Peptide Bond Formation in Nano-Reactors

by

Kelsey Hayes

Biology Departmental Honors Thesis

Appalachian State University

Submitted to the Biology Department
in partial fulfillment of the requirements for the degree of

Bachelor of Science

April 20th, 2018

Approved by:

Nicholas N. Shaw, Ph.D., Thesis Director

Andrew Bellemer, Ph.D., Second Reader

Lynn Siefferman, Ph.D., Departmental Honors Director

Abstract

Organic synthesis has many applications in the chemical field. A frequent reaction used in organic synthesis is amide bond formation. Amide bond formation is commonly achieved through the combination of an amine and an activated carboxylic acid. Amide bonds link together amino acids to form peptides in biological systems. Reports of amide bond formation through the direct coupling of carboxylic acids and amines are non-existent as the favored reaction between the amine and carboxylic acid is an equilibrium reaction. The forward reaction, bond formation, products water and is a condensation reaction. The reverse reaction is a hydrolysis reaction where water can hydrolyze the newly formed bond. Recently, we have developed a method for the direct coupling of the amine and carboxylic acid using nano-reactors. The nano-reactors have four characteristics that allows them to facilitate direct amide bond coupling: (1) Nano-reactors absorb organic molecules into central cavities. (2) Upon absorption of organic molecules, the nano-reactor swells 3-5 times its dry volume (3) The nano-reactor is hydrophobic, water does not enter the central cavity and water synthesized in the cavity migrates out of the cavity. The produced water is expelled from the hydrophobic nano-reactor driving the reaction forward. When applied to the synthesis of proteins, which has challenged chemists for over a century, trial syntheses conducted using nano-reactors and amino acids occur with high efficiency and yield. These advancements can be applied to peptide synthesis for the use of vaccines. The authors will present an optimized procedure for the direct coupling of amino acids and its subsequent use in the synthesis of several di-peptides.

Acknowledgements

My thesis advisor, Dr. Shaw, deserves my upmost thanks for all his time and investment into my development as a both a student and a scientist. This thesis would not be complete without him. I would like to thank Dr. Bellemer for agreeing to be my second reader and for taking the time to also invest in me as a student. Great professors such as these are the jewels of any university and make attending a joy. To the Biology Department, I would like to show my thanks for allowing me to have this opportunity to cross designations and explore both biology and chemistry through this project. I would like to thank the students of the Shaw Research Group for all their hard work and perseverance in learning more about science and pushing the boundaries of what we can do with science. Finally, I would like to thank my friends and family for supporting me through this process.

Table of Contents

Table of Figures

Table of Tables

Table of Schemes

Chapter 1: Applications to Biology

Vaccines were first documented and administered by Edward Jenner in 1776. Since then the methods for creating vaccines that are not only more effective and safe for the public has been on the scientific radar. One of the new advances and considerations for vaccines is the introduction of peptide vaccines. The majority of vaccines on the market today contain killed whole pathogens, viruses, and or purified proteins produced from bacteria in the lab. Some of the major concerns for traditional vaccines are the yield of purified proteins and the strong allergic responses for patients who receive these types of vaccines. The whole pathogens usually elicit longer lasting immunity, however they also can cause strong allergic or autoimmune responses in the patient (Skwarczynski et. al. 2016). Peptide-based vaccines have some advantages over whole pathogen vaccines. Peptide-based vaccines are more customizable and can be engineered to target several strains or different life stages of pathogens (Skwarczynski et. al. 2016). Peptide based antigens are also less likely to induce allergic reactions or autoimmune responses due to a lack of redundant elements (Skwarczynski et. al. 2016). Both biological and synthetic systems can be used to synthesize peptides and proteins.

Peptide/Protein Synthesis with Escherichia coli

Peptides, and whole proteins, can be produced and purified from *Escherichia coli* cultures. For this method plasmids are chosen to transform into the *E. coli* cultures that will then start synthesizing the peptide of interest. Culturing *E. coli* for this process is convenient due to extensive research and data collected on *E. coli* as a model organism. *E. coli* is also relatively easy to culture in the lab and cultures can achieve a high cell density quickly on readily available media. These aspects of *E. coli* allow for this method to be very cost effective and allow for a high protein purification yield due to the sheer number of cells in culture (Rosano 2014). In practice there are

many factors to consider when synthesizing proteins in *E. coli*. Before transforming a plasmid into the culture, an appropriate plasmid must be designed to express the peptide or protein of interest at the most effective yield. The replicon of the plasmid should contain an appropriate origin of replication region, promoter, the coding sequence, any affinity tags, sequence for tag removal, a terminator sequence, and a selection marker.

Through PCR and reverse transcription methods, mRNA from human or microbial cells can be reverse transcribed into a cDNA that is ready for insertion into a desired plasmid. The resulting cDNA's contain no introns and can be used as templates for protein/peptide expression in prokaryotic organisms. This allows for proteins/peptides to be expressed that do not need to be spliced, taking away one of the modifications needed. Restriction enzymes produce complementary sticky ends that can be ligated with ligase to produce a recombinant plasmid. This recombinant plasmid of interest should contain the cDNA insert and the plasmid vector chosen to insert into the *E. coli*. Plasmids up to 10,000 base pairs have been shown to be successfully transformed into *E. coil* without loss of transformation frequency (Inoue 1990).

Transformation of the plasmid into *E. coli* cultures has been optimized by Inoue et. al for the overall process. Certain strains of *E. coli* were presented as competent cells that are optimal for transformation including strains DH5α, JM109, HB101 (Inoue 1990). To make cells competent for transformation the cells are usually treated with a salt, like calcium chloride ($CaCl_2$) to prep the cells for transformation of the plasmid. The full role of $CaCl_2$ is not entirely understood, however $CaCl_2$ is thought to open pores in the membrane of the cells and provide a positive counter charge allowing for the intake of negatively charged plasmid DNA. Salts that can be removed from the *E. coli* include KCl, $CaCl_2$, $MnCl_2$, and K · acetate. Other salts are shown to be inhibitory for

transformation (Inoue 1990). A heat shock can be used to promote the plasmid DNA being taken up by the *E. coli* after the cells have been prepped with salts.

The first concern and method of controlling the amount of protein synthesized in the *E. coli* culture is the origin of replication in the recombinant plasmid. Origin regions control the copy number of the plasmid being synthesized which will affect the amount of protein/peptide synthesized. For example, ori pSC101 will make less than 5 copies of the plasmid while ori pUC will make 500-700 copies of the plasmid per cell in the culture (Rosano 2014). This is important for keeping the *E.coli* culture healthy and producing peptides for a longer period. Some proteins at higher concentrations can cause cell lysis or will elicit the protein degradation system in the cell (Rosano 2014). Both processes result in loss of the protein yield. Cell lysis will affect the surrounding cells as well by releasing proteases into the culture that could potentially cause a complete loss of the culture and synthesized protein. Besides preventing loss of peptide or protein yield, higher concentrations of the protein in solution will hinder correct folding of the protein/peptide or cause aggregation. The shape of a protein is a large factor in that proteins function. If a larger amount of protein is produced but is folded incorrectly, it will be ineffective in a vaccine. Protein aggregated out of solution will also be ineffective for vaccine use.

Promoters also play a large role in the expression level of proteins and the most commonly used is the *lac* promoter. This promoter, having been studied thoroughly, is under positive control due to catabolite repression. Lactose can be used to induct of the system resulting in protein production (Rosano 2014). This allows for control over the promoter but the yield level for the *lac* promoter is weak, therefore synthetic promotors are available that can combine the control of the lac promoter with the strength of another promoter (Rosano 2014). The default promoter is usually

the T7 system that allows for control over the system and allows for the protein/peptide of interest to consist of up to 50% of the total protein yield of the cell (Rosano 2014).

Plasmids should also include selection markers for distinguishing between bacteria successfully transformed with the plasmid and bacteria without the presence of a plasmid. To do this an antibiotic resistant gene is placed into the plasmid backbone. When bacteria are successfully transformed with the plasmid, those cells can grow on media containing that antibiotic. This is useful in collecting the transformed colonies; however, antibiotics are costly and getting rid of any resistance before the protein/peptide is inserted into the vaccine is a major concern (Rosano 2014). Some plasmid systems have been developed that are antibiotic free and use plasmid addiction to select for transformed cells. Plasmid addiction occurs when cells without the plasmid are unable to grow. This involves the deletion of an essential gene in the *E. coli's* genome and then introducing that gene back into the cell with the plasmid (Rosano 2014).

Affinity tags and their removal are also a concern for synthesizing proteins/peptides from *E. coli*. Affinity tags allow for the removal or purification of the protein/peptide out of the cells once the production process is over. Affinity tags can be utilized in western blots, antibody stain, and column chromatography. These tags should be relatively small as to not interfere with the tertiary folding of the protein of interest (Rosano 2014). Tags can be removed by chemical or enzymatic cleavage (Rosano 2014). Chemical cleavage allows for easy elimination, but the reaction conditions are harsh, limiting the number of proteins that can use this system and this cleavage can also result in unwanted protein modification (Rosano 2014). Removal of the affinity tag is a concern because the solubility of the desired protein/peptide is unpredictable (Rosano 2014). This varies from protein to protein and calls for multiple rounds of protein synthesis and purification.

Peptide/Protein Synthesis with Baculovirus Vectors

Baculoviruses are a part of the family *Baculoviridae* that infects invertebrate cells and have been used to infect mammalian cells in the lab. Baculoviruses can infect mammalian cells in the lab but are unable to reproduce in mammalian or vertebrate cells. Proteins/peptides can also be synthesized and purified using a baculovirus vector system. In this system the sequence for the gene is incorporated into in the baculovirus genome (Kost 2005). This vector allows for expression in insect and mammalian cells rather than bacteria culture like *E. coli*. Infection of insect cells will allow for 3 to 5 days of synthesis before cell lysis occurs (Kost 2005). Lysis of this system, like the lysis of the *E.* coli cells, results in the release of proteases that can chew up protein/peptide product before collection of the product. To avoid time limits on protein production, strains of the baculovirus have been engineered with a reduced capability for cell lysis (Kost 2005). These modified strains show an average of 53% less cell lysis after the 5 day mark compared to the parent virus (Kost 2005).

Expression in eukaryotic cells, usually in insect systems, can help to ensure that the protein/peptide product will be folded, modified, and trafficked correctly (Kost 2005). Co-expression of chaperone proteins with the protein of interest in this system aids in the correct folding of the protein/peptide (Kost 2005). This feature of the baculovirus system is an advantage over the *E. coli* system. Expression of proteins/peptides in *E. coli* would not have eukaryotic modifications. Examples of eukaryotic modifications include but are not limited to phosphorylation and glycosylation. This system is useful for proteins that require a large amount of specific folding and if several subunits are needed to form the active product. All of the subunits can be encoded for and expressed at the same time. Multiple cultures of baculaviruses can be generated so that each culture expressed a single protein at a specific concentration. Co-infection

of cells from the different cultures then allows for control over the expression of multiple proteins/peptides each at specific concentrations. This fine-tuning aspect of this system allows for a more flexible approach for expression of multiple proteins/peptides at the same time. Also, if co-expression of many extra proteins is necessary then co-infection can ensure that a single culture is not overloaded.

As mentioned the baculovirus can be used to make virus like particles (VLPs) which cannot replicate in mammalian cells but can induce an immune response by presenting antigens to the immune system (Kost 2005). This expression system is used to produce VLPs for the current vaccine Cervarix® to prevent infection by the human papillomavirus (Senger et. al. 2009). Of the 200 types of human papillomavirus (HPV) there are two main groups, cutaneous and mucosal, based on what is being infected (Senger et. al. 2009). Cutaneous HPV infects the skin causing skin warts or epithelial tumors; mucosal HPV causing genital warts and anogenital cancer (Senger et. al. 2009). Cervarix® prevents HPV types 16 and 18 which cause 70% of cervical cancer cases (Senger et. al. 2009). Very low yields of protein expression for certain types of HPV were a concern for the baculovirus expression system. Now with improved methods of expression, MultiBac system, these proteins can be increased and controlled by the polyhedron (polh) or the p10 promotor (Senger et. al. 2009). Large amounts of the proteins for different types of HPV are able to be produced using the optimizations presented by Senger et. al, however, the team did not research the genetic stability of the MultiBac baculoviruses (Senger et. al. 2009). Genetic stability and ensuring gene preservation is an important concern for large scale vaccine productions. For this optimized system to be considered for vaccine productions further studies must be made on the genetic stability.

Furthermore, baculoviruses can be used for gene delivery in mammalian cells by carrying mammalian cell- active expression cassettes; these viruses are referred to as BacMan viruses (Kost 2005). Many different cells types can receive genes from the BacMan virus making it an ideal system for gene therapy and targeted peptide delivery. The BacMan virus also cannot replicate in mammalian cells and is therefore unable to generate a new viral infection (Kost 2005). Studies using the BacMan method have reported significant toxicity in response to possible induction of inflammatory cytokines in response to viral delivery (Kost 2005). Creating strains that are less prone to complication of this kind are in the works (Kost 2005).

Peptide/Protein Synthesis with Transient Expression Systems

Transient expression systems first introduce the gene and then begin harvesting protein similar to other systems. The gene of interest is not incorporated into the host genome and may need to be reintroduced to the system multiple times. This system might be used in instances where the protein of interest may need to be expressed at precise levels to avoid toxicity. Generally, this system shows lower aggregation and lysis rates and thus fewer problems with producing "too much protein". Furthermore, this expression system, like the *E. coli*, system does use affinity tags and columns for protein harvesting. The use of columns is common and affinity tags can be inserted into the DNA without too much complication. Though not specified in Wurm et. al, similar concerns for affinity tag removal would most likely be present for this system like the concerns of tag removal in the *E. coli* expression system.

The methods for transient expression can be divided into two categories: viral vector and plasmid vector systems. Both systems have some advantages to them. These methods are relativity simple in terms of processes and techniques used. Production of protein occurs in a few days in a

wide range of host cell lines. Due to the shorter production time, shorter time between generation of vector and the product recovery, there is more genetic stability and consistency (Wurm 1999).

Expression in mammalian cells can be achieved by transient expression in viral vectors such as the adenovirus and vaccinia vectors. The adenovirus transcription virus has been studied thoroughly over the years allowing for the creation of heterologous expression systems (Wurm 1999). In these systems, early regions of the genome have been deleted to prevent viral replication. Replication is restored in trans by the host cell and the gene of interest is placed under control of a late promoter, allowing for expression after other viral machinery (Wurm 1999). This viral vector system is able to be increased to large scale production by propagation in suspension cell cultures. Vaccinia vectors are ideal for large scale infection in rates of 5000 virus particles per cell (Wurm 1999). This vector system is also placed under control of a late promoter. In both cases, late promoters are ideal for high levels of protein/peptide expression and earlier promoters can be chosen to decrease the level of protein expression. Vaccinia vectors can also target a large variety of hosts (Wurm 1999). A draw back to these systems include a high safety level for labs which may require more money and equipment driving up the cost of the protein/peptide product for vaccines.

Transient expression is also possible in plasmid vector systems. DNA delivery methods for mammalian cells using plasmid vectors are fairly cost effective and can be utilized for adherent or suspended cells. Cells in suspension with plasmid vectors have been able to reach protein production levels of 0.1-1 pg/cell/day and are ideal for production levels up to 10 L (Wurm 1999). After transfection cells can remain in a producing stage at a limited time span, 5- 10 days. This process is not suitable for scales beyond 10 L and would not be able to be used for large scale protein/peptide production for mass vaccines (Wurm 1999). Transfection levels have an average

rate of 70% which is reliable for protein product set ups. The problems with this system lie mainly in the vehicle systems used. Different vehicle systems include calcium-phosphate, polyethylenimine (PEI), and electroporation which have all been shown to be successful vehicles for the larger scale transient gene expression (Wurm 1999). Both calcium-phosphate and PEI methods rely on endocytosis to enter the cell. There is the assumption that large amounts of DNA are degraded before they make it to the cell (Wurm 1999). Electroporation allows for the pores of the cells to open due to an electrical shock. This vehicle method can cause a large portion of cells to die from electric discharge, sometimes resulting in the death of 50% or the cells being transformed (Wurm 1999).

A disadvantage of the transient expression system in cultivated mammalian cells at a large scale is that it is usually done in the presence of serum (Derouazi et. al.2004). Contamination of serum in a vaccine could lead to higher health risks. Also working in the presence of serum would require that an extra purification step must be taken to remove the serum from the synthesized protein. Extra steps raise the cost level of vaccine production. New experiments are being conducted with Chinese hamster ovary cells (CHO) and transient expression systems in a serum free environment. The CHO cells were transformed with green fluorescing protein (GFP) and the amount of fluorescence reached a maximum of 10 mg/L after 5 days' time (Derouazi et. al. 2004). Derouazi et. al's study shows that mammalian CHO cells can be cultured in the absence of serum and can be used to transiently express proteins.

Peptide/Protein Synthesis with Fully Synthetic Methods

Instead of utilizing a biological system to produce peptides/proteins, a fully synthetic approach can be used. The synthesis of a fully synthetic peptide is exclusively chemical which eliminates contamination of antigens from a biological system (Skwarczynski 2016). This leads to

the production of a much safer vaccine with lower risks of a severe allergic reaction. Furthermore, automated steps can be introduced with the use of a fully synthetic approach and can therefore cause the vaccine to be much more cost effective (Skwarczynski 2016). Finally, fully synthetic peptides can be stored under simple storage conditions, due to their water-solubility, and can be freeze dried (Skwarczynski 2016).

Solid phase synthesis is the current fully synthetic method of choice for synthesizing peptides. Developed by Professor Merrifield from The Rockefeller University in New York, this method won the Noble Prize in Chemistry in 1984. This method involves sequentially bonding individual amino acids, one monomer at a time. The first amino acid is covalently bound to the to the support material (Merrifield 1985). Once this first amino acid is anchored, the next amino acids can be subsequently added. The C terminus end of the anchored amino acid is deprotected and the next amino acid is added (Merrifield 1985). This process continues until the last amino acid in the peptide is added. After the complete peptide is synthesized, the covalent bond holding the initial amino acid to the support system is selectively cleaved (Merrifield 1985). This same general process can be applied for the synthesis of polynucleotides and polysaccharides.

The solid support system used are cylindrical beads consisting of co-polymerized styrene and 1% divinylbenzene (Merrifield 801). With the addition of organic solvents, the beads are able to swell up to 25-fold the size of the original dry bead which allows for solvents to penetrate to the interior of the bead (Merrifield 1985). The addition reactions are able to take place both on the surface of the bead and on the beads interior (Merrifield 1985). The combination of bead swelling and saturation of solvents in the beads allows for the reaction to proceed quickly (Merrifield 1985). Another advantage to peptide synthesis with this method is that solvation is increased, and peptides are less likely to aggregate. This occurs due to a mutual solvating effect that the bead and the cross-

linked polymer chain have on each other; thermodynamically making self-aggerates less favorable (Merrifield 1985).

New developments in nan-reactor technology has been made that could improve upon or become its own synthesis system. Nano-reactors will be explored more after a look a chemical reaction.

Chapter 2: Introduction to a Chemical Problem

Traditional Methods of Conducting Synthesis Reactions

A basic chemical synthesis reaction involves two molecules colliding with one another to form a new bond and subsequent new compound (Brown 2014). However, there are many variables that require optimization for a reaction to occur. These variables can be molecular: the number of bonds on each atom, any charges, steric hindrance, and the reactivity of the molecule's side chains. They can be chemical: does the reaction require heat or light? Does the reaction need help from a base or an acid? And they can be situational: how physically close are the molecules and how big is the reaction flask?

Traditionally there are methods that can be used to get a reaction to proceed. These methods can be physical (placing the reactants in a flask with a solvent and adding a stir bar) or chemical (affording energy with heat/light or the addition of an appropriate acid or base).

The Theory of Chemical Reactivity

Collision theory and transition state theory suggest that in order for a reaction to occur reactants must collide in the correct orientation and possess enough energy (activation energy) (Brown 2014). Collisions are less likely to occur when the reacting molecules are very far apart. This is the case for molecules in a traditional sized flask. The molecules will spend more time traveling within the flask without colliding with another reactant molecule. The molecules that do collide but are not correctly aligned will not form product and will bounce off each other (Brown 2014). The molecules that collide in correct alignment still require enough energy in the collision to break and form bonds.

Traditionally, scientists have been able to increase the probability of reactants colliding in the correct orientation and with enough energy by increasing the energy, temperature, of the

solution and/or increasing the concentration of the reactants in the reaction and/or increasing the reaction time. And while these methods are successful in increasing the likelihood of collisions, these methods (even when optimized) often fail to produce the desired product in quantitative yields efficiently (Kalinski et. al. 2010).

Problems with Traditional Reactions: Reaction Rate

How fast reactions proceed is very important to drug and vaccine manufacturing industries. The time it takes to conduct a reaction factors very heavily into the cost of these medications. Factors that impact the rate of a reaction are the physical state of the reactants, the concentrations of the reactants, the temperature of the reaction, and the use of a catalysts (Brown 2014).

Reaction rates are dependent on the physical state of the reactants. Reactants can be in solid, liquid, or gas form if the two reactants are in the same state the mixture is homogenous if the two reactants are in different states then the mixture is considered heterogenous (Brown 2014). Heterogenous mixtures can be limited by the amount of contact area available between the two reactants. For instance, if one of the reactants (A) is in a solid state and the other reactant (B) is in a liquid state there is an unequal distribution of reactant A to reactant B (Brown 2014). Reactant A could be stirred to increase the amount of contact with reactant B, but this can take a long time especially if reactant A is a compact solid rather than a powder.

Reaction rates are also dependent on the concentration of reactants. The rate can be increased with an increase in reactant concentration (Brown 2014). Increased concentration increases the probability that the two reactants will collide and increases the frequency of these molecular collisions, thus increasing the overall reaction rate.

Reaction rates are also dependent on the reaction temperature. In general, reaction rates increase if the temperature of the reaction is increased (Brown 2014). Typically, as a rule of thumb,

the reaction rate is doubled with every ten degrees increase in Celsius (Brown 2014). Increasing the temperature increases the energy of the reactants which increases molecular movement (Brown 2014). If the molecules in solution are moving at a higher rate, then there is a higher probability more molecules will collide at velocities with sufficient energy therefore increasing the reaction rate.

Reaction rates can be impacted by the presence of a catalyst. Catalysts are found in many cells and they are crucial for increasing biological reaction rates. Catalysts are not used up in a reaction, so they can be recycled and participate in many rounds of the reaction (Eisemann et. al. 2014). Catalysts effect the collisions of molecules by binding to one or more of the reactants and bringing them closer together in the correct orientation. This has a profound effect and can increase the rate of the reaction. The use of catalysts are limited as catalysts are reaction specific, they bind select molecules, they require specific temperatures and pH conditions in order to work, and their use are often cost prohibitive (Eisemann et. al. 2014).

Problems with Traditional Reactions: Reaction Yield

Reaction yield is another concern for chemical reactions both in the laboratory and in an industrial setting. Non-quantitative yields are a driving factor in the cost of drug and vaccine synthesis and can be a frustrating and costly component of laboratory research. A single reaction that yields the desired product in 80% yield, accepted as a great yield, represents a 20% loss of product. Furthermore, this becomes a huge deal for a drug synthesis that requires multiple steps. With each step there is a loss in yield and a loss in money. For instance, a three-step synthesis with a 20% loss in product in each round means that only 40% of the final product is made. If there is a purification step after the three-step synthesis, there is also no guarantee that you will be able to perfectly purify the product and get the entire 40%. After this you are looking at a lower yield of

20-30% of actual product you can use. At lower yields synthesis reactions like these are not feasible for industry level synthesis (Kalinski et. al. 2010). This monetary loss falls on the shoulders of the consumers of the medications who pay an *inflated* cost for the medication.

Problems with Traditional Reactions: Limited Scope and Applications

There exist sub-classes of reactions wherein reaction yields are incredibly low (<10%) or reactions do not progress – reactions in chemical equilibrium (Brown 2014). In these reactions, letting the reaction proceed for an extended period and/or under concentrated conditions should allow for more collisions which should then allow for more product formation, but often do not, which poses enormous problems for reactivity and yield. Chemical equilibrium occurs when opposing reactions are occurring in at the same rate so that there is a net change of zero in the concentration of both the reactants and products (Brown 2014). Not all chemical reactions can proceed in the forward and reverse directions, however, many drug synthesis reactions have at least one step that face equilibrium challenges (Brown 2014).

Equilibrium reactions are governed by Le Chatelier's Principle which states "If a system at equilibrium is disturbed by a change in temperature, pressure, or a component of concentration, the system will shift its equilibrium position so as to counteract the effect of the disturbance" (Brown 2014). Removing products from the equilibrium reaction drives the reaction towards production formation. Reactions that exist in equilibrium are hard to complete on a benchtop scale (Brown 2014). It is reasonable to suggest that scaling the reaction up to a large industrial scale would be even more difficult. Synthesis reactions on an industrial scale depend on increasing temperature and pressure to drive the reaction forward. This is a harder and more costly method. With the removal of product, the reaction is driven forward without adding external factors and in

theory could allow the reaction to proceed to a 100% yield. With constant removal of product, the reactants will keep forming product until all of reactants are converted into product.

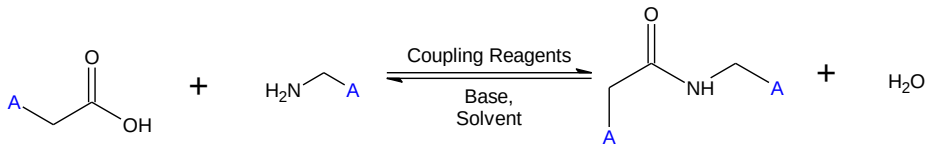


Figure 2.1. Reaction scheme for a generic direct amide bond formation reaction.

Direct amide bond formation is a reaction that exists at equilibrium in solution. **Figure 2.1** shows the general reaction scheme. This reaction is a condensation reaction and forms water as one of the products in the forward reaction. Water that is produced can hydrolyze the newly formed amide bond in the reverse reaction. Yields for this type of reaction in solution are extremely low (Brown 2014). Through traditional reaction methods, it is hard to remove water from the reaction. Industries and laboratories choose to increase the temperature to drive the reaction towards product formation (Kalinski 2014).

SOMS as Nano-Reactor for Synthesis – A New Way to Run Reactions

Swellable organically modified silica (SOMS), commercially known as Osorb®, is traditionally used for decontaminating water (Edminston 2009). As the name implies, SOMS' structure is a form of modified silica that takes the form of a hollow sphere. SOMS is porous and allows for molecules from the external environment to enter the central cavity. The porous nature of SOMS attributes to the swellable portion of the name. As molecules enter the central cavity, SOMS will expand, or swell, to accommodate the molecules (Edmiston et. al. 2009). SOMS can be manufactured to swell to different sizes. For example, some SOMS in the lab can swell to 3 times its original size (Edminston et. al 2011). The organic modifications of SOMS gives it a

hydrophobic nature that repels water from entering and drives it out of the central cavity if it is a side product (Edminston et. al.2009).

SOMS can be used as a nano-reactor because SOMS absorbs organic molecules into a central cavity (Shaw 2017). These molecules are free to react within the central cavity and then can be washed out with excess solvent. Secondly, SOMS swells to 3-5 times its original size when the organic molecule is adsorbed and absorbed (Shaw 2017). This characteristic allows SOMS to accommodate many molecules at once. Thirdly, selective removal of organic material from a multi-component mixture of organic material collapses the SOMS, concentrating the remaining organic material - upon complete removal of organic molecules, SOMS resumes its original size (Shaw 2017). which allows for SOMS to be reused. Lastly, SOMS is hydrophobic. Water cannot enter the central cavity and any water that is produced during the reaction migrates out of the center cavity (Shaw 2017). SOMS can be used to simplify traditional reactions through its nano-reactor characteristics.

Problems Solved with SOMS Nano-Reactors: Reaction Rate

SOMS aids in the collisions between the reactants thus effectively reducing reaction time. Adsorption and subsequent absorption of organic solvent and reactants causes the SOMS to swell and accommodate the molecules (Shaw 2017). Selective removal of solvent molecules collapses the SOMS forcing the reactants into a confined space and where they have less room to move around (Shaw 2017). This process is referred to as a flex or a single closing of SOMS. The confined space increases the probability of collisions between molecules effectively increasing the rate of the reaction.

As described, SOMS can swell to accommodate solvent and reactant molecules to enter the central cavity (Edminston et. al. 2009). Solvent is then selectively removed to condense the

SOMS, bringing the molecules in close proximity to one another. Solvent can be reintroduced to the SOMS to make it swell again and the solvent can be selectively removed again. This swelling and contracting called a 'flex' (Shaw 2017). The solvent removal process is both efficient and relatively fast. This allows a flex to be quick. **Figure 2.2** shows percent conversion, with and without the use of SOMS, of canola oil into biodiesel compared to the number of closes (Shaw 2017). **Figure 2.2** shows that 96% conversion is achieved after three flexing steps. A lower number of flex steps combined with the speed of each step effectively reduces the overall reaction time.

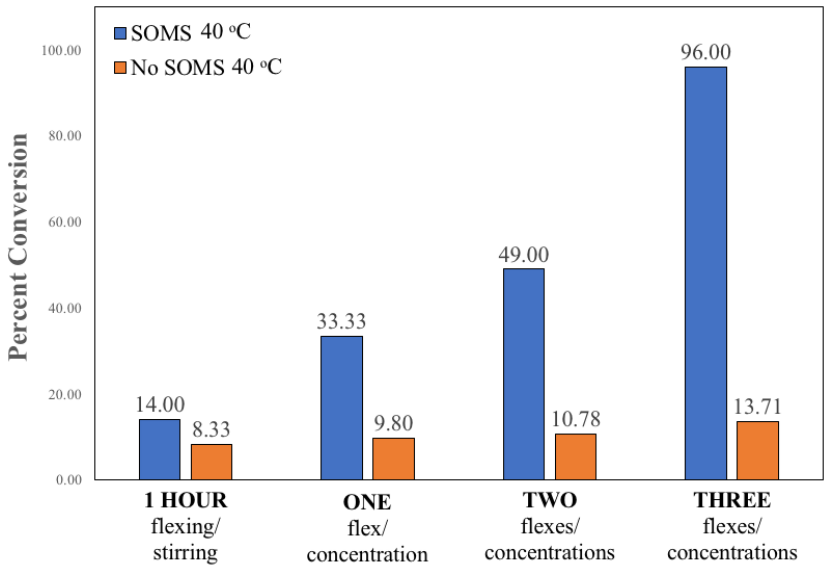


Figure 2.2. Comparison between a traditionally conducted imine condensation reaction (orange) and the same imine condensation reaction conducted in SOMS nano-reactors (blue).

A system can be devised to allow the reaction vessel to flex multiple times over the course of an hour (Shaw 2017). This flexing station takes advantage of reflux to selectively remove the solvent, close the SOMS, and then condense the solvent to return and allow the SOMS to swell again (Shaw 2017). Flexing strength can be increased if a more powerful system is used to selectively remove solvent (Shaw 2017). This is the case for flexes that occur on the rotary

evaporator. More of the solvent is removed and the SOMS closes completely. This increases the percent conversion per flex.

Problems Solved with SOMS Nano-Reactors: Reaction Yield

SOMS increases reaction yield. The confined space of the central cavity in SOMS can increase yield since the number of molecular collisions has increased. **Figure 2.3** demonstrates the increased reactivity with conversion of canola oil into biodiesel (Shaw 2017). Each close, or flex, of the SOMS increases the percent conversion.

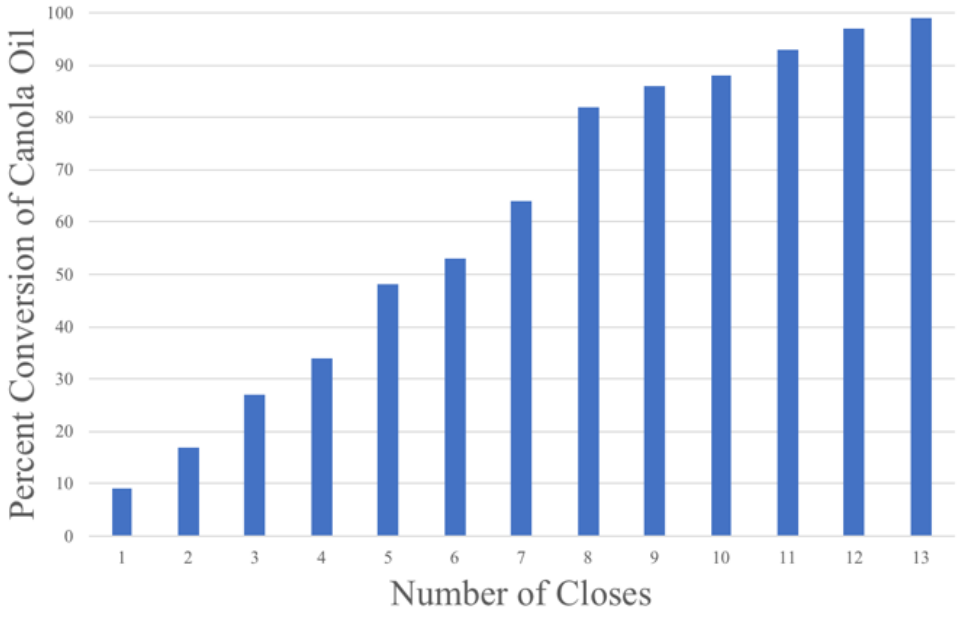


Figure 2.3. Dependence of reaction progress on the number of closes for the conversion of canola oil into biodiesel fuel.

In SOMS the molecules no longer have to travel great distances to find each other, as in a traditional reaction flask. More flexes provide multiple opportunities for the reactant molecules to collide and form product. After each sequential closing or flexing of SOMS, more product is formed (Shaw 2017).

Problems Solved with SOMS Nano-Reactors: Expanding Scope and Limitations

Amide bonds link amino acids in proteins and are referred to as peptide bonds. Peptide bond formation occurs when the carboxyl end of one amino acid interacts with the amino end of another amino acid. This reaction follows the general direct amidation scheme presented in **Figure 2.1**. As demonstrated in **Figure 2.1**, water is a product of peptide bond formation and is a condensation reaction. Water molecules can also cleave peptide bonds through hydrolysis. SOMS can be used to solve problems with the direct amidation or peptide bond formation in solution by preventing water from being initially present in the reaction and by driving out water produced from bond formation (Edminston et. al. 2009).

SOMS has a hydrophobic nature and will prevent any water from entering the central cavity (Shaw 2017). This prevents water from being present to immediately hydrolyze a newly formed peptide bond. As water is produced from peptide bond formation, the hydrophobic nature of SOMS will also drive it out of the reaction cavity (Shaw 2017). This prevents water from running the reverse hydrolysis reaction. Using Le Chatelier's principle, since water is a product of the reaction its removal decreases the concentration of product and will drive the equilibrium reaction towards product formation (Brown 2014).

The Hypothesis

Vaccinations are incredibly important. There is a sub-class of vaccinations that are made from peptides that offers incredible advantages. Unfortunately, a reliable way to produce peptide vaccinations does not exist. The Shaw Research Group recently discovered a method to conduct amide bonds using SOMS nano-reactors. We believe SOMS can be used to produce peptides – sequential coupling of amino acids through amide bonds. Thus, potentially offering a method to reliably produce peptide vaccinations.

Chapter 3: The Reaction – Experimental Design

In order to explore the synthesis of peptides in SOMS, amino acid reactants were chosen such that an amino acid with a reactive amino terminus (N-terminus) will react with an amino acid containing a reactive carboxyl terminus (C-terminus). This project focuses on addition of amino acids in an N-terminus to C-terminus orientation to form the desired dipeptides. The reactive N-terminal amino acid was chosen from a group of amino acids that would not evaporate during the deprotonation reaction. Aromatic side-chain amino acids – tryptophan, phenylalanine, and tyrosine were chosen. Furthermore, to control reactivity, the N-terminus end of the C-terminal reactive amino acids were purchased with a tert-butyoxycarbonyl (BOC) on the N-terminus amino group to protect it from reacting and ensuring that the carboxyl group was free to react. The C-terminus end of the N-terminal reactive amino acids were purchased with methyl ester on the C-terminus carbonyl group to protect it from reacting and ensuring that the amino group was free to react.

Methods – Priming

To avoid evaporation, tryptophan was chosen as the N-terminus reactive amino acid. When purchased, the N-terminus reactive amino acids, were purchased as N-terminus hydrochlorides that needed to undergo priming prior to coupling. To deprotonate, tryptophan methyl ester hydrochloride (Trp-OMe·HCl) was dissolved in 1 M sodium hydroxide (NaOH). NaOH was the chosen base because when it obtains a proton it produces water and charged sodium and chloride ions (NaCl). The dissolved Trp-OMe · HCl was added to a separatory funnel and washed with dichloromethane (DCM). The deprotonated tryptophan, no longer soluble in water was extracted with an organic solvent, DCM. This formed two distinct layers in the separatory funnel. The top layer contained the aqueous layer with water soluble salt (NaCl) and the bottom layer contained

the DCM and the deprotonated Trp-OMe. Washing with DCM was repeated two more times. The combined bottom layer was dried with sodium sulfate, ensuring that no water was present in the solution.

Methods – Rotary Evaporator

To selectively separate the DCM from deprotonated Trp-OMe, the solution was placed on a rotary evaporator to evaporate DCM. The rotary evaporator lowers the pressure in the round bottom flask which lowers the boiling point of a liquid. This allows for removal of the organic solvent without excessively heating the sample which could be harmful to the deprotonated Trp-OMe. The rotary evaporator also allows for minimal co-evaporation that can occur in other methods.

Methods – Coupling

The N-terminus reactive, deprotonated, Trp-OMe was combined with the C-terminus reactive, amino protected tryptophan (BOC-Trp), SOMS, and a small amount of DCM. This process was very important in the experimental design because the organic solvent dissolved both Trp-OMe and BOC-Trp and the solvent which allowed for the molecules to easily enter SOMS. SOMS swelled to accommodate the presence of molecules in the reaction cavity (step I, **Figure 3.1**) and once encapsulated inside SOMS, the molecules were much closer together. To bring the molecules even closer together, the reaction vessel was placed back on the rotary evaporator to remove DCM (step II, **Figure 3.1**). Removal of DCM allowed SOMS to shrink, driving the reactant amino acids closer together which increased the potential for collisions for product formation (steps III – IV, **Figure 3.1**). The shrinking or closing of the SOMS can be referred to as a flex. The wetting and rotary evaporation step was repeated three times to ensure that all reactant molecules entered SOMS and were converted into the desired dipeptide product. **Figure 3.1**

demonstrates the general mechanism of a SOMS reaction where reactants and solvent enter the central cavity of the nano-reactor, solvent is then removed to shrink SOMS (Shaw 2017). Multiple additions and extractions of solvent can occur in SOMS leading to the final product.

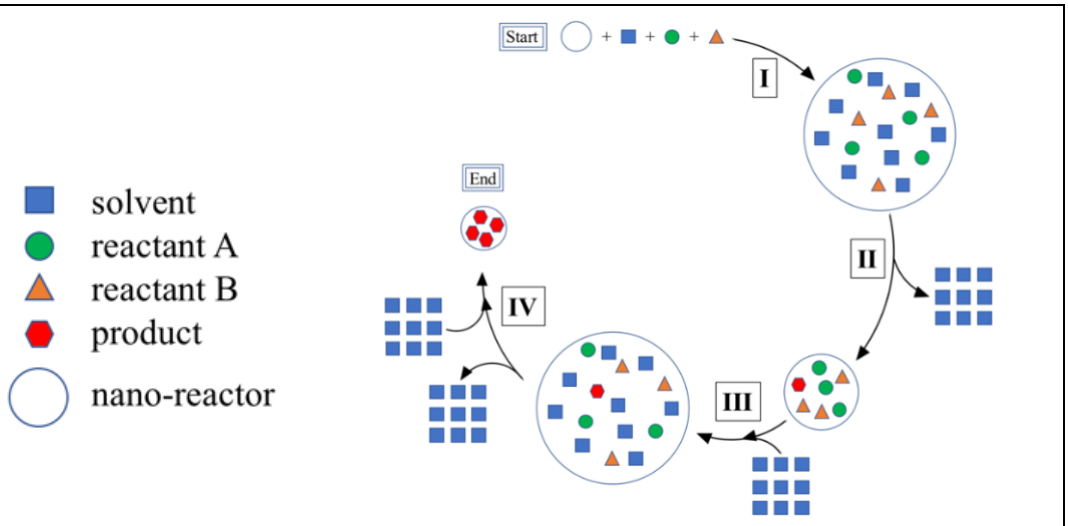


Figure 3.1. Diagram of a general SOMS reaction.

More important than the reaction procedure, is the ability of SOMS to repel water (Shaw 2017). The formation of a peptide bond is a condensation reaction which forms a molecule of water along with the desired dipeptide product. In this equilibrium reaction, the water molecule can hydrolyze a peptide bond converting the products back into their respective starting reactants. SOMS is hydrophobic and water molecules produced in the reaction are driven out of the reaction cavity (Shaw 2017). The removal of water prevents newly formed peptide bonds from being hydrolyzed and removal of product drives the equilibrium reaction towards more product formation.

Methods – Recovery of Product

After the coupling reaction the dipeptide was extracted from SOMS using DCM and vacuum filtration. To remove the dipeptide from SOMS, SOMS was flooded with copious amounts of DCM which flushed the dipeptide out of SOMS. Vacuum filtration separated the solid SOMS from the dipeptide dissolved in DCM. Flooding was repeated multiple times to ensure that all dipeptide product was flushed from SOMS. To recover the dipeptide from DCM, the resulting solution was placed on the rotary evaporator to remove DCM.

Methods – ^{1}H-NMR

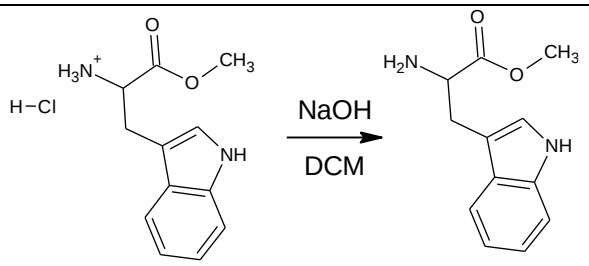
To demonstrate dipeptide formation, ^{1}H-NMR was utilized. The recovered dipeptide was suspended in deuterated chloroform ($CDCl_3$). The ^{1}H-NMR was collected on a 400 MHz Bruker nuclear magnetic resonance instrument. All ^{1}H NMR spectra contained organic contaminants which were accounted for during analysis.

Methods – Materials

Table 3.1. List of Chemicals and Compounds Used.

Name	Abbreviation	CAS #	Vendor
Sodium Hydroxide	NaOH	1310-73-2	FisherBiotech
Dichloromethane	DCM	75-09-2	Alfa Aesar
Sodium Sulfate	Na_2SO_4	7757-82-6	Alfa Aesar
Swellable Organically Modified Silica	SOMS	532987-74-9	ABS Materials
Tryptophan methyl ester HCl	Trp-OMe · HCl	7524-52-9	Acros Organics
BOC-Phenylalanine	BOC-Phe	13734-34-4	Chem-Impex INT'L INC.
BOC-Isoleucine	BOC-Ile	13139-16-7	Alfa Aesar
BOC-Alanine	BOC-Ala	15761-38-3	Acros Organics
BOC-Valine	BOC-Val	137334-41-3	Alfa Aesar

Synthesis of Deprotonated Tryptophan



Scheme 3.1. Synthesis of deprotonated tryptophan.

To a 250-mL beaker Trp-OMe · HCl (0.450 g) was dissolved in 1M NaOH (100 mL). The solution was transferred to a 500-mL separatory funnel and was washed with DCM (100 mL) 3 times. Each wash of DCM resulted in two distinct layers in the separatory funnel and the bottom layer was drained off. The combined bottom layers were dried with Na_2SO_4 and was decanted into a 100-mL round bottom flask (RBF). RBF was placed on the rotary evaporator to remove DCM (40℃ bath; 488 mbar) affording a clear oil: ^{1}H NMR (400 MHz, $CDCl_3$).

General Coupling Procedure – BOC-Phe-Trp-OMe example.

BOC-Phe (0.308 g), SOMS (2.511 g), and a small amount of DCM (1 mL) were added to the 100-mL RBF containing Trp-OMe (0.450 g). Followed by evaporation of the DCM (488 mbar, 40 ˚C bath). Aliquot addition (2x) of DCM (1 mL) was added to the reaction flask and evaporated (488 mbar, 40 ˚C). Dipeptide was extracted from SOMS by addition of DCM (10 mL) and vacuum filtration. The filtrate was placed in a 100-mL RBF and DCM (500 mbar, 40 ˚C) was evaporated to afford the desired product.

Synthesis of Phenylalanine – Tryptophan Dipeptide

BOC-Phe-Trp-OMe was afforded as a clear oil: ^{1}H NMR (400 MHz, CDCl3) δ 7.48 – 7.46 ppm, 7.35 – 7.33 ppm, 7.24 ppm, 7.22 ppm, 7.20 ppm, 7.16 ppm, 7.09 ppm, 5.11 – 4.34 ppm, 3.68 ppm, 3.66 ppm, 3.25 – 3.10 ppm, 2.97 – 2.90 ppm, and 1.39 ppm

Synthesis of Isoleucine – Tryptophan Dipeptide

BOC-Ile-Trp-OMe was afforded as a clear oil: ^{1}H NMR (400 MHz, CDCl3) δ 7.46 ppm, 7.31 ppm, 7.30 ppm, 7.06 ppm, 7.15 ppm, 5.20 – 5.21 ppm, 4.07 ppm, 3.65 ppm, 3.44 – 3.22 ppm, 1.78 ppm, 1.41 ppm, and 0.92 – 0.80 ppm,

Synthesis of Alanine – Tryptophan Dipeptide

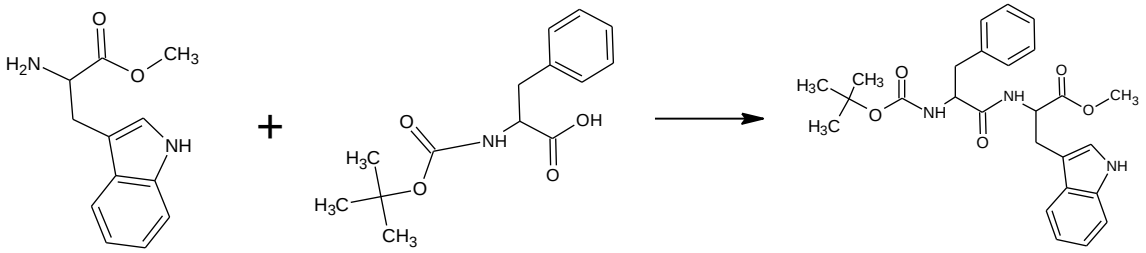
BOC-Ala-Trp-OMe was afforded as a white solid: ^{1}H NMR (400 MHz, CDCl3) δ 7.53 ppm, 7.35 ppm, 7.18 ppm, 7.13 ppm, 7.12 ppm, 4.18 – 3.96 ppm, 3.73 ppm, 3.37 – 3.17 ppm, 1.43 ppm, 1.36 ppm, and 1.37 ppm

Synthesis of Valine – Tryptophan Dipeptide

BOC-Val-Trp-OMe was afforded as a clear semi-solid: ^{1}H NMR (400 MHz, CDCl3) δ 7.37 ppm,7.20 ppm, 7.19 ppm, 7.06 ppm, 7.00 ppm, 5.39 – 4.95 ppm, 4.06 ppm, 3.67 ppm, 3.36 – 3.15 ppm, 2.07 ppm, and 0.90 – 0.73 ppm

Chapter 4: The Reaction – Proof of Principle and Scope

Proof of Principle – Synthesis of BOC -Phenylalanine- Tryptophan-OMe (BOC-Phe-Trp-OMe)



Scheme 4.1. Synthesis of BOC-Phe-Trp-OMe.

The synthesis of BOC-Phe-Trp-OMe was attempted in SOMS, **Scheme 4.1**. ^{1}H-NMR of the product mixture was obtained and compared to the respective starting materials BOC-Phe and the deprotonated Trp-OMe, **Figure 4.1**, **Table 4.1.**

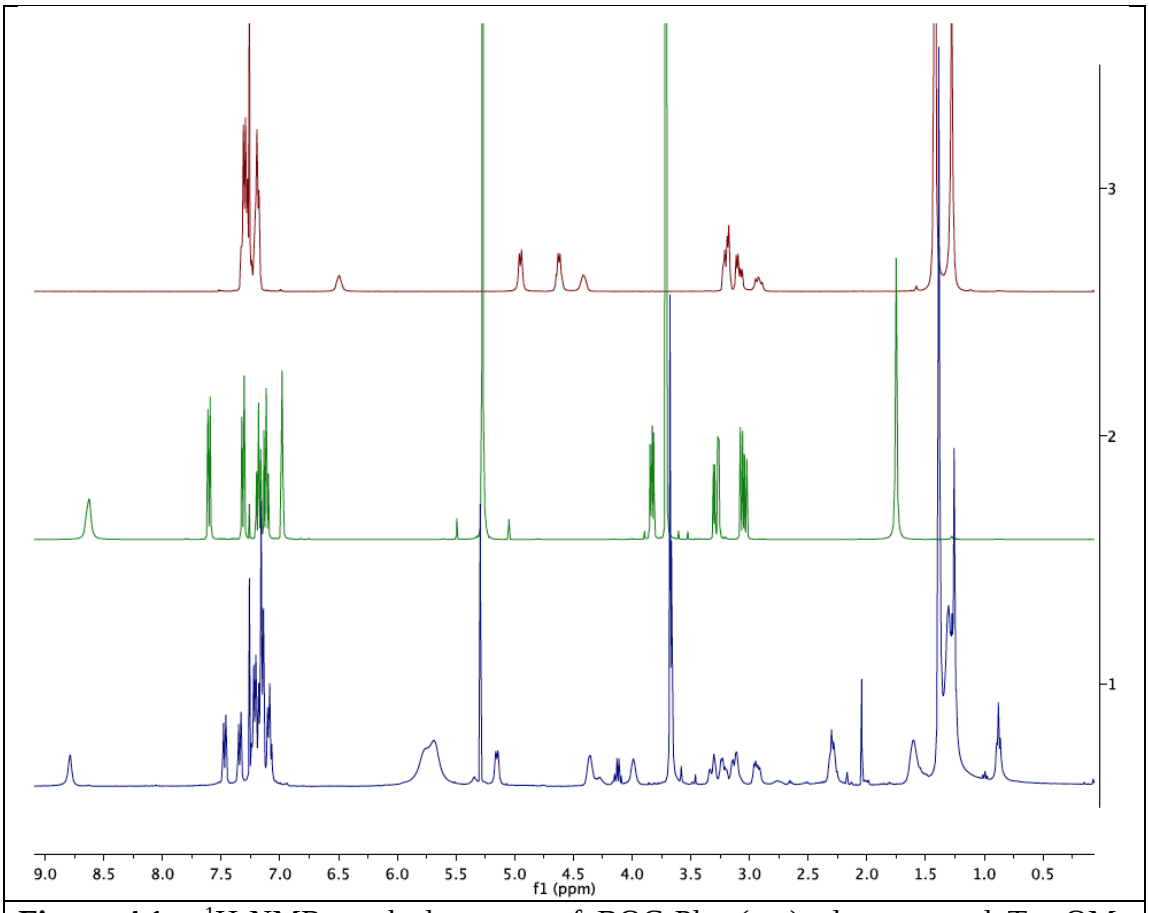
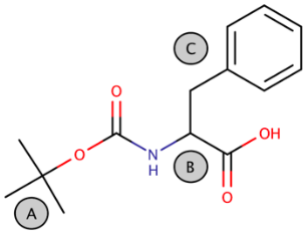
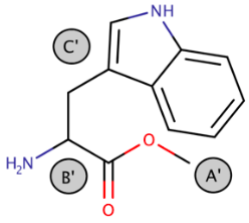
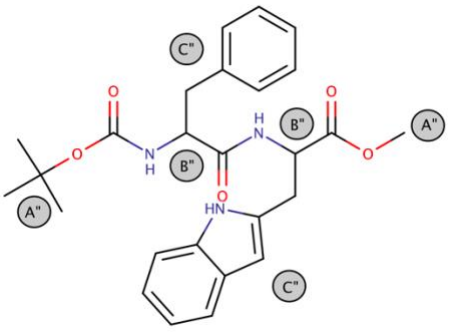


Figure 4.1. ^{1}H-NMR stacked spectra of BOC-Phe (top), deprotonated Trp-OMe (middle), and BOC-Phe-Trp-OMe (bottom).

The ^{1}H-NMR spectrum of Boc-Phe featured a total of eight distinct peaks (**Figure 4.1**). The backbone methine peak (peak B) is located at 3.12 - 3.05 ppm. The sidechain peaks (peaks C) are located at 3.12 – 3.05 ppm for the methane peaks and 7.18, 7.20, 7.27, 7.29, and 7.31 ppm for the aromatic peaks. The BOC protecting group peak is located at 1.42 ppm (peak A). The ^{1}H-NMR spectrum of the deprotected Trp-OMe was previously discussed.

Table 4.1. Summary of diagnostic BOC-Phe, Trp-OMe, and BOC-Phe-Trp-OMe ^{1}H-NMR data

	BOC-Phe		**Deprotonated Trp-OMe**		**BOC-Phe-Trp-OMe**	
	Peak Assignment	Position (ppm)	Peak Assignment	Position (ppm)	Peak Assignment	Position (ppm)
Protecting Groups	A	1.42	-	-	A”	1.39
	-	-	A’	3.71	A”	3.68
Backbone	B	4.97-4.63	-	-	B”	5.11-4.34
	-	-	B’	3.83	B”	3.66
Side Chain	C	3.12-3.05	-	-	C”	2.97-2.90
	C	7.18	-	-	C”	7.20
	C	7.20	-	-	C”	7.22
	C	7.27	-	-	C”	7.20
	C	7.29	-	-	C”	7.22
	C	7.31	-	-	C”	7.24
	-	-	C’	3.30-3.05	C”	3.25-3.10
	-	-	C’	7.30-7.28	C”	7.35-7.33
	-	-	C’	7.61-7.61	C”	7.48-7.46
	-	-	C’	7.18	C”	7.16
	-	-	C’	7.12	C”	7.09
	-	-	C’	6.98	C”	7.16

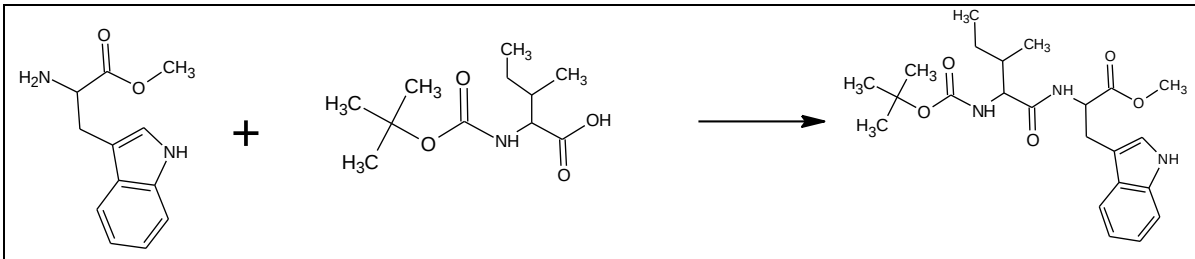
BOC-Phe	**Deprotonated Trp-OMe**	**BOC-Phe-Trp-OMe**
		

The ^{1}H-NMR spectrum of the product mixture from the synthesis of BOC-Phe-Trp-OMe featured sixteen peaks. Evidence of the amino acid di-peptide backbone is present as the ^{1}H-NMR spectrum features peaks at 5.11 – 4.34 ppm and 3.66 ppm. The ^{1}H-NMR spectrum also contains evidence of both phenylalanine and tryptophan side chains – the ^{1}H-NMR spectrum features peaks representative of the side chain of phenylalanine at 2.97 – 2.90 ppm, 7.20, 7.22 and 7.24 ppm as well as the side chain of tryptophan at 3.25 – 3.10 ppm, 7.35 – 7.33 ppm, 7.48 -7.46 ppm, 7.09, and 6.16 ppm. Furthermore, evidence of the presence of the BOC and methoxy protecting groups is clear as the ^{1}H-NMR spectrum features a peak at 1.39 ppm and 3.68 ppm, respectively.

Furthermore, presence of the peptide bond existing between the two amino acids is clear. A shift in peak position is indicative of peptide bond formation as the protons in the dipeptide are in different chemical environments than in their respective monomeric amino acids. A shift in the back bone methine peaks from 4.97-4.63 ppm (peak B) to 5.11-4.34 ppm (peak B") and from 3.83 ppm (peak B') to 3.66 ppm (peak B") suggest peptide bond formation. Protons that are adjacent to the peptide backbone also experience a shift in the ^{1}H-NMR spectrum. Peaks C and C' (3.12 - 3.05 ppm and 3.30 - 3.05ppm, respectively) are shifted from their respective starting material ^{1}H-NMR – these peaks appear at 2.97 - 2.90 ppm and 3.25 - 3.10 ppm). The remainder of the peaks do not experience significant changes to their chemical environment and appear in near identical places in both dipeptide and monomer ^{1}H-NMR spectra.

The presence and position of the peaks in the ^{1}H-NMR spectrum of BOC-Phe-Trp-OMe suggest successful dipeptide formation.

Expanding the Scope – Synthesis of BOC-Ile-Trp-OMe



Scheme 4.2. Synthesis of BOC-Ile-Trp-OMe.

The synthesis of BOC-Ile-Trp-OMe was attempted in SOMS, **Scheme 4.2**. ^{1}H-NMR of the product mixture was obtained and compared to the respective starting materials BOC-Ile and the deprotonated Trp-OMe, **Figure 4.2**, **Table 4.2**.

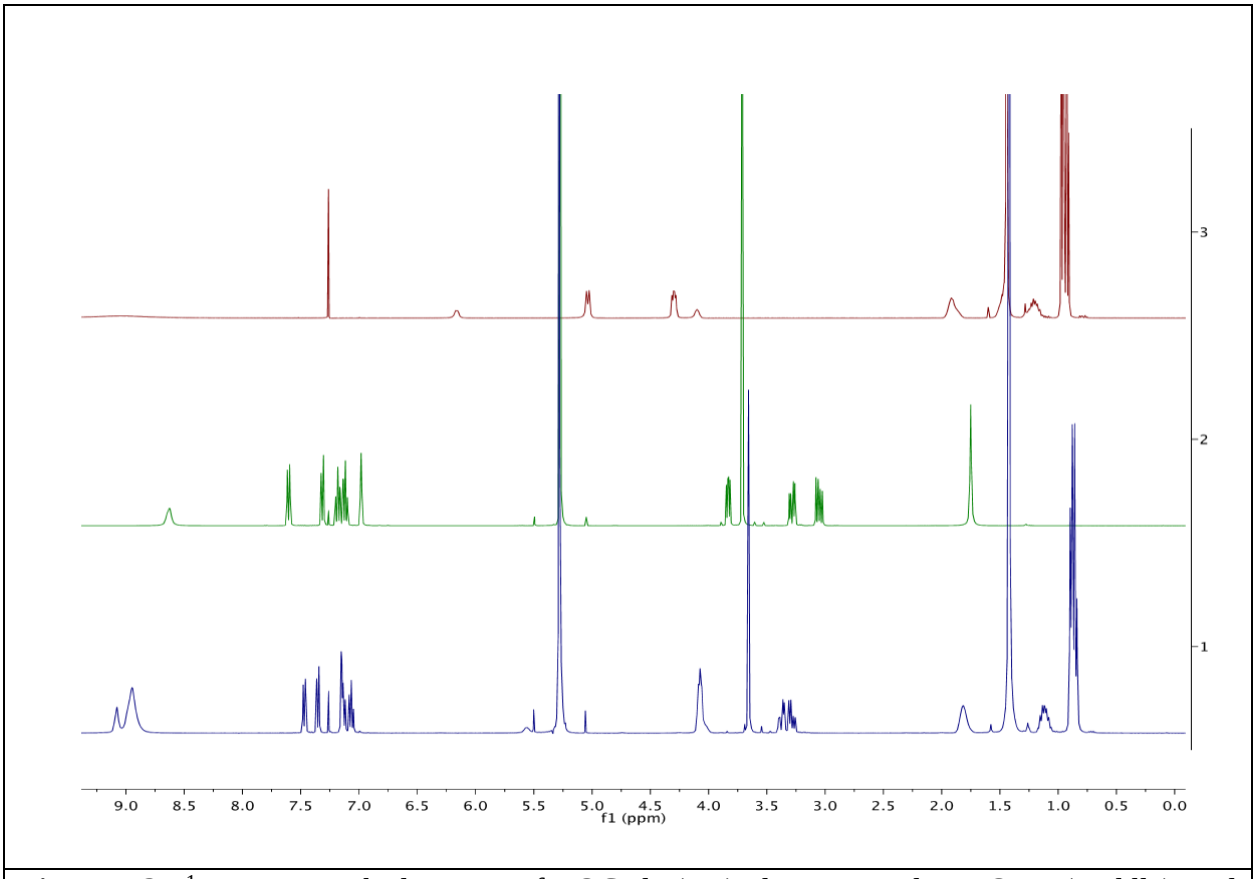
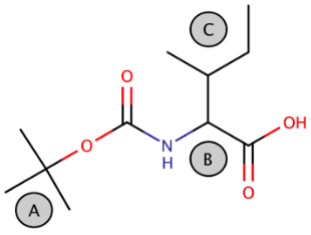
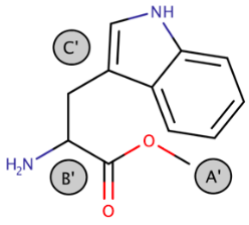
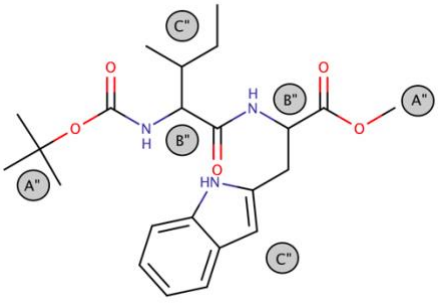


Figure 4.2. ^{1}H-NMR stacked spectra of BOC-Ile (top), deprotonated Trp-OMe (middle), and BOC-Phe-Ile-OMe (bottom).

The ^{1}H-NMR spectrum of Boc-Ile featured a total of four distinct peaks (**Figure 4.2**). The backbone methine peak (peak B) is located at 5.01 – 4.27 ppm. The sidechain peaks (peaks C) are located at 1.88 and 1.02 – 0.90 ppm for the methane peaks. The BOC protecting group peak is located at 1.42 ppm (peak A). The ^{1}H-NMR spectrum of the deprotected Trp-OMe was previously discussed.

Table 4.2. Summary of diagnostic BOC-Ile, Trp-OMe, and BOC-Ile-Trp-OMe ^{1}H-NMR data

	BOC-Ile		**Deprotonated Trp-OMe**		**BOC-Ile-Trp-OMe**	
	Peak Assignment	Position (ppm)	Peak Assignment	Position (ppm)	Peak Assignment	Position (ppm)
Protecting Groups	A	1.42	-	-	A”	1.41
	-	-	A’	3.71	A”	3.65
Backbone	B	5.01–4.27	-	-	B”	5.50-5.21?
	-	-	B’	3.83	B”	4.07
Side Chain	C	1.88	-	-	C”	1.78
	C	1.02–0.90	-	-	C”	0.92-0.80
	-	-	C’	3.30-3.05	C”	3.44-3.22
	-	-	C’	7.60	C”	7.46
	-	-	C’	7.30	C”	7.30
	-	-	C’	7.18	C”	7.31
	-	-	C’	7.12	C”	7.06
	-	-	C’	6.98	C”	7.15

BOC-Ile	**Deprotonated Trp-OMe**	**BOC-Ile-Trp-OMe**
		

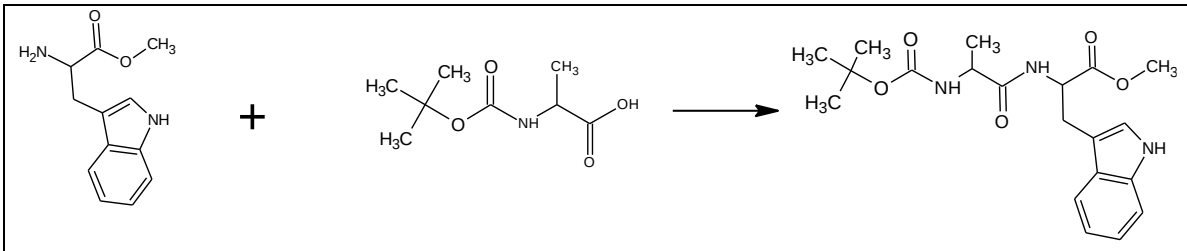
The ^{1}H-NMR spectrum of the product mixture from the synthesis of BOC-Ile-Trp-OMe featured twelve peaks. Evidence of the amino acid di-peptide backbone is present as the ^{1}H-NMR spectrum features peaks at 5.50 – 5.21 ppm and 4.07 ppm. The isoleucine part of the back bone is

slightly over shadowed by the DCM solvent peak. The peaks presence is apparent however the range is not as precise. The ^{1}H-NMR spectrum also contains evidence of both isoleucine and tryptophan side chains – the ^{1}H-NMR spectrum features peaks representative of the side chain of isoleucine at 1.88 ppm and 1.02 – 0.90 ppm as well as the side chain of tryptophan at 3.44 – 3.22 ppm, 7.46 ppm, 7.30 ppm, 7.31 ppm, 7.06 ppm, and 7.15 ppm. Furthermore, evidence of the presence of the BOC and methoxy protecting groups is clear as the ^{1}H-NMR spectrum features a peak at 1.41 ppm and 3.65 ppm, respectively.

Furthermore, presence of the peptide bond existing between the two amino acids is clear. A shift in peak position is indicative of peptide bond formation as the protons in the dipeptide are in different chemical environments than in their respective monomeric amino acids. A shift in the back bone methine peaks from 5.01 – 4.27 ppm (peak B) to roughly 5.50 - 5.21 ppm (peak B") and from 3.83 ppm (peak B') to 4.07 ppm (peak B") suggest peptide bond formation. Protons that are adjacent to the peptide backbone on the tryptophan amino acid also experience a shift in the ^{1}H-NMR spectrum. Peaks C' 3.30 - 3.05 ppm are shifted from their respective starting material ^{1}H-NMR – these peaks appear at 3.44 – 3.22 ppm. The remainder of the peaks do not experience significant changes to their chemical environment and appear in near identical places in both dipeptide and monomer ^{1}H-NMR spectra.

The presence and position of the peaks in the ^{1}H-NMR spectrum of BOC-Ile-Trp-OMe suggest successful dipeptide formation.

Expanding the Scope – Synthesis of BOC-Ala-Trp-OMe



Scheme 4.3. Synthesis of BOC-Ala-Trp-OMe.

The synthesis of BOC-Ala-Trp-OMe was attempted in SOMS, **Scheme 4.3**. ^{1}H-NMR of the product mixture was obtained and compared to the respective starting materials BOC-Ala and the deprotonated Trp-OMe, **Figure 4.3**, **Table 4.3**.

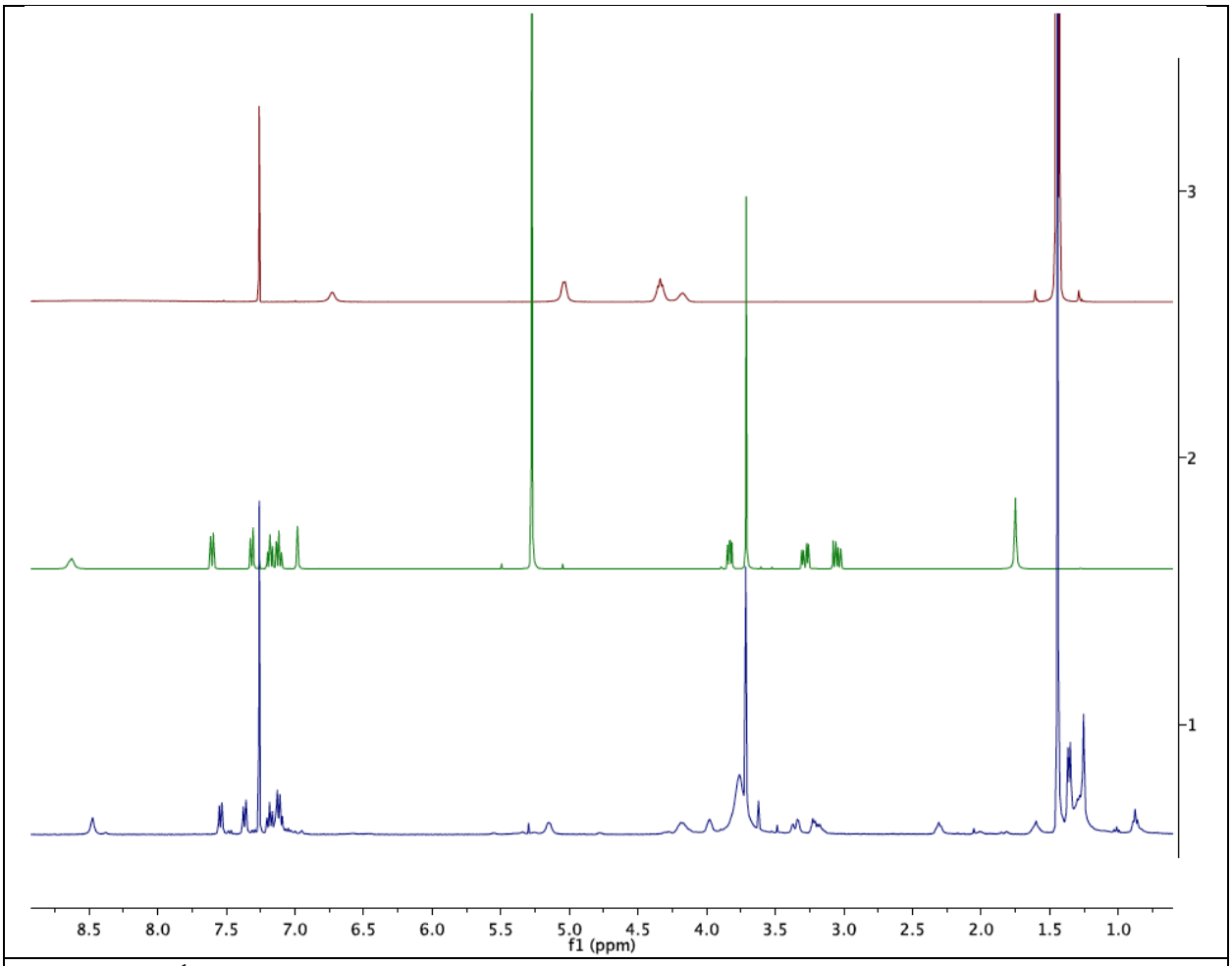
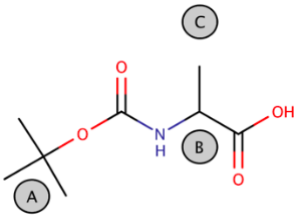
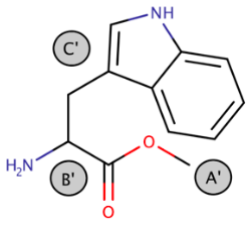
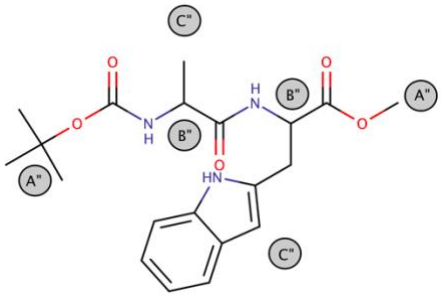


Figure 4.3. ^{1}H-NMR stacked spectra of BOC-Ala (top), deprotonated Trp-OMe (middle), and BOC-Phe-Ala-OMe (bottom).

The ^{1}H-NMR spectrum of Boc-Ala featured a total of eight distinct peaks (**Figure 4.3**). The backbone methine peak (peak B) is located at 4.30 - 3.18 ppm. The sidechain peaks (peaks C) are located at 1.43 ppm for the methane peaks. The BOC protecting group peak is located at 1.45 ppm (peak A). The ^{1}H-NMR spectrum of the deprotected Trp-OMe was previously discussed.

Table 4.3. Summary of diagnostic BOC-Ala, Trp-OMe, and BOC-Ala-Trp-OMe ^{1}H-NMR data

	BOC-Ala		**Deprotonated Trp-OMe**		**BOC-Ala-Trp-OMe**	
	Peak Assignment	Position (ppm)	Peak Assignment	Position (ppm)	Peak Assignment	Position (ppm)
Protecting Groups	A	1.45	-	-	A”	1.43
	-	-	A’	3.71	A”	1.36
Backbone	B	4.30-4.18	-	-	B”	4.18-3.96
	-	-	B’	3.83	B”	3.73?
Side Chain	C	1.43	-	-	C”	1.37
	-	-	C’	3.30-3.05	C”	3.37-3.17
	-	-	C’	7.60	C”	7.53
	-	-	C’	7.30	C”	7.35
	-	-	C’	7.18	C”	7.18
	-	-	C’	7.12	C”	7.12
	-	-	C’	6.98	C”	7.13

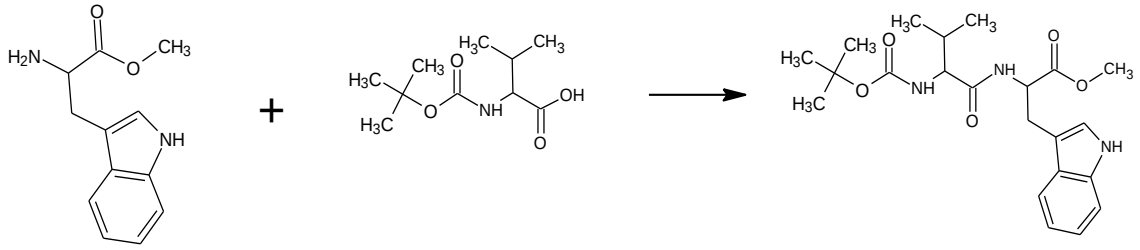
BOC-Ala	**Deprotonated Trp-OMe**	**BOC-Ala-Trp-OMe**
		

The ^{1}H-NMR spectrum of the product mixture from the synthesis of BOC-Ala-Trp-OMe featured eleven peaks. Evidence of the amino acid di-peptide backbone is present as the ^{1}H-NMR spectrum features peaks at 4.18 – 3.96 ppm and 3.73 ppm. The peak for the tryptophan carbon back bone is overshadowed by the alanine side chain peak. The ^{1}H-NMR spectrum also contains evidence of both alaine and tryptophan side chains – the ^{1}H-NMR spectrum features peaks

representative of the side chain of phenylalanine at 1.43 ppm as well as the side chain of tryptophan at 3.37 - 3.17 ppm, 7.53 ppm, 7.35 ppm, 7.18, 7.12 and 7.13 ppm. Furthermore, evidence of the presence of the BOC and methoxy protecting groups is clear as the ^{1}H-NMR spectrum features a peak at 1.43 ppm and 1.36 ppm, respectively.

Furthermore, presence of the peptide bond existing between the two amino acids is clear. A shift in peak position is indicative of peptide bond formation as the protons in the dipeptide are in different chemical environments than in their respective monomeric amino acids. A shift in the back bone methine peaks from 4.30 - 4.18 ppm (peak B) to 4.18 - 3.96 ppm (peak B") and from 3.83 ppm (peak B') to 3.73 ppm (peak B") suggest peptide bond formation. Protons that are adjacent to the peptide backbone also experience a shift in the ^{1}H-NMR spectrum. Peaks C" at 3.37 - 3.17 ppm is shifted from their respective starting material ^{1}H-NMR – these C' peaks appear at 3.30 - 3.05 ppm. Peaks C" (7.53 ppm, 7.35 ppm, 7.18 ppm, 7.12 ppm, and 7.13 ppm) are shifted from their respective starting material ^{1}H-NMR – these peaks appear at 7.60 ppm, 7.30 ppm, 7.18 ppm, 7.12 ppm, and 6.98 ppm respectively.

The presence and position of the peaks in the ^{1}H-NMR spectrum of BOC-Phe-Trp-OMe suggest successful dipeptide formation.



Scheme 4.4. Synthesis of BOC-Val-Trp-OMe.

The synthesis of BOC-Val-Trp-OMe was attempted in SOMS, **Scheme 4.4**. ^{1}H-NMR of the product mixture was obtained and compared to the respective starting materials BOC-Ala and the deprotonated Trp-OMe, **Figure 4.4**, **Table 4.4**.

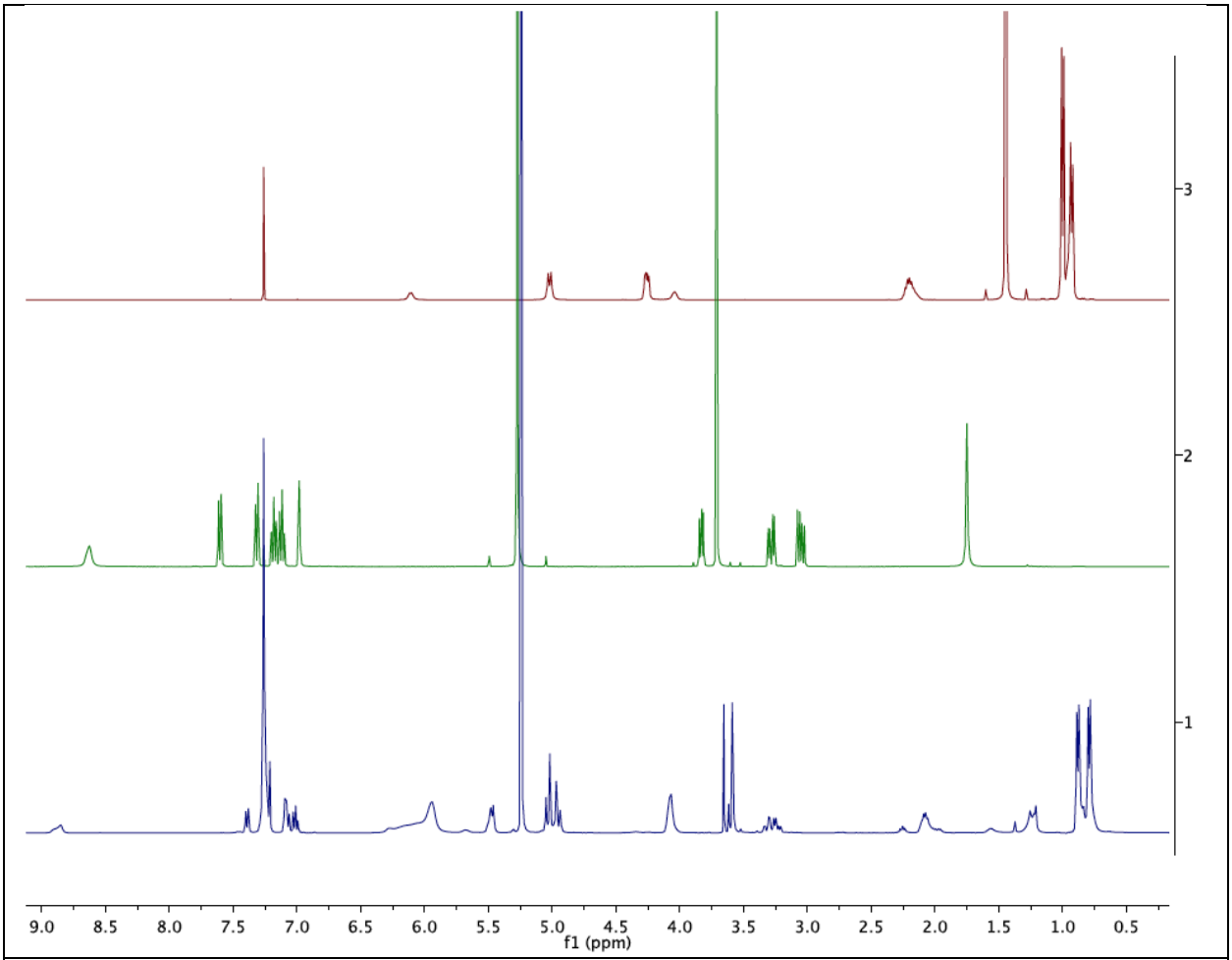
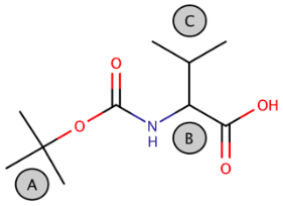
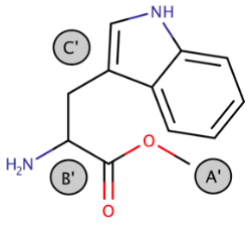
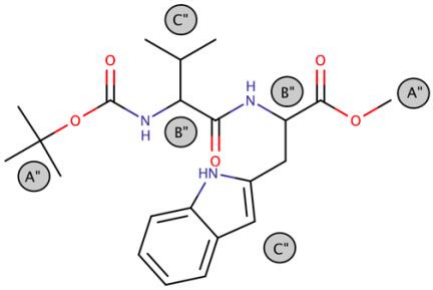


Figure 4.4. ^{1}H-NMR stacked spectra of BOC-Val (top), deprotonated Trp-OMe (middle), and BOC-Val-Trp-OMe (bottom).

The ^{1}H-NMR spectrum of Boc-Val featured a total of four distinct peaks (**Figure 4.4**). The backbone methine peak (peak B) is located at 5.02 – 4.26 ppm. The sidechain peaks (peaks C) are located at 2.20 ppm and 1.03 – 0.88 for the methane peaks. The BOC protecting group peak is located at 1.44 ppm (peak A). The ^{1}H-NMR spectrum of the deprotected Trp-OMe was previously discussed.

Table 4.4. Summary of diagnostic BOC-Val, Trp-OMe, and BOC-Val-Trp-OMe ^{1}H-NMR data

	BOC-Val		**Deprotonated Trp-OMe**		**BOC-Val-Trp-OMe**	
	Peak Assignment	Position (ppm)	Peak Assignment	Position (ppm)	Peak Assignment	Position (ppm)
Protecting Groups	A	1.44	-	-	A”	?
	-	-	A’	3.71	A”	3.67
Backbone	B	5.02-4.26	-	-	B”	5.39-4.95
	-	-	B’	3.83	B”	4.06
Side Chain	C	2.20	-	-	C”	2.07
	C	1.03 – 0.88	-	-	C”	0.90-0.73
	-	-	C’	3.30-3.05	C”	3.36-3.15
	-	-	C’	7.60	C”	7.37
	-	-	C’	7.30	C”	7.19
	-	-	C’	7.18	C”	7.06
	-	-	C’	7.12	C”	7.00
	-	-	C’	6.98	C”	7.20

BOC-Val	**Deprotonated Trp-OMe**	**BOC-Val-Trp-OMe**
		

The ^{1}H-NMR spectrum of the product mixture from the synthesis of BOC-Val-Trp-OMe featured eleven peaks. Evidence of the amino acid di-peptide backbone is present as the ^{1}H-NMR spectrum features peaks at 5.39 – 4.95 ppm and 4.06 ppm. The ^{1}H-NMR spectrum also contains

evidence of both valine and tryptophan side chains – the ^{1}H-NMR spectrum features peaks representative of the side chain of valine at 2.07 ppm and 0.90 – 0.73 ppm as well as the side chain of tryptophan at 3.36 - 3.15 ppm, 7.37 ppm, 7.19 ppm, 7.06 ppm, 7.00 ppm and 7.20 ppm. Furthermore, evidence of the presence of the methoxy protecting group is clear as the ^{1}H-NMR spectrum features a peak at 3.67 ppm. The BOC protecting group is not apparent on the dipeptide ^{1}H NMR spectra. All of the other peaks suggest di-peptide formation so it possible that the BOC was cleaved in peptide bond formation.

Furthermore, presence of the peptide bond existing between the two amino acids is clear. A shift in peak position is indicative of peptide bond formation as the protons in the dipeptide are in different chemical environments than in their respective monomeric amino acids. A shift in the back bone methine peaks from 5.02 - 4.26 ppm (peak B) to 5.39 - 4.95 ppm (peak B") and from 3.83 ppm (peak B') to 4.06 ppm (peak B") suggest peptide bond formation. Protons that are adjacent to the peptide backbone for the tryptophan amino acid also experience a shift in the ^{1}H-NMR spectrum. Peaks C" 3.36 – 3.15 is shifted from their respective starting material ^{1}H-NMR – these peaks appear at 3.30 – 3.05 ppm. The remainder of the peaks do not experience significant changes to their chemical environment and appear in near identical places in both dipeptide and monomer ^{1}H-NMR spectra.

The presence and position of the peaks in the ^{1}H-NMR spectrum of BOC-Val-Trp-OMe suggest successful dipeptide formation.

Chapter 5: Future Work and Applications

Peptide Bonds and Merrifield Synthesis

Peptide bonds form between the carboxyl end (C-terminus) of an amino acid and the amino end (N-terminus) of the next amino acid (Brown 2014). Peptide bond formation is a condensation reaction that releases water. A water molecule can also hydrolyze a peptide bond. This type of bond does not form in solution. The leading synthetic method of peptide synthesis is Merrifield synthesis (McMurry 2012). Updated papers document the procedure yields as 65 – 92% (Boojamra 2004). for each step. Merrifield synthesis has been optimized since its conception however there is still room for improvement (Boojamra 2004). Through the proposed method several steps can be eliminated effectively decreasing the amount of time that it takes to run both a single step and the entire reaction. Decreasing the number of steps can also increase production because less starting material is lost overall (Shaw 2017).

The proposed method is unique and is theorized to run at 100% yield. If complete closing is achieved, the nature of SOMS should allow for complete reactant conversion into product (Shaw 2017). Furthermore, if all reactant material is driven into the reaction cavity of SOMS then 100% product formation should be achieved (Shaw 2017). Current ^{1}H NMR evidence shows lack of starting material in the spectrum suggesting that the reaction ran to completion. Other research projects in the research group appear to have similar results. More experiments are needed to statistically prove 100% yield. A quantification study will be performed in the future to show exact product yields. This aspect could be applied to Merrifield synthesis. If only dipeptides were able to be synthesized in SOMS, they could still be introduced into Merrifield synthesis procedure rather than single amino acids. This could cut the number of steps and significantly increase yield.

One of the major drawbacks to Merrifield synthesis is that it cannot produce a high enough quantity of peptide/proteins to be used in therapeutics (Boojamra 2004; Merrifield 1985). This is something that the proposed procedure wishes to address. The focus of the thesis has been on dipeptide formation and at this stage even dipeptide formation in 100% yield could greatly improve the efficiency of Merrifield synthesis. If dipeptides were added as "monomers" in Merrifield Synthesis rather than single amino acids, then proteins of greater length in higher yields could be synthesized. With future work on the dipeptide library and in trimer formation this project could improve Merrifield synthesis or form a completely independent process that could change peptide and protein roles in therapeutics.

Future Work

Next steps in the research project would include expanding the dipeptide library to including all types of amino acids coupled with the initial deprotonated Tryptophan. Amino acids with charged side chains would be attempted to be coupled with an unprotected side chain and with a protected side chain. Once the library has been established further work into tripeptide formation would proceed. This work would focus on cleaving a protection group off the c-terminus and adding a deprotonated amino acid in SOMS. A more thorough study of protection cleavage would be needed to ensure that the peptide bond is not cleaved in the process. This procedure could be modified from the de-protection cleavage observed in Merrifield synthesis – where the protection group is removed without breaking the peptide bond (Merrifield).

Long term goals for the proposed method would include experimentally determining the upper limit for protein length synthesize inside of SOMS. Correct folding of proteins would potentially utilize chaperon proteins in solution and could be confirmed by x-ray crystallography.

Applications: Cancer and Vaccines

The war on cancer started in the 1970's and there have been may advances in treatment. However, cancer is still a prevalent disease and a leading cause of death (Weinberg 2013). Some types of cancers are caused by viruses. These viruses tend to use reverse transcription for replication and thus insert their genome into host cells and use host cell machinery for the replication process (Weinberg 2013). Several vaccines, like Cervarix® for HPV and the multi-epitope melanoma peptide vaccine, have peptides in them (Senger et. al 2009). A purely synthetic procedure for producing peptides destined for vaccines would potentially reduce the need for extensive purification and growth of organisms to produce these proteins. This would allow for a purer sample of peptide/protein for the vaccine potentially reducing side effects to impurities in the vaccine.

This method could also reduce costs in several areas. SOMS reactions are fast and are to run reactions at high yields (Shaw 2017). This cuts down on production time saving costs in energy, staffing the lab, and in reactant material. SOMS reactions do not need to run at very high temperatures or pressure again saving in energy costs (Shaw 2017). After a reaction, SOMS can be cleaned and reused for later experiments cutting costs for reaction materials.

References:

Brown, Theodore L. *Chemistry: the Central Science*. Prentice Hall, 2014.

Boojamra, Constantine G., Kristina M. Burow, and Jonathan A. Ellman. "An expedient and high-yielding method for the solid-phase synthesis of diverse 1, 4-benzodiazepine-2, 5-diones." *The Journal of Organic Chemistry* 60.18 (1995): 5742-5743.

Derouazi, Madiha, et al. "Serum-free large-scale transient transfection of CHO cells." *Biotechnology and Bioengineering* 87.4 (2004): 537-545.

Eisemann, Joan, and Kim, Sung Woo. *Animal Nutrition*. Kendall Hunt Publishing Company, 2014.

Edmiston, Paul L., and Laura A. Underwood. "Absorption of dissolved organic species from water using organically modified silica that swells." *Separation and Purification Technology* 66.3 (2009): 532-540

Edmiston, Paul L., et al. "Pilot scale testing composite swellable organosilica nanoscale zero-valent iron—Iron-Osorb®—for in situ remediation of trichloroethylene." *Remediation Journal* 22.1 (2011): 105-123.

Hopkins, Ralph F., Vanessa E. Wall, and Dominic Esposito. "Optimizing transient recombinant protein expression in mammalian cells." *Protein expression in mammalian cells: methods and protocols* (2012): 251-268.

Inoue, Hiroaki, Hiroshi Nojima, and Hiroto Okayama. "High efficiency transformation of Escherichia coli with plasmids." *Gene* 96.1 (1990): 23-28.

Kalinski, Cédric, et al. "On the industrial applications of MCRs: molecular diversity in drug discovery and generic drug synthesis." *Molecular diversity* 14.3 (2010): 513-522.

Kost, Thomas A., J. Patrick Condreay, and Donald L. Jarvis. "Baculovirus as versatile vectors for protein expression in insect and mammalian cells." *Nature biotechnology* 23.5 (2005): 567-575.

McMurry, Susan, and John McMurry. *Study Guide and Student Solution Manual, Organic Chemistry, Eighth Edition, John McMurry*. Brooks/Cole Cengage Learning, 2012.

Merrifield, Robert Bruce. "Solid phase synthesis (Nobel lecture)." *Angewandte Chemie International Edition in English* 24.10 (1985): 799-810.

Rosano, Germán L., and Eduardo A. Ceccarelli. "Recombinant protein expression in Escherichia coli: advances and challenges." *Recombinant protein expression in microbial systems* 7 (2014).

Shaw, Nicolas N. *Unpublished Works and Papers.* 2017

Senger, Tilo, et al. "Enhanced papillomavirus-like particle production in insect cells." *Virology* 388.2 (2009): 344-353.

Skwarczynski, Mariusz, and Istvan Toth. "Peptide-based synthetic vaccines." *Chemical Science* 7.2 (2016): 842-854.

Weinberg, Robert A. *The Biology of Cancer, Second Edition, Robert A. Weinberg.* Garland Science, 2013.

Wurm, Florian, and Alain Bernard. "Large-scale transient expression in mammalian cells for recombinant protein production." *Current Opinion in Biotechnology* 10.2 (1999): 156-159.