Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Robert Devin Boğ April 11th, 2020

Creation of a Functional Dynamic Chemical Network

by

R. Devin Boğ

David G. Lynn
Adviser

Department of Biology

David G. Lynn
Adviser

Gregg M. Orloff
Committee Member

W. Michael Caudle
Committee Member

2020

Creation of a Functional Dynamic Chemical Network

By

R. Devin Boğ

David G. Lynn
Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Biology

2020

Abstract
Creation of a Functional Dynamic Chemical Network
By R. Devin Boğ

Life is a coherent system of myriad chemical networks we refer to as *metabolisms*. Metabolisms can be understood as systems that channel energy to generate molecular information, including sequence-specific polymers. The *de novo* synthesis of dynamic chemical networks (DCNs) which capture metabolic function promises to reveal new insights into the fundamental molecular principles of biological function. In this work, new designs for DCNs with expanded function and efforts in their synthesis are described. In particular, the synthesis and generation of peptide aldehydes via solid-phase peptide synthesis is detailed.

Creation of a Functional Dynamic Chemical Network

By

R. Devin Boğ

David G. Lynn
Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Biology

2020

Acknowledgements

This research was supported by grants from NSF CHE-1507932 and NSF/DMR-BSF 1610377, and NIH Alzheimer's Disease Research Center: P50AG025688.

I'd like to thank my committee: Dr. David G. Lynn, Dr. Gregg M. Orloff, and Dr. W. Michael Caudle, for their time and advice. A thousand apologies for the lack of catering at the defense. I hope future readers of this thesis will never have to know the pains of living without an integrated platform for food delivery and online meetings.

I thoroughly thank Youngsun Kim for his camaraderie and direct mentorship, and the entirety of the Lynn Lab: Anthony Sementilli, Christella Dhammaputri, Ayanna Jones, Jipeng Yue, Nicholas Ingle, Regina Luu, and Daisy.

I'd like to thank all of my professors in the sciences and beyond who have pushed me to enjoy the fruits of rigorous intellectual pursuits. In particular, I'd like to thank Dr. Arri Eisen for helping me engage in undergraduate research.

I acknowledge the wide cadre of fellow Emory students and friends who have provided invaluable moral support throughout the process of writing this thesis: namely, Lauren Karr and Shivaang Chawla. I love you guys.

Mom, Baba and Deren, thanks for letting me come to Emory. I love you guys, too.

Table of Contents

INTRODUCTION

Life requires the continuous function of systems we refer to as metabolisms. Biology, as we know it, uses a limited group of polymer scaffolds to create functional molecules: nucleic acids, amino acids, fatty acids and carbohydrates. Out of these molecules, it creates metabolisms: complex molecular networks which are distinguished by their ability to adapt to their environment. Our group's work is focused on determining if there are other ways, beyond those which have been identified, to create such metabolisms; can we use systems that don't polymerize, for example, via amide bonds or phosphodiester bonds, to generate chemical networks that can change their chemical makeup in response to their environment? If we can create novel metabolic systems from the ground up, we might be able to better understand the design principles that underlie biology *and* create new functional materials outside of the natural paradigm [1].

Within biology, amino acid polymers are involved in catalysis, structure, and other cellular functions, while DNA stores information that can be accessed and expressed in a highly regulated manner. The ability of each polymer to have a different structure or function depends on the specific sequence of monomers which make up each polymer. Small changes to the sequence of a protein or a DNA strand can have large functional consequences. Thus, it is vital that there is a system for producing specific polymers in a consistent manner. Indeed, Nature achieves this consistency through template-directed polymerization [2]. DNA serves as a direct template for its own replication and for the polymerization of mRNA, which in turn directs the polymerization of proteins by ribosomes. These processes, collectively called the Central Dogma, rely on different molecular machineries, but use the physical characteristics of a template polymer (DNA or mRNA) to facilitate the formation of another, new polymer.

The Central Dogma is a metabolism itself. It is comprised of a system of enzymes that turn monomers into polymers depending on an external stimulus: a transcription factor, for example, might interact with DNA to recruit an RNA polymerase, directing the formation of an RNA polymer from nucleobases. Through these polymerization machineries, the Central Dogma releases the molecular information stored in DNA to create all other metabolisms.

The *de novo* creation of an artificial metabolism that functions in a manner similar to the Central Dogma, one that is able to store and transfer the information encoded in polymers, without the aid of the complex enzymatic machinery already developed in life, would be of significant import. A system able to synthesize specific polymers in response to environmental input would serve as a scaffold for the creation of materials and therapeutics able to respond to changing environments, an extension of the concept behind environmentally-dependent assembly currently seen in certain classes of designed molecules which hold promise for treating disease [18]. Similarly, the development of such systems provides insights and potential pathways for the appearance of biopolymers with specific sequences, a necessary step in the emergence of life. Research into the origins of life continues to better elucidate reasonable pathways toward networks which are able to undergo sequence-specific replication processes [1, 10], doing so as further evidence for the abiotic synthesis and presence of biopolymers mounts [11].

Multiple systems that have begun to recapitulate the process of template-directed polymer synthesis have been devised [12-16]. Broadly, these chemical networks attempt to manipulate non-covalent intramolecular interactions and covalent chemistry between molecules to generate and enrich a specific product by creating assembled templates for chemical ligation of different components. While these attempts have made great progress toward the design of self-replicating networks, many of them achieve such polymerization with modified monomers

that have reactive groups which are markedly different from those currently found in biology, and in some cases undergo polymerization chemistries which generate side products.

By slightly modifying peptide building blocks, our group was able to create a network that undergoes template-directed amplification of a specific polymer without the use of enzymes, leveraging physical characteristics of a molecular network to create monodisperse, extended structures which assemble through defined phase transitions [2]. In order to create a polymerizing system, Chen et al. first selected small peptide scaffolds which are able to form cross-beta architectures (**Figure 1**), to serve as monomers in on our polymerization system. Polymerization was facilitated by modifying the C-terminus of those peptides to an aldehyde and by the addition of an N-terminal side chain nucleophilic residue, resulting in sequences H-NF-CHO (NF) and H-NFF-CHO (NFF). These reactive ends condense quickly and easily with mild acid catalysis, forming a cyclic 6-membered N,N-acetal ring NF and NFF generated an initial library of different oligomers, observable by HPLC, which because of the polymerization, underwent a liquid-liquid phase separation (LLPS) (**Figure 2**). The resulting solute-rich particle phase, observable via transmission electron microscopy (TEM), both alters the degree of polymerization and creates a nucleus for template-directed reactions and amplification of a single library member within a fibrous array (**Figure 3**). The oligomers that form, importantly, are monodisperse — a paracrystalline fiber containing a single oligomer [2].

Intermediate dynamic states, like the visible particle phase, are essential for the assembly of ordered structures in aqueous environments [2]. The ability of this particle state to control the chemistry, leading to the selection of a specific polymer, and transfer information in these phases is profound [2]. While there may be a distribution of different molecules initially in any chemical network, during the particle phase, there is a *selection for the most fit replicating*

nucleus. Individual nuclei induce reactions that template structurally compatible growth, serving as a physical template for the formation of chain length specific polymerization, and possibly sequence specific polymerization [7]. This process represents a new approach to template-directed polymerization not seen in biology and defines a new approach to molecular information storage and amplification distinct from the processes used in the Central Dogma.

Our group became interested in how this new polymerization paradigm, characterized by the particle phase, could be adapted to create new sequence-specific polymers with different architectures. We conceptualized two systems designed to control key assembly steps: the initial access of the intermediate particle phase itself, and the dynamics of template-directed amplification that occur within it. In order to achieve the former goal, we sought to extend the acetal linkage to a 5-membered N,O-acetal, formed by the reaction of a C-terminal aldehyde and an N-terminal threonine. Through this, we hoped to develop the 5-membered linkage into a chemical tagging strategy that might control access to the solute-rich liquid phase. We found that TTF-CHO (TTF) is more polar and therefore does not access the solute-rich particle phase alone. When combined with TTFTTF-NH_2 (TTF_2), which also does not indepenently undergo a phase separation, they polymerize and access the phase transition and form monodisperse cross-beta fibers of the linear dimer formed by their combination, TTF*ox*TTFTTF, within days [9] (**Figure 4**).

In a complementary approach, to probe the dynamics of template-directed amplification, we designed a system we hypothesized would use a different physical template for extension. NFF and NF ostensibly form structures with cross beta sheet structures which extend from initial nuclei with the same architecture [2]. We supposed that this likely led to the limit on the length of the polymers observed, as beta sheets destabilize as their length increases [17]. However,

there are other structures found in biology with less restrictions on length: for example, the triple helix. We hypothesized that by using a peptide polymerization network with a sequence that promoted the formation of triple helices, we would create nuclei in the particle phase able to direct the amplification of much longer NFF-like modified-peptide polymers and would create structures with features similar to collagen.

In order to devise a collagen mimetic, we imagined the 6-membered ligation, formed via the asparagine and aldehyde reaction, serving as an isostere to hydroxyproline, a residue which plays an important role in stabilizing the triple helix of collagen. Our group previously synthesized the sequence NH_2-NPG-CHO (NPG), with both the N-terminal asparagine and the peptide aldehyde, predicting it would form extended structures in the form of NH_2-6-membered ring-PG-...-CHO, with the 6-membered ring taking the place of hydroxyproline, and not limited by beta-sheet length limitations [9]. This strategy mimics the X-Y-Gly motif found in collagen, in which X and Y are proline and hydroxyproline [5]. Indeed, the NPG system accessed the solute-rich liquid particle phase and assembled into bundled fibrils in 20mM NaOAc/AcOH buffer at pH 4.75 (**Figure 5)**. The 6-membered ring formed readily, with the smallest polymer found containing 30 amino acids, indicating 9 linkages had been formed. This result is in stark contrast to the maximum 2 linkages found in the NFF system [9].

We obtained the NPG peptide necessary for these experiments by liquid-phase synthesis without the use of an automated solid-phase peptide synthesizer (SPPS). Liquid-phase peptide synthesis is the synthesis of an amino acid performed *without* a “solid” support, a resin which allows for the isolation of a target peptide chain without having to extract it after each synthetic step. However, this method is significantly more labor intensive and time consuming than automated solid-phase synthesis. When a CEM Liberty Blue Microwave synthesizer became

available, we focused on developing a method for increasing our working amounts of this peptide with SPPS, which would allow us to extend our previous studies of the NPG system (**Outline 1**).

Such extension would involve repetition of the assembly studies performed previously, as well as seeding experiments, which involve templating assembly of NPG using a synthetic collagen as a "seed" to further elucidate NPG's structural similarity to collagen. As mentioned previously, solutes in water assemble via intermediate particles that are solute-rich in liquid-liquid phase aggregates. In this environment, monomers nucleate and propagate larger ordered structures, before emerging out into the solvent environment. We have shown distinct nuclei can template these peptides, forming a structurally complementary co-assembly. When these templates are present, the self-nucleating peptides tend to be quickly templated and grow on them, propagating as mixtures.

Further, we imagined the chemical dynamics of NPG might significantly impact the dynamics of the extracellular matrix (ECM), the chemical network outside of the cell which is maintained by a dynamic physical matrix of exchanging collagen fibers. Thus, we hypothesized changes in matrix dynamics may have a pronounced impact on organoid culture. These cultures have no vascular system to ensure efficient nutrient access. Indeed, in initial experiments, when 100 μM NPG is incubated with day 28 brain organoids, there is a significant reduction in caspase-3 activation (a marker of apoptosis) and an increase in proliferation marker Ki-67, consistent with improvement of the ECM [9]. Further analyses will be required to define the mechanisms underlying this increased viability [9]. Additional work with NPG would benefit from the increased yield available through SPPS.

RESULTS

Functionalization of TTF sequence scaffold. Interested in generating new structure and functionality with this cross-beta peptide scaffold, we imagined the 5-membered ring formed via threonine-aldehyde reaction might serve as an effective isostere of D-proline, an amino acid which has been shown to stabilize beta turns. Beta turns are a key component in beta hairpins structures, which have been shown to form extended functional assemblies [3]. Thus, we devised a system in which two monomers using our previously defined acetal chemistry could be used to form extended structures. Beta hairpin amyloids could be used to create arrays of binding pockets formed by interlaminate beta turns facing the outside of nanotube assemblies (**Figure 6**). Synthesis of the extended hydrophobic sequences necessary to create amyloidogenic beta-hairpins is generally discouraged by the poor solubility of such sequences. However, utilizing a system of *two* initial monomers using acetal linkage chemistry could circumvent this problem. Each monomer might be individually soluble until polymerized. If they became less soluble when polymerized, as in the TTF*ox*TTFTTF system, this would ostensibly drive the formation of the particle phase. Inspired by previously defined synthetic beta turn sequences, we conceptualized the network H-pY-$(TTF)_2$-V-CHO and H-TGT-$(TTF)_2$-NH_2, expecting the system to form the linear monomer H-pY-$(TTF)_2$-V*ox*GT-$(TTF)_2$ (*ox* = oxazolidine, the organic molecular class of the 5-membered linkage). The four-residue turn sequence is inspired by the MAX1 peptide code utilized to create highly stable beta hairpin hydrogels [3]. The N- and C-terminal charges were meant to improve solubility of the structure and potentially direct charge interaction at the leaflet interface leading to nanotube stabilization, as our group had seen during the design of an Aβ fragment congener [4]. We saw that the monomeric, non-polymerized version of this network was indeed insoluble. After synthesis and TFA cocktail cleavage, the

crude beta-turn functionalized peptide sequence H-pY-$(TTF)_2$-VdPGT-$(TTF)_2$-K-NH_2 did not dissolve in a variety of common solvent systems and precipitated in very low concentrations of buffer. This indicated this sequence is a prime target for using templated chemistry to modulate the physical characteristics of the system, allowing for the emergence of uniquely dynamic structure.

Synthesis of NPG. Generation of the peptide aldehyde after SPPS requires additional synthetic steps different from those used in liquid synthesis. Each of these steps represents a process which must be developed and optimized to maximize yield and purity.

The peptide aldehyde is imagined being generated by SPPS on a Weinreb AM resin (**Scheme 1)**. After automated synthesis, this peptide is first cleaved from the solid support by reaction with $LiAlH_4$ in THF at -80℃, generating the protected peptide aldehyde necessary for further chemistry.

Our first strategy focused on deprotecting the N-terminus and the asparagine first while keeping the peptide aldehyde protected from undergoing side reactions using a methanol protection strategy described in the literature [19]. After this, we planned to deprotect the asparagine residue using a reaction involving HCl and HFIP [20], deprotect Fmoc using a base such as triethylamine (TEA), and finally deprotect the aldehyde in assembly conditions.

First, we optimized the removal of NPG off of Weinreb AM resin, forming the peptide aldehyde. Multiple conditions were explored in order to maximize yield and purity, and the most notable changes in crude yield resulted from changing the temperature at time of addition of $LiAlH_4$. We evaluated the success of the reaction by performing thin layer chromatography (TLC) post-synthesis in 50% EA/hexane, as described above. An initial procedure, found in an Aaptec technical manual, directed us to perform the reaction at 0℃ [7]; however, performing the

reaction in a water ice bath resulted in very low concentrations of product. Thus, we moved to using an acetone/dry ice bath, lowering the initial temperature of the reaction to approximately -78℃. This increased the concentration of products as observed by TLC. We also attempted adjusting the base used to quench the reaction from $KHSO_4$ to $KHCO_3$, which appeared tolerated, but did not result in any significant change in yield. We noticed a significant increase in the concentration of the product when we used four molar equivalents of $LiAlH_4$ relative to peptide and allowed the reaction to warm from -78℃ to 0℃, observing visible product accumulating on the round-bottom flask after rotary evaporation.

However, this increased yield was also correlated with a change in the source of resin. While we had synthesized all batches of NPG peptide using a Weinreb AM resin, in initial trials, we had used a batch produced by Aaptec, while for the cleavage described above we used Novabiochem-made supports. Performing the cleavage reaction with Novabiochem resin, 4 equivalents of $LiAlH_4$ at -78℃ resulted in a significant amount of crude product, both visible to the eye after rotary evaporation and through concentrated signals on TLC. From these promising results, we decided to move forward with this reaction condition.

We then attempted to purify our product from the crude mixture. TLC with 50% EA reveals two significant spots in the range 0.2 - 0.5 R_f that are not well separated. NMR spectroscopy suggests that the analyte which produces the higher R_f signal is the desired product and that the other signal is an otherwise uncharacterized impurity (**Figure 7**). We attempted to use gradient elution to improve yield of the pure product. The product tends to elute after the column is washed with 250 mL of 30% ethyl acetate (EA)/hexane and a further 250 mL of 35% EA/hexane — however, only a small amount of the product is meaningfully separated from the impurity.

We attempted this elution strategy multiple times to no avail. In an attempt to improve the reaction, we turned our efforts toward an alternate synthetic strategy (**Scheme 2**). We opted to use a deprotected asparagine residue in the NPG sequence during solid-phase peptide synthesis, as opposed to the trityl-protected amide group we had used before. We hypothesized that such a change might affect the physicochemical properties of the peptides after cleavage, causing them to interact differently with the chromatographic system and yielding better separations. After resin cleavage, we could perform the same steps of aldehyde protection and Fmoc deprotection in order to generate the final product. This process involves one less step than the original synthetic scheme. We were, however, wary of using a strategy that did not involve a protected asparagine residue. Protecting groups are used in SPPS to prevent side reactions with coupling reagents that are used during the process; additionally, the unprotected, nucleophilic amide may interact with the electrophilic aldehyde and cause the monomers to partially polymerize; however, this is somewhat unlikely as the amide nitrogen is a weak nucleophile. It is unlikely that it would interact with the aldehyde at an acceptable rate to drive dominance of a non-acetal polymeric product. However, considering that we could generate a new peptide sequence via SPPS rapidly and further test whether or not we generated the target product after cleavage, we were interested in attempting this new strategy.

We synthesized the N-unprotected-PG sequence on Weinreb resin from Novabiochem and cleaved the peptide using 2 equivalents of $LiAlH_4$ at -78°C. We performed two syntheses using these parameters: one in which the cleavage reaction vessel was allowed to cool to -30°C before the reaction was quenched, and one in which the reaction was quenched at approximately 8°C. For the reaction that was allowed to cool, we observed a very complex TLC signal in a variety of mobile phases, including 50% EA/hexane and combinations of methanol/EA/hexane.

Different mobile phase concentrations revealed different discrete spots, potentially indicating future difficulties in purifying a single product. For the reaction that was not allowed to cool, we observed a discrete spot via TLC using 50% EA/hexane: however, higher quenching temperatures likely destabilize the important tetrahedral intermediate in Weinreb-resin mediated aldehyde formation [8], leading to over-reduction and the loss of our product. However, without separation of the crude mixture via column chromatography followed by NMR spectroscopy or mass spectrometry for validation, we cannot yet test this hypothesis.

DISCUSSION

We have thus far detailed two alternate synthetic routes to generate NPG. The products of the cleavage reaction in Scheme 1 have proven difficult to separate under an initial set of chromatographic conditions. Data from Scheme 2 is as of yet inconclusive and requires further structural characterization before its feasibility can be definitively stated. However, this work details an outline toward a complete synthetic process which is able to generate significant quantities of NPG, and perhaps other short peptide aldehydes, in a reasonable yield, purity, and time frame. Our work characterizing the removal of the SPPS-synthesized peptide from Weinreb resin has provided better knowledge of the parameters associated with this reaction and will prove useful for generating peptide aldehydes for studies of systems currently envisioned by our group, including TTF and NPG, as well as other potential studies involving peptide aldehydes, of broad utility due to their reactive properties, including in the design of protease inhibitors [21-22].

An extension of our inquiry into the NPG system [9] will yield important information about the dynamics of the phase-progressive assembly observed by our group. The seeding

experiments described in the Introduction will serve to strengthen our model of triple helical templated assembly. As well, further investigation into the mechanisms of organoid systems will be greatly facilitated by higher working amounts of NPG, allowing for more experiments to be performed in higher throughput.

Our research into the NPG and TTF systems define complementary and distinct strategies for controlling the transformation of molecular information via polymerization. The TTF scaffold represents a system by which assembly is dependent upon the formation of a polymer linkage, acting as a "switch" for the access of the particle phase [9]. The regulation of molecular systems is a fundamental aspect of biology: with TTF, the propagation of extended and specific fibrillar structures is controlled by modulating the specific physical characteristics of the monomer system. Non-membrane bound organelles, such as ribonucleoprotein granules, are found extensively in biology. Their structure and function depend on the tight regulation of phase networks perpetuated by intramolecular interactions. This TTF tagging strategy may represent a means by which extant protein networks can be modified, a suggestion further strengthened by the initially seen efficacy of NPG in interacting with living systems [9]. Understanding and validating the processes by which these systems form structural networks then serves a dual purpose: it will provide strategies for the *de novo* development of more varied and functional self-assembling structures that are, most importantly, dynamic and responsive to their environment, and also provide insight into the complexity of existing biological phenomena. Life is a dynamic system, and as scientific knowledge of the complexity of cell physiology grows, the next generation of therapeutics will be more effective if they are similarly dynamic: a logic that underlies current cancer treatments such as CAR T-cell therapy [23].

Specific molecular, metabolically-inspired systems, such as the ones presented in this work, may serve as scaffolds for further functionalization.

FUTURE AIMS

The completion of the synthesis of NPG is of primary importance. To solve the challenges associated with implementing Scheme 1, a different mobile phase may be able to effectively separate the desired product from impurities. While EA/hexane is a commonly used solvent system, other combinations involving mixtures of dichloromethane and methanol can be used. Chromatography could also potentially be delayed until after the second reaction, the protection of the aldehyde. After the aldehyde is protected, the desired product may have different physicochemical properties which causes it to interact with the chromatographic system differently, leading to more effective separation.

The products of the cleavage reaction in Scheme 2, utilizing the deprotected-asparagine NPG-resin peptide, also will need to be analyzed. TLC yielded a clear signal for the initial cleavage reaction without cooling pre-quench, with one major spot with Rf ~0.3. We will use column chromatography to separate this product and characterize it using NMR spectroscopy and MALDI mass spectrometry. If this spot is revealed to be the desired product, these conditions could be used in the synthetic process going forward.

Both Scheme 1 and Scheme 2 will also require further characterization of the synthetic pathway. For both, further optimization of the trityl and Fmoc deprotection reactions will likely be necessary. Once we have synthesized NPG, we will be able to extend our investigation of NPG assembly, seeding, and organoid dynamics as described within the introduction.

In order to interrogate seeding in particular, we plan to create a solution of a pre-assembled 2.5 mM collagen-mimetic with a 10 mM concentration of free, unassembled NPG. A positive result, faster assembly than seen without the seed, would indicate that our NPG structure is able to be templated by the pre-formed collagen mimetic. This would indicate that NPG shares a structural complementarity with a previously defined collagen-like protein, as this templating event would require NPG polymer to achieve a conformation necessary to propagate assembly from the collagen mimetic.

METHODS

General methods. All chemical reagents and solvents are purchased from Millipore-Sigma and Novabiochem (St. Louis, MO), Alfa-Aesar (Haverhill, MA) and Aaptec (Louisville, KY) without further purification. Figures not adapted with permission were created using ChemDraw Professional 16.0.

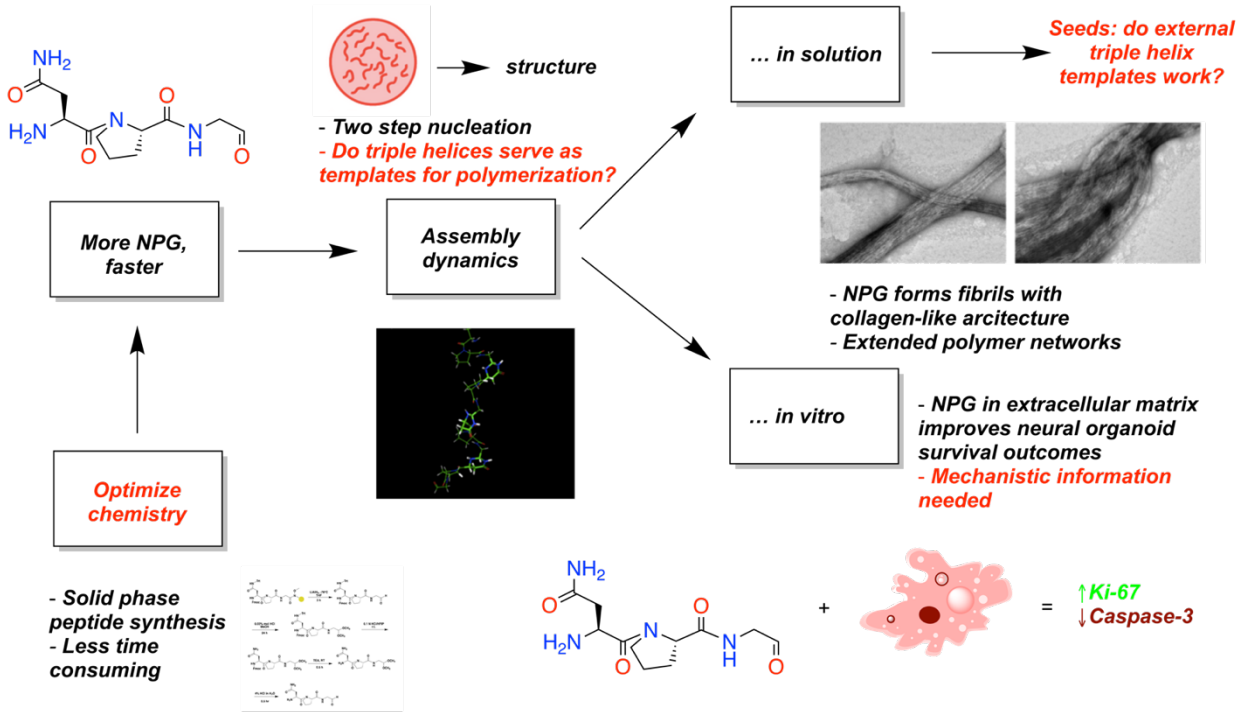
Nuclear magnetic resonance spectroscopy (NMR). All NMR experiments are performed on INOVA 400 MHz or INOVA 600 MHz NMR spectrometers. Analytes are dissolved in deuterated chloroform ($CDCl_3$) and placed into the spectrometer. Spectrometers are operated using VnmrJ software. Data is then analyzed using the MNOVA software suite.

Automated solid-phase peptide synthesis. All peptides are synthesized using Fmoc solid-phase synthesis protocols on a CEM Liberty Microwave Automated Peptide Synthesizer on Weinreb amide resin. All Fmoc amino acids were dissolved in a volume of dimethylformamide (DMF) calculated by the Liberty Blue software to create solutions of proper concentration. Peptides were then synthesized using default coupling methods on the Liberty Blue software, associated with each amino acid being added to the growing chain. Fmoc deprotection was

facilitated by a solution of 20% piperidine in DMF, while amino acid coupling occurred via automated addition of the peptide along with coupling reagents diisopropylcarbodiimide (DIC) and ethyl cyanohydroxyiminoacetate (Oxyma).

Cleavage from solid-phase peptide support. The resin was vacuum filtered and washed with DCM. Anhydrous THF was added to the resin under an inert N_2 atmosphere and was cooled to -80℃ in an acetone/dry-ice bath. One molar equivalent of 1.0 M $LiAlH_4$ was added and the reaction was stirred until the temperature of the ice bath reached 0℃. The reaction mixture was re-chilled to approximately -30℃ before being quenched with $KHSO_4$, allowed to warm to room temperature, and the resin was vacuum filtered and washed with ethyl acetate. The filtrate was then washed with saturated $KHSO_4$, $KHCO_3$ and NaCl solutions before drying the solution with Na_2SO_4 and subsequent further vacuum filtration. The final filtrate was then concentrated using rotary evaporation and analyzed via thin-layer chromatography (TLC) using a 50% ethyl acetate (EA)/hexane mobile phase. The crude product was purified via silica gel column chromatography, prepared with SiliaFlash and a gradient eluted by increasing the proportion of ethyl acetate in hexane from 30% to 40% in steps of 5%. Fractions were checked by TLC in 50% EA/hexane, and then combined, dried using rotary evaporation, desiccation overnight, and analyzed via NMR in deuterated chloroform.

OUTLINES, FIGURES & SCHEMES



Outline 1. Schematic visually detailing the logic and work described in this project, previous work, and future work. Some components of this figure are adapted from [2] and [9] with permission.

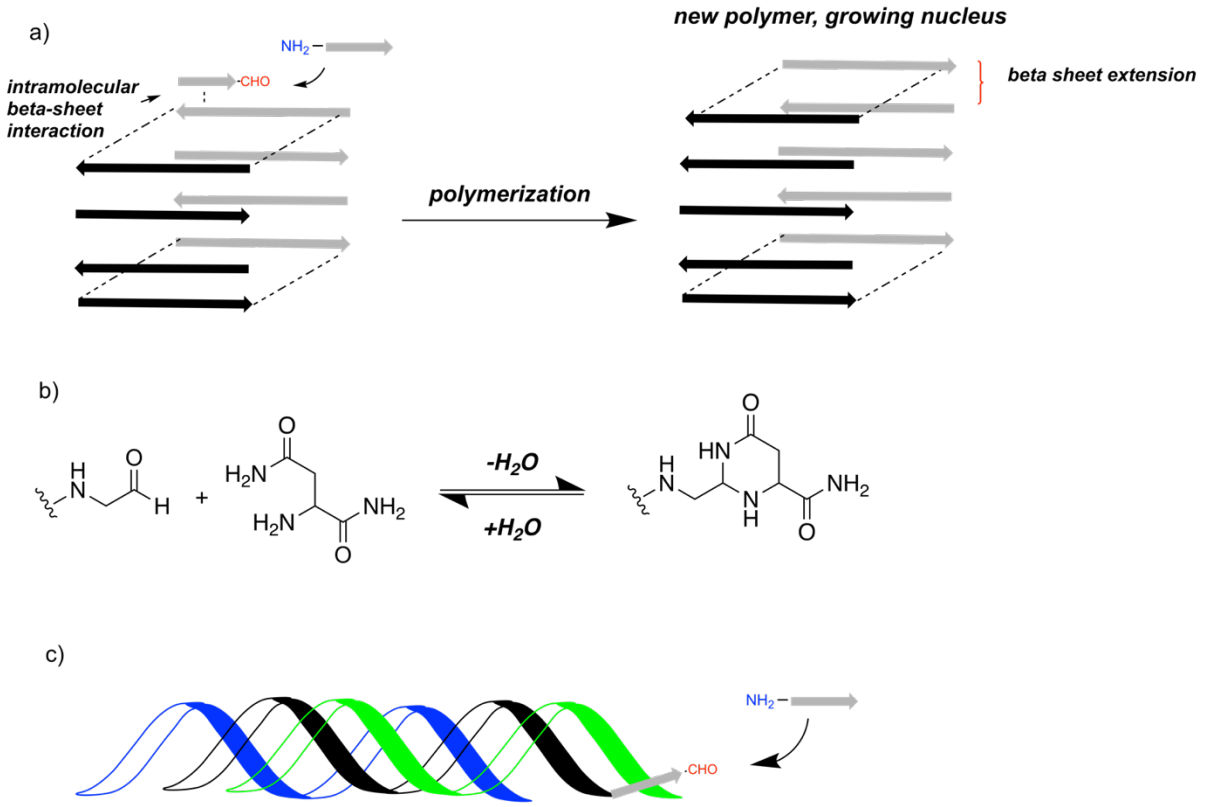


Figure 1. a) A model for the amplification of chain-length specific polymers via template-directed polymerization in a cross-beta architecture. A stable cross-beta structure made up of a specific polymer length serves as a nucleus that undergoes intramolecular beta-sheet characteristic interactions between different monomers and oligomers, stabilizing the formation of the polymer with the same length as those found in the cross-beta nucleus. That polymer lengthens the cross-beta structure. b) Abbreviated reaction scheme depicting the polymerization strategy used in the NFF and NPG systems. c) Simplified depiction of template-directed polymerization in a triple helical structure. This is analogous to template-directed synthesis in cross-beta sheets: the stabilizing forces are those proper to triple helices.

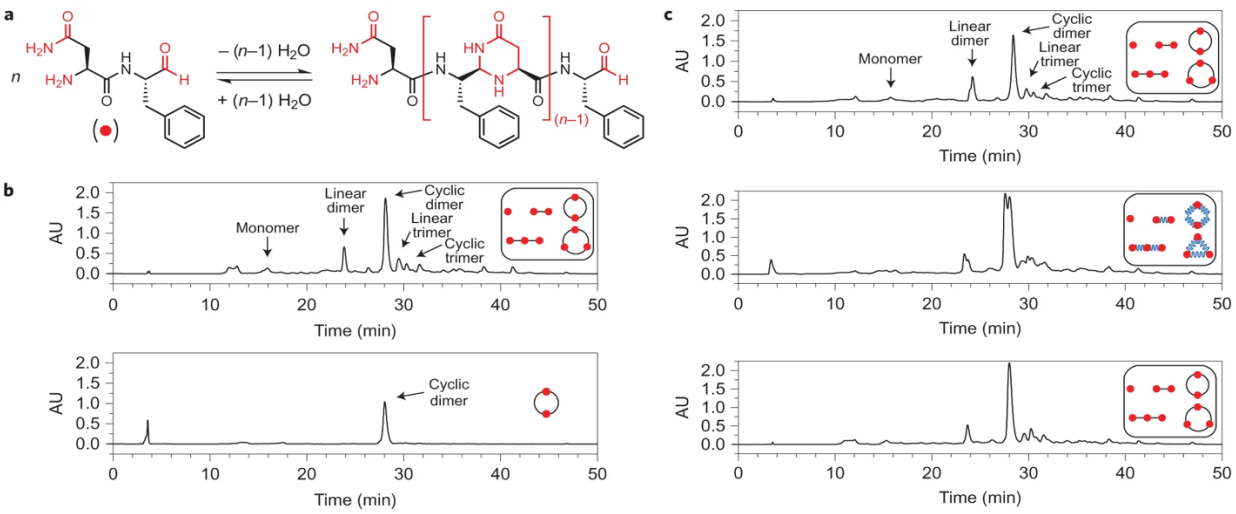


Figure 2. a, Condensation scheme of H-Asn-Phe-CHO (NF-CHO) peptides to give oligomers with N,N-acetal (4-pyrimidinone) linkages. b, HPLC of the NF-CHO (8 mM) dynamic network monitored at 222 nm: top, after 2 days in water/acetonitrile (3/2, v/v) at pH 4 and room temperature; bottom, the cyclic dimer collected under quasi-neutral pH is stable. c, Top: the new network generated from the isolated cyclic dimer incubated at pH 4; middle: heating the network to 40 °C results in isomerization of the acetal center (indicated by the squiggly blue lines); bottom: cooling back to room temperature restores the network distribution. The insets in b,c indicate the species detected by HPLC, where red dots represent the NF-CHO building block, and lines joining monomers represent the reversible acetal bond (py) resulting in formation of linear and cyclic oligomers. Adapted with permission from [2]. Copyright 2017 Nature Publishing Group.

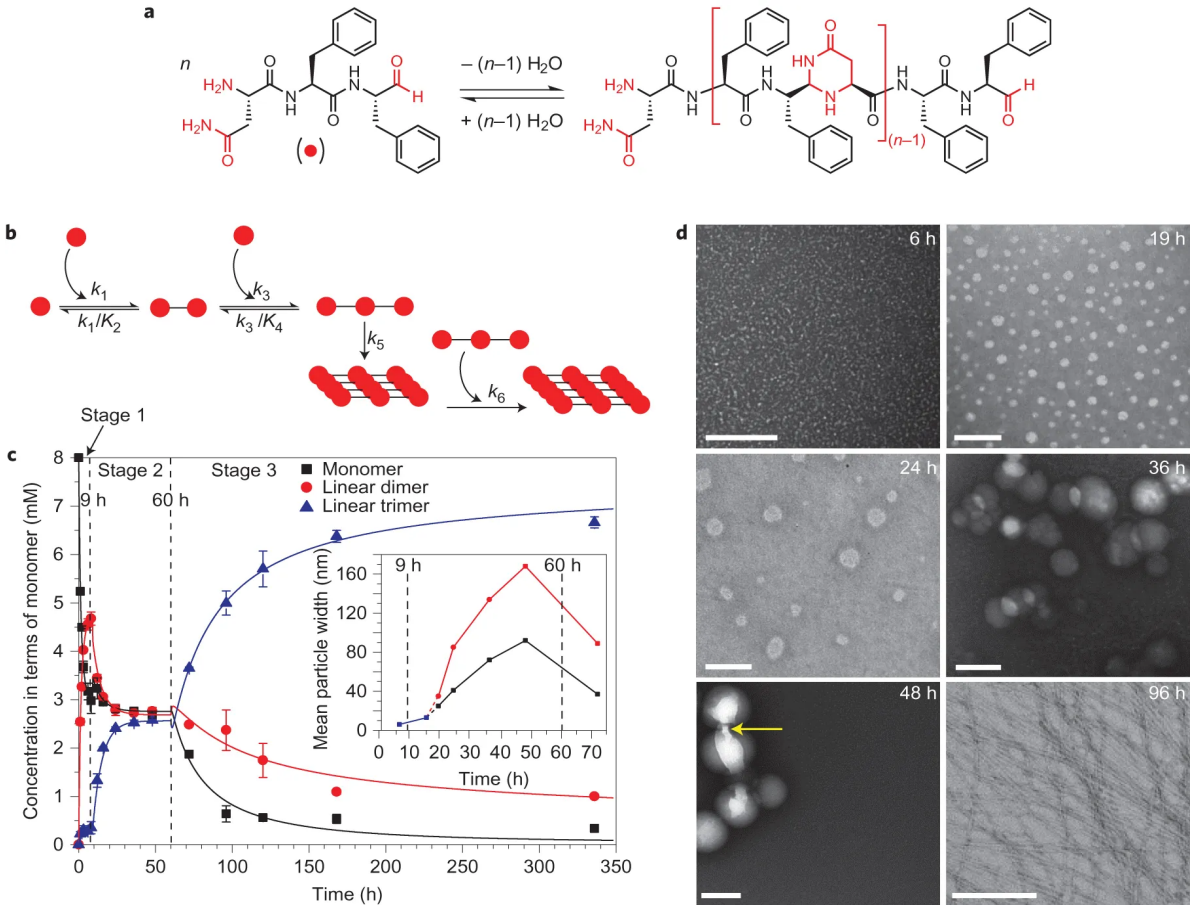


Figure 3. Condensation scheme of H-Asn-Phe-Phe-CHO (NFF-CHO). b, Kinetic model for the NFF-CHO network, where red dots represent the NFF-CHO building block, and lines joining monomers represent the reversible acetal bond (py) resulting in formation of linear oligomers. c, Kinetic fits (solid lines) of network member concentrations (symbols) for monomer (black), linear dimer (red) and linear trimer (blue). Inset shows the mean widths as visualized by electron microscopy of the two particle populations (black and red) over time. Error bars are the standard deviation of three trials and if not visible are the size of the data point. d, TEM micrographs of samples at the indicated time points (scale bars, 200 nm). Yellow arrow in the 48 h micrograph highlights the fibers emerging from the particles as seen in amyloid assembly. Adapted with permission from [2]. Copyright 2017 Nature Publishing Group.

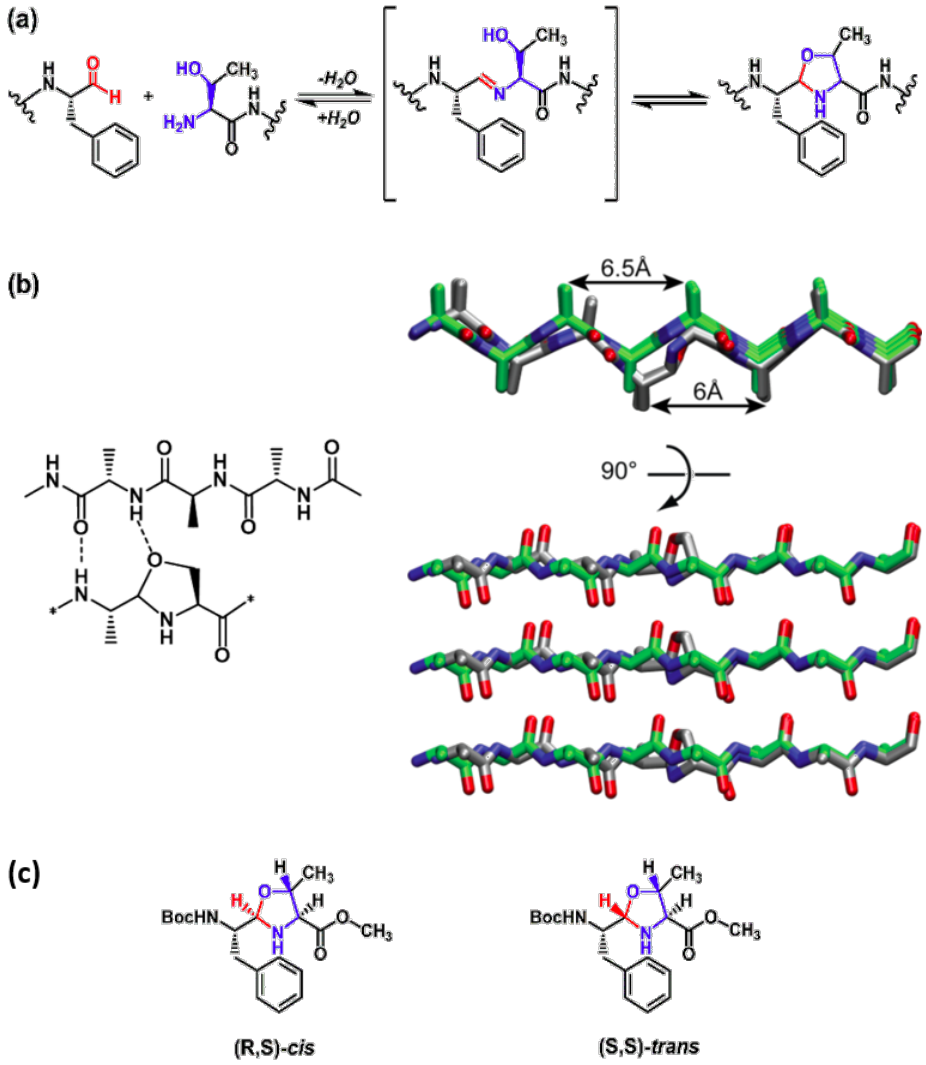


Figure 4. a) N,O-acetal formation (5-membered ring) via N-terminal threonine and C-terminal aldehyde reaction. b) Modelling of N,O-acetal ring formation suggests that the linkage is accommodated in beta-turn architectures. c) A model system used to deduce the stereochemistry of the formed N,O-acetal. Previous experiments allowed us to define the (R,S)-*cis* isomer as the energetically preferred configuration. Adapted with permission from [9].

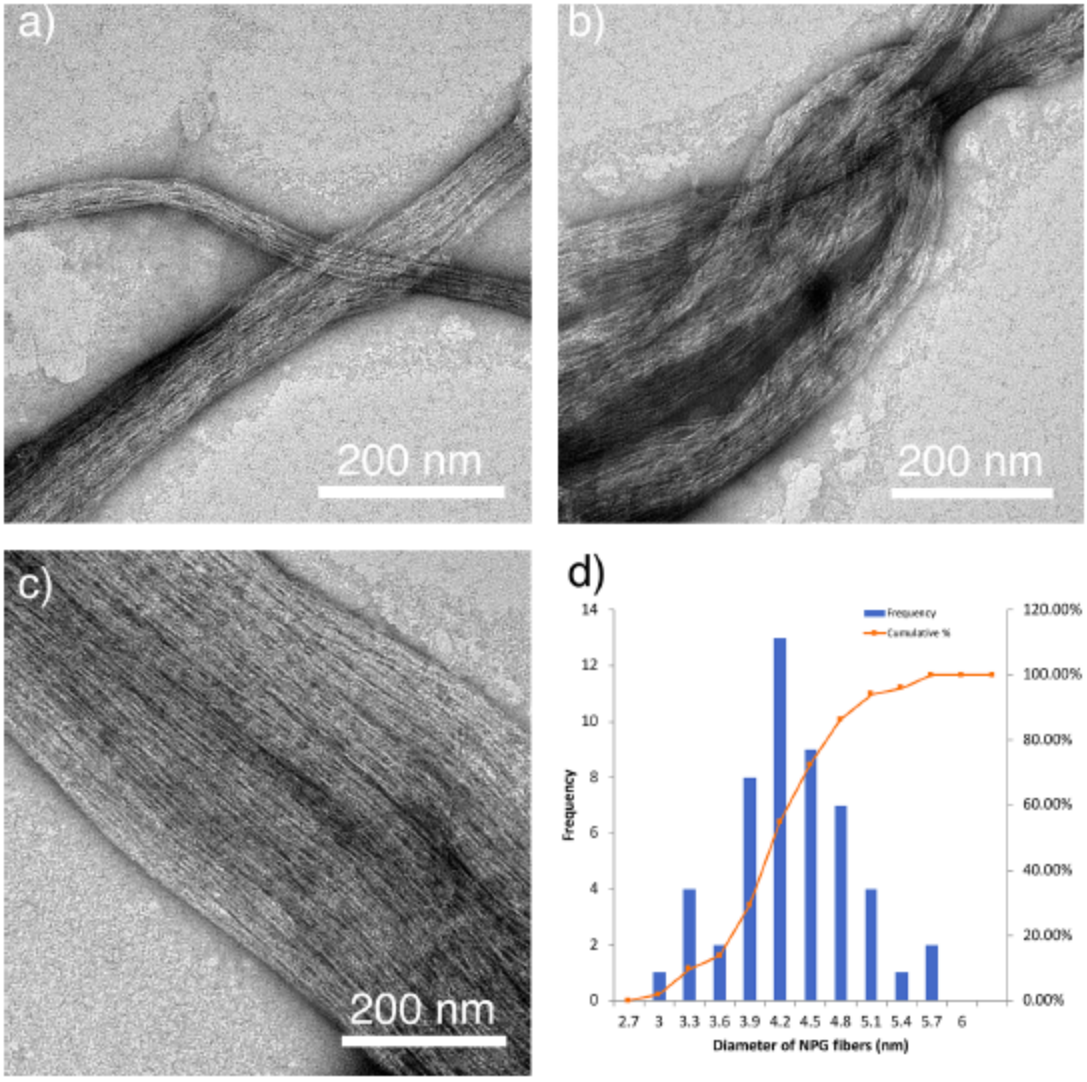


Figure 5. Representative TEM images (a-c) of bundled NPG at 10 mM concentration in 20 mM NaOAc/AcOH buffer, assembled in fibers illustrating the repeating diameters of individual thin fibrils of these bundles. d) Individual fibrils have average widths of 4.2 +/- 0.6 nm. Adapted with permission from [9].

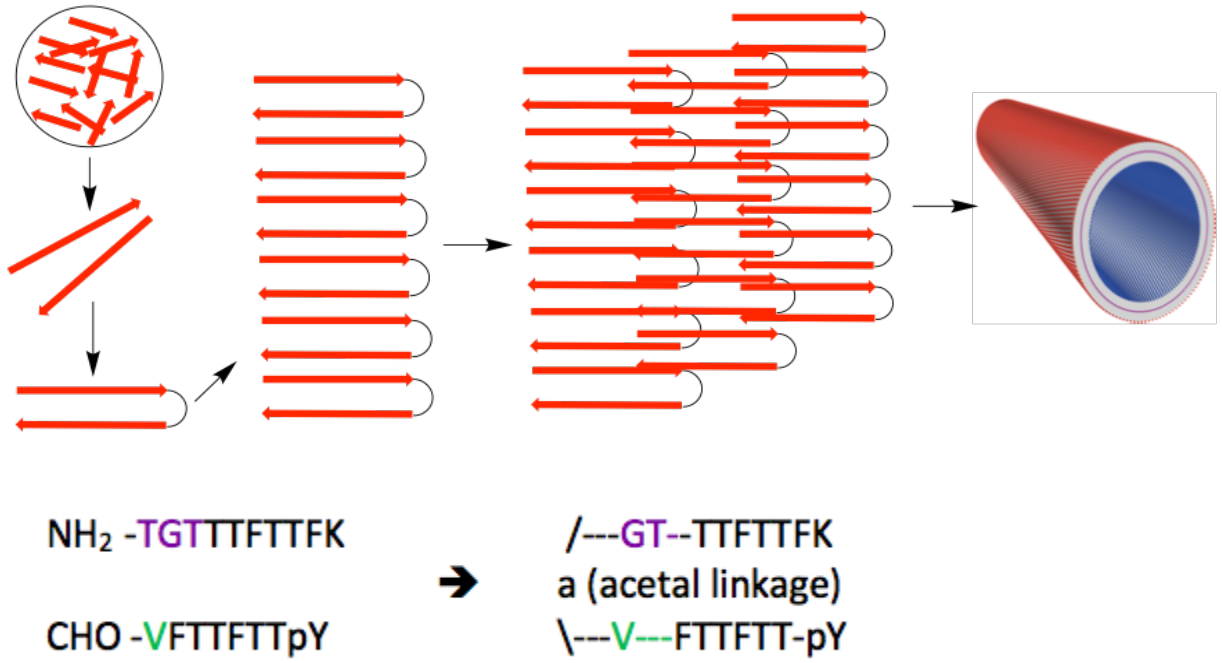


Figure 6. Schematic defining a hypothesized beta-hairpin assembly. Monomers would undergo LLPS, triggered by formation of an N,O-acetal. This acetal, when placed in the sequence context of a beta-turn in the place of D-proline due to its structural rigidity, would stabilize the formation of beta-hairpin monomers. These polymers could then assemble into extended beta-sheet structures, then engage in supramolecular interlaminate reactions with other beta-sheets to form extended nanotube structures.

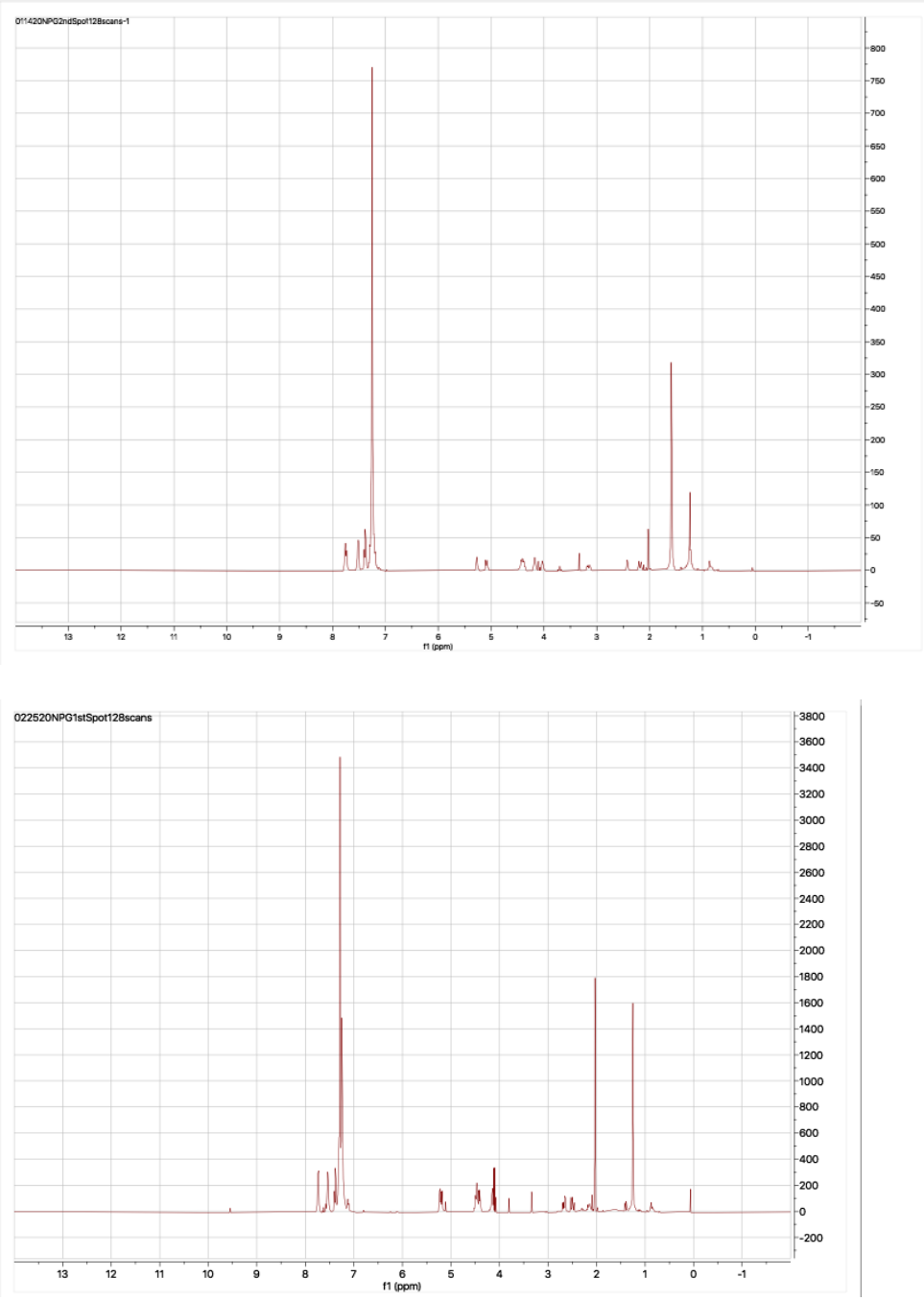
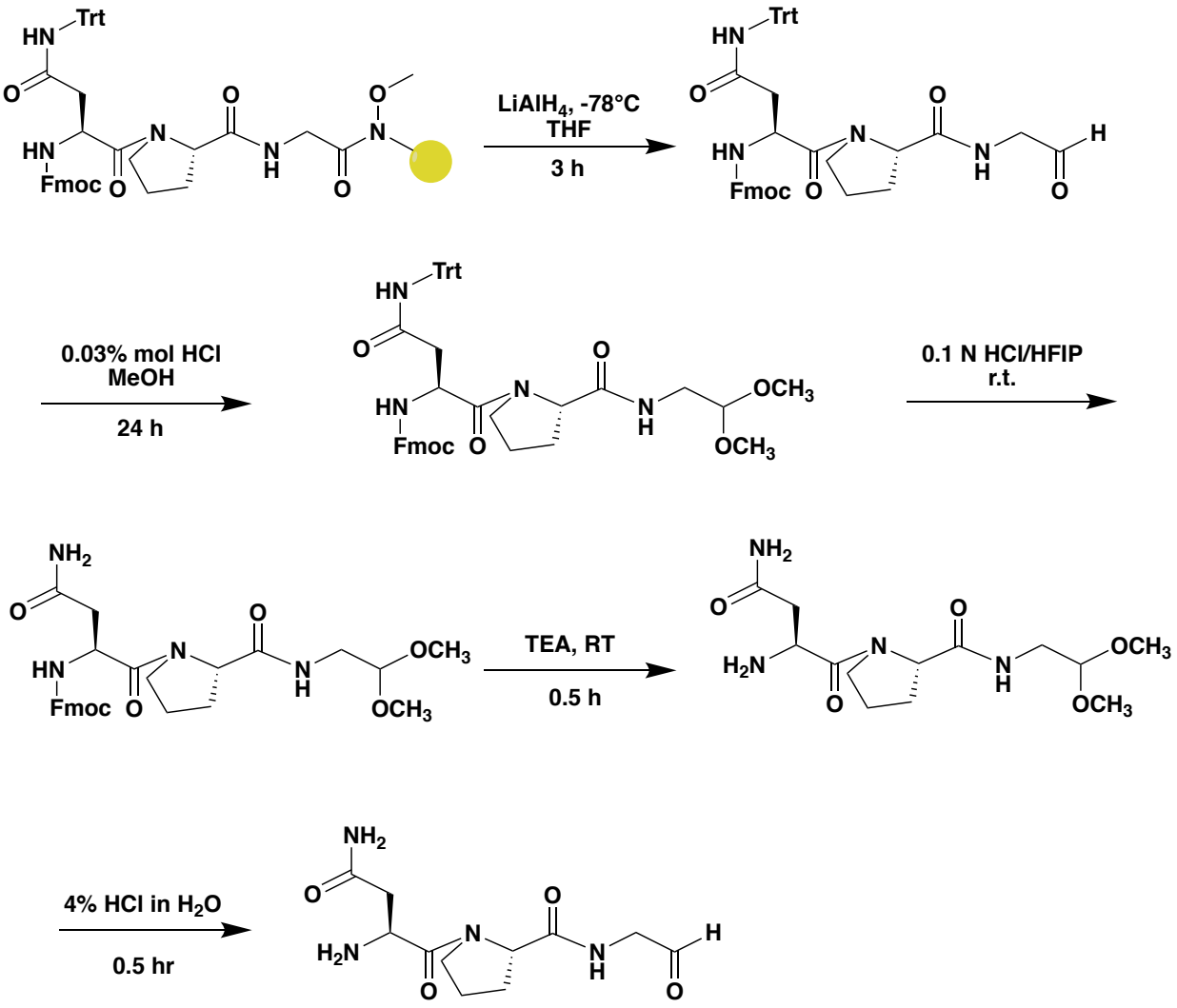
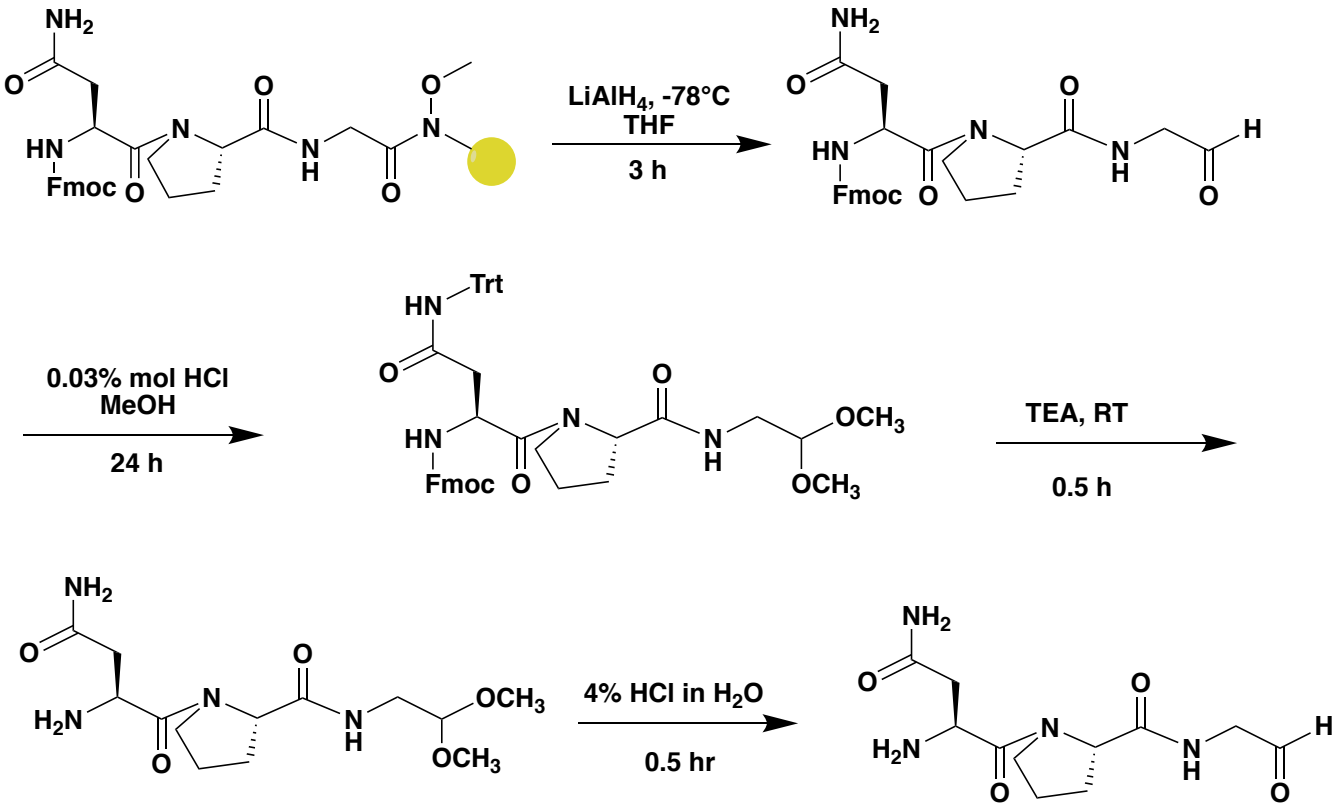


Figure 7. NMR spectra of suspected main impurity (top) and product (below). The spectra are primarily differentiated by a very low relative integration peak visible at 9.5 ppm on the bottom graph, the usual chemical shift of an aldehyde. Suspected sharp solvent peaks between 1 and 2 ppm differ between samples.



Scheme 1. Initial proposed scheme for the synthesis of H-NPG-CHO. H-N(trt)PG is generated via solid-phase peptide synthesis, attached to a Weinreb resin. It is cleaved via reaction with $LiAlH_4$, revealing the aldehyde. This aldehyde is then protected using methanol and low catalytic amounts of HCl. The trityl group is then removed from the asparagine residue using HCl in the solvent HFIP. Removal of the N-terminal Fmoc is facilitated by TEA. The final product can be generated by adding 4% HCl in H_2O.



Scheme 2. Alternate proposed scheme for the synthesis of H-NPG-CHO. H-N(unprotected)PG is generated via SPPS, attached to a Weinreb resin. It is cleaved via reaction with $LiAlH_4$, revealing the aldehyde. This aldehyde is then protected using methanol and low catalytic amounts of HCl. Removal of the N-terminal Fmoc is facilitated by TEA. The final product can be generated by adding 4% HCl in H_2O.

REFERENCES

1. Goodwin J.T.; and Lynn D.G. **2014** Alternative Chemistries of Life – Empirical Approaches. Emory University, ISBN: 978-0-692-24992-5. http://alternativechemistries.emory.edu/
2. Chen, C.; Tan, J; Hsieh, M-C; Pan, T.; Goodwin, J.T.; Mehta, A.K.; Grover, M.A.; and Lynn, D.G. **2017** Design of Multi-Phase Dynamic Chemical Networks, *Nature Chem*, DOI: 10.1038/NCHEM.2737.
3. Rughani R.V.; and Schneider J.P. **2008** Molecular Design of β-Hairpin Peptides for Material Construction. *MRS Bull*, 33(5): 530–535.
4. Li S.; Mehta A.K.; Sidorov A.N; Orlando T.M.; Jang Z.; Anthony N.R.; and Lynn D.G. **2016** Design of Asymmetric Peptide Bilayer Membranes. Journal of the American Chemical Society, 138 (10). 3579-86. doi:10.1021/jacs.6b00977. pdf
5. Brodsky B.; and Shah N.K. **1995** Protein motifs. 8. The triple-helix motif in proteins. *FASEB J.* DOI: 10.1096/fasebj.9.15.8529832.
6. Truong W.T.; Su Y.; Gloria D.; Braet F.; and Thordarson P. **2015** Dissolution and degradation of Fmoc-diphenylalanine self-assembled gels results in necrosis at high concentrations in vitro. *Biomaterials Science*. DOI: 10.1039/C4BM00244J.
7. Aaptec. Practical Guide to Solid Phase Peptide Synthesis. https://www.peptide.com/resources/solid-phase-peptide-synthesis/
8. Nahm S.; and Weinreb S.M. **1981** N-methoxy-n-methylamides as effective acylating agents. *Tetrahedron Letters*. 22(39), 3815-3818.

9. Tan J.; Zhang L.; Hsieh M.; Bai Y.; Young S.; Kim Y; Goodwin J.T.;, Grover M.A.; and Lynn, D.G. Chemical Control of Intrinsically Disordered Phase Transitions. *In submission*.

10. Higgs P.G.; and Lehman N. **2015** The RNA world: molecular cooperation at the origins of life. *Nat Rev Gen*, DOI: 10.1038/nrg3841.

11. McGeoch M.W.; Dikler S.; and McGeoch J.E.M. Hemolithin: A Meteoric Protein containing Iron and Lithium. *arXiv,* https://arxiv.org/abs/2002.11688

12. John M. Beierle, Yasuyuki Ura, M. Reza Ghadiri, and Luke J. Leman . Templated Self-Assembly of Dynamic Peptide Nucleic Acids. Biochemistry **2018,** *57* (1) , 160-172. DOI: 10.1021/acs.biochem.7b00656

13. Rubinov B.; Wagner N.; Matmor M.; Regev O.; Ashkenasy N.; and Ashkenasy G. **2012** Transient fibril structures facilitating nonenzymatic self-replication. *ACS Nano*, DOI: 10.1021/nn302223v.

14. Rubinov B.; Wagner N.; Rapaport H.; and Ashenkasy G. **2009** Self-Replicating Amphiphilic B-sheet peptides. *Angew. Chem., Int. Ed.* DOI: 10.1002/anie.200902790

15. Li X.; Zhan Z.Y.; Knipe R.; and Lynn D.G. **2002** DNA-catalyzed polymerization. *J Am Chem Soc*, DOI: 10.1021/ja017319j

16. Nowak P.; Colomb-Delsuc M.; Otto S.; and Li J. **2015** Template-Triggered Emergence of a Self-Replicator from a Dynamic Combinatorial Library. *J Am Chem Soc,* DOI: 10.1021/jacs.5b04380

17. Stanger H.E.; Syud F.A.; Espinosa J.F.; Giriat I.; and Muir T.; and Gellman S.H. **2001** Length-dependent stability and strand length limits in antiparallel β-sheet secondary structure. *PNAS,* DOI: 10.1073/pnas.211536998.

18. Lee S.; Trinh T.H.T; Yoo M.; Shin J.; Lee H.; Kim J.; Hwang E.; and Lim Y.; and Ryou C. **2019** Self-assembling peptides and their application in the treatment of diseases. *Int J Mol Sci,* DOI: 10.3390/ijms20235850.

19. Dong J.L; Yu L.; and Xie J. **2018** A Simple and Versatile Method for the Formation of Acetals/Ketals Using Trace Conventional Acids. *ACS Omega,* 3, 4974-4985.

20. Palladino P.; and Stetsenko D.A. **2012** New TFA-Free Cleavage and Final Deprotection in Fmoc Solid-Phase Peptide Synthesis: Dilute HCl in Fluoro Alcohol. *Organic Letters*, 14(24): 6346-6349.

21. Zhu L.; George S.; Schmidt M.F.; Al-Gharabli S.I.; Rademann J.; and Hilgenfeld R. **2011** Peptide aldehyde inhibitors challenge the substrate specificity of the SARS-coronavirus main protease. *Antiviral Res*, DOI: 10.1016/j.antiviral.2011.08.001

22. Melnyk O.; Fehrentz J.A.; Martinez J.; and Gras-Masse H. **2000** Functionalization of peptides and proteins by aldehyde or keto groups. *Biopolymers,* 55(2):165-86.

23. June C.H.; O'Connor R.S.; Kawalekar O.U.; Ghassemi S.; Milone M.C. **2018** CAR T cell immunotherapy for human cancer. *Science*, DOI: 10.1126/science.aar6711