Biomacromol. J., Vol. 6, No. 1, 17-24, July 2020.

Biomacromolecular Journal
www.bmmj.org

Overexpression and Purification of the Synthetic Human Proinsulin to Efficient Produce Glargine Insulin in *E. coli*

L. Feizi[a], A. Saeedinia[a,]*, M. Zeinoddini[a] and R. Gheitasi[b]

[a]*Faculty of Chemistry and Chemical Engineering, Malek Ashtar University of Technology, Iran*

[b]*Department of Immunology, Student Research Committee, School of Medicine, Hamadan University of Medical Sciences, Hamadan, Iran*

(Received 1 September 2020, Accepted 28 December 2020)

ABSTRACT

Glargine insulin, as a recombinant human insulin analog, is the first long-acting insulin analog. In glargine insulin structure, the asparagine amino acid is replaced with glycine amino acid at the end of the carboxyl chain A, and also two arginine amino acids are added to the end of the carboxyl chain B. The aim of this study is to design, clone and express the human modified proinsulin analog in Escherichia coli and purify it to produce glargine insulin. Initially, the nucleotide sequence of human proinsulin analog was designed based on the *E. coli* codon usage and Gene Bank data, and then syntactically constructed and cloned into the pBHA vector. Then, the modified proinsulin fragment was sub-cloned into the expression vector pET-21b(+) and the recombinant vector pET21b-proInsG was transformed into the *E. coli* BL21(DE3). Then, induction of expression in cells containing recombinant vector was done by IPTG. After the evaluation of expression, purification of recombinant protein was performed by using a nickel affinity chromatography and a batch system. Synthesis and cloning of modified proinsulin fragment and construction of recombinant vector were confirmed by DNA sequencing, specific PCR amplification and restriction enzyme mapping. The expression of recombinant protein was approximately 40% in mass form that was confirmed by using SDS-PAGE and western blot techniques. The purification of the recombinant modified proinsulin was successfully done with a purity of about 80%. This modified recombinant proinsulin protein can be used for efficient production of the insulin glargine analogue.

Keywords: *E. coli*, Insulin glargine, Modified proinsulin, Overexpression, Purification

INTRODUCTION

Diabetes is a metabolic disorder in the body [2]. In this disease, human body cannot produce insulin hormone, or resistance of the body to insulin does not let this hormone to have properly function [3]. Before developing biotechnology, needing insulin to treat patients with diabetes were obtained by extracting hormone from the pancreas of cattle or pigs. The first produced insulin was based on the bovine and swine insulin. Until the 1980s, semisynthetic insulin was not clinically available [4,5]. Human insulin has been slightly available since the 1960s. This insulin was extracted from the pancreas of human corpses and was used as reference material in radioimmunoassay tests or physicochemical identity testing [6,7]. The first completely chemical synthesis of human insulin was carried out in 1974 by Siber and coworkers, and biologically was quite similar to the natural hormone [8]. By developing the genetic engineering, human insulin produced by recombinant DNA technology entered the pharmaceutical market [9-13]. Human recombinant insulin for the treatment of diabetes is predominantly produced in *Escherichia coli* and *Saccharomyces cerevisiae*. Using the *E. coli* expressive system, insulin precursor is produced in inclusion body form, and biologically active polypeptides are ultimately obtained through the processes of solubility and protein refolding [14]. Yeast-based expression systems produce insulin precursors in the form of solution, this form of insulin can be removed from the surface soloution of the culture medium [15]. In addition to *E. coli* and yeast, expression systems of mammalian cells, plants, and transgenic animals were also used to produce large-scale recombinant insulin [16]. Initial recombinant insulins, with the same structural insulin, had limitations. They need

*Corresponding author. E-mail: ar_saeedi@mut.ac.ir

several times daily injections per day due to their low period of function in the body [17]. The first attempt to increase insulin function time was the use of various additives such as vasoconstrictor, but had little success. Understanding the structure of insulin made it possible to change its structure. Glargine insulin is the first long-acting insulin analog, which requires only one injection per day, and has been introduced to the medical community for several years. This insulin has been specifically developed to meet the body's basic needs for insulin. Glargine insulin has a long duration of activity and hasn't any peak area of effect, which makes it possible to be injected once a day; it also reduces the risk of nightly hypoglycemia associated with insulin-mediated effects. This insulin analog was developed by Aventis Pharmaceutical Company and was approved by the Food and Drug Administration and the European Medicines Agency in 2000. Insulin is composed of two polypeptide chains A with 21 amino acids and B with 30 amino acids and the C chain, which creates a binding between the A and B chains, is released after the proinsulin decomposition to the insulin [10]. Glargine insulin has been created by the replacing asparagine at the C-terminal of chain A with a glycine, as well as adding two arginines to the C-terminal of chain B. These changes increase the isoelectric point of insulin from 4.5 to neutral. Proinsulin as insulin precursor was produced in *E. coli* and finally formulated in solution type at pH = 4. Due to neutral pH, this insulin is deposited after subcutaneous injection in under the skin tissues. Recovery of insulin occurs slowly, resulting in a longer duration of action and sufficient time for secretion in the blood [18].

In this study, after designing and synthesis the desired sequence, the production of cytoplasmic human recombinant proinsulin performed in *E. coli* Bl21(DE3) and then the purification of the protein inclusion body form was done by the Batch method.

MATERIALS AND METHODS

Materials, Vectors and Strains

LB-Broth and LB-agar media were purchased from Srlchem Company. Ampicillin was purchased from the Roche Company. XhoI, NdeI and T4 DNA Ligase enzymes and PCR master mix were prepared from Fermentas Company. BSA, Coomassie Brilliant Blue, $NiSO_4$, Imidazole and Urea were purchased from Merck Company, and plasmid and gel extraction kits were purchased from GeneAll Company. The bacterial strains including *E. coli* DH5α and *E. coli* BL21(DE3), and pET-21b(+) plasmid were purchased from the Novagen Company.

Design and Synthesis of the Human Modified Proinsulin Encoding Sequence

The modified proinsulin encoding sequence was designed based on the optimal codecs of *E. coli* and the gene bank data for the proinsulin (NC-000011.1) and expression vector pET-21b(+) (GAP38373.1) and by using SnapGene 1.1.3 software. The synthesis, cloning in pBHA vector, and sequencing of the modified proinsulin were carried out by Eurofins MWG Company (Germany).

Construction of the Recombinant Expression Vector

To obtain the synthesized proinsulin, the vector containing the target fragment (pBHA-proInsG) was transformed into the *E. coli* DH5α competent cells, than DNA plasmid was extracted from target colony. At the next step, the proinsulin fragment was cut out from pBHA-proInsG plasmid using XhoI and NdeI. The pET-21b(+) expression vector was also digested by the same enzymes. After extraction of the pieces from gel agarose by using kit (GeneAll), the proinsulin fragment was ligated into vector, than ligation product was transformed into the *E. coli* DH5α competent cells. In order to screen the colonies containing pET21b-proInsG recombinant vector, the colony PCR reaction was performed using upstream 5′-GGTCATATGCACCATCACCACC ATCACGGATC-3′ and downstream 5′-GGACTCGAGATTAGCCGCAGTAGTTCTCCAGCT G-3′ specific primers. Following the culture of the target colonies and extraction of DNA plasmid, the restriction enzyme mapping was performed using NdeI, XhoI and PstI.

Expression of the Human Modified Proinsulin

After transformation of pET21b-proInsG recombinant vector into the *E. coli* BL21(DE3) competent cells, a colony was treated in 5 ml of LB broth containing ampicillin at a concentration of 50 μg ml^{-1} in an shaking incubator at 37 °C

with 200 g to overnight. For refreshing the culture, 5 µl of bacteria were sub-cultured in 5 ml of fresh medium under the same conditions until reach the OD = 0.7 at the 600 nm. After that, one ml of the medium was harvested, then bacteria were induced by addition 50 µl of 0.1 M IPTG with a final concentration of 1 mM. After 4 h of induction, one ml was harvested from the medium for SDS-PAGE analysis. Taken samples and also the residue of the culture medium were sedimented for 5 minutes at 9000 g and keeped in -20 °C. After SDS-PAGE analyzing and observing the proinsulin band in related sample after 4 h of induction, western blot analysis was performed to confirm the expression by using specific monoclonal antibody against the HisTag tail.

Purification of the Human Recombinant Proinsulin

First, to determine the rate of recombinant protein production in inclusion body form, the expression induction was carried out on 200 ml of culture medium. After sedimentation and washing, the cell precipitate was resuspended in 20 ml of lysis buffer containing 50 mM Tris, 8 mM EDTA, 8% sucrose, 3% Triton X100 and 1 mM PMSF with pH = 8. Subsequently, the cells were sonicated under 10-cycle, 200-wat for 7 min. In the next step, centrifugation was carried out at 3000 g for 3 min, and the supernatant was re-centrifuged at 10,000 g for 25 min. The precipitate was washed with 10 ml washing buffer containing 3 M urea three times. Finally, the final product was evaluated to ensure about the presence of recombinant protein in form of inclusion body by SDS-PAGE. In order to simultaneous purification and refolding, the inclusion bodies were dissolved with 700 µl of solution containing 10 mM Tris, 6 M Urea, 2.5 mM imidazole, 0.5 M NaCl, 10 mM 2-mercaptoethanol and 20% glycerol, at pH 7.8. Then the microtube was placed at room temperature for 60 min and to completely resolve the inclusion body, it was slowly inverted every 10 minutes. The solution containing the recombinant protein was centrifuged for 30 min at 8000 g to sediment the insoluble precipitates. Then, the clear supernatant was collected and used to purify the recombinant modified proinsulin protein. Purification of the recombinant protein was carried out using 1 ml of resin. After washing the resin with 1 ml of distilled water, 1 ml of 0.2 M $NiSO_4$ was added. After a few minutes, the supernatant was collected (at this point the resin turns blue). The collected supernatant containing recombinant protein was slowly mixed with nickeled-resin and placed at room temperature for 90 min and it was slowly inverted every 10 min. After 90 min, the supernatant were collected and washed with 1 ml of 6 to 1 M urea buffer containing 10 mM Tris, 0.5 M NaCl, 10 mM 2-mercaptoethanols and 20% glycerol, pH = 7.8, respectively, and finally with the 0 M urea solution containing 20 mM imidazole. At the next step, the recombinant proinsulin protein was purified with 100 and 200 mM imidazole buffer containing 10 mM Tris, 0.5 M NaCl, 10 mM 2-mercaptoethanols and 20% glycerol, pH = 7.8, respectively. The last two step supernatants were used for SDS-PAGE analysis.

RESULTES

Designing the Modified Proinsulin Encoding Sequence

Recombinant insulin is expressed as proinsulin and then cleaved to active insulin by enzymatic digestion with trypsin. The trypsin enzyme has the ability to cut peptide bonds at the end of positively charged carboxyl amino acids such as arginine and lysine. Since the proinsulin has several internal arginine and lysine residues, a number of impurities are generated following trypsin treatment. In this study, an asparagine residue at the carboxyl end of chain A was replaced with glycine at the position 21 (Gly^{A21}) in proinsulin sequence. Also, a lysine residue at the carboxyl end of chain C was replaced with alanine at the position 34 (Ala^{C34}). To prevent the trypsin enzyme function on a lysine residue at the position 29 of chain B (Lys^{B29}), the citraconic anhydride will be used to cover this amino acid before applying trypsin. The amino acid sequence of modified proinsulin designed in this study has been compared with the natural human proinsulin in Fig. 1.

Subcoloning the Modified Proinsulin Encoding Sequence

After designing, synthesis and cloning of modified proinsulin fragment and construction of recombinant vector pBHA-proInsG were confirmed by DNA sequencing, specific PCR amplification and restriction enzyme mapping (data not shown).

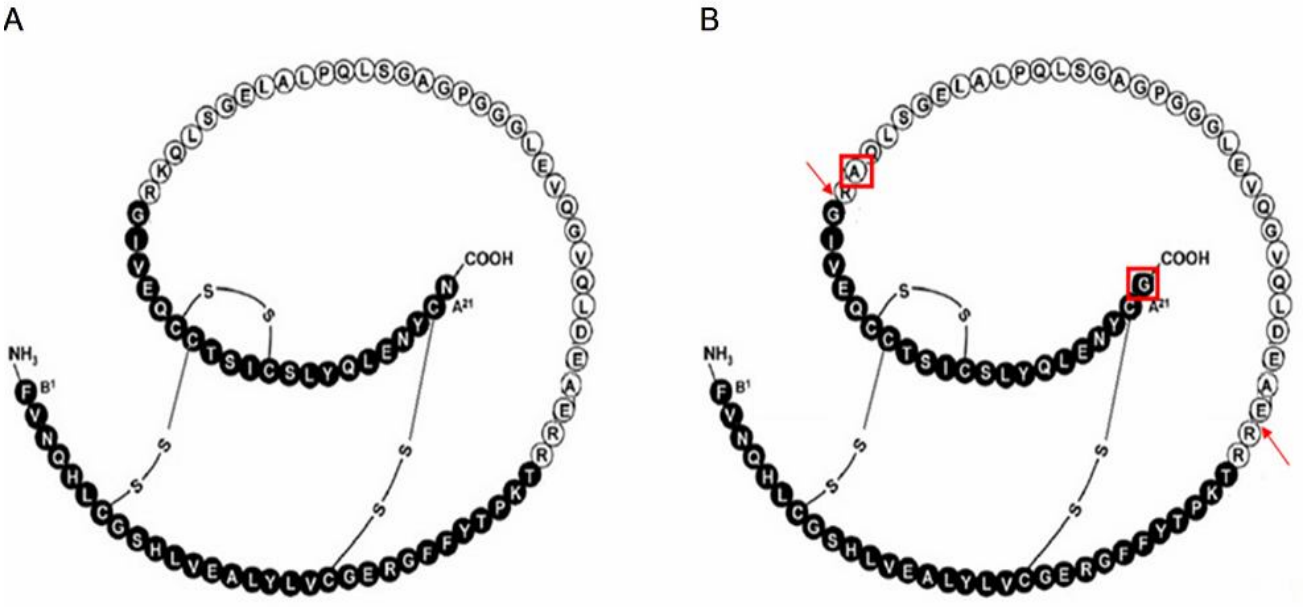

Fig. 1. Comparison of amino acid sequence of human natural (A) and modified (B) proinsulin. Trypsin cleavage sites for production of glargine insulin are shown by arrow. Replaced amino acids in modified proinsulin are shown by square.

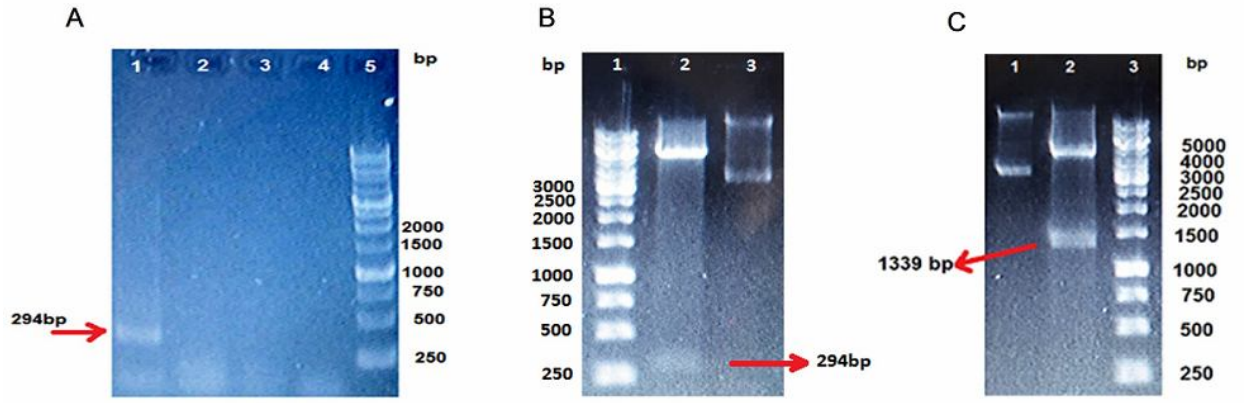

Fig. 2. Sub-cloning the modified proinsulin encoding sequence into pET-21b(+) vector. (A) Selection of the recombinant colony by colony PCR using specific primers: Lanes 1-4, PCR product from 4 different colonies; lane 5, 1kb DNA ladder. (B) Confirmation of the recombinant vector pET21b-proInsG by using NdeI and XhoI digestion: Lane 1, 1kb DNA ladder; Lane 2, digestion product; and Lane 3, undigested plasmid. (C) Confirmation of the pET21b-proInsG by using PstI digestion: Lane 1, undigested plasmid; Lane 2, digestion product; and Lane 3, 1kb DNA ladder.

In next step, the synthetic proinsulin fragment was sub-cloned in pET-21b(+). Sub-cloning of the fragment was confirmed by colony PCR amplification and restriction enzyme digestion (Fig. 2). As shown in Fig. 2A a 294 bp fragment (lane 1) has been amplified in PCR reaction using specific primers. The restriction mapping was performed by enzymatic digestion of extracted plasmids from selected colonies with NdeI and XhoI and PstI enzymes

(Figs. 2B and 2C). As it can be seen in Fig. 2B after digestion with NdeI and XhoI a 294 bp fragment has been cut out from the recombinant vector and the vector can be seen as a 5363 bp fragment (lane 2). Also, after digestion with PstI, that had a recognition site on vector and other on proinsulin encoding sequence, was formed two separate fragments of 1339 bp and 4320 bp fragment (Fig. 2C, lane 2).

Expression of the Human Modified Proinsulin

The modified proinsulin was expressed in *E. coli* BL21(DE3) by the induction with IPTG and the result was analyzed by SDS-PAGE (Fig. 3). As the Fig, 3 shows the presence of a 10.5 KDa recombinant protein is seen on the gel (lane 2). Also, western blotting analysis using anti His-Tag antibody confirmed the expression of the recombinant protein (Fig. 4, lane 2).

Purification of the Recombinant Modified Proinsulin

To determine the amount of recombinant protein produced in the form of inclusion body, after cellular sonication and upper phase separation, the precipitate was washed with washing buffer containing 3 M urea. SDS-PAGE analysis of the precipitate after 3 steps washing confirmed the significant expression of recombinant protein in inclusion body form (Fig. 5). Then, purification of refolded recombinant protein was carried out using a nickel affinity chromatography method through a batch system without using the column. SDS-PAGE analysis of the purified recombinant protein confirmed the significant purification of recombinant protein and its purity was measured about 80 % by using Image J Software (Fig. 6).

DISCUSSION AND CONCLUSION

Diabetes is developing rapidly around the world, and efficient methods for producing insulin are highly required. In 2018, it was estimated that 405.6 million people suffered from Type 2 diabetes, and this number is projected to increase to approximately 510.8 million by the year 2030. Based on these estimates, the global usage of insulin is estimated to rise from 516.1 million vials (1000 IU) to 633.7 million vials in 2030 (Basu *et al.* 2019) [19]. In a study conducted in 2020, A PCR-based strategy was employed for the cloning and verification of human insulin which eliminated the use of affinity tags since an untagged pET21b expression vector was employed [20]. Todays, protein production in form of inclusion body is one of the most effective methods of insulin production. Inclusion body strategy for the production of recombinant proteins has the main advantage that the gene product is typically protected from proteolysis and protein is produced with high level of expression. Of course, in this method require the technologies for protein refolding [21]. In the Institute of Biotechnology and Antibiotics a new bacterial host strain (Escherichia coli 20) and the pIBAINS expression vector were constructed that provide greater efficiency for the production of recombinant human insulin [22]. In the present study, the expression of proinsulin protein was approximately 40% in mass form. Excessive intracellular expression of proinsulin leads to accumulation and massification and dramaticly results in stabilization against proteolysis [23]. In another study in 1982 for the production of insulin, the production of recombinant protein in massive form was reported up to 30% of the total protein [24]. In this study, simultaneously purifying and refolding the recombinant protein were used in nickel affinity chromatography system. This method that requires solubility of inclusion bodies and refolding of recombinant proteins by reducing the urea concentration, was been used to purify the massive form of photoprotein aequorin in 2002 [25]. It should be mentioned that the use of these materials at high concentrations can also damage the structure of the protein [26].

In this study, with the primary aim of expressing and purifying proinsulin protein, the sequences were designed and synthesized according to the modified proinsulin sequence for the glargine analog. In this way, after the production of recombinant proinsulin and structural surveys, to achieve the ultimate aim, it is necessary to convert the produced protein to the mature insulin in purification column by the enzymatic reactions such as trypsin. Trypsin has the ability of cleavage specific peptide binding at the carboxylic end of arginine and at a lesser extent of lysine. This enzyme cleave C-terminal of the Arg^{B32} and Arg^{A0} for the production of glargine insulin. Nevertheless, it may cleave other Arg and Lys amino acids on proinsulin and

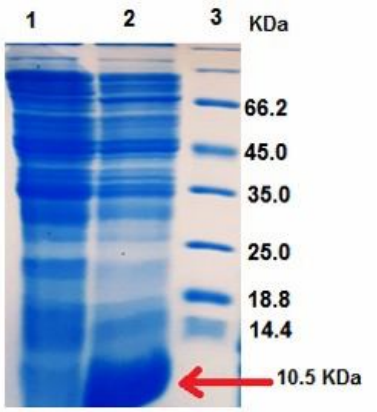

Fig. 3. SDS-PAGE analysis of the modified proinsulin expression. lane 1, uninduced sample; Lane 2, induced sample at 4 h after induction; and lane 3, protein molecular weight marker.

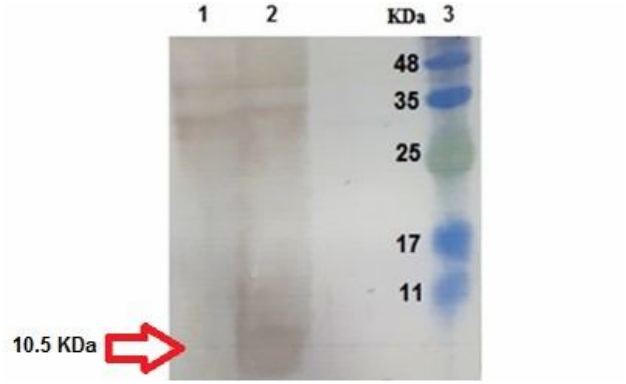

Fig. 4. Western blotting analysis of the modified proinsulin expression. lane 1, uninduced sample; Lane 2, induced sample at 4 h after induction; and lane 3, protein molecular weight marker.

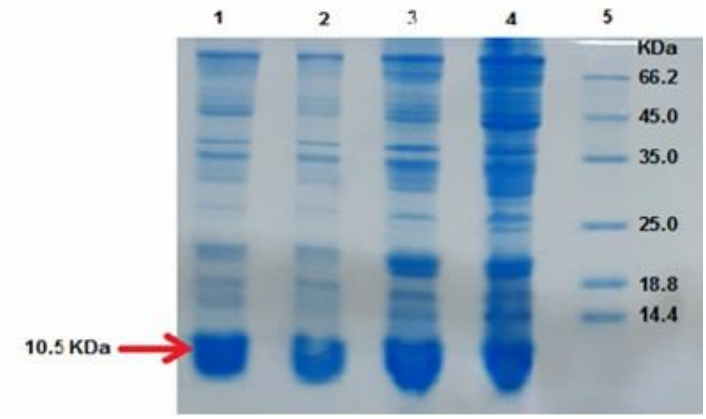

Fig. 5. SDS-PAGE analysis of inclusion body precipitate washed by 3 M urea solution. lane 1, sample after 3 times washing; lane 2, sample after 2 times washing; lane 3, sample after 1 time washing; lane 4, sample before washing; and lane 5, protein molecular weight marker.

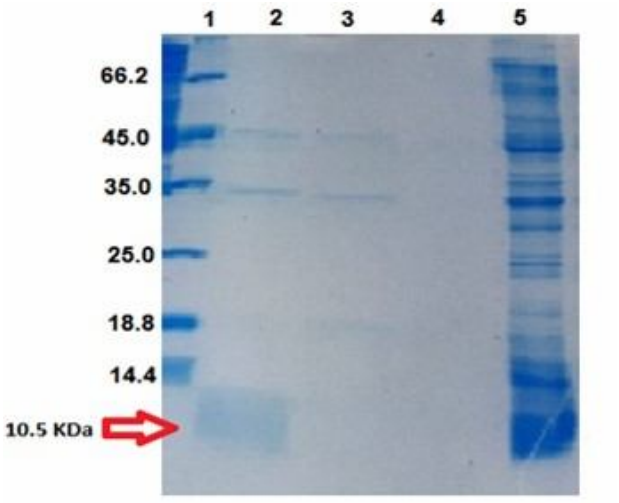

Fig. 6. SDS-PAGE analysis of the recombinant protein refolded by imidazole solution. lane 1, protein molecular weight marker; lane 2, sample with 200 mM imidazole; lane 3, sample with 100 mM imidazole; lane 4, sample with 20 mM imidazole; and lane 5, cellular precipitate sample.

produces insulin derivatives. For example, it may cut the C-terminal of Lys^{B29} and cause formation of a disthreonine. Some studies have used the acid citraconic anhydride to reduce the formation of these derivatives [27]. Acid citraconic anhydride, by citraconizing the amine groups of lysine amino acids, generates a negative charge and so the lysine is not known by trypsin. In this study, a lysine residue at the carboxyl end of chain C was also replaced with alanine at the position 34. In this way, the probability of producing another category of derivatives will also decrease.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

REFERENCES

[1] WHO Model List of Essential Medicines (2015). https://www.who.int/medicines/publications/essentialmedicines/en/

[2] E. Loghmani, in: J. Stang, M. Story (Eds.), Diabetes Mellitis: Type 1 and Type 2, University of Minnesota, Minneapolis, MN, 2005.

[3] R.A. DeFronzo, E. Ferrannini, P. Zimmet, G. Alberti, International Textbook of Diabetes Mellitus, John Wiley & Sons, 2015.

[4] J. Markussen, U. Damgaard, M. Pingel, L. Snel, A.R. Sørensen, E. Sørensen, Diabetes Care 6 (1983) 4.

[5] L. Thim, M.T. Hansen, A.R. Sørensen, FEBS Lett. 212 (1987) 307.

[6] Z. Gu, M. Weidenhaupt, N. Ivanova, M. Pavlov, B. Xu, Z.G. Su, J.C. Janson, Protein Expr. Purif. 25 (2002) 174.

[7] World Health Organization, Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications, 1999. https://apps.who.int/iris/handle/10665/66040

[8] F. Märki, W. Albrecht, Diabetologia 13 (1977) 293.

[9] D. Wong, The ABCs of Gene Cloning, Springer US, 2006.

[10] S.R. Joshi, R.M. Parikh, A. Das, J. Assoc. Physicians India 55 (2007) 19.

[11] Z. Vajo, W.C. Duckworth, Pharmacol. Rev. 52 (2000) 1.

[12] S.A. Tabish, Int. J. Health Sci. 1 (2007) V.

[13] S. Sahdev, S.K. Khattar, K.S. Saini, Mol. Cell. Biochem. 307 (2008) 249.

[14] J. Nilsson, P. Jonasson, E. Samuelsson, S. Stahl, M.

Uhlén, J. Biotechnol. 48 (1996) 241.

[15] T. Kjeldsen, Appl. Microbiol. Biotechnol. 54 (2000) 277.

[16] N.A. Baeshen, M.N. Baeshen, A. Sheikh, R.S. Bora, M.M.M. Ahmed, H.A.I. Ramadan, K.S. Saini, E.M. Redwan, Microb. Cell Fact. 13 (2014) 141.

[17] R. Hilgenfeld, G. Seipke, H. Berchtold, D.R. Owens, Drugs. 74 (2014) 911.

[18] G. Walsh, Appl. Microbiol. Biotechnol. 67 (2005) 151.

[19] S. Basu, J.S. Yudkin, S. Kehlenbrink, J.I. Davies, S.H. Wild, K.J. Lipska, J.B. Sussman, D. Beran, Lancet Diabetes Endocrinol. 7 (2019) 25.

[20] K. Govender, T. Naicker, J. Lin, *et al.,* AMB Expr. 10 (2020) 43.

[21] R. Rudolph, Protein Eengineering: Principles and Practice, John Wiley and Sons, 1996.

[22] M. Zieliński, A. Romanik-Chruścielewska, D. Mikiewicz, N. Łukasiewicz, I. Sokołowska, J. Antosik, A. Sobolewska-Ruta, A. Bierczyńska-Krzysik, P. Zaleski, A. Płucienniczak, Protein Expr. Purif. 157 (2019) 63.

[23] S.H. Shen, Proc. Natl. Acad. Sci. 81 (1984) 4627.

[24] D.C. Williams, R.M. Van Frank, W.L. Muth, J.P. Burnett, Science 215 (1982) 687.

[25] K. Glynou, P.C. Ioannou, T.K. Christopoulos, Protein Expr. Purif. 27 (2003) 384.

[26] A. Singh, V. Upadhyay, A.K. Upadhyay, S.M. Singh, A.K. Panda, Microb. Cell Fact. 14 (2015) 41.

[27] R.E. Zimmerman, D.J. Stokell, M.P. Akers, Glargine Proinsulin and Methods of Producing Glargine Insulin Analogs Therefrom. Google Patents, 2011.