HP world paper: March 19, 2015 – August 23, 2015

Contents

1 Introduction . . . 1
2 Flory problem . . . 2
3 We use hydrophobic effect to solve Flory problem . . . 4
3.1 We use HP model to represent prebiotic polymerization . . . 5
3.2 HP-foldamers can work as prebiotic catalysts . . . 5
4 Materials and methods . . . 7
4.1 Simulations . . . 7
5 Results . . . 9
6 Discussion . . . 10
6.1 Autocatalysis . . . 10
6.2 2D-3D . . . 11
A The model. Details. . . . 12
A.1 Kinetics of the simple model . . . 12
A.1.1 Steady State Kinetics . . . 12

COLOR CODE

Blue: **reformulate**

Red: **comments, todo**

1 Introduction

Functioning of modern organisms is not possible without nucleic acids and proteins. Cells produce these extremely long polymers with complex 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 whether one's approach is information first or metabolism first. 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 [4–10] is still in the future. One of the problems is that both RNA and peptides favor hydrolysis over polymerization in aqueous solutions, in other words in the prebiotic soup. Both polymers wouldn't survive if exposed to solution, especially at high temperature[11].

Several mechanism were proposed to increase yields of oligomers: adsorption to clays[7, 12], minerals[10, 13], evaporation of tidal pools[14], concentration in ice through eutectic melts [15], freezing conditions[16] or temperature cycles.

However the main problem with spontaneous polymerization processes is that it 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. The proof for polymer concatenation is present in [17]. This dynamics can be observed even in some autocatalytic systems [18].

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 [19, 20]

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. Our mechanism doesn't have to be the only one. It can be successfully applied if there another accelerating processes such as adsorption or wetting and drying cycles, for example. write about autocatalytic sets here and cite Kauffman, maybe also mention far from equilibrium statement. mention composomes, mention lipid vesicles [21**?**] 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

Production of long oligo-peptides and oligo-nucleotides has focused much interests[10, 15, 22–24].

Results vary. Maximum length of biopolymers reached (both nucleic acids and peptides) in the lab during non-enzymatic synthesis range from 2-3mers to 55[1] monomers long depending on conditions: Ferris et al. max 55, average 30[10], Rode et al. 2-3 mers [25], max 11, average 4 [15], [8].

In [15] nucleotides polymerization was considered. Authors used below zero temperature to enhance polymerization. They also provided data for other experiments with nucleotide polymerization. See fig 1

The problem is that in the systems with spontaneous polymerization distribution of chain lengths of the produced polymers has an exponential asymptote (see fig. 2). Depending on

[1] primary product contains 20-40-mers, mean chain length is expected to be 30[10]

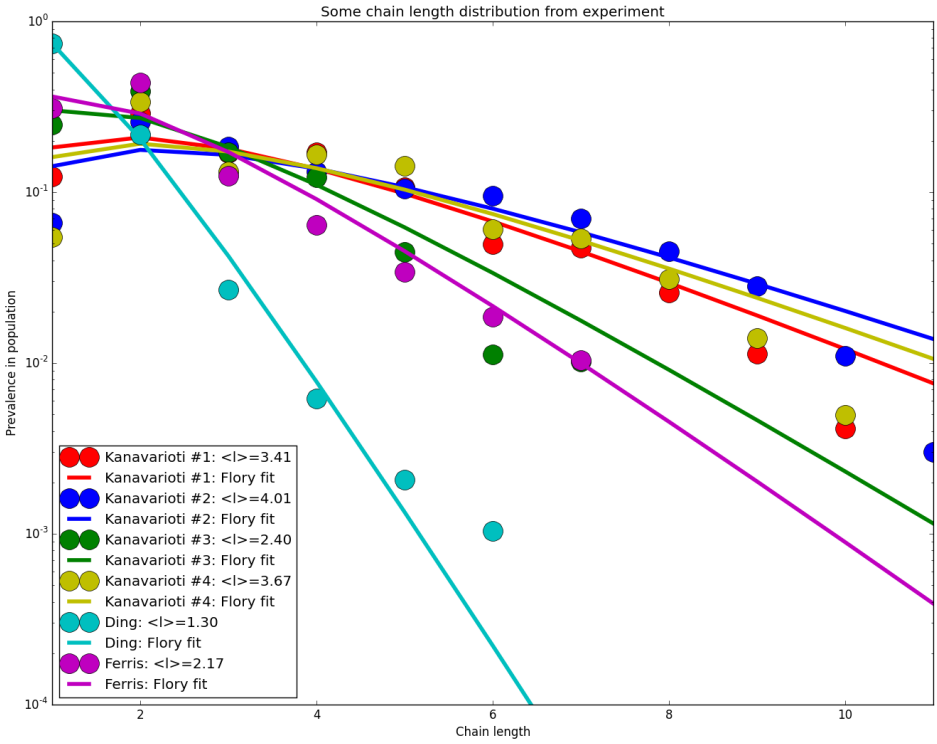

Fig. 1: Experimental relsults from [15] and their fitting to Flory distribution. **I left too much information here intentionally. We can decide, what to leave during the meeting. I want to Show that at least some experimental results match Flory distribution at longer chains: dashed – experiment, solid – fit.**

the actual system it can can follow for example actual exponential law $f(a) \propto a^l$[17, 26] or a Flory distribution $f(a) = a^2 l(1-a)^{l-1}$[27], where a is an empirical constant and l is a chain length. In either case, given low prebiotic concentrations of amino acids and nucleotides with estimations of submillimolar to submicromolar[15, 28, 29], longer chains would not be observable.

Figure 2 represents Flory distribution with different average chain length. This distribution gives the following polymer abundance for the blue line:

$$\frac{[10mers]}{[1mers]} \propto 10^{-4}, \qquad \frac{[20mers]}{[1mers]} \propto 10^{-9} \tag{1}$$

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.

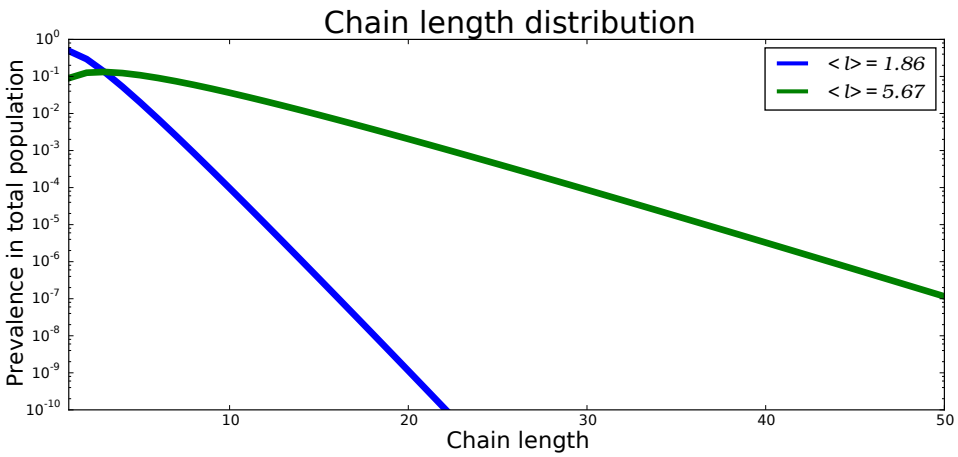

Fig. 2: Spontaneous polymerization gives Flory length distribution with exponential asymptote.

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. [30]. When hydrophobic molecules are placed in the water, water molecules form cage-like structures around them (see fig. 3). 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.

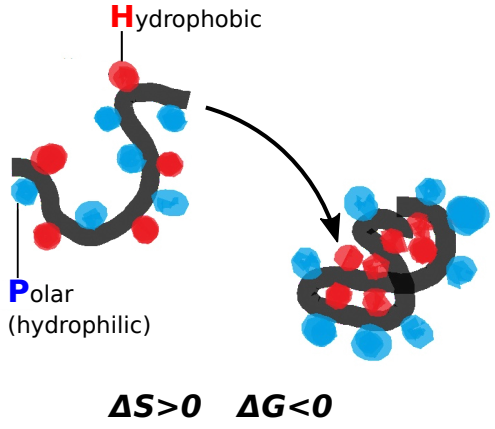

Fig. 3: Hydrophobic effect is an entropic effect originating from the disruption of hydrogen bonds between water molecules by the non-polar solute. When hydrophobic molecules are placed in the water, water molecules form cage-like structures around them. When hydrophobes come into contact, water molecules 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[31–35]. 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 conformations 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 considered being capable of folding (see fig. 4).

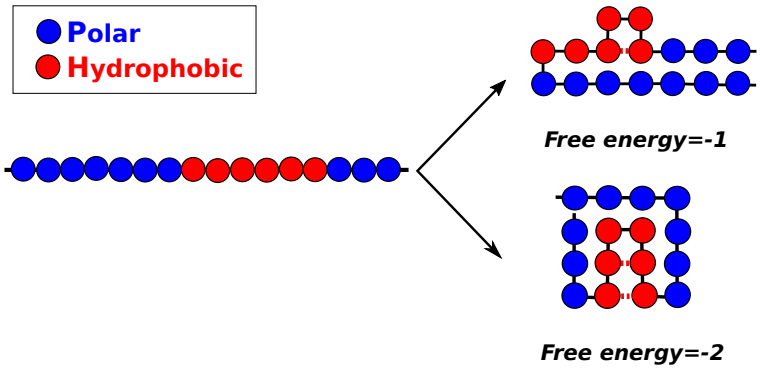

Fig. 4:

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[31].

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. 5).

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.5(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.5(b)). Of course, this is not a good rate enhancement, compared to $2 \cdot 10^7$-fold rate enhancement brought about by modern ribosomes[36]. 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.

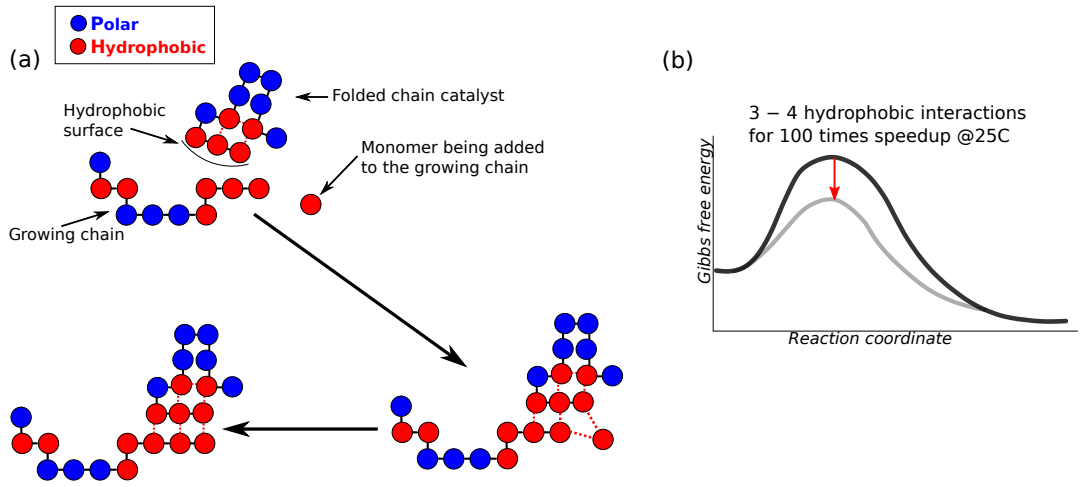

Fig. 5: Catalyst catalyzes a growing of an unfolded hp-polymer. Having just 3-4 hydrophobic contacts is enough to lower an activation barrier for $\propto 100$ times at room temperature.

4 Materials and methods

4.1 Simulations

To test our hypothesis we performed direct stochastic simulations on several. We used PDMmod method [**?**] 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[31, 37] [2]), so we had to limit maximum chain lengths to 25. We also try to keep total 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 either can mimic a protocell splitting and loose of materials due to it or in the case when system isn't bounded by any borders tha fact that some molecules will diffuse away. 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. Without loss of generality we can put this parameter equal 1; all other rates with be relative to the growth rate in this case.
- Hydrolysis has constant rate d_h per bond. Half-life time of hydrolysis bonds in neutral conditions and temperatures around room temperature are on the order of hundreds of years[3]. We test hydrolysis rate constants to be about $0.001 - 1$ of polymerization rate constants. This way we account for polymerization conditions, which happens on the order of days to years.
- We also import monomers into system with rate $a \gg 1$. It is safe to assume that we would have enough monomers in the system and import of monomers wouldn't be a bottleneck of reactions chain. Therefore we explore big values of $a \propto 10^2, 10^3\alpha$
- Dilution parameter d mimics cell division and loss of the matter because of that. From A.1.1 we see that total mass of the system is $M \propto \frac{a}{\alpha}$, $d \approx \alpha$ or $d \gg \alpha$ and $M \propto \frac{a}{\alpha}\frac{d}{2\alpha}$, $d \ll \alpha$. Therefore we explore valued of d from $\propto 0.01\alpha$ to $\propto 1\alpha$. Given values of a we'll explore various populations from $\propto 10^2$ to $\propto 10^5$ monomers per cell.
- (Fix this after discussion) Folding and unfolding reactions happen very quickly with the unfolding rate constants of $k_{unf} \gg \alpha$ and folding rate constant of $k_{unf} \cdot \exp(E_{native}/kT)$.

[2] Python implementation and description can be found here: http://hp-lattice.readthedocs.org/en/latest/

[3] Hydrolysis rate constants of oligopeptides in neutral conditions are of the order of $10^{-11} - 10^{-10}$: $1.310^{-10}M^{-1}s^{-1}$ for benzoylglycylphenylalanine ($t_{1/2} = 128y$)[38], $6.310^{-11}M^{-1}s^{-1}$ ($t_{1/2} = 350y$) for glycylglycine and $9.310^{-11}M^{-1}s^{-1}$ for glycylvaline [39].

E_h in our experiments is around $1-2$kT[]. k_{unf} we keep $\propto 10^2$, which gives us range of unfolding rates from a reaction per hours and days and range of folding rates from a reaction per hours to fractions of a second.

- Catalysis rate is proportional to the exponent of hydrophobic energy E_h and number of contacting hydrophobes n_c: $\alpha \cdot \exp(E_h \cdot n_c/kT)$. Number of hydrophobic contacts for the short HP-sequences is about $3-6$. With the hydrophobic energies of $1-2$kT this gives us catalysis rates around hours and days for one reaction.

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.

Constant name	Symbol	Normalized simulation value	Simulation value (deduced) per $1M$	Value from literature, per $1M$
Polymerization rate constant	α	1	$\propto 1\,month^{-1}$	??
Hydrolysis rate constant	d_h	$\propto 10^{-1}$–10^{-4}	$\propto 1\,month^{-1} - -10^{-3}year^{-1}$	$\propto 10^{-3}year^{-1}$ [24, 38, 39]
Dilution rate constant	d	$\propto 10^{-2}-1$	$\propto 0.1year^{-1} - -10^{-3}year^{-1}$	— Is used to keep model from overflowing
Monomer import rate constant	a	$\propto 10^2-10^3$	$\propto 1-10^2day^{-1}$	??
Number of rotational freedoms	z	$1.5-2.5$	$1.5-2.5$	??
Hydrophobic energy per kT	e_h	$1-2$	$1-2$	$0-3.3$ [41]

Tab. 1: !

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 6 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 $E_h = 1 - 3kT$

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 catalysis**??**. 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.

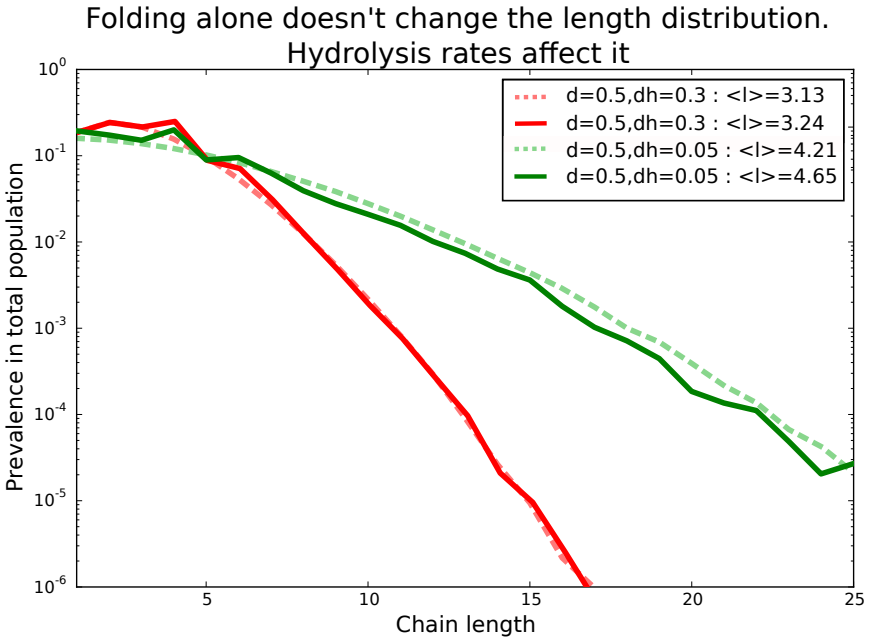

Fig. 6: Dashed lines represent polymerization without folding or catalysis. Solid lines correspond to simulations run with folding but without catalysis. For details of simulations see section 4.1, Experiment 2.

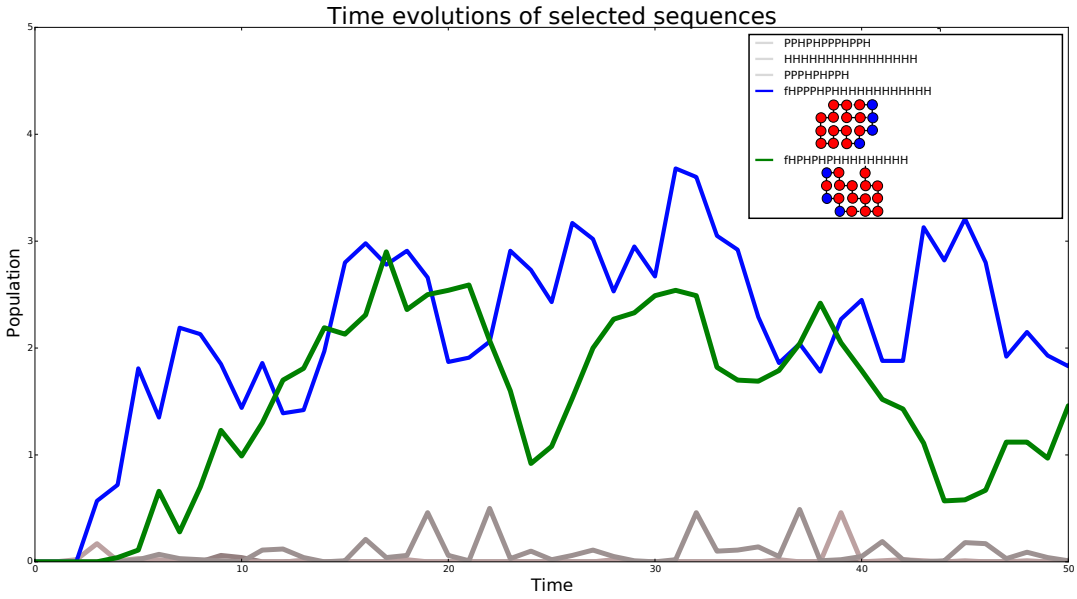

Fig. 7: Some examples of dominating autocatalytic sequences. Gray lines represent regular non-catalytic sequence. structures on the right are native structures of autocatalytic sequences.

6 Discussion

6.1 Autocatalysis

In our analysis of a polymerization prebiotic systems we applied a physical principle of hydrophobic interactions to an old idea of autocatalytic sets [42]. In one or another form

the idea of autocatalysis was applied to various systems []. They key difference with our work is that hydrophobic interaction provides a simple physical set up which produces non-linear dynamics with complex feedback. This enables system to develop a non-trivial selection mechanism.

6.2 2D-3D

- Folding and unfolding rates are likely underestimated in 2D case
- Overall length dependence is stepper in 2D case
- However cases are very similar and it's possible to do a mapping between them: there's a direct mapping between surface to volume ration in 3D case to perimeter to area ratio in 2D case

A The model. Details.

A.1 Kinetics of the simple model

This model was presented and studied thoroughly in [26, 43, 44]

We enumerate all the polymers, so that x_i is population of i^{th} monomer, and $x_{i'}$ is a population of its precursor.

Equations are:

$$\text{One mers:} \quad \dot{x_i} = a - 2\alpha x_i - dx_i \tag{2}$$

$$\text{2+ mers:} \quad \dot{x_i} = \alpha x_{i'} - (2\alpha + d)x_i \tag{3}$$

A.1.1 Steady State Kinetics

Steady state: $\dot{x_i} = 0$

$$\text{One mers:} \quad 0 = a - 2\alpha x_i - dx_i \tag{4}$$

$$\text{2+ mers:} \quad 0 = \alpha x_{i'} - (2\alpha + d)x_i \tag{5}$$

So we have:

$$\text{One mers:} \quad x_i = \frac{a}{2\alpha + d} \tag{6}$$

$$\text{2+ mers:} \quad x_i = \frac{\alpha}{2\alpha + d} x_{i'} \tag{7}$$

Therefore for every sequence of length l we get:

$$\boxed{x_l = \frac{a}{\alpha}\left(\frac{\alpha}{2\alpha + d}\right)^l} \tag{8}$$

Population of all the sequence of length l is therefore:

$$p_l = \frac{a}{\alpha}\left(\frac{\alpha}{2\alpha + d}\right)^l 2^l = \frac{a}{\alpha}\left(\frac{2\alpha}{2\alpha + d}\right)^l = \frac{a}{\alpha}\left(\frac{1}{1 + d/2\alpha}\right)^l \tag{9}$$

If we denote $x \equiv \frac{d}{2\alpha}$, population of all the sequences of length l will be:

$$\boxed{p_l = \frac{a}{\alpha}\left(\frac{1}{1 + x}\right)^l} \tag{10}$$

Total mass of all the sequences is:

$$M = \sum_{l=0}^{\infty} l p_l \tag{11}$$

$$M = \sum_{l=0}^{\infty} \frac{a}{\alpha} l \left(\frac{1}{1 + x}\right)^l \tag{12}$$

According to [45] the sum will be

$$M = \frac{a}{\alpha} \frac{\frac{1}{1+x}}{\left(1 - \frac{1}{1+x}\right)^2} = \frac{a}{\alpha}\left(\frac{1+x}{x}\right) \tag{13}$$

Therefore total mass is:

$$M = \frac{a}{\alpha}\left(1 + \frac{1}{x}\right) \tag{14}$$

Remember that $x = d/2\alpha$. It means that values of d $d \approx \alpha$ or $d \gg \alpha$ produce total masses

$$M \propto \frac{a}{\alpha}, \qquad d \approx \alpha \quad \text{or} \quad d \gg \alpha \tag{15}$$

while very small values of $d:\ d \ll \alpha$ produce total masses

$$M \propto \frac{a}{\alpha}\frac{d}{2\alpha}, \qquad d \ll \alpha \tag{16}$$

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.

[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] N. R. Pace. Origin of life - Facing up to the physical setting. *Cell*, 65(4):531–533, 1991.

[12] M. Rao, D. G. Odom, and J. Oró. Clays in prebiological chemistry. *Journal of Molecular Evolution*, 15(4):317–331, December 1980.

[13] J D Bernal. The Physical Basis of Life. *Proceedings of the Physical Society. Section B*, 62(10):597–618, October 1949.

[14] Kevin E. Nelson, Michael P. Robertson, Matthew Levy, and Stanley L. Miller. Concentration by Evaporation and the Prebiotic Synthesis of Cytosine. *Origins of Life and Evolution of the Biosphere*, 31(3):221–229, 2001.

[15] Anastassia Kanavarioti, Pierre-Alain Monnard, and David W. Deamer. Eutectic Phases in Ice Facilitate Nonenzymatic Nucleic Acid Synthesis. *Astrobiology*, 1(3):271–281, September 2001.

[16] Jeffrey L. Bada. How life began on Earth: a status report. *Earth and Planetary Science Letters*, 226(1-2):1–15, September 2004.

[17] 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.

[18] 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.

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

[20] David L Abel and Jack T Trevors. Three subsets of sequence complexity and their relevance to biopolymeric information. *Theoretical Biology and Medical Modelling*, 2(1):29, 2005.

[21] 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.

[22] Robert Shapiro. The improbability of prebiotic nucleic acid synthesis. *Origins of Life*, 14(1-4):565–570, 1984.

[23] André Brack. From interstellar amino acids to prebiotic catalytic peptides: A review. *Chemistry and Biodiversity*, 4(4):665–679, 2007.

[24] Grégoire Danger, Raphaël Plasson, and Robert Pascal. Pathways for the formation and evolution of peptides in prebiotic environments. *Chemical Society reviews*, 41(16):5416–29, August 2012.

[25] Bernd Michael Rode. Peptides and the origin of life. *Peptides*, 20(6):773–786, 1999.

[26] Martin A Nowak and Hisashi Ohtsuki. Prevolutionary dynamics and the origin of evolution. *Proceedings of the National Academy of Sciences*, 105(39):14924–14927, September 2008.

[27] Paul J Flory. *Principles of polymer chemistry*. Ithaca, NY : Cornell Univ., 1953.

[28] a. D. Aubrey, H. J. Cleaves, and Jeffrey L. Bada. The role of submarine hydrothermal systems in the synthesis of amino acids. *Origins of Life and Evolution of Biospheres*, 39(2):91–108, 2009.

[29] Antonio Lazcano and Stanley L Miller. The Origin and Early Evolution of Life: Prebiotic Chemistry, the Pre-RNA World, and Time. *Cell*, 85(6):793–798, June 1996.

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

[31] 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.

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

[33] 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.

[34] 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.

[35] RICHA AGARWALA, SERAFIM BATZOGLOU, VLADO DANČÍK, SCOTT E. DECATUR, 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.

[36] A. Sievers, M. Beringer, M. V. Rodnina, and R. Wolfenden. The ribosome as an entropy trap. *Proceedings of the National Academy of Sciences*, 101(21):7897–7901, May 2004.

[37] 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.

[38] Rebecca A R Bryant and David E Hansen. Direct Measurement of the Uncatalyzed Rate of Hydrolysis of a Peptide Bond. *Journal of the American Chemical Society*, 118(23):5498–5499, January 1996.

[39] Robert M. Smith and David E. Hansen. The pH-Rate Profile for the Hydrolysis of a Peptide Bond. *Journal of the American Chemical Society*, 120(35):8910–8913, September 1998.

[40]

[41] William C. Wimley and Stephen H. White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural Biology*, 3(10):842–848, October 1996.

[42] Stuart A Kauffman. Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119(1):1–24, March 1986.

[43] Hisashi Ohtsuki and Martin A Nowak. Prelife catalysts and replicators. *Proceedings. Biological sciences / The Royal Society*, 276(1674):3783–90, November 2009.

[44] Irene A Chen and Martin A Nowak. From Prelife to Life: How Chemical Kinetics Become Evolutionary Dynamics. *Accounts of chemical research*, 45(12), February 2012.

[45] I.S. Gradstein and I.M. Ryzhik. *Table of Integrals, Series, and Products*. New York, New York, USA, 1980.