An extended cellular automata method applied to the simulation of auto-catalytic and self-replicative molecular behavior

José Pereira
Instituto de Ciências Biomédicas de Abel Salazar, Porto, Portugal

1. The concept of cellular automata (CA)

In a classical cellular automata the properties of each cell in an array of many cells is varied (or not) in each time pulse depending on the properties of it's neighbors and a set of predefined rules. Some extended cellular automata methods look at the cells as moving objects which is equivalent to adding to a classical cellular automaton some "no property cells" (empty cells) and the "movement rule" (swap of properties from a normal cell to an empty cell).

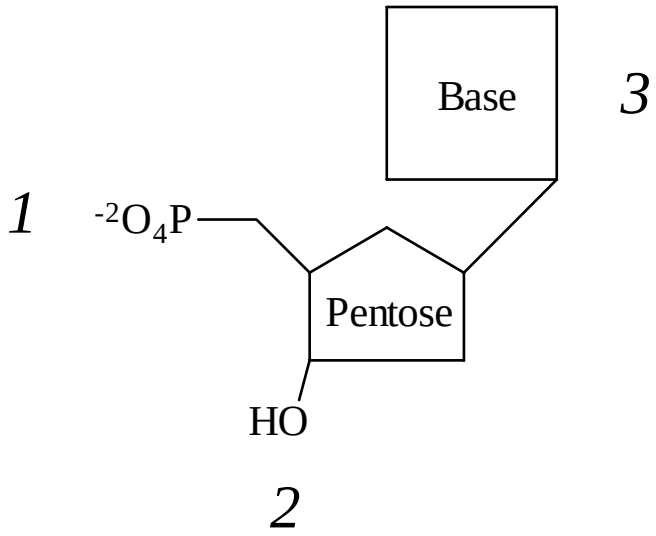
Cellular automata are used for a variety of simulations in different fields of research (physics, chemistry[1], biology, engineering) applied mainly to systems were many objects (molecules, cells, animals) interact following certain rules. The variety of cell types and interactions types can make the time evolution modellization of the whole very difficult by analytical means due to the resultant complexity.

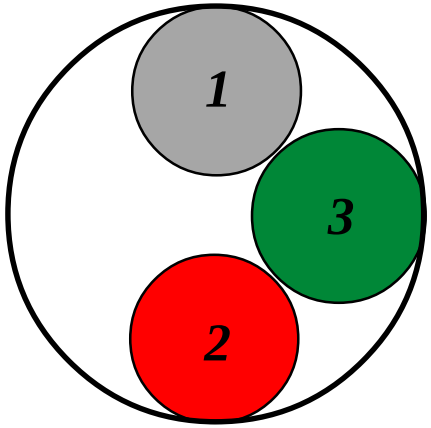
When trying to simulate a dynamic system, usually a set of differential equations have to be derived based on some hypothetical model and then fitted, by changing adequate parameters, to some varying property of the system. In the extended cellular automata method the system is approximately reproduced, based also on an hypothetical model, in a smaller scale and sometimes in a less dimensional space, the interaction rules being the adjustable parameters. This approach gives a direct insight over the evolution of the system, including spacial distribution of objects.

[1] H-H Chou, J.A. Reggia, R. Navarro-González and J. Wu, *Computers Chem.*, **18**, 33 (1994)

2. The application to nucleotide-nucleotide interaction

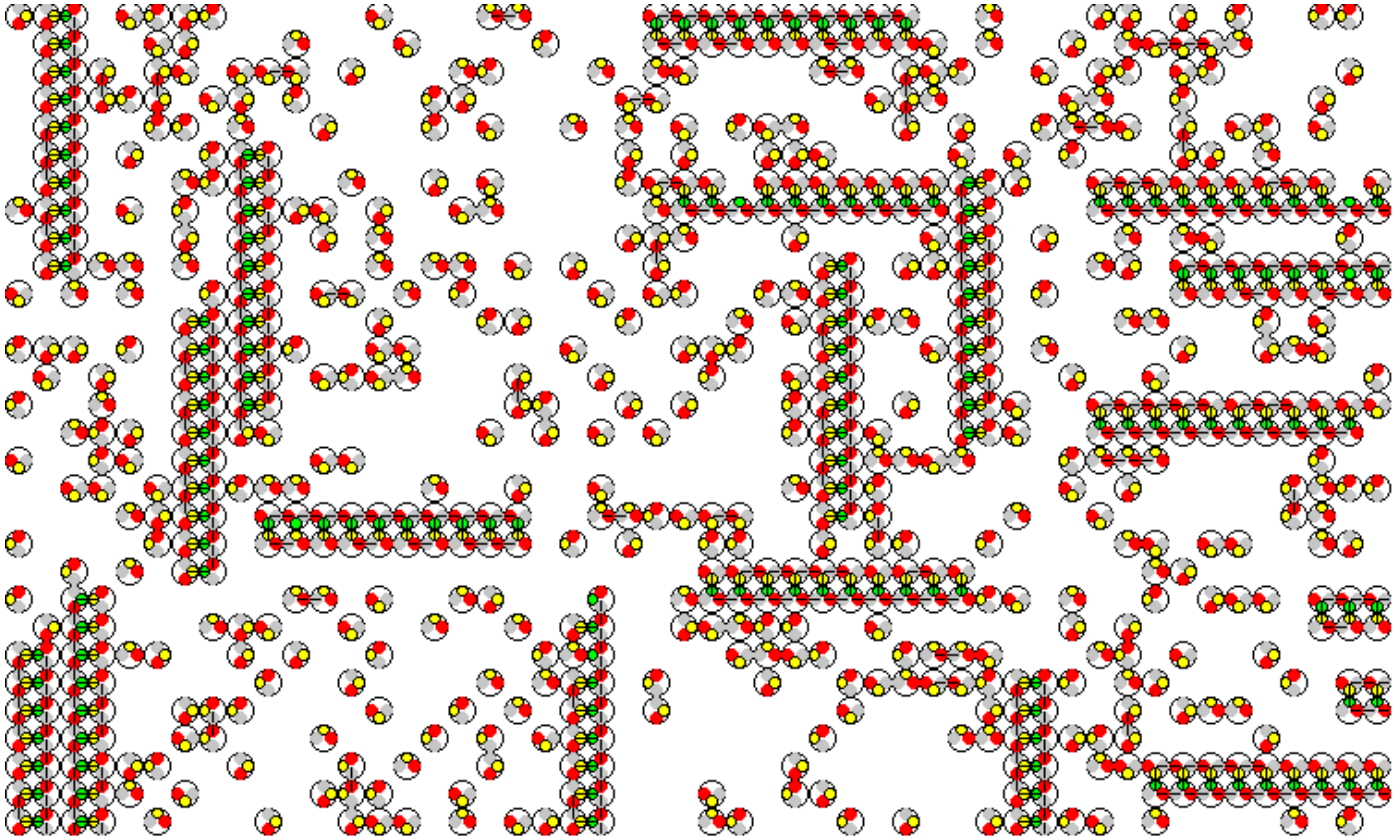
In this work the first level objects in the cellular automaton are high-level representations of nucleotides, with three sites for chemical bonding: the 3'-OH, the 5'-phosphate and the base hydrogen bond sites. Site 3 can have four different types, one for each nucleotide type: A,G,C,U(T). These objects can have four different orientations and move randomly in a 2-dimensional space. When a collision occur between two cells a bond may be formed if (1) the relative orientation of the cells aligns site 1 of one cell and site 2 of the other or two sites 3 of the right kind (A-U(T) or G-C) and if (2) a predefined threshold parameter is randomly exceeded (bond formation probability). These parameters are independently defined for 1-2, 3A-3U(T) and 3G-3C interaction. Polymers of any length may form and are translated and rotated as rigid units.





3. CA extract example

The small black line crossing over the cell color code indicates that it is bonded to the adjacent cell. In this example are represented two types of cells (green, type 3 and yellow, type 5) as well as some template type 3 decamers and associated polymers of type 5.



4. Basic assumptions of the CA program

1. The cell space is 100 by 100. Each single object has coordinates (x,y) with 1<x,y<100.

2. Rotation can be -90, 0, 90 or 180 degrees and translation is limited by one cell at a time in the eight directions possible. For multiple-cell objects (polymers) rotation is restricted to -90, 0 and 90 degrees.

2. In each cycle the objects (single cell or polymer) are chosen at random and a new orientation and position for that object is also randomly chosen. Polymers occupy more than one cell and all cells move in block to preserve the object structure.

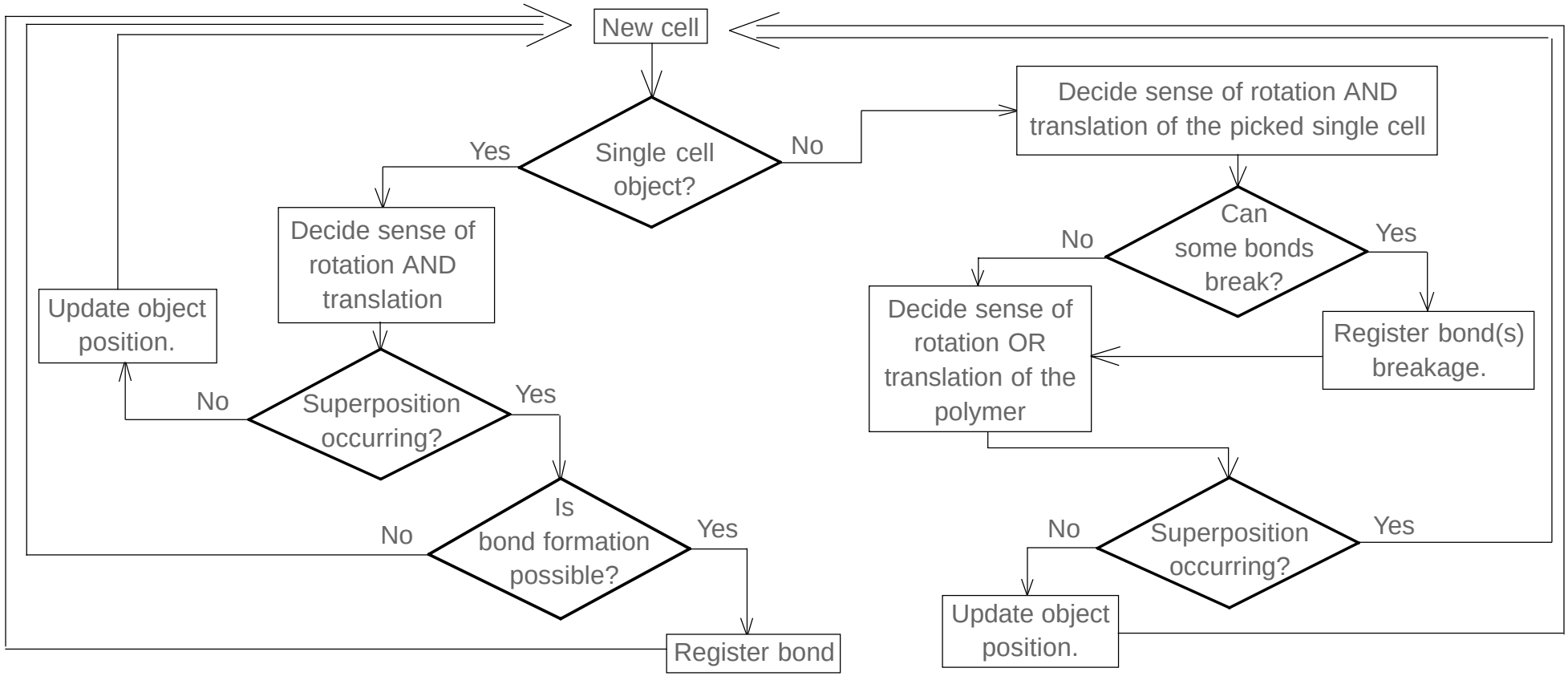
3. Superpositions are not allowed. Polymers don't translate or rotate if a superposition can occur.

4. Collisions exist when a single cell object moves and a superposition can occur. In that case a bond can be created if the orientation of the two cells is the correct one and the threshold (between 0 and 1) for that bond formation is randomly exceeded.

5. Bonds may break when a cell that is connected to others tries to rotate or move directly against a bond direction. In that case a threshold must also be randomly exceeded for the bond to break.

5. Structure of the CA program

The program was implemented in FORTRAN language and requires two input files. One file contains the bond formation and breakage threshold values. The other contains the population of objects with the x and y coordinates, orientation, and bonding status (bond present or not) of the three sites. The program performs a user designated number of cycles until it stops.



6. Dimer formation

The dimer formation is accomplished by setting a threshold of 0 for sites 1 and 2 bonding (the 5'-phosphate/3'-OH bond). Only two kinds of complementary monomers were present at the beginning, thus able of forming only the "hydrogen bond" connection (site 3).

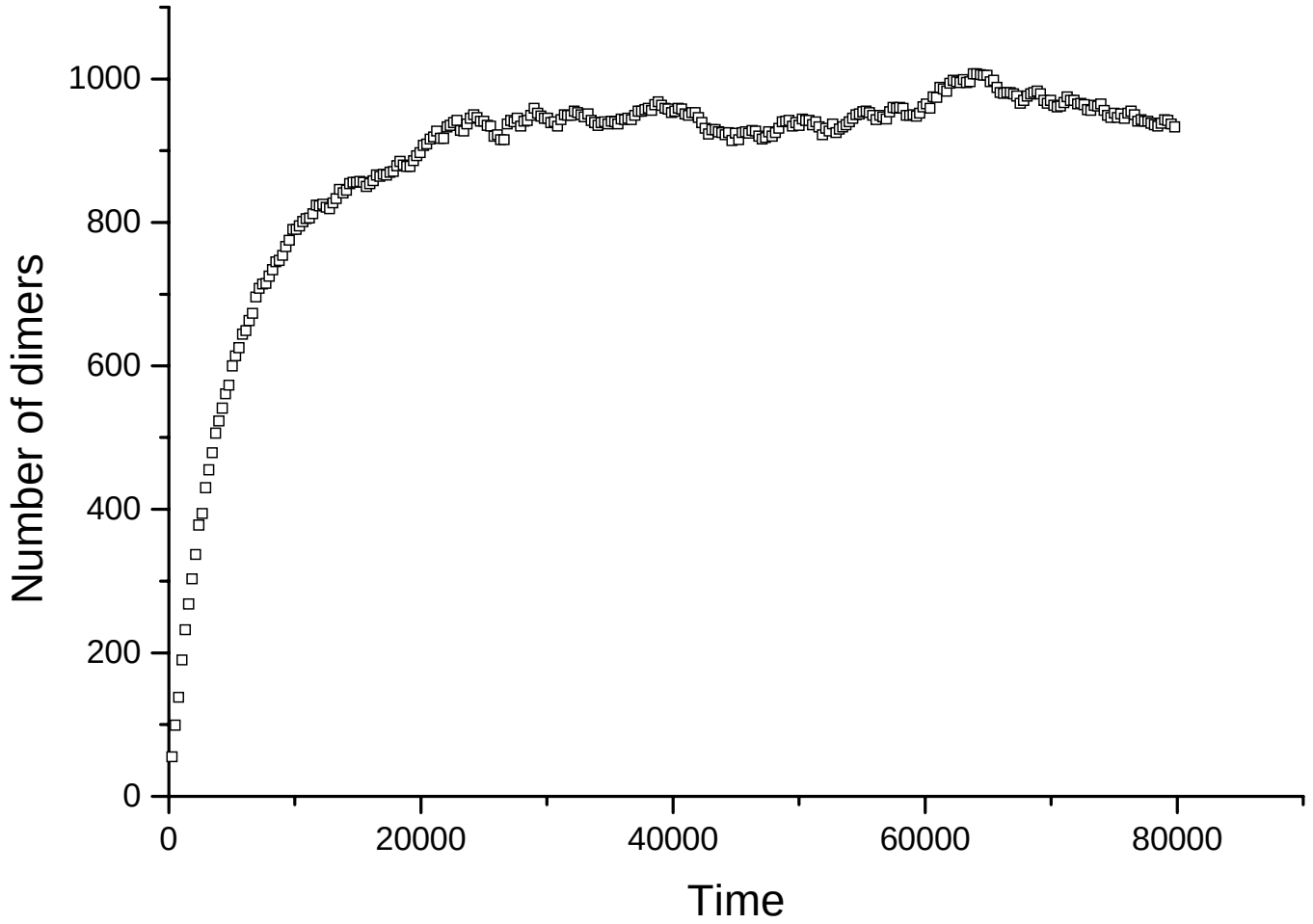
Dimer formation with "time" is represented in the next page. The number of dimer objects stabilizes in what can be identified as the equivalent of an equilibrium state.

This simple model allowed to observe that the number of cells has to be high to reduce the noise from the object counting. That noise is also minimized when the threshold parameters induce a slower evolution of the system, that is, maximizes the number of interactions between the cells.

An equilibrium constant can be calculated for this dimer formation reaction, just like for a normal chemical reaction, that is independent of the number of objects, it's concentration and it's initial proportion. It depends only on the bond forming and bond breaking probabilities set for the system.

Initial values were 2000 units of cell type 3 and 3000 units of cell type 5. At the end about 960 dimer objects were found as a equilibrium state for the system.

6. Dimer formation - Plot



7. Template-directed synthesis

Template-directed synthesis of polymers was accomplished in several cases of which only one is presented. Initially, 1000 decamers composed of type 3 cells are mixed with 4000 type 5 cells. In one experiment "hydrogen bonding" was allowed and so type 5 cells can cross-bond to the decamers which accelerates the appearance of type 5 cell dimers (a2), trimers (a3), tetramers (a4), pentamers (a5) and even longer polymers.

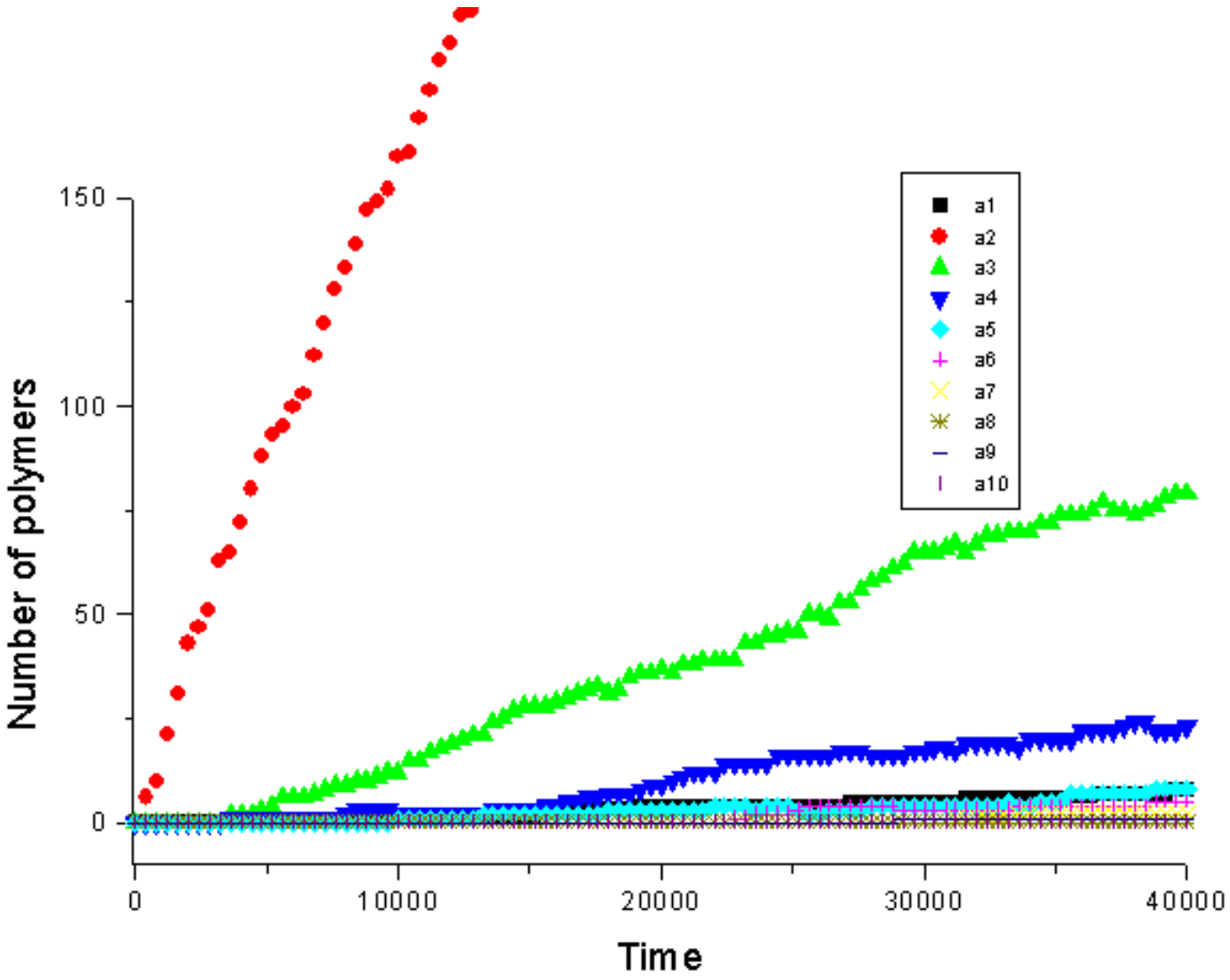
For a control experiment exactly the same initial system was used but with probability 0 for "hydrogen bonding". As can be seen in plot 2 the number of polymers higher than trimer (a3) is very low as the accelerator factor of the decamer template is not present.

This simulates qualitatively a series of experiments done with actual oligonucleotide template-directed synthesis, namely using polycytidylic acid (poly-C) as template and 2-MeImpG (5'-phosphate activated guanosine nucleotide) to form poly-G oligomers.[2]

This approach to the simulation of template-directed synthesis can be developed to give a useful tool for testing possibilities for the beginning of life in it's most eluding aspect: the appearance of consistent self-replication.

[2] A. Kanavarioti, C.F. Bernasconi, D .J. Alberas, E.E. Baird, *J. Am. Chem. Soc.*,**115**, 8537 (1993)
D. Sievers and G. von Kiedrowski, *Nature*, **369**, 221 (1994)

7. Template-directed synthesis - Plot 1



7. Template-directed synthesis - Plot 2 (Control)