Биотехнология. Теория и практика. 2014, №2, стр. 28-35
DOI: 10.11134/btp.2.2014.4

UDK 577.175.722

PRODUCTION IN BACTERIAL CELL A RECOMBINANT ANALOG OF PROINSULIN WITH THE PROPER DISULFIDE BONDS

S. Ponomarenko, Ch. Bouchaala

GlucoMetrix PVS GmbH, 14476, Potsdam, Am Mühlenberg, 11, BRD

ABSTRACT

Mechanism of molecular folding of polypeptides synthesized in animal cells, is fundamentally different from the set found in prokaryotes and yeasts, which are traditionally used in the biopharmaceutical industry for expression of recombinant human proteins. The recombinant proteins expressed in a bacterial host system are chemically refolded *in vitro* to form their correct conformations. Arrangement of disulfide bridges is essential for the process of molecular folding of polypeptides into their three-dimensional structure and, eventually, for the formation of native protein conformation required for biological activity.

Here is described a new technology of protein expression in bacteria, exploiting prokaryotic system to fold of recombinant polypeptides with formation of correct disulfide bonds. According this technology, human insulin - important therapeutic protein - can be obtained without chemical refolding from recombinant analog of preproinsulin synthesized in bacteria. This is realized by creation in host cell of polypeptide structure with correct thiol bonds in cystine, which are analogous of disulfide bridges found in transit form of proinsulin.

According to the new technology, recombinant protein is expressed at a concentration sufficient for biosynthesis of the heterogeneous membrane protein in *Escherichia coli* cells. This semifolded proteopeptide is protected from proteolytic degradation, and it may be isolated as a complex similar to the proinsulin transit form. Transit forms of proinsulin are required for further transformation of hetero-dimer molecules into the native proteohormone structure. The new method without chemical refolding prevents accumulation of isoforms, significantly reduces time and material costs by production of recombinant human insulin in its native conformation.

***Keywords: recombinant insulin, preproinsulin, refolding, membrane protein, ATP synthase*.**

INTRODUCTION

Biologically active insulin molecule [Protein Data Bank (PDB) 1A91) is a product of intracellular posttranslational transformation of linear preproinsulin polypeptide in pancreas cells [1-3].

In bacteria or in some other expression systems, which are commonly used in the biopharmaceutical industry, there are no conditions for proper post-translational folding of many therapeutic recombinant proteins, including insulin [4-5]. Accurate posttranslational folding of protein molecule in 3D structure is required for its stability, transport and functional activity. The folding process of the synthesized polypeptide into native protein molecule involves several transit states, which are preceded by the formation of alpha- and beta-structures [1]. To be effective in replace therapy the recombinant proteins produced as pharmaceuticals, should not differ in three-dimensional structure from the natural equivalent, produced in human cells.

In the industrial production of recombinant insulin expressed in bacterial cells is employed the chemical refolding method of isolated polypeptide [4-6] to achieve its correct three-dimensional

structure. This process of oxidative folding *in vitro* is hard to control, as the number of molecules with the correct cysteine-cysteine (Cys-Cys) bonds and proper 3D structure depends on a wide range of factors: the length and amino acids composition of the polypeptide; the number of Cys residues; pH buffer, its components, and ionic strength; temperature and speed of chemical reaction also from ingredients, conducting or increasing the process [5]. Refolding *in vitro* is a costly and time consuming stage in the fabrication of recombinant drugs [5-6]. Another limitation of the refolding method for recombinant insulin is contamination of the pharmaceutical active substance by isoforms, to remove which should apply additional treatment and purification.

In order to overcome these disadvantages of currently insulin production, a new bacterial expression system was designed by gene-engineering. It provides accumulation in host cells *Escherichia coli* (*E.coli*) preproinsulin analogue with Cys-Cys covalent bonds which are correct disulfide bridges in the molecule prohormone [1] and finally in the native hormone [2-3].

We describe a new technology in which a bacterial membrane-bound protein assists in expression of a chimeric polypeptide like preproinsulin analog, serves as a folding-helper and prevents building of isoforms as well as a method of isolation of recombinant human protein in the transit form necessary to obtain the active hormone substance without chemical refolding.

MATERIALS AND METHODS

Molecular biological and molecular genetic procedures were carried out as described by Sambrook et al. [7]. All reagents and solvents were analytical grade and are commercial available. All used restriction enzymes were from NEB (www.neb.com).

Recombinant gene cDNA was constructed *in silico* according to the codon usage of the host organism *E.coli* and especially to the codon usage within the *unc* operon (EMBL GenBank ID J01594). The *unc* operon consists of gene cluster coding for the eight types of subunits of the proton-translocating ATPase (EF_oEF_1).

After the chemical synthesis of DNA in EZBioLab (www.ezbiolab.com) a fragment of about 0.7 Kbp was inserted into multicloning site of the vector pET-21a (Novagen) and expression plasmid was named pBsrG2. Other commercial host vectors were used to find optimal expression level of the chimeric preproinsulin analog named here CpIA. As a result of the transformation into various competent cells (Table) several clones of *E.coli* cells were obtained. Plasmid DNA was isolated by anion exchange chromatography using Qiagen and MN Kits, the cloning was carried out using the method of Sambrook et al. [7]. *E.coli* cells were incubated under standard conditions: medium Luria Bertrani [7] at 37°C and 250 rpm. Conditions differ from this standard given in the text. Heterogeneous expression of the protein was induced by addition of 1 mM IPTG. After cultivation, the

cells were pelleted by centrifugation for 30 minutes at 2500 g and the precipitate was suspended in buffer TMG (50 mM Tris, pH 7.5; 5 mM $MgSO_4$; 10% Glycerol) to keep at -20°C. Cells with added lysozyme were incubated at room temperature (RT) and centrifuged 2500 g for 30 min. Cell pellet was collected and named P1. One volume of P1 was suspended in two volumes of 20 mM HEPES buffer pH 4.5 and incubated for 5 min at RT than 0.5 volume of 1M Tris pH 8 was added and the mix named MD was incubated for 5 min at 4°C. The mix MD was centrifuged for 10 min at 18 000 g and 4°C. The precipitate was named P2.

Purification of the recombinant protein was performed by metal-chelate chromatography (MAC) using His-tag. PAGE in the 12% or 16% polyacrylamide gels in the presence or absence of reducing agents was done accordingly described method [8]. Proteins were stained with Coomassie Brilliant Blue R250. Western Blot (WB) analysis was performed using reagents and PVDF membrane from BioRad and monoclonal antibodies against Tetra-His-Tag (Qiagen) or human insulin (Sigma-Aldrich) accordingly chemiluminiscence POD technology (www.roche.com).

Fraction MD or P2 were used for protein isolation. The protein CpIA was isolated by modified method as described [9] using 2% n-octyl-β-D-glucopyranosid (OG) or 2% N,N-Dimethyldodecylamine N-oxide (LDAO) with protease inhibitor by incubation in a thermomixer for 10 min at room temperature and 1 h at 4°C at 500 rpm. The incubation mix was centrifuged 10 min at 18000 g and 4°C. Supernatant of the lysate was centrifuged 40 min at 30 000 g and 4°C. The protein was diluted in SDS loading buffer. Commercial recombinant human insulin "Berlinsulin ® H Normal" (Berlin Chemie) was used as an insulin marker. Immuno-precipitation (IP) study was done accordingly manufacture guide of Sigma-Aldrich Protein G Immunoprecipitation Kit (PGIP). Protein concentration was measured by Bradford method.

RESULTS

The human hormone insulin is synthesized as a linear polypeptide preproinsulin (fig. 1) of 110 amino acid residues in the beta cells of the pancreas. During the maturation process the polypeptide sequence of preproinsulin and subsequently proinsulin is converted into active proteohormone composed of two subunits A and B, which are linked by two disulfide bridges. The third disulfide bridge forms a pair of cysteines Cys^{A6} and Cys^{A11} of subunit A [1-3]. All three covalent bonds are formed at the stage of proinsulin, and the connecting peptide C is involved in the formation of transition forms [1, 10], which, consecutively, provide the correct folding of the insulin molecule.

Expression vector

Our goal was to develop a method to express recombinant insulin in *E. coli*, using the intracellular mechanisms of molecular folding of polypeptides of the host cell to produce a human

protein with the correct disulfide bonds. An expression vector was constructed to synthesize a chimeric molecule CpIA (fig. 1) in *E. coli* cells. For molecular modeling of protein structure typically are used several software [11]. Biosynthesis of preproinsulin analog CpIA is coded by 0,5 Kbp gene *cpia*.

Recombinant protein length of 164 amino acid residues contains precursor (C') and C" peptide connecting the insulin A and insulin B subunits (fig. 2). The arrangement of the peptides A and B in CpIA differ from that in human preproinsulin (fig. 1). The both peptides C' and C" are truncated fragments of subunit c of bacterial ATP synthase (EMBL Seq. ID J01594) often named F_oc. These peptides provide creation of a hydrophilic loop in the cytoplasm the same as F_oc with formation of a hairpin like structure [12].

Fragment C" is a structure analog to the peptide C from human proinsulin, with nine identical amino acids (fig. 3). Peptide C" can be involved like its analog in protein folding and in formation of a covalent bond Cys^{A20}-Cys^{B19} (fig. 2).

Most part of polypeptides C' and C" like F_oc is represented by α-helices due to the dominant content of amino acid residues susceptible to the formation such kind of the secondary protein structure (fig. 2). Peptides C' and C" are responsible for providing the necessary orientation of the insulin subunits A and B, as well as to promote matching of SH-groups of suitable cysteine residues required for the formation of disulfide bonds (fig. 2). The N-flank of the C" polypeptide like connecting peptide C of proinsulin links to trypsin-digesting dipeptide, but in contrast to the analog its C-flank contains His-Tag built for subsequent protein purification by MAC.

The N-flank each of the A and B chains, fused into a polypeptide CpIA, starts with methionine. Processing CNBr, similar to the previously described method [5, 6] of producing a recombinant insulin, causing break of the peptide bonds between the C' and A, or between C" and B and gives dipeptide of 89 amino acids of common length (fig. 4).

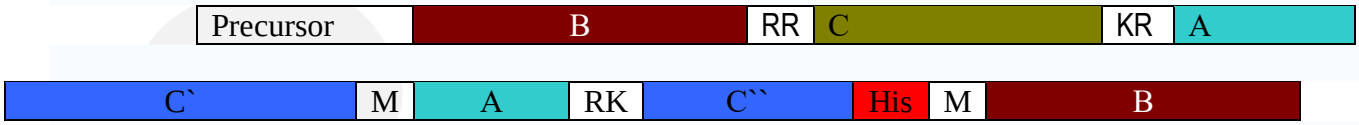
The constructed vector with inserted *cpia* gene was coexpressed with plasmid pAVD-10 (Lucigen) accordingly described method [13].

In other expression vector the termination flank of gene *cpia* is attached by truncated gene *uncF* from *unc* operon (EMBL GenBank ID J01594). The *E. coli* gene *uncF* codes the ATP synthase subunit b (F_ob). The F_ob is a membrane bound protein consisting of 156 amino acids including the single cysteine. The product of the truncated gene *uncF* was named here F_ob'.

F_ob' is the first half of the F_ob and contains 72 amino acids. To prevent formation of crosslinks between cysteins of the chimeric polypeptide CpIA and the subunit b its single cysteine was substituted with serine. Other substitution mutations were done in *uncF*: in start codon Ctg>Atg; restriction site AseI was removed. The complete chimeric gene about 0,7 Kbp was named *BsrG2* and

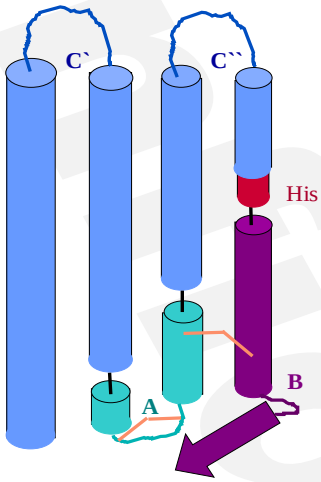
Биотехнология. Теория и практика. 2014, №2, стр. 28-35
DOI: 10.11134/btp.2.2014.4

digested by *Nde*I and *Xho*I to fuse with vector pET-21a (Novagen) to get an expression plasmid pBsrG2.



Linear model of preproinsulin and constructed chimeric polypeptide CpIA. A and B – insulin chains A and B; C – proinsulin connecting peptide; C` and C`` – truncated F_oc, His – HisTag; M – methionin; KR, RK and RR – connecting dipeptides

Fig. 1. Polypeptide model of the preproinsulin and CpIA



The polypeptide CpIA is about 16 KDa. A and B – insulin chains A and B, C' and C" – truncated F_oc, His – HisTag. Two Cys-Cys bonds are shown (orange) in the chimeric peptide

Fig. 2. Model of the constructed chimeric protein CpIA

Check of protein expression ability by the chimeric DNA construct

The *in silico* designed and chemically synthesized cDNA (the gene named *BsrG2*) is a fusion product of different genes (fig. 1) plus connected additional nucleotide sequence for truncated F_ob'.

Transcription and translation capability of the chimeric DNA designed *in silico* was at first proved using an *in vitro* protein translation system. The expression ability of the vector pBsrG2 was tested in two different "cell-free expression systems": EasyXpress® Protein Synthesis (Qiagen) and Purexpress® In Vitro Protein Synthesis Kit (NEB). Experiments were performed according to manufacturer protocols. Protein expression was detected by western blot. WB gives clear signal for CpIA with anti-Insulin antibodies (fig. 5). The CpIA was also detected in experiment with anti-His anti-bodies in blotting membrane, but together with other non-clear running bands.

Биотехнология. Теория и практика. 2014, №2, стр. 28-35
DOI: 10.11134/btp.2.2014.4

Figure 5 shows that the chimeric polypeptide CpIA comprising a membrane protein and insulin peptides could be synthesized in cell free expression system using a chimeric cDNA. Some vectors with *BsrG2* gene were transformed in a few host cells to achieve maximal expression (table 1).

```
>C":   RK-------GLAAIGAAIGIGILGGKFLEGAARQPDLIP
>PI-C:RREAEDLQVGQVELGGGPGAGSLQPLALEGSL-Q-----
       *          *     *    * * *     ***    *
```

Fig. 3. Alignment of the human proinsulin chain C (PI-C) and C" of the CpIA
A star indicates the identical amino acid

```
     1        10  14    20        30        40        50        61        72     79
EFoc MENLNMDLLYMAAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVGLGLYVMFAVA

CpIA MENLNMDLLYMAAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVGLGLYVMGIVEQC

     CTSICSLYQLENYCNRKGLAAIGAAIGIGILGGKFLEGAARQPDLIPHHHHNMFVNQHLCGSHLVEALYLVCGERGFFYTPKT
```

Fig. 4. Amino acids alignment of the EF_oc and CpIA
Shadows suggest the membrane helices of the F_oc and CpIA and red line shows eventually Cys-Cys bond

Investigation of chimeric peptide expression in bacterial cells

Recombinant *E. coli* clones were obtained after transformation of different host cells with plasmid vector pBsrG2 according to traditional procedures [7]. Optimal expression was achieved with *E.coli* c41 and c43 host cells (table 1). A c41 clone named TbciEc was selected for overproduction of the recombinant protein.

SDS-PAGE and WB analysis with anti-Insulin or anti-HisTag antibodies were used for detection of the chimeric polypeptide after expression in *E.coli* cells. Chimeric peptide CpIA was produced (fig. 6) in the bacterial cells TbciEc similar to the "cell free translation system" (fig. 5). Expression of the chimeric peptide in *E. coli* was a little induced with 1 mM IPTG. The recombinant protein production was higher at 38°C than at 25°C (fig. 7) with amount near described previously [9]. As expected from experiments with substitution of small amino acids in trans-membrane helix of EF_oc for tryptophan [14] which did not decline the expression level, the change in the proteolipid did not delay the CpIA synthesis.

It could further be demonstrated that coexpression of the chimeric peptide and the truncated subunit F_ob' enhances expression in the same order of magnitude as for the coexpression with full F_ob subunit as described previously [13].

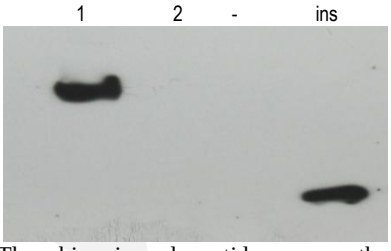
Examination of the expressed protein location

Биотехнология. Теория и практика. 2014, №2, стр. 28-35
DOI: 10.11134/btp.2.2014.4

For investigation of the CpIA intracellular setting the *E. coli* cells precipitate P2 (see methods) was suspended in buffer then sonificated. The mix was centrifuged 10 min at 18 000 g and 4°C. Supernatant and precipitate were used for PAGE and WB. Fig. 8 shows that CpIA was not find in supernatant or in cytosole fraction, but in the precipitate. Expressed chimeric peptide with fused insulin chains A and B forms the low spin vesicles in bacterial cells TbciEc which were similar to a membrane net described before for c41 and c43 host cells [13].

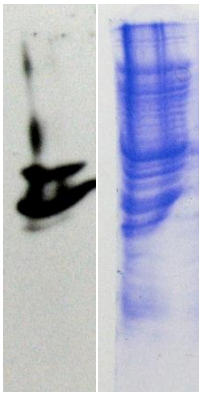
Control of Cys-Cys bond formation in the chimeric polypeptide

Formation of disulfide bonds in protein can be examined by PAGE. Reducing agents break up the S-S covalent bonds and the linearized protein moves more slowly compared with folded protein in non-reduced PAAG. To show arrangement of Cys-Cys bridges in the chimeric peptide the MD fraction was separated in two epis: one with 0,1 mM $CuCl_2$ and the other with 100 mM DTT. Sample with $CuCl_2$ was incubated 10 min at 4°C and with DTT at 65°C. The mix was centrifuged 10 min at 18 000 g and 4°C. Precipitates after $CuCl_2$ named PC (oxidized) and after DTT – Pr (reduced) were taken for non-reduced 12% SDS PAGE. Difference in polypeptide bands is shown in fig. 9: a part of the chimeric protein moves slowly, which shows that S-S bonds in that protein fraction were reduced after DTT treatment. It was shown such kind of a difference in electrophoretic movement for insulin folded in several levels [15] and for protein fused with insulin chain A with or without Cys-Cys bridges [16]. Investigation of disulfide linking in CpIA was done and data can be interpreted in the way that the chimeric polypeptide CpIA builds intracellular Cys-Cys bonds.



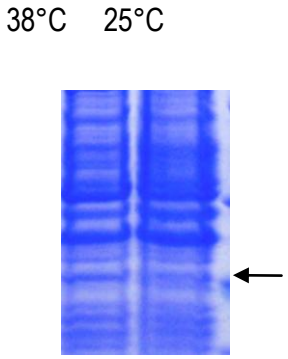
The chimeric polypeptide was synthesized in "cell free translation system" (See results). Anti-insulin antibody for detection of the chimeric CpIA in WB was used. 1 – CpIA; 2 –control; ins – insulin marker (commercial insulin "Berlinsulin® H Normal"). 12% SDS PAGE

Fig. 5. Synthesis of chimeric peptide CpIA with human insulin chain A and B



Expression of the CpIA in TbciEc clone. Left panel: WB with anti-Insulin antibodies. Right panel: 12% reduced PAAG stained with Coomassie Brilliant Blue R250 after protein blotting

Fig. 6. Expression of the recombinant chimeric protein CpIA in *E. coli* cells



Arrow shows the CpIA band

Fig. 7. Expression CpIA by 38°C and 25°C.

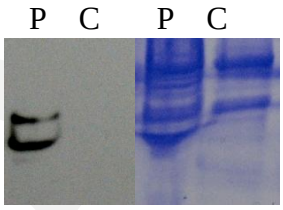
Биотехнология. Теория и практика. 2014, №2, стр. 28-35
DOI: 10.11134/btp.2.2014.4

Table 1. Velocity of cell duplication

Host cells	Duplication time, min	
	- IPTG	+IPTG
c41	22.4	63.3
c43	21.8	64.5
c41Lys	20.6	67.8
c43Lys	20.8	68.7
BL21	29.5	79.2
M15 [pREP4]	31.2	95.1
SG 13009	37.9	98.3

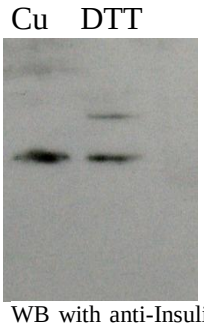
Time of duplication in min of different host cells transformed with vector containing gene *BsrG2*, cultivated at standard conditions with (+) or without (–) 1 mM IPTG. Host cells c41, c43, c41Lys and c43Lys from Lucigen, M15 [pREP4] and SG 13009 from Qiagen.

Биотехнология. Теория и практика. 2014, №2, стр. 28-35
DOI: 10.11134/btp.2.2014.4



12% SDS reduced PAAG. Left: WB with anti-Insulin anti bodies. Right: proteins stained with Coomassie Brilliant Blue R250 after blotting. P – membrane fraction; C – cytosol fraction

Fig. 8. Cell localization of the expressed chimeric peptide CpIA



WB with anti-Insulin antibodies. 12% nonreduced PAGE. Cells were incubated with $CuCl_2$ (Cu) or DTT

Fig. 9. Test of disulfide bonds formation in the chimeric peptide CpIA

Control of formation of the A^{20}-B^{19} disulfide bonding in the chimeric polypeptide

Cys^{A20} and Cys^{B19} build a link which connects C-flank α-helix of insulin chain A and the central α-helix of the chain B. Formation of Cys^{A20} -Cys^{B19} bond stimulates to develop a folding nucleus in insulin superfamily proteins [17].

As can be seen in Fig. 9 the chimeric protein CpIA builds some disulfide bonds. Proteolytic digestion was used to detect formation a Cys^{A20}-Cys^{B19} bond between peptide chains A and B. The largest fragment after trypsin treatment should be 107 amino acids in the presence of the Cys^{A20}-Cys^{B19} link (fig. 4) accordingly calculation of the size of CpIA potential digesting products. The Arg-22 of peptide B and connected dipeptide RK (fig. 4) lie in the trans-membrane domain and should not be affected by peptidase. After reduction of the Cys-Cys bonds the peptides A and B are not further linked resulting in a smaller than non-reduced peptide digestion pattern (fig. 4).

To investigate the Cys^{A20}-Cys^{B19} covalent bond both precipitates PC and Pr were washed in buffer TdB (20mM Tris, pH 8, 5mM $MgCl_2$, 150 mM KCl) and suspended in the same buffer. Trypsin solution was added in ratio 1:500 of total protein. The enzymatic treatment continued 60 min at room temperature. In each slot of the 12% SDS PAAG all samples were loaded with equal amount of protein. WB analysis with anti-insulin antibodies confirmed prevailing time dependent accumulation peptides about 11 KDa in oxidized sample (without DTT treatment). Very low amount of a peptide near 11 KDa was present in the DTT treated sample, resulting from the reduced chimeric peptide after digestion with trypsin (fig. 10). Anti-His antibodies gave signal for bands the same size.

This experiment demonstrates that the chimeric protein CpIA digested with trypsin, gives the expected restriction fragments and preferably forms a heterodimer containing insulin chain A and B including intra-chain Cys-Cys bond.

Биотехнология. Теория и практика. 2014, №2, стр. 28-35
DOI: 10.11134/btp.2.2014.4

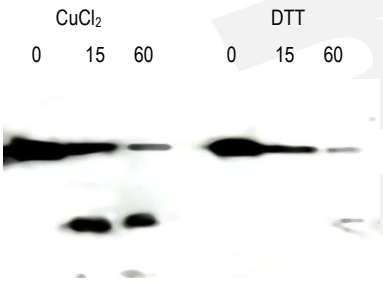
Control of the chimeric polypeptide semifolding

Proinsulin after its folding *in vitro* is processed by trypsin digestion [4, 6]. CpIA needs additional digest with CNBr. After treatment trypsin plus CNBr the semi-folded CpIA should give the biggest digesting product about 6 KDa (fig. 4) if Cys^{A20} matches Cys^{B19}. It is very critical the covalent bond would not break after protein solubilization. To confirm Cys-Cys bond present in semi-folded conformation after protein isolation the chimeric polypeptide CpIA was solubilized from PC and Pr fractions by modified method [9] and digested in two subsequently steps. The precipitate PC or Pr was resuspended in 1 mL buffer TEP: 20mM Tris pH 8, 1mM EDTA, 0,002% PMSF, 0,2% LDAO, incubated 10 min at 4°C, centrifuged 10 min at 18 000g. Lysis buffer: 20mM Tris pH 8, 0,001% PMSF, 2% LDAO was added to the precipitate.

The mix was incubated in a thermomixer for 10 min at room temperature and 1 h at 4°C at 500 rpm. The incubation mix was centrifuged 10 min at 18000 g and 4°C. Supernatant was centrifuged 40 min at 30 000 g and 4°C. Clear lysate was used for PIGP. Trypsin was added to protein in ratio 1:1000 and BrCN in concentration 60 mM per 10 mg protein to IP mix.

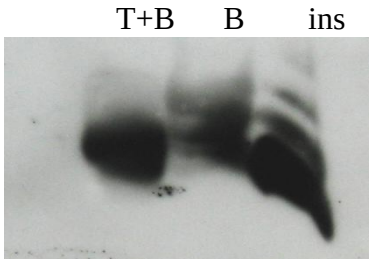
These experiments as described thus proved the stable and correct disulfide bond formation in the chimeric peptide CpIA (fig. 11).

As a consequence, the semi-folded recombinant polypeptide can be easily processed by used methods [1, 17] into the correct 3D conformation.



WB of the oxidized ($CuCl_2$) or reduced (DTT) chimeric protein. 12% SDS PAAG anti-Insulin antibodies. 0, 15, 60 – time of trypsin digestion in min

Fig. 10. Check Cys^{A20}-Cys^{B19} disulfide bond formation in the chimeric protein



16% SDS PAAG. WB with anti-insulin anti bodies. B – after digestion with BrCN. T+B –digestion products after BrCN+ trypsin, ins-commercial insulin "Berlinsulin® H Normal".

Fig. 11. Control of the chimeric peptide semifolding

Discussion of results

The hormone insulin regulates glucose metabolism in animal cells [2-3]. Humans with high glucose concentration in blood suffering from metabolic disorder named diabetes mellitus. Insulin has to be administered to patients with an inadequate supply of the hormone. Due to the constant

increasing of the number of insulin-dependent patients, the need for the recombinant human insulin rises significantly. Experts predict that in 2015 should be produced about 30 tons of recombinant insulin, which is five times more than in 2006. Therefore, development of new methods, to improve the quality and reduce the production costs of recombinant human hormone is extremely essential. Refolding of proteins is very expensive step in biopharmaceutical production. We have developed a new method without chemical refolding.

Accordingly our method the recombinant insulin can be obtained in *E.coli* cells as semi-folded proinsulin analog and isolated in transit-like form with the native disulfide bonds (fig. 2, 11) which are important for finally 3D structure of the proteohormone molecule. As the folding-helper was used the subunit F_oc from *E. coli* (fig. 1-2, 4), termed here EF_oc, it is named as well as proteolipid due its holding disproportionate high amount of hydrophobic amino acids (fig. 4).

Expression of the recombinant membrane proteins as membrane complex can avoid chemical refolding, but their yields are regularly too low. A rational strategy should be designed and optimal host find to increase the intracellular yield of the protein of interest. EF_oc is a suitable host for the expression of membrane or membrane-bound proteins for several reasons, which are discussed below.

Two antiparallel α-helices connecting by polar loop gives a hairpin form of the proteolipid proposed from nuclear magnetic resonance (NMR) analysis of EF_oc [12]. It is supposed 9-14 proteolipids build a membrane-traversing ring in F_o sector of different ATP synthases [8-20]. Enzyme ATP synthase (F_oF_1) consists from catalytic sector (F_1) and membrane-bound F_o, it catalysis intracellular ATP synthesis. Biosynthesis of ATP synthase subunits in *E.coli* is coded by polycistronic operon *unc* (EMBL J01594). A single gene *uncE* codes EF_oc subunit, which contains no cysteine. Different proteins participate in the proteolipid transport and membrane insertion [21].

The extramembranous loop surface of subunit F_oc exposed to the water phase of the cytoplasm is expanded through hydrophilic part L31-Q42 (Fig. 4). The loop may be mainly α-helical with a few randomly organized residues around proline-43, as deduced from NMR analysis of subunit EF_oc [12]. Genetically introduced two cysteins in EF_oc help to build functional dimers and multimers with Cys-Cys cross-links: A14C/V72C, A21C/I66C, A21C/M65C and M11C/M75C [18]. The chimeric protein CpIA like the EF_oc host-dimer and native proteolipid dimer analog [22] forms disulfide bond Cys^{A20}-Cys^{B19} (fig. 2, 4), because the introduced Cys^{A20} and Cys^{B19} of insulin lie correspondingly in positions alanine-14 (A14) and leucine-72 (L72) of EF_oc, which due α-helix of both strands of the hairpin able to form a S-S covalent bond [18]. EF_oc mutants with single substitution A14/C did not form and with V72/C accumulated very little homodimers [18]. So possibility to form of CpIA-homodimers due for pairing Cys^{A20}-Cys^{A20} or Cys^{B19}- Cys^{B19} is very low.

Consequently, fragments EF_oc help in the disulfide Cys^{A20}-Cys^{B19} pairing required for active insulin molecule (fig. 2, 10). This study has shown that the necessary disulfide bridges in the recombinant molecule CpIA, were formed in *E.coli* cells and kept during processing and after isolation of this protein (Fig. 11). This covalent bond is very important because it stabilizes the disulfide intermediate in proinsulin and in other proteins which belong to insulin family [17].

Linker of a few amino acid residues connecting the two subunits EF_oc generates a turn in dimer [8]. The EF_oc dimer in size (about 16 KDa) and structure is comparable with trans-membrane lipoproteid C from VoV1 type ATPase [22]. Its longer linker sequence (10 or 13 amino acids) or linker in fused EF_oc [8] gives greater flexibility than shorter but even tetra-peptide Gly-Ser-Ala-Gly (GSAG) is also functional in turn formation. Moreover one or more genetically introduced linkers GSAG, connecting the C-and N-flank of two proteolipids may give FoF1 enzyme with transmembrane complex containing from two till 14 hairpin repeating chains F_oc [19] and all type of ATP synthases were catalytic active. Tetropeptide TSIC of insulin chain A makes a turn [1] like linkers oligopeptides GSAG or GSLN in F_oc dimer [8] because connects two proteolipids in the same position. This TSIC turn locates in insulin intra-chain A loop which allows to form Cys^{A6}-Cys^{A11} bond (fig. 2) in periplasm area. Experimentally was shown that the intra-A chain Cys^{A6}-Cys^{A11} bond develops first and straightforward due the amino acid sequence during insulin precursor folding [15]. Formation of Cys-Cys bridge in insulin chain A was shown in soluble proinsulin transit form [Weiss et al., 2000] and in insulin chain A fused with subunit γ of ATP synthase [16]. Given experimental data illustrate that formation of the Cys^{A6}-Cys^{A11} bond in the conformational micro-domain of the CpIA (fig. 2) due proximity between sulfhydryls of Cys^{A6} and Cys^{A11} residues of intra-chain A loop is very possible (fig.10).

Protein of the interest was cut out from the host polypeptide by CNBr treatment, since both A and B subunits in CpIA start with methionine (fig. 4). His-Tag was inserted for the MAC purification and thorough washing of the bound to the matrix protein from the toxic CNBr. Tetra His-Tag locates before residue cF54 up the second trans-membrane domain (fig. 2, 4). Unfortunately, the His-Tag was hidden in soluble by LDAO or OG native CpIA, but after Gu/HCl solubilization the reduced protein binds to MAC and can be eluted from column.

For industrial production an expression system should produce sufficient quantity of recombinant protein. Post-transcriptional control of the polycistronic *unc* operon expression was investigated and shown that production of subunits was in ratio a: 2b: 10 (12) c of the ATP synthase F_o sector due high mRNA stability of the *uncE* gene [23]. Subunit F_oc was overexpressed in *unc* knock-out or other host *E.coli* cells [13] for various purposes. The recombinant cDNA as a product of two linked *uncE* genes produces also stable mRNA like the *uncE* or *Vma3* [22-23]. Expression of the membrane-bound EF_oc

was especially increased by coexpression with F_ob in host cells c41 or c43 [13]. The F_ob subunit is a membrane bound protein. It forms a stalk which can be divided into functional domains called: the membrane domain (bM1-I33), the tether domain (bE34-A61), the dimerization domain (bT62-K122), and the F_1-binding domain (bQ123-L156) described by Revington [24]. Two subunits F_ob build a dimer in enzyme complex EF_oF_1. A part of the N-terminal flank of F_ob locates in the plasmatic membrane near F_oc and the C-terminal flank binds to the catalytic sector of the ATP synthase.

Overexpressed F_ob polypeptides generate intracellular heavy membrane complexes, but non one of truncated 1-25, 1-34, 1-48 and 25-156 fragments led to formation of a large membrane network [13].

The chimeric polypeptide CpIA was coexpressed with subunit F_ob or its 72 amino acids N-flank fragment F_ob'. The recombinant F_ob' was effective in CpIA overproduction (Fig. 6) and subsequently in insulin expression as F_ob (data not shown). Perhaps the both dimerization and the F_1-binding domains are less required as the membrane and the tether domains [24] present in the truncated F_ob' to enhance synthesis of the recombinant membrane proteins.

The protein F_oc was steadily expressed in the *E.coli* cells and recombinant expression of its gene was greatly increased by coexpression with subunit F_ob [9, 13]. EF_oc dimers were expressed on the same level as their monomers [8]. Recombinant protein CpIA like the host F_oc polypeptide coexpressed with F_ob or F_ob' forms heavy membrane complexes (fig. 8), may be a membrane network similar to the one observed after coexpression F_oc with F_ob [13]. Formation of disulfide bridges between CpIA and subunit b' excluded, because the single cysteine in F_ob' was replaced by serine.

Folding-helper F_oc assists correct Cys^{A20}-Cys^{B19} cross-linking. It gives additionally ability to control structure of the polypeptide before protein isolation (fig. 10).

Cys^{B7} is in position Val-60 of the EF_oc and remains in the transmembrane helices of the chimeric polypeptide (Fig. 4), it is not able to form a disulfide bond with other cysteine residues of the CpIA, preventing the formation of the isomeric forms observed previously [10]. That is, the described method is different from the previous known ones in that it allows the recombinant precursor to create and control the correct disulfide bonds arrangement.

Chimeric recombinant protein CpIA was isolated from the host cells in a transit form similar to one described by Weiss [1, 10] with two disulfide bonds Cys^{A6}- Cys^{A11} and Cys^{A20} - Cys^{B19}, as shown in the model (fig. 2). The isolated semi-folded protein can finally fold under suitable conditions as described [1, 10] because the insulin A and B peptides hold sufficient structural information to form the native molecule [25].

The data indicate that the designed technology of gene-engineered bacterial system provides the necessary intracellular formation of disulfide bonds in the recombinant human protein (fig. 2, 10-11). New bacterial expression system provides production of comparatively high amount of recombinant human protein, presumably in one of the transit form. Biosynthesized semi-folded polypeptide has a structure that does not require a chemical treatment of molecules *in vitro* for subsequent folding. The method helps to avoid isoform accumulation and gives possibility for intermediate structure check. The advantage of the described here method for biopharma is that it takes less time and costs to get the folded recombinant protein than the currently used technologies.

Authors have no conflicts.

Acknowledgements: We thank CEO P.-P. Schikora for support of the project, Professors G. Groth (Heinrich Heine University Düsseldorf) and A. Pfützner (ikfe) for critical reading of the manuscript. The project was supported by ILB and ZAB (Brandenburg State). We thank Dr. M. Fuhrman for assistance in some WB experiments.

REFERENCES

1. Weiss M.A. Proinsulin and the Genetics of Diabetes Mellitus// J. Biol. Chem. -2009.- V. 284. N 29. – P. 19159–19163.
2. Ward C.W., Lawrence M.C. Landmarks in insulin research// Front.Endocrin. -2011.- V. 2. N 76. http://dx.doi.org/10.3389/fendo.2011.00076
3. Hoeg-Jensen T. Design of insulin variants for improved treatment of diabetes// In: Peptide and Protein Design for Biopharmaceutical Applications Editor Jensen K.J.-2009.- Chapter 7.- P. 249-286.
4. Kayser O. Grundwissen Pharmazeutische Biotechnologie// Verlag Teubner. Stuttgart. – 2002. 256 S.
5. Rudolph R., Lilie H., Schwarz E. In vitro Folding of Inclusion Body Proteins on an Industrial Scale.// In: Biothechnology. VCHVol. 5a Recombinant Proteins, Monoclonal Antibodies, and Therapeutic Genes. Vol. Eds. Mountain A., Ney U.M., Schomburg D.- 2000.- P. 111-123.
6. Petrides D., Sapidou E., Calandranis J. Computer-aided process analysis and economic evaluation for biosynthetic human insulin production – A case study// Biotech.Bioengin.- 1995.-V. 48. N 5.- P. 529-541.
7. Sambrook J., Fritsch E. F. Maniatis T. Molecular Cloning: a Laboratory Manual// 2nd edn.- Cold Spring Harbour Press, Cold Spring Harbour. NY.- 1989.
8. Jones P.C., Fillingame R.H. Genetic Fusions of Subunit *c* in the F0 Sector of H1-transporting ATP Synthase// J.Biol.Chem.-1998.- V. 273. N 45.- P. 29701–29705
9. Arechaga I., Butler P.G.G., Walker J.E. Self-assembly of ATP synthase subunit c rings// FEBS Letters.- 2002.- V.515.- P. 189-193
10. Qiao Z.-S., Min C.-Y., Hua Q.-X., Weiss M.A., Feng Y.-M. In Vitro Refolding of Human Proinsulin. Kinetic intermediates, putative disulfide-forming pathway, folding initiation site, and potential role of C-peptide in folding process// J.Biol.Chem. – 2003. - V. 278.-P. 17800-17809

11. Tramontano A. Protein Structure Prediction.// WILEY-VCH GmbH&Co KGaA Weinheim.- 2006. 208 p.
12. Girvin M.E., Rastogi V.K., Abildgaard F., Markley J.L., Fillingame R.H. Solution Structure of the Transmembrane H+-Transporting Subunit c of the F1Fo ATP Synthase// Biochemistry.- 1998.- V. 37.- N. 25.- P. 8817-9924
13. Schnick C., Forrest L.R., Sansom M.S., Groth G. Molecular contacts in the transmembrane c-subunit oligomer of F-ATPases identified by tryptophan substitution mutagenesis// Biochim Biophys Acta.-2000.- V.1459. N.1.-P.49-60..
14. Arechaga I., Miroux B., Karrasch S., Huijbregts R., de Kruijff B., Runswick M.J., Walker J.E. Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale over-production of the b subunit of F1Fo ATPase// FEBS Letters.- 2000.- V. 482.- P. 215-219
15. Yuan Y., Wang Z.-H., Tang J.-G. Intra-A chain disulphide bond forms first during insulin precursor folding// Biochem. J. -1999.-V. 343.-P. 139-144.
16. Ponomarenko S.V. Biochemical characteristics of Escherichia coli ATP synthase with insulin peptide A fused to the globular part of the gamma-subunit. Biochemistry (Mosc). - 2006.- V.71.-P. 1006-1012
17. Hua Q.-X., Mayer J.P., Wenhua J., Zhang J., Weiss M.A. The Folding Nucleus of the Insulin Superfamily// J.Biol.Chem. – 2006.-V. 281. N. 38.- P. 28131-28142
18. Jones P.C., Jiang W., Fillingame R.H. Arrangement of the Multicopy H1-translocating Subunit c in the Membrane Sector of the *Escherichia coli* F1F0 ATP Synthase// J.Biol.Chem. –1998.- V. 273.- N. 27.-P. 17178–17185
19. Mitome N., Suzuki T., Hayashi S., Yoshida M. Thermophilic ATP synthase has a decamer c-ring: Indication of noninteger 10:3 H^+/ATP ratio and permissive elastic coupling// PNAS.- 2004.- V. 101. N. 33.- P. 12159-12164
20. Hakulinen J.K., Klyszejko A.L., Hoffmann J., Eckhardt-Strelau L., Brutschy B., Vonck J., Meier T. Structural study on the architecture of the bacterial ATP synthase Fo motor// PNAS.- 2012. http://dx.doi.org /10.1073/pnas.1203971109
21. Revington M., McLachlin D.T., Shaw G.S., and Dunn S.D. The Dimerization Domain of the *b* Subunit of the *Escherichia coli* F1F0-ATPase // J.Biol.Chem. – 1999.- V. 274.- N. 43.- P. 31094–31101
22. Suzuki T., Ozaki Y., Sone N., Feniouk B., Yoshida M. The product of uncI gene in F1Fo-ATP synthase operon plays a chaperone-like role to assist c-ring assembly// PNAS.- 2007.- V. 104.- N. 52.- P. 20776-20781
23. Saroussi S., Nelson, N. The little we know on the structure and machinery of V-ATPase// J. Experimental Biology.- 2009.- V. 212.-P. 1604-1610. http://dx.doi.org/10.1242/jeb.025866
24. McCarty J.E.G._Expression of the *unc* Genes in *Escherichia coli*// J Bioenergetics Biomembranes.- 1988.- V. 20.- N. 1.- P. 19-39
25. Wang C.C., Tsou C.L. The insulin A and B chains contain sufficient structural information to form the native molecule.// *Trends Biochem Sci*.- 1991.- V. 16.- P. 279– 281.

ТҮЙІН

Жануарлардың жасушаларында түзілген полипептидтердің молекулалық фолдингінің тетігі адамның рекомбинантты ақуыздарын экспрессиялау үшін биофармацевтикалық өнеркәсіпте дәстүрлі түрде пайдаланылатын прокариоттар мен ашытқыларда белгіленгеннен принципті түрде ерекшеленеді. Бактериялық жүйеде экспрессияланған рекомбинантты ақуыздар дұрыс конформация қалыптастыру үшін *in vitro* химиялық рефолдингіне тартылады. Дисульфидтік кішкене көпірлер жасаудың полипептидтерді үш өлшемді құрылымға молекулалық орналастыру процесі үшін, ақыр аяғында ақуыздың оның биологиялық белсенділігі үшін қажетті табиғи конформациясын құру үшін шешуші маңызы бар.

Мұнда бактериялардың жасушаларындағы ақуызды экспрессиялаудың жаңа технологиясы сипатталады, ол рекомбинантты полипептидтерді молекулалық орналастыру үшін прокариоттық жүйені пайдаланады және дұрыс дисульфидтік байланыстардың пайда болуын қамтамасыз етеді. Осы технологияға сәйкес, адам инсулині – маңызды терапиялық ақуыз – препроинсулиннің бактерияда түзілген рекомбинантты аналогынан химиялық рефолдингсіз алынуы мүмкін. Бұл ие-жасушада цистиннің дұрыс тиолды байланыстарымен полипептидті құрылымды құру жолымен іске асырылады, олар проинсулиннің белгілі транзитті нысанындағы дисульфидтік кішкене көпірлерге ұқсас.

Биотехнология. Теория и практика. 2014, №2, стр. 28-35
DOI: 10.11134/btp.2.2014.4

Жаңа технологияға сәйкес рекомбинантты ақуыз *Escherichia coli* жасушаларындағы жарғақшалы ақуыздың әртекті түзілісі үшін барынша жоғары шоғырланымда бейнеленеді. Бұл ретте ішінара салынған протеопептид протеолиттік деградациядан қорғалған және проинсулиннің транзитті нысанына ұқсас кешен ретінде бөлініп шығарылуы мүмкін. Проинсулиннің транзитті нысандары гетеро-димердің молекулаларын одан әрі протеогормонның табиғи құрылымына айналдыру үшін қажет. Химиялық рефолдингсіз жаңа әдіс изоформалардың жинақталуын болғызбайды, адамның рекомбинантты инсулинін оның табиғи конформациясында өндіруге жұмсалатын уақытты және материалдық шығындарды айтарлықтай қысқартады.

Кілтті сөздер: рекомбинантты инсулин, препроинсулин, рефолдинг, жарғақшалы ақуыздар, АТФ-синтез.