Simple Evolution of Complex Crystal Species

Rebecca Schulman[1] and Erik Winfree[2]

[1] University of California Berkeley, Berkeley, CA 94720, USA
`rschulman@berkeley.edu`
[2] California Institute of Technology, Pasadena, CA 91125, USA
`winfree@caltech.edu`

Abstract. Cairns-Smith has proposed that life began as structural patterns in clays that self-replicated during cycles of crystal growth and fragmentation. Complex, evolved crystal forms could then have catalyzed the formation of a more advanced genetic material. A crucial weakness of this theory is that it is unclear how complex crystals might arise through Darwinian selection. Here we investigate whether complex crystal patterns could evolve using a model system for crystal growth, DNA tile crystals, that is amenable to both theoretical and experimental inquiry. It was previously shown that in principle, the evolution of crystals assembled from a set of thousands of DNA tiles under very specific environmental conditions could produce arbitrarily complex patterns. Here we show that evolution driven only by the dearth of one monomer type could produce complex crystals from just 12 monomer types. The proposed mechanism of evolution is simple enough to test experimentally and is sufficiently general that it may apply to other DNA tile crystals or even to natural crystals, suggesting that complex crystals could evolve from simple starting materials because of relative differences in concentrations of the materials needed for growth.

1 Introduction

A plausible hypothesis for the origin of life on Earth must explain both *spontaneous self-replication*, i.e. how a self-replicating system first emerged, and *open-ended evolution*, i.e. how Darwinian evolution of this system led to complex organisms. Spontaneous self-replication requires that the components of the genetic material and the environment for replication existed on the early Earth, ideally in abundance, and that the assembly of these components into a replicator could reasonably have occurred. Open-ended evolution requires at minimum that complex genomes exist which would be fitter than all simpler genomes under plausible environmental conditions. A major difficulty in origin of life research is that while many hypotheses can explain the process by which either spontaneous self-replication or open-ended evolution might have occurred, no one hypothesis yet gives a detailed picture of the sequence of events leading to both.

There are many simple systems capable of replication that may have existed on the early Earth, including fire, autocatalytic reaction cycles [31] or lipid vesicles [32,30] for which it is not clear how Darwinian evolution might produce

complexity. Similarly, there are complex systems, such as DNA combined with enzymes [21], ribozymes with specific oligonucleotide substrates [18], prions [17], DNA crystals [26], robots [16,14] or computer programs [1] for which complex evolution seems feasible, but how they might have arisen spontaneously is not clear.

Graham Cairns-Smith has proposed that polytypic clay crystals, which may have been common on the early Earth, could have been the first replicators [5,7]. If each layer within a clay crystal could contain monomers in any of several distinct arrangements (called layer types), then the sequence of layer types in a stack of layers would contain information. Crystal growth would propagate this information and crystals' occasional breaking would produce new growth fronts for its propagation, thereby amplifying it (Figure 1). Particular clay crystal sequences might promote catalytic functions that would be selected for, and eventually these crystal sequences would evolve into more complex sequences to promote more complex chemical functions, until the chemistry induced by the crystal's sequence became capable of self-replication on its own. Cairns-Smith called this scenario a "genetic takeover" [6].

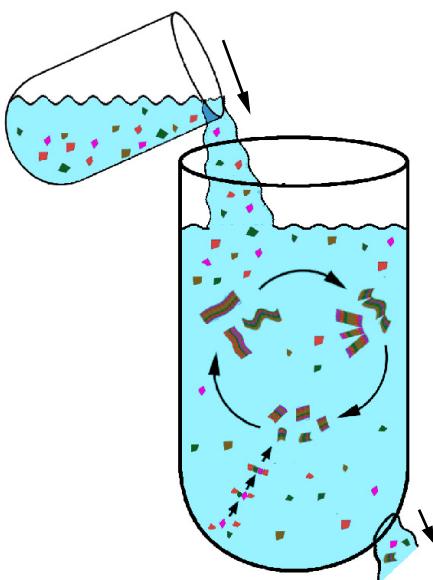To date no one has demonstrated clay crystal information replication [4] and there is no direct experimental evidence that particular clay sequences have properties that might make them particularly efficient replicators [8]. However, one reason for its continued discussion [22] is that both prebiotic replication of clay patterns and the idea that some complex clay structures could have selective advantages seem reasonable. Since there has been only limited investigation into either of these questions, it is worth investigating whether clay might be capable of spontaneous self-replication and open-ended evolution.

Fig. 1. Cairns-Smith's theory of crystal replication. Heterogeneous crystals propagate information consisting of the arrangement of monomers (shown as stripes) during growth. Crystal fragmentation creates new growth fronts to propagate the information. Monomers are replenished, and nucleation occasionally produces new crystals.

However, it is unclear more generally how *any* complex crystal could arise through Darwinian selection. Here we consider whether there are features of crystal growth dynamics, rather than of chemistry particular to clay crystals, that might result in the evolution of complex crystals. We use DNA tile crystals, a model system for investigating generic features of 2 dimensional crystal growth [26]. DNA tile crystals can be readily studied experimentally (e.g. [38]) and theoretically using quantitative models (e.g. [37]).

One might be skeptical that generic features of crystal growth alone could be responsible for complex crystal evolution: It is generally assumed that a complex sequence would evolve because the sequence imbues some chemical functionality or otherwise alters the environment in a way that improves its reproductive fitness. However, a complex crystal may be selected for simply because it uses more abundant raw materials in its growth than simpler crystals. The preference for the addition of monomers that form multiple contacts with an existing crystal can put complex constraints on the order and frequency with which a crystal uses available monomers. It has been argued that these constraints could cause the selection of complex crystals [28], but so far only for crystals containing thousands of monomer types and under special physical conditions that are of limited relevance to natural crystal growth.

We will show that similar constraints could produce complex evolution in DNA tile crystals containing just 12 monomer types and over a wide set of physical conditions. We use a cellular automaton model [35] to enumerate the set of crystal morphologies that could be produced by a given set of monomers (a *tile set*). We show that if the monomer compositions of the crystal morphologies produced by a given tile set satisfy a simple property, then in a simplified model of growth we would predict selection of complex crystals. Using a combinatorial search, we find tile sets that satisfy this property. We then use a kinetic simulation to show that for at least one tile set, this evolution also occurs in a more realistic crystal growth model. While our search focuses on crystals containing 12 monomer types, it is reasonable to suppose that even smaller tile sets can exhibit similar phenomena.

2   DNA Tile Crystal Replication and Evolution

DNA tile crystals consist of tile monomers [13], rigid assemblies of oligonucleotides with short, single-stranded DNA segments, *sticky ends*, by which they bind to other tile monomers. Tiles can assemble into rectangular lattices to produce two-dimensional crystals [38] and ribbons [19,27,40]. The ease of working with DNA tiles make them a model system for crystal growth: DNA crystal monomers (tiles) are easily created by designing the tiles' oligonucleotide sequences, and the growth pathways of DNA crystals are relatively well-understood at the monomer level [24,9,27].

An example set of DNA tile monomers, i.e. a tile set, is shown in Figure 2(a). The diagram describes the possible interactions between the tile types, which determine the crystal growth dynamics. Edges with the same color that fit together

like jigsaw puzzle pieces represent sticky ends with complementary sequences. The set consists of square *rule* tiles, and rectangular *top* and *bottom* edge tiles. These tiles can assemble to form a variety of patterned ribbon structures (Figure 2(d)). Close to the crystal melting temperature, the addition of a DNA tile to a crystal is energetically favorable only if it forms two or more bonds [36] (Figure 2(b)).
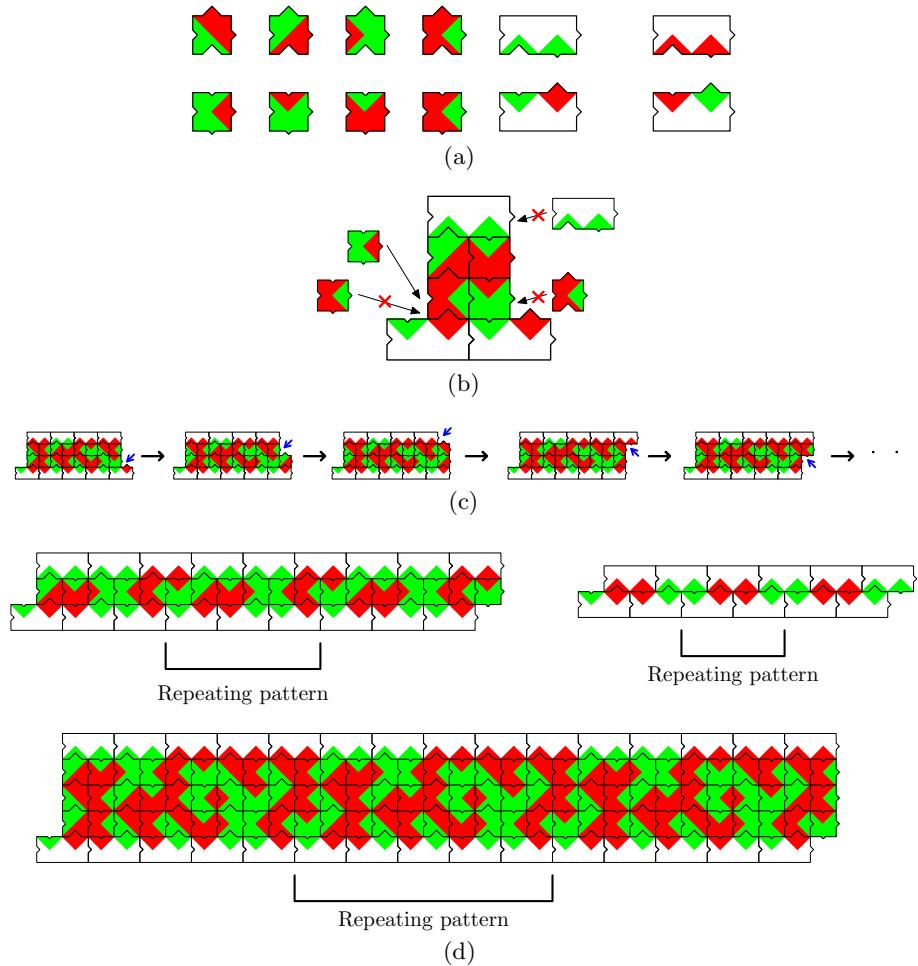

Fig. 2. An example zig-zag CA tile set. (a) The 12 tiles. Notches and colors represent the single-stranded sticky ends of each tile; interlocking edges of matching color have complementary sequences. Tiles cannot be rotated. **(b)** Close to the melting temperature tiles tend to attach to a crystal only where they match at least two edges. **(c)** Zig-zag crystal growth. At each step, a new tile may be added at the location designated by the small arrow. A simultaneous growth process also occurs on the crystal's left end. **(d)** Example assemblies of widths 2, 3 and 5 formed by the tiles in (a). The pattern repeated in each crystal is bracketed.

Under these conditions, for the tiles in Figure 2(a), a single column of tiles terminated by top and bottom edge tiles uniquely defines the preferred growth process and ribbon pattern (Figure 2(c)). Growth produces ribbons bearing repeating *wallpaper patterns* (Figure 2(d)). The wallpaper pattern particular to a crystal is its information, which would be replicated by the process Cairns-Smith proposed.

We are interested in whether evolution in this system might produce complex crystal forms. Crystal evolution occurs when (a) mistakes during crystal growth occur occasionally, producing mutations, and (b) some crystals replicate faster than others, i.e. are fitter. The fitness of a crystal is the geometric mean of crystal growth rate (in columns / time) and the per column rate of crystal fragmentation [28]. For a given tile set and crystal growth environment, we use mathematical models and simulations to estimate the growth and fragmentation rates of the crystals and thus their fitness.

A simplified version of the kTAM [25], a generalized crystal growth model applicable to DNA tile crystal growth [38,24,2] can be used to estimate crystal growth rates. We consider a version of this model in which (a) crystal growth proceeds exclusively by single tile addition (we ignore tile dissociation), (b) a tile may be added to a site if labels on at least two edges match those presented by the crystal at that site, and (c) monomer tiles arrive at potential binding sites with a frequency proportional to their concentration in solution. We assume that occasional violations of rule (b) produce mutations that introduce new crystal patterns into the population. Because we will show that growth rates can vary arbitrarily widely, we will for simplicity ignore the dependence of crystal width on fragmentation frequency and assume that fitness is proportional to crystal growth rate. However, under fluid shear, for example, breakage rates of crystals do decrease with crystal width [15].

Both the tile set, which is the alphabet that determines the types of ribbons that form, and physical conditions such as tile concentration determine the results of an evolutionary process. To ask whether crystal evolution could produce non-trivial genomes, we will examine the complexity of DNA tile crystals that are produced by an evolutionary process. For simplicity we measure complexity as a crystal's width. In Section 3 we develop a formalism that describes a family of tile sets we will consider. In Section 4 we then describe an environment, in terms of tile concentrations, where evolution of complex (wide) crystals could occur.

3   Binary Zig-Zag CA Tile Sets

To clearly enumerate the members of the family of tile sets we wish to consider as possible substrates for open-ended evolution, we define a cellular automaton variant, the *zig-zag cellular automaton* (or zig-zag CA). Computation on a zig-zag CA takes place on a lattice of defined width w and infinite length. The computation tape is a vertical column of width w, and the horizontal (lengthwise) axis of the lattice contains the computation history. Cell updates proceed alternately from the top edge to the bottom edge (downward), then from the
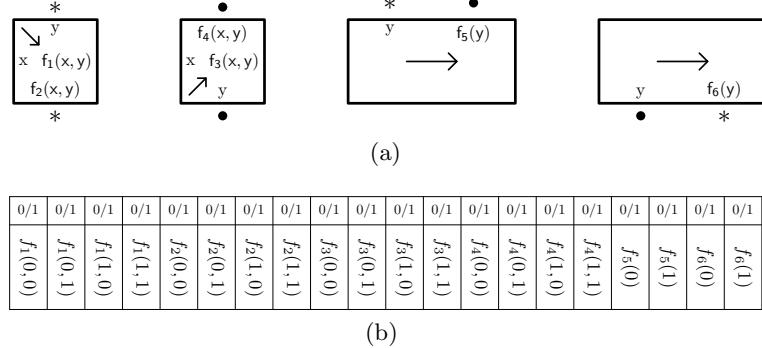
(a)

(b)

Fig. 3. Abstract representation of zig-zag CA tiles. (a) Tiles with each set of inputs x and y compute as growth proceeds down the ribbon (left) and an analogous group of tiles compute as growth proceeds up (second to left). As denoted by the dots and stars, the input and output alphabets are not interchangeable between upward and downward computing tiles. The two top and two bottom tiles each have one input and one output. In this paper we consider the case where $f_1...f_6$ have Boolean inputs and outputs. In other figures, green and red represent 0 and 1. **(b)** The table used to generate a unique numerical identifier for a binary zig-zag CA tile set. The values of each output of the functions is written in the order shown give the identifier as a binary number, from most significant bit (left) to least (right).

bottom edge to the top edge (upward). The input to each cell consists of two values from a fixed alphabet $\mathcal{A}$: a right value and either a top or bottom value. During updates proceeding downward, the output right value is a function f_1 ($\{\mathcal{A}, \mathcal{A}\} \to \mathcal{A}$) of the cell's input "right" value and of the cell above's input "bottom" value. The output bottom value is given by f_2 ($\{\mathcal{A}, \mathcal{A}\} \to \mathcal{A}$), which takes the same inputs. Correspondingly, during upward updates, two functions f_3 and f_4 that operate on the cell's input "right" value and the cell below's input "top" value to determine the new right and top values respectively.

When a downward updating process reaches the bottom edge, the last bottom value is used to determine the edge top value that the bottom-most cell will use on the first upward update. A fifth function, f_5 ($\mathcal{A} \to \mathcal{A}$) determines the new top value given the edge bottom value of the previous row. Likewise, f_6 ($\mathcal{A} \to \mathcal{A}$), determines the new bottom value from the edge top value at the end of an upward series of updates. Figure 3(a) shows how to construct a zig-zag CA tile set from the functions $f_1...f_6$. For $|\mathcal{A}| = 2$, the set of tiles encodes a *binary zig-zag CA*.

Because computation takes place on a finite-width lattice, there are only a finite number of tape states that are possible: 2^{w-1} on a tape of width w ($w-2$ cells have a right value and there is one top or bottom value). The tape's states, therefore, must repeat. If computation is logically irreversible, some states may be transient, i.e. on the path to a repeating cycle, but never themselves repeated.

(Details about the dynamics of finite cellular automata and illustrations of their state spaces, including cycles and transient states, are given in [39].) These repeating cycles are the wallpaper patterns in the zig-zag ribbon examples in Figure 2(d).

We will study the family of tile sets that implement binary zig-zag CA computation. Each tile set in the family contains 8 rule tiles (to encode the four possible binary inputs and their respective outputs in each of the up and down directions) and 4 edge tiles (two top tiles with the inputs 0 and 1 and their respective outputs and two bottom tiles with 0 and 1 inputs and their respective outputs), for a total of 12 tiles. Since a binary zig-zag CA tile set is defined by 4 two-input Boolean functions and 2 one-input Boolean functions (Figure 3(b)) and there are 16 two-input Boolean functions and 4 one-input Boolean functions, there are $16^4 \times 4^2 = 1048576$ binary zig-zag CA tile sets.

4 Predicting the Fitness of Zig-Zag CA Crystals

Growth of crystals is driven by diffusion of tiles into growth sites; the rate of tile attachment is thus proportional to tile concentration [20]. We consider a simple model of attachment rates where the time it takes a tile with concentration $[z]$ to attach is $\frac{1}{k_f[z]}$, where k_f is the forward rate constant of attachment [33], independent of tile type.

To examine how concentrations affect crystal growth rates and therefore crystal fitness, we consider the case where the rule tile concentrations are $[r]$, the concentrations of three of the four edge tiles (referred to as *common* edge tiles) are $[e]$, and the concentration of the fourth edge tile (referred to as the *rare* edge tile) is $[q]$. Let n be the number of rule tiles per column in the repeating pattern (cycle) of a crystal, c the number of common edge tiles used per cycle and u the number of rare edge tiles used per cycle. Since one edge tile is added per column, the average number of common edge tiles used per column is $\frac{c}{c+u}$ and the average number of rare edge tiles used per column is $\frac{u}{c+u}$. The average time to add a column of tiles, $\langle T \rangle$, is therefore:

$$\langle T \rangle = \frac{1}{k_f} \left(\frac{c}{(c+u)[e]} + \frac{u}{(c+u)[q]} + \frac{n}{[r]} \right). \tag{1}$$

When the concentration of all edge tiles is the same, i.e. $[q] = [e]$, $\langle T \rangle$ grows monotonically with the number of rule tiles in each column of the crystal. However, when $[q]$ is very small, the middle term dominates. A wider crystal with n_1 rows of rule tiles, c_1 common edge tiles, and u_1 rare edge tiles could grow faster than a thinner crystal with n_2, c_2, and u_2 rule, common edge and rare edge tiles, respectively, if $\langle T_1 \rangle < \langle T_2 \rangle$, i.e. the following equation is satisfied for $n_1 > n_2$:

$$(n_1 - n_2)\frac{1}{[r]} < \left(\frac{c_2}{c_2 + u_2} - \frac{c_1}{c_1 + u_1} \right) \frac{1}{[e]} + \left(\frac{u_2}{c_2 + u_2} - \frac{u_1}{c_1 + u_1} \right) \frac{1}{[q]}. \tag{2}$$

Thus when $\frac{1}{[q]} \gg \frac{1}{[e]}$, a wider crystal can grow more quickly than thinner ones if it uses particularly few rare edge tiles even though it must add more rule tiles per column. Further, for any case where $\frac{u_1}{c_1+u_1} < \frac{u_2}{c_2+u_2}$ it is possible to find an environment (in terms of $[r]$, $[e]$, and $[q]$) where the wider crystal grows more quickly.

5   Evolution of Binary Zig-Zag CA Crystals

We sought to determine whether any binary zig-zag CA tile sets had the property that $\langle T \rangle$, the average time to add a crystal row, could decrease with width when one edge tile type was rare. For each tile set, we enumerated the patterns formed of particular widths. We then tabulated the frequency with which each pattern used top edge tiles with a "0" input vs top edge tiles with a "1" input, i.e. *top edge zeros* and *top edge ones*. Likewise, we tabulated the frequency that each pattern used *bottom edge zeros* and *bottom edge ones*.

As shown in Section 4, if the rare edge tile is present at a sufficiently low concentration, a pattern using few rare edge tiles per column is selected for. If a pattern of width w uses no rare edge tiles, no matter what the concentration of the rare edge tile, no pattern of width larger than w would be selected for simply because $[q]$ is low. In order to search for tile sets in which open-ended evolution might be feasible, therefore, we specifically searched for tile sets where no pattern could ever eliminate the rare edge tile, but where progressively wider patterns used the rare edge tile progressively less frequently than thinner patterns. We call these *evolvable tile sets*.

We first surveyed the edge tile usage of patterns of widths 2 to 12 of 4 randomly chosen tile sets (Figure 4(a)). In the first example (tile set 962191), all patterns either use only top edge zeros or top edge ones. Thus, if either tile type were rare, the thinnest type of crystal that used only the other top edge tile type would be the fittest. The same is true for bottom edge ones and zeros. No concentrations of edge tiles would be expected to induce the evolution of crystals wider than these crystals. The concentration of edge tiles used by the crystals in the other 3 examples follow the same sort of pattern. Since we did not observe that any of the 4 randomly chosen tile sets have the capacity for open-ended evolution as a result of a edge tile concentration differences, it is unlikely that most binary zig-zag CA tile sets are evolvable.

To find evolvable binary zig-zag CA tile sets, we surveyed all tile sets and identified those for which there was at least one edge tile that when rare, produced a fitter crystal with each increase in width from 2 through 7 tiles. This search produced 6144 putative evolvable tile sets. The edge tile usage of all patterns up to width 12 produced by a selection of these tile sets is shown in Figures 4(b-c).

The rate at which rare tile type usage decreases with width determines the shape of the fitness landscape. In most putatively evolvable tile sets that we examined, the fittest pattern of a given width has a per-row usage of the rare edge tile of roughly either $\frac{1}{w}$ or $\frac{1}{2^w}$. In each case, it appears that one rare tile is used in each repeat unit of the fit pattern, and the repeat units have lengths either linear or exponential in the width.
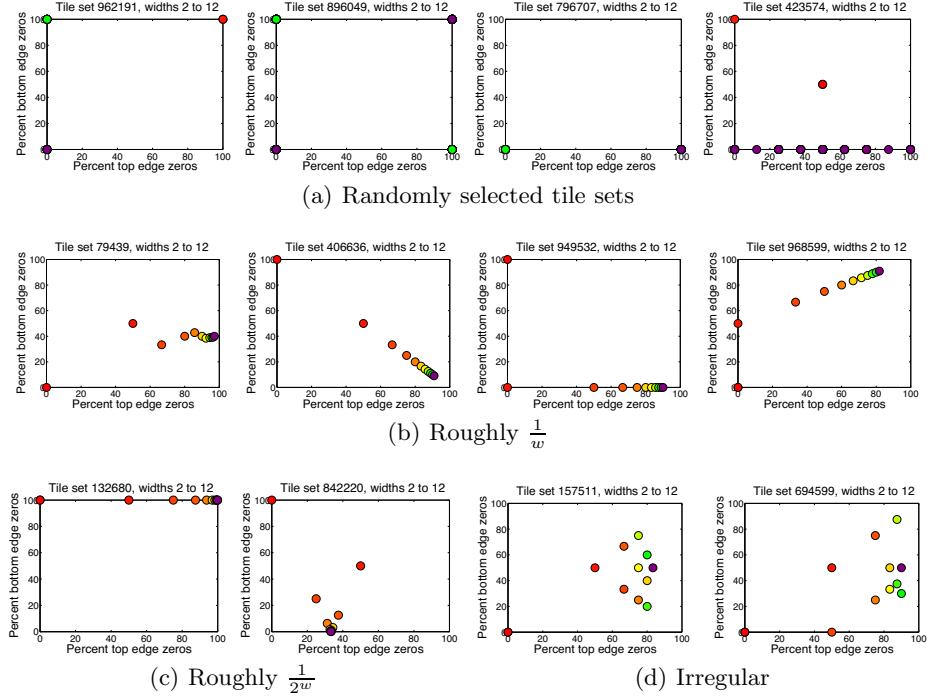
Fig. 4. Edge tile usage by zig-zag CA tile sets. The proportional usage of top and bottom zero and one tiles for each possible assembly of widths 2 to 12 for 12 representative tile sets. Each dot is one assembly type, and color indicates width, in rainbow order – red is width 2 and violet is width 12. The tile set number is computed as described in Figure 3(b) **(a)** 4 randomly selected tile sets. **(b-d)** Tile sets for which per-row usage of either or both top and bottom edge zero tiles decreases roughly as $\frac{1}{w}$ (b), $\frac{1}{2^w}$ (c), or irregularly (d) with width.

A second search of 100 random tile sets produced 7 tile sets for which fitter patterns appeared with some increases in width but not others. Figure 4(d) shows the usage patterns produced by two such tile sets. When the appropriate edge tile is rare, the first tile set produces a potentially fitter assembly type with assemblies of 1, 3, 5, 7, and 9 rows of rule tiles, and the second produces such new patterns with assemblies of 1, 2, 4, 5, and 10 tiles. Figure 5(c) shows some patterns produced by a rule of this type.

6   Evolution of Logically Reversible Zig-Zag CA Crystals

In the analysis in Sections 4 and 5, we assumed that the crystal's growth rate was the growth rate from its right end. However, in practice growth can occur from both the left and right ends of the crystals and the growth rate is the sum of the growth rates at each end.
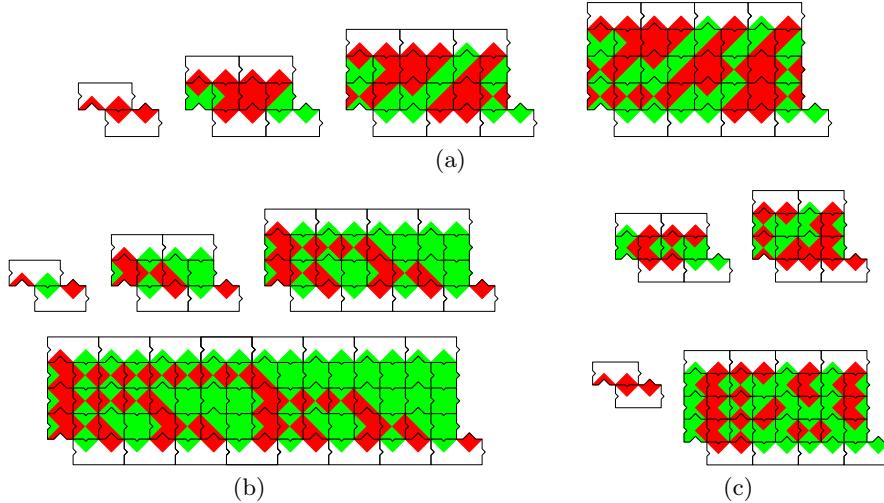
Fig. 5. Patterns of widths 2 to 5 of some representative evolvable tile sets.
(a) Patterns from tile set 968599. **(b)** Patterns from tile set 132680. Note that the patterns produced by the tiles count in binary [11]. **(c)** Patterns from tile set 694599.

Growth from the right end is equivalent to running the tile set's zig-zag CA forward. By definition there is exactly one tile that can attach at the growth site by two edges simultaneously (i.e. energetically favorably) and the attachment of this tile creates a new growth site where exactly one tile can attach favorably (Figure 2(c)). The rightward growth rate is therefore easy to approximate by computing the rates of attachment of each tile in this series.
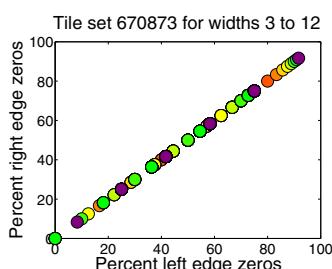

Fig. 6. Edge tile usage by a logically reversible zig-zag CA tile set. Format as in Figure 4.

Growth from the left end is equivalent to running the tile set's zig-zag CA backwards. Because a zig-zag CA is not guaranteed to be reversible, there may be no, one or multiple tiles that can fit at a given growth site. The dynamics of growth therefore can be more complex than in the rightward direction and we ignored these dynamics in Section 5. However, a subset of zig-zag CAs are logically reversible, i.e. there is always exactly one tile that can fit at the leftward growth site. A logically reversible zig-zag CA's left and right edges grow at the same speed. Our simple analysis of fitness landscapes based on growth rates is particularly accurate for these tile sets. Here we consider whether there are logically reversible zig-zag CA tile sets where a low concentration of an edge tile could drive evolution toward complex crystals, i.e. they are evolvable.

There are just 2304 types of logically reversible binary zig-zag cellular automata (an example being the tile set in Figure 2(a)). An exhaustive survey of these tile sets produced 16 for which each increase in width up to at least 12 could produce a new, fitter crystal (Figure 6).

For each of the 16 tile sets, the longest pattern of width w is $2w$ columns long, and with each increasing width the fittest pattern is the longest and uses just one rare edge tile.

Because the cellular automaton is logically reversible, there can be no transient states; every column is included in a repeating pattern. All patterns are have a short repeat length, so the number of patterns increases exponentially with width. As a result, crystal growth is susceptible to growth errors: most mistakes will change the pattern being copied. This is in contrast to the evolvable logically irreversible tile sets that we found in Section 5, which had only one pattern type for each width.

7   Kinetic Simulation of an Evolution Process

To determine whether the proposed selection pressure results in wider ribbons in a more realistic crystal growth model, we simulated crystal growth and fragmentation using a full model of kinetic tile assembly (kTAM [36]) with the software package xgrow [34]. We compared simulated evolutionary dynamics for two tile sets. According to the simple analysis in Section 4, evolution of crystals assembled from the tile set we simulated (shown in Figure 2(a)) is predicted to produce wider crystals when either top or bottom edge ones were rare. The second tile set differed from the first only in that the outputs of the top 2 tiles were swapped. This small change also changed the fitness landscape such that the fittest crystal is predicted to have no rule tile rows.

In the simulation, tiles reversibly attached to each other or to existing assemblies with a diffusion dependent forward rate ($k_f = 10^6/M/s$ [33]) and a backward rate $k_r = k_f e^{-\Delta G^\circ/RT}$ set by the ΔG of tile attachment, which was assumed to be strictly cooperative: $\Delta G^\circ = -17$ kcal/mol for an attachment by two bonds and $\Delta G^\circ = -8.5$ kcal/mol for attachment by one bond. The energy for two sticky end bond attachments was chosen in order to be close to experimental measurements [27]. The concentration of free tiles was held constant, and crystal shearing was initiated at a given tile with rate $\frac{1}{2500}$ seconds with probability approximately $e^{-l/4}$ where l is the length of a vertically or horizontally oriented shearing path. Rule tiles, top edge tiles and the bottom edge zero tile were present at 2.25 μM, and the bottom edge one tile was present at 0.025 μM. The simulation tracked the crystals in a volume of 10^{-14} liters, and serial dilution was modelled by removing each crystal from the simulation with probability 0.5 when the mixture reached 1000 crystals consisting of at least 2 monomers.

Figure 7 shows the rate at which assemblies of different widths arise in the two simulations. An energetic barrier to nucleation [29] meant that more than 99% of crystals originated as fragments sheared from another crystal rather than
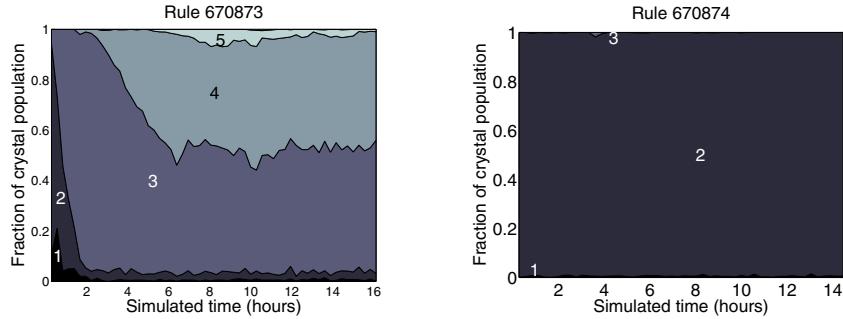
Fig. 7. Simulated evolution of zig-zag CA crystals. Numbers inside the graph indicate the width, including edge tiles. (a) Logically reversible tile set 670873, where the minimum possible usage of a rare edge tile increases with crystal width. (b) Logically reversible tile set 670874, where a crystal of width 2, which uses no rare edge tiles, is predicted to be fittest.

having spontaneously nucleated. The resulting dynamics show that wide crystals appear only for the tile set where this behavior was predicted by the analysis in Section 6. Equations 1 and 2 predict that for a fixed rare tile concentration, there should be an optimal width. Optimal widths for both tile sets in Figure 7 were observed.

8   Conclusions and Open Questions

This work suggests that the evolution of complex crystals from a simple set of monomers can be induced by differences in the monomer concentrations during crystal growth and replication. The components are simple in the sense that it would be feasible to synthesize their components and to watch their assembly and replication in the laboratory [3]. It will be important to determine how much the predictions made here hold up under a more realistic model of self-assembly. We must understand how growth errors and spontaneous nucleation of new crystals affect the evolutionary process. In any evolutionary process there is a bit-wise error rate (an error threshold) above which evolution becomes impossible [12]. There is reason to believe the tile sets that we study may be robust to many errors and therefore, that they can evolve even under imperfect assembly conditions. For example, in most of the logically irreversible cellular automata tile sets that we investigate, there is only 1 pattern that can be copied at a given crystal width. A mismatch error in such a tile set would simply change the segment of the pattern being copied at the growth front, rather than changing the pattern altogether.

Unfortunately, though, the highest acceptable bit-wise mutation rate and therefore the robustness of a sequence to mutation decreases with crystal width. Might we as a consequence expect a limit to the sizes of patterns that can be

copied under attainable physical conditions? Naively, the answer to this question seems to be yes. But it may be that for some tile sets where the number of patterns a tile set can copy grows sub-exponentially with width, the error rate for the whole pattern may not increase with width.

However these questions are answered, this work already suggests two important points. First, simple crystals are capable of complex evolution. More investigation is needed to determine whether natural crystals are capable of complex evolution; specific crystals such as clays will often have fewer monomer types, more complex dynamics, and a greater variety of specific and nonspecific interactions between monomer types. Since many of the details of these dynamics and affinities are still unknown in clay and other natural crystal systems, simpler DNA tile systems such as DNA nanotubes [23,40] may also be useful in further investigations.

Second, our analysis suggests a more important point about complex evolution in simple systems. While much attention has been given to the chemical functionality of sequences, it is generally assumed that the dynamics of assembling a sequence are of ancillary interest in a Darwinian evolution process. In biology, where sequence information is often stored in linear polymers such as genomic DNA and all monomers bind to their sequence neighbors using the same chemistry, this may be mostly true. But the constraints in crystal patterns make logic a fertile ground for complex evolution, and similar constraints are likely for other kinds of chemical replicators. In considering the replication in a more general class of chemical systems, therefore, it is worth considering the contribution of logical evolution to the dynamics of an evolutionary process.

References

1. Adami, C.: Introduction to Artifical Life. Springer, Berlin (1998)
2. Barish, R.D., Rothemund, P.W.K., Winfree, E.: Two computational primitives for algorithmic self-assembly: Copying and counting. Nano Letters 5, 2586–2592 (2005)
3. Barish, R.D., Schulman, R., Rothemund, P.W.K., Winfree, E.: An information-bearing seed for nucleating algorithmic self-assembly. Proceedings of the National Academy of Sciences USA 106(15), 6054–6059 (2009)
4. Bullard, T., Freudenthal, J., Avagyan, S., Kahr, B.: Test of Cairns-Smith's crystals-as-genes hypothesis. Faraday Discussions 136, 231–245 (2007)
5. Cairns-Smith, A.G.: The origin of life and the nature of the primitive gene. Journal of Theoretical Biology 10, 53–88 (1966)
6. Cairns-Smith, A.G.: Genetic Takeover and the Mineral Origins of Life. Cambridge University Press, Cambridge (1982)
7. Cairns-Smith, A.G.: The chemistry of materials for artificial Darwinian systems. International Reviews in Physical Chemistry 7, 209–250 (1988)
8. Cairns-Smith, A.G., Hartman, H.: Clay Minerals and the Origin of Life. Cambridge University Press, Cambridge (1986)
9. Chen, H.-L., Schulman, R., Goel, A., Winfree, E.: Reducing facet nucleation during algorithmic self-assembly. Nano Letters 7(9), 2912–2919 (2007)
10. Chen, J., Reif, J.H. (eds.): DNA 9. LNCS, vol. 2943. Springer, Heidelberg (2004)

11. Cook, M., Rothemund, P.W.K., Winfree, E.: Self-assembled circuit patterns. In: Chen and Reif [10], pp. 91–107

12. Eigen, M., McCaskill, J., Schuster, P.: Molecular quasi-species. Journal of Physical Chemistry 92, 6881–6891 (1988)

13. Fu, T.-J., Seeman, N.C.: DNA double-crossover molecules. Biochemistry 32, 3211–3220 (1993)

14. Griffith, S., Goldwater, D., Jacobson, J.M.: Self-replication from random parts. Nature 437, 636 (2005)

15. Hariadi, R.F., Yurke, B.: Elongational-flow-induced scission of DNA nanotubes in laminar flow. Physical Review E 82(4), 046307 (2010), http://pre.aps.org/abstract/PRE/v82/i4/e046307

16. Klavins, E.: Universal self-replication using graph grammars. In: 2004 International Conference on MEMS, NANO and Smart Systems (ICMENS 2004), pp. 198–204 (2004)

17. Li, J., Browning, S., Mahal, S.P., Oelschlegel, A.M., Weissmann, C.: Darwinian evolution of prions in cell culture. Science 327(5967), 869–872 (2010)

18. Lincoln, T.A., Joyce, G.F.: Self-sustained replication of an RNA enzyme. Science 323(5918), 1229–1232 (2009)

19. Mao, C., Sun, W., Seeman, N.C.: Designed two-dimensional DNA Holliday junction arrays visualized by atomic force microscopy. Journal of the American Chemical Society 121, 5437–5443 (1999)

20. Markov, I.V.: Crystal Growth for Beginners. World Scientific, Singapore (2003)

21. Mills, D., Peterson, N., Spiegelman, S.: An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. Proceedings of the National Academy of Sciences USA 58, 217–224 (1967)

22. Orgel, L.E., Crick, F.H.C.: Anticipating an RNA world. Some past speculations on the origin of life: Where are they today? FASEB Journal 7, 238–239 (1993)

23. Rothemund, P.W.K., Ekani-Nkodo, A., Papadakis, N., Kumar, A., Fygenson, D.K., Winfree, E.: Design and characterization of programmable DNA nanotubes. Journal of the American Chemical Society 126(50), 16344–16352 (2004)

24. Rothemund, P.W.K., Papadakis, N., Winfree, E.: Algorithmic self-assembly of DNA Sierpinski triangles. PLOS Biology 2, 424–436 (2004)

25. Rothemund, P.W.K., Winfree, E.: The program-size complexity of self-assembled squares. In: Symposium on Theory of Computing (STOC), pp. 459–468. ACM, New York (2000)

26. Schulman, R., Winfree, E.: Self-replication and evolution of DNA crystals. In: Capcarrère, M.S., Freitas, A.A., Bentley, P.J., Johnson, C.G., Timmis, J. (eds.) ECAL 2005. LNCS (LNAI), vol. 3630, pp. 734–743. Springer, Heidelberg (2005)

27. Schulman, R., Winfree, E.: Synthesis of crystals with a programmable kinetic barrier to nucleation. Proceedings of the National Academy of Sciences USA 104(39), 15236–15241 (2007)

28. Schulman, R., Winfree, E.: How crystals that sense and respond to their environments could evolve. Natural Computing 7, 219–237 (2008)

29. Schulman, R., Winfree, E.: Programmable control of nucleation for algorithmic self-assembly. SIAM Journal on Computation 39, 1581–1616 (2009)

30. Segré, D., Ben-Eli, D., Deamer, D.W., Lancet, D.: The lipid world. Origins of Life and Evolution of Biospheres 31(1-2), 119–145 (2001)

31. Wächtersäuser, G.: Before enzymes and templates: theory of surface metabolism. Microbiology and Molecular Biology Reviews 52(4), 452–484 (1988)

32. Walde, P., Wick, R., Fresta, M., Mangone, A., Luisi, P.L.: Autopoetic self-reproduction of fatty acid vesicles. Journal of the American Chemical Society 116, 11649–11654 (1994)
33. Wetmur, J.G., Fresco, J.: DNA probes: Applications of the principles of nucleic acid hybridization. Critical Reviews in Biochemistry and Molecular Biology 26(3-4), 227–259 (1991)
34. Winfree, E.: The xgrow simulator, `http://www.dna.caltech.edu/Xgrow/`
35. Winfree, E.: On the computational power of DNA annealing and ligation. In: Lipton, R.J., Baum, E.B. (eds.) DNA Based Computers. DIMACS, vol. 27, pp. 199–221. American Mathematical Society, Providence (1996)
36. Winfree, E.: Simulations of computing by self-assembly. Technical Report CS-TR:1998.22, Caltech (1998)
37. Winfree, E., Bekbolatov, R.: Proofreading tile sets: Error-correction for algorithmic self-assembly. In: Chen and Reif [10], pp. 126–144
38. Winfree, E., Liu, F., Wenzler, L.A., Seeman, N.C.: Design and self-assembly of two-dimensional DNA crystals. Nature 394, 539–544 (1998)
39. Wuensche, A., Lesser, M.: The Global Dynamics of Cellular Automata: An Atlas of Basin of Attraction Fields of One-Dimensional Cellular Automata. Perseus Books, Cambridge (1992)
40. Yin, P., Hariadi, R.F., Sahu, S., Choi, H.M.T., Park, S.H., LaBean, T.H., Reif, J.H.: Programming DNA tube circumferences. Science 321, 824–826 (2008)