ABSTRACT

of the dissertation for the philosophical doctor degree (PhD)
on specialty 6D070100 - "Biotechnology"

Izat Smekenov

Cloning and optimization of cellulases gene expression in industrial strains of *Saccharomyces cerevisiae*

General characteristics of the work. The dissertation work is devoted to the creation of recombinant industrial yeast strains using recombinant DNA technology and methods for cloning genes that can efficiently ferment cellulose to ethanol.

Relevance of the research topic.

Today, all fuel ethanol is produced biotechnologically by fermentation of sugars (sugarcane), or starch-containing raw materials (mainly corn). In the context of the threat of the global food crisis, this strategy is unacceptable because it leads to more build-up of food shortages and inefficient use of agricultural land. However, even in the United States, where corn is very cheap to maintain the profitability of fuel ethanol production requires government subsidies, and in order to ethanol could create price competition to petroleum products is necessary to further significant reduction in prices of bioethanol.

The main ways to reduce the cost of this product can be the replacement of raw materials for its production and a radical change in the technology of alcoholic fermentation. The replacement of raw materials is to use the renewable biomass of whole plants, both herbaceous and trees, including agricultural wastes (wheat straw, rice, etc.) instead of cereal grain to turn into ethanol.

Grain farming is the main branch of agriculture in Kazakhstan. Kazakhstan is one of the largest grain-producing countries in the world. In recent years, total grain crops occupy more than 80% of the sown area of crops. Cereal straw is the most important renewable energy source in the Republic of Kazakhstan. For example in 1990, production of straw was almost 37 million tons (http://www.kazee.kz). In present time most of the straw crops are burned directly in the fields. Burning of wheat straw results in large amounts air pollutants including particulate matter (PM10), CO and NO_2. Thus, finding an alternative way for disposal of surplus wheat straw is of high interest and an immediate necessity.

Changing technology alcoholic fermentation technology application expects consolidated bioprocess hydrolysis and fermentation (CBP), i.e. direct fermentation of cellulosic substrates into ethanol. The ideal solution in this matter would be to use microorganisms capable of simultaneously hydrolyzing biomass (cellulose-containing), and fermenting the resulting sugars to ethanol. However, such microorganisms having the above properties do not exist in nature.

Cellulose is the most abundant biological polymer and has a linear structure in which unbranched chains of β-glucose residues are linked by β-1,4-glucosidic bonds. The hydrolysis of cellulose with the formation of soluble sugars is carried out by a combination of the three main types of cellulases: (I) endo-1,4-β-glucanase (EC 3.2.1.4), which cleaves the internal β-1,4-glycosidic bonds; (II) exo-1,4-β-glucanases (EC 3.2.1.91), which cleave cellobiose from the ends of the polymer molecules of

native or partially hydrolyzed cellulose; (III) 1,4-β-glucosidase (EC 3.2.1.21), which breaks down cellobiose to glucose.

These enzymes are known as cellulases, which exhibit synergism for the complete hydrolysis of cellulose to soluble oligomeric and monomeric sugars. The key enzymes of the cellulase complex responsible for the deep hydrolysis of crystalline cellulose are exo-1,4-β-glucanases or cellobiohydrolases, the main product of which is cellobiose. Most of them belong to the family of 6 and 7 glycosyl hydrolases. Currently, most commercial cellulase preparations are produced using fungi of the genus *Trichoderma* and / or *Aspergillus*.

To date, the most widely used organism in ethanol production processes is *Saccharomyces cerevisiae* yeast. The use of *S. cerevisiae* yeast, also known as "baker's yeast", for the preparation of ethanol from hexose sugars is rooted in prehistoric times. Several microorganisms were examined for the production of ethanol from lignocellulose, including both bacteria and yeast. The yeast *S. cerevisiae* proved to be more reliable than bacteria with respect to tolerance to the final product ethanol and other compounds present in hydrolysates. In addition, *S. cerevisiae* yeast is included in the list of organisms "generally recognized as safe" (GRAS).

However, natural *S. cerevisiae* yeast does not contain 1,4-β-glycosyl hydrolase (cellulase) genes; therefore, they are not able to ferment cellulose contained in plant cell walls. But because of the high ability to ferment glucose to ethanol and withstand high ethanol in the medium, *S. cerevisiae* yeast is the most suitable candidate for the bioconversion of cellulose-containing biomass into biofuel. In recent years, efforts in the field of genetic engineering related to the creation of yeast, in particular *S. cerevisiae* capable of fermenting cellulose, have focused on the cloning of genes encoding cellulases. In this regard, fungal cellulase genes were cloned and expressed in strains of *S. cerevisiae*. In addition, a recombinant strain of *S. cerevisiae* was created that coexpresses three cellulite enzymes (exo-1,4-β-glucanase, endo-1,4-β-glucanase and 1,4-β-glucosidase) and is able to convert amorphous cellulose to ethanol . However, in all of the above works, episomal or plasmid vectors were used as expression vectors. It is known that yeast cells transformed with multicopy autonomous YEp vectors contain up to 40 copies of plasmids per cell. The continuous expression of a large number of foreign genes in *S. cerevisiae* can lead to depletion of energy resources and impaired cell metabolism. However, YEp vectors in absence of selective pressure are difficult to stably maintain within the cells, and they are often lost from the transformed cells during culture growth under non-selective conditions. From a practical point of view, for the stability of the protein, as well as to prevent the loss of the recombinant gene in the absence of a selective factor, it is more suitable to use integral vectors that ensure the introduction of genes into the yeast chromosome.

Yamada et al. used long terminal repeats of Ty elements, known as δ sequences, which are good targets for integrating genes by homologous recombination, since approximately 425 copies of this sequence are contained throughout the yeast genome. In this regard, δ integration vectors were used to stably increase the copy number of the target gene in *S. cerevisiae*. δ Integration vectors have been used to express human nerve growth factor, anticoagulant hirudin, glucoamylase. However, all of the above integration vectors contained auxotrophic marker genes such as LEU2-D, HIS3, TRP1

and URA3. The described approaches force one to work with a narrow circle of well characterized mutant yeast strains. Moreover, these yeast strains might not be suitable for industrial applications, since the industrial yeast strains do not contain internal genetic markers, such as the need for amino acids or nucleic acids. Much of this work has been done using laboratory strains due to their ease of genetic manipulation, but unlike industrial strains, they have not been selected for producing ethanol on a large scale.

Industrial strains have slightly more pronounced features necessary for the production of ethanol, such as a high growth rate, fermentation and production of ethanol, as well as resistance to high concentrations of ethanol and high temperature. However, during hydrolysis of biomass, lignocellulosic material often decomposes to growth inhibitory compounds such as 2-furaldehyde (furfural), weak acids, and phenolic compounds. These inhibitors are toxic to microbial metabolism and inhibit fermentation. Several studies have been devoted to the selection of industrial strains of *S. cerevisiae* yeast that are resistant to inhibitory compounds and capable of fermenting lignocellulosic sugars. However, resistance to inhibitory compounds can vary widely depending on yeast strains. Therefore, in order to convert sugars derived from biomass into ethanol, it is important to select yeast strains with an industrial background and high resistance to inhibitors. These naturally resistant strains can serve as a platform for the construction of recombinant strains capable of using cellulose as a carbon source for ethanol production.

In this regard, the creation of new recombinant industrial strains of microorganisms producing cellulases resistant to stress factors, as well as studying the properties of cellulolytic and related proteins, is a task of great scientific and practical importance.

Purpose of research. Creation of recombinant industrial yeast strains containing introduced genes encoding endo-1,4-β-glucanase, exo-1,4-β-glucanases, 1,4-β-glucosidase and a membrane cellodextrin transporter to efficiently ferment cellulose to ethanol.

To achieve this goal, the following tasks were formulated:

1. Comparative analysis of *Aspergillius niger* endo-1,4-β-glucanase gene expression in different expression vector systems;
2. Creation of industrial yeast strains *S. cerevisiae* with the *Thermoascus aurantiacus* β-glucosidase gene and evaluation of their ability to ferment cellobiose into ethanol;
3. Analysis of the glycosylation nature of TaBGL1 synthesized by various recombinant strains of *S. cerevisiae*;
4. Assessment of recombinant yeast strains *S. cerevisiae* resistance to individual stress factors;
5. Creation of an industrial yeast strain *S. cerevisiae* coexpressing *T. aurantiacus* β-glucosidase and *Neurospora crassa* membrane cellodextrin transporter, as well as analysis of its ability to grow and produce ethanol in a liquid medium with cellobiosis;
6. Creation of an industrial strain of *S. cerevisiae* stably coexpressing the *T. aurantiacus* β-glucosidase, *Lentinula edodes* exo-1,4-β-glucanases and *A. niger* endo-1,4-β-glucanase genes;

7. Investigation of the ability of a constructed industrial yeast strain *S. cerevisiae* to ferment cellulose to ethanol under anaerobic conditions.

Objects of research. Industrial yeast strains. Strain ATCC 24860 was obtained from the American Type Culture Collection (ATCC), strains YB-2625, Y-1528 and Y-2034 were obtained from the ARS culture collection at the National Center for Agricultural Use Research, Peoria, Illinois. The chosen strains of *S. cerevisiae* are widely used for the production of fuel and drinking ethanol based on starch-containing raw materials, and possess important features necessary for industrial use.The Y-1528 yeast strain has previously been recommended for fermentation of softwood hydrolysates owing to the rare ability, among yeasts, to co-ferment glucose and galactose. It has been shown that the Y-2034 yeast strain can ferment different sugars commonly present in various raw materials of food waste and ethanol fermentation. The other important factor for the selection of a potential host strain is the presence of an active and efficient pentose phosphate pathway linking the introduced xylose-to-xylulose pathway to glycolysis. *S. cerevisiae* ATCC 24860 and YB-2625 are supposed to the most efficient xylulose fermenters

Subject of study. Cloning and expression of cellulase genes in industrial yeast strains.

Methods of research. In the dissertation work, methods of molecular cloning, microbiology and biochemistry with some modifications indicated in the text of the work are used.

Scientific novelty of research

An integral vector was constructed for integration into the yeast HO locus containing the *T. aurantiacus* 1,4-β-glucosidase gene (*tabgl1*) with the yeast α-factor signal peptide under the constitutive GAPDH promoter and terminator control, and the geneticin resistance gene (G418). Four industrial recombinant strains of *S. cerevisiae* were first created for the fermentation of cellobiose after stable integration of the *tabgl1* gene encoding β-glucosidase from *Thermoascus aurantiacus* (TaBGL1). The obtained recombinant strains of *S. cerevisiae* secrete TaBGL1, which can hydrolyze cellobiose to glucose, followed by its fermentation to ethanol. It was found that the degree of glycosylation of secreted TaBGL1 depends on the yeast strains used and largely depends on carbon sources (cellobiose or glucose). Recombinant yeast strains showed high osmotolerance and resistance to various concentrations of ethanol, furfural and H_2O_2, as well as to high temperatures.

An integral vector was constructed for integration into the yeast HO locus, containing the gene for *T. aurantiacus* 1,4-β-glucosidase without the signal peptide of the yeast α-factor and the *Neurospora crassa* membrane cellodextrin transporter gene (*cdt1*) with green fluorescent protein (eGFP) under strong constitutive TEF1 and PGK1 promoters, and the gene for resistance to geneticin (G418). The ability of a recombinant strain to grow in a medium with cellobiose as the sole source of hydrocarbons is shown. The resulting recombinant strains of *S. cerevisiae* are able to convert cellobiose to ethanol due to the *cdt1* and *tabgl1* genes integrated into the chromosome.

An integral vector was constructed for integration into the δ-sequence of the yeast genome, carrying the *A. niger* endo-1,4-β-glucanase gene and two dominant exo-1,4-β-glucanase genes of *L. edodes* (*cel7A* and *cel6B*) with the marker gene for phleomycin

resistance (*sh ble*).

For the first time, a new stable yeast strain with the cellulite enzymes genes integrated into the yeast chromosomes was created. This strain provides continuous expression of cellulite enzyme genes and secretion of expression products into the environment. It was found that a recombinant strain of *S. cerevisiae* expressing cellulase genes is able to grow in a synthetic medium containing avicel or carboxymethyl cellulose (CMC) as the sole carbon source. The recombinant strain produced 5.62 g/L ethanol (22% of the theoretical maximum yield) using 5% Avicel as a substrate, while from 5% pre-processed wheat straw it produced 1.43 g/L ethanol (14% of the theoretical maximum yield).

Scientific and practical significance of the work.

Characterization and optimization of gene expression of thermostable cellulolytic fungi in yeast systems are of great theoretical importance for understanding the patterns of gene expression regulation. The practical work significance is due to the prospects of using recombinant yeast strains - carriers of cellulase genes, for the production of ethanol from renewable and cheap sources of cellulose (cereal straw, paper waste, sawdust and others).

Basic provisions for the defence:

- The expression of A. niger endo-1,4-β-glucanase in the episomal plasmid vector YEGAp significantly slows down the growth of yeast cell culture. Stable gene expression of this enzyme at the HO locus of the yeast chromosome does not cause growth inhibition, and the activity of the secreted enzyme stably remains during the entire incubation time (880 units).

- Recombinant strains expressing TaBGL1 grow well on cellobiose medium and consume all available sugar. The average ethanol yield from cellobiose is 7.6 g/L (77% of the theoretical maximum).

- The secretion efficiency and glycosylation pattern of TaBGL1 depends on the genetic background of the initial yeast strain and on the carbon source present in the culture medium.

- Created recombinant strains demonstrated industrially significant characteristics, such as growth rate, resistance to high temperatures, high secretory activity, high resistance to ethanol and high resistance to the effects of inhibitory compounds, usually found in lignocellulose hydrolysates.

- The recombinant strain Y-2034/*tabgl1•cdt1-egfp*, created by integrating the intracellular cellobiose fermentation pathways, efficiently expresses the membrane cellodextrin transporter *cdt1* and *tabgl1* genes under the control of the constitutive promoters TEF1 and PGK1, respectively.

- The recombinant strain Y-2034/*tabgl1 • cdt1-egfp* grows well on a medium with cellobiose and produces ethanol in an amount of 5.55 g/L from 2% cellobiose, which is 52% of the theoretical maximum.

- An integral vector based on the δ-sequence was constructed containing the *A. niger* fungus endo-1,4-β-glucanase gene and two *L. edodes* fungus exo-1,4-β-glucanase genes with the dominant marker marker of phleomycin resistance.

- An industrial recombinant strain of *S. cerevisiae* has been created that coexpresses the *T. aurantiacus* 1,4-β-glucosidase, *A. niger* endo-1,4-β-glucanase and

L. edodes exo-1,4-β-glucanases genes.

The possibility of direct conversion of lignocellulosic raw materials to bioethanol using the recombinant yeast strain Y-2034-TaBgl1/*eng1•cel6B•cel7A*, which produces 5.62 g/L ethanol from 5% avicel and 1.43 g/L from 5% pre-processed straw wheat, has been shown for 3 days.

Main research results and conclusions:

1. Expression of *A. niger* endo-1,4-β-glucanase (ENG1) in the episomal plasmid vector YEGAp significantly slowed down the growth of yeast cell culture. Moreover, stable gene expression of this enzyme at the HO locus of the yeast chromosome did not cause growth inhibition, and the activity of the secreted enzyme was stably maintained throughout the incubation time (880 units/mg).

2. ENG1 expressed on the basis of the pHO-KanMX4-pGAPDH-*eng1*-tGAPDH-HO integrative vector, in addition to its N-glycosylation, contained sugar residues at the O-glycosylation site. Whereas ENG1 expressed by an episomal plasmid was present in recombinant yeast only in N-glycosylated form. It has been suggested that the O-glycosylated form of ENG1 retains more stable activity during continuous cultivation of the recombinant yeast strain with an integral vector than only the N-glycosylated form secreted by the episomal vector strain.

3. It was shown that ENG1 had optimal catalytic activity at pH 6.0 and was stable in the range of pH 5.0–9.0. The optimal enzyme temperature for catalytic activity at pH 6.0 was 70°C. More than 95% of the initial activity of the enzyme was maintained after incubation of the enzyme at 60°C for 2 hours.

4. An integral vector was constructed for integration into the yeast HO locus containing the *T. Aurantiacus* 1,4-β-glucosidase gene (*tabgl1*) with the signal peptide of the yeast α-factor under the control of the constitutive GAPDH promoter and terminator, and the gene for resistance to geneticin (G418)

5. For the first time, four industrial recombinant strains of *S. cerevisiae* were created for the fermentation of cellobiose after stable integration of the *tabgl1* gene. The resulting recombinant strains of *S. cerevisiae* secreted *T. aurantiacus* 1,4-β-glucosidase (TaBGL1), which can hydrolyze cellobiose to glucose, followed by fermentation to ethanol. It was shown that recombinant strains produced an average of 7.6 g/L of ethanol (77% of theoretical maximum yield) using cellobiose as the sole source of hydrocarbon.

6. It was shown that the glycosylation of secreted TaBGL1 depends on the genetic background of yeast strains and to a large extent on the carbon sources present in the cultivation medium (glucose or cellobiose). Differences in the nature of the TaBGL1 N-glycosylation leads to a change in the specific enzymes activity, without significantly affecting biochemical properties such as thermal stability, pH stability, temperature and pH optimums of activity.

7. It was shown that the created recombinant strains exhibit high osmotolerance and resistance to various concentrations of ethanol, furfural and H_2O_2, as well as to high temperatures.

8. An integral vector was designed for integration into the yeast HO locus, containing the gene (*tabgl1*) without the α-factor signal peptide of yeast and the hybrid *N. crassa* membrane cellodextrin transporter gene (*cdt1*) fused to the green fluorescent

protein gene under strong constitutive TEF1 and PGK1 promoters as well as the geneticin resistance gene (G418).

9. For the first time, a recombinant strain Y-2034/*tabgl1•cdt1-egfp* was created that efficiently expresses the cellodextrin transporter *cdt1* and *tabgl1* genes under the control of constitutive promoters TEF1 and PGK1, respectively. It was shown that the recombinant strain Y-2034/*tabgl1•cdt1-egfp* grew well on a medium with cellobiose and produced ethanol in an amount of 5.55 g/L from 2% cellobiose, which is 52% of the theoretical maximum.

10. An integral vector was constructed based on the δ-sequence containing the *A. niger* endo-1,4-β-glucanase gene and two *L. edodes* exo-1,4-β-glucanase genes with the dominant marker gene for resistance to phleomycin.

11. For the first time, an industrial recombinant strain of *S. cerevisiae* was created, co-expressing the *T. aurantiacus* 1,4-β-glucosidase, *A. niger* endo-1,4-β-glucanase and *L .edodes* exo-1,4-β-gluconase genes.

12. The possibility of direct conversion of lignocellulosic raw materials to bioethanol under anaerobic conditions using the recombinant yeast strain Y-2034-TaBgl1/*eng1•cel6B•cel7A* was shown. This strain most effectively converted avicel into ethanol, while the concentration of ethanol reached 5.62 g/L in 3 days. During the conversion of pre-treated wheat straw, alcohol concentrations of 1.43 g/L were achieved.

Connection of work with the plan of state programs.

The dissertation work was performed under project No. 0144/GF "Creation of recombinant strains of microorganisms that efficiently express cellulase genes to produce biofuels from cellulose-containing raw materials" and No. 1324/GF "Development of a technology for producing bioethanol from cellulose raw materials using genetically modified industrial yeast strains" of the Ministry of Education and Science of the Republic of Kazakhstan.

Approbation of scientific work. The main results of the research were reported:

- International conference of students and young scientists "Farabi Alemi", Almaty, Kazakhstan, 2016-2019;
- International scientific-practical conference “The contribution of microbiology and virology to the modern bio-industry”, (2016, Almaty, Kazakhstan);
- International conference "European Biotechnology Congress" (2017, Dubrovnik, Croatia).

Publications. The main content of the dissertation were published in 22 publications, including 2 articles and 2 theses in journals with a non-zero impact factor included in the Web of Science or Scopus database, 6 articles in national scientific journals recommended by the Education and Science Monitoring Committee of the Ministry of Education and Science of the Republic of Kazakhstan, and 12 abstracts in materials of international conferences. Based on the results of the thesis, 1 patent for useful model of the Republic of Kazakhstan No. 2017/0230.2 “Integrative plasmid vector for gene expression in yeast” was obtained.

The structure of the dissertation.

The dissertation work is presented on 145 pages and consists of symbols and abbreviations, introduction, review of literature, materials and methods of research,

results and their discussion, conclusions, a list of references from 260 items of which 248 are in English. The scope of work includes 10 tables, 55 figures and 1 annexes.