

## 重组二硫键异构酶在大肠杆菌中表达条件的统计优化

罗曼, 关怡新\*, 姚善泾

(浙江大学化学工程与生物工程学系, 杭州 310027)

**摘要** 二硫键形成蛋白 A (disulfide bond formation protein A, DsbA) 是大肠杆菌周质胞腔中辅助多种含有二硫键的蛋白质正确折叠并具有生物学活性的一种二硫键异构酶。通过统计实验设计的方法将生产重组 DsbA 的发酵过程进行了优化。首先通过 Plackett-Burman 设计挑选出了对 DsbA 表达量影响较大的四个因素, 然后再利用杂合设计进行实验, 并通过拟合得到了响应曲面函数, 但该函数的驻点是鞍点, 因此不具有全局的极值。最后通过约束优化得到了较佳的实验点, 在该实验点下 DsbA 的表达量比基本培养条件下提高了 50.14%。

**关键词** DsbA, 统计优化, Plackett-Burman 设计, 杂合设计, 响应面法

**中图分类号** Q78, Q332

### Statistical Optimization of Protein Disulfide Isomerase Expression in *E. coli*

LUO Man, GUAN Yi-Xin\*, YAO Shan-Jing

(Department of Chemical and Biochemical Engineering, Zhejiang University, Hangzhou 310027, China)

**Abstract** Disulfide bond formation protein A, DsbA, existing in the periplasm of *E. coli*, is a protein disulfide isomerase (PDI), which assists folding of the disulfide bond containing proteins to achieve biological activities. In this paper, statistical experimental designs were applied to optimize the fermentation process of recombinant DsbA in *E. coli*. Firstly, four most important parameters, which affect the DsbA expression, were screened among 17 parameters by Plackett-Burman design. Then hybrid design was applied to fit a response surface function, where stationary point was a saddle point with neither global maximum nor minimum. Finally, the optimal settings of fermentation parameters were calculated using the nonlinear constrained optimization method. Verification experiment followed by SDS-PAGE showed that the settings resulted in 50.14% higher expression after the optimization.

**Key words** DsbA, statistical optimization, Plackett-Burman design, hybrid design, response surface methodology

二硫键的形成是蛋白质生物合成的一个重要步骤, 其中某些二硫键对蛋白质的三级结构的形成起着关键性的作用<sup>[1]</sup>。此外, 二硫键一般还与蛋白质是否能折叠和组装成为具有生物学活性的结构, 并稳定这种结构有着密切的联系<sup>[2]</sup>。

随着基因工程技术的飞速发展, 人们越来越多地由从生物组织中直接提取转向采用先克隆基因再在宿主菌中表达的方法获得大量的目标蛋白质。常用的宿主菌 *E. coli* 系原核生物, 在其中表达的重组蛋白质由于缺乏类似真核系统的翻译后加工和修饰的调控机制而形成没有生物学活性的蛋白质聚集体, 即包涵体 (inclusion body)。那些含有二硫键的重组蛋白质在 *E. coli* 还原环境中, 表达后不能立即形成正确的二硫键配对, 造成了分子内或分子间的二

硫键错配, 加速了包涵体的形成。如何将这些包涵体进行体外重折叠复性已成为基因工程下游过程的核心问题, 而在包涵体蛋白质体外复性的过程中, 二硫键的形成又是整个反应的限速步骤, 因而富含二硫键的重组蛋白质体外复性一直是研究的热点和难点问题。

收稿日期: 2004-08-25, 接受日期: 2004-11-25

国家自然科学基金资助项目 (No. 20476093)

\*联系人 Tel: 0571-87951982; Fax: 0571-87951015;

E-mail: guanyx@zju.edu.cn

Received: August 25, 2004; Accepted: November 25, 2004

Supported by National Natural Science Foundation of China (No. 20476093)

\* Corresponding author Tel: 0571-87951982; E-mail: guanyx@zju.edu.cn

Dsb (disulfide bond) 家族是一类通过相互协作完成对 *E. coli* 合成的分泌蛋白质中的二硫键氧化形成和还原异构的过程,并最终使之形成正确配对的蛋白质。它们具有共同的活性位点模体 (motif) Cys-X-X-Cys<sup>[3]</sup>。迄今为止,已经发现了 Dsb 家族的 7 个成员: DsbA ~ G, 其中 DsbA 于 1991 年由 Bardwell<sup>[4]</sup> 发现,是该家族中第一个被发现的蛋白质,它存在于 *E. coli* 周质腔 (periplasm) 中,具有最强的氧化性,分子量为 21 kD。最近,有文献报道了 DsbA 辅助变性蛋白质体外重折叠复性的研究,如碱性磷酸酶和牛核糖核酸酶<sup>[5]</sup>、牛胰蛋白酶抑制剂 (EPTI) 和  $\alpha$ -乳白蛋白<sup>[6]</sup>、D-3-磷酸甘油醛脱氢酶 (GAPDH) 和硫氰酸酶<sup>[7]</sup> 及 Ragi 双功能抑制子 (Ragi bifunctional inhibitor, RBI)<sup>[8]</sup> 等。

本文在参考单因素优化方法的<sup>[9]</sup>基础上,采用了统计的方法,对重组 DsbA 发酵过程中培养基成分以及操作条件进行了优化,并得到较高表达量的 DsbA,为进一步研究 DsbA 协助富含二硫键的重组蛋白质的体外复性的机理奠定基础。

## 1 材料和方法

### 1.1 材料

1.1.1 菌株和质粒 宿主菌为 *E. coli* BL21 (DE3), 对氨苄青霉素敏感 (Amp<sup>s</sup>)。质粒为 pAVD63, 其上含有编码 DsbA 的基因以及 Amp 抗性 (Amp<sup>r</sup>) 基因, DsbA 基因的表达由 lac 增强子/操纵子控制。该质粒为英国剑桥大学蛋白质工程中心 Fersht 教授惠赠。

1.1.2 培养基 种子培养基 2 × YT, 发酵培养基为在 LB + M9 的基础上优化得到。

1.1.3 试剂和仪器 无机盐为国产分析纯,其余均为上海生工生物工程公司进口分装。Ultrospec 3300 pro 分光光度计为瑞典 Amersham Biosciences 生产; Sorvall Super T21 为 SORVALL 生产; GD2000 凝胶图象处理系统和 Quantity One 软件购自美国 Bio-Rad 公司。

### 1.2 方法

1.2.1 培养条件优化 种子培养基中添加 Amp 至终浓度 50  $\mu$ g/ml, 500  $\mu$ l 甘油管 (20% 甘油) 接入, 200 r/min, 37  $^{\circ}$ C 摇床过夜培养 12 h。发酵培养基中添加 Amp 至终浓度 50  $\mu$ g/ml, 将种子液按 2% (V/V) 的接种量接入, 200 r/min, 37  $^{\circ}$ C 摇床培养至菌浓  $A_{600}$  约为 1.0, 加 IPTG 至终浓度 1 mmol/L 诱导 7 h 后放罐。在该操作条件下,以 500 ml 摇瓶装液量 150 ml 的 LB 培养基为基准,进行培养基组成和培养条件的

优化。

1.2.2 细胞生长的测定 发酵液适当稀释后在 600 nm 处比色 ( $A_{600}$ ) 测定细胞密度。

1.2.3 SDS-PAGE 聚丙烯酰胺凝胶电泳,分离胶浓度为 12%,浓缩胶浓度为 4%。

1.2.4 DsbA 表达量的确定 放罐后,4、8 000 r/min 离心 10 min,弃上清收集菌体,用发酵体积 1/10 的缓冲液重悬细胞,冰浴 3 s 工作,6 s 间歇,500 W 超声波破碎 90 次,4、14 000 r/min 离心 40 min 得破胞上清。用缓冲液稀释 5 倍后,进行 SDS-PAGE。采用考马斯亮蓝 R-250 染色方法<sup>[10]</sup> 处理后,用凝胶成像系统及软件分析 DsbA 的表达量。

### 1.3 实验设计

1.3.1 基于统计原理的实验设计方法 基于统计的实验设计其基本思想是 Pareto 原理<sup>[11]</sup>。影响一个过程的参数可以分为少许重要的 (the vital few) 和大多数不重要的 (the trivial many); 也有称做“80/20”原则,即通过改变 20% 的过程参数就可以使过程的性能提高 80%。本实验中,以 DsbA 表达量作为响应目标,先利用 Plackett-Burman 设计挑选出对响应影响较大的几个因素,进而利用响应曲面方法 (response surface methodology, RSM) 中的杂合设计进行实验,通过实验数据拟合响应面模型,最终确定最优实验点,并进行验证。

1.3.2 筛选实验——Plackett-Burman 设计 Plackett-Burman 设计由 Plackett 和 Burman 于 1946 年提出<sup>[12]</sup>,它是建立在平衡的非完全区组 (balanced incomplete block) 基础上的,通过  $N$  个实验可以研究  $N-1$  个变量 ( $N$  一般为 4 的倍数)。在实验过程中,通常会预留出伪变量 (dummy variables) 作为误差分析。每个变量有高低两个水平,分别以 +、- 标记,在整个 Plackett-Burman 设计中,每个变量取高低水平的值各  $N/2$  次,而且在某个因素取得高 (低) 水平时,其他各个因素取得高低水平各  $N/4$  次。在本实验中,对 DsbA 表达量产生影响的参数主要包括氮源、无机盐、Amp 浓度、诱导剂浓度、诱导时刻及培养时间等 17 个因素。因此,应选用  $N=20$  的 Plackett-Burman 设计,其中两个变量为伪变量。实验设计如 Table 1 所示,每一行代表一次实验,每一列表示了一个考察的参数或伪变量。各参数代表的变量及高低水平对应的真实值见 Table 2。以基本培养条件下的表达量为单位,上述培养条件下的表达量与之比较,其相对表达量,作为每种条件下的响应值。

1.3.3 优化实验——杂合设计 (hybrid design) 响

**Table 1** Matrix for Plackett-Burman experimental design

| Run No. | Variables/Levels |   |   |   |   |   |       |   |   |   |   |   |   |       |   |   |   |   |   |
|---------|------------------|---|---|---|---|---|-------|---|---|---|---|---|---|-------|---|---|---|---|---|
|         | A                | B | C | D | E | F | (G) * | H | J | K | L | M | N | (O) * | P | Q | R | S | T |
| 1       | +                | - | + | + | - | - | -     | - | + | - | + | - | + | +     | + | + | - | - | + |
| 2       | +                | + | - | + | + | - | -     | - | - | + | - | + | - | +     | + | + | + | - | - |
| 3       | -                | + | + | - | + | + | -     | - | - | - | + | - | + | -     | + | + | + | + | - |
| 4       | -                | - | + | + | - | + | +     | - | - | - | - | + | - | +     | - | + | + | + | + |
| 5       | +                | - | - | + | + | - | +     | + | - | - | - | - | + | -     | - | - | + | + | + |
| 6       | +                | + | - | - | + | + | -     | + | + | - | - | - | - | +     | - | + | - | + | + |
| 7       | +                | + | + | - | - | + | +     | - | + | + | - | - | - | +     | - | + | - | + | + |
| 8       | +                | + | + | + | - | - | +     | + | - | + | + | - | - | -     | + | - | + | - | - |
| 9       | -                | + | + | + | + | + | -     | + | + | - | + | + | - | -     | - | - | + | - | + |
| 10      | +                | - | + | + | + | + | -     | - | + | + | - | + | + | -     | - | - | - | + | - |
| 11      | -                | + | - | + | + | + | +     | - | - | + | + | - | + | +     | - | - | - | - | + |
| 12      | +                | - | + | - | + | + | +     | + | - | - | + | + | - | +     | + | - | - | - | - |
| 13      | -                | + | - | + | - | + | +     | + | + | - | - | + | + | -     | + | + | - | - | - |
| 14      | -                | - | + | - | + | - | +     | + | + | + | - | - | + | +     | - | + | + | - | - |
| 15      | -                | - | - | + | - | + | -     | + | + | + | + | - | - | +     | + | - | + | + | - |
| 16      | -                | - | - | - | + | - | +     | - | + | + | + | + | - | -     | + | + | - | + | + |
| 17      | +                | - | - | - | - | + | -     | + | - | + | + | + | + | -     | - | + | + | - | + |
| 18      | +                | + | - | - | - | - | +     | - | + | - | + | + | + | +     | - | - | + | + | - |
| 19      | -                | + | + | - | - | - | -     | + | - | + | - | + | + | +     | + | - | - | + | + |
| 20      | -                | - | - | - | - | - | -     | - | - | - | - | - | - | -     | - | - | - | - | - |

\* Parentheses denote for the dummy variables

**Table 2** The actual values of variables chosen for study versus the coded values

|     | Variables                           | Unit   | Low level        | High level       |
|-----|-------------------------------------|--------|------------------|------------------|
| A   | Tryptone                            | g/L    | 6                | 12               |
| B   | Yeast extract                       | g/L    | 6                | 12               |
| C   | Casamino acid                       | g/L    | 5                | 10               |
| D   | NaCl                                | g/L    | 4                | 8                |
| E   | Na <sub>2</sub> HPO <sub>4</sub>    | g/L    | 6                | 9                |
| F   | KH <sub>2</sub> PO <sub>4</sub>     | g/L    | 3                | 6                |
| (G) | Dummy <sub>1</sub>                  | —      | - 1              | 1                |
| H   | NH <sub>4</sub> Cl                  | g/L    | 4                | 8                |
| J   | MgSO <sub>4</sub>                   | mmol/L | 0                | 5                |
| K   | TES per fermentation volume (in ml) | μl     | 0                | 1                |
| L   | Amp                                 | μg/ml  | 50               | 100              |
| M   | IPTG                                | mmol/L | 0.5              | 1                |
| N   | Temperature                         |        | 33               | 37               |
| (O) | Dummy <sub>2</sub>                  | —      | - 1              | 1                |
| P   | Induction time                      | h      | 5                | 8                |
| Q   | Induction at A <sub>600</sub> =     | —      | 0.5 <sup>a</sup> | 1.4 <sup>b</sup> |
| R   | Shaking speed                       | r/min  | 180              | 220              |
| S   | Medium volume per 500ml shake flask | ml     | 150              | 200              |
| T   | Inoculum size                       | %      | 1                | 2                |

a. A<sub>600</sub> = 0.4 denotes for the prophase of log phase.b. A<sub>600</sub> = 1.4 denotes for the anaphase of log phase

应曲面法可以对人们感兴趣的受多个变量影响的响应问题进行建模与分析,并可以将该响应进行优化<sup>[13]</sup>.其中的杂合设计是由 Roquemore 于 1976 年首次提出<sup>[14]</sup>,这种设计可以通过最少的实验点来拟合

二阶的响应面模型,前面的因素具有 5 个水平,最后一个因素具有 4 个水平.本实验所采用的杂合设计是 416B 型,具体设计方案及各水平对应的变量真实值如 Table 3 所示.

**Table 3** Matrix for hybrid design and predicted values of the relative expression of DsbA

| Run No. | Variables/Levels            |  |                             |   |                             |                               |                             |                                | Relative expression of DsbA |           |
|---------|-----------------------------|--|-----------------------------|---|-----------------------------|-------------------------------|-----------------------------|--------------------------------|-----------------------------|-----------|
|         | NH <sub>4</sub> Cl          |  | IPTG                        |   | t                           |                               | Induction time              |                                | Experimental                | Predicted |
|         | X <sub>1</sub> <sup>a</sup> | x <sub>1</sub> <sup>b</sup> /g L <sup>-1</sup> | X <sub>2</sub> <sup>a</sup> | x <sub>2</sub> <sup>b</sup> /mmol L <sup>-1</sup> | X <sub>3</sub> <sup>a</sup> | x <sub>3</sub> <sup>b</sup> / | X <sub>4</sub> <sup>a</sup> | x <sub>4</sub> <sup>b</sup> /h |                             |           |
| 1       | 0                           | 3.5  | 0                           | 0.4   | 0                           | 39                            | 1.732                       | 14.20                          | 0.3965 ±0.0210              | 0.3965    |
| 2       | 0                           | 3.5  | 0                           | 0.4   | 0                           | 39                            | -0.270                      | 8.192                          | 0.9089 ±0.0061              | 0.9090    |
| 3       | -1                          | 2  | -1                          | 0.2   | -1                          | 36                            | 0.605                       | 10.81                          | 0.9936 ±0.0067              | 0.9476    |
| 4       | 1                           | 5  | -1                          | 0.2   | -1                          | 36                            | 0.605                       | 10.81                          | 0.8109 ±0.0028              | 0.8568    |
| 5       | -1                          | 2  | 1                           | 0.6   | -1                          | 36                            | 0.605                       | 10.81                          | 0.9079 ±0.0069              | 0.9538    |
| 6       | 1                           | 5  | 1                           | 0.6   | -1                          | 36                            | 0.605                       | 10.81                          | 0.8223 ±0.0210              | 0.7763    |
| 7       | -1                          | 2  | -1                          | 0.2   | 1                           | 42                            | 0.605                       | 10.81                          | 0.6058 ±0.0262              | 0.6517    |
| 8       | 1                           | 5  | -1                          | 0.2   | 1                           | 42                            | 0.605                       | 10.81                          | 0.7662 ±0.0059              | 0.7202    |
| 9       | -1                          | 2  | 1                           | 0.6   | 1                           | 42                            | 0.605                       | 10.81                          | 0.6127 ±0.0001              | 0.5667    |
| 10      | 1                           | 5  | 1                           | 0.6   | 1                           | 42                            | 0.605                       | 10.81                          | 0.5027 ±0.0115              | 0.5486    |
| 11      | 1.518                       | 5.777  | 0                           | 0.4   | 0                           | 39                            | -1.050                      | 5.851                          | 0.9841 ±0.0092              | 0.9841    |
| 12      | -1.518                      | 1.223  | 0                           | 0.4   | 0                           | 39                            | -1.050                      | 5.851                          | 0.7684 ±0.0050              | 0.7684    |
| 13      | 0                           | 3.5  | 1.518                       | 0.704   | 0                           | 39                            | -1.050                      | 5.851                          | 0.6873 ±0.0018              | 0.6873    |
| 14      | 0                           | 3.5  | -1.518                      | 0.097   | 0                           | 39                            | -1.050                      | 5.851                          | 0.6185 ±0.0020              | 0.6187    |
| 15      | 0                           | 3.5  | 0                           | 0.4   | 1.518                       | 43.55                         | -1.050                      | 5.851                          | 0.6627 ±0.0189              | 0.6626    |
| 16      | 0                           | 3.5  | 0                           | 0.4   | -1.518                      | 34.45                         | -1.050                      | 5.851                          | 1.0823 ±0.0078              | 1.0823    |

a : coded value

b : actual value

**1.3.4 验证实验** 通过将优化实验所得的实验结果与响应面模型进行拟合,便可回归得到模型中的各个系数,对该多元函数进行简单的性状分析便可确定出其极值点以及取得极值的相应的自变量的值.进一步按照计算所得到的参数进行实验,以验证模型的可靠性,并能够确定最后的优化结果.上述每个实验均重复 3 次,对应的响应值取实验结果平均值.

**1.4 数据处理**

**1.4.1 筛选实验的数据处理** 每个因素的效应可由下式进行计算:

$$E_{A_i} = \frac{R_{A_i(+)} - R_{A_i(-)}}{N/2} \quad (1)$$

式中  $E_{A_i}$  表示变量  $A_i$  的效应,  $N$  表示实验的次数,本实验中  $N$  为 20,  $R_{A_i(+)}$ 、 $R_{A_i(-)}$  分别表示变量  $A_i$  取高水平和低水平值时的响应.实验误差可由计算伪变

量的方差来估计,而标准差(S. E.)则由方差的平方根决定:

$$S. E. = \sqrt{V_{eff}} = \sqrt{\frac{(E_d)^2}{n}} \quad (2)$$

其中  $E_d$  为伪变量的效应,  $n$  是伪变量的个数.最后各因素的显著性水平 ( $P$  值) 则可以通过学生  $t$  检验 (student  $t$ -test) 进行计算:

$$t_{(N-2), A_i} = \frac{E_{A_i}}{S. E.} \quad (3)$$

**1.4.2 响应面的拟合与分析** 采用二次 Scheffe 模型将系统对变量的响应行为进行表征:

$$y = 0 + \sum_{i=1}^k i X_i + \sum_{i=1}^{j-1} \sum_{j=1}^k ij X_i X_j + \sum_{i=1}^k ii X_i^2 \quad (4)$$

式中  $y$  代表系统的响应,  $0$ 、 $i$ 、 $ii$  分别是偏移项、线性偏移和二阶偏移,  $ij$  是交互效应,  $X_i$  为各因素的水平值(即编码值).

将式(4)改写成矩阵形式如下:

$$y = y_0 + X^T b + X^T B X \quad (5)$$

其中  $X$  为自变量向量,  $b$  为线性偏移项向量,  $B$  为二阶偏移及交互效应项向量. 通过求解系数矩阵  $B$  的特征值, 由其符号就可以判断出驻点处函数的性状从而确定最优的实验点.

## 2 结果

### 2.1 重要影响因素的选取

按照 Table 1 进行 Plackett-Burman 设计实验, 所得 DsbA 的相对表达量的正态概率如 Fig. 1 所示. 可见参数 G、H、M、N、P 和 S 偏离直线较远, 其中 G、H、M 和 S 具有较大的负效应, N 和 P 具有较大的正效应. 一般认为显著性水平在 95% 以上的因素对系统的响应影响很大. 经过统计分析, 计算得到各因素的效应、 $t$  值、显著性水平等. 结果表明, 参数 H、M、N 和 P 对 DsbA 的相对表达量影响很大, 即分别为:  $\text{NH}_4\text{Cl}$  浓度、IPTG 终浓、温度以及诱导时间. 因此, 在后续的优化实验中, 应对这四个因素进行进一步的考察. 根据其他因素的效应的正负和大小, 可以确定优化实验中这些因素的取值为: 9 g/L 胰蛋白胨, 9 g/L 酵母抽提物, 10 g/L 酪蛋白氨基酸 (casamino acid), 4 g/L NaCl, 9 g/L  $\text{Na}_2\text{HPO}_4$ , 6 g/L  $\text{KH}_2\text{PO}_4$ , 装液量为 150 ml/500 ml, 2% (V/V) 接种量, TES 添加量 150  $\mu\text{l}$ /瓶,  $\text{MgSO}_4$  和 Amp 终浓分别为 3 mmol/L 和 100  $\mu\text{g/ml}$ , 在 220 r/min 摇床中培养至菌浓  $A_{600}$  为 1.4 时加入诱导剂.

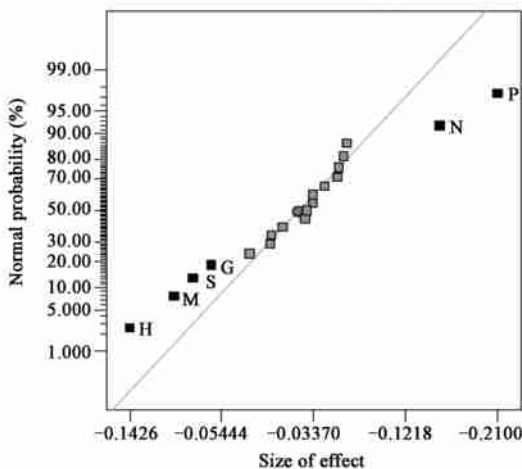


Fig. 1 Normal probability plot based on Plackett-Burman design experiments

### 2.2 响应曲面的拟合以及最佳操作点的确定

在筛选实验中, 确定了 4 个待优化的参数, 按照

Table 3 进行杂合设计实验, 仅需 16 次实验即可完成优化过程. 通过拟合, 得到二次响应曲面函数为:

$$\begin{aligned} y = & 0.8974 + 0.0087 X_1 - 0.0180 X_2 - 0.1336 X_3 \\ & - 0.0762 X_4 + 0.0150 X_1^2 - 0.0820 X_2^2 \\ & + 0.0133 X_3^2 - 0.1231 X_4^2 - 0.0217 X_1 X_2 \\ & + 0.0399 X_1 X_3 - 0.0594 X_1 X_4 - 0.0228 X_2 X_3 \\ & - 0.0387 X_2 X_4 + 0.0044 X_3 X_4 \end{aligned} \quad (6)$$

式中,  $X_1 \sim X_4$  分别代表培养基中  $\text{NH}_4\text{Cl}$  浓度、IPTG 终浓、温度以及诱导时间的编码值. 类似地, 响应曲面函数亦可由上述 4 种因素的真实值 ( $x_1 \sim x_4$ ) 形式给出:

$$\begin{aligned} y = & 3.8130 - 0.2384 x_1 + 3.8640 x_2 - 0.1801 x_3 \\ & + 0.2735 x_4 + 0.0066 x_1^2 - 2.0490 x_2^2 \\ & + 0.0015 x_3^2 - 0.0137 x_4^2 - 0.0722 x_1 x_2 \\ & + 0.0089 x_1 x_3 - 0.0132 x_1 x_4 - 0.0380 x_2 x_3 \\ & - 0.0645 x_2 x_4 + 0.0005 x_3 x_4 \end{aligned} \quad (7)$$

对模型的方差分析结果表明, 拟合的多元相关系数为 0.9686, 说明模型对实验点的拟合相当好. 另一方面, 由于大于  $F$  的概率小于 0.0001, 因而模型的效果十分显著, 即模型能够很好地说明实验考察区域的响应情况.

进一步对用参数真实值表示的响应曲面(7)进行分析, 求得其驻点为:

$$\begin{aligned} x_1 = & 36.6723, \quad x_2 = 0.8006, \\ x_3 = & -36.9201, \quad x_4 = -10.2674 \end{aligned} \quad (8)$$

然后再将响应曲面函数化为矩阵形式, 得到二次项系数矩阵的特征值为:

$$= -2.0504, -0.0694, 0.0635, 0.0017 \quad (9)$$

因为这四个特征值异号, 从而可以判断该驻点为鞍点 (saddle point), 因此, 该响应曲面函数既没有最大值, 也没有最小值. 为了更加直观地说明培养基中  $\text{NH}_4\text{Cl}$  浓度、IPTG 终浓、温度以及诱导时间四个变量对 DsbA 相对表达量的影响以及表征响应曲面函数的性状, 作出了以两两自变量为坐标的 3D 图及等值线, 如 Fig. 2 所示.

由于响应曲面函数的特殊性, 无法求取全局的最大值. 因而 DsbA 相对表达量最大的问题只能转化为一个带约束的优化问题. 考虑到函数中各个自变量的真实物理含义, 该带约束的优化问题可以表示为:

$$\begin{aligned} & \max\{y\}; \\ & \text{subject to: } x_1 = 0; \quad x_2 > 0; \quad x_3 \leq 33; \quad x_4 > 0 \end{aligned} \quad (10)$$



Fig. 2 3D and isoreponse contour plots of the response surface function with absent variables at their middle level

(a) IPTG-NH<sub>4</sub>Cl; (b) Temperature-NH<sub>4</sub>Cl; (c) Temperature-IPTG;

(d) Induction time-NH<sub>4</sub>Cl; (e) Induction time-IPTG; (f) Induction time-Temperature

这里考虑到 NH<sub>4</sub>Cl 的添加量为非负, IPTG 必须添加, 诱导时间必须为正值等因素. 值得指出的是, 由于温度过低将导致 *E. coli* 生长过于缓慢, 延滞期变长等现象, 从而会延长发酵生产的周期, 因此, 对于培养温度 ( $x_3$ ) 给出了 33 这一下限. 此带约束的优化问题的解为:

$$y_{\max} = y(0, 0.4885, 33, 9.4463) = 1.4864 \quad (11)$$

即在不添加 NH<sub>4</sub>Cl 的情况下, 33 培养至  $A_{600}$  为 1.4, 加入终浓度为 0.4885 mmol/L 的 IPTG 诱导 9.4463 h 后放罐, 可使 DsbA 的相对表达量达到最佳.

### 2.3 拟合优化的验证

根据筛选实验中确定的参数条件, 以及拟合优

化得到关键参数的最佳实验点, 确定了最终较优的发酵条件, 即培养基组成为: 9 g/L 胰蛋白胍, 9 g/L 酵母抽提物, 10 g/L 酪蛋白氨基酸, 4 g/L NaCl, 9 g/L Na<sub>2</sub>HPO<sub>4</sub>, 6 g/L KH<sub>2</sub>PO<sub>4</sub>. 装液量为 150 ml/500 ml, 添加 TES 150 μl/瓶、MgSO<sub>4</sub> 终浓至 3 mmol/L、Amp 终浓至 100 μg/ml, 2% (V/V) 接种量, 在 33, 220 r/min 摇床中培养至菌浓  $A_{600}$  为 1.4 时加入终浓度为 0.4885 mmol/L 的 IPTG 诱导 9.4463 h 放罐. 在该条件下, DsbA 的相对表达量为 1.5014, 与模型预测的相对误差为 ±1.0%. Fig. 3 为优化前后的破胞上清中 DsbA 表达量的比较电泳图.

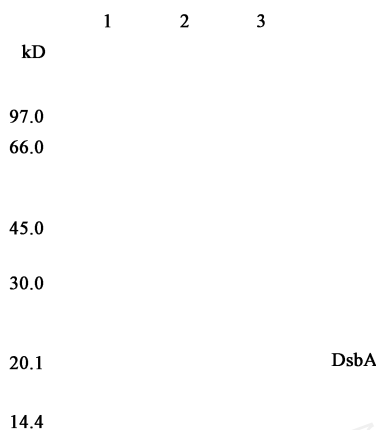


Fig. 3 SDS-PAGE analysis of DsbA expression with different medium and culture conditions

1:Marker; 2:Not optimized condition; 3:Optimized condition

### 3 讨论

Plakett-Burman 设计可以很好地从众多的因素中挑选出那些对系统响应影响较大的部分,本实验成功地考察了培养基成分(胰蛋白胨、酵母抽提物、无机盐等的浓度)和培养条件(培养温度、装液量、摇床转速等)等综合因素的影响.然而由于 Plakett-Burman 设计的高度分化(fractionated),本实验中所采用的设计实际上是  $2^{19}$ -完整析因设计的  $5 \times 2^{17}$  部分,主效应相互混杂(confounding),并且可能造成两因素间、三因素间,甚至更高阶的因素间的交互作用,这样就会给 Plakett-Burman 设计的分析带来很大的偏差.例如在本实验中,从筛选实验的结果中分析出培养温度对 DsbA 的相对表达量有正效应,即培养温度高对目标蛋白表达有利,然而在最后的优化实验中,通过对拟合的响应曲面函数的分析,却说明在 33 这样较低的温度下使 DsbA 的相对表达量达到最大.这种实验结果分析的偏差可以通过交互作用图(interaction plot)来解释,在交互作用图中,如果两条线平行则说明两因素间无交互作用,相反两条线若相交,则说明两因素间存在交互作用.作出培养温度同  $\text{NH}_4\text{Cl}$  浓度的交互作用图,如 Fig. 4 所示.可以看出两者的交互作用很大,这样交互作用的效应就会同主效应产生混杂.同时,从统计分析的角度来看,拟合得到的响应曲面函数(7)的各项回归系数的分析结果表明,  $\beta_{13}$  的显著性远大于  $\beta_3$ ,也就是说,两者交互作用对响应的影响远大于培养温度这个单因素对响应的影响,从而使得再筛选实验结果的分

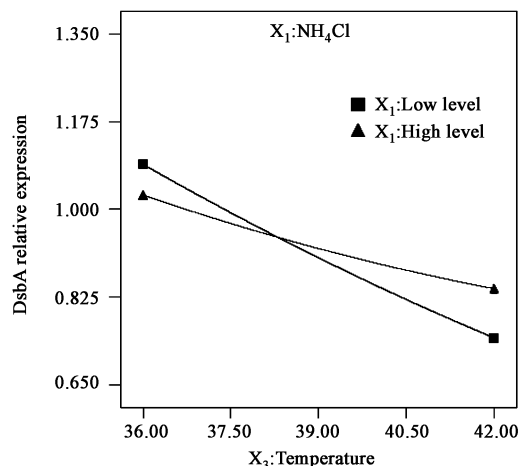


Fig. 4 The interaction plot for culture temperature and  $\text{NH}_4\text{Cl}$  concentration

析中得到了培养温度具有正效应的错误推断.为了考察这种两因素间的交互作用同主效应的混杂,可以将实验设计为次数为 2 的幂次数<sup>[15]</sup>或者按照 Plakett-Burman 设计矩阵 D 的对映形式,即按照矩阵 D,进行实验,再将两次实验结果一起分析<sup>[16]</sup>,便可以将上述的混杂区分开来.

响应面方法特别适合于一些具体机理尚不清楚,无法利用理论依据进行建模的系统,它可以在小的实验范围内通过对实验点数据的拟合,在一定程度上对该范围内系统的响应情况作出表征,从而研究者可以通过对响应曲面函数的数学分析来确定目标实验点.对于拟合二阶响应面模型的实验设计,不但要求设计中每个因素必须至少有三个水平,而且对其正交性(orthogonality)和可旋转性(rotatability)也有较高的要求.杂合设计与小型中心组合设计(small CCD)、正多边形设计具有同等的正交度,具有几乎最小的实验数,而且是可旋转或近似可旋转的<sup>[14]</sup>.杂合设计对离群值(outlier)和缺失数据十分敏感.尽管 Roquemore 早在 1976 年就给出了 3、4、6 以及 7 个因素的杂合设计的设计表,然而还未有报道将其用于优化实验,运用于本实验中,得到了十分良好的效果,也建立了对所研究的实验空间的响应值进行模拟与预测的模型.

除此之外,还可以看出,携带含有编码 DsbA 基因质粒的重组 *E. coli*,在生理及生长特性上,并无较大变化.实验中发现,碳源的添加虽不会影响菌体生长,却在很大的程度上抑制了目标蛋白 DsbA 的表达.在基本 LB 培养基中添加 10 g/L 甘油、5 g/L 葡萄糖、5 g/L 甘露醇、2.5 g/L 蔗糖以及 4 g/L 木糖,发现

除了蔗糖外,其余的发酵产品中 DsbA 的表达量均有较大的下降,其中以葡萄糖尤为显著.蔗糖作为一种非代谢糖,其存在只能对培养基的渗透压起到调节的作用,而不能被菌体生长所利用,然而其他的碳源却均会被代谢消耗.因此,这种添加碳源却造成目标蛋白表达量下降的情况的原因,推测是这些代谢糖在代谢途径上的某些产物,对基因的表达调控或者目标蛋白质的合成起到了一定的抑制作用而造成的.另外,Ramirez 等人<sup>[17]</sup>指出,在重组蛋白质的合成过程中,由于各类氨基酸在胞内的大量消耗,会造成胞内氨基酸池(amino acid pool)的耗尽,因而添加了酪蛋白氨基酸的培养基可以在一定程度上补足这种氨基酸的局部大量的消耗.

### 参考文献 (References)

- 1 Saitô N, Yura K, Fukuda Y. Some remarks on protein folding. In: Hatano M. ed. *Protein Structure Analysis, Folding and Design*. Japan: Japan Scientific Societies Press, 1990: 31
- 2 Jeffrey W N, Thomas E C. Reactivity and ionization of the active site cysteine residues of DsbA, a protein required for disulfide bond formation *in vivo*. *Biochemistry*, 1994, **33**: 5974 ~ 5983
- 3 Debarbieux L, Beckwith J. Electron avenue: pathways of disulfide bond formation and isomerization. *Cell*, 1999, **99**: 117 ~ 119
- 4 Bardwell J C A, McGovern K, Beckwith J. Identification of a protein required for disulfide bond formation *in vivo*. *Cell*, 1991, **67**: 581 ~ 589
- 5 Akiyama Y, Kamitani S, Kusukawa N, Ito K. *In vitro* catalysis of oxidative folding of disulfide-bonded proteins by the *Escherichia coli dsbA* (*pyfA*) gene product. *J Biol Chem*, 1992, **267** (31): 22440 ~ 22445
- 6 Zapun A, Creighton T E. Effect of DsbA on the disulfide folding of bovine pancreatic trypsin inhibitor and  $\alpha$ -lactalbumin. *Biochemistry*, 1994, **33**: 5202 ~ 5211
- 7 Zheng W D, Quan H, Song J L, Yang S L, Wang C C. Does DsbA have chaperone-like activity? *Arch Biochem Biophys*, 1997, **337** (2): 326 ~ 331
- 8 Maskos K, Huber-Wunderlich M, Gockshuber R. DsbA and DsbC-catalyzed oxidative folding of proteins with complex disulfide bridge patterns *in vitro* and *in vivo*. *J Mol Biol*, 2003, **325**: 495 ~ 513
- 9 黄鹤,杨晟,李仁宝,黄晓冬,袁中一. 重组青霉素 G 酰化酶在枯草芽孢杆菌中的表达条件优化. 中国生物化学与分子生物学报 (Huang He, Yang Sheng, Li Ren-Bao, Huang Xiao-Dong, Yuan Zhong-Yi. Optimization of recombinant penicillin G acylase expression in *Bacillus subtilis*. *Chin J Biochem Mol Biol*), 2001, **17** (2): 173 ~ 177
- 10 Sambrook J, Russell D W. *Molecular Cloning: A Laboratory Manual*, 3rd ed. New York: Cold Spring Harbor Laboratory Press, 2001
- 11 Haaland P D. *Experimental Design in Biotechnology*. New York: ASQC Quality Press, 1989: 6
- 12 Plackett R L, Burman J P. The design of optimum multifactorial experiments. *Biometrika*, 1946, **37**: 305 ~ 325
- 13 Montgomery D C. *Design and Analysis of Experiments*, 3rd ed. New York: J Wiley, 1991: 521 ~ 568
- 14 Roquemore K G. Hybrid designs for quadratic response surfaces. *Technometrics*, 1976, **18** (4): 419 ~ 423
- 15 Stowe R A, Mayer R P. Efficient screening of process variables. *Ind Eng Chem*, 1966, **58** (2): 36 ~ 40
- 16 Biles W E, Swain J J. *Optimization and Industrial Experimentation*. New York: J Wiley, 1980: 176 ~ 177
- 17 Ramirez D M, Bentley W E. Enhancement of recombinant protein synthesis and stability *via* coordinated amino acid addition. *Biotechnol Bioeng*, 1993, **41**: 557 ~ 565