F., and Swanson, W. J. (2009). Coevolution of interacting fertilization proteins. PLoS Genet *5*, e1000570.

Delucas, L. J., Hamrick, D., Cosenza, L., Nagy, L., McCombs, D., Bray, T., Chait, A., Stoops, B., Belgovskiy, A., William Wilson, W., Parham, M., and Chernov, N. (2005). Protein crystallization: virtual screening and optimization. Prog Biophys Mol Biol *88*, 285-309.

Diller, T. C., Shaw, A., Stura, E. A., Vacquier, V. D., and Stout, C. D. (1994). Acid pH crystallization of the basic protein lysin from the spermatozoa of red abalone (*Haliotis rufescens*). Acta Crystallogr D Biol Crystallogr *50*, 620-626.

Duplay, P., Bedouelle, H., Fowler, A., Zabin, I., Saurin, W., and Hofnung, M. (1984). Sequences of the malE gene and of its product, the maltose-binding protein of Escherichia coli K12. J Biol Chem *259*, 10606-10613.

E. Hochuli, W. Bannwarth, H. Döbeli, R. Gentz1 & D. Stüber (1988). Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. Nature Biotechnology *6*, 1321-1325.

Edward C. Metz, R. E. K., Hiroko Yanagimac H, and Stephen R. Palumbi (1994). Fertilization between closely related sea urchins Is blocked by incompatibilities during sperm-egg attachment and early stages of fusion. Biol. Bull. *187*, 23-34.

Florman, H. M., and First, N. L. (1988). The regulation of acrosomal exocytosis. I. Sperm capacitation is required for the induction of acrosome reactions by the bovine zona pellucida *in vitro.* Dev Biol *128*, 453-463.

Florman, H. M., and Storey, B. T. (1982). Mouse gamete interactions: the zona pellucida is the site of the acrosome reaction leading to fertilization in vitro. Dev Biol *91*, 121-130.

Foster, J. A., Friday, B. B., Maulit, M. T., Blobel, C., Winfrey, V. P., Olson, G. E., Kim, K. S., and Gerton, G. L. (1997). AM67, a secretory component of the guinea pig sperm acrosomal matrix, is related to mouse sperm protein sp56 and the complement component 4-binding proteins. J Biol Chem *272*, 12714-12722.

Fénelon, F. D. S. D. L. M. (1825). Lives of the ancient philosophers. 1-301.

Fridberger, A., Sundelin, J., Vacquier, V. D., and Peterson, P. A. (1985). Amino acid sequence of an egg-lysin protein from abalone spermatozoa that solubilizes the vitelline layer. J Biol Chem *260*, 9092-9099.

Hünefeld, F. L. (1840). Der chemismus in der thierischen organisation. 160-161.

Ruthven, G. H. L. A. C. R. J. (1956). The separation of substances and estimation of their relative molecular sizes by the use of columns of starch in water. Biochem J. 665-674.

Gahlay, G., Gauthier, L., Baibakov, B., Epifano, O., and Dean, J. (2010). Gamete recognition in mice depends on the cleavage status of an egg's zona pellucida protein. Science *329*, 216-219.

Galindo, B. E., Moy, G. W., Swanson, W. J., and Vacquier, V. D. (2002). Full-length sequence of VERL, the egg vitelline envelope receptor for abalone sperm lysin. Gene *288*, 111-117.

Galindo, B. E., Vacquier, V. D., and Swanson, W. J. (2003). Positive selection in the egg receptor for abalone sperm lysin. Proc Natl Acad Sci U S A *100*, 4639-4643.

Galkin, O., and Vekilov, P. G. (2000). Control of protein crystal nucleation around the metastable liquid-liquid phase boundary. Proc Natl Acad Sci U S A *97*, 6277-6281.

Geisse, S., and Henke, M. (2005). Large-scale transient transfection of mammalian cells: a newly emerging attractive option for recombinant protein production. J Struct Funct Genomics *6*, 165-170.

Gosavi, R. A., Mueser, T. C., and Schall, C. A. (2008). Optimization of buffer solutions for protein crystallization. Acta Crystallogr D Biol Crystallogr *64*, 506-514.

Goudet, G., Mugnier, S., Callebaut, I., and Monget, P. (2008). Phylogenetic analysis and identification of pseudogenes reveal a progressive loss of zona pellucida genes during evolution of vertebrates. Biol Reprod *78*, 796-806.

Graham, F. L., Smiley, J., Russell, W. C., and Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol *36*, 59-74.

Grayson, P. (2015). Izumo1 and Juno: the evolutionary origins and coevolution of essential sperm-egg binding partners. R Soc Open Sci *2*, 150296.

Gupta, S. K. (2015). Role of zona pellucida glycoproteins during fertilization in humans. J Reprod Immunol *108*, 90-97.

Han, L., Monné, M., Okumura, H., Schwend, T., Cherry, A. L., Flot, D., Matsuda, T., and Jovine, L. (2010). Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. Cell *143*, 404-415.

Han, L., Nishimura, K., Sadat Al Hosseini, H., Bianchi, E., Wright, G. J., and Jovine, L. (2016). Divergent evolution of vitamin B_9 binding underlies Juno-mediated adhesion of mammalian gametes. Curr Biol *26*, R100-1.

Hawks, J., Wang, E. T., Cochran, G. M., Harpending, H. C., and Moyzis, R. K. (2007). Recent acceleration of human adaptive evolution. Proc Natl Acad Sci U S A *104*, 20753-20758.

Hellberg, M. E., and Vacquier, V. D. (1999). Rapid evolution of fertilization selectivity and lysin cDNA sequences in teguline gastropods. Mol Biol Evol *16*, 839-848.

Hertz, T., Nolan, D., James, I., John, M., Gaudieri, S., Phillips, E., Huang, J. C., Riadi, G., Mallal, S., and Jojic, N. (2011). Mapping the landscape of host-pathogen coevolution: HLA class I binding and its relationship with evolutionary conservation in human and viral proteins. J Virol *85*, 1310-1321.

Huang, H. L., Lv, C., Zhao, Y. C., Li, W., He, X. M., Li, P., Sha, A. G., Tian, X., Papasian, C. J., Deng, H. W., Lu, G. X., and Xiao, H. M. (2014). Mutant ZP1 in familial infertility. N Engl J Med *370*, 1220-1226.

Inoue, N., Hamada, D., Kamikubo, H., Hirata, K., Kataoka, M., Yamamoto, M., Ikawa, M., Okabe, M., and Hagihara, Y. (2013). Molecular dissection of IZUMO1, a sperm protein essential for sperm-egg fusion. Development *140*, 3221-3229.

Inoue, N., Ikawa, M., Isotani, A., and Okabe, M. (2005). The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. Nature *434*, 234-238.

Jenkins, N., Murphy, L., and Tyther, R. (2008). Post-translational modifications of recombinant proteins: significance for biopharmaceuticals. Mol Biotechnol *39*, 113-118.

Jones, R., and Brown, C. R. (1987). Identification of a zona-binding protein from boar spermatozoa as proacrosin. Exp Cell Res *171*, 503-508.

Jovine, L., Darie, C. C., Litscher, E. S., and Wassarman, P. M. (2005). Zona pellucida domain proteins. Annu Rev Biochem *74*, 83-114.

Jovine, L., Janssen, W. G., Litscher, E. S., and Wassarman, P. M. (2006). The PLAC1-homology region of the ZP domain is sufficient for protein polymerisation. BMC Biochem *7*, 11.

Jovine, L., Qi, H., Williams, Z., Litscher, E., and Wassarman, P. M. (2002). The ZP domain is a conserved module for polymerization of extracellular proteins. Nat Cell Biol *4*, 457-461.

Kaji, K., Oda, S., Miyazaki, S., and Kudo, A. (2002). Infertility of CD9-deficient mouse eggs is reversed by mouse CD9, human CD9, or mouse CD81; polyadenylated mRNA injection developed for molecular analysis of sperm-egg fusion. Dev Biol *247*, 327-334.

Kaji, K., Oda, S., Shikano, T., Ohnuki, T., Uematsu, Y., Sakagami, J., Tada, N., Miyazaki, S., and Kudo, A. (2000). The gamete fusion process is defective in eggs of Cd9-deficient mice. Nat Genet *24*, 279-282.

Kresge, N., Vacquier, V. D., and Stout, C. D. (2000a). 1.35 and 2.07 Å resolution structures of the red abalone sperm lysin monomer and dimer reveal features involved in receptor binding. Acta Crystallogr D Biol Crystallogr *56*, 34-41.

Kresge, N., Vacquier, V. D., and Stout, C. D. (2000b). The high resolution crystal structure of green abalone sperm lysin: implications for species-specific binding of the egg receptor. J Mol Biol *296*, 1225-1234.

Kresge, N., Vacquier, V. D., and Stout, C. D. (2001a). Abalone lysin: the dissolving and evolving sperm protein. Bioessays *23*, 95-103.

Kresge, N., Vacquier, V. D., and Stout, C. D. (2001b). The crystal structure of a fusagenic sperm protein reveals extreme surface properties. Biochemistry *40*, 5407-5413.

Lee, Y. H., Ota, T., and Vacquier, V. D. (1995). Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol Biol Evol *12*, 231-238.

Lewis, C. A., Talbot, C. F., and Vacquier, V. D. (1982). A protein from abalone sperm dissolves the egg vitelline layer by a nonenzymatic mechanism. Dev Biol *92*, 227-239.

Lillie, F. R. (1914). Studies of Fertilization. VI. The Mechanism of Fertilization in *Arbacia*. J. Exp. Zool. *16*, 523-590.

Lou, D. I., McBee, R. M., Le, U. Q., Stone, A. C., Wilkerson, G. K., Demogines, A. M., and Sawyer, S. L. (2014). Rapid evolution of BRCA1 and BRCA2 in humans and other primates. BMC Evol Biol *14*, 155.

Luft, J. R., Wolfley, J. R., Said, M. I., Nagel, R. M., Lauricella, A. M., Smith, J. L., Thayer, M. H., Veatch, C. K., Snell, E. H., Malkowski, M. G., and Detitta, G. T. (2007). Efficient optimization of crystallization conditions by manipulation of drop volume ratio and temperature. Protein Sci *16*, 715-722.

Lyon, J. D., and Vacquier, V. D. (1999). Interspecies chimeric sperm lysins identify regions mediating species-specific recognition of the abalone egg vitelline envelope. Dev Biol *214*, 151-159.

Makrides, S. C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol Rev *60*, 512-538.

Messick, T., and Marmorstein, R. (2004). Crystallization of macromolecules. Curr Protoc Protein Sci *Chapter 17*, Unit 17.4.

Messier, W., and Stewart, C. B. (1994). Molecular evolution. Dissolving the barriers. Curr Biol *4*, 911-913.

Mintseris, J., and Weng, Z. (2005). Structure, function, and evolution of transient and obligate protein-protein interactions. Proc Natl Acad Sci U S A *102*, 10930-10935.

Miyado, K., Yamada, G., Yamada, S., Hasuwa, H., Nakamura, Y., Ryu, F., Suzuki, K., Kosai, K., Inoue, K., Ogura, A., Okabe, M., and Mekada, E. (2000). Requirement of CD9 on the egg plasma membrane for fertilization. Science *287*, 321-324.

Monné, M., Han, L., and Jovine, L. (2006). Tracking down the ZP domain: From the mammalian zona pellucida to the molluscan vitelline envelope. Semin Reprod Med *24*, 204-216.

Monne, M., Han, L., Schwend, T., Burendahl, S., and Jovine, L. (2008). Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. Nature *456*, 653-657.

Morgan, C. C., Loughran, N. B., Walsh, T. A., Harrison, A. J., and O'Connell, M. J. (2010). Positive selection neighboring functionally essential sites and disease-implicated regions of mammalian reproductive proteins. BMC Evol Biol *10*, 39.

Morse, D. E., Duncan, H., Hooker, N., and Morse, A. (1977). Hydrogen peroxide induces spawning in mollusks, with activation of prostaglandin endoperoxide synthetase. Science *196*, 298-300.

Muro, Y., Buffone, M. G., Okabe, M., and Gerton, G. L. (2012). Function of the acrosomal matrix: zona pellucida 3 receptor (ZP3R/sp56) is not essential for mouse fertilization. Biol Reprod *86*, 1-6.

Nikaido, H. (1994). Maltose transport system of Escherichia coli: an ABC-type transporter. FEBS Lett *346*, 55-58.

Nishimura, K., Han, L., Bianchi, E., Wright, G. J., de Sanctis, D., and Jovine, L. (2016). The structure of sperm Izumo1 reveals unexpected similarities with Plasmodium invasion proteins. Curr Biol

Ohto, U., Ishida, H., Krayukhina, E., Uchiyama, S., Inoue, N., and Shimizu, T. (2016). Structure of IZUMO1-JUNO reveals sperm-oocyte recognition during mammalian fertilization. Nature

Okumura, H., Kohno, Y., Iwata, Y., Mori, H., Aoki, N., Sato, C., Kitajima, K., Nadano, D., and Matsuda, T. (2004). A newly identified zona pellucida glycoprotein, ZPD, and dimeric ZP1 of chicken egg envelope are involved in sperm activation on sperm-egg interaction. Biochem J *384*, 191-199.

Okumura, H., Sato, T., Sakuma, R., Fukushima, H., Matsuda, T., and Ujita, M. (2015). Identification of distinctive interdomain interactions among ZP-N, ZP-C and other

domains of zona pellucida glycoproteins underlying association of chicken egg-coat matrix. FEBS Open Bio *5*, 454-465.

Parker, G. A., and Partridge, L. (1998). Sexual conflict and speciation. Philos Trans R Soc Lond B Biol Sci *353*, 261-274.

Rankin, T., Talbot, P., Lee, E., and Dean, J. (1999). Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. Development *126*, 3847-3855.

Riffell, J. A., Krug, P. J., and Zimmer, R. K. (2002). Fertilization in the sea: the chemical identity of an abalone sperm attractant. J Exp Biol *205*, 1439-1450.

Sakai YT, S. Y., Haino-Fukushima K (1982). Fine structural changes in the acrosome reaction of the Japanese abalone, Haliotis discus. Develop Growth Differ *5*, 273-282.

Schlaeger, E. J., and Christensen, K. (1999). Transient gene expression in mammalian cells grown in serum-free suspension culture. Cytotechnology *30*, 71-83.

Shaw, A., Fortes, P. A., Stout, C. D., and Vacquier, V. D. (1995). Crystal structure and subunit dynamics of the abalone sperm lysin dimer: egg envelopes dissociate dimers, the monomer is the active species. J Cell Biol *130*, 1117-1125.

Shaw, A., McRee, D. E., Vacquier, V. D., and Stout, C. D. (1993). The crystal structure of lysin, a fertilization protein. Science *262*, 1864-1867.

Swanson, W. J., and Vacquier, V. D. (1995). Liposome fusion induced by a M(r) 18,000 protein localized to the acrosomal region of acrosome-reacted abalone spermatozoa. Biochemistry *34*, 14202-14208.

Swanson, W. J., and Vacquier, V. D. (1997). The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. Proc Natl Acad Sci U S A *94*, 6724-6729.

Swanson, W. J., and Vacquier, V. D. (1998). Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. Science *281*, 710-712.

Swanson, W. J., and Vacquier, V. D. (2002). The rapid evolution of reproductive proteins. Nat Rev Genet *3*, 137-144.

Swanson, W. J., Yang, Z., Wolfner, M. F., and Aquadro, C. F. (2001). Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci U S A *98*, 2509-2514.

Szmelcman, S., Clément, J. M., Jehanno, M., Schwartz, O., Montagnier, L., and Hofnung, M. (1990). Export and one-step purification from *Escherichia coli* of a MalE-CD4 hybrid protein that neutralizes HIV in vitro. J Acquir Immune Defic Syndr *3*, 859-872.

Tanphaichitr, N., Carmona, E., Bou Khalil, M., Xu, H., Berger, T., and Gerton, G. L. (2007). New insights into sperm-zona pellucida interaction: involvement of sperm lipid rafts. Front Biosci *12*, 1748-1766.

Thakur, A. S., Newman, J., Martin, J. L., and Kobe, B. (2008). Increasing protein crystallization screening success with heterogeneous nucleating agents. Methods Mol Biol *426*, 403-409.

Tranter, R., Read, J. A., Jones, R., and Brady, R. L. (2000). Effector sites in the three-dimensional structure of mammalian sperm ß-acrosin. Structure *8*, 1179-1188.

Tsubamoto, H., Hasegawa, A., Nakata, Y., Naito, S., Yamasaki, N., and Koyama, K. (1999). Expression of recombinant human zona pellucida protein 2 and its binding capacity to spermatozoa. Biol Reprod *61*, 1649-1654.

Unger, T., and Peleg, Y. (2012). Recombinant protein expression in the baculovirus-infected insect cell system. Methods Mol Biol *800*, 187-199.

Vacquier, V. D., Carner, K. R., and Stout, C. D. (1990). Species-specific sequences of abalone lysin, the sperm protein that creates a hole in the egg envelope. Proc Natl Acad Sci U S A *87*, 5792-5796.

Vacquier, V. D., and Lee, Y. H. (1993). Abalone sperm lysin: unusual mode of evolution of a gamete recognition protein. Zygote *1*, 181-196.

Vacquier, V. D., and Swanson, W. J. (2011). Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. Cold Spring Harb Perspect Biol *3*, a002931.

Van Wettere, A., and Lewbart, G. A. (2007). Cytologic diagnosis of diseases of invertebrates. Vet Clin North Am Exot Anim Pract *10*, 235-54.

Von Eichborn, J., Günther, S., and Preissner, R. (2010). Structural features and evolution of protein-protein interactions. Genome Inform *22*, 1-10.

Walker, I. H., Hsieh, P. C., and Riggs, P. D. (2010). Mutations in maltose-binding protein that alter affinity and solubility properties. Appl Microbiol Biotechnol *88*, 187-197.

Wassarman, P. M. (1990). Profile of a mammalian sperm receptor. Development *108*, 1-17.

Wassarman, P. M. (2008). Zona pellucida glycoproteins. J Biol Chem *283*, 24285-24289.

Wassarman, P. M. (2014). Reproductive biology: Sperm protein finds its mate. Nature *508*, 466-467.

Wassarman, P. M., Jovine, L., and Litscher, E. S. (2001). A profile of fertilization in mammals. Nat Cell Biol *3*, E59-64.

Wassarman, P. M., and Litscher, E. S. (2001). Towards the molecular basis of sperm and egg interaction during mammalian fertilization. Cells Tissues Organs *168*, 36-45.

Wassarman, P. M., and Litscher, E. S. (2008). Mammalian fertilization: the egg's multifunctional zona pellucida. Int J Dev Biol *52*, 665-676.

Wolfner, M. F. (1997). Tokens of love: functions and regulation of drosophila male accessory gland products. Insect Biochem Mol Biol *27*, 179-192.

Wurm, F. M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol *22*, 1393-1398.

Yamagata, K., Murayama, K., Okabe, M., Toshimori, K., Nakanishi, T., Kashiwabara, S., and Baba, T. (1998). Acrosin accelerates the dispersal of sperm acrosomal proteins during acrosome reaction. J Biol Chem *273*, 10470-10474.

Yanagimachi, R. (1994). Fertility of mammalian spermatozoa: its development and relativity. Zygote *2*, 371-372.

Yanagimachi, R. (2011). Mammalian sperm acrosome reaction: where does it begin before fertilization. Biol Reprod *85*, 4-5.

Yang, and Bielawski (2000). Statistical methods for detecting molecular adaptation. Trends Ecol Evol *15*, 496-503.

Yang, Z., Swanson, W. J., and Vacquier, V. D. (2000). Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Mol Biol Evol *17*, 1446-1455.

Zhu, Y., Zhu, L. N., Guo, R., Cui, H. J., Ye, S., and Fang, Q. (2014). Nanoliter-scale protein crystallization and screening with a microfluidic droplet robot. Sci Rep *4*, 5046.