Intercalation Mediated Assembly and the First Informational Polymers of Life

Final Report

Nicholas V. Hud, PI
Period Covered: 5/5/2008 – 5/4/2012
Georgia Institute of Technology
Atlanta, GA, 30332-0400
Grant # NNX08AO14G

A. Objectives

We proposed to investigate how small molecules that intercalate nucleic acid bases can drive the assembly of nucleic acids, and how this action may have facilitated the abiotic formation of, and influenced the structure of, the first RNA-like polymers.

The specific objectives of this grant were:

1. To explore intercalation-mediated template-directed polymerization of nucleic acids with Watson–Crick base pairs.

2. To explore intercalation-mediated ligation, polymerization and *de novo* assembly of oligonucleotides through purine-purine base pairing.

3. To explore the relationship between backbone structure and facility of intercalation-mediated ligation and polymerization.

B. Studies and Results

Advances with intercalator-mediated polymerization.

A major objective of this project was to explore the ability for intercalators to promote the polymerization of short oligonucleotides that have reactive groups on both ends (e.g. *bi-functional* oligonucleotides). As noted in our proposal, we anticipated that bi-functional oligonucleotides, from the length of dinucleotides to octanucleotides, would tend to self-cyclize (Figure 1, upper route). We therefore proposed that a stable primer-template pair would be necessary to initiate a ligation reaction, but we still expected there to be a considerable loss of short oligonucleotides over the course of the reaction due to strand cyclization.

Experimental investigations carried out during the period of funding confirmed that intercalators could facilitate the assembly of short oligonucleotides for polymerization *and* prevent strand cyclization. More specifically, our investigations revealed that the action of intercalation increases the persistence length of short oligonucleotides by increasing base pair stability and rigidifying the helices, such that their reactive ends are only rarely close enough to react. In contrast, without intercalators present, these oligonucleotides exist as random coiled and highly flexible single strands. A consequence of this action by intercalators is that short oligonucleotides are able to assemble into long non-covalent duplex assemblies that promote ligation (Figure 1, lower route).

As a specific demonstration of the ability of intercalators to promote oligonucleotide polymerization and prevent cyclization, the tetranucleotide d(pCGTA) was treated with the phosphate activating reagent *N*-cyanoimidazole, in the absence and presence of ethidium, a well-characterized intercalator of Watson–Crick duplexes. The early products of the ligation reaction (cyclic tetranucleotide, linear octanucleotide and cyclic octanucleotide) were monitored by HPLC. In the absence of ethidium, the major product was observed to be the cyclic tetranucleotide. Only about one-fourth as much cyclic octanucleotide and a trace amount of linear octanucleotide was observed to form. The lack of appreciable linear octanucleotide formation indicated that octanucleotide cyclization occurs rapidly, compared to tetranucleotide ligation. In contrast, in the presence of ethidium, linear octanucleotide formation is greatly favored over tetranucleotide cyclization. It was also observed that the rate of octanucleotide formation in the presence of ethidium is slower than the tetranucleotide cyclization rate in the absence of intercalator. Thus, in addition to promoting linear product formation, ethidium inhibits strand cyclization, presumably by promoting assemblies with greater persistence lengths.

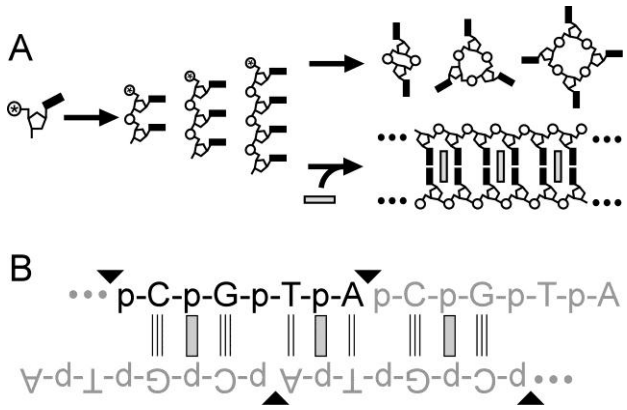

Figure 1. (**A**) Schematic illustration of the strand cyclization problem (*top route*) and the solution provided by intercalation (*bottom route*). In the absence of intercalators, the expected products of chemically-activated di-, tri- and tetra-nucleotides are cyclic oligonucleotides. In the presence of intercalators (shown as grey rectangle), the predicted products are long duplex polymers. (**B**) The extended, regularly nicked duplex resulting from the assembly of multiple copies of the tetranucleotide d(pCGTA) upon intercalation. Triangles indicate sites of backbone nicks, which can be closed by condensation after chemical activation of the terminal phosphate group.

We found that denaturing polyacrylamide gel electrophoresis (PAGE) and SYBR Gold staining was the best way to monitor the synthesis of longer products in intercalation-mediated ligation reactions. For example, the gel image shown in Figure 2 confirmed that d(pCGTA) does not ligate appreciably in the absence of an intercalator. While the tetranucleotide starting material was not observable by SYBR Gold staining, the observation of small amounts of cyclic octanucleotide and cyclic dodecanucleotide is consistent with results obtained by HPLC. In contrast, when ethidium was added to the same reactions, polymers of up to 100 nucleotides in length were observed (i.e. 24 linear couplings) (Figure 2). This analysis also illustrated that the ratio of linear to cyclic products increases with increasing ethidium concentrations for all polymer lengths (Figure 2). For example, no linear octanucleotide or linear dodecanucleotide is detected in the absence of ethidium (lane 1, Figure 2), while approximately equal amounts of linear and cyclic octanucleotide and dodecanucleotide products are observed when ethidium is present at a stoichiometry of one ethidium per tetranucleotide (lane 2, Figure 2). For higher ethidium to tetranucleotide stoichiometries, the relative amounts of cyclic products are far lower than linear products of the same nucleotide length (lanes 3-5, Figure 2)

By varying the concentration of tetranucleotides in our ligation experiments we able to assess the magnitude of the ligation enhancement provided by intercalators. Specifically, in the presence of 600 μM ethidium, polymers up to 100 nt are formed, even at strand concentrations as low as 5 μM, whereas in the absence of intercalators . Thus, *intercalators decreased the concentration required for d(pCGTA) to polymerize by more than a factor of 10,000*.

We also investigated the ability for intercalators to select oligonucleotides for ligation based upon a match between intercalator size and base pair geometry. The gel image shown in Figure 3 demonstrated that ethidium and proflavine, intercalators of Watson-Crick duplexes, are not effective midwives for the polymerization of the hexamer d(A_6). However, coralyne and aza3, molecules that stabilize A•A base pairs (see below), promote the polymerization of d(pA_6). This result provided important support for our hypothesis that intercalators, or molecular midwives, could have helped select the first base pairing motif of life. These results were published in the *Proceedings of the National Academy of Sciences USA* (43).

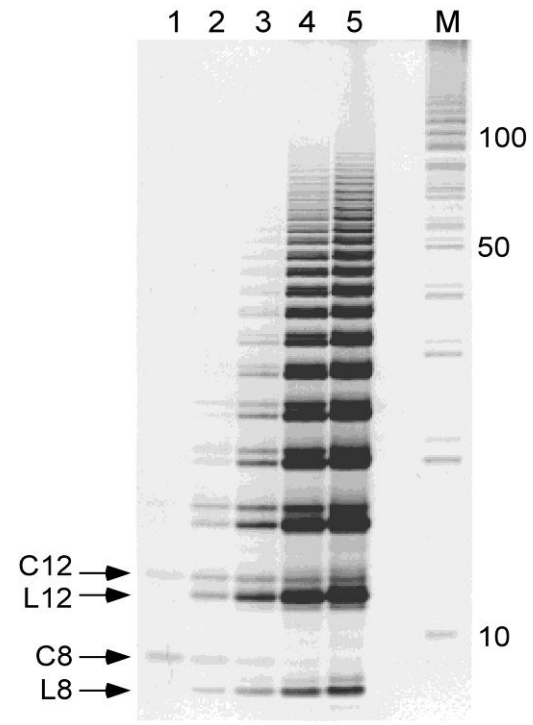

Figure 2. The addition of ethidium to the tetranucleotide d(pCGTA) prevents cyclization and promotes oligonucleotide polymerization. Lane 1: 200 μM d(pCGTA), condensed with 250 mM N-cyanoimidazole at 4 ºC for 72 h. As indicated by arrows, with no intercalator present, only small amounts of cyclic octanucleotide (C8) and cyclic dodecanucleotide (C12) are produced and almost no linear octanucleotide (L8) or linear dodecanucleotide (L12). Lanes 2-5 illustrate that increasing ethidium:substate stoichiometries promote linear polymerization far beyond that observed in lane 1, to *ca.* 100 nucleotides in length. Ethidium concentrations for lanes 2-5 were 100 μM, 200 μM, 400 μM and 600 μM, respectively. The tetranucleotide starting material and cyclic tetranucleotide products observed in HPLC analyses were not observed here, due to inefficient staining.

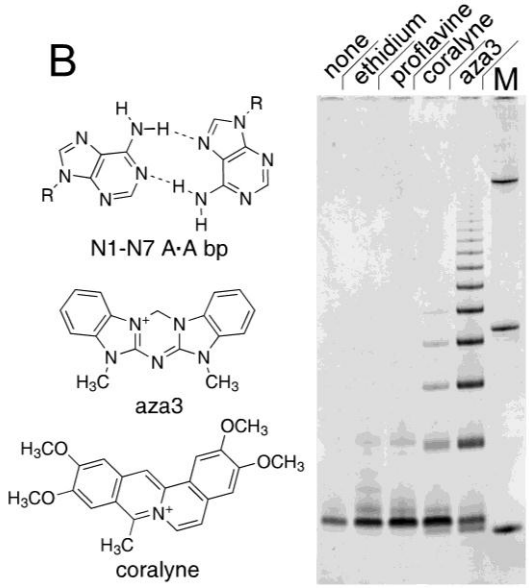

Figure 3. Analysis of products formed by the condensation of 500 μM d(pA_6) in the absence and the presence of various small molecules (750 μM when present). Neither ethidium or proflavine promotes the ligation of d(pA_6). In contrast, coralyne and aza3, which bind homo-A polymers, promote the ligation of d(pA_6).

Advances using intercalators stabilize to duplexes with non-Watson-Crick base pairs.

Our laboratory previously discovered that coralyne (shown in Figure 3) promoted duplex formation between two homo-adenine strands (44). To understand more about the nature of coralyne-mediated duplex formation, we initiated a complementary set of physical and chemical investigations, which included the use of NMR spectroscopy, molecular dynamics (MD) simulations, nucleobase substitutions, and poly(A) assembly by other small molecules that had structural similarities with coralyne. These studies provided valuable information that directed the intercalation-mediated d(pA_6) polymerization experiments discussed in the previous section. Briefly, a combination of MD and base substitution studies revealed that the A•A N1-N7 motif shown in Figure 3 is the base pair stabilized by coralyne binding. Our work with alternative coralyne-like molecules lead to the identification of aza3 (Figure 3) as a second molecule that could stabilize A•A duplexes (38). As shown in Figure 3, aza3 is even more effective than coralyne for promoting polymerization of d(pA_6).

Our studies of small molecule intercalation of poly(A) also revealed a remarkable form of cooperativity between duplex formation and small molecule binding (38). Briefly, we discovered that dilution of a completely assembled aza3-poly(A) sample produced a dissociation curve relating disassembly to sample concentration that was fit perfectly by the critical micelle concentration equation – an equation developed to explain the highly cooperative nature of micelle formation from fatty acids (45). This result lead us to define concentrations at which half of the nucleic acid is assembled by a given intercalator (C*) and the minimal concentration of assembly (or MCA), below which no nucleic acids are assembled by an intercalator (Figure 4). Our analysis showed that a minimum of six intercalators was necessary to initiate the assembly of poly(A). This six-order dependence on ligand binding explained the origin of the highly cooperative nature of ligand-poly(A) structure assembly. With regard to the origin of nucleic acids, we hypothesize that such all-or-nothing assemblies could have been a means by which early nucleic acids were transitioned between times of template-directed replication (above the C*) and times of ligand-free intramolecular structures (below the MCA), where intramolecular folding could promote proto-ribozyme activity. Our studies of intercalator-mediated poly(A) assembly resulted in two publications in *Nucleic Acids Research* (38, 46).

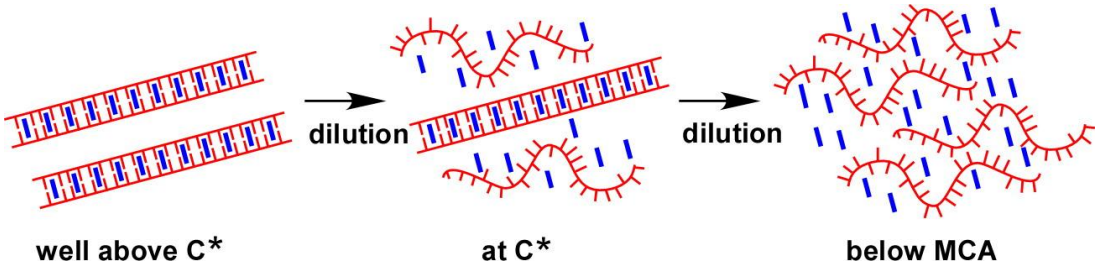

Figure 4. Representation of the highly cooperative disassembly of a poly(A)-intercalator complex upon dilution. The reverse process is achieved by concentrating the nucleic acids and intercalators.

Although our work with poly(A) intercalation-mediated assembly and polymerization provided valuable insights, this system would not likely have given rise to the first genetic polymers, as poly(A) cannot store information. This limitation, juxtaposed with our promising assembly results for duplexes with purine-purine base pairs, led us to investigate intercalation of

duplexes containing Watson-Crick-like base pairs of guanine•isoguanine and adenine•inosine. The possibility of purine-purine base pairs as the ancestors to contemporary Watson-Crick base pairs was proposed decades ago by several prominent origin of life researchers (47-50), and more recent studies from the Battersby and Switzer laboratories reemphasized the feasibility of this proposal by demonstrating the stability of purine-purine duplexes (51, 52). During the past funding cycle, we conducted a comprehensive investigation of small molecule binding to purine-purine duplexes with G•isoG and A•I base pairs. Results from these studies, published in *ChemBioChem*, demonstrated our ability to selectively stabilize duplexes containing these non-standard based pairs (53). These results placed us in excellent position to investigate intercalation-mediated assembly of informational polymers containing purine-purine base pairs.

Advances in understanding how backbone structure affects intercalation-mediated assembly.

There has been considerable discussion regarding the possibility that the first informational polymers had a backbone that was distinct from that of RNA (11). These well-founded hypotheses motivated us to seriously consider alternative backbones as part of our intercalation-mediated polymerization studies. As we moved into this area of research, we found it necessary to first demonstrate that alternative nucleic acid polymers could be intercalated. Because there is no unequivocal spectroscopic signature of nucleic acid intercalation, we were compelled to determine the high-resolution structure of an intercalator bound to a duplex with a non-natural backbone. For this study, we selected 2',5'-linked RNA, because 2',5'-linked RNA has been considered for sometime as a possible ancestor of natural (*i.e.*, 3',5'-linked) RNA (16, 27, 54-56). Additionally, this backbone is, arguably, the closest analog to natural RNA. Using NMR spectroscopy we determined the mode of interaction of proflavine bound to a 2',5' RNA duplex (Figure 5). In addition to providing the *first* structure of an intercalator bound to a non-natural duplex, our study revealed an intriguing pattern of backbone bond angles that, along with previous structures of natural RNA duplexes, provided insights regarding the requirements of the backbone for accommodating intercalation. Furthermore, thermodynamic studies of this system revealed that intercalation of 2',5'-RNA is enthalpically driven and entropically unfavored, whereas intercalation of 3',5'-RNA is almost completely driven by the favorable entropy of binding. This work was published in the *Journal of the American Chemical Society* (57).

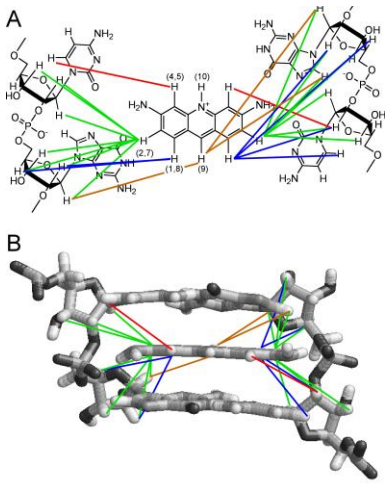

Figure 5. Diagram illustrating the NOEs observed between proflavine and the CpG steps of the 2′,5′ RNA duplex. The IUPAC acridine numbering is shown around the proflavine ring. (B) Diagram illustrating the 3D spatial arrangement of the NOEs shown in image A (viewed from the major groove). Structure shown is the lowest energy NMR-derived structure.

C. Significance of Studies

The research that has been accomplished over the duration of this grant has provided substantial support for our hypothesis that intercalators could have facilitated the assembly of the first RNA-like polymers. Perhaps the most significant result during this granting cycle is our discovery that intercalators provide a possible solution to the "strand cyclization problem", a problem that has frustrated the efforts of prebiotic chemists for decades. Our demonstration that

purine-purine base pairs can be stabilized by intercalators that are slightly larger than Watson–Crick base pairs presents additional possibilities regarding the origin of the first RNA-like polymers and the selection of the first base pairs. Finally, our demonstration that intercalators can act in a highly cooperative manner for driving the conversion of nucleic acids between a duplex state and a single stranded state suggests that intercalators might also have acted as prebiotic concentration-dependent switches that facilitated the cycling of early nucleic acids between two distinct modes, one for replication and one for enzymatic activity.

D. Student Participation

Over the duration of this grant, eight graduate students have participated in this research: Brian Cafferty, Denise (Enekwa) Okafor, Eric Horowitz, Ozgul Persil Çetinkol, Aaron Engelhart, and Ragan Buckley. Ms. Ragan Buckley was partially supported by a teaching assistanceship; Aaron Engelhart and Brian Cafferty were partially supported by Georgia Tech funds, which were provided in association with the PI serving as Associate Director of the Institute of Bioengineering and Biosciences. Denise (Enekwa) Okafor, a minority student that works primarily in the Williams laboratory, was supported part time for two years for her contributions to the molecular modeling of intercalated nucleic acids structures. Aaron Engelhart, Eric Horowitz and Ozgul Persil Çetinkol graduated with their Ph.D. degrees during the course of this grant, and have acknowledged NASA support in their theses. Four undergraduate students have also contributed substantially to these research projects, Michael Chen, Kaycee Quarles, Michael Smith and Benjamin Holladay. The impact of this research opportunity on these undergraduates is clear. Three out of four are in graduate school graduate school. Ms. Quarles is in the Chemistry Ph.D. program at Penn State U., Mr. Holladay is in the Physics Ph.D. program at UCSD, and Mr. Chen is attending Cambridge U. for his Ph.D.

E. Publications and Presentations

Publications produced with the support of the grant:

Özgül P. Çetinkol and N. V. Hud, Molecular recognition of poly(A) by small ligands: An alternative method of analysis reveals nanomolar, cooperative and shape-selective binding, *Nucleic Acids Res.* 37, 611–621 (2009).

Eric D. Horowitz, Seth Lilavivat, Benjamin W. Holladay, Markus W. Germann and Nicholas V. Hud, Solution Structure and Thermodynamics of 2',5' RNA Intercalation, *J. Am. Chem. Soc.*, (2009).

Heather D. Bean, David G. Lynn and Nicholas V. Hud, Self-Assembly and the Origin of the First RNA-Like Polymers, *In* Chemical Evolution II: From Origins of Life to Modern Society, L. Zaikowski and J. M. Friedrich, Eds., ACS Symposium Series, Vol. 1025, 109–132 (2009).

In Suk Joung, Özgül P. Çetinkol, Nicholas V. Hud* and Thomas E. Cheatham III*, Molecular dynamics simulations and coupled nucleotide substitution experiments indicate the nature of A·A base pairing and a putative structure of the coralyne-induced homo-adenine duplex, *Nucleic Acids Res.* 37, 7715–7727 (2009).

Eric D. Horowitz, Aaron E. Engelhart, Michael C. Chen, Kaycee A. Quarles, Michael W. Smith, David G. Lynn and Nicholas V. Hud. Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world. *Proc. Natl. Acad. Sci. USA.*, 107, 5288–5293 (2009).

Aaron E. Engelhart and Nicholas V. Hud, Primitive genetic polymers, *In* Origins of Cellular Life, D. Deamer and J. Szostak, Eds., Life Cold Spring Harbor Press, scheduled publication date of Nov. 2010.

Ragan Buckley, Denise C. Enekwa, Loren D. Williams and Nicholas V. Hud, Molecular recognitionof Watson-Crick-like purine-purine base pairs, *ChemBioChem* 12, 2155-2158 (2011).

Tatsuya Maehigashi, Chiaolong Hsiao, Kristen Kruger Woods, Tinoush Moulaei, Nicholas V. Hud and Loren Dean Williams, B-DNA Structure is Intrinsically Polymorphic: Even at the Level of Base Pair Positions, *Nucleic Acids Res.* 40, 3714–3722 (2012).

Aaron E. Engelhart, Brian J. Cafferty, C. Denise Okafor, Michael C. Chen, Loren Dean Williams, David G. Lynn, and Nicholas V. Hud, Nonenzymatic ligation of DNA with a reversible step and a final linkage that can be used in PCR, *ChemBioChem* 13, 1121–1124 (2012).

Invited seminars and oral presentations where supported work was presented by PI:

University of Ulster, Department of Pharmaceutical Chemistry, Coleraine, Northern Ireland, UK, September 2008.
Symposium in honor of Darwin's 200th birthday, Emory U. October 2008.
Department of Chemistry, University of Virginia, November 2008.
Department of Chemistry, Georgia State University, March 2009.
Georgia Academy of Science Annual Meeting, Keynote Speaker, April 2009.
Mississippi State University, Starkville Mississippi, May 2009.
Emergence in Chemical Systems 2.0, Anchorage, Alaska, June 2009.
Department of Biology, SUNY Buffalo, October 2009.
Gordon Research Conference on the Origin of Life, Galveston, TX, January 2010.
Darwin's living legacy: Conference on Evolution and Society, Alexandria, Egypt, November 2009.
Department of Chemistry, University of Southern California, CA, March 2010.
Department of Chemistry, University of Alabama Birmingham, AL, April 2010.
Symposium on Nucleic Acid Chemistry, Structure and Interactions, Slovenia, May 2010.
Self-Assembly in Biology and Materials Science, Huatulco, Mexico, June 2010.
International Chemical Congress of Pacific Basin Societies (Pacifichem), December 2010.
Astrobiology Program, University of Washington, Seattle, WA, April 2011.
17th Conversation on Biomol. Struct. and Dyn., Albany, NY, June 2011.
ISSOL – The International Astrobiology Society, Montpellier, France, July 2011.
Instituto de Investigación Biomédica (IRB), Barcelona, Spain, September 2011.
Department of Chemical and Biomolecular Nanotech, CSIC, Barcelona, Spain, October 2011.
Department of Bioorganic Chemistry, Universidad de Sevilla, Sevilla, Spain, November 2011.
Department of Chemistry, Université de Liege, Liege, Belgium, December 2011.
National Meeting of the American Chemical Society, San Diego, CA, March 2012.
ACS Mid-Hudson Regional Meeting, Bard College, Annandale-on-Hudson, NY, April 2012.
New York Center for Astrobiology, NASA NAI, April 2012.
Department of Chemistry, Rensselaer Polytechnic Institute, Troy, NY, May 2012.