5

Modeling and Simulation of Production of Metallothionein and Red Fluorescent Fusion Protein by Recombinant *Escherichia Coli* Using Graphical Programming

Ling Gao[1], Yilin Ren[2], Yao Ma[3], Jianqun Lin[3] and Jianqiang Lin[3]
[1]*School of Information Science and Engineering, Shandong Normal University, Jinan 250014*
[2]*School of Life Science, Shandong Normal University, Jinan 250014*
[3]*State Key Lab of Microbial Technology, Shandong University, Jinan 250100*
China

1. Introduction

Metallothioneins (MTs) are a family of low molecular weight, cysteine rich, metal binding proteins, and exist in a variety of organisms ranging from bacteria to mammals (Moffatt & Denizeau, 1997). MTs play significant roles in clearance of free radical against oxidative damage (Masters et al., 1994), trace element metabolism and heavy metal detoxification (Feldman & Cousins, 1976), cell metabolism (Cherian & Apostolova, 2000), disease processes (Sens et al., 2001; Velázquez et al., 1999), rare metal accumulation (Nordberg, 1998), heavy metal removal and environment protection *et al.* (Nordberg, 1998; Elinder et al., 1981). The important applications stimulate the interest of large scale production of MT. Compared with the conventional method by extraction from animal organs, MT production by using recombinant *E. coli* has obvious advantages in reducing cost, increasing efficiency and product quality. However, recombinant MT is easily hydrolyzed and unstable for the thiol group of cysteine when produced by recombinant microorganism (Saito & Hunziker, 1996). By integrating the MT gene with fusion tags, MT was successfully expressed in *E. coli* (Hong et al., 2001; Yang et al., 2007). On the other hand, rapid measurement of MT production is important for optimization of fermentation process. MT was reported to have light absorption at the wave length of 250 nm after binding with Cd (Onosaka & Cherian, 1982), which could be used for MT measurement. But, according to our experience, this method has low accuracy and poor reproducibility. In order to detect MT production in an easy and rapid way during fermentation process, in this study, MT gene is fused with red fluorescent protein (RFP) gene and coexpressed in recombinant *E. coli*. Modeling and simulation is a powerful tool for the analysis of the mechanism and dynamic responses of the fermentation process, and for process optimization and automatic control. The production of recombinant protein using recombinant *E. coli* involves the regulations of aimed gene transcription and translation, which process can be well described by using the white box structured model involing operon regulation and is done in this research. But, the structured model has the disadvantage of extreme complex not suitable for the purpose of process optimization.

Therefore, a grey box unstructured model is also constructured, which model is simple enough for process variable prediction in real applications but still sufficient for modeling the major regulation effects on the aimed genes of the recombinant strain. LabView is a data driven and graphical programming software tool, widely used in process simulation, data acquisition and automatic control. The MT production mathematical model developed using LabView is easily incorporated into the model based on-line optimization and automatic control system of MT fermentation process developed using LabView in the future. In this research, the structured and unstructured mathematical models and computer simulation of MT production and RFP fluorescence are made. The final developed mathematical model includes substrate consumption, the growth of the recombinant *E. coli* cells, the repression and induction of the aimed genes, the production of the aimed protein, and the maturation of the expressed RFP. The mathematical model consists of two levels: the intracellular level for the expression and induction of the aimed genes and the reactor level for the dynamic changes of substrate and cell concentrations as well as the fluorescence. This research shows the advantages and the usefulness of the data driven and graphical programming software of LabView in modeling and simulation of biological phenomena.

2. Experimental production of MT fusion protein

The human MT (hMT-1A) gene sequence was obtained from Genebank. Some rare codons for *E. coli* were modified, and the sequence was synthesized chemically (Su et al., 2009). The plasmid pGHM-RFP was used for MT production (Fig. 1). DsRed2 gene was cloned and ligated with hMT-1A gene to construct the fusion protein expression vector. In order to minimize the conformational change and final clearence of RFP, a Gly-Gly-Gly tri-peptide and thrombin restriction site were designed and inserted between DsRed2 and hMT-1A genes in pGHM-RFP construction. *E. coli* BL21(DE3) (Novagen, Germany) was used as the host cell to construct *E. coli*(pGHM-RFP). The plasmid pGHM-RFP has a *tac* promotor (Ptac), which is a combination of -10 reagion of *lac* promotor and -35 reagion of *trp* promotor (DE Boer et al., 1983).

Cultivations of *E. coli*(pGHM-RFP) were made at 37^{o}C, shaken at 180 r/min. The volume of 250 μl of seed culture was inoculated into 40 ml of flask containing 300 ml of LB medium (10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl, pH 7.0) and 100 μg/ml of ampicillin. Isopropyl-β-D-thiogalactoside (IPTG) was added to the final concentration of 0.2 mM when OD_{600} reached 0.9~1.1. Samples were made every hour and the cells were harvested by centrifugation at 7000×g for 10 min at 4^{o}C. Proteins were analyzed using 12% SDS_page, and stained by 0.25% Coomassie brilliant blue R250. The amount of MT produced was analyzed by ImageQuant TL (GE Healthcare, USA) using BSA (bovine serum albumin) as the standard. Fluorescence intensity was measured using RF-5301PC (Shimadzu, Japan). The results were shown in Fig. 2. The results showed that the fusion protein of GST-MT-RFP expressed immediately after IPTG induction, but the fluorescence was detected about 6 hours after the induction (Fig. 2). The fluorescence was detected about one hour after MT production reached the maximum, and then increased. It indicated that RFP fluorescence could be used as the suboptimal indicator for termination of MT fermentation process. In addition, the fluorescence intensity of RFP was quite strong, which ensured the accuracy of the detection. The possible reason for the delay of the appearence of RFP fluorescence should be that the maturation of red chromophore undergoes a period of oxidation, which costs some time (Verkhusha et al., 2004). The time delay of the RFP fluorescence made it

unsuitable for online monitoring and optimization of MT fermentation process, but could be used as the suboptimal indicator for termination of the MT fermentation process.

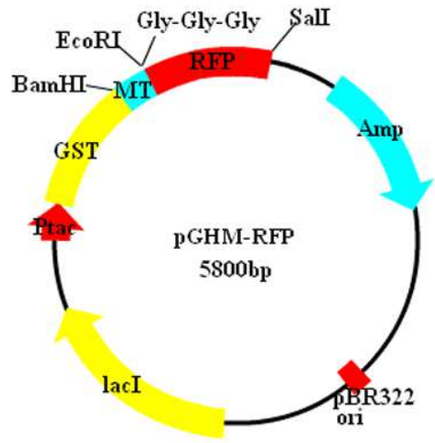


Fig. 1. The structure of plasmid pGHM-RFP.

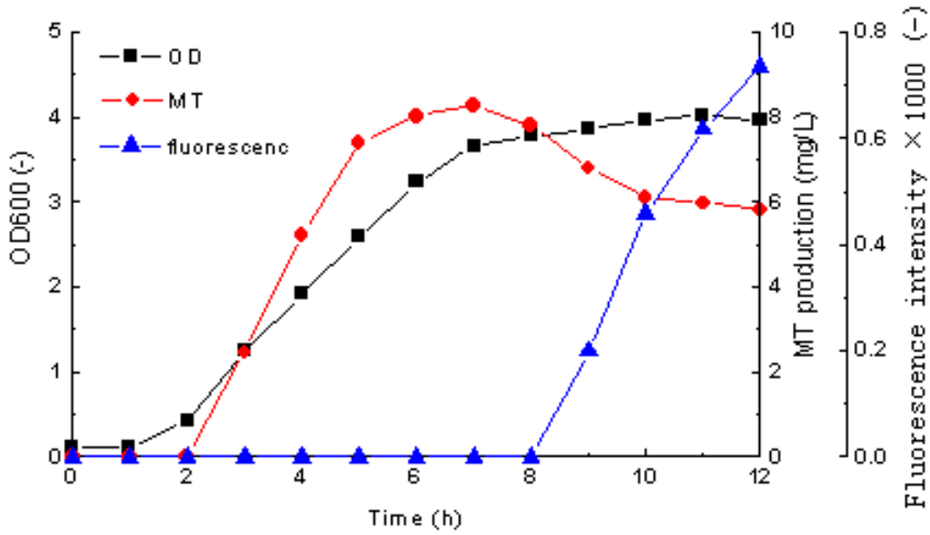


Fig. 2. The time courses of cell growth, MT fusion protein production and RFP fluorescence.

3. Structured model for plasmid based MT production

3.1 Modeling of operon regulation

The detailed structure and regulation mechanism of *tac* promotor-operon was shown in Fig. 3. From Fig. 3 for the control mechanism, it shows that the operon is coordinately controlled by two regulatory sites positioned upstream of the aimed genes, which are the regulatory sites of cAMP-CAP complex and the lacI repressor protein (Fig. 3). For the lacI repressor mechanism, when the active lacI repressor protein is binding to the operator gene, mRNA polymerase can not pass through the operator gene to transcribe the aimed gene. When the inducer molecules (lactose or its structure analogue IPTG) binds to lacI to inactivate the lacI repressor and make it get off from the operator, the mRNA polymerase can pass through the

operator gene to transcribe the aimed gene. On the other hand, the operon is also controlled by the mechanism of carbon catabolite repression. The binding of mRNA polymerase to the promotor gene (at the -10 and -35 reagions) is loose, needs to be strengthened by the cAMP-CAP complex. The catabolite activator (CAP) protein is activated when it binds to cAMP to form the complex. When glucose concentration is high in the medium, the intracellular cAMP concentration is low, as a result, the concentration of cAMP-CAP complex decreases, and the binding of mRNA polymerase to the promotor becomes loose and the aimed gene is untranscribed (Fig. 3).

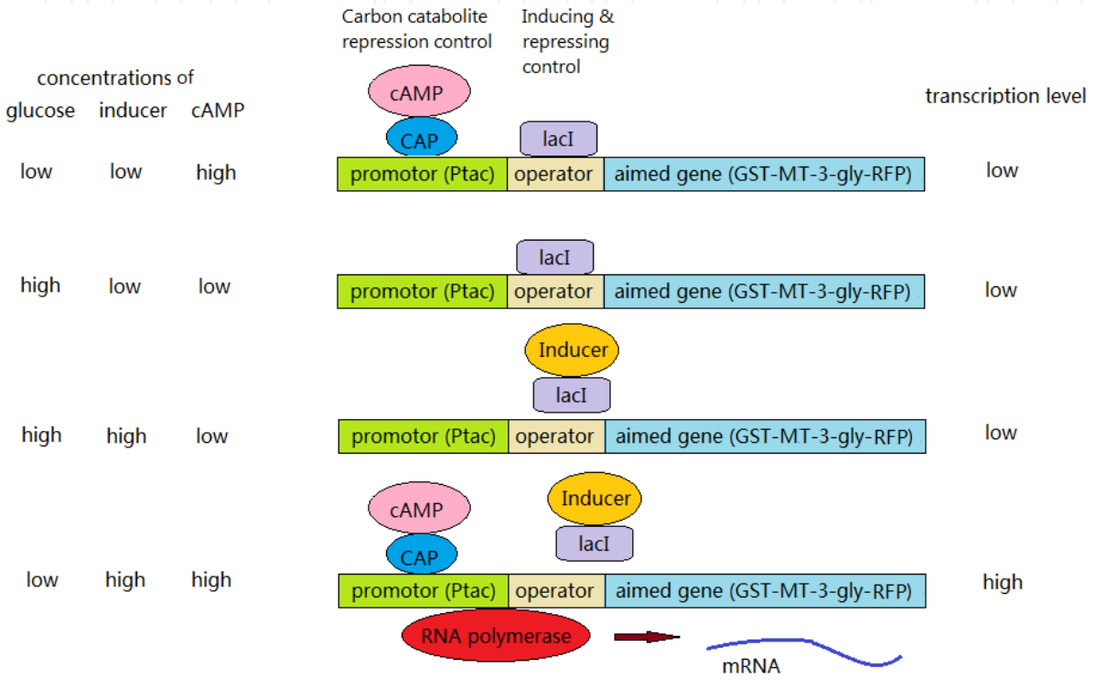


Fig. 3. The structure and regulation mechanism of the operon of pGHM-RFP (adapted from Abeles et al., 1992).

The induction and repression control mechanism of the operon ensures that the genes under control will not be transcribed without induction by the inducers. The binding of the repressor protein (lacI) to the inducer (lactose) or to the operator gene is described by Eqs. (1-2). The binding of the inducer to the repressor-operator complex is described by Eq. (3).

$$X_r + nS_{lac} \underset{k_{-1}}{\overset{k_1}{\leftrightarrow}} X_r nS_{lac} \tag{1}$$

$$X_o + X_r \underset{k_{-2}}{\overset{k_2}{\leftrightarrow}} X_o X_r \tag{2}$$

$$X_o X_r + nS_{lac} \underset{k_{-3}}{\overset{k_3}{\leftrightarrow}} X_o X_r nS_{lac} \overset{k_4}{\rightarrow} X_o + X_r nS_{lac} \tag{3}$$

Where, X_r , the repressor protein; S_{lac} , lactose; X_o , the operator gene; n , the number of the lactose binding sites of the repressor protein. Eq. (3) shows that the inducer-repressor complex of $X_r nS_{lac}$ has low binding affinity to the operator X_o , quickly falls off. Therefore, $\left[X_o X_r nS_{lac}\right]$ can be neglected in the following balances. Balances for the repressor, operator, and inducer are

$$\left[X_r\right]_t = \left[X_r\right] + \left[X_r nS_{lac}\right] + \left[X_o X_r\right] \tag{4}$$

$$\left[X_o\right]_t = \left[X_o\right] + \left[X_o X_r\right] \tag{5}$$

$$\left[S_{lac}\right]_t = \left[S_{lac}\right] + n\left[X_r nS_{lac}\right] \tag{6}$$

Where, the index *t* refers to the total concentration. As there are more free lactose molecules than that combined in the complex, the last term in Eq. (6) can be neglected:

$$\left[X_{lac}\right]_t = \left[S_{lac}\right] \tag{7}$$

From above equations, the following equations can be obtained:

$$\frac{dX_r}{dt} = k_{-1} \cdot X_r nS_{lac} - k_1 \cdot X_r \cdot nS_{lac} - k_2 \cdot X_o \cdot X_r + k_{-2} \cdot X_o X_r \tag{8}$$

$$\frac{dX_r nS_{lac}}{dt} = k_1 \cdot X_r \cdot nS_{lac} - k_{-1} \cdot X_r nS_{lac} + k_4 \cdot X_o X_r nS_{lac} \tag{9}$$

$$\frac{dX_o}{dt} = k_{-2} \cdot X_o X_r - k_2 \cdot X_o \cdot X_r + k_4 \cdot X_o X_r nS_{lac} \tag{10}$$

$$\frac{dX_o X_r}{dt} = k_2 \cdot X_o \cdot X_r - k_{-2} \cdot X_o X_r - k_3 \cdot X_o X_r \cdot nS_{lac} + k_{-3} \cdot X_o X_r nS_{lac} \tag{11}$$

$$\frac{dX_o X_r nS_{lac}}{dt} = k_3 \cdot X_o X_r \cdot nS_{lac} - k_{-3} \cdot X_o X_r nS_{lac} - k_4 \cdot X_o X_r nS_{lac} \tag{12}$$

As there are more repressor molecules than there are operators, and in this case the last term in Eqs. (4) and (8) can be neglected. The fraction of repressor-free operators,*Q1*, is obtained:

$$Q_1 = \frac{\left[X_o\right]}{\left[X_o\right]_t} = \frac{X_o}{X_o + X_r} \tag{13}$$

The carbon catabolite repression control mechanism of the operon ensures that the genes under control will not be transcribed when a preferred substrate is available, e.g., glucose. The mechanism of operator activation by CAP-cAMP complex can be described by reactions of (14) and (15):

$$X_{CAP} + mX_{cAMP} \underset{k_{-5}}{\overset{k_5}{\leftrightarrow}} X_{CAP} mX_{cAMP} \tag{14}$$

$$X_p + X_{CAP}mX_{cAMP} \underset{k_{-6}}{\overset{k_6}{\leftrightarrow}} X_p X_{CAP}mX_{cAMP} \tag{15}$$

The balances for CAP and promotor are

$$\left[X_{CAP}\right]_t = \left[X_{CAP}\right] + \left[X_{CAP}mX_{cAMP}\right] + \left[X_p X_{CAP}mX_{cAMP}\right] \tag{16}$$

$$\left[X_p\right]_t = \left[X_p\right] + \left[X_p X_{CAP}mX_{cAMP}\right] \tag{17}$$

From Eq. (14) to (17), Eq. (18) to (21) can be obtained.

$$\frac{dX_{CAP}}{dt} = k_{-5} \cdot X_{CAP}mX_{cAMP} - k_5 \cdot X_{CAP} \cdot mX_{cAMP} \tag{18}$$

$$\frac{dX_{CAP}mX_{cAMP}}{dt} = k_5 \cdot X_{CAP} \cdot mX_{cAMP} - k_{-5} \cdot X_{CAP}mX_{cAMP}$$

$$-k_6 \cdot X_P \cdot X_{CAP}mX_{cAMP} + k_{-6} \cdot X_P X_{CAP}mX_{cAMP} \tag{19}$$

$$\frac{dX_p}{dt} = k_{-6} \cdot X_p X_{CAP}mX_{cAMP} - k_6 \cdot X_p \cdot X_{CAP}mX_{cAMP} \tag{20}$$

$$\frac{dX_p X_{CAP}mX_{cAMP}}{dt} = k_6 \cdot X_p \cdot X_{CAP}mX_{cAMP} - k_{-6} \cdot X_p X_{CAP}mX_{cAMP} \tag{21}$$

From the above equations, the fraction of promoters being activated can be obtained:

$$Q_2 = \frac{\left[X_p X_{CAP}mX_{cAMP}\right]}{\left[X_p\right]_t} = \frac{X_p X_{CAP}mX_{cAMP}}{X_p + X_p X_{CAP}mX_{cAMP}} \tag{22}$$

The intracellular level of cAMP in Eq. (22) can be described by the empirical equation correlating X_{cAMP} and the extracellular concentration of the preferred energy source (S), eg. glucose:

$$X_{cAMP} = \frac{K}{K + S} \tag{23}$$

3.2 Modeling of plasmid copy number and mRNA polymerase limiting effect

The plasmid copy number (N_P) in the host cell is variant, which is reversely corelated with the specific growth rate (μ). The more plasmids it contains the slower it grows. The cell grows fastest ($\mu = \mu_{max}$) when it contains no plasmid. On the other hand, it is regarded that when N_P increases towards its maximum value of $N_{P,\max}$, the plasmid replication rate decreases. Then, the plasmid replication rate can be described by Eq. (24), which is decreased with the increases of the plasmid copy number and the specific growth rate.

$$r_P = k_{P.\max} \cdot N_P \cdot \left(1 - \frac{N_P}{N_{P\cdot\max}}\right) \cdot \left(1 - \frac{\mu}{\mu_{\max}}\right) - \mu \cdot N_P \tag{24}$$

Where, $k_{P.\max}$, the maximum plasmid replication rate; μ, the specific growth rate; $\mu_{\max}$, the maximum specific growth rate.

In the recombinant srtain that contains multicopy plasmid, the promotors in the plasmids and in the chromosome will compete for the mRNA polymerase. Therefore, the transcription rate is not increased linearly with the increase of gene copy number but in a decreased manner. The mRNA polymerase competing effect can be described by Eq. (25) (Lee & Bailey, 1984).

$$Q_3 = \left(1 - \frac{N_P}{N_{P.\max}}\right)^{\alpha} \tag{25}$$

3.3 Modeling of mRNA transcription and aimed protein synthesis

The intracellular aimed mRNA synthetic rate and degradation rate can be described by Eq. (26):

$$r_{mRNA} = k_{+mRNA} \cdot Q_1 \cdot Q_2 \cdot Q_3 \cdot N_P - \left(k_{-mRNA} + \mu\right) \cdot X_{mRNA} \tag{26}$$

Where, X_{mRNA}, the intracellular aimed mRNA; k_{+mRNA}, the constant for aimed mRNA synthesis; k_{-mRNA}, the constant for aimed mRNA degradation. The last term of Eq. (26) describes the degradation and cell growth dilution effects on intracellular mRNA. Then, the intracellular plasmid based MT fusion protein synthesis rate can be described by Eq. (27):

$$r_{MT} = k_{+MT} \cdot X_{mRNAk} - \left(k_{-MT} + \mu\right) \cdot X_{MT} \tag{27}$$

Where, X_{MT}, the intracellular MT fusion protein concentration; k_{+MT}, the intracellular MT fusion protein synthesis rate; k_{-MT}, the intracellular MT fusion protein degradation rate.

The structured model developed above is based on the mechanism of biological reactions and signal transductions, and is useful in analysis of the molecular and physiological dynamics and mechanisms. But, the structured model involved too many model parameters of intracellular reactions and signal transductions that are difficult to measure, which makes the model difficult for practical uses.

4. Unstructured model and simulation of MT production

As discussed in last section, the structured model is too sophisticated for bioprocess modelling and simulation in practical applications. In contrast, unstructured models without detailed intracellular mechanisms are convenient and widely used in this field. But, one major drawback of the unstructured model is that the unstructured black box model sometimes can not reflect the real dynamics of the bioprocess. For example, the Monod model and the Luedeking-Piret model are most often used unstructured models for the cell growth and product production, which are not sufficient to model the gene induction and repression effects. In this section, an unstructured model will be constructed for recombinant MT fusion protein production, and improvements are made in this model so

that it can describes the induction and repression effects on the aimed genes exerted by the bioreactor level process variables of the concentrations of related substrates in the culture medium.

In last section, the inducer and repressor control mechanism of the operon is summarized into Q_1, and the carbon catabolite repression control mechanism is summarized into Q_2 by using several intracellular reactions and signal transductions. Hear, unstructured model will be constructed for Q_1 and Q_2. In the experiments described in section 2, IPTG is used in replace of lactose as the inducer, which is unable to be degraded and has strong inducing effect than lactose. Therefore, the induction effect exerted by IPTG is in the extremes, which can be regarded as "1" or "yes" with IPTG induction, and "0" or "no" without IPTG induction. In addition, the signal transduction of the induction and repression effects have a time scale of several seconds from on-state to off-state or vise versa. Therefore, Q_1 can be set to "1" or "0" for the cases of IPTG addition or not described by Eq. (28), which is much simpler than the development of Eq. (13).

$$Q_1 = \begin{cases} 0 & \text{no IPTG} \\ 1 & \text{IPTG} \end{cases} \tag{28}$$

The carbon catabolite repression control mechanism is related with the concentration of cAMP, which is an activator of CAP protein and is reversely correlated with the energy charge of the cell. The relationships between substrate concentration and intracellular cAMP concentration can be described by Eq. (23). Similar with Q_1, Q_2 has the activated or non-activated states under high or low cAMP conditions. But, unlike IPTG which has an almost constant concentration during the cultivation process, the intracellular cAMP concentration is always changing with the changes of the substrate concentration resulted from the cell consumption or the feeding of the substrate. Therefore, different from Q_1 that has the discrete states of "0" or "1", Q_2 is described using unstructured Hill equation for activator as a function of intracellular cAMP concentration:

$$Q_2 = \frac{\beta \cdot {X_{cAMP}}^{\gamma}}{{k_{cAMP}}^{\gamma} + {X_{cAMP}}^{\gamma}} \tag{29}$$

Where, k_{cAMP}, β and γ are the constants. Combining Eq. (23) and (29), Q_2 becomes the function of substrate concentration (S) of the cultivation medium. The sharpness of Q_2 function can be adjusted by the parameters. With the consumption of substrate as shown in Fig. 4A, Q_2 increases from about 0.3 to 1. The intracellular plasmid replication rate is described by Eq. (24). Combing with Eq. (24), the total plasmid copy number can be described by Eq. (30).

$$\frac{d\left[N_P\right]_T}{dt} = r_P \cdot \left[X\right]_T \tag{30}$$

Where, X, the cell concentration; the footnote "T" indicates the total concentration in the fermentation broth. As discussed earlier, the transcription rate is not increased linearly with the increase of plasmid copy number, as the promoters of the plasmid and the chromosome compete for the mRNA polymerase. The effect of plasmid copy number on the transcription

rate is described by Q_3 of Eq. (25). Then, combining with Eq. (26), the total mRNA can be described by Eq. (31) using the unstructured model.

$$\frac{d[X_{mRNA}]_T}{dt} = r_{mRNA} \cdot [X]_T \tag{31}$$

Similarly, total MT production can be modelled by Eq. (32):

$$\frac{d[X_{MT}]_T}{dt} = r_{MT} \cdot [X]_T \tag{32}$$

The equations for the specific growth rate, cell growth, and substrate consumption are:

$$\mu = \frac{\mu_{\max} \cdot [S]_T}{k_m + [S]_T} \cdot \left(1 - \frac{[X]_T}{[X_{\max}]_T}\right) \tag{33}$$

$$\frac{d[X]_T}{dt} = \mu \cdot [X]_T \tag{34}$$

$$\frac{d[S]_T}{dt} = -\frac{1}{Y_{X/S}} \cdot \frac{d[X]_T}{dt} - \frac{1}{Y_{MT/S}} \cdot \frac{d[X_{MT}]_T}{dt} \tag{35}$$

Where, $X_{\max}$, the maximum cell concentration; k_m, the substrate affinity constant; S, the substrate concentration; $Y_{X/S}$, the cell yield from the substrate; $Y_{MT/S}$, the product yield from the substrate. The RFP fluorescence intensity curve is almost proportional to the MT production curve except that a time delay for the maturation of RFP, which can be modelled by Eq. (36):

$$\frac{d[F]_T}{dt} = k_F \cdot \frac{d[X_{MT}]_T}{d(t - t_L)} \tag{36}$$

Where, k_F, the constant; t_L, the time delay for the fluorescence. By using the unstructured model constructed in this section and the experimental data of section 2, some of the model parameters are calculated and the other unknown model parameters are optimized using genetic algorithm (GA) by minimizing the aimed function calculated from errors between model predictions and the measured data. The details of model parameter optimization using GA is introduced in section 5. By using the unstructured model constructed in this section and the model parameters, the production process of MT fusion protein is simulated, and the results are shewn in Fig. 4.

From Fig. 4, it shows that the model fits the experimental data very well, and is suitable for modelling the recombinant protein production. With the increase of cell concentration, the substrate concentration decreases (Fig. 4A). During the simulation, the initial and the maximum plasmid copy numbers are set to 10 and 30, respectively. The initial value for aimed mRNA is set to 0 with the aimed gene under non-induced state. As shown in Fig. 4B, the plasmid copy number decreases with the increase of cell growth before 5th h of

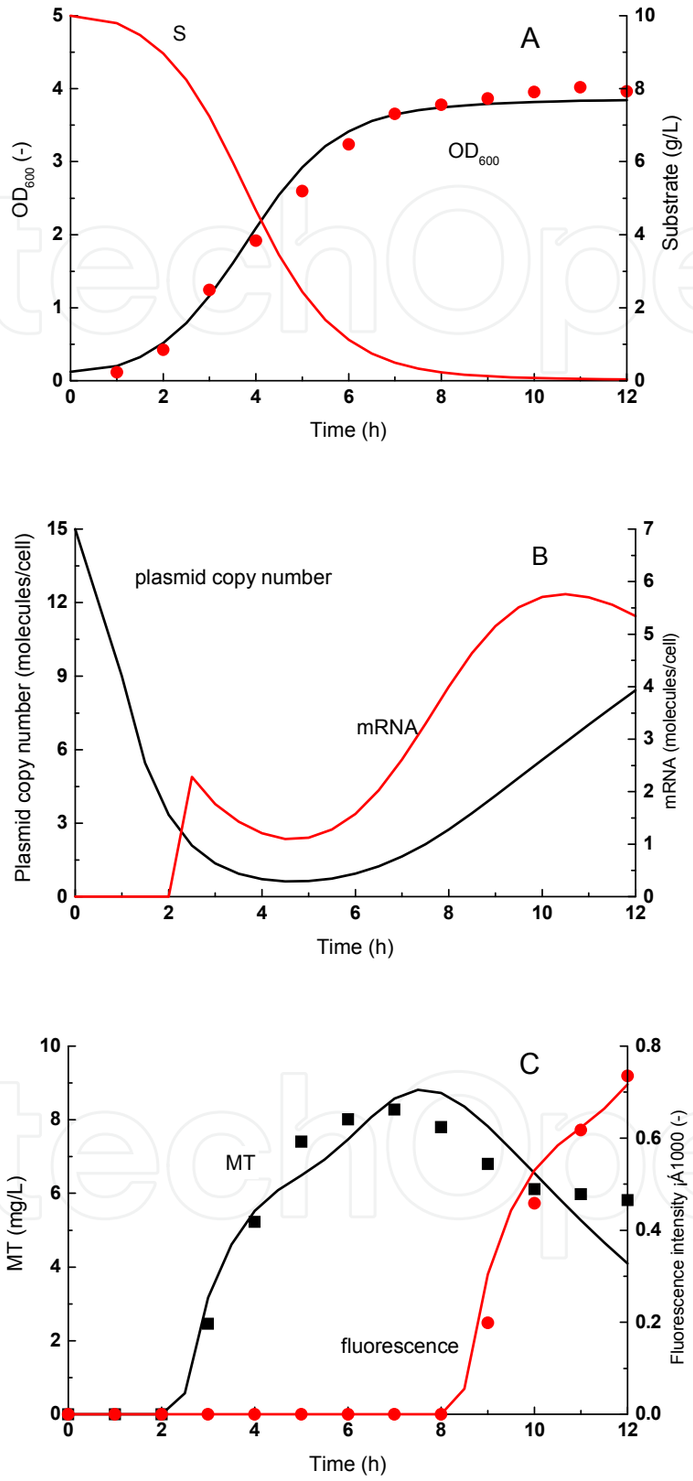


Fig. 4. Simulation of the production process of MT fusion protein. Lines, model predictions; dots, measured data.

cultivation (Fig. 4.A), which is modelled by Eq. (24). The number of intracellular aimed mRNA molecules increases immediately after the addition of the inducer of IPTG at 3rd h of the cultivation (Fig. 4B), modelled by Eq. (26). From Fig. 4, it shows that the aimed mRNA time profile is not proportional to the time profile of plasmid copy number, resulted from the effects of aimed gene regulation, mRNA polymerase competition, mRNA degradation *et al.*. The MT fusion protein production and the RFP fluorescence time courses are shown in Fig. 4C. MT production increases after addition of IPTG at 3rd h of cultivation and reaches the maximum value at 7th h, then decreases because of degradation. The RFP fluorescence is 6 h delay than MT production (Fig. 4C). The reason for the time delay of RFP fluorescence should be the time consumed for maturation of red chromophore which undergoes a time period of oxidation (Verkhusha, 2004).

5. Model parameters and the optimization using genetic algorithm (GA)

Some of the model parameters have clear physiological meanings and can be obtained by theoretical calculation or by experimental measurements. $X_{\max}$ and t_L can be obtained directly from Fig. 2; $\mu_{\max}$ can be calculated from the growth curve of Fig. 2; $Y_{X/S}$ can be calculated from the measured data; $Y_{MT/S}$ can be estimated from the biochemical pathways of protein synthesis; $N_{P.\max}$ is obtained from general knowledge about the plasmid copy number of *E. coli*; $k_{P.\max}$ is chosen arbitaryly as 2.5 times of $\mu_{\max}$. The other model parameter values are optimized using GA. The differential equations are solved using 4th order Runge-Kutta method.

Genetic algorithm is a search heuristic that mimics the process of natural evolution, which is proved useful in solving sophisticated optimization problems as well as in optimization of biological model parameters (Lin et al., 2004). A standard GA is used in the optimization, which has the operations of coding, initialization, mutation, hybridization, decoding, fitness calculation, selection, and reproduction procedures exerted on individuals (or chromosomes) in a population. The potential solution (individual) is coded as a binary vector, called a chromosome, the elements of which were called genes and situated in predefined positions, indicated as alleles (Fig. 5). One gene codes for one model parameter. The gene number n of one chromosome equals to the number of model parameters, so that each chromosome codes for all parameters of the model and enable to calculate the model prediction error with the set of model parameters coded by the chromosome. In GA, one population contains m individuals (or chromosomes) so that m model prediction errors can be calculated for the m chromosomes, and the fitness of each chromosome can then be calculated and used for the selection procedure. In this study, one gene is coded by a binary string of 10 bits. Then, n parameters are represented by a binary string of $n\times10$ bits (Fig. 5). The population of m chromosomes can be represented by a $m\times10n$ array. The initialization procedure is performed to randomly assign initial values of "0" or "1" to the $m\times10n$ array. The hybridization procedure is applied to two randomly selected chromosomes called parents, and two new chromosomes created by the exchange of one or more parts randomly selected from the parent chromosomes, which occurred based on a probability called the hybridization rate. The mutation procedure is applied by turning over one or more randomly selected bits in the $m\times10n$ array from "0" to "1" or from "1" to "0", which occurred based on a probability called the mutation rate. Decoding was performed to

transform the binary number of each gene in the chromosome to a decimal system using the following equation:

$$D_{mn} = D_{\min} + \frac{B_{mn}}{2^Z - 1} \cdot \left(D_{\max} - D_{\min} \right) \tag{37}$$

Where, m, the ordinal rank of the chromosome; n, the ordinal rank of the gene in the chromosome; Z, the digits of the binary vector for one gene; B, the binary value coded by the corresponding gene; D, the decimal value transformed from B; and $D_{\max}$ and $D_{\min}$, the decimal values of the maximum and minimum limits of the corresponding parameter. Selection and reproduction are performed to select individuals (chromosomes) to create a mating pool for reproducing offspring. The selection procedure is stochastic with the fitted chromosomes, judged by fitness (or objective function J), indicating a better chance of being selected. A low value of J is defined as a high fitness, with J defined as follows:

$$J = \frac{1}{R \cdot K \cdot L} \cdot \sum_{r}^{R} \sum_{k}^{K} \sum_{l}^{L} \left(\frac{Y_{rkl}^{*} - Y_{rkl}}{Y_{rkl}} \right)^2 \tag{38}$$

Where, R, the number of experiments; K, the number of measured variables; L, the number of measured points of one variable; Y_{rkl}^{*}, the model output; and Y_{rkl}, the measured value in the experiments. In order to give each variable the same importance in the optimization, the relative error $\left(Y_{rkl}^{*} - Y_{rkl} \right) / Y_{rkl}$ is used so that J had no unit. Another advantage using the relative error instead of the absolute error $\left(Y_{rkl}^{*} - Y_{rkl} \right)$ in calculating J is to enable the data points of one variable at different sampling time have the same importance in the optimization. A flow diagram of the GA used for the model parameter optimization is

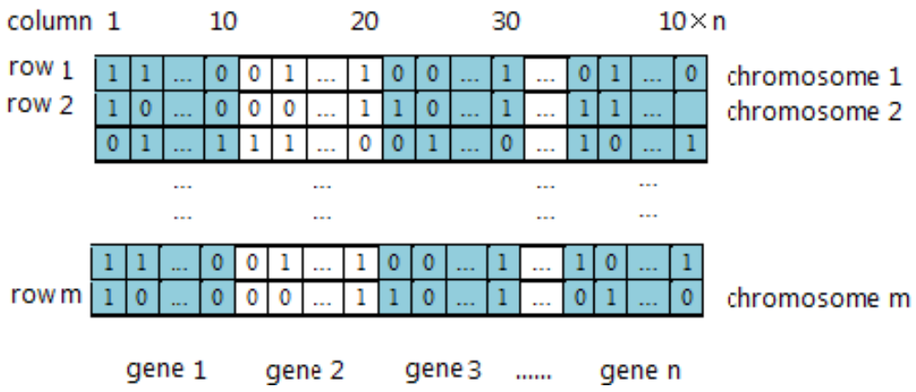


Fig. 5. The structure of the chromosome and the $m \times 10n$ array for m chromosomes of n genes.

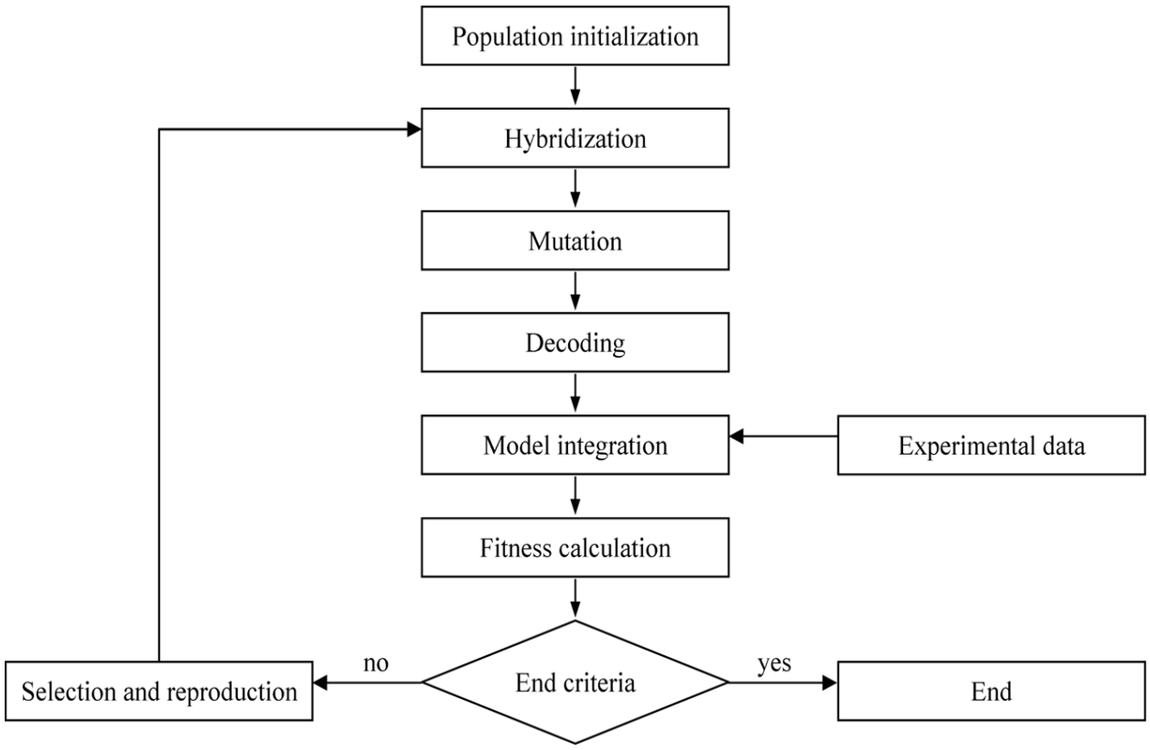
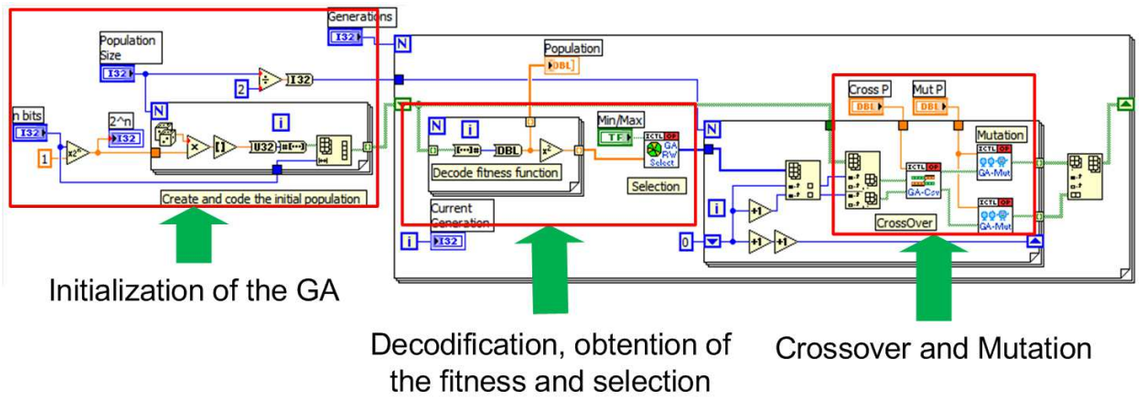


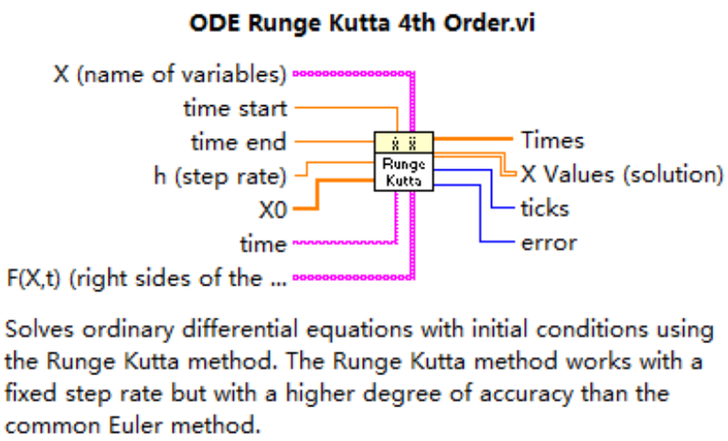
Fig. 6. The flow diagram of GA calculation in model parameter optimization.

shown in Fig. 6. In this research, the 4th order Runge-Kutta method is used to solve the differential equations and GA is used to minimize J in Eq. (38). A population size of 50, hybridization rate of 0.2, and mutation rate of 0.05 are used in GA optimization. GA programming using LabView for model parameter optimization was reported (Moore, 1995). In LabView, there are sub VIs of GA (located in: Control toolkit for LabView, ICTL) and also the 4th order Runge-Kutta method (located in: Functions/Mathematics/ Differential equations) (Fig. 7). In addition, the softwares written in MatLab, VisualBasic et al. can also be linked with LabView and used for the same purpose.

In optimization of the unknown model parameter values using GA, the known value model parameters are also refined in the mean time to eliminate the possible errores made in the measurements. Each of the measured value is set as the middle value of a small range and the value is searched within the small range by GA in the refining process to minimize the aimed function.



(A)



(B)

Fig. 7. (A) The VI of GA and (B) the VI image of 4th order Runge-Kutta method from the manuals of LabView.

6. Conclusion

The production of MT fusion protein is modelled using structured and unstructured models. The unstructured model developed in this research is suitable for bioprocess prediction and optimization. The data driven and graphical programming software of LabView has the advantages of high efficiency in programming, having many handy sub VIs for various applications, and especially easy in constructing data acquitation and on-line monitoring and control system. The mathematical models developed in LabView is easily incorporated into the future developed on-line model based predictive control system.

7. References

Abeles, R.H.; Frey, P.A. & Jencks, W.P. (*1992*). Biochemistry. *Jones* & *Bartlett Publishers*, *Boston, Mass,* p.383

Cherian, M. G. & Apostolova, M. D. (2000). Nuclear localization of metallothionein during cell proliferation and differentiation. *Cell Mol. Biol.*, 46, 347-356.

DE Boer, H.A.; Comstock, L. J. & Vasser, M. (1983). *Proc Natl Acad Sci USA*, 80:21-25.

Elinder, C. G.; Jönsson, L.; Piscator, M. & Rahnster B. (1981). Histopathological changes in relation to cadmium concentration in horse kidneys. *Environ. Res.*, 26, 1-21.

Feldman, S. L.& Cousins, R. J. (1976). Degradation of hepatic zinc-thionein after parenteral zinc administration. *Biochem. J.*, 160, 583-588.

Hong, S. H.; Toyama, M.; Maret, W. & Murooka, Y. (2001). High yield expression and single step purification of human thionein/metallothionein. *Protein Expr. Purif.*, 21, 243-250.

Lee, S. B. & Bailey, J. E. (1984). Genetically structured models for lac promotor-operotor function in the *Escherichia coli* chromosome and in multicopy plasmids: lac promotor function, *Biotechnol. Bioeng*., 26,1381-1389.

Lin, J.; Lee, S. M. & Koo Y. M. (2004). Model development for lactic acid fermentation and parameter optimization using genetic algorithm, *J. Microbiol. Biotechnol*., 14(6), 1163–1169.

Masters, B. A.; Kelly, E. J.; Quaife, C. J. ; Brinster, R. L. & Palmiter, R. D. (1994). Targeted disruption of metallothionein I and II genes increases sensitivity to cadmium. *PANS*, 91, 584-588.

Moore, J. H. (1995). Artificial intelligence programming with LabVIEW: genetic algorithms for instrumentation control and optimization, *Computer Methods and Programs in Biomedicine*., 47, 73-79.

Moffatt, M. and D. Denizeau (1997) Metallothionein in physiological and physiopathological processes. *Drug Metabolism Reviews* 29: 261-307.

Nordberg, M. (1998). Metallothioneins: historical review and state of knowledge. *Talanta*, 46, 243-254.

Onosaka, S. & Cherian, M. G. (1982). Comparison of metallothionein determination by polarographic and cadmium-saturation methods. *Toxicol. Appl. Pharmacol*., 63, 270-274.

Saito, S. & Hunziker P. E. (1996). Differential sensitivity of metallothionein-1 and-2 in liver of zinc-injected rat toward proteolysis. *Biochim. Biophys. Acta*, 1289, 65-70.

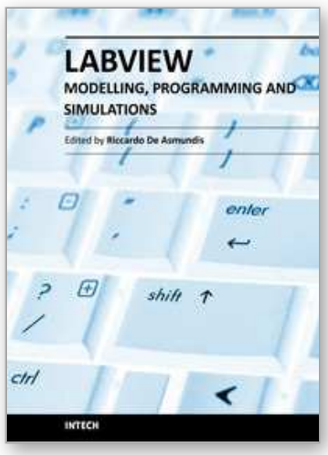
Sens, M. A.; Somji, S.; Garrett, S. H.; Beall, C. L. & Sens, D. A. (2001). Metallothionein isoform 3 overexpression is associated with breast cancers having a poor prognosis. *Am. J. Pathology*., 159, 21-26.

Su, Y.; Lin, J.; Lin, J. & Hao, D. (2009). Bioaccumulation of arsenic in recombinant *Escherichia coli* expressing human metallothionein. *Biotechnol. Bioprocess Eng.*, 14, 565-570.

Velázquez, R. A.; Cai, Y.; Shi, Q. & Larson, A. A. (1999). The distribution of zinc selenite and expression of metallothionein-III mRNA in the spinal cord and dorsal root ganglia of the rat suggest a role for zinc insensory transmission. *J. Neurosci*., 19, 2288-2300.

Verkhusha, V. V.; Chudakov, D. M.; Gurskaya, N. G.; Lukyanov, S. & Lukyanov, K. A. (2004). Common pathway for the red chromophore formation in fluorescent proteins and chromoproteins. *Chem. Biol*., 11, 845-854.

Yang, F.; Zhou, M.; He, Z.; Liu, X.; Sun, L.; Sun, Y. & Chen Z. (2007). High-yield expression in *Escherichia coli* of soluble human MT2A with native functions. *Protein Expr. Purif*., 53, 186-194.



Modeling, Programming and Simulations Using LabVIEW™ Software

Edited by Dr Riccardo De Asmundis

ISBN 978-953-307-521-1
Hard cover, 306 pages
Publisher InTech
Published online 21, January, 2011
Published in print edition January, 2011

Born originally as a software for instrumentation control, LabVIEW became quickly a very powerful programming language, having some characteristics which made it unique: simplicity in creating very effective User Interfaces and the G programming mode. While the former allows for the design of very professional control panels and whole applications, complete with features for distributing and installing them, the latter represents an innovative way of programming: the graphical representation of the code. The surprising aspect is that such a way of conceiving algorithms is extremely similar to the SADT method (Structured Analysis and Design Technique) introduced by Douglas T. Ross and SofTech, Inc. (USA) in 1969 from an original idea by MIT, and extensively used by the US Air Force for their projects. LabVIEW enables programming by implementing directly the equivalent of an SADT "actigram". Apart from this academic aspect, LabVIEW can be used in a variety of forms, creating projects that can spread over an enormous field of applications: from control and monitoring software to data treatment and archiving; from modeling to instrument control; from real time programming to advanced analysis tools with very powerful mathematical algorithms ready to use; from full integration with native hardware (by National Instruments) to an easy implementation of drivers for third party hardware. In this book a collection of applications covering a wide range of possibilities is presented. We go from simple or distributed control software to modeling done in LabVIEW; from very specific applications to usage in the educational environment.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Ling Gao, Yilin Ren, Yao Ma, Jianqun Lin and Jianqiang Lin (2011). Modeling and Simulation of Production of Metallothionein and Red Fluorescent Fusion Protein by Recombinant Escherichia Coli Using Graphical Programming, Modeling, Programming and Simulations Using LabVIEW™ Software, Dr Riccardo De Asmundis (Ed.), ISBN: 978-953-307-521-1, InTech, Available from: http://www.intechopen.com/books/modeling-programming-and-simulations-using-labview-software/modeling-and-simulation-of-production-of-metallothionein-and-red-fluorescent-fusion-protein-by-recom

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China

51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen

IntechOpen