Communicating Research to the General Public

At the March 5, 2010 UW-Madison Chemistry Department Colloquium, Prof. Bassam Z. Shakhashiri, the director of the Wisconsin Initiative for Science Literacy (WISL), encouraged all UW-Madison chemistry Ph.D. candidates to include a chapter in their Ph.D. thesis communicating their research to non-specialists. The goal is to explain the candidate's scholarly research and its significance to a wider audience that includes family members, friends, civic groups, newspaper reporters, program officers at appropriate funding agencies, state legislators, and members of the U.S. Congress.

Over 40 Ph.D. degree recipients have successfully completed their theses and included such a chapter.

WISL encourages the inclusion of such chapters in all Ph.D. theses everywhere through the cooperation of Ph.D. candidates and their mentors. WISL is now offering additional awards of $250 for UW-Madison chemistry Ph.D. candidates.

Wisconsin Initiative for Science Literacy

The dual mission of the Wisconsin Initiative for Science Literacy is to promote literacy in science, mathematics and technology among the general public and to attract future generations to careers in research, teaching and public service.

UW-Madison Department of Chemistry
1101 University Avenue
Madison, WI 53706-1396
Contact: Prof. Bassam Z. Shakhashiri
bassam@chem.wisc.edu
www.scifun.org

Case studies of cryptic proteins contributing to shape change in eosinophils

Workflows to investigate STAT3, LOCGEF and NHSL2

By

Keren B. Turton

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biomolecular Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 8/1/2018

The dissertation is approved by the following members of the Final Oral Committee:

Deane F. Mosher, Professor, Biomolecular Chemistry/Medicine

David A. Brow, Professor, Biomolecular Chemistry

Joshua J. Coon, Professor, Biomolecular Chemistry/Chemistry

Lixin Rui, Assistant Professor, Medicine

Michael D. Sheets, Professor, Biomolecular Chemistry

Chapter 7: How do cells of the immune system change shape and travel to sites where needed? Leveraging proteomics data to investigate eosinophil morphology

- for the Wisconsin Initiative for Science Literacy

1. Where do eosinophils fit in the immune system?

The human body is made up of hundreds of different types of cells, all with the same DNA but various functions and appearances. Cells can have distinct functions because of a process called differentiation, whereby a combination of specific proteins specify a multipotent stem cell, capable of becoming one of many different types of cell, along a certain trajectory. This is analogous to all the cells in the body being given a long instruction manual (the DNA), but each being instructed to read different, overlapping parts of it. Certain parts of the DNA are copied to RNA transcripts, which are in turn translated into proteins.

The cells of the immune system are called leukocytes, or white blood cells. Most of them are specified from stem cells in the bone marrow. Some white blood cells reside in tissues while others are found in the bloodstream or lymphatic system. One way of classifying components of the immune system is according to whether they elicit an innate or adaptive response. The **innate immune system** is responsible for first-line defense against invading pathogens. These pathogens could be viruses, bacteria or helminths (parasitic worms). Usually, the

TERMS

Adaptive immune system: specific response to pathogens, coordinated by antibodies

Cytokines: a group of small signaling molecules, some of which specify, activate and signal white blood cells (eg. Interleukins)

Cytoskeleton: protein scaffold that gives a cell its shape

Eosinophil: a white blood cell with toxic proteins stored in granules. It has a bi-lobed nucleus

Innate immune system: general first-line defense against pathogens, including white blood cells releasing toxic proteins and engulfing invaders

Neutrophil: a white blood cell with toxic granule proteins, and ability to engulf and digest pathogens. It has a multi-lobed nucleus

Pathogen: bacterial, viral or micro-organism that causes disease

Proteomics: study of the protein complement of cells, often by mass spectrometry

Rho GTPase: regulates proteins that regulate the cytoskeleton

Rho GEF: guanine nucleotide exchange factor – this protein increases the rate of GDP dissociation from GTPase, enhancing its activity

T helper cell type 2 (Th2): a white blood cell that coordinates signaling to recruit several types of other white blood cells, including eosinophils

body is able to mount an immediate, non-specific response through the innate immune system. This first line of defense is the equivalent of strafing: certain immune cells release toxic proteins which damage invaders, but also release signaling molecules (called **cytokines**) to recruit other immune cells. This may lead to inflammation and localized tissue damage as many other cells join the battle. Inflammation can lead to pain and swelling.

The cells primarily responsible for first-line innate immune defense are granulocytes (cells with collections of toxic proteins stored in granules), which includes **neutrophils** and **eosinophils**. Neutrophils are far more abundant than eosinophils, and in addition to releasing toxic proteins, they can engulf and digest (phagocytose) invading bacteria and viruses. Eosinophils have a supporting role, releasing both proteins that can digest viral nucleic acids, and also Major Basic Protein which is toxic to bacteria and helminths.

If given time to regroup and strategize, the body is able to mount tailored defenses against particular pathogens. These responses are more specific, and often more effective. This is part of the **adaptive immune system**. The adaptive immune system includes B cells, which preserve a memory of particular pathogens in surface **antibodies**, enabling rapid recognition and response in situations of re-infection. This is why vaccines are useful: if non-functional influenza virus proteins are injected into the bloodstream, B cells can be trained to recognize these proteins prior to exposure with the actual virus. This means the body will be able to mount a much quicker and more specific immune response.

T helper type 2 (Th2) cells, which coordinate other white blood cells, can recruit eosinophils to an anti-helminth immune response. They also release particular **cytokines** (signaling proteins) which promote development of eosinophils.

2. What are eosinophils (and why study them)?

As described above, eosinophils are part of the first-line defense against helminths in humans. They also contribute to antimicrobial defense by releasing toxic granule proteins. Mice, frogs, newts and zebrafish also all have eosinophil-like cells.

Although they can help immune response, eosinophils also play a role in certain immune disorders, like asthma. Eosinophils are normally resident in the gastrointestinal tract, spleen and lymph nodes. In asthma, these cells move into the lungs because of inappropriate signaling from Th2 cells. Release of toxic granule proteins leads to further tissue damage. Asthma is characterized by inflammation of the bronchi (airways in the lung), and can be life-threatening. It is usually triggered by allergens, which are erroneously recognized as hazardous by the immune system.

Eosinophils have a unique collection of proteins stored in granules. These proteins stain bright red in the presence of eosin, an acidic dye. The eosinophil gets its name from "eosin" + "phile" (Greek for *loving*). One of these proteins is Major Basic Protein, which is cytotoxic (kills cells). Another is RNase II, a protein which can digest viral RNA (the information storage molecule). The proteins are carefully packaged in granules, so that they can be easily released when needed, but don't harm the eosinophil during storage.

In most cells of the body, DNA is compartmentalized in a roughly spherical membrane-bound sac called the nucleus. Eosinophils have a distinctive bi-lobed nucleus (**Figure 7-1**), which looks like a pair of ear muffs. Neutrophils also have unusually-shaped nuclei, and these morphologies are thought to help the cells move out of blood vessels and into tissues. The process of moving between the endothelial cells that line the wall of the blood vessel is called extravasation.

To extravasate, eosinophils must change shape. They are initially prepared for extravasation and granule protein release by **cytokine** (signaling molecule) stimulation. Interleukin-5 (IL5) is a

cytokine that binds to receptors on the surface of the eosinophil. This event leads to a chain of signaling events, culminating in changes to cell shape and the positioning of the nucleus. This is accomplished by modifying proteins. These modifications change the set of proteins that can interact, leading to different functional outcomes.

Eosinophils can adopt a number of shapes. In our lab, we refer to these shapes as "balls", "acorns" and "pancakes". Eosinophils change from spherical (balls) to acorn-shaped following cytokine stimulation, with the nucleus pushed to the pointed rear of the cell (termed the "nucleopod"). Eosinophils can also flatten out on a surface (pancakes). We study eosinophils because of their clinical significance, but also because they are a useful model for researching rapid cell shape change and reorganization of cellular components (specifically the nucleus and granules).

3. How do cells change shape and move?

The **cytoskeleton** is the scaffold that determines a cell's shape, size and movement. A collection of regulatory proteins organize the scaffold proteins. There are three major (and several minor) scaffold proteins that together comprise the cytoskeleton. These are actin, microtubules and intermediate filaments. Having many protein components enables versatility in cell shape, and also means that dynamics of cell movement can be coordinated across the cell.

White blood cells migrate in response to protein signals in the surrounding fluid. There will be a gradient in the concentration of the signal, meaning the cell can decide whether to move toward or away from a signal. Eosinophils polarize (changing shape from balls to acorns) and migrate toward IL5, and other interleukins. Other factors (like substrate stiffness) can also influence cell movement. The substrate is the surface cells move across.

A migrating cell is polarized: there is a front edge and rear. The force which enables movement is generated at the rear of a migrating white blood cell. This is analogous to squeezing a tube of toothpaste: rearward squeezing pushes the contents of the cell forward. In eosinophils, the nucleus is propelled to the rear, and might contribute to force generation. Most white blood cells move by protruding arm-like projections at the front edge of the cell. These projections are full of a cytoskeleton structural protein called actin, and actin assembly is coordinated by **Rho GTPases** (discussed further below).

In a way, white blood cell movement is a bit like rolling a partially-deflated ball across a sticky surface. The surface is selectively "sticky" in that proteins on the surface of the cell's protrusions recognize proteins on the substrate and attach to them. Inflammation changes some of the proteins on the substrate surface, enhancing attachment of white blood cells.

The cell must coordinate attachment at the front of the moving cell, and detachment at the rear. Compressive forces at the rear of the cell are transmitted through the cytoskeletal scaffold, allowing forward protrusion along a sticky surface. Migration is also dependent on the substrate itself: a stiff substrate enables cells to stay attached for longer but also means slower migration. Regulatory proteins are crucial because they integrate information about the substrate, cell shape and signals to coordinate movement in the cell.

Eosinophils can crawl at a speed of about 0.8 mm/h (about 1.5 cell lengths per minute). Cells travel more quickly in the bloodstream, where they are part of the flow, not crawling along a substrate. White blood cells crawl along the endothelium (blood vessel wall) before squeezing between endothelial cells to migrate into tissues.

4. Why investigate the proteins STAT3, LOC100996504 and NHSL2?

In collaboration with Joshua Coon's lab on campus, we analyzed proteins in eosinophils before and after stimulation with cytokines. As described, cytokines elicit a rapid and profound shape change in eosinophils. Emily Wilkerson used mass spectrometry to compare the proteins (and their modifications) in cells before and after 5 minutes stimulation with cytokine interleukin-5. In her WISL thesis chapter[1], Wilkerson explains how mass spectrometry works, and gives an overview of the project. Briefly, proteins are composed of long chains of amino acid building blocks. There are 20 different amino acids, and the sequence of these residues in a protein will determine function. Different residues lead to proteins folding into different shapes, with differing properties for the residues, enabling them to perform unique functions.

The investigative approach the Coon lab used entails breaking open eosinophils, cutting the proteins into smaller pieces, ionizing the fragments and measuring their mass-to-charge ratios (m/z) using a mass spectrometer. From these data, they determined the relative abundances of about 7,000 different proteins in eosinophils. They were also able to catalogue changes in protein modifications. These datasets were foundational for determining potential proteins involved in eosinophil shape change. The most abundant protein in eosinophils is actin, one of the building blocks of the cytoskeleton.

We chose to investigate STAT3 protein because we found evidence of a variant form with a very small difference in the protein sequence. We wanted to know if the difference could be quantified. Our attention was drawn to a protein designated as "LOC100996504" in the dataset. The amount of this protein in eosinophils appeared to decrease after stimulation. The "LOC" part

[1] Wilkerson, E.M. (2017) *Chapter 5 A proteomics primer for non-specialists: for the Wisconsin Initiative for Science Literacy*. In: *Method optimization and application of mass spectrometry to the field of hematology* (thesis).

of this name is short for locus, followed by a number string – basically, it was a placeholder name for a protein of unknown function. We investigated this protein, and another protein called NHSL2, that nobody has studied before. We were curious about NHSL2 because within 5 minutes of IL5 stimulation it becomes highly modified, and migrates to the nucleopod of eosinophils.

5. Do cytokines influence the composition of protein variants in eosinophils? – STAT3

We investigated Signal transducer and activator of transcription 3 (STAT3) because it is an important middleman in cell signaling. External messages (like interleukin signals) communicate with cells by binding to receptors, triggering a relay of signals. Certain proteins act as signaling hubs, integrating disparate stimuli and coordinating cell responses. These responses can include morphology change, but also transcription of DNA. DNA is transcribed to RNA transcripts with the help of proteins called transcription factors. RNA is translated to protein. This process is how information coded in DNA is made tangible as molecules capable of performing cellular functions. STAT3 is one of the key transcription factors in eosinophils. It comes in two major forms, called STAT3α (alpha) and STAT3β (beta). They mostly have the same sequence, but STAT3α has a stretch of 55 amino acids that β lacks. This difference is due to a process called splicing: the *STAT3* gene (DNA) is transcribed to messenger (m)RNA. This RNA message has non-coding sequences, and parts of it can be cut out and the sequence stuck back together. This process (alternative splicing) is a key part of creating a repertoire of proteins with subtly different functions in cells.

When inspecting the eosinophil mass spectrometry data, we found that a form of STAT3 with *another* minor sequence variation (one amino acid residue missing) was present in eosinophils. The missing amino acid is a Serine, a type of amino acid that can be modified by adding a group called a phosphate. This addition changes the activity of some proteins. It's uncertain if it affects STAT3's function. We called the types of STAT3 variant S (with Serine) or ΔS (delta S, lacking

Serine). Was this difference significant? Did it only occur in the STAT3α form, or also in the β? Because the difference was too small to detect at the protein level, we decided to develop a method to distinguish the four STAT3 variants' RNA transcripts (Sα, Sβ, ΔSα and ΔSβ) and quantify their relative abundances. This method is called quantitative polymerase chain reaction (qPCR). We found that all four transcripts are present in eosinophils. It looked like the relative proportions were fairly consistent even when eosinophils were stimulated with various cytokines. Cytokine stimulation increased overall levels of STAT3.

6. ***What are the cell-specific regulatory features that enable eosinophils to change shape or move in unique ways? – LOC100996504***

Rho GTPase proteins regulate actin building blocks (which scaffold cell shape) and other cytoskeleton proteins. Rho GTPases have active and inactive forms. GTPases must be active to regulate actin. In turn, Rho GTPases are themselves regulated by three types of protein. The regulatory complement is analogous to a traffic light:

1) Proteins that enhance activation of GTPases - green
2) Proteins that slow down GTPase activity - orange
3) Proteins that deactivate GTPases - red

There are three Rho GTPases in eosinophils, and a host of regulatory proteins. Why are so many regulatory proteins needed? It is likely that the diversity enables *tissue-specific regulation*, since different cell types have different sets of regulators, and also different requirements for cell movement and internal trafficking of cellular components. Unique features of regulatory proteins enable *spatial* and *temporal* regulation in cells.

Our project focuses on a particular protein that enhances GTPase activation (green), RhoGEF18. We found a variant that is unique to white blood cells. Much of our work focused on demonstrating that this variant protein exists.

The eosinophil protein catalogue from the mass spectrometry project included the oddly-named "LOC100996504" protein (hereafter "LOC"). We were curious about its role in eosinophils because activation of the cells led to an apparent decrease of ~60% of LOC in the space of 5 minutes. By looking at the sequence of this protein and comparing it with translated DNA sequences, we were able to predict that LOC wasn't a discrete entity. Instead, it was fused to the front of a protein called RhoGEF18. We called this form of the protein "LOCGEF". We searched sequences from other species and validated that LOCGEF is conserved (the sequence is highly similar) in mice, frogs and other animals.

We used a variety of techniques to prove to our satisfaction that the fused protein exists in eosinophils and showed that LOCGEF is found only in white blood cells. We also used antibodies targeting RhoGEF18, or parts of LOC to investigate where the protein is found in eosinophils. Antibodies are able to specifically recognize particular proteins. It is possible to conjugate fluorescent dyes to antibodies so that the antibodies can be used for detection of proteins inside cells. We found that LOCGEF is dispersed in resting cells, but moves towards the nucleopod (the side of the cell with the nucleus) in activated eosinophils. If LOCGEF is white blood cell-specific, and involved in cell movement, it is a potential pharmacological target for diseases with inappropriate white blood cell migration, like asthma, allergies or leukemia.

7. How can scientists investigate completely unknown proteins? – NHSL2

NHSL2 stands for Nance-Horan Syndrome protein-like 2. Nobody had published anything about this protein when we started studying it, meaning that it was poorly-characterized. The

process of characterizing a protein often begins with the basics: working out the sequence of amino acids that make up the protein. The sequence may have a resemblance to sequences of well-studied proteins, which helps in generating hypotheses about its function. It is usually easier to work out the RNA sequence, which is translated to protein. Just like the proteome (all the proteins in a cell) can be determined using mass spectrometry, the transcriptome (all the RNA sequences) can be determined by using a method called RNA-Seq. There are many publicly available transcript datasets. We used these to come up with possible sequences for NHSL2. Our predicted sequences differed from sequences listed in online databases, so we sought to verify them using PCR (making many copies of specific nucleotide sequences) and sequencing the amplified products.

Once we were satisfied that we knew the sequences of the two major variants in eosinophils (NHSL2α and NHSL2β), we cut and pasted the sequences into circles of DNA called plasmids. Plasmids can be introduced into bacteria, and will instruct the bacteria to make the encoded protein. Being able to make protein in bacteria helps increase yield of that protein, and it means one can use the protein for experiments, after working out how best to enhance protein production. We were initially interested in NHSL2 because of the apparent increase in modifications to its amino acids after stimulation with IL5. Modifications like these in other proteins impact protein function and location in cells. Our bacterially-expressed proteins were used as size standards in experiments to see if the eosinophil's NHSL2 proteins are the same size as NHSL2α and β, which would corroborate our predictions about the unknown protein.

We determined that a highly-modified form of NHSL2 migrates to the nucleopod of activated eosinophils. We have not determined NHSL2's role in this process, but we suspect it has a role in regulating eosinophil shape change based on its significantly-altered state and localization.

8. Concluding remarks and outcomes of research

It is thanks to these large-scale datasets (like proteomics) that we can identify poorly-studied proteins as candidate regulatory proteins.

The STAT3 assay could be used to investigate the proportions of STAT3 variants in other cell types. STAT3 is hyperactive in many blood cancers. Our work with Lixin Rui's lab suggests that the levels of STAT3 S-β and STAT3 ΔS-β might be higher in certain types of lymphoma. The principles of assay development could be used in the design of assays to investigate other subtle transcript variants.

In demonstrating that LOCGEF is found in white blood cells, we developed an array of tools that might be useful for future work. We cloned and expressed LOCGEF and the shorter form of RhoGEF18. These could be used in assays to determine whether they differently regulate GTPases. We developed tools for expressing LOCGEF in other cell types. If it is only found in white blood cells, will there be an effect if it were made in liver cells?

Our work describing and characterizing NHSL2 has introduced it as a potential cytoskeletal regulator. Our work could serve as an example for how to go about characterizing previously-unknown proteins. The bacterially-expressed NHSL2 proteins can be used to generate antibodies for NHSL2 detection experiments in human cells. The proteins could also be used to determine which proteins NHSL2 interacts with, which is helpful in determining a protein's function.

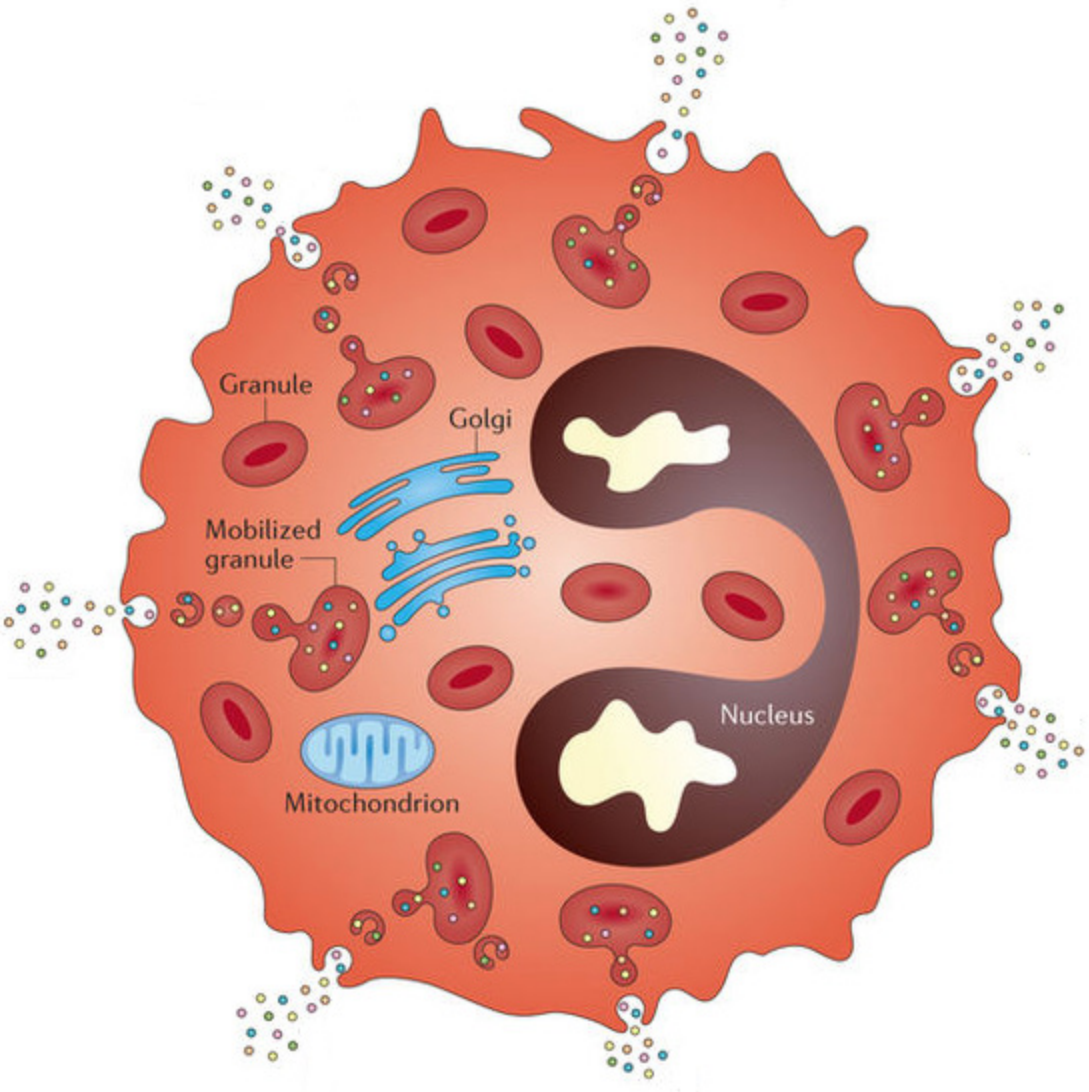

Figure 7-1. Schematic of an eosinophil, with bi-lobed nucleus and granules secreting proteins. Image modified from: Weller, P & Spencer, L. 2017. Functions of tissue-resident eosinophils. *Nature Reviews Immunology*. **17:**746-60.