TORC Bonding Pairs as an Alternative to Nucleobases in Self Replicating Polymers

Presented by Aevi van der Stok

In partial fulfillment of the requirements for graduation with the Dean's Scholars Honors Degree in Chemistry

Abstract:

The search for life on other planets is limited due to having only an incomplete knowledge of the origins of life on Earth as reference. While genetic information is stored and replicated by RNA and DNA using nucleobase chemistry here on Earth, this may not be the case on other planets with different environmental conditions. Tunable Orthogonal Reversible Covalent (TORC) bonds have promise in the creation of sequence-specific replicators because their orthogonality allows for the TORC bonding pairs to function similarly to nucleotide bases, defining and replicating the sequence, while their reversibility allows for duplexes to be separated for replication and their increased strength as covalent bonds would make these replicators more suitable to hot environments than the hydrogen bonding interactions observed in DNA replicators. This project represents the first step in the creation of sequence-specific TORC replicators by demonstrating the templating and synthesis of short peptide strands from a template strand using the hydrazone TORC bonding pair. In addition to this, macrocyclic products were also produced using these templates. This unexpected result holds promise for the creation of macrocyclic TORC replicators. The creation of macrocycles also has therapeutic applications, as macrocyclic peptides have useful properties but are generally difficult to synthesize, and the templates created in this project can quantitatively produce macrocyclic peptide products without any unwanted side products.

Table of Contents

BACKGROUND

One of the many challenges in the search for extraterrestrial life is that the only frame of reference we have for what constitutes "life" is what can be found on our own planet, and the origins of life on Earth are still not well understood.[1] By exploring and expanding upon our current definition of life, we can gain a greater understanding of how it came to be and what forms it could take on other planets. Currently, life is most often defined as something that has the capability to reproduce, sometimes with the added caveat that reproduction errors must be possible and reproducible themselves.[1] On Earth, we see this in the replication of nucleic acids such as DNA and RNA. In DNA the nucleotide bases pair up with each other through hydrogen bonding interactions, holding the double helix together while still being reversible enough that the strands can be separated for replication. However, one cannot assume that life on other planets would utilize the exact same chemistry as does on Earth. Different conditions and available compounds may affect how life could, or could not, potentially arise on another planet. For example, hydrogen bonding is an intermolecular force, not a true bond, and therefore it is rather weak and can be denatured by higher temperatures, rendering it less viable for planets significantly hotter than Earth. Therefore, the study and creation of sequence-specific replicators that utilize different chemistries from nucleic acids or are favorable in environments different from those found on Earth contribute to our understanding of the origins of life and the forms that life could theoretically take on other planets.

This project sought to explore alternatives to nucleobase chemistry in the synthesis of sequence-specific and self-replicating polymers, building upon previous research from the Anslyn group on tunable orthogonal reversible covalent (TORC) bonding pairs. The goal was to synthesize short peptide strands that can template and self-replicate like DNA, but that utilize dynamic covalent bonds instead of hydrogen bonding. Dynamic covalent bonds are of interest for this project because while they are strong covalent bonds, they can also be made highly reversible depending on the conditions.[5] This would allow for the creation of duplexes that are strongly bound to one another and unlikely to separate undesirably, but that could still be separated from one another for replication. In order to template and replicate specific sequences however, there must also be orthogonal pairing of these bonds.
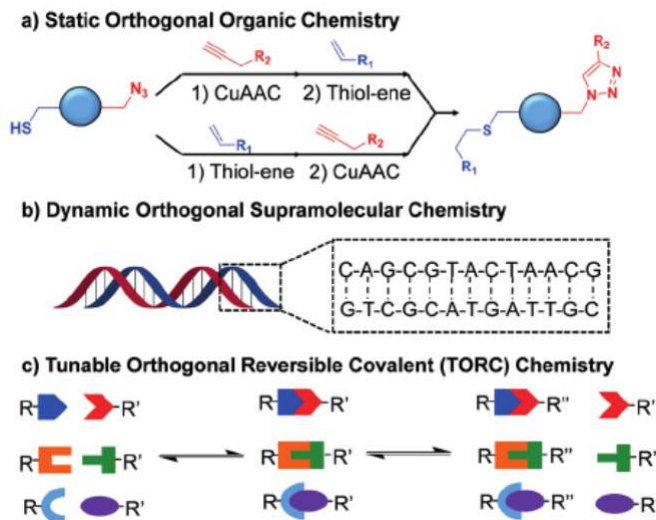
Figure 1: Orthogonality as observed in a) static orthogonal reactions b) DNA nucleobases c) TORC bonding pairs[4]
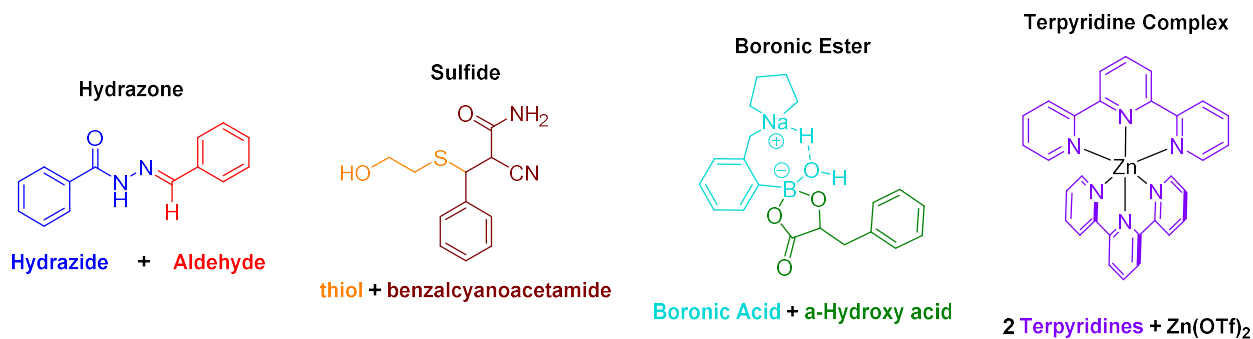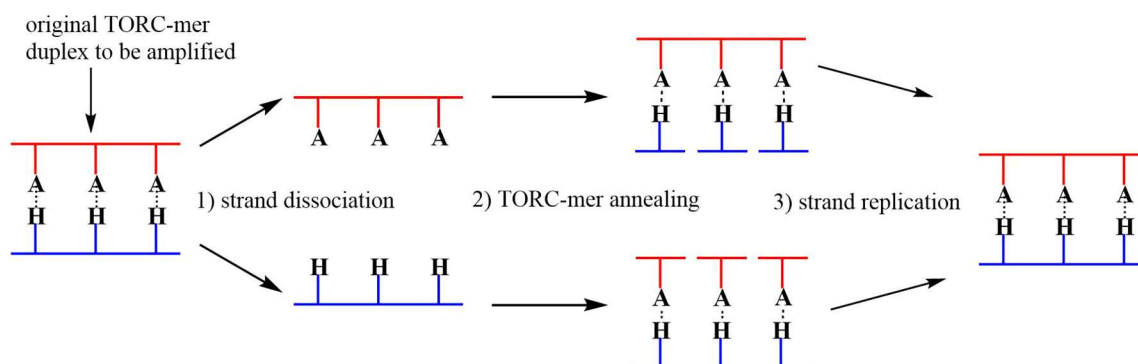

Figure 2: TORC Bonding Pairs

The Anslyn lab previously studied a set of four Tunable Orthogonal Reversible Covalent (TORC) bonding pairs, which can fulfill this purpose.[4,5] In addition to having the other qualities described for dynamic covalent bonds, they are also orthogonal.[4,5] The TORC bonds form only between bonding pairs, even if they are all mixed together in one solution.[4,5] This is crucial for the creation of sequence-specific replicators using dynamic covalent bonds, as the orthogonality of the TORC bonding pairs allows for them to fulfill a similar role as nucleotide base pairs, ensuring that the original sequence of monomers is reproduced in newly formed duplexes. The existence of four TORC bonding pairs, compared to only two nucleotide base pairs, also means that these replicators would be capable of higher information storage density than DNA. While future study is intended

to investigate the use of all four TORC bonding pairs in replicators, experiments so far focus specifically on the hydrazone TORC bonding pair, consisting of an aldehyde and a hydrazide.


Scheme 1: PCR-like replication of TORC-mer strand duplexes

In order to test the viability of TORC bonding pairs in self-replicating molecules, the goal of this project was to perform PCR-like replication of short peptide strands consisting of aldehyde and hydrazide TORC monomers, as shown in Scheme 1. The replication cycle would start with a duplex of two conjugate strands bound together in a solution of free monomers, then the strands would then be dissociated, allowing for the TORC monomers to anneal to the strands. Finally, the templated monomers would be coupled to one another, producing two copies of the original duplex. Further study would incorporate a mix of aldehyde and hydrazide TORCs on each strand, followed by the introduction of other TORC bonding pairs. Before attempting a full replication cycle however, the initial goal was to first successfully template hydrazide monomers onto aldehyde template strands of varying length in order to generate the hydrazide conjugate strands.

SYNTHESIS
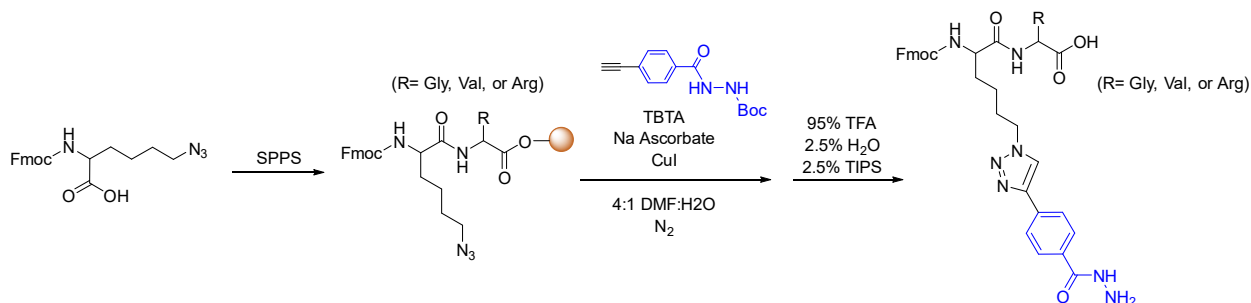
Scheme 2: Synthesis of hydrazide monomer (H) from azido lysine

5
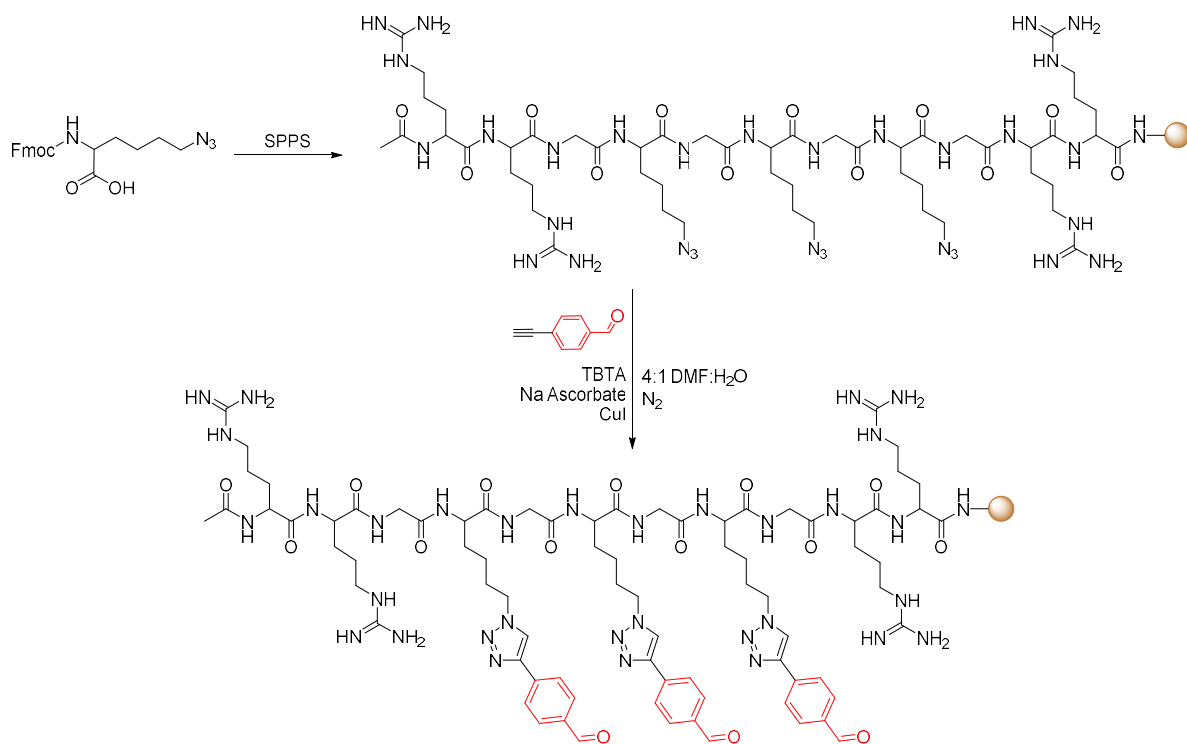
**Scheme 3:** Synthesis of aldehyde template strand (AAA)

The base structure of the monomers and strands were synthesized from azido lysine on a CEM Liberty Blue Automated Microwave Peptide Synthesizer using DIC and Oxyma as the activator and activator base and 20% piperidine in DMF as the deprotecting agent. The monomers were synthesized with Wang resin on a 0.25 mmol scale and the template strands with Rink Amide resin on a 0.1 mmol scale. In order for the TORC bonding groups to be correctly spaced and oriented, one amino acid was needed between each TORC bonding group. Therefore, an additional amino acid was added to the azido lysine to produce the base structure of the monomers, and the template strands were synthesized with a glycine between each azido lysine, as well as two arginines on each end to improve solubility. In later experiments, a second amino acid was added to the monomer, for reasons that will be explained in the results section.

The aldehyde and hydrazide TORCs were then added using copper click reactions. A catalytic solution was generated by mixing 5 mL of 40 mol% TBTA in DMF, 2 mL of 40 mol% sodium ascorbate in H_2O, and 1 mL of 20 mol% CuI in DMF, and stirred under nitrogen until evenly mixed. This catalytic solution was then added to a mixture of the resin bound monomer or

strand and the appropriate alkyne in 2 mL of DMF, and the reaction was stirred or agitated overnight. The copper click reactions are performed with the monomer or strand still on resin, as this allows for the product to be easily isolated by simply filtering off the catalytic solution and washing the resin with DMF to remove any excess reagent.

The hydrazide monomers needed to be off resin in order to be able to freely bind to the template and react with one another to form new strands, so after being washed and filtered they were cleaved from the resin using a solution of 95% trifluoracetic acid, 2.5% deionized water, and 2.5% triisopropyl silane. In earlier experiments the template strands were also cleaved from the resin after synthesis using a solution of 90% TFA and 10% deionized water, however in later experiments the templates were left on resin in order to create a pseudo dilution. The synthesis of a template strand with three aldehyde TORC bonding groups (AAA) is shown in Scheme 3, as that was the template length most frequently utilized, however template strands containing two or four aldehyde bonding groups (AA and AAAA) were also synthesized for these experiments by similar methods.

RESULTS
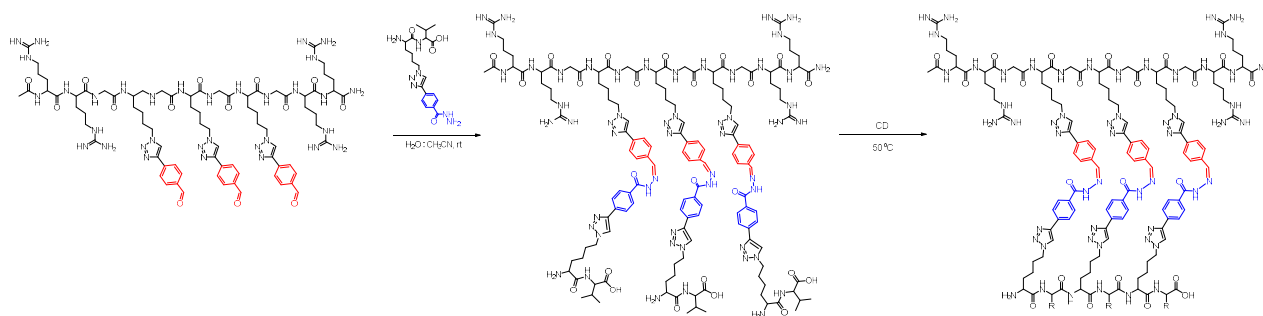
Scheme 4: Initial linear templating experiments

The initial templating experiments were performed by mixing the aldehyde template strands with a slight excess of hydrazide monomers in a mixture of water and acetonitrile, and allowing the mixture to template overnight. The monomers were coupled to one another at 50 °C with CDI as the coupling reagent. The products, already in HPLC water and acetonitrile, were then analyzed by LCMS, as well as MALDI-TOF.
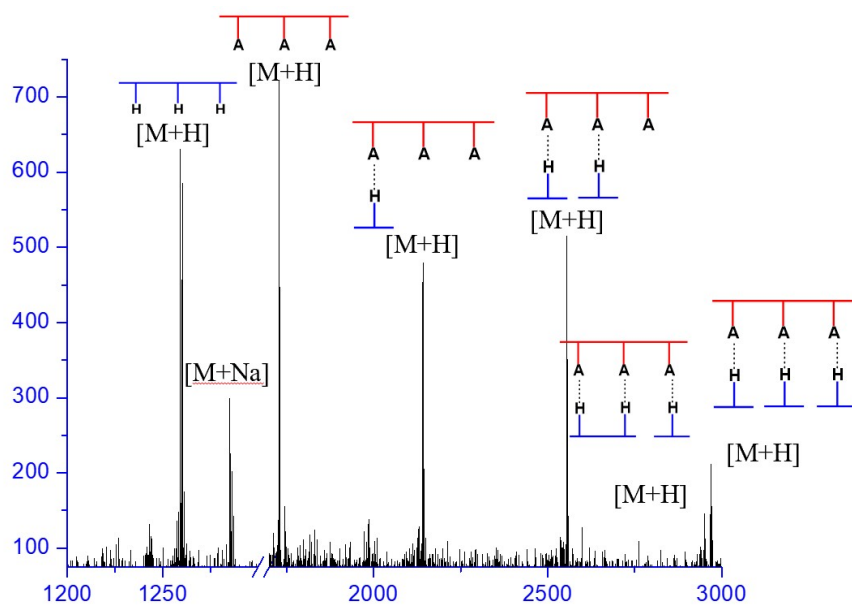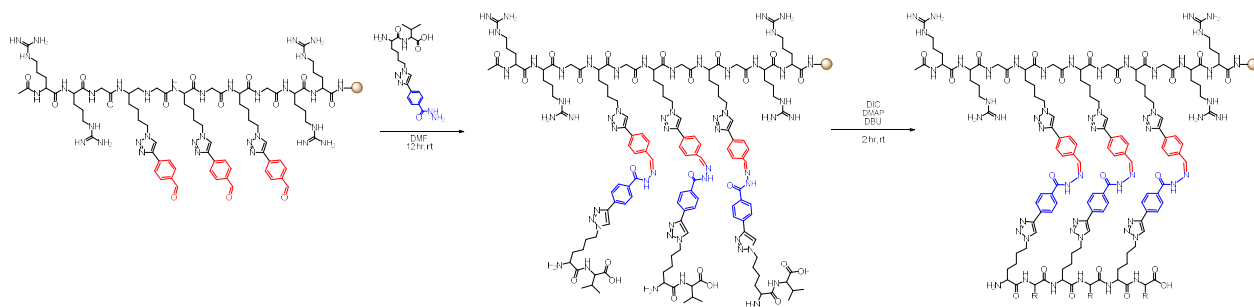
Figure 3: MALDI-TOF of templating experiment using AAA template and CDI coupling reagent

The preliminary results in templating hydrazide monomers on a triple aldehyde template strand were promising, with the desired duplexes being successfully produced. However, template strands with only one or two hydrazide monomers annealed or with all hydrazides annealed but not coupled to one another were also observed in the MALDI-TOF analysis of the products. Additionally, the hydrazide TORCS seemed to be reacting with the carbonyl on the CDI coupling reagent, causing them to precipitate out. In order to optimize the reaction, we tried adjusting various different conditions such as the coupling reagent, temperature, and reaction time. We also began to perform the templating experiments with the template strand still on resin, creating pseudo dilution conditions.

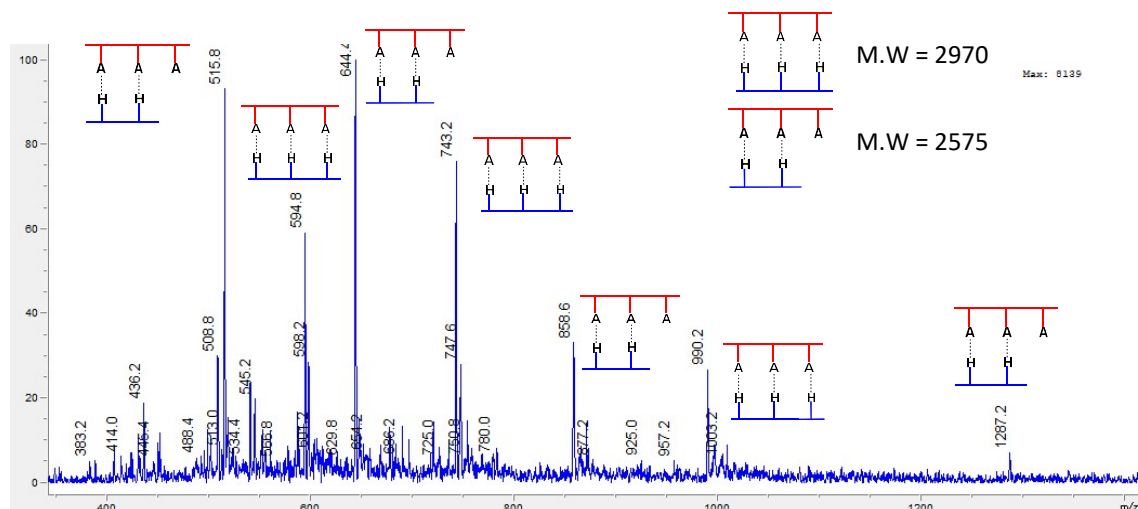**Scheme 5:** Adjusted linear templating experiments

Figure 4: LCMS of templating experiment using AAA template and DIC coupling reagent

Templating experiments with CDI, EDC, and DIC coupling reagents were attempted, with DIC generating the best results. The best reaction time using the DIC coupling reagent was determined to be a period of two hours, however the temperature of the reaction did not seem to significantly affect results. While changing the coupling reagent did improve results somewhat, the templating still did not go all the way to completion. LCMS analysis of the products showed the presence of duplexes which had a strand of only two hydrazides bound to them, as opposed to all three (Figure 4). We suspected that this could be due to steric hinderance from the valine side chain on the monomer, and decided to experiment with different monomer side chains in an attempt to achieve better yields.
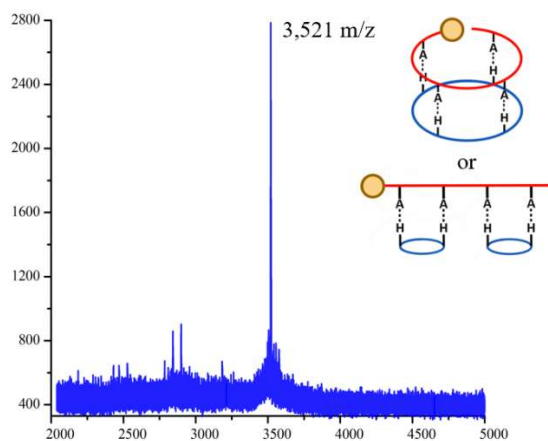

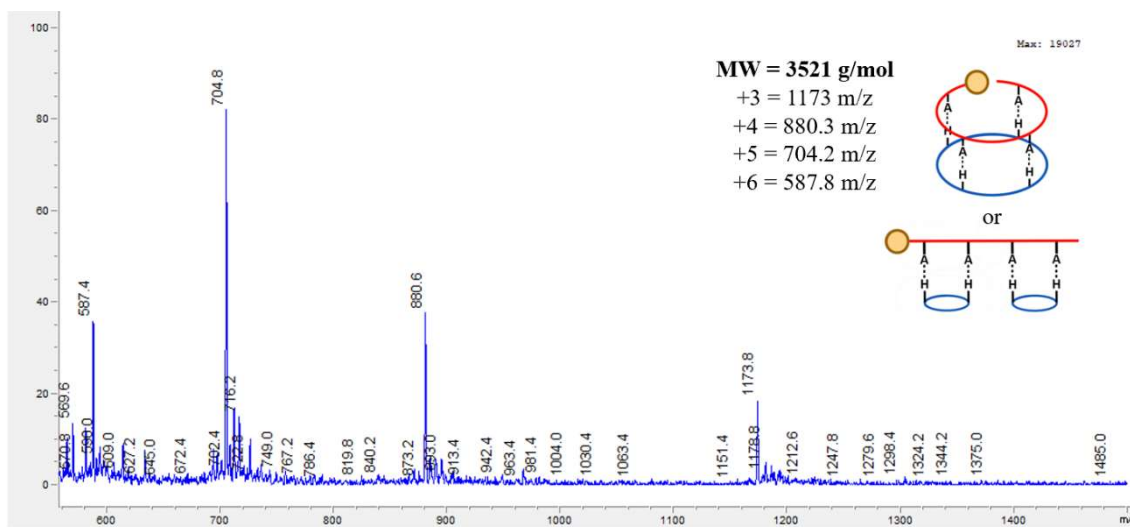**Figure 5:** MALDI-TOF of templating experiment using AAAA template and DIC coupling reagent

9

Figure 6: LCMS of templating experiment using AAAA template and DIC coupling reagent

The templating experiments were repeated with glycine as the side chain monomer, as it would not contribute any steric hinderance. This modification increased the product yield to over 80%, however something else unexpected occurred. LCMS and MALDI-TOF analysis (Figures 5 and 6) revealed that the hydrazide 'strands' annealed to the aldehyde template had lost the mass of four waters when coupled, creating four peptide bonds between the hydrazides, as opposed to losing the mass of three waters as expected, forming three peptide bonds. This means that instead of coupling in a solely linear fashion to create strands, the monomers were wrapping around to form an additional peptide bond, creating macrocycles. Considering that one of the possible long-term aims of this project was to create macrocyclic TORC replicators as well, we decided to pursue this further.
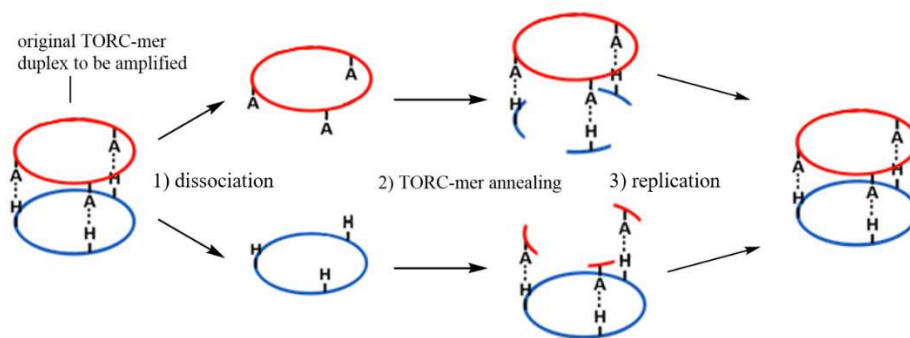
Scheme 6: PCR-like replication of TORC-mer macrocycle duplexes

10

Here the goals of the project shifted in order to determine if the creation of cyclic vs linear products could be controlled, followed by repeating the templating experiments to intentionally produce only the macrocyclic products. Templating experiments beginning with a macrocycle template could then be attempted, followed by PCR-like replication of the macrocycle duplex (Scheme 6). It was also necessary to determine whether or not the macrocycles being formed were larger four membered rings, or smaller two membered rings, as both have the same mass observed on the LCMS and MALDI-TOF (Figures 5 and 6). While the project moved on to pursue the macrocycle angle further, linear template strands were still used to create these macrocycles, as macrocycles are difficult to produce via conventional peptide synthesis methods.
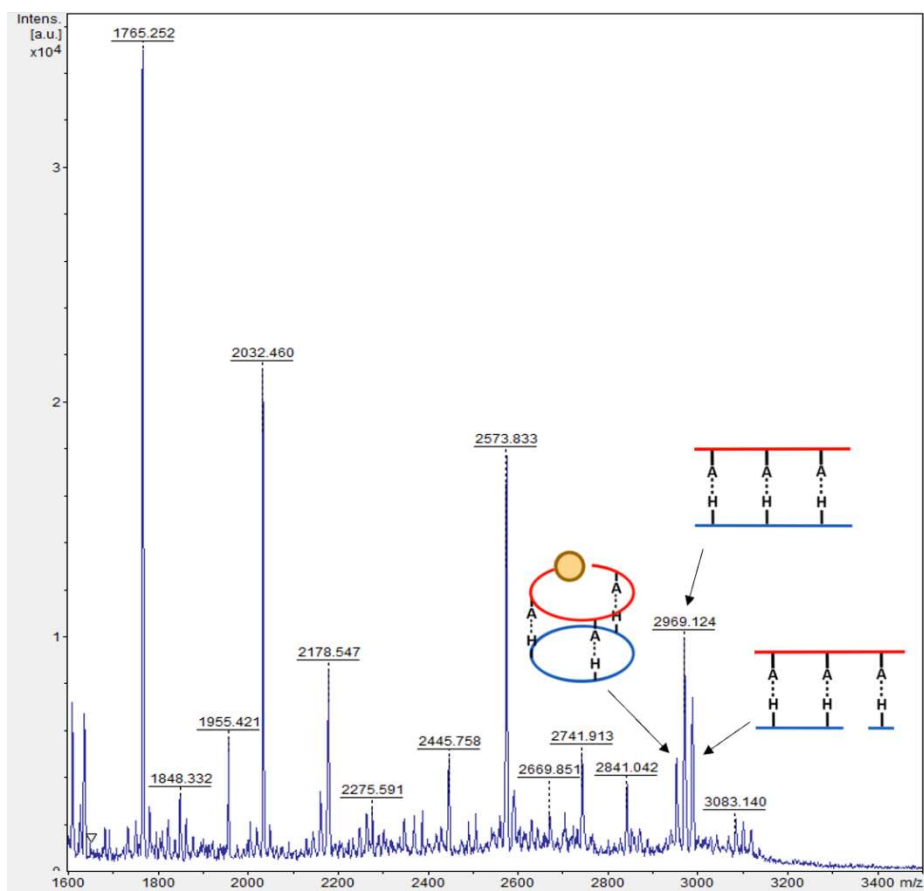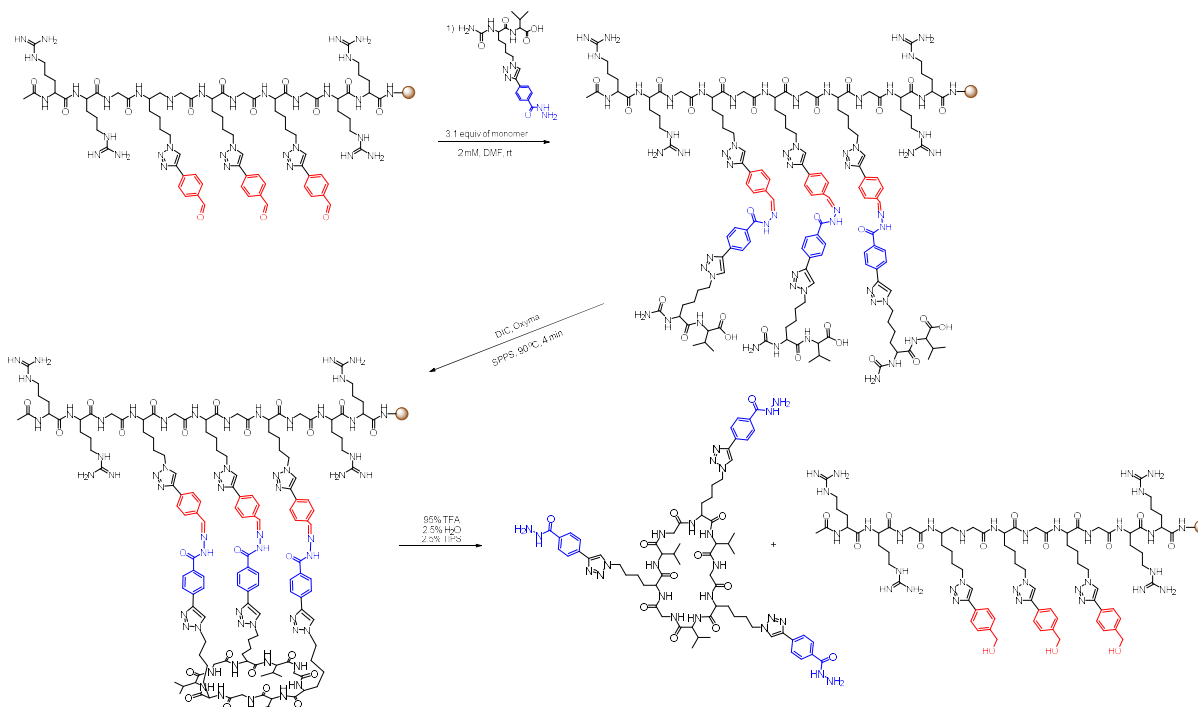

Figure 7: LCMS of macrocycle templating experiment using AAA template and DIC coupling reagent

The macrocyclic product (cHHH) was produced on the AAA template using monomers with glycine, indicating that the monomers were not just forming smaller two hydrazide macrocycles, but could produce larger macrocycles as well. However, the linear product and a

product with incomplete coupling between the hydrazides were also observed. The amino acid could not be further adjusted to improve yield, as glycine already has the lowest steric hinderance of the amino acids. However, adding a second amino acid onto the monomer, slightly increasing its length, did make the macrocyclic product more favorable. The coupling step was also optimized by performing it in the solid phase peptide synthesizer using DIC and Oxyma at 90 ºC.

Scheme 7: Macrocycle templating experiments

The final macrocycle templating experiments were performed by mixing the resin bound aldehyde template strands with a 2 mM solution of hydrazide monomers in slight excess in DMF, and allowing the monomers to anneal onto the template for 1 hour. The resin was then filtered and washed with DMF to remove excess monomer, and the N terminus of the monomers were deprotected with 10% DBU. The monomers were then coupled to one another in the microwave assisted solid phase peptide synthesizer at 90 °C for 4 minutes with DIC and Oxyma. The hydrazide product was released from the template strand using a solution of 95% trifluoracetic acid, 2.5% deionized water, and 2.5% triisopropyl silane. This also reduced the aldehydes on the template strands to alcohols, preventing the product from potentially binding back on to the template. While this is convenient for analyzing the product in this stage of the project, a better method of separating the two strands that does not reduce the aldehydes would be needed to

12

complete the full replication cycle. Products were analyzed using LCMS or extracted ion chromatography. As with the linear experiments, while only the template strand with three aldehydes (AAA) is shown in Scheme 7, similar methods were also used to conduct experiments with template strands that had two or four aldehyde TORCs (AA and AAAA).
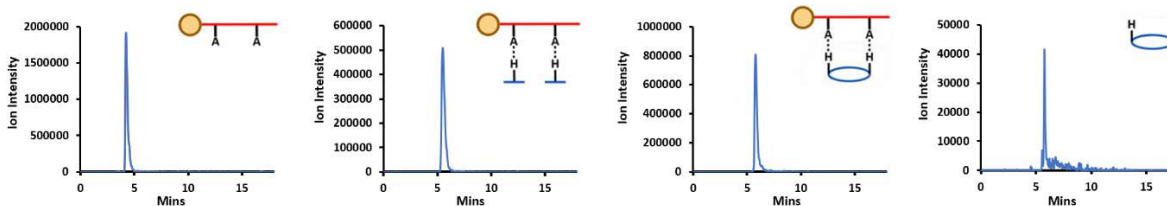

Figure 8: Extracted ion chromatographs for each step of templating experiment with AA template and DIC coupling reagent
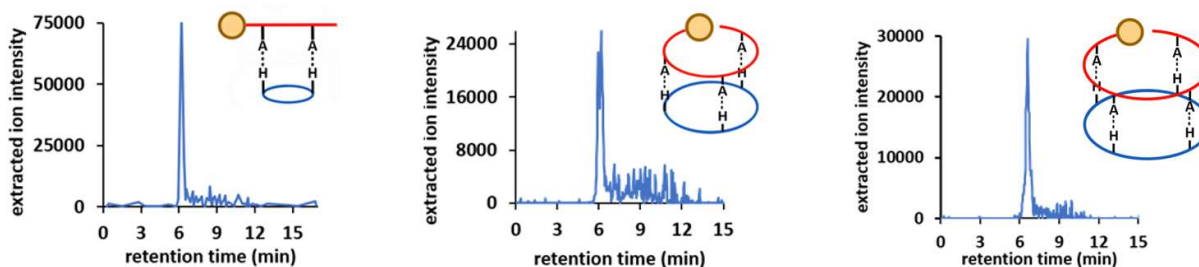

Figure 9: Extracted ion chromatograms of macrocycle templating experiment products

Using this procedure, the two hydrazide macrocycle (cHH) was successfully synthesized from the template without any undesired linear side product. The extracted ion chromatographs of each step of the templating experiment, shown in Figure 8, demonstrate that the hydrazide monomers successfully and completely annealed to the template, coupled to form the macrocycle, and were released from the template. This experiment was then repeated again with the AAA and AAAA templates, and the desired three hydrazide (cHHH) and four hydrazide (cHHHH) macrocyclic products were produced. These macrocycles were also synthesized quantitatively without any unwanted linear side products, nor was the formation of any smaller macrocycles observed.
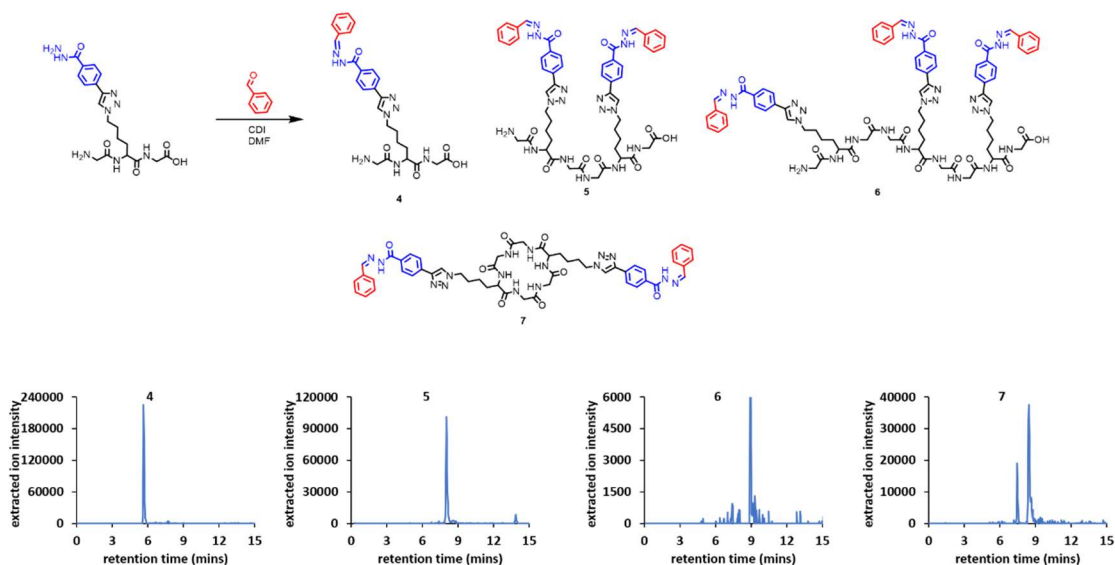
Figure 10: Extracted ion chromatograms of control study products

A control experiment was also performed to ensure that the creation of the macrocycles could be attributed to the template. In the control, the hydrazide monomers and benzaldehyde were subjected to the same reaction conditions as the templating experiments with the same coupling reagents, and allowed to react. This produced an assortment of products including a single hydrazide monomer bound to the benzaldehyde, a strand formed from two hydrazide monomers, a strand formed from three hydrazide monomers, and the two hydrazide macrocycle. The single hydrazide monomer bound with the benzaldehyde was the most prevalent product.

DISCUSSION

Our results demonstrate the successful templating and synthesis of linear or macrocyclic peptide strands using the hydrazone TORC bonding pair. While a full replication cycle for either the linear or macrocyclic duplexes has not yet been attempted, the templates successes in producing the correct products without unwanted side products suggest that replication should be possible. Had this project been able to continue, the next step would have been to synthesize macrocycles from the templates that could then be used as macrocyclic templates themselves. This would allow for the full PCR-like replication cycle of the macrocycle duplexes to be attempted. Additionally, the linear templating experiments could also be revisited and further optimized in order to try the full replication cycle with linear duplexes as well. It would also likely be necessary

to perform a series of melting studies in order to optimize the separation of the duplex for replication. Further study towards the creation of true sequence specific TORC replicators would involve repeating the replications with strands or macrocycles that have a mix of aldehyde and hydrazide TORCs. These experiments could then also be repeated with other TORC bonding pairs, eventually leading up towards the replication of strands or macrocycles that contain multiple different TORC bonding pairs.

The fact that macrocyclic products were accidentally produced instead of linear strands also has significance in understanding the origins of life. Some studies into RNA as a prebiotic replicator and form of genetic information storage have encountered similar problems with cyclic products competing with or preventing the formation of the desired linear products.[6] The same results appearing when working with replicators that do not utilize nucleobase chemistry could suggest that competing linear and macrocyclic products may be a reoccurring phenomenon in the origins of life, even when in different environments or utilizing different chemical interactions.

The synthesis of macrocyclic peptide produces using a linear template strand also has therapeutic applications. Cyclic peptides can exhibit improved binding affinity compared to linear peptides and are more resistant to proteases, however they are often difficult to synthesize.[2,3,8] The methods used in this project could be adapted and utilized to synthesize macrocyclic peptides more easily. Additionally, other studies have shown that peptides can also be cyclized or stabilized into a desired conformation by "stapling" the ends or parts of the peptide together via various chemical strategies, many of which require expensive or toxic metal catalysts.[3,7] TORC bonding pairs could also potentially be utilized in this manner to cyclize peptides by "stapling" the ends together via a hydrazide TORC on one end of the peptide and an aldehyde TORC on the other end.

ACKNOWLEDGEMENTS

Additionally, a special thanks to Elvira "Doc" de Pieri. You inspired me and started me on my journey into chemistry and I would not be who or where I am today without your guidance and encouragement. Although sadly you are no longer here for me to say this too, I hope you know how important you were to me and how thankful I am to have known you.

REFERENCES

1. Benner, S. A. *Astrobiology* **2010**, *10* (10), 1021.
2. Chow, H. Y.; Zhang, Y.; Matheson, E.; Li, X. *Chemical Reviews* **2019**, *119* (17), 9971.
3. Montgomery, J. E.; Donnelly, J. A.; Fanning, S. W.; Speltz, T. E.; Shangguan, X.; Coukos, J. S.; Greene, G. L.; Moellering, R. E. *Journal of the American Chemical Society* **2019**, *141* (41), 16374.
4. Reuther, J. F.; Dahlhauser, S. D.; Anslyn, E. V. *Angewandte Chemie International Edition* **2018**, *58* (1).
5. Seifert, H. M.; Trejo, K. R.; Anslyn, E. V. *Journal of the American Chemical Society* **2016**, *138* (34), 10916.
6. Sosson, M.; Pfeffer, D.; Richert, C. *Nucleic Acids Research* **2019**, *47* (8), 3836.
7. Tan, Y. S.; Lane, D. P.; Verma, C. S. *Drug Discovery Today* **2016**, *21* (10), 1642.
8. White, C.; Yudin, A. *Nature Chemistry* **2011**, *3*, 509.