Communicating Research to the General Public

The WISL Award for Communicating PhD Research to the Public launched in 2010, and since then over 100 Ph.D. degree recipients have successfully included 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—as well as their excitement for and journey through their area of study—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.

WISL encourages the inclusion of such chapters in all Ph.D. theses everywhere, through the cooperation of PhD candidates, their mentors, and departments. WISL offers awards of $250 for UW-Madison Ph.D. candidates in science and engineering. Candidates from other institutions may participate, but are not eligible for the cash award. WISL strongly encourages other institutions to launch similar programs.

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.

Contact: Prof. Bassam Z. Shakhashiri

UW-Madison Department of Chemistry

bassam@chem.wisc.edu

www.scifun.org

Substrate Multiplexed Screening Guides Development of Biocatalysts for Friedel-Crafts Alkylation of Aromatic Amino Acids

By

Peyton M. Higgins

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

University of Wisconsin-Madison

2023

Date of final oral examination: 7th July 2023

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

Andrew R. Buller, Professor, Chemistry

Tina Wang, Professor, Chemistry

Thomas Record, Emeritus Professor, Biochemistry

Amy Weeks, Professor, Biochemistry

Chapter 5

Evolving Enzymes to Aid Drug Discovery

Chapter 5: Evolving Enzymes to Aid Drug Discovery

5. 1. Introduction

One of the most common questions I'm asked when I say I'm in graduate school is "what do you study?" I never seem to answer this question the same way twice – how I describe my research depends entirely on the person who asked the question. I share a common scientific language with others who work in my field (protein engineering), but figuring out how to effectively discuss my research with others who haven't learned this particular dictionary of jargon has been a fun challenge. I opted to write this chapter to present my graduate research in the most accessible format I could. My goal was to explain the concepts that underlie my research projects in an approachable manner so that readers with any amount of scientific background could understand what I've been working on for the past five years (and why). These efforts were supported by the Wisconsin Initiative for Science Literacy (WISL) at UW-Madison. In particular, I'd like to thank Professor Bassam Shakhashiri, Elizabeth Reynolds, and Cayce Osborne for their guidance and support.

5. 2. Where do new medications come from?

The development of new medications has completely revolutionized modern medicine, from early antibiotics like penicillin to groundbreaking new cancer therapies or drugs that manage chronic health conditions. But where do new medications come from? Many drugs come directly from Nature. The antibiotic penicillin was originally discovered in a sample of mold, and a malaria-treating chemical called quinine was isolated from the bark of *Cinchona* trees in South America.[1] While naturally occurring molecules can serve as a promising starting point in pharmaceutical research, there are often modifications we can make to the structure of these chemicals to make them safer or more effective.

The process to develop a new drug typically follows three main phases (**Figure 1**).[2] In the initial discovery step, medicinal chemists make and screen thousands of different chemicals to search for those that are the best drug candidates. Making new chemicals can be a complex and demanding task. Organic chemists have spent centuries researching methods to synthesize chemicals, including most of the medications that are prescribed today. Medicinal chemists rely on these methods to synthesize large libraries of chemical compounds that are tested for their biological activity.

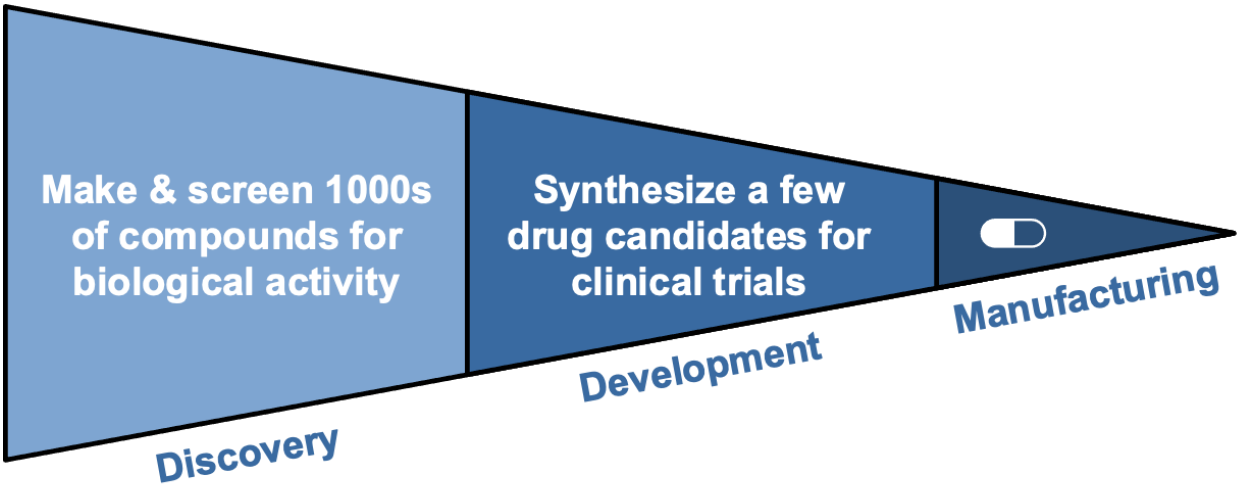

Figure 1. Development of new medications. Typically, thousands of chemicals are tested to eventually identify one that is suitable for clinical use.

Once a promising lead compound is identified, that chemical is put through increasingly rigorous tests meant to assess safety and efficacy in cellular or animal model systems. When we're lucky, a drug candidate progresses to clinical trials, where it is tested in human volunteers. At this stage, process chemists at pharmaceutical companies search for the most efficient method to synthesize a lead compound. Their task is to make large quantities of a single compound, rather than small amounts of many different compounds.

Much of the research conducted by organic chemists in academic labs is designed to support pharmaceutical research. By developing new methods for synthesizing chemicals, we can expand the pool of options medicinal chemists can sample from when developing new drugs.

New synthetic methods may also be useful to process chemists, especially when those methods are environmentally friendly and non-toxic. I've spent the past five years studying how we can use tools from biology to improve chemical synthesis methods.

5. 3. What are enzymes, and what are they used for?

As humans, we tend to think about DNA as affecting things we can observe about ourselves, like hair color or height. DNA contains all the instructions for life, including many things we can't see and typically don't have to think about. Our cells work nonstop to stay alive. They obtain energy from sugar and other nutrients. They defend themselves from toxic substances. They communicate with one another through chemical signals. All of these functions are ultimately guided by DNA. Since keeping cells alive is an extremely complex task, our genome (all the DNA in a cell) is complex too. You can think of DNA as an encyclopedia – it contains all the information you might need, but it's also unwieldy to carry around with you all the time.

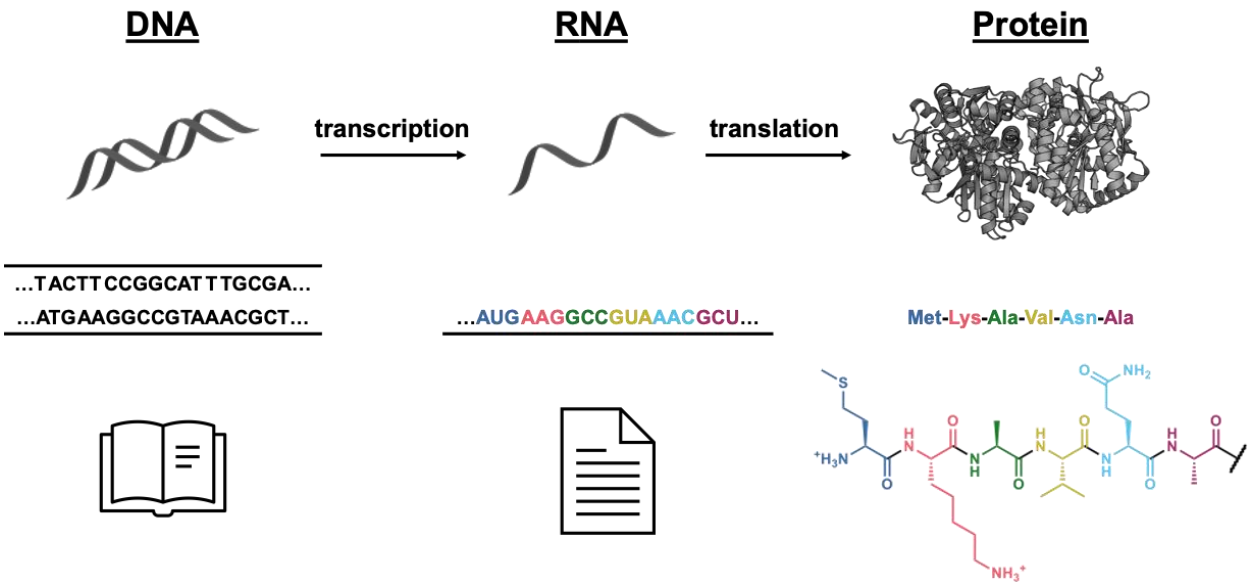

Figure 2. The central dogma of biology. DNA and RNA are used to store and transmit information using a four-letter code. RNA sequences are translated to form proteins out of 20 different amino acid building blocks. The structure of a protein determines its function.

Cells use a process called transcription to convert sections of DNA into a RNA, much like copying a single article out of an encyclopedia so you can refer to it easily when you want to learn about a particular topic (**Figure 2**). Both DNA and RNA transmit information; the four-letter codes that make up their sequences provide instructions for the next step in the central dogma of biology: translation into proteins. Instead of four letters, proteins are made up of 20 different building blocks, called amino acids. Each of these amino acids has a different chemical structure; when they're linked together to form a protein, that protein adopts a three-dimensional structure. The structure of a protein determines its function in a cell. Some proteins have structural roles and help cells keep their shape. Some proteins transport chemicals into or out of cells. Others protect or produce copies of DNA. Proteins that are missing or function incorrectly can cause various diseases, such as cystic fibrosis or sickle cell anemia. The key takeaway here is that the **sequence** of DNA affects the **structure** of the corresponding protein, which also affects that protein's **function**.

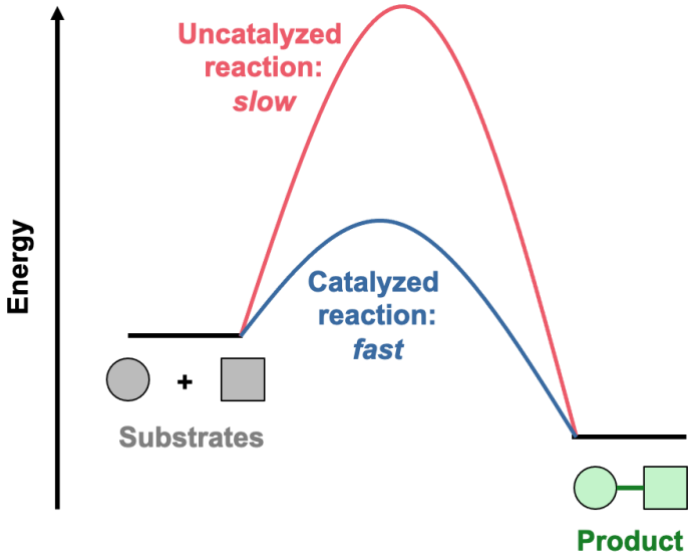

Figure 3. A catalyst increases reaction rate by reducing the amount of energy required to convert substrate(s) into a product.

My research has focused on a subset of proteins called enzymes. Enzymes are catalysts for chemical reactions within cells. Essentially, an enzyme binds to one or more chemicals, called

substrates, and reduces the energy required for a chemical reaction to take place (**Figure 3**). The same reaction might still be possible in the absence of the enzyme, but the enzyme's job is to speed up the rate of the reaction by positioning substrates in an orientation that promotes a desired reaction. Non-enzyme catalysts are also used by organic chemists for synthetic methods. In my research, I combine ideas from organic chemistry and biology to develop enzymes as biocatalysts.

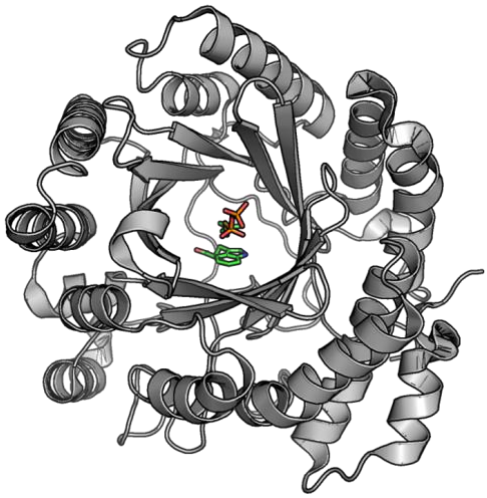

Figure 4. A cartoon-style model of an enzyme (named FgaPT2) is shown in gray. Two substrates (orange and green) are bound in the tunnel-like active site of the enzyme, which promotes formation of a new chemical bond between the two substrates.

There are several reasons enzymes make especially useful catalysts when compared to traditional organic catalysts.[3] Most organic catalysts are only active when dissolved in flammable or carcinogenic solvents, but enzymes operate in water. When two substrates are mixed together, there may be multiple types of bonds that can form between them to generate a product. A poorly selective catalyst could increase the rate of formation for several of those different bonds, resulting in a mixture of products. Enzymes are often highly selective, since they can use their three-dimensional structure to precisely position substrates in an "active site" to guide those substrates to form a single product (**Figure 4**). You can think of enzymes like machines in a factory; Nature has designed them to perform a particular function in a precise and reproducible way. When multiple enzymes work together, much like an assembly line, they can build some

truly impressive and complex molecules! Finally, enzymes make great catalysts because we can use the connection between sequence, structure, and function to change the behavior of enzymes, a process called directed evolution. We can use directed evolution to create new enzymes that catalyze the formation of new chemical products.

5. 4. What is directed evolution, and how does it relate to natural evolution?

Every time cells divide to produce new cells, they must make a copy of their entire DNA sequence. Every so often, the proteins that replicate DNA make a mistake, called a mutation. Sometimes mutations can be harmful and result in cancerous cells. At other times, a mutation might prove beneficial. Mutations that improve an organism's "fitness" are passed on to future generations. This gradual process of change caused by accumulation of beneficial mutations is called evolution through natural selection. A successful organism in this context isn't necessarily the smartest, strongest, or most advanced, but simply an organism that is well-adapted to its particular environment and therefore capable of passing its DNA down to offspring.

Humans have relied on principles of evolution for millennia by using artificial selection to develop new breeds of domesticated species like pigeons, sheep, or dogs. Development of the many agricultural crops we now rely on was also accomplished with artificial selection. In this context, "fitness" is defined by the humans selecting the successful organisms. We can select the most cold-hardy tomatoes or the friendliest puppies to gradually alter a species over the course of many generations.

Earlier I discussed how a gene's sequence affects the structure and function of the corresponding protein. By mutating a gene, we can change the way an enzyme functions, potentially granting it the ability to catalyze a new reaction (**Figure 5**). Unfortunately, natural evolution is slow. We don't want to spend decades (or more!) waiting for spontaneous mutations to accumulate. Fortunately, many methods for rapidly introducing mutations in the lab have been

developed. Some of these mimic natural evolution, where mutations arise randomly throughout a gene. Other methods target a specific site in a gene, which can be useful if we know that a particular portion of an enzyme interacts directly with substrates.

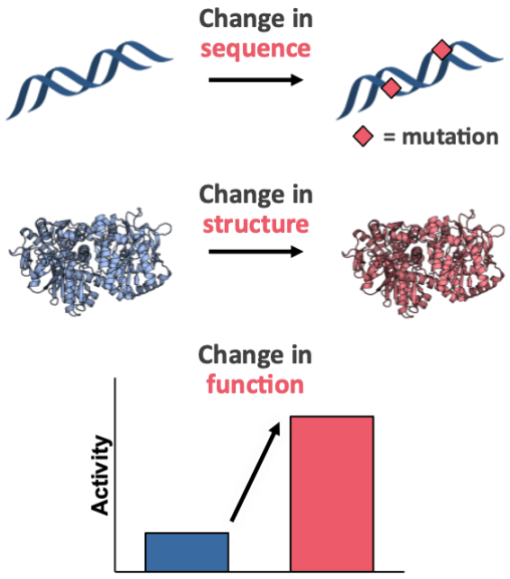

Figure 5. Underlying principles of directed evolution. By introducing changes to the gene for a particular enzyme, we can change that enzyme's function and increase its activity for catalyzing a desired chemical reaction.

While we know that DNA sequence and enzyme function are closely linked, enzymes are far too complex for us to reliably predict *how* a change in sequence will affect function. Instead, we need to make lots of different changes to a gene sequence and screen the resulting enzyme variants to assess how they've changed. There are many ways to tailor the screening process, and we have a saying in our field that "you get what you screen for."[4] If you want to select for an enzyme that's capable of catalyzing a specific reaction, you should screen the enzyme variants for their ability to produce the corresponding product. If you want to select for an enzyme that is stable at high temperatures, you need to measure the activity of the enzyme variants at high temperatures.

I was interested in screening for a special kind of enzyme that's especially useful for synthesizing new chemicals. Let's imagine a long hallway full of locked doors. Behind each door

is a different chemical product you're interested in making, and the key to unlock each door is an enzyme capable of forming that product. The most efficient way to access all those products is to find an enzyme that works like a master key and can unlock many doors. Extending the analogy, you might find an enzyme key that can open a door if you really force it or take a long time, but the goal of evolution is to find keys that work smoothly. Most enzymes have evolved naturally to unlock only one door, since enzymes that react with too many chemicals could have negative consequences for a cell. When we instead use enzymes in the lab to synthesize chemicals, we want them to provide access to as many different products as possible. We say that a catalyst capable of converting many similar substrates into their corresponding products has a large "substrate scope." When I joined my lab in 2018, the field of biocatalysis hadn't thought much about how we could intentionally screen for enzymes that have a large substrate scope. That was the challenge my graduate research aimed to tackle.

5. 5. How can substrate multiplexed screening evolve enzymes with a larger substrate scope?

To identify enzymes with a large substrate scope, we needed to develop a screening method that tested enzyme activity on multiple related substrates at once. We used an approach called substrate multiplexed screening (SUMS), where multiple substrates are added to a single reaction with each enzyme variant (**Figure 6**). We can analyze these reactions to measure the amount of product formed from each substrate. We're looking for enzymes that work like master keys, by unlocking activity with multiple substrates to form their corresponding products.

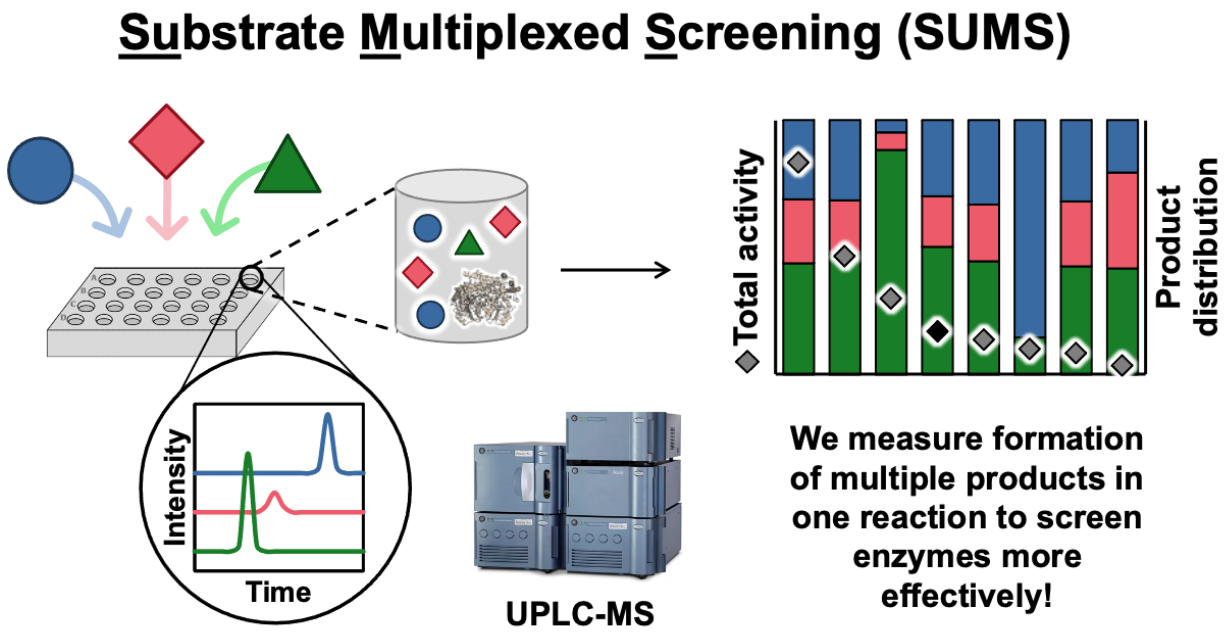

Figure 6. Overview of Substrate Multiplexed Screening (SUMS). We combine multiple substrates with each enzyme variant and measure the resulting products to assess the total activity and substrate scope of every enzyme in a screening library.

I used an enzyme called TrpB to show that substrate multiplexed screening can help us evolve enzymes with better substrate scopes. TrpB catalyzes the formation of a bond between two substrates, indole and serine. The product of this reaction is called tryptophan. Tryptophan is one of the twenty amino acids that are used to build proteins (and so is serine). If we instead use substrates that look like indole, but with extra atoms attached, in reactions with TrpB, we can generate products that are similar to tryptophan but may have different biological properties. For example, serotonin is a neurotransmitter and psilocin is a hallucinogenic drug; both of these molecules are made in nature by modifying tryptophan (**Figure 7**).

Figure 7. TrpB produces the standard amino acid tryptophan, which can be converted into related compounds such as serotonin and psilocin. Subtle structural differences dramatically alter the biological activity of these compounds.

Thanks to decades of study, we already know a lot about how TrpB functions. We know its three-dimensional structure and how it binds to substrates, which has aided past evolution efforts to expand the substrate scope of TrpB. Often, testing a new method on a system that is well-studied can make it easier to troubleshoot the new method. A lot of work goes into optimizing experimental setup to generate reproducible results, so it's helpful to already have a rough idea of what results to expect if an experiment is working. After optimizing substrate multiplexed reactions with a variant of TrpB named 2B9, I introduced mutations to the gene and screened the resulting enzyme variants to look for changes in how they functioned. One variant, named I102T after the mutation it possesses, behaved more like a master key by improving activity on multiple substrates, which was exactly what we were hoping to identify with substrate multiplexed screening!

While looking for a master key is the most straightforward way to expand the substrate scope of an enzyme, it's not the only way. Let's say we have a parent enzyme that unlocks three doors very well, and we identify an enzyme variant that doesn't work as well for those three products but unlocks a new door farther down the hall. That evolved enzyme has an altered substrate scope, also called a "change in specificity," that can prove beneficial for future steps in evolution.

Our starting enzyme, 2B9, preferred to form two products, called 2-Me-Trp and 7-Cl-Trp, and formed only small amounts of the other potential products (**Figure 8**). We identified another

mutation, H275R, that shifted the product distribution so that a different product, DIT, was favored instead. H275R was unlocking a door that had been relatively tricky for 2B9 to access! Unfortunately, the H275R mutation decreased the activity for all of the other reactions, so it wasn't a particularly useful catalyst. Since the H275R mutation affected activity on different substrates to different extents, we inferred that H275R was directly affecting how substrates bound to the enzyme and reacted. Our hypothesis was supported when we checked the location of the mutation in the enzyme's structure and found that it was in a part of the enzyme that plays a role in changing the shape of the enzyme to accommodate substrates and products.

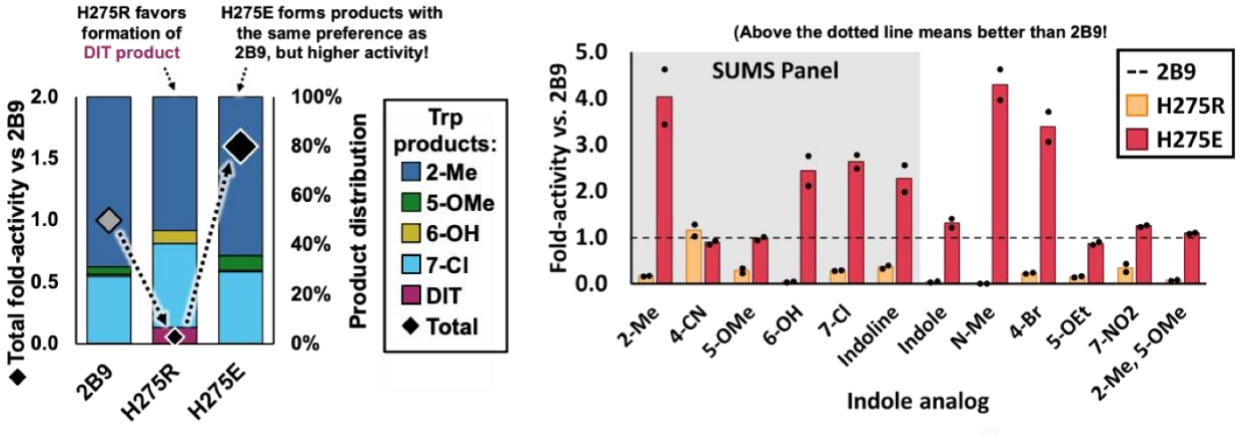

Figure 8. H275R is less active overall, but we were interested in it because its substrate scope was different from 2B9 (left). A different mutation at the same position, H275E, was more active than 2B9 across a large substrate scope (right).

We decided to try other mutations at the same site and found that one mutation, H275E, did improve activity with multiple substrates! Thanks to substrate multiplexed screening, we found a deactivating mutation (H275R) that affected substrate binding and used it as a stepping stone to uncover a master key variant with improved activity across many substrates. The catalytically useful H275E variant was teamed up with another enzyme called *Rgn*TDC (which was evolved by my labmate Dr. Allwin McDonald) to synthesize a variety of chemicals that resemble serotonin.[5]

Since that first project, I and others in my lab have continued using substrate multiplexed screening to investigate enzymes that catalyze other useful chemical reactions. I used the products from TrpB reactions to study a family of enzymes called prenyltransferases, which are

involved in the biological synthesis of many different antibiotic, antiparasitic, and anticancer chemicals. Prenyltransferases could be used to further modify tryptophan-containing chemicals to tune their biological activity for drug discovery. Our lab's ultimate goal is to develop enzymes with large substrate scopes that can be used to synthesize a large variety of chemicals that may prove useful in our search for new medications. My graduate research studying the use of substrate multiplexed screening for directed evolution has helped our lab to pursue that goal.

5. 6. References

1. Christensen, S. B. Natural Products That Changed Society. *Biomed. 2021, Vol. 9, Page 472* **9**, 472 (2021).

2. France, S. P., Lewis, R. D. & Martinez, C. A. The Evolving Nature of Biocatalysis in Pharmaceutical Research and Development. *JACS Au* (2022) doi:10.1021/JACSAU.2C00712/ASSET/IMAGES/LARGE/AU2C00712_0037.JPEG.

3. Sharma, S., Das, J., Braje, W. M., Dash, A. K. & Handa, S. A Glimpse into Green Chemistry Practices in the Pharmaceutical Industry. *ChemSusChem* vol. 13 2859–2875 (2020).

4. Buller, A. R. *et al.* Directed evolution mimics allosteric activation by stepwise tuning of the conformational ensemble. *J. Am. Chem. Soc.* **140**, 7256–7266 (2018).

5. McDonald, A. D., Higgins, P. M. & Buller, A. R. Substrate multiplexed protein engineering facilitates promiscuous biocatalytic synthesis. *Nat. Commun.* **13**, 5242 (2022).