Journal of Chromatography B, 908 (2012) 105–112

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Insulin related compounds and identification

Cheol-Ki Min[a], Jin-Won Lee[a], Chang-Kyu Kim[b], Seong Hwan Kim[c], Sang Yeol Lee[d], Sang-Myung Lee[e], Young-Jin Son[f,*]

[a] Department of Integrated Biotechnology, Sogang University, Seoul 121-742, Republic of Korea
[b] Department of Animal Biotechnology, College of Animal Bioscience and Technology, Konkuk University, Seoul 143-701, Republic of Korea
[c] Lab. of Chemical Genomics, Korea Research Institute of Chemical Technology, Yuseong-gu, Daejeon 305-600, Republic of Korea
[d] Department of Life Science, Gachon University, Sungnam, Kyunggi-do, 461-701, Republic of Korea
[e] Department of Chemical Engineering, Kangwon National University, Chuncheon, Gangwon-do 200-701, Republic of Korea
[f] Department of Pharmacy, Sunchon National University, Suncheon, Jeonnam 540-742, Republic of Korea

ARTICLE INFO

Article history:
Received 5 June 2012
Accepted 24 August 2012
Available online 16 September 2012

Keywords:
Insulin related compounds
Identification
Chromatography

ABSTRACT

Insulin-related compounds (IRCs), which originate during the expression and purification of human insulin using recombinant *Escherichia coli*, were purified and identified. We investigated the identity of IRCs and their origin. We also presented methods for inhibiting IRC formation. The strains used in this report were *E. coli* B5K and *E. coli* H27R. *E. coli* B5K had a 6-amino acid-fused peptide at the N-terminus of proinsulin, and *E. coli* H27R had a 28-amino acid-fused peptide at the N-terminus of proinsulin. We investigated the identity of IRCs and their origin by mainly using High Performance Liquid Chromatography (HPLC). The well-known IRCs, desamido human insulin and desthreonine human insulin, formed in both strains. In addition to these two IRCs, the B5K strain produced three different IRCs, $Arg^{A(0)}$-insulin (IRC 1), prepeptide-insulin (IRC 2), and $Glu^{A(22)}$-insulin (IRC 3). The amounts of IRC 1, IRC 2, IRC 3 were approximately 0.1–0.3% after final purification step. Among these IRCs, $Arg^{A(0)}$-insulin, prepeptide-insulin, and desthreonine insulin originated from incomplete enzyme reaction. $Glu^{A(22)}$-insulin was formed when we used a double stop codon during the expression of preproinsulin; that is, it was formed by the misreading of the first stop codon through the amber mutation. The major IRCs of the H27R strain were human insulin fragment (B1–B21) (IRC 4), and A9(Ser → Asn) amino acid single mutation human insulin (IRC 5), $Arg^{B(31)}$-insulin (IRC 6). Human insulin fragment (B1–B21) was formed by β-mercaptoethanol, which was added during refolding. It formed when the disulfide bonds between A-chain and B-chain of human insulin were cut by β-mercaptoethanol, followed by cleavage of the B-chain by trypsin and carboxypeptidase B. A9(Ser → Asn) amino acid single mutation human insulin originated from the mistranslation of A9 serine, such that asparagine was translated instead of serine. $Arg^{B(31)}$-insulin originated from incomplete enzyme reaction. The amount of IRC 4 was 10–15% after enzyme reaction. The amounts of IRC 5, IRC 6 were around 0.2% after final purification step. We present methods for inhibiting the formation of IRCs by controlling the amount of enzyme, controlling the rate of enzyme reaction, using a single stop codon, using hydrogen peroxide (H_2O_2) to inhibit β-mercaptoethanol, and modifying the A9 codon.

1. Introduction

The human insulin molecule, which is used for diabetes mellitus therapy, contains 51 amino acid residues in two polypeptide chains (A and B) linked by two inter-disulfide bonds and one intra-disulfide bond. Ever since the successful mass production of human insulin by recombinant *Escherichia coli* and yeast, various methods for mass production of insulin have been developed using new expression strains and production processes in order to increase production yield [1]. The purity of recombinant insulin must be more than 98%, with the insulin-related compound A21-desamido insulin accounting for less than 2%, in order for it to be used as medicine for diabetes. And high molecular weight proteins must be less than 1%. Immunogenic polypeptides must be less than 10 ppm. Bacterial endotoxins must be less than 10 EU/mg. Residual DNA must be less than 20 pg/human dose [2,3].

IRCs are produced when recombinant human insulin is expressed, purified and also during storage of insulin. The well-known insulin-related compounds are desamido insulin (A21, B3),

Abbreviations: IRCs, insulin related compounds; des-X insulin, desthreonine insulin and desamido insulin; IEX, ion exchange chromatography.

* Corresponding author at: 255 Jungangno, Suncheon, Jeonnam 540-742, Republic of Korea. Tel.: +82 61 750 3755; fax: +82 61 750 3708.
E-mail address: sony@sunchon.ac.kr (Y.-J. Son).

1570-0232/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jchromb.2012.08.028

mono-arginine insulin (B31), diarginine insulin (B31–32), desthreonine insulin, ethylester insulin, des-pentapeptide insulin, des-octapeptide insulin, insulin dimer, proinsulin, intermediary insulin, N-carbamoyl-Gly insulin, N-formyl-Gly insulin, N-carbamoyl-Phe insulin, and arginine-A0-insulin [4–8]. Among these IRCs, mono-arginine insulin (B31), diarginine insulin (B31–32), and arginine insulin (A0) are produced by the side reaction of trypsin and carboxypeptidase B, and desthreonine insulin is produced by over-reaction during the enzyme reaction [9,10]. Because the related compounds decrease the production yield, we could remove or decrease their formation by optimizing the enzyme reaction condition by adding citraconic anhydride and hydrogen peroxide during the enzyme reaction with trypsin and carboxypeptidase B [11–13]. There are different types of IRCs. They include deaminated human insulin, carbamylated human insulin, formylated human insulin, and so on. They originate from the chemical modification of amino acid residues. Many IRCs, such as A21-desamidoinsulin, can be separated by ion-exchange chromatography and preparative HPLC, whereas most impurities, which are chemically modified, are not easily purified and identified. It is therefore desirable to modify the production process in order to block the formation of chemically modified impurities [14,15].

Unknown IRCs can originate from strain development and modification of the production process and must be identified. The most famous IRC, A21 desamido insulin, can be detected by the USP analysis method for IRCs; however, many other IRCs cannot be detected by this method. Undetected IRCs may cause quality problems in recombinant human insulin, so they should be thoroughly analyzed and detected in order to block their formation and to remove them during the purification process. A variety of analytical methods, including capillary zone electrophoresis as well as typical RP-HPLC methods, have therefore been developed to detect many IRCs efficiently [16–20].

In this report, the types of IRCs that originated during the expression and purification of recombinant human insulin from *E. coli* strains containing two different plasmids were analyzed by two different RP-HPLC analytical methods. Inhibitory methods to decrease the IRCs were also developed through identification of the impurities and exploring the impurity formation. The results in this report may help to increase the production yield of recombinant human insulin from other strains by decreasing potential IRCs.

2. Materials and methods

2.1. Cell construction and related compounds purification

Two human insulin expression strains, *E. coli* JM109 harboring the pPT-B5Kpi plasmid (B5K strain), which expresses a 6-amino acid-fusion peptide (Met-Thr-Met-Ile-Thr-Lys) at the N-terminus in the B-chain and *E. coli* JM109 harboring the pPT-H27Rpi plasmid (H27R strain), which expresses a 28-amino acid-fusion peptide (Met-Thr-Met-Ile-Thr-Asn-Ser-Pro-Glu-Ile-Ser-His-His-His-His-His-His-His-His-His-His-Gln-Leu-Ile-Ser-Glu-Ala-Arg). The vector maps of the pPT-B5Kpi and pPT-H27Rpi plasmids and the cloning methods for their construction were described in our previous patent and paper [21,22].

Seed culture was carried out at 30 °C and 250 rpm in a 2-L flask containing 600 mL of LB medium inoculated with 5 mL of the culture stock for 18 h. In order to start the fermentation of human insulin expression strains, cultured cells in the 2-L flask were transferred into a 300-L fermentor. After 18-h fermentation, cells were harvested by centrifugation at 8000 rpm for 30 min, and then lysed mechanically with a homogenizer (10,000 psi) after adding 1% triton X-100 and 0.05% lysozyme. The inclusion bodies containing fused proinsulins were solubilized by raising pH until 10.0 with 4 M urea, and they were converted to S-sulfonated forms by oxidative sulfitolysis with 0.2 M sodium sulfite and 20 mM sodium tetrathionate [23]. The solution containing S-sulfonated proinsulin fusion protein was diluted in 50 mM glycine (pH 10.6), 0.3 M urea to a final protein concentration of 0.5 mg/mL. In order to exchange the sulfonate groups and induce disulfide bond formation, β-mercaptoethanol (0.75 equivalents of —SH per —SSO_3—) was added and these reaction mixtures were incubated at 12 °C under gentle stirring for 17 h. After refolding, trypsin and carboxypeptidase B were added to convert the fused proinsulin to insulin.

After the enzyme reaction, we obtained high quantity fractions of related compounds from ion exchange chromatography (IEX) and preparative HPLC, and then we carried out semi-prep HPLC with these fractions to remove unwanted related compounds. The resin of IEX was Macroprep High S cation exchange resin (Bio-rad, CA, USA), equilibrium buffer was 4 M urea, 0.25 M acetic acid, and elution buffer was 4 M urea, 0.25 M acetic acid, 1 M NaCl. The elution of human insulin was performed by NaCl concentration gradient. Preparative HPLC was carried out by C8 reverse phase resin (EKa chemical, Sweden). The buffer of preparative HPLC was 0.25 M acetic acid, and then human insulin was eluted by acetonitrile gradient.

To obtain high purity IRCs, semi-preparative HPLC was performed with C8 reverse phase column (Vydac, MN, USA). The next steps were a buffer change with molecular cutoff 10,000 membrane, concentration, and freeze-drying for 3 days, and we then analyzed the IRCs by several analytical methods.

Table 1
Chromatograph conditions for IRCs analysis.

Condition	Method 1	Method 2
HPLC system	Water alliance	Waters alliance
Column	Kromasil C4 (4.6 mm × 150 mm)	Vydac C18 (4.6 mm × 250 mm)
Buffer A	$NaH_2PO_4H_2O$ 0.05 M, $NaClO_4$ 0.1 M (pH 2.5)	Solvent[a] + acetonitrile (82:18)
Buffer B	100% acetonitrile	Solvent[a] + acetonitrile (50:50)
Run control	1 mL/min, DAD 214 nm (room Temp.)	1 mL/min, 214 nm (40 °C)
Elution	35 → 45% B during 28 min	(81% A + 19% B)/60 min

[a] Solvent: 28.4 g sodium sulfate anhydrous/1 L water, pH 2.3

2.2. Detection and analysis of IRCs

Two analytical methods were used to detect a variety of IRCs after the enzyme reaction and final purification step, because we could not analyze many IRCs with only one analytical method. The first analytical method (Method 1) used to analyze human insulin derivatives differed from the USP analytical method (Method 2). We detected desthreonine (B30) insulin, mono-arginine (A0) insulin, prepeptide proinsulin, and insulin B-chain fragment with Method 1, whereas it was impossible to detect these IRCs with the USP analytical method (Method 2). The strength of Method 1 was the simultaneous detection of insulin, proinsulin, and intermediates of the enzyme reaction, allowing determination of the state of the enzyme reaction.

The second analytical method (Method 2) is a typical analytical method for human insulin detection in USP. It is usually used to analyze human insulin and desamido human insulin. But we could not analyze desthreonine human insulin, which is a major derivative, by Method 2 during the enzyme reaction. The detailed conditions for the above-mentioned analytical methods are described in Table 1.

The peptide mapping of human insulin derivatives was carried out by USP human insulin peptide mapping method.

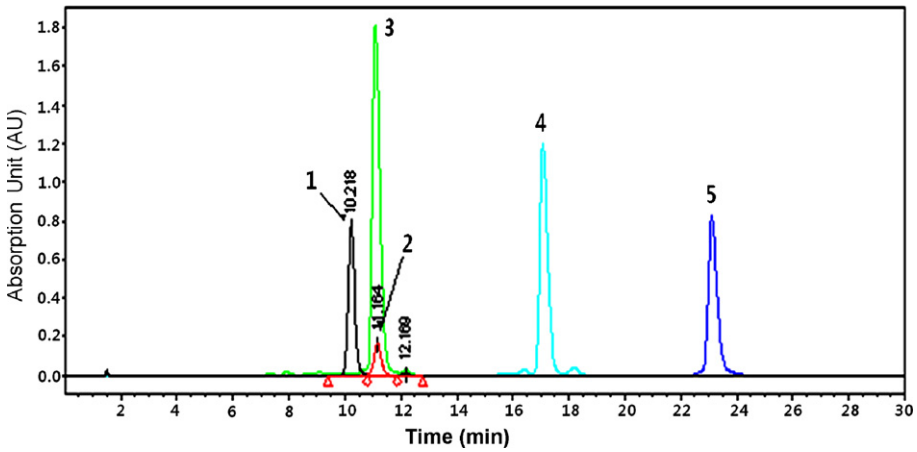

Fig. 1. RP-HPLC profile of insulin, des-X insulin, fused proinsulin and two IRCs for B5K strain by Method 1. Peak 1: insulin, peak 2: des-X insulin, peak 3: IRC 1, peak 4: IRC 2, peak 5: fused proinsulin. Des-X insulin contains desthreonine insulin and desamido insulin. The figure was the overlapping of 4 chromatograms. Peak 1 and peak 2 were detected simultaneously in a chromatogram. Other three peaks (peak 3, peak 4, peak 5) were detected independently by different analysis. And then four chromatograms were overlapped to compare the retention time of each peak.

2.3. Identification of IRCs

To identify IRCs that originated during the enzyme reaction and purification, we purified the IRCs to more than 97% purity and then analyzed the molecular weights, N-terminal sequences, and peptide maps. If necessary, we carried out HPLC analysis of the A and B chains after cleavage of the human insulin chain using DTT. The molecular weights and N-terminal amino acid sequences of IRCs were measured by MALDI-TOF (Bruker Daltonik Inc.) and protein sequencing system (PE Biosystems Inc.), respectively. Five amino acids in each N-terminal sequences were measured initially and, if necessary, the number of analyzed amino acids was increased. Peptide mapping was performed to detect the modification site in human insulin after purification. The human insulin standard and the IRCs were cleaved into four fragments with *Staphylococcus aureus* protease V8 and then compared. Finally, the A and B chains of the IRCs were compared with those of the standard to check for modifications in the A and B chains.

3. Results and discussion

3.1. HPLC profile of IRCs from the B5K strain

The derivatives of recombinant human insulin were side-products caused by chemical modification of amino acids during the purification process and by overcutting or undercutting of C-peptide during the enzyme reaction. Fig. 1 is the chromatogram for high purity IRC 1, IRC 2, and fused proinsulin after the enzyme reaction with fused proinsulin from the B5K strain. Methods 1 and 2 resulted in different chromatogram patterns for each sample. The typical impurity produced during the human insulin purification process, desamido human insulin, could be detected by Method 2 (USP Method). When we analyzed desamido human insulin with Method 2, we could get an exclusive peak for desamido human insulin. But the other major impurity, desthreonine human insulin, could not be detected by Method 2 (data not shown). The peak of desthreonine human insulin overlapped that of human insulin in the chromatogram of Method 2, whereas the peak of desthreonine human insulin did not overlap that of human insulin in the chromatogram of Method 1 (Desthreonine Method). But the drawback of Method 1 was that the exact quantity of desthreonine human insulin could not be measured with only Method 1. When we used Method 1, the peak of desthreonine human insulin was split from that of human insulin, but overlapped that of desamido human insulin. In other words, we had to use both Method 1 and Method 2 to analyze desamido human insulin and desthreonine human insulin. The overlapping peak (peak 2) for desthreonine and desamido human insulin was named des-X human insulin in Fig. 1; that is, des-X human insulin consists of desthreonine human insulin and desamido human insulin in the chromatogram by Method 1. Desamido human insulin is the derivative that is formed during the overall purification process. When we analyzed samples after the enzyme reaction step, the average amount of desamido human insulin was 1–2%. The main derivatives from the B5K strain after the enzyme reaction were des-X insulin (peak 2), IRC 1 (peak 3), IRC 2 (peak 4), and fused proinsulin (peak 5). We confirmed that peak 5 was fused proinsulin, because the retention time of the peak was that of the fused proinsulin and the amount of the peak gradually decreased during the enzyme reaction. Peak 5, the fused proinsulin, has no effect on the quality of final human insulin after the enzyme reaction because it can be removed by an ion exchange chromatography process during purification. IRC 1 and IRC 2 remained above 0.2% after purification, so methods to block their formation and to identify them need to be devised.

Fig. 2 shows the chromatogram of IRC 3, which is another impurity after fermentation and enzyme reaction with the same strain by Method 2. IRC 3 could not be separated by Method 1, so it must be analyzed by Method 2. Peak 1 was the human insulin standard, peak 2 was desamido insulin, and peak 3 was IRC 3. IRC 3 could not be removed by IEX and Prep HPLC, so the origin of its formation should be studied and its formation must be blocked.

3.2. Identification of IRC 1, 2, and 3

As mentioned previously, the main impurities, IRC 1, 2, and 3, originated from the B5K strain after fermentation and enzyme reaction. The amounts of IRC 1, IRC 2 were above 0.2%, respectively. The amount for IRC 3 was approximately 0.1–0.3% using Method 1 and Method 2. The origins of their formation were elucidated in order to block their formation. Table 2 shows the results of molecular weight analysis, N-terminal sequence analysis, and peptide mapping analysis for IRCs that were more than 97% pure after purification. The molecular weight of IRC 1 was 5964.21 by MALDI-TOF. It was higher than that of human insulin, 5807.58, by 156.63. The N-terminal sequences of IRC 1 were Arg-Gly-Ile-Val-Glu, and Phe-Val-Asn-Gln-His. IRC 1 had two N-terminal regions, similar to the A and B chains of human insulin. The difference compared with

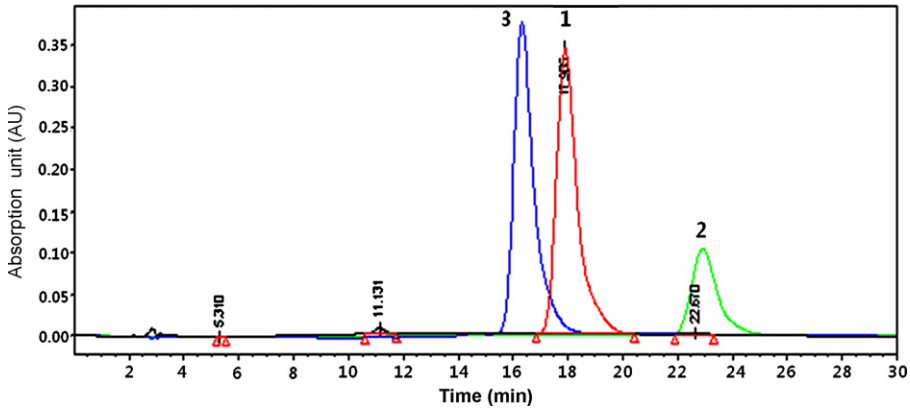

Fig. 2. RP-HPLC profile of insulin, desamido insulin, and IRC 3 for B5K strain by Method 2. Peak 1: insulin, peak 2: desamido insulin, peak 3: IRC 3. The figure was the overlapping of 3 chromatograms.

Table 2
Results of MALDI-TOF, N-terminal analysis, and peptide mapping for three IRCs (IRC 1, IRC 2, and IRC 3) of B5K strain.

Analysis	IRC1	IRC2	IRC3
MALDI-TOF	5964.21	6380.75	5937.84
N-terminal	Arg-Gly-Ile-Val-Glu Phe-Val-Asn-Gln-His	Gly-Ile-Val-Glu-Lys Thr-Met-Ile-Thr-Lys	Gly-Ile-Val-Glu-Lys Phe-Val-Asn-Gln-His
Peptide mapping	Modification in fragment IV (A1–A4)	–	Modification in A chain of fragment II (A18–A21)
Identification	$Arg^{A(0)}$-insulin	Fused insulin	$Gln^{A(22)}$-insulin

the A-chain (Gly-Ile-Val-Glu-Lys) and B-chain (Phe-Val-Asn-Gln-His) for human insulin was the 1st amino acid, the arginine, in the A-chain for IRC 1. If arginine binds to the A-chain of human insulin, the theoretical molecular weight of the modified human insulin is 5963.78. Thus it was confirmed that IRC 1 was modified human insulin containing an additional arginine at the N-terminus of the A-chain. Fig. 3 shows the peptide maps for human insulin and IRC 1. Fragments 1–3 of IRC 1 overlapped with those of human insulin (data not shown), whereas fragment 4 of IRC 1 did not overlap with that of human insulin. This result indicated that there was a modification in the A1–A4 amino acids in IRC 1. Consequently IRC 1 was $Arg^{A(0)}$-human insulin, which contained arginine at the A0 site in human insulin. It was an impurity caused by incomplete enzyme reaction. Its formation could be inhibited by increasing the amount of trypsin or by increasing the enzyme reaction time.

The molecular weight and N-terminal sequence of IRC 2 were assessed. The molecular weight of IRC 2 was 6380.75. The molecular weight of IRC 2 was higher than that of human insulin by 573.17. The N-terminal sequences of IRC 2 were Gly-Ile-Val-Glu-Lys, and Thr-Met-Ile-Thr-Lys. The A-chain of IRC 2 was identical to that of human insulin, whereas the B-chain of IRC 2 was different from that of human insulin. The N-terminal sequence in the B-chain of IRC 2 was from a fusion. It was considered that the first amino acid, methionine, of the fusion part was removed from the N-terminus of the B chain of IRC 2. Like IRC 1, IRC 2 therefore originated from an incomplete enzyme reaction. So, the generation of IRC 2 could be inhibited by modifying the conditions for the enzyme reaction. The methionine at the N-terminus in the B-chain of IRC 2 was removed by N-terminal methionine protease in *E. coli* [24].

The molecular weight of IRC 3 was 5937.84. The molecular weight of IRC 3 was higher than that of human insulin by 130.26, suggesting that one amino acid was attached to human insulin. The N-terminal sequences of the A-chain and B-chain of IRC 3 were identical to those of human insulin. The peptide mapping showed that fragment 2 (A18–A21 to B14–B21) of IRC 3 was different from that of human insulin (Fig. 4). To check for modification in the A or B chain, HPLC analysis was performed with the human insulin standard and IRC 3 after cleavage of the disulfide bond that linked

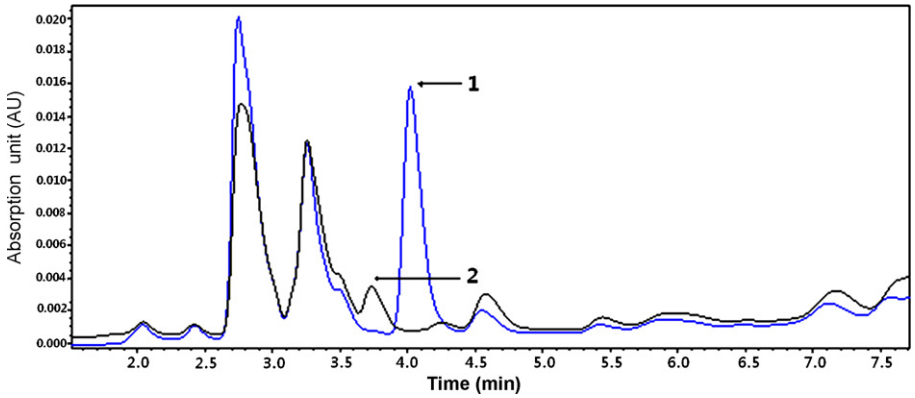

Fig. 3. The chromatogram after peptide mapping analysis with standard insulin and IRC 1. The figure was the overlapping of 2 chromatograms. Blue line is the chromatogram with standard insulin and black line is the chromatogram with IRC 1. Peak 1: fragment 4 of the standard of human insulin, peak 2: fragment 4 of IRC 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

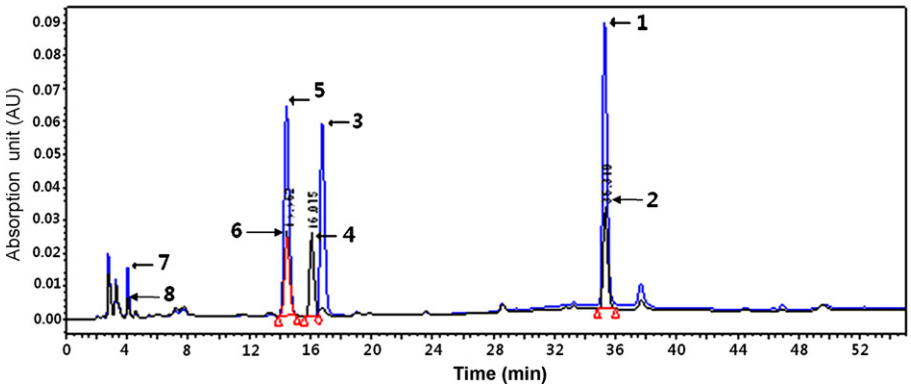

Fig. 4. The chromatogram after peptide mapping analysis with standard insulin and IRC 3. The figure was the overlapping of 2 chromatograms. Blue line is the chromatogram with standard insulin and black line is the chromatogram with IRC 3. The red peak was designated with red color automatically by the software of HPLC. Peak 1: fragment 1 of standard insulin, peak 2: fragment 1 of IRC 3, peak 3: fragment 2 of standard insulin, peak 4: fragment 2 of IRC 3, peak 5: fragment 3 of standard insulin, peak 6: fragment 3 of IRC 3, peak 7: fragment 4 of standard insulin, peak 8: fragment 4 of IRC 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the two chains together. The result showed that the peak for the A-chain of IRC 3 had a different retention time than that of human insulin (Fig. 5). Based on the above results, we could infer that an amino acid such as Glu, Gln, and Lys was attached at A18–A21 in the A-chain of IRC 3 because the molecular weights of these amino acids are around 148 Da. The preproinsulin expression gene of the B5K strain had two stop codons, amber (UAG) and ochre (UAA). This suggested that IRC 3 was $Gln^{A(22)}$-human insulin because the first stop codon was read through and then translated as glutamine by an amber mutation in the JM109 host strain [25]. We therefore constructed a preproinsulin expression gene containing one stop codon in order to prevent the expression of IRC 3.

3.3. HPLC profile of IRCs from H27R strain

One of the methods for increasing the human insulin expression yield and promoting efficient purification is using a fusion partner during human insulin expression [26–28]. We carried out experiments in which the length and type of fusion partner were modified to increase the preproinsulin expression level and purification yield and then constructed the H27R strain whose expression rate was 30% higher than that of the B5K strain [21]. The H27R strain expressed preproinsulin with a 28-amino acid fusion partner. It had one stop codon to prevent the expression of IRC 3. As mentioned in Section 1, most of the impurities are produced by incompletion or overreaction in the enzyme reaction. So we had to check for new derivatives during the enzyme reaction when we changed the expression strain. The IRCs from the B5K strain (IRC 1, IRC 2, and IRC 3) were not detected from the H27R strain, whereas new IRCs were formed. Fig. 6 shows the chromatogram of the H27R strain after the enzyme reaction by analytical Method 1. Peak 1 was human insulin, peak 2 was des-X human insulin, and peak 3 was IRC 4. As seen in Fig. 6, the chromatogram of the H27R strain after fermentation and enzyme reaction was different from that of the B5K strain. The IRCs from the B5K strain were not detected in the H27R strain, whereas a new impurity was formed. We named the new impurity IRC 4 as it was peak 4 in the chromatogram. The amount of IRC 4 was 10–15% after the enzyme reaction, so IRC 4 decreased the enzyme reaction yield significantly; however, this amount decreased gradually during the enzyme reaction. In order to inhibit IRC 4 formation, we identified the impurity and researched its formation.

Fig. 7 shows the chromatogram of final recombinant human insulin after IEX and preparative HPLC of the H27R strain using analytical Method 2. Peak 1 was human insulin, peak 2 was desamido human insulin, peak 3 was IRC 5, peak 4 was IRC 6, and peak 5 was IRC 7. And Fig. 8 shows the chromatogram of same samples for Fig. 7 using analytical Method 1. As we seen in Fig. 8, desamido human insulin overlapped with desthreonine human insulin, so we named this peak des-X insulin. The absorption unit (AU) of des-X insulin peak (around 0.010) in Fig. 8 increased compared with that of desamido insulin peak (around 0.003) in Fig. 7, because the desamido insulin and desthreonine insulin were overlapped. IRC 6 and IRC 7 could not be separated by Method 1. So, we had to use both chromatographic methods (Method 1 and Method 2) in

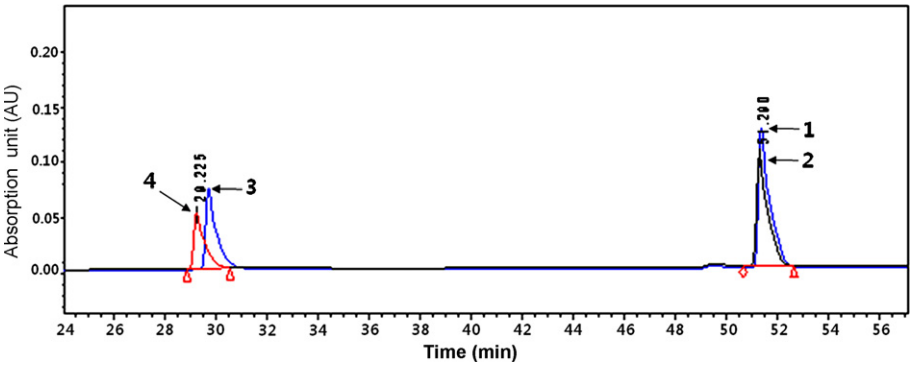

Fig. 5. HPLC profile of the standard of human insulin and IRC 3 after disulfide bond cleavage by Method 2. The figure was the overlapping of 2 chromatograms with the standard of human insulin and IRC 3. Blue line is the chromatogram with standard insulin and black line is the chromatogram with IRC 3. The red peak was designated with red color automatically by the software of HPLC. Peak 1: B chain of standard insulin, peak 2: B chain of IRC 3, peak 3: A chain of standard insulin, peak 4: A chain of IRC 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

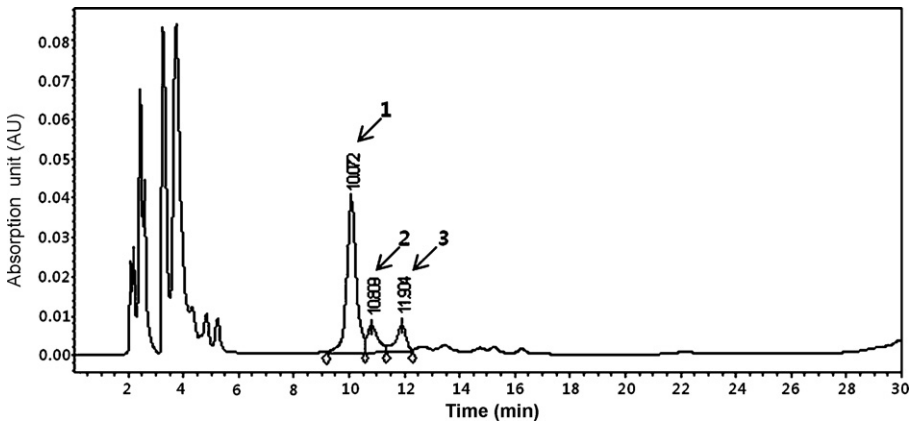

Fig. 6. RP-HPLC profile of the human insulin, des-X insulin and IRC 4 for H27R strain by Method 1 after enzyme reaction. Peak 1: human insulin, peak 2: des-X insulin, peak 3: IRC 4.

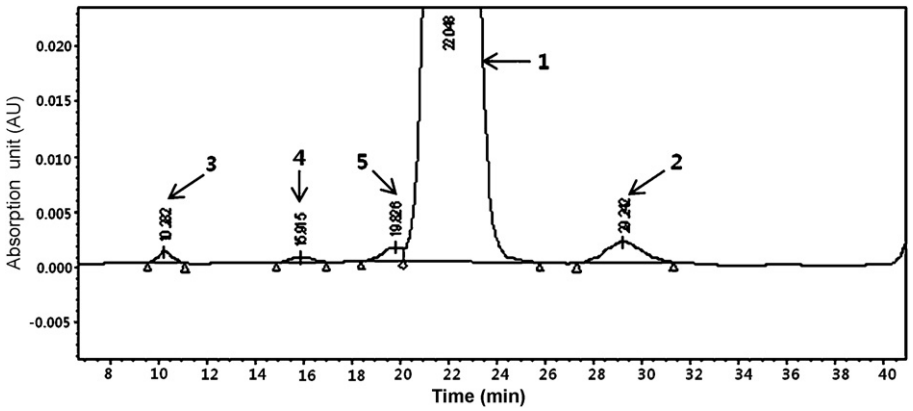

Fig. 7. RP-HPLC profile of human insulin, desamido insulin and IRCs for H27R strain by Method 2. Peak 1: human insulin, peak 2: desamido insulin, peak 3: IRC 5, peak 4: IRC 6, peak 5: IRC 7.

order to detect desamido insulin, desthreonine insulin, and IRCs without interference of other peaks. The amounts of IRC 5, IRC 6 were around 0.2% after final purification step. There was a great difference in the formation pattern for IRCs depending on fermentation batches and the pooling point during chromatography. But these were common IRCs except for IRC 7, which was not found in most batches. In our research, we analyzed the molecular weights, N-terminal sequences, and peptide maps of these IRCs, and then identified IRC 4, 5, and 6.

3.4. Identification of IRC 4, 5, and 6

Table 3 shows the molecular weights, N-terminal sequences, and peptide maps for IRC4, IRC 5, and IRC 6. The molecular weight

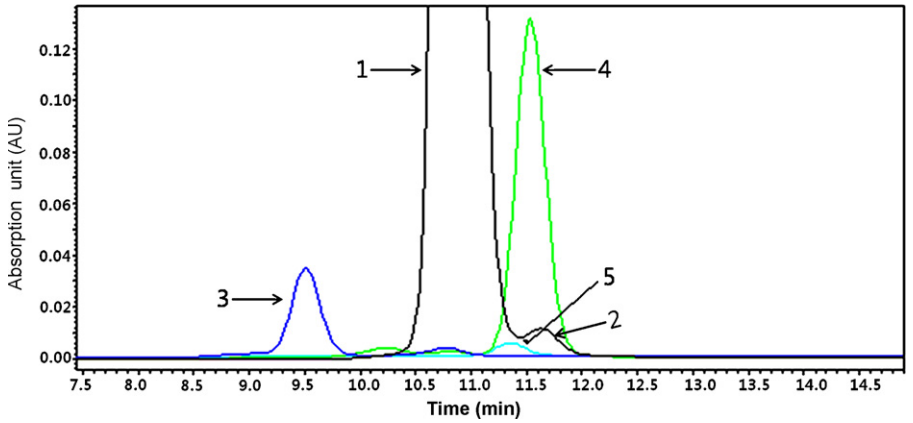

Fig. 8. RP-HPLC profile of human insulin, des-X insulin and IRC for H27R strain by Method 1. Peak 1: human insulin, peak 2: des-X insulin, peak 3: IRC 5, peak 4: IRC 6, peak 5: IRC 7. The figure was the overlapping of 4 chromatograms. Peak 1 and peak 2 were detected simultaneously in a chromatogram. Other three peaks (peak 3, peak 4, peak 5) were detected independently by different analysis. And then four chromatograms were overlapped to compare the retention time of each peak.

Table 3
Results of MALDI-TOF, N-terminal analysis and peptide mapping for three IRCs (IRC 4, IRC 5, and IRC 6) of H27R strain.

Analysis	IRC4	IRC5	IRC6
MALDI-TOF	2331.06	5834.22	5963.82
N-terminal	Phe-Val-Asn-Gln-His	Gly-Ile-Val-Glu-Gln-X[a]-X-Thr-Asn(or Ser)-Ile-X-Ser[b]	Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Thr-Arg[c]
Peptide mapping	–	Modification in A chain of fragment I (A5–A17)	Modification in B chain of fragment III (B22–B30)
Identification	B chain (B1–B21)	$Ser^{A(9)} \rightarrow Asn^{A(9)}$-insulin	$Arg^{B(31)}$-insulin

[a] X of N-terminal sequence meant that the amino acid was not analyzed.
[b] The result of N-terminal sequence of B chain for IRC 5 was omitted.
[c] The result of N-terminal sequence for IRC 6 was the sequence of fragment 3 for IRC 6 after peptide mapping cleavage.

of IRC 4 was 2331.06. It was smaller than human insulin by 3476.52. The N-terminal sequence was Phe-Val-Asn-Gln-His. Unlike human insulin, it had only one N-terminus. Human insulin has two N-termini, one from the A-chain and the other from the B-chain. The N-terminal sequence of IRC 4 was the same as that of the B-chain in human insulin. IRC 4 was the second major derivative; it comprised approximately 2% of the total amount of human insulin. IRC 4 was formed by cutting the disulfide bond between the A and B chains, followed by cleavage of the B chain of human insulin by trypsin and carboxypeptidase B during the enzyme reaction. B22-arginine can be cut by trypsin and then removed by carboxypeptidase B. The theoretical molecular weight of the human insulin fragment from B1 to B21 is 2332.02. So the molecular weight of IRC 4, 2331.06, was very similar to that of the B1–B21 human insulin fragment. We confirmed that IRC 4 was the B1–B21 human insulin fragment that was mentioned previously [13]. Even though the B22-Arg site is the cleavage site for trypsin, this Arg of proinsulin is not cut by trypsin due to steric hindrance. Misfolded preproinsulin is then denatured slowly by β-mercaptoethanol, exposing B22-Arg, so that trypsin can cut the B22-Arg cleavage site. In order to confirm this, we performed the enzyme reaction with purified preproinsulin without β-mercaptoethanol. Under these conditions, IRC 4 was not generated during the enzyme reaction. It was confirmed that β-mercaptoethanol induced the formation of IRC 4 by cleaving the chains in preproinsulin. Because IRC 4 originated due to β-mercaptoethanol, the formation of IRC 4 could be decreased by the addition of hydrogen peroxide [13]. Preproinsulin from the H27R strain was more hydrophilic than that from the B5K strain, so preproinsulin from H27R had a tendency for increased contact with β-mercaptoethanol, and then its hydrophilic character led to the cleavage of the disulfide bond of preproinsulin. Actually, the preproinsulin from the H27R strain was very sensitive to the amount of β-mercaptoethanol, and the yield for refolding was decreased with high concentrations of β-mercaptoethanol [21].

Aside from IRC4, the major IRCs of the H27R strain were IRC 5 and IRC 6. As seen in Table 3, the molecular weight of IRC 5 was 5834.22, which was higher than that of human insulin by 26.64. The N-terminal sequence showed that the 9th amino acid of the A-chain in IRC 5 was changed from serine to asparagine. To confirm the result of the N-terminal sequence analysis, peptide mapping analysis was performed with IRC 5, and the sequence was compared with that of standard human insulin. According to the peptide mapping, fragment 1 (A5–A17, B1–B13) of IRC 5 was different from that of human insulin (data not shown). In order to assess the modification in the A-chain (A5–A17) or B-chain (B1–B13), the disulfide bonds of IRC 5 were cut with DTT, and then the sample was analyzed by HPLC. Fig. 9 shows the chromatogram after cutting the disulfide bonds of the standard human insulin and IRC 5. The peak of the B-chain for IRC 5 overlapped that of human insulin, whereas the peak of the A-chain for IRC 5 had a different retention time compared with that of human insulin. Thus there was a modification in the A-chain of IRC 5. We confirmed that IRC 5 had an asparagine at the 9th amino acid instead of serine. In this assay, the size difference in the molecular weight of IRC 5 compared with human insulin was 27.22, which was similar to the size difference between serine and asparagine, 27. The cause of mistranslation from serine to asparagine was presumed to be the similarity of the codons of AGC for serine and AAC for asparagines [29,30].

As seen in Fig. 10, peptide mapping for IRC 6 showed that fragments1, 2, and 4 of IRC 6 were identical to those of human insulin, whereas fragment 3 was different from that of human insulin. That meant there was a modification in B22–B30. In order to identify the modification of IRC 6, the N-terminal sequence of fragment 3 of IRC 6 was analyzed. The N-terminal sequence of IRC

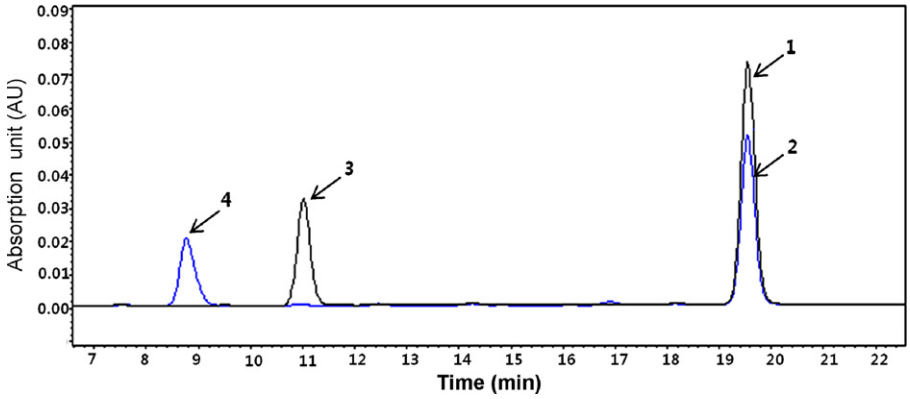

Fig. 9. HPLC profile of the standard of human insulin and IRC 5 after disulfide bond cleavage by Method 2. The figure was the overlapping of 2 chromatograms. Black line is the chromatogram with standard insulin and blue line is the chromatogram with IRC 5. Peak 1: B chain of the standard of human insulin, peak 2: B chain of IRC 5, peak 3: A chain of the standard of human insulin, peak 4: A chain of IRC 5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

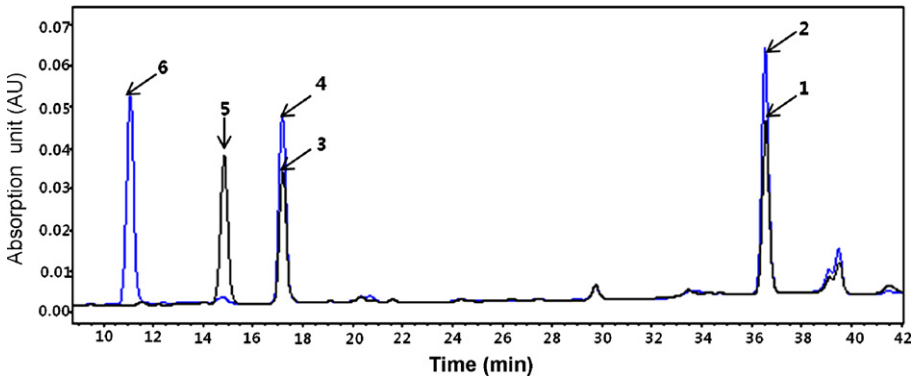

Fig. 10. The chromatogram after peptide mapping analysis with standard insulin and IRC 6. The figure was the overlapping of 2 chromatograms. Black line is the chromatogram with standard insulin and blue line is the chromatogram with IRC 6. Peak 1: fragment 1 of standard insulin, peak 2: fragment 1 of IRC 6, peak 3: fragment 2 of standard insulin, peak 4: fragment 2 of IRC 6, peak 5: fragment 3 of standard insulin, peak 6: fragment 3 of IRC 6. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

6 was Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Thr-Arg. Compared with standard human insulin, IRC 6 contained one more arginine at the C-terminus of the B-chain. The molecular weight of IRC 6 was 5963.82, which is higher than that of human insulin by 156.24. The theoretical molecular weight is 5963.21 if one arginine attaches to human insulin. IRC 6 had one arginine at the C-terminus of the B-chain. This derivative originated due to an incomplete enzyme reaction.

3.5. Formation and inhibition of IRCs

Among the six previously mentioned IRCs, IRC 1, IRC 2, and IRC 6 were formed by an incomplete enzyme reaction. The formation of this type of IRCs could be inhibited by increasing the amount of added enzyme, increasing the reaction time, and increasing the reaction temperature. But modification of the enzyme reaction conditions might trigger an overreaction of the enzyme reaction. Indeed, desthreonine human insulin and other IRCs could be formed and increased by modifying the enzyme reaction conditions. So, the enzyme reaction conditions should be carefully controlled in order to inhibit many other types of IRCs. IRC 3 was formed by read-through of the 1st translation stop codon in JM109. We removed the amber stop codon in order to inhibit IRC 3 formation. And IRC 5 originated due to mistranslation of the 9th amino acid, a serine codon. We therefore used a different codon to designate serine, in order to inhibit IRC 5 formation. IRC 4 was formed by the reducing agent that was used during the refolding step. We decreased the amount of the reducing agent and added hydrogen peroxide to block the reducing agent in order to inhibit IRC 4 formation.

4. Conclusions

Insulin-related compounds (IRCs) originating from two different human insulin expression strains were analyzed by two analytical HPLC methods in this study. One was a typical HPLC method (Method 2) in USP; the other was a specified HPLC method (Method 1) to detect desthreonine human insulin. Among well-known IRCs, desthreonine human insulin could not be analyzed by Method 2 in USP. We introduced Method 1 to detect and remove desthreonine human insulin during purification. In addition to desthreonine human insulin and desamido human insulin, there were three major unknown IRCs from the B5K strain and three IRCs from the H27 strain. All six major unknown IRCs were identified by analytical techniques, such as MALDI-TOF, N-terminal sequencing, peptide mapping, and traditional HPLC techniques after purification. We then presented methods for inhibiting the formation of these IRCs.

Acknowledgements

This research was supported by Suncheon Research Center for Natural Medicines, and in part by Sunchon National University Research Fund in 2012.

References

[1] E.P. Kroeff, R.A. Owens, E.L. Campbell, R.D. Johnson, H.I. Marks, J. Chromatogr. 461 (1989) 45.
[2] British Pharmacopoeia 2012, H.M. Stationary Office, London, 2012.
[3] USP 35-NF 30, Philadelphia, Rockville, MD, USA, 2012.
[4] V.E. Klyushnichenko, D.M. Koulich, S.A. Yakimov, K.V. Maltsev, G.A. Grishina, I.V. Nazimov, A.N. Wulfson, J. Chromatogr. A 661 (1994) 83.
[5] R.G. Nielsen, G.S. Sittampalam, E.C. Rickard, Anal. Biochem. 177 (1989) 20.
[6] C. Yomota, Y. Yoshii, T. Takahata, S. Okada, J. Chromatogr. A 721 (1996) 89.
[7] Z.H. Yang, X. Du, J.G. Tang, Appl. Biochem. Biotechnol. 76 (1999) 107.
[8] P. Moslemi, A.R. Najafabadi, H. Tajerzadeh, J. Pharm. Biomed. Anal. 33 (2003) 45.
[9] N.V. Sergeev, N.S. Glukhova, I.V. Nazimov, V.A. Guliaev, I.A. Donetskii, A.I. Miroshnikov, Bioorg. Khim. 26 (2000) 516.
[10] B.H. Frank, R.E. Heiney, W.F. Prouty, M.R. Walden, EP 0264250 B1, 1987.
[11] N.V. Sergeev, N.S. Gloukhova, I.V. Nazimov, V.A. Gulyaev, S.V. Shvets, I.A. Donetsky, A.I. Miroshnikov, J. Chromatogr. A 907 (2001) 131.
[12] Y.J. Son, C.K. Kim, Y.B. Kim, D.H. Kweon, Y.C. Park, J.H. Seo, Biotechnol. Prog. 25 (2009) 1064.
[13] Y.J. Son, C.K. Kim, B.T. Choi, Y.C. Park, J.H. Seo, J. Microbiol. Biotechnol. 18 (2008) 983.
[14] D. Gusarov, V. Nekipelova, V. Gusarova, V. Lasman, D. Bairamashvili, J. Chromatogr. B: Analyt. Technol. Biomed. Life Sci. 877 (2009) 1216.
[15] I.S. Johnson, Diabetes Care 5 (Suppl. 2) (1982) 4.
[16] B.S. Welinder, S. Linde, B. Hansen, J. Chromatogr. 265 (1983) 301.
[17] A. Peter, G. Szepesi, L. Balaspiri, K. Burger, J. Chromatogr. 408 (1987) 43.
[18] R.L. Patience, L.H. Rees, J. Chromatogr. 352 (1986) 241.
[19] B.S. Welinder, H.H. Sorensen, B. Hansen, J. Chromatogr. 361 (1986) 357.
[20] V.E. Klyushnichenko, S.A. Yakimov, A.M. Arutyunyan, A.E. Ivanov, K.V. Maltsev, A.N. Wulfson, J. Chromatogr. B: Biomed. Appl. 662 (1994) 363.
[21] C.K. Min, Y.J. Son, C.K. Kim, S.J. Park, J.W. Lee, J. Biotechnol. 151 (2011) 350.
[22] S.Y. Lee, S.J. Oh, C.K. Kim, Y.J. Son, K.H. Park, C.K. Min, B.M. Choi, T.W. Kang, J.W. Kim, WO 2004/044206 A1, 2004.
[23] J.S. Patrick, A.L. Lagu, Anal. Chem. 64 (1992) 507.
[24] P.E. Devlin, R.J. Drummond, P. Toy, D.F. Mark, K.W. Watt, J.J. Devlin, Gene 65 (1988) 13.
[25] M. Nilsson, M. Ryden-Aulin, Biochim. Biophys. Acta 1627 (2003) 1.
[26] Y. Kang, J.W. Yoon, J. Biotechnol. 36 (1994) 45.
[27] J. Nilsson, P. Jonasson, E. Samuelsson, S. Stahl, M. Uhlen, J. Biotechnol. 48 (1996) 241.
[28] C.W. Cho, S.H. Park, D.H. Nam, Biotechnol. Bioprocess Eng. 6 (2001) 144.
[29] T.L. Calderone, R.D. Stevens, T.G. Oas, J. Mol. Biol. 262 (1996) 407.
[30] X.C. Yu, O.V. Borisov, M. Alvarez, D.A. Michels, Y.J. Wang, V. Ling, Anal. Chem. 81 (2009) 9282.