

Describing Protein Synthesis and a Cell Cycle of an Imaginary Cell Using a Simple Artificial Chemistry

Kazuto Tominaga

Tokyo University of Technology, School of Computer Science
1404-1 Katakura, Hachioji, Tokyo 192-0982 Japan
tomi@acm.org

Abstract. Recently, artificial chemistries are used as good tools to study self-replication. One current trend in the use is to incorporate a continuous space, but it would make implementation of a simulator a complex and time-consuming task. We take another approach, which is to study self-replication using an artificial chemistry without a spatial structure. In this paper, we used a simple artificial chemistry based on pattern matching and recombination of stacks of strings to describe a system that models an imaginary self-replicating cell. This cell produces proteins according to a coding table, and copies its chromosomes. Its functional protein components, named ribosome and polymerase, coordinate using small elements similar to signal molecules in natural cells to give the cell a cycle.

1 Introduction

Self-replication has been of great interest in the field of artificial life [1], and a number of systems have been developed and studied, especially those based on cellular automata. Recently, the field of artificial chemistry emerged [2] and self-replication is being studied with several artificial chemistries [3–6].

One approach in this direction is to provide an artificial chemistry with the capability to express spatial aspects of what is modelled. Examples of such artificial chemistries are JohnnyVon [5] and Hutton's two-dimensional artificial chemistry [6]. These systems are exploring the properties of self-replication in the context of spatial dynamics. Another approach is to study self-replication using an artificial chemistry without a spatial structure, such as the work with SAC [4].

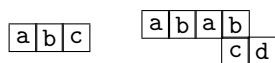
We take the latter approach in this paper: we describe an imaginary cell that produces protein and has a cell cycle using a simple artificial chemistry without a spatial structure [7] as an attempt to reveal essential properties of self-replication that are independent from particular spatial structures.

2 An Artificial Chemistry for Stacks of Character Strings

This section briefly explains a simple artificial chemistry based on pattern matching and recombination of stacks of strings [7], which we are going to use in the following discussion.

2.1 Elements and objects

An *element* is a character; it corresponds to an atom or a group of atoms in nature. An *object* (corresponds to a molecule) is a character string or a stack of strings, such as those depicted below.



These objects are denoted by the string notations $0\#abc/$ and $0\#abab/3\#cd/$, respectively; the numbers represent displacements of the line relative to the first line.

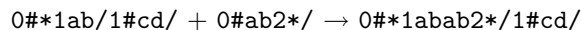
2.2 Patterns

A *pattern* matches (or does not match) an object, and it can utilize two kinds of *wildcards*. An *element wildcard*, which is denoted by a number such as 1, matches any element. A *sequence wildcard*, denoted by a number and an asterisk such as 2^* and $*2$, matches any sequence of zero or more elements; the position of an asterisk represents the direction in which the sequence can extend.

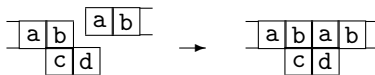
A pattern is denoted in a similar way to an object. For example, the pattern $0\#a1c/$ matches the object $0\#abc/$; the pattern $0\#*1ab/1\#cd/$ matches the object $0\#abab/3\#cd/$. Note that the displacements are meaningful and that the length of a sequence wildcard is treated as zero in the notation for patterns.

2.3 Recombination rules

A *recombination rule* transforms a group of objects into a group of objects, conserving elements just like a chemical reaction does. It is expressed in a manner similar to chemical formulae, but in terms of patterns. An example rule is



which is illustrated as follows.



If this rule is applied to the objects $0\#abab/3\#cd/$ and $0\#abc/$, the object $0\#abababc/3\#cd/$ is produced and the reactants disappear.

2.4 Sources and drains

Objects are kept in the *working multiset*. A *source* is defined as an object, and it supplies objects of the specified form to the multiset one at a time without limit. A *drain* is defined as a pattern, and it eliminates objects matched by the pattern, one at a time, from the multiset.

2.5 Dynamics

A *system* is a construct $\langle \Sigma, S, D, R, P_0 \rangle$ where Σ is a set of elements, S is a multiset of sources, D is a multiset of drains, R is a set of recombination rules, and P_0 is the *initial working multiset*, which specifies objects in the working multiset at the initial state. The system is interpreted nondeterministically as follows.

1. Initialize the working multiset to be P_0 .
2. Do one of the following operations.
 - Apply one recombination rule to a collection of objects.
 - Operate one source.
 - Operate one drain.
3. Go to Step 2.

3 An Imaginary Cell

Let us suppose a world where the following materials exist.

- Amino acids. There are twelve kinds of amino acids and are denoted (as elements in the artificial chemistry) by **R**, **B**, **D**, **P**, **a**, **b**, **c**, **d**, **e**, **f**, **g**, and **h**, respectively.
- Bases. There are four kinds of bases named **T**, **C**, **A** and **G**.
- Miscellaneous materials. Let us have two miscellaneous materials named **p** and **q**.

Now let us have an imaginary cell. The cell consists of the following components.

- A ribosome. This is a protein that produces proteins specified by chromosomes. A ribosome is expressed as an object of the form **RB** in the artificial chemistry.
- A DNA polymerase. This is a protein that copies chromosomes, and denoted by **DP**.
- Two chromosomes. One specifies ribosomes, and the other specifies DNA polymerases.

In this world, a codon comprises two bases instead of three (which is the number of bases in a codon in natural cells). The coding table of this world is shown in Table 1. For example, the codon **TA** specifies the amino acid **D**.

Table 1. Coding table.

		Second base			
		T	C	A	G
First base	T	R	B	D	P
	C	a	b	c	d
	A	e	f	g	h
	G	-	-	-	-

The sequence of amino acids for a ribosome RB is coded as the base sequence TTTC. The chromosome of ribosome is given as an object $0\#pdCTTTTCq/$ (like single-stranded DNA), where p designates the beginning of the gene, and q the end. The sequence CT, which specifies the amino acid a , controls the behaviour of the cell, and so the amino acid d in the chromosome (see Section 4.3). Having an “amino acid” in a chromosome might seem strange; we will discuss it in Section 6.

Similarly, the sequence for a DNA polymerase DP is coded as TATG, and its chromosome is $0\#paATTATGq/$.

4 Descriptions for the Cell

Now we are going to define a system that describes our cell. To let the system supplied enough materials for the cell to operate, we first furnish the system with the following sources of materials.

Sources: $0\#p/$, $0\#q/$, $0\#R/$, $0\#B/$, $0\#D/$, $0\#P/$, $0\#a/$, $0\#b/$, $0\#c/$, $0\#d/$, $0\#e/$, $0\#f/$, $0\#g/$, $0\#h/$, $0\#T/$, $0\#C/$, $0\#A/$, $0\#G/$

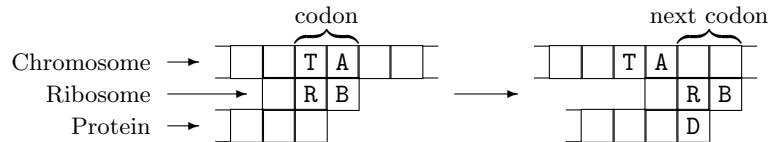
The initial working multiset has one cell, composed of a ribosome, a DNA polymerase, and their chromosomes.

Initial working multiset: $0\#aRB/$, $0\#eDP/$, $0\#pdCTTTTCq/$, $0\#paATTATGq/$

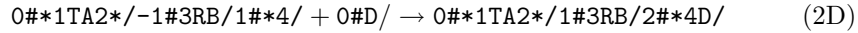
We will give recombination rules to implement protein synthesis and a cell cycle in the following sections. Shown with each rule is its rule number (such as “2D”) in the complete description of the system, which is given in Appendix A.

4.1 Protein Synthesis

In this cell, the ribosome attaches directly to a chromosome to produce proteins. The below picture shows a ribosome adding an amino acid D to the protein under synthesis according to the codon TA.



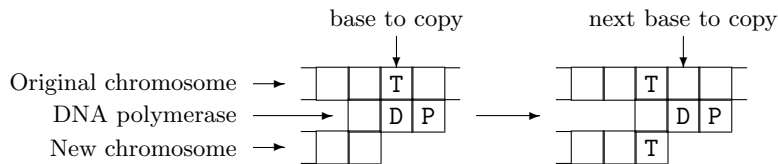
The ribosome moves rightward relative to the chromosome as the production proceeds. This behaviour of ribosome is described as the following recombination rule.



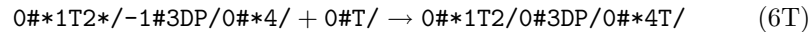
A ribosome has one extra amino acid on its left end such as *a* in $0\#aRB/$; the wildcard 3 in the rule is the slot for it. Giving twelve such rules for the set of amino acid implements the coding table (Table 1).

4.2 Copying Chromosomes

A DNA polymerase copies a chromosome by attaching to it. Shown in the picture below is the process of copying the base T.



The polymerase moves rightward relative to the attached chromosome as the copying proceeds. The below rule describes this process.



We give the system such rules for the bases C, A and G as well.

4.3 Cell Cycle

Our cell consists of one ribosome, one DNA polymerase and two chromosomes. If synthesis of proteins and copying of chromosomes occur independently from each other, it might happen, for example, that many proteins are produced without copying any chromosome; there should be some control over protein synthesis and chromosome copying to produce one new cell at a time.

We introduce four phases to the system (Figure 1). Each phase produces a component of the new cell as follows.

Phase 1: The ribosome produces a new polymerase.

Phase 2: The polymerase produces a new chromosome of polymerase.

Phase 3: The polymerase produces a new chromosome of ribosome.

Phase 4: The ribosome produces a new ribosome.

Each phase is characterized by which protein is active and what it does. An active protein is surrounded by a box in Figure 1. In Phase 1, the ribosome is active to produce a new polymerase $0\#eDP/$. We give each protein amino acids to represent its states; let us call them *tags*. For example, the amino acid *a* is

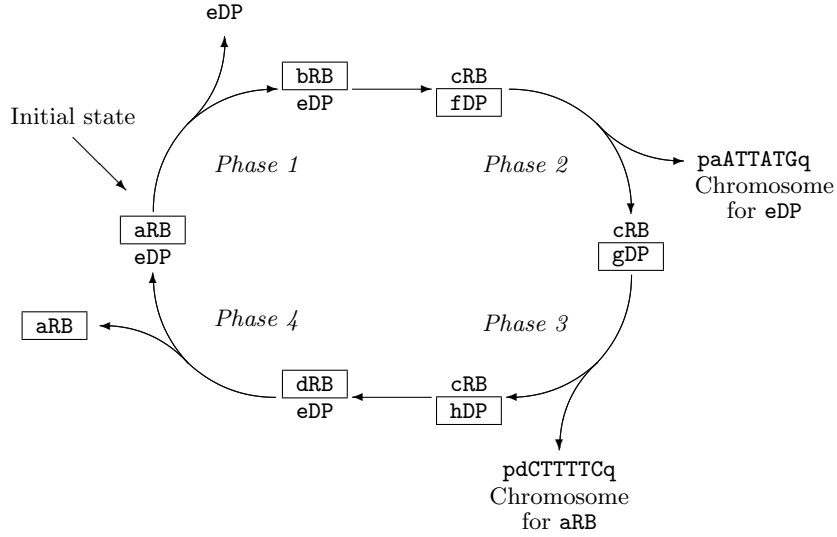
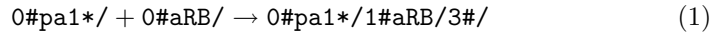


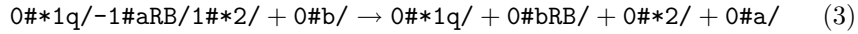
Fig. 1. The cell cycle of the system.

the tag for ribosomes that represents the ribosome is going to produce a new polymerase. A ribosome tagged with a is of the form $0\#aRB/$.

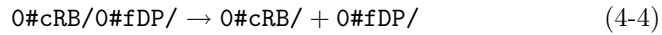
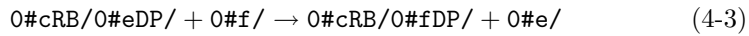
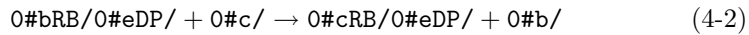
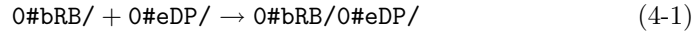
The proteins' activities are controlled as follows. First, the cell is in the initial state shown in the figure. The ribosome with the tag a is active and going to produce a new polymerase. It attaches to the chromosome of polymerase by the rule below and starts the synthesis of protein.



When the ribosome completes the production of polymerase, it comes to have the tag b by the rule below



and hands over its activity to the polymerase by the following rules; this finishes Phase 1 and starts Phase 2.



Similar rules are given for each connection of phases.

At the end of Phase 4, all the components of the new cell have been produced by this cycle. Therefore, the production of the new cell completes at this point.

The ribosome of the new cell, which has just been produced, is of the form $0\#aRB/$, so it starts a new cell cycle in the new cell. At the same time, the original ribosome comes to take the form $0\#aRB/$ and starts the next cell cycle.

5 Sample Execution

We are developing a simulator for the artificial chemistry. Figure 2 shows window snapshots from an execution of the described system. The total number of objects in the working multiset is shown at the top of each window. Since our simulator cannot handle sources yet, materials are given in the initial multiset.

Figure 2a shows the initial state. Each line in the table represents one form of object shown in the right column, and the number of objects of that form is shown in the left column. (Note that not all lines in the table are shown in the figure.) There are components of one cell in Figure 2a, namely, one ribosome, one DNA polymerase and two chromosomes. Figure 2b shows the ribosome producing a new polymerase. In Figure 2c, the ribosome is handing over its activity to a polymerase. The polymerase copies the chromosome of itself (Figure 2d). At the end of cell cycle, two sets of cell components are obtained (Figure 2e). They start new cell cycles and work independently from each other (Figure 2f).

The simulator was implemented with Objective-C. The execution above took about four minutes on PowerPC G4 867MHz from Figure 2a to 2e with approximately 87,000 random collisions of objects.

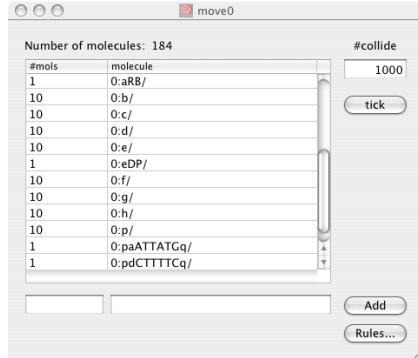
6 Discussion

The processes of replicating chromosomes and protein synthesis in this system are designed based on direct analogies from natural cells. It is easy to extend the coding table to use three bases for a codon instead of two, with increase of the number of amino-acid adding rules up to 64 ($= 4^3$).

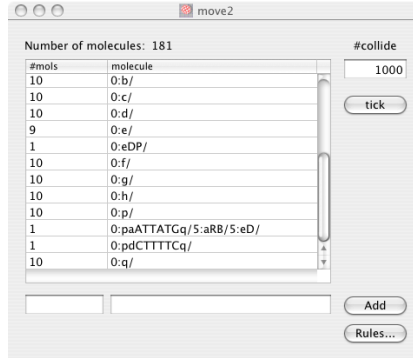
The cell cycle is implemented by the tags attaching to the proteins. In a natural cell, signalling is used to coordinate various activities of molecules, organelles and the cell itself. The tags in this system serve as signal molecules. The mechanism of this cell's cycle is by far different from the natural one, but signalling should also play a role in the natural cell cycle.

The notion of individual cell is obscure in this system since the components of a cell are implemented as separate objects. For example, if there are multiple cells in the working multiset, any ribosome can interact with any chromosome. Except for this, we think the system describes a cell that has protein-synthetic capability and a cell cycle quite well with the simple artificial chemistry without spatial structures. This cell cycle solved the problem of uncontrolled production of proteins or replication of chromosomes, which can be found in a study of self-replication with SAC [4].

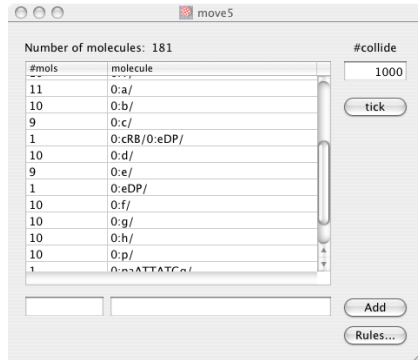
In this description, chromosomes contain materials other than bases, namely, the elements p , q , a and d . The element p serves as a promoter that specifies



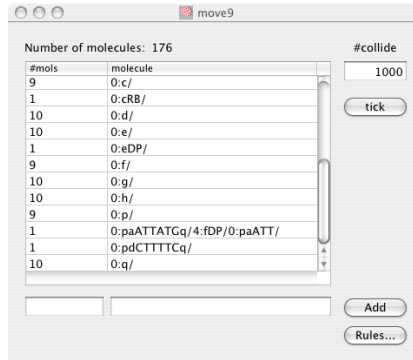
a. Initial state



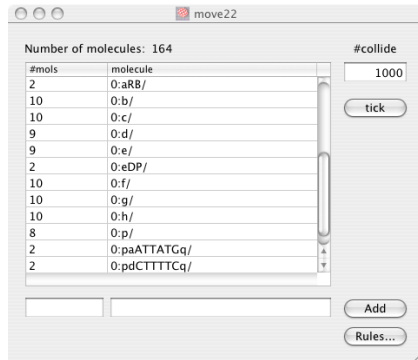
b. Protein synthesis



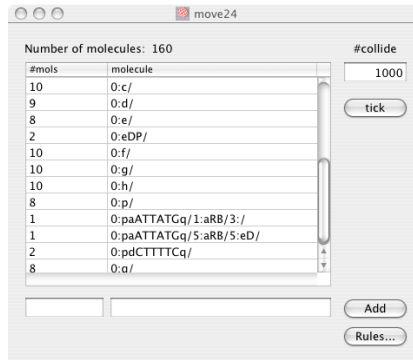
c. Handing over the activity



d. Chromosome copying



e. Two cells



f. Working independently

Fig. 2. Sample execution.

the beginning of a gene; it can be replaced with a particular sequence of bases by giving an appropriate set of rules. The element q indicates the end of a gene, so introducing a stop codon can eliminate it. The elements a and d serve as activators for genes and as the initiators of chromosomal replication; they could also be replaced with sequences of bases to which an activator protein and an initiator protein attach, by giving appropriate rules and the new types of proteins. Therefore, the system could be designed to have only bases in the chromosomes.

If one is to furnish the cell with additional components so as to have more complex behaviour take place, incorporating some notion of compartment would be necessary such as membranes used in Lattice Artificial Chemistry [8] and Hutton's two dimensional artificial chemistry [6]; this will be our future work.

7 Conclusions

We have given the description of a system that models the protein synthesis and cell cycle of an imaginary cell using an artificial chemistry. The described system has the following properties: a protein is produced according to a sequence of codons in a chromosome; and the activities of functional proteins are controlled by attaching elements that are similar to signal molecules, initiator and activator proteins in natural cells. We think the system models some part of the self-replicating behaviour of natural cells quite well with the simple artificial chemistry having no spatial structure.

As future work, modelling other activities of natural cells using this artificial chemistry may also be possible and will be interesting. Enhancements to the artificial chemistry, especially incorporating a notion of compartment, would enable it to explore more in the properties of self-replication.

References

1. Sipper, M.: Fifty years of research on self-replication: An overview. *Artificial Life* **4** (1998) 237–257
2. Dittrich, P., Ziegler, J., Banzhaf, W.: Artificial chemistries — a review. *Artificial Life* **7** (2001) 225–275
3. Hutton, T.J.: Evolvable self-replicating molecules in an artificial chemistry. *Artificial Life* **8** (2002) 341–356
4. Suzuki, H., Ono, N.: Universal replication in a string rewriting system. In: Proceedings of the Fifth International Conference on Humans and Computers (HC-2002). (2002)
5. Smith, A., Turney, P., Ewanschuk, R.: Self-replicating machines in continuous space with virtual physics. *Artificial Life* **9** (2003) 21–40
6. Hutton, T.J.: A functional self-reproducing cell in a two-dimensional artificial chemistry. In: Proceedings of the 9th International Conference on the Simulation and Synthesis of Living Systems (ALIFE9). (2004) 444–449
7. Tominaga, K.: A formal model based on affinity among elements for describing behavior of complex systems. Technical Report UIUCDCS-R-2004-2413, Department of Computer Science, University of Illinois at Urbana-Champaign (2004)

8. Ono, N., Ikegami, T.: Selection of catalysts through cellular reproduction. In: Proceedings of the 8th International Conference on the Simulation and Synthesis of Living Systems (ALIFE8). (2002)

A The System Description

Elements: p, q, R, B, D, P, a, b, c, d, e, f, g, h, T, C, A, G.

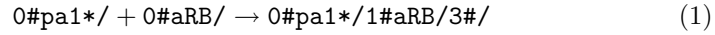
Sources: 0#p/, 0#q/, 0#R/, 0#B/, 0#D/, 0#P/, 0#a/, 0#b/, 0#c/, 0#d/, 0#e/, 0#f/, 0#g/, 0#h/, 0#T/, 0#C/, 0#A/, 0#G/.

Drains: none.

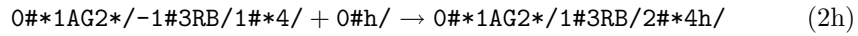
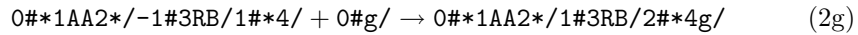
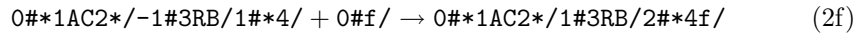
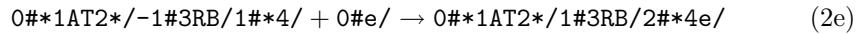
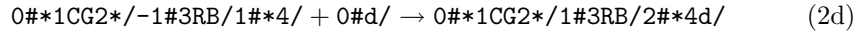
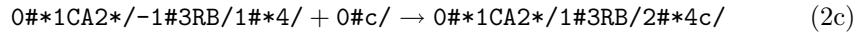
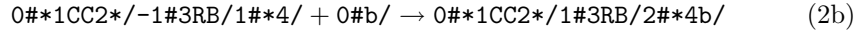
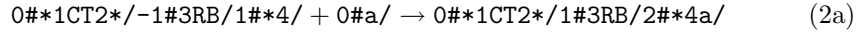
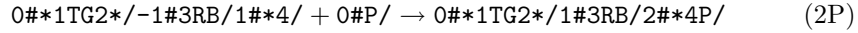
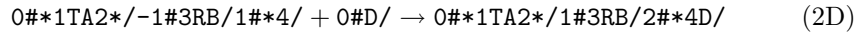
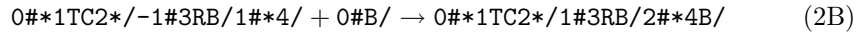
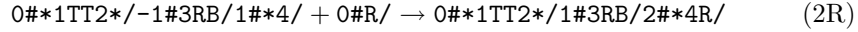
Initial working multiset: 0#aRB/, 0#eDP/, 0#pdCTTTCq/, 0#paATTATGq/.

Recombination rules

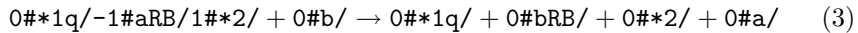
Starting production of polymerase



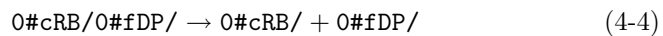
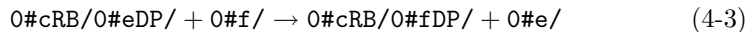
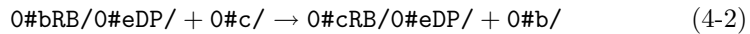
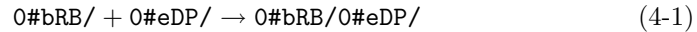
Adding amino acids to the protein



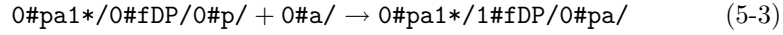
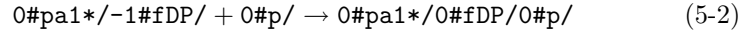
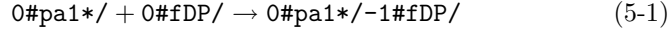
Finishing the production of ribosome



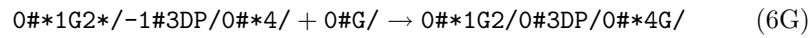
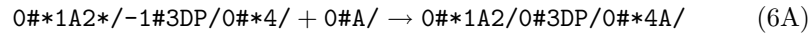
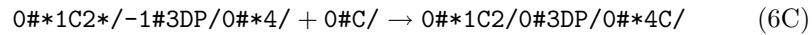
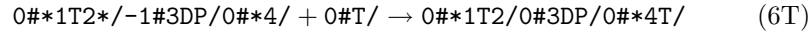
Transiting from Phase 1 to 2



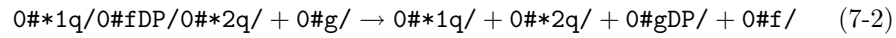
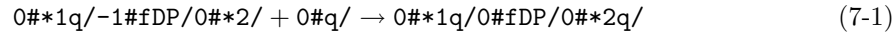
Starting copying chromosome of polymerase



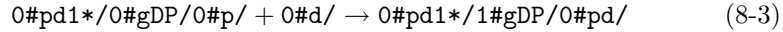
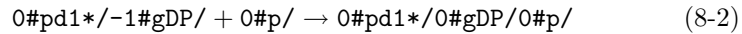
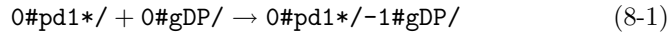
Adding bases to the chromosome



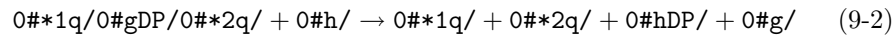
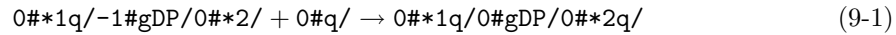
Finishing copying chromosome of polymerase and moving on to Phase 3



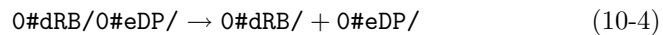
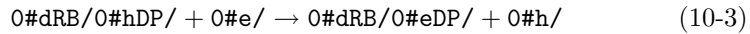
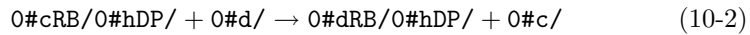
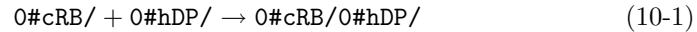
Starting copying chromosome of ribosome



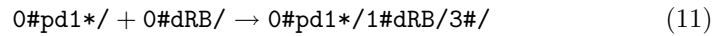
Finishing copying chromosome of ribosome



Transiting from Phase 3 to 4



Starting production of ribosome



Finishing the production of ribosome and restart the cell cycle

