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 50 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

Investigating the O-GlcNAc cycling enzymes: substrate recognition, enzymatic activity, and biological functions

By

Dacheng Fan

A dissertation submitted in partially fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Pharmaceutical Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

2020

Date of final oral examination: 7/15/2020

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

Jiaoyang Jiang, Associate Professor, School of Pharmacy

Lingjun Li, Professor, School of Pharmacy and Department of Chemistry

Charles Lauhon, Associate Professor, School of Pharmacy

Jennifer Golden, Assistant Professor, School of Pharmacy

Wei Xu, Professor, Oncology

Chapter 7. A magic candy for proteins: protein modification by a unique sugar unit

For Wisconsin Initiative for Science Literacy

I have written this chapter to explain my research to a broad, non-scientific audience. The ultimate goal of conducting scientific research is for benefiting the general public. Thus, it is imperative that we communicate our research with others. What are the subjects we are studying? Why are they worth being studied? And what we have learnt from the study? I want to thank the Wisconsin Initiative for Science Literacy at UW-Madison for providing this opportunity, and for sponsoring and supporting the creation of this chapter, especially Professor Bassam Shakhashiri for his encouragement and guidance during the process, and Editor Elizabeth Reynolds for her help on revising this chapter.

What are proteins?

Proteins are essential building blocks to our bodies. In many ways, they are similar to Legos: Lego models are built by assembling pieces of bricks, and proteins are made by linking amino acids, which are the small structural units of proteins (**Figure 7.1**). There are twenty different amino acids to build tens of thousands of human proteins[1]. Determined by gene, the amino acid sequence of each protein is unique and defines the identity of the protein. Just as Lego models can consist of either ten pieces of bricks or a thousand pieces of bricks, the sizes of proteins vary from dozens of amino acids to tens of thousands of amino acids[1]. Instead of arranging bricks into a straight line, Lego models are designed with certain shapes to create a specific structure. Proteins also have three-dimensional shapes resulting from the proper folding of the long amino acid chains (**Figure 7.1**). These shapes are vital for the normal

functions of proteins.

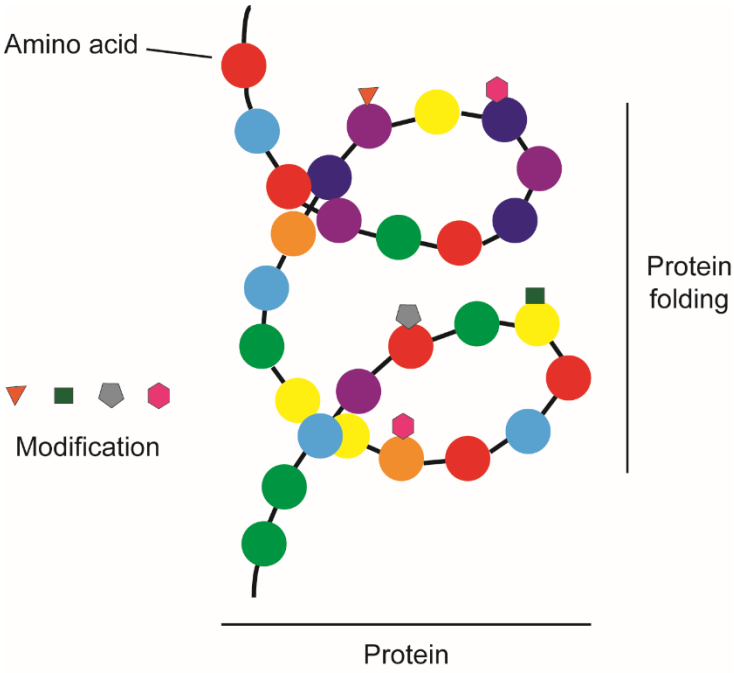

Figure 7.1. A cartoon illustration of a protein. Proteins are made by linking different amino acids (circles with different colors) and folding into a three-dimensional structure. Proteins can be modified by chemical groups on certain amino acids (see Section "**What is protein modification?**").

Why are proteins important?

Proteins are critical for living organisms: they carry out functions such as building muscle structures and serve as a major nutrient source. On the cellular and molecular level, proteins play a crucial role in almost all biological processes. Each protein has designated duties in a cell, just as people are assigned various specific jobs within a large company. Below are a few examples.

Metabolic reactions which convert food to energy for maintaining the normal physiology of cells are performed by a subgroup of proteins that are capable of accelerating reactions called enzymes. The breakdown of glucose to produce energy for cells, for instance, involves ten enzymes working sequentially in a pipeline. There are also transporter proteins responsible for the proper localization of particular materials. They work like vehicles to transport materials to the desired locations within a cell.

Certain proteins serve as the building blocks to construct and maintain the structure of cells. Proteins are also largely responsible for the storage and transmission of genes. Extremely long DNA chains are tightly, yet orderly, packed in the cell nucleus by a unique protein named histone, and unwound and replicated for the newly divided cell again by a number of other proteins.

Besides the above examples, proteins perform many more biological processes, with a highly sophisticated division of responsibilities. Combined together, the diversified proteins cooperate perfectly with each other to maintain the physiology of cells and organisms under normal conditions.

What is protein modification?

Besides the amino acid sequence and the three-dimensional structure, proteins have another important characteristic called protein modification. In protein modification, specific amino acids of proteins are modified by certain chemical groups (**Figure 7.1**), just like putting stickers onto particular bricks of a Lego model. Various types of protein modifications exist, and they can occur on different amino acids of the same protein simultaneously.

While stickers on Lego models are mainly for decoration, protein modifications possess much more significance. Again, think of proteins in a cell as people in a large company, where many employees in the company are issued with particular cards to access certain workplaces for their job while others do not need access cards to work. The same principle applies to proteins, as some proteins are ready to function as soon

as they are produced, and other proteins require particular modifications to perform their duties in cells. These modifications could affect the stability, localization, and function of the proteins, thereby regulating a variety of biological functions in cells and playing important roles in a number of diseases[2], making them popular research targets for understanding the biology for certain diseases, and for developing treatments for these diseases.

Which type of protein modification does our lab study?

Of the various types of protein modifications, our lab is especially interested in a unique modification with the sugar unit O-linked *N*-acetylglucosamine (O-GlcNAc /ō-'glo͞ok,nak/) that derives from glucose and contains carbon, hydrogen, oxygen, and nitrogen atoms (**Figure 7.2a**). This modification is termed O-GlcNAcylation (/ō-,glo͞oknasə'lāsh(ə)n/), and the modified proteins are called protein substrates, which are protein molecules upon which enzymes act to attach the modification.

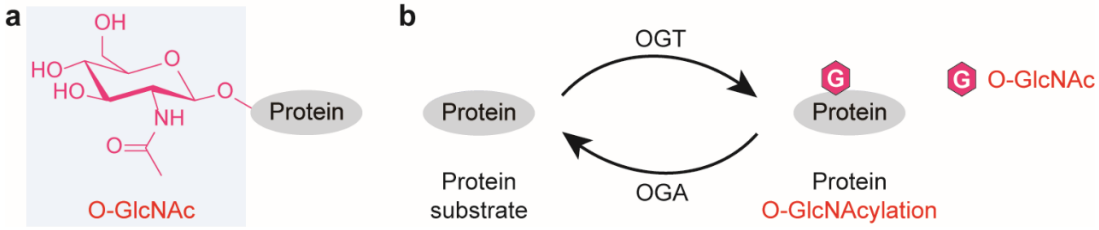

Figure 7.2. Illustration of protein O-GlcNAc modification. (**a**) The chemical structure of O-GlcNAc linked to a protein. (**b**) Protein O-GlcNAcylation is regulated by OGT and OGA. OGT is responsible for modifying protein substrates with the sugar unit O-GlcNAc, while OGA removes the sugar modification off protein substrates.

Why are we interested in O-GlcNAcylation?

Multiple molecules of the same protein substrate exist simultaneously within a

cell, and the modification happens on some of the protein substrate molecules while leaving other molecules unmodified. The relative ratio of the modified protein substrate to the entirety of this protein (modified + unmodified) is defined as the modification level of this protein substrate. For example, if there are ten molecules of protein X in a cell, among which three are modified while the remaining seven are not, the modification level of protein X is 30%. The modification levels of different proteins can vary significantly, and each should be kept at a certain range to maintain proper cell function and organism health. The overall modification levels of the whole protein population needed for the normal physiology is called homeostasis.

Given the fact that thousands of human proteins involved in diversified biological processes are modified by the sugar unit O-GlcNAc[3], the homeostasis of this sugar modification on the entire protein population is crucial for cells and organisms to maintain normal physiological conditions, and its dysregulation is closely related to a number of the most deleterious diseases including diabetes, cardiovascular disorder, neurodegenerative diseases, and different types of cancers[3]. For example, this modification is highly involved in insulin function, and aberrant modification could affect insulin resistance[4]. In terms of neurodegenerative diseases, this sugar modification on a particular protein could limit Alzheimer's disease progression[5], giving birth to a promising drug candidate that is currently in early clinical trials[6]. This modification on another protein could block its gathering, which is an important contributor for Parkinson's disease[5]. Furthermore, this sugar modification on the entire protein population generally promotes tumor growth and metastasis, and its

dysregulation is found in almost all types of cancers[7]. With all of the evidence above, insights into the underlying mechanism through which the homeostasis of this sugar modification is achieved could inspire novel treatments for such diseases.

How is O-GlcNAc modification regulated?

Despite the large population of protein substrates, only a single pair of proteins regulate this dynamic sugar modification: O-GlcNAc transferase (OGT) modifies its protein substrates with the sugar unit, and O-GlcNAcase (OGA) removes the sugar unit off protein substrates[3] (**Figure 7.2b**). You can think of this as OGT putting stickers onto the Lego model and OGA peeling those stickers off.

Out of thousands of protein substrates, how do OGT and OGA recognize and decide which proteins to modify? And to what levels do they modify these proteins in order to achieve the homeostasis of this sugar modification? Answering these questions would provide valuable information for better learning the relationship between the modification and certain diseases, as well as for further developing potential diagnostic and therapeutic approaches. Unfortunately, the exact mechanism through which OGT and OGA regulate this modification remains largely elusive.

Our lab tries to reveal the mechanism of regulating this sugar modification primarily through three aspects: substrate recognition, enzymatic activity, and biological functions (**Figure 7.3**). Biological function means the functions of this sugar modification on substrate proteins. The other two terms, substrate recognition and enzymatic activity, are explained below.

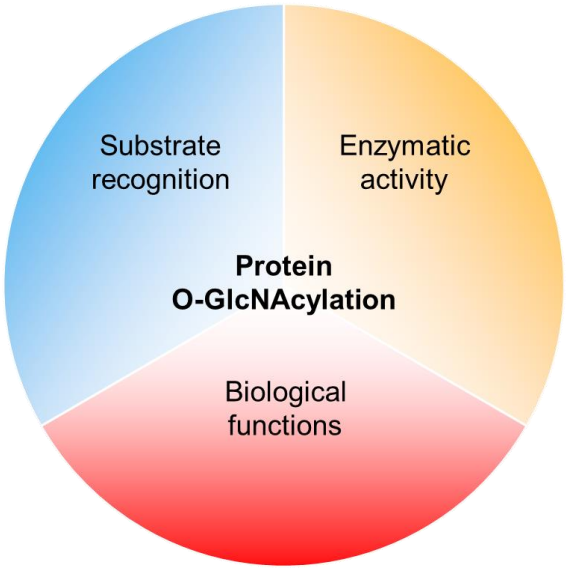

Figure 7.3. Our lab studies protein O-GlcNAcylation through three aspects: substrate recognition, enzymatic activity, and biological functions.

What is substrate recognition? People recognize each other through multiple aspects including visual, verbal, and physical contacts. Proteins, on the other hand, recognize each other by making direct contact[8]. But how exactly do proteins do this? Think of connecting two Lego models together: you only need to directly link a few key bricks of both models instead of linking every brick of one model to the other. Similarly, proteins recognize each other through the contacting of certain critical amino acids (**Figure 7.4**). Discovering key amino acids of OGT and OGA for their substrate recognition would provide us clues to the mechanism with which they recognize their protein substrates. This would help us understand how OGT and OGA distinguish their various protein substrates to control the homeostasis of this modification on the entire protein population.

Generally speaking, enzymatic activity is the ability of an enzyme to accelerate the reaction, such as OGT's ability to modify its protein substrates with the sugar unit. Ideally, we would like to develop chemicals that alter OGT activity in cells. Then, we could analyze the effect of treatment with those chemicals, which could provide us with

valuable information about OGT's functions. Eventually, this could help us develop novel therapies for the diseases in which OGT plays an important role.

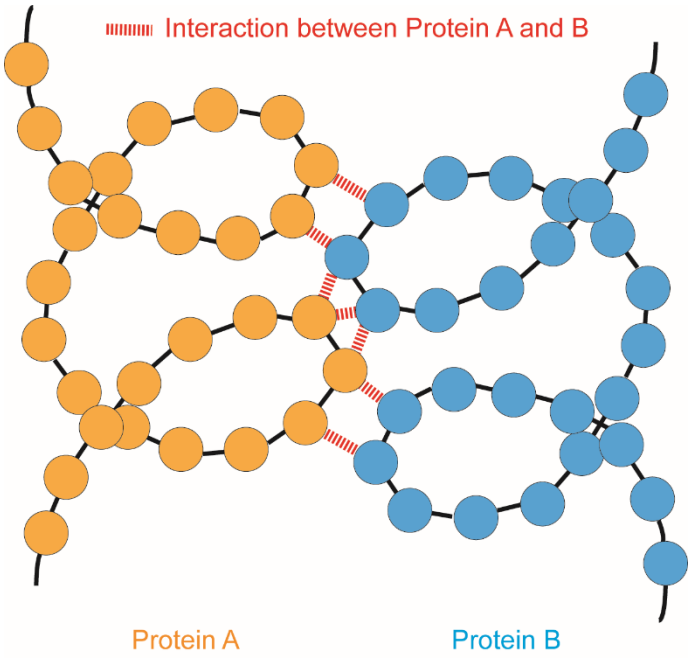

Figure 7.4 Proteins recognize each other through the contacting of certain amino acids. Protein A (in yellow color) and Protein B (in blue color) recognize each other through the interaction between several amino acids (red dash line).

What have I learnt about O-GlcNAcylation during my PhD study?

During my PhD study, I have worked on several projects in collaboration with my colleagues. We have developed a reliable, rapid, and economical experimental method to identify amino acids of OGT that are important for its substrate recognition[9]. We have applied this novel technique and discovered two amino acids that are essential for OGT interacting with different protein substrates with varied significance[9]. Considering OGT is a protein that contains more than 1,000 amino acids, this assay offered an efficient option to quickly screen those amino acids involved in the substrate recognition of OGT. In another ongoing study focusing on a single amino acid of OGA which is mutated in a number of different human cancer tissues, we found the amino acid affects a small population of proteins by altering their O-GlcNAcylation levels

and/or their interaction with OGA. This disrupts the normal stability and cellular functions of certain proteins. This particular study would reveal insights into the relationship between the sugar modification and cancer, and may inspire relative clinical applications.

In an effort to manipulate the enzymatic activity of OGT, we developed an OGT inhibitor that could specifically target OGT and inhibit its activity in cells[10]. This inhibitor not only serves as a tool to study the function of OGT in cells, but also provides a new strategy to study other enzymes similar to OGT.

Another ongoing project aims to study the biological functions of this sugar modification on OGA, the enzyme that regulates this modification. It is interesting to find out how this sugar modification affects OGA, and as a result, how OGA reacts to this effect to regulate this modification as a feedback. In addition to this particular sugar modification, the OGA protein also bears other types of modifications in human cells. Since each type of protein modification has its own functional significance, we are curious about the interplay between different types of modifications on OGA. We would like to learn how the different modifications on OGA affect each other, how they affect OGA functions in cells, and consequently, how they influence cells. We expect this study to offer new knowledge about the regulation of OGA and O-GlcNAc.

References

1. Ponomarenko, E. A. *et al*. The size of the human proteome: The width and depth. *Int. J. Anal. Chem.*, (2016).

2. Duan, G., & Walther, D. The roles of post-translational modifications in the context

of protein interaction networks. *PLoS Comput. Biol.* **11**, (2015).

3. Hardiville, S. & Hart, G. W. Nutrient regulation of signaling, transcription, and cell physiology by O-GlcNAcylation. *Cell Metab.* **20**, 208–213 (2014).

4. Ma, J. & Hart, G. W. Protein O-GlcNAcylation in diabetes and diabetic complications. *Expert Rev. Proteomics* **10**, 365–380 (2013).

5. Wani, W. Y., Chatham, J. C., Darley-Usmar, V., McMahon, L. L., & Zhang, J. O-GlcNAcylation and neurodegeneration. *Brain Res. Bull.* **133**, 80-87 (2017).

6. Selnick, H. G. *et al*. Discovery of MK-8719, a potent O-GlcNAcase inhibitor as a potential treatment for tauopathies. *J. Med. Chem.* **62**, 10062-10097 (2019).

7. Singh, J. P., Zhang, K., Wu, J. & Yang, X. O-GlcNAc signaling in cancer metabolism and epigenetics. *Cancer Lett.* **356**, 244–250 (2015).

8. De Las Rivas, J., & Fontanillo, C. Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput. Biol.* **6**, (2010).

9. Hu, C. W. *et al*. Electrophilic probes for deciphering substrate recognition by O-GlcNAc transferase. *Nat. Chem. Biol.* **13**, 1267-1273 (2017).

10. Worth, M. *et al*. Targeted covalent inhibition of O-GlcNAc transferase in cells. *Chem. Commun.* **55**, 13291-13294 (2019).