Belmont University

Belmont Digital Repository

Belmont University Research Symposium (BURS)

Special Events

2022

Synthesis of Selenotryptophan for Protein Elucidation

James P. Reiselman
Belmont University, parker.reiselman@pop.belmont.edu

Follow this and additional works at: https://repository.belmont.edu/burs

Part of the Amino Acids, Peptides, and Proteins Commons, Heterocyclic Compounds Commons, Inorganic Chemicals Commons, Organic Chemicals Commons, and the Polycyclic Compounds Commons

Recommended Citation

Reiselman, James P., "Synthesis of Selenotryptophan for Protein Elucidation" (2022). *Belmont University Research Symposium (BURS)*. 106.
https://repository.belmont.edu/burs/106

This Oral Presentation is brought to you for free and open access by the Special Events at Belmont Digital Repository. It has been accepted for inclusion in Belmont University Research Symposium (BURS) by an authorized administrator of Belmont Digital Repository. For more information, please contact repository@belmont.edu.

Synthesis of [6,7]SeTrp for use in Multiple Isomorphous Replacement X-ray Crystallography

Protein structure, function, and elucidation techniques

A protein's function is fundamentally linked to its structure allowing them to carry out different chemical reactions depending on their substrate and their role in biochemical pathways. Structurally, they are primarily made of amino acid peptide chains which then are packed into sub-units that oligomerize into a larger quaternary structure, which is the primary mode for which proteins function by undergoing conformational change[5] (Figure 1). With many different types of proteins ranging from enzymes, hormonal proteins, antibodies, storage proteins, transport proteins, etc. there is a large range of structural differences that are simply too complex to model with chemical techniques traditionally used to determine simpler molecules.

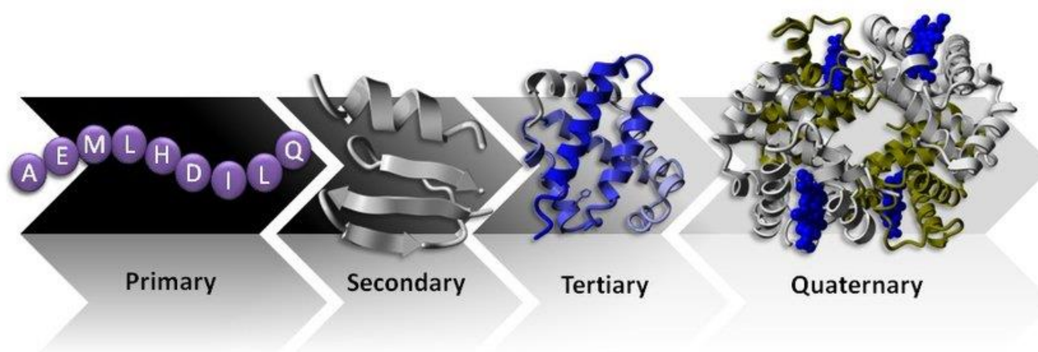


Figure 1. Levels of protein structure. Notice the secondary alpha helices and beta sheets forming the larger quaternary structure.

Even so, the process of initial protein analysis typically involves mass spectrometry, melting point and denaturation analysis, and peptide mass fingerprinting to link a specific peptide chain to previously identified proteins in organisms from a large database to document chemical properties of the protein in question[5]. The amino acid sequence of a protein informs how it is structured, and because the structure determines the function, a database with previously mapped amino acid sequences is commonly used to help identify certain aspects of the protein's structure[4]. NMR spectroscopy can be useful in determining the structure of smaller proteins (under 20 kDa) by combining the data with prior knowledge of the amino acid sequence to reliably determine the 3-D structure[4]. However, the amino acid sequence is not always known, and membrane proteins have trouble forming a stable crystal structure, therefore light-scattering analyses like Batch Dynamic Light Scattering (DLS + SLS) have been used to elucidate the folding patterns of proteins in purified solutions of the protein[5]. X-ray crystallography is now the modern and preferred method for structural analysis because it produces a useful electron density map that can be computationally processed into a 3-D structure instead of dense graphical data from NMR or DLS that has room for resolution and interpretational problems[4]. A concentrated beam of x-rays is shot into the crystal protein and the diffraction pattern data can be used to calculate atomic positions in the crystal with great accuracy[4]. Once the crystal has been formed, its structure can be determined, but large quantities of protein is needed and it is often intensely time consuming to find the proper conditions for crystal formation, especially when working with membrane integrated proteins[4].

The X-ray Crystallography Phase Problem and MIR

Within the large and growing field of x-ray crystallography, there are certain problems that arise due to the intrinsic properties of the crystal being measured. One notable problem is with the measured diffraction data obtained from the x-rays bouncing off the crystal structure back into the detector. Sometimes, these signals can overlap with each other and destructively interfere making the signal weaker and difficult to map[6] (Figure 2a). The smaller the protein, the easier it is to mathematically deduce the information lost within the interfered diffraction data by relating stronger phases with each other. Furthermore, the absence of diffraction patterns form certain portions of the protein can also be a big clue as to the structural position and with all this data, an electron density map can be constructed for determining the protein's structure[6]. As the number of atoms grows, however, the relationships between the phases also become weaker due to the increased complexity of the protein[6]. Molecular replacements in the protein and phase improvements can help cut through the noise made by overly complex or weak diffraction patterns[6].

A quite common method to overcome this phase problem is known as Multiple Isomorphous Replacement (MIR). An isomorphous molecule is one that keeps its crystal structure even with a replacement atom in its structure. Typically, a metal chalcogen, which are group 16 elements like oxygen, sulfur, selenium, and tellurium, is appended into the structure of an amino acid within the protein. Once the metal-containing protein is crystallized and the metal atom can be located, the interference between the native structure signal and the heavy-atom containing signal can be analyzed for constructive or destructive interference[6] (Figure 2b). This gives something to compare the native diffraction phase to the higher magnitude heavy atom containing phase[6]. Selenium and tellurium are typically the metals that incorporate themselves the best into the molecular structure of amino acids.

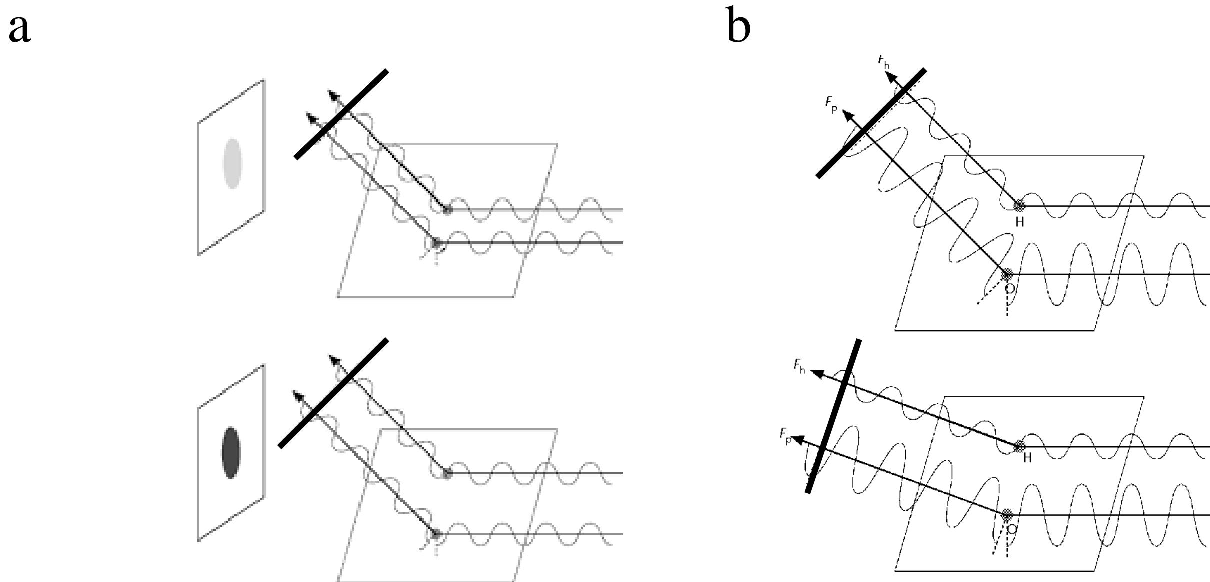


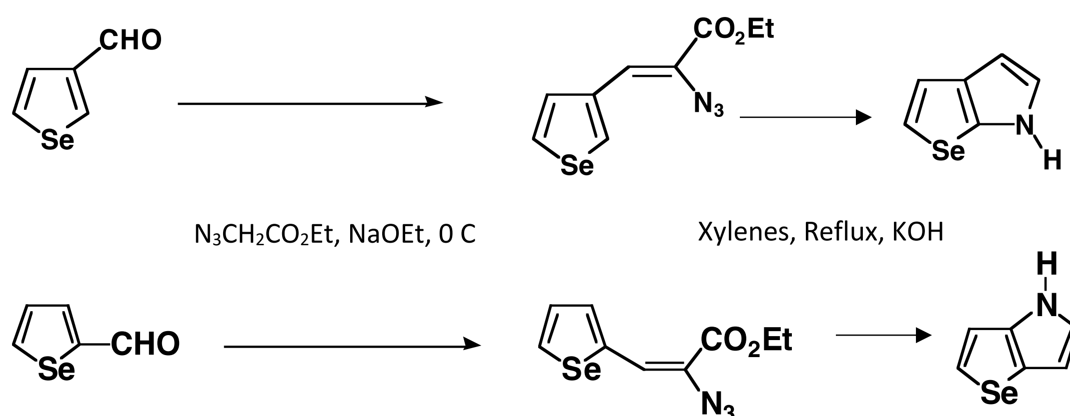
Figure 2. [2a] Constructive and destructive diffraction pattern phases causing the beam signal to be stronger or weaker. [2b] Larger magnitude heavy atom diffraction pattern (bottom) interfering with native signal (top). Notice the heavy atom phase compounded with the native phase always gives a strong enough signal to detect even with destructive interference.

Synthesis and laboratory use of [2,3-b]Selenolopyrrolylalanine

Heavy atom containing amino acid analogues are sometimes a necessary tool in accurately determining a protein's structure. Chemical protein synthesis and semisynthesis is becoming more common for studying unknown proteins partially because it allows the incorporation of these synthetically produced amino acids. Leucine and alanine are both found in over 7% of proteins in vertebrates and thus already have well established synthetic routes which include selenium inserts[7]. However, depending on the amino acid sequence for the protein, a rare amino acid insert may also be required to further map its structure. Tryptophan, being the rarest amino acid making up approximately 1% of globular proteins, serves many important functions in certain proteins (anchor points for some integrated membrane proteins) and the precursor to serotonin, a biologically essential neurotransmitter[1,8]. Replacing a part of the tryptophan's aromatic ring with a chalcogen atom leads to the formation of two preferred products in the 2,3 or 3,2 position. The thienopyrrolylalanine[I,II] and selenopyrrolylalanine(**[6,7]SeTrp**)[III,IV] isomers (Figure 3) have been synthesized and incorporated successfully into recombinant proteins via Trp-auxotrophic (inability to produce tryptophan) E Coli[1]. The selenolopyrrolylalanine derivatives were found to have exceptional stability within the parent protein as well as some thermodynamic differences[1]. The resonance stabilization provided by the selenium atom and the aromaticity of the indole makes a chalcogen atom much less susceptible to outside oxidation compared to other dialkylselenides[1]. Furthermore, due to their stability *when incorporated,* selenolopyrrolylalanine derivatives will be more reliable than other more common chalcogen-based amino acids[1] (L-selenocysteine or L-selenomethionine). Considering the rarity of tryptophan, studying its molecular replacement's biological activity hasn't been widely explored and the reliable synthesis of [2,3-b]selenolopyrrolylalanine(**[6,7]SeTrp**)[III] could prove useful for such research as well as future MIR elucidations.

Figure 3. Some synthetic chalcogen analogues from native tryptophan. Notice the incorporated chalcogen atom is so large that it takes the place of two carbons in the indole ring.

Tellurium would also help in ring stabilization; however, this synthesis will study the selenium derivative because there are many more reported studies on the chemistry of the selenophene than the tellurophene. Once a reliable pathway is established with good yields, further experimentation with telluropyrrolylalanine synthesis would provide a theoretically more stable derivative because of the trend of stability chalcogens follow. The synthetic pathway for formation of **III** + **IV** has been reported from an initial selenophene carbaldehyde and a subsequent annulation under Pomeranz-Fritsch cyclization conditions[2] (Scheme 1). This method approaches the problem of cyclizing from the nitrogen side of the molecule rather than from the selenide and, depending on the position of the carbaldehyde group on the selenophene pyrrole, the [2,3-b] or [3,2-b] product is readily formed. The utility of this method is mostly based off the already integrated selenium pyrrole which allows the reaction to proceed based off reagents suggested in the literature. However, the process is lengthy and cannot be done on large scales.

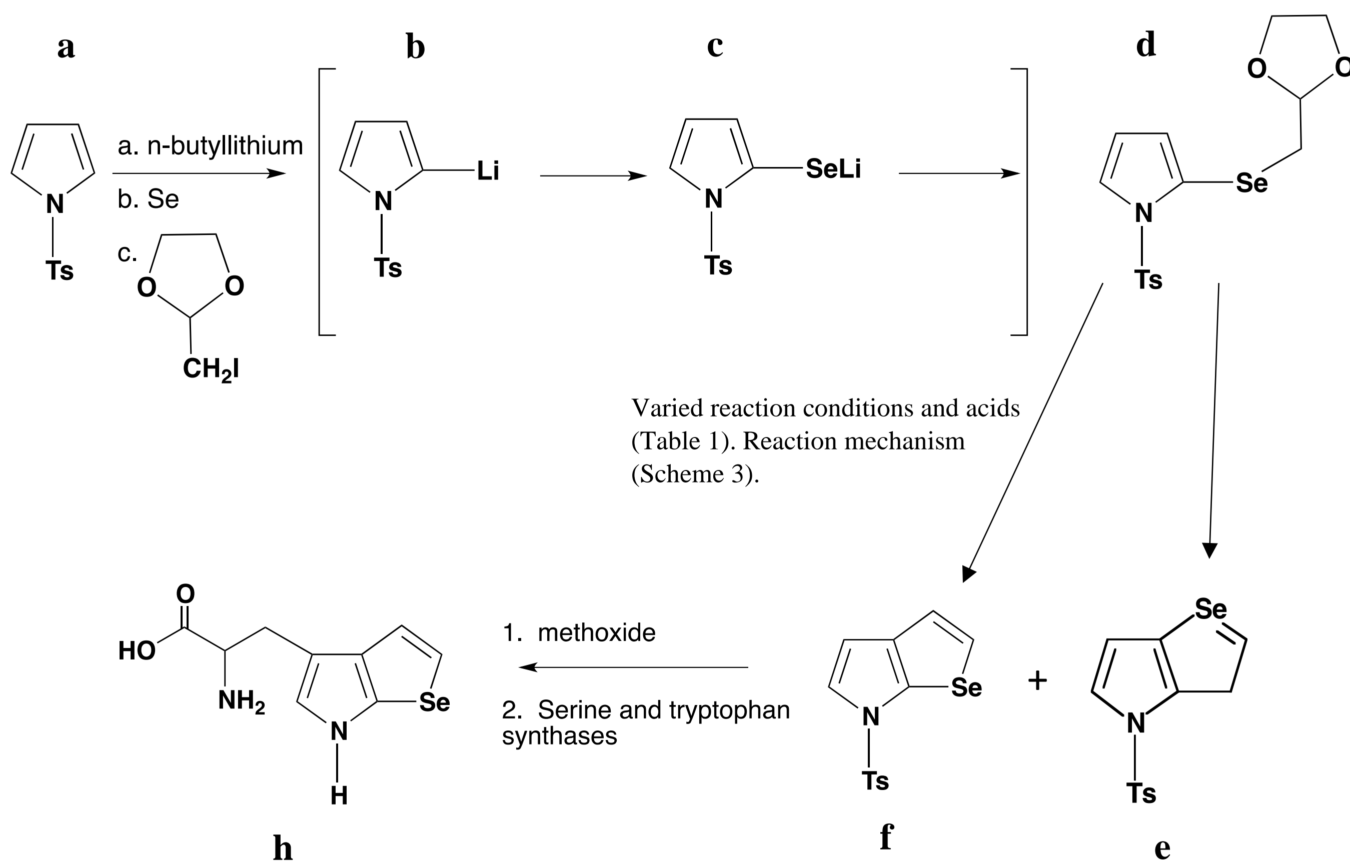


Scheme 1. Overall formation of both selenopyrrole isomers from differently substituted selenophene carbaldehyde starting material.

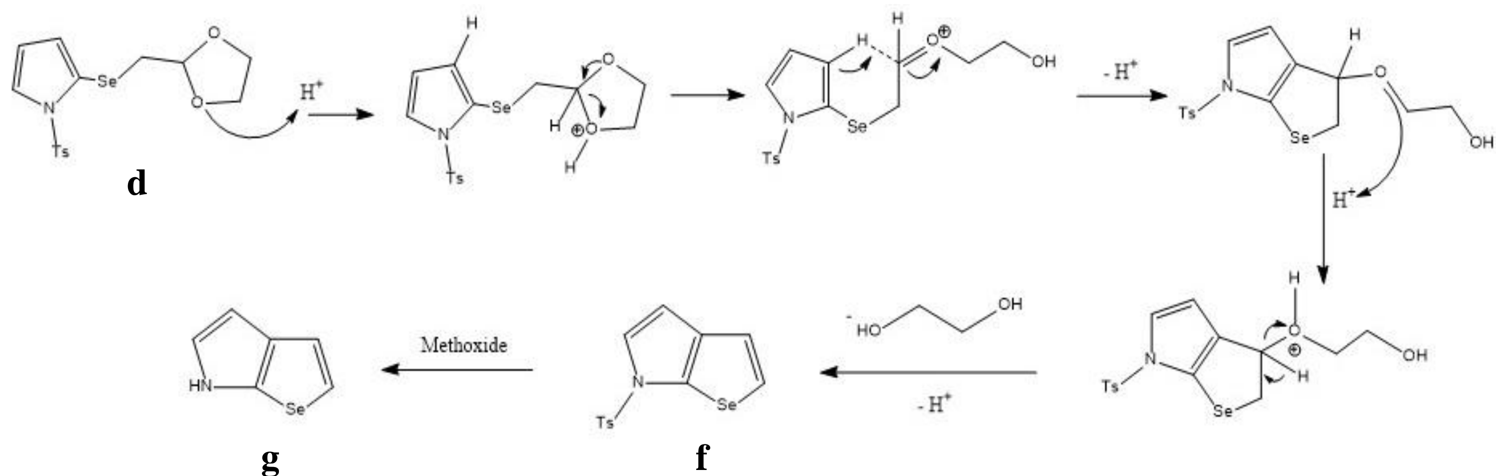
The goal of this newer synthesis is to test different lewis acids and conditions (Table 1) during the annulation step of the [2,3-b]selenolopyrrole to increase yield and sample size as well as mitigate the formation of the [3,2-b]selenopyrrole variant[2] **e**. There is a possibility of forming **e** from the minor product of the n-butyllithium treatment, N-tosyl, 3-selenopyrrole, but we can ignore this because it is a very small percentage of the product formed. The [2,3-b]selenophene has been recorded to be more sensitive to air, light, and temperature than the [3,2-b] isomer. The synthetic pathway being explored in this proposal will approach the annulation from the selenide side of the molecule rather than the nitrogen side from a commercially available tosylated pyrrole **a**, which opens the possibility of diverse reaction conditions that could lead to higher yields of **f** and lesser yields of **e** (Scheme 2). Following the Pomeranz-Fritsch mechanism, an acetal in the form of 2-(iodomethyl)-1,3-dioxolane is appended onto the selenium atom after undergoing metalation. Once compound **d** is formed, a range of lewis acids (Table 1) will be used under varied reaction conditions to find the best yields of the desired annulated product (**f**). The mechanism for the ring closing of **d** can be seen in Scheme 3. It is important that we have enough starting material to make an appropriate amount of **d** for carrying out the [six] planned lewis acid tests.

Table 1. Lewis acid and reaction conditions for formation of modified indole ring.

Acid	Conditions
Amberlyst	Reflux 24 h
$TiCl_4$	-78 °C
$SnCl_4$	-78 °C to RT
$ZnCl_2$	-20 °C to RT
$ZnCl_2$	120°C
$CeCl_3$	120°C



Scheme 2. Reaction scheme for formation of [6,7]SeTrp (**h**).



Scheme 3. Pomeranz-Fritsch reaction mechanism from formed N-tosyl, 2-(1,3-dioxolanyl) methaneselenopyrrole (**d**).

Every reagent required for this reaction is available online for purchase except for the 2-iodomethyl-1,3-dioxolane. Therefore, the 2-bromomethyl variant will be purchased making it necessary to perform a simple Finkelstein transhalogenation reaction on the terminal bromide group and replace it with the necessary iodide atom before beginning the main reaction involving the tosyl pyrrole[3]. This reaction occurs under the presence of acetone and a salt containing the replacement anion. To shift the equilibrium in favor of the product, precipitation of the bromide salt occurs while in acetone (Figure 4). It has been noted in previous syntheses that the use of 2-bromo-1,3-dioxolane leads to more unwanted side products and lesser yields making a simple substitution for the bromine variant unfavorable[2].

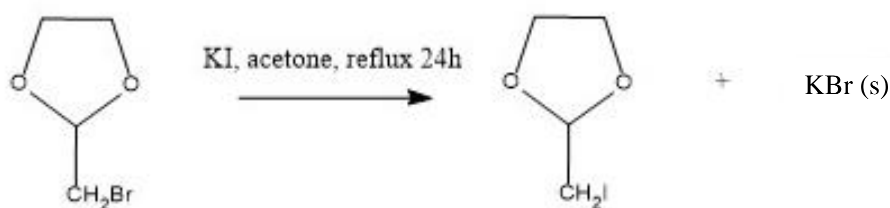


Figure 4. Finkelstein equilibrium reaction equation.

<u>Experimental procedure[2,3]</u>:

Formation of 2-iodomethyl-1,3-dioxolane:

Store-bought 2-bromomethyl-1,3-dioxolane will be dissolved in acetone. An excess amount of NaI will then be added and refluxed for 24h. The newly formed 2-iodomethyl-1,3-dioxolane mixture will then be separated from the precipitated NaBr using ether and stored for later use.

Formation of N-tosyl, 2-(1,3-dioxolan-2-yl)methaneselenopyrrole (d):

The N-tosyl pyrrole (**a**) will be dissolved in THF and cooled with dry ice to -22 C. N-Butyllithium will be added dropwise while maintaining the temperature between -20 and -40 C by adding more dry ice. This suspension will be stirred for approximately 50 min to form **b**. The solution will then be warmed to -15 C and elemental selenium will be added to perform a metalation reaction. The solution will be left to settle and warmed to 0 C (**c**). TLC analysis will then be performed. The solution will then be cooled down to -65 C with the addition of more dry ice and the previously prepared 2-iodomethyl-1,3-dioxolane will be added. The solution will then be allowed to stir overnight at room temperature. The resulting mixture will be separated using ether and water. The ether phase will be washed with water, dried with sodium sulfate, and evaporated with rotovap. Further purification will be performed using silica gel chromatography to obtain **d**. H NMR

Formation of N-tosyl, Seleno[2,3-b]pyrrole (f):

[Six] equal aliquots of **d** will be measured for the six lewis acids being tested. A mixture of clay, a lewis acid, and the previously formed N-tosyl, 2-(1,3-dioxolan-2-yl)methaneselenopyrrole (**d**)**,** dissolved in chlorobenzene, will be refluxed or cooled following the conditions listed in Table 1. This solution will then be filtered, and the solvent is removed. The residue will be purified with silica gel chromatography. Once the yield of **f** is confirmed, it will be recorded and the next lewis acid will be tested. H NMR

Formation of Seleno[2,3-b]pyrrole (g) and [6,7]SeTrp (h):

Sodium methoxide will be added to the purified **f** in methanol to remove the tosyl group forming **g**. The resulting **g** will then be introduced to tryptophan synthase from E. coli to append the free alanine onto the modified indole forming **h**.

Discussion

The study of protein's quaternary 3-D structure provides useful insight into the function of the protein. Measuring the diffraction patterns from native x-ray crystallography data and amplified data (Figure 2) from non-natural amino acids incorporated into the primary structure of the protein can provide valuable information on the secondary, tertiary, and final quaternary structure that gives the protein its primary role within a given pathway in the body. More specifically, the L-tryptophan molecule, involved in essential pathways involving NAD^+/NADP synthesis as a substrate, serotonin synthesis (leading to dopamine, norepinephrine, and other neurotransmitters), and for use in the structure of proteins, can be a necessary component in these larger oligomeric protein structures. Mammals cannot produce their own L-tryptophan for use in the body and that is why it is the least common amino acid found in mammals (by concentration). Yet, we cannot live without it, which provides many good reasons for studying its role in biology.

Methods are constantly improving in the field of biochemistry and chemistry to analyze protein's chemical activity when exposed to certain substrates and metabolic enzymes because knowing the components of a metabolized

substrate is useful information when elucidating proteins. Synthesis of the [6,7]SeTrp analogue isn't limited to use in X-ray crystallography, and more specifically multiple isomorphous replacement. The selenium atom incorporated into the indole ring has similar aromatic properties to native L-tryptophan mitigating any unwanted interactions the larger metal atom may have when integrated into a protein's structure. The metabolism of [6,7]SeTrp would also be an interesting pathway to explore because of the changes tryptophan undergoes while being introduced to these various proteins that effectively oxidize the amine group on the alanine . And, because of their relatively similar stability, analyzing the metabolites of [6,7]SeTrp and comparing them to those of L-tryptophan through the pathway below (Figure 5) could show reactions that occur on the selenopyrrole side of the molecule that could alter the chemistry of the oxidation of the amine group on the other side.

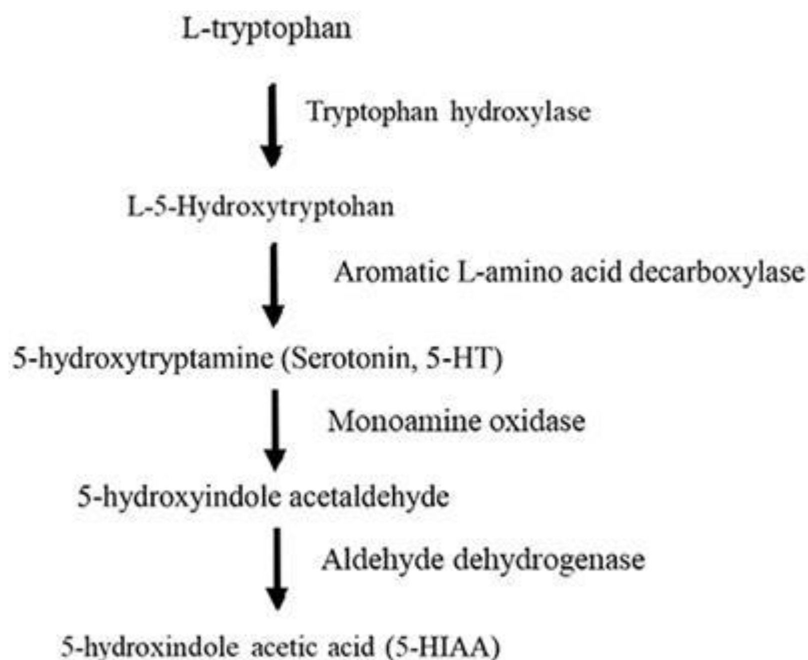


Figure 5. Tryptophan metabolism pathway involving serotonin.

Another possible route of exploration would be synthesizing the previously mentioned [6,7]TeTrp isomorph. Tellurium has shown to have similar stability to selenium being the heaviest of the non-radioactive chalcogens and would theoretically give a better electron density map compared to selenium when used in MIR crystallography due to its massiveness.

References:

1. Moroder, L, Musiol, H-J. Amino acid chalcogen analogues as tools in peptide and protein research. *J Pep Sci*. 2020; 26:e3232. https://doi.org/10.1002/psc.3232 (Accessed March 16, 2021).
2. hatch selenophene manuscript
3. hatch powerpoint
4. Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. Analyzing Protein Structure and Function. Available from: https://www.ncbi.nlm.nih.gov/books/NBK26820/
5. ATA Scientific Instruments. Protein Analysis Techniques Explained. https://www.atascientific.com.au/3-protein-analysis-techniques/ (Accessed March 16, 2021).
6. Cowtan, Kevin, Phase Problem in Xray Crystallography, and its Solution. *Encyclopedia of Life Sciences.* 2001. http://people.bu.edu/mfk/restricted566/phaseproblem.pdf (Accessed March 17, 2021).
7. *M. Beals, L. Gross, S. Harrell.* AMINO ACID FREQUENCY. http://www.tiem.utk.edu/~gross/bioed/webmodules/aminoacid.htm (Accessed March 10, 2021).
8. Richard DM, Dawes MA, Mathias CW, Acheson A, Hill-Kapturczak N, Dougherty DM. L-Tryptophan: Basic Metabolic Functions, Behavioral Research and Therapeutic Indications. International Journal of Tryptophan Research. January 2009. doi:10.4137/IJTR.S2129. https://journals.sagepub.com/doi/10.4137/IJTR.S2129#articleCitationDownloadContainer (Accessed March 16, 2021).