HP world paper: March 19, 2015 – July 17, 2015

Contents

COLOR CODE

Blue: **reformulate**

Red: **comments, todo**

1   Introduction

Functioning of modern organisms is not possible nucleic acids and proteins. Cells produce these extremely long polymers with cell machinery like ribosomes and polymerases – also extremely long polymers. Therefore question "how long polymers can be produced prebiotically?" is crucial for the origin of life research. we know that amino acids can be produced prebiotically [1] and are abundant in stony meteorites [2]; significant progress has been made in synthesis of single nucleotides [3]. **However a discovery of a mechanism of prebiotic production of biochemically long polypeptides or nucleic acids despite certain success [4, 5, 6, 7, 8, 9, 10] is still in the future.**

Spontaneous polymerization processes fall under what we call Flory problem (for details see sec.2): chain length follow exponential distribution and desirably long chains are present in negligible quantities. While increasing equilibrium constant by means of catalysis or changing conditions to dry medium would lead to longer chain lengths, it will still give exponentially decreasing distribution of length. We show the proof of it for the case of monomer addition

in A. The proof for polymer concatenation is present in [11]. This dynamics can be observed even in some autocatalytic systems [12].

This suggests that first catalysts emerged from mixture of short sequences, which were both not diverse and had poor information content. Therefore the problem of short chains also brings with itself problems of information production and complexity emergence [13, 14]

We sought a simple structure based mechanism, which would solve Flory problem, which would be able to select for certain sequences and amplify their chain lengths.

We present here a physical mechanism, which give rise to heavy tail chain lengths distribution and selects the sequences based on the physical principle: hydrophobic interaction. write about autocatalytic sets here and cite Kauffman, maybe also mention far from equilibrium statement. mention composomes, mention lipid vesicles [15, ?] Our *in-silico* experiments demonstrate that binary polymers capable of primitive folding and hydrophobic interaction is a working mechanism. Not only we get longer chains and selection. Our systems gives reasonable computational diversity of the sequences.

2   Flory problem

In the systems with spontaneous polymerization chain lengths of the produced polymers follow exponential (Flory) distribution (see fig. 1 and appendix section A). Probabal prebiotic concentrations of a.a here [?] Experimental results for RNA yield:

- max 10mers [10] in solution

- up to 55mers on mineral surfaces [10] (primary product contains 20-40 mers)

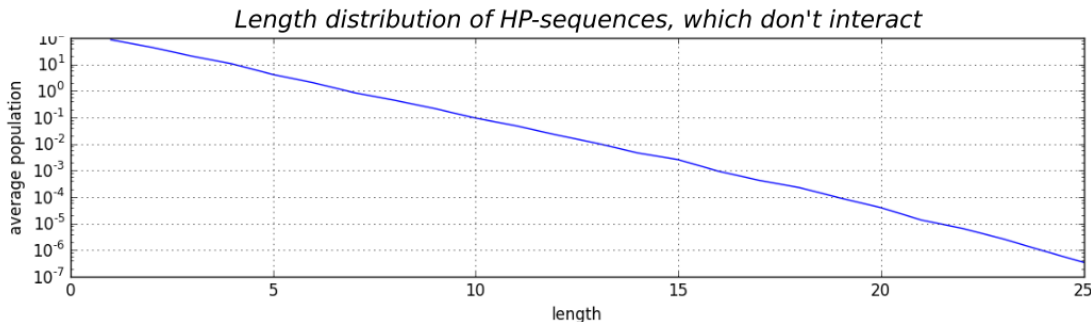Experimental results for peptides yield:

-

Fig. 1: We are dealing with an exponential length distribution. Model described in A

add another line with lower degradation constants Let us consider set up, which gives figure 1. This particular system is a throughput system. We have import of monomers and a sink. Polymers polymerize by monomer addition from one side. All bonds of the polymers are subject to random hydrolysis.

This model gives the following polymer abundance for the blue line:

$$\frac{[10mers]}{[1mers]} \propto 10^{-3}, \qquad \frac{[20mers]}{[1mers]} \propto 10^{-7} \qquad (1)$$

For such a system, if we start with nano-molar concentrations of monomers, 40-mers will have concentrations $\propto 10^{-23}$ mol/L, which is just a few molecules per liter.

We worked through several models of spontaneous polymerization, and all of they yield exponential or near-exponential length distribution.

3   We use hydrophobic effect to solve Flory problem

Hydrophobic effect is an effect which lays in the core of protein folding. It is mostly an entropic effect originating from the disruption of hydrogen bonds between water molecules by the nonpolar solute. [16]. When hydrophobic molecules are placed in the water, water molecules form cage-like structures around them (see fig. 2). When hydrophobes come into contact, water molecules get released; this increases entropy and therefore decreases free energy. This decreases in free energy allows proteins keep a tight hydrophobic core.
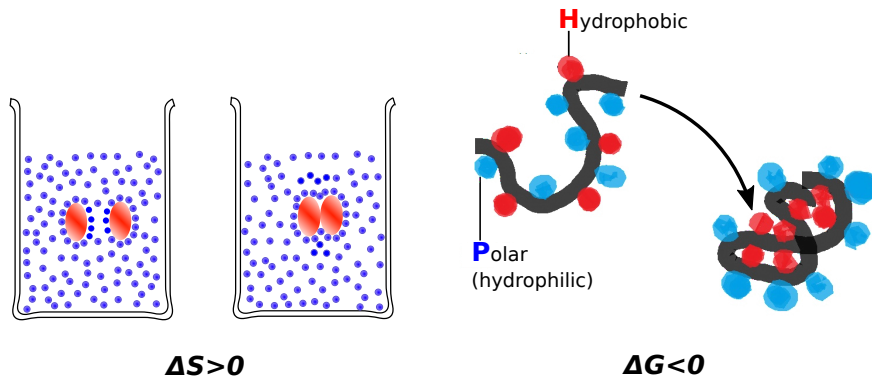

Fig. 2: Hydrophobic effect is an entropic effect originating from the disruption of hydrogen bonds between water molecules by the non-polar solute. When hydrophobic molecules are placed in the water, water molecules form cage-like structures around them. When hydrophobes come into contact, water molecules get released; this increases entropy and therefore decreases free energy.

3.1   We use HP model to represent prebiotic polymerization

To describe effect of hydrophobic interaction on prebiotic polymerization, we adopt the HP model – one of the simplest models of proteins; it's well studied and sequence space is well understood[17, 18, 19, 20, 21]. While initially HP model was introduced as a model for proteins, we are indifferent to exact chemical nature of the prebiotic polymers and consider only principles of spontaneous polymerization.

HP model features:

- It is a two dimensional square lattice model of protein folding

- It has 2 types of monomers: hydrophobic (H) and polar (P).

- HP model has a folding code: presence of hydrophobic interaction makes some confor-
  mations of the same chain more energetically favorable than others. Moreover certain
  chains will have a unique conformation, which delivers free energy minimum. This
  conformations are called native states, and sequences, which have native state, are con-
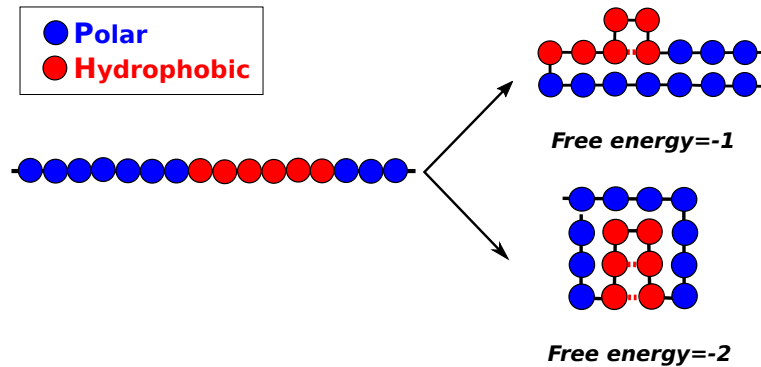 sidered being capable of folding (see fig. 3).

Fig. 3:

Because of the ability to form hydrophobic contacts in water, even hetero-polymers that
are as simple as HP-polymers will often fold up, even as short chains. While short HP
chains will not necessarily have great stability qualities (they will often be fairly amorphous
"oil-drop"-like balls that are ensembles of conformations), some HP sequences will fold more
uniquely than others. Latter will spend most of the time in the native state. And, what is
important, it has been shown that a relatively large fraction of sequence space will fold to
compact structures or compact ensemble structures[17].

3.2   HP-foldamers can work as prebiotic catalysts

Our central premise is that the same promiscuous hydrophobic interactions that can cause
random HP heteropolymers to collapse into compact, folded, structures. However hydrophobic
interaction will also cause polymer-polymer attraction and binding between molecules. In
some cases, a folded HP-polymers can provide a hydrophobic "landing site" for another HP
polymer and/or another H monomer (see fig. 4).

When a folded chain has exposed hydrophobic monomers on its surface it can attract
another chain with hydrophobes as well as activated hydrophobic monomer. Interaction
between 3 of them localizes growing chain and next monomer (fig.4(a)). In addition to
that, hydrophobic interaction also lowers activation barrier of the polymerization reaction,
accelerating reaction this way.

One hydrophobic interaction is about $1 - 2kT$. Given that rate of catalysis is proportional to an exponent of the activation barrier, 3-4 hydrophobic interactions are enough to increase polymerization rate ≈ 100 times (fig.4(b)). Of course, this is not a good rate compared to modern biochemical rates. However the very first catalysts don't have to be very efficient: their purpose is to create a driving force of evolution.

HP-catalysis drives addition of hydrophobes to hydrophobes. A seemingly logical conclusion would be that one will end up with purely hydrophobic polymers. However this is not true. Sequences capable of catalysis must have a relatively stable structure. Purely hydrophobic sequences don't have this property: they have very many conformational states with the same low free energy. Therefore they will spend a lot of time jumping between those states and their bonds will be affected by hydrolysis. They also will not be able serve as catalysts. Sequences with $50 - 80\%$ of hydrophobes, on the other hand, will have the most stable structures; they will be protected from hydrolysis and will be able to localize growing chain with the next added monomer.
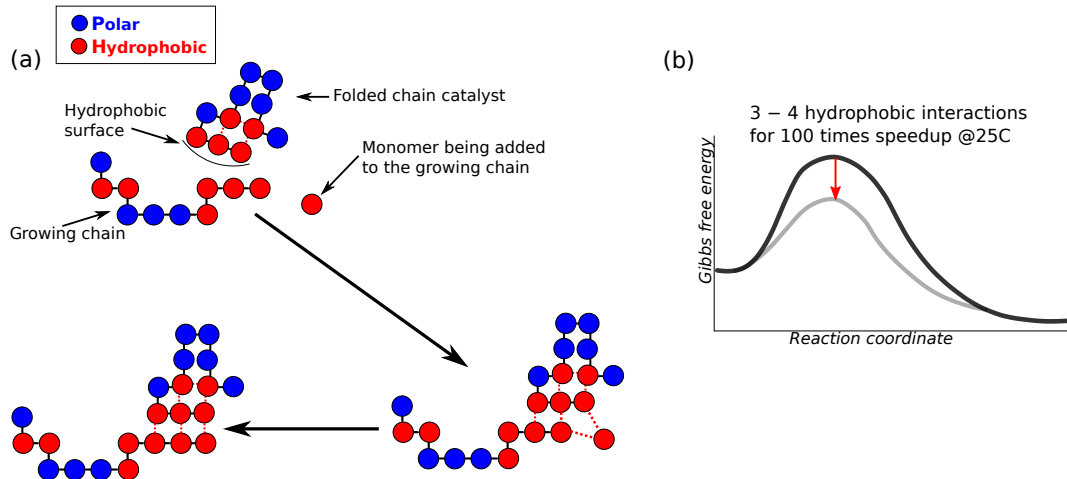

Fig. 4: Catalyst catalyzes a growing of an unfolded hp-polymer. Having just 3-4 hydrophobic contacts is enough to lower an activation barrier for $\propto 100$ times at room temperature.

4 Materials and methods

4.1 Simulations

To test our hypothesis we performed direct stochastic simulations on several. We used PDMmod method [**?**][1]. Stochastic simulations keep track of each molecular specie in the system. However simulations are limited due to computational reasons. First of all we have to explore conformational space of every polymer. This task is NP-hard (we use HPSandbox algorithm[17, 22] [2]), so we had to limit maximum chain lengths to 25. We also try to keep total

---

[1] C++ library and description can be found here: https://github.com/abernatskiy/pdmmod
[2] Python implementation and description can be found here: http://hp-lattice.readthedocs.org/en/latest/

number of species in the low thousands, to avoid computational costs. We do it by introducing dilution parameter d: molecules are being removed from the system with probabilities $\propto d$. This mimics a protocell splitting and loose of materials due to it. Total number of molecules varies from simulation to simulation, however it mostly holds in the region insert.

    We start our simulations with a small pool of monomers, usually below 100 molecules.

- We assume that there are enough of activated monomers in the system, so that their concentrations are constant. This way we don't have to track them in the simulations.

- Polymers can therefore spontaneously grow with the rate α, which we put equal 1 for simplicity; all other rates are relative to the growth rate.

- Activated monomers decay into regular ones with the rate $a \gg 1$. This imitates input of food into the system.

- Dilution parameter d varies from $\propto 0.01$ to $\propto 1$.

- Hydrolysis has constant rate d_h per bond; it varies from $\propto 0.01$ to $\propto 1$.

- Folding and unfolding reactions happen very quickly with the rates $k_{unf} \gg 1$ and $k_{unf} \cdot \exp(E_{native}/kT)$ correspondingly.

- Catalysis rate is proportional to the exponent of hydrophobic energy E_h and number of contacting hydrophobes n_c: $\alpha \cdot \exp(E_h \cdot n_c/kT)$

- Some experiments also include aggregation reactions for the long hydrophobic chains.

We looked at the lengths distribution in steady state. In order to account for stochastic effects we took average over several realizations. We also looked at the time evolutions of specific chains to investigate correlations between sequences and internal dynamics. The simulations were performed on Computing Cluster of Laufer Center. See appendix for simulation details.

Experiment 1. Reproduction of Flory distribution. We started simulations with small pool of monomers (20 H and 20 P). We ran 30 identical simulations for 200 s each, with measurements taken every 0.1s. Steady state is being achieved around 30-50s. To calculate length distribution, we took one trajectory and calculated average over time over all time points after 100s; so we got 1000 time points for every chain length, over which we averaged. The rate of conversion of activated monomers into regular ones is $a = 100$. We took dilution rate of $d = 0.5$. We ran experiments for 2 hydrolysis rates: $d_h = 0.3$ and $d_h = 0.03$. We varied hydrolysis and dilution rates. Experiments with $d_h = 0$ reproduce accurate exponential curves; adding hydrolysis, however, slows down distribution around short lengths. This effect is due to constant concentration of activated monomers: there's no competition for "food". This enriches population of short chains, however doesn't affect longer chains significantly, leaving their distribution nearly exponential.

Experiment 2. Study how folding affects length distributions. In addition to the parameters of the experiment 1 we also added non-zero hydrophobic energy and introduced folding and unfolding reactions. Hydrophobic energy is taken $E_h = 2kT$ and rate of unfolding is $k_{unf} = 100$. We varied parameters around given values and didn't notice qualitative changes of the system's behavior. From the figure 5 in section 5 we can see that presence of folding doesn't affect length distribution significantly.

Experiment 3. Introduction of HP-catalysis. In addition to folding in this *in-silico* experiment we introduced interaction between proteins. All parameters are as above. We varied parameters of the simulations, and noticed significant stability of the length distribution towards change of d_h and d. distribution is sensitive towards hydrophobic energy, as expected. Chain length distribution has a noticeably non-exponential behavior in the region when $E_h = 1 - 3kT$

Experiment 4. HP-catalysis and aggregation. Short description

Experiment 5. Study of internal dynamics. Short description

5 Results

Simulation: folding. Presence of folding reactions and absence of catalysis ones relationship between abundance and length at steady state follows exponential distribution for longer chains and is slower than exponential for short chains. While presence of folding makes some of the folded sequences get higher than average for their length populations, these populations are never more than few-fold of average population of regular sequences and don't change nature of the distribution.

Simulation: folding and HP-catalysis. Presence of catalysis in the system skews the distribution significantly, and while it leaves average chain length about the same it brings substantial excess of long sequence compared to the cases without catalysis6. The system is fairly stable towards hydrolysis and dilution parameters. It allows for 1 order of magnitude change in those parameters without significant change in the behavior of the system. Sequences responsible for the skew of the distribution are few in numbers they all are catalysts and have long stretches of hydrophobs, which also means that they are products of catalysis. In the figure 7 there examples of several sequences. The lines represent average over 30 time evolutions. For this particular experiment, concentration of monomers at steady state is $\propto 100$. Most of the longer sequences have average populations $\ll 1$. However for the most of the chain lengths there are few sequences, which dominate populations significantly.

Some discussion on what it means and that it is still not enough for a proper inheritance and evolution, however it allows us to get to longer sequences, some of which can be "magical".
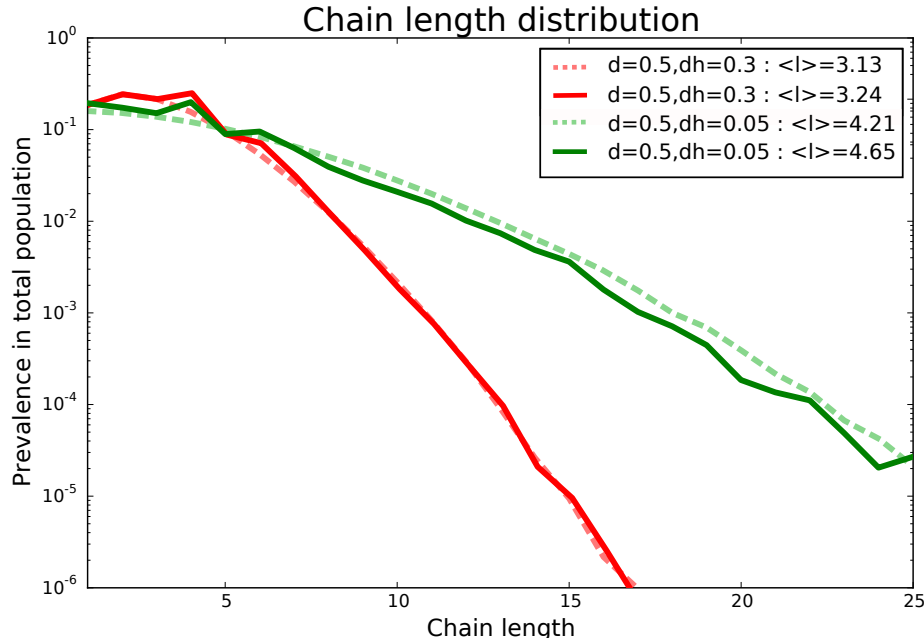
6 Discussion

Fig. 5: Dashed lines represent polymerization without folding or catalysis. Solid lines correspond to simulations run with folding but without catalysis. For details of simulations see section 4.1, Experiment 2.

A   Simulation details

References

[1] S. L. Miller. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science*, 117(3046):528–529, May 1953.

[2] Mark A. Sephton. Organic compounds in carbonaceous meteorites. *Natural Product Reports*, 19(3):292–311, May 2002.

[3] Matthew W Powner, Béatrice Gerland, and John D Sutherland. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, 459(7244):239–42, May 2009.

[4] Everett L Shock. Stability of peptides in high-temperature aqueous solutions, 1992.

[5] R Bruce Martin. Free Energies and Equilibria of Peptide Bond Hydrolysis. *Biopolymers*, 45:351–353, 1998.

[6] M. PAECHT-HOROWITZ, J. BERGER, and A. KATCHALSKY. Prebiotic Synthesis of Polypeptides by Heterogeneous Polycondensation of Amino-acid Adenylates. *Nature*, 228(5272):636–639, November 1970.
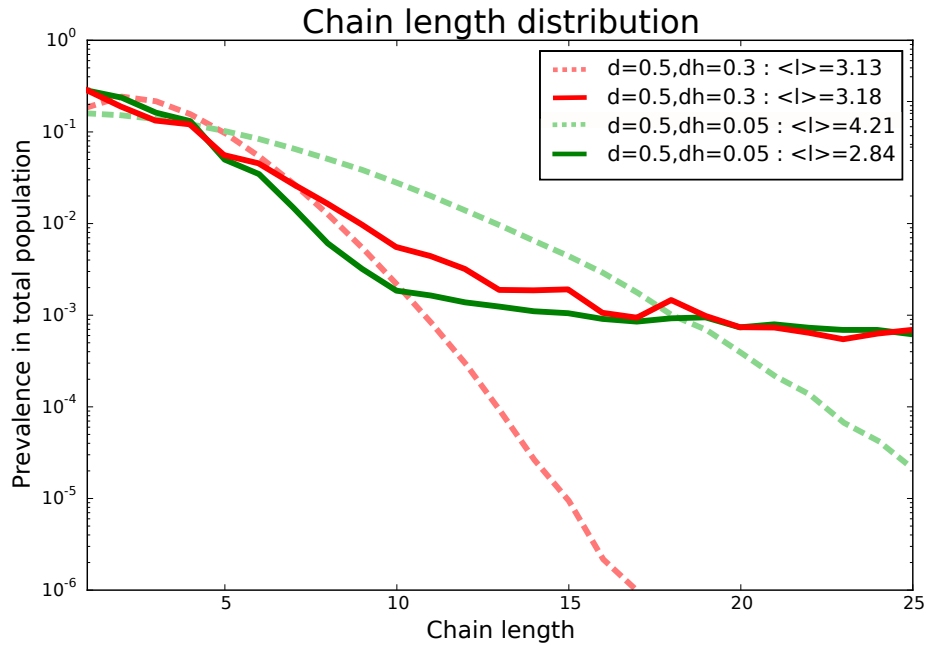
Fig. 6: Dashed lines represent polymerization without folding or catalysis. Solid lines correspond to simulations run with folding and catalysis reactions on. For details of simulations see 4.1, Experiment 3.
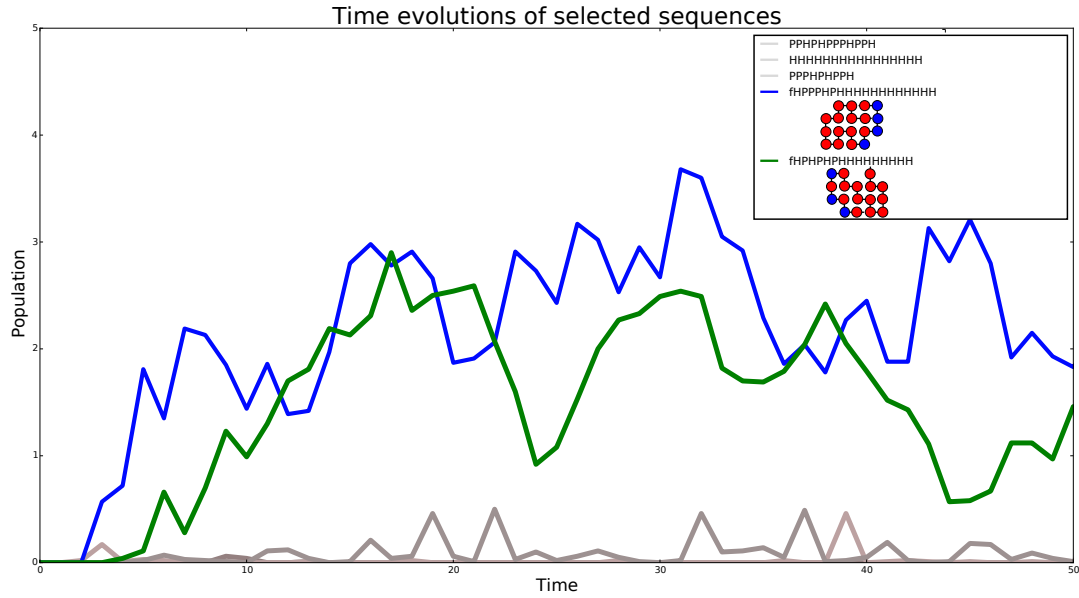

Fig. 7: Some examples of dominating autocatalytic sequences. Gray lines represent regular non-catalytic sequence. structures on the right are native structures of autocatalytic sequences.

[7] Jean-François Lambert. Adsorption and polymerization of amino acids on mineral surfaces: a review. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life*, 38(3):211–42, June 2008.

[8] Luke Leman, Leslie E Orgel, and M Reza Ghadiri. Carbonyl sulfide-mediated prebiotic formation of peptides. *Science (New York, N.Y.)*, 306(5694):283–6, October 2004.

[9] Leslie E Orgel. Prebiotic chemistry and the origin of the RNA world. *Critical reviews in biochemistry and molecular biology*, 39(2):99–123, 2004.

[10] J P Ferris, A R Hill, R Liu, and Leslie E Orgel. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, 381(6577):59–61, May 1996.

[11] Julien Derr, Michael L Manapat, Sudha Rajamani, Kevin Leu, Ramon Xulvi-Brunet, Isaac Joseph, Martin A Nowak, and Irene A Chen. Prebiotically plausible mechanisms increase compositional diversity of nucleic acid sequences. *Nucleic acids research*, 40(10):4711–22, May 2012.

[12] Meng Wu and Paul G Higgs. Origin of self-replicating biopolymers: autocatalytic feedback can jump-start the RNA world. *Journal of molecular evolution*, 69(5):541–54, November 2009.

[13] G.F. Joyce. Nonenzymatic Template-directed Synthesis of Informational Macromolecules. *Cold Spring Harbor Symposia on Quantitative Biology*, 52:41–51, January 1987.

[14] David L Abel and Jack T Trevors. No Title. *Theoretical Biology and Medical Modelling*, 2(1):29, 2005.

[15] Pier Luigi Luisi, Peter Walde, and Thomas Oberholzer. Lipid vesicles as possible intermediates in the origin of life. *Current Opinion in Colloid & Interface Science*, 4(1):33–39, February 1999.

[16] Todd P. Silverstein. The Real Reason Why Oil and Water Don't Mix. *Journal of Chemical Education*, 75(1):116, January 1998.

[17] Kit Fun Lau and Ken A Dill. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules*, 22(10):3986–3997, October 1989.

[18] Hue Sun Chan and Ken A Dill. Sequence space soup of proteins and copolymers. *The Journal of Chemical Physics*, 95(5):3775, 1991.

[19] D W Miller and Ken A Dill. A statistical mechanical model for hydrogen exchange in globular proteins. *Protein science : a publication of the Protein Society*, 4(9):1860–73, October 1995.

[20] K Yue and Ken A Dill. Forces of tertiary structural organization in globular proteins. *Proc Natl Acad Sci U S A*, 92(1):146–150, 1995.

[21] RICHA AGARWALA, SERAFIM BATZOGLOU, VLADO DANČÍK, SCOTT E. DE-CATUR, SRIDHAR HANNENHALLI, MARTIN FARACH, S. MUTHUKRISHNAN, and STEVEN SKIENA. Local Rules for Protein Folding on a Triangular Lattice and Generalized Hydrophobicity in the HP Model. *Journal of Computational Biology*, 4(3):275–296, January 1997.

[22] Ken A. Dill, Sarina Bromberg, Kaizhi Yue, Hue Sun Chan, Klaus M. Ftebig, David P. Yee, and Paul D. Thomas. Principles of protein folding - A perspective from simple exact models. *Protein Science*, 4(4):561–602, December 2008.