ANNOTATION

research work for the degree of
Doctor of Philosophy (PhD) in the specialty "8D05105 - Biotechnology"

Aisha Maratovna Shaizadinova

"Study of CRISPR/Cas novel programmable RNA-directed endonuclease for diagnostic applications"

General characterization of the dissertation research. This dissertation work is devoted to the characterization of new programmable RNA-directed CRISPR/Cas programmable endonucleases, MbCas12a and MeCas12a, and the development of molecular diagnostics based on these enzymes.

Relevance of the study. In recent years, CRISPR/Cas technologies have been the object of active research, due to which their application in various fields of science and medicine has significantly expanded. CRISPR/Cas systems were originally discovered as an immune mechanism of archaea and bacteria used by them to recognize and destroy foreign nucleic acids. This mechanism involves two major components: the CRISPR gene cluster that stores information about viral or plasmid invasions, and Cas proteins that provide a molecular mechanism of immunity that is activated upon re-invasion of a foreign nucleotide sequence. Subsequently, CRISPR/Cas systems have been widely used in genome editing, which has opened new horizons in genetic engineering, including the development of gene therapy methods and epigenetic research.

One of the most promising applications of CRISPR/Cas is molecular diagnostics of infectious diseases. Due to the ability of CRISPR/Cas to detect and destroy foreign DNA and RNA sequences with high accuracy, these systems have become the basis for diagnostic platforms capable of detecting even small concentrations of pathogens. These methods are highly accurate, fast and easy to use, and some platforms do not require expensive equipment, making them particularly relevant to global health challenges.

Thus, the use of CRISPR/Cas systems in molecular diagnostics represents an innovative approach that opens new perspectives in the fight against infectious diseases and promotes the development of new diagnostic methods.

Objects of study. The objects of the study are RNA-guided endonucleases Cas12a derived from *Moraxella bovis* and *Moraxella equi* strains; synthetic guide crRNAs; and DNA fragments of the pathogens *E. coli*, *S. aureus*, and *Alternaria* spp.

Study Subject. Structural and functional characterization of MbCas12a and MeCas12a endonucleases and development of effective methods of molecular diagnostics based on them.

Study Objective. Characterization of novel programmable RNA-directed CRISPR/Cas endonucleases and the use of these systems in the accurate identification of pathogen DNA sites in molecular diagnostics.

Objectives of the study:

1. Preparation and structural and functional study of recombinant Cas12a endonucleases from *Moraxella bovis* and *Moraxella equi* strains.
2. Design of guide RNAs (crRNAs), determination of the optimal PAM sequence for MbCas12a and MeCas12a endonucleases, preparation of the CRISPR-Cas12a ribonucleoprotein complex, and optimization of conditions for their operation under *in vitro* conditions.
3. Homology modeling and study of mutant forms of Cas12a enzymes to identify the active center of the endonuclease.
4. Development of diagnostic systems based on LAMP/CRISPR-Cas12a and RPA/CRISPR-Cas12a.

Research Methods. The study employed methods of molecular biology (including DNA amplification, cloning, and transformation), biochemistry (expression and purification of recombinant proteins), genetic engineering (construction of plasmid vectors and recombinant systems), and

bioinformatics (sequence analysis, homology modeling, and in silico structure prediction). Structural and functional analysis of Cas12a proteins was performed using molecular visualization tools and software packages for 3D modeling and active site annotation.

Scientific novelty of the study. The scientific novelty of the work lies in the characterization and obtaining of new previously unstudied analogues of Cas12a, which are programmable RNA-directed endonucleases of CRISPR/Cas systems. Also scientific novelty lies in the development of highly sensitive and specific methods of molecular diagnostics based on the obtained enzymes in combination with preliminary isothermal amplification to enhance detection efficiency.

Practical significance of the study. The practical significance of this work lies in the creation of highly effective and accurate diagnostic platforms capable of reliable detection of pathogens and infectious diseases, which contributes to a significant improvement of early diagnosis and treatment methods in clinical practice, as well as in the field.

Theoretical significance. The theoretical significance of the work consists in the expansion of knowledge about the mechanisms of CRISPR/Cas endonucleases, their interaction with DNA and the possibilities of their application in biotechnology.

The main provisions of the thesis put forward for defense:

1. Functionally active recombinant proteins MbCas12a and MeCas12a from *Moraxella bovis* and *Moraxella equi* are presented as novel Cas12a homologs with confirmed in vitro nuclease activity in complex with guide crRNAs. The optimized expression and purification protocol ensures the production of highly purified proteins.
2. The structural parameters of guide RNAs**,** including a length of 46–47 nucleotides, a target region of 23–24 nucleotides, and the PAM sequence “TTTN,” determine the high specificity of interaction with the DNA target and the effective performance of MbCas12a and MeCas12a endonucleases. The biochemical reaction conditions — 2 μM crRNA, 1 μM Cas12a, incubation temperature of 37 °C, and incubation time of 15–30 minutes — ensure efficient formation of the ribonucleoprotein complex and maximum endonuclease activity for nucleic acid cleavage.
3. The functional importance of the RuvC and Nuc catalytic domains in MbCas12a and MeCas12a enzymes was confirmed by homology modeling and site-directed mutagenesis. Critically important amino acid residues, including those in RuvC1, RuvC2, RuvC3, and Nuc regions, determine the enzyme’s activity in DNA cleavage. Mutations at these positions result in complete or partial loss of endonuclease activity.
4. The developed and experimentally validated molecular diagnostic method based on the MbCas12a endonuclease in combination with isothermal amplification (LAMP, RPA) enables specific and sensitive detection of DNA targets. The sensitivity of the diagnostic systems reaches 10^1 to 10^4 copies of DNA, demonstrating their practical applicability for rapid and field diagnostics of microbial and fungal pathogens such as *E. coli*, *S. aureus*, and *Alternaria* spp.

Main research results and conclusions

1. Genes encoding Cas12a endonucleases from *Moraxella bovis* and *Moraxella equi* were obtained. The Cas12a (*Moraxella bovis*) gene was cloned into an intermediate vector for sequencing, production, and storage. The Cas12a (*Moraxella equi*) gene was synthesized with a codon-optimized sequence. Both proteins, MbCas12a and MeCas12a, were expressed in *E. coli* cells, and purification conditions were optimized. Purity and identity were confirmed by SDS-PAGE and chromatography. Experimental studies showed that both endonucleases exhibited specific DNA cleavage activity in the presence of crRNA, confirming their functionality under in vitro conditions.

2. The optimal PAM sequence for MbCas12a and MeCas12a was determined, revealing that the TTTN motif, where "N" denotes any nucleotide, provides the highest DNA cleavage activity. Screening of PAM variants (TTTT, TTTA, TTTC, TTTG) showed comparable activity levels, confirmed by densitometric analysis. These findings indicate high tolerance and adaptability of CRISPR/Cas12a systems to various PAM configurations without significant loss of efficiency.

3. Design, synthesis, and purification of guide RNAs (crRNAs) were performed. crRNA design was based on identifying direct repeats in the genomes of *Moraxella bovis* and *Moraxella equi*, as well as target regions in pathogenic DNA sequences. Synthesized crRNAs were tested for their

efficiency in directing MbCas12a and MeCas12a to cleave target DNA. The experiments confirmed that these crRNAs effectively guide the endonucleases to specific DNA sites for accurate and specific cleavage. Additionally, crRNAs were synthesized using specific primers, and experiments were conducted to determine the optimal length and truncation of the 5′ handle. It was shown that both the target region length and the handle length significantly influence Cas12a activity. Reducing the target region from 23 to 15 nucleotides led to a statistically significant reduction in activity, with residual activity of 3–8% according to densitometric analysis. The optimal target length was 23–24 nucleotides. Similarly, shortening the 5′ handle below 15 nucleotides resulted in a marked drop in activity, especially at 13 nucleotides or less.

4. Cas12a endonuclease activity and specificity were evaluated against various substrates. Reaction conditions were optimized, including temperature, duration, component concentrations, and buffer composition. For cis-activity assays, optimal concentrations were 2 μM crRNA and 1 μM Cas12a. DNA cleavage began as early as 5 minutes after incubation under optimal conditions, though 30-minute incubations were used for reproducibility. For trans-activity, 100 nM crRNA and 100 nM Cas12a in 1× NEBuffer 2.1 were used. Fluorescence-based CRISPR/Cas12a assays showed that incubating the MbCas12a-crRNA complex at 25 °C for 15 minutes followed by a 30-minute trans-cleavage reaction at 37 °C allowed detection using a UV transilluminator.

5. Homology modeling of MbCas12a and MeCas12a identified catalytically important amino acids in the RuvC and NUC domains. Site-directed mutagenesis and functional assays confirmed the roles of these residues. Mutations in RuvC1 (D874A/D864A), RuvC2 (E968A/E958A), RuvC3 (D1218A/D1208A), and NUC (R1181A/R1171A) domains resulted in complete or partial loss of activity. Densitometric and statistical analyses confirmed significant differences. Complete loss of function was observed in mutants D874A, E968A, D1218A (MbCas12a) and D864A, E958A, D1208A (MeCas12a), while R1181A and R1171A caused partial reduction, indicating their differentiated roles. RuvC domains primarily mediate cleavage of both DNA strands, whereas the Nuc domain residues contribute to single-strand cleavage.

6. Conditions for loop-mediated isothermal amplification (LAMP) and recombinase polymerase amplification (RPA) were optimized. For RPA, temperature (37–39 °C) and time (20–30 minutes) were optimized along with specific primers targeting DNA. For LAMP, the optimal conditions included 62 °C for 45 minutes, using three pairs of specialized primers.

7. A CRISPR/Cas12a-based immunochromatographic lateral flow assay using gold nanoparticles and ssDNA probes labeled with fluorescein and biotin was developed. The reaction was conducted at 37 °C for 30 minutes, followed by visual detection on a lateral flow strip, allowing rapid and efficient identification of target DNA sequences.

8. A molecular diagnostic method for *Escherichia coli* was developed based on the combination of loop-mediated isothermal amplification (LAMP) and CRISPR/MbCas12a technology. The method demonstrated detection sensitivity down to 1 plasmid DNA copy and 10 genomic DNA copies per reaction. Specificity was ensured through the use of target-specific crRNAs and the collateral activity of MbCas12a, allowing reliable differentiation of *E. coli* from other mastitis-causing pathogens such as *Staphylococcus aureus* and *Streptococcus agalactiae*. The assay was validated on clinical samples of mastitic milk and showed reproducible detection of *E. coli*, confirming its applicability for field-level veterinary diagnostics.

9. A detection method for *Staphylococcus aureus* was proposed based on the combination of recombinase polymerase amplification (RPA) and CRISPR/MbCas12a, using crRNA targeting the *nuc* gene. The system enabled detection of down to 10^2 copies of DNA with no cross-reactivity to other pathogens. The method was validated on 13 clinical isolates, including mastitic milk samples, and demonstrated stable performance, high accuracy, and a total analysis time of approximately 60 minutes, making it suitable for rapid on-site veterinary diagnostics.

10. A method for detecting *Alternaria spp.* was developed using CRISPR/MbCas12a combined with preliminary DNA amplification via RPA. The detection limit reached 100 DNA copies in wheat grain samples. Specificity testing confirmed no detection signal with DNA from *Bipolaris sorokiniana*, *Nigrospora oryzae*, *Fusarium equiseti*, and *Fusarium acuminatum*. Application to

naturally infected samples confirmed the method's suitability for rapid phytopathogen screening in agricultural biotechnology and phytosanitary control.

11. Statistical analysis of the sensitivity data of the developed molecular diagnostic methods, as well as the statistical processing of the results obtained by densitometric analysis of electropherograms, showed that the activity of the Cas12a endonuclease significantly depends on the testing conditions, including the structure of the guide RNA, mutations in the catalytic domains, and the concentrations of the reaction components. Significant differences were established using one-way analysis of variance (ANOVA, $p < 0.05$) and the non-parametric Kruskal–Wallis test.

12. A portable CRISPR-Kaz device for fluorescence-based pathogen detection was developed. Designed for low-infrastructure and field use, the device includes a 3D-printed housing with wells for standard tubes, a glass platform for even UV light distribution, and an integrated UV source. The compact design ensures ease of use and rapid pathogen detection via CRISPR/Cas12a.

Relation to the plan of the main scientific works. The thesis was carried out in the Laboratory of Molecular Biotechnology of "National Center of Biotechnology" LLP within the framework of grant AP09259771 "Characterization of novel programmable RNA-guided endonucleases CRISPR/Cas systems for diagnostic applications", BR10764944 "Development of analytical control methods and monitoring of food safety products", AP19579275 "Development of modern molecular diagnostics of crop pathogens based on the CRISPR/Cas12a method ".

Approbation of research work. The results of the research work were presented at the Scientific and Technical Council of the Faculty of Biology and Biotechnology, meetings of the Department of Genetics and Molecular Biology of Al-Farabi KazNU, scientific seminars of the Laboratory of Molecular Biotechnology of "National Center of Biotechnology" LLP. The results of the dissertation work are included in the reports of research projects AP09259771 "Characterization of novel programmable RNA-guided endonucleases CRISPR/Cas systems for diagnostic applications", BR10764944 "Development of analytical control methods and monitoring of food safety", AP19579275 "Development of modern molecular diagnostics of crop pathogens based on the CRISPR/Cas12a method".

Publications. The main results of the dissertation work ar e presented in 7 printed works, including 3 articles in journals with non-zero impact factor included in the database Web of Science or Scopus, 3 articles in the republican scientific editions recommended by KKSON MES RK, 1 patent was obtained. Normative and technical documentation (NTD) was issued.

Doctoral student's personal contribution. Doctoral student's personal contribution includes planning and conducting all experiments, data analysis, writing scientific articles, processing of research results.

The scope and structure of the thesis. The dissertation consists of an introduction, literature review, description of methods, research results and discussion of findings, conclusion, reference list and appendices. The thesis is set out on 163 pages and contains 16 tables, 88 figures, 162 sources, 9 appendices.