The Origins and Essential Nature of Biopolymers

Calvin Runnels, Kathryn A. Lanier, Justin K. Williams, Jessica C. Bowman, Anton S. Petrov

and Loren Dean Williams*

School of Chemistry and Biochemistry; Georgia Institute of Technology, Atlanta, GA 30332, USA

*Correspondence: ldw@gatech.edu (L.D. Williams).
Abstract

Life as we know it requires three basic types of polymers: polypeptide, polynucleotide, and polysaccharide. Here we evaluate both universal and idiosyncratic characteristics of biopolymers and incorporate this information into a model that explains much about their origins. We observe that all three biopolymer types are pre-organized, conditionally self-complementary, thermodynamically unstable yet persistent because of kinetic trapping, with chiral monomers and directional chains. All three biopolymers are synthesized by dehydration reactions that are catalyzed by molecular motors driven by hydrolysis of phosphorylated nucleosides. All three biopolymers can access specific states that protect against hydrolysis. These protected states are folded, using complementary interactions among small numbers of recurrent folding elements within a given biopolymer type, or assembled, in complementary heterogeneous associations between biopolymer types. Self-association in a hydrolytic environment achieves self-preservation. Heterogeneous association achieves is mutualism-based partner-preservation. In sum, it seems most probable that life’s polymers arose synchronously via co-evolution in a chemical environment where molecular persistence depended on folding and assembly. We believe that an understanding of the structure, function and origins of any given type of biopolymer requires the context of other biopolymers.
Polymers are large molecules formed by covalently linking small monomers in chain-like structures. Polyethylene, a polymer with molecular formula \((-\text{C}_2\text{H}_4-)_n\) and molecular weight around 5 million Daltons, is used to make bottles and bags. Living systems universally employ three types of polymers. The biopolymers DNA and RNA (polynucleotides), protein (polypeptide) and polymerized sugars (polysaccharide) are winners of an evolutionary process that occurred during the earliest phases of life on Earth. These biopolymers have special properties that distinguish them from other polymers.

**Biopolymers:**

(i) spontaneously fold and assemble into precise and highly elaborate yet fragile structures with meager stability,

(ii) pre-pay the free energy of folding and assembly during monomer synthesis and polymerization, and

(iii) spontaneously degrade in the aqueous environments characteristic of biological systems, but can persist via kinetic trapping.

The three biopolymer types differ profoundly in their structures, properties, and functions. Polypeptide and polynucleotide dominate the functional and informational machineries of life, while polysaccharide is important in physical structure, energy storage, and recognition. Yet biopolymers share many critical properties including chirality, chain directionality, pre-organization and conditional self-complementarity, synthesis by condensation dehydration using phosphorylated intermediates, chemical instability in aqueous media, and persistence via kinetic traps whose depths are increased by folding and assembly. A recognition of the universalities and distinctive characteristics of DNA, RNA, protein, and polysaccharide is a necessary prerequisite for modeling their origins and early evolution.
**Table 1. Biopolymer Attributes**

<table>
<thead>
<tr>
<th>ATTRIBUTE</th>
<th>POLYNUCLEOTIDE</th>
<th>POLYPEPTIDE</th>
<th>POLYGLUCOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary proficiency</td>
<td>maintain, record, and transduce information, catalyze several chemical reactions</td>
<td>catalyze and regulate chemical reactions, provide physical structure</td>
<td>provide physical structure, energy storage and recognition</td>
</tr>
<tr>
<td>Conditional complementarity</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Condition for complementarity</td>
<td>nucleotide sequence</td>
<td>amino acid sequence</td>
<td>stereochemistry (β– vs. α–anomers)</td>
</tr>
<tr>
<td>Number of folding elements</td>
<td>four</td>
<td>one</td>
<td>one</td>
</tr>
<tr>
<td>Folding element(s)</td>
<td>nitrogenous bases</td>
<td>peptide linkage</td>
<td>cyclic glucose</td>
</tr>
<tr>
<td>Enzymatic capability</td>
<td>moderate</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>Sidechain diversity</td>
<td>low: four planer nitrogenous bases</td>
<td>high: twenty amino acid sidechains</td>
<td>N/A: no sidechains</td>
</tr>
<tr>
<td>Sidechain complementarity</td>
<td>yes, base pairing</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Backbone self-complementarity</td>
<td>no: anionic, self-repulsive backbone</td>
<td>yes: neutral, cohesive backbone</td>
<td>yes: neutral, cohesive backbone</td>
</tr>
<tr>
<td>Hydrogen bonding</td>
<td>unipolar, planar</td>
<td>unipolar, planar</td>
<td>bipolar, non-planar</td>
</tr>
<tr>
<td>Net hydrogen bond polarity</td>
<td>large excess of acceptors over donors</td>
<td>equivalent number of acceptors and donors</td>
<td>slight excess of acceptors over donors</td>
</tr>
<tr>
<td>Linearity</td>
<td>always linear</td>
<td>always linear</td>
<td>sometimes linear</td>
</tr>
<tr>
<td>Strand directionality</td>
<td>yes, 5’ to 3’</td>
<td>yes, N to C</td>
<td>yes, 1 to 4</td>
</tr>
<tr>
<td>Secondary structure</td>
<td>helices, stem-loops, pseudoknots</td>
<td>α-helices and β-sheets</td>
<td>elongated fibers</td>
</tr>
<tr>
<td>Conformational constraints</td>
<td>“rigid nucleotides”, planar bases</td>
<td>planar peptide, allowed regions of φψ space</td>
<td>conformational preferences within and between cyclic glucose</td>
</tr>
<tr>
<td>Self-destruct mechanism</td>
<td>RNA: yes (2’ hydroxyl); DNA: no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Folding cofactors</td>
<td>cations</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Degradation by hydrolysis</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Polymerization by</td>
<td>proteins dependent on divalent cations</td>
<td>ribozyme RNP complexes dependent on divalent cations</td>
<td>proteins dependent on divalent cations</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------------------------------</td>
<td>------------------------------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Polymerization mechanism</td>
<td>condensation dehydration, with phosphorylated intermediates and retention of phosphate</td>
<td>condensation dehydration, with phosphorylated intermediates</td>
<td>condensation dehydration, with phosphorylated intermediates</td>
</tr>
</tbody>
</table>

**Figure 1.** Net reactions for biopolymer formation by condensation dehydration and biopolymer degradation by hydrolysis. a) Protein. b) RNA. c) Polysaccharide. All biopolymers are chiral and directional with distinctive ends. Chiral centers (stars) and strand directionalities (arrows) are indicated. Blue boxes indicate atoms involved in the synthesis/degradation reactions.
Figure 2. *Intermediates in the biosynthesis of a) protein and b) polyglucose.*

**Chemical Cousins**

The universalities of chemistry, structure, biosynthesis, and reactivity shared by biopolymers (Table 1) can inform us about their essential natures and origins. We believe that an understanding of the structure, function, and origins of any given type of biopolymer requires the context of other biopolymers. In contrast to standard treatments, we first focus on phenomena that are common to all biopolymers before narrowing our field of view to the unique properties of each type.

*Biosynthesis.* All biopolymers are formed by condensation dehydration reactions (Figure 1) that link well-defined and modest sets of monomers. Proteins are formed by condensation of twenty types of amino acids. Polynucleotides are formed by condensation of four types of nucleotides. Cellulose, the most abundant polymer in the biosphere, is formed by condensation of glucose (McNamara et al. 2015). Complex cell-surface polysaccharides contain fewer than
twenty different monosaccharides (Gabius et al. 2011). Here we will limit our discussion of polysaccharides to polyglucose, encompassing cellulose, glycogen, amylose, amylopectin, and chitin. However, our conclusions apply to polysaccharides in general.

The synthesis of each biopolymer type utilizes phosphorylated or pyrophosphorylated intermediates (Figures 1 and 2) in reactions catalyzed by processive divalent cation-dependent motors. In translation, the motor is the ribosome (Trappl and Polacek 2011). In replication, the motor is DNA polymerase (Steitz 1999). In transcription, the motor is RNA polymerase (Fuchs 1976). In cellulose synthesis the motor is glycosyl transferase (Kang et al. 1984; McNamara et al. 2015; Morgan et al. 2016). RNA and DNA retain a phosphate during polymerization, forming anionic phosphodiester linkages, while other polymers eliminate phosphate groups and form neutral linkages.

Living Dangerously

Here we address the paradox that life’s polymers are chemically unstable in the aqueous environment required for their function.

Fold. All biopolymers spontaneously fold in aqueous media. DNA, RNA, protein, and polysaccharide form elaborate and highly specific three-dimensional structures characterized by specific intramolecular interactions, low configurational entropy and assignment of functional groups to exact locations in three-dimensional space. These low entropy states, stabilized by self-interactions, are called “folded.” Finely controlled molecular interactions allow proteins to fold into domains (Porter and Rose 2012), or into fibers (Shoulders and Raines 2009), composed primarily of α-helices and β-sheets (Figure 3) (Pauling et al. 1951; Pauling and Corey 1951; Eisenberg 2003). RNAs fold into large (Woodson 2011) sometimes monolithic (Ban et al. 2000) structures composed of duplexes, tetraloops, junctions, and pseudoknots (Moore 1999). Complementary DNA sequences fold to double helices of (Watson and Crick 1953) approaching a meter in length with billions of base pairs. Polyglucose forms microfibrils, which are supramolecular assemblies of indeterminate length containing multiple parallel chains (Valeri 2010; Cosgrove 2014).

Assemble. Biopolymers form complex heterogeneous assemblies with specific structures and intermolecular interactions. The prokaryotic ribosome is a spontaneous assembly of 3 large rRNAs and around 50 rProteins. Nucleosomes, which are specific to eukaryotes and some archaea,
are protein–DNA complexes. Protein-saccharide assemblies are critical in cell-cell communication, cell adhesion, and host–pathogen interactions.

_Degrade_. Biopolymers are ephemeral. Biopolymers hydrolyze in aqueous media and suffer a variety of other chemical assaults _in vivo_ and _in vitro_, spontaneously degrading to the level of monomers and beyond. In water, degradation is _always_ favored in the thermodynamic sense. However, biopolymers can persist via kinetic trapping; rates of degradation are sometimes slow. Folding and assembly increase the depth of the kinetic traps, which decreases rates of hydrolysis and other chemical degradation (McKinley et al. 1983; Prusiner 1998; Nahvi et al. 2002; Shoulders and Raines 2009; van der Lee et al. 2014). Folding depopulates states along degradation reaction coordinates. Assembly confers similar protection to the regions of the molecules directly implicated in inter-molecular association. The areas shielded from attack via assembly are called “footprints” (Draganescu and Tullius 1998). Biopolymers can be described as self-protective (by intramolecular folding) and partner-protective (by intermolecular assembly), which can delay but not avert the ultimate fate of any biopolymer – chemical degradation.
Figure 3. Self-complementarity of peptide linkages dominates protein secondary structure. a) Antiparallel β-sheet. b) Parallel β-sheet, c) α-helix. d) A globular protein showing α-helices (violet) and β-sheets (green). e) An amyloid fibril showing the dominance of β-sheets. Hydrogen bonding polarities (arrows) are indicated. Each peptide linkage donates one hydrogen bond and accepts one hydrogen bond.
Biopolymer persistence is programmed. Biopolymers are pre-organized for folding and assembly and thus can persist, via kinetic trapping, in a hydrolytic environment. Pre-organization for folding and assembly is achieved by geometrically arrayed self-complementary molecular interactions privileged by rotameric and steric restraints. Proteins are restrained by the planarity of the peptide linkage and by \( \phi \phi \) restraints (Pauling and Corey 1951; Ramachandran and Sasisekharan 1968). Polynucleotides are restrained by “rigid nucleotides” (Sundaralingam and Westhof 1979) and planarity of bases. Polysaccharides are restrained by conformational preferences within and between sugars (Stick and Williams 2010). Thus, biopolymers retain a kinetic and thermodynamic propensity to fold even at high temperature or in the presence of chemical denaturants. When the temperature is lowered or the denaturant is removed, folding is spontaneous.

The Anfinsen Postulate. The folded state of a protein, known as the native state, was said by Anfinsen to be determined only by amino acid sequence (Anfinsen et al. 1961). The native state was said to be unique, stable, kinetically accessible and at a free energy minimum.

We believe that the Anfinsen model should be understood as an imprecise description of the essential nature of proteins. The essential nature of proteins is found in the self-complementarity of the backbone and in the dominance of cohesive-interactions of the backbone. In the interior of the native folded protein studied by Anfinsen, essentially all backbone hydrogen bond donors interact with backbone hydrogen bond acceptors, and vice versa. The intramolecular interaction space within Anfinsen’s folded protein is dominated by cohesive-interactions of the backbone.

Anfinsen empirically described the behavior of an isolated protein under a specific set of non-denaturing conditions - in dilute solution. His model can correctly describe protein folding under those conditions. However, backbone cohesion dominates protein folding under all conditions. The Anfinsen model does not predict amyloids (Figure 3d) and appears to be partially falsified by their observation. Amyloids are protein assemblies composed primarily of \( \beta \)-sheets. \( \beta \)-sheet is the ancestral mode of polypeptide self-interaction (Kovacs et al. 2017; Lupas and Alva 2017). Amyloids grow by inducing the conversion of \( \alpha \)-helices to \( \beta \)-sheets during formation of hyperstable fibrils. At high in vivo concentrations, \( \beta \)-sheet is the default mode of self-interaction of the backbone of essentially any amino acid sequence (Fändrich and Dobson 2002; Pedersen et al. 2010). Amino acid sequence, we believe, should be considered a second order perturbation.
of cohesive backbone interactions. Amyloids and isolated globular proteins follow the same organizing principles; both demonstrate the dominance of cohesive backbone interactions under all non-denaturing conditions.

**Complements to the Chef**

*Complementarity.* All biopolymers are conditionally self-complementary. Self-complementarity refers to the proficiency for preferential self-binding, which is the ability to attract and associate with self to the exclusion of non-self. Three-dimensional structures of folded DNA, RNA, protein, and polysaccharide reveal extensive networks of highly specific molecular interactions in which biopolymers complement themselves.

Self-complementarity of biopolymers is conditional and can be switched between “on” or “off” states. For DNA and RNA, self-complementarity is conditional on nucleotide sequence, which acts as the controlling switch for the formation of elongated DNA duplexes or RNA stem-loops. The sequence r(CGA---UCG) can form an intramolecular stem-loop or an intermolecular dimer while r(CGA---CGA) cannot. This complementarity is controlled by direct molecular interactions between nitrogenous bases. Complementarity within or between polynucleotide strands is switched on and switched off by changing the sequence.

The backbone of polypeptide is intrinsically self-complementary. Polypeptide selectively adheres to itself; arrays of hydrogen bond donors and acceptors are geometrically matched within α-helices and β-sheets (Figure 3). The cohesive interactions of protein, like those of polynucleotides, are modulated by sequence, but at a higher level of control and sophistication. Proline can switch the self-complementarity to the off state by unbalancing the ratio of hydrogen bond donors to acceptors. Using mechanisms that are more subtle and less direct than those of polynucleotides, sequence modulates protein self-assembly. An amphipathic α-helix can be relatively stable on the surface of a globular protein, forming local self-complementary hydrogen bonding interactions, but in isolation of the globular domain will aggregate and convert to β-sheet, forming alternative non-local self-complementary interactions.

Polyglucose, like DNA, RNA, and protein, is conditionally self-complementary. Cyclic glucose is intrinsically self-complementary, as observed in Figure 4, where essentially all hydrogen bonding functionalities of each glucose are positively engaged with those of other glucose moieties. Cellulose and chitin form stable intra-chain interfaces stabilized by hydrogen bond
donors and acceptors in large complementary arrays. The anomeric linkage provides an on/off switch for self-complementarity. β-Anomers such as cellulose and chitin are self-complementary. α-Anomers such as glycogen and amylose are not.

Specific molecular interactions observed in folded DNA, RNA, or protein, utilize unipolar hydrogen bonds such as those of keto oxygens, amide and imine nitrogens, and polarized amino groups (Figure 3). The molecular interactions of polysaccharides are dominated by hydroxyl groups (Figure 4). Hydroxyl groups are bipolar, with the ability to both donate and accept hydrogen bonds.

Figure 4. In cellulose, complementary arrays of hydrogen bond donors and acceptors stabilize the assembly of (1-4) polymerized glucose. The β-anomer but not the α-anomer enables complementary glucose-glucose interactions.

Separated at Birth

Although DNA, RNA, protein, and polysaccharide have many chemical and structural similarities, they are distinguished by obvious differences. The backbone of protein is neutral,
cohesive, and self-complementary, enabling formation of hydrophobic cores where water is excluded. The backbones of RNA and DNA are anionic and self-repulsive. RNA folds to globular structures with wet, salty cores while DNA tends not to form globular structures at all. Polyglucose forms dry but hydrophilic cores stabilized by the vastness of the contact area. Protein and polysaccharide folding are generally independent of cofactors. RNA and DNA folding always requires cationic cofactors. The specific ordering of sidechains along monotonous backbones of RNA, DNA, and proteins is an important device for modulating and manipulating conformation and molecular interaction. Protein sidechains are many and chemically diverse. RNA and DNA sidechains are few and are chemically homogeneous. Polysaccharides lack sidechains altogether but can be branched and chemically modified. RNA, DNA, protein, and some types of polyglucose (cellulose, amylose and chitin) are linear, while glycogen (animals), and amylopectin (plants) are branched. Each of the linear biopolymers folds to form helical structures (Pauling et al. 1951; Pauling and Corey 1951; Watson and Crick 1953).

*Adding it up.* The net hydrogen bonding polarities of polypeptides sum to zero, with equivalent numbers of hydrogen bond donors and acceptors. Polyglucose has a slight excess of hydrogen bond acceptors over donors. RNA, and especially DNA, have a large excess of acceptors over donors. Polypeptide has been designed by nature to self-assemble predominantly via backbone interactions. Polynucleotides have been designed by nature to self-assemble predominantly via sidechain interactions.

*Functional Distance.* Is it possible to relate the functional roles of biopolymers to their structures? First, one must attempt to accurately describe biological functions. What does each biopolymer type do? There are no bright lines—functional roles are not rigidly proscribed by polymer type. There is enormous diversity in the chemical transformations of biological systems, and the great majority are catalyzed and regulated predominantly by proteins. Protein contributes enzymes, fibers, adhesives, pumps, pores, switches, and receptors. RNA performs far more limited but nonetheless critical catalytic functions. By contrast, DNA appears to be exclusively informational. On the whole, polynucleotides maintain, record, read and transmit sequence information. Polysaccharides contribute structural elements along with energy storage capabilities and elaborate recognition functions.

Ribozymes (Kruger et al. 1982; Guerrier-Takada et al. 1983), which are RNA-based “enzymes,” have correctly assumed a great deal of symbolic significance and importance in
discussions of fundamentals of biology and the origin of life. However, thus far there has been no observation of a biological RNA-only ribozyme that is formally enzymatic; there are no RNA-only biological ribozymes that turn over (Kruger et al. 1982; Hutchins et al. 1986; Prody et al. 1986), and thus formally there are no known biological RNA-only enzymes. All protein-independent biological ribozymes discovered thus far perform suicide phosphoryl transfer functions. By contrast, highly abundant and critically important ribonuclear protein ribozymes (protein-assisted ribozymes) do turn over and are thus fully enzymatic. These RNP ribozymes include the ribosome (Khaitovich et al. 1999), RNase P (Guerrier-Takada et al. 1983) and the spliceosome (Brody and Abelson 1985). No catalytic function of polysaccharide has been observed thus far, to our knowledge.

**Fraternal Twins: DNA and RNA**

DNA and RNA both self-assemble to form double-helical structures with central cores of paired and stacked nucleobases, framed by external, anionic backbones. DNA and RNA appear similar in chemical representations, differing only by a single atom on the backbone and by a methyl group on one base.

The 2' hydroxyl group profoundly influences folding, providing a nucleation hook for base-backbone association, thus fostering diverse loops and junctions. The imbalance of hydrogen bond acceptors over donors of DNA is partially relaxed in RNA by the 2' hydroxyl group, which provides a locus for intramolecular cohesion. The most frequent base-backbone assembly of rRNA is the GNRA tetraloop. There are over 40 examples of this motif in the large ribosomal subunit of prokaryotes (Hsiao et al. 2009). These structures, and many other non-helical structures, are stabilized by intramolecular interactions between 2’ hydroxyl groups and RNA bases (Figure 5b). These base-backbone interactions promote folding of RNA into local stem-loops, which are often further stabilized by tertiary interactions (Figure 5c). DNA, by contrast, is generally restrained to base-base associations, forming long, monotonous double helices (Figure 5d).
Figure 5. The impact of the 2’ hydroxyl group on polynucleotide reactivity and structure. a) Reactivity. RNA holds a gun to its own head. The 2’ oxygen is a nucleophile that is poised to attack the adjacent phosphorous atom, cleaving the RNA backbone. b) Assembly. The 2’ hydroxyl group nucleates folding of complex structures by enabling hydrogen bonding between the backbone and bases, as demonstrated in the GNRA tetraloop. The 2’ hydroxyl of a guanine forms a hydrogen bond with the N7 of an adenine. In addition, the N1 and N2 of the guanine form a hydrogen bond with a phosphate oxygen of the backbone. c) Complexity. RNA folds into elaborate three dimensional structures. d) DNA folds to long double helices. In panels a and b, hydrogen bonding groups that do not form hydrogen bonds are omitted for clarity.

Profound differences in reactivity distinguish RNA from DNA. DNA stays safe while RNA “holds a gun to its own head” (Hud 2010). Each 2’ hydroxyl of RNA is poised to attack an adjacent phosphorus atom and cleave the backbone (Figure 5a). Thus, RNA is far more labile than DNA, with a shorter chemical lifetime. The rate of RNA self-cleavage is modulated by local structure,
flexibility, pH, and interactions with cations. The DNA backbone is much less reactive, persisting in a deeper kinetic trap than the backbone of RNA.

**Sibling Rivalry: Polypeptide vs. Polynucleotide**

We believe that polypeptide and polynucleotide can be understood using the concept of “molecular Yin and Yang.” In some respects, these two polymers appear to be polar opposites. Protein folds via a cohesive backbone, with sidechains directed away from the folding core, while polynucleotide has a self-repulsive backbone and assembles with sidechains directed inward. In other respects, there are close analogies in protein and polynucleotide. Both are linear, directional and chiral polymers formed by condensation dehydration reactions. Both fold via hydrogen bonding of planar elements, which contain functionalities with fixed hydrogen bonding polarities, relative positions and orientations. Both contain biological information represented by sequences of linked monomer units. The Yin and Yang of polypeptide and polynucleotide (Figure 6) informs our understanding of the form, function and evolution of biopolymers, which is only clear in context of all biopolymers.
**Figure 6.** Yin and Yang. Polynucleotide has a self-repulsive backbone and assembles with sidechains directed inward. Protein assembles with a cohesive backbone, with sidechains directed outward.

**Nature Chose Phosphate**

Westheimer suggested that phosphates dominate molecular biology because phosphate is a kinetically trapped, tunable, water-soluble leaving group that can be linked to small molecules, conferring anionic charge and blocking their transit across membranes (Westheimer 1987). While correct, this analysis omits the importance of phosphate in mechanochemical coupling.

All biopolymerization reactions utilize phosphorylated or pyrophosphorylated intermediates (see above, Figures 1 and 2) in reactions catalyzed by processive enzymes. Phosphorylated intermediates appear to be necessary for the mechanochemical coupling required for processive polymerization. The polymerases that make DNA, RNA, protein, and polysaccharide are motors. Translocation is energy-driven; the nascent polymer translocates relative to the polymerization enzyme. Mechanochemical coupling in motor proteins is commonly
linked to association/dissociation of phosphate because phosphate has “claws” that reach out in three-dimensions; phosphate can grab onto and deform proteins. The strength, directionality, and unipolarity of hydrogen bonding and electrostatic interactions between phosphate and protein link phosphate association/dissociation to protein conformation (Rice et al. 1999; Wittinghofer 2016). This coupling of directed molecular displacement (work) to association/dissociation of phosphate, which is in turn linked to pyrophosphate hydrolysis, has been characterized in myosin and kinesin, in the ribosome and in DNA, RNA and cellulose polymerases (Wang et al. 1998; Morin et al. 2015; Morgan et al. 2016; Arias-Gonzalez 2017). During polymerizations of DNA, RNA, protein, and polyglucose, translocations are structurally and energetically coupled to phosphate association/dissociation.

**Molecules in Mutualism**

Mutualisms are everywhere in the biosphere and are fundamentally important in evolution, ecology, and economy (Moran 2006; Bronstein 2015; Douglas 2015; Gray 2017). We have previously proposed that formalisms for describing mutualisms on levels of cells, organisms, and ecosystems also apply to biopolymers (Lanier et al. 2017). Biopolymers are *Molecules in Mutualism*. The mutual benefit, exchange of proficiencies, persistence, interdependence, and co-evolution that characterize relationships on cellular, organismal, and ecological levels have direct parallels in the behaviors of biopolymers.

A mutualism is a persistent and intimate interaction that benefits multiple interactors (Douglas 2015). Mutualism is reciprocity; an interactor proficient in obtaining certain benefits confers those on other interactors, which reciprocate by conferring different or complementary benefits (Schwartz and Hoeksema 1998). Because mutualisms are prolonged and intimate, partners in mutualism influence each other’s evolution. Evolutionary change of one partner triggers change of the other. Co-evolution increases the space available for phenotypic exploration and stabilizes the mutualism.

**Levels of Mutualism.** Mutualisms are understood to operate on levels of cells, organisms, ecosystems, and even societies and economies. The eukaryotic cell is a culmination of mutualism between simpler prokaryotic cells (Sagan 1967; Poole and Gribaldo 2014; Gray 2017). The majority of land plant families are mycorrhizal; this plant-fungi mutualism is traceable to the origins of land plants (Wang and Qiu 2006). Flowering plants such as the fig (Ficus spp., Moraceae)
and insects such as the fig wasp (Agaonidae, Chalcidoidea) form obligate mutual relationships
(Machado et al. 2005). The wasp depends on the fig for food and the fig depends on the wasp for
pollination. Pollen-bearing female wasps initiate seed production in the fig by distributing pollen.
The fig provides each wasp larva with a fig seed, which is required for wasp development.
Essentially every species on Earth is involved in mutualisms.

Polypeptide synthesizes polynucleotide (polymerases) and polynucleotide synthesizes
protein (ribosome). Translation of mRNA to protein represents a transduction of information
between two inverse but related forms. Furthermore, during the essential steps of this
transduction process, coding is performed by proteins (aaRS enzymes that charge tRNAs), while
decoding is carried out by RNA (mRNA and rRNA) in the ribosome.

Molecules in Mutualism predicts:

- survival – extant biopolymers are more persistent than competing polymer types,
  which are now extinct,
- co-evolution – biopolymers co-evolved and created each other,
- fitness – biopolymers are more fit in combination than in isolation,
- distance – each biopolymer type has distinct proficiencies,
- innovation – proficiencies of one type of biopolymer release constraints on
  partner biopolymer types,
- robustness – the backbone structures of biopolymers have been fixed for billions
  of years; biopolymers in mutualism are the seminal and most ancient mutualism
  in the biological world.

The large phylogenetic distance common in species-level mutualisms has parallels in
chemical differences between, for example, RNA and protein. A biopolymer that assembles
primarily with its sidechains exchanges proficiencies with a polymer that assembles primarily with
its backbone.

**Origins of Biopolymers**

Darwinian processes require biopolymers, which must therefore pre-date natural
selection. The universality of (i) synthesis by condensation and degradation by hydrolysis, (ii)
self-complementarity and pre-organization, (iii) folding-conferved protection from chemical
degradation, and (iv) molecular mutualisms is consistent with biopolymer origins in a shared
milieu driven by ‘self-interest’ and co-evolution. Self-interest is equivalent to chemical persistence, a survival strategy with demonstrated utility in extant biological systems (DeWoody et al. 2008).

The best-supported models suggest that biopolymers arose in a geo-chemical environment of alternating synthesis and degradation, driven by cycling water activity (Forsythe et al. 2015). Synthesis is favored during low water activity (day) and degradation is favored in high water activity (night). Molecular persistence depended on folding and assembly. Self-complementary in such an environment is simply an expression of molecular self-interest. Heterogeneous associations are expressions of partner-protection in the context of molecular mutualisms.

The nominal stability of biopolymers suggests unfolding and disassembly conferred some advantage. The ability to unfold and disassemble provides the mechanisms for prospecting for new folds and new partners during geochemical cycling. Extremely stable folds and assemblies could form molecular dead-ends.

This model suggests that biopolymers were chemically selected in a sea of chemical diversity in which competing polymers, including racimates, polyesters, heterogeneous polymers, etc., failed to complete because of their inability to fold and assemble with appropriate stability, and were thus forced to extinction. One cannot exclude the possibility of ancestral polymers that dominated in initial stages, and were partially repurposed by more successful second generation polymers.

**Conclusion**

Although biopolymer types are traditionally studied and taught in isolation of each other, we believe that DNA, RNA, polypeptide, and polysaccharide are best understood in the context of their shared attributes, key differences, mutualism relationships and origins via co-evolution. Recognition of universalities highlights the most important differences between biopolymers, including charge of the backbone, diversity of sidechains, and functional distance, and highlights the special relationships between subsets of biopolymers (DNA vs. RNA, polypeptide vs. polynucleotide). Only by examining biopolymers in context of each other can we hope to achieve a complete understanding of the fundamental molecules of life.
Acknowledgements. The authors thank Drs. Gary Schuster, Nicholas Hud and Roger Wartell for helpful discussions. This work was funded in part by NASA Exobiology grants NNX16AJ28G and NNX16AJ29G. The authors declare they have no conflict of interest.
References


