

# COMPUTER SIMULATION OF A LIVING CELL: PART I

R. WEINBERG\* and M. BERKUS†

*Department of Computer and Communication Sciences, University of Michigan (USA)*

(Received: 10 June, 1970)

## SUMMARY

*This paper, to be presented in two parts, is concerned with the problem of a realistic and useful simulation of a living cell, the simple unicellular bacterium Escherichia coli, the program being written in Fortran IV for an IBM 360/67 computer. The object is to represent the cell in such a way as to simulate its growth in real environments of a living cell, as well as during changes from one chemical environment to another. This simulation could also supply useful information for answering current questions in biology.*

*In Part 1 the authors consider the criteria for the success of the simulation, its reality and potential usefulness, and describe the method of simulation, as well as that of allosteric inhibition.*

## SOMMAIRE

*Cette étude s'occupe du problème d'une simulation réaliste et utile d'une cellule vivante: la bactérie simple et unicellulaire Escherichia coli, le programme étant écrit en Fortran IV pour un ordinateur IBM 360/67. Le but c'est de représenter la cellule de telle façon à simuler la croissance de celle-ci dans le milieu véritable d'une cellule vivante ainsi que pendant les changements d'un milieu chimique à un autre. Cette simulation pourrait aussi fournir des renseignements utiles en répondant à des questions courantes dans la biologie.*

## 1. CRITERIA FOR SUCCESS

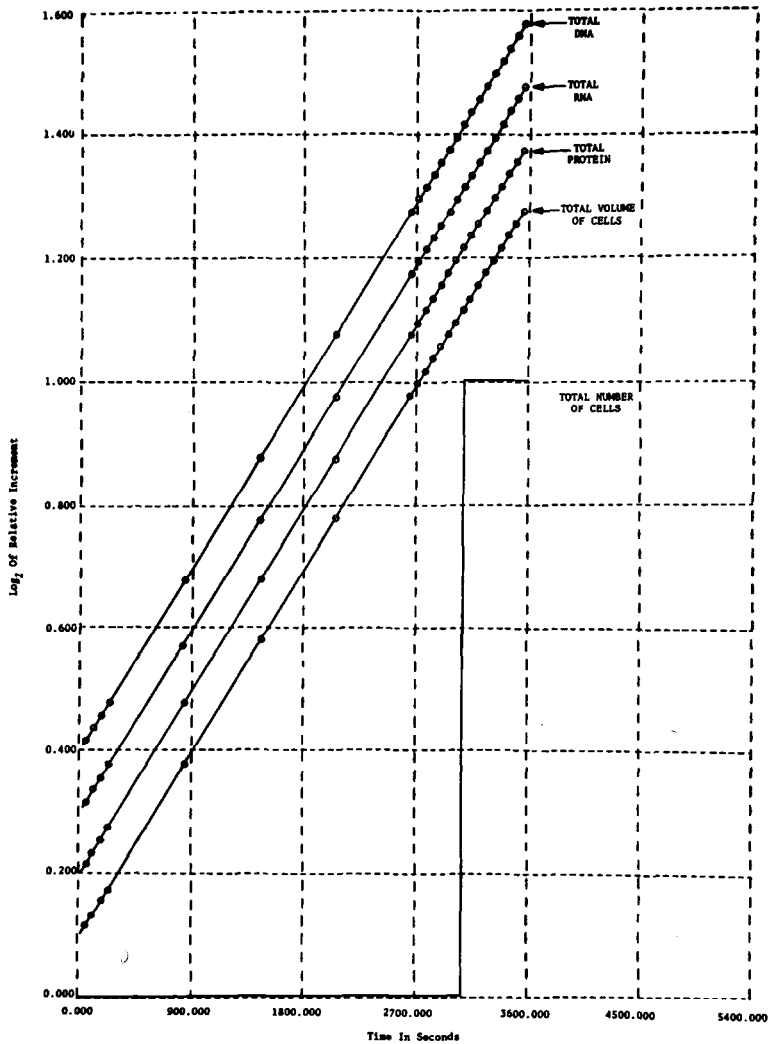
### *Reality of the Simulation*

(a) The simulated cell has realistically simulated growth curves of a living cell in the real environments used to adjust the parameters in the simulation. Since a

---

\* Supported by the Department of Health, Education and Welfare, NIH, Bethesda, Maryland (USA).

† Present address: Dept of Statistics and Computer Science, Kansas State University, Manhattan, Kansas (USA).



Logarithmic Increase in Cell Mass over Time Stepwise Increase in Number of Cells

Comparative magnitude of various quantities are a function of the scaling factors used in order to plot all quantities on one graph. A cell doubles after a DNA replication cycle. The doubling takes a relatively short time, as indicated by the sudden, stepwise increase in 'TOTAL NUMBER OF CELLS', where as 'TOTAL DNA' increases throughout the replication cycle.

Fig. 1. Logarithm of various quantities in a growing culture.

vast amount of experimental data on *E. coli* are available in the literature, it was easy to check the behavior of the simulated cell against data from the real world. The simulated cell produced chemicals and cell mass at logarithmic rates, and duplicated in stepwise fashion, just as the real cell does (Fig. 1). Since the simulated cell produced these realistic growth curves from a complex interaction of many equations, the growth curves are a good preliminary confirmation of the simulation.

(b) The simulated cell should realistically simulate growth during changes from one chemical environment to another. Since the control equations were fitted for growth in different fixed environments, but not for behavior during a shift in environmental conditions, the behavior of the simulated cell during a change in its environment comprises a test of new conditions not used to adjust the parameters in the simulation. Although the simulated cell could adapt to changes in its environment, further work is necessary to resolve artificial fluctuations in ATP and ADP concentrations following the shift.

(c) The simulated cell should predict real-cell behavior in environments not used to adjust the parameters in the simulation. The most stringent test would be a comparison with an experimental situation in which the pools represented in the simulation are available as experimental measurements with which the simulation must agree. Growth in a new, completely defined chemical medium by the real cell will be used as the data with which the simulated cell must agree. The response of the program changes both in different simulated environments, and as a function of modifications of the hypotheses used to formulate the equations in the program. This output response serves as data for analysing the interactions occurring among the hypothetical subsystems on which the program is based. Interesting and valuable works have been programmed for solving theoretical mathematical equations in the study of feedback controls of individual pathways in metabolism (Koch, 1967; Garfinkel, 1966) and to study repression or induction of one gene (Griffith, 1968). Garfinkel has simulated the glycolytic pathway in elegant detail for higher protists. Heinmetz (1964) has solved differential equations on an analog computer in order to study repression of one enzyme. I am synthesising in one program for a digital computer interacting systems important and necessary for replication and growth of a real, specific living cell, *Escherichia coli* (Weinberg, 1968a, b, c; Weinberg and Berkus, 1969). My simulation includes homomorphic representations of many biochemical pathways, gene controls, and mechanisms for cell replication, permitting studies of interactions of these vital organisational entities in ways not intended or possible with simulations limited to at most a few metabolic pathways (e.g. Garfinkel, 1966).

Simulations have been written for abstract systems which resemble living cells in their general organisation. These simulations were not models of any particular cell, and were not designed so that the output of the simulation could be checked against experimental data (Heinmetz, 1966; Yeisley and Pollard, 1964; Tsanev and

Sendov, 1966; Stahl, 1967). Yeisley and Pollard have simulated production of metabolic intermediates, emphasising organisation of sequences rather than including feedback, repression, and interaction of metabolic pathways to control enzyme activity and production. Heinmetz has simulated a general model for an abstract cell on an analog computer with no attempt at comparison of the results of the simulation with data from the real world (1966). Tsanev has simulated an abstract system for regulation of multiplication of a diploid cell, again with emphasis on making a workable model rather than on comparison of predictions of the simulation with real experimental data. Although simulating a real cell requires careful literature searches, and painstaking accuracy as to metabolic topology and concentrations, my realistic simulation of *E. coli* can be used in ways neither envisaged nor accessible for the abstract models mentioned. I can predict experimental results for real organisms, and can improve my simulation as I use it by correcting discrepancies between the simulation's predictions and real experimental results.

#### *Future Usefulness of the Computer Simulation of a Living Cell*

I will try to show that the simulated living cell is useful by demonstrating that it can be applied to current questions in biology. I will choose questions concerning interrelationships among metabolic pathways and control mechanisms which are important in the establishment of a realistic and stable simulation. The answers to the questions will therefore serve the two-fold purpose of illustrating the simulation's usefulness, and making it even more useful.

(1) Typical questions relating to the interactions among metabolic pathways and cellular control mechanisms are: Can the simulated ATP/ADP controls, along with the simulated operon, replicon and allosteric inhibition account for the stability of a real cell in changing environments? Is the simulated cell's stability lost when the ADP/ATP controls are relaxed? Does charged and uncharged TRNA as a control on protein synthesis lead to simulated results corresponding to the difference between normal cells and relaxed mutants?

(2) The simulation should add information about the concentrations of major cell constituents during complex changes in environmental conditions. Since all of the 22 pools used to represent metabolism are represented at all times during the simulation, the concentrations of these pools will be calculated in every program run. Many of these concentrations are not measurable in some of the relevant biological experiments, although the concentrations may influence important factors being studied. Availability of realistic approximations of all twenty-two concentrations during many different conditions will be useful for theoretical analyses of metabolism.

(3) The simulation should be useful for theoretical studies not amenable to direct experimental analysis in the biological laboratory. Adding new metabolic pathways, or changing the efficiency of old metabolic pathways, is easily implemented in the simulation, with automatic prediction of the concentrations of the 22 pools of biochemicals represented. Viability of the simulated cell may be defined as the ability to duplicate at a rate similar to the duplication rate of a real cell in a similar environment. Many possible alterations can be analysed as to effect, and some chosen for their feasibility for experimental analysis using criteria such as the viability of the altered cell, or the simplicity of the alteration in terms of genetic changes necessary to produce the alteration.

(4) The computer simulation should be a working simulation of a living cell with evolutionary potentialities for the simulation as well as for the model being simulated.

Living organisms adapt to their environments in two ways; during evolution and at the level of the individual organisms. At the individual level, an organism adapts over its own life span by changing its phenotype in response to information it receives from its external and internal environments. The simulated cell accomplished this type of phenotypic adaptation when it maintained stability in changing simulated chemical environments (Weinberg, 1968a, b, c).

In an evolutionary sense, populations of organisms adapt over many cycles of reproduction, with more fit organisms selected for greater ability to survive and to produce offspring. A scheme for evolutionary adaptation of the simulated cell will be outlined (Weinberg and Berkus, 1969), and its validity supported by theoretical as well as practical considerations (Holland, 1968a, b; 1969a, b). DNA of a cell can be represented by an array in which is stored information concerning the various rate constants in the equations representing a simulated cell. Evolving DNA can be simulated by modifying arrays representing the DNA of different cells in a population. The evolutionary operators crossing over, inversion, mutation and selection can be simulated by a genetic program written to modify the contents of the DNA arrays. Fundamental theories of population genetics by Fisher (1958), Kimura (1965) and Holland (1968a, b; 1969a, b) assure one of the effectiveness of this type of modification in achieving desired goals. An example of a startling and interesting prediction arising from Holland's theory is the global optimisation available to the reproductive plan through the use of local optimisation techniques. The theorem makes heavy use of a series of developments in economics (Gale, 1967). Not only does a reproductive plan exist for such an accomplishment, but many different goals may be reached by using the same plan over much of the evolutionary path. The turnpike theorems from economics which Holland relates to reproductive plans indicate the existence of such robust plans. The extraordinarily efficient sampling and search technique of natural populations becomes apparent within the framework of Holland's formal adaptive system. This efficient sampling

technique has profound implications as to the kinds of task possible to an evolving population.

## 2. COMPUTER SIMULATION OF A LIVING CELL

We are trying to write a realistic and useful computer simulation of a living cell. The enormous advances in molecular biology in the past fifteen years have greatly increased our understanding of the basic sub-systems in the living cell. DNA was found to be the molecule on which the cell records information to be passed on

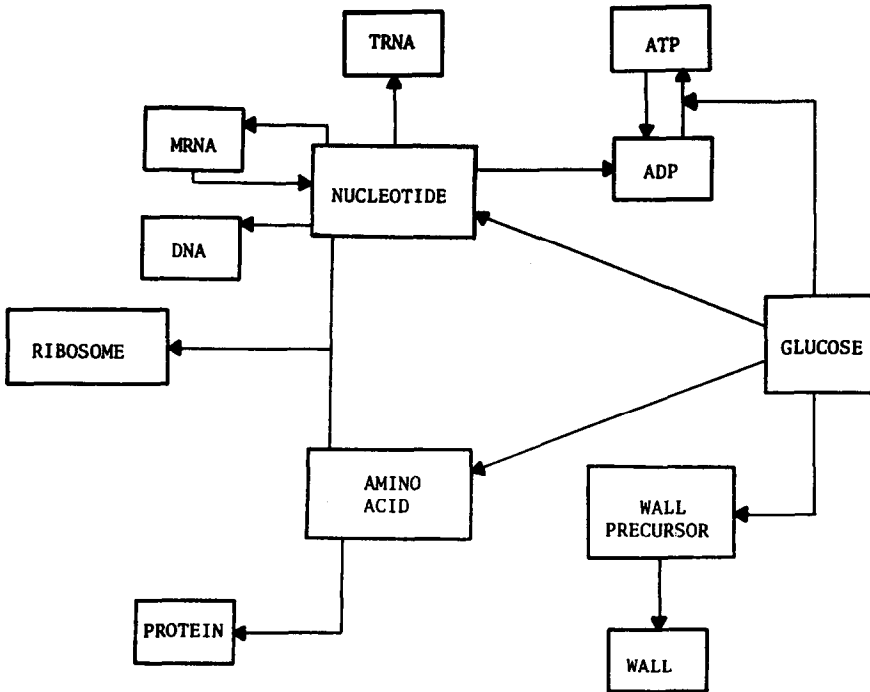


Fig. 2.1. Flow of materials.

to the next generation. M. H. F. Wilkins, James D. Watson and Francis H. C. Crick established the precise structure of the DNA molecule. The molecular code used by DNA to record the information to be passed on has been successfully analysed. Viable hypotheses have been formulated for translation of genetic information into cellular activity and structure. Experiments have yielded support for specific theories concerning the control of cellular activity at the level of DNA function. Control mechanisms have also been discovered which control a given cellular process by modifications of molecules far removed from DNA. Biology is

entering a new phase. The interactions of the various subsystems analysed by the molecular biologists are becoming objects of study. These subsystems interact among themselves, and with the environment external to the cell. Current research on regulation of DNA replication (Clark, 1968) and regulation of enzyme activity (Atkinson, 1966; Murray and Atkinson, 1968), as well as numerous experimental and theoretical studies of molecular interactions within the cell (Davis *et al.*, 1968) have made it possible to model a living cell in which both metabolism and cell replication are represented.

In a study of interacting systems, the large, high-speed digital computer is an enormous aid. One can write down hypotheses in the form of logical and analytic equations. These equations form the program. Environmental conditions can be given to the computer as input data. If the program represents a living cell, the program generates cell behavior as output.

In order to analyse interactions among DNA, RNA, protein and the activities of self-reproduction and metabolism in a living cell, a simulation of the simple unicellular bacterium *Escherichia coli* was written as a program in Fortran IV for an IBM 360/67 digital computer (Weinberg and Berkus, 1969). The cell was represented in terms of (1) the concentrations of 22 pools of chemicals (Fig. 2.1) and (2) equations representing functional relationships among pools (Figs. 2.2 and 2.3). These equations were used to predict the change in concentration and total amount of chemicals in each pool over time (Figs. 2.3, 2.4 and 2.5).

Homomorphic mappings are often used from the real world onto the models used for computer simulations. One must be aware of the approximate nature of these homomorphisms, and consequent limitations on the predictive ability of the simulation. An example of a domain for such a mapping would be the 22 pools of chemicals in the simulation of a living cell used to represent corresponding pools in a real living cell. I hoped that these pools would create equivalence classes of chemicals with substitution property over time. The substitution property is usually approximate, and limits the domain in which the predictions of the simulation are accurate. This approximation does not necessarily invalidate the model, but should be kept in mind while using it. One can often use imperfect predictions from imperfect models in a Bayesian sense, and work with the new probabilities one obtains to select possible courses of action.

The model of the cell was simplified in order to make the computer simulation manageable. The chemical constituents of the cell were lumped into the aforementioned pools (Fig. 2.1). The pool of proteins was further divided into different enzyme groups: each enzyme group was associated with one group of chemical reactions responsible for converting one pool into another pool in the simulated cell. For example the enzyme group EK2 catalysed the production of amino acids from carbohydrates. The messenger RNA pool was subdivided into a separate messenger RNA for each enzyme group. Variables used in the Fortran program for the simulation of a living cell are defined in the appendix (see Part 2).

```

91 DDNA = K(6)★NUC★DT★EK(6)★ATP
   DDNA1 = K(6)★(T/DBLE)★.5★IN1★NUC★DT★EK(6)★ATP
   DDNA2 = K(6)★(T/DBLE)★.5★IN2★NUC★DT★EK(6)★ATP
   DDNA3 = K(6)★(T/DBLE)★.5★IN3★NUC★DT★EK(6)★ATP
C★★★★ 40 MINUTES TO REPLICATE .5 ★ DNA
   DO 100 I = 1,10
100 DRNK(I) = (K8K(I)★NUC★DNA★EK(8)★ATP - KDRNK★RNK(I))★DT
C★★★★ DMRNA = DRNK(1) + DRNK(2) + DRNK(3) + DRNK(4) + DRNK(5)
   + DRNK(6)
1 + DRNK(8) + DRNK(9) + DRNK(10) + DRNK(7)
C★★★★ DTRNA = K(10)★NUC★DT★EK(10)★ATP
   DRIB = K(9)★NUC★AA★DT★EK(9)★ATP
   DRNA = DMRNA + .25★DTRNA + .75★DRIB
   DWALL = K(4)★GLUC★DT★EK(4)★ATP
C★★★★ DO 101 I = 1,10
101 DEK(I) = K(7)★AA★(RNK(I)/MRNAO)★DT★EK(7)★ATP
C★★★★ DPRTN = DEK(1) + DEK(2) + DEK(3) + DEK(4) + DEK(5) + DEK(6)
   + DEK(8) + DEK(9) + DEK(10) + DEK(7)
C★★★★ DNUC =
1 + K(1)★GLUC★DT★EK(1)★ATP - (2.5E4/660.)★DTRNA
2 - (2.E6/660.)★DRIB - K(5)★NUC★DT★EK(5)★ATP
C★★★★ DAA = K(2)★GLUC★DT★EK(2)★ATP - 1.E6/102.★DRIB
   - (4.E4/102.)★DPRTN
   DATP = K(3)★GLUC★DT★ATP★EK(3) - DNAP★DDNA
   - MRNAP★DMRNA
1 - TRNAP★DTRNA - BIBP★DRIB - PRTNP★DPRTN
   - WALLP★DWALL
2 - (AAP★K(2)★GLUC★EK(2)★ATP
   + NUCP★K(1)★GLUC★EK(1)★ATP
   + 2★K(5)★NUC
3 ★EK(5)★ATP )★DT
C★★★★ DADP = -DATP + K(5)★NUC★DT★EK(5)★ATP
   DVOL = K(14)★WALL★DT
C★★★★ INCREASE IN VOLUME PER UNIT INCREASE IN CELL WALL

```

Fig. 2.2. *Differential Equations*: quantity to the left of = is the change in amount of the substance; e.g. DDNA represents the change in the amount of DNA in one time increment DT. The differential equation underlying the first equation is

$DDNA = K(6)★NUC★EK(6)★ATP★DT$  for a discrete time interval DT. As DT approaches 0, we get the underlying continuous differential equation

$$\lim_{DT \rightarrow 0} \frac{D(DNA)/DT}{DT} = d(DNA)/dt = K(6)★NUC★EK(6)★ATP$$



```

ADJST = VOL0/(VOL0 + DVOL)
C★★★★
VOLN= VOL★EXP (DT★ K(14)★WALL/VOL)
C★★★★
DNA = (DNA + DDNA)★ADJST
DNA1 = (DNA1 + DDNA1)★ADJST
DNA2 = (DNA2 + DDNA2)★ADJST
DNA3 = (DNA3 + DDNA3)★ADJST
DNA1T = DNA1★VOL
DNA2T = DNA2★VOL
DNA3T = DNA3★VOL
MULT = 0.
      IF (DNA3T - 1.000010) 151,151,150
150 MULT = MULT + 2.
      GO TO 153
151 CONTINUE
      IF (DNA3T - .1) 153,153,152
152 MULT = MULT + 1.
153 CONTINUE
      IF (DNA2T - 1.000010) 155,155,154
154 MULT = MULT + 2.
      GO TO 158
155 CONTINUE
      IF (DNA2T - 0.1) 157,157,156
156 MULT = MULT + 1.
157 CONTINUE
      IF (DNA1T - 1.000010) 159, 159,158
158 MULT = MULT + 2.
      GO TO 161
159 CONTINUE
      IF (DNA1T - 0.1) 161, 161,160
160 MULT = MULT + 1.
161 CONTINUE
DIN = MULT★KIN★(AA/AA0)★DT
IN = IN + DIN
DO 102 I = 1,10
102 RNK(I) = (RNK(I) + DRNK(I)) ★ ADJST
C★★★★
MRNA = (MRNA + DMRNA)★ADJST
TRNA = (TRNA + DTRNA)★ADJST
RIB = (RIB + DRIB)★ADJST
RNA = (RNA + DRNA)★ADJST
WALL = (WALL + DWALL)★ADJST
C★★★★
DO 103 I = 1,10
103 EK(I) = (EK(I) + DEK(I)) ★ ADJST
C★★★★
PRTN = (PRTN + DPRTN)★ADJST
NUC = (NUC + DNUC)★ADJST
AA = (AA + DAA)★ADJST
ATP = (ATP + DATP)★ADJST
ADP = (ADP + DADP)★ADJST

```

Fig. 2.3. New Concentrations. The amount of new material made during one time increment  $DT$  is added to the amount of old material present at the beginning of the time increment, and the new concentration is obtained by adjusting for the increase in cell volume during the time increment.

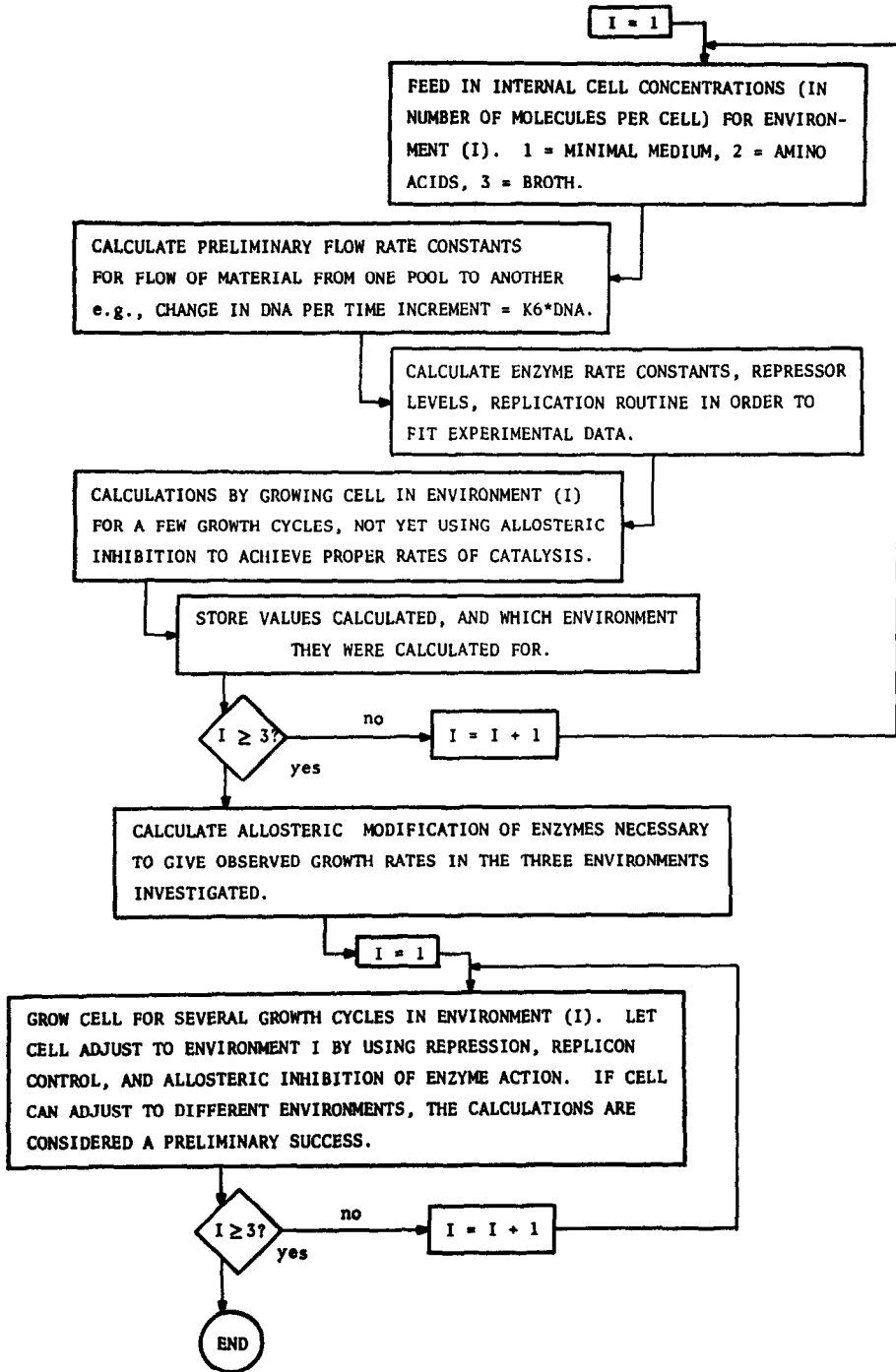


Fig. 2.4. Summary of program.

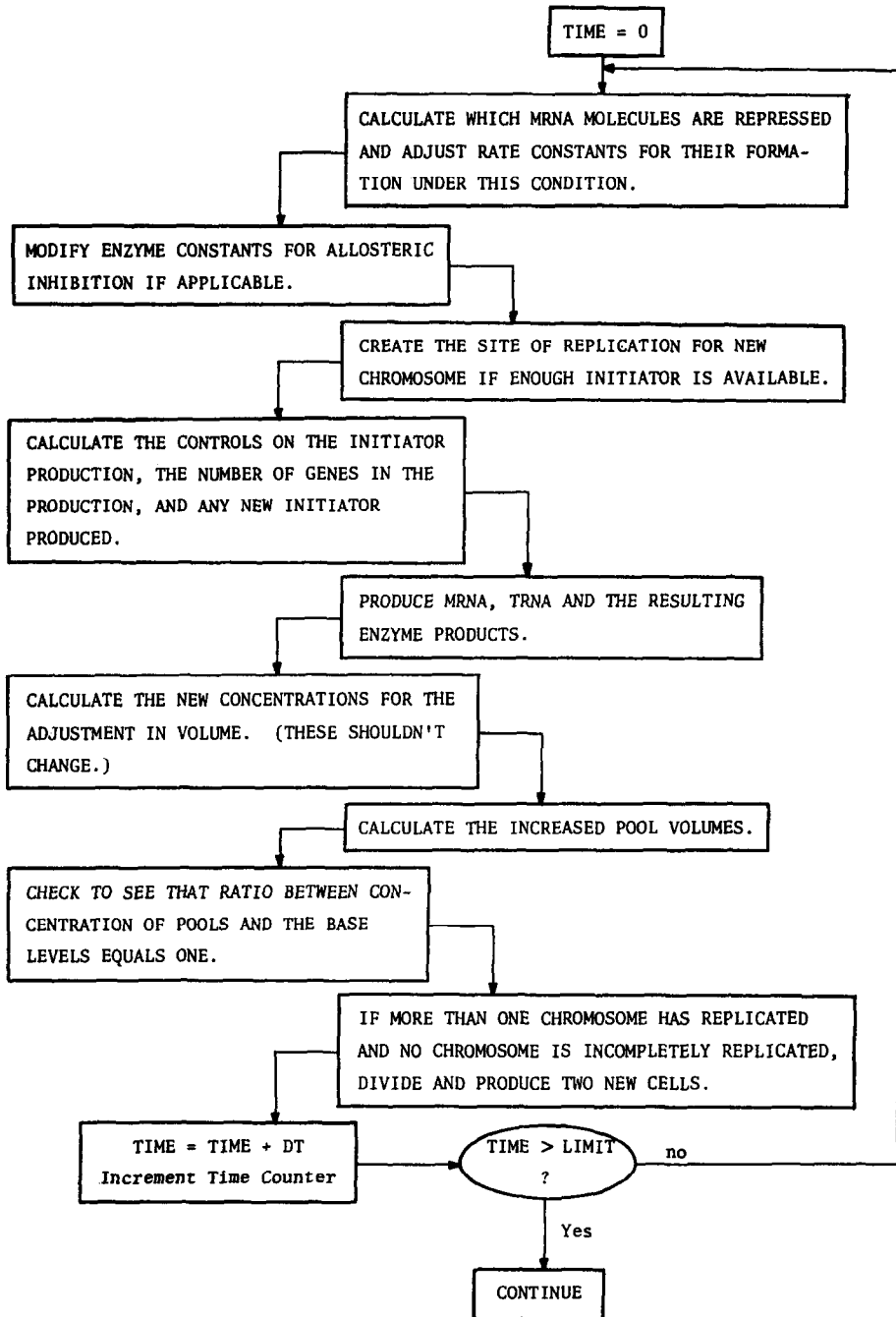


Fig. 2.5. Growth cycle.

```

1      T = 3000.
C★★★★ GENERATION TIME 50 MINUTES
      NO = 1.
      DT = 1.
      CHRMO = 2.
      FACTR = 2.
C★★★★ 2 CHROMOSOMES AT TIME 0, BOTH REPLICATE IMMEDIATELY
C★★★★ DBLE = 50★60
C★★★★ 50 MINUTES FOR 1 CHROMOSOME TO REPLICATE
      DNA1Z = 1.
      DNA2Z = 1.
      DNA3Z = 0.
      DNAO = DNA1Z + DNA2Z + DNA3Z
      IN1Z = 1.
      IN2Z = 1.
      IN3Z = 0.
      IN11Z = 0.
      IN21Z = 0.
      IN31Z = 0.
      INZ = 0.
      MRNAO = 1.E3
      TRNAO = 4.E5
      PRTNO = 1.E6
      NUCO = 1.2E7
      RIBO = 1.5E4
      RNAO = MRNAO + .25★TRNAO + .75★RIBO
C★★★★ CONSIDER WEIGHT OF RNA 1E6
      AAO = 3.E7
      GLUCO = 40.E-03★1.E-15★2.25★(1/180.)★6.02E23
      WALLO = 2.25E8
      ATPO = 173.E-14★6.02E23★1.E-06
      ADPO = 23.E-14★6.02E23★1.E-06
      LN2 = ALOG(2.)
      DO 2 II = 1,14,1
2      PRDCO(II) = PRDCO(II)★LN2
      RNAO = RNAO★LN2
      VOLO = 1.
      CAA = 0
      BROTH = 0
      COUNT = 0.
      IF(CNTRL.EQ.1)CRAZY=0
3      DNAO = DNA1Z + DNA2Z + DNA3Z
      RNA = RNAO
      DO 4 II = 1,14
4      PRDC K(II) = PRDCO(II)
      DNA1 = DNA1Z
      DNA2 = DNA2Z
      DNA3 = DNA3Z
      DNA1T = DNA1★VOL
      DNA2T = DNA2★VOL
      DNA3T = DNA3★VOL
      IN1 = IN1Z
      IN2 = IN2Z
      IN3 = IN3Z
      IN11 = IN11Z
      IN21 = IN21Z
      IN = INZ
      IN31 = IN31Z

```

Fig. 2.6. Cell in Environment 1 (Minimal Medium).

The environment simulated was a chemically defined liquid growth medium, at a temperature of 37°C with an abundance of oxygen. The changes in the environment simulated were changes in the chemical constituents of the media. Three different media were 'fed' to the simulated cell: (1) a medium containing glucose, ammonium salt, and minerals (Fig. 2.6), (2) a medium containing glucose, minerals, and amino acids (Fig. 2.7), (3) a medium containing glucose, minerals, amino acids and nucleosides (Fig. 2.8). With successive additions of amino acids, and amino acids and nucleosides the cell grew faster since it had fewer molecules to make on its own. This agreed with experimental data from literature on real cells (Maaloe and Kjeldgaard, 1966). The equations in the simulation were tested by comparing the growth of the simulated cell with that of a real cell (Figs. 2.4 and 2.5). The simulated cell and the real cell both took 50 minutes to reproduce in mineral medium containing ammonium and glucose, 28 minutes in medium to which amino acids had been added, and 25 minutes in medium containing both amino acids and nucleosides.

The number of DNA molecules in the simulated cell at the beginning of a cycle of cell reproduction was 2 for mineral medium, and 3 for both amino acid medium and amino acid nucleoside medium. This amount of DNA agrees with data for real cells (Lark, 1966).

One can see from Fig. 2.9 that the cell could use allosteric modification of its enzyme pools to calculate rate constants which agreed rather well with the rate constants necessary for realistic growth. All of the rate constants agreed in the first

```

MRNA0 = 1.1★MRNA0
TRNA0 = .9★TRNA0
ATP0 = 1.1★ATP0
PRTN0 = 1.1★PRTN0
CAA = 1
C★★★★
VOL0 = 92.2/50.7
DNA1Z = 1./VOL0
DNA2Z = 1./VOL0
DNA3Z = 1./VOL0
T = (60./2.14)★60.
CHRM0 = 3.
FACTR = 3.
DBLE = T
IN3Z = 1./VOL0
IN31Z = 1.
RIB0 = ((2.14 - 1.20)/(2.4 - 1.20)★((250./135.)★1.5E4 -
1 1.5E4) )★LN2
RNA0 = MRNA0 + .25★TRNA0 + .75★RIB0
AA0 = 1.5E8★LN2
C★★★★ ADDITION FOR SOLVE
NUC0 = 1.E7
ADP0 = 38./62.)★ADP0
GLUC0 = GLUC0★VOL0★3.
WALL0 = WALL0★VOL0

```

Fig. 2.7. Cell in Environment 2 (Minimal Medium + Amino Acids). These are the quantities which change upon addition of amino acids to minimal medium.

```

TRNA0 = 2.★TRNA0
MRNA0 = MRNA0★1.1
PRTN0 = PRTN0★1.1
BROTH = 1
CAA = 1
C★★★★
VOL0 = 117./50.7
DNA1Z = 1./VOL0
DNA2Z = 1./VOL0
DNA3Z = 1./VOL0
T = 25.★60
DBLE = T
NUC0 = 6.E8★LN2
C★★★★
ADDITION FOR SOLVE
AA0 = 3.E8
RIB0 = (250./135.)★1.5E4★LN2
RNA0 = MRNA0 + .25★TRNA0 + .75★RIB0
ATP0 = 1.1★ATP0
ATP0 = (59./62.)★23.E-14★6.02E23★1.E-6★LN2 ★1.1
GLUC0 = 40.E-3★1.3-15★2.25★(1./180.)★6.02E23★LN2 ★2.
WALL0 = 2.25E8★VOL0★LN2

```

Fig. 2.8. Cell in Environment 3 (Broth). These are the quantities which are different in broth than in minimal medium.

C ( 1) = 1.469818E-12	C ( 1) = 1.835376E-12
C ( 2) = 2.183038E-12	C ( 2) = 2.294225E-12
C ( 3) = 1.177201E-10	C ( 3) = 1.177268E-10
C ( 4) = 6.903309E-12	C ( 4) = 6.907760E-12
C ( 5) = 4.533463E-15	C ( 5) = 4.531281E-15
C ( 6) = 1.110285E-20	C ( 6) = 1.110301E-20
C ( 7) = 1.539183E-15	C ( 7) = 1.539176E-15
C ( 8) = 9.812282E-17	C ( 8) = 9.812185E-17
C ( 9) = 2.775713E-24	C ( 9) = 2.775785E-24
C(10) = 1.539183E-15	C(10) = 1.539060E-15
Rate constants calculated at start of program	Rate constants by simulating 'Allosteric Modification'

Fig. 2.9. The Cell Successfully Adjusted its Enzyme Rate Constants by Allosteric Modification of its Enzymes. E indicates multiplication by a power of 10, e.g. 1.469818E-12 = 1.469818 · 10<sup>-12</sup>.

figure with the proper rate constants when the calculation was done by the equations representing allosteric modification for the simulation, and by observation of experimental data for proper rate constants. However, the rate of production of ATP and ADP changed when the allosterically modified enzymes were used in the simulation. This is shown in Figs. 2.10 and 2.11. RADP represents the ratio of

```

RDNA = 9.999995E-01 RMRNA = 9.999996E-01
RNUC = 1.000000E 00 RAA = 9.999992E-01
RDNA1 = 9.999999E-01 RDNA2 = 9.999999E-01
RTRNA = 9.99998E-01 RRIB = 9.999996E-01 RWALL = 1.000000E 00 RPRTN
= 1.000000E 00
RATP = 9.999976E-01 RADP = 1.000014E 00
RRNA = 9.999992E-01

```

Fig. 2.10. Ratio of concentrations of chemicals at the end of one time step to base concentrations cell should maintain, for rate constants calculated at start of program. E indicates multiplication by a power of 10, e.g. 9.999995E-01 = 9.999995 · 10<sup>-1</sup>.

RDNA = 9.999990E-01 RMRNA = 9.999993E-01  
 RNUC = 1.000413E 00 RAA = 1.000221E 00  
 RDNA1 = 9.999999E-01 RDNA2 = 9.999999E-01  
  
 RTRNA = 9.999995E-01 RRIB = 9.999992E-01 RWALL = 1.000000E 00 RPRTN  
 1.000000E 00  
 RATP = 6.329882E-01 RADP = 3.760564E 00  
 RRNA = 9.999984E-01

Fig. 2.11. Ratio of concentrations of chemicals at the end of one time step to base concentrations cell should maintain, for rate constants obtained by simulating allosteric modification. E indicates multiplication by a power of 10, *e.g.* 9.999990E-01 =  $9.999990 \cdot 10^{-1}$ . A ratio of 1 indicates success. Note that all ratios are 1 (approximately) with the important exceptions of ADP and ATP using allosteric modification.

ADP at the end of one time increment to the ratio the simulated cell should maintain for proper growth in minimal medium. RADP should be  $1.000000 \pm 0.00001$  if the cell is growing correctly. RADP = 3.760564 instead of 1.000000 (Fig. 2.11), and RATP = 0.6329882 instead of 1.000000 (Fig. 2.11). The ratios of other chemical pools are approximately 1 when allosteric modification of enzyme pools is used to simulate real control of metabolism (Fig. 2.11).

The artificial fluctuation of ATP concentration and ADP concentration did not agree with data for real cells. The errors may have been introduced by approximating changes in concentrations of the pools during a one-second time interval by the rate of change of the pools at the beginning of the one-second interval. The fault in the program is being corrected by introducing predictor corrector methods (Schied, 1968), by reducing the size of the time interval used for calculating successive concentrations for the pools in the simulation, and by using evolutionary techniques (section 4, Part 2).

This error illustrates one of the many constraints imposed upon one in realising a computer simulation. The hardware and software one uses to obtain the advantages of rapid computational ability and large memory have limitations. The computer uses digital arithmetic to accomplish calculations involving real numbers. In mapping the field of real numbers onto a finite field one may lose track of some of the implications in terms of numerical errors. One may lose the associative and distributive properties of the real field one is attempting to simulate. Several common types of errors must be acknowledged, and their baleful influence avoided where possible. Roundoff error results from approximating a real number of a finite string of digits. Truncation error results from approximating an infinite series by a finite series in a numerical technique. Experimental error may be implicit in the data one analyses from the real world (Hildebrand, 1956).

The simulated cell produced chemicals and cell mass at a logarithmic rate, but duplicated in a stepwise fashion (Fig. 1.1) just as the real cell does. Since the simulated cell produced these smooth growth curves from a complex interaction of many equations, the growth curves are a good preliminary confirmation of the models used to write the simulation.

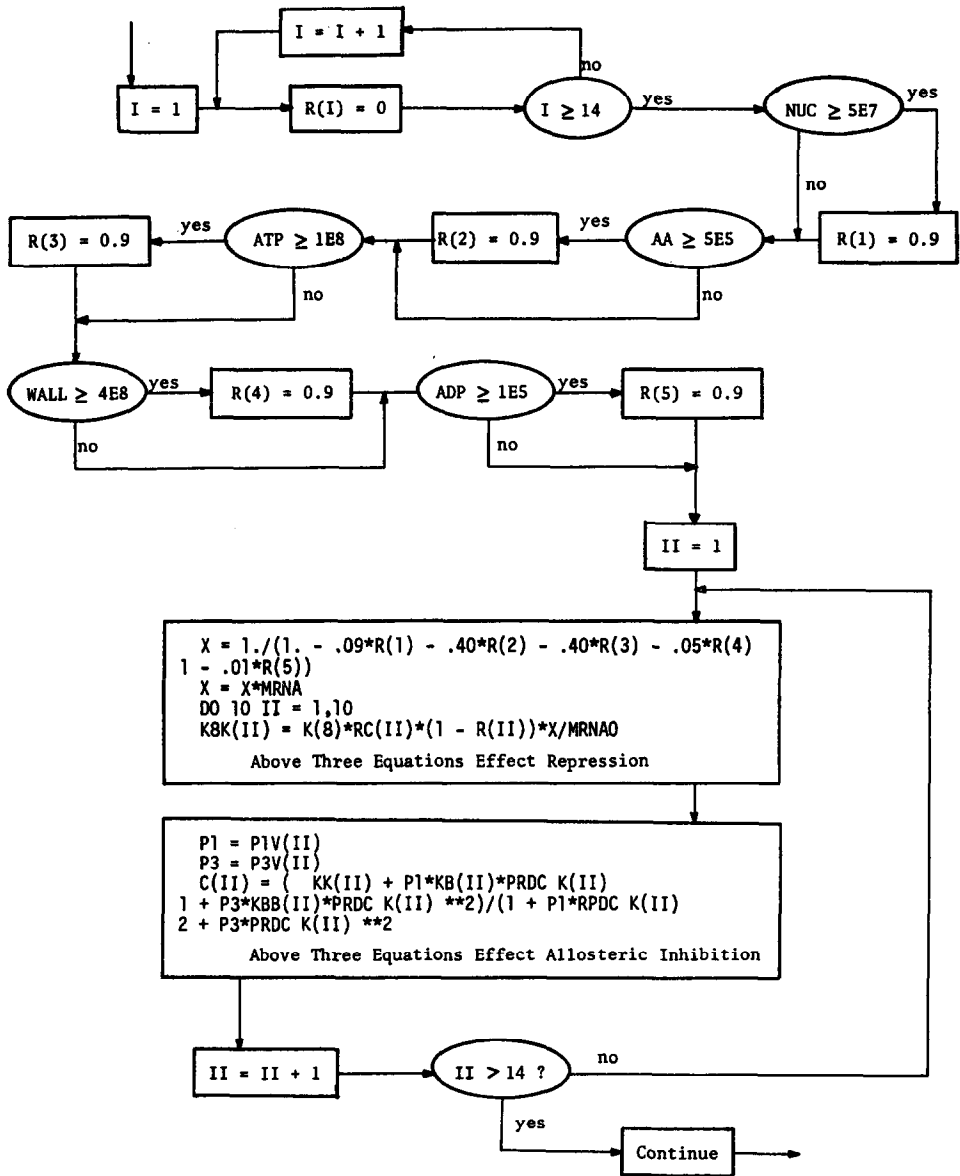


Fig. 2.12. Repression and allosteric inhibition. Repression is obtained by adjustment of KK8K (INTGR). Allosteric inhibition is obtained through adjustment of C (INTGR).



The simulated cell employed repression to control the production of its enzymes (Fig. 2.12). Repression operated at the DNA level. For example EK2 was the enzyme pool needed for producing amino acids from carbohydrates. EK2 was produced under control of DNA by way of the RNA pools as long as the amino-acid pool concentration was below a certain critical level. DNA directed the production of messenger RNA specific for the production of EK2. EK2 was produced by hooking together amino acids attached to transfer RNA. This hooking was done by messenger RNA, attached to ribosomes. If the amino-acid level rose above the critical level, production by DNA of messenger RNA responsible for EK2 production was sharply curtailed; the messenger RNA already present rapidly decayed, and almost no new messenger RNA for EK2 production was produced.

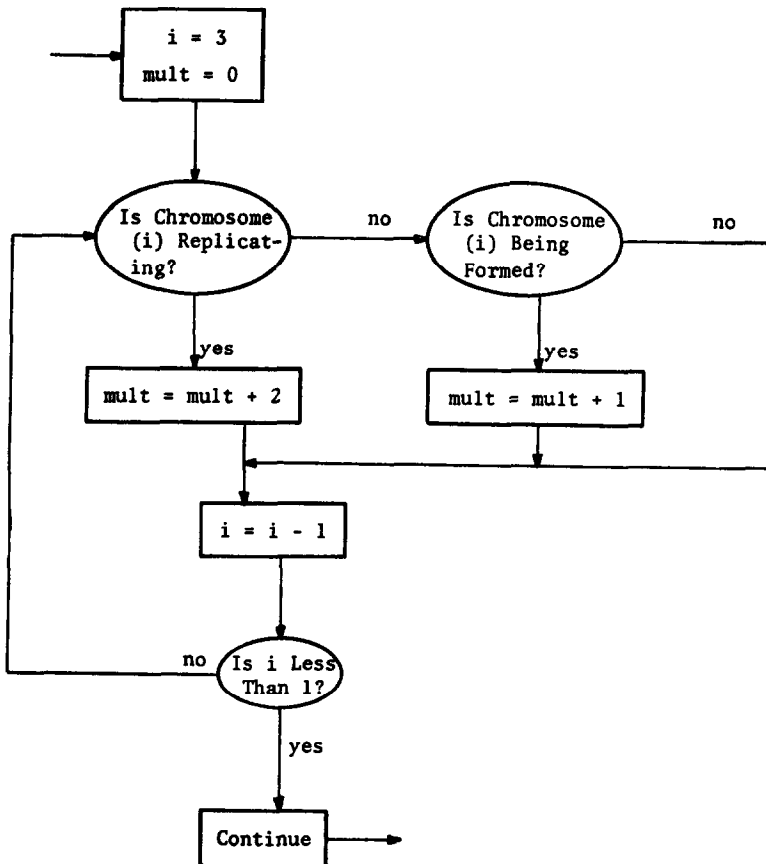


Fig. 2.13. Number of genes producing initiator substance into cytoplasm. Number of genes = MULT. The number of genes increases by one for every chromosome which has multiplied greater than ten percent of its length.

In DNA replication a new DNA molecule was produced using the old DNA molecule as a model. In cell division, a cell split into two new cells by forming a cell wall partition through its center. DNA replication (Figs. 2.13, 2.14, 2.15 and 2.16) and cell division (Figs. 2.17 and 2.18) were represented as two separate though interdependent processes.

```

MULT = 0.
  IF (DNA3T - 1.000010) 151,151,150
150 MULT = MULT + 2.
  GO TO 153
151 CONTINUE
  IF (DNA3T - .1) 153,153,152
152 MULT = MULT + 1.
153 CONTINUE
  IF (DNA2T - 1.000010) 155,155,154
154 MULT = MULT + 2.
  GO TO 158
155 CONTINUE
  IF (DNA2T - 0.1) 157,157,156
156 MULT = MULT + 1.
157 CONTINUE
  IF (DNA1T - 1.000010) 159,159,158
158 MULT = MULT + 2.
  GO TO 161
159 CONTINUE
  IF (DNA1T - 0.1) 161,161,160
160 MULT = MULT + 1.
161 CONTINUE

```

Fig. 2.14. Number of genes producing initiator substance into the cytoplasm. Number of genes = MULT. The number of genes increases by one for every chromosome which has multiplied for greater than ten percent of its length.

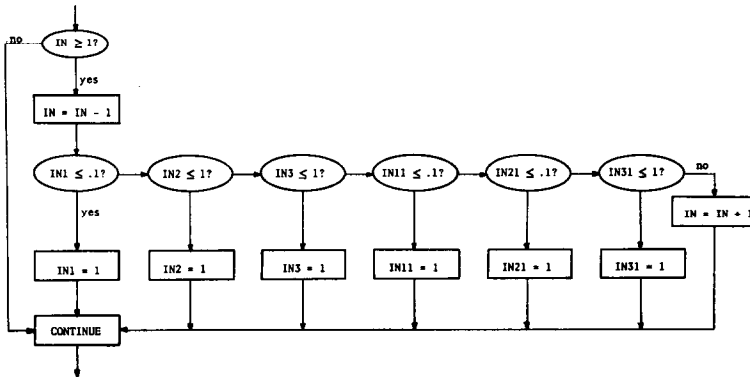


Fig. 2.15. Creation of new replication sites when cytoplasmic initiator substance is high enough (greater than one), e.g. IN1 will be the replication site created if IN is greater than or equal to one. If IN1 is already created, other sites get a chance.

```

          IF (IN - 1.) 69,55,55
55      IN = IN - 1.
56      IF(IN1 - 1.) 57,57,58
57      IN1 = 1.
          GO TO 69
58      IF (IN2 - .1) 59,59,60
59      IN2 = 1.
          GO TO 69
60      IF (IN3 - 1.) 61,61,62
61      IN3 = 1.
          GO TO 69
62      IF (IN11 - .1) 63,63,64
63      IN11 = 1.
          GO TO 69
64      IF (IN21 - .1) 65,65,66
65      IN21 = 1.
          GO TO 69
66      IF (IN31 - .1) 67,67,68
67      IN31 = 1.
68      IN = IN + 1
69      CONTINUE

```

Fig. 2.16. Creation of new replication sites when cytoplasmic initiator substance is high enough (greater than one), *e.g.* IN1 will be the replication site created if IN is greater than or equal to one. If IN1 has already been created, other sites get a chance to appear.

```

          IF (DNA1T - 2.) 81,76,76
76      IF (1.000010 - DNA2T) 77,77,78
77      IF (DNA2T - 2.) 81,78,78
78      IF (1.000010 - DNA3T) 79,79,80
79      IF (DNA3T - 2.) 81,80,80
80      NO = 2★NO
          IN = IN/2.
          IN1 = IN11
          IN2 = IN21
          IN3 = IN31
          IN11 = 0.
          IN21 = 0.
          IN31 = 0.
          COUNT = 0.
          DT = 1.
          VOL = VOL/2.

```

Fig. 2.17. Cell division occurs if conditions are appropriate.

DNA was simulated as a circular molecule. Before initiating DNA replication, an empty site for attachment of the newly produced DNA to the cell membrane had to be available. If an empty DNA site was present on the cell membrane, a DNA molecule already present in the cell would begin replication of a new DNA molecule which would attach to the empty DNA site. Once a DNA molecule started to replicate, its rate of replication was dependent on energy available as ATP (adenosine triphosphate), and on the availability of enough nucleotides to form the new DNA.

New DNA sites of attachment on the membrane were produced under control of the initiator gene (Fig. 2.13 and 2.14). High concentration of amino acid stimulated the rapid production of new sites, and therefore allowed more DNA molecules to replicate at the same time. This led to faster growth. This faster growth expressed

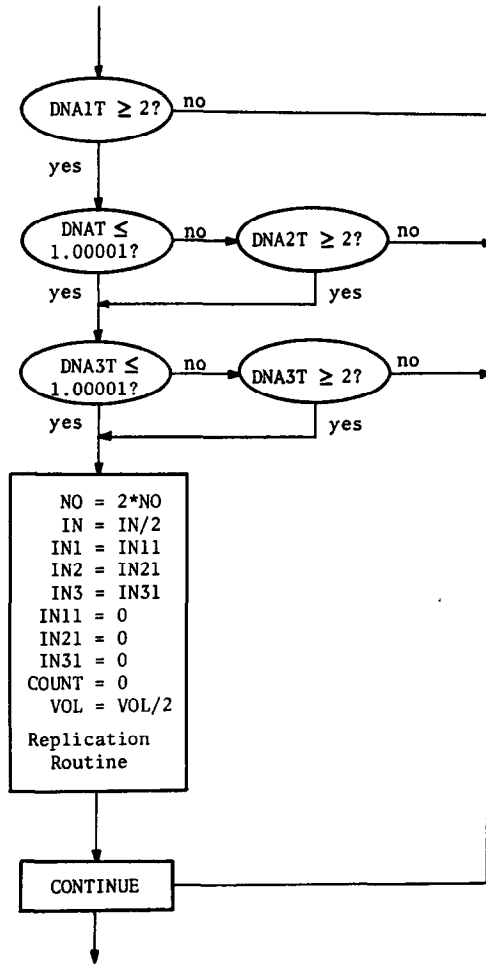


Fig. 2.18. Cell division occurs if conditions are appropriate.

the relationship between the number of replicating DNA molecules and metabolic conditions.

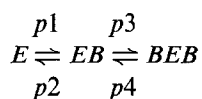
Whenever any DNA molecule had completely replicated, a decision was made by the simulated cell as to whether it should divide (Fig. 1.2). If the cell could divide without breaking any DNA strands in the process of replication, it did divide. Otherwise it waited until interfering strands completed their replication, and then looked again to see if the road was clear for division. This model of DNA replication and cell division produced simulated division rates similar to division rates of real cells growing in real media (Davis *et al.*, 1968).

Of course if the amino-acid concentration fell too low, insufficient amino acids were available for hooking together into the EK2 enzyme; production of all enzymes was blocked in the event of extreme scarcity of amino acids.

The simulated cell employed feedback inhibition to control the activity of the enzymes already present (section 3). For example EK2, the enzyme for production of amino acids from carbohydrates, appeared in three different forms: pure enzyme, enzyme with one molecule of amino acid attached to it, and enzyme with two molecules of amino acid attached to it. These three forms of EK2 had different catalytic ability. The relative amount of EK2 in each form determined the activity of the EK2 present in the cell in terms of its efficiency in converting carbohydrate into amino acids. The percentage of EK2 in each of the three forms was determined by the number of amino-acid molecules per cell volume unit (one cell volume unit was taken as the volume of a cell growing rapidly in mineral glucose medium with ammonium salt). The higher the amino-acid concentration in the simulated cell, the greater was the percentage of EK2 in its low-activity form, and the less effective was the EK2 in production of amino acids from carbohydrates.

### 3. COMPUTER SIMULATION OF ALLOSTERIC INHIBITION

#### *Feedback Inhibition Used in Simulation of a Living Cell*



Three different forms of enzyme:

$E$  = concentration of pure enzyme,

$EB$  = concentration of enzyme with one molecule of product  $B$  attached,

$BEB$  = concentration of enzyme with two molecules of product  $B$  attached.

$p_1$ ,  $p_2$ ,  $p_3$ , and  $p_4$  are rate constants for the rate of formation of various forms of the enzyme as a function of the concentrations of other forms of the enzyme, and concentrations of product  $B$ .

$$\frac{d(E)}{dt} = -p_1 \cdot E \cdot B + p_2 \cdot EB$$

$$\frac{d(EB)}{dt} = p_1 \cdot E \cdot B - p_2 \cdot EB - p_3 \cdot EB \cdot B + p_4 \cdot BEB$$

$$\frac{d(BEB)}{dt} = p_3 \cdot EB \cdot B - p_4 \cdot BEB$$

At equilibrium,

$$\frac{d(E)}{dt} = \frac{d(EB)}{dt} = \frac{d(BEB)}{dt} = 0.$$

The total concentration of enzyme is equal to the sum of the concentrations of its three alternate forms:

$$E_{total} = E + EB + BEB$$

Manipulating the equations for equilibrium conditions one can obtain the following expressions for  $E$ ,  $EB$  and  $BEB$ , where new rate constants are defined as follows:

$$P1 = \frac{p1}{p2}$$

$$P3 = \frac{p3}{p4}$$

Expressions for  $E$ ,  $EB$ , and  $BEB$ :

$$E = E_{total} \cdot \frac{1}{(1 + P1 \cdot B + P3 \cdot B^2)}$$

$$EB = P1 \cdot B \cdot E$$

$$BEB = P3 \cdot B^2 \cdot E$$

Manipulations leading to expressions for  $E$ ,  $EB$  and  $BEB$ , at equilibrium,

$$p1 \cdot B \cdot E = p2 \cdot EB$$

$$p3 \cdot B^2 \cdot E = p4 \cdot BEB$$

$$EB = (p1/p2) \cdot B \cdot E = P1 \cdot B \cdot E$$

$$BEB = (p3/p4) \cdot B^2 \cdot E = P3 \cdot B^2 \cdot E$$

$$E = E_{total} - EB - BEB = E_{total} - \frac{p1}{p2} \cdot B \cdot E - \frac{p3}{p4} \cdot B^2 \cdot E$$

Dividing out  $E$  from both sides, one obtains

$$1 = \frac{-p1}{p2} \cdot B - \frac{p3}{p4} \cdot B^2 + E_{total}/E$$

Manipulating, and substituting  $P1 = p1/p2$  and  $P3 = p3/p4$  one obtains

$$E = E_{total} \cdot \frac{1}{1 + P1 \cdot B + P3 \cdot B^2}$$

which gives the three expressions we started out to obtain. Setting equal the two equations giving  $d(B)/dt$  one obtains

$$K(k, j) \cdot E_{total} \cdot A = [KK \cdot E + KB \cdot EB + KBB \cdot BEB] \cdot A$$

Substituting in expressions for  $E$ ,  $EB$  and  $EEB$  in terms of  $E_{total}$ , one obtains

$$K(k, j) \cdot E_{total} \cdot A = A \cdot \frac{[E_{total}]}{1 + P1 \cdot B + P3 \cdot B^2} \cdot [KK + KB \cdot 1 P \cdot B + KBB \cdot P3 \cdot B^2]$$

Cancelling out  $A$  and  $E_{total}$  from both sides of the equation, and re-arranging, one gets

$$KK + P1 \cdot B \cdot KB + P3 \cdot B^2 \cdot KBB = K(k, j) \cdot (1 + P1 \cdot B + P3 \cdot B^2).$$

Since  $B$  is simply a product of the reaction  $A \rightarrow B$  the concentration of product ( $k$ ) in environment ( $j$ ) is  $PRDC(k, j)$ . One similarly indexes the  $KK$ ,  $KB$  and  $KBB$  series so that

$KK(k)$  = the rate constant  $KK$  associated with product  $PRDC(k)$

$KB(k)$  = the rate constant  $KB$  associated with product  $PRDC(k)$

$KBB(k)$  = the rate constant  $KBB$  associated with product  $PRDC(k)$

$PRDC(k, j)$  =  $B$  for product ( $k$ ) in environment ( $j$ ).

Substituting in the indexes for purposes of iterative calculations, one obtains the equation

$$KK(k) + P1 \cdot PRDC(k, j) \cdot KB(k) + P3 \cdot PRDC(k, j)^2 \cdot KBB(k) = K(k, j) \cdot (1 + P1 \cdot PRDC(k, j) + P3 \cdot PRDC(k, j)^2).$$

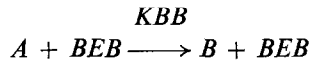
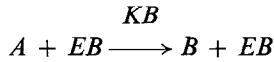
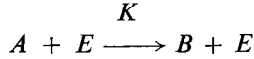
For each product  $k$ , one obtains three linear equations for the three unknowns  $KK(k)$ ,  $KB(k)$  and  $KBB(k)$  for the three values of the environment  $j = 1$ ,  $j = 2$ , and  $j = 3$ .  $j = 1$  indicates mineral glucose environment,  $j = 2$  indicates amino acids have been added (indicated by the program variable  $CAA = 1$ ), while  $j = 3$  indicates a broth environment (indicated by the program variable  $BROTH = 1$ ). There are three rate constants for different forms of an enzyme catalysing a reaction.

$KK$  = rate constant of pure enzyme  $E$ .

$KB$  = rate constant of enzyme  $EB$  which has one molecule of  $B$  attached to it.

$KBB$  = rate constant of enzyme  $EBB$  which has two molecules of  $B$  attached to it.

These rate constants lead to the following form for catalysis by different forms of the enzyme, and the differential equation for the chemical reaction catalysed by the enzyme.



$$\frac{d(B)}{dt} = KK \cdot E \cdot A + KB \cdot EB \cdot A + KBB \cdot BEB \cdot A = (KK \cdot E + KB \cdot EB + KBB \cdot BEB) \cdot A$$

(See Fig. 3.1.)

In terms of the rate constants calculated for each environment (Fig. 3.2)  $K(k, j)$  = rate constant for production of product  $k$  in environment  $j$  where

$$\frac{d(B)}{dt} = K(k, j) \cdot E_{total} \cdot A$$

```

43 DO 44 II = 1,14
44 R(II) = 0.
    IF (NUC - 5.E7) 46,45,45
45 R(1) = .9
46 IF (AA - 5.E7) 48,47,47
47 R(2) = .9
48 IF (ATP - 1.E8) 50,49,49
49 R(3) = .9
50 IF (WALL - 4.E8) 52,52,51
51 R(4) = .9
52 IF (ADP - 1.E5) 54,53,53
53 R(5) = .9
54 CONTINUE
X = 1./(.1 - .09★R(1) - .40★R(2) - .40★R(3) - .05★R(4)
1 - .01★R(5))
X = X★MRNA
DO 10 II = 1,10
K8K(II) = K(8)★RC(II)★(1 - R(II))★X/MRNA0
P1 = P1V(II)
P3 = P3V(II)
C(II) = ( KK(II) + P1★KB(II)★PRDC K(II)
1 + P3★KBB(II)★PRDC K(II) ★★2)/(1 + P1★PRDC K(II)
2 P3★PRDC K(II) ★★2 )
10 WRITE(6,114)II,C(II),II,K8K(II)
114 FORMAT(4H0 C(12,2H)= 1PE13.6,5H K8K(12,2H)= 1PE13.6)

```

Fig. 3.1. Repression and Allosteric Inhibition. Repression is obtained by an adjustment of  $K8K(II)$ . Allosteric inhibition is obtained through an adjustment of  $C(II)$ .



```

K(5) = ( (ADP0/T + ATP0/T)/NUC ) ★LN2
K(6) = (DNA0/NUC0)/T ★LN2
KDRNK = LN2/60.
K(8) = (LN2★MRNA0/T + KRNK★MRNA0)/(NUC0★DNA0)
K(10) = (TRNA0/NUC0)/T ★LN2
K(9) = RIB0 /((NUC0★AA0★T ) ) ★LN2
K(4) = WALL0/(GLUC0★T ) ★LN2
K(7) = PRTN0/(AA0★T ) ★LN2
KIN = (FACTR★CHRM0 - INZ)/(CHRM0★DBLE★IN1Z+IN2Z)
C★★★★★ 1ST DBLE MINUTES, CHRM0 GENES MAKE IN - INZ
C★★★★★ KIN IN IN PER GENE PER SEC
DDNA = K(6)★NUC
DMRNA = K(8)★NUC★DNA - KDRNK★MRNA
DTRNA = K(10)★NUC
DRIB = K(9)★NUC★AA
DWALL = K(4)★GLUC
DPRTN = K(7)★AA

C★★★★★
1 K(1)=(LN2★NUC0/T + (2.5E9/660.)★DDNA+(1.E6/660.)★DMRNA
2 + (2.5E4/660.)+DTRNA + K(5)+NUC
3 +(2.E6/660.)★DRIB)/GLUC0

C★★★★★
IF(BROTH.EQ.1) K(1) = 0.1★K(1)

C★★★★★
DNUC = K(1)★GLUC - (2.5E9/660.)★DDNA - K(5)★NUC
1 -(1.E6/660.)★DMRNA
2 -(2.5E4/660.)★DTRNA - (2.E6/660.)★DRIB
K(2)=(LN2★AA0/T + (1.E6/660.)★DRIB
1 + (4.E4/102.)★DPRTN)/GLUC0

C★★★★★
IF((CAA.EQ.1).OR.(BROTH.EQ.1)) K(2) = 0.1★K(2)

C★★★★★
IF(ABS(COUNT).GT.1.AND.COUNT.LT.9.9) GO TO 110
DAPO2 = (LN2★15.★1.E-8★6.02E23★38./6.)
1 / ( 3600.★22.4★1.E3★1.E3 )
K(3) = DAPO/GLUC0
DNAP = (6.E4/.0033)★(2.5/2.)
MRNAP = 7.5E4/12.5
TRNAP = (7.5E4/12.5)★(2.5E4/1.E6)
PRTNP = (2.12E6/1.4E3)★(4.E4/6.E4)
RIBP = (7.5E4/12.5)★(2.E6/1.E6) + (2.12E6/1400.)★(1.E6/6.E4)
WALLP = (2.E8/2.25E8)★(6.5E4/32.5)★(150./2.E6)
1 +(2.5E8/2.25E8)★(8.75E4/1.25E4)★(750./1.E3)
NUCP = (K(3)★GLUC - DNAP★DDNA - MRNAP★DMRNA
- TRNAP★DTRNA
1 - RIBP★DRIB - PRTNP★DPRTN - WALLP★DWALL - ATP0/T★LN2
2 - 2★K(5)★NUC)
3 /(K(2)★GLUC + K(1)★GLUC + K(4)★GLUC)
AAP = NUCP
WALLP = WALLP + NUCP

110 CONTINUE
IF(ABS(COUNT-4.).LE.0.1.OR.ABS(COUNT-7.).LE.0.1)
2 K(3) = (ATP/T★LN2+DNAP★DDNA + MRNAP★DMRNA
+ TRNAP★DTRNA +
3 RIBP★DRIB + PRTNP★DPRTN + WALLP★DWALL
+ AAP★K(2)★GLUC +
4 NUCP★K(1)★GLUC + 2★K(5)★NUC)/GLUC
DATP = K(3)★GLUC - DNAP★DDNA - MRNAP★DMRNA
1 - TRNAP★DTRNA - RIBP★DRIB - PRTNP★DPRTN
2 - WALLP★DWALL - AAP★K(2)★GLUC - NUCP★K(1)★GLUC
- 2★K(5)★NUC
K(5) = (ADP★LN2/T + DATP)/NUC
DADP = -DATP + K(5)★NUC
K(14) = VOL0/(T ★WALL0) ★LN2
DVOL = K(14)★WALL

```

Fig. 3.2. Preliminary Rate Constants K(1), . . . , K(14) for Flow Rates Between Pools. These rate constants are later used to calculate enzyme rate constants.

$B$  = amount of  $PRDC(k, j)$  = amount of product  $k$  in environment  $j$ .

$$E_{total} = E + EB + BEB$$

$d(B)/dt$  = change in amount of  $B$  as a derivative with respect to time. Given the three equations in three unknowns  $KK(k)$ ,  $KB(k)$  and  $KBB(k)$  (Fig. 3.3) one solves for the unknowns in the SOLVE routine of the program (Fig. 3.4). Trial values of

```

16   DO 13 ID = 1,10
      XK(ID,J) = C(ID)
      XEK(ID,J) = EK(ID)
      PRDC (ID,J) = PRDC K(ID)
13   XK8K(ID,J) = K8K(ID)

```

Fig. 3.3. Storing different variables (which variable is indicated by the ID subscript) for the cell in different environments (which environment is indicated by the J subscript).

```

      DO 12 L = 1,10
      IX = L
      P1 = 10./PRDC0(L)
      P3 = 100./PRDC0(L)★★2
      P1V(L) = P1
      P3V(L) = P3
C★★★
17  REPLACE INTERNAL FUNCTION SOLVE(IX)
      DO 14 II = 1,3,1
      A2(II) = P1★PRDC (IX,II)
      A3(II) = P3★PRDC (IX,II)★★2
14  SUM(II) = XK(IX, II) ★ (1. + A2(II) + A3(II))
      DUM1 = A2(3) - A2(1)
      DUM2 = A2(2) - A2(1)
      DUM3 = A3(2) - A3(1)
1  KBB(IX) = (SUM(3) - SUM(1)) - (DUM1/DUM2)★(SUM(2) - SUM(1))
      /(A3(3) - A3(1) - (DUM1/DUM2)★DUM3)
      KB(IX) = (SUM(2) - SUM(1) - DUM3★KBB(IX))/DUM2
      KK(IX) = SUM(1) - A3(1)★KBB(IX) - A2(1)★KB(IX)

```

Fig. 3.4. Calculations for allosteric inhibition necessary to fit data for real cells.

$P1$  and  $P3$  are used in the equations. All other quantities are available after data are collected for the simulated cell growing in each of its three environments.