

DOI: 10.51790/2712-9942-2020-1-1-4

АВТОМАТНАЯ МОДЕЛЬ РЕПЛИКАЦИИ МОЛЕКУЛЫ ДНК**Г. Е. Деев, С. В. Ермаков***Обнинский институт атомной энергетики, Национальный исследовательский ядерный университет «МИФИ», г. Обнинск, Российская Федерация, georgdeo@mail.ru, ermakov@iate.obninsk.ru*

Аннотация: показано, что репликация молекулы ДНК может быть промоделирована как вычисление, проводимое простейшими абстрактными автоматами с одним состоянием — перестановочными автоматами.

Ключевые слова: абстрактный автомат, В-схема, ДНК — дезоксирибонуклеиновая кислота, репликация, комплементарность.

Благодарности: работа выполнена за счет гранта РФФИ (проект № 20-07-00862).

Для цитирования: Деев Г. Е., Ермаков С. В. Автоматная модель репликации молекулы ДНК. *Успехи кибернетики*. 2020;1(1):26–30. DOI: 10.51790/2712-9942-2020-1-1-4.

AUTOMATA-BASED MODEL OF DNA REPLICATION**Georgiy E. Deev, Sergey V. Ermakov***Obninsk Nuclear Energy Institute, National Research Nuclear University MEPhI, Obninsk, Russian Federation, georgdeo@mail.ru, ermakov@iate.obninsk.ru*

Abstract: it is shown that DNA replication can be simulated as computation by the simplest abstract single-state automata: the permutation automata.

Keywords: abstract automaton, B-scheme, DNA (deoxyribonucleic acid), replication, complementarity.

Acknowledgements: this study is supported by RFBR, project No. 20-07-00862.

Cite this article: Deev G. E., Ermakov S. V. Automata-Based Model of DNA Replication. *Russian Journal of Cybernetics*. 2020;1(1):26–30. DOI: 10.51790/2712-9942-2020-1-1-4.

Абстрактные автоматы

Определение автомата носит комплексный характер. Автоматом называется составной объект, включающий в себя следующие компоненты:

1. *Множество* $Z_k = \{0, 1, \dots, k-1\}$ — входной алфавит автомата. Он состоит из k цифр k -ичной системы счисления. Последняя цифра « $k-1$ » должна восприниматься как единый цельный символ, несмотря на то, что он состоит, по видимости, из трех отдельных символов: k , $-$, 1 . Это объясняется тем, что мы не знаем, как пользователи k -ичной системы счисления договорятся об обозначении последней цифры.

2. *Множество* $Z_l = \{0, 1, \dots, l-1\}$ — выходной алфавит автомата. Он состоит из l цифр l -ичной системы счисления. Относительно цифры « $l-1$ » то же замечание, что и относительно цифры « $k-1$ ».

Таким образом, входной и выходной алфавиты автомата могут быть разными.

3. *Множество* $Q = \{q_1, q_2, \dots, q_n\}$ — множество состояний автомата. Состояния автомата могут обозначаться любыми символами. От обозначения состояний автомата (и их физической реализации) принцип функционирования автомата не зависит.

4. *Функцию переходов автомата* $q' = \theta(q, x)$, показывающую, в какое «новое» состояние q' переходит автомат из «старого» состояния q под воздействием входного сигнала x .

5. *Функцию выходов автомата* $y = \sigma(q, x)$, показывающую, какой выходной сигнал y будет выработан автоматом из состояния q под воздействием входного сигнала x .

Везде выше ($q', q \in Q, x \in Z_k, y \in Z_l$).

Кратко, желая указать компоненты автомата A , пишут:

$$A = A(Z_k, Z_l, Q, \theta(q, x), \sigma(q, x)).$$

Обычно рассматриваются автоматы с конечными множествами Z_k, Z_l, Q . Такие автоматы называются *конечными*.

Перечисления элементов, составляющих автомат, недостаточно для задания автомата. Требуется еще указать, как функционирует автомат.

Автомат функционирует в дискретном времени, для обозначения моментов которого, не ограничивая общности, будем использовать множество целых чисел. Когда наступает целочисленный момент времени, автомат на мгновение «оживает», выполняет свое действие и «замирает» до наступления следующего целочисленного момента времени. При этом он подчиняется *уравнениям функционирования* следующего типа:

$$A: \begin{cases} q(t+1) = \theta(q(t), x(t)), \\ y(t) = \sigma(q(t), x(t)). \end{cases} \quad (t \in Z = \{\dots - 2, -1, 0, +1, +2, \dots\}). \quad (1)$$

Возможны уравнения функционирования иного типа.

При постановке автомата в то или иное состояние в качестве начального он будет реализовывать то или иное отображение.

Действие автомата A , поставленного в начальное состояние q , на объект \bar{x} записывается в виде $\bar{y} = A\bar{x}|q$.

Если у автомата все состояния неэквивалентны, то автомат может реализовать столько различных отображений, сколько у него состояний. Богатство реализуемых отображений определяет возможности автомата.

Перестановочные автоматы с одним состоянием

Нам понадобятся только перестановочные автоматы с одним состоянием. Перестановочным автоматом называется автомат, функция выходов которого получается путем перестановки букв, входящих в выходной алфавит. Количество перестановочных автоматов с одним состоянием для четверичной системы счисления Z_4 равно $4! = 24$.

Перестановочные автоматы с одним состоянием задаются таблицами типа 1.

Таблица 1

	q	0
x		
0		0, y_0
1		0, y_1
2		0, y_2
3		0, y_3

В таблице 1 $y_i \in Z_4$ и различны. Среди таких автоматов есть автомат, задаваемый табл. 2.

Таблица 2

	q	0
x		
0		0, 0
1		0, 1
2		0, 2
3		0, 3

Этот автомат реализует тождественное отображение (тождественное вычисление)

$$x_r \dots x_2 x_1 x_0 \rightarrow x_r \dots x_2 x_1 x_0,$$

т.е. любое число $x_r \dots x_2 x_1 x_0$ переводит в себя. На языке подстановок он реализует тождественную подстановку

$$\begin{pmatrix} 3 & 2 & 1 & 0 \\ 3 & 2 & 1 & 0 \end{pmatrix}.$$

Произвольный автомат, задаваемый табл. 1, реализует произвольную подстановку

$$\begin{pmatrix} 3 & 2 & 1 & 0 \\ y_3 & y_2 & y_1 & y_0 \end{pmatrix}.$$

Известно, насколько богата содержанием алгебраическая теория подстановок. Как видно, все факты из теории подстановок можно изложить в терминах перестановочных автоматов с одним состоянием.

Биологическая трактовка функционирования перестановочных автоматов

Достойную внимания трактовку находят себе перестановочные автоматы с одним состоянием в генетике.

Сведения из генетики, использованные нами, следуют изложению в книге [1].

Молекула ДНК, носитель наследственной информации, согласно модели Уотсона-Крика, представляет собой две сцепленные спиралевидные последовательности нуклеотидов (двойная спираль) четырех типов: аденин (А), гуанин (Г), тимин (Т), цитозин (Ц). Порядок следования нуклеотидов вдоль молекулы может быть любым [1, с. 22]. Количество нуклеотидов вдоль спирали велико. Из произвольности порядка следования нуклеотидов и их количества в молекуле следует огромное количество вариантов наследственной информации [6, с. 237]. Несмотря на то, что последовательность нуклеотидов вдоль одной спирали может быть произвольной, в связанной с ней параллельной спирали последовательность нуклеотидов не может быть произвольной и определяется распределением нуклеотидов в первой спирали. Это зависимое расположение нуклеотидов в параллельной спирали определяется *законом комплементарности*, состоящим в следующем: между нуклеотидами должно быть взаимно однозначное соответствие:

$$A \leftrightarrow T, \quad G \leftrightarrow C. \quad (2)$$

В среде, в которой живут молекулы ДНК, возможен режим, при котором спирали ДНК расходятся, после чего на каждой из них происходит наращивание потерянных нуклеотидов по принципу комплементарности. В результате разошедшиеся спирали восстанавливаются в точности идентичными той, что разделилась, и вместо одной молекулы ДНК получаются две, ничем не отличимые друг от друга молекулы ДНК (не отличающиеся также и от первоначальной молекулы ДНК). Этот процесс удвоения молекулы ДНК называется *репликацией*, и он (и аналогичные ему, отличающиеся в деталях, но не в принципе) лежит в основе роста организма.

Сказанное иллюстрируется рисунками 1 и 2, вместе с подписями, взятыми из [1].

Описанный процесс восстановления молекулы ДНК может быть промоделирован с помощью рассмотренных подстановочных автоматов с одним состоянием.

Действительно, как при репликации, так и при других процессах, которые можно истолковывать как снятие информации, происходит перемещение вдоль молекулы некоторого объекта, который воспринимает входную для него информацию, реагирует на нее, вырабатывая свою выходную информацию, которая используется при строительстве молекулы.

Если перевести это на язык В-схем, то ситуацию можно представить в виде схемы на рис. 3.

Закон комплементарности представлен связями внутри прямоугольника. Точками слева показан возможный вариант последовательности нуклеотидов. Последовательность точек слева, рассматриваемая сверху вниз, означает, что этот участок ДНК имеет верхним звеном нуклеотид гуанин (1), ниже — нуклеотид тимин (2), еще ниже — нуклеотид аденин (0), далее — нуклеотид цитозин (3). Справа почти все точки пустые, что означает, что комплементарное построение еще отсутствует.

Автомат А, перемещаясь вдоль левой половины молекулы ДНК, своим левым входом воспринимает входную информацию и по ней вырабатывает выходной сигнал, комплементарный с входным.

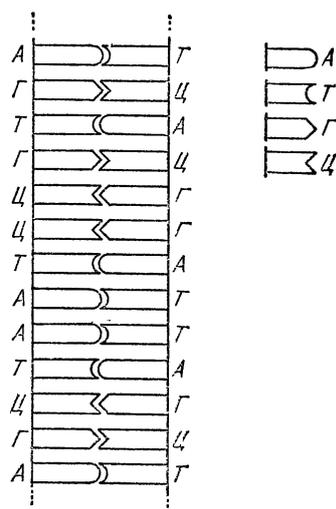


Рис. 1. Молекула ДНК похожа на веревочную лестницу, состоящую из перекладин двух типов — пар нуклеотидов А–Т и Г–Ц

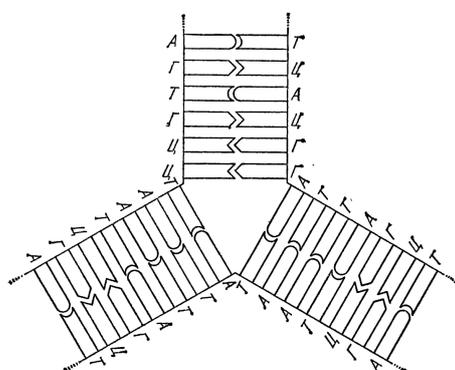


Рис. 2. Так, согласно Уотсону и Крику, происходит процесс репликации ДНК, в результате которого из исходной молекулы, изображенной на рис. 1, получаются две абсолютно такие же молекулы

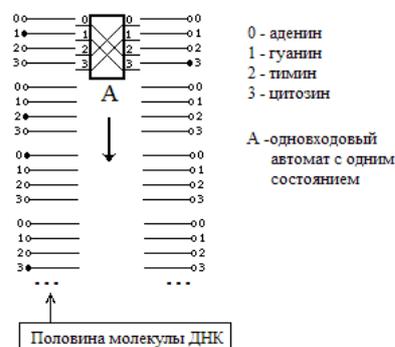


Рис. 3. Автоматная модель процесса репликации молекулы ДНК

На рис. 3 отражена ситуация, когда автомат уже завершил выработку одного комплементарного выхода, т. е. одного комплементарного нуклеотида. А именно, проходя верхний левый нуклеотид, по своему левому входу по линии 1 он воспринял входную информацию, «узнал» нуклеотид гуанин, идущий под номером 1 и согласно связям, имеющим место в автомате (см. рис. 3 и рис. 4), выработал реакцию в виде сигнала 3, что на языке генетики означает присоединение к молекуле ДНК нуклеотида цитозина, комплементарного нуклеотиду гуанину. Одноходовый автомат, осуществляющий комплементарное соединение, можно представлять себе как обычный одноходовый автомат с типичной *B*-схемой, содержащей четыре элемента &. Это автомат A_{10} (см. рис. 4.) (автомат A_{10} взят из общей теории перестановочных автоматов, изложенной в [4]).

Полевая структура молекул слева (рис. 4), вступая во взаимодействие с полевой структурой молекул «прямоугольника», создает такую новую полевую структуру, которая по вариационному принципу притягивает к себе только один из четырех возможных нуклеотидов, а именно тот, который комплементарен левому нуклеотиду. После притягивания создается такая новая полевая структура, которая выталкивает молекулярное образование, представленное прямоугольником, и принуждает его к дальнейшему продвижению вдоль молекулы ДНК. После выталкивания комплементарные нуклеотиды склеиваются, а «прямоугольник» переходит на следующую позицию и там осуществляет последующее взаимодействие с нуклеотидом слева, и все повторяется. При этом надо отметить, что полевая структура образования «левый нуклеотид — прямоугольник — правый нуклеотид» такова, что прямо-

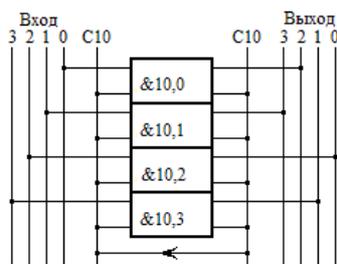


Рис. 4. Схема автомата A_{10} , осуществляющего комплементарное соединение

угольник выталкивается всегда в одну сторону и схватывается следующим в молекуле левым нуклеотидом. Этим обеспечивается направленность движения прямоугольника вдоль молекулы. В принципе, все эти полевые структуры могут быть рассчитаны. Эффект выталкивания, причем всегда в одну сторону, может быть также обнаружен путем вычислений. Можно вычислить время, затрачиваемое на этот процесс, и тем самым оценить быстродействие такой биологической вычислительной структуры. Как видно, простейший биологический автомат с одним состоянием существует и определяет всю нашу жизнь, достигающую невероятных степеней сложности. Следует обратить внимание на то, что математика допускает на этом элементарном абстрактном уровне различные вариации, скажем, не 4-, а 6-нуклеотидные образования. Однако в земной природе они не реализуются. Возникает вопрос: по какой причине? Неужели виной всему вариационные принципы?

Как бы там ни было, биологические процессы кибернетического типа могут быть промоделированы В-компьютерами *непосредственно*, как показано в тексте, без какого-либо программного сопровождения, в отличие от существующих компьютеров, которые для аналогичного моделирования требуют создания специального программного обеспечения.

Заключение

Показано, что жизненные процессы, происходящие на уровне молекул ДНК, могут быть промоделированы абстрактными автоматами, реализованными в виде В-схем. Четырехбуквенные В-схемы автоматов наилучшим образом соответствуют вариационному принципу оптимизации, имеющему место в природе.

ЛИТЕРАТУРА

1. Франк—Каменецкий М. *Королева живой клетки*. М.: АСТ-ПРЕСС; 2010. 272 с.
2. Шредингер Э. *Что такое жизнь?* АСТ; 2018. 288 с.
3. Деев Г. Е. *Абстрактные вычислительные устройства*. М.: Энергоатомиздат; 2007. 332 с.
4. Деев Г. Е. *Теория вычислительных устройств*. Ч. 1. Конспект лекций. Обнинск; 2009.
5. Робертис Э. де, Новинский В., Саэс Ф. *Биология клетки*. М.: Мир; 1973. 488 с.
6. Уотсон Дж. *Молекулярная биология гена*. М.: Мир; 1978. 720 с.
7. Колмогоров А. Н. *Автоматы и жизнь. Математика — наука и профессия*. М.: Наука; 1988. С. 43–62.