0273-2289/1559-0291 (Online)/06/135/0059/$30.00

Sequential and Simultaneous Statistical Optimization by Dynamic Design of Experiment for Peptide Overexpression in Recombinant *Escherichia coli*

Kwang-Min Lee,[†,1] Chang-Hoon Rhee,[†,1] Choong-Kyung Kang,[2] and Jung-Hoe Kim*[,1]

[1]Cellular Metabolic Engineering Lab., Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong Yuseong-gu Daejeon, 305-701, South Korea, E-mail: Klee0922@yahoo.co.kr; [2]KoBioTech Co., Ltd., 713-12 Gojan-dong Nam-gu Inchen, 405-821, South Korea

Received July 25, 2005; Revised January 8, 2006; Accepted January 26, 2006

Abstract

The production of recombinant anti-HIV peptide, T-20, in *Escherichia coli* was optimized by statistical experimental designs (successive designs with multifactors) such as 2^{4-1} fractional factorial, 2^3 full factorial, and 2^2 rotational central composite design in order. The effects of media compositions (glucose, NPK sources, $MgSO_4$, and trace elements), induction level, induction timing (optical density at induction process), and induction duration (culture time after induction) on T-20 production were studied by using a statistical response surface method. A series of iterative experimental designs was employed to determine optimal fermentation conditions (media and process factors). Optimal ranges characterized by %T-20 (proportion of peptide to the total cell protein) were observed, narrowed down, and further investigated to determine the optimal combination of culture conditions, which was as follows: 9, 6, 10, and 1 mL of glucose, NPK sources, $MgSO_4$, and trace elements, respectively, in a total of 100 mL of medium inducted at an OD of 0.55–0.75 with 0.7 m*M* isopropyl-β-D-thiogalactopyranoside in an induction duration of 4 h. Under these conditions, up to 14% of T-20 was obtained. This statistical optimization allowed the production of T-20 to be increased more than twofold (from 6 to 14%) within a shorter induction duration (from 6 to 4 h) at the shake-flask scale.

[†]Coauthors.
*Author to whom all correspondence and reprint requests should be addressed.

Index Entries: Anti-HIV peptide; statistical optimization; dynamic experimental design.

Introduction

T-20 is a promising new experimental antiretroviral and the most prominent of a new class of anti-HIV drug, the fusion inhibitors that inhibit HIV entry into host cells *(1)*. It is a synthetic 36 amino acid peptide corresponding to residues 127–162 in the ectodomain of the transmembrane segment of the HIV envelope glycoprotein (gp41) *(2,3)*. This peptide targets the fusion phase of HIV infection, in which the HIV particle fuses or binds with a new target cell such as a CD4 T-cell *(4,5)*. T-20 is considered to interfere with the conformational changes in gp41 surface molecule to prevent fusion of viral and host cell membranes *(6,7)*.

T-20 is the most complex synthetic peptide ever chemically manufactured at a large scale, requiring an unprecedented complexity of manufacturing process that involves 106 production steps, as opposed to 8–10 steps, for typical synthetic process (www.fuzeon.com). Therefore, manufacturing companies have focused attention on biologic production for therapeutic large chemicals such as peptides and proteins by gene-cloning technique. Biologic production could be a good alternative solution to the chemical one for mass production of pharmaceutical products to satisfy the balance of large supply and demand.

Culture media and bioprocess optimization are vital to the long-term manufacturing success of a product. Productivity can substantially affect the cost and profit, i.e., the ultimate commercial success of the product. To reach optimal production conditions, bioprocess research requires effective problem-solving methods, because the conditions involve adjustment of multiple parameters and complications that inhibit application of engineering principles *(8,9)*. Additional obstacles for bioprocess research include the lack of an accurate mathematical model to describe the whole process, high noise levels, interactions among variables, and complex biochemical reactions. These conditions call for a good strategy to deal with such a complicated system.

Statistically designed experiments use a small set of carefully planned experiments. This method is more satisfactory and effective than other methods such as classic one-factor-at-a-time or mechanistic approach, because it can deal with many variables simultaneously with a low number of observations, saving time and cost. The statistical experimental design provides a universal language of statistics with which people from different areas such as academia, engineering, business, and industry can communicate with each other *(10,11)*.

The purpose of the present study was to find optimal culture conditions for foreign peptide overexpression in *Escherichia coli* at the shake-flask scale by employing a series of statistical experimental designs involving several variables in order to maximize the production and productivity of T-20.

Materials and Methods

Microorganism

A genetically modified recombinant *E. coli* strain, BL21(DE3)/pET23a-G3T20, obtained from Inje University, Chooncheon, Korea was used throughout this research. The strain was maintained by culturing in Luria-Bertani (LB) medium containing 5 g/L of yeast extract, 10 g/L of NaCl, and 10 g/L of tryptone with 100 mg/L of kanamycin.

Culture Media

Seed cultures were prepared by conducting growth in a 250-mL shake flask containing 50 mL of LB medium supplemented with 100 µg/mL of kanamycin at 37°C for 12 h in a rotary shaker at 250 rpm. Fermentation assay media for T-20 production were prepared by mixing concentrated stock solutions into the 250-mL shake flask and adding distilled water to make a total volume of 50 mL. Assay media varied with treatments from the experimental design matrix (Tables 1–3). Major stock components of the assay media used for the treatments (formula) were as follows: glucose (20%), 200 g/L of glucose; NPK sources (1 *M*), 136 g/L of KH_2PO_4, 132 g/L of $(NH)_2HPO_4$, 19.2 g/L of citric acid; $MgSO_4$ (0.1 *M*), 26.4 g/L of $MgSO_4 \cdot 7H_2O$; trace element (100X, per liter of 5 *M* HCl), 10 g/L of $FeSO_4 \cdot 7H_2O$, 2 g/L of $CaCl_2 \cdot H_2O$, 2.2 g/L of $ZnSO_4 \cdot 7H_2O$, 0.5 g/L of $MnSO_4 \cdot 4H_2O$, 1 g/L of $CuSO_4 \cdot 5H_2O$, 0.1 g/L of $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, 0.02 g/L of $Na_2B_4O_7 \cdot 10H_2O$; isopropyl-β-D-thiogalactopyranoside (IPTG) (0.1 *M*), 23.83 g of IPTG. All media were sterilized at 121°C for 15 min and cooled to room temperature prior to use.

Expression of Recombinant Peptide T-20

Batch fermentation for T-20 expression was carried out in 250-mL shake flasks containing 50 mL of assay media supplemented with 100 mg/mL of kanamycin. Fermentation was initiated by inoculating 1 mL of seed culture cultured in LB for 12 h into the flasks variously formulated by the design. The assay cultures were grown at 37°C in a shaker at 250 rpm. For the induction of T-20 gene expression, IPTG was simultaneously added into the flasks at an optical density (OD) of 0.6 at 600 nm for the center point (OD variations were according to the formulations) to a final concentration of 0.3–0.9 m*M* (depending on treatments). Samples were collected at regular 2-h intervals and analyzed immediately after they were taken from the cultures.

Experimental Design

The factor treatments and experimental settings were performed by dynamic experimental design (design of experiment [DOE] combined with repeated measures). Design Expert (version 6.09, Stat-Easy, Minneapolis, MN) was employed for building and analyzing experimental designs.

Table 1
Design Matrix and Responses for 2^3 Factorial Design

ID	Factor 1 (A:glucose) (mL)	Factor 2 (B:N,P,K) (mL)	Factor 3 (C:IPTG) (m*M*)	Factor 4 (D:duration) (h)	Response 1 (OD [abs.])	Response 2 (T-20 [%])
1	6	6	0.3	2	0.695	8.09
2	12	6	0.3	2	0.505	7.79
3	6	12	0.3	2	0.715	6.29
4	12	12	0.3	2	0.63	5.86
5	6	6	0.9	2	0.745	7.38
6	12	6	0.9	2	0.85	6.82
7	6	12	0.9	2	0.82	6.87
8	12	12	0.9	2	0.62	6.22
9	9	9	0.6	2	0.66	8.01
10	9	9	0.6	2	0.785	7.21
1	6	6	0.3	4	0.8	8.72
2	12	6	0.3	4	1.15	8.85
3	6	12	0.3	4	0.99	5.52
4	12	12	0.3	4	0.91	5.33
5	6	6	0.9	4	0.88	8.62
6	12	6	0.9	4	0.955	8.60
7	6	12	0.9	4	1.13	8.31
8	12	12	0.9	4	0.9	7.41
9	9	9	0.6	4	0.895	10.33
10	9	9	0.6	4	1.035	9.41
1	6	6	0.3	6	1.245	5.93
2	12	6	0.3	6	1.04	6.50
3	6	12	0.3	6	1.38	4.80
4	12	12	0.3	6	1.095	4.00
5	6	6	0.9	6	1.12	5.74
6	12	6	0.9	6	1.215	6.03
7	6	12	0.9	6	1.455	6.42
8	12	12	0.9	6	1.405	5.33
9	9	9	0.6	6	1.32	7.36
10	9	9	0.6	6	1.52	7.17
1	6	6	0.3	8	1.765	2.70
2	12	6	0.3	8	1.76	3.40
3	6	12	0.3	8	1.86	2.78
4	12	12	0.3	8	1.645	2.87
5	6	6	0.9	8	1.93	2.65
6	12	6	0.9	8	1.865	3.13
7	6	12	0.9	8	2.065	3.97
8	12	12	0.9	8	1.83	3.85
9	9	9	0.6	8	1.925	3.97
10	9	9	0.6	8	1.805	3.55

Table 2
Induction-Timing (OD 0.40–0.60) Covariate Effect on %T-20[a]

Source	df	Seq SS	Adj SS	Adj MS	*F*	*p* value	Covariated with
Induction OD	1	4.620	0.024	0.024	0.01	0.939	Glucose
Induction OD	1	4.620	0.228	0.228	0.06	0.814	NPK
Induction OD	1	4.620	0.256	0.256	0.07	0.795	IPTG
induction OD	1	4.620	4.620	4.620	1.60	0.214	Duration

[a]df, degrees of freedom; Seq SS, sequential sum of squares; Adj SS, adjusted sum of squares; Adj MS, adjusted mean squares.

Table 3
Induction-Timing (OD 0.55–0.75) Covariate Effect on %T-20[a]

Source	df	Seq SS	Adj SS	Adj MS	*F*	*p* value	Covariated with
Induction OD	1	2.43	2.15	2.15	0.16	0.694	NPK
Induction OD	1	2.43	13.47	13.47	0.99	0.327	IPTG
induction OD	1	2.43	2.43	2.43	0.55	0.465	Duration

[a]df, degrees of freedom; Seq SS, sequential sum of squares; Adj SS, adjusted sum of squares; Adj MS, adjusted mean square.

To optimize the composition of the assay medium and the fermentation process for T-20 production, a series of statistical designs was created prior to performing experiments. On the basis of the results of pilot experiments (data not shown), four potentially significant factors were selected initially and varied according to the design layout obtained from Design Expert. Three experimental designs were used for the whole optimization process: 2^{4-1} fractional factorial design, 2^3 factorial design, and 2^2 factorial rotational central composite design.

Factors initially considered significant were the amount of glucose, NPK, $MgSO_4$, and trace elements as media formulation factors; and timing, concentration, and duration of induction as process factors. Table 4 shows the experimental factor range of actual values and coded levels of variables (factors) used for initial screening in the 2^{4-1} fractional factorial design. Based on analysis of this design and an additional process factor of IPTG, a new series of trials was performed in the second 2^3 factorial design (Table 5). To fit a second-order polynomial model, a rotational central composite design with two factors selected from the previous 2^3 factorial design was conducted to find optimal conditions (Table 6).

The novel characteristic of this experimental design and modeling is that it is a combination of factorial design and repeated measurement (induction duration) design, which is able to show statistical effects among media and the dynamic status of the process simultaneously from a single DOE.

Table 4
Actual Values and Coded Levels of Factors for 2^{4-1} Fractional Factorial Design

Factor	Coded symbol	Units	Coded level/actual values[a] –1	0	+1
Glucose (20%)	A	mL	5	10	15
NPK (1 *M*)	B	mL	5	10	15
$MgSO_4$ (0.1 *M*)	C	mL	5	10	15
Trace elements	D	mL	0.5	1	1.5
Induction duration	E	h	2 4	6	8

[a]Code for high = (high – average)/step size, and code for low = (low – average)/step size, in which step size = factor range/2.

Table 5
Actual Values and Coded Levels of Factors for 2^3 Factorial Design

Factor	Coded symbol	Units	Coded level/actual values –1	0	+1
Glucose (20%)	A	mL	6	9	12
NPK (1 *M*)	B	mL	6	9	12
IPTG (0.1 *M*)	C	mL	0.3	0.6	0.9
Induction duration	D	h	2 4	6	8

Table 6
Actual Values and Coded Levels of Factors for 2^2 Central Composite Design

Factor	Coded symbol	Units	Coded level/actual values –1	0	+1
NPK (1 *M*)	A	mL	6	8	10
IPTG (0.1 *M*)	B	mL	0.3	0.5	0.7
Induction duration	C	h	2 4	6	8

Analytical Methods

Cell growth was monitored by measuring OD at 600 nm. For the qualitative analysis of T-20, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on a 15% (w/v) gel system. Gels were stained using the Coomassie brilliant blue R method for protein *(12)*. Peptide T-20 was quantitatively analyzed by measuring image profiles of scanned gel with ImageJ 1.29x (National Institutes of Health, Bethesda, MD). All measurements for OD were duplicated, and two center points were used to estimate experimental errors and variations among T-20 production.

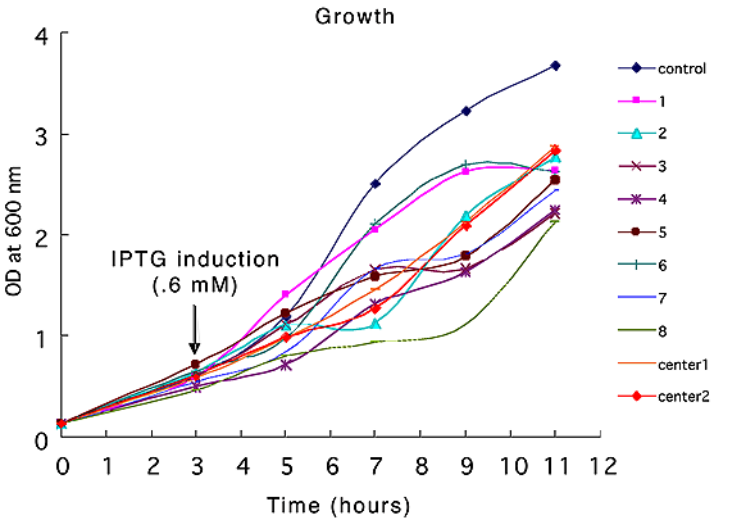

Fig. 1. Time courses of cell growth from differently formulated cultures for 2^{4-1} fractional factorial design.

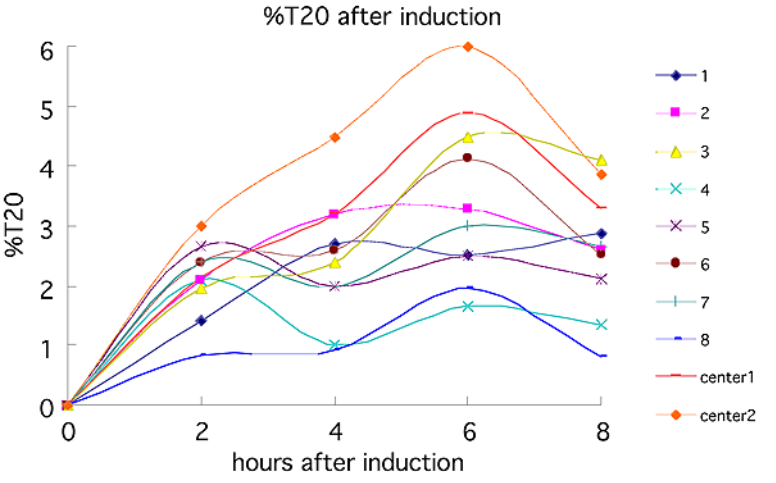

Fig. 2. Time courses of %T-20 from differently formulated cultures for 2^{4-1} fractional factorial design.

Results

2^{4-1} Fractional Factorial Design for Factor Screening

Figures 1 and 2 present cell growth and T-20 production time sequences, respectively. Figure 3 demonstrates that different concentrations of protein were expressed under the different media culture conditions. Figure 4 shows an image profile of the scanned SDS-PAGE gel for

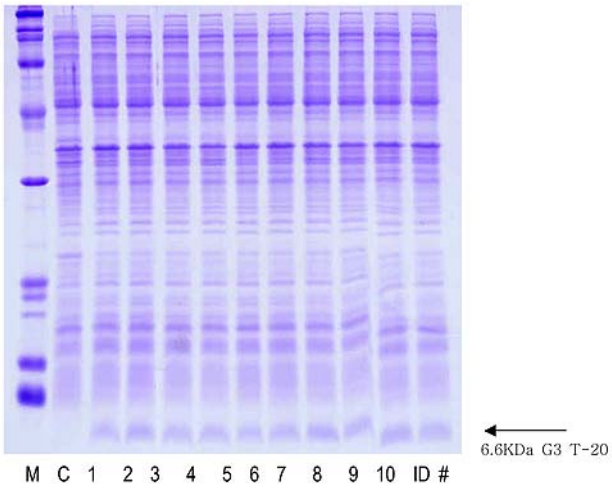

Fig. 3. SDS-PAGE analysis for total cell protein containing T-20 at 6 h after induction. Lane M, size marker; lane C, control without induction; lanes 1–8, design ID; lanes 9 and 10, center points (ID 0) of 2^{4-1} fractional factorial design.

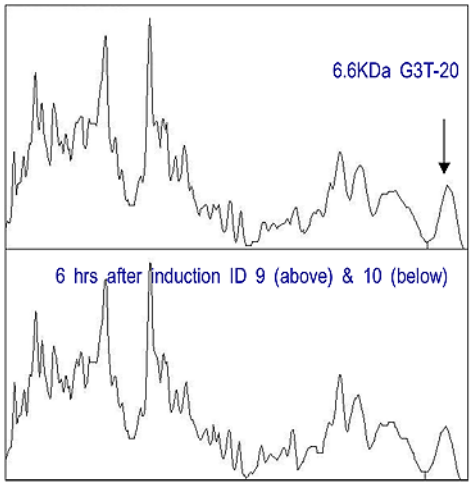

Fig. 4. T-20 quantification by ImageJ. The ratio of the area indicated by the arrow to the whole one was calculated for %T-20.

T-20 quantification. The area with the arrow indicates the proportion of T-20 peptide to the whole protein.

A 2^{4-1} fractional factorial design was used to screen the most significant factors from the initially selected factors. One can estimate the significance of factor effects on the response by making normal (or half normal) plots of the effects as shown in Fig. 5. The squares plotted on the graph correspond to estimates of the model effects, including all possible interactions between factors. Those effects that can be ignored, such as D, AC, C, and AD, are

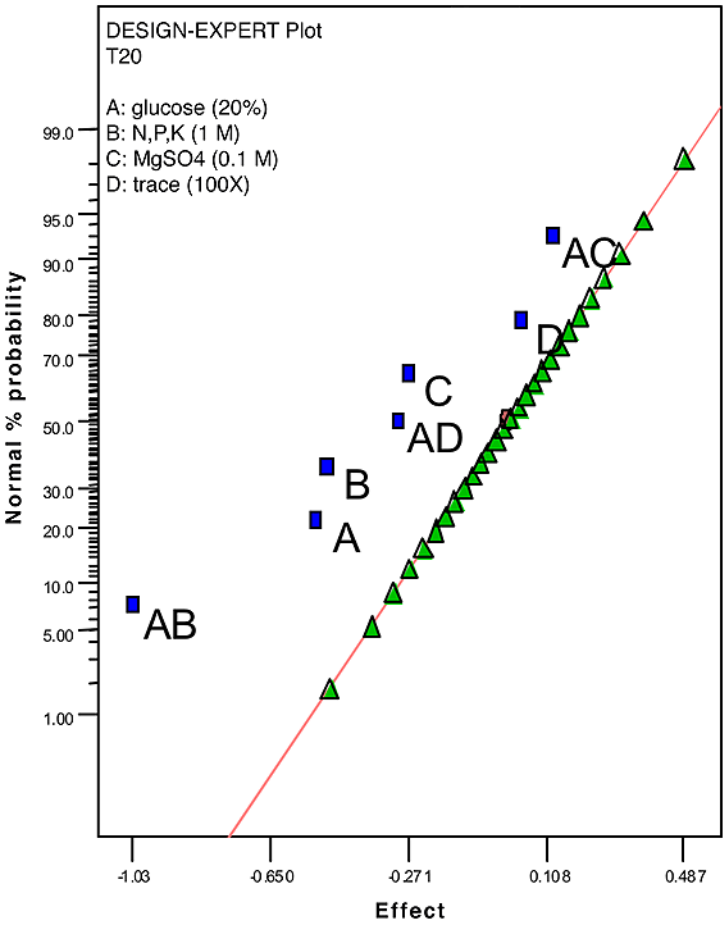

Fig. 5. Normal probability plot of effects obtained from factor screen for %T-20 production of 2^{4-1} fractional factorial design.

required to be parallel with or close to pure errors that are represented by triangles on the straight line. However, significant effects can be picked out as the isolated squares such as B, A, and AB that do not line up along the straight line, their distance from the line depending on the magnitude of their effects. These plots are used to detect the possible significant factors or interactions effects prior to analysis of variance (ANOVA) and regression analysis. This suggests that lower levels (or ranges) of A and B than those of this design should be adjusted to get higher T-20 production than that of this design.

2^3 Factorial Design for Approaching Optimal Region

A 2^3 factorial design was created based on information from the previous design. A smaller scaling factor range than the one applied for screening design was used to narrow down the potential optimal region. $MgSO_4$ and trace elements were fixed at the center point of the previous design,

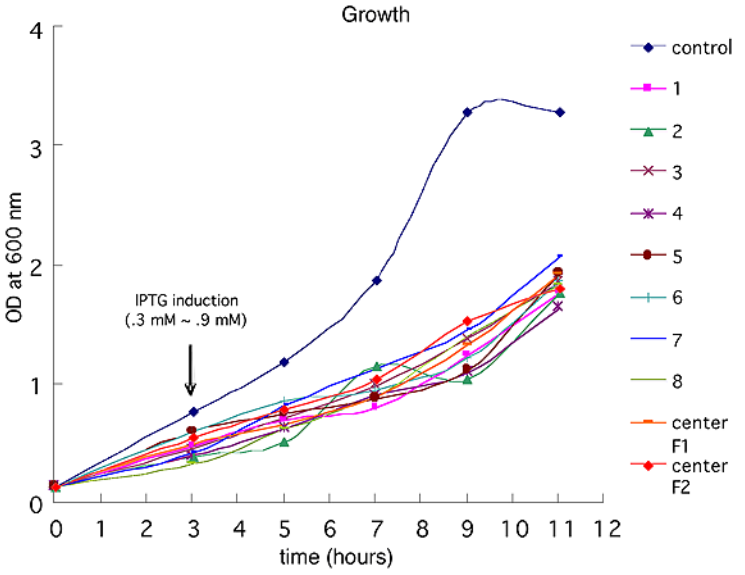

Fig. 6. Time courses of cell growth from differently formulated cultures for 2^3 factorial design.

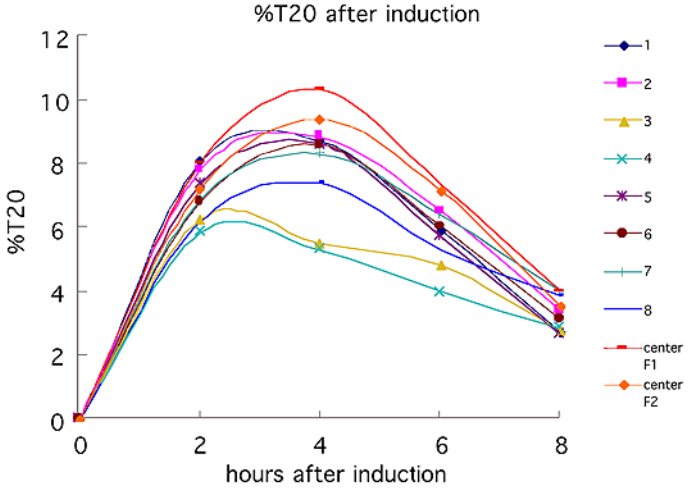

Fig. 7. Time courses of %T-20 from differently formulated cultures for 2^3 factorial design.

owing to their insignificant effects on T-20 production. Table 1 presents the new experimental layout and results. Figures 6 and 7 show cell growth and T-20 production time courses, respectively. Figures 8 and 9 introduce SDS-PAGE gel analysis for T-20 identification and an image profile of the gel for T-20 quantification, respectively. The area under the arrow in Fig. 9 appears to be larger than the one in Fig. 4, implying that the proportion of T-20 peptide to the whole protein was increased.

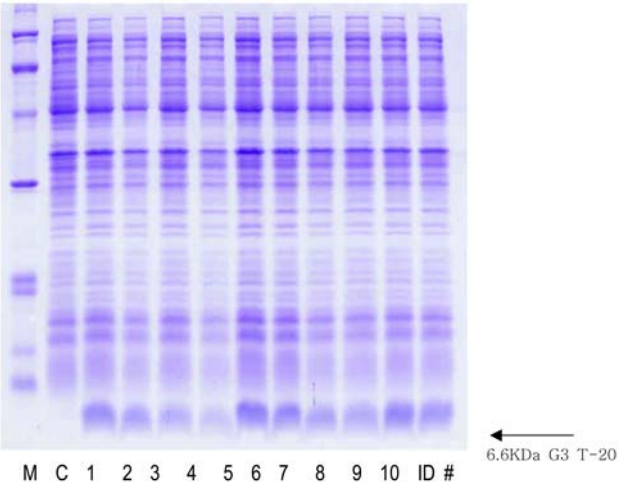

Fig. 8. SDS-PAGE analysis for total cell protein containing T-20 at 4 h after induction. Lane M, size marker; lane C, control without induction; lanes 1–8, design ID; lanes 9 and 10, center points (ID 0) of 2^3 factorial design.

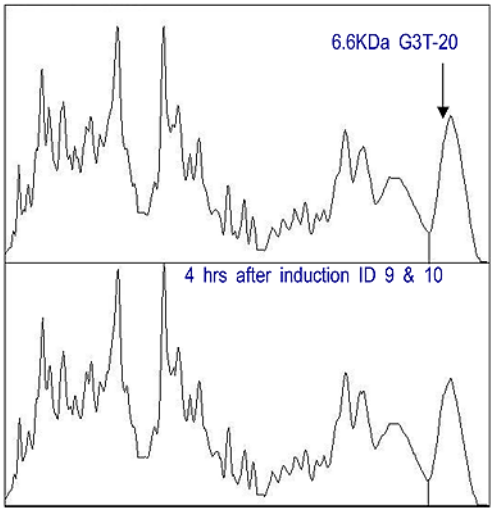

Fig. 9. T-20 quantification by ImageJ. The ratio of the area indicated by the arrow to the whole one was calculated for %T-20.

The normal plot in Fig. 10 indicates that B, C, and BC are significant factors for T-20 production. As shown in Table 7, the factors B ($p = 0.000652$), C ($p = 0.0536$), D ($p < 0.0001$), and BC ($p = 0.00143$) turned out to be significant at a 10% significance level, which means levels of these variables strongly influence T-20 production. However, A ($p = 0.482$) and other interactions showed insignificant effects on T-20 production. A and D showed a quadratic effect. The main effect of B was negative, and C and BC gave positive effects, as seen in the regression coefficients in Table 8. This sug-

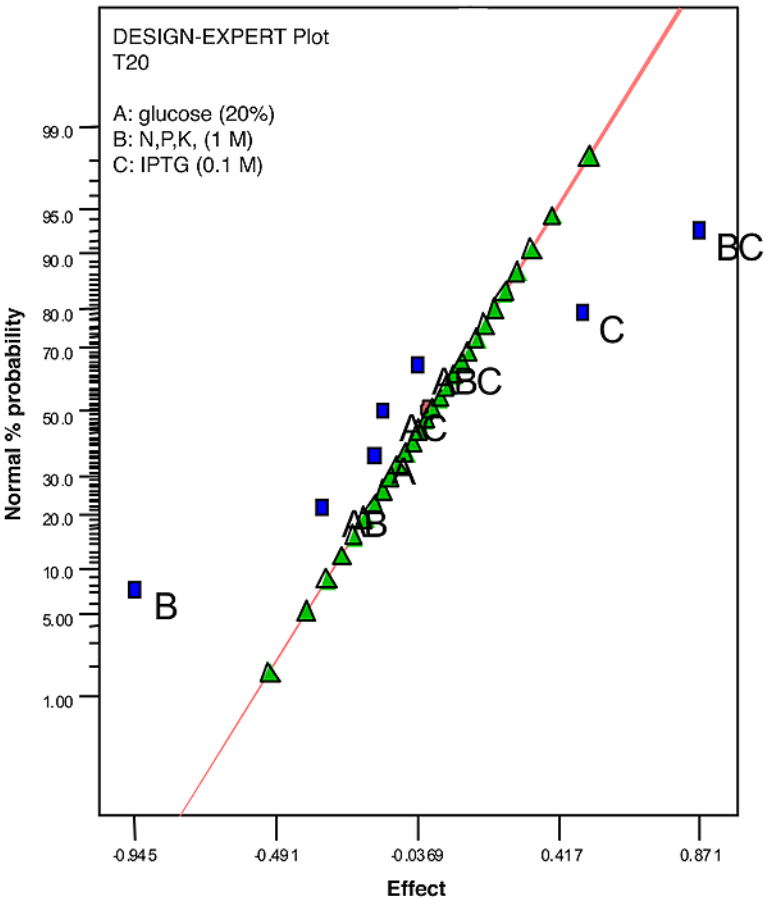

Fig. 10. Normal probability plot of effects on %T-20 production of 2^3 factorial design.

gests that a lower level (or range) of B and a higher level of C should be performed to obtain more T-20 production in this experimental range.

Interestingly, C (duration time) gave rise to significant negative effects on T-20 production, as opposed to the case of two previous designs, which showed significant positive effects on T-20 production (ANOVA data not shown). This is owing to the fact that a shorter induction period (4 h) was required to reach optimal production than that of the former designs (6 h), leading to increased productivity of T-20.

One of the most important factors in determining efficient recombinant protein production is the timing of induction. Because the media compositions varied as different factor combinations, the OD of the culture at induction process was different among treatments (OD range of 0.4–0.6). Thus, the effects of induction timing (OD at induction stage) on T-20 production should be investigated to determine how it influences T-20 production. The statistical method used for determining the induction timing effect on %T-20 was the general linear model, which takes induction timing as a

Table 7
ANOVA for %T-20 From 2^3 Factorial Design[a]

Source	SS	df	MS	*F* value	Prob > *F*
Model	156.484	12.000	13.040	27.097	<0.0001
A	0.245	1.000	0.245	0.509	0.482
B	7.144	1.000	7.144	14.845	0.000652
C	1.960	1.000	1.960	4.073	0.0536
D	90.896	1.000	90.896	188.876	<0.0001
A^2	10.640	1.000	10.640	22.109	<0.0001
D^2	34.170	1.000	34.170	71.002	<0.0001
AB	0.905	1.000	0.905	1.880	0.182
AC	0.171	1.000	0.171	0.356	0.556
AD	0.531	1.000	0.531	1.104	0.303
BC	6.073	1.000	6.073	12.619	0.00143
BD	3.558	1.000	3.558	7.394	0.0113
CD	0.192	1.000	0.192	0.399	0.533
Residual	12.994	27.000	0.481		
Lack of fit	12.144	23.000	0.528	2.486	0.195
Pure error	0.849	4.000	0.212		
Corrected total	169.478	39.000			
SD	0.694			R^2	0.923
Mean	6.095			Adj R^2	0.889
CV	11.382			Pred R^2	0.843
PRESS	26.583			Adeq precision	16.831

[a]SS, sum of squares; df, degrees of freedom; MS, mean square; SD, standard deviation; CV, coefficient of variation; Adj, adjusted; Pred, predicted; Adeq, adequate.

Table 8
Regression Analysis for %T-20 From 2^3 Factorial Design[a]

Factor	Coefficient estimate	df	SE	95% CI: low	95% CI: high	VIF
Intercept	8.282	1.000	0.281	7.705	8.858	
A-glucose (20%)	–0.088	1.000	0.123	–0.339	0.164	1.000
B-N,P,K (1 *M*)	–0.473	1.000	0.123	–0.724	–0.221	1.000
C-IPTG (0.1 *M*)	0.248	1.000	0.123	–0.004	0.499	1.000
D-time	–2.022	1.000	0.147	–2.324	–1.721	1.000
A^2	–1.289	1.000	0.274	–1.852	–0.727	1.000
D^2	–2.080	1.000	0.247	–2.586	–1.573	1.000
AB	–0.168	1.000	0.123	–0.420	0.083	1.000
AC	–0.073	1.000	0.123	–0.325	0.178	1.000
AD	0.173	1.000	0.165	–0.165	0.510	1.000
BC	0.436	1.000	0.123	0.184	0.687	1.000
BD	0.447	1.000	0.165	0.110	0.785	1.000
CD	0.104	1.000	0.165	–0.234	0.441	1.000

[a]df, degrees of freedom; SE, standard error; CI, confidence interval; VIF, variance inflation factor.

covariate. Table 2 shows that induction timing had insignificant effects on T-20 production as it was covariated with glucose, NPK, IPTG, and induction duration time ($p = 0.939$, $p = 0.814$, $p = 0.795$, and $p = 0.214$, respectively) within an OD range of 0.4–0.6.

2^2 Central Composite Design for Optimization

To approach the optimum response region of T-20 production, significant variables (NPK and IPTG) selected from the previous design were further investigated. Other factors such as glucose, $MgSO_4$, and trace elements were kept at the center point applied for the previous design, because no more variations were expected from these variables around their center levels, which were assumed to be optimized or near optimized.

Table 9 provides the new experimental layout and results. Figures 11 and 12 present cell growth and T-20 production, respectively. Figures 13 and 14 present SDS-PAGE analysis and T-20 quantification by Image J, respectively. The growth and protein band patterns were very similar to those of two previous designs. However, based on three graphs (Figs. 2, 7, and 12) of T-20 production, the proportion of T-20 peptide to the whole protein became enhanced as optimal experiments continued.,

According to Tables 10 and 11, only B (IPTG) showed significant positive effects on T-20 production ($p = 0.085$) at the 10% significance level. Table 3 shows that covariate of induction timing had insignificant effects on T-20 production as it was covariated with NPK, IPTG, and induction duration time ($p = 0.694$, $p = 0.327$, and $p = 0.465$, respectively) within an OD range of 0.55–0.75.

The final optimal setting for T-20 production was calculated from the quadratic response surface model from this design (Fig. 15). Based on the results from this study, it was concluded that the maximum 14% T-20 production was obtained from the optimal culture conditions (as below) determined by the sequential and simultaneous statistical experimental designs, i.e., 9, 6, 10, and 1 mL of glucose, NPK sources, $MgSO_4$, and trace elements, respectively, in a total of 100 mL of medium inducted at an OD of 0.55–0.75 with 0.7 m*M* IPTG in a 4-h induction duration time.

Figure 16 demonstrates the reliability of the model for the central composite design. This normal probability plot of the residuals was utilized to determine whether the residuals follow a normal distribution, which is the most important assumption for statistical modeling and model adequacy checking. No significant violations of the model assumptions were found in this residual analysis. Therefore, the modeling (Fig. 12) was able to be used for further studies without any bias.

Adequacy of the model for predicting optimization of T-20 production was tested by using a mixture design (simplex centroid). A verification experiment was performed under selected optimal conditions modified by the proportion of each component to fit the mixture design. Experimental data from the mixture design were within the predicted range of 95% con-

Table 9
Design Matrix and Responses for 2^2 Central Composite Design

ID	Factor 1 (A:NPK) (mL)	Factor 2 (B:IPTG) (mL)	Factor 3 (C: duration) (h)	Response 1 (OD, [abs.])	Response 2 (T-20 [%])
1	6.00	0.30	2.00	0.97	8.16
2	10.00	0.30	2.00	1.02	7.72
3	6.00	0.70	2.00	1.09	9.79
4	10.00	0.70	2.00	0.85	6.09
5	5.17	0.50	2.00	0.90	7.97
6	10.83	0.50	2.00	0.84	7.17
7	8.00	0.22	2.00	1.03	10.50
8	8.00	0.78	2.00	0.86	9.29
9	8.00	0.50	2.00	0.88	8.58
10	8.00	0.50	2.00	0.89	8.98
1	6.00	0.30	4.00	1.22	9.38
2	10.00	0.30	4.00	1.38	12.23
3	6.00	0.70	4.00	1.34	14.21
4	10.00	0.70	4.00	1.09	13.05
5	5.17	0.50	4.00	1.11	14.20
6	10.83	0.50	4.00	1.24	12.89
7	8.00	0.22	4.00	1.37	10.79
8	8.00	0.78	4.00	1.15	7.05
9	8.00	0.50	4.00	1.12	9.28
10	8.00	0.50	4.00	1.16	7.90
1	6.00	0.30	6.00	1.19	6.58
2	10.00	0.30	6.00	1.82	7.52
3	6.00	0.70	6.00	1.57	9.13
4	10.00	0.70	6.00	1.43	6.55
5	5.17	0.50	6.00	1.19	8.90
6	10.83	0.50	6.00	1.64	8.03
7	8.00	0.22	6.00	1.61	8.34
8	8.00	0.78	6.00	1.44	5.82
9	8.00	0.50	6.00	1.42	6.29
10	8.00	0.50	6.00	1.48	4.92
1	6.00	0.30	8.00	2.15	5.20
2	10.00	0.30	8.00	2.33	5.75
3	6.00	0.70	8.00	2.39	5.32
4	10.00	0.70	8.00	2.21	3.26
5	5.17	0.50	8.00	1.76	4.60
6	10.83	0.50	8.00	2.31	5.08
7	8.00	0.22	8.00	2.41	4.51
8	8.00	0.78	8.00	2.15	2.92
9	8.00	0.50	8.00	2.19	3.04
10	8.00	0.50	8.00	2.19	2.88

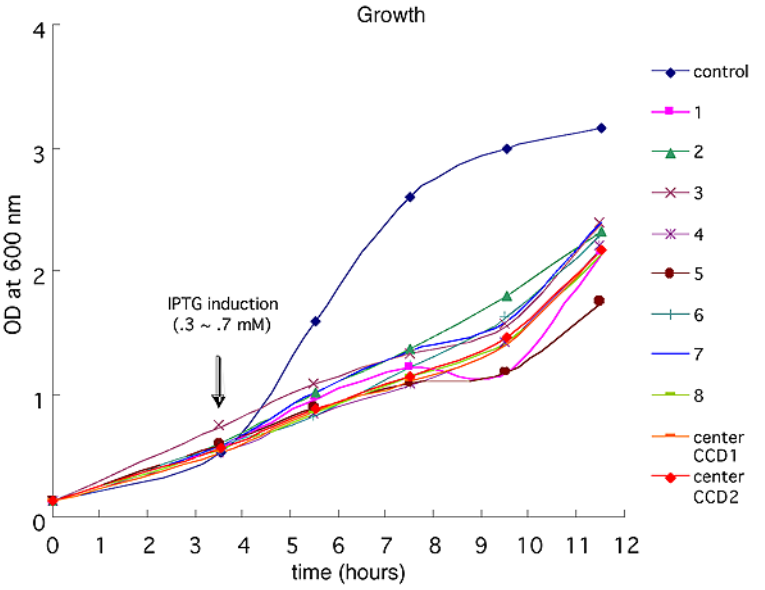

Fig. 11. Time courses of cell growth from differently formulated cultures for 2^2 central composite design.

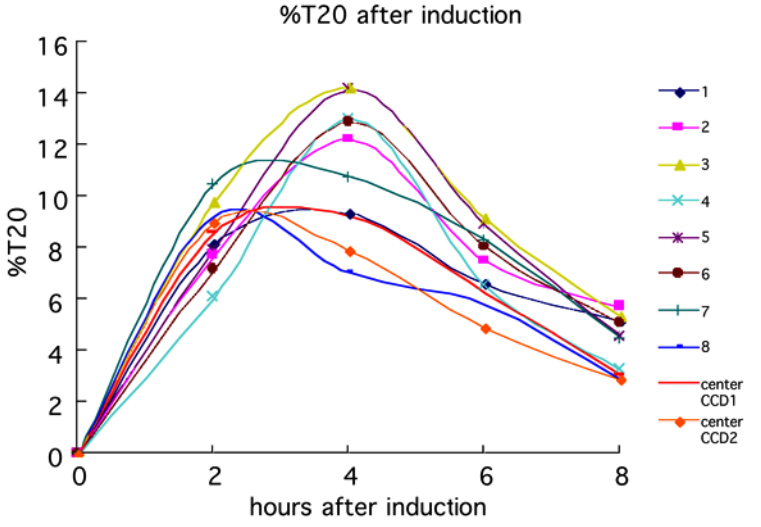

Fig. 12. Time courses of %T-20 from differently formulated cultures for 2^2 central composite design.

fidence level of the optimal point, which confirms the validity of the response model and the success of optimization.

Discussion

The goal for optimal production of recombinant proteins is to produce the highest amount of functional product per unit volume per unit time.

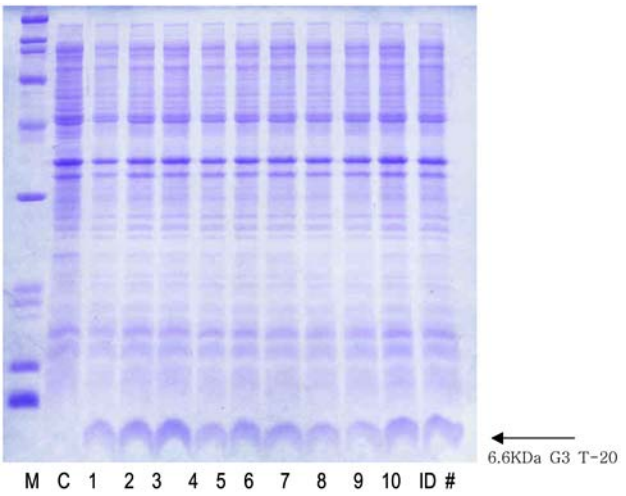

Fig. 13. SDS-PAGE analysis for total cell protein containing T-20 at 4 h after induction. Lane M, size marker; lane C, control without induction; lanes 1–8, design ID; lanes 9 and 10, center points (ID 0) of 2^2 central composite design.

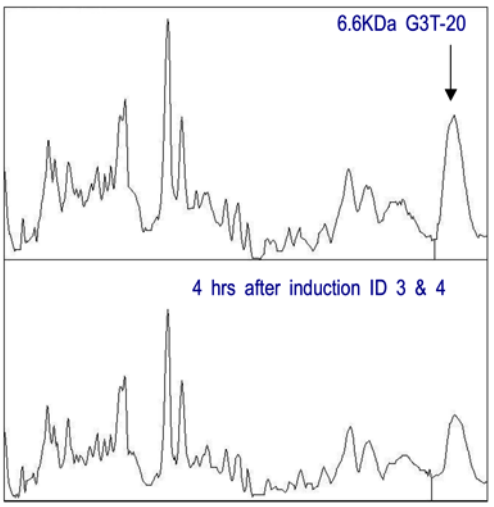

Fig. 14. T-20 quantification by ImageJ. The ratio of the area indicated by the arrow to the whole one was calculated for %T-20.

For any fermentation system, the level of intracellular accumulation of a recombinant protein depends on the final cell density and the specific activity of the protein, with the level of accumulation relative to total protein. Several strategies are typically considered for optimizing the production of recombinant protein, such as media formulation, fermentation mode, strain development, metabolic approach, and expression system control *(5,13–20)*.

Table 10
ANOVA for %T-20 From 2^2 Central Composite Design[a]

Source	SS	df	MS	*F* value	Prob > *F*
Model	296.111	16.000	18.507	8.530	<0.0001
A	0.355	1.000	0.355	0.163	0.690
B	7.030	1.000	7.030	3.240	0.0850
C	63.710	1.000	63.710	29.363	<0.0001
A^2	21.683	1.000	21.683	9.994	0.00436
B^2	4.361	1.000	4.361	2.010	0.170
C^2	79.076	1.000	79.076	36.446	<0.0001
AB	11.204	1.000	11.204	5.164	0.0327
AC	0.701	1.000	0.701	0.323	0.575
BC	0.761	1.000	0.761	0.351	0.559
A^3	0.132	1.000	0.132	0.061	0.807
B^3	9.775	1.000	9.775	4.505	0.0448
C^3	28.139	1.000	28.139	12.969	0.00151
A^2C	4.253	1.000	4.253	1.960	0.175
AC^2	0.295	1.000	0.295	0.136	0.716
B^2C	0.133	1.000	0.133	0.061	0.807
BC^2	0.689	1.000	0.689	0.318	0.578
Residual	49.903	23.000	2.170		
Lack of fit	47.912	19.000	2.522	5.066	0.0633
Pure error	1.991	4.000	0.498		
Corrected total	346.014	39.000			
SD	1.473			R^2	0.856
Mean	7.747			Adj R^2	0.755
CV	19.014			Pred R^2	0.524
PRESS	164.562			Adeq precision	11.540

[a]SS, sum of squares; df, degrees of freedom; MS, mean square; SD, standard deviation; CV, coefficient of variation; Adj, adjusted; Pred, predicted; Adeq, adequate.

Much effort has been aimed at improving recombinant expression systems by manipulating the gene dose, transcriptional controlling machinery, and translation process to maximize the recombinant proteins, and little is known about the effects of media composition on the expression of recombinant protein *(15)*. However, it is well known that the production of secondary metabolites in bacterial strains depends on the composition of the medium *(5,20,21)*. In spite of this knowledge, little attention has been paid to the effects of medium formulation on the production of recombinant proteins.

Past research on microbial culture in the biotechnology industry for several decades made clear that there is no universal medium for the culture. Cells generated by fusion, transformation, or transfection after selection and cloning demonstrate different phenotypes, physically and physiologically. As a consequence, no single set of culture conditions can be applied to different kinds of cells.

Table 11
Regression Analysis for %T-20 From 2^2 Central Composite Design[a]

Factor	Coefficient estimate	df	SE	95% CI: low	95% CI: high	VIF
Intercept	8.243	1.000	0.597	7.008	9.477	
A-NPK	–0.358	1.000	0.885	–2.190	1.474	11.563
B-IPTG	1.594	1.000	0.885	–0.238	3.425	11.563
C-time	–6.914	1.000	1.276	–9.553	–4.274	16.674
A^2	1.089	1.000	0.344	0.376	1.802	1.225
B^2	0.488	1.000	0.344	–0.224	1.201	1.225
C^2	–3.164	1.000	0.524	–4.248	–2.080	1.000
AB	–0.837	1.000	0.368	–1.599	–0.075	1.000
AC	0.199	1.000	0.349	–0.524	0.921	1.000
BC	–0.207	1.000	0.349	–0.930	0.516	1.000
A^3	0.128	1.000	0.521	–0.949	1.206	10.000
B^3	–1.105	1.000	0.521	–2.183	–0.028	10.000
C^3	4.220	1.000	1.172	1.796	6.644	12.674
A^2C	0.647	1.000	0.462	–0.309	1.603	2.625
AC^2	–0.216	1.000	0.586	–1.428	0.996	2.563
B^2C	0.114	1.000	0.462	–0.842	1.070	2.625
BC^2	–0.330	1.000	0.586	–1.542	0.882	2.563

[a]df, degrees of freedom; SE, standard error; CI, confidence interval; VIF, variance inflation factor.

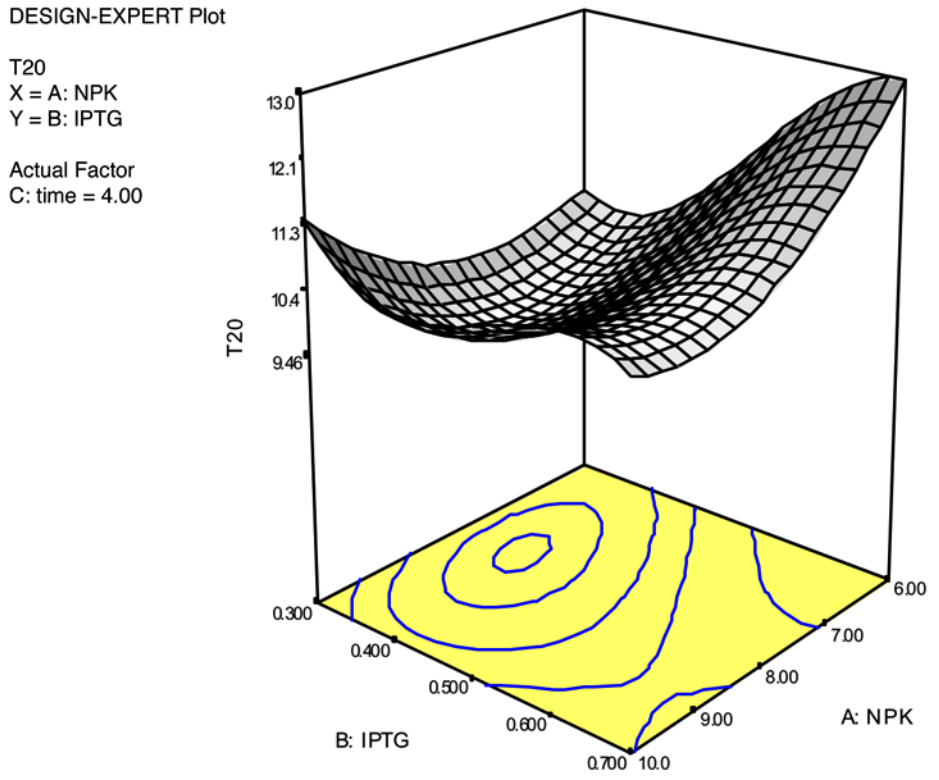

Fig. 15. Three-dimensional model for %T-20 production of 2^2 central composite design at 4 h after induction.

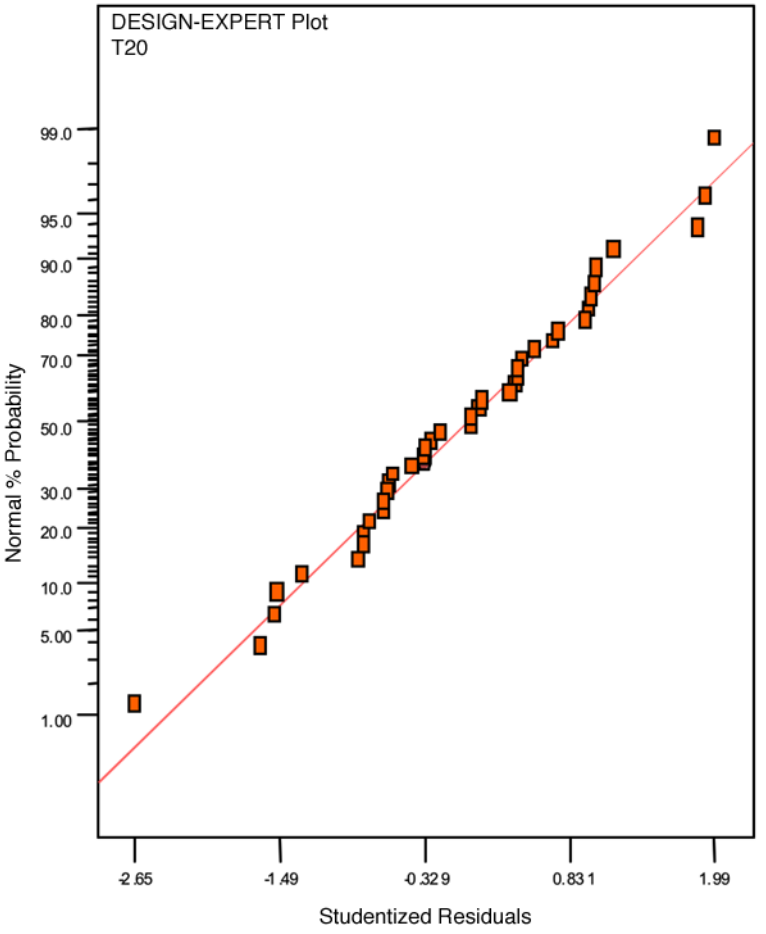

Fig. 16. Residual analysis of %T-20 production of central composite design for model verification.

The goal of media development is to determine the conditions for a particular cell and for a particular end point of production of interest. The optimization process provides the cell with the suitable components and cofactors in order for them to be available at the right time and in the right amount for various metabolic processes involved in target production. In addition, optimization of nutrient requirements results in efficient utilization of energy. Therefore, media and process optimization identifies the best growth and production environment that leads to optimal cell performance in a completely defined medium and process.

Media consist of a variety of components. Thus, a high-throughput screening and developing approach is necessary for optimal performance and for reducing time and cost. Once this approach is determined, it will be universal and can be applied to any kind of cell or culture system. A statistical approach to optimizing the medium and process has been a well-established method. Statistical optimization coupled with experimental

design proved to be a powerful tool for a situation in which several variables need to be considered and yet the relative importance of each is still not known *(10,11,21)*. This statistical optimization technique can be employed for any biotechnology process as well as any cell culture medium.

In the present study, the amount of glucose and NPK source appear to be the most significant factors for the effect of variation on T-20 production. This may be owing to the fact that carbon, nitrogen, and buffer sources are critical factors in determining the balance of cell growth and protein production. Induction duration and IPTG concentration were also significant factors in determining the optimal conditions for T-20 production in that factors relevant to the induction process can influence activation or modulation of the expression of protein. More sophisticated experiments would be necessary to understand the relation between induction and expression optimization. The combined effects of media and process factors were also analyzed in our study. NPK was positively related (interacted) to IPTG and induction duration, as shown in the second design.

Data analysis of cell growth and T-20 production by Pearson's correlation method showed interesting results that high cell densities do not necessarily correlate with high T-20 production (data not shown). Retardation of growth may be caused by metabolic burden (metabolic load) imposed on the host cell during gene expression. Metabolic load can be defined as the portion of a host cell's resources that is required to maintain and express foreign gene *(22)*. The magnitude of the metabolic load owing to the overproduction of foreign protein depends on the expression rate, the size and number of cloning vector, and the composition of the medium *(22,23)*. Thus, the presence of a metabolic load may result in a variety of physiologic and physical alterations to a host cell. Several strategies are being investigated to avoid problems involved in metabolic load and to monitor the signal process for the metabolic load *(23)*. Study for recombinant expression processes and productions should be focused on the metabolic capacity of the host cell.

Cells are exposed to environmental changes as a result of a great variety of metabolic activities and adjust to factors that influence the activities. Coordination of these significant factors makes the cells work properly by providing a unique set of conditions for the specific activities. Overexpression of recombinant protein results in the depletion of building blocks and energy. Although the supply of the depleting nutrients to support the overexpression seems necessary, it does not always guarantee the enhancement of target protein. A high concentration of a certain component may inhibit the synthesis of specific metabolite required for protein production. Therefore, an efficient medium formulation needs to be developed, and a statistical approach allows a rational choice of each nutrient of the culture. This statistical technique makes it possible to predict the best corresponding concentration of a component within experimental ranges.

We have explained herein the applicability of statistical theory to the optimization of recombinant protein expression. We have concluded that the statistical optimization approach with response surface model was a reliable technique to establish the optimal culture conditions for the expression of recombinant protein required for subsequent scale-up to bioreactor production.

Acknowledgment

This research was supported by a grant from the Korea Science and Engineering Foundation (KOSEF).

References

1. Richman, D. (1998), *Nat. Med.* **4(11),** 1232–1233.
2. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997), *Cell* **89,** 263–273.
3. Kilby, J. M., Hopkins, S., Venetta, T. M., et al. (1998), *Nat. Med.* **4(11),** 1302–1307.
4. Rice, W. G., Supko, J. G., Malspeis, L., et al. (1995), *Science* **270,** 1194–1197.
5. Vanot, G., Valerie, D., Guilhem, M. C., Phan-Tan-Luu, R. and Comeau, L. C. (2002), *Appl. Microbiol. Biotechnol*. **60,** 417–419.
6. Rimsky, L. T., Shugars, D. C., and Matthews, T. (1998), *J. Virol.* **72,** 986–992.
7. Wei, X., Ghosh, S. K., Taylor, M. E., et al. (1995), *Nature* **373,** 117–122.
8. Lee, K.-M. and Gilmore, D. F. (2005), *Appl. Biochem. Biotechnol*. **133,** 113–148.
9. Lee, K.-M. and Gilmore, D. F. (2005), *Process Biochem.* **40,** 226–249.
10. Montgomery, D. C. (2004), *Design and Analysis of Experiments*, 6th ed., John Wiley & Sons, New York.
11. Myers, R. H. and Montgomery, D. C. (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments,* 2nd ed., John Wiley & Sons, New York.
12. Bollag, D. M., Rozycki, M. D., and Edelstein, S. J. (1996), *Protein Methods,* 2nd ed., John Wiley & Sons, New York.
13. Buzzini, P. (2000), *J. Ind. Microbiol. Biotechnol.* **24,** 41–45.
14. Coninck, J., Bouquelet, S., Dumortier, V., Duyme, F., and Verdieer-Denantes, I. (2000), *J. Ind. Microbiol. Biotechnol.* **24,** 285–290.
15. El-Heolw, E. R., Abdel-Fatth, Y. R., Ghanem, K. M., and Mohamad, E. A. (2000), *Appl. Microbiol. Biotechnol.* **54,** 515–520.
16. Hounsa, C. G., Aubry, J. M., Dubourguier, H. C., and Hornez, J. P. (1996), *Appl. Microbiol. Biotechnol*. **45,** 746–770.
17. Souza, M. C. O., Roberto, I. C., and Milagres, A. M. F. (1999), *Appl. Microbiol. Biotechnol*. **52,** 768–772.
18. Toksoy, E., Onsan, Z. I., and Kirdar, B. (2002), *World J. Microbiol. Biotechnol*. **18,** 23–27.
19. Tong, W. Y., Yao, S. J., Zhu, Z. Q., and Yu, J. (2001), *Appl. Microbiol. Biotechnol*. **57,** 674–679.
20. Trupkin, S., Levin, L., Forchiassin, F., and Viale, A. (2003), *J. Ind. Microbiol. Biotechnol.* **30,** 682–690.
21. Xie, L., Hall, D., Eiteman, M. A., and Altman, E. (2003), *Appl. Microbiol. Biotechnol.* **63,** 267–273.
22. Glick, B. R. (1995), *Biotechnol. Adv*. **13(2),** 247–261.
23. Cserjan-Puschmann, M., Kramer, M., Duerrschmid, E., Striedner, G., and Bayer, K. (1999), *Appl. Microbiol. Biotechnol*. **53,** 43–50.