A Novel Cellular-Level Numerical Model of the Lac Operon in *E. Coli*

Alex Huaqiang Leo, Sun Gang, Paul Decelles, and Fu Qian

Abstract—**Based on cellular biology and mathematic modeling, the present paper put forth a flow chart to depict the metabolic pathway of the lac operon structural gene expression of an Escherichia coli (E. coli) cell, and further introduces eight difference equations of the pathway, a mathematic model, for cellular-level computation. In detail, the model can compute the molecular activities and regular patterns of the mRNA, repressor, inducer, β –galactosidase, lactose, and Adenosine 5'-triphosphate (ATP) of an E. coli cell, which was not all available by traditional means and ways. The model itself and its state-variables construct a dynamics, and then numerically describe the gene expression and growth of the cell supplied with mere lactose. The model also established a basis to quantitatively predict activities of the cell. The agreement between the outcomes of the model and traditional biological experiment justified the rightness and achievement of the model.**

Index Terms—***E Coli*, Lac Operon, Gene Expression, Mathematic Cellular-Level Model.**

I. Introduction

The control of gene expression has been a major issue in biology. One basic way that gene expression can be controlled by regulating transcription. In prokaryotes, our basic understanding of how transcriptional control of gene expression began with Jacob and Monod's 1961 model of the Lac operon in which a regulator protein controls transcription by binding to a target region called the operator. In the absence of this binding, RNA transcription is initiated when an RNA polymerase is able to bind to a region, called the promoter, just upstream from the operator. The RNA polymerase travels down stream and transcribes a series of structural genes coding for enzymes in a particular metabolic pathway. The operon system acts as a molecular switch involving positive or negative control (Lewis 2004). Recently, attempts have been made to model operon systems in order to understand the dynamics of these feedback systems. In 2003, Yidirim modeled the Lac operon as a series of coupled differential equations in terms of chemical kinetics. Yidrim and Mackay note that their formulation assumes that they are dealing with a large number of cells and that the dynamics of small numbers of individual cells might be quite different. Meanwhile, it is hardly to obtain analytic solutions by using concentrations, average-values, as the variables of dynamics in differential equations, thus Laplace transform is necessary to solve the equations.

Whereas, in the present study, it was independently attempted to model the Lac operon, based on the molecules and their activities on the lac operon of an *E. coli*. By various ways and means in engineering, a flow chart (Figure 1) on the metabolic pathway of the lac operon activities of the *E. coli* was established. In terms of the amount of each relevant molecule present in the cell, a set of difference equations, mathematic model, comes on the basis of the flow chart. Since in form of difference equation, the numerical solution of molecular state-variable are available, by recurring the state-variables upon given a set of initial state or condition intuitionally. As a core governing the state-variables of the *E. coli* cell's molecules on the lac operon, the model could be developed into another set of formula for colony, with regard to cellular divisions in compliance with normal distribution in mathematics. The set of formula may well model a colony of the bacteria by halving the values of state-variable in various periods, simulating cellular division. Finally, the formula could be transformed into corresponding differential equations.

Manuscript received March 1, 2008. 8

F.A. Alex H. Leo, PhD, Baker University, 8001 College Blvd., Suite 100, Overland Park, KS 66210, USA, email: ahliu@comast.net)

S.B. Sun Gang, PhD, corresponding author, Mechanics & Engineering Science Dep., Fudan Univ., Shanghai 200433, China, Tel: 0086-21-65642740,Fax: 0086-21-65642742; email:gang_sun@fudan.edu.cn;

T.C. Paul Decelles，PhD, Math & Science Department, Johnson County Community College, 12345 College Boulevard, Overland Park, KS, 66210, USA

F.D. Fu Qiang, PhD, School of Science, University of Science and Technology of PLA, Nanjing, China, 211101, Tele: 0086-25-264-2852, e-mail: njfqhp@jlonline.com

II. List of Variables

$mR(n)$__ amount of mRNA polymers at step n;
τ_R__ half-life of mRNA polymer;
r_{TRSCRB}__ transcriptional rate of mRNA in the presence of allolactose.
$G(n)$__ amount of β-galactosidase molecules at step n;
τ_G__ half-life for β-galactosidase;
r_{TRSLTN}__ translational rate of β-galactosidase, permease and transacetylase molecules per mRNA polymer;
$L(n)$__ amount of lactose molecules transported by the Permease at step n in the cell;
$L_{ext}(n)$__ amount of lactose molecules outside of the cell at step n; $L_{ext}(0)$ is the amount of lactose molecules supplied at the beginning;
$Max\ r_{TRSPT}$__ the maximum rate of lactose transported in by each permease molecule;
$A(n)$__ amount of ATP molecules, produced from the lactose at the step n;
τ_A__ half-life of ATP;
r_{meta}__ rate of a β-galactosidase molecule breaking down the lactose into the glucose, and thus into the ATP.

Δ__ set step length , *e.g.* Δ=1(minute), with regard to the half-life and the creating-rate for the substance presented by the variable.

III. Determining *mR(n+1)*

An *E. coli* cell contains about 3,000 copies of holoenzyme, a form of RNA polymerase (Russell 113). The DNA-length of lacZYA is (Lewis, 281),

3510+780+825 =5115 bp

It was found that mRNA half-lives were globally similar in nutrient-rich media, determined from a least-squares linear fit. Approximately, 80% of half-lives ranged between 3 and 8 min in M9 + glucose medium, and 99% of the half-lives measured were between 1 and 18 min. In the LB, 99% of half-lives were between 1 and 15 min. The mean half-life was 5.7 min in M9 + glucose and 5.2 min in the LB (Bernstein 2002). The transcription of a polymerase approaches at a rate of average 30~50 nucleotides per second (Russell 115). But in some defined media, the general time was approximately tripled (Bernstein). The lac mRNA is extremely unstable, and decays with a half-life of only ~3 minute (Russell 115). Another article indicates the average half-life of an mRNA in *E. coli* is about 1.8 minutes. As lactose is no longer present, the repressor will be activated, and thus bind to the operator, making the transcription action stopped;

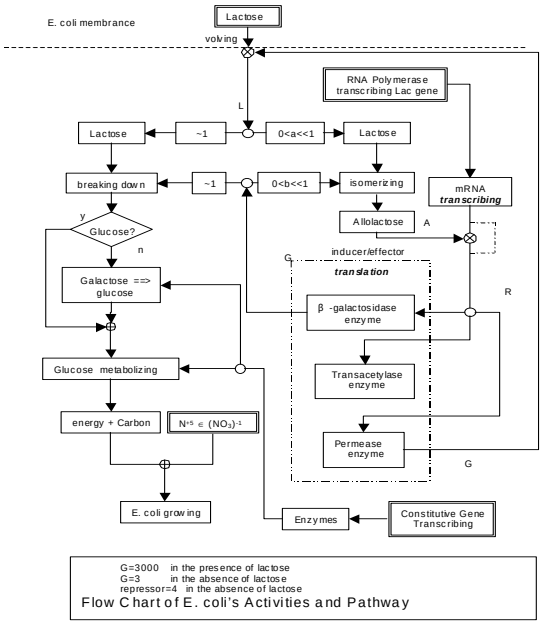

Figure 1. The flow chart of the activity pathways of an *E. coli*

Considering the above conditions, mR(n+1), r_{TRSCRB}, and τ_R are able to be determined in the presence of lactose. The variable, mR(n+1), equates, elapsing the period of Δ, the amount remained of mR(n) through the halving attenuation, plus the amount produced by transcription triggered in the presence of the allolactose from the lactose.

Therefore, a polymerase needs 1.705~2.842 minutes for transcribing an mRNA, or 0.5865~0.3519 copies of mRNA come per min per polymerase. So to speak, 1759~1055 copies of mRNA are maximally created per minute, with regard to 3000 copies of the polymerases within an *E. coli*.

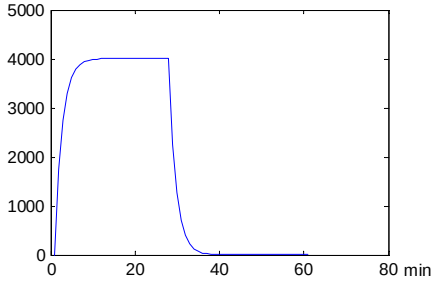

Figure 2. A simulated curve of *mR(n)*, which indicates the amount of the mRNA molecules of the cell.

$$r_{TRSCRB}=1759\sim586 \text{ mRNA / min}$$
$$\tau_R=5 \text{ min}$$

That is, the resulting formula is as follows:

$$mR(n+1)= e^{\frac{-0.69\Delta}{\tau_R}} \times mR(n) + r_{TRSCRB} \times \Delta \quad (1)$$

IV. Determining *G(n+1)*

Given supplying *E. coli* cells lactose, β-galactosidase enzyme will appears in minutes, so do lac permease in membrane and transacetylase, a third protein. The level of β-galactosidase enzyme can accumulate to 10% of cytoplasmic protein (Terry 2002). Once a ribosome approaches away of the initiative site on an mRNA, another one will locate at the initiative site. Thus, many ribosomes may well simultaneously be translating each mRNA. An average mRNA has a cluster of 8~10 ribosomes, named as polysome, for synthesizing protein (Russell 153).

Where cells of *E. coli* are grown in the absence of lactose, there is no need of β-galactosidase. An *E. coli* cell contains 3~5 molecules of the enzyme (Russell 2004). That is, a dozen of the repressors bind and unbind rather than just bind and stay on the mRNA. In a fraction of a second, after one repressor unbinds and before another binds, an mRNA polymerase could initiate a transcription of the operon, even in the absence of lactose. (Russell 447). In 2-3 minutes after adding lactose, there soon are 3,000~5,000 molecules of β-galactosidase per *E. coli* cell (Russell, Lewis). The β-galactosidase in the *E. coli* is more stable than the mRNA, whose half-life is more than 20 hours. The half-life of the permease is more than 16 hours (Lewis 284). So that, the β-galactosidase activity remains at the induced level for longer. To add or change codons at the 5' end of the gene to encode N-terminal amino acid within the DNA-lacZYA will provide greater resistance to β-galactosidase production (Southern Illinois Univ.). As a porter, the permease compound can be to concentrate lactose against the gradient across the cellular membrane in ad hoc manner for E. coli. The concentration of the substrate against a gradient can be achieved up to 10^3 to 10^6-fold. The existing β-galactosidase molecules will be diluted out by cell division (Terry 2002). Whereas, replicating the DNA of an *E. coli* needs 84 minutes

(Russell 80), the half-life is 100 minutes for simplicity, or 30 minutes for conservation. This point of view could be used to derive a model in colony of the bacteria.

The DNA-length of lac_ZYA (Lewis 281):
3510+780+825=5115 bp;
The mRNA-length of lac_ZYA:
5115÷3-1=1705-1=1704 amino acids.

Protein synthetic rate is 350-400 amino acids (1/4~1/5 β-galactosidase) per minute per ribosome (Terry 2002). There are approximate 8 ribosomes attached to an mRNA That is, for a ribosome to make a β-galactosidase needs 4~5 minutes (say 4.5 minutes) by running all the way of the mRNA, and there are 1.8 copies of the β-galactosidase translated per min per mRNA.

As a function of mR(n), G(n+1) is the amount of β-galactosidase, same as permease in count. G(n+1),then r_{TRSLTN} and τ_G are able to be determined. G(n+1) equates, elapsing the period of Δ, the β-galactosidase amount remained of G(n) through the halving attenuation, plus the amount produced by translation of mR(n).

τ_G =1000 minutes
r_{TRSLTN} =8÷4.5=1.8
G(0)=3

That is, the resulting formula is as follows:

$$G(n+1) = e^{\frac{-0.69\Delta}{\tau_G}} \times G(n) + r_{TRSLTN} \times \Delta \times mR(n) \quad (2)$$

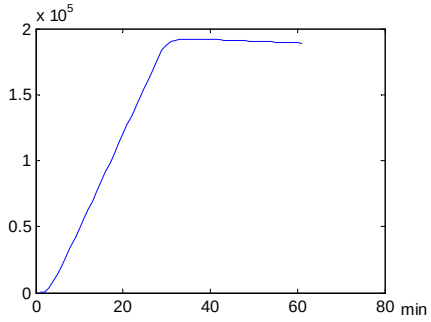

Figure 3. The simulated curve of G(n), the permease or β-galactosidase within the cell.

V. Determining L_{EXT}(N+1) and L(N+1)

Lac permease is a highly lipophilic protein with an apparent subunit molecular weight of about 30,000 (Ullmann 2001). The lactose transport-in rate of permease is 120 nmol per mg per minute. That is, 2.4 copies of lactose are transported into the cell by a permease per minute.

r_{meta}, the rate of lactose breaking down into glucose by G(n); 2 min is for to convert a lactose into 72 ATPs (Russell). Therefore, r_{meta} is 0.5, and max r_{TRSPT} is 2.4.

r_L is the rate of transporting lactose into the cell by the permease, presented by G(n). L(n+1) is the amount of the lactose at step n+1 in the *E. coli.* For the period of Δ, L(n+1) equates to the amount of the lactose brought-in by the permease, presented by G(n), but subtracting the amount metabolized into glucose. Each glucose molecule, in turn, is immediately transformed into 72 ATPs.

That is, the resulting formula is as follows:

$$L_{ext}(n+1) = L_{ext}(n) - \max r_{TRSPT} \times \Delta \times G(n) \times f(L_{ext}(n) - L(n)) \quad (3)$$

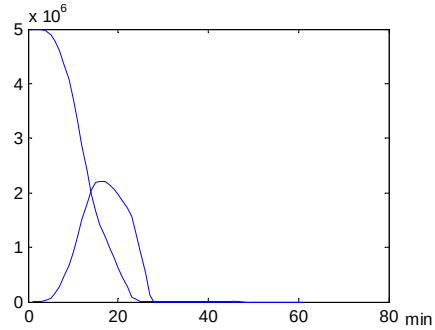

Figure 4. The simulated curves of the L_{ext}(n) and L(n), the external and internal lactose molecules of the cell.

$$L(n+1) = L(n) + \max r_{TRSPT} \times \Delta \times G(n) \times f(L_{ext}(n) - L(n)) - r_{meta} \times \Delta \times L(n) \quad (4)$$

$$f(x) = \begin{cases} 2 - e^{\frac{-x}{\lambda}} & when \quad x \geq 0 \\ e^{\frac{x}{\lambda}} & else \end{cases} \quad (5)$$

in which, λ is a constant, whose magnitude close to that of the lactose supplied.

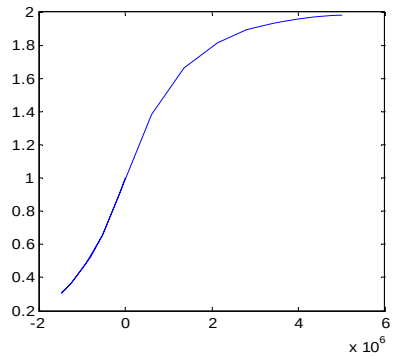

Figure 5. The simulated function curve of the f(x)
.

VI. Determining A(N)

ATP half-life, τ_A, is about 15 minutes in water at room temperature, estimating 2 min for to convert a copy of lactose into 72 ATPs. The energy resource for *E. coli* growth is from the lactose in form of ATP at 72 in scale. The number of ATP for the transportation-in of a lactose by a permease in energy equates about one. Therefore, the transcription of an mRNA needs 5115 ATPs (in forms of ATP, UTP, CTP, or GTP), and the translation of an mRNA needs 1700 ATPs.

Given the consumption of energy, comprising transcribing, translating, as well as transporting lactose into the cell, count 80% of total energy, as is the primary life activity, then a coefficient d for energy distribution set at 0.8. The remaining energy is for the other metabolic activities.

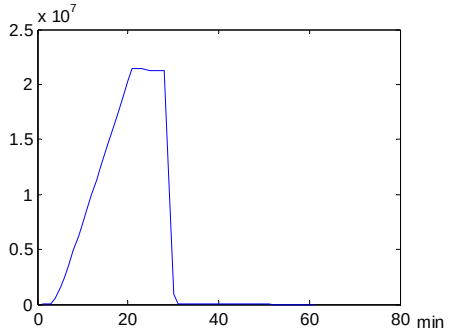

Figure 6. The simulated curve of the ATP molecules consumed within the cell.

With regard to the r_{meta}, G(n), Δ, steps n+1 and n, the energy generated by the glucose is distributed among producing mRNA and β-Galactosidase, and bringing the lactose into *E. coli* cell.

An energy-distributional equation is as follows:

$$\begin{aligned} A(n+1) = {} & e^{\frac{-0.69\Delta}{\tau_A}} A(n) \\ & + 72\Delta \times r_{meta} \times G(n) \\ & - \Delta[5115 r_{TRSCRB} \\ & + 1700 r_{TRSLTN} \times mR(n) \\ & + \max r_{TRSPT} \\ & \times f(L_{ext}(n) - L(n)) \times G(n)] \end{aligned} \quad (6)$$

In the above equation, first term is accumulated energy before step n; second term is the energy created from the lactose, and third term is the consumed for the three activities. If the third is greater or equal to the second, there will not be energy accumulation. An comprehensive energy distributive coefficient equation, c(n), can be determined as following,

$$c(n) = \begin{cases} e(n) & when \quad e(n) \le 1 \\ 1 & else \end{cases} \quad (7)$$

$$\begin{aligned} e(n) = {} & 72\Delta \times r_{meta} \times G(n) \times d \\ & \div [5115 r_{TRSCRB} + 1700 r_{TRSLTN} \times mR(n) \\ & + r_{TRSPT} \times f(L_{ext}(n) - L(n)) \times G(n)] \end{aligned} \quad (8)$$

The domain of *c(n)* is within [0,1].

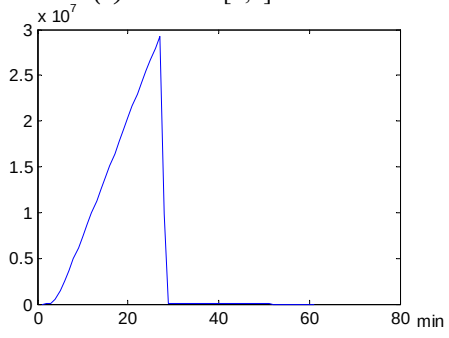

Figure 7. The simulated function curve of the ATP molecules created within the cell.

If the energy is inadequate, the energy -distributional coefficient equation comes, and thus the equations (1) ~ (4) are modified into energy -constrained equations (1)' ~ (4)'.

$$\begin{aligned} & mR(n+1) = \\ & e^{\frac{-0.69\Delta}{\tau_R}} mR(n) + r \times_{TRSCRB} \Delta \times c(n) \end{aligned} \quad (1)'$$

$$\begin{aligned} & G(n+1) = \\ & e^{\frac{-0.69\Delta}{\tau_G}} G(n) + r_{TRSLTN} \\ & \times \Delta \times m \times R(n) \times c(n) \end{aligned} \quad (2)'$$

$$\begin{aligned} L_{ext}(n+1) = {} & L_{ext}(n) \\ & - \max r_{TRSPT} \times \Delta \times G(n) \\ & \times f(L_{ext}(n) - L(n)) \times c(n) \end{aligned} \quad (3)'$$

$$\begin{aligned} L(n+1) = {} & L(n) \\ & + \max r_{TRSPT} \times \Delta \times G(n) \\ & \times f(L_{ext}(n) - L(n)) \\ & \times c(n) - r_{meta} \times \Delta \times L(n) \end{aligned} \quad (4)'$$

VII. The System Structure

Equations (1) ~ (8), or Equations (1)' ~ (4)' in place if necessary, are integrated into a systematic function Equation (9), with respect to the initial condition presented as Equation (10).

$$\mathbf{X(n+1)} = \mathbf{T(X(n))} + \mathbf{U} \quad (9)$$

$$\mathbf{X(n)} = \begin{bmatrix} R(n) \\ G(n) \\ L_{ext}(n) \\ L(n) \\ A(n) \end{bmatrix} \quad (10)$$

They are solved numerically by MATLAB through setting the following initial condition:

$$\mathbf{X(0)} = [0 \quad 3 \quad supply \quad 0 \quad 0]'$$

$$\mathbf{U} = [r_{TRSCRB}\Delta \quad 0 \quad 0 \quad 0 \quad 0]'$$

VIII. In Comparison to Traditional Biological Experiment

A traditional experiment culturing E. coli in tubes offered the following graph (Terry 2002),

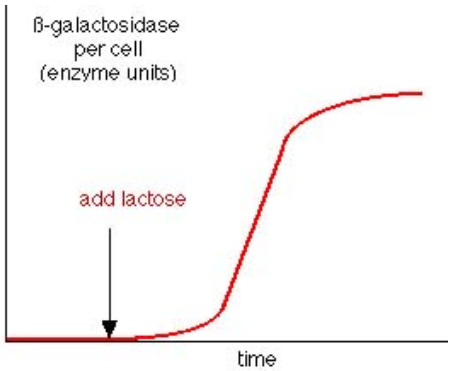

Figure 8. A classic experimental data of *E. coli*.

In comparison of the outcomes of the presented equations and the traditional experimental shown as Figure 2, the results are consistent in a mutual time scale; and so does mRNA in Figure 1. In addition, more information on the cell, such as external and internal lactose, ATP consumption, etc in a cell of E. coli are available by the model, but not in the other methods.

IX. Conclusion

By computational simulation, the Figures 2 to 7 were obtained. Supplied with mere lactose as nutrition, an *E. coli* cell of the Operon activities can be described by the eight difference equations, or a mathematical model. The Relevant outcomes of the model are highly consistent with traditional experimental results. The mathematic model can be developed into a general method to govern the growth of an *E. coli* colony by gene expression. The model is justified through the traditional biological experiment.

X. Discussion

This paper discusses the growth of an *E. coli* cell supplied with mere lactose. A model for a colony of the bacteria could be developed by using the presented model as a core, and then plus cellular divisions of the colony periodically. In addition, effects of parameters λ and d in the equations (5) and (8) should be studied further. The rate or behavior pattern of the permeases transporting lactose across the membrane of an E. coli cell is rather approximate. Finally, it is worth to mention that the simulated numerical solutions of the model vary slightly with ⊿ value. The phenomenon probably is of the theoretical approximate hypothesis in the Section 4.

Since on individual cellular level, a colony model could be developed rather to describe and simulate more complex activities and regular patterns of the representative substances of the colony, by considering individual cell division, and could further be developed into a general quantitative analysis for cellular biology in high precise.

Acknowledgment

This paper was based on the authors' background of computer science, Electrical Engineering, and Genetics. In the process of implementing, great supports in various aspects come from Mr. Jian Liu, Ph D, Honeywell Inc. (Kansas, U S), Ms. Ting Wu, MD., Mr. Guo Chi Zhang, MD., and Ms. Yan Hong respectively. Herein, I thank the persons for their generous supports again.

References

[1] Bernstein, Jonathan A. et al, .2002 July 23. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. USA*. 99 (15): 9697–9702. http://www.pnas.org/cgi/content/full/99/15/9697

[2] Huan, Lan et al, 2002. Changing the lactose permease of Escherichia coli into a galactose-specific symporter, *PNAS*. Vol. 99, no. 10, 6613-6618. *http://www.pnas.org/cgi/reprint/99/10/6613.pdf*.

[3] Russell, Peter J., 2002. *iGenetics, Chapter 5,16*. Benjamin Cumming, San Francisco, CA.

[4] Russell, Peter J., 2002. *iGenetics, Chapter 5,16*. Benjamin Cumming, San Francisco, CA. Southern Illinois University, 2003. BIOTECHNOLOG.Y. *MICR 421*, chapter 6. http://www.science.siu.edu/microbiology/micr421/chapter6.html.

[5] Terry, Thomas M., 2002. Regulation of Metabolism. *Microbiology*, Chapter 12 (p. 237-250), 4th Ed. *http://www.biologie.uni-hamburg.de/b-online/library/micro229/terry/229sp00/lectures/regulation.html*.

[6] Ullmann, Agnes, 2001. Escherichia coli lactose Operon, *Encyclopedia of Lifesciences*. http://216.239.41.104/search?q=cache:mzCMeayarnMJ:www.cellcycle.bme.hu/oktatas/mikrofiz/extra/lac%2520operon.pdf+how+weight+lac+permease+is+in+e.+coli&hl=en&ie=UTF-8.

[7] Weinglass, Adam B. et al, 2001. Manipulating Conformational Equilibria in the Lactose Permease of Escherichia coli, *Encyclopedia of Life sciences*. http://darwin.bio.uci.edu/~bardwell/weinglass02jmb.pdf.

[8] Yidirim, Necmettin et al, Feedback Regulation in the Lactose Operon: A Mathematical Modeling Study and Comparison with Experimental Data, *Biophysical Journal*, Volume 84, May 2003, 2841-2851.