

## Optimizing expression of antiviral cyanovirin-N homology gene using response surface methodology and protein structure prediction

H. Lotfi<sup>1,2,3</sup>, M. A. Hejazi<sup>3</sup>, M. K. Heshmati<sup>4</sup>, S. A. Mohammadi<sup>2,5</sup>, N. Zarghami<sup>1,2\*</sup><sup>1</sup> Drug Applied Research Center, Tabriz University of Medical Sciences, Tabriz, Iran<sup>2</sup> Department of Medical Biotechnology, Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran<sup>3</sup> Department of Genomics, Branch for the Northwest and West Region, Agricultural Biotechnology Research Institute of Iran (ABRII), Agriculture research, Education and Extension Organization (AREEO), Tabriz, Iran<sup>4</sup> Department of Food Science and Technology, Faculty of Agriculture, University of Tabriz, Iran<sup>5</sup> Department of Plant Breeding and Biotechnology, Faculty of Agriculture, University of Tabriz, Tabriz, IranCorrespondence to: [zarghami@tbzmed.ac.ir](mailto:zarghami@tbzmed.ac.ir)

Received July 25, 2017; Accepted September 8, 2017; Published September 30, 2017

Doi: <http://dx.doi.org/10.14715/cmb/2017.63.9.17>

Copyright: © 2017 by the C.M.B. Association. All rights reserved.

**Abstract:** Cyanovirin-N (CVN) is well known as an anti-HIV protein. The efficient production of low cost microbicides for preventing the HIV-infection has lately become a requirement worldwide. The aim of the present study was to optimize the expression of antiviral Cyanovirin-N homology gene found in the indigenous strain of *Nostoc ellipsosporum* LZN using Response Surface Methodology (RSM) and Protein Structure Analysis. Optimization of three induction factors (IPTG concentration (0.1, 0.55 and 1mM), temperature for bacterial growth (20, 28.5 and 37°C) and induction time (4, 10 and 16h) was done using RSM and Box-Behnken Design. Total RNA extraction was performed and mRNA levels were quantified in each experimental design by one-step SYBR qPCR. Protein structure was predicted using I-TASSER server. The full-length sequence of LZN-CVN gene is 306 bp in length, due mostly to five mutations. RSM analysis showed that the optimum condition to obtain maximum fold change was a concentration of 0.6mM IPTG, temperature set to 29°C and a 12h long induction time. The extracted protein from periplasmic fraction (8 kDa) was verified via SDS-PAGE. The high percentage of LZN-CVN similarity was demonstrated with PDB (Protein Data Bank) accession code of 2rp3A (CVN domain B mutant) and the ligand binding sites were related to N42, V43, D44, G45, S52, N53 and E56 residues. Different expression systems could assist in the development of anti-HIV proteins in a large scale. The LZN-CVN protein was successfully expressed in the *E.coli* system. RSM could be applied to a series of mathematical and statistical methods for modeling and analysis of responses which are influenced by various variables of interest.

**Key words:** Antiviral; Cyanovirin-N; Gene expression; *Nostoc ellipsosporum*; Response Surface Methodology; Sequencing.

### Introduction

Cyanovirin-N (CVN) was discovered in *Nostoc ellipsosporum* as a stable and effective antiviral protein especially against human immunodeficiency virus (HIV) (1). Moreover, its effect against the Ebola virus (2), influenza virus (3), simian immunodeficiency virus (4) feline immunodeficiency virus (5), hepatitis C virus, measles virus, HSV-1 and HSV-6 (herpes simplex virus type) (6-8) was reported. The key characteristics of the viruses responsible for their sensitivity to CVN are associated with surface glycoproteins, such as N-linked high mannose oligosaccharides, which provides unique sites for binding with the CVN (9). The antiviral potency of CVN is related to its high affinity for binding HIV-gp120 which subsequently inhibits the binding of virion-gp120 to host CD4 receptor (5, 10). The ability of CVN as a powerful HIV-inactivation protein along with the related remarkable physicochemical properties led to its use as an interesting anti-HIV medication (11). Challenges for applying new approaches to produce recombinant proteins (r-proteins) or pharmaceuticals are still present (12, 13).

Besides, the production of active recombinant CVN (rCVN) in large scale will be achieved via different ex-

pression systems with relative efficiency as well as low cost approaches. Bacterial expression systems suitable for producing rCVN include: *Escherichia coli* (14-16), *Streptococcus gordonii* (17), *Lactobacillus jensenii* (18) *Lactococcus lactis* and *Lactobacillus plantarum* (19). *Pichia pastoris* (20) and the transgenic plant *Nicotiana tabacum* (21, 22), *Althaea officinalis* (23), soybean seeds (24), and the endosperm of rice (25) were all used as a eukaryote expression system. It is noticeable that the function of rCVN is sensitive to forming disulfide bond and modifying side chains by expression hosts (26).

Furthermore, by choosing the best expression system, mathematical and statistical methods will also be developed to produce the recombinant protein of interest with an adequate expression level. Recently, the RSM has been in use for optimizing the variables which influence the expression level (such as inducer concentration, growth temperature, time, pH, cell density, media components and etc.).

RSM applying statistical tools and the experimental design of the selected factors could qualify the significant level of variables in an expected response (27-31). This study was aimed to isolate the nucleotides of the CVN gene from the indigenous *N.ellipsosporum* strain

LZN (named as LZN-CVN). Then, the gene was cloned in pET expression plasmids and the expression level in *E. coli* (BL21) was optimized by qPCR and RSM to further investigate its activity as anti-HIV protein. In addition, the related protein structure was predicted using I-TASSER server.

## Materials and Methods

### Microorganism preparation

*N. elliposporum* strain LZN (Persian Type Culture Collection (PTCC: 1659)) used in this work was retrieved from the Iranian Research Organization for Science and Technology (IROST), Tehran, Iran. This strain had been isolated from soil, Babol, Iran.

### Growth medium conditions

The medium (blue-green nitrogen-fixing medium without  $\text{NaNO}_3$ ) components for growing the Nostoc included 0.04g of  $\text{K}_2\text{HPO}_4$ , 0.075g of  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.036g of  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ , 0.006g of Citric acid, 0.006g of Ferric ammonium citrate, 0.001g of EDTA, 0.02g of  $\text{Na}_2\text{CO}_3$  per liter and 1 ml of Trace Metal Mix A5 (2.86g of  $\text{H}_3\text{BO}_3$ , 1.81g of  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ , 0.222g of  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.39g of  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ , 0.079g  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , 49.4 mg Co  $(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$  per liter). The medium was autoclaved at  $121^\circ\text{C}$  for 15 minutes and adjusted to a final pH of 7.1. The culture was maintained in the medium at  $25^\circ\text{C}$  and sub-cultured at regular intervals. The survival of the strain was monitored with light microscopy.

### Genomic DNA extraction

CTAB protocol was used for DNA extraction (32). The quantity and quality of the isolated DNA was determined via electrophoresis on 1% agarose gel.

### Primer design, PCR amplification and sequencing of Cyanovirin-N gene

The CVN sequence from *N. elliposporum* was obtained from the submitted sequence in GenBank. Two pairs of specific primers were designed using Oligo7 software (Forward primer: 5' CTTGGTAAATTCTCC-CAGACCTGC 3' and reverse primer: 5' TTCGTATTT-CAGGGTACCGTCGATG 3'). CVN gene was amplified by high-fidelity PCR master mix (total volume was 20 $\mu\text{l}$ ). The protocol of PCR amplification included: initial denaturation ( $94^\circ\text{C}$  for 3 min) which was followed by 35 amplification cycles ( $94^\circ\text{C}$  for 45 s,  $54^\circ\text{C}$  for 1min, and  $72^\circ\text{C}$  for 1min) and  $72^\circ\text{C}$  for 5min as a final extension cycle. Then 1% agarose gel electrophoresis was used for analyzing the quantity and quality of the PCR products. The purified products were sequenced by Macrogen Company (Korea). Analysis and sequence alignment were carried out by Chromaspro2 software (version 2.0.1. available at <http://technelysium.com.au/wp/chromaspro/>) and BLAST service at the National Center for Biotechnology Information (NCBI) Genbank, respectively.

### Cloning of LZN-CVN

The LZN-CVN sequence was cloned between *NcoI* and *XhoI* restriction sites without the stop coding sequences to link the gene to the 6-histidine tag of pET22b vector (Novagen). Expression host in this

study was *E. coli* BL21 (DE3) strain. The transformed bacteria were kept in 25% glycerol at  $-70^\circ\text{C}$ .

## Experimental design and data analysis

A Box-Behnken design (BBD) was adopted to optimize three main expression variables (IPTG concentration ( $X_1$ ), temperature ( $X_2$ ) and induction time ( $X_3$ )) necessary for recombinant protein production. The experimental BBD design included 17 runs of three parameters at three levels ( $-1$ ,  $0$  and  $+1$ ). The coded and natural values are explained in table 1, and the combinations of variable levels used in BBD are listed in table 2. The aim of RSM was to investigate the relative contributions of  $X_1$ ,  $X_2$  and  $X_3$  to the response under experiments (recombinant protein production was followed by maximum mRNA level and fold change). The effects of the parameters on the response was determined using second-order polynomial model. The following model indicates the effect of the parameters (the linear, quadratic, and cross-product terms) in equation 1 (Eq.1): Eq.(A1):

$$Y = \beta_0 + \sum \beta_i X_i + \sum \beta_{ii} X_i^2 + \sum \beta_{ij} X_i X_j$$

where  $Y$ ,  $X_i$  and  $X_j$  are the fold change and the coded independent parameters, and  $\beta_0$  and  $\beta_i$  are the constant term and the coefficient of the linear terms, respectively.

**Table 1.** Independent parameters and their values for optimizing the expression conditions.

Independent variables	Symbols	Range and levels*		
		-1	0	+1
IPTG concentration (mM)	$X_1$	0.1	0.55	1
Temperature ( $^\circ\text{C}$ )	$X_2$	20	28.5	37
Induction Time (h)	$X_3$	4	10	16

\*-1, +1 and 0 were considered as a lower, higher value and central point condition in the experiment, respectively.

**Table 2.** Box-Behnken Design with the experimental and predicted values of the response.

Run order	Un- Coded value			Expe. fold change	Pred. fold change
	$X_1$ (mM)	$X_2$ ( $^\circ\text{C}$ )	$X_3$ (h)		
1	0(0.55)	0(28.5)	0(10)	5.3	5.348
2	0(0.55)	0(28.5)	0(10)	5.7	5.348
3	0(0.55)	0(28.5)	0(10)	5	5.348
4	0(0.55)	-1(20)	-1(4)	0.33	-0.045
5	-1(0.1)	1(37)	0(10)	1.78	1.4725
6	1(1)	0(28.5)	1(16)	4.7	3.8925
7	-1(0.1)	0(28.5)	-1(4)	0.81	1.6175
8	0(0.1)	1(37)	-1(4)	0.07	-0.045
9	1(1)	-1(20)	0(10)	1.2	1.4725
10	0(0.55)	-1(20)	1(16)	1.06	2.23
11	1(1)	1(37)	0(10)	1.65	1.4725
12	1(1)	0(28.5)	-1(4)	0.44	0.1225
13	0(0.1)	0(20)	0(10)	5.24	5.348
14	-1(0.1)	0(28.5)	1(16)	2.08	2.3975
15	0(0.55)	1(37)	1(16)	2.91	2.23
16	0(0.55)	0(28.5)	0(10)	5.5	5.348
17	-1(0.1)	-1(20)	0(10)	1.26	1.4725

Finally,  $\beta_{ij}$  and  $\beta_{ji}$  are the coefficient of the quadratic terms, and the crossproduct coefficient, respectively

A Response-surface analysis of the final experimental results was performed using the commercial statistical software (Design Expert, version 10.0.1, Stat Ease Inc., Minneapolis, USA). The analysis of variance (ANOVA) indicated the significant terms for the response in the model. The adequacy of the model was specified via analysis of the model, lack-of fit test and coefficient of determination ( $R^2$ ). Regression coefficient was found between the experimental and predicted responses. The generations of response surfaces, contour and 3D plots as well as the numerical optimization were performed with the help of Design Expert software.

### Isolation of total RNA and qPCR

The total RNA extraction was followed according to the protocol of RNX-Plus kit (Sinaclone Co.). The quality of the extracted RNA was determined using the absorbance ratio ( $A_{260nm}/A_{280nm}$ ) and 2% agarose gel electrophoresis. Then, the RNA extractions were treated with Thermo Scientific DNase I according to the Manufacturer's guidelines. Also, the production of cDNA was performed using Thermo Scientific RevertAid H Minus Reverse Transcriptase and random hexamer primers. Quantification of mRNA levels was determined by one-step Thermo Scientific Maxima SYBR Green qPCR Master Mix. The primer sequences used for the amplification of LZN-CVN were 5'CTTGG-TAAATTCTCCCAGACCTGCTACAACCTC3' and 5'TCAGCAGCCAGCCAGTTCGG 3', as a forward and reverse primers, respectively. The forward and reverse primers for 16sRNA were 5'GTGTTGTGAAA-TGTTGGGTTAAGTCCCGC3' and 5'TTTATGAG-GTCCGCTTGCTCTCGC 3'. Based on these primers, 219-bp and 218-bp amplicons were amplified for LZN-CN and 16s rRNA, respectively. The mRNA level was normalized to the 16sRNA mRNA level as an internal standard. The presented mRNA expression level was calculated through  $2^{-\Delta\Delta CT}$  method.

### Preparation of periplasmic fraction and SDS-PAGE

Induction by IPTG was performed based on the parameter levels for maximum fold change which was obtained from RSM analysis. In the expression conditions, recombinant *E. coli* BL21 (DE3) was cultivated in LB (Luria-Bertani) medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) at 37°C and supplemented with 100 µg/mL ampicillin. Cell density,  $OD_{600}$ , before the induction was 0.6-0.8.

The bacterial cells were centrifuged at 5000g for 10 min and periplasmic fraction was extracted via osmotic shock method. In summary, the cell pellets were re-suspended in a solution containing sucrose (20%), Tris-HCl (0.2 M), lysozyme (0.05%), and EDTA (1 mM). Then the suspension was incubated (15 minute at room temperature), cold water (equal volume) was added and the mixture incubated on ice (15 min). Finally, centrifugation ( $13000 \times g$  at 4°C) was done to obtain the periplasmic fraction. The supernatant was separated and analyzed as a periplasmic fraction by SDS-PAGE (33).

The proteins from the periplasmic fractions were desalted, dia-filtrated and concentrated with Amicon® Ultra Centrifugal Filters (Molecular Weight Cutoff:

3KDa), then analyzed using 17% SDS-PAGE (34). The gel was stained with a dye including 50% methanol, 10% acetic acid and 0.1% Coomassie brilliant blue R-250 and de-stained with a solution containing 30% ethanol and 10% acetic acid. The tagged recombinant protein was purified using Ni-NTA His-bind resin (Novagen).

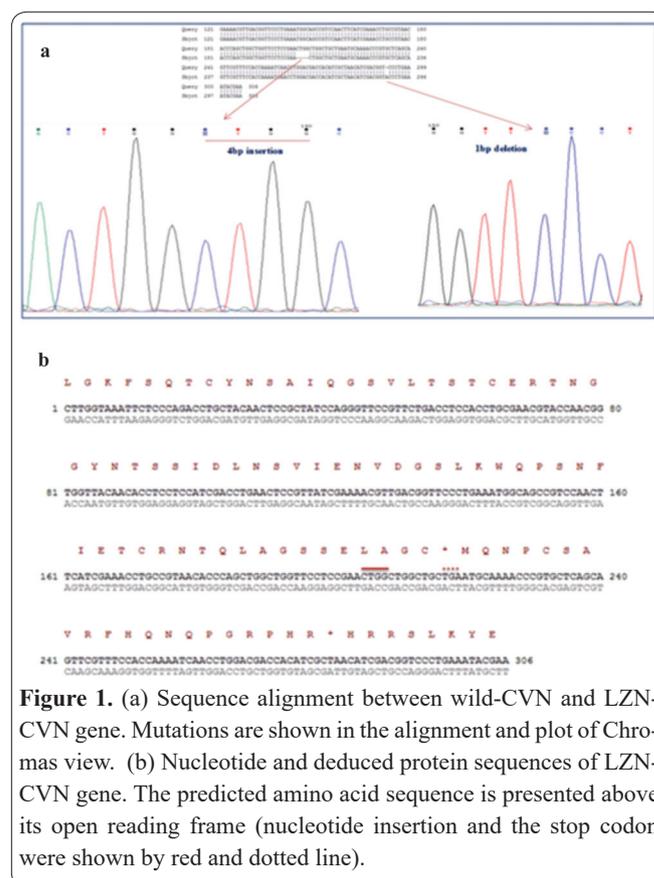
### Prediction of LZN-CVN homology protein structure

Primary sequence analysis molecular weight, overall charge, and charge variation of LZN-CVNH was conducted according to theoretical isoelectric point (pI) via ProtParam using ExPASy server online (<http://www.expasy.ch/tools/protparam.html>). Prediction of secondary structure and function of LZN-CVNH and its functionality was obtained by I-TASSER (Iterative Threading ASSEMBly Refinement available at <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (35).

## Results

### Identification of LZN-CVN gene

The survival of strain LZN was observed after 20 days of incubation by cell monitoring with light microscopy. Using specific primers, a single amplicon of ~303bp was produced from *N. elliposporoum* strain LZN (data not shown). Sequence comparison of LZN-CVN gene (recently submitted as GenBank KX673892) with wild-CVN showed that five mutations (4-bp insertion: CTGG (nt 205-208) and 1-bp deletion: A<sub>289</sub>) had changed the full length of wild-CVN from 303bp to 306 bp. The remaining nucleotides were exactly identical to the CVN gene. These mutations are shown in alignment results and plot of Chromas view (Figure.1a). The predicted amino acid residues are displayed above the corresponding open reading frame (Figure.1b). The LZN-



**Figure 1.** (a) Sequence alignment between wild-CVN and LZN-CVN gene. Mutations are shown in the alignment and plot of Chromas view. (b) Nucleotide and deduced protein sequences of LZN-CVN gene. The predicted amino acid sequence is presented above its open reading frame (nucleotide insertion and the stop codon were shown by red and dotted line).

CVN gene encodes a shorter protein of 72 amino acid residues (nt 1-219) compared with wild-CVN which encodes a 101-amino-acid protein. This reduction was due to the insertion mutations which had changed the corresponding open reading frame.

**RSM analysis of BBD design**

The model was obtained as a result of fitting Eq. (1) to the experimental data listed in Table 2. This model was tested for adequacy and fitness through analysis of variance. Table 3 lists the related model regression coefficients and the corresponding  $R^2$  for the response being studied. In the present study, the quadratic model was proposed for fold change response which includes linear, interaction and quadratic terms. The ANOVA results suggested that the mentioned model possessed non-significant lack of fit at 95% CI ( confidence interval), thus indicating a highly significant p value ( $< 0.001$ ) of the quadratic model for the response (Table 3). Consequently, responses could be predicted using an appropriately accurate model. The high  $R^2$  value ( $> 0.977$ ) of the fold change response displayed a good correlation between the experimental data and the data predicted by the proposed model (Table 3).

Quadratic equation describing the effect of different applied expression parameters on fold change level, based on the coded level of the parameters, are given as Eq.2:

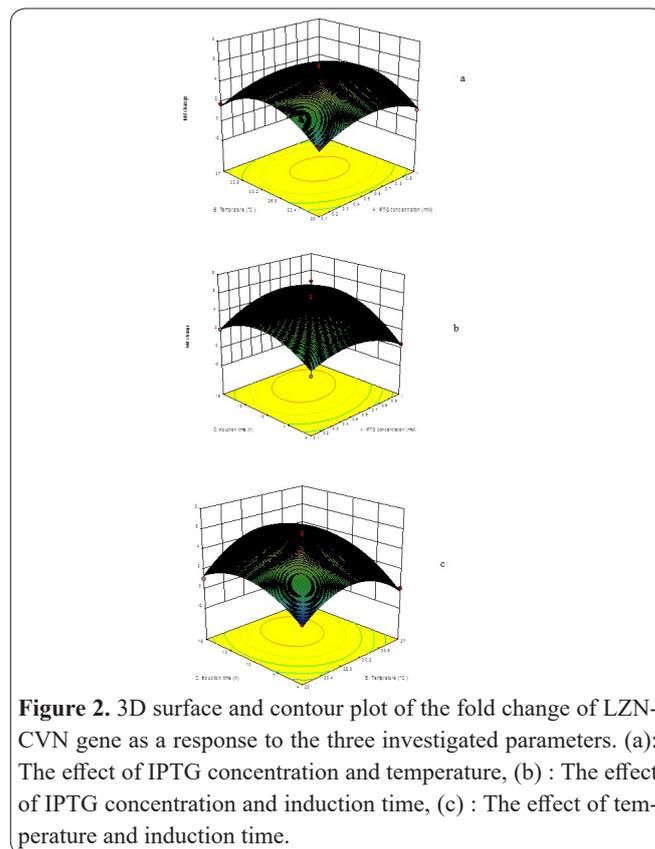
$$\text{Eq.(A2): } Y = 5.348 + (1.1375X_3) + (0.7475X_1X_3) - (1.48025X_1^2) - (2.39525X_2^2) - (1.86025X_3^2)$$

It can be concluded from the ANOVA results that fold change was significantly ( $p < 0.001$ ) affected by the linear terms of induction time rather than the other factors, as well as a quadratic effect of three terms ( $p < 0.001$ ) (Table 3). Table 3 also demonstrates the significant contribution of interaction term and the interaction effect of IPTG concentration-induction time ( $p < 0.01$ ) for fold change.

The factor combination effects on the response were determined via generation of response surface

and contour plots for models. Figure 3 shows the fold change as a function of IPTG concentration, temperature and induction time. The Minimum fold change (0.07) was observed at IPTG concentration = 0.1mM , temperature = 37°C and induction time = 4h, while, the maximum fold change was found at IPTG concentration = 0.55mM , temperature = 28.5°C and induction time = 10h. It can be indicated that the low level of IPTG and short induction time together decreased the fold change due to the low level of themRNA produced during the growth. (Figure. 2 a, b, c).

The optimum condition for processing the maximum fold change was determined based on the selected



**Figure 2.** 3D surface and contour plot of the fold change of LZN-CVN gene as a response to the three investigated parameters. (a): The effect of IPTG concentration and temperature, (b) : The effect of IPTG concentration and induction time, (c) : The effect of temperature and induction time.

**Table 3.** ANOVA results listed, showing the variables as linear, quadratic and interaction terms on each response variable and coefficients for the prediction model.

Source of variation	Fold change			
	DF	Estimated coefficients	Sum of squares	F-value
Model	9		68.34	34.19***
$\beta_0$	-	5.348	-	-
$\beta_1$ (IPTG concentration)	1	0.2575	0.53	2.38 <sup>ns</sup>
$\beta_2$ (Temperature)	1	0.32	0.81	3.68 <sup>ns</sup>
$\beta_3$ (Time after induction)	1	1.1375	10.35	46.61***
$\beta_{12}$	1	-0.0175	0.001	0.005 <sup>ns</sup>
$\beta_{13}$	1	0.7475	2.23	10.06**
$\beta_{23}$	1	0.5275	1.11	5.01 <sup>ns</sup>
$\beta_1^2$	1	-1.48025	9.22	41.54***
$\beta_2^2$	1	-2.39525	24.15	108.77***
$\beta_3^2$	1	-1.86025	14.57	65.61***
Residual	7		1.55	-
Lack of Fit	3		1.27	6.01 <sup>ns</sup>
Pure Error	4		0.28	-
Cor. Total	16		69.90	-
Correlation coefficient ( $R^2$ )	-		0.977	-

\* Significant at  $p < 0.05$ . \*\* Significant at  $p < 0.01$ . \*\*\* Significant at  $p < 0.001$ . <sup>ns</sup> Non-significant.

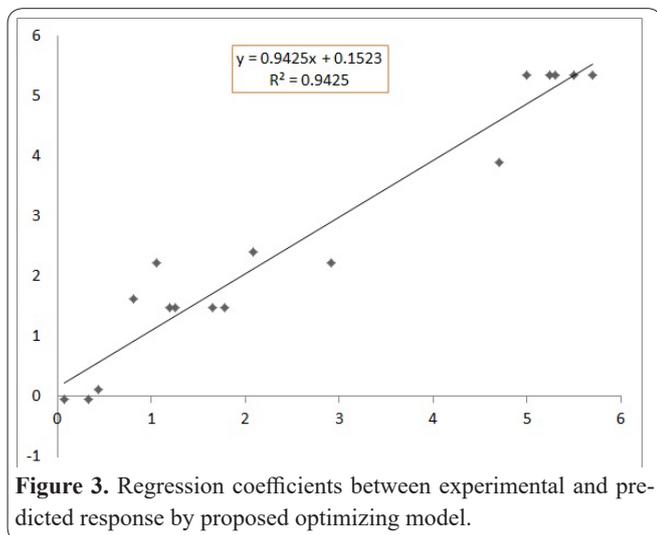


Figure 3. Regression coefficients between experimental and predicted response by proposed optimizing model.

ranges of the three parameters including 0.6mM IPTG concentration, 29°C growth temperature and 12h induction time. This condition had the maximum desirability value (0.980). A very strong correlation ( $R^2 = 0.94$ ) was found between the experimental response and the response predicted by the proposed optimizing model (Figure. 3).

#### SDS-PAGE of periplasmic fraction

The LZN-CVN gene was expressed in *E.coli* (DE3) within the final optimized condition to obtain a high level of mRNA expression, as well as a high fold change. In the periplasmic fraction, the mentioned protein with a molecular weight of 8 KDa (LZN-CVN-6His tag) was found showed an efficient expression using secretory signal peptide (pelB) (Figure. 4).

#### Structure of LZN-CVN homology protein

The LZN-CVN gene encodes a shorter protein of 72 amino acid residues (nt 1-219) compared with wild-CVN which encodes a protein containing 101aa. This short protein has a predicted isoelectric point (pI) of 4.66 and a molecular mass of 7.6684 KDa while the pI and molecular mass for wild-CVN were calculated 4.94

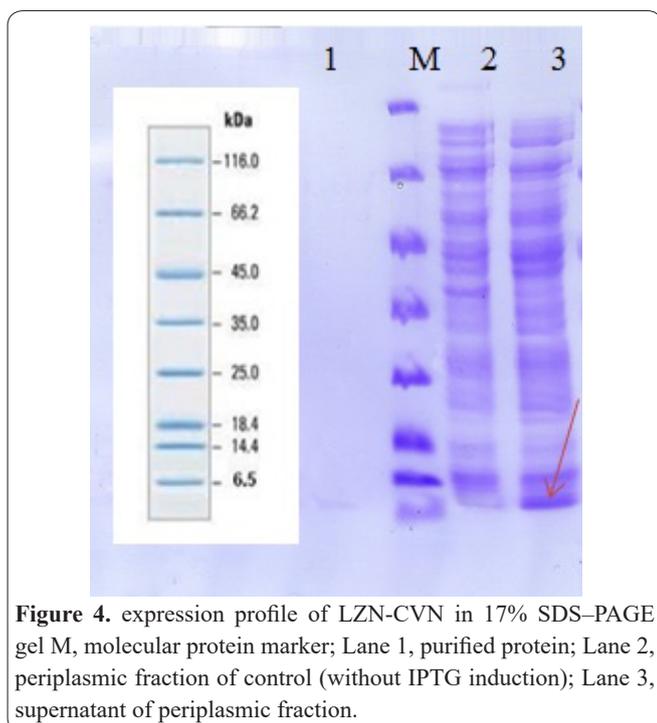


Figure 4. expression profile of LZN-CVN in 17% SDS-PAGE gel M, molecular protein marker; Lane 1, purified protein; Lane 2, periplasmic fraction of control (without IPTG induction); Lane 3, supernatant of periplasmic fraction.

and 11.013 KDa, respectively. Considering its amino acid sequence, Ser happens to be the most abundant (15.3%). The frequency of acidic and basic amino acids of this new protein was 8.33% and 5.52%, respectively. The N-terminal of the sequence examined is Leu and the estimated half-life is 5.5 hours (mammalian reticulocytes, in vitro), 3 min (yeast, in vivo) and 2 min (*E.coli*, in vivo). The instability index (II) was computed to be 26.25 which classifies the protein as stable.

Considering the results of I-TASSER, predicted secondary structure server, the LZN-CVN protein consists of 4.16% helices (H), 44.44% strands and 51.38% coils. The higher the score gets, the more confident the prediction becomes. In I-TASSER, the B-factor value was deduced from the PDB as the template proteins in combination with the sequence profile were obtained from sequence databases. B-factor was defined as a value to show the degree of the inherent thermal mobility of protein residues/atoms. Modeling of the query sequence was started in I-TASSER using the structure templates identified by LOMETS (meta-server threading approach) in PDB library (36). The 10 best templates with the highest significance (A Normalized Z-score of 1 and greater, suggests good alignment and vice versa) in the threading programs were chosen (see supplementary file). The high percentage of LZN-CVN identity with PDB accession code was related to 2rp3A (CVN domain B mutant) (Z-score = 2.89) (37). In the 2rp3A the binding site on domain B was eliminated and the protein was completely inactive against HIV and able to bind to Man-3 and Man-9 with regard to their similar dissociation constants. The remaining nine identical PDB accession codes were related to various mutants of wild-CVN including 1N02 (38), 2PYS (39), 3LHC (40), 2RP3 (37) and 2L2F (in *Gibberella zeae* CVNH) (41). The top five final structural conformation models were generated by I-TASSER with the help of SPICKER program in which model 1 was depicted as the best model (Figure 5). Overall, models with a C-score higher than  $-1.5$  had an accurate folding. The C-score for predicted model of LZN-CVN was calculated 0.52 (TM-score =  $0.78 \pm 0.09$  and Estimated RMSD =  $2.4 \pm 1.8 \text{ \AA}$ ). Following the assembly of the first structure simulation, the TM-align structural alignment program was used to match this model to all protein structures from PDB library. As a result, the top 10 closely relevant proteins from the PDB library were reported (see supplementary file). Because of structure similarity, these ten proteins

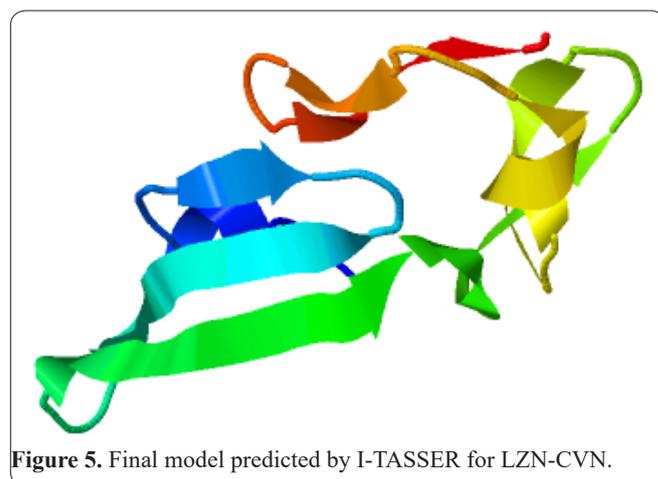


Figure 5. Final model predicted by I-TASSER for LZN-CVN.

usually possess similar function to the specific target. Then, the finding were used in the predicted function section by COACH (Meta-server approach / outcome is a combination of biological functions that were annotated from three programs include COFACTOR, TM-SITE and S-SITE) to define the function of the protein of our interest. Similarly, a significant protein structurally close to the LZN-CVN in the PDB is 2PY5B. This mutation of CVN showed that the presence of intact domain A and B as oligomannoside-binding sites which are necessary to construct domain-swapped dimers, is critical for anti-HIV activity (39). LZN-CVN have similarity to 5K79 which was recently identified in the cyanobacterium *Cyanothece7424*. The assay of its anti-HIV activity showed that the inhibition potency is approximately four fold more than CV-NP51G which makes it a valuable microbicidal lectin (42). 2L9Y is another CVNH-LysM lectin which showed similarity to LZN-CVN and carbohydrate specificities for its two domains were discovered (43). 2JZK (*Tuber borchii* CVNH) (44), 2L2F (*Gibberella zeae* CVNH) (41) and 2Y1S (Microvirin lectin) (45) are listed as other similar proteins to LZN-CVN. The predicted function for LZN-CVN suggested that ligand binding sites with the highest confidence were related to N42, V43, D44, G45, S52, N53 and E56 residues. A possible ligand name is sucrose and the predicted structure is 3HP8 (Crystal structure of a designed cyanovirin-N homolog lectin; LKAMG which binds to sucrose) (46). Associated gene ontology (GO) terms of LZN-CVN were 0030246 (carbohydrate binding), 0050688 (regulation of defense response to virus) and 0030246 (carbohydrate binding) with measurement of similarity to 2PY5 (Mutant of CVN) (39).

## Discussion

In this study, it can be inferred that five nucleotide mutations were found in CVN gene which had been isolated from *N. ellipsosporum* LZN in indigenous strain originated from Babol, Iran. The mutations were identified when the sequencing and blast analysis results were compared with that of the wild-CVN gene. It was found that CVN gene length had been changed from 303bp to 306 bp. The obtained results indicated that four insertions could change the open frame reading of the CVN protein. As a main finding, because of the insertion mutations, a stop codon had occurred and the coding protein was shorter than wild-CVN protein (72 aa vs. 101aa). Previously, CVN had been expressed in different heterologous systems (bacteria, yeasts and transgenic plants) and its anti-HIV activity was assayed. Nevertheless, the production of various recombinant proteins in *E.coli*, was frequently referred to as a low cost heterologous expression system with high amount of recombinant protein production. In addition, *E.coli* managed to grow rapidly with the high cell densities on cheap substrates, its genetics characters traits were well-determined and several cloning vectors are now available, especially genetically engineered mutant strains that are considered possible hosts for producing different proteins (47).

In the present study, LZN-CVN gene was cloned into pET22 which had a signal peptide (pelB (pectate lyase B) to secrete the expressed protein into the periplasmic

space, mainly aimed to form disulfide bonds and correct foldings. Several remarkable advantages of secretory production of r-proteins were confirmed in contrast to cytosolic production. For example, secreted r-protein has an intact N-terminal which is the same as that of a natural gene followed by cleavage of specific signal peptide. Also, the protease activity in the periplasmic space is much less than that of the cytoplasmic space. So, purification of r-protein in the periplasm is simple because of the lower contamination in the produced protein. Another advantage of periplasmic expression was the ability to form correct disulfide bond when provided with suitable oxidative conditions (48).

various signal sequences were used to increase the efficiency of secretory r-protein production in the *E.coli* system such as PelB, endoxylanase, outer membrane protein A (OmpA), maltose-binding protein (MalE), alkaline phosphatase (PhoA), murein lipoprotein (Lpp) and heat-stable enterotoxin 2(StII) (49). Despite the forementioned advantage of the secretory production of r-proteins in *E.coli*, some limitations were still present which made it more challenging. Some of the limitations were associated with the incomplete processing of secretory signal peptides, changes in the secretion efficiency of r-protein due to their different characteristics, low level or undetectable secreted protein, inclusion bodies formation after using strong promoters in the cytosol and periplasm spaces, and also incorrectly formed disulfide bonds (50). It was confirmed that anti-HIV activity and stability of CVN required proper formation of disulfide bond (1). Also, in LZN-CVN amino acid sequence, four cysteine residues were present and formation of the same disulfide bond would be accrued which was necessary for investigating its anti-HIV in the future studies. In this study, LZN-CVN was incorporated along with pelB to localize the high amounts of the expressed protein in the periplasm space and to ensure the correct formation of disulfide bonds. Mori et al. reported that replacement of ompA with pelB signal peptides could increase the CVN expression yield in the periplasmic space (51). Then, the increased yield of CVN was evaluated via forming an inclusion body in the *E.coli* cytoplasm (52). The yield of periplasmic expression was a single and homogeneous CVN, while, the yield of expression type as an inclusion body was heterogeneous CVN, full-length monomeric, dimeric with the N-terminal residue deleted indicating the weakness of this type of expression. Expression of CVN was investigated by various vectors in *E.coli*, but the low amount of the final protein and the formation of inclusion bodies were unacceptable; so the attentions were focused on chaperone-fusion expression methods. In this system, a soluble form of CVN was produced via application of 6-histidine and small ubiquitin-related modifier (SUMO) tags in the cytoplasm of *E. coli*. Similarly, rapid purification of intact soluble CVN was done using two downstream approaches of affinity chromatography (6-His tag) and cleavage by SUMO-protease (53). Our results confirmed the successful purification above by 6-His tag which was observed in the C-terminal of the expressed LZN-CVN. Barrientos et al. showed the identical anti-HIV activity of CVN in both assayed proteins, with or without C-terminal His-tag, and also demonstrated that the added residues had no

influence on the thermodynamic parameters (54).

Expression of r-proteins in *E.coli* is affected by various parameters including the promoter of vector, induction time, temperature and inducer concentration (55). In previous studies, in order to produce high amounts of r-protein different IPTG concentrations, , were used as an inducer for the promoter of pET system (56-58). Like IPTG, lactose is an inducer too, which could promote cell growth as a carbon source in the synthesis process of r-protein (14, 16). Unfortunately, none or little reports compared the expression amount of CVN or an r-protein with lactose and IPTG or combination of them as in inducer.

In addition to selecting expression host and the corresponding vectors or purification processes by cloning design, application an experimental design techniques seems to be necessary. The design methods mentioned could recognize the effect of the main parameters during protein production and their interactions to predict the best production, increasing the final protein expression level and decreasing the manufacturing costs for bioreactor use. RSM could apply a series of mathematical and statistical methods for modeling and analysis of the responses that were influenced by various variables of interest (59). The main advantage of RSM is the reduction of the number of experiments conducted to find statistically reliable results. RSM is commonly used for response optimization and selecting acceptable operating conditions to obtain a set of target specifications (27, 28, 60).

In the present study, RSM was performed and the validation of parameters (IPTG concentration, temperature and time induction) was done for protein production conditions. The obtained results led us to have LZN-CVN protein produced in LB medium at 29 °C for 12 h with 0.6 mM IPTG, as a feasible process in optimized expression conditions. It is proved that a higher level of r-protein was produced when the IPTG concentration was reduced (56, 61). In our results, low fold change was related to high levels of IPTG (run 11, 12 and 9). But in the run 6, high levels of IPTG concentration was applied and compared with the other runs, a high level of fold change was achieved in a different induction time and temperature. As a result, suitable inducer concentration were not specific for producing all r-proteins, hence optimization of expression parameters were suggested for each experiment to obtain the high level and best concentration of r-protein.

*E.coli* is a common host for protein expression in the biotechnology process, through which controlling of strong promoters (such as lac) could be regulated by IPTG induction at different levels (62). Finding the optimum concentration of IPTG is an essential step, because a high concentration results in plasmid segregation and severe reduction of cell growth rate (especially during the production of toxic proteins) or some proteases and soluble forms. Meanwhile slow induction with low level may actually increase the solubility and activity of proteins (63-66). However, low IPTG concentrations may fail to deliver an effective full induction (67). In molecular biology the standard concentration of IPTG is 1mM (68) and in various studies 0.1 mM was considered as a low inducer concentration (66, 69). Reducing the concentration of IPTG is also appealing since it is an

expensive compound and is potentially toxic (70, 71). In our study, the optimum concentration of IPTG was 0.6mM. Besides the concentration, induction temperature and time are also two main variables that affect the final cell growth paving the way for achieving high cell concentration and the accumulation of recombinant protein (72, 73). A higher cell growth potentially causes a higher recombinant protein synthesis and formation of inclusion bodies (IBs) (74). The optimum temperature for growing *E.coli* is 37°C, but in this induction temperature, the expressed protein could be accumulated as IBs, while a 7°C decline , would lead to the production of soluble and active protein at 30°C. Also, long induction time (24h) at lower than 37°C - about 15-20°C - could provide an optimal condition for the production of soluble protein (75). Also, further isolation, purification, and denaturation processes are necessary to recover the proteins from the IB form, but in this condition the final soluble yield is low and the activity of protein may be altered. It was proposed that the protein expression rate as a soluble form could be reduced in low temperature but formation of IBs did not occur(76, 77). Similarly, soluble production will be achieved in optimized cell growth phases. Cell growth was restricted during the protein expression due mostly to energy expenditure for other cell process as well. In this case, the desirable induction time is in the exponential phase (especially in the middle of that phase) of cell growth not stationary phase since the metabolic activity is complete and also growth rate is rising (78). In our study, the optimum induction temperature and time was calculated about 29°C and 12h.

In addition to CVN, a family of cyanovirin-N homology (CVNH) has recently been recognized. The new researches focused on the properties of CVNHs include: structural information analysis, spectrum of antiviral activity, carbohydrate-binding specificities and prediction of structure-function interactions (44). CVNHs were identified in various organisms such as, filamentous ascomycetes, seedless plants (*Ceratopteris thalictroides*, *Ceratopteris richardii*), *Tuber borchii* Vitad., and *Neurospora crassa* (44, 79, 80). These findings elucidated that CVNHs had common fold structure in compared to CVN and especially the anti-HIV domain was conserved in the various organs and their anti-HIV potential could be sustained during evolution.

This paper illustrates the novel evolutionary mutated gene of CVN with four insertions and one deletion of nucleotide sequences. This sequence leads to production of a shorter protein (named LZN-CVN) which is different from that of wild-CVN. The BBD method and RSM successfully optimized the production of LZN-CVN in *E. coli* through modifying the three variables, IPTG concentrations, temperature, and induction time in 17 experimental runs. As with the main result, it was found that LZN-CVN production in *E. coli* was influenced significantly by the interaction effect of IPTG concentration and induction time. The obtainable LZN-CVN here is an important induction time protein to assay its functionality in the future experiments.

An on-line server of I-TASSER established at the KU Center for bioinformatics was used. The obtained results suggested that four insertions could change the open frame reading of the CVN protein. As a main fin-

ding, because of insertion mutations a stop codon had occurred resulting in expression of a protein shorter than the CVN reference protein (72 aa vs. 101aa). Differences in the 72 amino acid sequence of the wild-CVN are related to A71 and E72 positions which had been replaced with G71 and C72 in LZN-CVN, respectively. Furthermore, it was suggested to function as a carbohydrate binding agent and its similarity to CVN homology proteins from PDB library was clarified. These differences in LZN-CVN gene lead to production of new protein in which the foldings of the LZN-CVN might be changed. This feature may also give it a different affinity for mannose disaccharide ligands on gp120 HIV. Our results provide a deeper insight to LZN-CVN function as anti-HIV candidate.

### Acknowledgments

This is a report of database from Ph.D thesis of H.Lotfi registered in Drug Applied Research Center, Tabriz University of Medical Sciences (Grant No: 93.113, Thesis No : 93.4-9.6).

### Competing Interests

The authors declare that they have no conflict of interest.

### References

- Gustafson KR, Sowder RC, 2nd, Henderson LE, Cardellina JH, 2nd, McMahon JB, Rajamani U, et al. Isolation, primary sequence determination, and disulfide bond structure of cyanovirin-N, an anti-HIV (human immunodeficiency virus) protein from the cyanobacterium *Nostoc ellipsosporum*. *Biochem Biophys Res Commun* 1997;238:223-8.
- Barrientos LG, O'Keefe BR, Bray M, Sanchez A, Gronenborn AM, Boyd MR. Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. *Antiviral Res* 2003;58:47-56.
- O'Keefe BR, Smee DF, Turpin JA, Saucedo CJ, Gustafson KR, Mori T, et al. Potent anti-influenza activity of cyanovirin-N and interactions with viral hemagglutinin. *Antimicrob Agents Chemother* 2003;47:2518-25.
- Boyd MR, Gustafson KR, McMahon JB, Shoemaker RH, O'Keefe BR, Mori T, et al. Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob Agents Chemother* 1997;41:1521-30.
- Dey B, Lerner DL, Lusso P, Boyd MR, Elder JH, Berger EA. Multiple antiviral activities of cyanovirin-N: blocking of human immunodeficiency virus type 1 gp120 interaction with CD4 and coreceptor and inhibition of diverse enveloped viruses. *J Virol* 2000;74:4562-9.
- Buffa V, Stieh D, Mamhood N, Hu Q, Fletcher P, Shattock RJ. Cyanovirin-N potently inhibits human immunodeficiency virus type 1 infection in cellular and cervical explant models. *J Gen Virol* 2009;90:234-43.
- Balzarini J, Van Damme L. Microbicide drug candidates to prevent HIV infection. *Lancet* 2007;369:787-97.
- Helle F, Wychowski C, Vu-Dac N, Gustafson KR, Voisset C, Dubuisson J. Cyanovirin-N inhibits hepatitis C virus entry by binding to envelope protein glycans. *J Biol Chem* 2006;281:25177-83.
- Barrientos LG, Matei E, Lasala F, Delgado R, Gronenborn AM. Dissecting carbohydrate-Cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral entry inhibition. *PEDS* 2006;19:525-35.

- Mori T, Boyd MR. Cyanovirin-N, a potent human immunodeficiency virus-inactivating protein, blocks both CD4-dependent and CD4-independent binding of soluble gp120 (sgp120) to target cells, inhibits sCD4-induced binding of sgp120 to cell-associated CXCR4, and dissociates bound sgp120 from target cells. *Antimicrob Agents Chemother* 2001;45:664-72.
- Fischetti L, Barry SM, Hope TJ, Shattock RJ. HIV-1 infection of human penile explant tissue and protection by candidate microbicides. *AIDS* 2009;23:319.
- Boes A, Reimann A, Twyman RM, Fischer R, Schillberg S, Spiegel H. A Plant-Based Transient Expression System for the Rapid Production of Malaria Vaccine Candidates. *Methods Mol Biol* 2016:597-619.
- Sullivan CJ, Pendleton ED, Sasmor HH, Hicks WL, Farnum JB, Muto M, et al. A cell-free expression and purification process for rapid production of protein biologics. *Biotechnol J* 2016;11:238-48.
- Boyd MR, Gustafson KR, McMahon JB, Shoemaker RH, O'Keefe BR, Mori T, et al. Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob Agents Chemother* 1997;41:1521-30.
- Colleluori DM, Tien D, Kang F, Pagliei T, Kuss R, McCormick T, et al. Expression, purification, and characterization of recombinant cyanovirin-N for vaginal anti-HIV microbicide development. *Protein Expr Purif* 2005;39(2):229-36.
- Gao X, Chen W, Guo C, Qian C, Liu G, Ge F, et al. Soluble cytoplasmic expression, rapid purification, and characterization of cyanovirin-N as a His-SUMO fusion. *Appl Microbiol Biotechnol* 2010;85:1051-60.
- Giomarelli B, Provvedi R, Meacci F, Maggi T, Medaglini D, Pozzi G, et al. The microbicide cyanovirin-N expressed on the surface of commensal bacterium *Streptococcus gordonii* captures HIV-1. *Aids* 2002;16:1351-6.
- Liu X, Lagenaur LA, Simpson DA, Essenmacher KP, Frazier-Parker CL, Liu Y, et al. Engineered vaginal lactobacillus strain for mucosal delivery of the human immunodeficiency virus inhibitor cyanovirin-N. *Antimicrob Agents Chemother* 2006;50:3250-9.
- Ndesendo VM, Pillay V, Choonara YE, Buchmann E, Bayever DN, Meyer LC. A review of current intravaginal drug delivery approaches employed for the prophylaxis of HIV/AIDS and prevention of sexually transmitted infections. *AAPS PharmSciTech* 2008;9:505-20.
- Mori T, Barrientos LG, Han Z, Gronenborn AM, Turpin JA, Boyd MR. Functional homologs of cyanovirin-N amenable to mass production in prokaryotic and eukaryotic hosts. *Protein Expr Purif* 2002;26:42-9.
- Colgan R, Atkinson CJ, Paul M, Hassan S, Drake PM, Sexton AL, et al. Optimisation of contained *Nicotiana tabacum* cultivation for the production of recombinant protein pharmaceuticals. *Transgenic Res* 2010;19:241-56.
- Drake PM, Barbi T, Sexton A, McGowan E, Stadlmann J, Navarre C, et al. Development of rhizosecretion as a production system for recombinant proteins from hydroponic cultivated tobacco. *FASEB J* 2009;23:3581-9.
- Drake PM, de Moraes Madeira L, Szeto TH, Ma JK. Transformation of *Althaea officinalis* L. by *Agrobacterium rhizogenes* for the production of transgenic roots expressing the anti-HIV microbicide cyanovirin-N. *Transgenic Res* 2013;22:1225-9.
- O'Keefe BR, Murad AM, Vianna GR, Ramessar K, Saucedo CJ, Wilson J, et al. Engineering soya bean seeds as a scalable platform to produce cyanovirin-N, a non-ARV microbicide against HIV. *Plant Biotechnol J* 2015;13:884-92.
- Vamvaka E, Evans A, Ramessar K, Krumpel L, Shattock R, O'Keefe B, et al. Cyanovirin-N produced in rice endosperm offers

- effective pre-exposure prophylaxis against HIV-1BaL infection in vitro. *Plant Cell Rep* 2016;35:1309-19.
26. Bewley CA, Gustafson KR, Boyd MR, Covell DG, Bax A, Clore GM, et al. Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nat Struct Mol Biol* 1998;5:571-8.
27. Zhang Y, Wei X, Lu Z, Pan Z, Gou X, Venkitasamy C, et al. Optimization of culturing conditions of recombinant *Escherichia coli* to produce umami octopeptide-containing protein. *Food Chem* 2017;227:78-84.
28. Patil MD, Shinde KD, Patel G, Chisti Y, Banerjee UC. Use of response surface method for maximizing the production of arginine deiminase by *Pseudomonas putida*. *Biotechnol Rep* 2016;10:29-37.
29. Mohajeri A, Pilehvar-Soltanahmadi Y, Abdolalizadeh J, Karimi P, Zarghami N. Effect of Culture Condition Variables on Human Endostatin Gene Expression in *Escherichia coli* Using Response Surface Methodology. *Jundishapur J Microbiol* 2016;9(8).
30. Li R-F, Wang B, Liu S, Chen S-H, Yu G-H, Yang S-Y, et al. Optimization of the Expression Conditions of CGA-N46 in *Bacillus subtilis* DB1342 (p-3N46) by Response Surface Methodology. *Interdiscip Sci* 2016;8:277-83.
31. Bhattacharjee K, Joshi SR. A selective medium for recovery and enumeration of endolithic bacteria. *J Microbiol Methods* 2016;129:44-54.
32. Su Y, Wang T, Yang W, Huang C, Fan G. DNA extraction and RAPD analysis of *Podocarpus*. *Acta Sci. Nat. Univ. Sunyatseni* 1997;37:11-8.
33. French C, Keshavarz-Moore E, Ward JM. Development of a simple method for the recovery of recombinant proteins from the *Escherichia coli* periplasm. *Enzyme Microb Technol* 1996;19:332-8.
34. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970;227:680-5.
35. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nature methods* 2015;12(1):7-8.
36. Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res* 2007;35:3375-82.
37. Matei E, Furey W, Gronenborn AM. Solution and crystal structures of a sugar binding site mutant of cyanovirin-N: no evidence of domain swapping. *Structure* 2008;16:1183-94.
38. Barrientos LG, Louis JM, Ratner DM, Seeberger PH, Gronenborn AM. Solution structure of a circular-permuted variant of the potent HIV-inactivating protein cyanovirin-N: structural basis for protein stability and oligosaccharide interaction. *J Mol Biol* 2003;325:211-23.
39. Fromme R, Katiliene Z, Giomarelli B, Bogani F, Mc Mahon J, Mori T, et al. A monovalent mutant of cyanovirin-N provides insight into the role of multiple interactions with gp120 for antiviral activity. *Biochemistry* 2007;46:9199-207.
40. Matei E, Zheng A, Furey W, Rose J, Aiken C, Gronenborn AM. Anti-HIV activity of defective cyanovirin-N mutants is restored by dimerization. *J Biol Chem* 2010;285:13057-65.
41. Matei E, Louis JM, Jee J, Gronenborn AM. NMR solution structure of a cyanovirin homolog from wheat head blight fungus. *Proteins* 2011;79:1538-49.
42. Matei E, Basu R, Furey W, Shi J, Calnan C, Aiken C, et al. Structure and Glycan Binding of a New Cyanovirin-N Homolog. *J Biol Chem* 2016.
43. Koharudin LM, Viscomi AR, Montanini B, Kershaw MJ, Talbot NJ, Ottonello S, et al. Structure-function analysis of a CVNH-LysM lectin expressed during plant infection by the rice blast fungus *Magnaporthe oryzae*. *Structure* 2011;19:662-74.
44. Koharudin LM, Viscomi AR, Jee JG, Ottonello S, Gronenborn AM. The evolutionarily conserved family of cyanovirin-N homologs: structures and carbohydrate specificity. *Structure* 2008;16:570-84.
45. Shahzad-ul-Hussan S, Gustchina E, Ghirlando R, Clore GM, Bewley CA. Solution structure of the monovalent lectin microvirin in complex with Man(alpha)(1-2)Man provides a basis for anti-HIV activity with low toxicity. *J Biol Chem* 2011;286:20788-96.
46. Koharudin LM, Furey W, Gronenborn AM. A designed chimeric cyanovirin-N homolog lectin: structure and molecular basis of sucrose binding. *Proteins* 2009;77:904-15.
47. Choi JH, Keum KC, Lee SY. Production of recombinant proteins by high cell density culture of *Escherichia coli*. *Chemical Engineering Science* 2006;61:876-85.
48. Makrides SC. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev* 1996;60:512-38.
49. Choi JH, Lee SY. Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl Microbiol Biotechnol* 2004;64:625-35.
50. Jeong KJ, Lee SY. Excretion of Human  $\beta$ -Endorphin into Culture Medium by Using Outer Membrane Protein F as a Fusion Partner in Recombinant *Escherichia coli*. *Appl Environ Microbiol* 2002;68:4979-85.
51. Bewley CA. Solution structure of a cyanovirin-N:Man alpha 1-2Man alpha complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure* 2001;9:931-40.
52. Colleluori DM, Tien D, Kang F, Pagliei T, Kuss R, McCormick T, et al. Expression, purification, and characterization of recombinant cyanovirin-N for vaginal anti-HIV microbicide development. *Protein Expr Purif* 2005;39:229-36.
53. Gao X, Chen W, Guo C, Qian C, Liu G, Ge F, et al. Soluble cytoplasmic expression, rapid purification, and characterization of cyanovirin-N as a His-SUMO fusion. *Appl Microbiol Biotechnol* 2010;85:1051-60.
54. Barrientos LG, Louis JM, Botos I, Mori T, Han Z, O'Keefe BR, et al. The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure* 2002;10:673-86.
55. Correa A, Oppezzo P. Tuning different expression parameters to achieve soluble recombinant proteins in *E. coli*: advantages of high-throughput screening. *Biotechnol J* 2011;6:715-30.
56. Cao W, Li H, Zhang J, Li D, Acheampong DO, Chen Z, et al. Periplasmic expression optimization of VEGFR2 D3 adopting response surface methodology: antiangiogenic activity study. *Protein Expr Purif* 2013;90:55-66.
57. Collins T, Azevedo-Silva J, da Costa A, Branca F, Machado R, Casal M. Batch production of a silk-elastin-like protein in *E. coli* BL21(DE3): key parameters for optimisation. *Microb Cell Fact* 2013;12:21.
58. Papanepoytou CP, Rinotas V, Douni E, Kontopidis G. A statistical approach for optimization of RANKL overexpression in *Escherichia coli*: purification and characterization of the protein. *Protein Expr Purif* 2013;90:9-19.
59. Gunst RF. Response surface methodology: process and product optimization using designed experiments. Taylor & Francis Group 1996.
60. Gholami Tilko P, Hajihassan Z, Moghimi H. Optimization of recombinant  $\beta$ -NGF expression in *Escherichia coli* using response surface methodology. *Prep Biochem Biotechnol* 2016:1-8.
61. Larentis AL, Argondizzo AP, Esteves Gdos S, Jessouron E, Galler R, Medeiros MA. Cloning and optimization of induction conditions for mature PsaA (pneumococcal surface adhesin A) expression in *Escherichia coli* and recombinant protein stability during long-term storage. *Protein Expr Purif* 2011;78:38-47.
62. De León A, Jiménez-Islas H, González-Cuevas Ma, de la Rosa APB. Analysis of the expression of the *Trichoderma harzianum* ech42 gene in two isogenic clones of *Escherichia coli* by surface

- response methodology. *Process Biochem* 2004;39:2173-8.
63. Manderson D, Dempster R, Chisti Y. A recombinant vaccine against hydatidosis: production of the antigen in *Escherichia coli*. *J Ind Microbiol Biotechnol* 2006;33:173-82.
64. Islam RS, Tisi D, Levy MS, Lye GJ. Framework for the rapid optimization of soluble protein expression in *Escherichia coli* combining microscale experiments and statistical experimental design. *Biotechnol Prog* 2007;23:785-93.
65. Beigi L, Karbalaee-Heidari HR, Kharrati-Kopaei M. Optimization of an extracellular zinc-metalloprotease (SVP2) expression in *Escherichia coli* BL21 (DE3) using response surface methodology. *Protein Expr Purif* 2012;84:161-6.
66. Einsfeldt K, Severo Júnior JB, Corrêa Argondizzo AP, Medeiros MA, Alves TLM, Almeida RV, et al. Cloning and expression of protease ClpP from *Streptococcus pneumoniae* in *Escherichia coli*: Study of the influence of kanamycin and IPTG concentration on cell growth, recombinant protein production and plasmid stability. *Vaccine* 2011;29:7136-43.
67. Urban A, Ansmant I, Motorin Y. Optimisation of expression and purification of the recombinant Yol066 (Rib2) protein from *Saccharomyces cerevisiae*. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003;786:187-95.
68. Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*: Cold spring harbor laboratory press 1989.
69. Shafiee F, Rabbani M, Jahanian-Najafabadi A. Optimization of the Expression of DT386-BR2 Fusion Protein in *Escherichia coli* using Response Surface Methodology. *Adv Biomed Res* 2017;6:22.
70. Volonte F, Marinelli F, Gastaldo L, Sacchi S, Pilone MS, Pollegioni L, et al. Optimization of glutaryl-7-aminocephalosporanic acid acylase expression in *E. coli*. *Protein Expr Purif* 2008;61:131-7.
71. Pan H, Xie Z, Bao W, Zhang J. Optimization of culture conditions to enhance cis-epoxysuccinate hydrolase production in *Escherichia coli* by response surface methodology. *Biochem Eng J* 2008;42:133-8.
72. Donovan RS, Robinson CW, Glick BR. Optimizing the expression of a monoclonal antibody fragment under the transcriptional control of the *Escherichia coli* lac promoter. *Can J Microbiol* 2000;46:532-41.
73. Ran H, Wu J, Wu D, Duan X. Enhanced Production of Recombinant *Thermobifida fusca* Isoamylase in *Escherichia coli* MDS42. *Appl Biochem Biotechnol* 2016;180:464-76.
74. Khan MA, Sadaf S, Sajjad M, Akhtar MW. Production enhancement and refolding of caprine growth hormone expressed in *Escherichia coli*. *Protein Expr Purif* 2009;68:85-9.
75. Schein CH. Production of soluble recombinant proteins in bacteria. *Nature Biotechnology* 1989;7:1141-9.
76. Chen Y, Xing X-H, Ye F, Kuang Y, Luo M. Production of MBP-HepA fusion protein in recombinant *Escherichia coli* by optimization of culture medium. *Biochem Eng J* 2007;34:114-21.
77. Volontè F, Marinelli F, Gastaldo L, Sacchi S, Pilone MS, Pollegioni L, et al. Optimization of glutaryl-7-aminocephalosporanic acid acylase expression in *E. coli*. *Protein Expr Purif* 2008;61:131-7.
78. Donovan RS, Robinson CW, Glick B. Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. *J Ind Microbiol* 1996;16:145-54.
79. Sun J, Su Y, Wang T. Expression, purification and identification of CtCVNH, a novel anti-HIV (Human Immunodeficiency Virus) protein from *Ceratopteris thalictroides*. *Int J Mol Sci* 2013;14:7506-14.
80. Koharudin LMI, Viscomi AR, Jee J-G, Ottonello S, Gronenborn AM. The Evolutionarily Conserved Family of Cyanovirin-N Homologs: Structures and Carbohydrate Specificity. *Structure* 2008;16:570-84.