Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced 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 or dissertation in whole or in part in all forms of media, now or hereafter known, 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 or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Ting Pan Date

EXTENDING THE CENTRAL DOGMA: DEVELOPING NOVEL INFORMATIONAL BIOPOLYMERS

By

Ting Pan
Master of Science

Chemistry

_____

Dr. David G. Lynn
Advisor

_____

Dr. Frank McDonald
Committee Member

_____

Dr. Emily Weinert
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____ Date

EXTENDING THE CENTRAL DOGMA: DEVELOPING NOVEL INFORMATIONAL BIOPOLYMERS

By

Ting Pan

B.S., Nanjing University, 2011

Advisor: David G. Lynn, Ph.D.

An abstract of

A thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science

in Chemistry

2017

Abstract

EXTENDING THE CENTRAL DOGMA: DEVELOPING NOVEL INFORMATIONAL BIOPOLYMERS

By Ting Pan

In biology, the highly accurate replication of sequence information encoded on informational biopolymers is accomplished by template-directed polymerization with sophisticated polymerases. Decades of efforts have been put into creating simple chemical systems to mimic such processes. By replacing phosphodiester with reversible imines, we created a dynamic nucleoside chemical network (DCN) as a very simple model that were managed by sophisticated biological catalyst. We have successfully achieved template-directed polymerization to synthesize amine nucleoside polymers (ANPs) with high fidelity. The diversity of ANP-based DCNs provides for potential emergent functions. Leveraging the catalytic activity of the amine backbone, we developed dANP-based ANPzymes as novel gene regulation platforms. We further constructed DCNs with full potential to respond to different environmental inputs and templates. Finally, taking advantage of dynamic chemical networks and templating effects, we set the stage for re-directing information transfer behavior in chemical systems through computational modeling of those processes.

EXTENDING THE CENTRAL DOGMA: DEVELOPING NOVEL INFORMATIONAL BIOPOLYMERS

By

Ting Pan

B.S., Nanjing University, 2011

Advisor: David G. Lynn, Ph.D.

A thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science

in Chemistry

2017

Table of Contents

List of Figures

List of Abbreviations

AFM	Atomic Force Microscopy
ANP	Amine Nucleoside Polymer
BODIPY	Boron-dipyrromethene
CD	Circular Dichroism
DCN	Dynamic Chemical Network
DCM	Dichloromethane
DIPEA	N,N-diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethyl Sulfoxide
DT	Decision Tree
EDT	1,2-Ethanedithiol
EM	Expected Maximization
FMOC	9-Fluorenylmethoxycarbonyl
FT-IR	Fourier Transform Infrared Spectroscopy
HPLC	High Performance Liquid Chromatography
KNN	K-Nearest Neighbors
ICA	Independent Component Analysis
LNA	Locked Nucleic Acid
MD	Molecular Dynamics
NMR	Nuclear Magnetic Resonance
NN	Neuro-Networks
NOE	Nuclear Overhauser Effect
SVM	Supported Vector Machine
TEM	Transmission Electron Microscopy
TFA	Trifluoroacetic Acid
TLC	Thin Layer Chromatography
RAE	ReliefAttributeEvl

PCA	Principle Component Analysis
PEI	Polyehtylene Imine
PNA	Peptide Nucleic Acid
RP	Random Projection

Chapter 1: Amine Nucleoside Polymer Enzyme (ANPzyme) as a Novel Gene Therapy Platform

1.1 Introduction

1.1.1 Gene Therapy for Neuro-degenerative Diseases

Gene therapy is now widely studied in the laboratory as a promising approach to control gene expression. Compared with traditional drugs which target proteins including enzymes and receptors, gene therapy target mRNA/Pre-mRNA to block the translational process. The first case of gene therapy was reported in 1978 when Zamecnik and Stephenson [1] first developed a DNA tridecamer to target a complementary mRNA sequence of Rous sarcoma virus 35S RNA to inhibit virus replication and cell transformation. Research on gene therapy, including design of oligonucleotides for gene regulation, viral/non-vial delivery vectors for gene therapeutics and imaging techniques for down regulation, have increased exponentially during the past two decades, with hundreds of clinical trials under different phases. Yet only one drug, Fomivirsen, which is used to treat cytomegalovirus retinitis(CMV) in immunocompromised patients, was approved by FDA twenty years after first discover [2].

Currently, well-established gene therapeutics include synthetic oligonucleotides such as antisense oligonucleotides, small RNA interference, and antagomirs, aptamers,

DNA/RNAzymes, exon-skipping and splice-switching compounds, and immunostimulatory nucleic acid [3] They share common upstream mechanism by blocking mRNA translation into proteins, but the downstream regulation strategies vary by reagents, chemical modifications, targeting approaches, and regulation of target mRNA.

1.1.2 Challenges for Current Gene Therapeutics

Promising as they are, current development of gene regulation methodologies mainly has been hampered by several obstacles. First of all, DNA/RNA therapeutic reagents as well as chemically modified oligonucleotides [4], are hardly be taken up by the cell. Both viral and non-viral vectors have been developed to improve cellular uptake. Secondly, chemical and biological stability limits the use of oligonucleotides for therapy. Chemical modifications can improve stability of the backbone to the acidic environment of cell. Yet this approach can limit target recognition, such as RISC and RNase H recognition, which are critical for targeting mRNA. Third, off target effect has become a major concern recently along with immune response, toxicity, and some other complicated side effects.

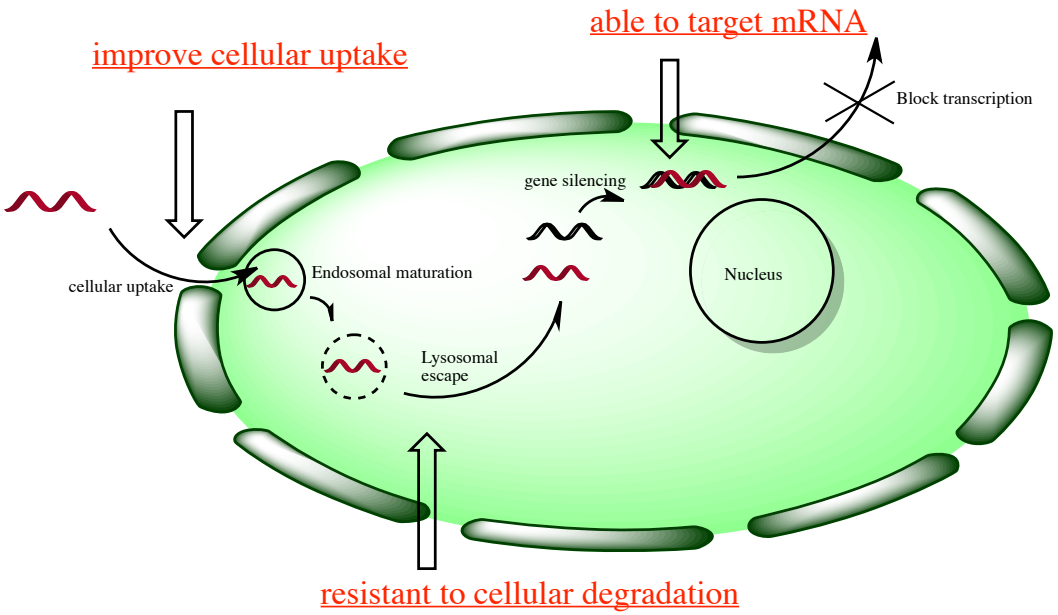


Figure 1 Desired improvements of current gene therapeutics.

1.1.3 Development of Amine Nucleoside Polymers(ANPs)

Amine nucleoside polymers(ANPs) are originally developed in our group [5] in the 1990s to explore reversible polymerization of synthetic monomers as mimic DNA replication with wide potential applications. We attempt to mimic the phosphodiester formation along complementary nucleic acid by designing an initial coupling reaction followed by subsequent trapping of the thermodynamic product. The thymidine imine functionalized with amines and aldehydes are then trapped by hydride to form amine nucleoside polymer (**Figure 2**). Standard protocol has been established in our group to obtain sequence specific ANP oligomer with DNA template directed reactions [6].

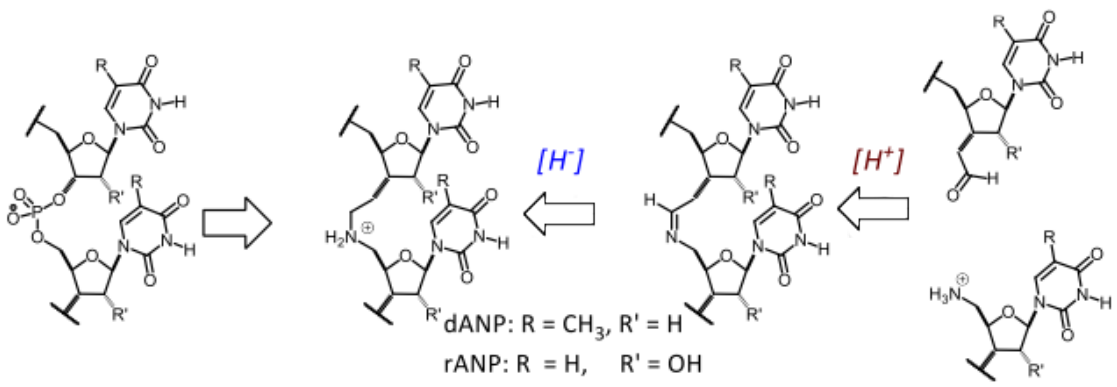


Figure 2 Scheme of amine nucleoside polymers(ANPs).

In contrast to the native phosphodiester backbone, the unnatural amine backbone of ANP carry several unique advantages including higher binding affinity for native nucleotides due to electronic interaction, higher chemical stability to cellular degradation, and potentially higher cellular uptake due to positively charged backbone, thus making the ANPs a promising platform for gene regulation.

1.1.4 Utilizing the ANPs as Novel Genetic Regulations

Utilizing ANPs for therapeutics requires overcoming the barriers mentioned, as well as the more critical issues associated with modified oligonucleotides targeting downstream regulation. Though most gene therapeutics take effect by blocking access to mRNA, the downstream regulation strategies vary. Take two most well studied methods, antisense oligonucleotides(AONs) and small RNA interference(siRNA) as examples, AONs target and cleave mRNA by recognition via RNase while siRNA is recognized by RNA-induced silencing complex(RISC). Therefore, very limited chemical modification can be accessed

in order to achieve activities. The recognition of RNase H and RISC involves backbone structure, base structure, and secondary structure recognition [7], complicating the use of ANPs.

DNAzymes, however, are relatively more accessible to chemical modifications. Recognizing target mRNA directly by two binding arms, DNAzyme does not require the involvement of other protein complexes to achieve high fidelity. This quality opens a window for modified oligonucleotides design. A modified DNAyme 10-23 was reported by Wengel at. el. [8] with replacement of the two binding arms to lock nucleic acid (LNA). LNA nucleoside have proved effective in improving cellular stability as well as catalytic efficiency [9].

The goal of developing an ANPzyme is to expand the chemical modifications of oligonucleotides for gene therapy, and therefore, provide therapeutics that are more accessible to more permeable, chemically stable, and target specific gene expression networks.

1.1.5 ANP ligation enzyme

Development of an ANP ligation enzyme was first attempted by a previous group member, Dr. Jingdong Ye [10], who found that an amine linkage modified DNA was able to catalyze the ligation phosphodiester bond. By modifying the amine nitrogen with chemical moieties, this approach creates a hydrophobic pocket that assists the ligation in the presence of divalent metal. Similarly, there have been attempts to create artificial enzyme mimics of

the RNase H/RISC catalytic site since they share the cleavage domain [7]. It is proposed that in the RNase catalytic domain, the substrate is surrounded by four carboxylates as source of general acid base chemistry and coordinated with two divalent metals to activate the nucleophile, which is claimed to be a water molecule in this model [11]. Based on this model, catalytic systems have been developed on a general acid/base mechanism with the presence of divalent metals. A synthetic DNAzyme reported by Perrin and his coworkers [12] explored the possibility of this design. With imidazole modification on one base and primary lysine modified on the adjacent base, RNA hydrolysis was catalyzed, providing evidence for a designed artificial modified oligonucleotide enzyme.

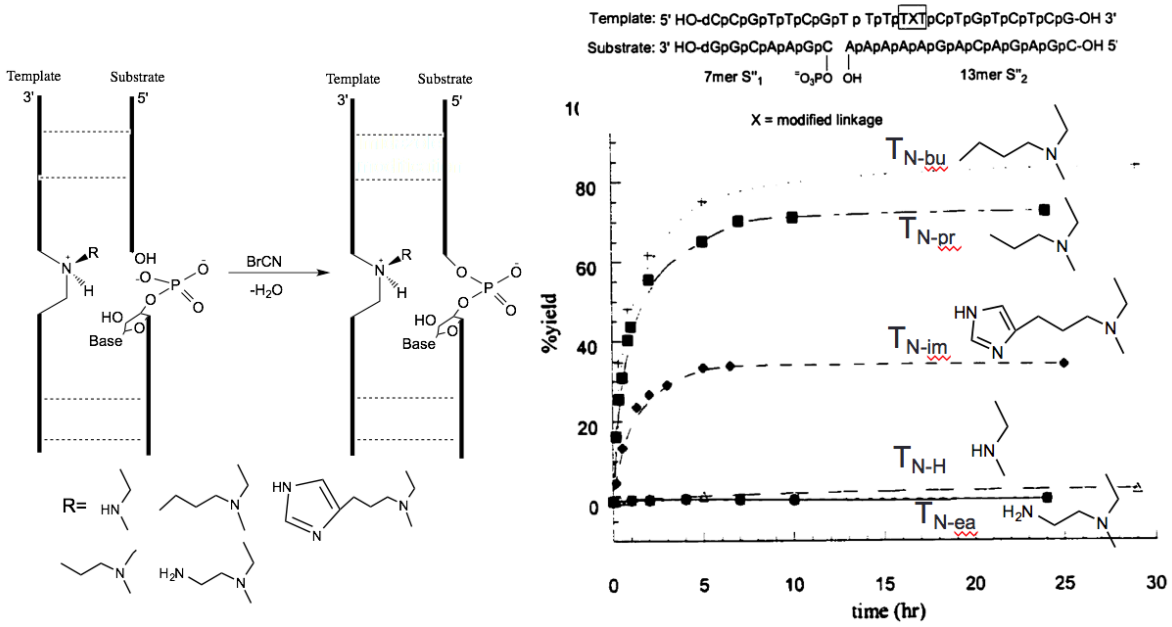


Figure 3 An ANP ligation enzyme.

In this thesis, we have developed an ANPzyme and demonstrated gene regulation taking advantage of the positively charged amine backbone to improve cellular uptake and

resistance to degradation. The designed catalytic site exploits the amine backbone, utilizing general acid/base catalysis with the presence of divalent metals. This study then aims to develop a novel gene regulation platform that can overcome some of the hurdles we currently face, including improving cellular uptake, resistance to cellular degradation and catalytic amplification. We focused on obtaining ANP oligomers at the first stage and then demonstrated improved cellular uptake in Jurkat cells. An ANPzyme model capable of targeting mRNA was designed as proof of principle.

1.2 Results & Discussion

1.2.1 Obtain ANP Oligomers and Characterize Their Biological Properties

While different strategies are under development by chemists in polymer synthesis, nature has developed a way to control efficiency and fidelity through macromolecule-templated synthesis, such as the transcription and translation processes. These well-developed natural processes enable complicated reactions in one pot with high sequential fidelity. As shown in **Figure 4**, DNA-template organic synthesis has been explored to template small molecules into complex polymers, and we have established a protocol to obtain ANP via DNA-template polymerization with modified ANP nucleosides. Previous work indicates that the ANP oligomers are formed in solution through a step-wise, geometric progression in the presence of a template DNA strand. This step-wise templated oligomerization mechanism provides for the step-growth production of oligomers of lengths of at least 32

base pairs. We will therefore synthesize ANP oligomers of n = 8, 16, and 32 oligomers in length using monomers synthesized from commercially available nucleosides.

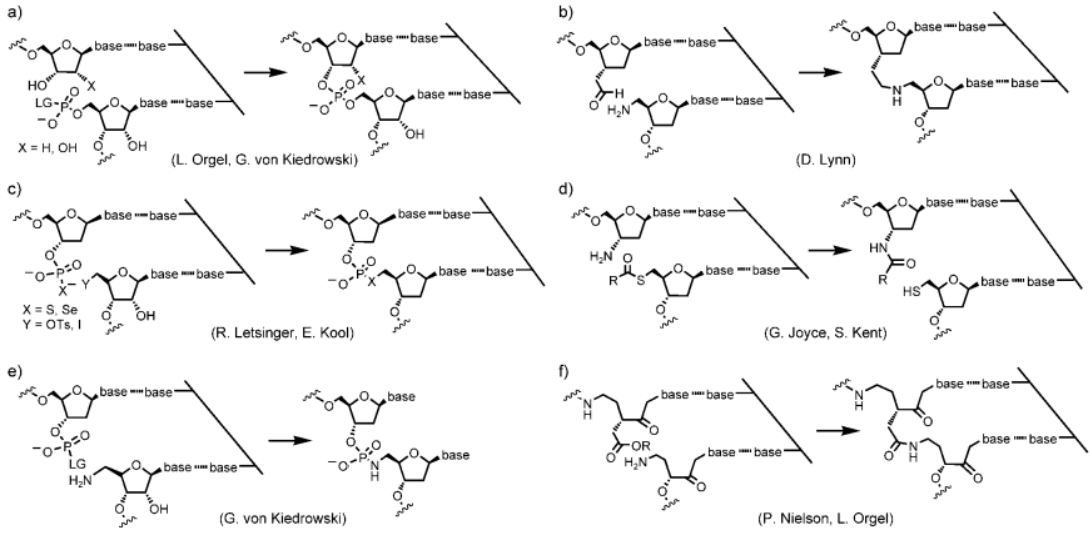


Figure 4 Representative DNA-templated syntheses of oligonucleotide analogues [13].

We further aim to develop a new strategy to construct ANP oligomers via solid phase synthesis. Compared to DNA-template polymerization, solid phase synthesis not only provides a more standardized, larger scale protocol, but also makes it possible to synthesize ANP oligomers in any lengths we desire. And therefore this strategy can serve as an excellent complimentary method to obtain ANP oligomers. This approach will allow gram-scale synthesis of ANP without the presence of DNA template with high fidelity.

The synthesis of ANP monomer is challenging considering all the protection/deprotection strategies and steric effects. We started with thymidine modified ANP monomers since the reagents are commercially available, cheap, and have fewer side reactions to avoid. Yet

synthetic chemistry for developing all four ANP nucleosides are under development in the lab. Currently, we are going to use thymidine ANP as a proof of concept. As a starting step, we prepared ANP oligomers follow DNA-template reactions established by Dr. Xiaoyu Li and continue to characterize ANP/DNA complex formation [14].

1.2.1.1 Synthesize ANP Monomers

ANP monomers are synthesized via modification of thymidine. Two different geometries are designed here in consideration of their potential difference in polymerization abilities and binding affinity to complimentary DNA strand. Previous work indicates that the ANP oligomers are formed in solution through a step-wise, geometric progression in the presence of a template DNA strand. The step-wise template oligomerization mechanism provides for the production of oligomers of lengths in multiples of eight and we prepared ANP oligomers of n=8 monomers in length as a model system for subsequent studies. Both geometries were synthesized as outlined below (**Figure 5**).

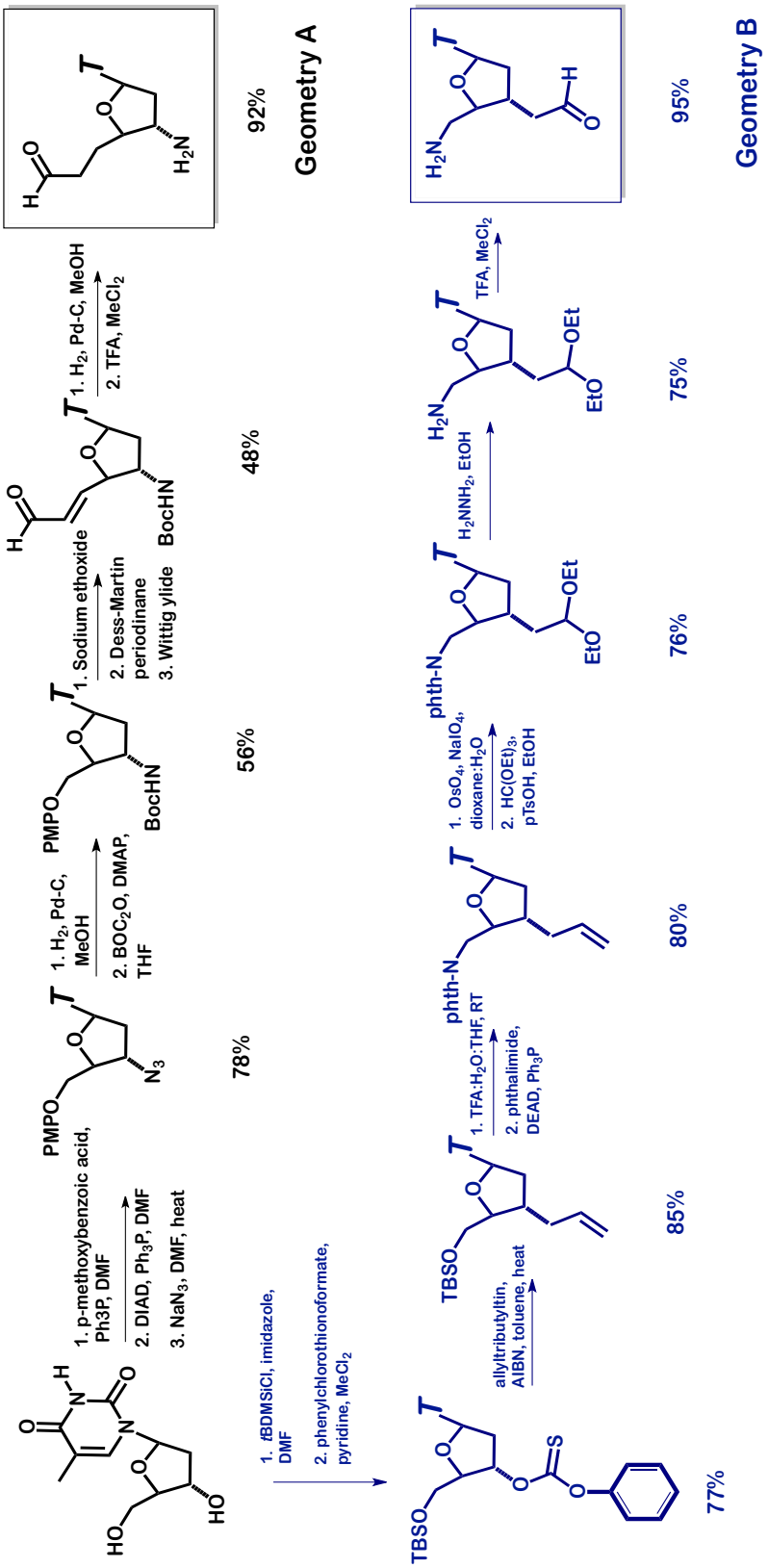


Figure 5 Obtain ANP oligomers by DNA template reactions.

While the ANP oligomers are obtained from DNA template reaction established by Xiaoyu Li, different polymerization and analytical conditions, including reaction temperature, buffer, pH and HPLC gradient are screened for the different monomer geometries.

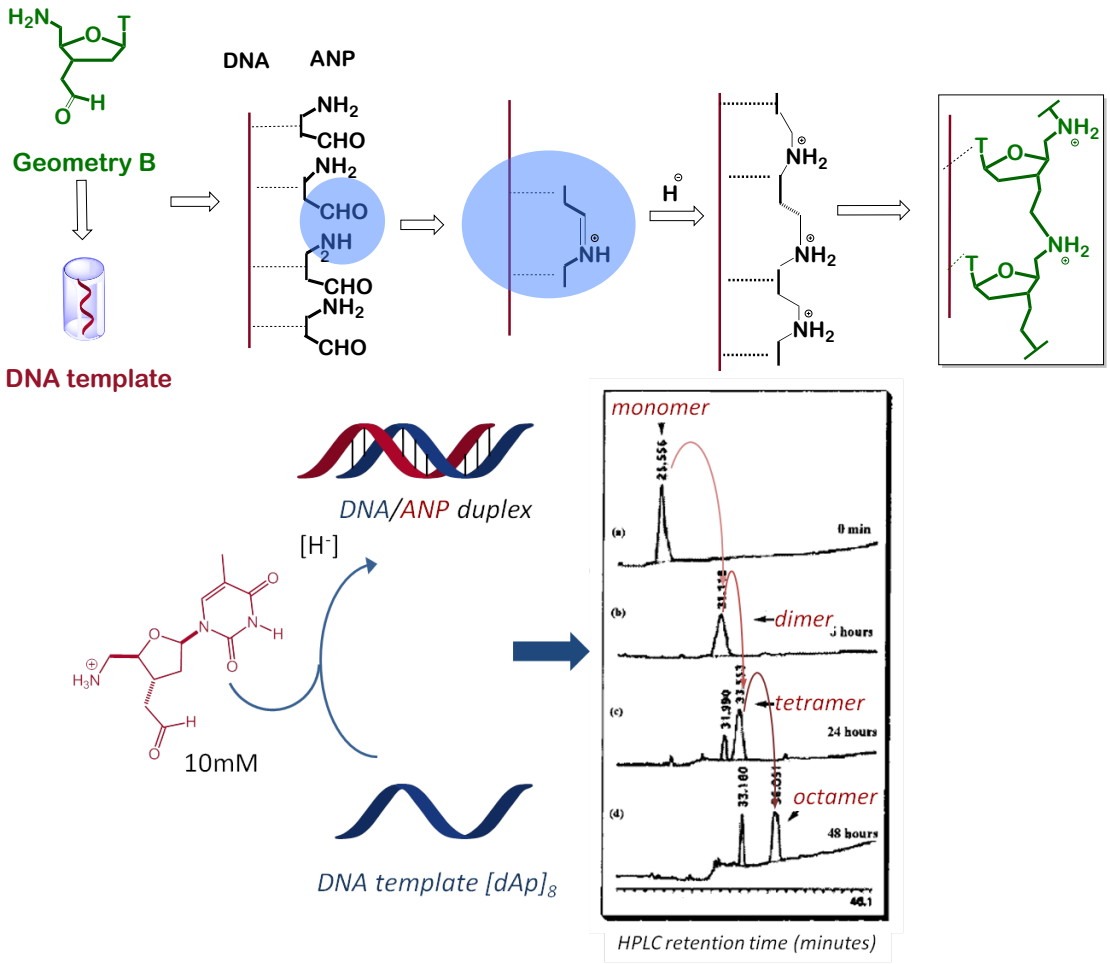


Figure 6 DNA-template ANP polymerization characterized by HPLC.

1.2.1.2 Synthesis of ANP Dimers

ANP dimers with amide linkage are synthesized for two aims. The dimer will have higher affinity for the template and thus improve polymerization efficiency. Also, the amide

linkage is more rigid than amine linkage and is therefore known to improve binding affinity. On the other hand, synthesis of ANP dimer through amide condensation provides us with a model system for solid phase synthesis. By studying the model systems, information on coupling efficiency, protection/deprotection chemistries can be obtained, and therefore, make it possible for solid phase synthesis of ANP oligomers.

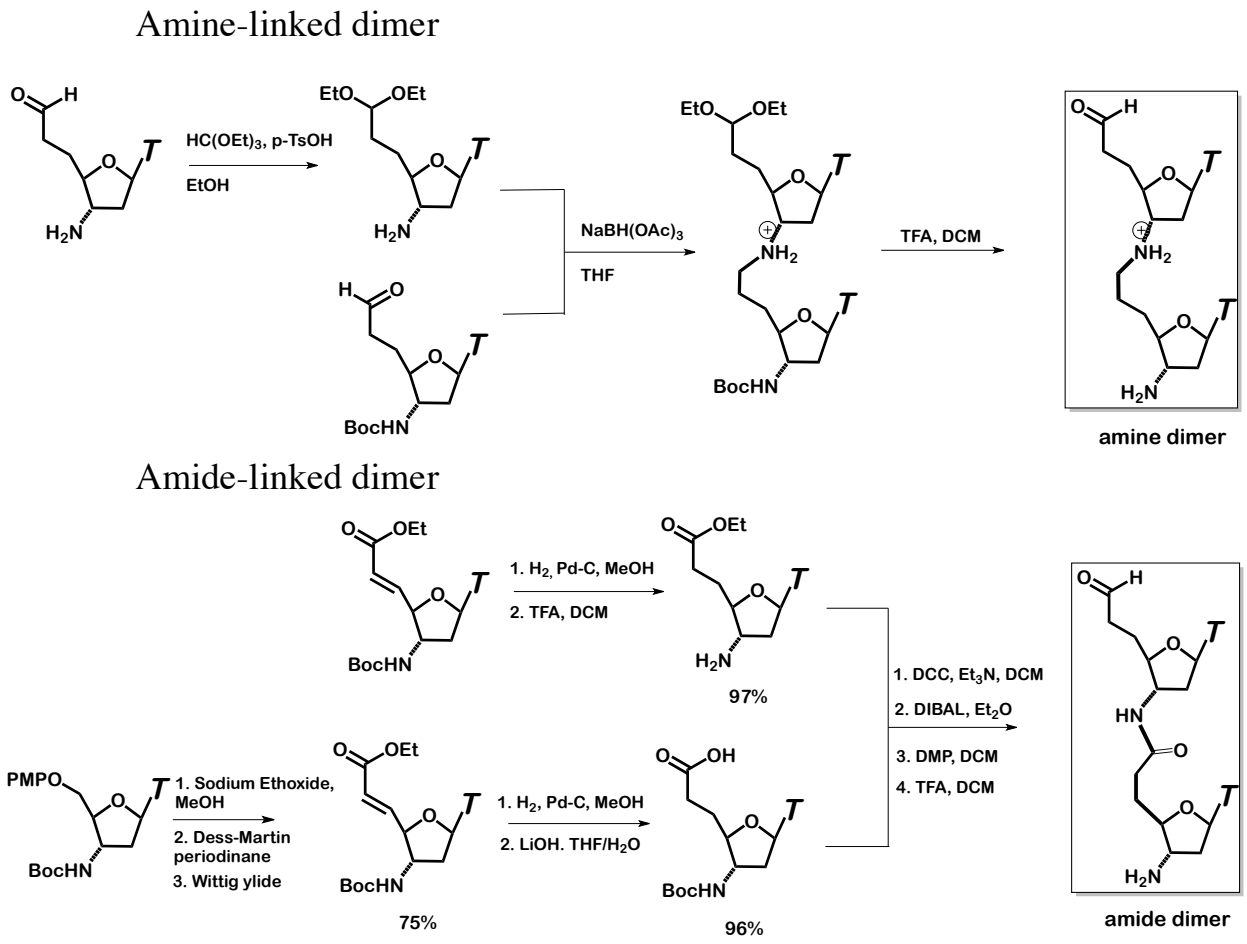


Figure 7 Synthesis of ANP dimer in two geometries.

1.2.1.3 Scaling Up via Solid phase Synthesis

Solid phase synthesis was conducted with peptide synthesized from 5' to 3'. Monomers are attached to resin in the 5'end while deprotected amine in the 3' end reacts with another monomer. The polymerized product was reduced to amine bond to yield the 8-base pair final product in gram-scale (**Figure 8)**.

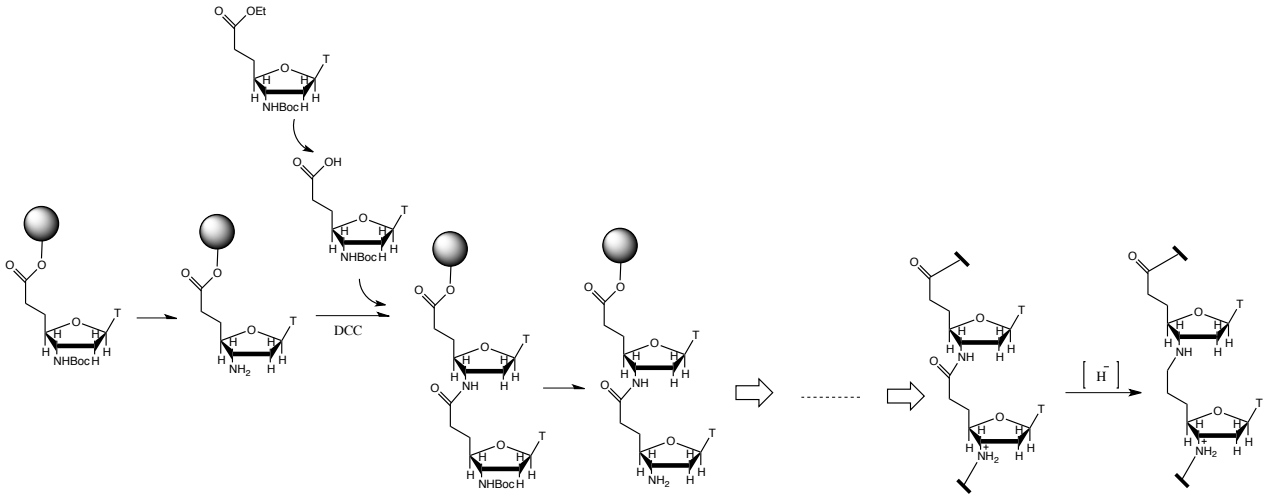


Figure 8 Solid-phase synthesis of ANP.

1.2.1.4 Characterize DNA/ANP Duplex Structures

Stability of the ANP/DNA complex will impact mRNA targeting and gene regulation efficiency. We obtained evidence that ANP and DNA form 1:1 complexes with Job-plots, and also detected melting temperature for a ANP/DNA 32mer (**Figure 9)**.

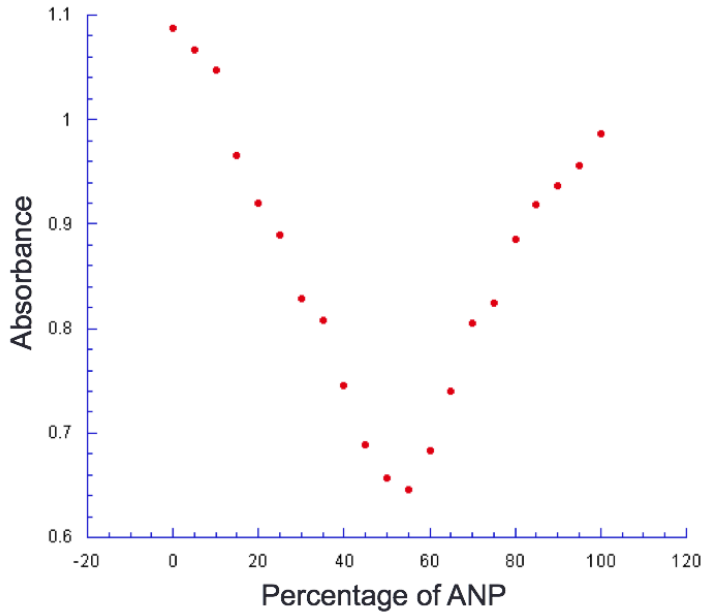


Figure 9 Absolute absorbance plotted as a function of different mole fraction of the DNA template with the 32-mer ANP product.

1.2.2 Prove Cellular Uptake in Jurkat Cells

1.2.2.1 Modified Nucleoside Polymers Improve Cellular Uptake

Sei *et. el.* [15] has reported an Peptide Nucleic Acid (PNA) 22mer that is able to penetrate the cell by itself and demonstrate gene mutation. Yet the delivery of naked nucleotides involves low efficiency, cellular degradation and other side effects while delivery with modified nucleotides tend to be more commonly used in recent years. Delivery of modified nucleotides includes conjugation to lipophilic moieties, peptides, cell-specific receptor ligands, and polycationic polymers [16].

Currently the most commonly used non-viral delivery vectors includes nucleic complex with cationic lipids, cationic polymers, biodegradable cationic polysaccharides, and cationic polypeptides, via electrostatic interactions. It is widely believed that delivery efficiency can be improved by changing the charge ratio associated with nucleic acids and condensing agents. Among these polycationic polymers, polyethylene imine (PEI) is one of the most commonly used. Its amine backbone has been proved able to condense nucleic acid, endosome release and exhibit low toxicity.

1.2.2.2 ANPs demonstrated improved cellular uptake

We have shown preliminary results of cellular uptake of ANPs by confocal microscopy of thymidine-dANP octamer bound to a complementary DNA sequence containing a 5′-end labeled with fluorescein (Fl). The confocal micrographs (**Figure 10**) are of 2D superpositions of 3D stacks taken through the middle of cultured Jurkat cells cultured with the ANP/DNA duplex (left) or the labeled dA octamer strand alone. This experiment suggests that not only are the ANPs sequestered, but they are able to serve as cellular carriers for other cargo, including their complementary DNA sequence. The reasons for the apparent clustering of the fluorescence may be membrane or endosomal sequestering in these early time points, further emphasizing the importance of the zwitterionic duplex in traversing the cell membrane. We are now poised to focus on characterizing the rates, cellular localization and lifetime of the full length strands by labeling the ANP oligomer itself.

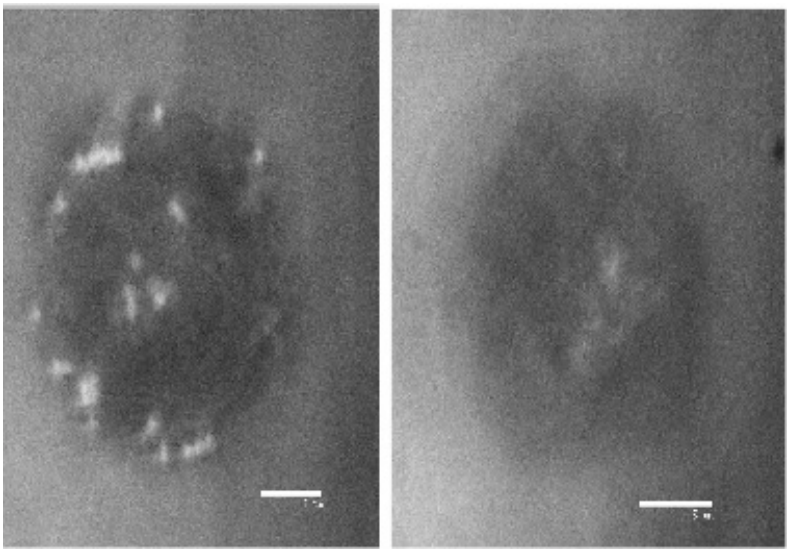


Figure 10 Absolute absorbance plotted as a function of different mole fraction of the DNA template with the 32-mer ANP product.

1.2.2.3 Label ANP Building Blocks with BODIPY Dye

We choose to label the aldehyde in the 5'end of the monomer/dimer to ensure specificity. The fluorescent dye we choose here, boron-dipyrromethene (BODIPY) has been proved efficient for in vivo imaging in living cell [17]. BODIPY has the advantage of long excitation wavelength, large stokes shift, chemically stable, accessible to chemical modification and high quantum yield, which is an excellent choice in this case. The synthesis of BODIPY is shown in **Figure 11**. We incorporated the fluorescent dye into the ANP oligomer by adding dye-labeled monomer to the templated reaction, in a ratio of 1:(n-1) to unlabeled monomer. The dye-labeled monomer reacts at a similar rate as for the unlabeled monomer, and yielded full-length ANP oligomers with the fluorescein label incorporated at the terminus. Product was characterized by HPLC and mass spectrometry.

To further demonstrate improved cellular uptake, cellular localization of BODIPY labeled ANP oligomer will be determined in Jurkat cells following previous protocols.

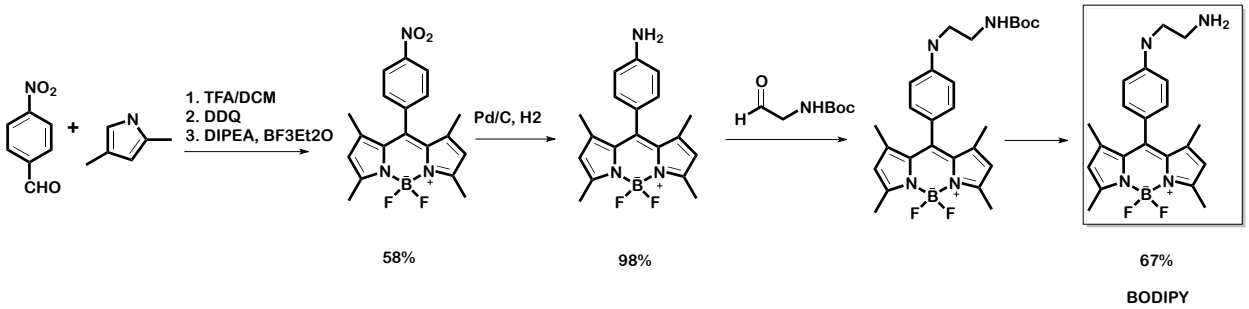


Figure 11 Synthesis of BODIPY dye.

1.2.3 Develop an ANPzyme Able to Cleave mRNA

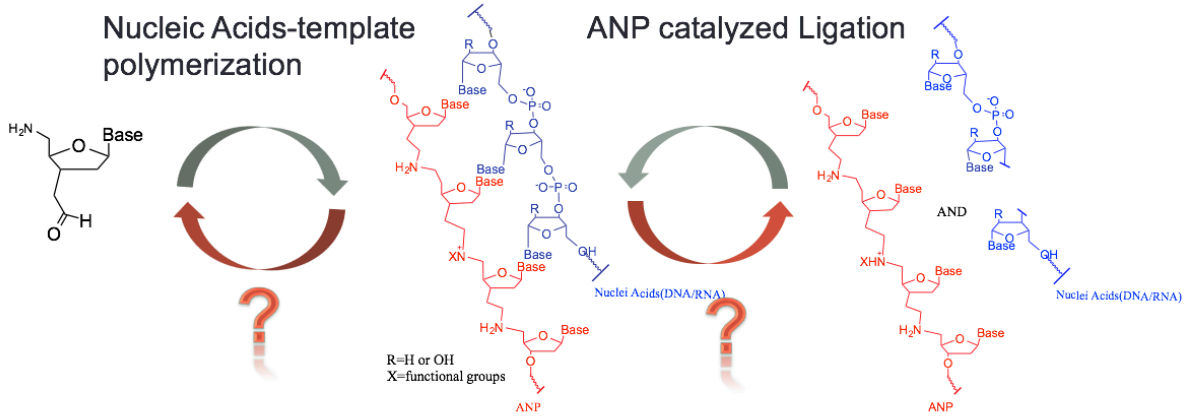


Figure 1-12 Schema of information encoding/decoding pathway.

we have demonstrated information encoding from simple units to complicated biopolymers via template-directed polymerization and ligation, we'd like to explore the reverse process: information decoding & re-coding, which is also critical in natural information flow

process. This can be achieved by template-directed, target-specific cleavage of the information carrier biopolymer.

1.2.3.1 Chemically modified nucleotides as artificially enzymes

As a next step goal, we aim developed an ANPzyme that is able to cleave mRNA upon specific recognition. Though cellular stability of those gene therapeutics can be largely improved by chemically modified nucleotides such as locked nucleic acid, 2-methyoxyl nucleic acid, and etc. [15] (**Figure 13**).

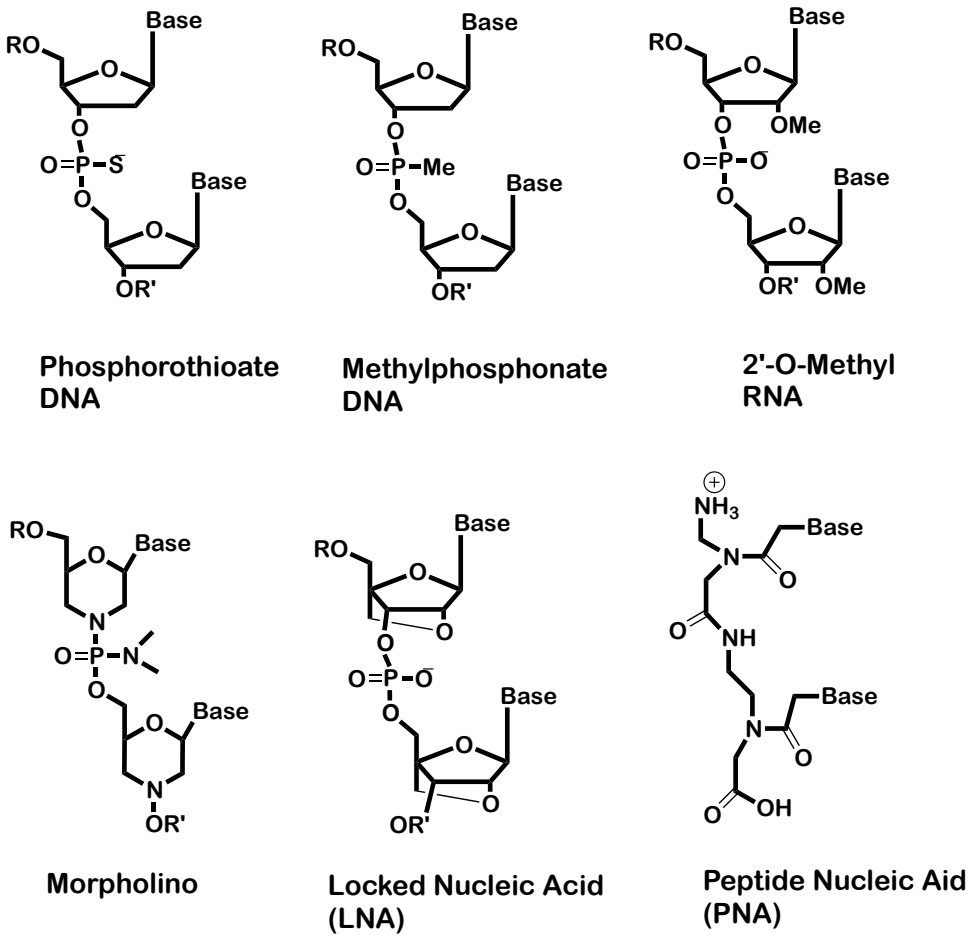


Figure 13 Chemically modified nucleotides.

However, one critical limitation is that majority of those modified nucleotides can not be recognized by RNase H or RISC, and therefore can not demonstrate gene regulation. In order to utilize modified nucleotides, either new recognition enzymes or specific binding by base pairing has to be screened, of which the latter seems more promising at current stage. What's more, RNase H and RISC recognition usually allow a certain extent of mismatches which will cause off target effects [18].

Currently, off target effect has become a major concern in gene regulation, especially for antisense oligonucleotides, RNA interference and steric blocking pathway, which are most commonly adopted [19, 20]. Therefore, it is important to develop a gene therapeutic that allow both chemical modification and ensure target specificity.

Though various DNAzyme has been developed during the past decades, the mechanism of its catalytic site has not been clearly revealed, therefore makes it challenge to develop artificial enzymes. In this paper, our idea mainly focused on mimicking RNase H cleavage mechanism, which utilizes a general base and acid. We aim to mimic this cleavage process my modifying ANP with an imidazole base while the proton is provided by the backbone.

1.2.3.2 Overall Design of ANPzyme

The design of ANPzyme follows RNase H mechanism, which utilizes an imidazole base and primary amine as proton source. The nucleobase (thymidine as an example here) is modified with an imidazole base while the amine backbone as a proton source. Ideally, the

imidazole will be held proposed to the RNA phosphodiester linkage and assist the cleavage of the phosphodiester backbone. The designed ANPzyme was obtained by amide coupling of two modified DNA fragment, and then followed by reduction to yield amine linkage (**Figure 14**).

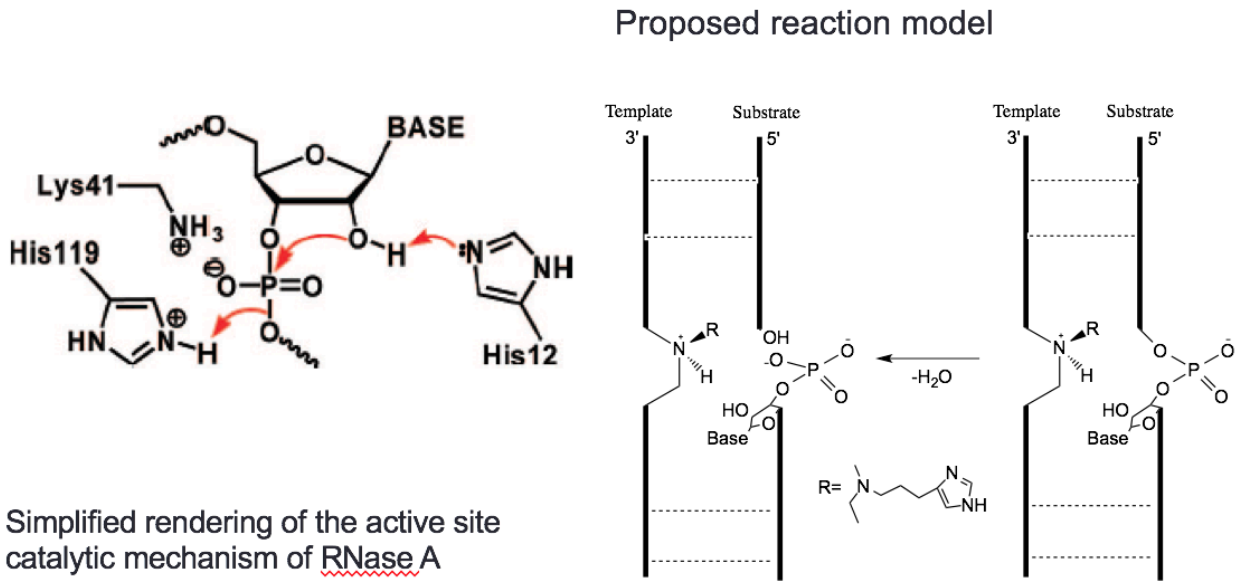


Figure 14 An overall design of ANPzyme which mimics RNaseA.

1.2.3.3 Design of Catalytic Site

One critical point is to identify the position of the imidazole, making sure it is inserted inside the base pairing instead of hanging outside the duplex. A computational model was built to estimate the position of the imidazole base. Two key factors were studied in this model: the orientation of the side chain and the length of the side chain.

A previous study on confirmation of ANP/DNA duplex was conducted by previous group members [10, 14]. According to a simulation model built by molecular dynamics and NMR

constraints, ANP/DNA duplex adopted similar B-form confirmation as native DNA duplex, while ANP/DNA showed looser binding and more internal space at modification site compared to unmodified DNA duplex (**Figure 15**). This evidence indicated that ANP/DNA duplex could potential provide enough room for modified side chain at catalytic site.

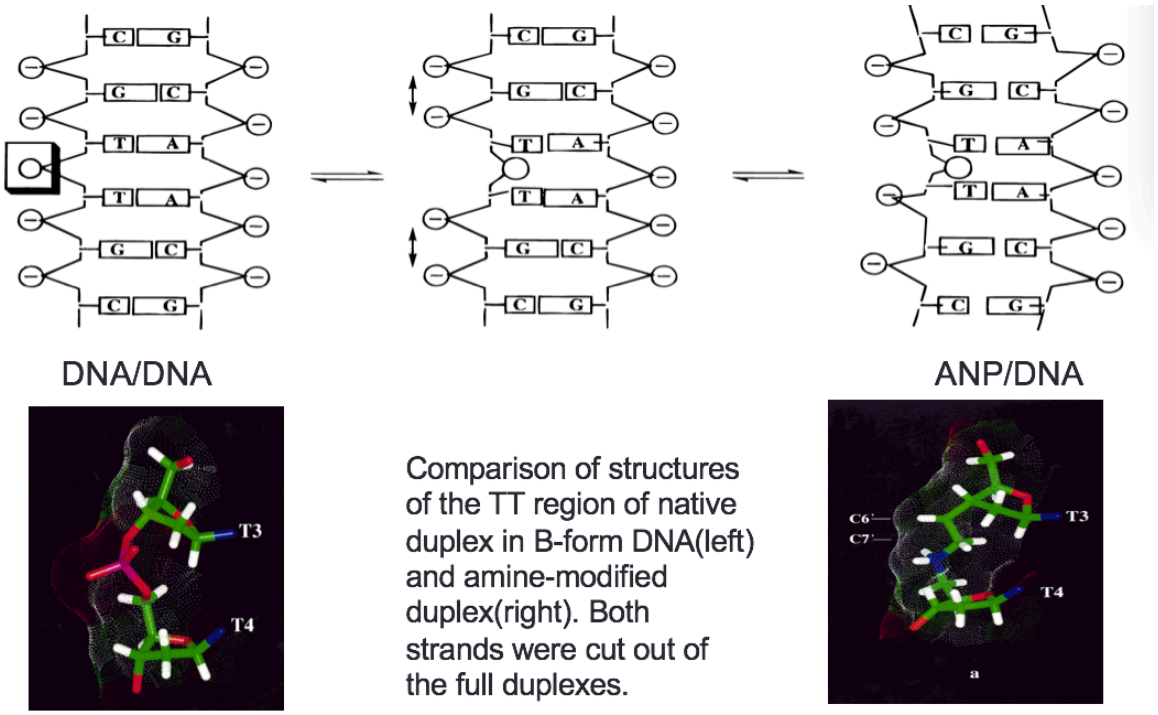


Figure 15 Comparison of structures of the TT region of native duplex in B-form DNA(left) and amine-modified duplex(right).

1.2.3.4 Visualization of Side Chain Orientation

Further study of catalytic side chain position at catalytic site was conducted by molecular dynamics simulation. For each potential side chain length, I conducted a conformational search of proposed structure within 7.5 KJ of energy range. This search resulted in thousands of potential orientation of the side chain at designed catalytic site. Simulated data was modeled with PCA and k-means clustering to yield four distinct groups. As shown

in results, one particularly promising group was identified. In this group, the side chain is within reach distance of the catalytic substrate.

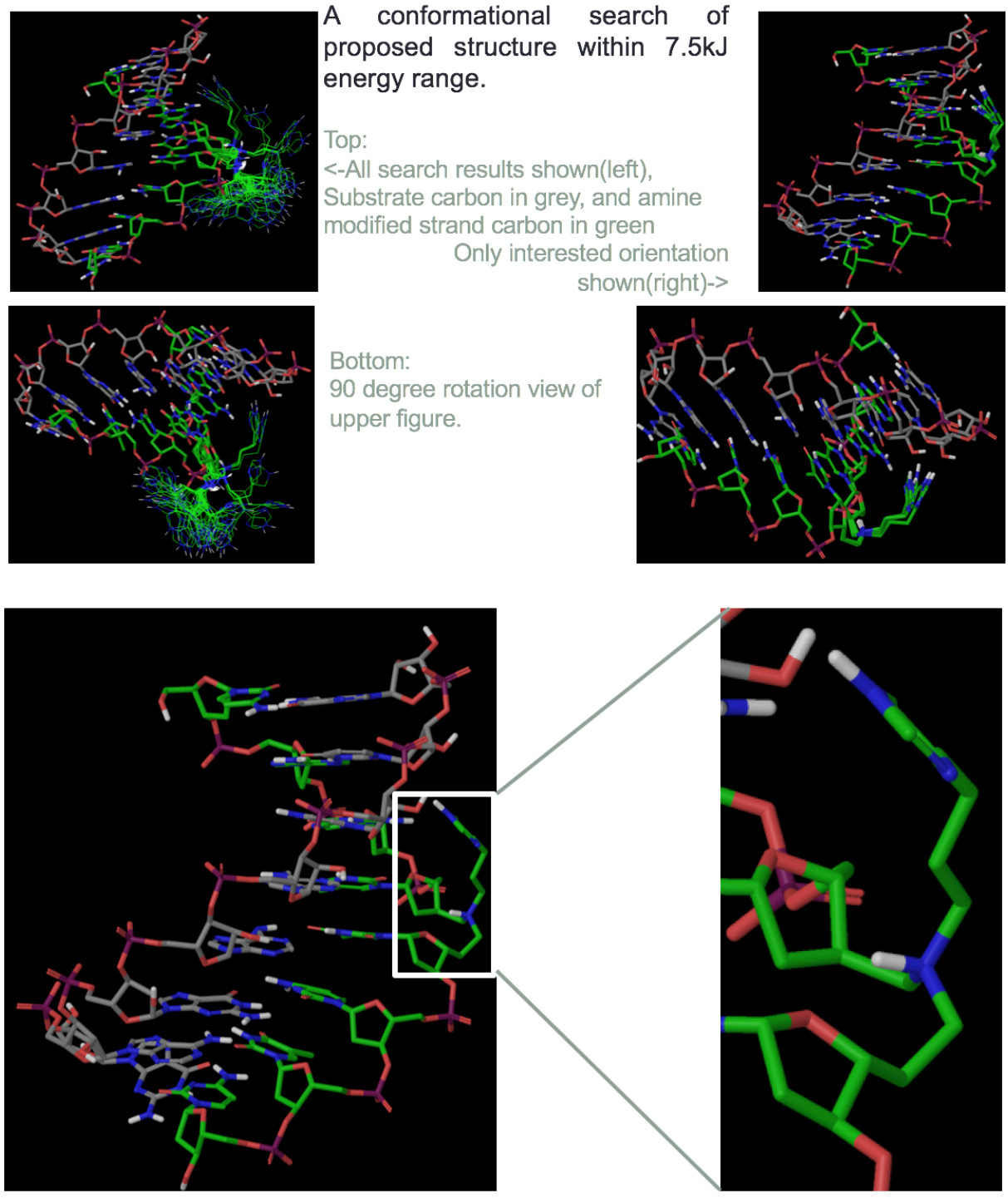


Figure 16 Clustering results of conformational search(top) & identified side chain group within reaching distance of cleaving substrate(bottom).

1.2.3.5 Data Visualization and Analytics by Dimension Reduction and Clustering

The conformational search simulation yielded huge data set with high dimension, which made it difficult to study & identify desired group by traditional statistical methods. And therefore, I researched two clustering methods: K-Means clustering and Expectation Maximization (EM) clustering [21] and four dimension-reduction methods [22] : principle component analysis (PCA), independent component analysis (ICA), random projection (RP), and ReliefAttributeEvl (RAE) to achieve better separation of simulated groups. The combination of PCA and K-Means clustering has given the best results. Original dataset was preprocessed in Weka (Java) while further visualization and analysis were achieved by Python.

PCA can serve as the best representation of nonlinear models as it provides a natural criterion to select the modes as they represent the energy contained in the data. In other words, PCA minimizes the covariance of the data, yielding orthogonal vectors of high energy contents in terms of the variance of the signals. However, PCA is not directly related to the structural frequencies, as it is simply the energy of the signal measured in Frobenius norm, not necessarily in terms of physical energy. As a result, PCA is among the most popular dimension reduction algorithms, and can work decently on almost all kinds of data that contains Gaussian components.

I applied PCA to my original data (n dimensions), and it returns n principal components, which are ordered based on the variance it can explain (**Figure 17**). The variance of each principal component is the corresponding eigenvalue.

It appeared that the eigenvalue decreases significantly for the data as I go from the first principal component to the last component, indicating that it was able to extract useful information to much few components, hence it should not compromise the performance after dimension reduction.

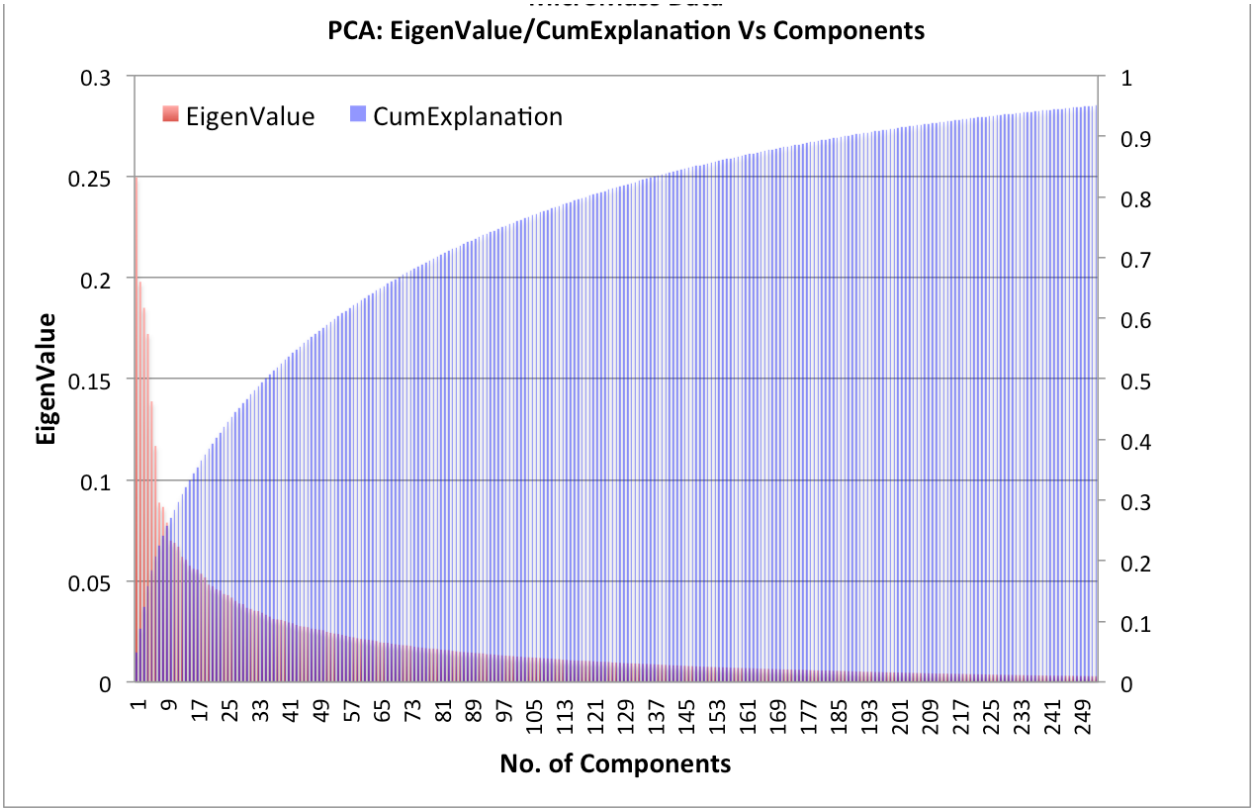


Figure 17 Principle component analysis of simulation data to achieve dimension reduction.

For both clustering algorithms, it's important to select the number of clusters: K. Intuitively, K should be the output number of the classes so that each instance can belong to an output class. To support that assumption, I evaluated K within a range (K=2:30). As shown below, if K is larger than the output class (K=2), then the extra classes will not be classified to any classes, and the improvement on accuracy was not significant. So K=20 was selected.

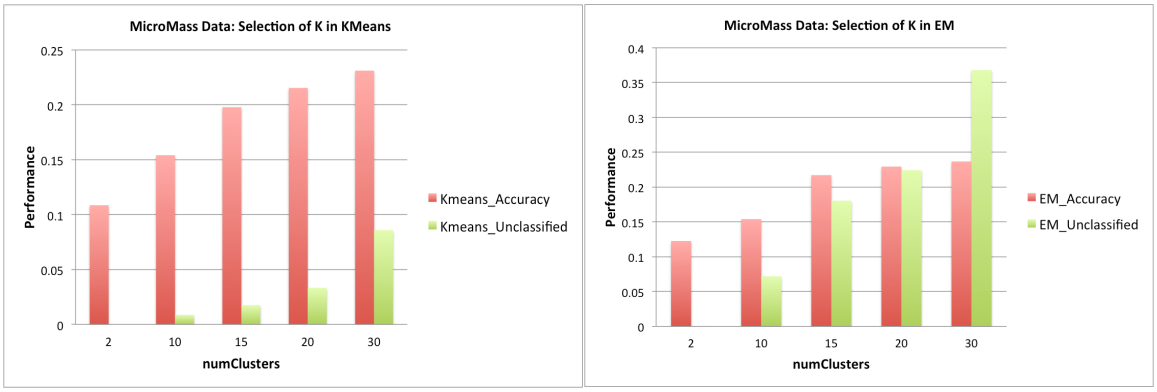


Figure 18 Selection of K in K-Means and EM clustering.

As shown in the statistics above, K-Means and EM behaved very similarly on clustering the datasets. K-Means used 'hard clustering', which tries to identify the boundary clearly while EM used a 'soft clustering' approach, whose boundaries are often blurred. However, in my dataset because both are too difficult to cluster without any dimension reduction, I was not able to see the differences in boundaries (**Figure 19**).

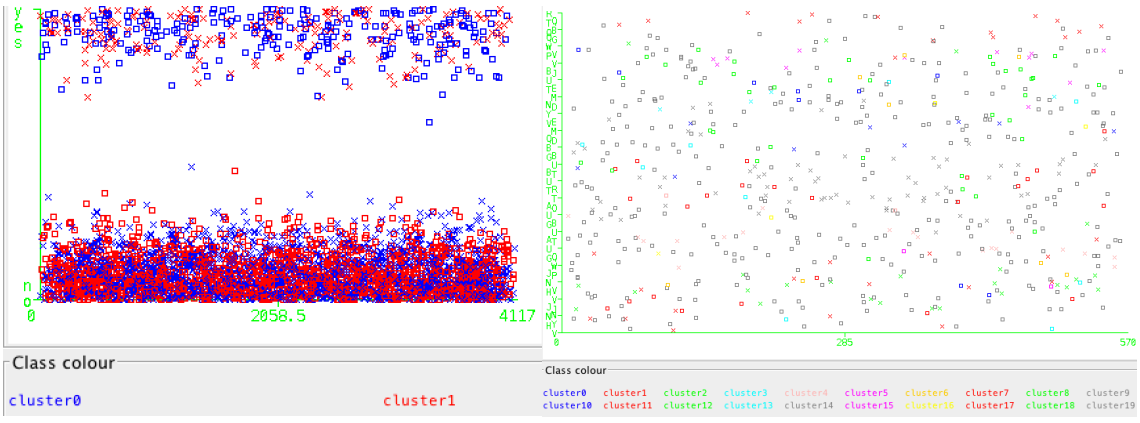


Figure 19 Comparison of clustering boundary in K-Means and EM clustering.

Since K-Means clustering is easily caught by local optima, and that it randomly selects its initial centroids, so the initialization method will affect its performance. I explored four initialization methods: Random, k-means++, Canopy and Farthest first, and the results showed that they don't make a big difference but the Farthest first method behaves slightly better (result for Mass data not shown). This is probably because both datasets are widely dispersed. However, in such a situation, the result can be highly variable due to it's selection of the farthest bound. And therefore I chose a random method for all the following experiments.

EM is susceptible to false minimum, too, because it relies on the random selection of a good starting point. This can be solved by random restart. The other problem of EM is that it doesn't converge but the performance can be improved over more iterations. The default number of 100 is used for all the following experiments.

1.2.3.6 Synthesis of ANPzyme Catalytic Site

The synthesis of ANPzyme catalytic site followed previous steps of ANP dimer synthesis. Incorporation of side chain was achieved by conjugating a modified imidazole to the backbone via imine condensation and reduction (shown in **Figure 20**). This method allows variation of side chain catalytic type and chain length.

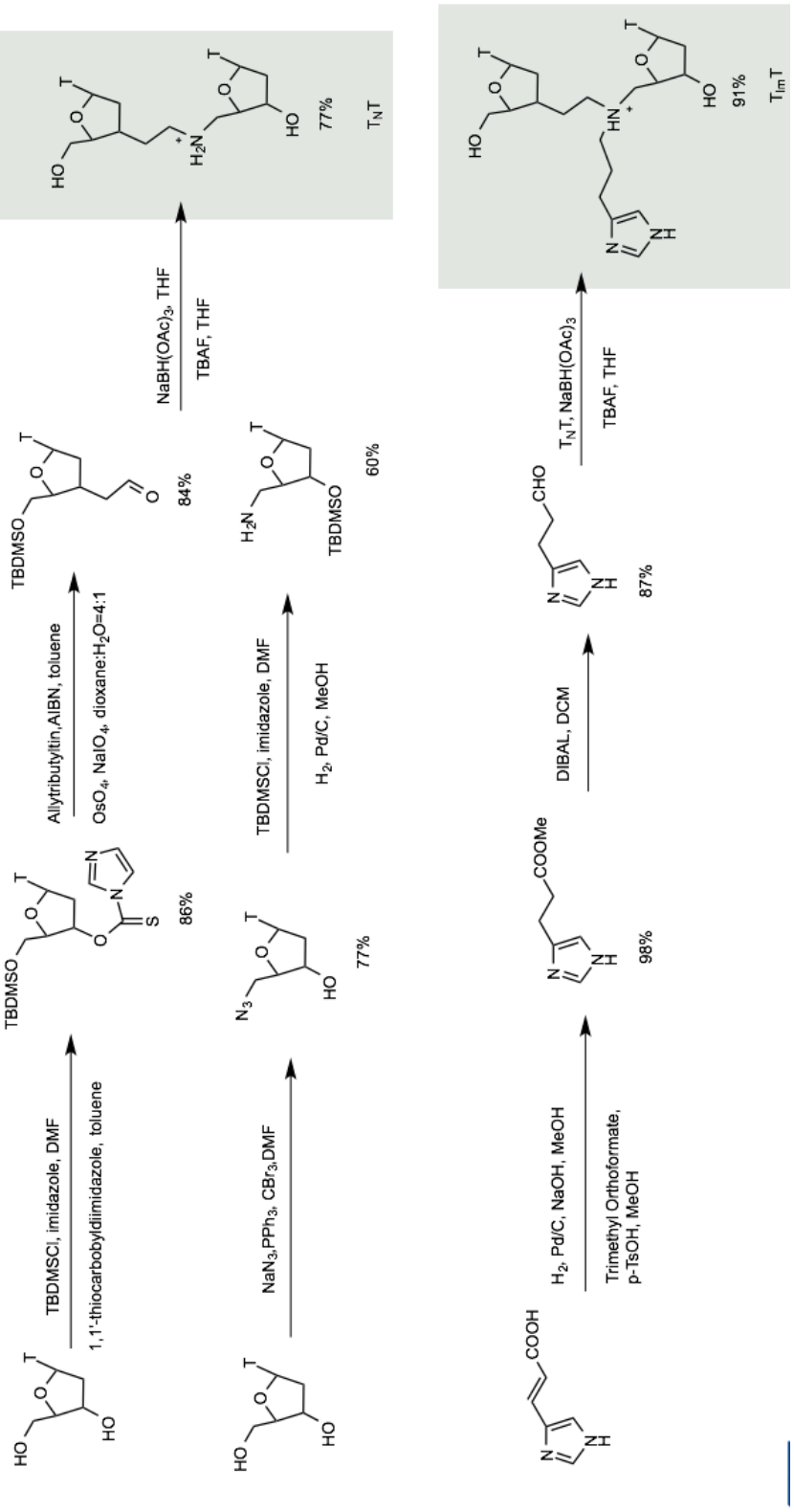


Figure 20 Synthesis of potential ANPzyme core catalytic site.

1.2.3.7 Demonstrate Catalytic Activity

Proof of ANPzyme catalytic activity was conducted in vitro. Reaction condition including concentration, solvent, buffer, pH and temperature was determined. Reaction results was analyzed by native PAGE. A 40 base pair length substrate of 5'-TCT CTC GCT TCT CTC TCT CTC TTT rArA TTT CTC T-3' DNA/RNA mixture was chosen, which will yield a 31 base pair detectable products 5'-TCT CTC CGT GCT CTC TTT rA-3' strand and another 9 base pair undetectable strand 5'-rA TTT CTC T-3'. Reaction was conducted with 1μM substrate concentration in 150 mM NaCl, 50mM HEPES, pH 7.0 buffer at 4 ^{0}C. As shown in PAGE result, only in presence of ANPzyme, it was able to yield cleavage product while even with both modified dimer TnT and free imidazole, there's no reaction. It indicates that both the catalytic group and it's positioning near the substrate are critical for the cleavage reaction. Time and concentration dependence of the enzyme was studied by observing ANPzyme concentration variation from 1μM to 1000 μM . While ANPzyme showed almost linearly increasing catalytic efficiency, product concentration reached plateau after 30 hours.

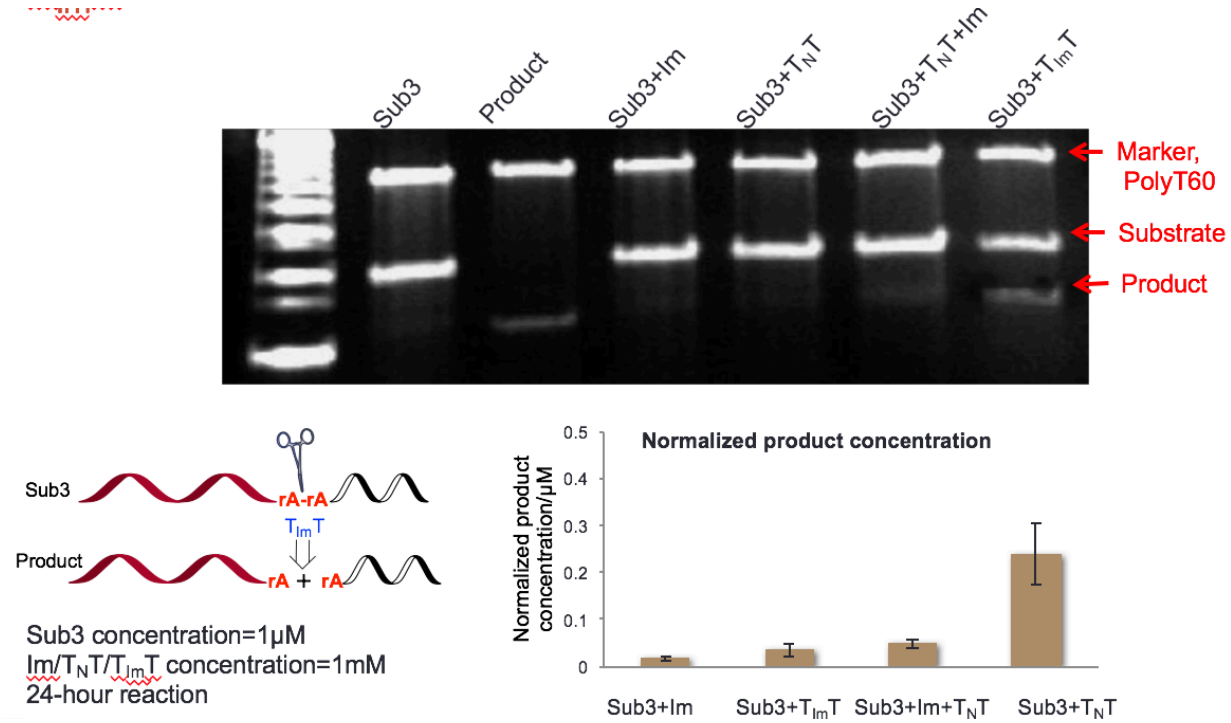


Figure 21 ANPzyme cleavage assay. Top: reaction results visualized by PAGE; Bottom left: schema of cleavage reaction; Bottom right: normalized product concentration.

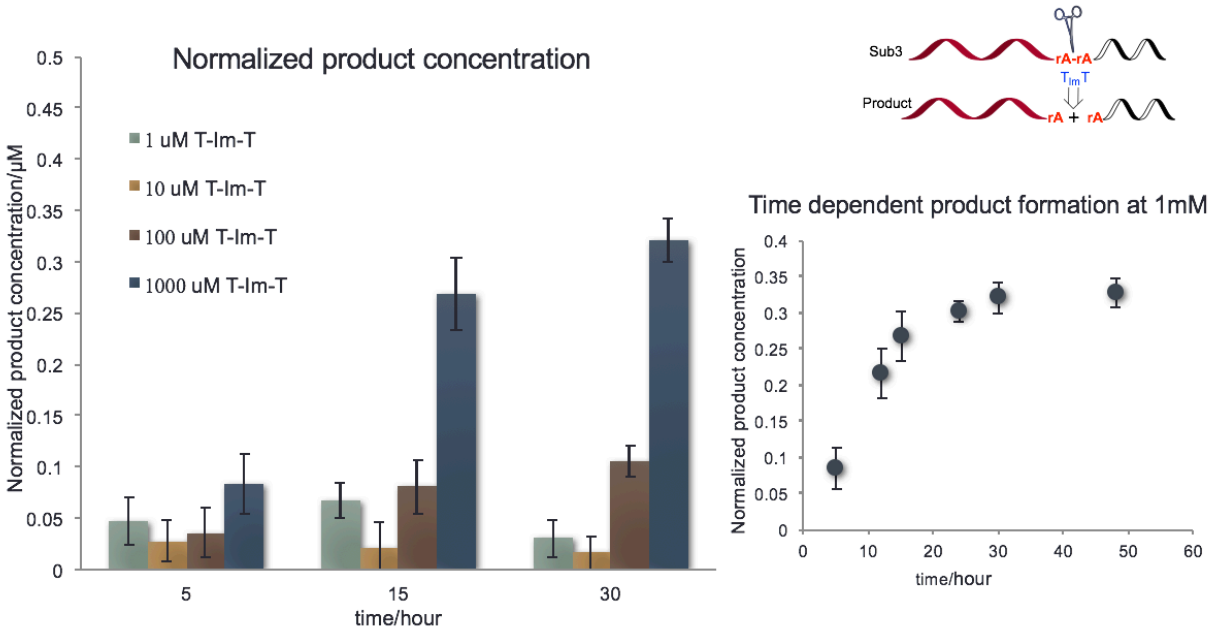


Figure 22 Time and concentration dependence of cleavage reaction.

1.3 Conclusion & Future Directions

In this study, a novel ANPzyme as a gene regulation platform is proposed and developed. We demonstrated the catalytic efficiency of designed ANPzyme and supported those results by computational simulation. A next step goal would be to optimize catalytic efficiency of the designed ANPzyme and show in vivo catalytic activities. As studied above, the most critical effect of catalytic efficiency is the positioning of the functional side chain. Based on our model, the side chain position can be further improved by increase side chain length by 1 carbon and fix the orientation of the side chain by more rigid amide conjugation (**Figure 23**). The success of this design established the ability to convey information flow from native biopolymers to synthetic biopolymers and vice versa. This has greatly empowered our ability to further edit/control genetic encoding of known knowledge to the growth, development, functioning and reproduction of living organisms, and established a promising platform for gene therapy.

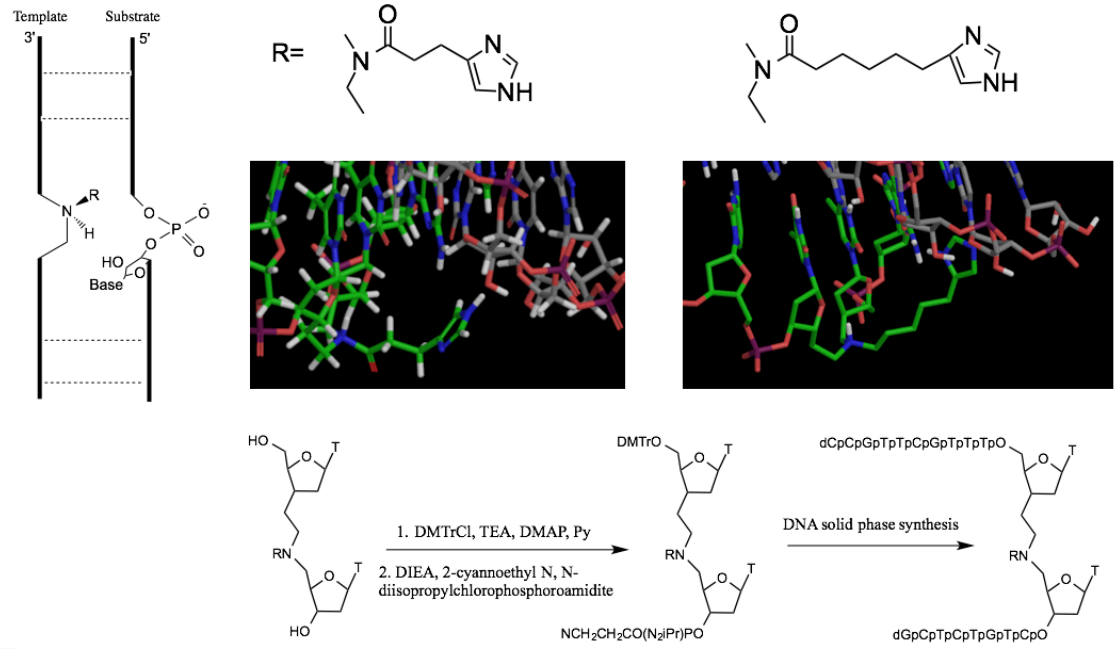


Figure 23 Improving catalytic efficiency by side chain modification.

Chapter 2: Understanding Conformational Evolution of Dynamic Chemical Library by Computational Simulation

2.1 Introduction

2.1.1 Nucleic Acid Dynamic Chemical Networks with reversible imine linkage

As discussed in chapter 1, the first dynamic chemical network (DCN) of imine reversible linkage was achieved in our lab in 1992 [5]. In the presence of the DNA hexamer template dGCAACG, a complementary 5'-amine trimer 5'-$^{+}H_3N$-dTGC, competes with non-complementary tetramer sequence of 5'-$^{+}H_3N$-dTTTT, for the coupling of the complementary 3'-aldehye trimer (**Figure 24**). The nature of imine lability provides equilibrium between coupled and uncoupled substrates, allowing for thermodynamics of substrate-template association to direct product formation. The DNA template biased the production of the complementary hexamer 30-fold over the non-complementary heptamer at 0°C.

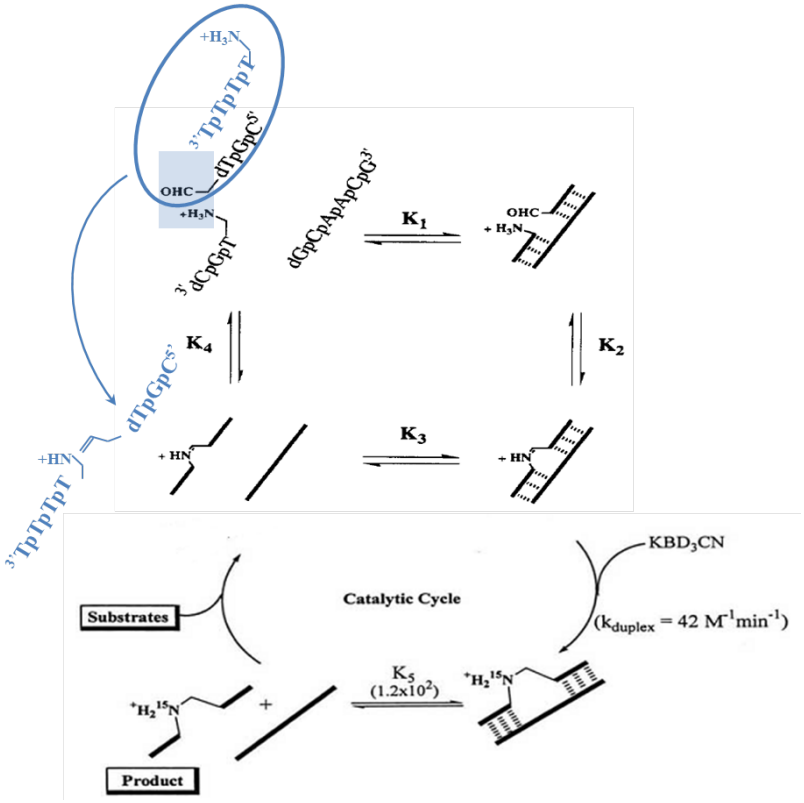


Figure 24 DNA template-directed synthesis in imine-linked dynamic network from mono-functionalized substrates (figure modified from [23] [24]).

In 2002, the imine linked network was extended from mono-functionalized substrates with amine or aldehyde to bis-functionalized substrates 5'-H_2N-dT-3'-CH_2CHO, where the equilibrium of imine formation is driven by the octameric DNA template $(dAp)_8$ under thermodynamic control [25]. In this DCN, the imine coupling among monomers in aqueous solution is surprisingly robust and achieved the first chain-length-specific and sequence specific reading of a DNA template through monomer polymerization. The reaction demonstrate a new step-growth kinetics mechanism, and extended high fidelity copying of a 32-mer DNA sequence (**Figure 25**) [26]. Following genotypic information transferred to the imine linked polymer, the product polymers are irreversibly reduced by sodium borohydride to the amine nucleoside polymers (ANP). While reductive amination is

synthetically useful to generating ANP, this reduction method does not maintain a dynamic network.

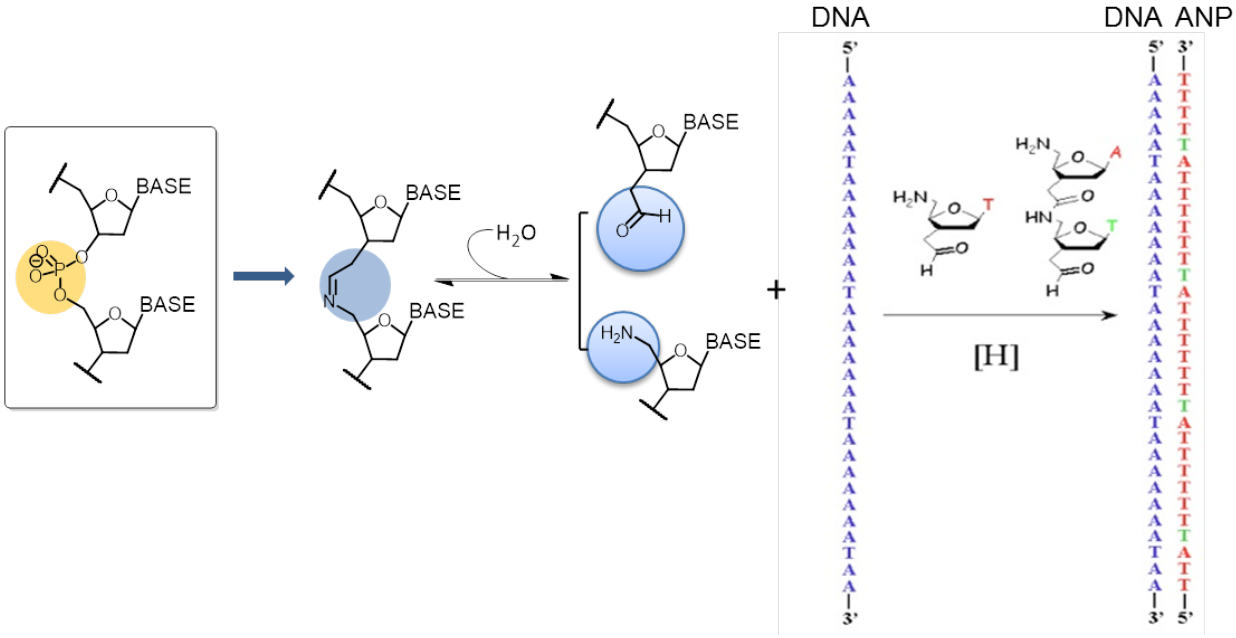


Figure 25 DNA template-directed synthesis in imine-linked dynamic network from bis-functionalized substrates. (figure modified from reference [26]).

If the imine can be stabilized through an additional reversible reaction instead of irreversible reduction, then the dynamics in the network will be maintained for responsiveness towards external stimuli. Incorporating this imine condensation reaction in an RNA backbone positions the 2'-OH as an intramolecular trap for the intermediate imine, and unlike the ANP, the resulting N,O-acetal would maintain reversibility through a condensation/hydrolysis manifold.

2.1.2 Extending Reversible Acetal linkages to Peptide Dynamic Chemical Networks

This reversible acetal linkage was further extended from nucleic acid biopolymers to peptide, reasoning that not only nucleic acid polymers can be connected through kinetically accessible linkages for replacement of phosphodiester bond, but the peptides could also be engineered with similar reversible linkages. The hypothesis is that the peptide dynamic network linked by acetal linkages will be able to respond towards external stimuli and provide opportunity for selection of desired oligomer and progressively generated assembly, taking one step closer to an evolvable system.

Simple change in the oxidation state of the C-terminal amino acid to an aldehyde allows coupling of another amino acid's N-terminal amine functional group as imine linkage for generating dynamic networks. Taking advantage of the greater inventory of amino acids with various side chains, the imine linkage can be further stabilized through intramolecular trapping by residue side chains. Side chain of hydroxyl functional group including serine (Ser), threonine (Thr) and homoserine (Hse) residues are positioned to trap imine intermediate as five-membered or six-membered ring N,O-acetal (oxazolidine) respectively, cysteine (Cys) side chain thiol group as a five-membered ring N,S-acetal (thiazolidine) and Asparagine (Asn) as a six-membered ring N,N-acetal (tetrahydropyrimidone) (**Figure 26**). The coupling reaction can be further diversified via other amino acids such as aspartic acid (Asp), histidine (His), etc.

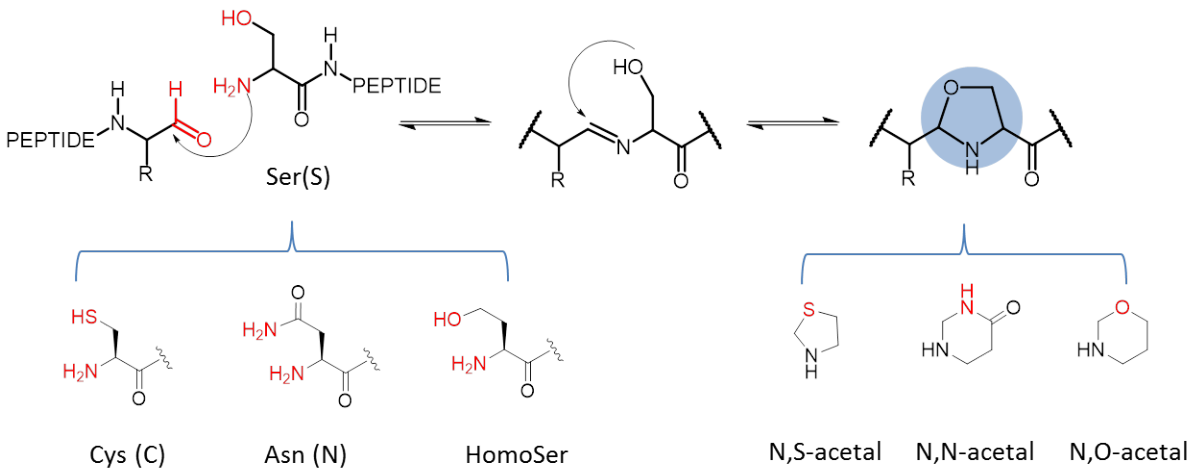


Figure 26 Structural illustration of reversible acetal linkages formed by trapping of imine intermediates via amino acid side chains.

To explore the generality of these linkages and examine if these reactions are robust to construct dynamic network, I use mono-functionalized substrates of Ser, Cys, Asn residues (providing amine and additional traps for imine intermediates) and phenylalanine (Phe) residue with modified aldehyde group (-CHO), *N*-Boc-Phe-CHO, to study N,O-acetal, N,S-acetal and N,N-acetal, respectively. Although any amino acid's C-terminal can be modified with aldehyde functional group theoretically, the choice of Phe is due to the great potential of the phenyl ring for promoting higher ordered structures in peptides assemblies [27-31].

2.1.3 Understanding Stereochemistry in Acetal Chiral Centers

These acetal condensation reactions create new chiral centers at the acetal carbons with two possible isomers: (S,S) *trans* and (R,S) *cis* isomers (**Figure 27**).

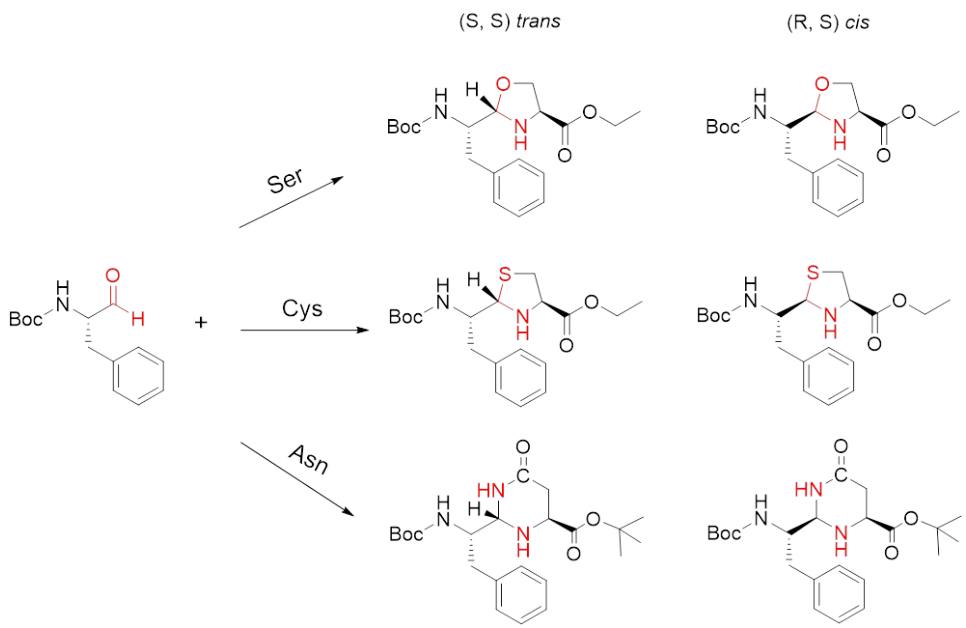


Figure 27 (S,S) trans isomer and (R,S) cis isomer of N,O; N,S; N,N-acetal products generated from N-Boc-Phe-CHO with L-Ser ethyl ester, L-Cys ethyl ester and L-Asn tert-butyl ester, respectively.

Previous researches [32-34] have shown that acid-catalyzed acetalization of serine, threonine and cysteine esters gives rise to *ca*. 1:1 mixtures of *cis*- and *trans*- substituted heterocycles in various solvents, e.g. acetone, acetic acid, DMSO and pyridine, etc. However, when acetals are acylated, pure *cis*- diastereomer result (**Figure 28**). The difference between the heterocycles with and without acylation on the ring is due to the well-known allylic 1,3- strain exerted by an amide group [35] [36]. The energy required to rotate an amide bond is 15-20 kcal/ mol, much larger than the conformational strain (e.g. switching substituents in the axial or equatorial positions).

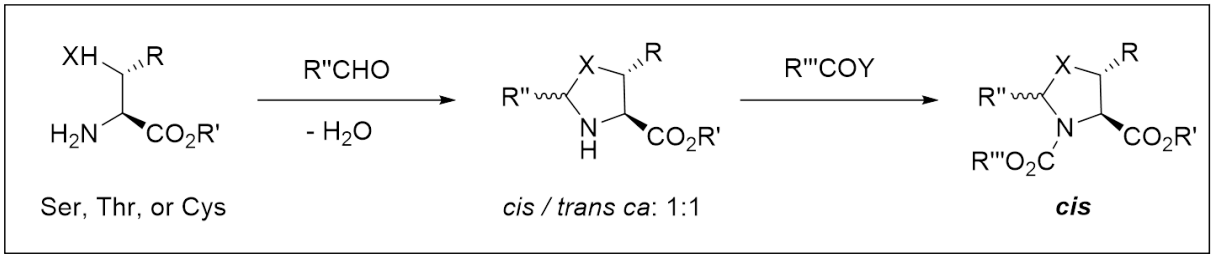


Figure 28 Chemical structure of acetals derived from amino acids Ser, Thr and Cys [36]. Without acylation on the acetal ring, Ser, Thr and Cys esters gives rise to *ca*: 1:1 mixtures of *cis*- and *trans*-substituted heterocycles.

2.1.4 Study Structural Information by Molecular Dynamics

From complex societies to single molecular networks, nature has demonstrated variety approaches to drive energy flow across an extraordinary range of length scales via self-organizing behaviors. Chemical systems capture these dynamics by interleaving the physical process of super-molecular assemblies with the rate of chemical syntheses. The discovery of analog evolution in protein mis-folding diseases allows combining the physical dynamics of prion-like folding with synthetic polymerization networks to create a system capable of accessing self-organizing behaviors.

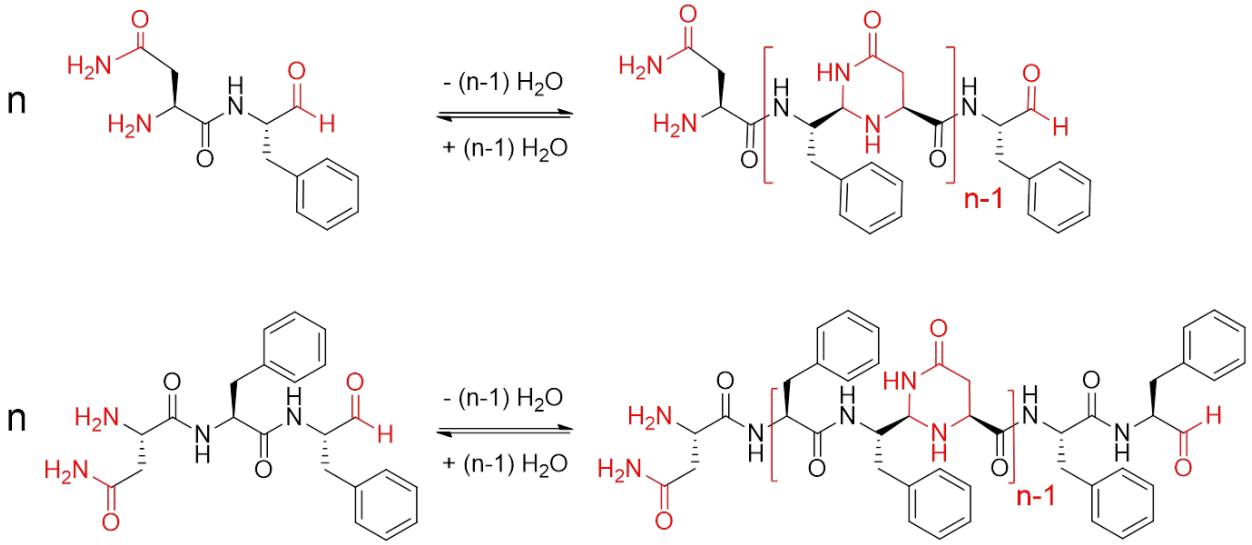


Figure 29 Chemical exchange of NF-CHO and NFF-CHO networks with representations for possible forms, linked via N,N-acetal.

When NF and NFF DCN are compared side by side, despite of the same chemical linkage, at 8 mM monomer, the behavior is different for both networks. As shown in **Figure 30** for the NF-CHO and NFF-CHO network in 40% acetonitrile, during the early stage, the composition of both networks differs by emergence of the cyclic dimer in NF-CHO network. Yet the entire network ultimately transitions almost exclusively to the linear trimer for both networks. During the course of network maturation there is a steady growth of particles that appear at early stage and are eventually consumed by the final fibers. However, for NFF-CHO network, the particles appear as early as 6 hours while in NF-CHO network, they don't appear until at 48 hours. The particles follow an Ostwald-like thermodynamic growth coincident with the chemical transitions to the paracrystalline fiber phase. Significant evidence now supports these particles as on pathway for cross-β assembly, and indeed seeds prepared from the mature fibers bypass the phase transitions and catalyze assembly within hours.

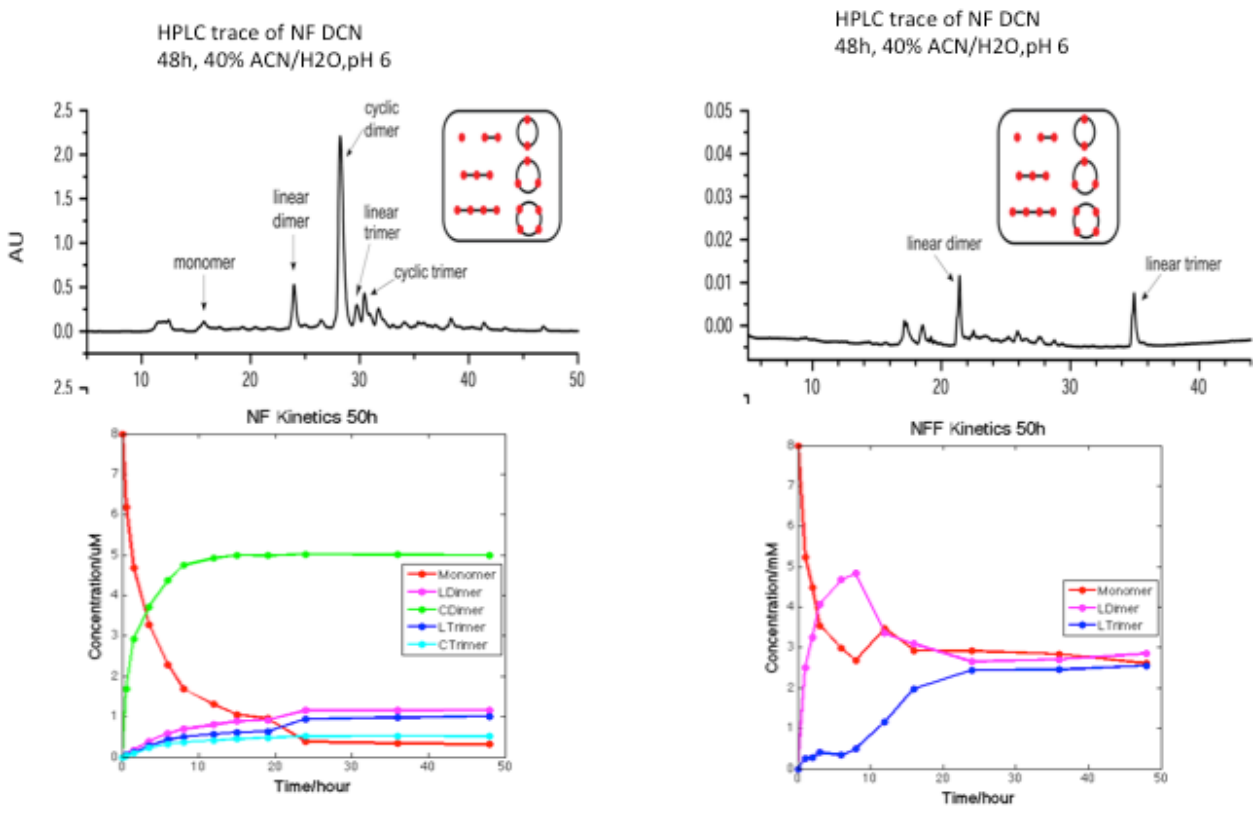


Figure 30 NF-CHO (left) and NFF-CHO (right) networks: HPLC trace, composition analyses and kinetics.

To better understand how similar chemical structures of NF-CHO and NFF-CHO monomers evolve distinctly in those networks, I have studied the structural formation by molecular dynamics (MD) simulations and investigated these super-molecular assemblies by Isotope-edited Fourier transform infrared spectroscopy (IE-FTIR) and solid-state NMR.

2.2 Results and Discussion

2.2.1 Understanding Conformational Evolution of Dynamic Chemical Network by Molecular (MD) Simulations

As a complementary approach to experimental characterizations, MD simulations are

powerful for revealing both inner-molecular and inter-molecular interactions. Preliminary MD simulations with explicit solvent were performed to evaluate stability of H-NFaNF-CHO and H-NFFaNFF-CHO linear dimers. Linear dimers were selected as basic studying units because the composition of the NFF network is dominated by linear dimer at early stage while the NF network is dominated by the cyclic dimer.

As shown in **Figure 31**, the NF/NFF linear dimers were constrained in a 10x10x10 A box at room temperature and energy-minimized under the OPLS force field before simulation. Conformation snapshots were taken every 4.8 ps for a total of 24ns. Our results suggest that both NF linear dimer and NFF linear dimer are structurally restricted in explicit solvent, and that they tend to be tightly packed at a U shape for higher order assembly.

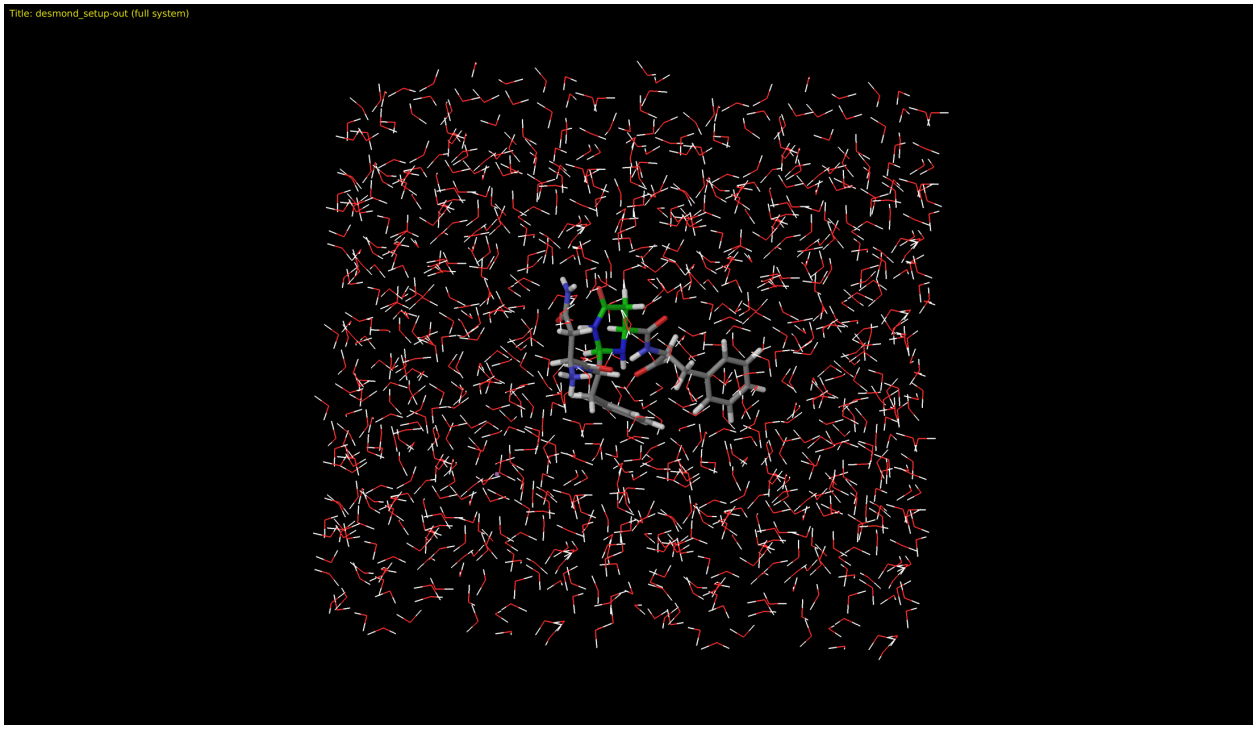


Figure 31 Initial simulation set-up.

Based on the preliminary MD simulation and experimental investigations, we then performed MD simulations by conformational searches for lowest-energy conformation. Subsequent MD simulations were conducted with lowest-energy conformation under the OPLS force field in explicit water constrained by a 20x20x20 box, which gives a calculated concentration comparable to experimental set-up. Conformation snapshot was taken every 4.8 ps for a total of 60ns. Simulation results do not position the heterocycle optimally for H-bonding between adjacent peptide stands. Our results suggest that the assemblies do not appear to reflect the degree of thermodynamic stability usually observed for peptide amyloids (**Figure 32**).

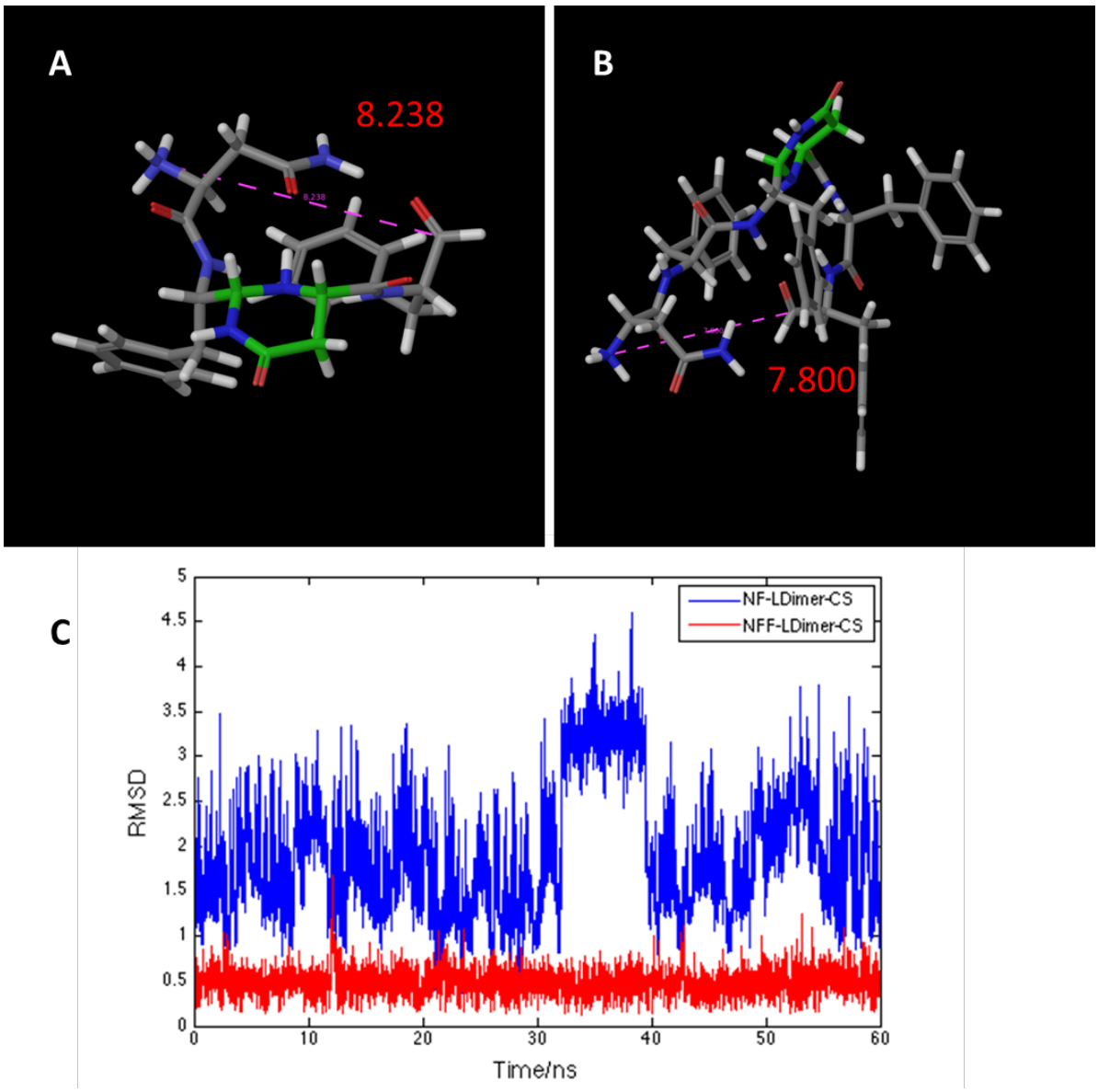


Figure 32 Examples of molecular dynamics derived structures of linear dimers of (A) NF-CHO and (B) NFF-CHO. (C) Root-mean-square deviation (RMSD) of the chain ends of the linear dimer in NF-CHO (blue) and NFF-CHO (red). Both linear dimers H-NFpyNF-CHO and H-NFFpyNFF CHO display "horseshoe" structures with similar distances between the two reactive ends, indicated by purple dashed lines at 8.238 Å and 7.800 Å, respectively. Although the distances are similar, the linear dimer H-NFpyNF-CHO displays greater flexibility as the H-NFFpyNFF-CHO structure packs a more tightly.

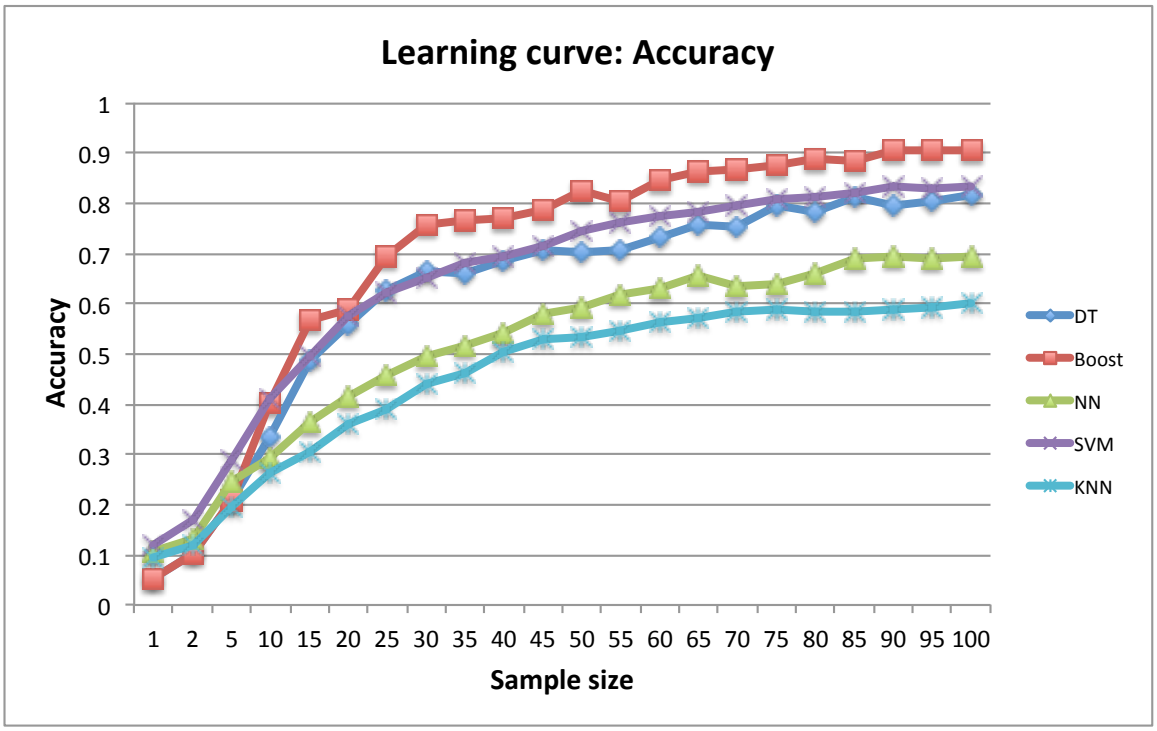
2.2.2 Analyze DCN Composition by Machine Learning Algorithms

While the DCN produces complicated systems, challenging to analyze individual components, the ability to identify certain chemicals from the library makes it critical to study the evolvement of the network. While the system is way to complicated for traditional analytical methods such as HPLC, NMR and MS, computational methods, especially machine learning techniques have made it possible to predict potential components in the library, and thus decode this black box of high dimension DCN. Since we already have a library of mass spectrum of known substrates, I was able to utilize segmented MS information to predict major components of the DCN library.

The processed MS Data has 1300 attributes, which are highly likely to make some algorithms very slow (Neural Networks(NN), and Supported Vector Machine(SVM), for example). In addition, most of the attributes contain spectrum values that has no physical meanings, and the variation across instances are very small. Therefore, I appied principal component analysis (PCA) to extract the most important information. Specifically, I keep the principle components which can jointly explain 90% of the covariance. After the dimension reduction, 197 attributes are acquired for my analysis.

Decision Tree (DT), Boosting, and SVM all performed quite well with datasets that has large amount of attributes. NN was not able to handle such feature, as expected. However, I was surprised to find that KNN didn't perform in this case. It's probably because that there's too little instance for it to learn, and therefore the more neighbors, the more confused the learner become. In summary, when handling datasets with large amounts of

attributes, it is important to properly reduce the amount of attributes. In addition, more instances should be obtained if possible. SVM can handle both cases quite well yet the time cost is significant. The following figure shows a prediction results for a simple NF DCN. It was able to achieve 90% percent accuracy with boosted decision trees (**Figure 33**).



	DT	Boosting	NN	SVM	KNN
Mass_Accuracy	0.818763	0.905429	0.695271	0.833625	0.600701

Figure 33 Supervised learning results of NF DCN based on MS database.

2.2.3 Characterize the Supramolecular Assemblies

2.2.3.1 Preliminary exploration of secondary structure by Isotope-edited FTIR

Isotope-edited FTIR analyses with model peptide systems provide sufficient time

resolution to follow cross-β peptide assembly pathways, and reflect the sum of strand orientations across dynamically exchanging peptide conformational populations. As shown in **Figure 34,** progressive phase transitions of specific paracrystalline morphologies by using IE-IR spectra of Aβ (16-22) congener peptide with antiparallel in- and out-of-register and parallel homogeneous amyloid assemblies in order to evaluate time-dependent changes of the Aβ (16-22) E22Q intermediates.

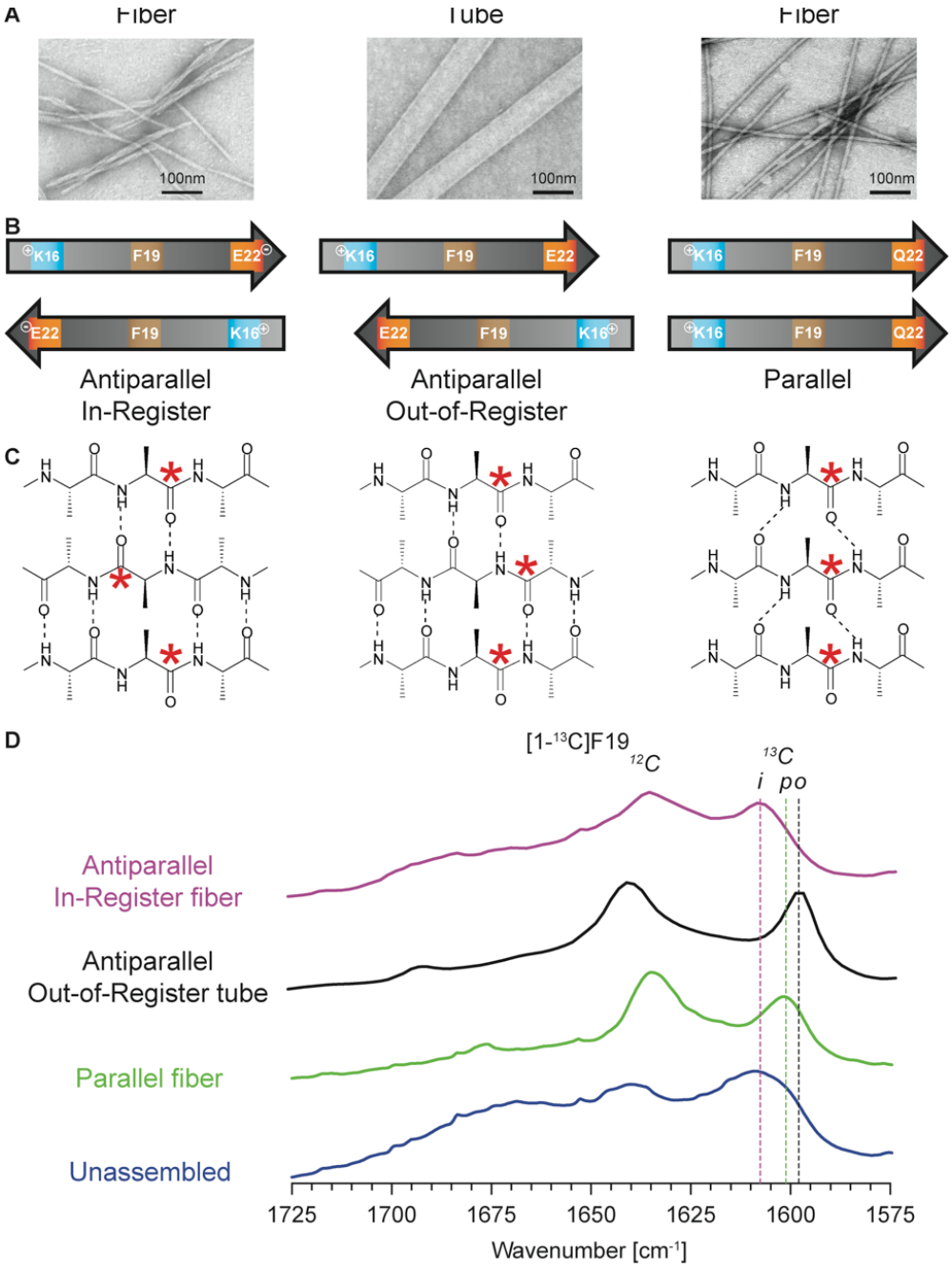


Figure 34 IR spectra of [1-^{13}C]^{19}F Aβ(16-22) have a ^{12}C/^{13}C splitting of 43 cm-1 for antiparallel out-of-register β-sheets, while antiparallel in-register β-sheets have a ^{12}C/^{13}C splitting of 31 cm-1 with a ^{12}C/^{13}C band intensity for both antiparallel assemblies of 0.91. In contrast, [1-^{13}C]^{19}F Aβ(16-22)E22Q assembles into fibers with parallel β-strand orientation and have a ^{12}C/^{13}C splitting of 33 cm-1 and a ^{12}C/^{13}C amide I band intensity ratio of 0.77. For comparison, the IR spectrum of unassembled [1-^{13}C]^{19}F Aβ(16-22) peptide was replicated by dissolving the peptide in HFIP to inhibit peptide aggregation and assembly, and the broad ^{12}C and ^{13}C bands are consistent with unordered peptide conformations.

This approach provides valuable insights into the transition phase of the dynamic chemical networks, and resolves the conformational phylogeny, of the initial particle phase. The factors that effect the conformational mutations that allow transitions between different amyloid polymorphs, and the environmental context that is so important to the evolution of these forms.

NF-CHO and NFF-CHO networks initially form particles, and then mature to fibers. When seeded with E22L nanotubes, the fibers transit to nanotubes. These transition signatures were followed by IE-IR to assign morphology and registry together with other characterization methods including solid-state NMR. Since the network consists of various specious, it's important to decompose the spectrum in order to assign each conformational population. Initial measurement of FTIR on NFF networks contains C^{13} enriched NFF monomers and NFF peptides are shown in **Figure 35**.

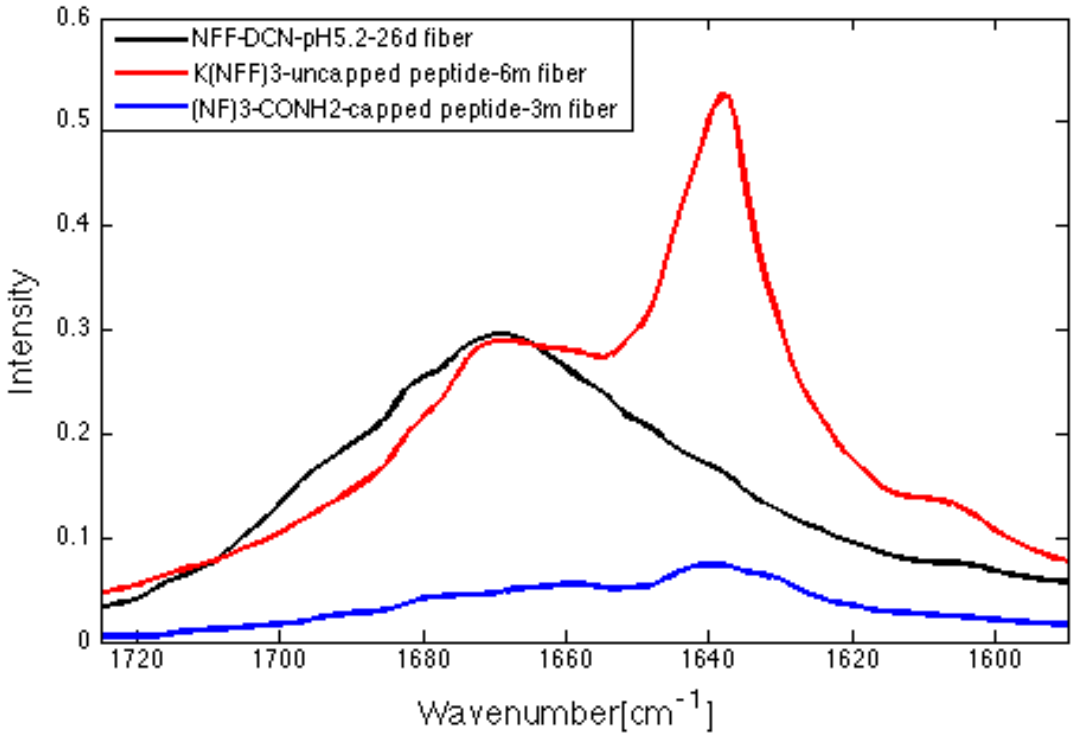


Figure 35 FTIR spectrum of assembled fibers. black: NFF network assembly; red: (NFF)3 peptide assembly; blue: (NF)3 peptide assembly.

2.2.3.2 Morphology determination by solid-state NMR

IE-FTIR provides insights into the secondary structure, yet it does not provide precise morphologies. Solid-state NMR analyses allows for stereochemical assignments via incorporation of ^{13}C at the aldehyde carbon. Given the repeating nature of the 9-mer H-NFFnFFnFF-CHO assembly, we predict that both parallel and anti-parallel β-strand arrangements are possible, and our NMR methods will allow the assignments to be made by precise ^{13}C-^{15}N and ^{13}C-^{13}C distance measurements between backbone carbon and nitrogens with $^{13}C(^{15}N)$ REDOR73,74 and ^{13}C DQF-DRAWS solid-state NMR experiments.

Incorporating [1-^{13}C]Phe and [^{15}N]Phe into the NFF-CHO monomer breaks the structural

symmetry and thus allows us to access the distance of ^{13}C-^{15}N distances from measured dipolar couplings by solid state NMR to precisely determine peptide registry. Our initial molecular models (**Figure 36**) suggest that for NFF-CHO trimers (9mer peptides) with a parallel β-strand arrangement will have an intermolecular [1-^{13}C] F2 distance of 4.8 Å between adjacent, H-bonded peptides, whereas the anti-parallel arrangement will have the carbonyl of F2 directly H-bonded to the amide nitrogen of F8, both of which are generated from the same NFF-CHO monomer.

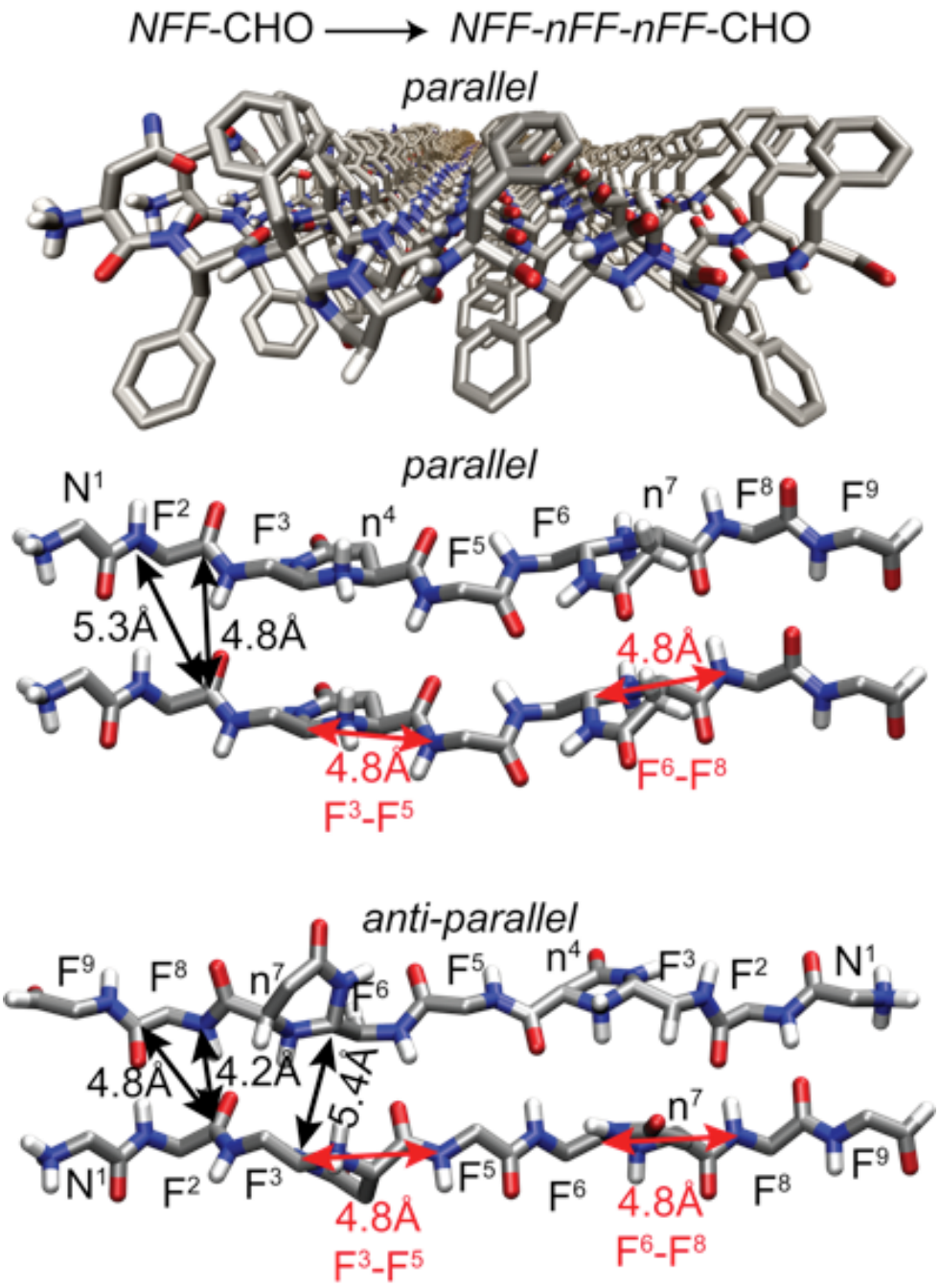


Figure 36 Molecular models of NFF-CHO as 9-mer oligomers arranged as parallel (top) and anti-parallel (bottom) β-strands. Models are a result of short (100-200ps) molecular dynamics simulations with explicit solvent and 64 peptides arranged into H-bonded β-sheets with 2-4 sheets stacked on top of each. Peptides from the non-solvent exposed regions were superimposed and the average structure was used as starting points for new simulations.

Mix acetyl capped CH3CO–N[1-^{13}C] FF-CHO with N[^{15}N] FF-CHO monomers. This

resulted in two unsymmetrical oligomers: one that has ^{13}C enriched carbonyl only at F2 and ^{15}N enriched amide nitrogen at both positions F5 and F8 (CH3CO-N[1-^{13}C]FFn[^{15}N]FFn[^{15}N]FF-CHO); and the other only has the uncapped monomer with ^{15}N enriched Phe (N[^{15}N]FFn[^{15}N]FFn[^{15}N]FF-CHO) (**Figure 37**).

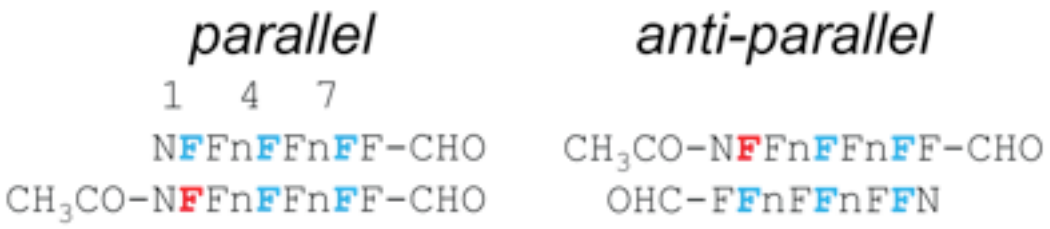


Figure 37 Solid-state NMR labeling scheme to determine NFFnFFnFF β-sheet registry. [1-^{13}C] residues are colored red and [^{15}N] residues blue. Parallel assemblies will have a ^{13}C-^{15}N distance ~1Å longer than anti-parallel assemblies (**Figure 36**).

2.3 Conclusion & Future Directions

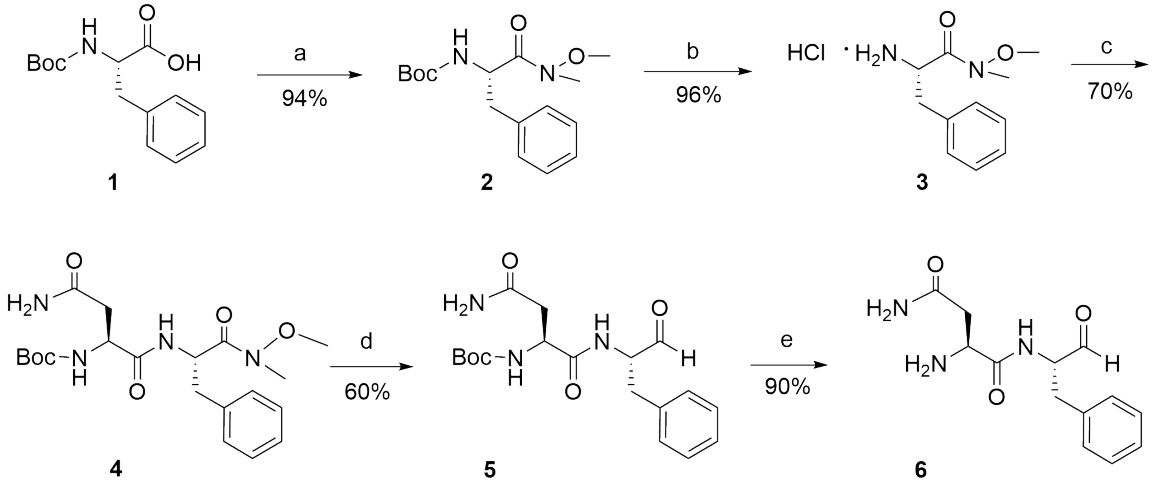
In this chapter, I have designed and developed dynamic chemical networks that showed two-phase polymerization. My work focused on characterizing the formation of network components and their higher dimension structures in both local phase and global phase via computational methods including molecular dynamics and machine learning algorithms. My future goals aim to further explore surface properties of the supermolecular assemblies, and extend the surface for functionality. While the growing ends of the network supramolecular assemblies template supramolecular growth, the assembly creates new surfaces that may also contribute to network dynamics. I am particularly interested in secondary nucleation events occurring on the fiber surfaces and have explored the scope of these possibilities initially with peptide assemblies.

2.4 Methods

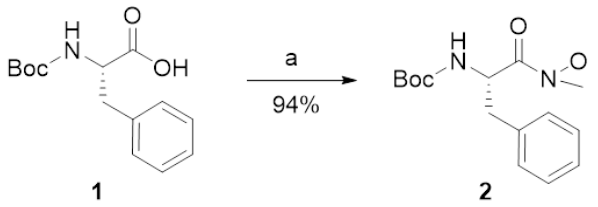
2.4.1 Materials

All commercially available chemicals are purchased from Sigma-Aldrich, AnaSpec and Nova Biochem. Anhydrous solvents are either dried over molecular sieve (4 Å) that had been pre-treated overnight at 300 ºC or purchased from EMD or Acros organics. HPLC grade acetonitrile and water are obtained from Sigma-Aldrich and/or Fisher Scientific. TLC plates are purchased from EMD (silica gel 60 F_{254}). Fmoc-amino acids, resins and solid phase peptide synthesizer reagents are purchased from AnaSpec. Distilled deionized water for sample preparation is obtained from EMD chemicals Inc.

2.4.2 Synthesis of NF-CHO



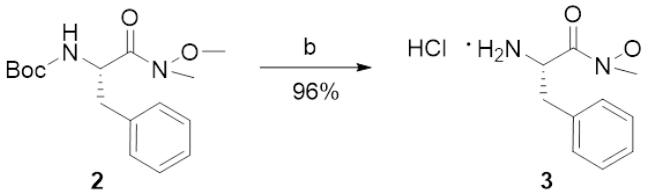
a. Preparation of the Boc-Phe-weinreb amide **2**



1 2

N-Boc-L-Phe (2.65 g, 10 mmol) is dissolved in anhydrous dichloromethane (DCM), and then 1,1-carbonyldiimidazole (1.78 g, 11 mmol) is added. The resultant mixture is stirred for an hour at room temperature. Subsequently, N,O-dimethylhydroxylamine hydrochloride (1.07 g, 11 mmol) is added and the reaction mixture (light yellow suspension) is stirred and allowed to proceed overnight. Then the solvent is removed by vacuum at 25 °C and the resulting residue is purified by extraction, which is dissolved in ethyl acetate (100ml) and wished successively with 2×100ml 1M HCl, 2×100ml saturated aqueous $NaHCO_3$ and 2×100ml brine. Finally, the organic layer is dried with anhydrous Na_2SO_4 and evaporation of the solvent yielded product **2**. Product is visualized by UV on TLC plate (hexane/ ethyl acetate (1/1) as solvent system, $R_f \sim 0.46$). ^{1}H NMR (600 MHz, Chloroform-d) δ 7.30 –7.14 (m, 5H, C_6H_5), 5.17 (d, J = 8.8 Hz, 1H, CH (α)), 4.94 (q, J = 7.1 Hz, 1H, NH), 3.65 (s, 3H, O-CH_3), 3.16 (s, 3H, N-CH_3), 3.05/2.87 (dd, 2H, CH_2(β)), 1.38 (s, 9H, $C(CH_3)_3$). MS (ESI) m/z $(M+H)^+$:309.1803 Calculated $(M+H)^+$: 308.1808 ($C_{16}H_{25}N_2O_4$)

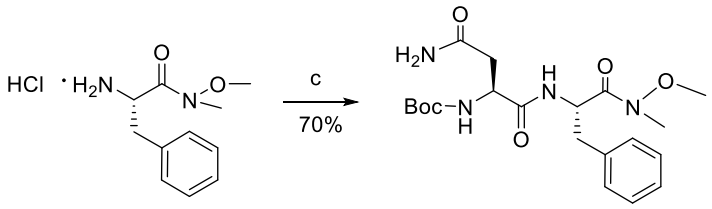
b. Removal of the t-butoxycarbonyl protecting group of **2**



2 3

The white solid **2** (2.90 g, 9.4 mmol) prepared above is dissolved in 10 ml ~ 20 ml 4M HCl in dioxane (purchased from Aldrich), and then the resultant mixture is stirred at room temperature for 0.5hr to 1hr at room temperature. Solvent is removed with vacuum at room temperature, leaving the product (**3**) of the hydrochloride salt as thick waxy oil. ^{1}H NMR (600 MHz, Chloroform-*d*) δ 8.83 – 8.29 (s, 3H, NH_3^+), 7.47 – 7.05 (m, 5H, C_6H_5), 4.74 (q, *J* = 5.9 Hz, 1H, CH (α)), 3.62 (s, 3H, O-CH_3), 3.40/3.28 (m, 2H, CH_2(β)), 3.10 – 3.07 (s, 3H, N-CH_3). MS (ESI) m/z $(M+H)^+$: 209.1284 Calculated $(M+H)^+$: 209.1285 $C_{11}H_{17}N_2O_2$

c. Coupling Boc-Asn to **3**

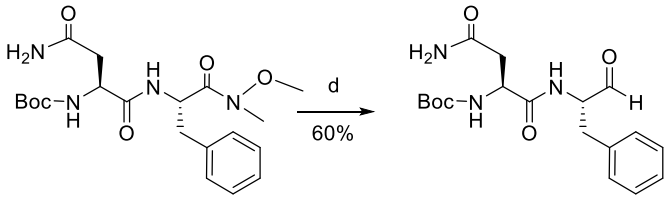


3 **4**

The waxy oil **3** (2.20 g, 9 mmol) is dissolved in anhydrous DCM (50-80 ml) before Boc-L-Asn (2.30 g, 9.9 mmol) and a two-fold excess of trimethylamine (2.5 ml, 18 mmol) are added. The reaction mixture is cooled to 0°C in an ice bath prior to the addition of 1-Ethyl-3(3-dimethylaminopropyl) carbodiimide (EDC) (1.90 g, 9.9 mmol), and the reaction mixture is stirred at 0°C under N_2 for 2hrs and then stirred at room temperature for overnight. The solvent is removed in vacuo at room temperature and the resulting residue dissolved in ~100ml ethyl acetate and wished successively with 2×100ml of 1M HCl, 2× 100ml of 4% $NaHCO_3$ and 2×100ml of brine. The excess of reagents of Boc-L-Asn and EDC are wished away by aqueous layer. The organic layer is dried with anhydrous sodium sulfate, decanted, and the solvent removed in vacuo to give white solid 4. UV visualization

on florescent TLC plates (ethyl acetate eluent, $R_f \sim 0.19$) is possible. ^{1}H NMR (600 MHz, Chloroform-d) δ 7.42 (d, J = 8.3 Hz, 1H), 7.32 – 7.16 (m, 5H), 5.96 (d, J = 9.4 Hz, 2H, NH_2), 5.49 – 5.45 (m, 1H, NH), 5.16 (q, J = 7.2 Hz, 1H, CH(α)), 4.47 (d, J = 8.5 Hz, 1H, CH(α)), 3.65 (s, 3H, O-CH_3), 3.15 (s, 3H, N-CH_3), 3.07 (dd, J = 13.6, 6.0 Hz, 1H), 2.97 – 2.83 (m, 2H, CH_2(β)), 2.55 (dd, J = 15.5, 6.1 Hz, 1H), 1.44 (s, 9H, $C(CH_3)_3$). MS (ESI) m/z $(M+Na)^+$: 445.2054, Calculated $(M+Na)^+$: 445.2057 ($C_{11}H_{16}N_2O_2Na$)

d. Reduction of the Boc-Asn-Phe-weinreb amide **4**

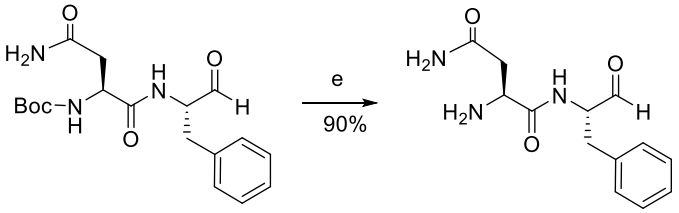


4 **5**

White solid **4** (2.66 g, 6.3 mmol) is dissolved in 60~80 ml anhydrous THF under N_2 and cooled to -78 °C in a dry ice and acetone bath before 1.5 equivalent of lithium aluminum anhydride (1.0 M) in THF (9.5 ml, 9.5 mmol) is added drop-wise via syringe. The temperature is raised to 0 °C with an ice bath for 30min and then cooled again to -78°C before the reaction is quenched with 10 ml of 1M aqueous solution of potassium bisulfate ($KHSO_4$). The mixture is allowed to warm to room temperature and the aluminate residue extracted several times with ethyl acetate. These combined extracts are wished with 2×50ml $NaHCO_3$ and 2×50ml brine before the organic phase is dried with anhydrous Na_2SO_4, decanted, and taken to dryness in vacuo at <20 °C to provide waxy solid 5. ^{1}H NMR (600 MHz, THF-d8) δ 9.46 (s, 1H, CHO), 7.94 (d, J = 7.2 Hz, 1H, NH), 7.28 – 7.09 (m, 5H, C_6H_5), 6.83/ 6.31 (s, s, 2H, NH_2), 6.45 (m, 1H, NH), 4.37 (q, J = 6.7 Hz, 1H, CH(α)),

4.20 (q, J = 6.8 Hz, 1H, CH(α)), 3.14/2.64 (dd, dd, 2H, CH_2(β)), 2.95 (ddd, 2H, CH_2(β)), 1.39 (s, 9H, $C(CH_3)_3$). MS (ESI) m/z $(M+Na)^+$:386.1698, Calculated $(M+Na)^+$: 386.1670 ($C_{18}H_{25}N_3O_5Na$)

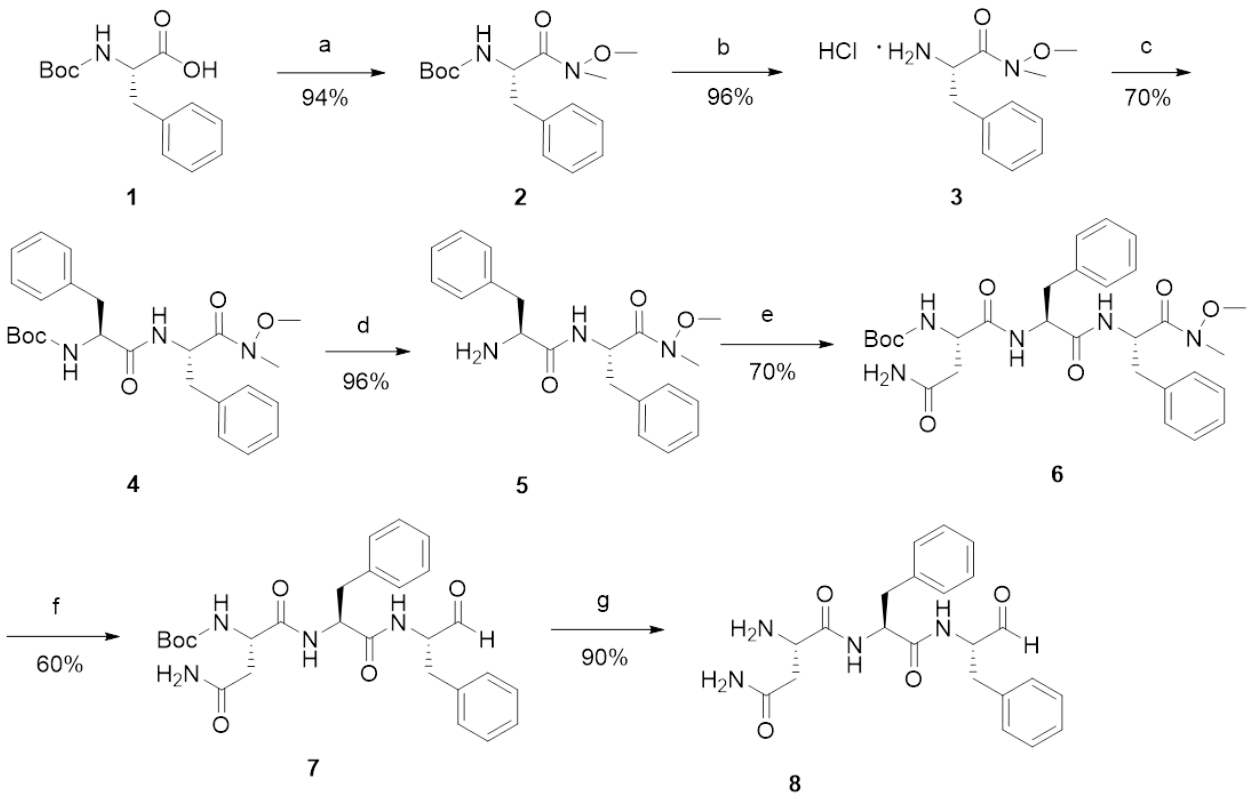
e. Deprotection of Boc group of **5** to give rise to the free monomer **6**



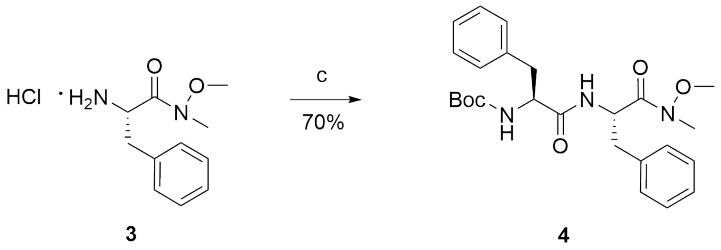
5 **6**

The waxy solid **5** (1.37 g, 3.78 mmol) is dissolved in 10 ml trifluoroacetic acid under N_2, stirred at room temperature for 1hr, and the solvent removed in vacuo to give thick waxy solid **6**. MS (ESI) m/z $(M+H)^+$: 264.1344 Calculated $(M+H)^+$: 264.1343 ($C_{13}H_{18}N_3O_3$)

2.4.3 Synthesis of NFF-CHO



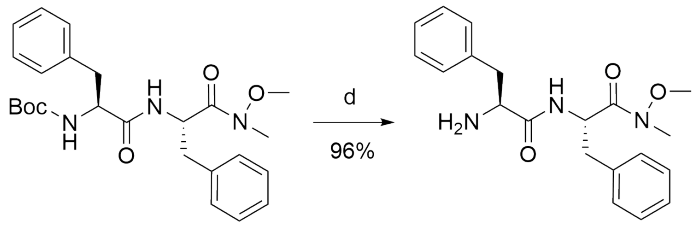
f. Coupling Boc-Phe residue to **3**



The waxy oil **3** (2.20 g, 9 mmol) is dissolved in anhydrous DCM (50-80 ml) before adding Boc-L-Phe (2.63 g, 9.9 mmol) and trimethylamine (2.5 ml, 18 mmol). The reaction mixture is cooled to 0°C in an ice bath before adding 1-Ethyl-3(3-dimethylaminopropyl) carbodiimide (EDC) (1.90 g, 9.9 mmol) stirred for 2hrs at 0°C under N_2 followed by overnight at room temperature. The solvent is evaporated *in vacuo* at room temperature,

and the resulting residue is dissolved in ~100ml ethyl acetate and wished successively with 2×100ml of 1M HCl, 2× 100ml of 4% $NaHCO_3$ and 2×100ml of brine. The excess of reagents of Boc-L-Asn and EDC are wished away by aqueous layer. The organic layer is dried with anhydrous sodium sulfate, decanted, and the solvent removed in vacuo to give white solid **4**. ^{1}H NMR (600 MHz, Chloroform-*d*) δ 7.40 – 7.07 (m, 10H, 2 C_6H_5), 6.55 (d, *J* = 8.3 Hz, 1H, NH), 5.19 (s, 1H, NH), 4.94 (m, 1H, CH (α)), 4.36 (m, 1H, CH (α)), 3.61 (s, 3H, O-CH_3), 3.13 (s, 3H, N-CH_3), 3.08 – 2.78 (m, 4H, 2 CH_2(β)), 1.39 (s, 9H, C(CH_3)$_3$). MS (ESI) m/z $(M+H)^+$: 456.2497 Calculated $(M+H)^+$: 455.2496 ($C_{25}H_{34}N_3O_5$)

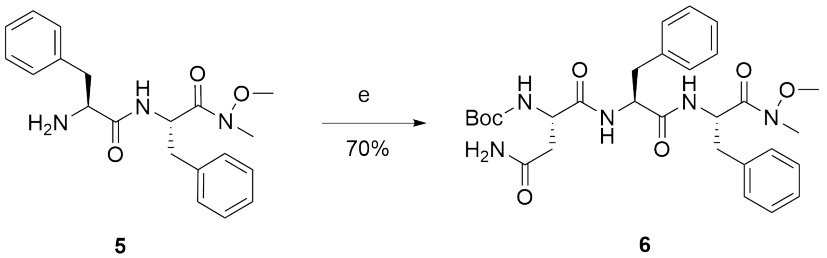
g. Removal of the t-butoxycarbonyl protecting group of **4**



4 **5**

The white solid **4** (2.87 g, 6.3 mmol) is dissolved in 10~20 ml 4M HCl in dioxane (Aldrich), and the resultant mixture stirred at room temperature for 0.5hr to 1hr. The solvent is removed directly *in vacuo* to give **5** hydrochloride as a thick waxy solid. ^{1}H NMR (600 MHz, Chloroform-*d*) δ 8.64 (s, 2H, NH_2), 7.84 (s, 1H, NH), 7.40 – 7.12 (m, 10H, 2 C_6H_5), 5.21 – 4.95 (m, 1H, CH (α)), 4.35 (s, 1H, CH (α)), 3.50 (s, 3H, O-CH_3), 3.41 (m, 2H, CH_2(β)), 3.19 (m, 2H, CH_2(β)), 3.03 (s, 3H, N-CH_3). MS (ESI) m/z $(M+H)^+$: 356.1976 Calculated $(M+H)^+$: 356.1969 ($C_{20}H_{26}N_3O_3$)

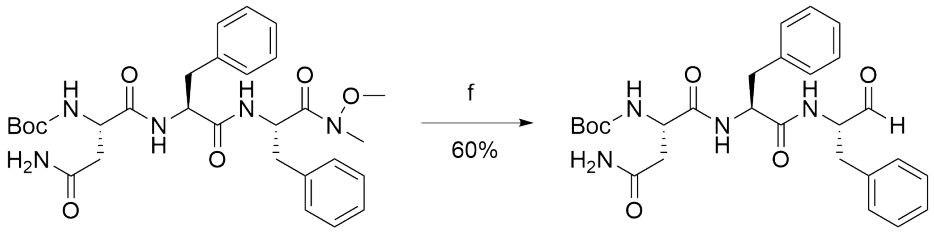
h. Coupling of Asn to **5**



5 6

The thick waxy solid **5** (2.13 g, 6.0 mmol) is dissolved in anhydrous DCM (50-80 ml) before Boc-L-Asn (1.53 g, 6.6 mmol) and trimethylamine (1.7 ml, 12 mmol) are added. The reaction mixture is cooled to 0°C in an ice bath before EDC (1.27 g, 6.6 mmol) is added and the mixture stirred for 2hrs at 0°C and overnight at room temperature. When the solvent is removed *in vacuo*, the resulting residue can be dissolved in ~100ml ethyl acetate and wished successively with 2×100ml of 1M HCl, 2× 100ml of 4% $NaHCO_3$ and 2×100ml of brine. The organic layer is dried with anhydrous sodium sulfate, decanted, and the solvent removed in vacuo to give white solid 6. ^{1}H NMR (600 MHz, Chloroform-*d*) δ 7.30 – 7.13 (m, 10H, 2 C_6H_5), 7.11 – 6.98 (m, 2H, NH_2), 6.23 (s, 1H, NH), 5.92 – 5.87 (m, 1H, NH), 5.59 (s, 1H, NH), 5.19 (q, *J* = 7.2 Hz, 1H, CH (α)), 4.70 (p, *J* = 8.6, 7.2 Hz, 1H, CH (α)), 4.47 (q, *J* = 6.6 Hz, 1H, CH (α)), 3.64 (s, 3H, O-CH_3), 3.11 (s, 3H, N-CH_3), 3.09 – 2.88 (m, 4H, 2 CH_2(β)), 2.81 (dd, *J* = 16.5, 4.0 Hz, 1H), 2.51 (dd, *J* = 15.3, 7.6 Hz, 1H), 1.42 (s, 9H, $C(CH_3)_3$). MS (ESI) m/z $(M+H)^+$: 570.2932, Calculated $(M+H)^+$: 570.2922 ($C_{29}H_{40}N_5O_7$)

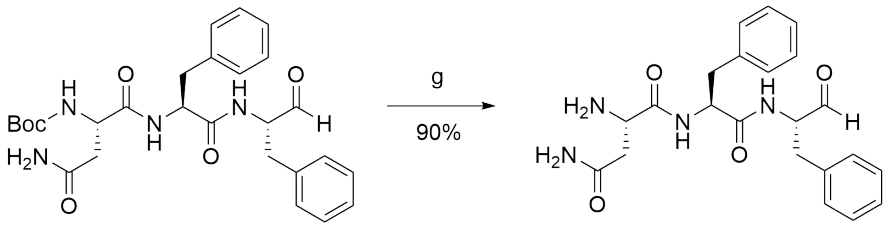
i. Reduction of the Boc-Asn-Phe-Phe-weinreb amide



6 **7**

White solid **6** (4.2 mmol) is dissolved in 60 ml anhydrous THF under N_2 and cooled to -78°C in a dry ice / acetone bath before LAH (1.0 M) in THF (6.3 ml, 6.3 mmol) is added dropwise via syringe. When addition is complete the acetone bath is replaced with a 0 °C ice bath for 30min and the mixture is cooled again to -78°C and quenched quickly with 10 ml of 1M aqueous potassium bisulfate ($KHSO_4$). This mixture is allowed to warm to room temperature, extracted several times with ethyl acetate, and the combined extracts wished with 2×50ml $NaHCO_3$, 2×50ml brine, dried with anhydrous Na_2SO_4. The solvents are evacuated by vacuum under 20 °C, providing products 7 as waxy solid. ^{1}H NMR (600 MHz, DMSO-d_6) δ 9.32 (s, 1H, CHO), 8.55 (d, J = 7.5 Hz, 1H, NH), 7.95 (d, J = 8.0 Hz, 1H, NH), 7.32 (s, 1H, NH), 7.30 – 7.13 (m, 10H, 2 C_6H_5), 6.93 – 6.87 (m, 2H, NH_2), 4.46 (m, 1H, CH (α)), 4.23 (2H, 2 CH (α)), 3.10 (m, 1H, CH_2(β)), 2.94 (dd, J = 13.8, 5.1 Hz, 1H, CH_2(β)), 2.84 – 2.69 (m, 2H, CH_2(β)), 2.39 (m, 1H, CH_2(β)), 2.31 (dd, J = 15.1, 7.9 Hz, 1H, CH_2(β)), 1.35 (s, 9H, $C(CH_3)_3$) MS (ESI) m/z $(M+Na)^+$:533.2376 Calculated $(M+Na)^+$:533.2371 ($C_{27}H_{34}N_4O_6$)

j. Deprotection of Boc group to give rise to the free monomer NFF-CHO **8**



7 8

The waxy solid **7** (1.29 g, 2.52 mmol) is dissolved in 10 ml trifluoroacetic acid (bubbled with N_2 gas for at least 15 min before usage) under N_2, stirred at room temperature for 0.5hr to 1hr, and the solvent removed in vacuo to give **8**·hydrochloride as a thick waxy solid. MS (ESI) m/z $(M+H)^+$:411.2031 Calculated $(M+H)^+$: 411.2027 ($C_{22}H_{27}N_4O_4$)

2.4.4 NMR Analysis

^{1}H NMR data are recorded on an INOVA 600 or an INOVA 400 NMR spectrometer (equipped with Bore Oxford super conducting magnet). The solvents Chloroform-d and THF-d8 are purchased from Cambridge Isotope Laboratories, Inc. The choice of the solvent is determined by the solubility of the compounds.

2.4.5 Dynamic Peptide Network Preparation

After final deprotection of the precursor of the network, Boc-NF-CHO by trifluoroacetic acid, resulting network building blocks of NH_2-NF-CHO are dried under vacuum at room temperature. Dried NH_2-NF-CHO are stored with Argon protection at -20°C. To prepare the dynamic network, NH_2-NF-CHO are dissolved to a concentration of 8mM in water/acetonitrile (3/2, v/v) at ambient temperature under N_2 protection. All the solvents are flushed with N_2 gas for at least 15 min before usage. Dissolution is assisted by ~2 minutes of continuous vortexing, followed by ~15 minutes of bath sonication until solution became clear. The pH of the solution is adjusted by titrating aliquots of 10mM NaOH and

the final pH value is determined by pH meter (Fisher Scientific Accumet Basic AB15 pH meter).

2.4.6 HPLC and LC-MS Analyses

HPLC analyses are performed on Waters Delta 600 equipped with a photodiode array UV/Vis detector at room temperature using a reversed-phase HPLC column (Kromasil 100-5C18, 4.6 × 250mm). Solvent A: water (0.1 vol % trifluoroacetic acid). Solvent B: acetonitrile (0.1 vol % trifluoroacetic acid). UV absorbance is monitored at 258nm (for Phenyl ring side chain absorption) and 222nm (for amide bond absorption). Flow rate is 1.0 mL/min. Gradient is from 10% acetonitrile to 90% acetonitrile, 2% acetonitrile/ min. LC-MS analyses are performed on Waters Synapt G2 MS/Acquity UPLC system. Positive-ion mass spectra are obtained using electrospray ionization.

Time (mins)	Solvent A	Solvent B	Note
0	90%	10%	Gradient starts
40	10%	90%	Gradient ends
41	0%	100%	Cleaning starts
48	0%	100%	Cleaning ends
49	10%	90%	Re-equilibrium starts
56	10%	90%	Re-equilibrium ends

2.4.7 Transmission Electron Microscopy and Electron Diffraction

Aliquots (10 μl) of sample solutions are dropped on TEM grids (200 mesh copper grid covered with a thin carbon film, purchased from Electron Microscopy Sciences) for 3 minutes before excessive solution is blotted with filter paper. Uranyl acetate (10 μl of 5%

solution) is added for 3 minutes for negative staining. Extra fluid is blotted with filter paper. The grids are analyzed on a Hitachi H-7500 transmission electron microscope with a LaB6 emission filament at an accelerating voltage of 75 kV.

2.4.8 Microwave Assisted Solid-Phase Peptide Synthesis

Peptides H-NFNF-NH_2 and H-NFNFNF-NH_2 are synthesized on a Liberty CEM Microwave Automated Peptide Synthesizer utilizing FMOC-Rink Amide MBHA Resins purchased from AnaSpec. All FMOC protected amino acids are from Anaspec, and other chemicals from Sigma-Aldrich or Fisher Scientific. Each peptide synthesis is performed at 0.1 mmol using a 45 mL reaction vessel at a scale of 0.1mmol. FMOC-Rink Amide MBHA Resin is initially swollen using ~7 ml dimethylformamide (DMF) for 15 minutes. FMOC deprotection is achieved by addition of 20% piperdine 0.1M N-Hydroxybenzotriazole (HOBt) in DMF with microwave power set to maintain temperature between 45-55°C for 180 sec, followed by 3X flushing with DMF. Each coupling step is performed using 0.1M FMOC protected amino acid, and activated with 0.1 M 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluoro-phosphate (HBTU), and 0.2M N,N-Diisopropylethylamine (DIEA) in DMF. Coupling temperatures are maintained between 75-82°C by optimizing microwave power for 300 sec. After coupling, the resin is rinsed with three aliquots of DMF. Peptides are cleaved from the resin using trifluoroacetic acid/thioanisole/1,2- ethanedithiol/anisole (90: 5 : 3 : 2, v/v/v/v) at room temperature for 3 hrs. The cleaved peptide- TFA solution is filtered, and precipitated by dropwise addition of cold (-20°C) diethyl ether. Precipitated product is centrifuged at 3000 rpm for 15 min, and the pellet is subjected to 3 additional rounds of ishing with cold diethyl ether, followed by desiccating overnight. Dried peptides are dissolved in minimal volume of 40%

acetonitrile / 60% water and purified by RP-HPLC (Water Delta 600) using a C18-reverse phase column with an acetonitrile-water (0.1% TFA) gradient. The molecular weight of each peptide is verified by mass spectrometry.

Purified peptides are dissolved in water/acetonitrile (3/2, v/v) with 0.1% TFA for nanofiber assemblies. Dissolution is assisted by ~2 minutes of continuous vortexing, followed by ~15 minutes of sonication until solution became clear.

References

1. Stephenson, M.L. and P.C. Zamecnik, *Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide*. Proceedings of the National Academy of Sciences of the United States of America, 1978. **75**(1): p. 285-288.
2. Roush, W., *Antisense Aims for a Renaissance*. Science, 1997. **276**(5316): p. 1192.
3. Deleavey, Glen F. and Masad J. Damha, *Designing Chemically Modified Oligonucleotides for Targeted Gene Silencing*. Chemistry & Biology, 2012. **19**(8): p. 937-954.
4. Cheng, L., K.G. Abhilash, and R. Breslow, *Binding and biomimetic cleavage of the RNA poly(U) by synthetic polyimidazoles*. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(32): p. 12884-12887.
5. Goodwin, J.T. and D.G. Lynn, *Template-directed synthesis: use of a reversible reaction*. Journal of the American Chemical Society, 1992. **114**(23): p. 9197-9198.
6. Li, X., et al., *DNA-Catalyzed Polymerization*. Journal of the American Chemical Society, 2002. **124**(5): p. 746-747.
7. Nowotny, M., et al., *Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis*. Cell, 2005. **121**(7): p. 1005-16.
8. Vester, B., et al., *LNAzymes: Incorporation of LNA-Type Monomers into DNAzymes Markedly Increases RNA Cleavage*. Journal of the American Chemical Society, 2002. **124**(46): p. 13682-13683.

9. Vester, B. and J. Wengel, *LNA (Locked Nucleic Acid): High-Affinity Targeting of Complementary RNA and DNA*. Biochemistry, 2004. **43**(42): p. 13233-13241.

10. Ye, J., Y. Gat, and D.G. Lynn, *Catalyst for DNA Ligation: Towards a Two-Stage Replication Cycle*. Angewandte Chemie International Edition, 2000. **39**(20): p. 3641-3643.

11. Tadokoro, T. and S. Kanaya, *Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes*. FEBS Journal, 2009. **276**(6): p. 1482-1493.

12. Thomas, J.M., J.-K. Yoon, and D.M. Perrin, *Investigation of the Catalytic Mechanism of a Synthetic DNAzyme with Protein-like Functionality: An RNaseA Mimic?* Journal of the American Chemical Society, 2009. **131**(15): p. 5648-5658.

13. Li, X. and D.R. Liu, *DNA-Templated Organic Synthesis: Nature's Strategy for Controlling Chemical Reactivity Applied to Synthetic Molecules*. Angewandte Chemie International Edition, 2004. **43**(37): p. 4848-4870.

14. Luo, P., et al., *Analysis of the Structure and Stability of a Backbone-Modified Oligonucleotide: Implications for Avoiding Product Inhibition in Catalytic Template-Directed Synthesis*. Journal of the American Chemical Society, 1998. **120**(13): p. 3019-3031.

15. Murtola, M., M. Wenska, and R. Stromberg, *PNAzymes That Are Artificial RNA Restriction Enzymes*. Journal of the American Chemical Society, 2010. **2010**(132): p. 6.

16. Honaker, M.T., et al., *Ensemble Perspective for Catalytic Promiscuity: Calorimetric Aanlysis of The Active Site Conformational Landscape of a*

Detoxification Enzyme. Journal of Biological Chemistry, 2011. **286**(49): p. 42770-42776.

17. Weinstain, R., et al., *Fluorescent ligand for human progesterone receptor imaging in live cells*. Bioconjugate Chemistry, 2013. **24**(5): p. 766-71.
18. Niu, Y., et al., *Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos*. Cell, 2014. **156**(4): p. 836-43.
19. Wiedenheft, B., S.H. Sternberg, and J.A. Doudna, *RNA-guided genetic silencing systems in bacteria and archaea*. Nature, 2012. **482**(7385): p. 331-8.
20. Bao, C., et al., *Optical Fusion Assay Based on Membrane-Coated Spheres in a 2D Assembly*. Journal of the American Chemical Society, 2013. **135**(33): p. 12176-9.
21. Jung, Y.G., M.S. Kang, and J. Heo, *Clustering performance comparison using K-means and expectation maximization algorithms*. Biotechnology, Biotechnological Equipment, 2014. **28**(sup1): p. S44-S48.
22. Blum, M.G.B., et al., *A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation*. Statistical Science, 2013: p. 189-208.
23. Goodwin, J.T. and D.G. Lynn, *Template-Directed Synthesis: Use of a Reversible Reaction*. Journal of the American Chemical Society, 1992. **114**(23): p. 9197-9198.
24. Zhan, Z.J. and D.G. Lynn, *Chemical Amplification through Template-Directed Synthesis*. Journal of the American Chemical Society, 1997. **119**: p. 12420-12421.
25. Li, X., et al., *DNA-Catalyzed Polymerization*. Journal of the American Chemical Society, 2002. **124**(5): p. 746-747.

26. Li, X., et al., *Step-growth control in template-directed polymerization*. Heterocycles, 2011. **82**(2): p. 1477-1488.

27. Mehta, A.K., et al., *Facial Symmetry in Protein Self-Assembly*. Journal of the American Chemical Society, 2008. **130**: p. 9829-9835.

28. Frederix, P.W., et al., *Virtual Screening for Dipeptide Aggregation: Toward Predictive Tools for Peptide Self-Assembly*. Journal Physical Chemistry Letter, 2011. **2**(19): p. 2380-2384.

29. Scanlon, S. and A. Aggeli, *Self-assembling peptide nanotubes*. Nano Today, 2008. **3**(3-4): p. 22-30.

30. Reches, M. and E. Gazit, *Designed aromatic homo-dipeptides: formation of ordered nanostructures and potential nanotechnological applications*. Physical Biology, 2006. **3**: p. S10-S19.

31. Adler-Abramovich, L., et al., *Thermal and Chemical Stability of Diphenylalanine Peptide Nanotubes: Implications for Nanotechnological Applications*. Langmuir, 2006. **22**: p. 1313-1320.

32. Szilagyi, L. and Z. Gyorgydeak, *Comments on the putative stereoselectivity in cysteine-aldehyde reactions. Selective C(2) inversion and C(4) epimerization in thiazolidine-4-carboxylic acids*. J. Am. Chem. Soc., 1979. **101**(2): p. 427-432.

33. Seebach, D., et al., *Preparation of Oxazolidine-Containing Peptides: Unusual effects in Rh(III)-catalyzed acetalizations of aldehydes with urethane-protected serine and threonine esters and with dipeptides containing serine or threonine residues at the N-terminus*. Helvetica Chimica Acta, 1994. **77**(5): p. 1313-1330.

34. Seebach, D., et al., *Stereoselektive Alkylierung an C(α) von Serin, Glycerinsäure, Threonin und Weinsäure über heterocyclische Enolate mit exocyclischer Doppelbindung*. Helvetica Chimica Acta, 1987. **70**(4): p. 1194-1216.

35. Hoffmann, R.W., *Allylic 1,3-Strain as a Controlling Factor in Stereoselective Transformations*. Chemical Reviews, 1989. **89**: p. 1841-1860.

36. Seebach, D., Lamatsch, B., Amstutz, R., Beck, A. K., Dobler, M., Martin Egli,, et al., *Structure and reactivity of five- and six-ring N,N-, N,O-, and O,O-acetals: A lesson in allylic 1,3-strain*. Helvetica Chimica Acta, 1992. **75**: p. 913-934.