Large-Scale Studies of the Protein Biochemistry Regulating Meiotic Exit and Fertilization

Large-scale studies of the protein biochemistry regulating meiotic exit and fertilization

A dissertation presented

by

Marc S. Presler

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

April 2018

Dissertation Advisor: Dr. Marc W. Kirschner Marc S. Presler

Large-scale studies of the protein biochemistry regulating meiotic exit and fertilization

Abstract

Dynamic post-translational modifications are central to coordinating complex biological processes. This is especially evident after fertilization, where the modification or destruction of the egg's existing proteins is sufficient to guide the embryo through early embryogenesis. In particular, protein phosphorylation and degradation drive the rapid release of the meiotic arrest and numerous other events, such as the blocks to polyspermy. Despite decades of study, the full extent of protein loss and phospho-regulation following fertilization is still unknown. This is largely because comprehensive and quantitative analysis of such biochemical regulation is currently difficult. The following work examines protein and phosphosite dynamics of the fertilization response by mass spectrometry-based proteomics in *Xenopus* eggs, which have long served as an informative cell cycle and developmental model. To interpret the function of the biochemical changes, it is informative to observe quantitative features like absolute rates and the stoichiometry of modifications. This motivated the development of two analytical approaches: the first to estimate protein concentration from label-free proteomics, the second to calculate the stoichiometry of phosphosites from multiplexed proteomics with confidence intervals. The data suggest that the rapid transition after fertilization is achieved by two contrasting and parallel programs: 1) protein degradation is limited to a few low abundance targets. However, this degradation promotes extensive dephosphorylation that occurs over a wide range of abundances during meiotic exit. 2) A large amount of protein is released from the egg into the medium just after fertilization, most likely related to the blocks to polyspermy. Concomitantly, there is a substantial increase in phosphorylation, which is likely tied to the signal transduction downstream of the calcium-activated kinases. The analytical approaches developed and demonstrated here are broadly applicable to many classes of post-translational modification and for studies of dynamic biochemical regulation of diverse biological systems.

Contents

Acknowledgements

Science is a human endeavor. Everything that I have accomplished is linked to my environment, my colleagues, and my friends. None of this is possible without their intellect, patience, and kindness.

I thank my advisor Marc Kirschner for his support and guidance during my graduate work. Marc's inquisitiveness and curiosity is infectious. It has been my privilege to perform my graduate studies here.

I thank Tim Mitchison, Galit Lahav, Mike Springer, Willi Haas, Chris Field, Sean Megason, Leon Peshkin, and Angela DePace for valuable discussions and for their thoughtfulness. I also thank my examining committee Hazel Sive, Jarrod Marto, and Michael Blower for their insightful and incisive commentary that greatly improved the dissertation. I thank Davi Van Vactor, who advocated for me and encouraged me throughout my early graduate school years. I am fortunate to have had a program head that cared so deeply about each student. I thank Diane Shakes for cultivating my interest in science, and for the ensuing years of careful direction and inspiration. You are such a role model; it is hard to overstate your influence on me. I thank Allon Klein for the countless hours of work-shopping proposals, fine-tuning paragraph structures, and polishing up scientific ideas right up until a desperate sprint to catch the last M2 of the night. To Martin Wühr, I owe you so much. I hope the work of this thesis does justice by all the time, effort, and generosity you gave to me.

I thank Becky Ward for countless conversations from immunology to Dorothy Parker, and of course for introducing me to some fabulous eating establishments. I thank Scott Gruver for elevating our space with good humor, intellect, and honesty. The lab would never have felt like home without you as my baymate. To Seungeun, so rarely does one find such brilliance, humility, and grace in one person. You would always spot the hole in my argument and smile kindly as I realized you were right. To Christine Hagan, I so admired your meticulousness and skill; you saved many experiments with you willingness to answer even the most trivial biochemistry questions. To Jessica Gray, Tao Wu, Wenzhe Ma, Doaa Megahed, Ying Lu, Bai Luan, Ana Hernandez, Bob Freeman, and Rick Deibler— I learned something from you in every single conversation we had. To Kate Hodgins, your guidance is irreplaceable. I cannot imagine my time at Harvard without your perspective. To Meghan Van Orden, thank you for all the great chats and the help corralling schedules and planning events.

To Ylaine Gerardin, John Ingraham, Matt Sonnett, Clarissa Scholes, Peter Koch, Chiara Ricci-Tam, Anna Green, Ilana Kelsey, Huixin Xu, Hattie Chung, Jose Reyes, Ivana Cvijovic, Laura Stone, Tami Lieberman, Javier Pineda, Peg Coughlin, Jennifer Yeh, Jeehae Park, Naren Tallapragada, James Briggs, Jue Wang, Adam Riesselman, Ed Hendel, Zach Collins, James Pelletier, Ivana Gasic—you challenged me to think deeper about my work and were quick to explain or discuss any problem I came to you with. You were so often my teachers. My education in is incomplete without you.

To Elaine, your quirky wit and artistic sensibility are one-of-a-kind. I would look up to you even if you weren't an inch taller than me (which of course you are, and you'll never let me forget it!). To Jamie, you are like your dancing— graceful, natural, and always in time. Your ability to appreciate people as they are and deal with those that differ from you is special; our friend-group would have fallen apart without you. To John, your insight and sense of humor are one-of-a-kind. I will never forget our adventures to Foxboro in the freezing New England winter, talking about science on the long train ride. To Crystal, I've always been struck by your fortitude and sensitivity; I felt understood and heard whenever we spoke. To Evan, thank you for tolerating all the times I turned your name into odd little songs; you are without fail kind and thoughtful. To Deepali, you are so genuine and full of light; you so often provided some perspective I was missing. To Katy, I am so lucky we met in Boston; it was always effortless to spend time together and I wouldn't trade a moment of it. I would not have made it far if not for you.

To Evi, I thank you for your originality, compassion, and the ever appropriate yet unrestrained outbursts. You were a wonderful colleague and an even better friend. This thesis and my graduate school experience were so improved by our friendship.

To Claire, it is difficult to express the joy I have felt becoming close to you. I admire you so much for your effortless meticulousness, adaptability, and warmth. You are at once exceptionally capable and loving, and all that I do is improved by your presence. I would come to graduate school all over again simply for the chance that we might meet again on a cool and lovely night in Maine.

To my parents, I only become more grateful to you as the years go on. You thought nothing of yourselves, and exude warmth and strength. I have been given perhaps the greatest gift that a parent can give a child— to never doubt, even for a second, that I was loved and cared for. You are the ink this thesis is written with. The pages are blank without you.

To Monica Brezinski Potkay and Jerry A. Wald

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

Little Gidding
T.S. Eliot, 1942

Chapter 1:

Introduction

1.1 Preface

1.1.1 *Ex ova, modern biologia*— from the egg comes modern biology

In 1691, William Harvey published *Exercitationes de generatione animalium* ("On Animal Generation"). Studies of deer, rabbits, and other animals led him to postulate *Ex Ovo Omnia*—all life comes from eggs—from the noblest inhabitants of the English forest to the humble worms in the soil. This work built on the much older observations of chick embryos in Aristotle's Book IV of the *History of Animals*; some documentation of embryos are older still, such as the *Garbha Upanishad* from Indian antiquity [1]. While a more molecular perspective would not follow until later centuries, these works speak to the central role that embryos hold in the scientific heritage.

The utility of studying eggs is notably showcased by one of the earliest events in embryology—fertilization. In 1919, embryologist Frank Lillie stated that "there is perhaps no phenomenon in the field of biology that touches so many fundamental questions as the union of the germ cells in the act of fertilization" [2]. Indeed, decades of careful observation of amphibian embryos and others revealed that the fusion of the sperm and egg initiates an impressive transition. A single, localized event at the membrane of the egg triggers the blocks to polyspermic fertilization, the initiation of the cell cycle and completion of meiosis, and the beginning of morphogenetic process of embryonic development

The study of the molecular mechanisms driving these events has been fruitful for revealing general principles of biological regulation. This is true most famously for the cell cycle, where the core components and mechanisms are conserved throughout eukaryotes; the same is true for many other processes in the egg as well. Thus, just as all animals come from eggs, much of modern molecular biology can trace its roots back to the study of eggs.

Discoveries in embryology are particularly enabled and limited by technology, such as the quality of microscopes or molecule probes. For example, even in a long-studied system like the amphibian egg, it is still unclear how the diverse processes that occur after fertilization are mechanistically linked. One reason is the difficulty of quantitative and comprehensive biochemical measurements. This thesis relies on egg's strength as a classic model (in this case, *Xenopus laevis*) to develop and apply new technologies that allow for large-scale biochemical studies during fertilization and other biological settings.

1.2 Principles of the protein-based regulation during fertilization

1.2.1 Fertilization: a unifying event in metazoan life cycles

Despite the astonishing diversity of metazoans in form and features, all animals begin as single cells called eggs. These shared beginnings include the remarkable events that follow fertilization. Some of the earliest descriptions of fertilization date back to microscopy and cytological studies in the late 19th and early 20th century [2]. The egg is described as initially held in a state of stasis [3]. Upon fusion of the egg and the sperm, the egg undergoes an abrupt transition that triggers several large-scale morphological changes. These include the visible swelling of the outer regions of the egg to form the so-called “fertilization envelope” [3] and later, the fusion of the maternal and paternal genomes, and the rapid cleavage divisions of the egg [4, 5]. Though the transition from the state of stasis is abrupt, it was understood early on that the fundamental character of the egg had changed in some irreversible manner. The search for the molecular basis of these changes motivated research programs for the next century.

I chose this area of study for similar reasons that inspired the early cytology work in eggs: a sense that the processes underlying these earliest steps in embryonic development were not just important for their own sake, but would reflect general and conserved biological principles. The following chapter reviews the advancements in the cell biology of fertilization, the principles of regulation that such work indeed revealed, and some of the elusive question that I sought to answer.

1.2.2 Fertilization relieves a cell cycle arrest

The cell cycle, which directs the process of cell division, plays a major role in gametogenesis and the early embryogenesis. The state of stasis described in the late 19th century is an actively maintained arrest of the cell cycle. Gametogenesis in animal eggs features a special cell cycle called meiosis, where the resulting germ cells (i.e., sperm or egg) are haploid. The physiological cell cycle arrests allow time for the gamete-specific morphogenesis (e.g., substantial growth in the egg) to occur and help to set the appropriate cell cycle state to allow for fertilization [6]. In amphibian, fish, and mammalian eggs, the cell cycle is initially arrested in metaphase of the first meiosis, which is relieved by progesterone exposure in a process called oocyte maturation. After maturation, the egg is again arrested in metaphase II until fertilization. Eggs are not fertilized at the same cell cycle stage in all animals. Bivalve clams, for example,

are fertilized during the prophase I of meiosis [7] whereas sea urchins complete meiosis II and are fertilized during interphase [8].

Fertilization restarts the cell cycle of the egg through initiating entry into anaphase and exit from meiosis (Figure 1.1). The release from the arrest is calcium sensitive [9, 10], and triggers the sister chromatids, initially aligned at the spindle midzone, to separate and move towards the spindle poles [11]. The molecular strategies of the cell cycle are broadly conserved for regulating cell division across eukaryotes.

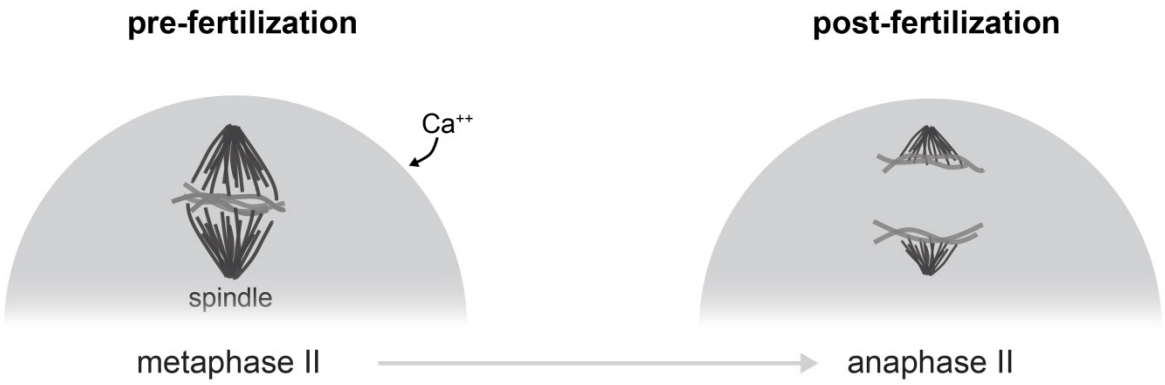

Figure 1.1 The meiotic arrest in egg cells. Prior to fertilization, eggs are arrested most commonly in metaphase of the second meiotic division, where chromosomes are aligned at the spindle midzone. Fertilization leads to elevated calcium levels, which promote the alleviation of the arrest and entry into anaphase. The sister chromatids separate, and half of the chromosomes are extruded as the second polar body (not shown), leaving a haploid genome. Rapid mitotic divisions follow ~one hour later.

1.2.3 Strategies to prevent multiple fertilization events

Fertilization is a unique process in development that employs additional regulation beyond the cell cycle. It is critical that the embryo contain the proper number of chromosomes after fusion with the sperm. In many organisms, the two major strategies for avoiding aneuploidy are the fast and slow blocks to polyspermy, which ensure that only one sperm enters the egg. There are other strategies, such as the polyspermic fertilization in urodeles and some birds which allow multiple sperm to enter the egg [12] and ensure euploidy by other means. In most animals, including frogs, fish, and mice, the strategy is instead to prevent any additional sperm entry after the initial fertilization.

The mechanism of a block to polyspermy was described in early observations as a “wave of negativity” the spread across the cortical (i.e., outer) surface of the egg minutes after fertilization [13, 14]. The fusion of the sperm causes a transient calcium wave to be generated in the egg cytoplasm [15-17], propagating from the animal pole (i.e., top) of the egg leading to the contraction of the cortex (Figure 1.2) [18, 19]. The main mechanism to prevent polyspermy is a large-scale secretion event, caused by the fusion of the cortical granules docked on the surface membrane of the egg. The contents of the cortical granules promote the elevation of the vitelline membrane, a protective meshwork surrounding the egg. This creates the “perivitelline space,” providing a physical barrier for further sperm entry (Figure 1.2) that is identical to the early observations of the “fertilization envelope.” The secretory events are typically called the “slow block” to polyspermy. After the blocks to polyspermy, the outer membrane is considered “hardened” to additional sperm binding.

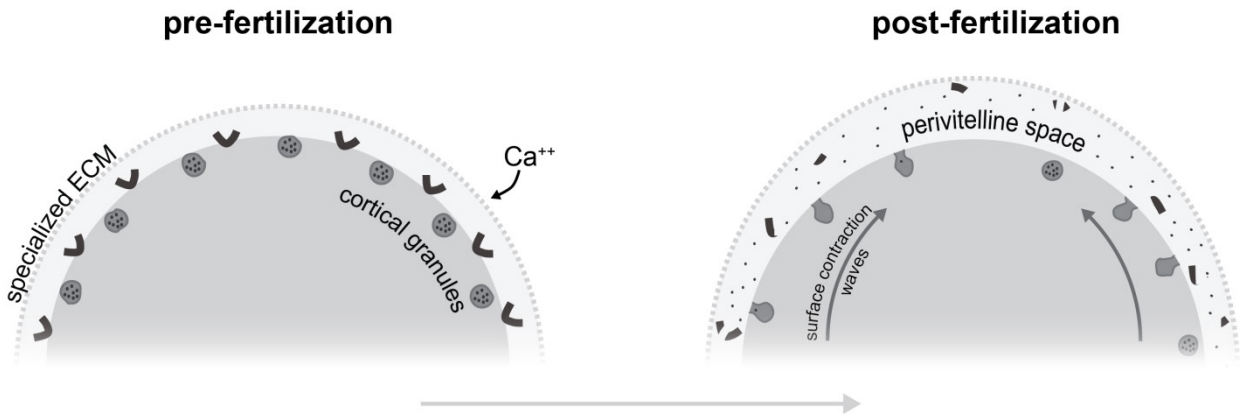

Figure 1.2 The blocks to polyspermic fertilization. A specialized extra-cellular matrix coats the outside of the egg that serves as a protective physical barrier (i.e., vitelline membrane) and enhances sperm-binding. After elevated calcium levels are triggered by sperm entry, specialized vesicles called cortical granules fuse with the outer membrane. The expelled contents cause the perivitelline space to become hydrated, which elevates vitelline membrane to create a perivitelline space. Additional changes occur, such as the destruction of sperm binding sites and the dramatic surface contraction waves on the outer membrane. These cortical changes prevent multiple fertilization events to ensure proper ploidy of the embryo during development.

The blocks to polyspermic fertilization occur in parallel with the events of the cell cycle and meiotic exit. Each program proceeds at different rates, though they are downstream of the common trigger of elevated calcium (Figure 1.3). This combined response to fertilization is called egg activation.

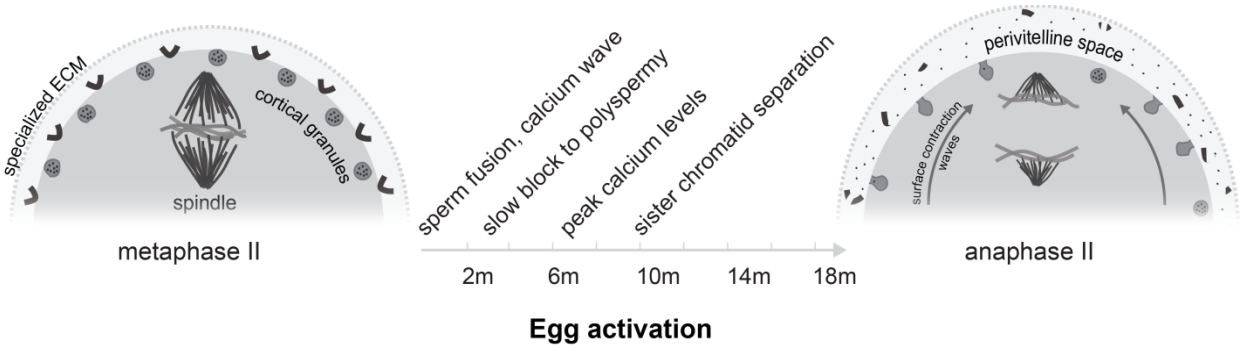

Figure 1.3 The parallel processes of meiotic exit and the blocks to polyspermy are called egg activation. Egg activation is caused by fertilization, but eggs can be activated independently of the sperm, for example with electric shock. The blocks to polyspermy occur within minutes of activation. The initial spike in calcium concentration is transient, and reaches peak concentrations 5-7 minute post-fertilization. Sister chromatid separation begins at approximately ten minutes. The majority of the events of egg activation are completed within 20-30 minutes.

1.2.4 The protein biochemistry regulating egg activation

Many transitions of cell state in embryogenesis are linked with large-scale changes in the transcriptional state of the cell. However, the contribution of transcription is minimal in the earliest steps of embryogenesis. During meiosis, oocytes and eggs are largely transcriptionally quiescent, as shown by experiments performed in *Drosophila*, *Xenopus*, and mice [20-24]. *Xenopus* embryos are viable for several hours into development (until the mid-blastula transition) without zygotic transcription [23]. The nucleus is required for the cleavage divisions, but the periodicity of the cell cycle can be maintained in its absence. This was shown elegantly by constricting a recently fertilized embryo into nucleated and enucleated fragments. The nucleated fragment continued to cleave, while the enucleated fragment showed surface contractions at the same period as cleavages [25]. This emphasizes the importance the maternal dowry of protein and RNA, as well as the sufficiency of post-transcriptional control for the early cell cycle.

Rather than transcription of new RNA, protein biochemistry of the existing molecules in the egg regulates the transition that occurs during egg activation. Specifically, the cell cycle and blocks to polyspermy are driven by the degradation, release, and phosphorylation of proteins (Figure 1.4).

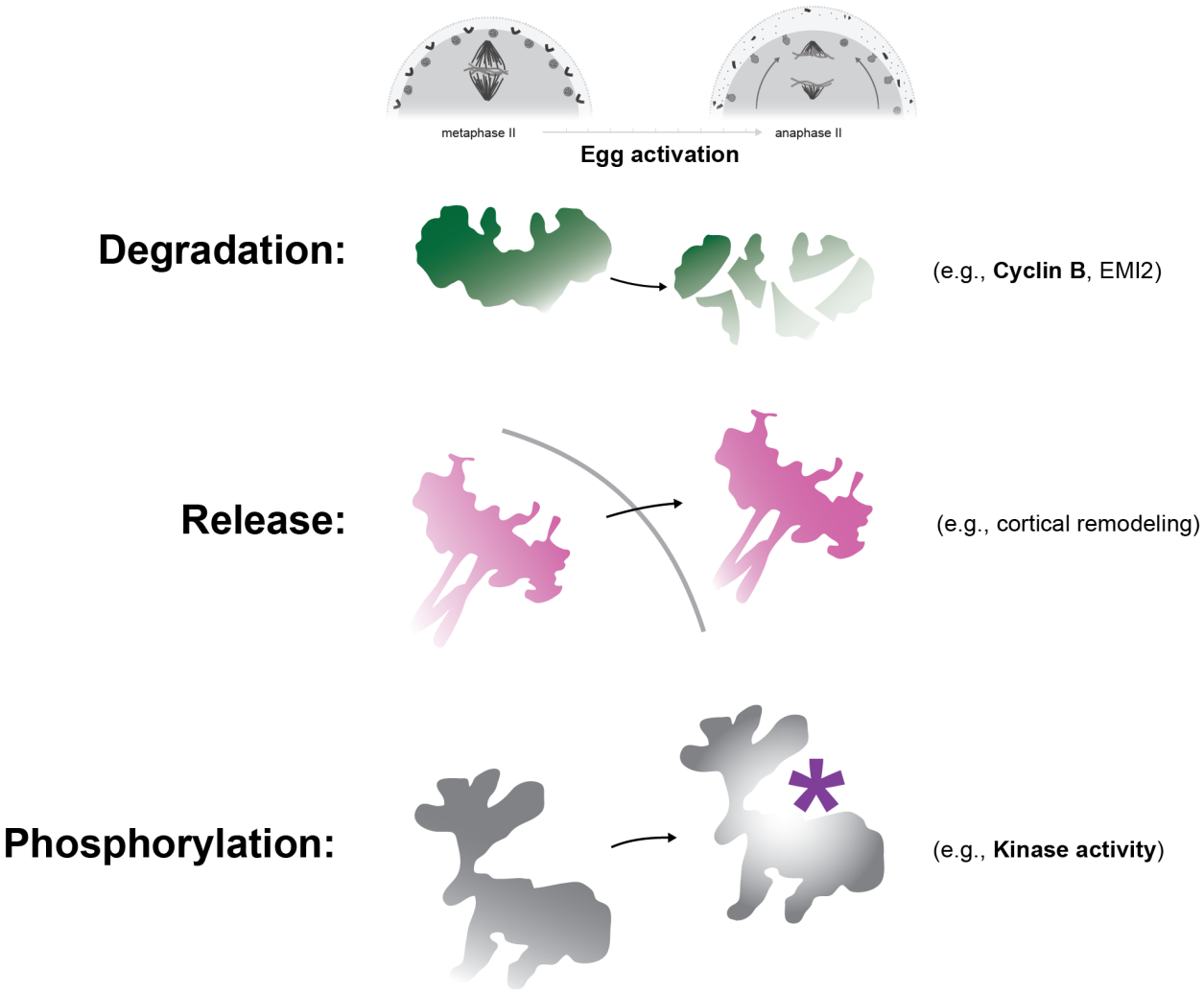

Figure 1.4 Protein biochemistry regulates egg activation. The contribution of new transcription is minimal in the egg. Instead, the transition of cell state after fertilization occurs through the regulation of existing molecules in the cell. The three primary principles of protein regulation are protein degradation, release, and phosphorylation. Degradation (e.g., proteins like Cyclin B and EMI2) are important for promoting meiotic exit. The release of proteins is critical for the cortical remodeling that occurs to inhibit multiple fertilization events. Protein phosphorylation affects the activities and function of many hundreds of proteins through the regulated activity of kinases and phosphatases.

For example, the degradation of proteins (e.g., Cyclin B, EMI2) are necessary to promote entry into anaphase and completion of the meiotic cell cycle [26]. The cortical remodeling events that prevent polyspermic fertilization are largely accomplished by the release of proteins by the cortical granules (Figure 1.2). The mechanisms of degradation and release each contribute to the early observations of the

irreversibility of fertilization. Once important components are lost by either mechanism, the cell cannot easily recover the previous cell state. For both the cell cycle and the blocks to polyspermy, the regulation of kinase activity is essential (e.g., Cyclin-dependent kinases, MAPK response, and calcium-sensitive kinases [15-17, 27, 28]). These kinases are all regulated, directly or indirectly, by the calcium wave.

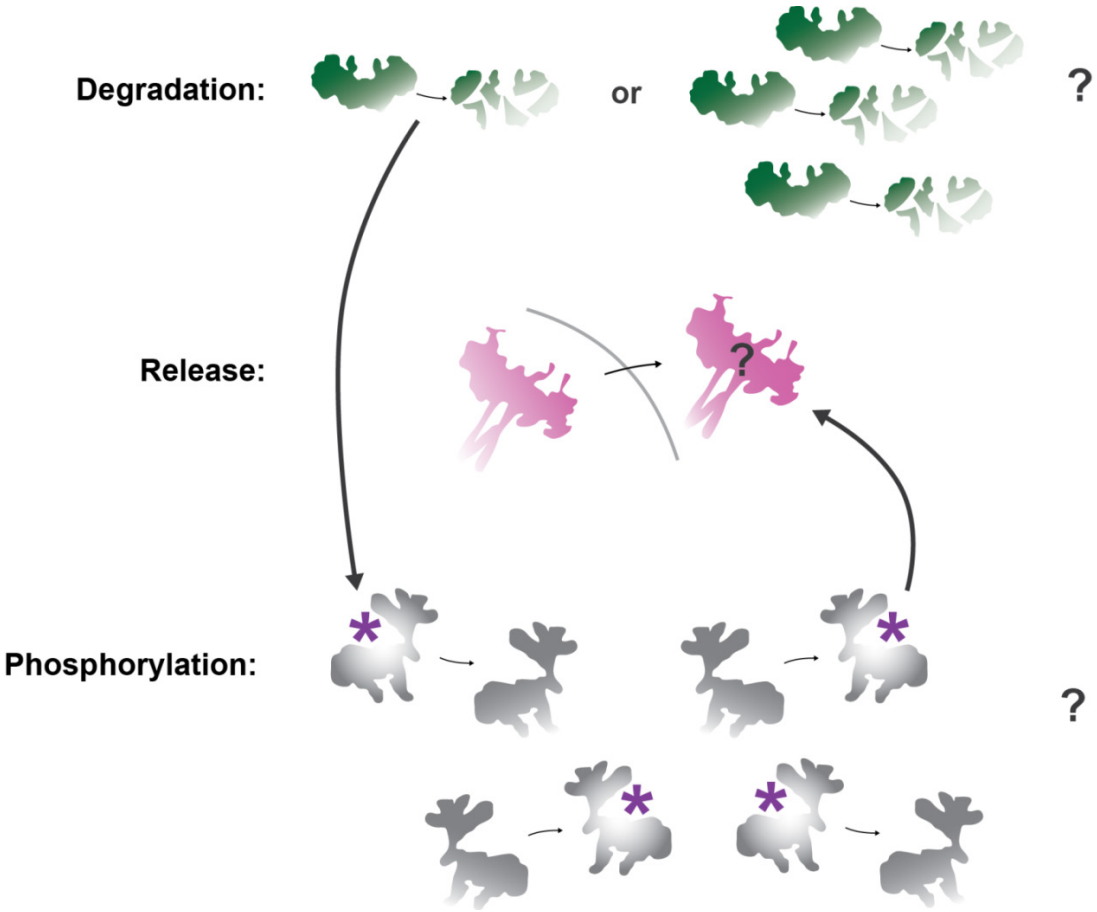

Figure 1.5 Graphical representation of the major biochemical questions in egg activation. Is protein degradation limited to a few important targets, or many dozens? What are the identities and function of the released proteins for preventing polyspermy? What are the functions and dynamics of the many hundreds of protein phosphorylation events that occur in the cell, and how are these modifications linked with the other biochemical mechanisms that control the essential and rapid steps of egg activation?

1.2.5 Many features of the biochemical regulation of egg activation remain unknown

Though many aspects of the biochemical activity occurring of egg activation are known, many fundamental features of the regulation have been difficult to determine from conventional biochemical

experiments (Figure 1.5). While the importance of protein degradation is appreciated, whether a handful or many dozen degradation targets is sufficient to regulate meiotic exit is not known. Thus, the “minimal” set of degradation targets that is sufficient to regulate the cell cycle is still unknown. Further, several important secreted proteins have been identified as part of the blocks to polyspermy, but the understanding of the released proteins, their function, and their upstream signaling components is still limited. The activity of kinases involved with the cell cycle and calcium signaling are known to affect the function of numerous proteins after fertilization [29]. The identities and function of these sites are still largely undiscovered even three decades after the initial observations.

The biochemical programs that drive egg activation are not occurring in isolation, but influence each other in a dynamical fashion. For example, protein degradation promotes meiotic exit by negatively regulating the cell cycle kinases, which results in a large scale loss of phosphorylation in the egg [26, 30, 31]. This connection between degradation and phosphorylation is fundamental to the irreversibility of meiotic exit, as phospho-regulation is powerful but inherently reversible. Without linking the kinase activity to a more “permanent” step like protein degradation, anaphase can be readily reverted to metaphase before the completion of M-phase exit [32]. Phosphorylation is important for protein release as well, as agonists to certain kinases are sufficient to trigger the secretion events of the cortical response [16]. Egg activation therefore depends not just on the individual components that are degraded, released, or phosphorylated, but ultimately on their coordinated and cohesive interaction in a dynamic manner.

Measuring dynamics of many components in the cell is necessary to explore the coordination of biochemical programs in the egg. Beyond relative change, it is often helpful to when comparing reactions of diverse molecules to understand complex reactions in terms of absolute measurements. This way, the parallel biochemical programs can be understood and compared in terms of absolute rate, flux, and stoichiometry to components and of modifications. These types of measurements have not been made previously because of the difficulty of protein quantification in general. Comprehensive and absolute protein quantification is even more difficult.

The remaining two sections of this chapter provide 1) additional molecular details of the protein biochemistry that regulates egg activation, and 2) an introduction to the new technologies that will enable the exploration of the questions outlined in this first section.

1.3 The molecular mechanisms that drive egg activation

1.3.1 The biochemical control of the metaphase to anaphase transition

A conserved feature of M-phase exit is the transition from metaphase to anaphase promoted by targeted protein degradation (Figure 1.6). The destruction of Cyclin-B1/2 is catalyzed by the anaphase promoting complex (APC/C), an E3 ligase. Once activated, the APC/C mediates Cyclin B degradation through ubiquitination [30, 31]. Cyclin-B1/2 degradation promotes M-phase exit by inhibiting the activity of Cyclin-dependent Kinase 1 (Cdk1) [27]. The APC/C also regulates the degradation of Securin [33], which promotes sister chromatids separation via release of the protease Separase [34].

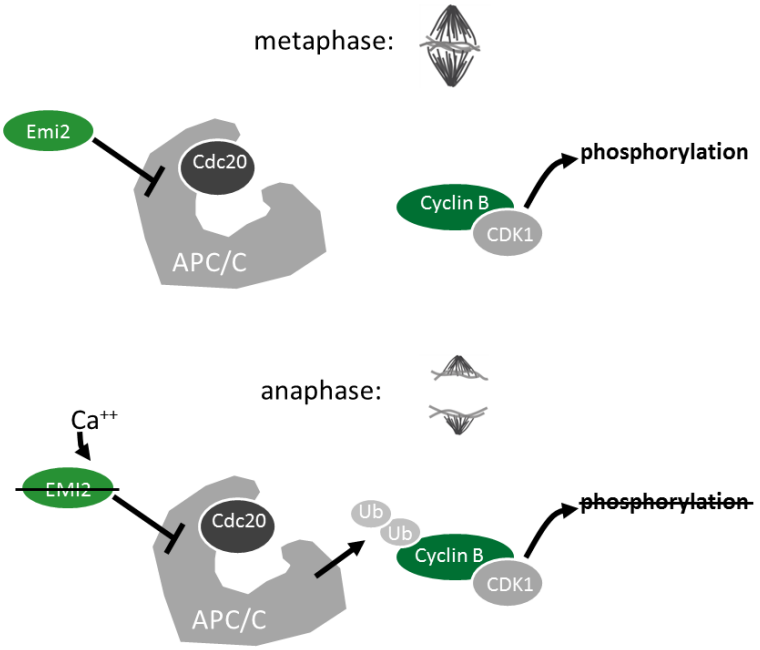

Figure 1.6 The molecular mechanism of the metaphase to anaphase transition in egg. In metaphase, Cdk1 activity is high when bound to its regulatory cofactor Cyclin B. The activity of the APC/C is inhibited by EMI2. During egg activation, EMI2 is degraded in a calcium-dependent manner, which permits the activity of the APC/C to ubiquitinated Cyclin B, specifically with the Cdc20 cofactor. Following the destruction of Cyclin B by the proteasome, Cdk1 activity is inhibited, leading to loss of phosphorylation of its targets and entry into anaphase.

1.3.2 Discovery of kinase and phosphatase activity as cell cycle regulators

The regulation of protein function through phosphorylation is essential to the cell cycle. It was first noted that catalysis could drive cell cycle progression after the discovery of an activity called Maturation

Promoting Factor (MPF) which was sufficient to induce oocyte maturation [35, 36]. The enzymatic component of MPF was later identified to be the stably-expressed kinase *cdc2*, also called Cdk1 [37-39]. The loss of Cdk1 activity corresponds with the bulk loss of phosphorylation, consistent with the negative regulation of kinase activity during meiotic exit [26].

The Protein Phosphatase 2A (PP2A) [40, 41] was recently implicated in reversing important M-phase phosphorylation sites. Inhibiting PP2A-B55 will stabilize a mitotic state indefinitely in *Xenopus* egg extract [42]. Additionally, the phosphatase Calcineurin is also required in the *Xenopus* for CSF release, though the mechanisms and targets remain unknown [43]. Overall, phosphatase activity remains less well understood than the kinase counterpart [44, 45].

The identities of the vast majority of phosphoproteins that are dephosphorylated in the egg remain undiscovered [29]; there is a strong expectation that this regulation reflects the reversal of Cdk1 phosphorylation.

1.3.3 Protein degradation controls cell cycle kinase activity

Protein degradation is of particular importance in meiotic exit. By observing protein synthesis in cleaving sea urchin eggs with radiolabeled methionine, a protein that oscillated with a similar frequency to the cleavage divisions was discovered and named Cyclin [46]. It was later shown that the synthesis of the protein Cyclin B is sufficient to drive M-phase progression [47], while stabilized versions of the protein arrested the egg cytoplasm in metaphase [48]. Cyclin B was shown to interact as a regulatory cofactor with Cdk1 ([37-39]), linking the oscillatory kinase activity with the level of Cyclin B. These observations illustrated the necessity of protein degradation to drive cell cycle progression.

1.3.4 M-phase arrests are mediated by the Anaphase Promoting Complex

A common feature of the cell cycle in somatic and germ cells is the metaphase arrest. Increased activity of the APC/C is required to relieve this arrest, but APC/C activity is controlled by at least two distinct mechanisms. Somatic cells control the entry into anaphase with the Spindle Assembly Checkpoint (SAC) [49], while vertebrate eggs instead inhibit anaphase with the so-called Cytostatic Factor (CSF) Arrest [10, 50].

In somatic cells, the Spindle Assembly Checkpoint (SAC) functions to arrest cell cycle progression until the chromosomes are aligned at the metaphase plate. The primary biochemical

mechanism of the SAC arrest is performed by the assembly of the Mitotic Checkpoint Complex (MCC) [51], which is comprised of the proteins BUB1 [52], BUBR1, CDC20 [53], and the “closed” form of MAD2 [54]. The MCC helps the cell “sense” incomplete chromosome congression, in part by dynamically assembling at unattached kinetochores. Once assembled, the MCC inhibits the activity of the Anaphase Promoting Complex by binding in the pocket required for the ubiquitination of its targets. The interaction of the MCC and the APC/C is now understood in exquisite structural detail [55]. The stability of the MCC undergoes additional regulation from proteins such as the deubiquitinating enzyme USP44 [56]. Entry to anaphase is inhibited until all kinetochores are correctly bound to microtubules. When the SAC is satisfied through the binding of kinetochores to microtubules and the alignment of chromosomes at the anaphase plate, APC/C activity is increased through the disassembly of the MCC. This allows for the binding of Cdc20 [57, 58] without the other complex members. The Cdc20 cofactor confers specific binding to degradation targets, allowing for the ubiquitination of targets such as Cyclin B and Securin, which the promote anaphase entry through negatively regulating Cdk1 activity and allowing for the separation of sister chromatids, respectively.

The SAC is not active in metaphase II in amphibians. Instead, the arrest depends on a different process discovered by Yoshio Masui, which he termed Cytostatic Factor (CSF) [59]. The CSF arrest in eggs is relieved by a transient calcium wave downstream of fertilization [60]. Activated by the calcium waves the Calcium/calmodulin dependent Protein Kinase (CaMKII) catalyzes a priming phosphorylation on the APC/C inhibitor XErp1/Emi2 is phosphorylated at threonine 195 [9, 10]. The kinase Polo-like Kinase 1 (PLK1, Plx1 in *Xenopus*) then phosphorylates the EMI2 phosphodegron, leading to Emi2 destruction by the SCF/cullin ubiquitin ligase β-TRCP [61]. As EMI2 is thought to function as a stoichiometric inhibitor, its degradation activates the APC/C [62, 63].

The precise molecular identity of the “Cytostatic Factor” remains elusive. The calcium-dependent destruction of a protein called Mos was a strong initial candidate. Mos degradation downregulates the mitogen-activated protein kinase (MAPK) cascade, which was shown to be an important step in egg activation in several organisms [28]. However, the direct involvement of EMI2 in inhibiting the APC suggests that it may be the most likely candidate instead. It may be that there is no single molecular identity of the cytostatic arrest, rather a set of molecules like EMI2 and Mos that converge on cell cycle

kinase regulation downstream of calcium signaling [10], assisted by E3 ligases (such as β-TRCP and the APC/C).

1.3.5 The effect of different Anaphase Promoting Complex cofactors on the cell cycle

The two activators of the APC/C are Cdc20 and Cdh1. The release of metaphase arrest typically involves Cdc20, which confers specificity for the "early" substrates of the APC/C. These include Cyclin-B1, Cyclin-B2, Geminin [64], and Securin [33, 65]. Cdh1 targets "later" substrates like PLK1 [66], Aurora A/B [67, 68], and Cdc20 [57, 69, 70]. The degradation of these targets promotes the transition from M-phase to G2. However, the cleavage divisions are a special, abbreviated cell cycle with no G1 or G2 phase (Figure 1.7) until after the mid-blastula transition [23]. Accordingly, Cdh1 is not expressed in

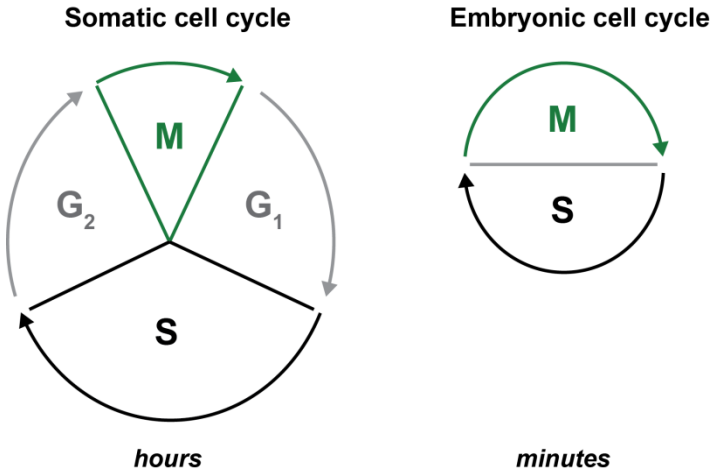

Figure 1.7 Comparing somatic and embryonic cell cycles. The somatic cell cycle occurs over many hours and features two Gap phases (G_1 and G_2) separating the cell division (M phase) and DNA replication steps (S phase). There are several checkpoints, including the DNA Damage and Spindle Assembly Checkpoints. Embryonic cell cycles occur within 30 minutes, and alternate quickly between DNA replication and division with no active checkpoints or Gap phases. Figure adapted from [71].

The *Xenopus* egg [72] or is present at low levels [73]. Egg extracts are not competent to degrade Cdh1 targets without supplementation of additional Cdh1 protein [74]. The Cdh1-specific substrates of the APC/C are therefore thought to be stable until later in development. The early cell cycle typically involves only Cdc20 [50] in the simplified embryonic cell cycle. It is worth noting that the list of known Cdh1

substrates continues to grow [75]. These substrates show broad effects in the organism, including proteins like microcephalin [76], which is involved in neurological malformations. There are some roles outside of M-phase for Cdc20, such as effects on the primary cilia [77]. Yet overall, the number of substrates remains small [66, 78, 79] and the Cdc20 targets function primarily as core cell cycle components. The cleavage divisions represent a pared-down cell cycle that is built for speed rather than complexity of regulation.

1.3.6 The "simplified" cell cycle in the early embryo

The streamlined cell cycle in amphibian embryos recalls discussions of whether to conceptualize cell cycle regulation as resembling "dominoes" or "clocks" [80]. Early models inspired by screens in yeast by Hartwell and others [81-83] described the cell cycle as a linear pathway where the next step was dependent on the product of the previous step (i.e., dominos). The "bouncing egg" experiments [25] in the amphibian embryo countered this view. The cell cycle oscillator clearly continued to function by switching between two states in a fixed period (i.e., clocks) even in the absence of the "essential" components identified in the yeast screens (e.g., DNA, cytoskeleton, etc.). The resolution is that the cell cycle is built around a relatively simple and conserved core oscillator. Additional feedback controls (i.e., check points) are built into this simple system (e.g., to impose an arrest).

The amphibian embryo during the cleavage divisions, proceeding with only the Cdc20-based activity and no gap phases (Figure 1.7), certainly represents a kind of minimal oscillator. It is of considerable interest to characterize this minimal set of Cdc20 substrates, which may represent the core components of the cell cycle, along with any post-translational changes (mainly phosphorylation) that occur downstream.

1.3.7 Fertilization as a cell cycle model: the advantages of synchrony

Fertilization is a complex process that has fascinated biologists in its own right for many decades. There are essential events like the blocks to polyspermy (discussed below) which have garnered considerable study. However, it has provided an important system to study of the cell cycle. The features of the fertilization response make the egg uniquely tractable for biochemical studies of the cell cycle in particular. In many species (sea urchins, frogs, etc.), eggs are available externally and in large quantities, allowing for sufficient material for study. It is also possible to achieve exquisite synchrony of the cell cycle

progression. The synchrony is enabled by two important features: 1) physiological cell cycle arrests and 2) the ability to “activate” eggs in bulk to reinitiate the cell cycle.

While fertilization of multiple eggs is not inherently synchronous (males fertilize the female’s eggs gradually during amplexus), *in vitro* fertilization is a common trigger to provide a relatively synchronous activation of many thousands of eggs. A jelly coat that is required for fertilization envelops the egg [84]. Substantial asynchrony can occur from the rate of sperm penetration through the jelly coat [85]. However, the process of egg activation can be induced parthenogenetically (i.e., without fertilization) by treatment with calcium ionophore [86] or electrical shock [87], neither of which is dependent on sperm penetration or the jelly coat. This allows for increased synchrony as well as more efficient sample collection. The jelly coat must be removed with alkaline cysteine solution [88] before proceeding with downstream sample preparation. In the case of artificially induced activation, the jelly can be removed ahead of time.

1.3.8 The validity of artificial activation in amphibian eggs

There is a reasonable concern that the eggs activated with artificial methods do not reflect the natural cell cycle events. There are many events in early embryogenesis that require the material contributed from the sperm. Syngamy and proper formation of the diploid nucleus cannot occur without the egg receiving the haploid paternal genome. In amphibians, the mitotic spindle cannot form without the centrosomes that are provided by sperm entry [89]. Finally, the early steps of dorsalization are biased by the point of sperm entry to direct the cortical movements of the cytoskeleton [90]. Symmetry breaking will fail without fertilization (though rescue of this eggs could be possible with the “tipping” experiment [91]).

The core cell cycle oscillator, however, proceeds essentially unperturbed. The period of the embryonic cell cycle is not dependent on the nucleus, the cytoskeleton, or transcription, so there is a strong expectation that many of the early events will proceed as normal. Indeed, to our knowledge, the early morphological and molecular events are equivalent between fertilization and parthenogenetic activation [92, 93].

There is a notable difference regarding MPF oscillations. In physiological fertilization, the embryo enters M-phase ~90 minutes post-fertilization (a longer pause than found in later cleavages), where Cyclin B synthesis causes increased MPF activity. While activated eggs show similar meiotic exit trends, the increase in MPF activity occurs earlier compared to the fertilized embryo (i.e., skipping some of the

longer delay seen in fertilized eggs) [92]. This deviation could affect comparison between the fertilized and activated eggs for affect later events, such as re-entry to M-phase and mitotic spindle formation [93]. However, it is still possible to conservatively say that the events occurring within 30 minutes of egg activation in artificially activated eggs are equivalent to the natural dynamics during fertilization. These first 30 minutes include meiotic exit and the early egg activation response (cortical granule release, surface contraction waves, discussed below), which span the interests of this thesis.

1.3.9 Calcium signaling during egg activation

Nearly all the molecular downstream events of fertilization are initiated using calcium release as a second messenger system. Fertilization triggers a calcium wave that typically originates from the site of sperm entry [94-98], though waves still propagate from the animal pole in parthenogenetically activated eggs [99]. Calcium peaks in cytosolic concentration around 5 minutes post-fertilization and declines thereafter [100]. The wave is generated by release of calcium from the endoplasmic reticulum, using activity from the Src kinase family and IP3 receptor. The initial rise of calcium promotes more release from the ER in a positive feedback loop / wave propagation model. The ions are quickly reabsorbed [15]. While intracellular calcium is important for most organisms, calcium from outside the egg can be used, such as in snails and worms [101].

It is evident that the calcium wave is necessary for both the cell cycle and more general egg activation, as EGTA blocks the cortical reaction entirely and egg extracts will not cycle [102, 103]. Two important sources of this molecular dependency is that 1) the calcium-sensitive kinase CaMKII [17, 104] is required to phosphorylate EMI2, which promotes its degradation and the subsequent activation of the APC/C, and 2) Agonists of the calcium-sensitive PKC [16] are sufficient to trigger the cortical response. Finally, the phosphatase Calcineurin is required for meiotic exit [43] and increased reactive oxygen species from the mitochondria are also downstream of the calcium wave [105].

1.3.10 The fast and slow blocks to polyspermic fertilization

The fast block is an ionic membrane depolarization event that happens within milliseconds of sperm fusion. In sea urchins, this is dependent on an increase in intracellular sodium [101], both through increased permeability and through exchange of H^+ ions, which later results in alkalinization of the

cytoplasm with similar timing as the calcium wave [106]. While sodium is directly involved in the fast block to polyspermy, downstream events require extracellular sodium as well [107].

The slow block to polyspermy occurs within minutes of sperm binding and triggers the cortical reaction (Figure 1.8). This is the fusion of specialized vesicles called cortical granules with the outer membrane, mediated by the SNARE complex and the calcium-sensitive synaptotagmin [108]. This results in the secretion of the cortical granule contents.

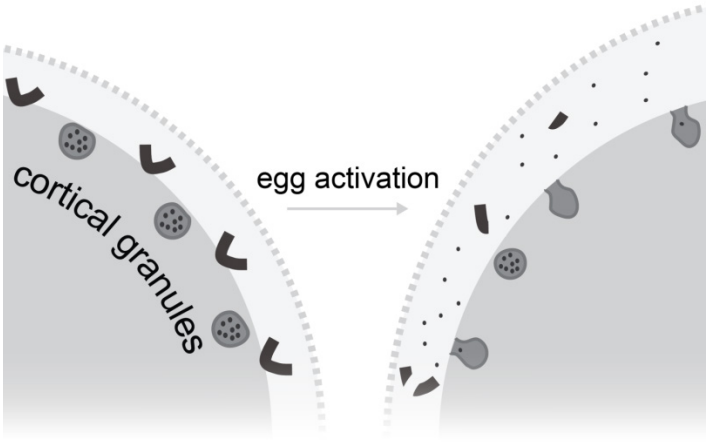

Figure 1.8 Illustration of a "zoomed in" view of the cortical changes during the blocks to polyspermy. Vesicles called cortical granules are docked with the outer membrane prior to fertilization. The cortical granules fuse with the outer membrane in a calcium-sensitive manner within two minutes of egg activation and expel their contents. One component is a set of highly charged proteins that cause an osmotic potential to hydrate the perivitelline space by lifting the vitelline membrane (shown as a dotted line). Sperm binding sites in the extracellular space are destroyed, mainly by the release of the cortical granule proteases, to inhibit further sperm entry.

A major component secreted by fusion of the cortical granules is the lectin family of proteins, which are highly glycosylated [109]. This helps to hydrate the space just outside of the egg, lifting the specialized extracellular matrix (called vitelline membrane) and creating a perivitelline space. This is thought to physically exclude additional sperm from access to the cell membrane.

The sperm binding components of the vitelline membrane are the Zona Pellucida (ZP) family of proteins [110]. Cortical granule proteases are released upon egg activation, mainly Ovostacin [111] and

Ovochymase [112]. These cleave the ZP components to inhibit binding of the sperm. While the ZP binding of sperm is important, the identity of the sufficient sperm receptor has proved elusive.

The first essential sperm receptor protein was only recently identified in mice, called Juno [113]. Rather than being cleaved, the receptor is exocytosed as an exosome after fertilization. The relationship in timing and function of this mechanism to cortical granule fusion is not known. A recent study in *Xenopus* identified the presence of a substantial amount of exosomes in the perivitelline space following fertilization [114]; other proteins at the cell surface may follow this mechanism as well. The cortical reaction results in a large increase in the protein outside the egg in the first few minutes post-fertilization, though the exact amount and identities of these proteins have been difficult to determine [16, 115].

The cortical reaction, including the membrane depolarization, perivitelline swelling, exocytosis of sperm receptors, and cortical proteolysis are the molecular mechanisms underlying the “membrane hardening” and “wave of negativity” that prevents multiple sperm entry that was first described nearly one hundred years ago.

1.3.11 Protein synthesis following egg activation

A substantial amount of the protein content of the egg is lost during egg activation through secretion, proteolysis, or degradation. The extent to which new protein production occurs in parallel is unclear. Protein synthesis following egg activation is reported in flies, sea urchins, and mice [116-119]. However, the work in flies is inferred from occupancy of polysomal RNA, sea urchin fertilization occurs when the eggs are already in interphase, and the maternal to zygotic transition in mice occurs much more quickly than in non-mammals. Whether bulk synthesis of new proteins occurs as a direct feature of egg activation or associated with later events in development is not clear.

1.4 Quantitative mass spectrometry as a tool for large-scale biochemistry

1.4.1 Mass spectrometry and the early embryo

Understanding the coordination of the protein biochemistry that regulates the early embryo requires making quantitative and dynamic measurements of multiple proteins and their modifications. However, protein levels and the dynamics of post-translational modifications have been difficult to

quantify in a comprehensive manner. Standard tools to detect proteins like antigen-based probes are time-consuming to produce and often do not give reliable quantification, especially for specific modifications [120]. New technologies, such as quantitative mass spectrometry-based proteomics, allow for the identification and quantification of many thousands of proteins and their post-translational modifications. This technology therefore offers an opportunity to observe the biochemical events in the egg at a larger scale than previously possible.

Amphibian eggs in particular offer unique compatibility with mass spectrometry approaches, as the eggs are large and develop externally. For example, *Xenopus laevis* egg provide sufficient material (~30µg of non-yolk protein per egg [121]) for routine proteomic studies that are more difficult to acquire in other systems. Importantly, obtaining reliable large-scale biochemical measurements are of general interest to many biological processes. Using these technologies in a well-studied system such as egg activation offers a useful situation to apply and develop novel but yet unproven technologies.

1.4.2 Identifying peptide sequences using mass spectrometry

The first protein to be sequenced was Insulin by Frederick Sanger in 1951 [122, 123]. For the next 40 years, biochemist typically relied on Edmund degradation and limited proteolysis approaches to identify the amino acid sequences of purified proteins [124] [125]. Identifying multiple proteins in complex mixtures proved much more challenging. Mass spectrometry became increasingly used for protein identification with the advent of electrospray in the early 1990s [126]. An important advancement was the ability to map peptide sequencing using genomic information to help identify spectra [127, 128]. Genomes enabled the production of theoretical spectra from the sequence information. Rather than relying on the highest quality spectra to directly determine the peptide sequences *de novo*, “noisier” spectra could be used to identify sequences by similarity to the theoretical spectra. This framework allowed for proteins to be more efficiently identified, setting up a new capability for “shotgun” proteomics [129, 130].

1.4.3 Approaches for multiplexed mass spectrometry

While identifying proteins is useful, the raw signal from mass spectrometry is not inherently quantitative. This is due to several factors, including 1) different ionization efficiencies of different molecules, and 2) variability in detector response and ion subsampling. This means that even differences of intensity of the same molecules between two separate analyses often are of technical origin, rather

than experimental [131]. In principle, multiplexed measurements eliminate this problem by analyzing peptides from different conditions in the same instrument run. It was clear from older mass spectrometry studies of small molecules that the ratios between the signals of the different isotopes are highly quantitative [132]. One of the first uses of this principle of encoding different conditions with stable isotopes was performed by incorporating heavy and light amino acids into tissue culture cells (SILAC) [131]. A similar approach called reductive dimethylation was developed to directly label isolated proteins (rather than labeling through feeding cells) [133]. Each technique gives a peptide peak a distinctive mass shift to be measured separately and the ratio of the two conditions assessed within the same spectra (Figure 1.9).

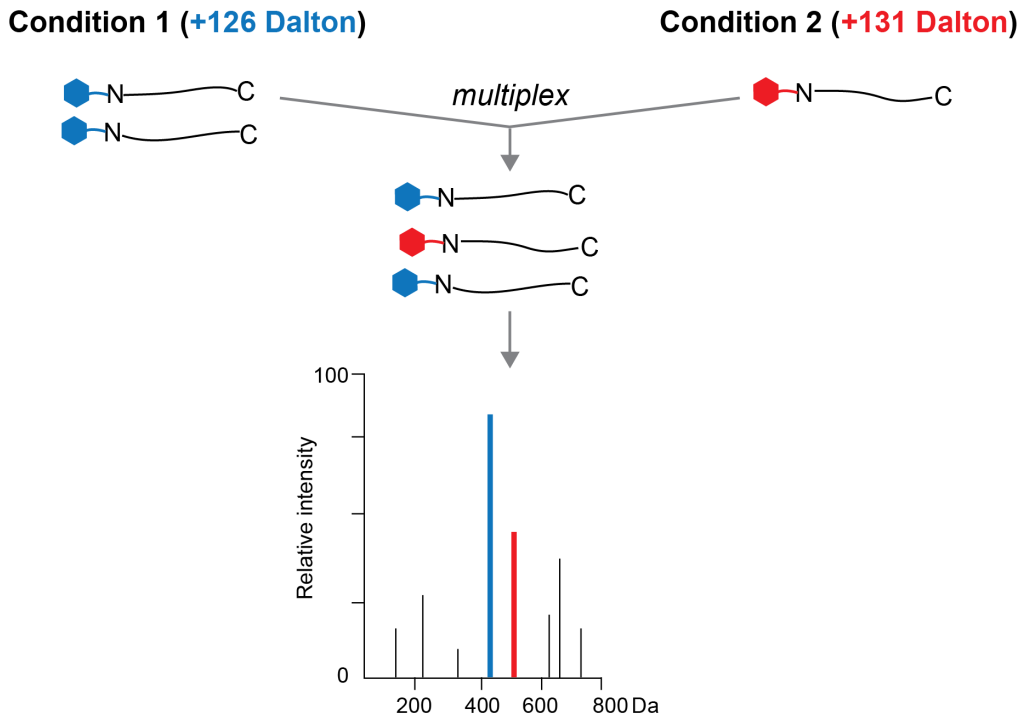

Figure 1.9 Multiplexed proteomics with mass barcodes. The condition of the peptide is encoded with a unique mass barcode that is chemically ligated to peptides after a digestion step. The conditions can then be mixed and analyzed simultaneously. Each condition corresponds to its own peak with a characteristic mass shift. The intensity differences between each peak reflect the quantitative relative change of the peptide between each condition. For simplicity, two conditions are shown here. Currently, ten barcodes are available for multiplexing analysis.

More recent labeling technologies includes Tandem Mass Tags (TMT) [134-136], which like dimethylation, use heavy isotopes to encode each condition and are chemically ligated to isolated peptides. The main advancement of TMT are 1) the capability to encode 10 conditions in one experiment, and 2) that each mass tag is identical in mass until it is fragmented in the gas phase. Whereas the first mass scan (MS1) becomes increasingly complicated with each additional condition in SILAC, TMTs allows all conditions to "stack" as a single peak in the first scan (Figure 1.10). iTRAQ reagents have also been developed, which operate under identical principles, but with a smaller molecular structure [137].

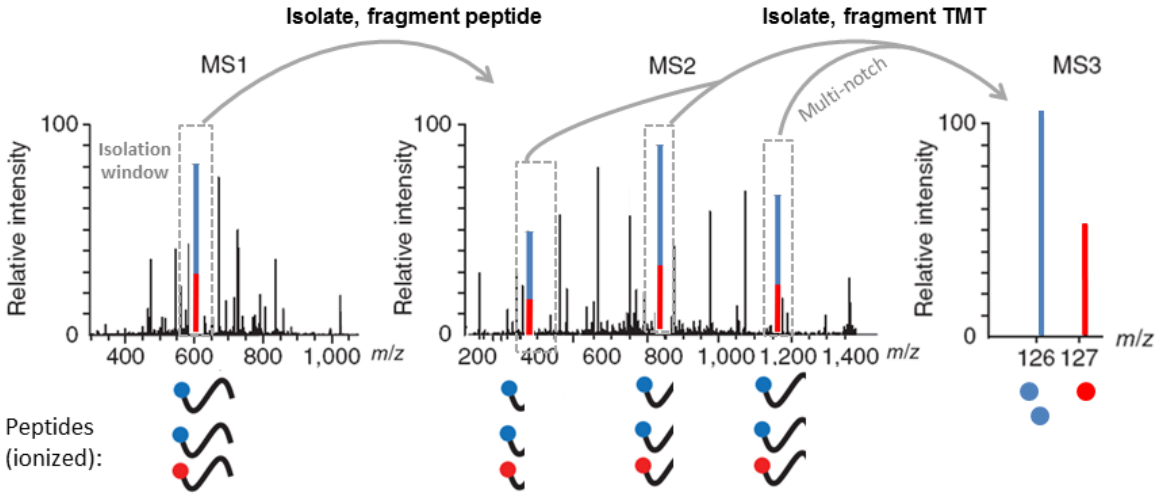

Figure 1.10 Overview of MultiNotch MS3 quantification using Tandem Mass Tags. Ionized peptides are tagged with different TMT reagents (blue, red) and measured in the MS1. Initially, the different conditions are indistinguishable in mass. The target peptide peak (i.e., precursor ion) is isolated, fragmented, and measured in the MS2. These steps are sufficient to deduce the sequence of the peptide. Finally, peaks from the MS2 are isolated and fragmented again (the MS3 spectra). Multiple peaks are isolated to boost the signal (i.e., Multi-notch). The fragmentation step liberates the reporter ions, revealing their mass differences (measured at low m/z). The relative intensity of the peaks encodes the ratio of change.

1.4.4 Quantitative proteomics with MultiNotch MS3

The quantification in this thesis primarily uses the MultiNotch MS3 method (Figure 1.10) [138]. TMT labeled peptides are fractionated using an on-line liquid chromatography system. Using electrospray ionization, the peptides are converted to gas phase for analysis. The sequence of each peptide is deduced from two scans that determine the mass/charge ratio of the ions. The first scan determines the

intact ionized peptides, or the precursor ion. This is known as the MS1. A specific peak is then isolated, fragmented by Collision-Induced Dissociation (CID), and measured in the MS2. The fragmentation is optimized to break the just one peptide bond, randomly distributed across the peptide backbone (to generate the 'y' and 'b' ions). The peaks of these fragments provide a "finger print" of the intact precursor ion. The sequence is surmised analytically by similarity to theoretical spectra (as in Section 1.4.2). The mass of peptides from each condition is identical during the previous steps. For improved signal to noise, an additional ion isolation step is performed called the MS3 [139]. Here, several peaks from the MS2 are selected and fragmented with High Collision Dissociation (HCD). The "reporter ions" are released in this step, which encode each condition with a unique mass. To increase sensitivity, multiple peaks from the MS2 are isolated for the MS3 analyses; this approach is called MultiNotch MS3.

The primary analytical tools in the mass spectrometer are the dual pressure linear ion trap and the Orbitrap. The Orbitrap generates a quadro-logarithmic electric potential, leading to "orbital trapping" of the ions around a central spindle structure [140]. The axial frequency of the orbit is proportional to the ion mass, which can be deconvolved through a Fourier transform with high accuracy [141]. The linear ion trap works differently, where the mass is determined by detecting the ions after the "trapping" becomes unstable under shifting electric fields [142]. It is more sensitive, but provides lower mass resolution.

The duty cycle in mass spectrometry is the ion path required to identify and quantify each measured species. A simplified duty cycle for the MS3 method is that a high resolution MS1 is taken in the Orbitrap, followed by a "low resolution" MS2, and finally a high resolution scan is taken in the Orbitrap to resolve the <1 Dalton mass differences. The rate of the duty cycle and the ion injection time are the primary settings that determine the overall sensitivity of proteomics.

1.4.5 Ratio distortion from non-target peptide "interference"

Initial implementations of TMT quantification simultaneously quantified the peptide fragments and the reporter ion fragments in the MS2. One important advantage of this approach is a shortened duty cycle, where each species could be identified and quantified with only one isolation step. However, the isolation window for a target peptide is not perfectly selective for the desired mass (Figure 1.11). Many low abundance, non-target peptides are co-isolated with the target peptide. After fragmentation, the reporter ions from both the target and non-target peptides are quantified together. As most proteins tend

to be stable across conditions, this has the effect of suppressing the ratio of change for proteins which are different in each condition [143]. By adding this additional selection of the fragments from the MS2, the non-target peptides are “diluted,” allowing for less ratio distortion in the MS3. However, taking a single peak from the MS2 mean that only a small subset of the total ions are used for quantification, which decreases sensitivity. “MultiNotch” selection (i.e., multiple isolation windows) was therefore developed to boost the signal of the reporter ions in the MS3 through isolating multiple fragments in the MS2.

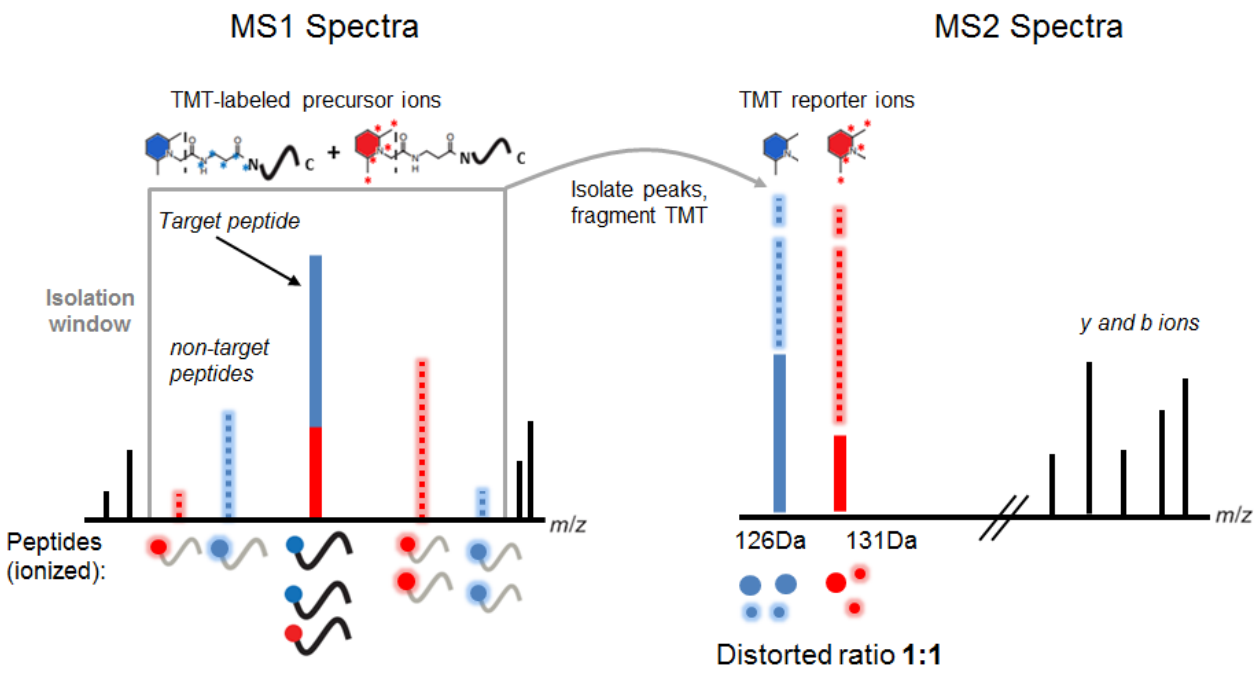

Figure 1.11 Ratio distortion through non-target peptide interference. A target peak is depicted with two TMT barcodes (blue and red). When isolated for the MS2 spectra, there are several low abundance non-target peptides that are co-isolated. After fragmentation, the reporter ions are liberated from the peptide, revealing their mass differences. However, because the reporter ions are separate from the parent peptide, it is not possible to distinguish signal from the target and non-target peptides. The effect is to suppress the true ratio (in this case, 2:1) between conditions toward 1:1.

1.4.6 Challenges of quantifying phosphorylation with proteomics

A phosphorylated peptide can be identified by the same process as described for the standard sequencing of peptides. The phospho-group gives a characteristic mass shift which can be used to identify modified proteins. The difficult task is to localize the modified residue, which often requires a "lucky" fragmentation pattern to isolate the exact position. This has been addressed using a probabilistic method to assign the most likely modified residue on a peptide [144]. Another issue is that the phosphorylation bond is more fragile than the peptide bond. This often leads to a neutral loss, where the phospho-group is preferentially fragmented over the peptide backbone, leading to low efficiency formation of the y and b ions needed to sequence the protein [145]. One concern specific to TMT analysis was that the neutral loss peak (the intact precursor with the removed phosphate group from fragmentation in the MS2) would concentrate non-target peptides, resulting in more interference if selected for the MS3. This fortunately is not the case, and isolation the neutral loss peak (to maximize reporter signal in the MS3) proved to give accurate quantification [146].

The most fundamental challenge for phospho-MS is the low abundance of phosphorylated peptides in complex mixtures. With an enrichment step [147-149], it is feasible to measure relative changes in a large number of phosphoproteins. A typical yield of phospho-enrichment is 1-2% of the starting mass. Obtaining sufficient material for mass spectrometry is a major limitation for large-scale quantification of phospho-regulation.

There are other pitfalls of interpretation for phosphorylation. Even if a peptide is detected and quantified, there are potential challenges in interpreting the direction of a given phosphosite trend (e.g., the decrease of a single modified site can result from the increase in the doubly modified form of that same peptide) [150]. To avoid this and other issues of interpretation, it is essential to consider the stoichiometry of the modification. Several approaches are available [151-153], though none of these are able to determine occupancies of peptides measured with multiplexed MS or that have multiple phosphorylated residues.

1.5 Summary of Thesis

The processes occurring after fertilization have served as an important cell cycle model, as well as informing other important mechanisms in cell biology (e.g., protein secretion, calcium signaling). The system has been particularly revealing toward the regulation of protein activity through protein degradation and phosphorylation. Nevertheless, the full extent of protein degradation in the egg, the components of the block to polyspermy, and the functional substrates of regulated phosphorylation are far from elucidated. Proteomics is a promising tool to further understand these processes, though the technology requires improvement to give reliable and interpretable biological insights.

In this thesis, we were inspired by the promise of mass spectrometry to study biochemical regulation in the *Xenopus* egg. However, the feasibility of proteomics in frog was at first in question. One major concern was the lack of a published genome at the inception of the project. Colleagues in the lab developed an approach to build a protein reference set from *de novo* RNA sequencing data in lieu of a genome. Chapter 2 details the study that explored the composition of the frog egg with proteomics. We were able to convert the summed ion signal for every detected protein into an estimate of absolute concentration in the egg [154]. This work served as a foundation of absolute measurements for the rest of the thesis.

In Chapter 3, my colleagues and I recognized that a fundamental problem in the interpreting phospho-proteomics data was that measurements do not typically account for site occupancy. While other techniques existed to calculate phospho-stoichiometry, none were compatible with multiplexed data or reported error in the estimate. Using the principle of conservation between peptide forms, we developed a regression-based approach to infer to stoichiometry of sites from the relative change of the modified and unmodified forms. The approach was the first to be compatible with single and multiply phosphorylated peptides, as well as sites that are dynamic or stable across conditions.

Finally, in Chapter 4, I used the highly synchronous events of fertilization to study protein degradation and phosphorylation following meiotic exit and egg activation. The developments of the previous two chapters were applied to investigate the extent and absolute rates of protein loss through degradation and release from the egg. These include the discovery of factors in the slow block to polyspermy and insights to the cell cycle degradation targets. Aided by the phospho-occupancy algorithm,

I investigated aspects of post-translational modification regulation, such as the reversal of mitotic phosphorylation, calcium signaling, and nuclear pore assembly [155].

This thesis introduces new biological findings about fertilization and cell cycle progression, but it also introduces new methods for studying the biochemistry of early development. In *A History of Embryology*, Joseph Needham noted that scientific advancement in embryology in particular has relied on careful and cumulative observation [156]. Major advancements have therefore often come with new tools (e.g., better microscopes, preservations reagents, sectioning techniques). In this thesis, I aimed to continue in this tradition by bringing new ways of observing embryos in the modern era, a task that has been continually productive for broader biology.

Chapter 2:

Deep proteomics of the *Xenopus laevis* egg using an mRNA-derived reference database

Martin Wühr*, Robert M Freeman*, Marc S. Presler, Marko E. Horb, Leonid Peshkin, Steven P. Gygi, Marc W. Kirschner

*These authors contributed equally

2.1 Attributions

The following chapter is reformatted from a paper published in *Current Biology* on July 7th, 2014 describing the protein and RNA composition of the unfertilized *Xenopus* egg. I am second author on the paper, behind the co-first authors Drs. Martin Wühr and Robert Freeman. Drs. Freeman, Wühr, and Peshkin worked to develop the RNA reference database. The early RNA sequencing experiments were performed under the guidance of Dr. Marko E. Horb. Dr. Freeman led the development of the reciprocal BLAST technique to properly set the transcript reading frame without 6-frame translation.

The proteomics experiments were performed with my own hands. Prior to these experiments, I developed improvements to the sample preparation protocols for mass spectrometry with guidance from Dr. Wühr. These steps were implemented the first time in these experiments (mainly streamlining the protocol, improving the protease digestion step and chaotropic resuspension techniques). I was then closely involved with writing the manuscript, including producing substantial sections of the text.

2.2 Abstract

Mass spectrometry based proteomics enables the global identification and quantification of proteins and their post-translational modifications in complex biological samples. However, proteomic analysis requires a complete and accurate reference set of proteins, and is therefore largely restricted to model organisms with sequenced genomes. Here, we demonstrate the feasibility of deep genome-free proteomics using a reference proteome derived from heterogeneous mRNA data. We identify more than 11k proteins with 99% confidence from the unfertilized *X. laevis* egg and estimate protein abundance with approximately two-fold precision. Our reference database outperforms the provisional gene models based on genomic DNA-sequencing and references generated by other methods. Surprisingly, we find that many proteins in the egg lack mRNA support and many of these proteins are found in blood or liver, suggesting that they are taken up from the blood plasma, together with yolk, during oocyte growth and maturation, potentially contributing to early embryogenesis. To facilitate proteomics in non-model organisms, we make our platform available as an online resource which converts heterogeneous mRNA

data into a protein reference set. Thus, we demonstrate the feasibility and power of genome-free proteomics while shedding new light on embryogenesis in vertebrates.

2.3 Introduction

Recent advancements in mass spectrometry-based proteomics now enable global identification and quantification for up to ~10K proteins in a single experiment, along with associated post-translational modifications [157-159]. The capability to identify proteins and measure their expression levels in an unbiased manner on a proteome-wide scale can revolutionize many areas of biology. However, many of the most interesting biological problems are best studied in non-standard organisms: limb regeneration in axolotl, [160] red blood cell development in ice fish, [161] or craniofacial developmental disorders in Darwin's finches [162]. To understand how different processes evolved it will be important to compare proteomic composition and dynamics in species from diverse clades.

Unfortunately, proteomics is currently very difficult in organisms without well-annotated genomes. In current approaches, proteins are digested with proteases, and the peptides are ionized, fragmented, and detected via MS/MS fragmentation spectra. In principle, these spectra contain sufficient information to deduce a peptide's amino-acid sequence. However, this approach is only feasible for subsets of spectra with exceptional quality. The number of interpretable spectra is significantly increased by matching MS/MS spectra with theoretical spectra generated from all proteins encoded in the studied species. This set should be both complete and accurate to achieve maximum sensitivity and specificity. The paucity of high quality reference databases is the main reason that MS-based proteomics is currently limited largely to species with well-annotated gene models.

Despite the rapid decrease in sequencing costs, obtaining genome-based protein reference sets for new organisms is time intensive and expensive. Creating accurate gene models for a new species relies on faithfully assembling a genome from short-read sequencing data and training gene predictors. Both processes are often met with bioinformatics and species-specific challenges. For example, the size and polyploidy of some species' genome, e.g. lungfish, axolotl or *Amoebae* [163-165], make sequencing challenging for the foreseeable future. In contrast, deep coverage RNA-seq is cost-effective, and protein-coding transcripts can be reconstructed using established tools and published protocols for any species

[166]. Some attempts have been made to generate a protein reference database by 6-frame translations of mRNA [167, 168]. Unfortunately, the majority of the obtained protein sequences are biologically irrelevant, unnecessarily increasing the search space for spectral matching, and therefore decreasing sensitivity while increasing the need for computational time and resources.

One under-exploited model for proteomic experiments is the African clawed frog *Xenopus laevis* [47, 169-171]. Large amounts of material required for deep proteomic experiments (> 100 µg of protein) can be obtained easily from *X. laevis* samples, which would be very hard or impossible to obtain in other model organisms (e.g., staged embryonic time series or undiluted, metaphase-arrested cytoplasm called egg extract). However, *X. laevis* has rarely been used for MS due to the lack of a released genome, likely due to the difficulty associated with sequencing quasi-tetraploid genomes [172].

Here, we demonstrate for the *X. laevis* egg that genome-free proteomics is feasible at remarkable depth, and that we can extract biological insight from this proteomics data. For our genome-free protein reference set we combine multiple sources of mRNA information and use knowledge of sequence similarity to proteins from related species for reading frame detection, frame-shift correction and annotation. In proteomic experiments, our database outperforms alternative approaches and even the latest rounds of preliminary gene models based on the unreleased genome. With more than 11k proteins identified with 99% confidence, this is by far the deepest proteomic study in *X. laevis*, and one of the deepest analyses performed in any organism. By enumerating the ~11k proteins in the *Xenopus* egg and measuring the concentration of each to approximately 2-fold precision, we have produced a valuable resource for the *Xenopus* community. Lastly, we offer the means for researchers to upload and convert mRNA data into a protein reference database for their own proteomic experiments in any organism.

2.4 Results

2.4.1 Objective evaluation of protein reference databases

To construct the best possible reference database for proteomics, we sought a method to evaluate and compare different reference versions objectively. We reasoned that for a given set of peptide fragment spectra, the number of confidently identified peptides is an objective measure of the

quality of that reference. To test this assumption, we collected spectra from a trypsin-digested *S. cerevisiae* lysate and searched them against a standard collection of all yeast proteins. We chose yeast, the first sequenced eukaryote, [173] because its gene models are exceptionally well annotated. We filtered the spectra, which were matched to peptides (peptide spectrum matches (PSMs)), to 0.5% false discovery rate (FDR) using the target decoy strategy [157, 174, 175]. Protein grouping was performed with maximum parsimony, with an additional filtering step to 1% FDR at the protein level [129, 130, 157, 176]. We then modified the yeast reference set to simulate the effects of searching spectra against low-quality references. First, we randomly removed half or three quarters of the yeast proteins in the reference database. The number of PSMs, unique peptides, and proteins approximately scales with the number of proteins in the reference database (Figure 2.1). To test whether irrelevant data would affect the number of identified peptides, we added shuffled human protein sequences to the yeast reference. As expected, the number of identified peptides and proteins is reduced due to the higher chance of false positive matches. To simulate a protein reference database with highly fragmented proteins, we bisected each protein from the reference at a random position. With this reference, the number of identified peptides slightly decreased, likely due to the removal of tryptic peptides at the fragmentation site. However, the number of identified "proteins" increased substantially (Figure 2.1) as some fragmented proteins were identified once per fragment. As an orthogonal test for the validity of peptide identification as a benchmark for the protein reference set quality, we searched MS spectra obtained from a *X. laevis* sample against the gene models from various species. As expected, the number of identified peptides decreases with evolutionary distance, likely reflecting the lower number of exactly matched peptides in the databases. Thus we conclude that proteomic data can be used to evaluate the relative quality of a reference protein dataset. More specifically, the number of identified peptides, but not the number of identified proteins, can be used as an objective benchmark to compare different reference sets.

2.4.2 Deriving an mRNA-based protein reference database

For proteomic experiments with *X. laevis*, we needed to obtain a comprehensive, artifact-free reference protein database without using a genome. To guide our approach, we evaluated the success of each processing step by the number of identified peptides when searching our reference against MS/MS

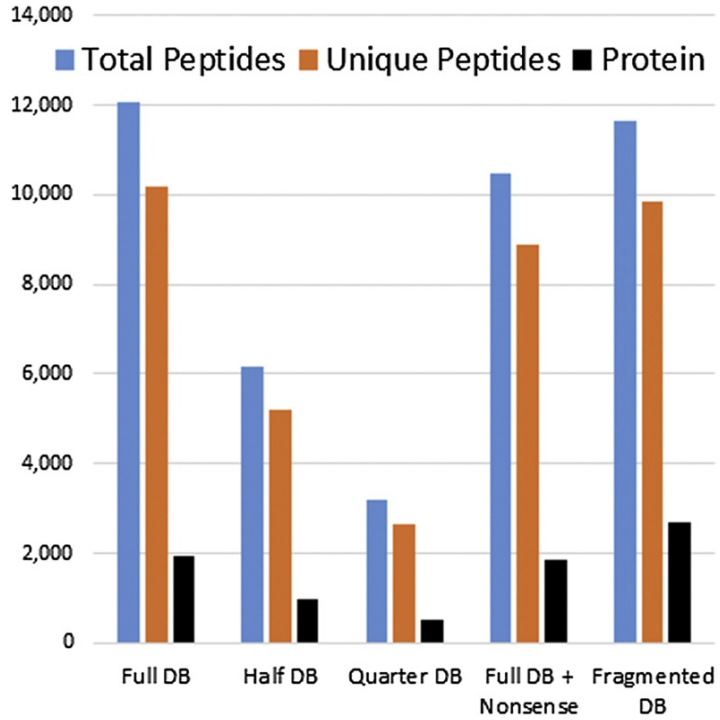

Figure 2.1 MS data can be used to evaluate relative reference database quality. Spectra from a tryptic digest of yeast lysate were searched against the standard yeast protein database (Full DB). Shown are the number of total peptide spectral matches (blue), unique peptides (orange), or proteins (black) that were confidently identified. To simulate "poor" reference databases, we removed half (half DB) or three quarters of proteins (Quarter DB) from the reference database. The number of identified PSMs and unique peptides scale approximately with the number or proteins in the database. To test how the addition of nonsense sequences would affect the number of identified peptides, we added randomized human proteins to the full yeast database (Full DB + Nonsense). The numbers of peptides and proteins are negatively affected. To simulate a reference database in which proteins are fragmented, we divided at a random position every protein in the reference into two proteins. While the number of identified peptides slightly decreases, the number of identified proteins substantially increases.

data from tryptic peptides of a *X. laevis* egg lysate. With this information we could evaluate alternative approaches while constructing the database, and choose the best possible option to improve our reference incrementally.

An overview of the process we used to generate our reference database, herein called PHROG (Proteomic Reference with Heterogeneous RNA Omitting the Genome), is shown in Figure 2.2. We combined information from publically available mRNA data as well as our own RNA-seq data, which we collected to study mRNA dynamics during early development. First, mRNA data from four sources (two RNA-seq *de novo* assemblies, transcripts from Genbank, and assembled contigs from the *Xenopus* Gene Indices) were combined, cleaned, and the repeats masked. The pre-processed transcripts were then clustered and assembled using parameters to maximize assembly, minimize spurious transcript fusions, and to collapse homeoalleles that are present in the quasi-tetraploid *X. laevis*. The assembled transcripts were compared in all six reading frames using BLASTX against proteins from six vertebrates in order to reveal the most likely translation frame, allowing us to bypass the introduction of large numbers of irrelevant protein sequences when using a six-frame translation. BLASTX alignments were also used to detect and correct for frameshifts that occurred due to sequencing errors. We then translated all transcripts in the BLASTX-hinted frame without regard to start and stop codons, ignoring translations from transcripts with a BLASTX E value >1E-5, as we hardly ever detected proteins from proteins with the indicated threshold. This hinted translation/filtering method significantly outperformed alternative translation attempts. To determine if 'hidden' or never-before-seen proteins might be present in these removed transcripts, we used *bona fide* protein coding transcripts as an initial training set and used the HMM-based translation method. Adding these trained translations back did not increase our peptide discovery numbers. Next, the translated sequence's N-termini and C-termini were trimmed to remove sequence ends which were either up or downstream of start and stop codons or for which would produce fractions of tryptic peptides and therefore could not be detected via standard MS searches. The removal of invalid sequences from the database increased the number of identified peptides. After eliminating redundancies, our final database PHROG contained 79,214 proteins (Figure 2.2). Finally, to facilitate interpretation of identified protein sequences, we assigned protein names and gene symbols using a modified reciprocal best BLAST hit approach based on a target reference of curated human proteins.

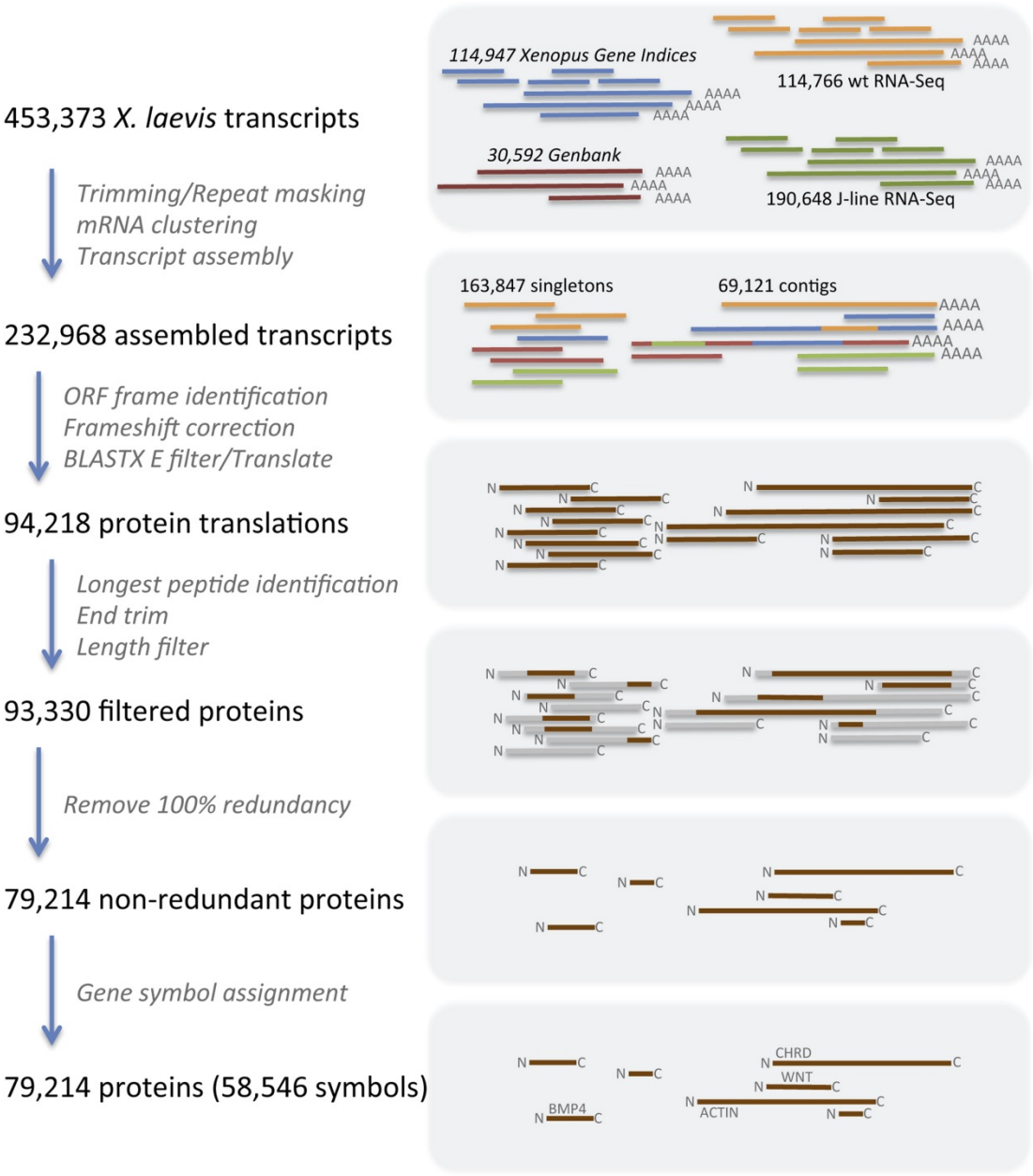

Figure 2.2 Overview of the Steps for Constructing the High-Quality Protein Reference Set PHROG. Transcripts from four different sources were combined, trimmed and cleaned using SeqClean, masked using RepeatMasker, and clustered and assembled using TGICL/CAP3. The assembled transcripts were aligned against a collection of model vertebrate proteins using BLASTX. The results were used for identifying the correct translation frame, for frameshift correction (if appropriate), and for removing sequences without significant similarity to known proteins. Once translated using BioPerl, the longest peptide for each protein is identified, and the ends are trimmed to match tryptic peptides. The collection is processed to remove 100% redundant proteins using CD-HIT, and gene symbols are assigned to the remaining members using the reciprocal or single best BLAST hit against human proteins. The numbers indicate the numbers of transcripts or proteins in each group.

A summary of the composition of our database, and its performance for a proteomic experiment compared to alternative reference sets is shown in Table 2.1. Judged by the number of identified peptides, via MS, our database outperforms the protein reference from Xenbase, the gene models from *X. tropicalis*, a 6-frame translated database, and even the gene models from the unreleased genome assembly v7.0 (kindly provided by Dan Rokhsar). One alternative to PHROG is using a better-annotated reference set from a related species (e.g., *X. tropicalis*). However, when using mass spectrometry a single amino acid mismatch makes it impossible to identify a peptide. Using the *X. laevis* published proteins from Xenbase identifies ~70% more peptides compared to the *X. tropicalis* reference (Table 2.1). The preliminary gene models provide a significant improvement for peptide identification over previously known proteins.

Table 2.1 Comparison of different reference databases and their performance in one example run of a tryptic digested X. laevis egg lysate. The first two rows compare the number of protein entries and amino acid numbers in the different reference databases. Next are the number of PSMs and unique peptides identified, using the different reference databases. Last are the identified number of proteins with maximal parsimony.

	X. trop. Gene Models	*X. laevis* Xenbase	*X. laevis* Gene Models	PHROG	PHROG + *X. laevis* Gene Models	PHROG Six-Frame	PHROG RNA-Seq Only
Proteins in database	43,455	34,178	44,159	79,214	123,373	610,557	71,716
Amino acid in database	22,546,772	14,676,179	15,683,803	25,605,893	41,289,696	76,509,919	24,281,510
PSMs	9,300	16,142	17,354	18,867	19,030	17,564	17,156
Unique peptides identified	7,847	13,381	14,531	15,894	16,043	14,791	14,510
Proteins identified	1,850	2,505	2,969	3,130	3,176	3,098	2,923

Surprisingly, even with latest assembly of the genome, our mRNA based approach identifies ~10% more peptides. When we combine PHROG with the preliminary gene models as protein reference, we only identify an additional ~1% of peptides compared to using PHROG alone. The PHROG 6-frame translated reference database is much larger than all other databases, and identifies ~10% fewer peptides compared to PHROG, likely because of additional false positive hits with irrelevant database entries, which hurts sensitivity.

One major advantage of our approach is that we combine mRNA information from various sources, thereby maximizing coverage. Besides our own RNA-seq data, we used publicly available mRNA sources for *X. laevis*, including expressed sequence tags, which are available for many nonstandard model organisms in large quantities[177]. To demonstrate that the mRNA-based proteomics approach is also feasible without public mRNA data, we created a reference relying only on our own RNA-seq data. This database identifies 90% of peptides that the PHROG identifies and approximately the same number of peptides as the *X. laevis* preliminary gene models (Table 2.1).

2.4.3 Deep Genome-Free Proteomics Demonstrated on the *X. laevis* Egg

To demonstrate the power of the genome-free proteomics approach, we determined the proteomic content of the metaphase-arrested *X. laevis* egg. To obtain the deepest possible coverage, we digested the proteins with both LysC and trypsin or with LysC alone, fractionated each sample with a medium pH reverse-phase column, and analyzed the fractions with liquid chromatography followed by MS (LC-MS). The acquired spectra were searched against our PHROG reference set, the preliminary gene models, and Xenbase protein database for comparison. The results are summarized in Figure 2.3. By using Xenbase's GenBank proteins known at the time of this writing, we identified 97,999 unique peptides. With the *X. laevis* 7.0 gene models, we identified 26% more peptides. With our PHROG reference, we identified 143,476 unique peptides, an increase of 46% over Xenbase. When we matched these peptides to the minimal number of proteins and filtered to 1% FDR on the protein level, we identified 6,455 proteins from Xenbase, 9,720 proteins with the genome, and 11,103 proteins from PHROG (Figure 2.3B). Unexpectedly, the relative increase of proteins when comparing PHROG to Xenbase is larger than the relative increase in unique peptides. We believe that this is mostly due to an overrepresentation of the highest-abundant proteins in Xenbase (i.e., the proteins for which most MS/MS spectra will be collected).

In contrast, PHROG seems to allow us to identify many lower-abundant proteins, which would be missed with the Xenbase reference set. Furthermore, PHROG might identify multiple splice forms or proteins with slightly different sequences (e.g., alloaleles), which may be missing in Xenbase. Importantly, the numbers obtained with the very stringent filtering criteria used here indicate that this study is among the deepest proteomic analyses ever performed on any species.

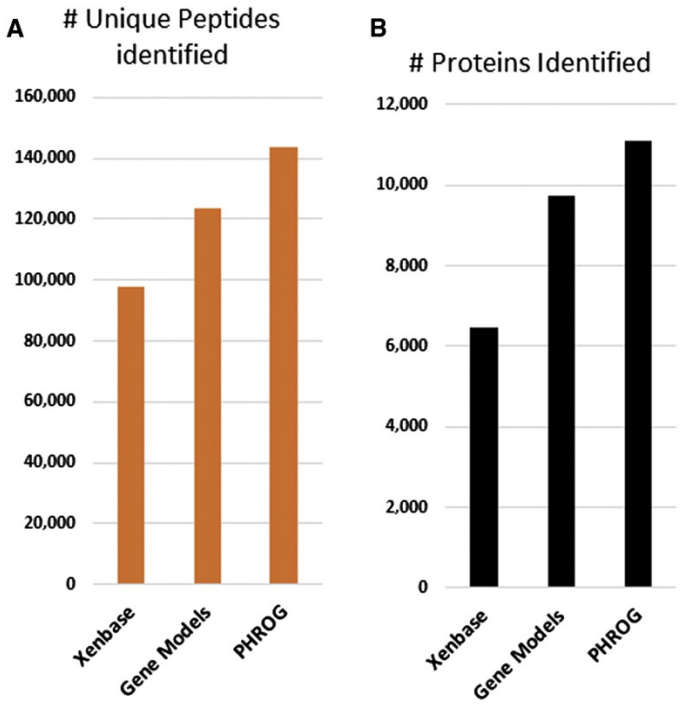

Figure 2.3 Comparison of Protein Reference Databases for the Fractionated X. laevis Egg Sample. A) Number of unique peptides identified with 0.5% FDR on the peptide level. PHROG significantly outperforms the publically available proteins from Xenbase and even the preliminary gene models from the 7.0 genome assembly as reference database. B) Comparison of the number of proteins identified in the egg, with additional filtering to 1% FDR at the protein level and maximal parsimony.

2.4.4 Estimation of the Concentration of Individual Proteins in the *X. laevis* Egg

Beyond providing a comprehensive list of identified proteins, we also wanted to estimate each protein's concentration. The difficult-to-predict ionization efficiency of peptides prevents us from directly

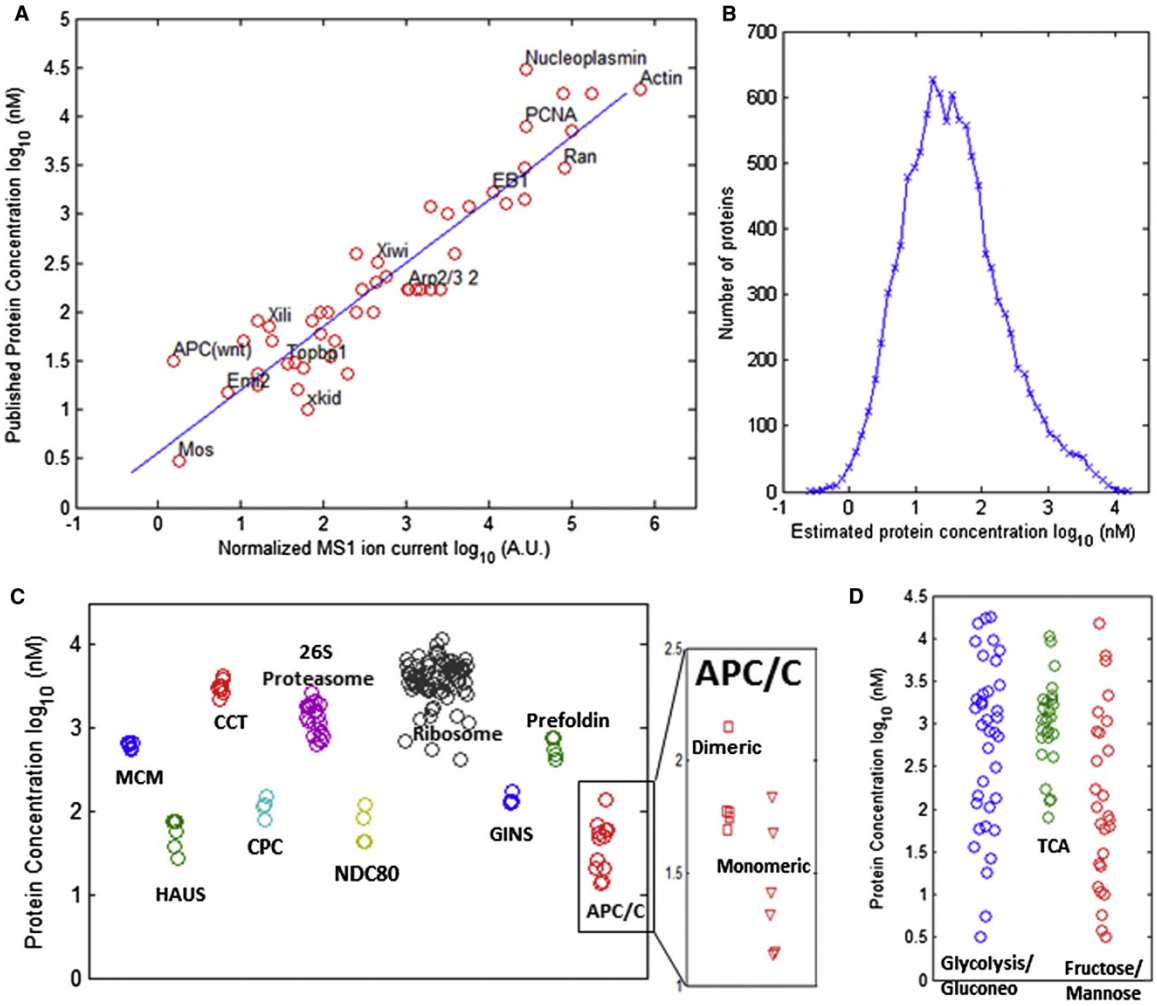

Figure 2.4 Estimation of Protein Abundance in the Xenopus Egg. A) Previously published protein concentrations for 49 proteins versus measured ion current in MS1 spectrum normalized to protein length. The Pearson correlation is 0.92. On average, the predicted protein concentration is approximately 2-fold different from the reported protein concentration. B) Histogram of concentration for all identified proteins regressed from normalized MS1 ion current. Median concentration of measured proteins is approximately 30 nM. C) Estimated concentration for subunits of stable complexes is similar. For the APC/C, we additionally distinguished between subunits that were reported to be dimeric (square) or monomeric (triangle) within the complex. Although our accuracy is not good enough to separate the two populations, the estimated concentrations for dimeric subunits tend to be higher than those for monomeric subunits. D) Concentrations for enzymes of a metabolic pathway can vary widely. For each metabolic pathway, the predicted concentrations of its members are plotted (based on the Kyoto Encyclopedia of Genes and Genomes).

measuring absolute protein abundance via MS. However, we can estimate each protein's concentration by summing up the ion current in the MS1 spectrum for all peptides of a protein and normalizing by the number of theoretical tryptic and LysC peptides [178]. We collected published concentrations for 50 proteins in *Xenopus* egg extract from the literature and plotted the concentration against the normalized ion current (Figure 2.4A). The detected proteins with published concentrations range over four orders of magnitude from 30 μM for Nucleoplasmin [179] to 3 nM for the MAPKKK Mos [180]; from our panel, we only failed to detect the 20 pM Axin [181]. The Pearson correlation for published protein concentration and normalized ion current in log-log space is 0.92 (Figure 2.4A). We confirmed that we did not overfit our data by performing a 10-fold cross-validation, obtaining essentially the same result (data not shown). Using this correlation, we regressed the protein concentration for all detected proteins (Figure 2.4B). With this approach, the estimated protein concentration differs on average by 1.9-fold compared to the published protein concentrations. The histogram for all estimated protein concentrations shows a median of ~30 nM (Figure 2.4B).

As an additional resource, we provide the protein concentrations summed by their assigned human gene symbols. Several distinct *Xenopus* proteins were mapped to the same human gene symbol. This is because similar but distinct proteins in *X. laevis* matched the same human gene during gene symbol assignment. The search results from the preliminary genome indicate that we identified nearly 10,000 distinct *X. laevis* genes (gene models do not contain splice variants).

For further validation, we asked whether subunits of stable protein complexes tend to have similar predicted concentrations. For ten stable complexes [182-184], we plotted the concentration of the subunits for each complex identified via the assigned gene symbols. Remarkably, the complexes' subunits cluster around similar concentrations, as shown in Figure 2.4C. At first glance, the anaphase promoting complex (APC) subunits are scattered relatively widely. However, some of the APC subunits are known to be dimeric, whereas some are monomeric [183]. Our precision is not good enough to separate these populations, but the dimeric subunits tend to have higher concentrations than the monomeric (Figure 2.4C). Interestingly, when we perform a similar analysis with components of metabolic pathways, the component's concentrations often vary by many orders of magnitude (Figure 2.4D).

2.4.5 Relationship of mRNA Abundance and Protein Abundance

Given our previous work in *Xenopus* transcriptomics [185], we sought to understand the relationship between mRNA and protein abundance. Using standard methods to estimate the abundance of the RNA-seq transcripts, we calculate the Pearson correlation of mRNA and protein abundance to be 0.32, whereas the Spearman correlation is 0.30 (in log-log space); these values are low compared to previous studies in tissue culture cells [158, 178, 186]. Unlike tissue culture cells, the *X. laevis* egg, which originates from the oocyte, emerges with a potentially different proteome and transcriptome after maturation. Although the correlation of protein and mRNA abundance is weak, we are more likely to observe the corresponding protein the more abundant the mRNA is (Figure 2.5A).

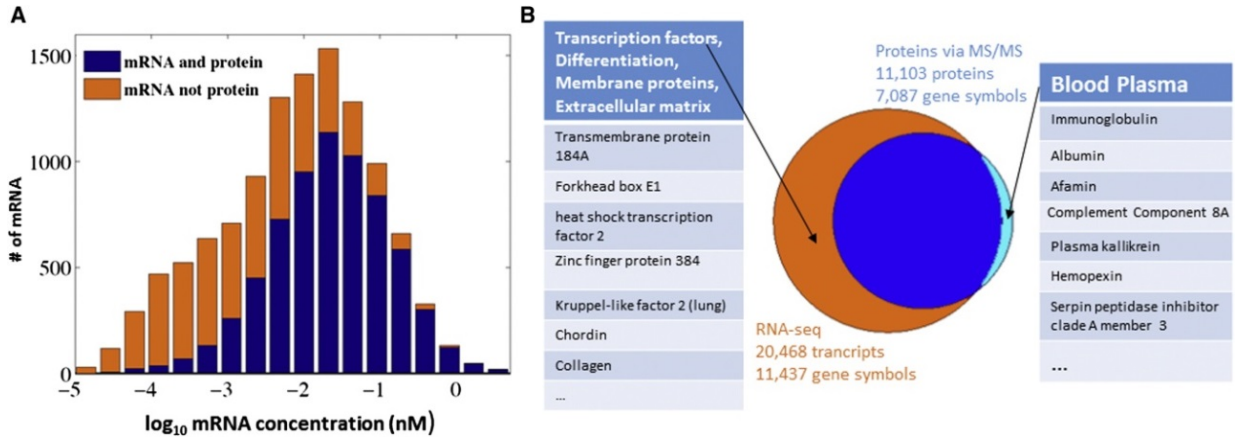

Figure 2.5 mRNA and Protein Abundance. A) Histogram of mRNA levels in the egg. mRNA for which the protein was also detected is colored blue. Orange indicates that only mRNA was detected. The median of mRNA concentration is approximately 1,000-fold lower than the median for protein abundance. Although we see only a weak correlation between mRNA and protein abundance (0.32 Pearson correlation), the lower the mRNA concentration, the less likely we are to detect the corresponding protein. B) mRNA and protein were matched via assigned gene symbols. MS is able to identify approximately 60% of all gene symbols for which we could detect mRNA. The proteins that we cannot detect via MS are overrepresented by transcription factors, proteins involved in differentiation, and transmembrane proteins. On the contrary, for ~350 gene symbols, we could identify only proteins, but not mRNA. This group is highly enriched for blood plasma and liver proteins and was likely endocytosed during oocyte growth.

We asked whether there were systematically overrepresented classes of genes that could only be seen via RNA-seq [187]. After mapping 4,675 gene symbols to our RNA-seq data, we found that membrane proteins (2,013 gene symbols), proteins involved in cell differentiation (894 gene symbols), transcription factors (316 gene symbols), and extracellular matrix proteins (189 gene symbols) are significantly overrepresented in the mRNA-only set (E values < 1×10^{-10}). Membrane proteins are known to be harder to detect via MS than soluble proteins, but we currently cannot distinguish whether membrane proteins are overrepresented as RNA because of MS sensitivity issues or because they are not expressed in the egg and are stockpiled for later translation. The same is true for the typically low-abundant transcription factors. For proteins used in differentiation and for extracellular matrix proteins, it seems more likely that the mRNA is present in the egg and will be expressed only during later stages of development.

With the current state of technology, RNA-seq is more sensitive than protein detection via MS. Therefore, we were surprised to find 368 proteins for which we could not find any mRNA support. After running gene set enrichment analyses with these proteins, we found that they were significantly enriched for blood plasma and liver proteins (Figure2.5B). During oocyte maturation, the yolk protein vitellogenin is synthesized in the liver and transported via the blood plasma to the oocyte, where it is endocytosed [188, 189]. We conclude that many proteins, besides vitellogenin, are also likely to be taken up via endocytosis from the blood plasma during oogenesis. Metabolic labeling experiments in the 1960s noted a small uptake of serum proteins in whole ovary but did not identify any of them [190]. It will be important to evaluate the intracellular role of these proteins during embryonic development.

2.5 Discussion

We present here the deepest proteomic study ever performed on *X. laevis* and one of the deepest performed on any organism. We identified ~11,000 proteins and estimated each protein's concentration, ranging more than four orders of magnitude, with an approximate average error of 2-fold. It might be possible to further improve protein concentration predictions by combining normalized ion current with peptide detectability prediction algorithms [191-193]. Our results will be a highly valuable resource for the *Xenopus* egg extract community for data mining, planning new experiments, and

complementing previous knowledge. For the development community, it begins to define the dowry of the egg and widens the opportunity for study of translational control, fertilization, and the maternal-zygotic transition. The large amount of material obtainable from *Xenopus* eggs and embryos, coupled with this new resource, should encourage the use of proteomics in development.

We started working on *X. laevis* proteomics in 2011 without access to a genome. We wanted to take advantage of proteomics in this unique system and had to develop the genome-free methods presented in this study out of necessity. Although this was intended as a preliminary effort, we were surprised by how well the approach worked, especially because we can compare it now to the early gene models. Ultimately, a high-quality genome with well-annotated gene models will likely provide the highest-quality reference set possible for RNA and protein analysis. However, reference sets based on mRNA are much cheaper and faster to obtain than gene models from genomic data. Based on this study, we now believe that mRNA-derived proteomic data could assist in building gene models that are more accurate by using identified peptide sequences to confirm exons. Recent studies suggest that even for model organisms with well-annotated genomes (e.g., rat or mouse), utilizing gene models based on RNA-seq evidence increases the information that can be gained from proteomic experiments [194, 195]. Furthermore, the relative quality of gene models, generated with different parameters, could be evaluated and potentially improved by utilizing the number of identified peptides from a proteomic experiment as a benchmark.

The proteomic data from the *X. laevis* egg illustrate the feasibility of genome-free proteomics, which can be extended to any nonstandard organism. One advantage of our methods is that mRNA data can be combined from heterogeneous sources. For many species, multiple expressed sequence tags and some full-length sequence information are available (e.g., http://compbio.dfci.harvard.edu/tgi/tgipage.html) [177]. However, only relying on our RNA-seq data, genome-free proteomics is possible. Approximately 10% of unique peptide data were lost by only using RNA-seq data from embryonic development; however, based on the findings in this paper, one could likely minimize this loss by adding mRNA data from the adult liver. We have integrated our series of pipeline scripts into an online resource that creates a high-quality protein reference database from

heterogeneous mRNA sources, and that resource can be found at http://kirschner.med.harvard.edu/tools/mz_ref_db.html.

Amino acid sequence information alone is not very informative. Rather, one needs to integrate that information with previous knowledge of proteins and their functions, e.g., which proteins bind to form a complex, which proteins are part of a metabolic pathway, or simply what the protein's name from which one can access the literature is. For nonstandard model organisms, it is unlikely that there is much previous knowledge of proteins from that species. However, by relating sequence similarity to human proteins, one can assign proteins to gene symbols and then interpret protein levels for development. One unexpected finding in *Xenopus* is that many proteins, which could be identified by MS, had no observable mRNA in the egg. We found that these were almost certainly proteins produced in the liver and endocytosed from blood. We also found mRNAs without protein, and this suggests that certain transcripts may be stockpiled in the egg for translation at later stages of development.

This study demonstrates the power of genome-free proteomics, and our online tool increases the scope of proteomic experiments. Knowledge of the level of protein expression can offer new insight into molecular regulation and provides a valuable resource for both biochemical and developmental work in *Xenopus*.

2.6 Methods

2.6.1 Sample Preparation for MS

The research with *X. laevis* was performed under the oversight of the Harvard Medical Area Institutional Animal Care and Use Committee. Female *X. laevis* were induced with 700 U HCG. After 14 hr, eggs were harvested, washed with 1× MMR, and dejellied with Cysteine (2% w/v) (pH 8.0). Sixty eggs were flash frozen with liquid nitrogen. Eggs were lysed with 250 mM sucrose, 1% NP40 substitute (Sigma), 5mM EDTA (pH 7.2), 1 Roche complete mini tablet (EDTA-free), 20 mM HEPES (pH 7.2), 10 μM Combretastatin 4A, and 10 μM Cyochalasin D. For lysis, eggs were vortexed at maximum speed for 10 s, pipetted ten times up and down with a 200 μL pipette tip, incubated on ice for 10 min, and again vortexed for 10 s. Lysates were clarified by centrifugation at 4,500 RCF at 4°C for 4 min in a tabletop centrifuge. The cytoplasmic and lipid layers were mixed by gentle flicking and removed from the pelleted yolk. To the

lysate, HEPES (pH 7.2) was added to 100 mM, and SDS was added to 2% (w/v). The sample was reduced with 5 mM DTT for 20 min at 60°C and then alkylated with 15 mM NEM for 20 min at room temperature (RT). Excess NEM was reacted with an additional 5 mM DTT at RT. Proteins were isolated by methanol-chloroform precipitation [196]. The protein pellet was resuspended (~5 mg/mL) in 6 M Guanidine HCl in 50 mM HEPES (pH 8.5) and sonicated for 5 min. The sample was diluted to 2 M Guanidine with 50 mM HEPES (pH 8.5) and digested with LysC (Wako Chemicals) at 20 ng/μL at RT for 14 hr. Next, we diluted Guanidine HCl with 50 mM HEPES (pH 8.5) to 0.5 M and digested further with 10 ng/μL of sequencing grade trypsin (Roche) at 37°C for 8 hr or LysC at an additional 20 ng/μL at RT. Samples were subjected to C18 solid-phase extraction (SPE) (SepPak, Waters) to desalt and isolate peptides. To reduce sample complexity, ~1mg LysC peptides and 0.5 mg LysC/trypsin peptides were resuspended in a 10 mM sodium carbonate buffer (pH 8.0) and then fractionated by medium pH reverse-phase HPLC (Zorbax 300Extend-C18, 4.6 mm × 250 mm column, Agilent) using an Acetonitrile gradient from 6%–31%. With a flow rate of 0.8 mL/min, fractions were collected into a 96-well plate every 38 s and then pooled into 24 fractions by combining alternating wells from each column of the plate. Each fraction was dried and resuspended in 20 μL of 1% phosphoric acid. Peptides from each fraction were desalted and extracted once more with reverse-phase purification and resuspended in 10 μL 1% formic acid. Approximately 4 μL per fraction was analyzed by LC-MS.

2.6.2 Estimation of Protein Concentration

Published protein concentrations were collected from the literature. To obtain the MS1 ion current, we divided the MS1 precursor peptide intensities by the corresponding noise value (Thermo raw file). This signal to noise ratio is a proxy for the number of charges in an Orbitrap analyzer [141]. To convert charges into ion current, we divided by the MS1 ion-injection time. For each PSM, we recorded the maximum ion current during a peptide's elution. These ion currents were summed for all PSMs that matched a protein [178] and normalized to the number of theoretically calculated tryptic plus LysC peptides, with at least 7 amino acids and at most 25 amino acids (missed cleavages were not allowed for theoretical peptides). The published protein name was searched on the Human Genome Organization gene name database to assign gene names (http://www.genenames.org/). If multiple proteins that had

been matched with the same gene symbol were found in the MS data set, their MS1 ion currents were summed. On occasion, multiple gene symbols were combined.

2.6.3 PHROG Final Build

X. laevis transcripts from GenBank, *X. laevis* Gene Indices version 11 [177, 197, 198], and the de novo assemblies from the wild-type and J line RNA-seq data were combined (ensuring unique identifiers), cleaned and trimmed using SeqClean (http://compbio.dfci.harvard.edu/tgi/software/), and masked for common repeat motifs using RepeatMasker (http://www.repeatmasker.org) with its default libraries. The cleaned sequences were clustered with TGICL [198], using default parameters (93% identity) but requiring a 100 bp overlap, and assembled using CAP3 [199] with default parameters (92% identity). The contigs and singletons were searched against a small database of model chordate proteins (*H. sapiens*, *M. musculus*, *G. gallus*, *D. rerio*, *X. tropicalis*, and *X. laevis*) using BLASTX [200], and the full BLASTX reports were parsed for strand, translation frame, expectation (E) value, bit score, and alignment coordinates of both query and subject. Before translation, the parsed data were processed to select transcripts that show possible frame shifts, as determined by translation frames of the high-scoring pairs (HSPs); the sequences of such transcripts were adjusted to compensate for and to retain the translation frame of the best HSP. All transcripts (corrected and not corrected) that showed conserved alignments ($E \leq 1 \times 10^{-5}$) were fully translated, without regard to the best open reading frame (ORF), in the hinted frame; those above this E value were discarded. The translated proteins were subsequently processed as follows: (1) the longest peptide from the full translation was retained; (2) protein ends were trimmed to reflect potential trypsin-digested peptides; and (3) any resulting protein fragments <7 amino acids were discarded. Finally, the remaining proteins were processed by CD-HIT [201], with a threshold of 100%, to collapse the group into a nonredundant data set. The alternative references were generated as follows: we performed the six-frame translation of the PHROG according to Evans et al. [167] by using the transcripts after TGICL/CAP3 clustering and assembly but prior to any filtering and/or trimming. We performed the HMM-based translation of PHROG on the same transcripts by using TransDecoder from the Trinity suite, translating on the positive strand only with a minimum size of 24 amino acids. The best-guess translation was performed using Virtual Ribosome [202], using parameters to translate on any

strand and return the longest ORF. All translations were also processed by CD-HIT with a threshold of 100% [201]. Scripts are available at http://kirschner.med.harvard.edu/tools/mz_ref_db.html.

2.6.4 Acknowledgements

We would like to thank R. Harland, D. Rokhsar, and the *X. laevis* genome consortium for making the preliminary gene models of *X. laevis* available. We thank Ramin Rad for his programming help and the RITG team for their HPC assistance. Thanks to Woong Kim, Robert Everley, and Joao Paulo for help with mass spectrometers and to the S.P.G. computer room for bioinformatics support. This work was supported by NIH grants R01GM103785, R01HD073104, P40OD010997, and R01DK077197. We thank MBL for their support of this project.

2.6.5 Accession Numbers

The MS proteomics data have been deposited to the ProteomeXchange Consortium [203] via the PRIDE partner repository with the data set identifier PXD000926.

Chapter 3:

Calculating phosphosite occupancy with quantitative mass spectrometry

Marc S. Presler, Elizabeth Van Itallie, Allon M. Klein,
Martin Wühr, Marc W. Kirschner

3.1 Attributions

The following chapter is an extension of the analytical section of a paper published in *PNAS* on October 27th, 2017 [155]. The biological results from the same paper are presented more fully in Chapter 4. I will describe the approached to calculate the occupancy of phosphorylated residues from measurements performed with quantitative mass spectrometry. The conceptualization of the algorithm was developed in close collaboration between myself, Drs. Allon Klein and Martin Wühr. Dr. Klein suggested the vectorized formulation of the problem, which led us to realize the occupancy could be obtained with a regression-based approach. Dr. Wühr was involved in many aspects of the conceptualization, including pointing out that standard least squares regression would be insufficient, which directed us toward a better method.

I performed all the coding to implement the approach and drove the effort to calculate confidence intervals from the fitting data, with consultation with all authors listed. Elizabeth Van Itallie participated in extensive discussions to strategize the implementation and also helped to better define the mathematical relationships that are the foundation of the occupancy calculation. We also collaborated to determine the experimental conditions of the phosphatase experiment used to induce dynamics to calculate occupancy for stable sites. The experiments to develop the protocol were done in her hands. The experiments for the paper were done in my hands. All the authors listed assisted in editing the writing of the manuscript. I performed the data analysis, compiled the figures, and wrote the text.

3.2 Abstract

Fertilization and egg activation induces a rapid transition that is directed primarily by the activity of several kinases. This inspired our interest in performing large-scale, quantitative measurements of phosphorylation with mass spectrometry. However, phosphoproteomics has presented many technical challenges. A parallel motivation was that using the *Xenopus* egg as a model was instructive for developing new analytical tools for mass spectrometry more generally. After obtaining preliminary time series data over, we were confronted with problems of interpreting the relative phosphorylation trends. We decided that measuring the stoichiometry (or percent occupancy) of the phosphorylation sites was

important for surmising their function. Unfortunately, inferring stoichiometry from mass spectrometry data is not possible without spike-in standards because unmodified and modified forms ionize at different efficiencies [204]. The difference in efficiency makes the ratios between the raw signal of each form uninterpretable.

We developed a method to instead infer the phospho-occupancy from the relative changes of the modified and unmodified form with a regression-based approach. Unlike previous techniques, this allowed the occupancy calculation for sites measured over multiple conditions. The regression framework enables the determination of confidence intervals for the occupancy value. Additionally, it allows the extension of the method to multi-dimensional space to include multi-phosphorylated peptides. Finally, these approaches typically cannot estimate the occupancy of stable sites (as the method relies on relative change between conditions as the primary input). We integrated a temperature labile phosphatase to induce dynamics to stable sites.

The approach developed here represents the first analytical and experimental framework that is compatible with phosphopeptides that 1) are measured under multiple conditions, 2) are dynamic or stable, 3) are multiply phosphorylated, and all while reporting confidence intervals.

3.3 Introduction

Dozens of covalent post-translational modifications regulate protein stability and activity. In animal development, the processes occurring during egg activation are particularly dependent on phosphorylation. Measurements of phosphosites are commonly performed with antigen-based assays. Phospho-specific antibodies are laborious to produce [120]. It is therefore rare that these reagents offer a scalable approach for measuring many parallel biochemical events. Additionally, studies that rely on antibodies are most often understandably qualitative. Quantitative measurements of protein concentration and degree of modification are critical for interpreting biological function. Radiolabeling assays can be quite quantitative [26] and detect the change of many modified sites, but the identify of individual residues in this case is often difficult to ascertain.

Mass spectrometry-based proteomics offers a method to broadly detect the relative changes of specific phosphorylation sites. While promising, one weakness of mass spectrometry is that the measurements do not reflect absolute concentration of the detected peptides (as discussed in Chapter 1.4). Absolute concentrations are useful for calculating the stoichiometry of the modification (also expressed as a percent, termed site occupancy). The inability to measure occupancy is a particular problem for phosphorylation, as the modification is substoichiometric. Specific residues are rarely modified in a binary fashion across a population of a given protein. Instead, a given site is modified at varying degrees of site occupancy. This information is critical to interpreting the function of any dynamic or static modification.

The relative trends of phospho-forms that can be readily obtained through proteomics have utility, but they are often difficult to interpret with measuring the occupancy. For example, a 2-fold increase in the occupancy of a residue from 1% to 2% appears identical to the same fold increase from 50% to 100% (Figure 3.1). These two cases can have very different functional implications.

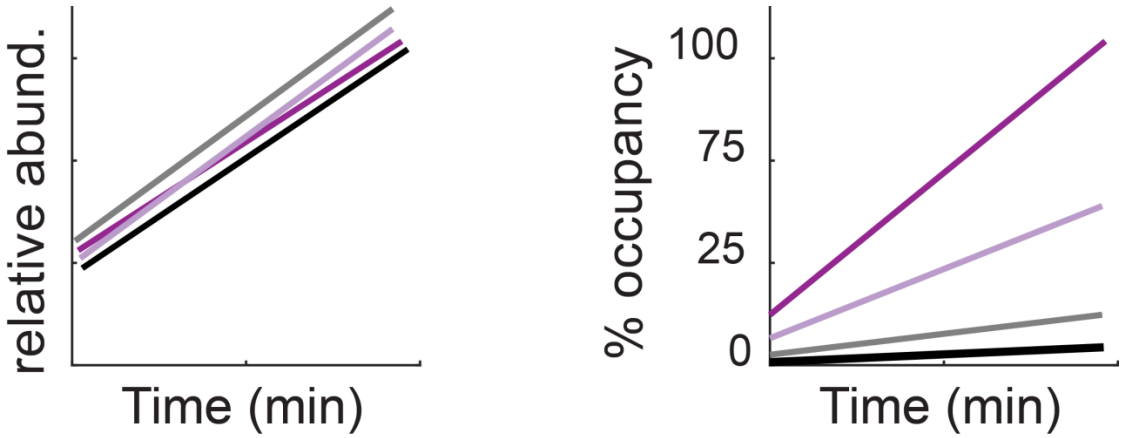

Figure 3.1 Identical relative trends can mask large difference in absolute change in proteomics measurements. The left illustrates the time series of four hypothetical phosphosites that show a ~10-fold increase over time, but which can change from 0.1%-1%, or from 10-100% occupancy (right panel). Therefore, important information is lost when measuring only relative changes for post-translational modifications.

It is not possible to directly determine stoichiometry from simple ratios of MS measurements without expensive spike-in standards [205]. This is because the unmodified and modified forms of a peptide ionize at different efficiencies and hence the inter-form ratio is distorted (Figure 3.2) without standards. Several approaches are available [151-153], but none of these are able to determine occupancies of peptides measured with multiplexed MS or that have multiple phosphorylated residues.

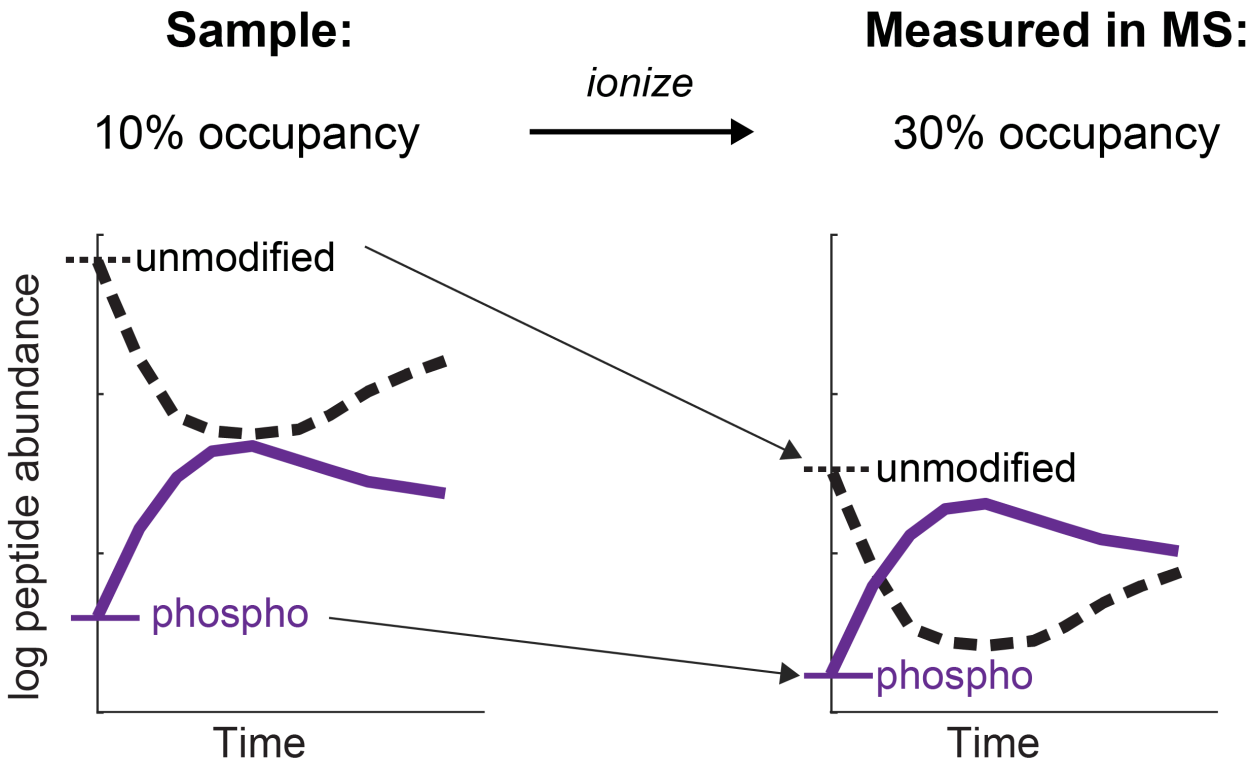

Figure 3.2 Differential ionization efficiencies distort occupancy values. Phosphosite occupancy cannot be directly calculated from the raw signal from the mass spectrometer because the inter-form ratios are distorted due to the differential ionization efficiencies of the peptide forms. This is depicted here as the unmodified form ionizing less efficiently than the modified form. If the percent occupancy is calculated from the MS data, there can be substantial error (in this case, from 10% starting occupancy to ~30%. In reality, the error can be substantially larger). However, the intra-form ratios are preserved. In other words, only the starting point is shifted. The amount that each form changes is the same, despite the ionization error. This is why multiplexed measurements still give quantitative information. The ratio between each form is nevertheless unreliable.

3.4 Results

3.4.1 Overview of method for evaluating phosphosite occupancy

It is nonetheless possible to infer the absolute stoichiometry of a site by invoking the principle of mass conservation, which states that the total abundance of each peptide equals the sum of all modified and unmodified forms of the peptide. This principle results in one constraint – a simple linear equation – for each pair of conditions in the experiment. If protein levels change, the equations must be scaled accordingly. By solving these conservation equations, we can infer the occupancy ratio from the measured relative intra-form ratios of change between conditions. Approaches invoking mass conservation were previously reported [151, 206] for the simple case of two phospho-forms and two conditions, giving two linear equations to be solved for the stoichiometries of each form. Here, we extend the solution and broaden its utility: (i) to include cases where the number of conditions is larger than the number of observed phospho-forms; (ii) to establish a statistical measure of confidence in the results; and (iii) to cases with multiple phospho-forms, which were not calculable by previous approaches.

Briefly, the method takes advantage of the fact that when the number of conditions exceeds the number of possible phospho-forms, the system is over-determined (Figure 3.3) and the problem becomes one of multivariate linear regression. Thus, an estimated solution to the system can be determined through slope-fitting.

3.4.2 Defining the graphical solution to the occupancy calculation

This section demonstrates the calculation of phosphosite stoichiometry from relative trends for an arbitrary number of experimental conditions and phospho-forms. The following set of equations show the generalized geometric relationship between the measured relative ratios of phospho-form change and their unknown absolute values. In other words, it conceptualizes the known and unknown quantities as a system of vectors that can be visualized graphically (Figure 3.3). From this representation, it becomes clear how the solution with many conditions becomes one of regression.

Let the total number of species of a peptide be M. In other words, if the peptide has three different residues phosphorylated, then $M = 4$, accounting for the modified forms plus the unmodified form. The species are the forms P, consisting of the phosphorylated residues $j = 0,1, 2, \ldots M-1$, where $j = 0$ is the unmodified form, $j = 0$ is the first phosphorylated form, etc. The ratios of change are the TMT-MS data

from any experimental condition *i* normalized by the data at the reference condition, set here as *i* = 1 (Eq. 1).

$$\text{Measured:} \qquad \frac{P_i^0}{P_1^0}, \frac{P_i^1}{P_1^1}, \ldots, \frac{P_i^j}{P_1^j} \qquad (1)$$

Let the ratio of change of the total protein T for a given peptide be $c_{i,1}$ (Eq. 2). This is either measured directly, or must be assumed as 1.

$$\text{Measured:} \qquad c_{i,1} = \frac{T_i}{T_1} \qquad (2)$$

Lastly, any condition and the reference condition, the sum of all phosphorylated forms is either equal or related by the change in protein level $c_{i,1}$ (Eq. 3).

$$\text{Conservation constraint:} \qquad \left(\sum_{j=0}^{M-1} P_1^j\right) c_{i,1} = \sum_{j=0}^{M-1} P_i^j \qquad (3)$$

Substituting Eq. (1) and Eq. (2) into Eq. (3) yields Eq. 4.

$$\text{Eq. (1) and (2) in Eq. (3):} \qquad 0 = \sum_{j=0}^{M-1} P_1^j \left(\frac{P_i^j}{P_1^j} - c_{i,1}\right) \qquad (4)$$

To visualize the system, it is helpful to rewrite Eq. (4) in vector form as Eq. (5) to reveal a geometric relationship in the data.

$$\text{Eq. (4) in vector form:} \qquad 0 = [P_1^0, P_1^1, \ldots P_1^{M-1}] \cdot \left(\left[\frac{P_i^0}{P_1^0}, \frac{P_i^1}{P_1^1}, \ldots \frac{P_i^{M-1}}{P_1^{M-1}}\right] - \vec{1}c_{i,1}\right) \qquad (5)$$

Eq. (5) shows that the vector of absolute values $[P_1^0, P_1^1, \ldots P_1^{M-1}]$ is orthogonal to the *M*-1 dimensional subspace containing the measured relative values $\left(\left[\frac{P_i^0}{P_1^0}, \frac{P_i^1}{P_1^1}, \ldots \frac{P_i^{M-1}}{P_1^{M-1}}\right] - \vec{1}c_{i,1}\right)$, as the dot product is zero. When *M* = 2, this subspace is a 1-dimensional line containing the point (1, 1) scaled by $c_{i,1}$ (Figure 3.3).

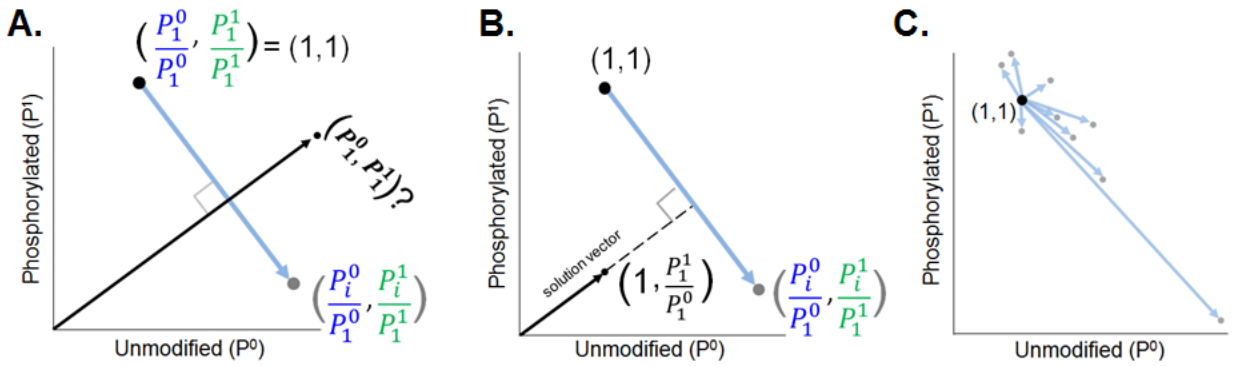

Figure 3.3: Graphical representation of the relationship of relative ratios to absolute phosphoforms. A) The MS data for the unmodified (P^0) and phosphorylated (P^1) forms is normalized to the reference time point 1 (Eq. 1) and plotted as coordinates on the P^0 vs P^1 plane. The subspace defined by the point (1, 1) and the point $\left(\frac{P_i^0}{P_1^0}, \frac{P_i^1}{P_1^1}\right)$ (shown as a light blue line) is orthogonal to the vector containing the absolute values of P^0 and P^1 forms at the reference time point (Eq. 4, 5; i.e., the dot product is zero). The values of P^0 and P^1 are not known.

B) The absolute values of the species $[P_1^0, P_1^1]$ lay on a line with the slope of $\frac{P_1^1}{P_1^0}$ passing through the origin. This can be seen graphically by normalizing the vector $[P_1^0, P_1^1]$ by it first element, to give the vector $\left[1, \frac{P_1^1}{P_1^0}\right]$ shown in black. The P^1/P^0 ratio is therefore the negative inverse of the slope of the line containing the points defined by the known ratios, and is sufficient to calculate stoichiometry.
C) Demonstrating the graphical representation of an over determined system when the number of measurements is greater than the number of phospho-forms. Each additional condition gives another vector (or linear equation). In this case, ten conditions give 9 vectors from the reference condition. The solution to this system can be estimated with regression (i.e., fitting a line).

Since our aim is to solve for stoichiometry, it is sufficient that all possible absolute values lie on a line passing through the origin with a slope of $\frac{P_1^1}{P_1^0}$ (for intuition, see Figure 3.3B). Stoichiometry is calculated by rewriting $\frac{P_1^1}{P_1^0+P_1^1}$ in terms of $\frac{P_1^1}{P_1^0}$, which is $(\frac{P_1^1}{P_1^0} / (1 + \frac{P_1^1}{P_1^0}))$. For *m* > 2, this subspace is an *M-1* dimensional plane, the orthogonal vector to this plane that is used to calculate the stoichiometry. When the number of conditions is greater number of conditions is greater than the number of species (*N* > *M*), the system is over-determined (Figure 3.3C). In this case, the unique *M-1* dimensional subspace can be estimated by minimization. The details of minimization are discussed in the following section.

3.4.3 Implementation of the regression-based stoichiometry calculation

We provide an algorithm in MATLAB to calculate the site stoichiometry of a single phosphosite when multiple conditions create an over-determined system (https://github.com/marcpresler/OccupancyMS). For the purposes of this study, we assumed that proteins were stable, as the data in Chapter 4 show that protein dynamics are overwhelmingly flat during egg activation (i.e., $c_{i,1}$ = 1). The main functions are to 1) estimate the *M-1* dimensional subspace through minimization and 2) establish confidence intervals for that estimate. For simplicity, we initially discuss the case where *M* = *2*. Input to the algorithm is quantitative mass spectrometry data for a P^0 and matching P^1 peptide for all measured conditions. The data for each species is normalized to the initial reference condition. The solution is the stoichiometry for that reference condition.

The data can be visualized as a P^0 vs P^1 plot that has one point per condition (Figure 3.4). Because of the normalization, the reference condition will lie at the point (1, 1) (Figure 3.3), scaled by $c_{i,1}$. When the number of conditions *N* = *M*, minimization is not necessary as the subspace is uniquely defined (in this case, a line). However, with more conditions, there is no longer a single solution (i.e., the system is over-determined). We can estimate the most likely subspace fitting a linear model to the data by minimizing the total least-squares residuals (Figure 3.4A). The ability to incorporate many measurements under different conditions enables a more accurate calculation of occupancy than one measurement alone. The "over-determination" of the system is key to reporting confidence intervals. We are able to estimate error in the fit coefficients using bootstrapping [207]. Solving for multi-site occupancy is not possible if the number of conditions is less than the number of phospho-forms, as the system is underdetermined. Therefore, our ability to calculate multisite stoichiometries is enabled by measuring multiple conditions by multiplex-MS.

We chose total least-squares (TLS) regression because we expect measurement errors in both the P^0 and P^1 ratios. To satisfy the orthogonal relationship defined in Eq. 5, the $\frac{P_1^1}{P_1^0}$ ratio is the negative inverse of the minimized slope of the subspace. Once obtained, we calculate stoichiometry (see end of previous section) and express in terms of percent occupancy of the modified residue (or the stoichiometry x 100). A steep fit slope will give a low occupancy because the orthogonal P^1/ P^0 ratio is therefore small,

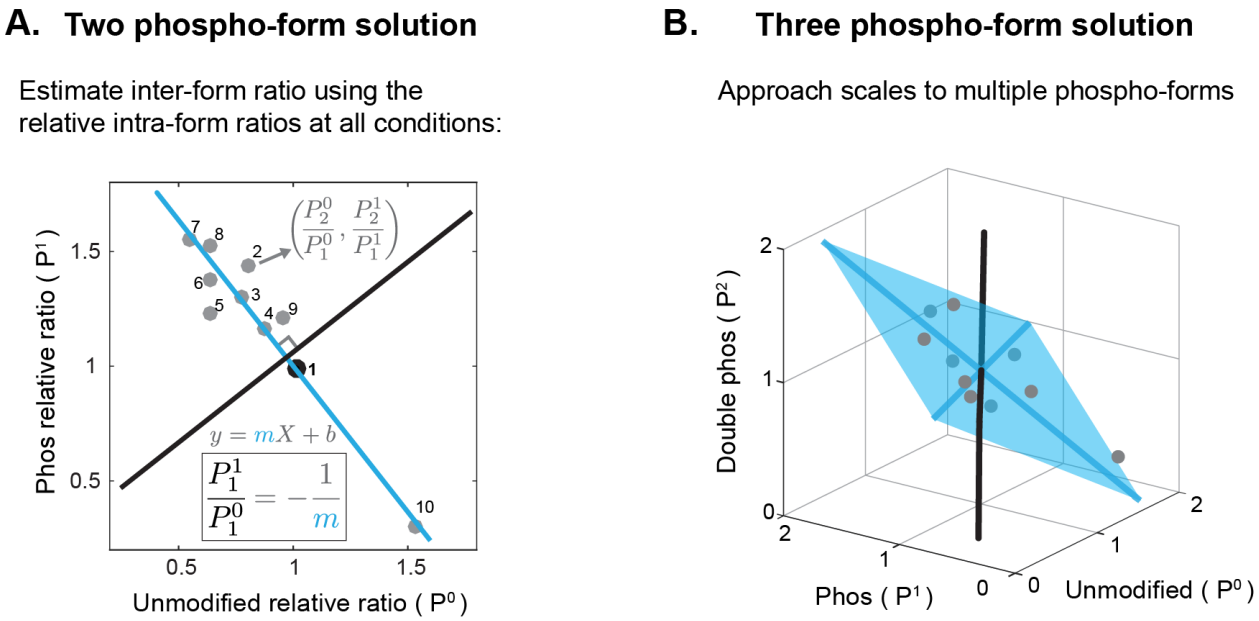

Figure 3.4 Calculation of phospho-stoichiometry from multiplexed phospho-dynamics measurements. A) Estimating the solution for the over-determined system from multiplexed-MS data. For each condition, the measured intra-form ratios of change of the unmodified (P^0) and single phosphorylated (P^1) forms define coordinates in a 2D plane. The relative ratio is defined by the reference point, which in this case is point 1 (black point). The solution to the over-determined system can be estimated by regression. The unknown inter-form P^1/P^0 ratio is the negative inverse of the fit line slope (i.e., is the orthogonal slope. B) For more than two P forms, the known ratios define a higher dimensional plane (e.g., a plane is fit in 3D space from the ratios of the P^0, P^1, and double form P^2). The plane is visualized as the blue shaded area spanned by two vectors. The solution (the black vector) is orthogonal to the plane, as in panel A. These vectors are calculated by principal component analysis (see text). Data points are plotted in grey, blue-shaded points are behind the plane.

while a shallow fit slope gives a high occupancy, as the orthogonal P^1/P^0 ratio is large. To obtain the stoichiometry for the rest of the conditions, the reference condition is switched and the analysis is repeated until the all conditions have been solved (see Figure 3.5).

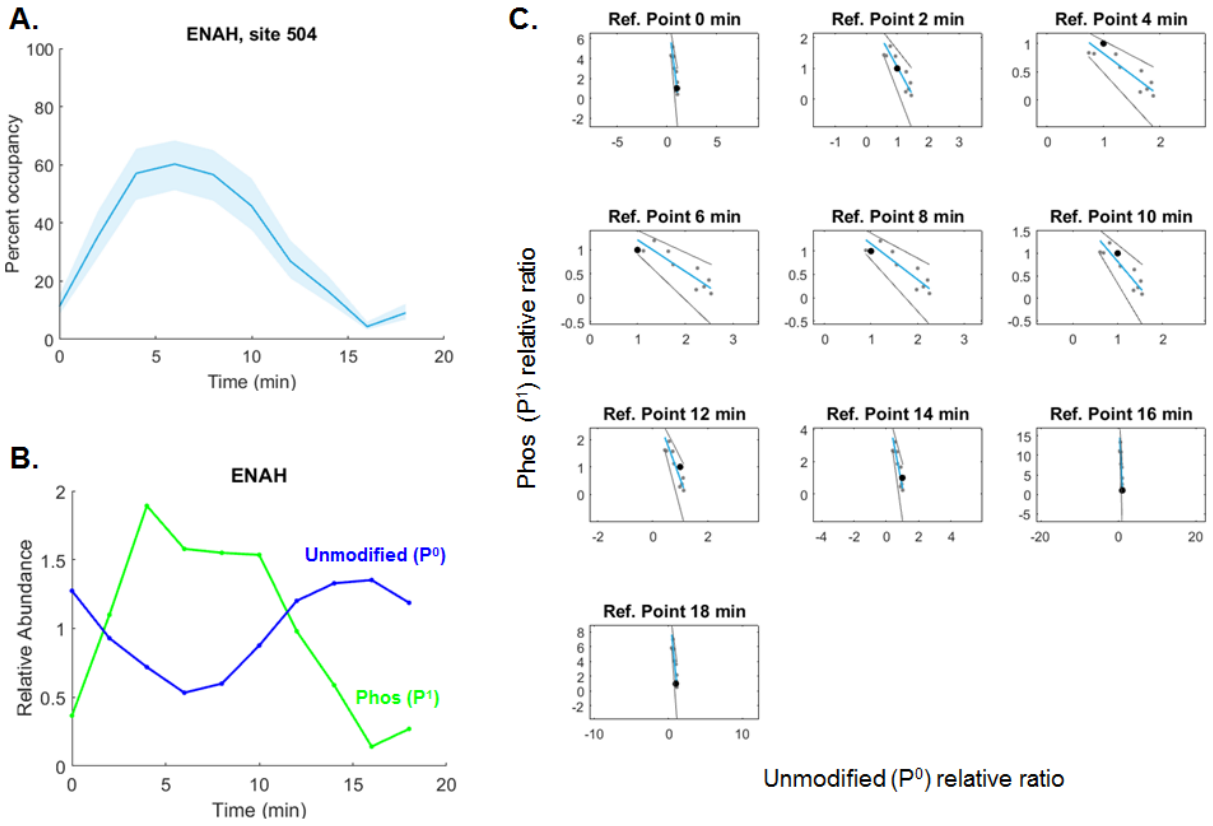

Figure 3.5 Calculating phospho-stoichiometry for multiplexed measurements. Visualization of how stoichiometry is solved for a single site with multiple time points. A) Example time series data of phosphosite stoichiometry of the residue 504 of the protein ENAH. The occupancy with 95% confidence intervals (shaded) is shown in light blue. B) The same data, but plotting the unmodified (P^0) and phosphorylated (P^1) trends (blue and green, respectively) that are used to calculate the stoichiometry. C) Estimating stoichiometry for the whole time course. Each subplot is comparable how to the visualization in Figure 3.3. To obtain the P^1/P^0 ratio for the whole time course, the reference time point was switched and the analysis is repeated for each time point. The reference point (1,1) is plotted in black, while the rest of the data are plotted in grey. The optimal line, obtained with Total Least Squares regression, is shown in light blue. The 95% confidence intervals are obtained by bootstrapping at each time point and are plotted in gray. Note how the angle of the slope corresponds with the shifting stoichiometry over the time course. The negative inverse of these lines (i.e., orthogonal) gives the P^1/P^0 ratio (as in Figure 3.3B) and its confidence intervals.

The approach extends to fitting a hyperplane in higher dimensional space for more than two phospho-forms (Figure 3.4B). For $M > 2$, the $M-1$ dimensional subspaces are estimated using Principle Component Analysis (PCA), which provides a method to minimize the total least squares residuals in an arbitrary number of dimensions through Singular Value Decomposition. The last principle component is the orthogonal vector to the $M-1$ plane.

The phospho-form ratios are obtained by projections onto the P^0 dimension or by finding the vector contained by the span of the orthogonal vector whose elements sum to 1 (satisfying Eq. 3). The TLS and PCA package [208] is adapted from File ID: #31109 on MathWorks File Exchange. Note that for $M > 2$, this approach assumes that we can detect all phospho-forms significant to the calculation. For example, a doubly phosphorylated species ($M=4$) will theoretically have two single forms. We typically detect only one of these forms and must calculate in $M=3$. This assumes the other form does not exist, or that it has a very small occupancy.

3.4.4 Calculating confidence intervals for the stoichiometry estimates

To aid in interpreting the quality of the estimated stoichiometry calculation, we sought to calculate a confidence interval of the minimized slope coefficient. Standard techniques for calculating confidence intervals rely on normally distributed error [209], which does not hold for total least squares minimizations. We therefore used bootstrapping [207] to estimate the confidence interval of the minimization. Specifically, the confidence interval is calculated from the distribution of slopes that is produced by iteratively fitting data subsampled with replacement. We used the MATLAB function 'bootci' and the bias corrected and accelerated percentile method to calculate confidence intervals. We performed 10,000 iterations per condition.

For validation, we calculated an upper-bound estimate of the confidence interval by regressing against error in the Y-axis and then the X-axis separately, and taking the wider confidence interval of the two for each axis [210, 211]. As expected, the bootstrapped confidence intervals were nearly always narrower than the upper-bound estimate (Figure 3.6).

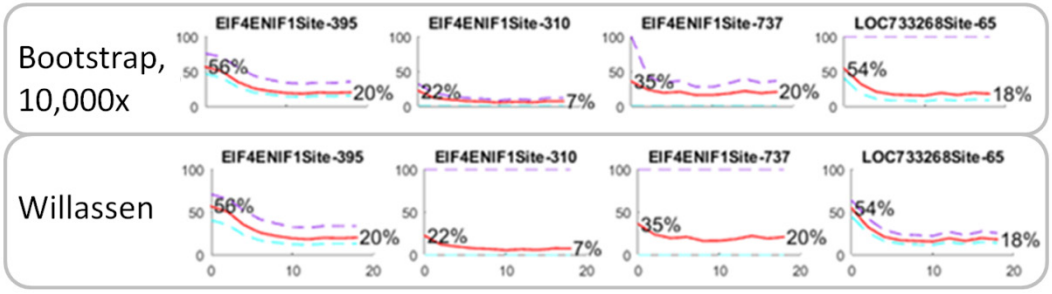

Figure 3.6 Bootstrapped confidence intervals compared to Willassen confidence intervals. Comparison of two different methods for calculating confidence intervals. Optimal estimate is shown in red, upper confidence intervals are shown in violet, and lower confidence intervals in cyan. For confidence interval validation, we minimized in the X and Y axis separately. Taking the wider confidence interval of the two gives an over-estimate of the 95% confidence interval of regressing in both axis [210, 211] (i.e., the Willassen solution). Bootstrapping confidence intervals are almost always narrower than the overestimate from the Willassen solution, as expected. The Willassen solution is only narrower when the fit of the line is determined by only one or two dynamic points. A very different slope results in the situation where these points are not sampled during bootstrapping, widening the resulting distribution. For the Willassen calculation, the confidence intervals are unaffected by one or two points defining the line as there is no subsampling.

Rarely, the subsampling with replacement will result in only one condition selected. In this case, fitting is not possible, so a random value between 0 and 100 is assigned. If a trend has repeated values in either the P^0 or P^1 forms (e.g., multiple zeros), a line with the slope of 0 or undefined can occur if only these points are subsampled during the bootstrapping, and fitting will fail. In these cases, the code returns either 0 or 100% occupancy automatically, depending whether the slope is undefined or zero, respectively. For the data where the initial best-fit slope results in an occupancy estimate that falls outside the 0 to 100% bounds, the confidence intervals are automatically set to 0 and 100% to reflect the high amount of error. This bootstrapping approach scales to *n*-dimensional space for multi-phosphorylated species.

Corresponding P^0 and all P^j forms must have a matching Protein ID (i.e., same gene symbol and isoform, if relevant) and an identical sequence to the peptide used to quantify the phosphosite occupancy. A given phosphosite may contain several peptides (e.g., missed cleaved species). Any of these peptides are used as a match for the P^0 species and are summed. The user must make sure that the each condition is appropriately normalized before import. Because of reference condition normalizations (Eq.

1), any value of 0 in the data is replaced with 1E-9 as to not result in undefined ratios. All replicates are averaged into one trend.

3.4.5 Experimentally inducing change to enable stable site occupancy calculation

For stable sites, the minimization is unreliable and gives wide confidence intervals (Figure 3.7). To extend the use of this approach for stable sites, developed a method to artificially induce induced dynamics with a nonspecific phosphatase treatment of two conditions [152]. We used a heat labile phosphatase, which allowed us to inactivate the phosphatase activity for subsequent experimental steps (See Methods, implemented in Figure 3.8).

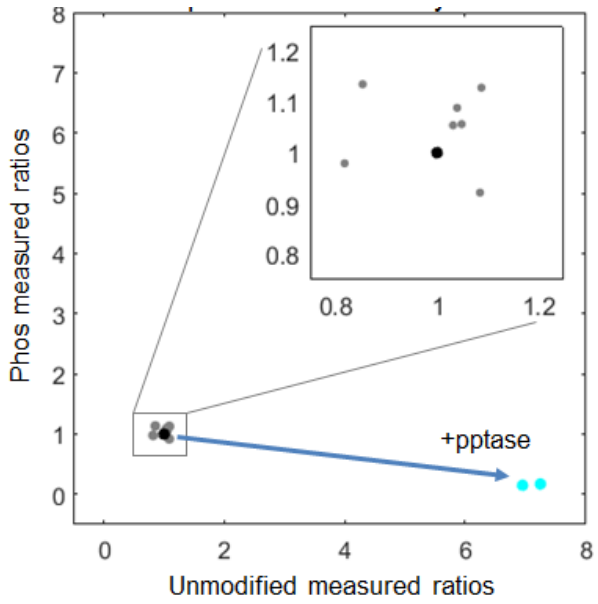

Figure 3.7 Visualization of data on the unmodified vs phosphoform plane for a stable site as in Figure 3.4A. The bold point is the normalized reference time point (see Figure 3.3A) and the cyan points indicate phosphatase-treated conditions. With low dynamics, perfect data would collapse onto a single point, rather than forming a line. Noise will cause scatter (see inset), but no clear line will be formed. The phosphatase treated conditions allow for reliable minimization, though the fit is defined nearly exclusively by these two points. For the confidence of the untreated conditions, the noise dominates when the phosphatase-treated conditions are not sampled during the bootstrapping. This can lead to two qualitatively similar stable sites giving very different confidence interval widths in the untreated conditions. For example, in perfect data, even very small fluctuations between the P^0 and P^1 forms are expected to be reciprocal to satisfy the conservation constraint. A "noisy" stable trend would have small fluctuations in P^0 and P^1 that change in the same direction. When the phosphatase conditions are not sampled during bootstrapping, the fit is extremely sensitive to this type of noise.

3.4.6 Applying occupancy calculation

In Chapter 4, we detail the experimental steps taken to study phosphorylation in early development. In brief, we have applied the method described in this chapter to the phosphorylation dynamics for the twenty minutes following egg activation. This was possible for a subset of the sites where we detect the unmodified and modified forms. We did not scale the conservation equations in this case. The short duration of the experiment makes it unlikely that a plurality of the proteins is changing in abundance. Indeed, direct measurement of the proteins show that of the nearly 9,000 proteins quantified, the overwhelming majority are stable in the twenty minutes after egg activation (See Chapter 4). These experiments included the phosphatase treatments as described above to enable the quantification of sites that are stable and difficult to estimate otherwise.

Figure 3.8A shows the relative dynamics of phosphosites on the kinases NEK3 and PAK2, as well as transcription factors/nucleotide-binding proteins SOX3 and YBX2. The amount of relative change for each set is nearly identical. However, YBX2 and NEK3 sites change phosphorylation for only a small fraction of their residues, whereas the residues on SOX3 and PAK2 change substantially. These are selected examples of many important regulatory proteins with discrepancies between relative and absolute changes (Appendix 1, Figure S5).

A proof-of-principle multi-site phosphorylation stoichiometry calculation is shown for a CaMKII-γ peptide (Figure 3.8B). The plot displays the unmodified (P^0), single (P^1), and double phosphorylated (P^2) forms that were detected and used to estimate the occupancy of each residue. We estimate the occupancy of the P^0 form with high confidence; the confidence intervals for the P^1 and P^2 forms are wide but non-overlapping, so we are able to conclude that the P^1 form is more occupied than the P^2. This highlights the utility of confidence intervals, especially as error increases in higher dimensional spaces.

A general consideration when applying this method is that it is most effective under conditions where the relative changes are substantial.

In total, we calculate confident stoichiometries (+/-25% occupancy of 95% confidence intervals) for ~500 sites (15% of dataset); ~150 of these are dynamic. With these data, we are able to compare stable versus regulated sites, which was difficult with previous approaches. For increasing, decreasing, or stable phosphosites (classified by their relative trends), we show the cumulative distribution of their

occupancy at 0 minutes (Figure 3.8). The majority of sites that increase have a low initial occupancy. The distribution of sites that decrease is skewed toward higher occupancy, though there is more density of lower occupied sites than reported in previous studies [151, 212, 213]. Interestingly, the stable site distribution is the most highly occupied class. However, the difference between the distribution of stable sites and deceasing sites is perhaps more subtle than it appears.

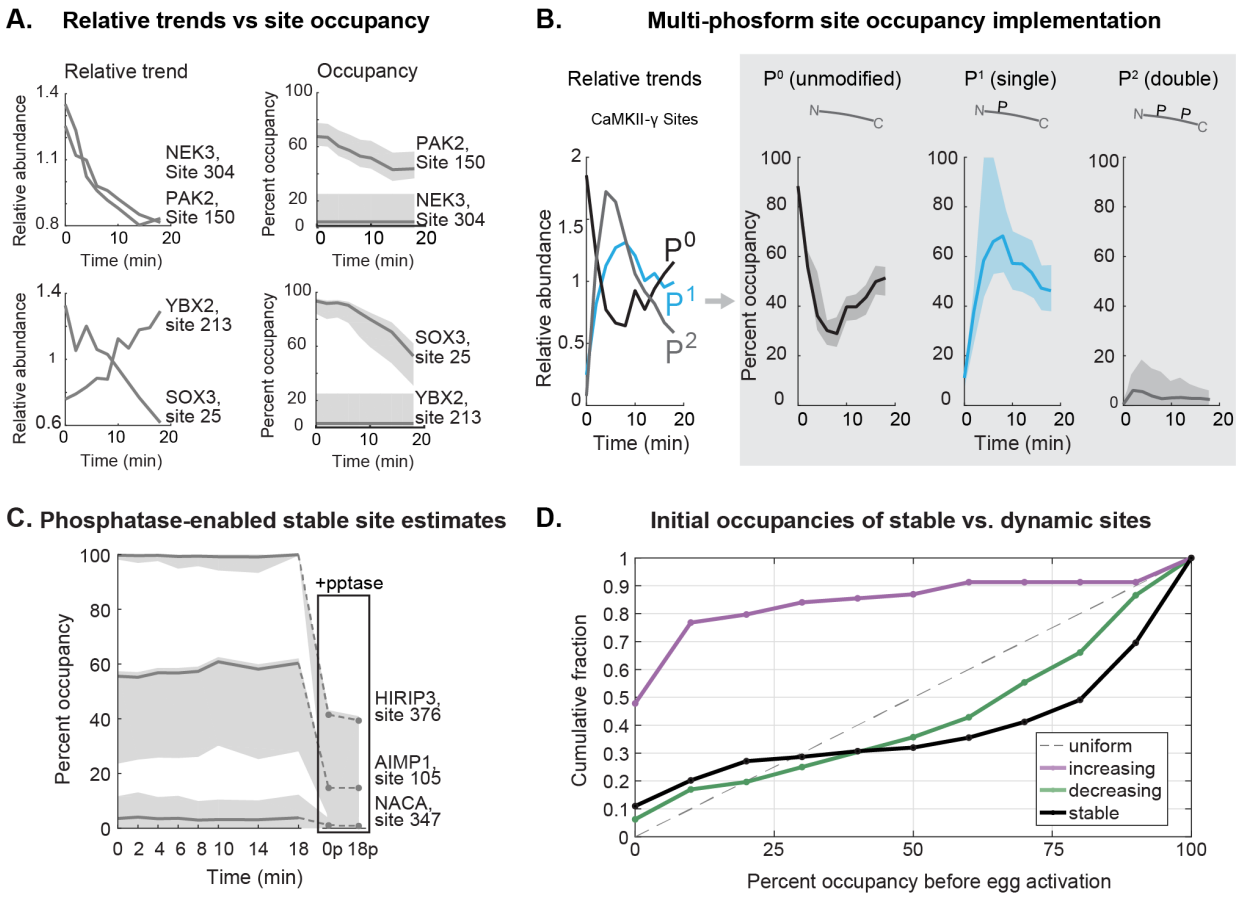

Figure 3.8 Occupancy of phosphorylation sites in time series post-activation. A) Time series of two kinases (NEK3, PAK2) and two transcription factors/nucleotide binding proteins (SOX3, YBX2) demonstrating that similar amounts of relative changes can give very different occupancy changes (shaded area is the 95% confidence interval). B) Phospho-occupancy time series of a multi-phosphorylated CaMKII-γ peptide. C) Reliable estimation of stable site occupancy is enabled by inducing dynamics with phosphatase treatment. Examples of stoichiometry estimated with phosphatase treatment are shown here. Treated conditions are replicates of the 0 and 18 minute time points (boxed). D) Cumulative distributions of the initial phosphosite occupancies at 0 minutes (unactivated egg) classified by whether the sites increase, decrease, or are stable after egg activation. A uniform distribution would lie on the dotted line.

The apparent lack of intermediate occupancies is largely a product of applying a confidence interval cutoff to the data. The cutoff improves data quality, but disproportionally filters stable sites (Figure 3.9).

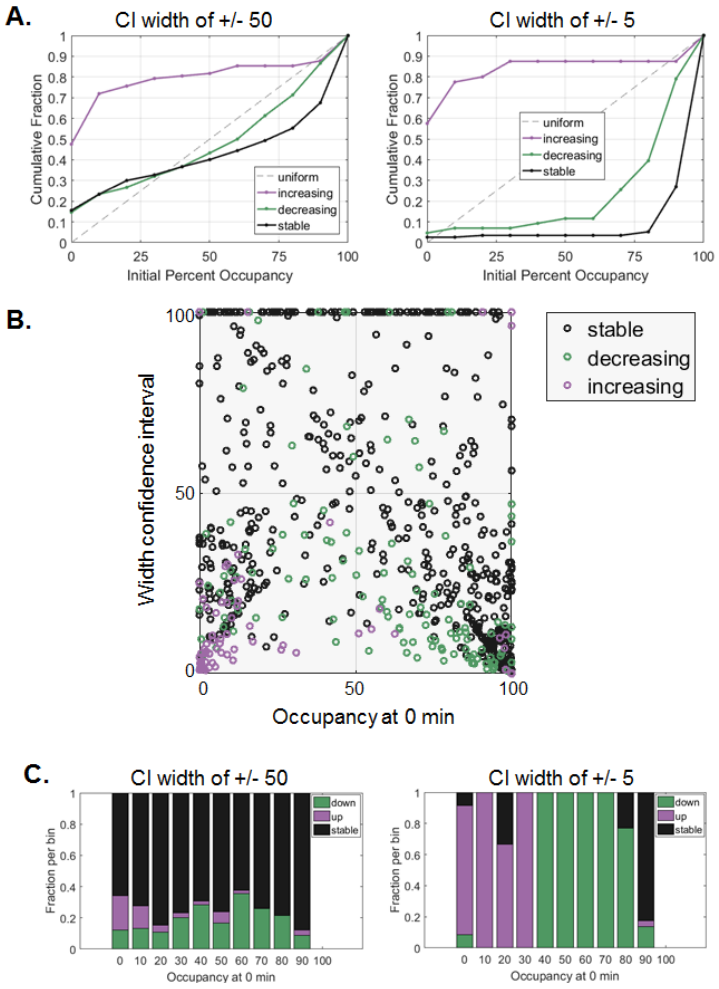

Figure 3.9 Intermediate occupancies are depleted with narrow confidence interval cutoffs. A) Cumulative distributions of the occupancies at 0 minutes for phosphosite trends which increase (purple), decrease (green), or are stable (black) with no cutoff confidence interval width (left) and a very stringent cutoff of +/- 5% (right). Narrow confidence interval cutoffs bias the distributions against intermediate occupancies. This is because the confidence intervals are limited to 0 and 100. Sites with intermediate occupancies have more potential for wide confidence intervals as they are less restricted by this limit. This is a problem for both dynamic and stable sites. However, because the estimate for stable sites is dominated by noise of the untreated conditions, the stable sites are particularly susceptible to this systematic bias against intermediate occupancies when applying cutoffs. B) To demonstrate, a scatterplot of the initial occupancy of a phosphosite at zero minutes is plotted against the width of the confidence interval for all sites where stoichiometry was calculated. The stable sites indeed show many sites with intermediate occupancies, but these sites often have more error (i.e., lie in the top half of the plot). C) The same data as B is plotted as a bar graph showing the fraction of each trend per unit occupancy. Comparing the left and right panels, the intermediate occupancies for stable sites (black) are depleted at a higher fraction relative to the other trends (purple and green) if the confidence interval thresholds are set too aggressively. We therefore set the threshold for this study at +/-25%.

Lastly, for corroboration of the approach, we used the matching phosphatase treated and untreated time points [152] to calculate occupancy using the previously reported two-condition analytical method [151]. We compared these results to the occupancies derived from applying our regression-based approach to the independently measured endogenous dynamics. The results are generally consistent (Figure 3.10).

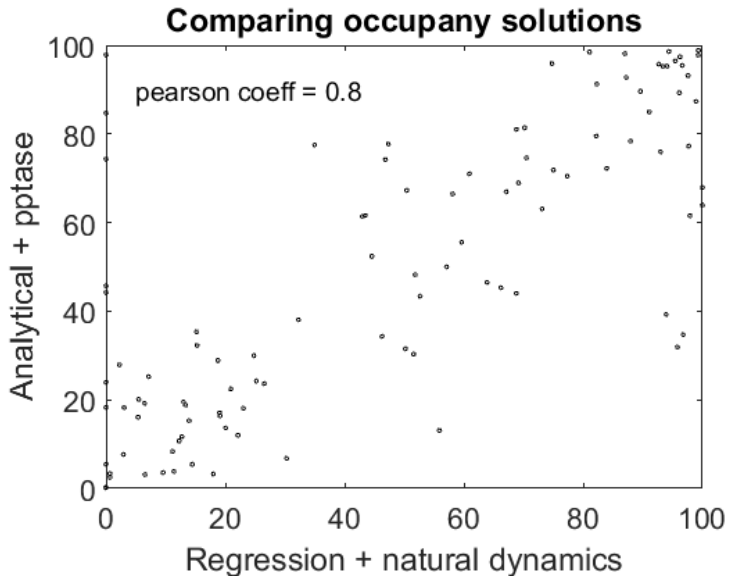

Figure 3.10 Corroboration of the regression-based occupancy approach. Phospho-specific antibodies were not available for the majority of detected phosphosites. To corroborate our regression methods, we relied on previously validated approaches to calculating stoichiometry for two conditions (one algebraic approach using natural dynamics from SILAC as input [151] and one with phosphatase treatment [152]). We used the treated and untreated phosphatase conditions as analogous to a SILAC pair and calculated the solution with the algebraic approach (Y axis). We compared the same sites that were measured in independent replicates and calculated their stoichiometry with our regression approach (X axis). In principle, these approaches should give similar answers despite different source of dynamics (i.e., natural vs induced) and method for calculating the solution (algebraic vs regression). The two approaches agree relatively well. Without an orthogonal experiment, it is difficult to say which method is "right" when the data disagree. However, it is reassuring that the two approaches largely agree.

3.5 Discussion

The utility of measuring phospho-occupancy is well-recognized, and several approaches have been developed to obtain stoichiometry from mass spectrometry data [151, 152, 206]. However, current approaches are not able to estimate occupancy for multi-phosphorylated peptides, stable sites, or sites measured with multiplexed proteomics. More fundamentally, these methods do not give statistical information (i.e., confidence intervals), which makes interpreting the estimates difficult. Our approach provides a unified analytical and experimental framework to address these limitations. Though we developed this approach to exploit the power of multiplexed proteomics, the advancements are compatible with other methodologies. For example, label-free phospho-proteomics typically have lower measurement precision but are conducive to high-throughput formats [214]. The additional conditions, in principle, give increased statistical power with the use of the regression framework. The higher sensitivity of label-free and other techniques [175] could mitigate a limitation of this approach, which is the low efficiency of measuring both the unmodified and phosphorylated peptides. Other methods for calculating phospho-stoichiometry, such as motif-targeting proteomics [215], could also incorporate this framework. Finally, the principles of the calculation could be extended to acetylation or ubiquitination, though we expect more success if occupancies are often >1%. These approaches can increase the power of the proteomics for analysis of posttranslational modification in diverse settings.

3.6 Experimental procedures— phosphatase treatment

To phosphatase treat some samples without affecting the untreated conditions after multiplexing or at any subsequent steps, we made use of a thermally unstable phosphatase, which can easily be inactivated. We used the temperature labile Shrimp Alkaline Phosphatase (Affymetrix, product #78390). To minimize volume added to the samples, the phosphatase was concentrated to ~4U / µL on a 5K Amicon Ultra filter, spun at 4°C. Since the supplied buffer contains Tris, which can interfere with TMT labeling, we exchanged this buffer with 10mM EPPS at pH 8.0 and stored the enzyme in 50% glycerol. We see no activity loss from the buffer exchange, as assayed by p-nitrophenyl phosphate (pNPP) hydrolysis. Following completion of trypsin digestion, we allowed the temperature to cool to RT then

added 10mM MgCl 0.1mM EDTA, 5mM EPPS pH 8.5. To the two phosphatase treated conditions, we added enzyme at a ratio of 3:1 peptide to phosphatase units and incubated for 12 hours at RT without purifying the peptides away from the proteases. The incubation is done in the presence of proteases as the dephosphorylated peptide creates a better substrate and allows for proteolysis of an otherwise missed cleavage caused by the charge of the phospho-group. Allowing for further cleavage of dephosphorylated peptides mimics the effect of an endogenous phosphatase more closely; otherwise the forms with missed cleavages will have lower abundance in the untreated samples and confound analysis. We determined that at RT, LysC does not cleave the phosphatase, while Trypsin does but at a reduced rate. The 3:1 ratio used gives sufficient activity while avoiding peptides from the proteolysis of the phosphatase from dominating the MS signal. After incubation, samples were lyophilized to dryness, and resuspended per the TMT-MS protocol discussed above. Importantly, all samples were incubated for 5 minutes at 65° C to destroy phosphatase activity in a water bath (inhibition under these conditions was confirmed beforehand with pNPP activity assays). Samples were cooled to RT before proceeding with the MS sample preparation.

3.7 Acknowledgements

I would like to acknowledge the insights of Dr. Katy Hartman, John Ingraham, and Dr. Randy King toward developing the calculation. Dr. Katy Harman assisted us in working through the early algebraic formulations of the problem. Dr. Hartman helped convince us that the general principle of measuring multiple conditions with TMT-MS conferred the analytical power to estimate stoichiometry for peptides with multiple phosphorylation sites. We may have abandoned the project too early without the clarity given by these comments. The decision to estimate confidence intervals from bootstrapping came from conversation with John Ingraham, who was also generally helpful in discussing the mathematics of the algorithm. Dr. King clearly articulated the shortcomings of interpreting post-translational changes with inferring the occupancy of the change fairly early in the project. These comments were highly influential in my decision to prioritize seeking a solution to the occupancy problem.

Chapter 4:

Proteomics of fertilization and meiotic exit in *Xenopus laevis*

Marc S. Presler, Elizabeth Van Itallie, Ryan Kunz, Peg Coughlin, Leonid Peshkin, Steven P. Gygi, Martin Wühr*, Marc W. Kirschner*

*These authors contributed equally

4.1 Attributions

The following chapter is focuses on the biological insights from a paper published in *PNAS* on October 27th, 2017 [155]. The analytical section of the paper is described in Chapter 3, which is applied here. The work focuses on the protein degradation, secretion, and phosphorylation of proteins measured in time series after egg activation. The project was conceptualized by Dr. Marc W. Kirschner, Dr. Martin Wühr, and myself. I conducted the majority of the experimental work, including the all of the electroactivation experiments, mass spectrometry sample prep, and phospho-enrichments. Elizabeth Van Itallie performed the experimental work to develop the phosphatase treatment conditions. Peg Coughlin provided technical and conceptual assistance to the electron microscopy. Dr. Ryan Kunz and Dr. Steven P. Gygi assisted with analytical conceptualization, obtaining the proteomics data, developing the initial phospho-enrichment protocols. Dr. Peshkin provided bioinformatics insights on analyzing the phosphorylation data and general bioinformatics help. Dr. Wühr offered close consultation and guidance throughout the project, and also performed the ubiquitin pull downs. All authors contributed to editing the manuscript, especially Dr. Wühr and Evi Van Itallie. I performed the data analysis, compiled the figures, and wrote the text.

4.2 Abstract

Fertilization releases the meiotic arrest and initiates the events that prepare the egg for the ensuing developmental program. Protein degradation and phosphorylation are known to regulate protein activity during this process. However, the full extent of protein loss and phospho-regulation is still unknown. We examined absolute protein and phosphosite dynamics of the fertilization response by mass spectrometry-based proteomics in electro-activated eggs. To do this, we developed a new approach for calculating the stoichiometry of phosphosites from multiplexed proteomics that is compatible with dynamic, stable and multi-site phosphorylation. Overall, the data suggest that degradation is limited to a few low abundance proteins. However, this degradation promotes extensive dephosphorylation that occurs over a wide range of abundances during meiotic exit. We also show that eggs release a large amount of protein into the medium just after fertilization, most likely related to the blocks to polyspermy. Concomitantly,

there is a substantial increase in phosphorylation likely tied to calcium activated kinases. We identify putative degradation targets as well as new components of the slow block to polyspermy. The analytical approaches demonstrated here are broadly applicable to studies of dynamic biological systems.

4.3 Significance Statement

Protein phosphorylation and degradation drive critical events in early embryogenesis and the cell cycle; however, comprehensive and accurate analysis of these changes is currently difficult. Using a mass spectrometric approach, we present a quantitative view of the protein and posttranslational economy of the fertilization response in the frog egg. Protein degradation affects a small but very important class of proteins, while regulatory phosphorylation and protein release occur on a far larger scale. We have developed new, broadly applicable analytical methods for phosphorylation that provide absolute quantification with confidence intervals for improved interpretability of post-translational modification analysis.

4.4 Introduction

For decades, the highly synchronous events of fertilization have provided a useful system for the study of many aspects of cellular regulation, especially protein degradation and phosphorylation. The destruction of Cyclin-B and other proteins is catalyzed by the anaphase promoting complex (APC/C), which promotes M-phase exit [27]. The two activators of the APC/C are Cdc20 and Cdh1. In the egg, the cell cycle typically involves only Cdc20 [50]. While the list of known Cdh1 substrates continues to grow [75], the Cdc20 target list remains small [66, 78, 79]. It is of considerable interest to characterize the minimal set of Cdc20 substrates that powers the early cell cycle. Kinase activity is equally important to this regulation. Cyclin-B degradation promotes mitotic exit by inhibiting the activity of Cyclin-dependent Kinase 1 (Cdk1). There is a bulk loss of phosphorylation following egg activation [26]. The identities of the vast majority of these phosphoproteins remain undiscovered. However, there is a strong expectation that this regulation reflects the reversal of Cdk1 phosphorylation. Fertilization employs additional regulation not common to other cell cycles. There are timed waves of phosphorylation that correspond to the release

of cortical granules just after fertilization as part of the slow block to polyspermy. This release is calcium sensitive, and may reflect increases in the activity of Protein Kinase C [15, 16] and CaMKII [17]. An account of the secreted proteins, their function, and their upstream signaling components is limited. To investigate these unknown aspects of degradation, release, and modification of proteins at fertilization comprehensively, we employed mass spectrometry (MS) in *Xenopus laevis* eggs. Recent advances in multiplexed proteomics allow quantitative comparisons of multiple conditions using Tandem Mass Tags (TMT) [134-136]. Our recent work demonstrates the power of proteomics in *Xenopus* for studies of early development [154, 216, 217]. Though fertilization has been studied before with MS [116, 218-220], the studies were either qualitative or did not measure phosphorylation. With an enrichment step [147-149], it is feasible to measure relative changes in a large number of phosphoproteins. This paper introduces new biological findings about fertilization and cell cycle progression, but it also introduces new methods for measuring absolute stoichiometry of phosphosites, widely applicable to MS protein modification studies.

4.5 Results

4.5.1 Quantitative proteomics of egg activation and meiotic exit

The large (1.2 mm) *Xenopus laevis* egg offers superb synchrony and sufficient material (~30µg of non-yolk protein per egg [121]) for proteomic and phosphoproteomic studies. To capture the progression of the rapid fertilization response (Figure 4.1A), metaphase-arrested eggs were activated with electro-shock and snap-frozen every two minutes in sets of 30 (Figure 4.1B). We chose electrical activation over sperm addition to maximize time resolution by avoiding the asynchrony of sperm penetration through the jelly coat [85]. The early morphological and molecular events to our knowledge are equivalent between fertilization and parthenogenetic activation [92, 93], hence we use the terms interchangeably here. More than 99% of eggs activate with a standard deviation of <15 sec using this approach (Appendix 1, Figure S1). Time points were analyzed by MultiNotch MS3 [138]. Phosphopeptides were enriched on a titanium dioxide column; the flow-through was used for protein-level MS. We multiplexed the TMT-labeled time points before enrichment to eliminate errors that arise from variation between columns. Protein-level MS was performed with four biological replicates, and phospho-level MS was performed in biological triplicate. We quantified ~8,700 proteins (80% detected in two or more replicates) and ~3,500 phosphosites

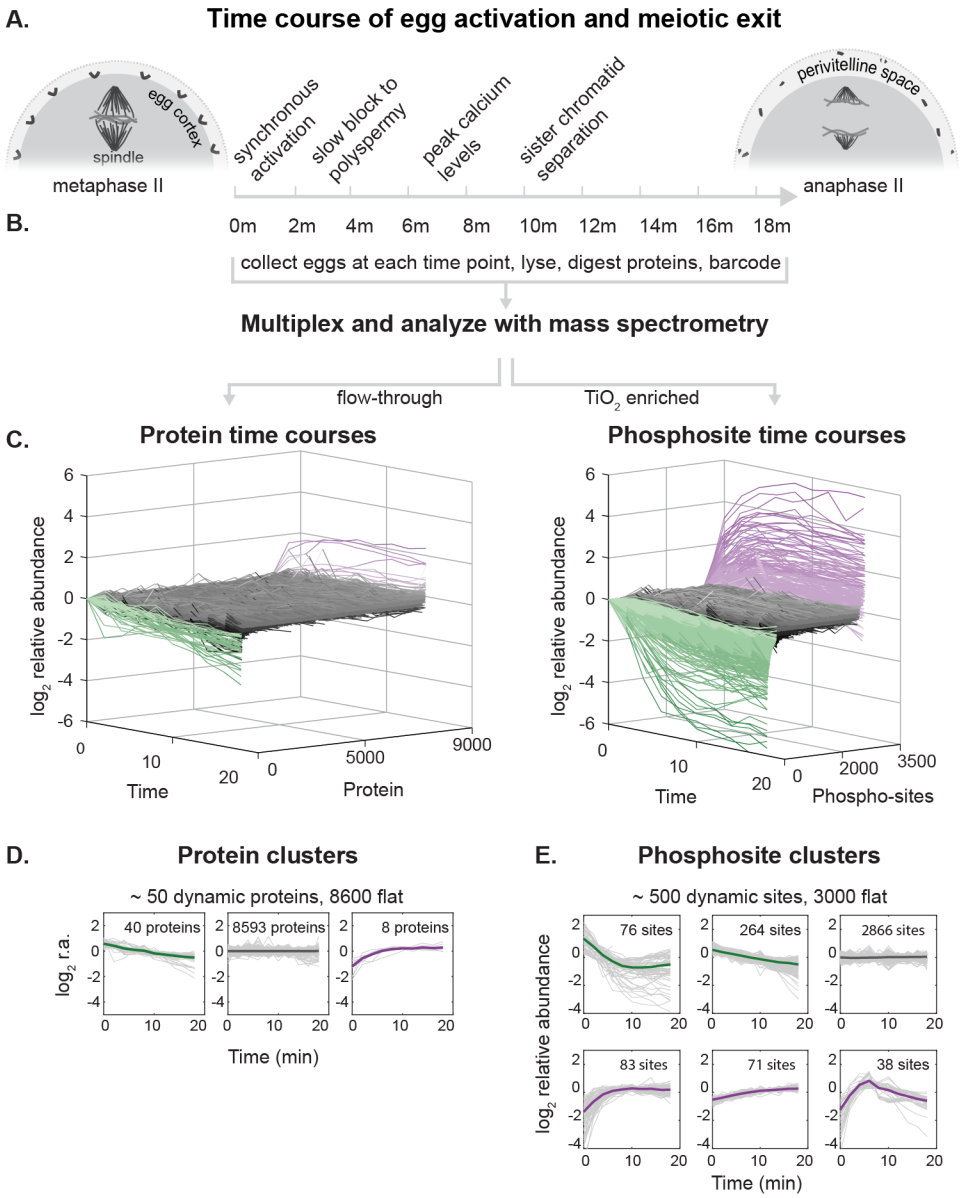

Figure 4.1 Experimental overview. A) Fertilization and egg activation release metaphase-arrested oocytes into anaphase and initiate the slow block to polyspermy, sister chromatid separation, remodeling of the specialized egg extracellular matrix and inflation of the perivitelline space, among other processes. We mimic fertilization and trigger egg activation via electric shock to maximize synchrony. B) After activation, eggs were collected every two minutes, lysed, and digested with proteases. Samples were barcoded using Tandem Mass Tags (TMT) and multiplexed. Protein and phosphosite dynamics were measured with mass spectrometry-based proteomics. Phosphorylated peptides were enriched using TiO_2, and the column flow-through peptides were used for protein analysis. C) Modified waterfall plots displaying the trend of every protein and phosphosite in the dataset, normalized to the first time point and ordered first by clusters (see below) and then in ascending order within each cluster. D, E) K-means clustering to summarize dynamic classes of the protein (D) and phosphorylation time series (E) (see Appendix 1 Figure S8-9, Methods). Bold line represents the centroid of the cluster while the gray lines are individual time courses (normalized to the time course means).

(40% in two or more replicates) on 1,700 proteins (Figure 4.1C). Less than 1% of detected proteins showed abundance changes. 40 proteins decrease in abundance over time, while 8 unexpectedly show an apparent increase (Figure 4.1D) (discussed below). Protein phosphosites are notably more dynamic, with ~15% of modified residues changing (Figure 4.1E). There is clear evidence for parallel dephosphorylation and phosphorylation. The phosphosite changes overwhelmingly occur on stable proteins, thus they reflect actual changes in phosphorylation level rather than protein level.

4.5.2 Protein loss following fertilization

We sought to determine whether additional proteins beyond the known cell cycle targets are degraded. We identify 29 unique proteins by gene symbols that significantly decrease in abundance. To establish statistical significance, we calculated a randomization-based False Discovery Rate (FDR) (Appendix 1, Figure S2) and set a cutoff of 1%.

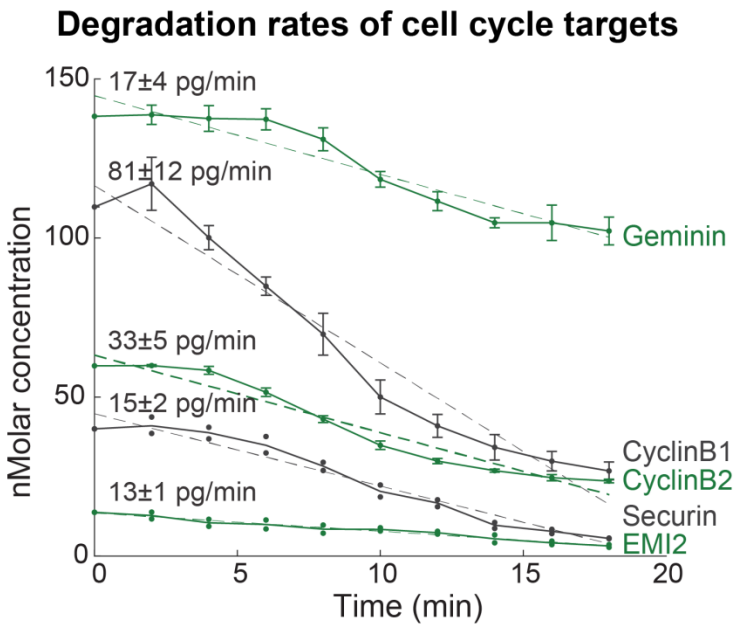

Figure 4.2 Absolute rates of protein loss for degradation targets. Plot shows a time series of the 5 proteins identified as significantly decreasing which are known cell cycle degradation targets, plotted by their absolute changes. Error bars reflect standard error (SEM) for proteins detected in at least three biological replicates. For proteins detected in fewer than three replicates, all data available is shown as points. Dashed lines represent a linear fit to the approximate zero-order kinetics (labeled with the slope and 95% confidence interval).

Specifically, we find known APC/C^{Cdc20} substrates, including Cyclin-B1, Cyclin-B2, Geminin [64], Securin [33, 65], and the β-TRCP substrate EMI2 [61] decrease significantly (Figure 4.2). Previously, we estimated the absolute abundances of ~11K proteins in the frog egg [154], which allowed us to estimate absolute rates of degradation for these proteins. We see a small delay of ~2-4 minutes between fertilization and initiation of degradation of the APC/C substrates.

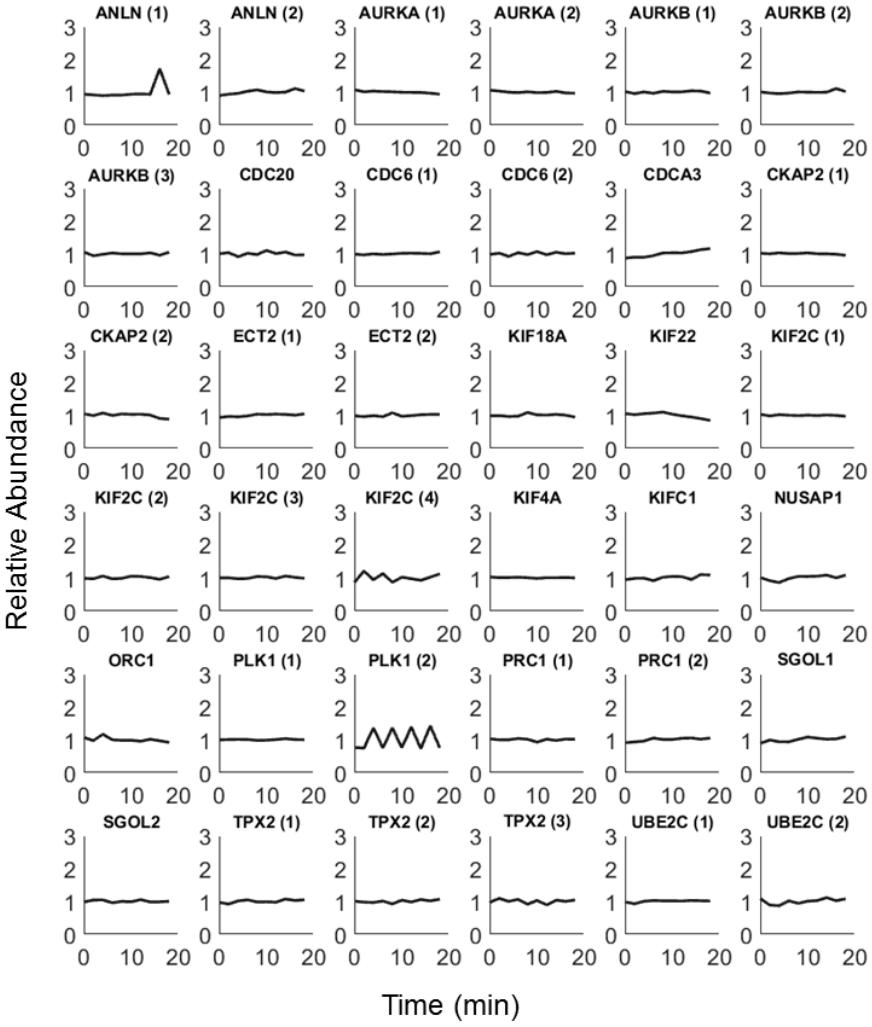

Figure 4.3 Known APC/C^{Cdh1} substrates are stable post-fertilization. Time series data of known targets of the APC/C^{Cdh1} following egg activation. The substrates are stable (i.e., do not pass the FDR threshold of 1% for decreasing, see Appendix 1 Figure S2). Repeated gene symbols indicate multiple protein IDs, which are likely alloaleles or splice variants. The proteins are CDC20, Plk1, Aurora A/B (see citations in main text), ANLN [221], Cdc6 [222], , CDCA3 (TOME1) [223], Ckap2 [224], Ect2[225], KIF18A, KIF2C, KIF4A, KIFC1[75], NUSAP1 [226], ORC1 [227], Ase1/PRC1 [228], Shugoshin [229], TPX2 [230], UBE2C [231]. KIF22 [232] is the only substrate that shows some marginal reduction in abundance, but it does not pass our FDR threshold and was previously shown in the literature to not interact with CDC20.

EMI2, whose destruction is required for APC/C activity, declines without delay. It is reduced by 30% 4 minutes after activation, when Cyclin-B1/2 and Securin begin to decline. EMI2 is stoichiometric with the least abundant components of the APC/C [154]. Therefore, a ~30% decline of EMI2 leads to a 30% increase in the maximal activity of APC/C. This activity must therefore be sufficient to initiate degradation. All observed putative APC/C^{Cdh1} substrates, such as CDC20 [57, 69, 70], PLK1 [66], and Aurora A/B [67, 68], are stable (Figure 4.3).

The total loss of the known APC/C^{Cdc20} substrates is ~3,000 pg (or 250 nM), and they are degraded at a rate of ~150 pg/min. From single molecule measurements of the proteasome [233], we estimate the reactions responsible for protein degradation are approximately diffusion-limited for the reaction (see Methods). Since it is possible to competitively delay substrate degradation by adding peptide fragments of Cyclin B [30, 64, 234, 235], the reaction in the egg may be near saturation. The overall estimated rates may be affected by the cell cycle metachrony (i.e., spatial heterogeneity) of the large *Xenopus* eggs. The cell cycle state should equilibrate between the top and bottom of the egg by ~20 minutes given the speed of the traveling “waves” [99, 236, 237] and other previous measurements [238]. Therefore, the end points we report are likely to be accurate. The half times may be overestimated if there is substantial delay of degradation in the vegetal pole, though the exclusion of the cytoplasm by yolk may partially compensate for this phenomenon.

4.5.3 Classifying candidate novel degradation targets

Annotation of the decreasing proteins suggested that some candidate proteins were released from the cell rather than degraded. To test this hypothesis, we harvested the egg media in time series after egg activation (Figure 4.4A). Proteins that appeared in the media and disappeared from the cell were classified as released (Figure 4.4B and Appendix 1, Figure S3). Release from the egg is the major source of protein loss rather than degradation (Figure 4.5). We found only six proteins that decreased abundance with no evidence for release (see Appendix 2, Table S1 and Methods section for full classifications and rationale). To determine whether these proteins are degraded through the ubiquitin-pathway, we performed ubiquitin pull-downs with a di-glycine antibody [239] and analyzed the samples by MS. Of the six proteins that were not previously known degradation targets, two were ubiquitinated. A *Xenopus* homolog of the gene SSX2IP (distinguished as SSX2IP-L) is the most promising candidate for a

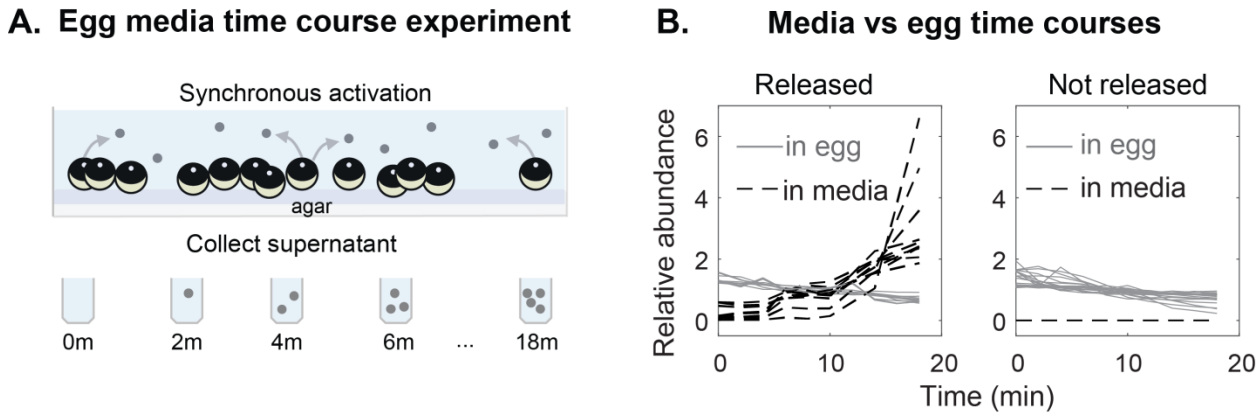

Figure 4.4 Classifying the mechanism of protein loss following egg activation. A) Experimental design to test for proteins released by eggs upon fertilization as an explanation for protein loss. Released proteins will increase in the media fraction over time. B) Time series for proteins detected in both the egg and supernatant or the egg alone. These data were used to classify whether proteins were lost by release rather than degradation (by direct evidence or annotation, see text). This class comprised all but two of the significantly decreasing proteins besides the known targets.

novel APC/C substrate (Figure 4.5). SSX2IP is a centrosomal protein [240] with roles in the primary cilium [241]. It is present at similar abundance to Cyclin B (60nM) and contains a confident D-box prediction [78]. Ciliary defects occur when Cdc20 is depleted [77]. SSX2IP may provide a mechanism for linking Cdc20 to the ciliary phenotypes. The other candidate is NPDC1, a neuronal protein with some association with cell cycle components [242] containing a weak D-box motif. However, it is localized to synaptic vesicles in neurons, which may suggest association with exocytosis machinery rather than degradation (though we do not detect it in the media). At the level of sensitivity in this study, the number of putative APC/C^{Cdc20} substrates is indeed very small. All of them are below 100nM in the egg. To estimate the sensitivity of our study, we compared the proteins identified using TMT-MS to those identified by the more sensitive (but semi-quantitative) label-free approach in the frog egg [154] (Appendix 1, Figure S4). At 50nM, we detect 95% of the proteins with TMT-MS as with label-free, and still detect 50% at 10nM. While we cannot rule out additional substrates, the data does suggest that any we did not detect are at low concentrations.

4.5.4 Proteins released from the egg

By contrast to the relatively small mass and small number of proteins degraded after fertilization, there is a substantial reduction in cytosolic protein due to release into the medium. We detect the expected cortical granule components like an Interlectin2 homolog, as well as the proteases

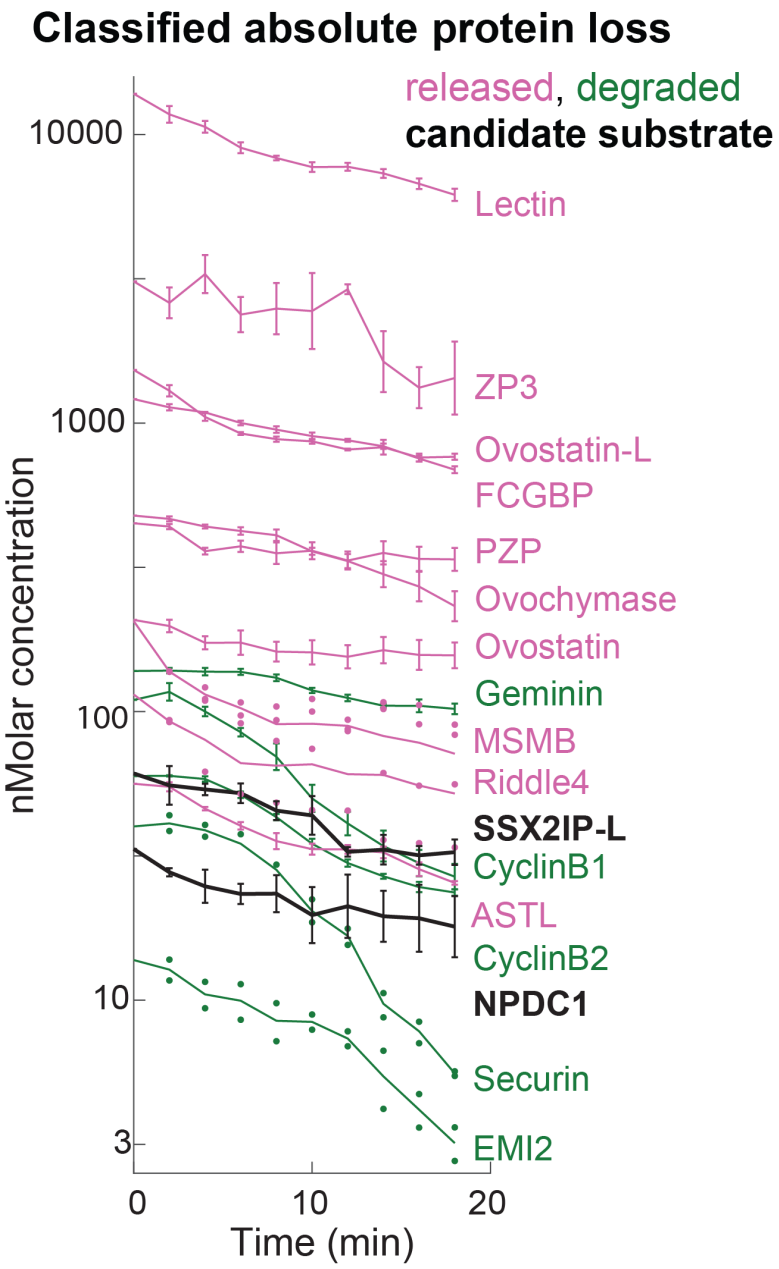

Figure 4.5 Protein loss occurs by degradation and release. Plot shows time series of all significant protein loss plotted by their absolute abundance (log10 transformed) and classified by the mechanism of loss (by release degradation, or candidate degradation target). Most released proteins were classified as such by showing corresponding loss in the egg and increase in the media. A few are surmised by their annotation, even if they are not directly detected in the media given the evidence for the general phenomenon. Error bars reflect standard error (SEM) for proteins detected in at least three biological replicates. For proteins detected in fewer than three replicates, all data available is shown as points. SSX2IP-L and NPDC1 homologs (black) are putative new degradation targets. These are proteins which are significantly decreasing after egg activation, ubiquitinated, and not secreted or annotated as secreted, among other criteria (see Methods).

Ovostacin [111] and Ovochymase [112] whose function is thought to be inhibition of further sperm binding by cleaving the extracellular Zona Pellucida (ZP) gene family [110]. We detect ZP2, 3 and 4 homologs [243] as released, which are likely to be cleaved peptide fragments that have diffused into the media (Appendix 1. Figure S3). Therefore, at least two mechanisms are responsible for the release of protein from the interior of the egg and from its surface: exocytosis of cortical granules and proteolysis of the specialized extracellular matrix.

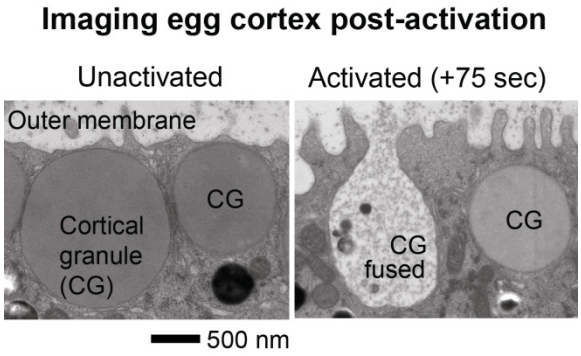

Figure 4.6 Electron microscopy images of cortex of the Xenopus laevis *egg. The cortical granules (CG) are specialized exocytic vesicles that are docked on the outer membrane of the egg before fertilization. After egg activation, cortical granules fuse with the outer membrane and expel their contents. Images show the fusion of a cortical granule within 75 second of egg activation. Lower magnification fields are presented in Figure 4.7.*

We would not detect change for proteins that are trapped in the perivitelline space. We detect several proteins previously unknown to be released from the egg at fertilization, including several protease inhibitors. One examples is Ovostatin (AM2 homolog) [244], which was previously shown to have anti-trypsin activity in the sea urchin egg [245]. There are additionally a small number of other annotated protease inhibitors that decrease in the egg but are not detected in the medium. We surmise that these are released as well, rather than degraded (Appendix 2, Table S1). One example is Riddle4 [246] with known functions reported later in development [247]. While the secretion of proteases is a well-recognized mechanism, the release of endogenous protease inhibitors in response to fertilization was not known. Another previously unknown released protein is a FCGBP homolog (Figure 4.5), which is similar to Zonadhesin [248, 249] that binds to the ZP proteins. It is likely extracellular and liberated by proteolysis.

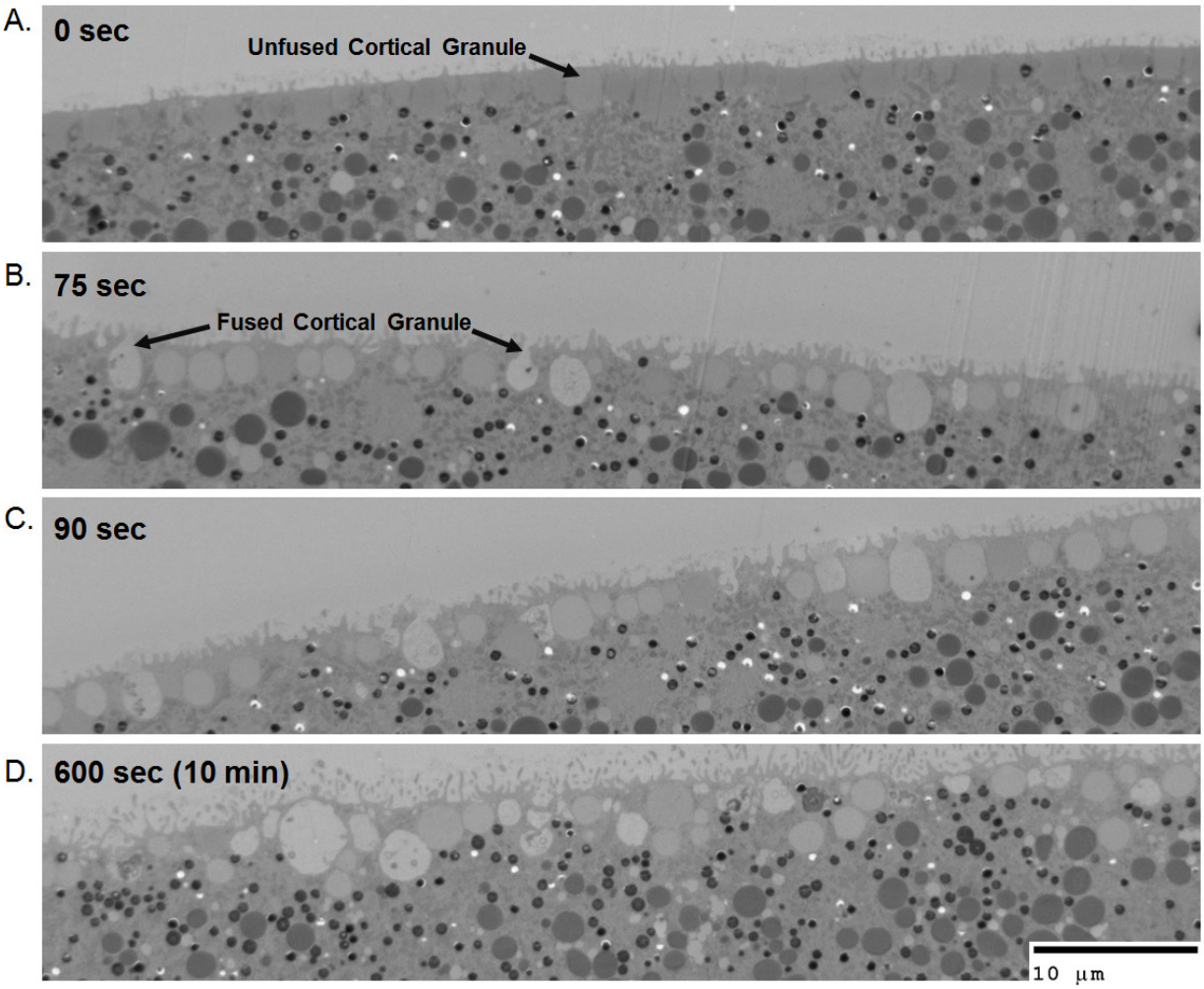

Figure 4.7 Thin sections of the outer membrane at the animal pole of Xenopus laevis *eggs at 690x magnification. A) Outer membrane of the egg before activation occurs. No fused cortical granules are seen here. B) 75 seconds post egg activation. C) 90 seconds post egg activation. D) 10 minutes post egg activation. ~50-60% of cortical granules depicted have released their contents through fusion with the outer membrane. At this time, fused cortical granules appear either as invaginations of the membrane or as low-density spheres (when the actual fusion event is not visible in the section). The vitelline membrane is not present in these images (Methods).*

To compare the measured release of protein with the major morphological events of fertilization, the cell cortex of activated eggs was imaged by electron microscopy (EM) every 15 seconds (Figure 4.6). The release of cortical granules shows a 60 second delay, but between 75 and 90 seconds ~40% of the vesicles fuse with the outer membrane. By 10 min, the majority (~60%) of the vesicles have fused (Figure 4.7). This is consistent with the generally faster loss rates of released proteins (Figure 4.5) and with previous work [16].

4.5.5 Observed protein increase is a phosphorylation artifact

A cluster of 8 proteins appear to accumulate post-fertilization (Figure 4.1D), 7 of which show a significant increase in abundance after the FDR analysis. Many of these proteins are typically stable and not expected to be synthesized during anaphase (e.g., NUP35, a nucleoporin [250]), which led us to search for an alternative explanation. Since protein trends are determined from the sum of all unmodified peptides, we examined the individual peptide trends to determine whether they all showed the same unexpected increase.

We found that these proteins did in fact have stable as well as increasing peptides (e.g., Figure 4.8). This discrepancy can occur from the loss of a modification on that particular peptide. The relative abundance of the unmodified peptide must increase if the modified form decreases, due to conservation. Indeed, for three of these proteins, we can directly show the loss of phosphorylation that causes the increase in the unmodified peptide (Figure 4.8). Overall, we have evidence that 6 of the 7 apparently increasing proteins have phosphopeptides showing dramatic decreases in phosphorylation. Therefore, we conclude that the increasing protein trends are most likely caused by dephosphorylation rather than rapid protein synthesis. We controlled for phosphorylation artifacts in protein loss as well, and found evidence that this impacted only two proteins (Appendix 2, Table S1). We found that analyzing the data with metrics that reduce the effect of outliers (e.g., median) mitigates but does not eliminate specious trends caused by dynamic modifications. This is because multiple peptides are often affected by modifications (e.g., Figure 4.8C), as is the case for the majority of the proteins discussed here.

4.5.6 Protein phosphorylation dynamics following egg activation

To explore the function of the regulated phosphosites, we performed motif and gene set enrichment analysis (GSEA) on the dominant trends of the dataset: increasing and decreasing

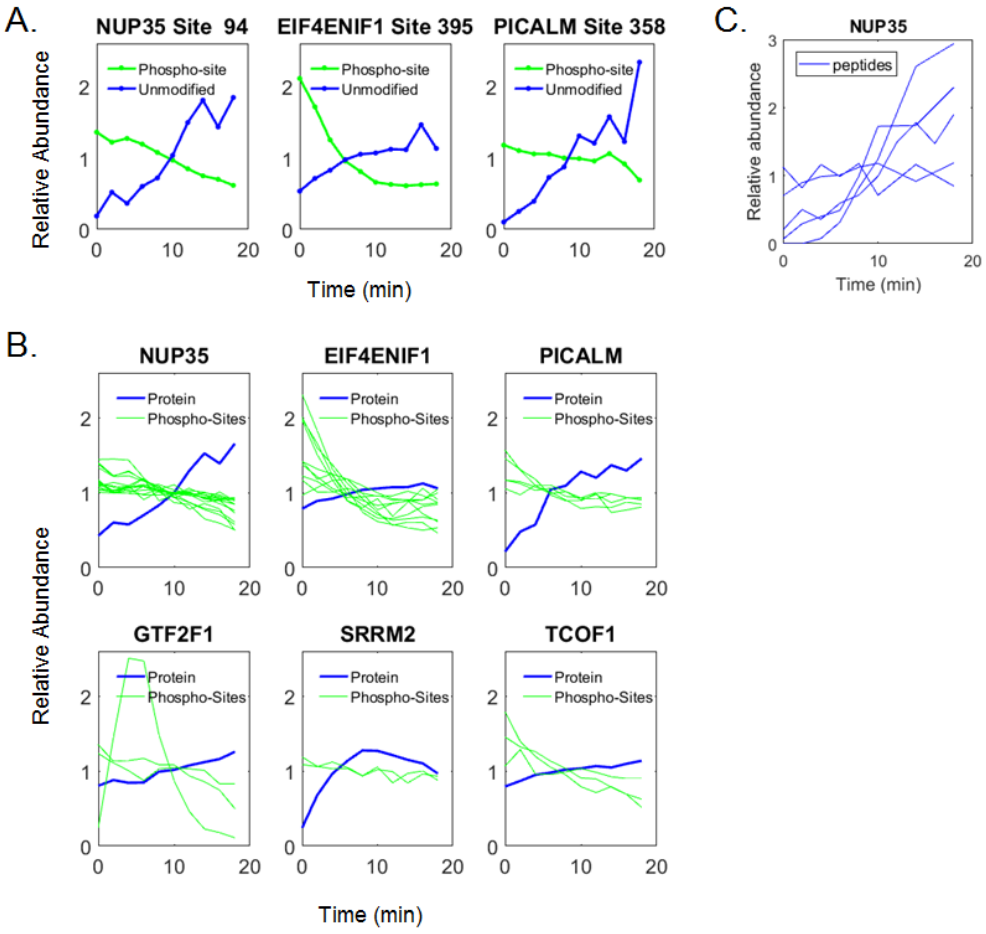

Figure 4.8 Evidence that dephosphorylation causes specious increasing protein trends. A small number of proteins significantly increase after fertilization at 1% FDR. Rather than synthesis, this can occur on stable proteins because of the loss of a modification on a particular peptide.
A) Times series of an unmodified peptide (blue) that increases in signal with matching a corresponding phosphosite residue (green) that decreases in signal for three of the proteins with apparently increasing proteins. This directly shows that the increasing signal is due to dephosphorylation, not protein synthesis.
B) Time series of 6 of the 7 proteins (blue line) which apparently increase that have multiple dephosphorylated sites (green) detected on the proteins. The top 3 are the same proteins as Figure 4.8A, but with additional sites shown. For the bottom three we do not detect exact matches of reciprocal modified and unmodified peptides. We do see multiple dephosphorylated sites for each protein is consistent with the phenomenon discussed in Figure 4.8A. For the 7th protein that increases (POLR2A, a RNA polymerase II subunit), we detect only one phosphosite, which is stable. There are many other polymerases dephosphorylated in the dataset, it is likely POLR2A is dephosphorylated as well.
C) All the peptides for NUP35 identified in the study. This is an example of how the unmodified peptides (one of them shown in Figure 4.8A) are dynamic from the loss of a modification, while the other two reflect the true flat trend of the overall protein. In this case and several others, using the median value at each time point, or even the median fold change, would still show a specious increase, as multiple peptides are affected by the same phenomenon from the loss of a modification.

stoichiometries. While the trends we detect are a subset of the total sites in the egg, their signatures are nevertheless revealing of the classes of regulation that are occurring. We were careful to exclude misleading trends that occur during multisite phosphorylation where, for example, the increase of a single form is actually the loss of a double form (Figure 4.9, Methods).

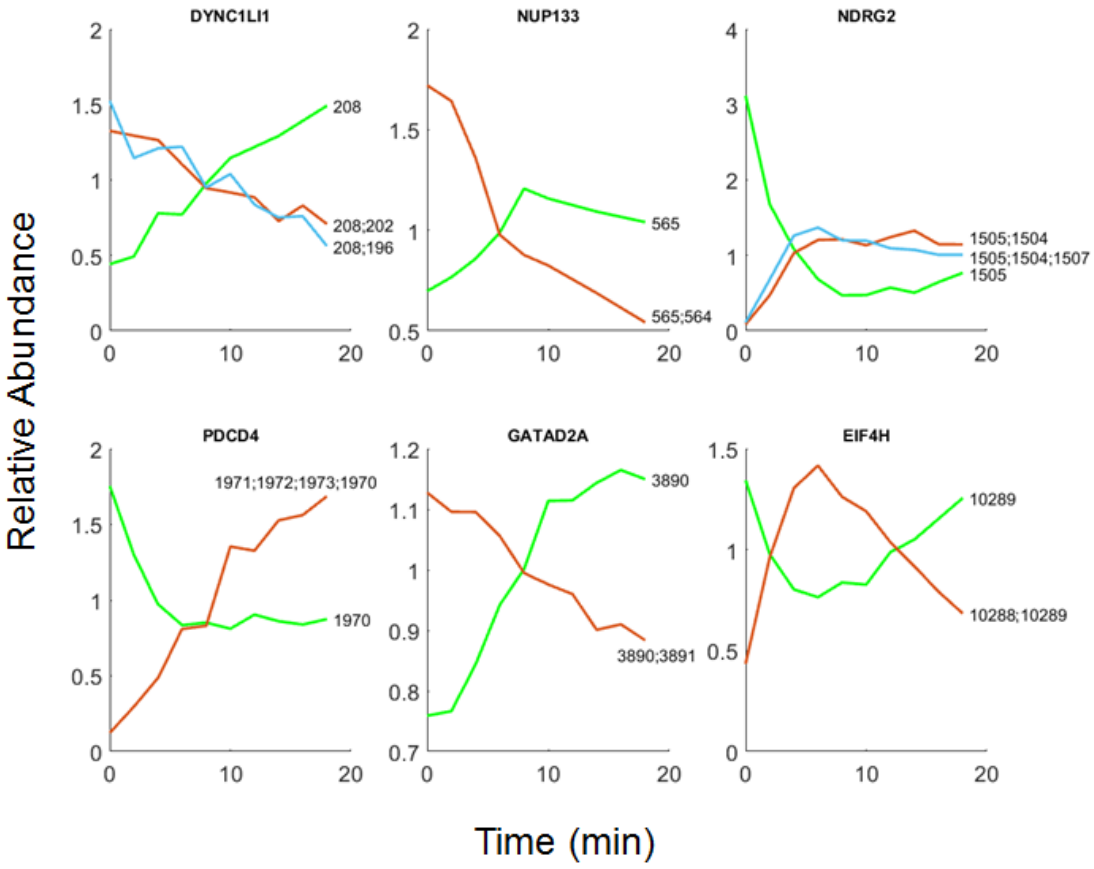

Figure 4.9 Specious trends occur between single and multi-phosphorylated forms of peptides. Figure displays time series of multi-site phosphorylated peptides for the single (green, first multi-site species (i.e., a double site, orange), and then the second (if present, blue) forms. Just as in Figure 4.8, an otherwise non-dynamic singly phosphorylated form can appear to rise in abundance if the doubly phosphorylated form decreases due to conservation. This means the true direction of change may be difficult to interpret for single sites that have multiple other sites on the same peptide. We detect this phenomenon for ~10% of the dynamic phosphosites, which we correct for during clustering by discarding the single form and keeping the most-modified form (i.e., if the single form is phosphorylated the multi-site should also rise, whereas if the double site is dephosphorylated, the non-dynamic single site will increase).

Since relative trends are sufficient for motif and GSEA, we used all the data to increase statistical power. The minimal motifs for Cdk1 and MAP kinases (S/T-P) explain ~70% of the decreasing sites (Figure 4.10A). We show several examples of the absolute dynamics for these phosphosites, calculated from

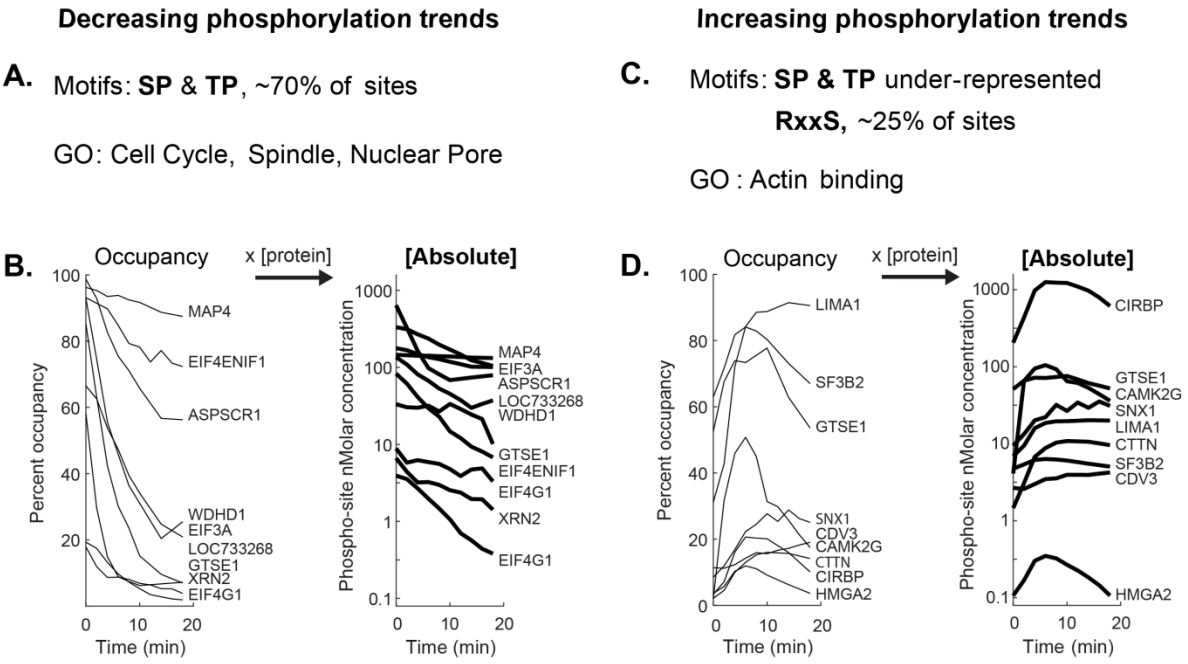

Figure 4.10 Phosphosite dynamics following egg activation. Analysis of the phosphorylation trends (increasing and decreasing) following egg activation with individual examples. A) Motif enrichment analysis (p <<0.01) results and gene set enrichment analysis (GSEA) of GO terms (p<0.01) results for dephosphorylated proteins. B) Time series of selected proline-directed phosphorylation plotted as occupancy and absolute dynamics (log10). All trends shown pass a 95% confidence interval width threshold of ±25%. C) Motif analysis results as in Figure 4.10A for proteins with increasing phosphorylation trends and GSEA results. D) Examples of increasing phosphorylation (plotted as in Figure 4.10B).

our estimated site occupancies and protein concentrations [154]. The majority of these show substantial loss of phosphate, which is consistent with the expected reversal of mitotic phosphorylation (Figure 4.10B). The rates of the trends align with cell cycle events. The majority of proteins are dephosphorylated with a halftime of ~10 min (some with <5 min) with minimal delay after egg activation. This matches the timing of sister chromatid separation in anaphase [93, 251] and the half time of CyclinB1/2 degradation

(Figure 4.2). GSEA of the dephosphorylated proteins are annotated with cell cycle, spindle, and nuclear pore functions (Figure 4.10A, specific examples in Appendix 1, Figure S6).

In parallel, we also see a strong signal for increasing phosphorylation. The (S/T-P) motifs are under-represented in this class. Instead, the minimal motif for the calcium-sensitive CaMKII is enriched, explaining 25% of the sites (Figure 4.10C). Many of the increasing phosphorylation trends are fast, with half-times of 5 min or ~2 min (Figure 4.10D). These are matched temporally with the calcium wave, cortical actin rearrangement [252-254], and exocytosis events (Figure 4.6). Indeed, the phosphorylated proteins are enriched for actin-binding GO terms; all of the members of this set are also annotated as "Cell Periphery" (Figure 4.10C). Given the localization of these events to the cortex, their fast rates, and the calcium sensitivity of cortical granule exocytosis [16], the increasing phosphorylation may in part promote the release of protein into the media.

4.5.7 Nucleoporin dephosphorylation corresponds to order of NPC assembly

There are many intriguing vignettes in the data. One example is the differential rates of dephosphorylation of nuclear pore sub-complexes (Figure 4.11 and Appendix 1, Figure S7). During entry into mitosis, nucleoporin phosphorylation promotes Nuclear Pore Complex (NPC) disassembly [255]. The sequence of post-mitotic NPC reassembly may be controlled by differential reversal of the mitotic phosphorylation [256]. Our observations give some support for this hypothesis. Rates of dephosphorylation on nucleoporin phosphosites cluster by regions [257] of the NPC (Figure 4.11). These rates mostly conform to the sequence of assembly [258, 259]. The fast dephosphorylation (~2-5 min half times) of inner basket and nuclear ring components (e.g., NUP153) is consistent with their early roles in NPC assembly [260]. The slower dephosphorylation (~10 min half time) of the Core Scaffold components (e.g., NUP188) is consistent with their later recruitment (Appendix 1, Figure S7). One notable exception is that while the incorporation of NUP98 is reported to be an intermediate assembly step, it is dephosphorylated rapidly (4 min half time). Phosphorylation of NUP98 was previously shown to be rate limiting for NPC disassembly [261]; it may be that its dephosphorylation is required or limiting for reassembly as well. There are two important caveats to the kinetic discrepancies: 1) nucleoporins or nucleoporin partial complexes may function outside the nuclear pore (e.g., NUP98 at the spindle) [262, 263]; 2) the nuclear envelope in the egg is packaged in structures called annulate lamellae, which also

repackage dynamically after fertilization [264]. Our observations may reflect this repackaging, rather than the canonical post mitotic reassembly. For example, the large relative change but small stoichiometric decrease of the transmembrane nucleoporin TMEM48 may indicate a separate pool of molecules, perhaps on the annnulate lamellae (Figure 4.11). Nevertheless, these differential rates provide new information on how NPC reassembly may be regulated in the egg and by extension in other circumstances.

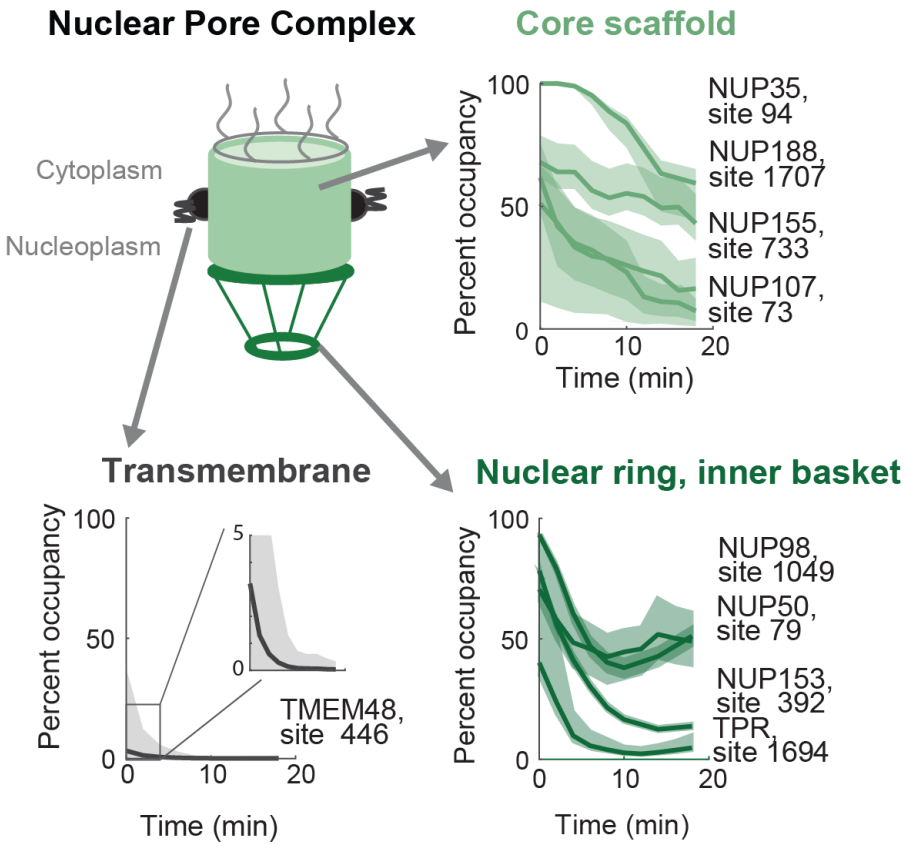

Figure 4.11 Nucleoporins show differential dephosphorylation corresponding to nuclear pore sub-complexes. Figure shows time series of nucleoporin phosphosite dynamics plotted by their respective sub-complexes in terms of occupancy (shaded area are 95% confidence intervals). The slower dephosphorylation (~10 min half time) of the core scaffold and cytoplasmic components is consistent with their later recruitment. The fast dephosphorylation (~2-5 min half times) of inner basket/nuclear ring components (e.g., NUP153) is consistent with their early roles in NPC assembly of binding chromatin and nuclear envelope association. An exception is that trans-membrane nucleoporin recruitment is a secondary step, yet nearly all the sites that we detect in this region show fast dephosphorylation. Another exception is the scaffold component NUP107, which is recruited early, but shows slow dephosphorylation.

4.5.8 Ca^{++} sensitive responses show substantial differences in occupancy

The data also offer insights into the pattern of calcium-initiated signaling. The calcium wave peaks in cytosolic concentration around 5 minutes post-fertilization and declines thereafter [100]. Figure 4.12 shows a set of relative phosphorylation changes from a larger cluster (Figure 4.1E) that correlates with the calcium wave.

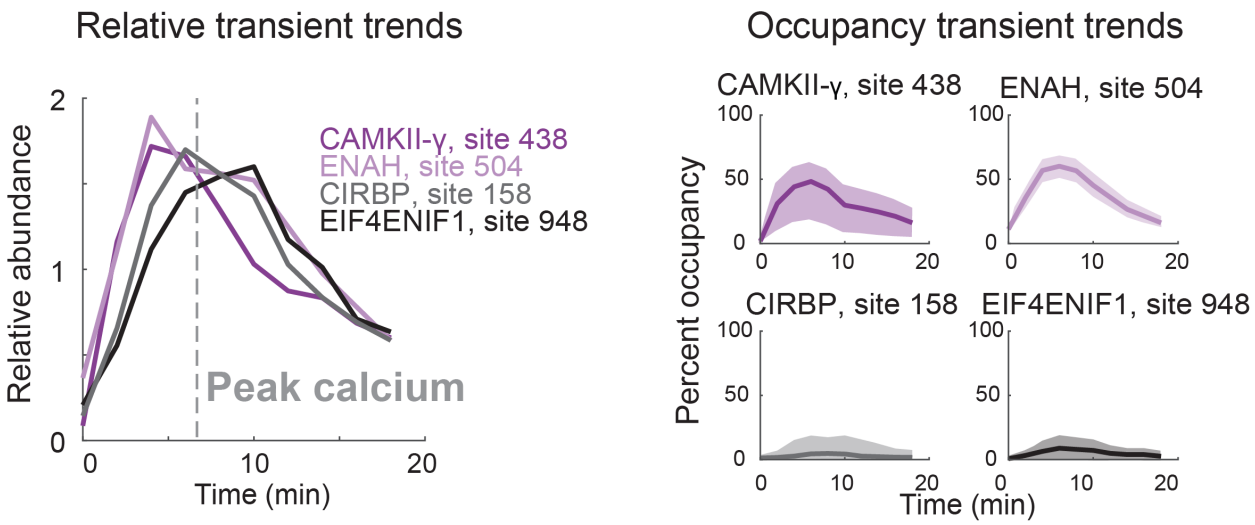

Figure 4.12 Relative and phospho-occupancy time series of proteins showing transient trends corresponding to the peak of calcium concentrations. Shaded area represents the 95% confidence intervals. While the relative trends are essentially identical, the change in percent occupancy of the phosphosites are very different.

A prominent example is CaMKII-γ, which is important for egg activation in the mouse [265], though the role of its phosphorylation is unclear. Three additional proteins with phosphosites that correlate with the calcium wave are shown (Figure 4.12). While the relative changes are nearly overlapping, the stoichiometries are very different. The two translational regulators (CIRBP, EIF4ENIF1) show phosphosites that change on a small fraction (<2%) of the proteins. In contrast, over half of the molecules of the gamma subunit of CaMKII and ENAH (an Ena/Vasp actin binding protein) are phosphorylated, supporting the hypothesis that the increasing phosphorylation class is related to calcium signaling and the remodeling of cortical actin. A possible explanation for the low stoichiometry of the changes on CIRBP

and EIF4ENIF1 is that this site of modification plays no significant role in regulation; perhaps these sites are modified through promiscuous activity of calcium-dependent kinases (e.g., CaMKII, PKC). Alternatively, the modification may be highly localized.

4.5.9 Absolute changes of protein phosphorylation, degradation, and release

The data allow comparison of the absolute changes in the diverse yet connected processes occurring in parallel at fertilization (Table 4.1). We estimate the total change resulting from protein degradation as ~300 nM (~0.01% of the total protein mass), which is dwarfed by a nearly 50-fold higher loss (~1%) from protein release (roughly consistent with older estimates [16, 115]). The total change in phosphorylation, which occurs on a diverse set of proteins from the cell cycle machinery to calcium-sensitive kinases, is ~7,000nM. This number is an underestimate, as we capture only a fraction of the total dynamics. Nevertheless, this is an order of magnitude higher than the changes due to protein degradation. Unlike protein degradation, where there was a strong correlation between low abundance and instability (Figure 4.5), we see large changes in phospho-occupancy for proteins that span a 1000-fold range of concentration (Figure 4.10B, D). This is striking evidence that phosphorylation has the capacity to change the activity of many abundant proteins within minutes. Protein degradation, even with an active E3 ligase like the APC/C, can only work so quickly on a small number of low abundance proteins, which seem to be core to cell cycle progression.

Table 4.1 Measured absolute changes after egg activation

Class of dynamic	Absolute change, nM	Fraction of proteome, %
Secretion	13,000	~1
Protein degradation	300	0.01
Phosphorylation	6,800	0.3
Decreased	4,200	—
Increased	2,600	—

4.6 Discussion

We measured the absolute protein and phosphosite dynamics at fine time resolution following fertilization. The data reveal a small number of degraded substrates and substantial protein dephosphorylation related to the cell cycle. We also detect release of protein related to the slow block to polyspermy and parallel increases in phosphorylation, which may be linked to the calcium wave. To aid in interpreting the data, we developed a new method for estimating phosphosite stoichiometry using multiplexed MS.

The phosphorylation data are compatible with a recent cell cycle model, which describes a quantitative model for the cell cycle oscillator in the *Xenopus* egg [266]. Earlier experiments demonstrated the oscillation [25, 80], but a satisfactory explanation of how it oscillated depended on several features, including a time delay (about 15min) between activation of Cdk1 and activation of APC/C and then a very small delay in the destruction of targets once APC/C is activated (<2min). Our measurements are strikingly consistent with fast Cdk1 inhibition, as we see immediate commencement of dephosphorylation at hundreds of proline-directed sites with no obvious delay. A discrepancy is that we see a small delay (~2-4 min) in the degradation of APC/C substrates like CyclinB1/2, which is presumably required to inactivate Cdk1. The rates of these early events may be underestimated if there is sufficient metachrony of the cell cycle in the large frog egg [99, 238], but the ordering of the observations discussed here should be preserved.

The substrate specificity of the APC/C is thought to be conferred by the cofactors Cdh1 and Cdc20. We show that all detected APC/C^{Cdh1} substrates are indeed stable in the frog egg, where Cdh1 is not expressed. The inhibition of APC/C^{Cdh1} promotes S-phase entry [267] in the conventional cell cycle; the absence of Cdh1 and resulting stability of its substrates may assist in the bypassing of the G1 phase to S phase during the cleavage divisions in the early embryo.

All detected degradation targets (e.g., Cyclin B1/2, EMI2) are present at low abundance, suggesting that the amount of protein degradation is small overall. However, if there is a global increase in protein synthesis at equilibrium with protein degradation, we would underestimate the flux of degradation. Protein synthesis at egg activation is reported in flies and mice [116-118], but our data shows no evidence for a burst of synthesis of any of the 8,700 proteins. Our previous work also showed

that new proteins comprise a small fraction of the frog embryo even hours post fertilization [216]. In principle, 2μM of new protein could accumulate in ~20 mins (~7μM/hr, a rate measured later in development) [268]. If this occurred during the first cell cycle, we would expect to identify many of these proteins given our sensitivity limits (Appendix 1, Figure S4).

In addition to measuring the extent of protein degradation and synthesis following fertilization, our proteome-wide approach unveiled unexpected components of the slow block to polyspermy. Early studies established that protease activity is essential for this event; the inhibition of proteases leads to polyspermy [269]. Paradoxically, we found that eggs release multiple endogenous protease inhibitors in response to fertilization. Though protease inhibitors were recently found in the perivitelline space of *Xenopus* eggs [114], their function was not addressed. There are several possible functions for released inhibitors: 1) they could prevent activity of the proteases inside the cortical granules, 2) they might control the activity of the proteases post-release, 3) they may act to block the sperm acrosome reaction [270].

Our application of quantitative proteomics to fertilization highlighted several general considerations for interpreting MS data. We concluded that phosphorylation dynamics can confuse the analysis of protein trends, giving the appearance of change in stable proteins. Dynamics in any high stoichiometric modification (e.g., acetylation) could also cause this same ambiguity. Using the median of a protein's peptides can mitigate but not will not eliminate specious trends caused by dynamic modifications. On the phosphosite level, a source of ambiguity is that a *decreasing* trend of a singly-phosphorylated form results from an *increasing* multi-phosphorylated form (Figure 4.9). This general possibility must be considered when interpreting the direction of a given phosphosite trend. This phenomenon was discussed previously [150]; our data provides clear examples.

4.7 Methods

4.7.1 Egg activation procedure

Female *X. laevis* were induced with 700 U HCG. After 14 hrs, eggs were harvested, washed with 1× MMR (Marc's Modified Ringers), de-jellied with cysteine (2% w/v, pH 8.0), and then kept in dishes with agar beds to avoid the eggs sticking to the dish. Eggs were placed in open-faced gel box with a 3% agar bed in 0.1X MMR and electro-shocked by applying an electric field of ~3 Volts/cm for one second. Thirty

eggs were collected every two minutes until 18 minutes. The excess media was removed and samples were flash frozen in liquid nitrogen. For the phosphatase treatment experiment, the 12 and 16 min time points were replaced with phosphatase-treated replicates of the 0 min and 18 min time points. Each biological replicate was performed on separate days with different frogs (termed EA, EAp2, EApR3, EAp5b, where EA stands for "Electro-Activation" and the 'p' denotes an experiment where phospho-enrichment occurred). The research with *X. laevis* was performed under the oversight of the Harvard Medical Area Institutional Animal Care and Use Committee.

4.7.2 Egg media time series

Egg media was collected at the similar time points as in Figure 4.1B following electro-activation, except with the 0 min time point taken in triplicate and 12 and 16 min time points skipped. The media was dried to concentrate the sample and processed for MS analysis. The experiment was performed as described above with the following exceptions: 1) Eggs were placed in 0.01X MMR, to reduce the salt concentrations after dry-down. 2) An additional ~10mM of calcium was added to the media to assist in any wound healing caused by the electroshock. 3) The electric field was doubled to ~6 Volts/cm to account for less conductive media, as eggs did not activate at 3 Volts/cm. 4) ~1,000 eggs were used instead of 300 to increase the signal of protein accumulating in the media. After electroshock, the media was kept well-mixed by gently rocking the gel-box on the long axis by hand (several methods were tested by mixing with dye). Each time point was taken by extracting 2 mL of media and flash freezing with liquid nitrogen. We replaced the volume of media removed for harvesting at each time point to keep overall volume constant throughout the experiment, and normalized afterward to account for the removal of mass in the media with each time point taken. We lyophilized the collected media samples to prevent unwanted proteolysis from the released proteases. The experiment was done in triplicate with three different clutches on the same day, and then processed in parallel as described in the MS sample preparation section below, except without the alkylation and protein precipitation steps.

4.7.3 General sample preparation for mass-spectrometry

Samples were prepared essentially as previously described [154, 217]. Thirty eggs per time point were snap frozen and lysed with 250 mM sucrose, 1% NP40 substitute (Sigma), 5mM EDTA (pH 7.2), 1 Roche complete mini tablet (EDTA-free), 20 mM HEPES (pH 7.2), 10 μM Combretastatin 4A, and 10 μM

Cyochalasin D while frozen [154]. All buffers were made with HPLC water. Eggs were lysed by pipetting up and down forty times with a 200 µL pipette tip, vortexed at maximum speed for 10 seconds, incubated on ice for 10 minutes, and again vortexed for 10 seconds. Lysates were clarified by centrifugation at 4,000 RCF at 4°C for 4 minutes in a tabletop centrifuge. The cytoplasmic and lipid layers were mixed by gentle flicking and removed from the pelleted yolk. To the lysate, HEPES (pH 7.2) was added to 100 mM, and SDS was added to 2% (w/v) to provide additional buffering capacity and to denature the sample. The samples were reduced with 5 mM DTT for 20 minutes at 60°C. Cysteines were alkylated with 15 mM NEM for 20 minutes at room temperature (RT). Excess NEM was reacted with an additional 5 mM DTT at RT. Proteins were isolated by methanol/chloroform precipitation [196]. The protein pellet was resuspended (~5 mg/mL) in 6 M guanidine HCl and 50mM EPPS, pH 8.5 with gentle pipetting and heated to 60°C for five minutes. For approximately 100-400 µg protein per condition, the sample was diluted to 2 M guanidine with 5mM EPPS, pH 8.5 and digested with LysC (Wako Chemicals) at 20ng/µL at RT for 14 hours. Next, we diluted guanidine HCl to 0.5 M with 5mM EPPS, pH 8.5 and digested further additional LysC at 20 ng/µL for ~15 minutes at RT, then added 10 ng/µL of sequencing grade Trypsin (Roche) and co-incubated at 37°C for 8 hours in an incubator. For each sample ~100-400 µg of peptides were dried down in a SpeedVac and resuspended with 100-150 µL of 500 mM EPPS, pH 8.0 respectively. If sample was not at pH ~8.0, an additional ~25-50mM of HCl was added. For labeling, we added 15-50 µL of TMT stock solution (0.2 mg/40 µL ACN) to each sample and incubated at RT for 2 hours (10 µg:1.5 µL peptide to TMT). Thereafter, we quenched the reaction with 10 mM Hydroxylamine for 15 minutes. All conditions were combined, acidified by addition of phosphoric acid to 5%, and were clarified by spinning at 21K RCF for 20 minutes. Samples were subjected to C18 solid-phase extraction (50mg, SPE) (SepPak, Waters) to desalt and isolate peptides. To reduce sample complexity, peptides were resuspended in a 10 mM sodium carbonate buffer (pH 8.0), then fractionated by medium pH reverse-phase HPLC (Zorbax 300Extend-C18, 4.6 X 250 mm column, Agilent) using an acetonitrile gradient from 5% - 39%. With a flow rate of 0.8 mL/min, fractions were collected into a 96 well-plate every 38 seconds, and then pooled into 24 fractions by combining alternating wells from each column of the plate. Each fraction was dried and resuspended in 20 µL of 1% phosphoric acid. Peptides from each fraction were desalted and extracted

once more with reverse-phase purification, resuspended in 10 µL 1% formic acid. ~4 µL per fraction were analyzed by LC-MS.

4.7.4 Phospho-peptide enrichment

We chose to multiplex peptides before the enrichment column to improve the data quality. There is a tradeoff of decreased yield and therefore depth, as it is not economical to label more than a few milligrams of material. A typical yield of 50-80 µg of peptides eluted from the column, with a median specificity of ~80%. We used 2.5-4 mgs of TMT-labeled peptides per replicate. These were desalted and dried down into a pellet, which was resuspended in Binding Buffer (85% 11.1M Lactic Acid, 50% acetonitrile, pH ~1.5-2.0) at 10 mg/mL and spun for 30 minutes at 21K RCF at RT. 5 µm Titanium Dioxide microspheres (GL Sciences 5020-75000) were added at 8:1 peptide to bead ratio (by weight) and washed 3X with 1 mL of Binding Buffer. Peptides were added to dried Titanium Dioxide beads and incubated for 1 hr at RT with constant agitation. PTFE membrane filters (0.2 µm low-binding hydrophilic, Millipore UFC30LG25) were washed three times with Binding Buffer, and then used to filter the TiO_2 microspheres, spinning at no more than 10K rcf until dryness. The first flow-through was used for a second serial enrichment as described below (A and B enrichments). The final flow-through was saved and used as a protein replicate. The A and B enrichments were each washed twice with two column-volumes of 50% acetonitrile, 0.1% Trifluoroacetic Acid and then twice with 25% acetonitrile, 0.1% Trifluoroacetic Acid. Phosphopeptides were eluted from the beads 3 times with 1 column volume of Elution Buffer (200mM of KH_2PO_4, pH 10) each. Beads were incubated in Elution Buffer for 5 minutes with occasional agitation before spinning the beads to dryness; the 3 eluates were combined. The A and B enrichments remained separated. All samples were acidified with 10% formic acid and 0.5% TFA and each desalted on C18 SepPak. We found if more than ~15 mg of beads were used in one filter, the specificity of the enrichment suffered as the column clogs and causes the washes and elution to be less efficient. Therefore, for the second two experiments, we split the sample onto 10 columns processed in parallel for the A enrichment. Since the peptides are pre-labeled and already multiplexed, the measured peptide ratios are unaffected by splitting the samples. For the B enrichment, the washing steps were performed in a single Eppendorf tube and the beads were washed by pelleting (a practical step to reduce the use of filter tubes). The A and B enrichments and the flow-through were desalted again. A typical yield of 50-80 µg of peptides

eluted from the column for the A and B enrichments combined, with a median specificity of ~80% between biological replicates. The flow-through sample was fractionated with HPLC as in the main text. For the enriched samples, peptides were subjected to an offline Strong Cation Exchange fractionation adapted from Dephoure and Gygi [271]. Peptides were re-supsended in Buffer A (7mM KH2PO4, pH 2.65 and 30% ACN) and loaded onto 20um, 30 angstrom PolySulfoethyl A beads (Nest Group, BMSE2003). Salt cuts were made by mixing appropriate ratios of Buffer A with Buffer B (7mM KH2PO4, pH 2.65, 30% ACN, and 350 mM KCl) for Enrichment A: 0, 5, 10, 20, 40, 60, and 350mM, for Enrichment B: 0, 5, 10, 40, 350. Fractions were desalted on STAGE tips and analyzed by mass spectrometry.

4.7.5 LC-MS analysis

Analysis performed essentially as previously described, with the spectra mapped to the PHROG reference database [154, 217]. The main modification was the use of 5-notch MS3 on the protein-level and 3-notch MS3 on the phospho-level [146]. Phosphorylation searches were performed with differential modification of +79.9663304104 on Serine and Threonine. Tyrosine was excluded as we found the majority of the identified tyrosine sites were erroneous identifications. Phosphosites were localized with in-house software based on Ascore [144]. Peptides with multiple phosphosites are reported as "composite sites" and noted with a ";" delimiting each modified residue. All searches (for the protein replicates, phospho replicates, media, ubiquitin pull downs) were mapped to a single Protein Map and subjected to one Linear Discriminator/Protein Sieve Analysis to control the false discovery rate.

4.7.6 Classification of released proteins and novel APC/C substrates

We classified a protein as released (rather than degraded) either by direct experimental evidence, annotation, or evidence in the literature. In most cases, a protein that decreased in the egg could be clearly detected as increasing in abundance in the media. There were several cases where a protein decreased in the egg that we did not detect in the media, but these were clearly annotated as secreted proteins or were of a similar class as proteins for which we had direct evidence. For example, we infer the Exocyst gene family are released into the media because vesicle trafficking and components of the exocytosis machinery were detected recently in the perivitelline space in frogs [114]. To be classified as a novel APC substrate, we imposed the following criteria: the protein passes a 1% FDR threshold for decreasing abundance, had no evidence (direct or from literature) of release from the egg, no evidence of

a specious trend from phosphorylation (this was established by looking for reciprocal phosphorylations, or more generally, whether all the detected peptides behaved similarly), and was ubiquitinated.

4.7.7 Phospho-stoichiometry regression-based algorithm

The full principle and methodology is included in Chapter 3.

4.7.8 False discovery rate (FDR) calculation

The approach to calculate the FDR is visualized in Appendix 1 Figure S2. In brief, the experimental distances are measured to a 10-dimensional, idealized degradation vector and then randomized iteratively. In practice, the data were measured in up to four replicates with ten conditions each. To account for this, all protein trends were represented as 40-dimensional vectors. The idealized vector was repeated four times to match this. Proteins were ranked according to their distance from the repeated idealized vector by Cosine distance. While non-linear degradation trends exist in the data, they are much closer to the idealized vector than the flat trends. Thus, they are still ranked highly, with a small underestimation in distance. If a protein was missing from a given replicate, the mean trend of that replicate (which is essentially flat) was substituted for the missing data in the 40-dimensional experimental vector. The data were then randomized 10,000 times with the cosine distance measured each iteration. The cumulative distributions of distances for all proteins (experimental) and pooled randomized trends were used to calculate the FDR at a given cosine distance as the idealized vector, or (Randomized/(Experimental + Randomized)) *100. Since the randomized values are pooled we apply one cutoff to the entire dataset, so multiple hypothesis correction is not required. The FDR is calculated for each Protein ID, and then combined by gene symbol afterward.

4.7.9 EM imaging time series

Eggs were prepared and electro-activated as described in the main text. Time points collected were 0 sec, 8 sec, 15 sec, then at 15 second intervals until 105 seconds, then 2 min, 4 min, and 10 min. 8 sec time point was the earliest condition we were able to collect. 3-5 eggs were collected in 200 µL of volume with a cut pipette tip. The eggs and buffer were added to 2X fixative solution (see below) and quickly mixed to final 1X concentration. We are confident the fixation occurred quickly at the cortex, where we focused our imaging. Eggs were fixed with 3% gluteraldehyde in 0.1X MMR for 30 minutes. The fixed eggs were cut and the animal and vegetal poles separated and placed into fix in separate tubes

to maintain orientation in later steps. In most cases, halving the eggs led to the loss of the vitelline membrane, and the extracellular morphology is therefore not present in the images. Eggs were rinsed 2 times in 0.1X MMR and 3 times in 0.05M cacoydylate buffer pH 7.0. They were then post-fixed with 1% osmium in 0.8% $K_3Fe(CN)_6$ in cacodylate buffer on ice in a fume hood for 1 hour [272]. They were then washed 3 times with cacodylate buffer and 2 times in distilled water prior to staining overnight in 1% aqueous uranyl acetate in the dark. Samples were washed in distilled water prior to dehydration in an ethanol series using the progressive lowering of temperature method followed by exchanges of 100% ethanol and propylene oxide at room temperature [273]. Infiltration was done in 2:1, 1:2 mixtures of propylene oxide epon araldite for 1 hour each step followed by 100% epon araldite for 2 hours. Samples were embedded and polymerized at 65°C for 48 hours. The animal poles of the eggs were thin-sectioned on a Reichert Ultracut S microtome and the sections viewed on a Technai G2 Spirit Bio TWIN Transmission Electron Microscope and imaged with an AMT 2k CCD camera.

4.7.10 Data normalization

Every protein or phosphosite was required to have a summed TMT Signal to Noise value across the ten TMT channels of greater than or equal to 386 on the Thermo Fusion or Lumos, or 189 for the Orbitrap Elite. Each condition was then normalized to the condition median to correct for pipetting errors. Each trend was then normalized to its mean. Replicates trends were averaged. For proteins and phosphosites that appear in the phosphatase-treated experiment, the 12 min and 16 min time points are missing. In this case, these time points are replaced by averaging the two flanking conditions. For the supernatant experiments, median normalization per time point is not appropriate, as the time series are not expected to be flat. We therefore estimated an accumulation curve and normalized each trend to this curve. Similarly, median normalization is not appropriate for the phosphatase treated samples on the phospho-level as the signal is expected to be lower than the rest for these conditions. We normalize by the mean of the averages of all untreated conditions. This correction is unnecessary on the protein level for the phosphatase-treated conditions, as the majority of proteins are not affected by the treatment.

4.7.11 Establishment of absolute changes of protein and phosphorylation level

Proteins were matched by gene symbol to our previous dataset of protein abundance in the *Xenopus* egg. For the protein loss, the abundance of candidates that passed the 1% FDR threshold were

multiplied by the fold change from the first condition, which is the unactivated egg that is equivalent to the sample used in the previous study. Securin was not seen in our previous absolute abundance data set. We estimated its abundance from its stoichiometric binding partner, Separase. The rates of degradation for the known cell cycle targets were calculated with a linear fit using Matlab "polyfit" function. Confidence intervals were calculated using File ID: 39126 on The MathWorks File Exchange. For obtaining absolute values of phospho site, the same procedure was followed for proteins for which we could estimate phospho-stoichiometry. Here, the abundance was multiplied by the percent occupancy at each time point.

4.7.12 Estimating maximal degradation rates of the APC/Proteasome

The overall rate of ubiquitination-dependent degradation is dependent on the rates of ubiquitination, deubiquitination, and proteasome reaction. It is difficult to estimate the K_{cat} for the APC or deubiquitination activity. However, we can approximate the rate of the proteasome reaction. We estimate $1/ K_{degradation}$ as ~80 seconds, given that $1/ K_{degradation} = N \times (1/ ([Proteasome] \times K_{on}) + 1/K_{cat}$. N, number of times a substrate must bind before degradation occurs, is ~6; K_{on} was measured at ~10^5 M^{-1} s^{-1}; Kcat is ~20 seconds per substrate bound [233]. The proteasome concentration is ~1μM in the frog egg [154]. For $[Substrate_{Ubiq}]$, though we cannot infer from rates 1 and 2, we can estimate that 1% or at most 10% of the substrate concentration is modified at steady state [274]. The total concentration of known APC substrates is ~400nM, therefore $[Substrate_{Ubiq}]$ = 4-40 nM. The rate of degradation is ~3-30nM /min (or $[Substrate_{Ubiq}] / K_{degradation}$). Our estimated rate of degradation rate from the data is ~14 nM/min, or 150 pg/min (Figure 4.2), and the original concentration estimates contain ~2 fold error [154]. With these parameters, an estimate of the upper bound of the degradation rate overlaps with the measured range

4.7.13 K-means clustering and multi-site artifact correction

K-means clustering was implemented in MATLAB using the 'kmeans' function with K = 56, which was chosen because we found it gave stable clusters. The analysis was performed on the mean normalized, replicate-averaged data for both protein and phosphorylation data. The 56 clusters were manually collapsed into 7 clusters for phosphorylation and 4 clusters for the protein that qualitatively represented the data. The last cluster was reserved for sites that were consistently clustered individually, unstably, or into clusters with less than 10 protein/sites. For the purposes of display, these are not included in Figure 4.1 but are provided in Figure S8 and Figure S9 and were included in any statistical

analyses. These cases constitute <1% of the data. The modified Waterfall plots in Figure 4.1 are normalized to the first time point and ordered by their cluster and then then ascending within each cluster by the value of the last time point. For the phosphorylation trends, we corrected for possibly misleading trends of single sites with corresponding composite sites. A double site and corresponding single site can often show the same reciprocal behavior as shown between unmodified and singly modified species (see Figure 4.9). Therefore, a trend that appears as phosphorylation is really dephosphorylation of the double form. To avoid being misled by this trend, we systematically identified single sites with reciprocal trends to their corresponding composite sites (as evaluated by a Spearman correlation of <-0.5). These sites were removed from the analysis, and the most-modified composite site was kept, as this is the form most likely to give the true direction of the sites. This correction totaled ~50 dynamic sites. This phenomenon has been previously discussed in the literature [150].

4.7.14 Motif and Gene set enrichment analysis

Using localized sites with possible artifacts from reciprocal trends removed, we evaluated the enrichment of various motifs -S#/T#-P- or -RxxS- using regular expression searches in each cluster discussed above and in Figure 4.1E. A p-value was calculated with the Fisher's exact test [275]. Gene Symbols classified by their artifact corrected K-means clusters were used for gene set enrichment analysis performed with WebGestalt [187]. Enrichment was assessed with the hypergeometric test, and multiple-hypothesis testing correction was performed with the Benjamini-Hochberg procedure [276].

4.8 Acknowledgements

This work was supported by NIH grants HD091846, HD073104, GM103785 to MWK, GM39565 to TJM, and Burroughs Wellcome Fund and Mallinckrodt Award to AMK. We thank Chris Rose and Joao Paulo for help with mass spectrometers and the Gygi lab computational team for bioinformatics support, Ying Lu, Tao Wu, Tim Mitchison, and Angela DePace for helpful discussion, Clarissa Rhodes, and Becky Ward for critical comments on the manuscript. We thank the PRIDE team for proteomic data distribution (accession number PXD006639) [277].

Chapter 5:

Discussion

5.1 General overview: main findings and advancements

The events in early embryonic development are regulated primarily by the biochemical activity of proteins. In the minutes after fertilization, phosphorylation and protein degradation are two of the most prominent regulatory mechanisms for coordinating the rapid release of the meiotic arrest and the blocks to polyspermy. While many individual components have been identified over the past several decades, the extent of protein degradation required to drive the cell cycle is unknown. Similarly, the identities and function of thousands of modified proteins are also still not known. This is largely because comprehensive and quantitative analysis of such biochemical regulation is currently difficult.

We performed experiments to measure protein and phosphosite dynamics following fertilization using mass spectrometry in *Xenopus* eggs. To gain more insight into the function of these trends, we developed new techniques to estimate the absolute concentration of proteins in the egg and assess the site occupancy of phosphorylated residues over time during egg activation. Though applied in the egg, these approaches are widely applicable to studies of protein biochemistry in diverse systems.

The data reveal that from the common trigger of fertilization, two parallel and contrasting biochemical programs are initiated (Figure 5.1). A small number of low abundance substrates are degraded, but this promotes the substantial protein dephosphorylation of many proteins across a broad range of concentrations related to the cell cycle exit. In contrast, increased concentration of many phosphorylation sites likely promotes the cortical response of the egg, including the release of many high abundance proteins as part of the blocks to polyspermy.

There are many individual molecular findings in the data as well. Specifically, we demonstrate a limited number of new substrates of the Cdc20-associated Anaphase Promoting Complex, along with the striking stability of other known Cdh1 targets in the simplified embryonic cell cycle. We identify new components of the block to polyspermy, hundreds of putative Cdk1 targets and their dynamics, as well as calcium-sensitive responses with differential stoichiometry changes. We also present evidence that the reassembly of the nuclear pore complex is directed by post-translational mechanisms, specifically the differential dephosphorylation of sub-complexes that conform to their sequence of assembly.

As opposed to other developmental stages, the absolute measurements of proteins dynamics during egg activation highlight the power of protein biochemistry to autonomously set the state of a cell and direct major transitions with minimal input from the genome.

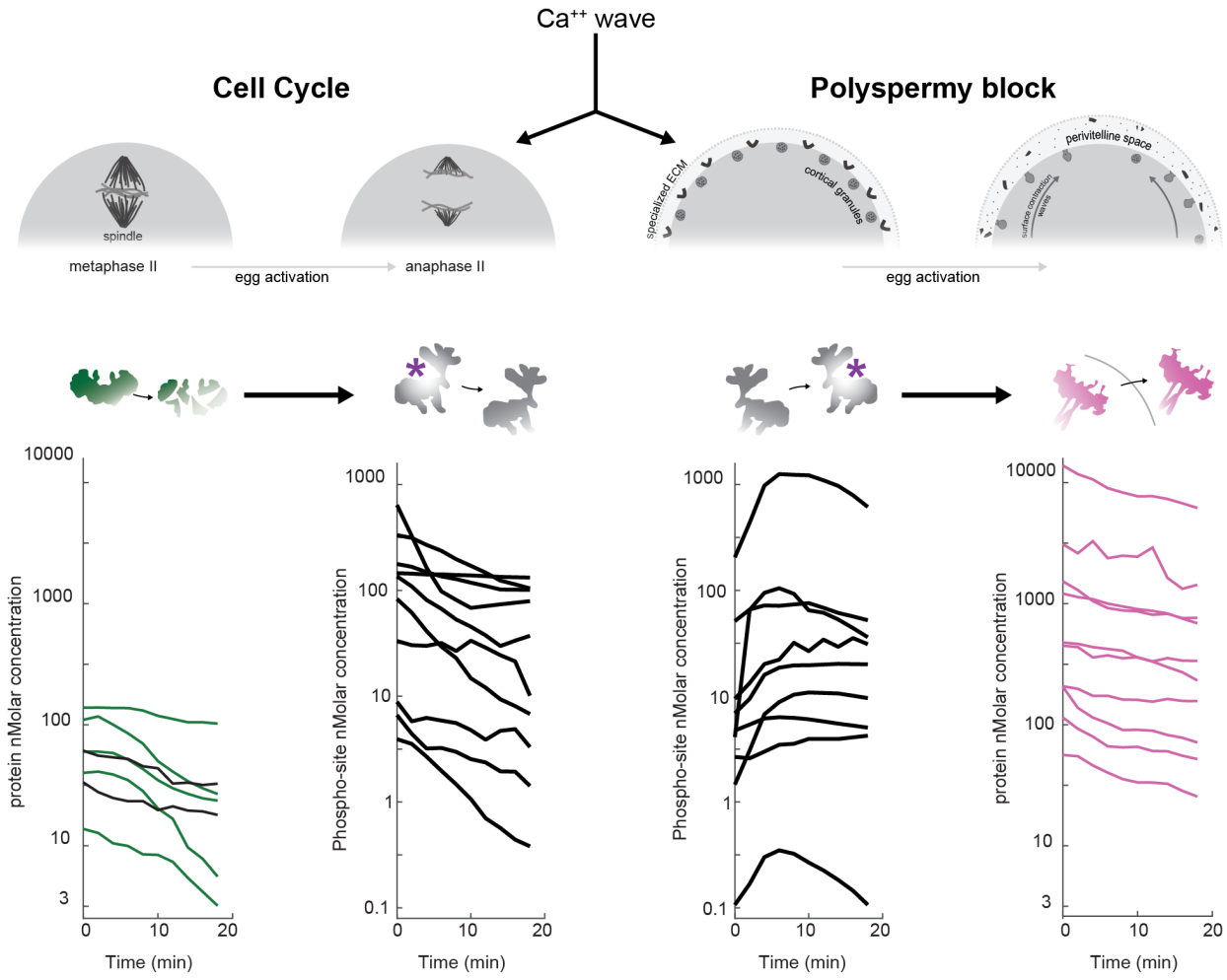

Figure 5.1 Graphical summary of the absolute measurements enabled by the methodologies developed in this dissertation. A pair of parallel biochemical programs reflects the differential regulation of the two major processes occurring during egg activation: the loss of a small amount of low abundance proteins promotes the large-scale dephosphorylation of sites. Contrastingly, increased phosphorylation, likely downstream of the calcium wave, promotes changes at the cortex in the egg. These include the secretion of several highly abundant proteins. Increasing and decreasing phosphorylation each occur over a wide range of abundances.

5.2 Insights into the biochemical regulation of meiotic exit and fertilization

5.2.1 Examining features of the biochemical switch at meiotic exit

The comprehensive and quantitative phosphorylation and protein dynamics data gathered in this dissertation presents an opportunity to examine models of the cell cycle that were built on measuring only a few components due to technological limitations. One such model predicts that proper cell cycle oscillations require a very small delay in the destruction of targets once APC/C is activated (<2min) [266]. In agreement with these predictions, we indeed observe the dephosphorylation of many proline-directed sites with no obvious delay upon activation.

A discrepancy is that we see a small delay (~2-4 min) in the degradation of APC/C substrates like Cyclin B1/2. This means for some putative Cdk1 sites, their dephosphorylation occurs before Cdk1 is inhibited by Cyclin B degradation. Others have noted related phenomena in their own experiments (personal communication, Ferrell lab). One explanation is that ubiquitination of Cyclin B is sufficient to partially inhibit Cdk1 activity before Cyclin B is degraded. Monoubiquitination of Cyclin B is sufficient in some cases to trigger degradation [274]; it is possible this type of modification is inhibitory as well. A clear expectation is that some reduction in Cdk1 activity should be measurable in the presence of ubiquitinated Cyclin B even if degradation is blocked with proteasome inhibitors (though proteasome inhibitors are sufficient to block the full completion of M-phase exit [32]).

Another possible source of an early inhibition of Cdk1 before Cyclin B degradation is via rapid phosphorylation by Wee1. There is evidence that the “equilibrium” between Cdk1 and Wee1 activities is shifted to control the length of the first and second cell cycles [278]. This hypothesis would predict that the earliest substrate dephosphorylation is triggered by partial inhibition of Cdk1, but full meiotic exit requires Cyclin B degradation. Phosphatase regulation may also contribute to the early loss of Cdk1 phosphorylation sites. The activity of the phosphatases PP2A and Calcineurin is required to exit meiosis [42, 43]. Their increased activity may drive the reaction to net dephosphorylation for some sites early on.

One consideration is that many of the proline-directed sites that are dephosphorylated early could be MAPK rather than Cdk1 substrates. The differences could be determined by assessing if the early sites are sensitive to treatment with the Cyclin-dependent kinase inhibitor (e.g., roscovatin [279]).

5.2.2 Enumerating the role of APC/C cofactors in cell cycle regulation

The specificity of the APC/C is thought to be conferred by the cofactors Cdc20 and Cdh1. We found that all detected APC/C^{Cdh1} substrates are stable in the frog egg, while all the detected Cdc20 substrates are clearly degraded. This finding strongly demonstrates the *in vivo* specificity of Cdc20.

The stability of the Cdh1 substrates (e.g., PLK1) is consistent with previous work [62], including experiments that showed egg extracts are not able to degrade APC/C^{Cdh1} targets without supplementing the extract with additional Cdh1 protein [74]. However, the role of Cdh1 *in vivo* was much more ambiguous. A set of contradictory papers initially reported Cdh1 as not present in the egg [72] or present at very low levels until after the midblastula transition [280]. Another group argued that Cdh1 was not just expressed but required for the early divisions in the frog embryo [73].

Our data helps to resolve this dispute; Cdh1 is not detected in the egg in our proteomics data. Even if it is present at low levels, it is not functional, as our measurements show all detected targets of APC/C^{Cdh1} are strikingly stable during meiotic exit. From this data and previous measurements, it is reasonable to expect that APC/C^{Cdh1} remains inactive until after the cleavage division, though Cdh1 may still be required in another unknown role.

The lack of APC/C^{Cdh1} activity may actually encourage the abbreviated nature of the embryonic cell cycle. In the somatic cells, inhibition of APC/C^{Cdh1} assists in the exit from G1 and entry into S-phase [267]. If APC/C^{Cdh1} is indeed inactive during the cleavage divisions, the resulting stability of its substrates may assist in the bypass of G1 phase to S phase in the early embryo (Figure 1.7).

Accordingly, Cdh1 may help reintroduce the gap phases during the midblasula/maternal to zygotic transition. The emergence of slower cell cycles at the midblastula transition is correlated temporally with zygotic synthesis many proteins, including Cdh1. Just as interphase extracts are competent to degrade Cdh1 substrates when concentrations are increased artificially [74], the endogenous synthesis of Cdh1 should be sufficient to drive APC/C^{Cdh1} activity. The degradation of APC/C^{Cdh1} targets perhaps assists in imposing gap phases, regulated by increasing Cdh1 levels.

The mouse embryo offers an apt comparison, as Cdh1 is already expressed in the egg. All the known Cdh1 targets are degraded during murine meiotic exit [281]. The early divisions in the mouse embryo proceed at a leisurely rate more similar to that of later cell divisions in the frog. Unlike externally

laid eggs, the primary function of the early murine cell divisions is to build the extra-embryonic tissues that support the embryo throughout development. There is a period of more rapid divisions that occurs in the pregastrula mouse embryo, which more closely resemble the fast cleavage divisions of the frog, fish, and flies. The divisions last two hours, with short gap phases (<30 minutes) [282]. If Cdh1 assists in shortening gap phases as suggested, one expectation would be that Cdh1 is partially degraded or otherwise inactivated to stabilize APC/C^{Cdh1} substrates during this period.

5.2.3 Endogenous degradation rates and consequences of spatial asynchrony

We were able to report the endogenous, absolute degradation rates for the cell cycle targets (Cyclin B1/2, Securin), which have been difficult to obtain from *in vivo* settings. It is important to note that the rates of these early events may be underestimated if there is sufficient metachrony of the cell cycle across the 1.2 millimeter frog egg. Indeed, there are spatial differences in the cytoplasm of the frog egg [99, 238]. An extreme example of spatial heterogeneity occurs in flies, where Cyclin B is degraded locally around the DNA, but is incompletely destroyed elsewhere in the cytoplasm [283, 284]. The same phenomenon of highly localized degradation does not appear to occur in frogs. Still, the spatial asynchrony of the egg's cell cycle state may retard the rates reported in this thesis (Figure 4.2), but the overall ordering of the observations discussed here should be preserved. Nevertheless, it may be necessary to confirm the rates by measuring protein degradation only the top half of the egg. This experiment would report degradation rates most proximal to the spindle more accurately. The importance of the chemical waves in the cell cycle should not be underestimated. Calcium signaling and Cdk1 loss proceed as trigger waves that coordinate events in the frog eggs faster than diffusion alone [99, 236]

5.2.4 Novel components of the slow block to polyspermy

In addition to measuring the extent of protein degradation and synthesis following fertilization, our proteome-wide approach unveiled unexpected components of the slow block to polyspermy. Studies in the late 1980s established that protease activity is essential to achieve euploidy after fertilization. This was shown by treating eggs with protease inhibitors [269]. Paradoxically, we found that eggs release multiple endogenous protease inhibitors along with the cortical granule proteases in response to fertilization. Protease inhibitors were recently found in the perivitelline space of *Xenopus* eggs in later developmental stages [114], though their function was not clear. Our data show they are released during

egg activation with similar kinetics as other known secreted factors (e.g., lectins, proteases). There are several possible functions for released inhibitors:

1) They could prevent activity of the proteases inside the cortical granules. The proteases are partitioned from the rest of the cytoplasm in vesicles, possibly in high concentration. It may be important to keep the activity low to avoid proteolysis of the cortical granule contents.

2) They may form a simple inhibitory circuit with the proteases to control or limit the extracellular proteolytic activity. It is general wisdom that embryos are sensitive to over-digestion during protocols to enzymatically remove the vitelline membrane; the danger may be the same from the release of endogenous proteases as well. It may be of particular importance to preserve some structure of the egg extracellular matrix. For example, in the mouse embryo the Zona Pellucida (ZP) is not shed until after implantation in the uterine wall [285]. The ZP shedding is proteolytic [286]. The inhibitors may prevent over-digestion of the ZP or other structures. The ZP does not shed in the same way in amphibians, but the vitelline membrane structure is still preserved until hatching.

3) They may act to block the sperm acrosome reaction [270]. Sperm binding and paternal membrane fusion requires proteolysis. It may be that the protease inhibitors are specifically preventing the efficiency of the acrosome reaction. If this were the case, fertilization of one egg could reduce fertilization rates of nearby eggs in the clutch. The proteins may diffuse too slowly through the jelly coat for this to be a factor, or the dilution effects are sufficient such that there is no effect. However, it may be relevant for *in vitro* fertilization procedures in fertility clinics, where multiple eggs are often fertilized together.

The fact that the released proteins diffused so readily through perivitelline membrane is somewhat unexpected, as the dilution of factors out of the perivitelline space may decrease their potency. If the mechanism of these molecules is to inhibit the extraneous sperm, the diffusion away from the egg is less surprising (and may be advantageous).

5.2.5 Confronting conflicting data on post-fertilization protein synthesis

Protein synthesis is thought to globally increase in response to egg activation in flies, sea urchins, and mice [116-119]. In contrast, our data shows no evidence for an increase in level of any individual protein. If there were indeed large scale protein synthesis, one consequence is the possible

misinterpretation of the degradation data. The detected degradation targets (e.g., Cyclin B1/2, EMI2) are few in number and present at low abundance. Here we can be confident that the total change in protein level due to protein degradation is small overall. However, if there had been a global increase in protein synthesis balanced by protein degradation, we would massively underestimate the flux of degradation.

If such a global increase were occurring, some individual proteins would likely increase in abundance; it is possible these proteins cannot accumulate sufficiently to be detected by mass spectrometry, particularly those at low abundance. However, 2μM of new protein could accumulate in principle over ~20 minutes given a rate of ~7μM/hr (which has been measured later in development [268]). Assuming a similar rate occurred during the early cell cycle, we would actually expect to identify many of these proteins given our sensitivity limits (Appendix 1, Figure S4). Moreover, our previous work that measured protein levels well beyond egg activation showed that new proteins comprise a small fraction of the frog embryo even hours post fertilization [216]. Our data suggests that overall, protein synthesis is not occurring at high levels during meiotic exit in the frog egg, and only occurs at low levels at later stages.

The impression that eggs are metabolically activated by fertilization traces largely back to the sea urchin. It is important to note that sea urchin eggs are fertilized in interphase [8], having already completed meiosis, as opposed to the meiosis II arrest in frogs. Identifying the conserved response to "egg activation" requires considering the broader developmental context.

A similar argument for considering the developmental stage applies to mouse eggs, where zygotic transcription is initiated shortly after fertilization. In this case, events that occur many hours into development in frog eggs occur shortly after or even in parallel with egg activation in mice. The protein synthesis in mouse fertilization may be reflective of the earlier initiation of the maternal to zygotic transition rather than a direct response to egg activation. Large increases in the occupancy of polysomal RNA have also been reported in *Drosophila* egg activation [116]. The eggs of flies and frogs are much more similar in cell cycle stage and developmental timing than mice. The discrepancy here is not so easily resolved. It is worth noting that egg activation can proceed independently of fertilization far more readily in insects [108, 287]. There is some possibility that this leads to some temporal asynchrony in fly eggs when studying the earliest steps after egg activation. In other words, parthenogenetically activated

eggs may occur at a higher frequency in flies. These eggs may have a “head start” and show signal for later events like protein synthesis.

Frogs, of course, may simply be an exception to a more general rule. The synthesis of new, zygotically-encoded proteins is an important and general step in all animal development, and may be especially critical for species with smaller eggs and more modest maternal dowries. However, a broad increase in protein synthesis may not be a conserved response to egg activation. Instead, it may be a more general early developmental event that shifts in time depending on the strategies of the particular organism.

5.3 Implications for mass spectrometry-based proteomics

5.3.1 Enabling proteomics without well-annotated genomes

Much of the results presented in this thesis involve interpreting aspects the function of protein biochemistry by estimating absolute concentrations of proteins and absolute rates of change. This would not have been possible without the early work of the project detailed in Chapter 2. The TMT-based methods were nascent at the time, but more fundamentally, it was not clear that large scale proteomics was possible in *Xenopus* without a well-annotated genome. Our lab developed a transcript-based reference set from RNA sequencing as a functional substitute.

A key insight in building this transcript reference (called PHROG) was the use of a BLAST search against known vertebrates to infer the correct reading frame of the *Xenopus* RNA sequences for *in silico* translation [154]. The *in silico* translations are then used to create the theoretical spectra used to identify proteins. While other methodologies existed to convert *de novo* RNA sequencing datasets for use in mass spectrometry, these approaches generated 6-frame translations of the transcripts rather than assessing the correct frame. This approach massively increased the search space and led to a significant reduction in the identification of peptide sequences. Using the reciprocal BLAST technique to set the correct frame, we were able to obtain one of the deepest proteomic datasets to date.

The reference set (PHROG) outperformed the genome in successful peptide identifications for several years. Proteomics offers a rigorous test of genome quality by comparing results from transcript

versus genome-derived references. Approaches of this type are often termed proteogenomics [288]. Beyond enabling our own experiments, this approach demonstrates the feasibility of performing high quality mass spectrometry in non-standard model organisms, which have so often been critical for studies of embryonic development (*Ascaris*, sea urchin, star fish, etc.). There are also many genomes that are highly repetitive in sequence, such as the malaria parasite *Plasmodium falciparum* [289]. These features make whole genomes difficult to assemble. Similar techniques to the ones used in our work could assist in bringing more reliable proteomics to such organisms.

5.3.2 Measuring absolute concentrations of proteins with mass spectrometry

The calculation of absolute rates in the egg illustrates the usefulness of developing proteomics methods in a well-studied biochemical system like the *Xenopus* egg. Over twenty years, many quantitative western blots were performed to calculate the concentration of proteins in the egg extracts. We were able to use this information to build a standard curve to estimate the concentration of all the detected proteins in the egg using the summed ion current of the peptides. Other organisms may not have such information available to them. A standardized mixture of proteins at known concentrations is available from Sigma, called the Universal Proteomics Standard Sets (UPS1 and UPS2). The reagent was originally produced as a troubleshooting tool. In principle, the standard could be adapted as a spike-in to regularly estimate absolute concentrations in proteomics samples more generally (personal communication, Martin Wühr and Matthew Sonnett). However, the reagent is too expensive for regular use. An alternative may be to use a spike-in of clarified *Xenopus* egg lysate, where the concentrations of ~11,000 proteins were estimated. A "*Xenopus* Standard" would be cheap to produce and contain proteins over a wide range of abundances. Importantly, the peptide sequences would be largely orthogonal to avoid confounding peptide identification in the target proteome.

5.3.3 Identifying artifacts caused by post-translational modifications

Our application of quantitative proteomics to fertilization highlighted several general considerations for interpreting MS data. We concluded that phosphorylation dynamics can confuse the analysis of protein trends, giving the appearance of change in stable proteins. This is due to conservation of forms (i.e., a decreasing phosphosite yields an increasing unmodified peptide). Dynamics in any high stoichiometric modification (e.g., acetylation) could also cause this same ambiguity.

Using the median of a protein's peptides can mitigate, but not eliminate, specious trends caused by dynamic modifications. Situations often arise where all of the detected peptides for a given protein are dynamically phosphorylated. Conversely, it is also common that only one peptide is detected. In our data, had we aggressively disregarded any data with only one peptide, considered the median fold change per protein, and taken the median of the biological replicates, we would have still erroneously identified 4 (instead of 7) proteins as significantly increasing when they were actually stable. While this error rate is small compare to the entire dataset, if we had been interested explicitly in synthesized proteins, our false discovery rate would have been nearly 100% without accounting for specious trends from phosphorylation.

On the phosphosite level, a source of ambiguity is that a *decreasing* trend of a singly-phosphorylated form results from an *increasing* multi-phosphorylated form (Figure 4.9). This general possibility must be considered when interpreting the direction of a given phosphosite trend. This phenomenon was discussed previously [150]; our data provides clear examples. In the future, it may be necessary to perform "top down" proteomics for post-translational studies (i.e., measuring intact proteins). However, there are significant technical problems, such as how to get site specific information.

5.3.4 Advancing the measurement of phospho-occupancy with proteomics

The utility of measuring phospho-occupancy is well-recognized [151, 152, 206]. However, the commonly applied approaches are not able to estimate occupancy for multi-phosphorylated peptides, stable sites, or sites measured with multiplexed proteomics. More fundamentally, these methods do not give statistical information (i.e., confidence intervals), which makes interpreting the estimates difficult. Our approach provides a unified analytical and experimental framework to address these limitations. Though we developed this approach to exploit the power of multiplexed proteomics, the advancements are compatible with other methodologies. For example, label-free phospho-proteomics typically have lower measurement precision, but are conducive to high-throughput formats [214]. The additional conditions, in principle, give increased statistical power with the use of the regression framework. The higher sensitivity of label-free and other techniques [175] could mitigate a limitation of this approach, which is the low efficiency of measuring both the unmodified and phosphorylated peptides. Other methods for calculating phospho-stoichiometry, such as motif-targeting proteomics [215], could also incorporate this framework.

5.3.5 Compatibility of occupancy approach with ubiquitination, acetylation, and new modifications

The phospho-calculation relies on the principle of conservation, which holds true for any modification. Therefore, the algorithm is compatible with any modification where the unmodified and modified forms are detected. If the modification is not dynamic, changes can be induced enzymatically or chemically. Two modifications of immediate interest are acetylation and ubiquitination. Acetylation may be more successful because of the generally higher stoichiometry ratios. Ubiquitination is often low occupancy, which will make the measurements more sensitive to noise. More importantly, ubiquitin chains are cleaved during the digestion step preceding proteomic analysis. While it may be possible to calculate ubiquitin stoichiometry, it will not be clear whether the site was mono- or polyubiquitinated.

Beyond the known modifications, it is possible to search for novel modifications in mass spectrometry data. This can be done by "scanning" with a wide precursor mass tolerance to screen for new modifications as detected by specific mass shifts [290]. Coupled with the occupancy calculation, it is possible to not just discover new modifications, but determine whether they exhibit significant dynamics in stoichiometry. This can be done simply by reanalyzing existing data, from this thesis or others. These approaches can increase the power of the proteomics for analysis of posttranslational modification in diverse settings.

5.4 Future directions

5.4.1 The function of the egg activation secretome

There are several new molecules identified as released from the egg (e.g., the protease inhibitors, also ZP binding proteins). The experiment was originally intended to rule out degradation as a mechanism of loss for certain candidate proteins. While the data was informative, a full fractionated proteomics experiment to identify the secreted components was never performed. I believe this experiment is feasible, relatively low cost, and would reveal a more complete understanding of the proteins released during egg activation. Additional biochemical studies of these molecules would answer several outstanding questions raised by the work, such as: 1) Do the protease inhibitors affect the activity

of the cortical granule proteases or the sperm acrosome proteases, or both? 2) Does freely diffusing fraction of proteins released after egg activation reduce the fertilization rate of other eggs?

The proteins identified as released from the egg in this study were found as decreasing in the egg and increasing in the egg media. This means that released proteins that do not diffuse through the vitelline membrane are excluded from identification scheme. There are assuredly proteins that are too large in effective size to diffuse into the media, such as proteins localized to exosomes [114].

One example is the protein Juno, which is an essential sperm receptor in mice that eluded discovery for decades, in part due to its unexpected mechanism of loss (receptor exocytosis rather than endocytosis) [113]. There are likely to be other similar molecules. The *Xenopus* homolog to Juno has not been identified. Additionally, there may be other important receptors mediating the transmembrane signaling during egg activation. It is still not known what components interact with the sperm-binding receptor to induce calcium release. They may be found as binding partners with Juno in the exosomes.

The contents of the perivitelline space could be harvested by manually removing the vitelline membranes, activating the egg, and collecting the media for analysis with mass spectrometry. For the perivitelline proteins, it will also be informative to determine which are in exosomes instead of freely diffusing. This could be done by fractionating the perivitelline fluid by filtering or size-exclusion chromatography.

5.4.2 Uncoupling the slow block to polyspermy from meiotic exit

The proteomics time series data contained sufficient information to resolve the signatures of multiple biological programs occurring in parallel. Mainly, the events were divided into those that related to meiotic exit or the block to polyspermy. We were able to resolve the molecular signals of each program using both the protein and phosphorylation dynamics to a large extent with data analysis alone. Protein degradation seemed to be exclusive to the cell cycle, whereas protein release was presumably exclusive to the block to polyspermy. On the phosphorylation level, the loss of phosphorylation appeared to be a part of the cell cycle exit, whereas increased phosphorylation occurred on proteins seemingly involved in the slow block to polyspermy or calcium signaling. While these patterns are likely to be broadly true, it would be stronger to test this experimentally by triggering one program and not the other. One approach is to treat eggs with an agonist to the calcium sensitive Protein Kinase C (PKC), as evidence has shown

that PKC activity is sufficient to drive the cortical reaction without triggering meiotic exit [16]. This observation is interesting in its own right and worthy of reexamination with standard molecular probes that were developed over the last three decades. If PKC agonist can indeed trigger the cortical reaction independent of the cell cycle, it will be possible to isolate the main biochemical events required for the slow block to polyspermy.

The converse experiment is of interest as well: triggering meiotic exit without the cortical response. This can be done by injecting the egg with cell cycle kinase inhibitors, but it is difficult to perform the injections without inducing the rest of egg activation (i.e., calcium signaling and the cortical response). It is possible to suppress the calcium wave during injection by substituting the sodium cation with potassium in the egg media (personal communication, Marc Kirschner). Alternatively, simply adding cell cycle kinase inhibitors to the media may achieve the required perturbation. One experiment is to test whether cells that have been chemically induced into interphase are still competent to trigger the cortical response. Performing these experiments with similar technology as developed in this thesis will provide comprehensive data to test the individual pathways of egg activation.

There are many potentially interesting results, for example to find selected phosphorylation trends that increase after inducing meiotic exit, or perhaps secretion events triggered by the inhibition of Cdk1. Almost all the known programs downstream of egg activation are sensitive to calcium levels, but whether there is a further dependency between the block to polyspermy and the cell cycle is not known.

5.4.3 Exploring egg isoforms specific to species with large eggs

We identified a *Xenopus*-specific isoform of the centrosomal protein SSX2IP as a novel degradation target. The hit seemed plausible from the proteomics data and especially promising after considering the literature. Human SSX2IP has known functions at the primary cilium [240, 241]. Work from our lab showed that ciliary defects occur when Cdc20 is depleted [77], and SSX2IP contains a confident D-box prediction [78]. However, we found that the two orthologs most similar to human SSX2IP in *Xenopus* were stable during egg activation. Only the *Xenopus*-specific ortholog is potentially degraded. The main difference between the isoform sequences is a long, ~300 amino acid extension at the C-terminus of the protein in the *Xenopus*-specific version. At first, we supposed this was an artifact of the transcript-based database. However, the extended C-terminus indeed maps to several other genes in

other amphibians and fish. In other words, the ortholog is only found in species with large oocytes. Our egg proteome data has identified several such isoforms, including an egg-specific version of a motor protein with functions at the spindle [291]. Large cells must complete similar tasks to cells that are orders of magnitude smaller with essentially the same components. Using the techniques from this thesis, it will be possible to systematically identify proteins that have evolved to function in big eggs. SSX2IP may be one example of many that are revealing of biochemical adaptation to extreme cell biological conditions (in this case, large size).

5.4.4 Investigating a novel "checkpoint" mechanism during the first mitotic interphase

A remarkable feature of the cell cycle in *Xenopus* eggs is its autonomous nature. The core cell cycle components function in the absence of many other cellular components, as seen in the regular oscillations of the cell cycle in unfertilized activated eggs. These match the 30 minute cell cycle as seen in the cleavage divisions in the fertilized egg.

One major exception is the timing of the first mitotic interphase. This special phase of the cell cycle takes much longer than the rest of the cleavage divisions that follow. During this time, the paternal and maternal pronuclei migrate to the center of the egg, undergo syngamy to produce the diploid nucleus, and build the large mitotic spindle. Interestingly, the activated egg, which lacks the paternal genome and centrosomes, skips this long delay and proceeds to the more rapid oscillations [92]. I believe this phenomenon hints at a unique "check point" that is most likely different from the standard check point machinery (i.e., the spindle assembly checkpoint and others, which are not active).

Proteomics is one way to compare the differential protein and phosphorylation states between the fertilized and activated eggs. There are other experiments to perform in addition to (or in parallel with) mass spectrometry measurements. It may be that the presence of the paternal DNA or the centrosome is sufficient to induce the delayed cell cycle. Just as interphase extracts are competent to degrade Cdh1 substrates, the addition of the missing factor may be enough to trigger the machinery. Male pronuclei (i.e., androgenetic activation), isolated diploid nuclei, and purified centrosomes could be injected individually and in combination. It may be that timing is critical as well. If the late injection of these factors is not sufficient, then the delay may be induced by an earlier signal. It could also be that the long

interphase is directly dependent on sperm binding. This could be tested by fertilizing the egg with irradiated sperm. Cytoplasmic transfer could also be used to test the inductive nature of the mechanism.

5.5 Future directions addendum– the gastrula timer

The cell cycle is a fundamental timing mechanism in development. I have often wondered whether there are more proteins that oscillate in addition to the known set of Cyclins, Securin, etc. Our data suggests there may be very few. Nevertheless, a related question remains: are there other timers in early development with similar, autonomous features to the cell cycle?

One possibility is that of a gastrulation timer. Over the past century, there are scattered hints in the literature from Holtfreter and many others that suggest the initiation of gastrulation is unexpectedly independent of many of the other developmental events. The idea of an autonomous gastrulation timer has intrigued me since conversations about the phenomenon of pseudogastrulation (discussed below) with my advisor Marc Kirschner early in my graduate work. From reading papers and personal communications, the main limitation in studying the phenomenon has been technological. Any mechanism of timing appears to be strictly biochemical in nature, at a time of development where screening is difficult due to the maternal deposits of protein and RNA. Modern technologies are perhaps advanced enough to gather molecular evidence necessary to identify the mechanism for such a phenomenon. Here, I briefly review the evidence that additional autonomous timers in development may exist, specifically as they pertain to gastrulation.

5.5.1 Overview of the gastrula timer

During gastrulation, embryos execute a coordinated series of morphogenic movements to establish the animal body plan. The timing of gastrulation is typically seen as strictly dependent on the succession of events following fertilization. However, parthenogenetic egg activation and embryonic manipulation demonstrate that some features of gastrulation are independent of cell division, DNA replication, and RNA transcription. One impressive observation is the process of “pseudogastrulation,” first characterized in the frog *Rana pipiens*, where non-fertilized, single cell eggs still undergo properly timed gastrulation-like surface membrane movements of epiboly and partial invagination following egg

activation. This was a single-cell phenomenon, and it does not resemble the multicellular movements like those of convergent extension. Nevertheless, the similarity of the morphological movements to gastrulation despite the single cell state is striking.

A molecular marker was identified by assaying the synthesis of the extracellular matrix protein fibronectin, which is normally coincident with gastrulation onset. Fibronectin is still expressed at the correct time in activated eggs [268]. Limited chemical perturbation experiments performed in the *Rana pipiens* based on pseudogastrulation (but no molecular markers) suggest that the molecules constituting the "timer" may be maternally deposited and engaged by egg activation. However, the mechanisms of action for gastrulation onset have eluded identification for nearly a century. Early development is largely non-transcriptional, therefore the timing mechanisms are likely reflected in translational or post-translational (e.g., phosphorylation) control.

5.5.2 Known timers in development

Biological timers establish time intervals to allow for proper sequence of events. The circadian clock [292], the larval-dauer transition in nematodes [293], the somite clock [294], and oscillation dependent differentiation of progenitor neurons [295] are examples of important molecular timers. Most are thought to function with complex feedback loops between transcription and translation. However, early development is largely transcriptionally quiescent, and therefore must rely on translational or post-translational control. A key example of a non-transcriptional oscillator in early development is the cell cycle. Another example is the timing of the onset of zygotic transcription: the mid-blastula transition (MBT). The MBT is marked by asynchronous cell divisions and the maternal-to-zygotic transition where endogenous RNA expression begins [23, 169]. The timer is dependent on the nuclear-to-cytoplasmic ratio through titration of cytoplasmic factors by replicating DNA [296].

5.5.3 Evidence for additional non-transcriptional timers in early development

Following the MBT, the embryo establishes the three germ layers through a series of coordinated cellular movements called gastrulation. Though gastrulation is a multi-cellular process whose execution requires zygotic transcription, evidence from the older embryological literature suggests its regulation is surprisingly independent of preceding events like embryonic cleavage and the MBT. This was most dramatically observed in 1943 by Holtfreter in non-fertilized parthenogenetically activated eggs from the

frog *Rana pipiens* [297]. Due to the lack of paternal DNA or a centrosome, activated eggs lack cell division, DNA replication, and the MBT. However, single-celled, activated eggs still initiate many morphological aspects of gastrulation at the proper time in a process called “pseudogastrulation,” such as epiboly-like expansion of the animal cortex that causes a buckling at the normal site of invagination. This observation was confirmed in the 1960s and 70s in several species of amphibians [298-302](personal communication to Nieuwkoop from E.C. Boterenbrood [303]). Similarly, acellularized fruit fly embryos were recently shown to induce properly timed hydrodynamic flow that occurs in physiological, cellularized gastrulation [304]. The earliest observation of a pseudogastrulation-like phenomenon was in the activated eggs of the annelid worm in 1902 [305]. Experiments in fertilized embryos are consistent with the autonomous nature of gastrulation onset. When the nuclear-to-cytoplasmic ratio is reduced by placing a diploid nucleus into a quartered egg, the mid-blastula transition occurs two hours early [306, 307]. However, gastrulation onset occurs at the same time as in unperturbed embryos.

5.5.4 The dependencies of the gastrula timer

The molecular identity and mechanism(s) of the gastrula timer are not known. However, embryologists have conducted some perturbations that disrupt the onset of gastrulation. Consistent with the results from activated eggs, small molecule inhibitors of DNA replication do not alter the timing of gastrulation [308, 309]. The morphological movements in activated eggs are sensitive to cytoskeletal poisons; whether these drugs perturb the timing or the execution of the movements directly is not clear [300, 309]. In the embryo, treatment with nocodazole blocks gastrulation, but taxol treatment does not [310]. Fusing two sea urchin embryos at different ages showed no effect on the timing of gastrulation. Therefore, inter-cell communication is likely not involved [311]. These perturbations argue for a timer that is comprised of intracellular, maternally-deposited factors independent of transcription and cell division, but retesting with molecular measurements is needed.

Additionally, the cell cycle dependency of the gastrula timer is unclear from the existing literature. There is weak evidence that gastrulation is sensitive to broad spectrum kinase inhibitors [308], as well as protein synthesis inhibitors [309], which may disrupt the cell cycle. Further testing is required to understand whether the gastrula timer acts autonomously from the cell cycle.

While many other developmental events, e.g. cell division, do not occur in parthenogenetically activated eggs, the major events of fertilization are still initiated. This means the gastrula timer is a mechanism triggered by the events following egg activation and connects gastrulation to fertilization over many hours. Importantly, the cell cycle and non-cell cycle events in egg activation are regulated by phosphorylation. Though there are many modifications, measuring phospho-protein dynamics is likely to hold insight to the gastrula timer mechanism.

5.5.5 Proposed experiments

The following section details experiments proposed in a previous grant detailed here for posterity to provide a sense of the types of experiments that would give insight to the problem.

"I will confirm that maternal DNA does not perturb expression of fibronectin in X*enopus* by enucleating oocytes, followed by activation. To test whether the paternal DNA has any effect, I will in parallel perform androgenic activation where sperm DNA is injected into the egg following enucleation [312]. To confirm the independence from transcription, I will perturb the activated eggs with the transcriptional inhibitor alpha-amanitin. Together, these perturbations are expected to show no effect on the expression of fibronectin in activated eggs, which argue strongly for maternally-deposited factors as timer components. To test the dependence on the cytoskeleton, I will also treat with microtubule and actin poisons, Colchicine and Cytochalasin D, respectively.

To test for cell cycle dependence, the most physiological experiment is to treat embryos with hydroxyurea, which will inhibit DNA synthesis and arrest the cell cycle near the MBT [313], then assay fibronectin expression in the embryo. Hydroxyurea seems to have no ill-effect on early embryogenesis until the MBT (personal communication, Marc Kirschner). I will also use the activated egg system to block the cell cycle in both metaphase and interphase. First, I will arrest in M-phase by injecting non-degradable Cyclin-B[234]. As a parallel approach, I will also inject EGTA that inhibits the calcium sensitive release from egg-specific "CSF" metaphase arrest. To arrest in interphase, I will inject p27/Xic1, which is a protein inhibitor of Cyclin-dependent Kinases [314]. However, the proteins may not be stable over the hours leading to gastrulation. If p27/Xic1 is degraded, I will instead inject Roscovitine [279] (small molecule CDK inhibitor) in both activated eggs and into the blastocoel of fertilized embryos. The clearest interpretation will be if multiple inhibitions of the cell cycle failing to ablate fibronectin expression,

suggesting cell cycle independence. However, ablating gastrulation onset through cell cycle manipulation must be interpreted cautiously. For example, a mitotic arrest generally inhibits translation, which may explain the lack of a fibronectin expression. Results from multiple methods used for arresting in interphase will help uncouple the many “global” effects of different cell cycle states. If the timer is not cell cycle dependent, we will inject the egg with protein synthesis, protein degradation, kinase, and phosphatase inhibitors.

Finally, I will perturb with two simple embryological manipulations: cytoplasmic pinching to show potential sub-cellular localization to the process [306]. The animal pole region may, for example continue to express fibronectin properly after activation, where the vegetal region will not. I will also use cytoplasmic transfers from different “aged” eggs to distinguish between production/depletion models. If a factor is depleted, transfers from an older egg to a younger egg should have no effect on timing, whereas if a factor is produced, cytoplasm from an older egg should trigger earlier expression of fibronectin. This is also a test of potency of the activity, since only relatively small amounts of cytoplasm can be transferred.”

The general class off experiments described in this section can coupled with or supplemented by proteomic time series data to identify protein trends that are maintained in the activated egg over the first 12 hours of development. This is one example of how the technology in this thesis could be used to bring molecular detail to interesting phenomenon in development where little was previously known.

5.6 Conclusion

In the late 19th century, a series of cytological studies allowed for the formulation of new theories about the specific mechanisms underlying the events after fertilization [2]. These included speculations regarding a “block to polyspermy” and the basic requirements for the early cell divisions. The extraordinary century of investigation that followed was successful at identifying many molecular components and central principles underlying the individual programs of egg activation. These were broadly relevant to the rest of biology. However, fertilization induces an impressive transition in the cell that requires the changes of the activity, localization, and general function of thousands of different molecules with minimal input from the genome. How these are integrated in the cell is still largely not understood. Making observations to understand the cell at this level of broad-scale biochemistry remains challenging even with contemporary technological capabilities.

The data in this thesis provide some insight on how the embryo integrates the diverse biochemical programs in egg activation. A small number of low abundance molecules are targeted for degradation. As these changes affect highly active kinases, a relatively modest change to the overall proteome is amplified massively by the change in phosphorylation state of many targets as part of the cell cycle. In parallel, increasing phosphorylation directs the changes at the cell cortex, including the extensive secretion events that block further fertilization.

The cell cycle and blocks to polyspermy are thought to operate largely independently from each other. It may be significant that though the two programs rely on phosphorylation, they employ opposite regulatory strategies (the inactivation of kinase and loss of modification for the cell cycle, versus the activation of kinase activity and the increase in modification for the other). Despite their independence, the cell cycle and block to polyspermy are induced through the same second messenger signaling of the calcium wave. The clarity of the biochemical regulation discussed here can be improved by incorporating the data with dynamical modeling principles like flux analysis and bistability, or physical principles like chemical waves. While these have gained prominence in recent years, there has been limited opportunity to explore such models that capture more than a few components. The methods developed in this dissertation to determine the absolute dynamics of the protein biochemistry were indispensable to our own interpretation of the data, as they well may be for future applications.

We chose to study egg activation in the frog egg because of its general importance, but also because enough was known to orient us while applying a new tool. By analogy, I never viewed quantitative mass spectrometry as a "screening assay." Rather, I hoped it would offer a new way to "see" the cell, and direct us toward fundamental questions not so dissimilar in scope to those enabled by the late 19th century microscopy. There are questions of this kind that could already be advanced by the approaches used here. Determining the timing mechanism of gastrulation is one example; enumerating the features of the early interphase checkpoint is yet another example.

It is worth noting that similar questions were posed in older literature and eventually abandoned. This is in large part due to technological constraints at the time, but also because of changing research interests. In an interview with John Gerhart in the late 1980s, Pieter Nieuwkoop lamented the increased influence of molecular biology perspectives at the expense of the events in the embryonic cytoplasm that help build the whole organism [315]. In fairness, these events have been difficult to pursue even with the most mature tools and accessible organisms. It is reasonable for biologists to gravitate toward simpler systems with better experimental control.

These types of conflicts are not new. According to Lillie, even in the early 20th century "the tendency toward excessive simplification of the physicochemical school is constantly being checked by the biological school, and the conservatism of the latter has been more than once rudely shaken by the former [2]." As perturbation and measurement techniques become increasingly sophisticated, the conflict between tractable and natural will become less pronounced. With these new advances, confronting biochemical regulation in complex, intact living systems will only become more feasible.

My hope is that this thesis has contributed in some way toward that new (and yet quite old) direction of scientific exploration.

Appendix 1: Supplementary Figures

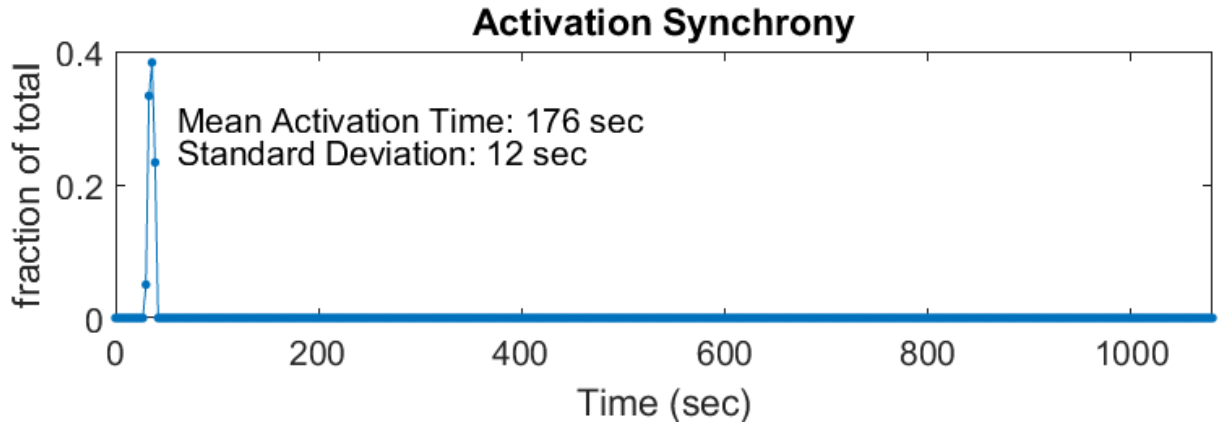

Figure S1 Synchrony of electrical activation. Distribution of egg activation following electro-shock. The scale of the X-axis is 18 minutes (displayed in seconds). Of the 400 eggs used in each experiment, ~70 were filmed per experiment in parallel with the collection of the time course. Images taken every 5 seconds (see Movies 1, 2, 3 and 4 filmed for each replicate; 14 frames/sec). The movies show de-jellied Xenopus eggs mostly oriented with the pigmented animal pole facing upward. The calcium waves and cortical contractions are visible. Of the ~250 eggs filmed across all replicates, only two failed to activate. The distribution of egg activation is shown above, assayed by the calcium wave (surface contraction wave) crossing the midpoint of the animal pole. Given the resolution of the movie, we conservatively estimate that all eggs which activate do so with a standard deviation of less than 15 seconds. Eggs that are inverted so the vegetal (i.e., white) side is facing upward are not quantified.

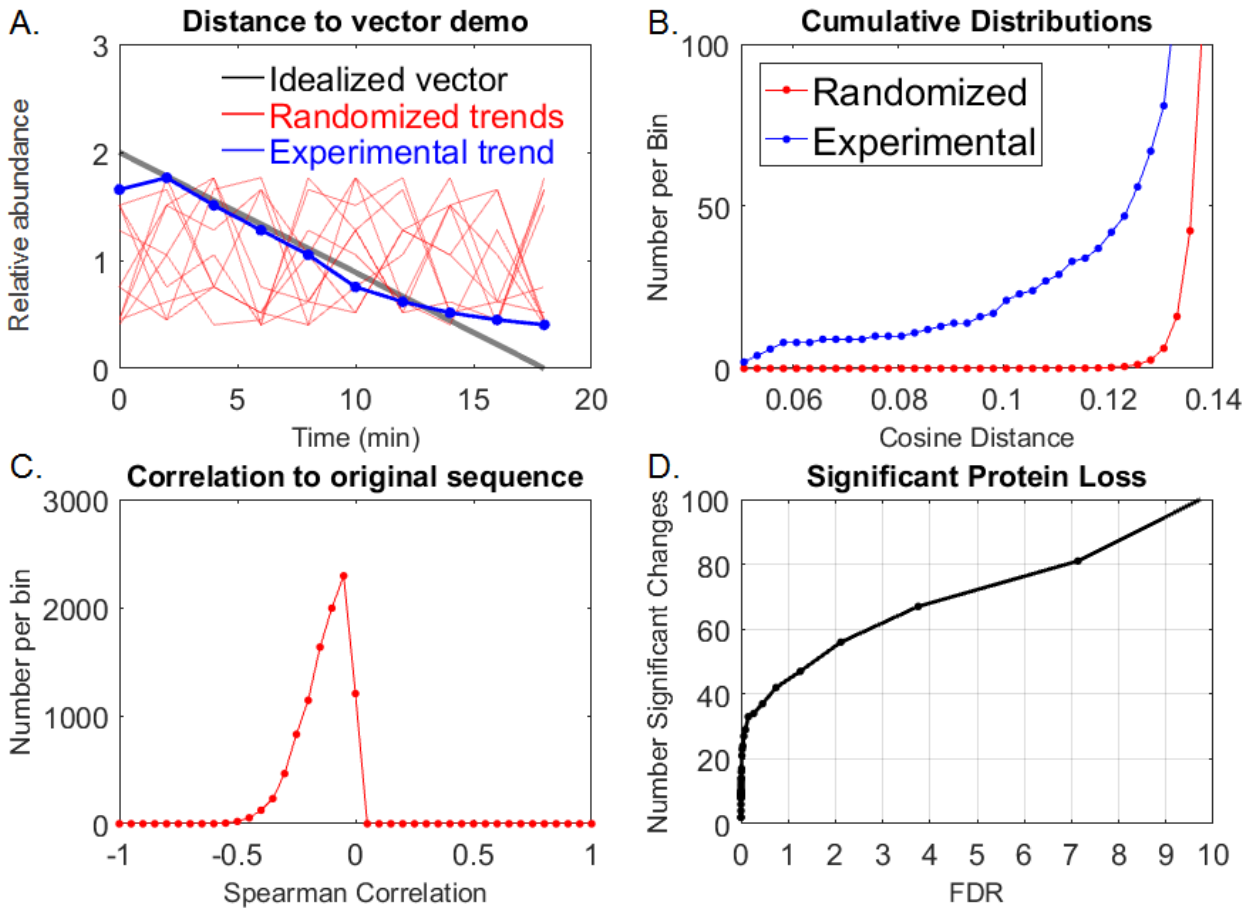

Figure S2 Estimating a false discovery rate for protein loss. Principles and visualization of the FDR calculation. See Supplementary Methods for further explanation of the implementation.
A) Plots to demonstrate the geometric distance measurements used in calculating false discovery rate. Here, an idealized 10-element degradation vector (gray, mean normalized to match the experimental data) is shown. The Cosine distance of the experimental trends (blue, in this case, Cyclin-B1) was measured to the idealized vector. The time sequence of the experimental trend was then iteratively randomized (red) and Cosine distances re-calculated each time.
*B) The cumulative distributions of distances for all proteins (experimental) and pooled randomized trends from the 10,000 iterations. These distributions are used to calculate the FDR at a given cosine distance to the idealized vector (Randomized/(Experimental + Randomized) *100). Since the randomized values are pooled, we apply one cutoff to the entire dataset. Multiple hypothesis correction is therefore not required.*
C) Distribution of the correlation of the randomized values to the original time sequence. The original time sequence is "0, 2, 4,..18min", which is scrambled during the randomizations. We excluded any randomized sequence that was positively correlated with the original sequence (e.g., "0, 4, 2*,...18min", which is essentially identical to the original order). This was done to avoid overestimating the false discovery from randomly recreating the original order of the time series and counting this as a "false discovery."*
D) Number of proteins identified as significantly decreasing from the distributions in Figure S2B. The FDR is calculated by Protein ID, then combined by gene symbol afterward (i.e., we identify multiple "forms" of Cyclin B1, which may be splice isoforms of pseudoalleles, which are later averaged to one gene symbol). We identify 41 protein ID's which significantly decrease which map to 29 unique gene symbols.

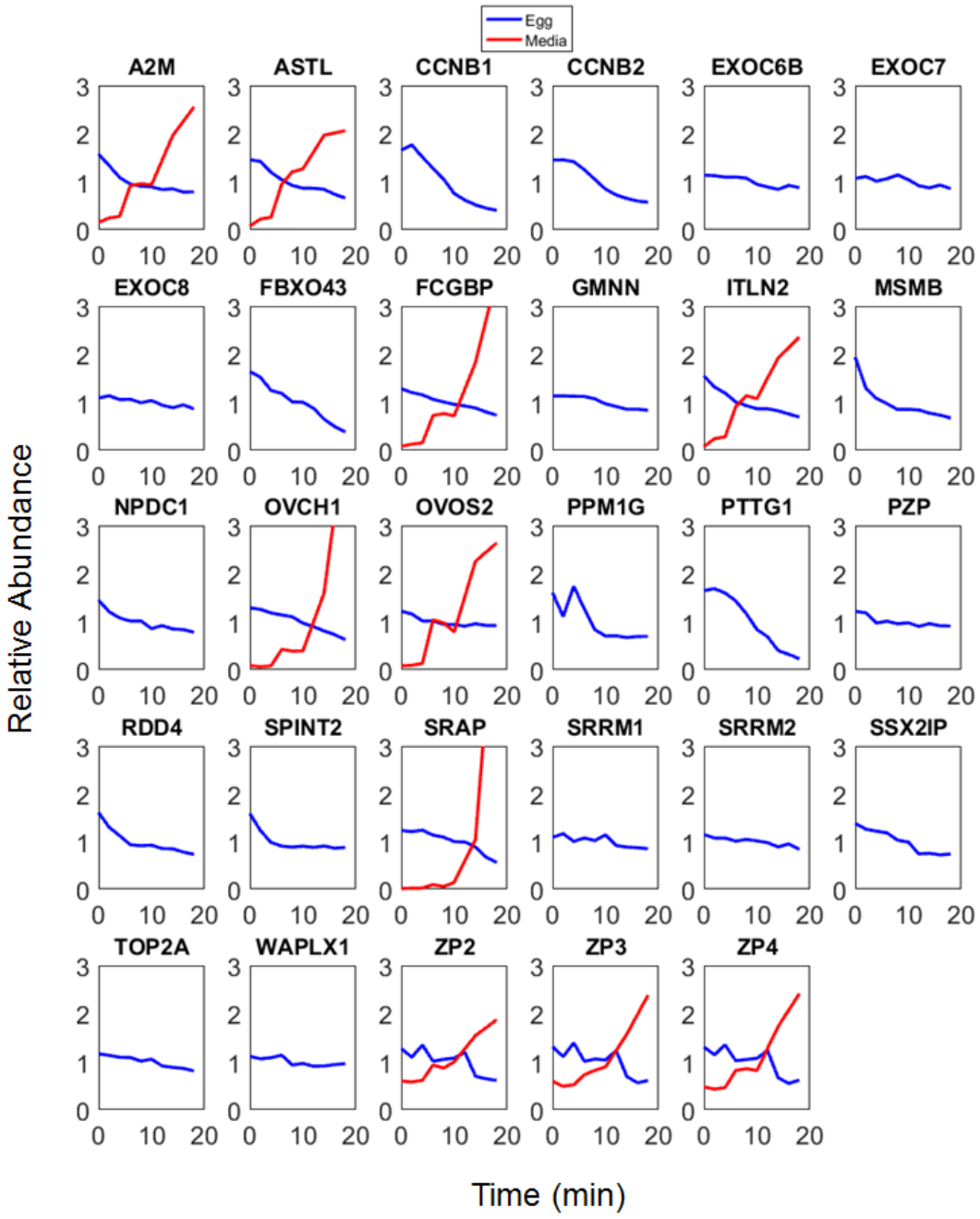

Figure S3 Degradation and release of proteins at 1% FDR. Time series data of the 29 genes which significantly decrease at 1% FDR following egg activation. CCNB1, CCNB2, GMNN, FBX043 (Emi2), and PPTG1 (Securin) are known APC/C substrates. Blue trends decrease abundance in the egg. Red trends increase in the supernatant; if no red trend is shown, it was not detected in the egg media. These are further classified in Appendix 2, Table S1. For ZP2, 3, and 4, the jagged decrease from the egg data is not seen in its appearance into the supernatant. Therefore, the noise in the ZP trends is consistent with a technical rather than biological explanation. It is not clear why ZP proteins are more prone to this noise than other released proteins. We suspect this is because the ZP proteins are released by proteolysis rather than exocytosis. This causes greater inconsistencies in the removal of ZP fragments from the egg surface during harvesting.

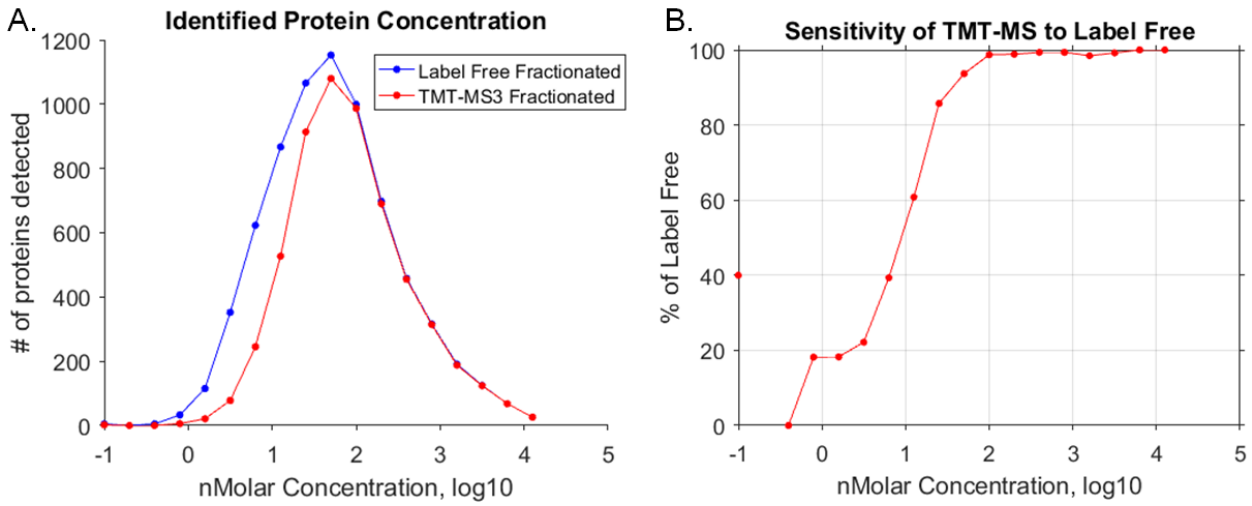

Figure S4 Sensitivity of label-free and multiplex TMT-MS3 methods.
A) Histograms comparing the distribution of abundances of proteins detected by label-free (blue) and fractionated TMT-MS3 (red) methods in the Xenopus laevis egg (all four replicates are pooled for the TMT distributions). The label-free method is the most sensitive technique available, but still does not detect the entire proteome. This is evident from by the fact that we can detect many more mRNA species via sequencing than protein species via MS. Additionally, there are proteins that are known to be expressed but are not detected in the label-free dataset (e.g., Axin). The fractionated TMT-MS3 method used in this study to quantify the proteins captures a subset of the label-free distribution, but with substantial overlap. B) Plot comparing the percentage of proteins identified in the TMT-MS compared to the label-free as a function of concentration. The fractionated TMT data is essentially equivalent to 100nM, and at 50nM, ~95% of proteins identified with the label-free method are also seen with TMT-MS. All known APC substrates are at least 60nM. At 10nM, we detect 50% of the proteins with TMT-MS that we do with label-free. Though detection is less reliable at these lower concentrations, EMI2 (10nM), for example, is still detected in 2 of the 4 replicates. The 40% overlap seen at 0.1 nM (-1 in log10) is due to under-sampling, as the numbers are very small at those concentrations.

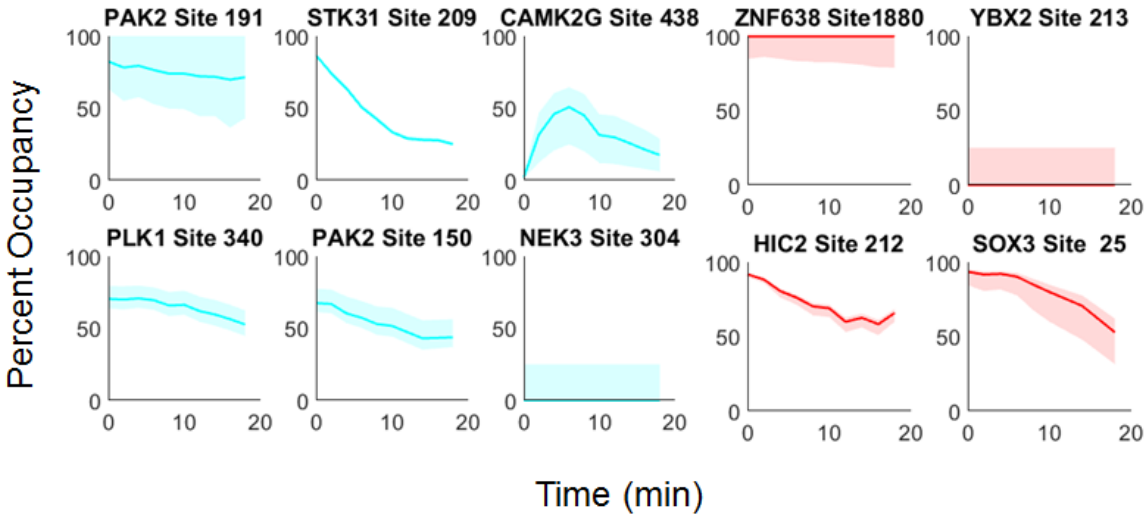

Figure S5 Panel of kinase and transcription factor phospho-occupancy dynamics. Time series of phosphosite occupancy dynamics on kinases (cyan) and transcription factors (red) following egg activation. Shaded areas are 95% confidence intervals.

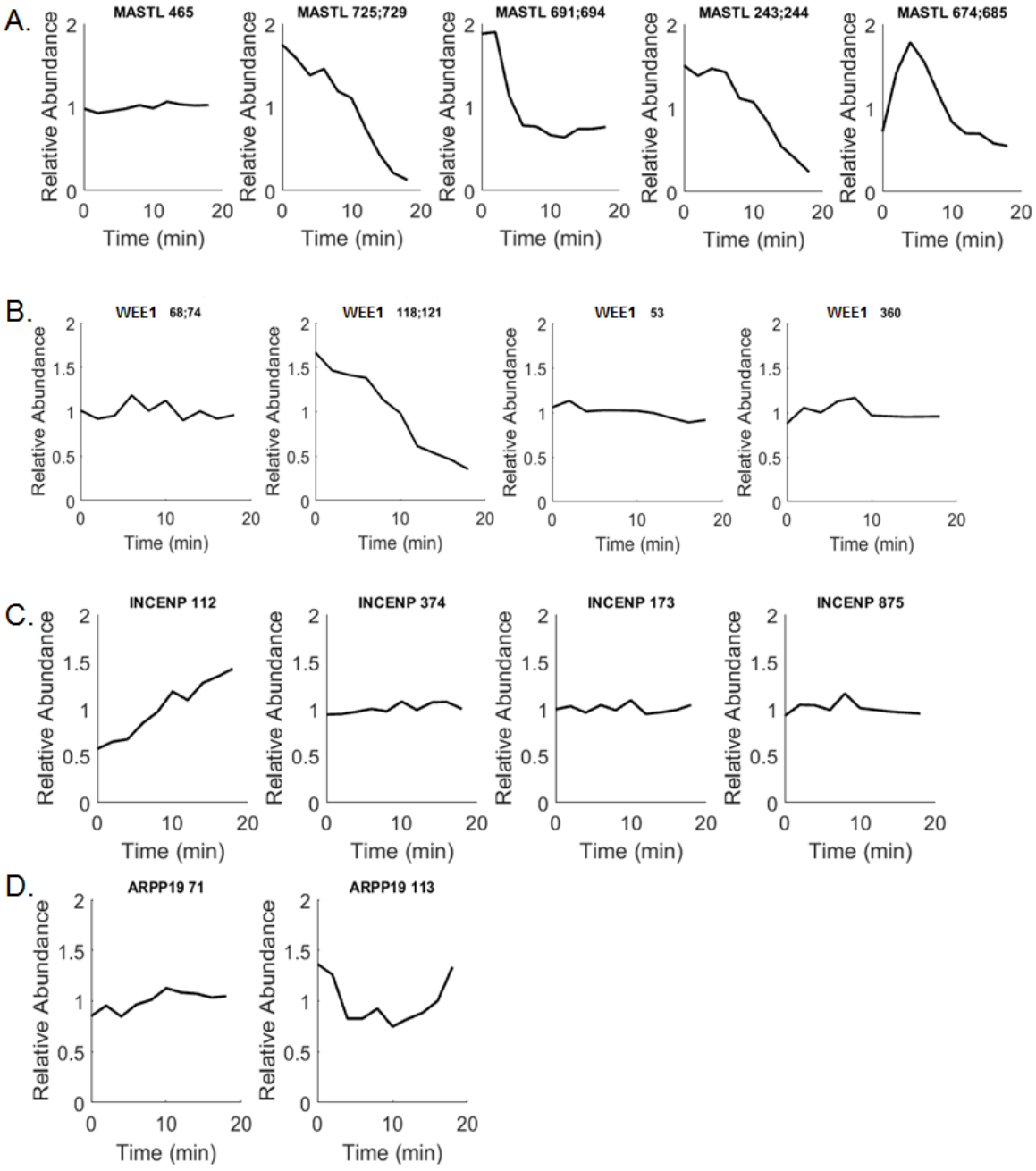

Figure S6 Phosphosite dynamics on selected cell cycle proteins. Time series of phosphosite relative abundance on known cell cycle regulators or targets. Site number indicated in title. A) Greatwall Kinase, B) WEE1 Kinase, C) INCENP, D) ENSA (PP2A inhibitor).

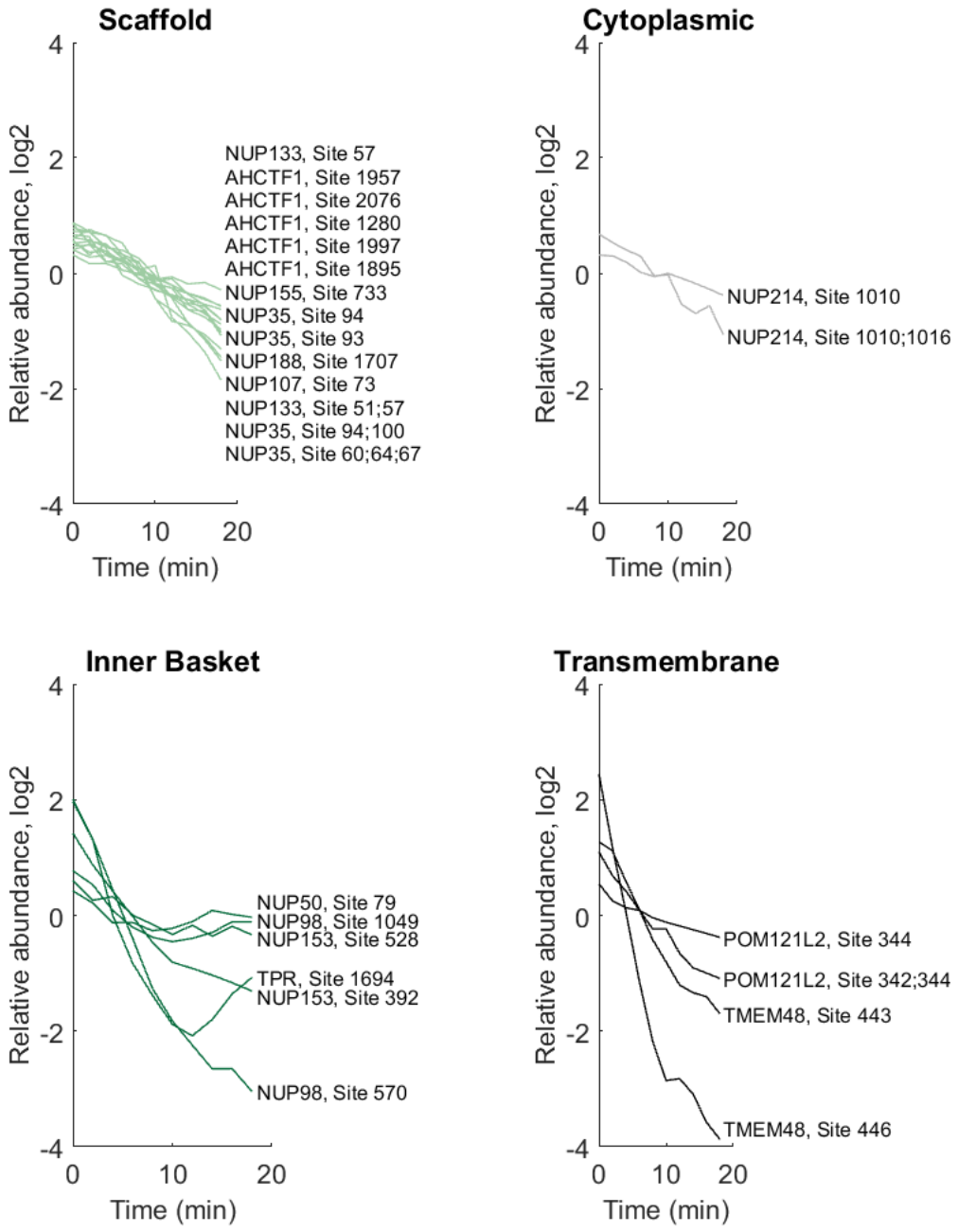

Figure S7 Nucleoporins show differential dephosphorylation corresponding to nuclear pore sub-complexes. Time series of nucleoporin phosphosite dynamics plotted by their respective sub-complexes. Many of the proteins have multiple phosphosites, indicated by multiple entries in the legend. The phospho-occupancy plots for the subset of these sites that are available are shown in Fig. 6E. The slower dephosphorylation (~10 min half time) of the core scaffold (e.g., NUP188) and cytoplasmic components is consistent with their later recruitment. The fast dephosphorylation (~2-5 min half times) of inner basket/nuclear ring components (e.g., NUP153) is consistent with their early roles in NPC assembly of binding chromatin and nuclear envelope association. An exception is that trans-membrane nucleoporin recruitment is a secondary step, yet nearly all the sites that we detect in this region show fast dephosphorylation. Another exception is the scaffold component NUP107, which is recruited early, but shows slow dephosphorylation.

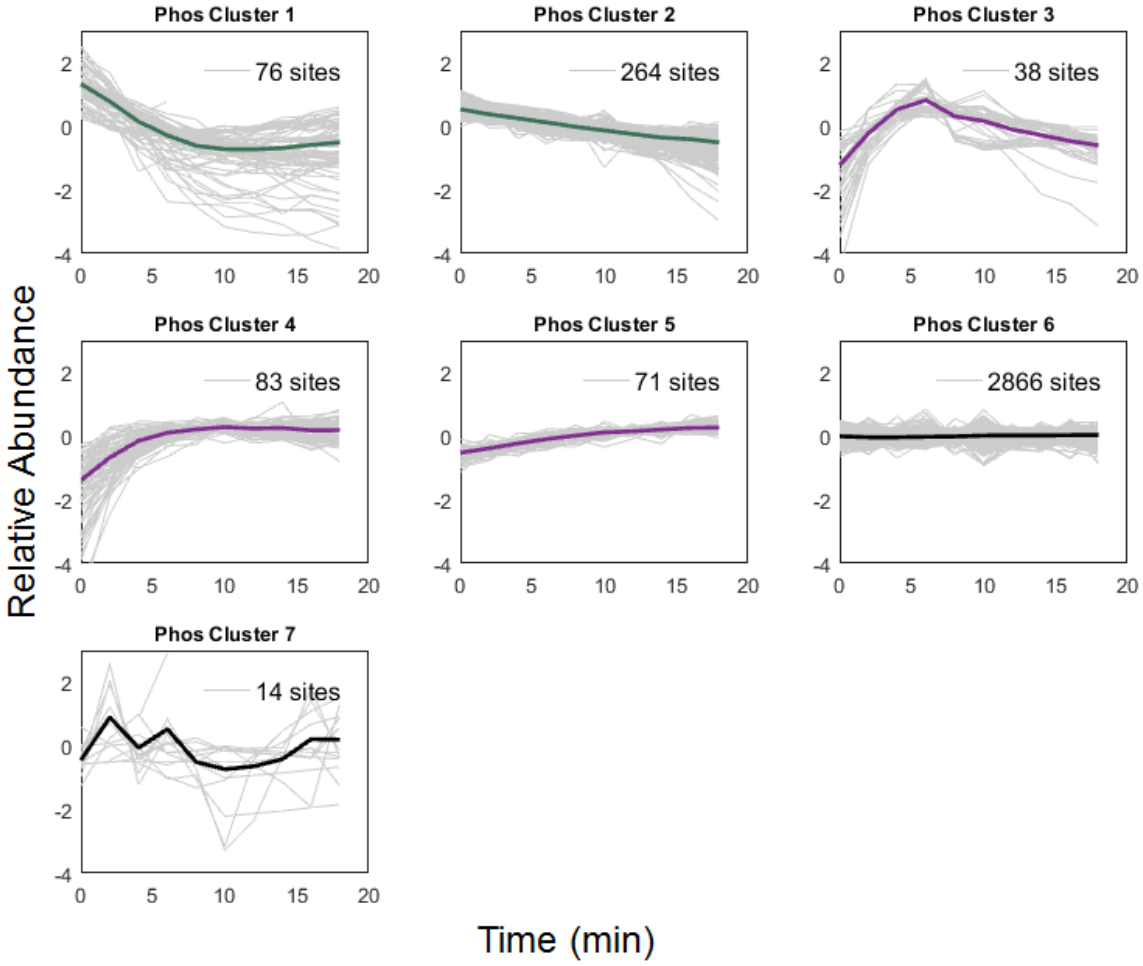

Figure S8 K-means clustering of phosphorylation dynamics. Time series of phosphosite cluster dynamics, including a 7th cluster of "noisy" trends excluded for display purposes only. These are included in all statistical analyses. Artifacts from single site dynamics are removed from clusters (as shown in Figure 4.9). See Methods for further explanation of collapsing the clustering.

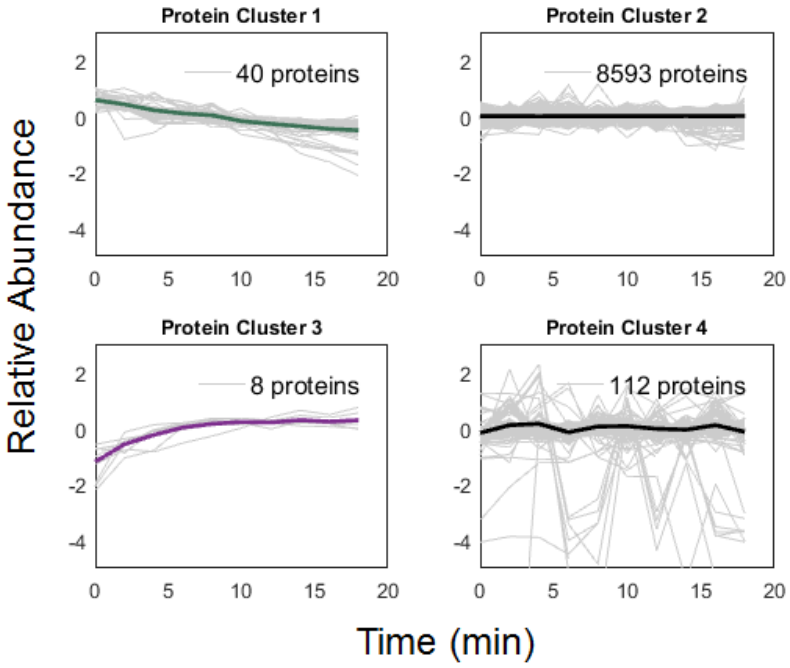

Figure S9 K-means clustering of protein dynamics. Time series of protein cluster dynamics, including a 4th cluster of "noisy" trends excluded for display purposes only. These are included to all statistical analyses (e.g., the FDR analysis). See Methods for further explanation of clustering.

Pseudo-code for computing occupancy with confidence intervals from multiplexed-MS data

Load data

%%Load the .xlxs or other file containing TMT signal for phospho sites and matched unmodified peptides%%

For each site in the data set:

Replace any zeros with 1e-9

For each condition:

%% Calculate occupancy value %%
Set the current condition as the reference
Normalize phospho site and unmodified data to the reference
Pass ratios to total least squares regression function and fit slope
Calculate occupancy from negative inverse of slope

End

%% Calculate confidence interval for estimated occupancy value %%
If, for all conditions, the best-fit occupancy is between 0 and 100
For each condition
Check for repeated values within phospho site and unmodified peptide data and flag
For each bootstrap iteration
Randomly choose conditions with replacement for fitting
Check that the same condition has not been chosen every time. If it has, choose random integer between 0 and 100 and move to next iteration
Create new input values using the conditions chosen above (If flag is present, check to make sure the new input does not define a line with infinite or zero slope and if so set occupancy to 0 or 100, respectively and move to next iteration)
Pass new input ratios to total least squares regression function and fit slope
Calculate occupancy from negative inverse of slope
End
Use the 'bias corrected and accelerated percentile method' determine the upper and lower 95% confidence intervals
End
Else
Report upper and lower confidence interval bounds as 0 and 100, respectively, for all conditions

End
End

Appendix 2: Supplementary Table

Table S1: Mechanism of loss classification

Gene Symbol	Classification
NPDC1	Candidate degradation target (Ubiquitinated, 1% FDR, no evidence for release)
SSX2IP-L	Candidate degradation target (Ubiquitinated, 1% FDR, no evidence for release)
CCNB1	Known degradation target
CCNB2	Known degradation target
FBXO43	Known degradation target
GMNN	Known degradation target
PTTG1	Known degradation target
TOP2A	Not ubiquitinated
SRRM1	Not ubiquitinated, also likely specious trend from phospho (by similarity to SRRM2)
PPM1G	Likely specious trend from phosphorylation, not ubiquitinated
SRRM2	Likely specious trend from phosphorylation, not ubiquitinated
A2M	Experimentally confirmed released
ASTL	Experimentally confirmed released
FCGBP	Experimentally confirmed released
ITLN2	Experimentally confirmed released
OVCH1	Experimentally confirmed released
OVOS2	Experimentally confirmed released
SRAP	Experimentally confirmed released
ZP2	Experimentally confirmed released
ZP3	Experimentally confirmed released
ZP4	Experimentally confirmed released
MSMB	Secreted annotation
PZP	Secreted annotation
RDD4	Secreted annotation
EXOC6B	Likely released (similar class reported in perivitelline space, Danilchick et al, 2017)
EXOC7	Likely released (similar class reported in perivitelline space, Danilchick et al, 2017)
EXOC8	Likely released (similar class reported in perivitelline space, Danilchick et al, 2017)
SPINT2	Likely released (similar to other protease inhibitors secreted)
WAPLX1	Likely released (by annotation)

References

1. Lopata, A., *History of the Egg in Embryology.* Journal of Mammalian Ova Research, 2009. **26**(1): p. 2-9.

2. Lillie, F.R., *Problems of fertilization*. 1919. Chicago, Ill. The University of Chicago press.

3. Loeb, J., *Artificial Parthenogenesis and Fertilization*. 1913, Chicago: The University of Chicago Press.

4. Scheer, U., *Historical roots of centrosome research: discovery of Boveri's microscope slides in Wurzburg.* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1650).

5. Boveri, T., *Ueber die Natur der Centrosomen. Zellen-Studien 4*. 1900, Germany: G. Fischer.

6. Clift, D. and M. Schuh, *Restarting life: fertilization and the transition from meiosis to mitosis.* Nat Rev Mol Cell Biol, 2013. **14**(9): p. 549-62.

7. Colas, P. and F. Dube, *Meiotic maturation in mollusc oocytes.* Semin Cell Dev Biol, 1998. **9**(5): p. 539-48.

8. Davidson, E.H., B.R. Hough-Evans, and R.J. Britten, *Molecular biology of the sea urchin embryo.* Science, 1982. **217**(4554): p. 17-26.

9. Rauh, N.R., et al., *Calcium triggers exit from meiosis II by targeting the APC/C inhibitor XErp1 for degradation.* Nature, 2005. **437**(7061): p. 1048-52.

10. Schmidt, A., et al., *Cytostatic factor: an activity that puts the cell cycle on hold.* J Cell Sci, 2006. **119**(Pt 7): p. 1213-8.

11. Farr, K.A. and O. Cohen-Fix, *The metaphase to anaphase transition: a case of productive destruction.* Eur J Biochem, 1999. **263**(1): p. 14-9.

12. Iwao, Y., *Egg activation in physiological polyspermy.* Reproduction, 2012. **144**(1): p. 11-22.

13. Just, E., *The fertilization reaction in Echinarachnius parma. I. Cortical response of the egg to insemination.* Biol. Bull., 1919. **36**: p. 1-10.

14. Byrnes, W.M., *Ernest Everett Just, Johannes Holtfreter, and the origin of certain concepts in embryo morphogenesis.* Mol Reprod Dev, 2009. **76**(10): p. 912-21.

15. Tokmakov, A.A., et al., *Calcium signaling and meiotic exit at fertilization in Xenopus egg.* Int J Mol Sci, 2014. **15**(10): p. 18659-76.

16. Bement, W.M. and D.G. Capco, *Activators of protein kinase C trigger cortical granule exocytosis, cortical contraction, and cleavage furrow formation in Xenopus laevis oocytes and eggs.* J Cell Biol, 1989. **108**(3): p. 885-92.

17. Markoulaki, S., S. Matson, and T. Ducibella, *Fertilization stimulates long-lasting oscillations of CaMKII activity in mouse eggs.* Dev Biol, 2004. **272**(1): p. 15-25.

18. Terasaki, M. and C. Sardet, *Demonstration of calcium uptake and release by sea urchin egg cortical endoplasmic reticulum.* J Cell Biol, 1991. **115**(4): p. 1031-7.

19. Jaffe, L.F., *Sources of calcium in egg activation: a review and hypothesis.* Dev Biol, 1983. **99**(2): p. 265-76.

20. Zalokar, M., *Autoradiographic study of protein and RNA formation during early development of Drosophila eggs.* Dev Biol, 1976. **49**(2): p. 425-37.

21. McKnight, S.L. and O.L. Miller, Jr., *Ultrastructural patterns of RNA synthesis during early embryogenesis of Drosophila melanogaster.* Cell, 1976. **8**(2): p. 305-19.

22. Brown, D.D. and E. Littna, *Rna Synthesis during the Development of Xenopus Laevis, the South African Clawed Toad.* J Mol Biol, 1964. **8**: p. 669-87.

23. Newport, J. and M. Kirschner, *A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage.* Cell, 1982. **30**(3): p. 675-86.

24. Chavez, D.J. and J.V. Blerkom, *Persistence of embryonic RNA synthesis during facultative delayed implantation in the mouse.* Dev Biol, 1979. **70**(1): p. 39-49.

25. Hara, K., P. Tydeman, and M. Kirschner, *A cytoplasmic clock with the same period as the division cycle in Xenopus eggs.* Proc Natl Acad Sci U S A, 1980. **77**(1): p. 462-6.

26. Karsenti, E., R. Bravo, and M. Kirschner, *Phosphorylation changes associated with the early cell cycle in Xenopus eggs.* Dev Biol, 1987. **119**(2): p. 442-53.

27. Izawa, D. and J. Pines, *How APC/C-Cdc20 changes its substrate specificity in mitosis.* Nat Cell Biol, 2011. **13**(3): p. 223-33.

28. Masui, Y., *The elusive cytostatic factor in the animal egg.* Nat Rev Mol Cell Biol, 2000. **1**(3): p. 228-32.

29. Stukenberg, P.T., et al., *Systematic identification of mitotic phosphoproteins.* Curr Biol, 1997. **7**(5): p. 338-48.

30. King, R.W., et al., *A 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B.* Cell, 1995. **81**(2): p. 279-88.

31. Glotzer, M., A.W. Murray, and M.W. Kirschner, *Cyclin is degraded by the ubiquitin pathway.* Nature, 1991. **349**(6305): p. 132-8.

32. Potapova, T.A., et al., *The reversibility of mitotic exit in vertebrate cells.* Nature, 2006. **440**(7086): p. 954-8.

33. Zou, H., et al., *Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis.* Science, 1999. **285**(5426): p. 418-22.

34. Ciosk, R., et al., *An ESP1/PDS1 complex regulates loss of sister chromatid cohesion at the metaphase to anaphase transition in yeast.* Cell, 1998. **93**(6): p. 1067-76.

35. Masui, Y. and C.L. Markert, *Cytoplasmic control of nuclear behavior during meiotic maturation of frog oocytes.* J Exp Zool, 1971. **177**(2): p. 129-45.

36. Reynhout, J.K. and L.D. Smith, *Studies on the appearance and nature of a maturation-inducing factor in the cytoplasm of amphibian oocytes exposed to progesterone.* Dev Biol, 1974. **38**(2): p. 394-400.

37. Booher, R. and D. Beach, *Involvement of cdc13+ in mitotic control in Schizosaccharomyces pombe: possible interaction of the gene product with microtubules.* EMBO J, 1988. **7**(8): p. 2321-7.

38. Draetta, G., et al., *Cdc2 protein kinase is complexed with both cyclin A and B: evidence for proteolytic inactivation of MPF.* Cell, 1989. **56**(5): p. 829-38.

39. Booher, R.N., et al., *The fission yeast cdc2/cdc13/suc1 protein kinase: regulation of catalytic activity and nuclear localization.* Cell, 1989. **58**(3): p. 485-97.

40. Wurzenberger, C. and D.W. Gerlich, *Phosphatases: providing safe passage through mitotic exit.* Nat Rev Mol Cell Biol, 2011. **12**(8): p. 469-82.

41. Mochida, S. and T. Hunt, *Protein phosphatases and their regulation in the control of mitosis.* EMBO Rep, 2012. **13**(3): p. 197-203.

42. Mochida, S., et al., *Regulated activity of PP2A-B55 delta is crucial for controlling entry into and exit from mitosis in Xenopus egg extracts.* EMBO J, 2009. **28**(18): p. 2777-85.

43. Mochida, S. and T. Hunt, *Calcineurin is required to release Xenopus egg extracts from meiotic M phase.* Nature, 2007. **449**(7160): p. 336-40.

44. Bouchoux, C. and F. Uhlmann, *A quantitative model for ordered Cdk substrate dephosphorylation during mitotic exit.* Cell, 2011. **147**(4): p. 803-14.

45. Visconti, R., et al., *The end of mitosis from a phosphatase perspective.* Cell Cycle, 2013. **12**(1): p. 17-9.

46. Evans, T., et al., *Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division.* Cell, 1983. **33**(2): p. 389-96.

47. Murray, A.W. and M.W. Kirschner, *Cyclin synthesis drives the early embryonic cell cycle.* Nature, 1989. **339**(6222): p. 275-80.

48. Murray, A.W., M.J. Solomon, and M.W. Kirschner, *The role of cyclin synthesis and degradation in the control of maturation promoting factor activity.* Nature, 1989. **339**(6222): p. 280-6.

49. Lara-Gonzalez, P., F.G. Westhorpe, and S.S. Taylor, *The spindle assembly checkpoint.* Curr Biol, 2012. **22**(22): p. R966-80.

50. Tunquist, B.J. and J.L. Maller, *Under arrest: cytostatic factor (CSF)-mediated metaphase arrest in vertebrate eggs.* Genes Dev, 2003. **17**(6): p. 683-710.

51. Chao, W.C., et al., *Structure of the mitotic checkpoint complex.* Nature, 2012. **484**(7393): p. 208-13.

52. Zhang, Y. and E. Lees, *Identification of an overlapping binding domain on Cdc20 for Mad2 and anaphase-promoting complex: model for spindle checkpoint regulation.* Mol Cell Biol, 2001. **21**(15): p. 5190-9.

53. Musacchio, A. and E.D. Salmon, *The spindle-assembly checkpoint in space and time.* Nat Rev Mol Cell Biol, 2007. **8**(5): p. 379-93.

54. De Antoni, A., et al., *The Mad1/Mad2 complex as a template for Mad2 activation in the spindle assembly checkpoint.* Curr Biol, 2005. **15**(3): p. 214-25.

55. Alfieri, C., et al., *Molecular basis of APC/C regulation by the spindle assembly checkpoint.* Nature, 2016. **536**(7617): p. 431-436.

56. Stegmeier, F., et al., *Anaphase initiation is regulated by antagonistic ubiquitination and deubiquitination activities.* Nature, 2007. **446**(7138): p. 876-81.

57. Fang, G., H. Yu, and M.W. Kirschner, *Direct binding of CDC20 protein family members activates the anaphase-promoting complex in mitosis and G1.* Mol Cell, 1998. **2**(2): p. 163-71.

58. Yu, H., *Cdc20: a WD40 activator for a cell cycle degradation machine.* Mol Cell, 2007. **27**(1): p. 3-16.

59. Masui, Y., *A cytostatic factor in amphibian oocytes: its extraction and partial characterization.* J Exp Zool, 1974. **187**(1): p. 141-7.

60. Meyerhof, P.G. and Y. Masui, *Ca and Mg control of cytostatic factors from Rana pipiens oocytes which cause metaphase and cleavage arrest.* Dev Biol, 1977. **61**(2): p. 214-29.

61. Hansen, D.V., J.J. Tung, and P.K. Jackson, *CaMKII and polo-like kinase 1 sequentially phosphorylate the cytostatic factor Emi2/XErp1 to trigger its destruction and meiotic exit.* Proc Natl Acad Sci U S A, 2006. **103**(3): p. 608-13.

62. Descombes, P. and E.A. Nigg, *The polo-like kinase Plx1 is required for M phase exit and destruction of mitotic regulators in Xenopus egg extracts.* EMBO J, 1998. **17**(5): p. 1328-35.

63. Hansen, D.V., et al., *Plk1 regulates activation of the anaphase promoting complex by phosphorylating and triggering SCFbetaTrCP-dependent destruction of the APC Inhibitor Emi1.* Mol Biol Cell, 2004. **15**(12): p. 5623-34.

64. McGarry, T.J. and M.W. Kirschner, *Geminin, an inhibitor of DNA replication, is degraded during mitosis.* Cell, 1998. **93**(6): p. 1043-53.

65. Cohen-Fix, O., et al., *Anaphase initiation in Saccharomyces cerevisiae is controlled by the APC-dependent degradation of the anaphase inhibitor Pds1p.* Genes Dev, 1996. **10**(24): p. 3081-93.

66. Sivakumar, S. and G.J. Gorbsky, *Spatiotemporal regulation of the anaphase-promoting complex in mitosis.* Nat Rev Mol Cell Biol, 2015. **16**(2): p. 82-94.

67. Floyd, S., J. Pines, and C. Lindon, *APC/C Cdh1 targets aurora kinase to control reorganization of the mitotic spindle at anaphase.* Curr Biol, 2008. **18**(21): p. 1649-58.

68. Stewart, S. and G. Fang, *Destruction box-dependent degradation of aurora B is mediated by the anaphase-promoting complex/cyclosome and Cdh1.* Cancer Res, 2005. **65**(19): p. 8730-5.

69. Prinz, S., et al., *The regulation of Cdc20 proteolysis reveals a role for APC components Cdc23 and Cdc27 during S phase and early mitosis.* Curr Biol, 1998. **8**(13): p. 750-60.

70. Shirayama, M., et al., *The Polo-like kinase Cdc5p and the WD-repeat protein Cdc20p/fizzy are regulators and substrates of the anaphase promoting complex in Saccharomyces cerevisiae.* EMBO J, 1998. **17**(5): p. 1336-49.

71. Tang, Z.I., *The Domino and Clock Models of Cell Cycle Regulation.* Nature Education, 2010. **3**(9): p. 56.

72. Lorca, T., et al., *Fizzy is required for activation of the APC/cyclosome in Xenopus egg extracts.* EMBO J, 1998. **17**(13): p. 3565-75.

73. Zhou, Y., et al., *The APC regulator CDH1 is essential for the progression of embryonic cell cycles in Xenopus.* Biochem Biophys Res Commun, 2002. **294**(1): p. 120-6.

74. Pfleger, C.M. and M.W. Kirschner, *The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1.* Genes Dev, 2000. **14**(6): p. 655-65.

75. Singh, S.A., et al., *Co-regulation proteomics reveals substrates and mechanisms of APC/C-dependent degradation.* EMBO J, 2014. **33**(4): p. 385-99.

76. Hainline, S.G., et al., *The Drosophila MCPH1-B isoform is a substrate of the APCCdh1 E3 ubiquitin ligase complex.* Biol Open, 2014. **3**(7): p. 669-76.

77. Wang, W., T. Wu, and M.W. Kirschner, *The master cell cycle regulator APC-Cdc20 regulates ciliary length and disassembly of the primary cilium.* Elife, 2014. **3**: p. e03083.

78. Liu, Z., et al., *GPS-ARM: computational analysis of the APC/C recognition motif by predicting D-boxes and KEN-boxes.* PLoS One, 2012. **7**(3): p. e34370.

79. Lim, H.J., et al., *The G2/M regulator histone demethylase PHF8 is targeted for degradation by the anaphase-promoting complex containing CDC20.* Mol Cell Biol, 2013. **33**(21): p. 4166-80.

80. Murray, A.W. and M.W. Kirschner, *Dominoes and clocks: the union of two views of the cell cycle.* Science, 1989. **246**(4930): p. 614-21.

81. Hartwell, L.H., *Cell division from a genetic perspective.* J Cell Biol, 1978. **77**(3): p. 627-37.

82. Hartwell, L.H. and T.A. Weinert, *Checkpoints: controls that ensure the order of cell cycle events.* Science, 1989. **246**(4930): p. 629-34.

83. Lee, M. and P. Nurse, *Cell cycle control genes in fission yeast and mammalian cells.* Trends Genet, 1988. **4**(10): p. 287-90.

84. Olson, J.H. and D.E. Chandler, *Xenopus laevis egg jelly contains small proteins that are essential to fertilization.* Dev Biol, 1999. **210**(2): p. 401-10.

85. Bernardini, G., et al., *Xenopus spermatozoon: correlation between shape and motility.* Gamete Res, 1988. **20**(2): p. 165-75.

86. Steinhardt, R.A. and D. Epel, *Activation of sea-urchin eggs by a calcium ionophore.* Proc Natl Acad Sci U S A, 1974. **71**(5): p. 1915-9.

87. Picheral, B. and M. Charbonneau, *Anuran fertilization: a morphological reinvestigation of some early events.* J Ultrastruct Res, 1982. **81**(3): p. 306-21.

88. Sive, H., Grainger, Robert M, Harland, Richard M, *Early Development of Xenopus Laevis: A Laboratory Manual.* 2000: Cold Spring Harbor Laboratory Press.

89. Boveri, T., *Zellen – Studien Heft 2, Die Befruchtung und Teilung des Eles von Ascaris megalocephala II*. 1888, Jena, Germany: Verlag von Gustav Fischer.

90. Gerhart, J., et al., *Cortical rotation of the Xenopus egg: consequences for the anteroposterior pattern of embryonic dorsal development.* Development, 1989. **107 Suppl**: p. 37-51.

91. Scharf, S.R. and J.C. Gerhart, *Determination of the dorsal-ventral axis in eggs of Xenopus laevis: complete rescue of uv-impaired eggs by oblique orientation before first cleavage.* Dev Biol, 1980. **79**(1): p. 181-98.

92. Gerhart, J., M. Wu, and M. Kirschner, *Cell cycle dynamics of an M-phase-specific cytoplasmic factor in Xenopus laevis oocytes and eggs.* J Cell Biol, 1984. **98**(4): p. 1247-55.

93. Ubbels, G.A., et al., *Evidence for a functional role of the cytoskeleton in determination of the dorsoventral axis in Xenopus laevis eggs.* J Embryol Exp Morphol, 1983. **77**: p. 15-37.

94. Fontanilla, R.A. and R. Nuccitelli, *Characterization of the sperm-induced calcium wave in Xenopus eggs using confocal microscopy.* Biophys J, 1998. **75**(4): p. 2079-87.

95. Nusco, G.A., et al., *Ca(2+) response to cADPr during maturation and fertilization of starfish oocytes.* Biochem Biophys Res Commun, 2002. **290**(3): p. 1015-21.

96. Shen, S.S. and W.R. Buck, *Sources of calcium in sea urchin eggs during the fertilization response.* Dev Biol, 1993. **157**(1): p. 157-69.

97. Stricker, S.A., V.E. Centonze, and R.F. Melendez, *Calcium dynamics during starfish oocyte maturation and fertilization.* Dev Biol, 1994. **166**(1): p. 34-58.

98. Whitaker, M., *Calcium at fertilization and in early development.* Physiol Rev, 2006. **86**(1): p. 25-88.

99. Chang, J.B. and J.E. Ferrell, Jr., *Mitotic trigger waves and the spatial coordination of the Xenopus cell cycle.* Nature, 2013. **500**(7464): p. 603-7.

100. Nuccitelli, R., D.L. Yim, and T. Smart, *The sperm-induced Ca2+ wave following fertilization of the Xenopus egg requires the production of Ins(1, 4, 5)P3.* Dev Biol, 1993. **158**(1): p. 200-12.

101. Gilbert, S., *Developmental Biology*. 2000, Sunderland, MA: Sinauer Associates.

102. Lohka, M.J. and Y. Masui, *Effects of Ca2+ ions on the formation of metaphase chromosomes and sperm pronuclei in cell-free preparations from unactivated Rana pipiens eggs.* Dev Biol, 1984. **103**(2): p. 434-42.

103. Kline, D., *Calcium-dependent events at fertilization of the frog egg: injection of a calcium buffer blocks ion channel opening, exocytosis, and formation of pronuclei.* Dev Biol, 1988. **126**(2): p. 346-61.

104. Lorca, T., et al., *Calmodulin-dependent protein kinase II mediates inactivation of MPF and CSF upon fertilization of Xenopus eggs.* Nature, 1993. **366**(6452): p. 270-3.

105. Han, Y., et al., *Ca(2+)-Induced Mitochondrial ROS Regulate the Early Embryonic Cell Cycle.* Cell Rep, 2018. **22**(1): p. 218-231.

106. Shen, S.S. and L.J. Burgart, *Intracellular sodium activity in the sea urchin egg during fertilization.* J Cell Biol, 1985. **101**(2): p. 420-6.

107. Herbert Schuel, R.S., Pramila Dandekar, Jeffrey Boldt, Robert G. Summers, *Sodium Requirements in Hardening of the Fertilization Envelope and Embryonic Development in Sea Urchins.* Biol. Bull. , 1982. **162**: p. 202-213.

108. Horner, V.L. and M.F. Wolfner, *Transitioning from egg to embryo: triggers and mechanisms of egg activation.* Dev Dyn, 2008. **237**(3): p. 527-44.

109. Wyrick, R.E., T. Nishihara, and J.L. Hedrick, *Agglutination of jelly coat and cortical granule components and the block to polyspermy in the amphibian Xenopus laevis.* Proc Natl Acad Sci U S A, 1974. **71**(5): p. 2067-71.

110. Dunbar, B.S., et al., *The mammalian zona pellucida: its biochemistry, immunochemistry, molecular biology, and developmental expression.* Reprod Fertil Dev, 1994. **6**(3): p. 331-47.

111. Burkart, A.D., et al., *Ovastacin, a cortical granule protease, cleaves ZP2 in the zona pellucida to prevent polyspermy.* J Cell Biol, 2012. **197**(1): p. 37-44.

112. Lindsay, L.L. and J.L. Hedrick, *Isolation and characterization of ovochymase, a chymotrypsin-like protease released during Xenopus laevis egg activation.* Dev Biol, 1995. **167**(2): p. 513-6.

113. Bianchi, E., et al., *Juno is the egg Izumo receptor and is essential for mammalian fertilization.* Nature, 2014. **508**(7497): p. 483-7.

114. Danilchik, M. and T. Tumarkin, *Exosomal trafficking in Xenopus development.* Genesis, 2017. **55**(1-2).

115. Wolf, D.P., *The cortical response in Xenopus laevis ova.* Dev Biol, 1974. **40**(1): p. 102-15.

116. Kronja, I., et al., *Widespread changes in the posttranscriptional landscape at the Drosophila oocyte-to-embryo transition.* Cell Rep, 2014. **7**(5): p. 1495-508.

117. Cascio, S.M. and P.M. Wassarman, *Program of early development in the mammal: post-transcriptional control of a class of proteins synthesized by mouse oocytes and early embryos.* Dev Biol, 1982. **89**(2): p. 397-408.

118. Latham, K.E., et al., *Quantitative analysis of protein synthesis in mouse embryos. I. Extensive reprogramming at the one- and two-cell stages.* Development, 1991. **112**(4): p. 921-32.

119. Malkin, L.I., P.R. Gross, and P. Romanoff, *Polyribosomal Protein Synthesis in Fertilized Sea Urchin Eggs: The Effect of Actinomycin Treatment.* Dev Biol, 1964. **10**: p. 378-94.

120. Mandell, J.W., *Immunohistochemical assessment of protein phosphorylation state: the dream and the reality.* Histochem Cell Biol, 2008. **130**(3): p. 465-71.

121. Gurdon, J.B., Wakefield, L., *Microinjection of amphibian oocytes and eggs for the analysis of transcription.*, in *Microinjection and Organelle Transplantation Techniques* A.G. J. E. Celis, A. Loyter, Editor. 1986, Academic Press: London.

122. Sanger, F. and H. Tuppy, *The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates.* Biochem J, 1951. **49**(4): p. 463-81.

123. Sanger, F. and H. Tuppy, *The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates.* Biochem J, 1951. **49**(4): p. 481-90.

124. Mann, M., *The Rise of Mass Spectrometry and the Fall of Edman Degradation.* Clin Chem, 2016. **62**(1): p. 293-4.

125. Cleveland, D.W., et al., *Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis.* J Biol Chem, 1977. **252**(3): p. 1102-6.

126. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

127. Mann, M. and M. Wilm, *Error-tolerant identification of peptides in sequence databases by peptide sequence tags.* Anal Chem, 1994. **66**(24): p. 4390-9.

128. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* J Am Soc Mass Spectrom, 1994. **5**(11): p. 976-89.

129. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry.* Anal Chem, 2003. **75**(17): p. 4646-58.

130. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.

131. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.

132. Browne, T.R., et al., *Kinetic equivalence of stable-isotope-labeled and unlabeled phenytoin.* Clin Pharmacol Ther, 1981. **29**(4): p. 511-5.

133. Hsu, J.L., et al., *Stable-isotope dimethyl labeling for quantitative proteomics.* Anal Chem, 2003. **75**(24): p. 6843-52.

134. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal Chem, 2003. **75**(8): p. 1895-904.

135. McAlister, G.C., et al., *Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses.* Anal Chem, 2012. **84**(17): p. 7469-78.

136. Werner, T., et al., *High-resolution enabled TMT 8-plexing.* Anal Chem, 2012. **84**(16): p. 7188-94.

137. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.

138. McAlister, G.C., et al., *MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes.* Anal Chem, 2014. **86**(14): p. 7150-8.

139. Ting, L., et al., *MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics.* Nat Methods, 2011. **8**(11): p. 937-40.

140. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.* Anal Chem, 2000. **72**(6): p. 1156-62.

141. Makarov, A. and E. Denisov, *Dynamics of ions of intact proteins in the Orbitrap mass analyzer.* J Am Soc Mass Spectrom, 2009. **20**(8): p. 1486-95.

142. Second, T.P., et al., *Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures.* Anal Chem, 2009. **81**(18): p. 7757-65.

143. Ow, S.Y., et al., *iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly".* J Proteome Res, 2009. **8**(11): p. 5347-55.

144. Beausoleil, S.A., et al., *A probability-based approach for high-throughput protein phosphorylation analysis and site localization.* Nat Biotechnol, 2006. **24**(10): p. 1285-92.

145. Lehmann, W.D., et al., *Neutral loss-based phosphopeptide recognition: a collection of caveats.* J Proteome Res, 2007. **6**(7): p. 2866-73.

146. Erickson, B.K., et al., *Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer.* Anal Chem, 2015. **87**(2): p. 1241-9.

147. Kettenbach, A.N. and S.A. Gerber, *Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: application to general and phosphotyrosine-specific phosphoproteomics experiments.* Anal Chem, 2011. **83**(20): p. 7635-44.

148. Thingholm, T.E., et al., *Highly selective enrichment of phosphorylated peptides using titanium dioxide.* Nat Protoc, 2006. **1**(4): p. 1929-35.

149. Fila, J. and D. Honys, *Enrichment techniques employed in phosphoproteomics.* Amino Acids, 2012. **43**(3): p. 1025-47.

150. Solari, F.A., et al., *Why phosphoproteomics is still a challenge.* Mol Biosyst, 2015. **11**(6): p. 1487-93.

151. Olsen, J.V., et al., *Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis.* Sci Signal, 2010. **3**(104): p. ra3.

152. Wu, R., et al., *A large-scale method to measure absolute protein phosphorylation stoichiometries.* Nat Methods, 2011. **8**(8): p. 677-83.

153. Curran, T.G., et al., *MARQUIS: a multiplex method for absolute quantification of peptides and posttranslational modifications.* Nat Commun, 2015. **6**: p. 5924.

154. Wuhr, M., et al., *Deep Proteomics of the Xenopus laevis Egg using an mRNA-Derived Reference Database.* Curr Biol, 2014. **24**(13): p. 1467-75.

155. Presler, M., et al., *Proteomics of phosphorylation and protein dynamics during fertilization and meiotic exit in the Xenopus egg.* Proc Natl Acad Sci U S A, 2017. **114**(50): p. E10838-E10847.

156. Wellner, K., *"A History of Embryology (1959), by Joseph Needham"*, in *Embryo Project Encyclopedia*. 2010. p. http://embryo.asu.edu/handle/10776/2031.

157. Huttlin, E.L., et al., *A tissue-specific atlas of mouse protein phosphorylation and expression.* Cell, 2010. **143**(7): p. 1174-89.

158. Nagaraj, N., et al., *Deep proteome and transcriptome mapping of a human cancer cell line.* Mol Syst Biol, 2011. **7**: p. 548.

159. Beck, M., et al., *The quantitative proteome of a human cell line.* Mol Syst Biol, 2011. **7**: p. 549.

160. Kragl, M., et al., *Cells keep a memory of their tissue origin during axolotl limb regeneration.* Nature, 2009. **460**(7251): p. 60-5.

161. di Prisco, G., et al., *Tracking the evolutionary loss of hemoglobin expression by the white-blooded Antarctic icefishes.* Gene, 2002. **295**(2): p. 185-91.

162. Abzhanov, A., et al., *Bmp4 and morphological variation of beaks in Darwin's finches.* Science, 2004. **305**(5689): p. 1462-5.

163. Thomson, K.S., *An attempt to reconstruct evolutionary changes in the cellular DNA content of lungfish.* J. Exp. Zool., 1972. **180** p. pp. 363-371.

164. Straus, N.A., *Comparative DNA renaturation kinetics in amphibians.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 799-802.

165. McGrath, C.L. and L.A. Katz, *Genome diversity in microbial eukaryotes.* Trends Ecol Evol, 2004. **19**(1): p. 32-8.

166. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.

167. Evans, V.C., et al., *De novo derivation of proteomes from transcriptomes for transcript and protein identification.* Nat Methods, 2012. **9**(12): p. 1207-11.

168. Looso, M., et al., *Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo stable isotope labeling-based mass spectrometry.* Mol Cell Proteomics, 2010. **9**(6): p. 1157-66.

169. Newport, J. and M. Kirschner, *A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription.* Cell, 1982. **30**(3): p. 687-96.

170. Desai, A., et al., *The use of Xenopus egg extracts to study mitotic spindle assembly and function in vitro.* Methods Cell Biol, 1999. **61**: p. 385-412.

171. Wuhr, M., et al., *A model for cleavage plane determination in early amphibian and fish embryos.* Curr Biol, 2010. **20**(22): p. 2040-5.

172. Hughes, M.K. and A.L. Hughes, *Evolution of duplicate genes in a tetraploid animal, Xenopus laevis.* Mol Biol Evol, 1993. **10**(6): p. 1360-9.

173. Goffeau, A., et al., *Life with 6000 genes.* Science, 1996. **274**(5287): p. 546, 563-7.

174. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nat Methods, 2007. **4**(3): p. 207-14.

175. Wuhr, M., et al., *Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster.* Anal Chem, 2012. **84**(21): p. 9214-21.

176. Chavda, V., *A greedy heuristic for the set-covering problem.* Math. Oper. Res., 1979. **4**: p. pp. 233-235.

177. Quackenbush, J., et al., *The TIGR gene indices: reconstruction and representation of expressed gene sequences.* Nucleic Acids Res, 2000. **28**(1): p. 141-5.

178. Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-42.

179. Mills, A.D., et al., *An acidic protein which assembles nucleosomes in vitro is the most abundant protein in Xenopus oocyte nuclei.* J Mol Biol, 1980. **139**(3): p. 561-8.

180. Huang, C.Y. and J.E. Ferrell, Jr., *Ultrasensitivity in the mitogen-activated protein kinase cascade.* Proc Natl Acad Sci U S A, 1996. **93**(19): p. 10078-83.

181. Lee, E., et al., *The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway.* PLoS Biol, 2003. **1**(1): p. E10.

182. Lawo, S., et al., *HAUS, the 8-subunit human Augmin complex, regulates centrosome and spindle integrity.* Curr Biol, 2009. **19**(10): p. 816-26.

183. Zhang, Z., et al., *Recombinant expression, reconstitution and structure of human anaphase-promoting complex (APC/C).* Biochem J, 2013. **449**(2): p. 365-71.

184. Ruepp, A., et al., *CORUM: the comprehensive resource of mammalian protein complexes--2009.* Nucleic Acids Res, 2010. **38**(Database issue): p. D497-501.

185. Yanai, I., et al., *Mapping gene expression in two Xenopus species: evolutionary constraints and developmental flexibility.* Dev Cell, 2011. **20**(4): p. 483-96.

186. Tian, Q., et al., *Integrated genomic and proteomic analyses of gene expression in Mammalian cells.* Mol Cell Proteomics, 2004. **3**(10): p. 960-9.

187. Wang, J., et al., *WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W77-83.

188. Opresko, L.K. and R.A. Karpf, *Specific proteolysis regulates fusion between endocytic compartments in Xenopus oocytes.* Cell, 1987. **51**(4): p. 557-68.

189. Opresko, L., H.S. Wiley, and R.A. Wallace, *Differential postendocytotic compartmentation in Xenopus oocytes is mediated by a specifically bound ligand.* Cell, 1980. **22**(1 Pt 1): p. 47-57.

190. Wallace, R.A. and D.W. Jared, *Studies on amphibian yolk. 8. The estrogen-induced hepatic synthesis of a serum lipophosphoprotein and its selective uptake by the ovary and trasformation into yolk platelet proteins in Xenopus laevis.* Dev Biol, 1969. **19**(5): p. 498-526.

191. Arike, L., et al., *Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli.* J Proteomics, 2012. **75**(17): p. 5437-48.

192. Vogel, C. and E.M. Marcotte, *Label-free protein quantitation using weighted spectral counting.* Methods Mol Biol, 2012. **893**: p. 321-41.

193. Fusaro, V.A., et al., *Prediction of high-responding peptides for targeted protein assays by mass spectrometry.* Nat Biotechnol, 2009. **27**(2): p. 190-8.

194. Low, T.Y., et al., *Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis.* Cell Rep, 2013. **5**(5): p. 1469-78.

195. Menschaert, G., et al., *Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events.* Mol Cell Proteomics, 2013. **12**(7): p. 1780-90.

196. Wessel, D. and U.I. Flugge, *A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids.* Anal Biochem, 1984. **138**(1): p. 141-3.

197. Lohse, M., et al., *RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W622-7.

198. Pertea, G., et al., *TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets.* Bioinformatics, 2003. **19**(5): p. 651-2.

199. Huang, X. and A. Madan, *CAP3: A DNA sequence assembly program.* Genome Res, 1999. **9**(9): p. 868-77.

200. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.

201. Li, W. and A. Godzik, *Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.* Bioinformatics, 2006. **22**(13): p. 1658-9.

202. Wernersson, R., *Virtual Ribosome--a comprehensive DNA translation tool with support for integration of sequence feature annotation.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W385-8.

203. Vizcaino, J.A., et al., *ProteomeXchange provides globally coordinated proteomics data submission and dissemination.* Nat Biotechnol, 2014. **32**(3): p. 223-6.

204. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6940-5.

205. Chaube, R., *Absolute quantitation of post-translational modifications.* Front Chem, 2014. **2**: p. 58.

206. Steen, H., et al., *Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS.* Proc Natl Acad Sci U S A, 2005. **102**(11): p. 3948-53.

207. Tu, J.S.a.D., *The Jackknife and Bootstrap*. 1st ed. 1995: Springer-Verlag.

208. Bednarova, I.P.a.D., *Total Least Squares Approach to Modeling: A Matlab Toolbox.* Acta Montanistica Slovaca, 2010. **15**(2): p. pp. 158-170.

209. Ness, C.-L.C.a.J.V., *Structural and Functional Models Revisited* in *Recent Advances in Total Least Squares Techniques and Errors-In-Variables Modeling* S.V. Huffel, Editor. 1997, Society for Industrial and Applied Mathematics p. 37.

210. Willassen, Y., *Testing hypotheses on the unindetifiable structural parameters in the classical 'errors-in-variables- model with application to Friedman's permanent income model* Economics Letters, 1984. **14**: p. 221-228.

211. Moran, P.A.P., *Estimating structural and functional relationships.* J. Multivariate Anal., , 1971. **1**: p. 232-255.

212. Sharma, K., et al., *Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling.* Cell Rep, 2014. **8**(5): p. 1583-94.

213. Carpy, A., et al., *Absolute proteome and phosphoproteome dynamics during the cell cycle of Schizosaccharomyces pombe (Fission Yeast).* Mol Cell Proteomics, 2014. **13**(8): p. 1925-36.

214. Humphrey, S.J., S.B. Azimifar, and M. Mann, *High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics.* Nat Biotechnol, 2015. **33**(9): p. 990-5.

215. Tsai, C.F., et al., *Large-scale determination of absolute phosphorylation stoichiometries in human cells by motif-targeting quantitative proteomics.* Nat Commun, 2015. **6**: p. 6622.

216. Peshkin, L., et al., *On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development.* Dev Cell, 2015. **35**(3): p. 383-94.

217. Wuhr, M., et al., *The Nuclear Proteome of a Vertebrate.* Curr Biol, 2015. **25**(20): p. 2663-71.

218. Kronja, I., et al., *Quantitative proteomics reveals the dynamics of protein changes during Drosophila oocyte maturation and the oocyte-to-embryo transition.* Proc Natl Acad Sci U S A, 2014. **111**(45): p. 16023-8.

219. Krauchunas, A.R., V.L. Horner, and M.F. Wolfner, *Protein phosphorylation changes reveal new candidates in the regulation of egg activation and early embryogenesis in D. melanogaster.* Dev Biol, 2012. **370**(1): p. 125-34.

220. Guo, H., et al., *Phosphoproteomic network analysis in the sea urchin Strongylocentrotus purpuratus reveals new candidates in egg activation.* Proteomics, 2015. **15**(23-24): p. 4080-95.

221. Zhao, W.M. and G. Fang, *Anillin is a substrate of anaphase-promoting complex/cyclosome (APC/C) that controls spatial contractility of myosin during late cytokinesis.* J Biol Chem, 2005. **280**(39): p. 33516-24.

222. Petersen, B.O., et al., *Cell cycle- and cell growth-regulated proteolysis of mammalian CDC6 is dependent on APC-CDH1.* Genes Dev, 2000. **14**(18): p. 2330-43.

223. Ayad, N.G., et al., *Tome-1, a trigger of mitotic entry, is degraded during G1 via the APC.* Cell, 2003. **113**(1): p. 101-13.

224. Seki, A. and G. Fang, *CKAP2 is a spindle-associated protein degraded by APC/C-Cdh1 during mitotic exit.* J Biol Chem, 2007. **282**(20): p. 15103-13.

225. Liot, C., et al., *APC(cdh1) mediates degradation of the oncogenic Rho-GEF Ect2 after mitosis.* PLoS One, 2011. **6**(8): p. e23676.

226. Li, L., et al., *NuSAP is degraded by APC/C-Cdh1 and its overexpression results in mitotic arrest dependent of its microtubules' affinity.* Cell Signal, 2007. **19**(10): p. 2046-55.

227. Araki, M., et al., *Degradation of origin recognition complex large subunit by the anaphase-promoting complex in Drosophila.* EMBO J, 2003. **22**(22): p. 6115-26.

228. Juang, Y.L., et al., *APC-mediated proteolysis of Ase1 and the morphogenesis of the mitotic spindle.* Science, 1997. **275**(5304): p. 1311-4.

229. Eshleman, H.D. and D.O. Morgan, *Sgo1 recruits PP2A to chromosomes to ensure sister chromatid bi-orientation during mitosis.* J Cell Sci, 2014. **127**(Pt 22): p. 4974-83.

230. Stewart, S. and G. Fang, *Anaphase-promoting complex/cyclosome controls the stability of TPX2 during mitotic exit.* Mol Cell Biol, 2005. **25**(23): p. 10516-27.

231. Rape, M. and M.W. Kirschner, *Autonomous regulation of the anaphase-promoting complex couples mitosis to S-phase entry.* Nature, 2004. **432**(7017): p. 588-95.

232. Feine, O., et al., *Human Kid is degraded by the APC/C(Cdh1) but not by the APC/C(Cdc20).* Cell Cycle, 2007. **6**(20): p. 2516-23.

233. Lu, Y., et al., *Substrate degradation by the proteasome: a single-molecule kinetic analysis.* Science, 2015. **348**(6231): p. 1250834.

234. Holloway, S.L., et al., *Anaphase is initiated by proteolysis rather than by the inactivation of maturation-promoting factor.* Cell, 1993. **73**(7): p. 1393-402.

235. van der Velden, H.M. and M.J. Lohka, *Mitotic arrest caused by the amino terminus of Xenopus cyclin B2.* Mol Cell Biol, 1993. **13**(3): p. 1480-8.

236. Ishihara, K., et al., *Organization of early frog embryos by chemical waves emanating from centrosomes.* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1650).

237. Bischof, J., et al., *A cdk1 gradient guides surface contraction waves in oocytes.* Nat Commun, 2017. **8**(1): p. 849.

238. Rankin, S. and M.W. Kirschner, *The surface contraction waves of Xenopus eggs reflect the metachronous cell-cycle state of the cytoplasm.* Curr Biol, 1997. **7**(6): p. 451-4.

239. Kim, W., et al., *Systematic and quantitative assessment of the ubiquitin-modified proteome.* Mol Cell, 2011. **44**(2): p. 325-40.

240. Barenz, F., et al., *The centriolar satellite protein SSX2IP promotes centrosome maturation.* J Cell Biol, 2013. **202**(1): p. 81-95.

241. Klinger, M., et al., *The novel centriolar satellite protein SSX2IP targets Cep290 to the ciliary transition zone.* Mol Biol Cell, 2014. **25**(4): p. 495-507.

242. Sansal, I., et al., *NPDC-1, a regulator of neural cell proliferation and differentiation, interacts with E2F-1, reduces its binding to DNA and modulates its transcriptional activity.* Oncogene, 2000. **19**(43): p. 5000-9.

243. Lindsay, L.L., J.C. Yang, and J.L. Hedrick, *Identification and characterization of a unique Xenopus laevis egg envelope component, ZPD.* Dev Growth Differ, 2002. **44**(3): p. 205-12.

244. Nagase, H. and E.D. Harris, Jr., *Ovostatin: a novel proteinase inhibitor from chicken egg white. II. Mechanism of inhibition studied with collagenase and thermolysin.* J Biol Chem, 1983. **258**(12): p. 7490-8.

245. Yamada, Y. and K. Aketa, *Ovostatin, an endogenous trypsin inhibitor of sea urchin eggs: purification and characterization of ovostatin from eggs of the sea urchin, Strongylocentrotus intermedius.* Gamete Res, 1988. **19**(3): p. 265-75.

246. The UniProt, C., *UniProt: the universal protein knowledgebase.* Nucleic Acids Res, 2017. **45**(D1): p. D158-D169.

247. Lim, J.C., et al., *Expression and localization of Rdd proteins in Xenopus embryo.* Anat Cell Biol, 2014. **47**(1): p. 18-27.

248. Johansson, M.E., K.A. Thomsson, and G.C. Hansson, *Proteomic analyses of the two mucus layers of the colon barrier reveal that their main component, the Muc2 mucin, is strongly bound to the Fcgbp protein.* J Proteome Res, 2009. **8**(7): p. 3549-57.

249. Gasper, J. and W.J. Swanson, *Molecular population genetics of the gene encoding the human fertilization protein zonadhesin reveals rapid adaptive evolution.* Am J Hum Genet, 2006. **79**(5): p. 820-30.

250. Toyama, B.H., et al., *Identification of long-lived proteins reveals exceptional stability of essential cellular structures.* Cell, 2013. **154**(5): p. 971-82.

251. Shao, H., et al., *Xenopus oocyte meiosis lacks spindle assembly checkpoint control.* J Cell Biol, 2013. **201**(2): p. 191-200.

252. Sun, Q.Y. and H. Schatten, *Regulation of dynamic events by microfilaments during oocyte maturation and fertilization.* Reproduction, 2006. **131**(2): p. 193-205.

253. Bement, W.M., A.M. Sokac, and C.A. Mandato, *Four-dimensional imaging of cytoskeletal dynamics in Xenopus oocytes and eggs.* Differentiation, 2003. **71**(9-10): p. 518-27.

254. Sokac, A.M., et al., *Cdc42-dependent actin polymerization during compensatory endocytosis in Xenopus eggs.* Nat Cell Biol, 2003. **5**(8): p. 727-32.

255. Guttinger, S., E. Laurell, and U. Kutay, *Orchestrating nuclear envelope disassembly and reassembly during mitosis.* Nat Rev Mol Cell Biol, 2009. **10**(3): p. 178-91.

256. Antonin, W., J. Ellenberg, and E. Dultz, *Nuclear pore complex assembly through the cell cycle: regulation and membrane organization.* FEBS Lett, 2008. **582**(14): p. 2004-16.

257. Raices, M. and M.A. D'Angelo, *Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions.* Nat Rev Mol Cell Biol, 2012. **13**(11): p. 687-99.

258. Dultz, E., et al., *Systematic kinetic analysis of mitotic dis- and reassembly of the nuclear pore in living cells.* J Cell Biol, 2008. **180**(5): p. 857-65.

259. Mackay, D.R. and K.S. Ullman, *Coordinating postmitotic nuclear pore complex assembly with abscission timing.* Nucleus, 2011. **2**(4): p. 283-8.

260. Meszaros, N., et al., *Nuclear pore basket proteins are tethered to the nuclear envelope and can regulate membrane curvature.* Dev Cell, 2015. **33**(3): p. 285-98.

261. Laurell, E., et al., *Phosphorylation of Nup98 by multiple kinases is crucial for NPC disassembly during mitotic entry.* Cell, 2011. **144**(4): p. 539-50.

262. Franks, T.M. and M.W. Hetzer, *The role of Nup98 in transcription regulation in healthy and diseased cells.* Trends Cell Biol, 2013. **23**(3): p. 112-7.

263. Cross, M.K. and M.A. Powers, *Nup98 regulates bipolar spindle assembly through association with microtubules and opposition of MCAK.* Mol Biol Cell, 2011. **22**(5): p. 661-72.

264. Sutovsky, P., et al., *Assembly of nuclear pore complexes and annulate lamellae promotes normal pronuclear development in fertilized mammalian oocytes.* J Cell Sci, 1998. **111 (Pt 19)**: p. 2841-54.

265. Backs, J., et al., *The gamma isoform of CaM kinase II controls mouse egg activation by regulating cell cycle resumption.* Proc Natl Acad Sci U S A, 2010. **107**(1): p. 81-6.

266. Yang, Q. and J.E. Ferrell, Jr., *The Cdk1-APC/C cell cycle oscillator circuit functions as a time-delayed, ultrasensitive switch.* Nat Cell Biol, 2013. **15**(5): p. 519-25.

267. Hsu, J.Y., et al., *E2F-dependent accumulation of hEmi1 regulates S phase entry by inhibiting APC(Cdh1).* Nat Cell Biol, 2002. **4**(5): p. 358-66.

268. Lee, G., R. Hynes, and M. Kirschner, *Temporal and spatial regulation of fibronectin in early Xenopus development.* Cell, 1984. **36**(3): p. 729-40.

269. Schuel, H. and R. Schuel, *Benzohydroxamic acid induces polyspermic fertilization in the sea urchin Arbacia punctulata.* Cell Biol Int Rep, 1987. **11**(3): p. 189-96.

270. Penn, A. and B.L. Gledhill, *Acrosomal proteolytic activity of amphibian sperm. A direct demonstration.* Exp Cell Res, 1972. **74**(1): p. 285-8.

271. Dephoure, N. and S.P. Gygi, *A solid phase extraction-based platform for rapid phosphoproteomic analysis.* Methods, 2011. **54**(4): p. 379-86.

272. McDonald, K., *Osmium ferricyanide fixation improves microfilament preservation and membrane visualization in a variety of animal cell types.* J Ultrastruct Res, 1984. **86**(2): p. 107-18.

273. Carlemalm, E., et al., *Low temperature embedding with Lowicryl resins: two new formulations and some applications.* J Microsc, 1985. **140**(Pt 1): p. 55-63.

274. Dimova, N.V., et al., *APC/C-mediated multiple monoubiquitylation provides an alternative degradation signal for cyclin B1.* Nat Cell Biol, 2012. **14**(2): p. 168-76.

275. Fisher, R., *On the interpretation of X^2 from contingency tables, and the calculation of P.* Journal of the Royal Statistical Society, 1922. **85**(1): p. 87-94.

276. Benjamini, Y.H., Yosef, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.

277. Vizcaino, J.A., et al., *2016 update of the PRIDE database and its related tools.* Nucleic Acids Res, 2016. **44**(D1): p. D447-56.

278. Tsai, T.Y., J.A. Theriot, and J.E. Ferrell, Jr., *Changes in oscillatory dynamics in the cell cycle of early Xenopus laevis embryos.* PLoS Biol, 2014. **12**(2): p. e1001788.

279. Meijer, L., et al., *Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdc2, cdk2 and cdk5.* Eur J Biochem, 1997. **243**(1-2): p. 527-36.

280. Papin, C., et al., *XCdh1 is involved in progesterone-induced oocyte maturation.* Dev Biol, 2004. **272**(1): p. 66-75.

281. Chang, H.Y., M. Levasseur, and K.T. Jones, *Degradation of APCcdc20 and APCcdh1 substrates during the second meiotic division in mouse eggs.* J Cell Sci, 2004. **117**(Pt 26): p. 6289-96.

282. O'Farrell, P.H., J. Stumpff, and T.T. Su, *Embryonic cleavage cycles: how is a mouse like a fly?* Curr Biol, 2004. **14**(1): p. R35-45.

283. Edgar, B.A., et al., *Distinct molecular mechanism regulate cell cycle timing at successive stages of Drosophila embryogenesis.* Genes Dev, 1994. **8**(4): p. 440-52.

284. Huang, J. and J.W. Raff, *The disappearance of cyclin B at the end of mitosis is regulated spatially in Drosophila cells.* EMBO J, 1999. **18**(8): p. 2184-95.

285. Dickmann, Z. and R.W. Noyes, *The zona pellucida at the time of implantation.* Fertil Steril, 1961. **12**: p. 310-8.

286. Yoshinaga, K., *A sequence of events in the uterus prior to implantation in the mouse.* J Assist Reprod Genet, 2013. **30**(8): p. 1017-22.

287. Doane, W.W., *Completion of meiosis in uninseminated eggs of Drosophila melanogaster.* Science, 1960. **132**(3428): p. 677-8.

288. Ruggles, K.V., et al., *Methods, Tools and Current Perspectives in Proteogenomics.* Mol Cell Proteomics, 2017. **16**(6): p. 959-981.

289. Weber, J.L., *Interspersed repetitive DNA from Plasmodium falciparum.* Mol Biochem Parasitol, 1988. **29**(2-3): p. 117-24.

290. Chick, J.M., et al., *A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides.* Nat Biotechnol, 2015. **33**(7): p. 743-9.

291. Nguyen, P.A., C.M. Field, and T.J. Mitchison, *Prc1E and Kif4A control microtubule organization within and between large Xenopus egg asters.* Mol Biol Cell, 2018. **29**(3): p. 304-316.

292. Ueda, H.R., et al., *System-level identification of transcriptional circuits underlying mammalian circadian clocks.* Nat Genet, 2005. **37**(2): p. 187-92.

293. Fielenbach, N. and A. Antebi, *C. elegans dauer formation and the molecular basis of plasticity.* Genes Dev, 2008. **22**(16): p. 2149-65.

294. Baker, R.E., S. Schnell, and P.K. Maini, *A clock and wavefront mechanism for somite formation.* Dev Biol, 2006. **293**(1): p. 116-26.

295. Imayoshi, I., et al., *Oscillatory control of factors determining multipotency and fate in mouse neural progenitors.* Science, 2013. **342**(6163): p. 1203-8.

296. Collart, C., et al., *Titration of four replication factors is essential for the Xenopus laevis midblastula transition.* Science, 2013. **341**(6148): p. 893-6.

297. Holtfreter, J., *A study of the mechanics of gastrulation. Part 1.* Journal of Experimental Zoology, 1943. **94**(3): p. 261-318.

298. Smith, L.D. and R.E. Ecker, *Uterine suppression of biochemical and morphogenetic events in Rana pipiens.* Dev Biol, 1970. **22**(4): p. 622-37.

299. Wiblet, M. and E. Baltus, *[Protein kinase of ovary, ovarian oocytes and oocytes during maturation and ovulation in vitro of Xenopus laevis].* Arch Int Physiol Biochim, 1973. **81**(4): p. 813.

300. Huff, R.P., J.T., *PRODUCTION OF A CLEAVAGE-INITIATING FACTOR BY ARTIFICIALLY ACTIVATED EGGS OF RANA PIPIENS.* Texas Journal of Science, 1965. **17**(2): p. 206.

301. Smith, J.C. and J.E. Howard, *Mesoderm-inducing factors and the control of gastrulation.* Dev Suppl, 1992: p. 127-36.

302. Smith, J.C., et al., *The Xenopus animal pole blastomere.* Bioessays, 1987. **7**(5): p. 229-34.

303. Nieuwkoop, P.D., A.G. Johnen, and B. Albers, *The epigenetic nature of early chordate development : inductive interaction and competence*. Developmental and cell biology series. 1985, Cambridge Cambridgeshire ; New York: Cambridge University Press. ix, 373 p.

304. He, B., et al., *Apical constriction drives tissue-scale hydrodynamic flow to mediate cell elongation.* Nature, 2014. **508**(7496): p. 392-6.

305. Lillie, F.R., *Differentiation without cleavage in the egg of the annelid Chaetopterus pergamentaceous* Wilhelm Roux Arch. Entwicklungsmech.Org., 1902. **14**(3-4): p. 477-499.

306. Kobayakawa, Y. and H.Y. Kubota, *Temporal pattern of cleavage and the onset of gastrulation in amphibian embryos developed from eggs with the reduced cytoplasm.* J Embryol Exp Morphol, 1981. **62**: p. 83-94.

307. Chulitskaia, E.V., *Desynchronization of cell divisions in the course of egg cleavage and an attempt at experimental shift of its onset.* J Embryol Exp Morphol, 1970. **23**(2): p. 359-74.

308. Takagi, M., T. Shimoda, and A. Shinagawa, *Dependence of the timing system regulating the onset of gastrulation on cytoplasmic, but not nuclear, activities in the Xenopus embryo.* Dev Growth Differ, 2005. **47**(6): p. 415-22.

309. Malacinski, G.M., H.M. Chung, and B. Youn, *Further characterization of the effects of ultraviolet irradiation of the amphibian egg.* Experientia, 1978. **34**(7): p. 883-4.

310. Kwan, K.M. and M.W. Kirschner, *A microtubule-binding Rho-GEF controls cell morphology during convergent extension of Xenopus laevis.* Development, 2005. **132**(20): p. 4599-610.

311. Mita, I. and N. Satoh, *Timing of gastrulation in fused double-embryos formed from eggs with different cleavage schedules in the starfish, Asterina pectinifera.* J Exp Zool, 1982. **223**(1): p. 67-74.

312. Rugh, R., *The frog; its reproduction and development*. 1952, Philadelphia,: Blakiston. x, 336 p.

313. Howe, J.A. and J.W. Newport, *A developmental timer regulates degradation of cyclin E1 at the midblastula transition during Xenopus embryogenesis.* Proc Natl Acad Sci U S A, 1996. **93**(5): p. 2060-4.

314. Carruthers, S., J. Mason, and N. Papalopulu, *Depletion of the cell-cycle inhibitor p27(Xic1) impairs neuronal differentiation and increases the number of ElrC(+) progenitor cells in Xenopus tropicalis.* Mech Dev, 2003. **120**(5): p. 607-16.

315. Gerhart, J., *The epigenetic nature of vertebrate development: an interview of Pieter D. Nieuwkoop on the occasion of his 70th birthday.* Development, 1987. **101**: p. 653-657.