SAINT PETERSBURG STATE UNIVERSITY

Printed as a manuscript

Nikolaev Dmitrii Mikhailovich

RATIONAL DESIGN OF ARCHAERHODOPSIN-3-BASED FLUORESCENT GENETICALLY ENCODED VOLTAGE INDICATORS

Scientific specialty 1.4.4. Physical Chemistry

Submitted for the degree of Candidate of Chemical Sciences

Translated from Russian

Thesis supervisor:
Doctor of Chemical Sciences,
Ryazantsev Mikhail Nikolaevich

Saint Petersburg — 2025

Contents

Introduction

Relevance of the topic. The study of biological processes at the cellular level requires the use of tools and techniques to monitor changes in the physical properties of cells. Membrane potential, which is the difference in electrical potential between the inside and outside of the cell membrane, is a key property of biological cells. In a healthy state, cells maintain a constant non-zero value of membrane potential, called the resting potential. Changes in membrane potential occur in a wide range of biological processes, such as the activity of electrically excitable cells [1], transitions between stages of the cell cycle [2], transformation of cells into tumorous states [3], cell death, and numerous other cellular activities [4].

Protein-based genetically encoded voltage indicators (GEVIs), which exhibit voltage dependent fluorescence intensity, are finding an increasing number of applications as tools for monitoring changes in cell membrane potential during biological processes [5–7]. Such sensors are introduced into cells using genetic engineering techniques, and the fluorescence intensity of sensor-containing cells is recorded during the process under investigation. The observed changes in fluorescence intensity are directly related to changes in cell membrane potential, allowing researchers to perform real-time monitoring of voltage transients using optical methods. Currently, GEVIs are widely used to study the activity of neurons and cardiomyocytes [8–10].

The main characteristics that determine the applicability of GEVIs are signal intensity, the kinetics of the signal response to voltage changes, the voltage sensitivity of the signal, which is estimated as the percentage change in fluorescence intensity with a voltage increase of 100 mV, and the proximity of the sensor absorption band to the window of optical transparency of biological tissues, which lies in the wavelength range approximately from 650 nm to 900 nm [11]. A promising class of GEVIs showing fast kinetics, high voltage sensitivity of fluorescence and absorption band maxima at wavelengths $>$ 580 nm are sensors based on microbial rhodopsin archaerhodopsin-3 [12]. The main drawback of archaerhodopsin-3-based sensors is the low fluorescence intensity, and research is still ongoing to find new brighter variants [9; 13–15]. However, the development of novel archaerhodopsin-3-based sensors is hampered by the lack of data on the molecular mechanisms that regulate the fluorescence intensity of microbial rhodopsins.

Elaboration of the topic. To date, all fluorescent GEVIs based on microbial rhodopsins are archaerhodopsin-3 variants with amino acid substitutions. Archaerhodopsin-3 is a microbial rhodopsin with a high expression level and good membrane localization in eukaryotic cells, in particular neuronal cells. The protein showed a linear dependence of fluorescence intensity on voltage changes in the range from -150 mV to +150 mV [16]. When the voltage was increased by 100 mV, the recorded percentage increase in fluorescence intensity was 30%. The recorded response time of protein fluorescence to changes in membrane potential was approximately 0.6 ms [16]. Due to these properties, archaerhodopsin-3 was selected as a promising protein for the development of molecular tools for the visualization of membrane potential changes in biological processes. However, three major drawbacks of wild-type archaerhodopsin-3 as a fluorescent voltage indicator have been identified. 1. Low intensity of the fluorescence signal; the reported fluorescence quantum yield of the protein was only 0.01% [17]. 2. The absorption band of the protein with a maximum at 556 nm does not overlap with the optical transparency range of biological tissues [12]. 3. When the protein is exposed to light, it transports protons across the cell membrane, which leads to non-negligible changes in the membrane potential [16].

Since the discovery of the voltage dependent fluorescence of archaerhodopsin-3, studies by several research groups have focused on the search for new archaerhodopsin-3 variants with improved properties for applications as membrane potential sensors, mainly with brighter fluorescence [9; 13; 14; 18]. The lack of information on the molecular mechanism responsible for the brightness enhancement prevented the application of the rational design approach to the development of new sensors. Therefore, only directed evolution was used, as this approach does not require information about the protein structure and the molecular mechanisms that regulate protein properties. In directed evolution, the search for proteins with improved properties is performed iteratively by sequentially generating libraries of protein mutants with a stochastic distribution of amino acid replacements and selecting the best protein variants in one generation to generate the next library of protein mutants. As a result of directed evolution studies, a series of archaerhodopsin-3 variants with enhanced voltage dependent fluorescence were proposed. Although the new proteins showed advantageous properties compared to wild-type archaerhodopsin-3, there is still room for improvement – the absorption maxima of the obtained proteins do not exceed 633 nm and their fluorescence quantum yields do not exceed 0.8% [19]. Therefore, further studies are needed to develop new archaerhodopsin-3 mutants with improved properties, especially increased fluorescence intensity.

The aim of the thesis study was to perform the rational design of novel archaerhodopsin-3-based fluorescent genetically encoded voltage indicators with enhanced fluorescence.

To achieve the stated goal, the following **objectives** had to be accomplished:

1. Development of reliable computational models of rhodopsins for predicting the three-dimensional structure of the protein and calculating its absorption maximum based on the corresponding amino acid sequence.
2. Revealing the mechanism responsible for the significant increase in fluorescence intensity of archaerhodopsin-3 mutants from the Archers group compared to wild-type archaerhodopsin-3.
3. Performing the rational design of novel fluorescent archaerhodopsin-3-based genetically encoded voltage indicators.

Scientific novelty.

1. New mutants of archaerhodopsin-3 with enhanced brightness compared to variants of fluorescent genetically encoded voltage indicators from the Archers group published by the time of the thesis study.
2. For the first time it has been shown that amino acid substitutions found in the genetically encoded voltage indicators from the Archers group cause stabilization of the protein state with the protonated counterion D222, leading to a significant increase in Archers fluorescence compared to wild-type archaeorhodopsin-3.
3. A new adaptation of the method of constructing quantum mechanics/molecular mechanics protein models, in which the amino acid environment of a cofactor described at the quantum mechanical level of theory is constructed by taking into account averaging over protein configurations existing in thermodynamic equilibrium.

Theoretical and practical significance. The developed archaerhodopsin-3-based fluorescent genetically encoded voltage indicators can be used as tools to study the electrical activity of biological cells. The data obtained on the molecular mechanism that causes a significant increase in the fluorescence intensity of archaerhodopsin-3 mutants from the Archers group compared to wild-type archaerhodopsin-3 can be used to develop new fluorescent sensors of cell membrane potential based on microbial rhodopsins. The methods developed to construct improved computational models of rhodopsins and the corresponding code can be used to study a wide range of different proteins.

Approbation of the research. The results obtained within this study were published in 6 articles:

1. **Nikolaev, D. M.**, Mironov, V. N., Metelkina, E. M., Shtyrov, A. A., Mereshchenko, A. S., Demidov, N. A., Vyazmin S. Yu., Tennikova T.B., Moskalenko S.E., Bondarev S.A., Zhouravleva G.A., Vasin A.V., Panov M.S., Ryazantsev, M. N. (2024). Rational Design of Far-Red Archaerhodopsin-3-Based Fluorescent Genetically Encoded Voltage Indicators: from Elucidation of the Fluorescence Mechanism in Archers to Novel Red-Shifted Variants. ACS Physical Chemistry Au: 347-362. https://doi.org/10.1021/acsphyschemau.3c00073. Journal Quartile – Q1, Journal Impact Factor – 3.7.
2. **Nikolaev, D. M.**, Shtyrov, A. A., Vyazmin, S. Y., Vasin, A. V., Panov, M. S., Ryazantsev, M. N. (2023). Fluorescence of the retinal chromophore in microbial and animal rhodopsins. International Journal of Molecular Sciences, 24(24), 17269. https://doi.org/10.3390/ijms242417269 Journal Quartile – Q1, Journal Impact Factor – 6.2.
3. **Nikolaev, D. M.**, Mironov, V. N., Shtyrov, A. A., Kvashnin, I. D., Mereshchenko, A. S., Vasin, A. V., Panov M.S. Ryazantsev, M. N. (2023). Fluorescence Imaging of Cell Membrane Potential: From Relative Changes to Absolute Values. International Journal of Molecular Sciences, 24(3), 2435. DOI: 10.3390/ijms24032435 Journal Quartile – Q1, Journal Impact Factor - 6.2.
4. **Nikolaev, D. M.**, Manathunga, M., Orozco-Gonzalez, Y., Shtyrov, A. A., Guerrero Martínez, Y. O., Gozem, S., Ryazantsev M.N., Coutinho K., Canuto S., Olivucci, M., 2021. Free Energy Computation for an Isomerizing

Chromophore in a Molecular Cavity via the Average Solvent Electrostatic Configuration Model: Applications in Rhodopsin and Rhodopsin-Mimicking Systems. Journal of Chemical Theory and Computation, 17(9), pp. 5885–5895. DOI: 10.1021/acs.jctc.1c00221. Journal Quartile – Q1, Journal Impact Factor - 6.0.

5. Shtyrov, A. A., **Nikolaev, D. M.**, Mironov, V. N., Vasin, A. V., Panov, M. S., Tveryanovich, Y. S., Ryazantsev, M. N., 2021. Simple Models to Study Spectral Properties of Microbial and Animal Rhodopsins: Evaluation of the Electrostatic Effect of Charged and Polar Residues on the First Absorption Band Maxima. International Journal of Molecular Sciences, 22(6), p. 3029. DOI: 10.3390/ijms22063029. Journal Quartile – Q1, Journal Impact Factor - 6.2.
6. **Nikolaev, D. M.**, Shtyrov, A. A., Mereshchenko, A. S., Panov, M. S., Tveryanovich, Y. S., Ryazantsev, M. N., 2020. An assessment of water placement algorithms in quantum mechanics/molecular mechanics modeling: the case of rhodopsins' first spectral absorption band maxima. Physical Chemistry Chemical Physics, 22 (32), pp. 18114-18123. DOI: 10.1039/D0CP02638G. Journal Quartile – Q2, Journal Impact Factor - 3.7.

The findings of the investigation were reported and discussed at the following conferences:

- Nikolaev D.M., Ryazantsev M.N. (2024) Rational design of novel archaerhodopsin-3-based fluorescent genetically encoded voltage indicators. Lomonosov-2024, April 12-26 2024, Moscow.
- Nikolaev D.M., Metelkina E.M., Panov M.S., Ryazantsev M.N. (2025) Rational design of archaerhodopsin-3-based fluorescent genetically encoded voltage indicators. Optogenetics+ 2025, March 3-5 2025, St Petersburg.

Methodology and research methods. An approach that integrates the methods of directed evolution and rational design has been used in the development of novel sensors of cell membrane potential. Rational design is an approach to protein engineering in which the choice of amino acid substitutions to modify protein properties or functions is based on data about the protein structure and the molecular mechanisms that determine its properties or functions. Directed evolution is an approach to protein engineering in which protein properties are modified by mimicking the process of natural evolution. A library of protein variants with a stochastic distribution of amino acid substitutions is generated, and the optimized property is measured for each protein in the library. The protein variants with the largest improvement in the optimized properties are selected and used as the basis for generating a new library of protein variants. The process is iterative and stops when several iterations have failed to improve the protein properties.

Rational design and directed evolution each have their own advantages and disadvantages. The advantage of rational design is efficiency: a small number of amino acid substitutions selected on the basis of structure and mechanism analysis are introduced by site-directed mutagenesis and lead to the desired change in protein properties. A disadvantage of rational design is that it is often difficult to implement. Highly accurate computational models of the protein are required to determine the molecular mechanism responsible for the change in protein properties. In turn, directed evolution does not require any prior information beyond the amino acid sequence, but it does require searching through too many variants. Besides, the probability of finding the best protein variant using directed evolution is not very high.

The combination of rational design and directed evolution, which combines their strengths, has been implemented as follows. The set of proteins resulting from directed evolution was divided into two groups – those with improved properties and those without. Then, using experimental methods and computational modeling, a comparative analysis of the two groups of proteins was carried out, and the structural feature responsible for the improvement of the properties of the proteins from the second group is determined. The obtained information allows introducing the amino acid substitutions necessary to modify a specific structural feature and, accordingly, to modify the optimized property of the protein – to carry out rational design. Both computational modeling and experimental methods, including steady-state absorption spectroscopy, Raman spectroscopy, fluorescence spectroscopy, and fluorescence microscopy, have been used in this study.

Personal contribution of the author. The author participated in the development of the research plan, construction of computational models of microbial rhodopsins and their analysis, development of the adaptation of the method of construction of quantum mechanics/molecular mechanics models of proteins, in which the amino acid environment of the cofactor described by quantum mechanical methods is generated by averaging over the configurations existing in thermodynamic equilibrium, measurement of absorption and fluorescence spectra of proteins, fluorescence quantum yields, and preparation of publications on the results of the research. The protein synthesis was carried out by colleagues, the Raman spectra measurements were carried out by the staff of the SPbU Resource Center.

Thesis structure. The dissertation consists of an introduction, four chapters, and a conclusion. Chapter 1 provides a literature review on the dissertation topic. Chapter 1.1 summarizes the current classes of fluorescent genetically encoded sensors of cell membrane potential. Chapter 1.2 gives an overview of the archaerhodopsin-3-based voltage indicators published at the time of this thesis study and provides the necessary information on the structure of archaerhodopsin-3. Chapter 1.3 presents an analysis of literature data on fluorescence in microbial rhodopsins, showing that neutralization of the amino acid environment near the Schiff base in these proteins leads to a significant increase in fluorescence brightness. Chapter 2 describes the methods used in the thesis study. In particular, Section 2.1 describes the general approach to protein design. Section 2.2 describes computational methods, including those developed in the thesis study. Chapter 2.3 describes the experimental methods. Chapter 3 describes the part of the thesis study aimed at determining the molecular mechanism responsible for the significant increase in fluorescence intensity of the archaerhodopsin-3-based genetically encoded voltage indicators from the Archers group compared to wild-type archaerhodopsin-3. Chapter 4 describes part of the dissertation study aimed at the rational design of new archaerhodopsin-3-based fluorescent genetically encoded voltage indicators with enhanced fluorescence and absorption band maxima shifted towards the optical transparency window of biological tissues.

The main scientific results:

1. Improved computational models for predicting the three-dimensional structure and calculating the spectral properties of rhodopsins starting the amino acid sequence. In particular, a new adaptation of the method of constructing quantum mechanics/molecular mechanics protein models, in which the amino acid environment of a cofactor described at the quantum mechanical level of theory is constructed by taking into account averaging over protein configurations existing in thermodynamic equilibrium (pages 18116 - 18120 in the article [20], author contribution - construction of

rhodopsin models, analysis of results; pages 3031-3043 in the article [21], author contribution - construction of protein models, analysis of results; pages 5888-5892 in the article [22], author contribution - adaptation of the ASEC-FEG methodology, writing the code, construction of rhodopsin models, analysis of results).

2. Evidence has been obtained that a sufficient condition for obtaining mutants of microbial rhodopsins with increased intensity of the fluorescent signal is thermodynamic stabilization of the protein form with the absence of charged residues near the retinal Schiff base. Evidence has been obtained that in proteins from the Archers group, introduced amino acid substitutions lead to thermodynamic stabilization of the protein form with protonated counterion D222, which causes a significant increase in Archers fluorescence compared to the wild-type protein (pages 17271-17284 in the article [23], author contribution – analysis of literature data, revealing the correlations; pages 348-355 in the article [19], author contribution – construction and analysis of rhodopsin computational models, measurement of spectral and photophysical properties of the proteins, analysis of experimental results).
3. New mutants of archaerhodopsin-3 were obtained. The proteins demonstrate increased brightness compared to known variants of fluorescent genetically encoded voltage indicators from the Archers group. The fluorescence intensity of proteins depends on the magnitude of the cell membrane potential. The absorption maxima of the new proteins are closest to the window of optical transparency of biological tissues in comparison with all archaerhodopsin-3-based fluorescent genetically encoded voltage indicators published at the time of the thesis study (page 2431 in the article [24], pages 355-357 in the article [19],author contribution – construction and analysis of rhodopsin computational models, measurement of spectral and photophysical properties of the proteins, analysis of experimental results).

Thesis statements to be defended:

1. Improved computational models for predicting the three-dimensional structure and calculating the spectral properties of rhodopsins starting the amino acid sequence.

2. The main factor responsible for the significant increase in fluorescence intensity of microbial rhodopsins is the stabilization of the protein state with the absence of charged residues near the retinal Schiff base. In the proteins of the Archer group, the increase in brightness is achieved by the stabilization of the protein state with the protonated counterion D222.
3. New mutants of archaerhodopsin-3 with increased brightness compared to known variants of fluorescent genetically encoded voltage indicators from the Archers group. The fluorescence intensity of the proteins depends on the magnitude of the cell membrane potential. The absorption maxima of the new proteins are closest to the window of optical transparency of biological tissues compared to all archaerhodopsin-3-based fluorescent genetically encoded voltage indicators published at the time of the thesis study.

Chapter 1. Literature Review

1.1 General information on fluorescent genetically encoded sensors of cell membrane potential

Studying biological processes at the cellular level requires tools and methods to record changes in the physical properties of cells during the process studied. One of the most important characteristics of the biological cell is the membrane potential – the difference in electrical potential between the inside and outside of the cell membrane. Cells in a healthy state maintain a constant value of the membrane potential that is different from zero – the resting potential. Changes in membrane potential occur in a variety of biological processes, such as the activity of electrically excitable cells [1], transitions between cell cycle phases [2], transformation of cells into a tumorigenic state [3], and many other processes [4]. There are two general classes of methods – electrophysiological and optical – for recording changes in the potential of the cell membrane. Electrophysiological methods use microelectrodes placed at the cell membrane to record changes in membrane potential. Electrophysiological methods can record changes in membrane potential with a high degree of accuracy. However, there are several limitations to their use. First, microelectrodes damage the cell membrane. This can lead to artifactual changes in potential [25]. Second, electrophysiological methods allow only a limited number of cells to be measured, ranging from a single cell when using the patch-clamp method to several hundreds of cells when using microelectrode arrays [26]. In addition, there are technical difficulties in getting the microelectrodes to the correct cells, especially during the *in vivo* studies [27].

Optical methods use sensors to detect changes in membrane potential. The sensors are organic molecules, proteins, or nanoparticles whose optical properties change when there is a change in the external electric field, and for most of the currently developed sensors, the voltage dependent optical property is fluorescence intensity. Such sensors are inserted into the cell membrane. Changes in their optical properties are then recorded during the process being studied. Optical methods do not limit the number of cells from which signals can be recorded. The methods and sensors developed so far allow signal recording with different spatial and temporal resolutions. The time scale of the processes that can be studied with optical methods ranges from a few milliseconds to several days [24]. Currently, the recording of membrane potential changes using fluorescent sensors is widely used in biological and biomedical research to study the activity of neurons in the brain, the activity of cardiomyocytes, and other processes [8–10].

Organic dyes and genetically encoded sensors are the two most widely used classes of fluorescent membrane potential sensors. Most organic dyes are characterized by high fluorescence intensity and fast kinetics, with response times of some dyes not exceeding fractions of a millisecond [28–30]. The main disadvantages of organic dyes include the inability to selectively introduce sensors into a population of cells of a specific type, phototoxicity, and low photostability. The lack of selectivity does not allow to register a signal only from the desired cell type. The sensors are introduced either into one cell or into all cells of the biological object studied. Phototoxicity leads to changes in the physiological properties of the investigated cells under prolonged exposure to light. Low photostability makes it difficult to use organic dyes in long-term experiments.

Genetically encoded membrane potential sensors are proteins or multiprotein constructs. Genetic engineering methods are used for the targeted delivery of these sensors to cell populations of the cell type of interest. Compared to organic dyes, most genetically encoded sensors are characterized by low or no phototoxicity and higher photostability [11; 31]. Based on the protein domain responsible for the voltage sensitivity, genetically encoded sensors can be divided into two major classes.

The first class of sensors uses a voltage sensitive domain of the voltage dependent phosphatase from *Ciona intestinalis* [5] or *Gallus gallus* [32]. This domain is composed of four transmembrane helices. The fourth helix (S4) contains a large number of charged residues and changes its position with changes in membrane potential. The fluorescent moiety is a fluorescent protein or a pair of fluorescent proteins linked to the voltage sensing domain by a peptide linker. Three types of constructs in this class have been developed to date.

- The fluorescent protein binds to the end of the S4 helix through a peptide linker. Reorganization of the charged S4 upon membrane depolarization leads to a decrease in brightness of the fluorescent protein [5] (Figure 1.1a).

- The circularly permutated fluorescent protein binds to the loop between helices S3 and S4 via a peptide linker. Reorganization of the loop upon membrane depolarization leads to a decrease in brightness of the fluorescent protein [33].
- A pair of fluorescent proteins capable of resonance energy transfer binds to the domain via a peptide linker [34] (Figure 1.1b). Reorganization of the domain upon membrane depolarization leads to a change in the relative position of the two fluorescent proteins and consequently a change in the amount of resonance energy transfer between them. A decrease in energy transfer results in an increase in the recorded fluorescence of the donor and a decrease in the recorded fluorescence of the acceptor. The ratio of the signal intensities of the donor and acceptor is recorded during the experiments.

Genetically encoded sensors from the second class use a microbial rhodopsin as a voltage sensitive part. Two types of sensors have been developed for this class.

- A microbial rhodopsin with a voltage dependent concentration of the protonated chromophore protein state is linked via a peptide linker to a fluorescent protein whose fluorescence spectrum overlaps with the absorption spectrum of the protonated chromophore rhodopsin [35]. (Figure 1.1b). Depolarization of the membrane leads to an increase in the concentration of protonated chromophore rhodopsin, an increase in resonance energy transfer from fluorescent protein to rhodopsin, and a decrease in the intensity of the recorded fluorescence signal.
- A microbial rhodopsin with voltage dependent fluorescence intensity [16] (Figure 1.1c). Such rhodopsins can be used as sensors without additional protein domains.

First and second class sensors using fluorescent protein constructs are characterized by high fluorescence intensity, but relatively low signal response rates to membrane potential changes and not very high signal voltage sensitivity [36; 37]. It should also be noted that all fluorescent proteins used in such sensors are activated by light in the blue or green spectral range, which is not optimal for studying biological objects.

On the other hand, fluorescent microbial rhodopsins offer several advantages over fluorescent protein-based sensors:

- Fast kinetics of signal response to voltage changes, response times of some sensors do not exceed one millisecond;
- High sensitivity of the recorded fluorescence signal to changes in membrane potential;
- High level of expression and plasma membrane localization in eukaryotic cells, particularly, in neurons;
- Possibility of activation by light with wavelengths in a range > 600 nm.

The main disadvantage of all known sensors based on microbial rhodopsins is the low fluorescence intensity [9; 13; 14]. This thesis research was aimed at the development of novel sensors based on microbial rhodopsins with increased fluorescence intensity.

1.2 Fluorescence sensors of cell membrane potential based on microbial rhodopsins

The results of the first experiments demonstrating the possibility of using microbial rhodopsins as fluorescent sensors of membrane potential were obtained by the group of Prof. Adam E. Cohen at Harvard University for a proteorhodopsin with the amino acid substitution D97N (PROPS sensor) [12]. Prokaryotic *Escherichia coli* cells expressing the protein were used for the experiments. Changes in the membrane potential of the bacterial cells during their life activity were correlated with the recorded changes in the fluorescence signal. A significant increase in fluorescence intensity was observed when the membrane potential of bacterial cells was brought to zero by the addition of the CCCP ionophore. A similar effect was observed after cell death caused by sodium azide or ultraviolet radiation.

Due to the impossibility of expressing PROPS in eukaryotic cells, further studies were performed with another microbial rhodopsin – archaerhodopsin-3 and its amino acid substituted variants. The group of Prof. Adam E. Cohen demonstrated the possibility of using wild-type archaerhodopsin-3 as a fluorescent sensor of the membrane potential [16]. In eukaryotic HEK293 cells, a linear relationship was found between the fluorescence intensity recorded from protein-expressing cells and the voltage applied to the cell in the range of -150 mV to +150 mV [16]. The

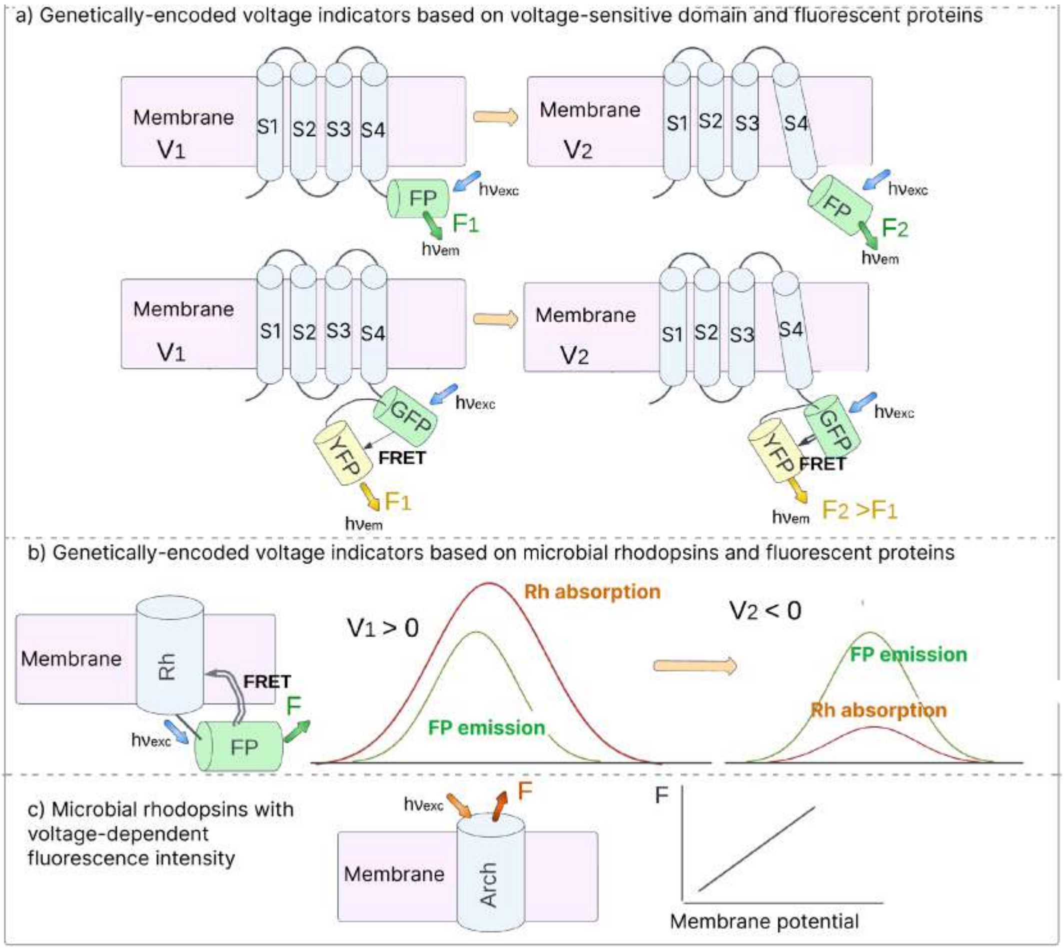


Figure 1.1 — Main types of fluorescent genetically encoded voltage indicators. a) A construct consisting of a voltage sensitive transmembrane domain, the position of helix S4 of which changes with alterations in membrane potential, and a fluorescent protein (*upper*) or a pair of fluorescent proteins between which resonance energy transfer is possible (*bottom*). b) A construct consisting of a microbial rhodopsin in which the concentration of the protonated chromophore form is regulated by a membrane potential and a fluorescent protein. c) A microbial rhodopsin with voltage dependent fluorescence.

fluorescence intensity increased by about 30% when the potential was increased by 100 mV [16]. Archaerhodopsin-3 showed high expression and localization in the membrane of eukaryotic cells, as well as rather high signal photostability [16]. The recorded response time of archaerhodopsin-3 fluorescence to changes in membrane potential was less than 0.5 ms [38]. These properties made archaerhodopsin-3 a promising protein for use as a membrane potential sensor. However, three significant drawbacks of this protein as a sensor were also identified. First, the protein exhibits low fluorescence intensity. The recorded fluorescence quantum yield of the protein was only 0.01% [17]. Second, under light exposure, archaerhodopsin-3 transports protons across the cell membrane. This causes a change in the membrane potential of the cell [16]. The main function of the protein is to transport protons from the cytoplasm to the extracellular environment under exposure to light. Third, the activation spectrum of the protein is not optimal for studying biological objects. The absorption band of archaerhodopsin-3 does not overlap with the window of optical transparency of biological tissues, which is in the range of $\sim$ 650 - 900 nm. The absorption maximum of the protein is 556 nm [16].

Further studies have been carried out to develop variants of archaerhodopsin-3 with amino acid substitutions that have more optimal properties for use as a fluorescent sensor of membrane potential. In order to describe the results of these studies, it is first necessary to give a brief description of the structure of archaerhodopsin-3, its functions, and to list the key amino acids involved in the function of the protein that have a significant impact on its properties.

Archaeorhodopsin-3 (Arch) is a member of the microbial rhodopsin family. Microbial rhodopsins are a family of light-activated membrane proteins found in all kingdoms of living nature, including viruses [39–43]. These proteins have multiple functions, acting as photosensors, light-activated ion pumps, and ion channels. The photosensitive moiety of microbial rhodopsins is the cofactor retinal, which is bound to a conserved lysine residue on the seventh helix of the protein via a Schiff base. In the most thermodynamically stable conformation of the protein, the Schiff base is protonated. Absorption of a photon brings the chromophore into the electronically excited state ($S_0 \to S_1$). The subsequent return of the chromophore to the ground state can occur by either non-radiative pathways or by photon emission. The non-radiative transition occurs through the conical intersection where the photoreaction branches (Figure 1.2). The first pathway results in the return of the chromophore to its original state, while the second pathway results in its photoisomerization, which triggers a series of processes necessary for rhodopsin to perform its biological function. The amino acid environment of the chromophore promotes photoisomerization because this process is directly related to the natural functions of the protein, and the quantum yields of retinal photoisomerization in microbial rhodopsins are very high. For example, for bacteriorhodopsin, the quantum yield of photoisomerization is 0.64% [44; 45]. Fluorescence, on the other hand, is a side process for rhodopsins, and the fluorescence quantum yields of microbial rhodopsins are very small, in the tenths and hundredths of a percent. The only exception to this rule is the recently discovered neorhodopsin from *Rhizoclosmatium globosum*, which has a fluorescence quantum yield of 20% [46].

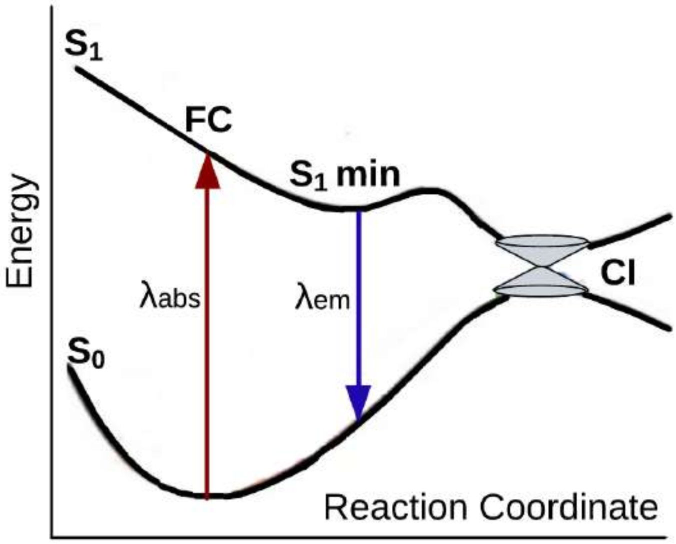


Figure 1.2 — Profiles of the potential energy surfaces of the ground state (S_0) and electronically excited state (S_1) of the microbial rhodopsin chromophore describing the photoisomerization reaction. The absorption of a photon puts the chromophore in the electronically excited state ($S_0 \rightarrow S_1$ transition). From the Franck-Condon (FC) region, the chromophore relaxes to a minimum on the potential energy surface of the S_1 state (S_1 min). From the minimum there is either a transition to the ground state S_0 with photon emission or another transition to the conical intersection (CI). At the conical intersection an non-radiative transition to the S_0 potential energy surface occurs and the reaction path branches, leading either to a successful isomerization or back to the original form of the chromophore.

The function of microbial rhodopsins is considered using the example of the archaerhodopsin-3 homolog - bacteriorhodopsin from archaea *Halobacterium* *salinarum*, which is a well-studied representative of this protein family. The main function of bacteriorhodopsin is to transport a proton across the cell membrane [47]. Upon photon absorption, the chromophore photoisomerizes from the all-*trans* form to the 13-*cis* form, triggering a series of conformational changes in the protein. These changes pass through several metastable states and culminate in the return of the protein to its original state. The entire cyclic process triggered by the absorption of a photon is called a photocycle. During the bacteriorhodopsin photocycle, a proton moves from the cytoplasm to the extracellular medium.

Figure 1.3 shows the three-dimensional structure of bacteriorhodopsin near the chromophore in the most thermodynamically stable state (ground state). In this state, the Schiff base is protonated. In close proximity to the Schiff base is a counterion complex consisting of two aspartic acids, D85 and D212 [39; 48]. This complex is in the unprotonated state, i.e. it carries a negative charge. The proton released into the extracellular environment during the photocycle is located on glutamic acid E204, which forms a hydrogen bond with E194. E204 and E194 form a proton release group located at the extracellular periphery of the protein (Figure 1.3). After photoisomerization and initial reorganization of the protein structure, a metastable state known as the L-intermediate is formed in which the chromophore is in the 13-*cis* form (Figure 1.4) [47]. The protonation states of the titratable residues in the L-intermediate are reported to be identical to their protonation states in the ground state of the protein photocycle [39].

The next step is the conversion of the protein into the M-intermediate, which is accompanied by the release of the proton from the chromophore; the proton is transferred to the counterion complex. At the same time, the proton of the glutamic acid E204 is transferred from the proton release group to the extracellular environment. In the next step, during the transition from the M- to the N-intermediate, the chromophore captures a proton. According to available experimental data, this proton is transferred from aspartic acid D96 [39]. The chromophore then undergoes thermal isomerization from the 13-*cis* form to the all-*trans* form during the transition to the O-intermediate [48]. A proton is transferred from the counterion complex consisting of D85 and D212 to E204 of the proton release group [48] during the thermal transition from the O-intermediate to the ground state of the protein photocycle, which lasts a few milliseconds. This transfer, according to the data available so far, is carried out by a chain of water molecules that connects D212 and E204 in the O-intermediate [49]. Thus, the proton transfer involves protonating and deprotonating several key amino acid residues of the protein and chromophore, as well as proton transfer along the water chains.

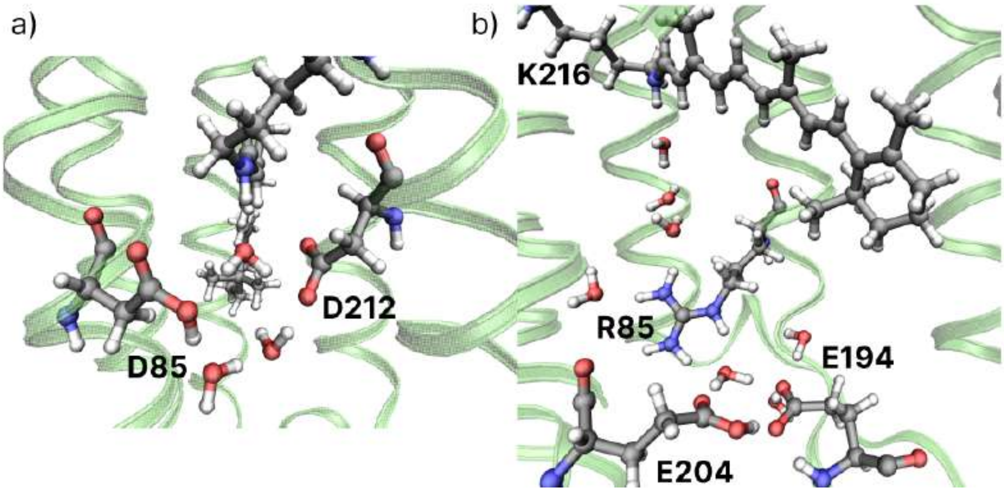


Figure 1.3 — The three-dimensional structure of bacteriorhodopsin in its ground state. a) Protein region near the Schiff base and counterions D85 and D212. b) Protein region near the proton release group E204-E194.

In initial studies, the choice of amino acid substitutions to improve the properties of archaerhodopsin-3 as a fluorescent membrane potential sensor was based on the similarity of this protein's photocycle to that of bacteriorhodopsin [50]. As mentioned above, wild-type archaeorhodopsin-3 has three major drawbacks that limit its applicability as a sensor. One of these is the generation of a proton current upon exposure to light. In order to disable this function, a version of the protein with the amino acid substitution D95N (D85N in the bacteriorhodopsin numbering) was obtained, i.e. one of the counterions was replaced by a polar amino acid, asparagine [16]. The resulting Arch D95N protein did not generate a proton current under illumination. The absorption maximum of the protein was shifted to 585 nm [16]. Another advantage of Arch D95N compared to the wild-type protein was the increased sensitivity of the fluorescence intensity to changing potential. When the potential was increased by 100 mV, the fluorescence intensity of the protein increased by 60%. This is twice as much as for the wild-type protein [16]. However, the fluorescence response time to potential change was significantly increased by the amino acid substitution. A value of 42 ms was obtained for Arch D95N [16], whereas the response time for the wild-type protein did not exceed 0.5 ms. It was further shown that for Arch D95N the kinetics of the fluorescence response to voltage

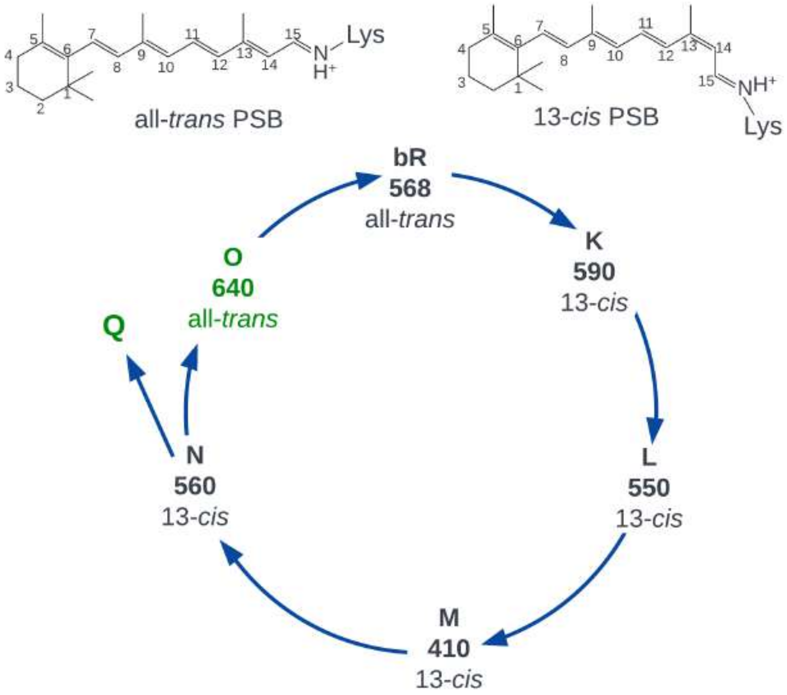


Figure 1.4 — The scheme of bacteriorhodopsin photocycle.

changes is not described by a simple exponential relationship and consists of two components, a fast component (20% of the response) with a time constant of less than 1 ms and a slow component (80% of the response) with a time constant of 36 ms [38]. The recorded value of the fluorescence quantum yield of Arch D95N was only 0.04% [16]. Further attempts were made to create new sensor variants by replacing not only the counterion D95, but also the aspartic acid D106, from which a proton is transferred to the Schiff base during the protein photocycle [6]. The proposed variants Arch D95N/D106N and Arch D95Q/D106N showed no improvement in signal intensity and response times were in the ∼ 5 - 15 ms range.

To make further progress, the mechanism of the voltage dependence of fluorescence in Arch was investigated. The results of the experiments led to the hypothesis that the recorded fluorescence is emitted by a protein photocycle intermediate with an absorption maximum at 570 nm – the Q-intermediate (Figure 1.4) [38]. Fluorescence activation requires the absorption of three photons – the first photon initiates the photocycle, the second photon transfers the N-intermediate to the Q-intermediate, and finally the third photon converts the Q-intermediate to the electronically excited state, from which relaxation occurs with the emission of a photon. The appearance of fluorescence in the photocycle intermediate is consistent with the nonlinear increase in fluorescence intensity with increasing illumination intensity. It should be noted that the proposed scheme explains the differences in the two fluorescence quantum yield values obtained for archaerhodopsin-3 [16; 17]. When the protein was excited at the absorption band maximum, a value of 0.01% [17] was obtained – this value corresponds to the ground state of the photocycle. Under intense irradiation at 633 nm, a value of 0.09% was obtained [16] – it corresponds to the Q-intermediate. The described scheme allowed the authors to propose the mechanisms of fluorescence voltage dependence in Arch. According to the proposed scheme, the fluorescence intensity is determined by the concentration of Q-intermediate, which is directly proportional to the concentration of N-intermediate. In turn, the N-intermediate is in equilibrium with the M-intermediate, which is regulated by the electric field. In both N- and M-intermediates, the chromophore is in the 13-*cis* form, but its protonation states differ [38]. The membrane potential can shift the equilibrium between forms with protonated (N) and non-protonated (M) Schiff bases.

The information obtained on the fluorescence voltage dependence was not sufficient for the rational design of brighter archaerhodopsin-3 mutants, and further steps were taken using directed evolution. The group of Prof. Adam E. Cohen used directed evolution to develop two Arch-based sensors, QuasAr1 and QuasAr2, each containing five amino acid substitutions relative to Arch. For both proteins, a linear increase in fluorescence intensity with increasing illumination intensity was observed. Based on these data, it was concluded that the fluorescence of these proteins occurs upon photoexcitation of the ground state of the protein photocycle rather than the photocycle intermediate. Both proteins had coincident band maxima of the fluorescence excitation and fluorescence spectra – 590 nm and 715 nm, respectively [13]. A linear increase in protein fluorescence intensity with increasing voltage was observed in the range from -100 mV to +50 mV. QuasAr2 had a greater voltage sensitivity of fluorescence – when the voltage was increased by 100 mV, the fluorescence intensity increased by 90%, while for QuasAr1 this value was only 32% [13]. At room temperature, the response time of QuasAr1 to a voltage change was only 0.053 ms, while for QuasAr2 the response kinetics were described by a biexponential relationship with time constants of 1.2 ms (68% contribution) and 11.8 ms (32% contribution), respectively [13]. QuasAr1 and QuasAr2 also showed no proton current under light exposure, no phototoxicity, and high photostability of the signal – the decay times under illumination with 300 W/cm^2 light at a wavelength of 640 nm were 440 s and 1020 s, respectively [13].

The group of Prof. Frances H. Arnold also performed a directed evolution of archaerhodopsin-3 [14]. A variant of archaerhodopsin-3 with Arch D95E/T99C amino acid substitutions was chosen as the starting protein for directed evolution. These amino acid substitutions were previously obtained by the same group as a result of directed evolution of *Gloebacter violaceus* rhodopsin aimed at red-shifting the absorption band of the protein [51]. For *Gloebacter Violaceus* rhodopsin, the aforementioned amino acid substitutions (in the numbering of archaerhodopsin-3) resulted in a shift of the absorption band at +80 nm and the appearance of fluorescence with a quantum yield comparable to that of QuasAr1. Directed evolution of archaerhodopsin-3 has proposed several fluorescent mutants (Archers), for which a high sensitivity of fluorescence to pH was found [14]. For two proteins, Arch D95E/T99C (Archer1) and Arch D95E/T99C/A225M (Archer2), additional experiments were performed in nematode *C. elegans* neurons [52]. Both proteins were shown to be characterized by high voltage sensitivity of fluorescence – when the potential was increased by 100 mV, the fluorescence intensity increased by 85% and 60% for Archer1 and Archer2, respectively.

In other work by the group of Professor Edward S. Boyden at the Massachusetts Institute of Technology, directed evolution of archaerhodopsin-3 was performed on several parameters – fluorescence brightness, membrane localization efficiency, fluorescence voltage sensitivity, and response time [18]. The first step was to generate a library of QuasAr2 variants with a stochastic distribution of additional amino acid substitutions. During the screening step, the signal from cells of the HEK293 cell line expressing the proteins was recorded. Cells showing increased brightness and high levels of protein localization in the plasma membrane were selected in each round of the procedure. The fluorescence voltage sensitivity of the sensors was also tested by measuring the percentage increase in fluorescence intensity at a fixed value of electrical potential applied to a cuvette containing cells. Several generations of directed evolution resulted in the selection of two variants, named Archon1 and Archon2, which were observed to have high signal brightness and fast response kinetics. Increasing the voltage by 100 mV resulted in a 43% and 19% increase in fluorescence intensity for Archon1 and Archon2, respectively [18]. For the proteins, a linear increase in fluorescence intensity with increasing illumination intensity was found, suggesting the involvement of a single, thermodynamically most stable state of the protein rather than photocycle intermediates. Archon1 showed high photostability – after 900 seconds of continuous illumination at 800 mW/mm^2, the signal decreased by only 5% [18].

The next step in improving the properties was taken for QuasAr2 – it was necessary to improve the degree of plasma membrane localization of the protein. By stochastic search of amino acid substitutions, a variant named QuasAr3 was found to have a high expression level and good localization in the plasma membrane of neurons [53]. The only amino acid substitution that distinguishes QuasAr3 from QuasAr2 is K171R, which is located at the ubiquitin interaction site on the intracellular loop. Incorporation of an additional amino acid substitution V49A (paQuasAr3) into QuasAr3 resulted in the ability to increase fluorescence intensity threefold when illumination with 488 nm blue light was added to illumination with 640 nm light. The mechanism responsible for this effect is currently unknown.

To enhance the effect of increased fluorescence intensity under simultaneous illumination with red and blue light, the next step of directed evolution was performed on paQuasAr3. A NovArch variant, paQuasAr3 with additional amino acid substitutions of V209I and I213T [15], was obtained. When illuminated with 640 nm light, the same fluorescence intensity was observed for paQuasAr3 and NovArch, but the addition of blue light increased the fluorescence of NovArch sevenfold. NovArch also possessed a red-shifted fluorescence excitation band maximum relative to QuasAr1/2 and Archon1/2, which was located at 620 nm and its position did not change upon the addition of blue light [15].

Finally, an additional directed evolution improvement was performed on Archon1 [9]. Two parameters – fluorescence brightness and percentage increase in fluorescence intensity at the onset of an action potential – were improved. Screening was performed in the HEK293T cell line. The action potential was generated by voltage gated sodium and potassium channels embedded in these cells. The channelrhodopsin was also incorporated into the cells to activate the action potential by illuminating the cells with blue light – under blue light, the protein transported ions, resulting in the voltage change necessary to activate the voltage gated sodium and potassium channels and thus generate the action potential. This study resulted in two variants, called QuasAr6a and QuasAr6b, that were 1.7 and 2 times brighter than Archon1. The QuasAr6a also demonstrated high voltage sensitivity of fluorescence – a 70% increase in fluorescence intensity upon a 100 mV voltage increase. Fast kinetics of fluorescence response to voltage change were recorded for the protein. The activation time constant was $\tau(\text{on}) = 1.8 \pm 0.5$ ms, and the deactivation time constant was $\tau(\text{off}) = 1.3 \pm 0.5$ ms [9]. Although QuasAr6b was brighter than QuasAr6a, its signal was less sensitive to voltage changes – only a 24% increase in fluorescence was observed with a 100 mV voltage increase. For both proteins, the authors demonstrated a linear dependence of fluorescence intensity on voltage over the range -70 mV to +30 mV, as well as the absence of proton current upon illumination with red or blue light[9].

Figure 1.5 shows the overall evolution of the Arch-based voltage sensors, and Table 1 compares their properties. Note that all sensors were obtained by directed evolution. Although several studies on the mechanisms of potential-dependent fluorescence have been performed [54], their results could not be used to propose new Arch variants with improved performance. It should also be noted that despite the significant improvement in the optical properties of the new Arch variants compared to wild-type archaerhodopsin-3, the available variants still exhibit too weak fluorescence, making them difficult to use *in vivo*. The aim of this work was to determine the mechanism responsible for the increase in fluorescence and to use this information for the rational design of new brighter Arch-based sensors. As a starting point, literature data on fluorescence in microbial rhodopsins were analyzed and summarized below.

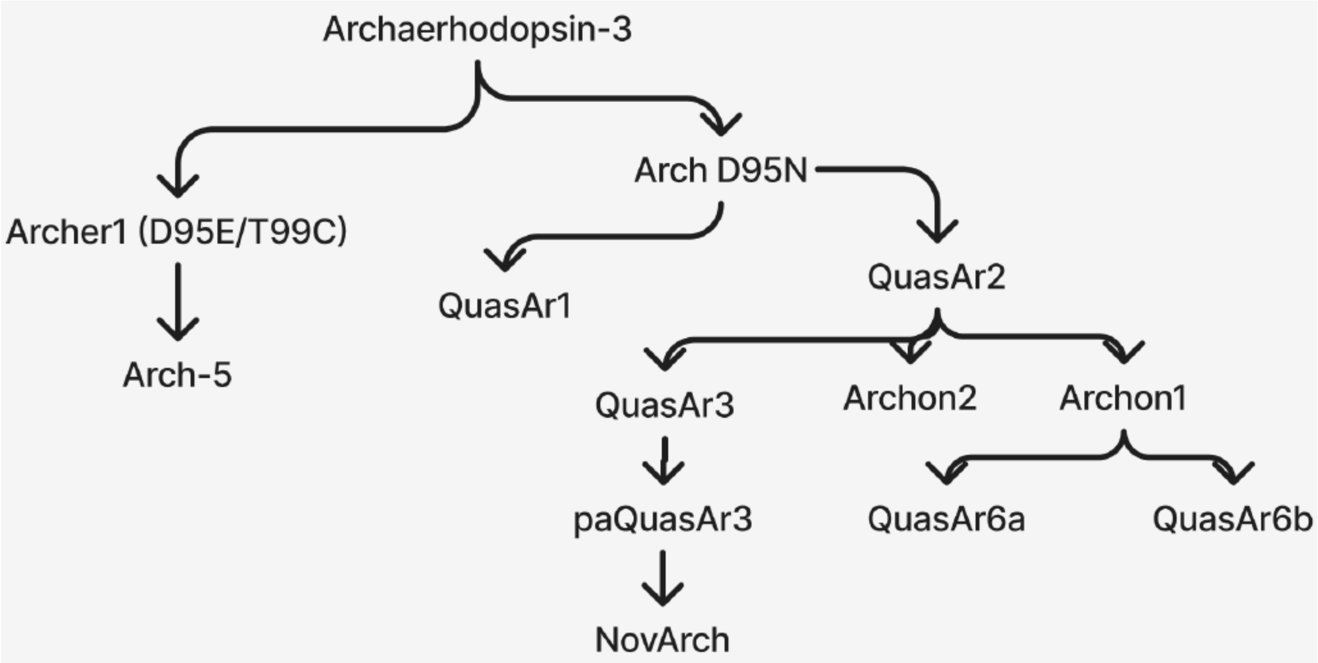


Figure 1.5 — The archaerhodopsin-3-based fluorescent sensors of cell membrane potential obtained to date.

Table 1 — Characteristics of archaerhodopsin-3-based genetically encoded voltage indicators published to date. λ_{max} - absorption band maximum, λ_{em} - fluorescence band maximum, Φ - fluorescence quantum yield, $\Delta F/F$, 100 mV – percentage change in fluorescence intensity with a 100 mV change in voltage.

Protein	λ_{max}	λ_{em}	Φ,%	$\Delta F/F$, 100 mV, %
Arch	558[16]	687[16]	0.01[16]	30[16]
QuasAr1	585[13]	715[13]	0.8[13]	32[13]
QuasAr2	587[13]	715[13]	0.4[13]	90[13]
Archon1	585[18]	735[18]	0.5[18]	80[18]
Archon2	586[18]	735[18]	-	19[18]
NovArch	-	725[15]	-	41[15]
Archer1	626[14]	729[14]	0.33[14]	85[14]
Arch-5	622[14]	745[14]	0.87[14]	-

1.3 Analysis of literature data on fluorescence in microbial rhodopsins

Figure 1.6a shows the dependence of the logarithms of the fluorescence quantum yields (log Φ) on the absorption maxima (λ_{max}) of the microbial rhodopsins [23]. The following color scheme is used in the figure. Data for wild-type microbial rhodopsins measured at pH $\approx$ 7 are shown as blue dots. Data for wild-type microbial rhodopsins measured at acidic or basic pH are shown in green. Green dots also indicate data for microbial rhodopsins with amino acid substitutions, except for light variants of microbial rhodopsins obtained by directed evolution, which are shown in brown. Finally, the light blue dot in the figure represents the O-intermediate of the bacteriorhodopsin photocycle (Figure 1.4).

The fluorescence quantum yields of microbial rhodopsins vary over a wide range. Figure 1.6a clearly shows the trend of increasing fluorescence quantum yield with increasing λ_{max}. This fact suggests that the fluorescence quantum yield and the absorption maximum are controlled by the same factors. The Φ values for most wild-type microbial rhodopsins at neutral pH are in the same order of magnitude, hundredths of a percent. The fluorescence quantum yield increases significantly for acidic forms of bacteriorhodopsin (bR blue, Φ=0.45%) and sensory rhodopsin (SRI pH6, Φ=0.13%), reaching the next order of magnitude. The same increase was found for the O and Q intermediates of the bacteriorhodopsin photocycle with quantum yields of ≈0.1% and 0.7%, respectively. Since λ_{max} was not measured for the Q intermediate, the corresponding point is not shown in Figure 1.6a. The fluorescence quantum yields for three bright microbial rhodopsin variants obtained by directed evolution (Archer1, QuasAr1, and GR D121A) are in the range of tenths of a percent.

A similar pattern can be observed for the fluorescence lifetimes of microbial rhodopsins. Figure 1.6b shows the dependence of the logarithms of the fluorescence lifetimes $log(\tau_F)$ on the absorption maxima. The points corresponding to the mutants of archaerhodopsin-3 (QuasAr1 and Archon2), the O-intermediate of the bacteriorhodopsin photocycle, the bacteriorhodopsin variants with a substituted counterion (bR D85S, bR D85N), and the values obtained at acidic pH (SRI pH=6, bR R82Q pH=4) are located above the points corresponding to the wild-type rhodopsins for which measurements were performed at neutral pH. The bright archaerhodopsin-3 mutants QuasAr1 and Archon2 with the highest τ_F values reported so far are at the top of the figure.

Differences in the interaction of the chromophore with its environment can explain the observed diversity of fluorescence properties of microbial rhodopsins containing the same chromophore. Many studies [21; 55–57] have shown the influence of the amino acid environment of a chromophore on the position of its absorption band maximum. The electrostatic and steric interactions of the opsin residues with the chromophore were found to be the most important factors. The most important factor determining the position of the absorption band maximum is the electrostatic effect of counterions. These are negatively charged residues located near the positively charged C=NH+ group of the chromophore. The results of both experimental and computational modeling studies show that replacement of the counterion by a neutral residue or protonation of the counterion leads to a significant red shift of the absorption band maximum [21]. The available experimental data [58–61] also show the key role of the electrostatic effect of counterions for the fluorescence properties of rhodopsins. The data presented in Figure 1.6 show an increase in fluorescence quantum yields and fluorescence lifetimes when moving from the ground states of wild-type microbial rhodopsins to protein variants in which a neutral amino acid replaces the counterion or to protein states in which the counterion is protonated.

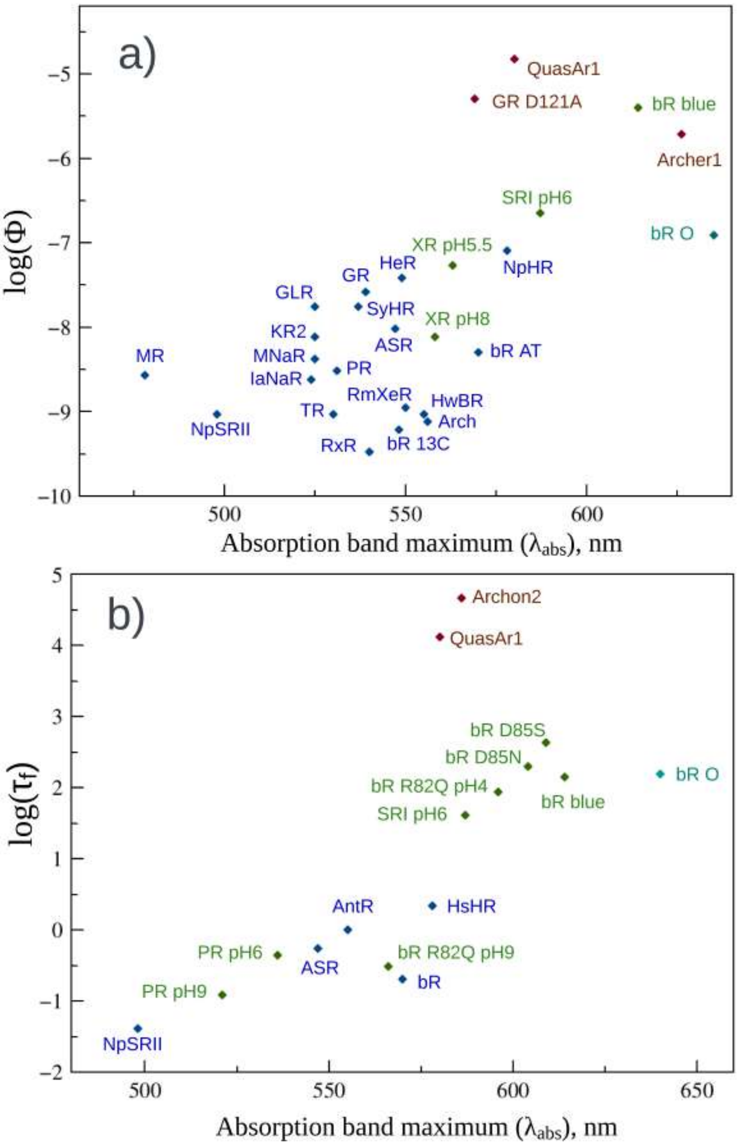


Figure 1.6 — a) The logarithms of fluorescence quantum yields of microbial rhodopsins plotted as the function of absorption maxima. b) The logarithms of fluorescence lifetimes of microbial rhodopsins plotted as the function of absorption maxima.

Available data on the pH dependence of fluorescence properties also support the relationship between counterion protonation state and fluorescence. For bacteriorhodopsin, a significant increase in fluorescence intensity was found at acidic pH values, with the most intense fluorescence occurring at a pH of about 1.7 (an increase of about 15-fold compared to neutral pH) [60]. The observed pH dependence of fluorescence intensity correlates with the pH dependence of the absorption band maximum – the maximum shifts to longer wavelengths upon acidification. In contrast, only small changes in the fluorescence intensity and the position of the absorption band maximum were observed in the alkaline pH range [60]. These results are in agreement with the pKa value of approximately 2.7 obtained for the bacteriorhodopsin counterion [62]. In a recent study, time-resolved spectroscopy revealed two components in the excited-state decay of bacteriorhodopsin at low pH < 4: ≈ 0.5 ps and 7.8 ps [63]. The slow component was hypothesized to be related to the protonated counterion form of the protein. For *Gloebacter violaceus* rhodopsin [64], xanthorhodopsin [65], *Exiguobacterium sibiricum* rhodopsin [66], an increase in emission intensity with decreasing pH was also found. For *Gloebacter violaceus* rhodopsin [64], proteorhodopsin (PR) [67; 68] and *Krokinobacter* rhodopsin 2 [69] an increase of the excited state lifetime upon acidification was reported.

In all cases, the transition from low to high fluorescence quantum yields or excited state lifetimes occurred in the pH range close to the pKa of the counterion. Furthermore, it was shown that the point of this transition can be altered by tuning the pKa of the counterion [59; 66]. For example, the pKa of the D85 counterion in bacteriorhodopsin was increased to ≈ 7.5 by replacing the positively charged arginine residue R82 with a [59] glutamine residue. For bacteriorhodopsin with the amino acid substitution R82N, a significant increase in excited state lifetime was observed when pH was decreased from 9.6 (0.6 ps) to 4.4 (equal contribution of lifetime components 2.0 ps and 7.0 ps), whereas for the wild-type protein such an increase in excited state lifetime was observed only at pH < 3, consistent with a counterion pKa of ≈ 2.7 [59]. In another study, different magnitudes of change in fluorescence intensity with decreasing pH were observed for *E. sibiricum* wild-type rhodopsin and its variant with the H57M amino acid substitution [66]. In the wild-type protein, the pKa of the D85 counterion was estimated to be less than 2.0. Lowering the pH from 7.0 to 5.0 resulted in a twofold increase in fluorescence. The pKa(D85) was increased to 6.3 by replacing the histidine residue, which interacts with the D85 counterion, with a methionine residue. By lowering the pH from 9.0 to 4.5 [66], a 100-fold increase in fluorescence intensity was observed for the protein with the amino acid substitution.

An increased counterion pKa was also observed for the O-intermediate of the bacteriorhodopsin photocycle. As mentioned above, the O-intermediate occurs in the last stage of the photocycle (Figure 1.4) and converts to the protein ground state in $\sim$ 10-30 ms. [39]. The main difference between the O-intermediate and the ground state is the protonation state of the two groups, the counterion complex consisting of aspartic acid residues D85 and D212 and the proton release group consisting of glutamic acid residues E194 and E204 [39]. In the O-intermediate, the proton is located on the counterion complex; during the transition from the O-intermediate to the ground state of the photocycle, it is transferred to the proton release group. A recent study has suggested that the proton transfer may be mediated by a chain of water molecules connecting the counterion D212 and the glutamic acid residue E204 [49]. Although direct experimental measurements of the pKa of the counterion for the O-intermediate have not been reported, bacteriorhodopsin titration showed that proton detachment from the proton release group leads to an increase in the pKa of the counterion to 7.5 [70], so this value for the pKa of the counterion can be assumed for the O-intermediate. In agreement with this assumption, the fluorescence quantum yield (0.1%) [71] and the excited state lifetime (9 ± 2 ps [71]) of the O-intermediate at neutral pH are closer to the values obtained for bacteriorhodopsin (pKa(D85) $\approx$ 2.7 [62]) at pH $\approx$ 2 (blue form) [60; 72] than at neutral pH.

Another fluorescent state, the Q-intermediate, was detected by time-resolved fluorescence spectroscopy as a state with long-lived fluorescence (62 ± 2 ps) [71]. The fluorescence quantum yield of this state was estimated to be 0.7% [71], which is 7 times higher than the value obtained for the O-intermediate. The Q-intermediate is formed by photoexcitation of the N-intermediate [73] and does not participate in the main photocycle of bacteriorhodopsin. The structure and protonation states of the titratable residues in the Q-intermediate have not been determined. However, the observed long-lived fluorescence suggests that the counterion in this intermediate is protonated. Archaerhodopsin-3, *N. pharaonius* halorhodopsin, *Krokinobacter* rhodopsin 2 (KR2), and *Rubrobacter xylanophilus* rhodopsin (RxR) [17] also showed long-lived fluorescence of photocycle intermediates. At low light source intensities, only weak fluorescence was observed, which was attributed to the ground state of the protein photocycle. However, under intense illumination, which causes the accumulation of photocycle intermediates, the fluorescence signal increased significantly. This intense fluorescence was characterized by a biexponential decay with faster and slower components τ_1 and τ_2. For the rhodopsins studied, the τ_1 and τ_2 components ranged from 5.2 to 9.6 ps and 24 to 60 ps, respectively [17]. The obtained fluorescence lifetimes are similar to the excited state lifetimes found for the O and Q intermediates of the bacteriorhodopsin photocycle and can be attributed to the O and Q intermediates of the studied proteins.

Thus, according to the available literature data, in order to increase the fluorescence intensity and shift the absorption band to the long wavelength region of the spectrum, it is necessary to stabilize a state of the protein that lacks negatively charged residues near the C=NH+ group of the chromophore – counterions. If such counterions are present, they should be replaced by neutral amino acid residues, or it is necessary to increase the pKa of the counterions to values above the physiological pH by introducing other amino acid substitutions. Prior to this dissertation study, the mechanism of fluorescence enhancement in bright mutants of archaerhodopsin-3 (Archer1, QuasAr1/2, and Archon1/2), which are now used as genetically encoded voltage indicators, was not fully understood. In this dissertation study, the mechanism leading to increased fluorescence in the Archer group of membrane potential sensors was determined based on the results of experimental studies, computational modeling, and literature data described in this chapter. The information obtained was used to guide the design of novel membrane potential sensors, mutants of archaerhodopsin-3 from the Archer group. The results of this study are described in chapters 3 and 4 of this dissertation.

Chapter 2. Methods

2.1 Approach to protein design

Novel archaerhodopsin-3-based genetically encoded voltage indicators were developed using a combination of directed evolution and rational design approaches. Directed evolution is an approach to protein engineering in which the modification of protein properties is accomplished by mimicking the process of natural evolution [74; 75]. A library of protein variants is created. Each variant has a stochastic distribution of amino acid substitutions. Based on the screening of the obtained library, the protein variants with the greatest improvement in modifiable properties are selected and used to generate a new library of protein variants. The process is iterative. It is terminated when switching to a new library does not result in improved protein properties over several iterations. The advantage of directed evolution is the ability to obtain improved variants of a protein without knowing its three-dimensional structure and the mechanism regulating the optimized property. However, the method requires a very large number of variants, and the probability of obtaining an optimal combination of amino acid substitutions is not very high.

Rational design is an approach to protein engineering in which data on the three-dimensional structure of the protein and the molecular mechanisms that determine its properties or functions guide the selection of amino acid substitutions to modify the properties or functions of a protein. The advantage of rational design is its efficiency. A small number of amino acid substitutions, selected on the basis of structure and mechanism analysis, are introduced by site-directed mutagenesis and result in the desired modification of protein properties. The disadvantage of rational design is that it is difficult to implement in most cases. To determine the molecular mechanism responsible for the change in protein properties, highly accurate computational models of the protein are required.

The combination of rational design and directed evolution approaches, taking advantage of their respective strengths, was realized as follows. The set of proteins resulting from directed evolution was divided into two groups – those with improved properties and those without (Figure 2.1). Then, a comparative analysis of the two groups of proteins was performed using experimental methods and computational modeling. The structural feature whose modification led to the improvement of the properties of the proteins from the second group was determined. The subsequent rational design of proteins with improved properties was based on this structural characterization information.

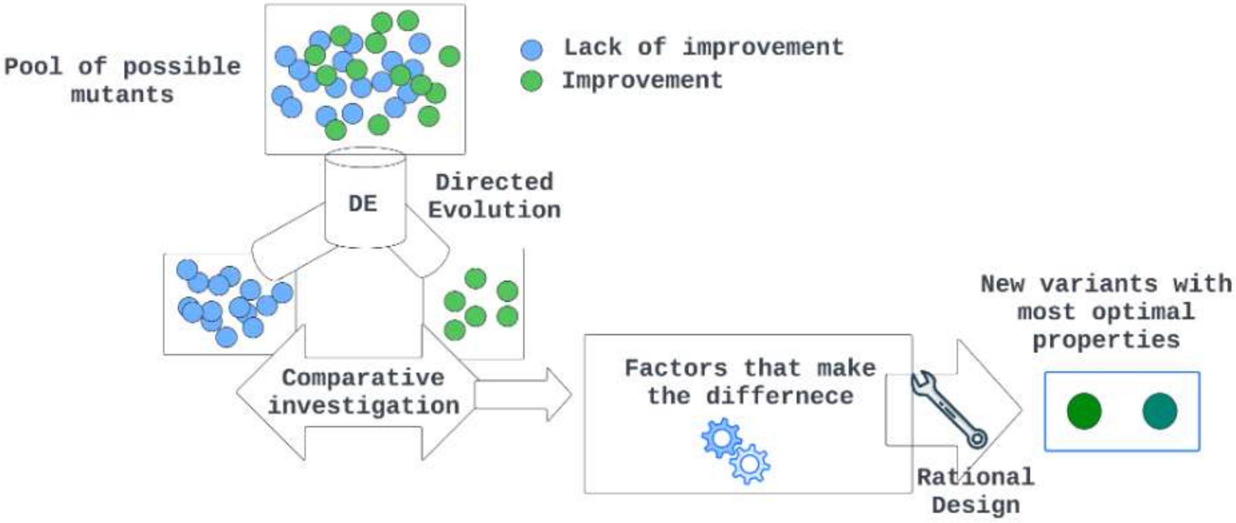


Figure 2.1 — The protein engineering approach that combines directed evolution and rational design.

2.2 Computational modeling

In this thesis study, computational modeling techniques were used to determine the structural feature whose alteration results in increased fluorescence of archaerhodopsin-3 mutants from the Archers group relative to the wild-type protein. Three-dimensional models were constructed and absorption maxima were calculated for all proteins considered in this study.

2.2.1 Constructing three-dimensional models and calculating spectral properties of microbial rhodopsins and their mutants.

For wild-type archaerhodopsin-3, bacteriorhodopsin, and the O-intermediate of bacteriorhodopsin with amino acid substitution L93A, models obtained by X-ray diffraction analysis were used as initial three-dimensional structures (archaerhodopsin-3: PDB ID 6GUX, resolution 1.3 Å [76], bacteriorhodopsin: PDB ID 5ZIM, resolution 1.25 Å [77], O-intermediate of bacteriorhodopsin with amino acid substitution L93A: PDB ID 3VI0, resolution 3.3 Å [78]).

For mutants of archaeorhodopsin-3, no experimental data on the three-dimensional structure were available. Therefore, the initial three-dimensional structures of the proteins were obtained by homology modeling. In homology modeling, the known three-dimensional structure of a homologous protein is used as an initial template. The differences between the amino acid sequences of the desired protein and the template protein are determined prior to 3D structure prediction by aligning the amino acid sequences of the two proteins. An example of the alignment of the amino acid sequences of two proteins is shown in Figure 2.2. The degree of homology (similarity) of two proteins is determined by calculating the ratio of the number of identical amino acids at identical sites in the alignment to the total number of sites in the alignment. The higher the value of the amino acid sequence identity of the target protein and the template, the higher the degree of homology and the higher the probability of obtaining the correct three-dimensional structure by homology modeling.

The amino acid sequence correspondence information of the target protein and the template protein, as well as the 3D structure information of the template protein, are used by homology modeling algorithms to generate an initial 3D model of the target protein[79]. To find differences in the three-dimensional structure caused by differences in the amino acid sequences of two proteins, a search for the most thermodynamically stable conformation of the target protein is performed, which often involves generating a large number of possible configurations and then evaluating the thermodynamic stability of each of them. This results in the final three-dimensional structure of the target protein. To date, a large number of homology modeling methods have been developed. To select the most optimal

```
4GD3_ChainA__   ----------------------VVSHYVFEAPVRIWHWLTVLCMAVLMVTGYFIGKPLPS
2ZT9_ChainA__   MANVYDWFEERLEIQAIAEDVTSKYVPPHVNIFYCLGGITLVCFLIQFATGFAMTFYYKP
                                                       *  *      **


4GD3_ChainA__   VSGEATYLFYMG--------YIRLIHFSAGMVFTVVLLMRIYWAFVGNRYSRQGVWYEIR
2ZT9_ChainA__   TVAEAYSSVQYIMNEVNFGWLIRSIHRWSASMMVLMMILHVFRVYLTGGFKKP-------
                   **                ** **


4GD3_ChainA__   WYLFLPIAQAAMFGYFLMSVFMIITGFALYSEHSQYAIF------------APFRYVVEF
2ZT9_ChainA__   -------RELTWVSGVILAVITVSFGVTGYSLPWDQVGYWAVKIVSGVPEAIPVVGVLIS
                                   *     *   **                     *   *


4GD3_ChainA__   FYWTGGN------SMDIHSWHRLGMWLIGAFVIGHVYMALREDIMSD--
2ZT9_ChainA__   DLLRGGSSVGQATLTRYYSAHTFVLPWLIAVFMLFHFLMIRKQGISGPL
                    **            * *        *          *    *

Sequence identity: 10.04%
Matched Positions: 72.05%
```

Figure 2.2 — Example of alignment of two amino acid sequences.

method, a comparative analysis of three widely used methods implemented in Medeller [80], Rosetta [81], and I-TASSER [82] programs was performed as well as three methods of amino acid sequence alignment implemented in the programs MP-T [83], AlignMe [84] and MUSTER [85]. All unique structures of microbial rhodopsins obtained by X-ray diffraction analysis at the time of the study (2017) were used as a test set. From this set, pairs of rhodopsins with different degrees of homology were obtained, and predictions of the three-dimensional structure of the first protein in the pair were made based on the three-dimensional structure of the second protein, and vice versa. The quality of the obtained structures was evaluated by comparing the obtained model with the corresponding crystallographic structure of the protein. Based on the obtained data, the algorithm for generating three-dimensional structures implemented in the I-TASSER program and the algorithm for aligning amino acid sequences implemented in the AlignMe program were selected as the most reliable and used within the project. The results of the study were published [86].

For all models, both crystallographic structures and those predicted by homology modeling, the positions of water molecules in protein cavities were predicted. Water molecules in protein cavities can play important functional roles, such as participating in proton transfer and stabilizing a particular protein conformation by hydrogen bonding with protein residues [87; 88]. Often, information about all the water molecules in the protein cavities is not included in protein structures obtained by X-ray diffraction analysis. Both insufficient resolution and the presence of mobile water molecules that are not localized at a particular position in the protein may account for the absence of some water molecules in the crystallographic structures [89; 90]. The protein structures that are obtained by homology modeling do not contain any water molecules. Therefore, predicting the position of water molecules in protein cavities is a necessary step in constructing a complete model of protein three-dimensional structure. The process can be divided into two steps. 1. Search for possible positions for water molecules. 2. Estimating the probability of finding a water molecule in each of the positions found. The search for possible positions for water molecules consists of finding all cavities in the protein structure that are large enough to accommodate one or more water molecules. For example, in the method implemented in the Dowser++ program [91], the protein structure is divided into a set of cubic regions, and in each region a standard molecular docking procedure is started, in which a water molecule acts as a ligand.

The second step is to estimate the probability of finding water at the found position in the protein. This requires calculating or estimating the Gibbs free energy to transfer a water molecule from aqueous medium to a given position in the protein cavity. There are approaches that include the calculation of this property by means of Free Energy Perturbation (FEP), thermodynamic integration and their analogues [92; 93]. Since the calculation of the free energy requires the enumeration of a large number of configurations of the molecular system, which is done by the molecular dynamics method or the Monte Carlo method, these methods require significant computational resources. Therefore, they are rarely used in practice. The use of empirical Gibbs free energy estimation functions is a better approach. Widely used software packages such as Dowser [94] and WaterDock [95] implement empirical functions for free energy estimation. In order to select the optimal method, a comparative study of four widely used algorithms implemented in the software packages Dowser [94], Dowser++ [91], WaterDock [96], WaterDock-2.0 [97] was conducted. A set of all unique rhodopsin crystallographic structures known at the time of the study (2020) was used. The total number of predicted water molecules in protein cavities, the ratio of the total number of predicted water molecules to the number of water molecules in crystallographic structures, and the ability to predict functionally significant clusters consisting of several water molecules connected by hydrogen bonds to each other and to neighboring residues were used as metrics. Based on the comparison, the method implemented in Dowser++ was selected as the most successful. An additional parametrization for the Dowser++ algorithm was performed in the study [20].

In the next step, the protein structure was optimized using a hybrid quantum mechanics/molecular mechanics (QM/MM) method. A two-level ONIOM (QM:MM-EE) scheme was used (QM=B3LYP/6-31g*; MM = Amber99 for amino acids and ions, TIP3P model for water molecules, EE is electron attachment) [98]. The quantum mechanical part of the model included 49 chromophore atoms and one NH group covalently bonded to the chromophore through a lysine residue. The molecular mechanical part included the rest of the protein and all the water molecules. A linker hydrogen atom was used to describe the boundary between the QM and MM parts of the model. The Gaussian09 [99] software package was used for all QM/MM calculations.

The resulting models were used to calculate the absorption band maximum of the protein. The QM/MM method SORCI+Q(6,6)/6-31g*/Amber99 was used to calculate the vertical transition energy ($S_0 \rightarrow S_1$). The calculation was done in ORCA 3.0.3 [100]. The methods used to construct rhodopsin models and calculate absorption band maxima have been previously verified in a large number of studies, including those performed in this work [20; 57; 101].

2.2.2 Taking into account the mobility of protein environment in rhodopsin models

Standard three-dimensional models of proteins consider only one conformation of the protein. However, the geometry of the cofactor and its spectral properties can be affected by the mobility of the protein at room temperature. To account for protein mobility in the thermodynamic equilibrium state for rhodopsins, a model of the average configuration of the amino acid environment of the cofactor was proposed and programmatically implemented. This model generates an electrostatic field in the cofactor region that is averaged over the existing configurations of the amino acid environment (Average Solvent Electrostatic Configuration (ASEC)) [102]. To build an ASEC model, a set of different configurations of the amino acid environment of the cofactor must be generated. These configurations are then averaged together. Due to the fact that the cofactor is most affected by the motion of its immediate environment, the ASEC model was built only for residues and water molecules whose distance from any of the chromophore atoms did not exceed 8 Å. A molecular dynamics method was used to generate the configurations using the Amber99 [103] molecular mechanical force field. When calculating the molecular dynamics trajectory, the position of the chromophore atoms and the position of all atoms of residues and water molecules more than 8 Å away from any of the chromophore atoms were fixed. Molecular dynamics calculations were performed using the Gromacs 2019.1 [104] software package. From the resulting molecular dynamics trajectory, 100 uncorrelated protein environment configurations were excised and structurally overlapped to construct an ASEC model. Thus, 100 pseudoatoms corresponding to 100 different configurations were used in the ASEC model to describe each of the atoms surrounding the chromophore (8 Å from the chromophore). The parameters of the atoms that determine the non-valent interactions - the Coulomb and van der Waals interactions - were recalculated to obtain standard values for the interaction energies between the environment and the chromophore. Residues and water molecules whose distance from any of the chromophore atoms was greater than 8 Å were added to the ASEC model without modification.

To account for the effect of the dynamics of the protein environment on the chromophore structure, a QM/MM optimization of the chromophore geometry in the protein environment described by the ASEC model was performed. The quantum mechanical part of the model included 49 chromophore atoms and one NH group covalently attached to the chromophore through a lysine residue. The mobile molecular mechanics part included the rest of the side chain of the lysine residue. A linker hydrogen atom was used to describe the boundary between the QM and MM parts of the model. The remaining environment of the chromophore was described as point charges of atoms and pseudo-atoms of the ASEC environment. The B3LYP/6-31g*/Amber99 method was used for the calculations. The software package Molcas 7.2 [105] combined with the software package Tinker 5.1 [106] was used for the QM/MM optimization. After the geometry optimization the calculation of the vertical transition energy $S_0 \rightarrow S_1$ for a chromophore surrounded by ASEC model point charges was performed, the calculation method was SORCI+Q(6,6)/6-31g*, the calculation was performed in ORCA 3.0.3 [100].

Previous theoretical work has shown that the potential energy gradient of a molecular system in the averaged environment of the medium is equal to the free energy gradient (FEG) of the molecular system [107]. Thus, when optimizing the chromophore geometry in the ASEC environment, its Gibbs free energy is optimized, and the resulting QM/MM ASEC-FEG models of rhodopsins correspond to minima on the Gibbs free energy surface of the molecular system. The resulting models can be used, for example, to calculate the difference in thermodynamic stability of two forms of rhodopsin that differ in chromophore conformation. Various methods can be used to calculate the free energy difference between conformations, such as the [108] free energy perturbation method. To verify the QM/MM ASEC-FEG models of rhodopsins, absorption band maxima and Gibbs free energy difference calculations were performed for visual rhodopsin and bathorhodopsin – the first intermediate in the visual rhodopsin photocycle – as well as the transition state for the reaction corresponding to the thermal transition between these forms. Calculations were also performed for the all-*trans* and 13-*cis* forms of *Anabaena* sensory rhodopsin and the transition state corresponding to the thermal isomerization reaction. The QM/MM ASEC-FEG models allowed us to obtain good agreement of the calculated absorption band maxima with experiment, and also good agreement of the Gibbs energy differences between the rhodopsin forms. Thus, the developed QM/MM ASEC-FEG models of rhodopsins were verified, and the verification results and details of the developed method were published [22].

2.3 Experimental methods.

The synthesis of the rhodopsins studied in this work was performed by colleagues of the author of the thesis, and the synthesis methods are described in detail in the paper published based on the results of the thesis study [19].

Steady-state UV-visible absorption spectra were recorded from purified proteins dissolved in Tris buffer (10 mM Tris, 200 mM NaCl, 0.15%

n-dodecyl-β-d-maltopyranoside (DM), pH=6.5) at room temperature on a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific). Dark-adapted samples were prepared by incubation in the dark for 1 hour prior to measurements. Light-adapted samples were prepared by irradiation with a 5 mW 532 nm laser for 10 minutes immediately prior to measurements.

Raman spectra of purified proteins dissolved in Tris buffer were recorded at room temperature on a SENTERRA Express Raman spectrometer (Bruker) equipped with a 50× objective using 100 mW excitation at 785 nm and 40 mW excitation at 532 nm. Excitation and fluorescence emission spectra were recorded using a Fluorolog-3 spectrofluorimeter (Horiba Jobin Yvon) for purified protein dissolved in Tris buffer.

The Quanta-ϕ integrating sphere of the Fluorolog-3 spectrofluorimeter (Horiba Jobin Yvon) was used to measure the fluorescence quantum yield of Arch D95E/T99C. The obtained fluorescence quantum yield (0.33%) is in agreement with previously published data of other authors. For the other proteins studied, the comparative method was used, with Arch D95E/T99C as reference. For each protein and reference, solutions with different protein concentrations corresponding to OD(λ_{max}) = 0.15, 0.1, 0.07, 0.05 and 0.03 were obtained. For each solution, the fluorescence spectrum was recorded under excitation at the wavelength of the absorption band maximum. The fluorescence spectra were obtained on a Chirascan-plus spectrofluorimeter (Applied Photophysics). The integral fluorescence intensity of each solution was determined by calculating the area under the wavelength dependence curve of the fluorescence intensity. The dependence of the integral fluorescence intensity on OD(λ_{max}) was plotted for each protein and reference and approximated by a linear function. The following equation was used to derive fluorescence quantum yield values from the obtained slope angles of the linear approximations (S): $\Phi_{protein} = \Phi_{D95E/T99C} \cdot (S_{protein}/S_{D95E/T99C})$, where $\Phi_{protein}$ is the fluorescence quantum yield of the studied protein, $\Phi_{D95E/T99C}$ is the fluorescence quantum yield of Arch D95E/T99C, $S_{protein}$ and $S_{D95E/T99C}$ are slope angles for linear approximations of the dependence of the integral fluorescence intensity on OD(λ_{max}) for the studied protein and Arch D95E/T99C, respectively.

In order to assess fluorescence voltage dependence of the new proteins, fluorescence spectra were recorded from live *E. coli* cells expressing sensors and from the same cells after their death from antibiotic or sodium azide action. The observed increase in the integral fluorescence intensity after cell death was then correlated with the dependence of fluorescence on the value of the cell membrane potential.

Chapter 3. The structural determinant of the significant increase in fluorescence intensity of archaerhodopsin-3-based genetically encoded voltage indicators from the Archers group compared to the wild-type archaerhodopsin-3

3.1 Introduction

The main objective of the thesis study was to develop new archaerhodopsin-3-based sensors of cell membrane potential with enhanced fluorescence signal. A combined approach of directed evolution and rational design methods was used to solve the problem. We considered a set of proteins, including wild-type microbial rhodopsins and mutants of archaerhodopsin-3 with increased fluorescence intensity, obtained by directed evolution in the work of Prof. Frances H. Arnold's group [14]. The proteins were divided into two groups – those with increased fluorescence intensity and those without. In the next step, a comparative analysis of these two groups of proteins was performed using experimental methods and computational modeling. The structural feature that was associated with the increase in fluorescence intensity was found. The structural information obtained was used for the subsequent rational design of proteins with increased signal intensity. This part of the study is described in Chapter 4 of this dissertation.

3.2 Differences in spectral properties of the wild-type archaerhodopsin-3 and mutants of archaerhodopsin-3 from the Archers group

According to the available information on the fluorescence of microbial rhodopsins and their spectral properties, one of the main factors determining the fluorescence quantum yield and the position of the absorption band maximum of these proteins is the electrostatic field generated by residues near the chromophore. Changing the electrostatic field is achieved by introducing or removing polar or charged residues, or by changing the protonation state of residues. Another spectral feature of microbial rhodopsins, which is determined by the electrostatic field near the chromophore and for which a correlation with the position of the absorption band maximum is observed, is the frequency of symmetric C=C ethylenic stretch vibrations of the chromophore (Figure 3.1) [109]. In Raman spectra, this oscillation corresponds to an intense band with a maximum located in the 1500-1540 cm^{-1} region for the chromophore in the protonated state [110–112] and near 1565 cm^{-1} for the chromophore in the non-protonated state [112].

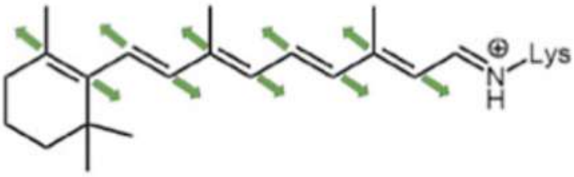


Figure 3.1 — Schematic representation of symmetric C=C ethylenic stretch vibrations of the chromophore.

The measured steady-state absorption spectra of dark-adapted samples of wild-type archaerhodopsin-3 and protein variants with amino acid substitutions D95E/T99C, P60L/D95E/T99C, V59A/D95E/T99C, D95E/T99C/P196S, V59A/P60L/D95E/T99C/P196S, as well as Raman spectra of dark-adapted and light-adapted forms of the proteins revealed the difference between the spectral characteristics of the Arch mutants and the wild-type protein. Namely, the absorption band maxima (λ_{max}) of Arch mutants are located in the range of 622-628 nm, while the absorption band maximum of the wild-type protein is located at 556 nm (dark-adapted state). For the Arch mutants, the maximum of the Raman spectral band corresponding to the symmetric C=C ethylenic stretch vibrations ($\lambda_{C=C}$) was shifted to the red region with respect to the wild-type protein (1508 to 1516 cm^{-1} for the mutants, 1527 cm^{-1} for the wild-type protein). All the data obtained are presented in Table 2 and Figures 3.2, 3.3. The data suggest a significant difference in the electrostatic fields generated by residues of opsin near the chromophore for the studied fluorescent mutants of archaerhodopsin-3 and the wild-type protein.

The spectral properties of archaerhodopsin-3 and its mutants from the Archers group were compared with the spectral properties of other microbial rhodopsins and intermediates of microbial rhodopsin photocycles for which literature data were available. All values of $\lambda_{C=C}$ and λ_{max} were plotted – $\lambda_{C=C}$ on the y-axis and λ_{max} on the x-axis (Figure 3.4). The points in the resulting graph fall into

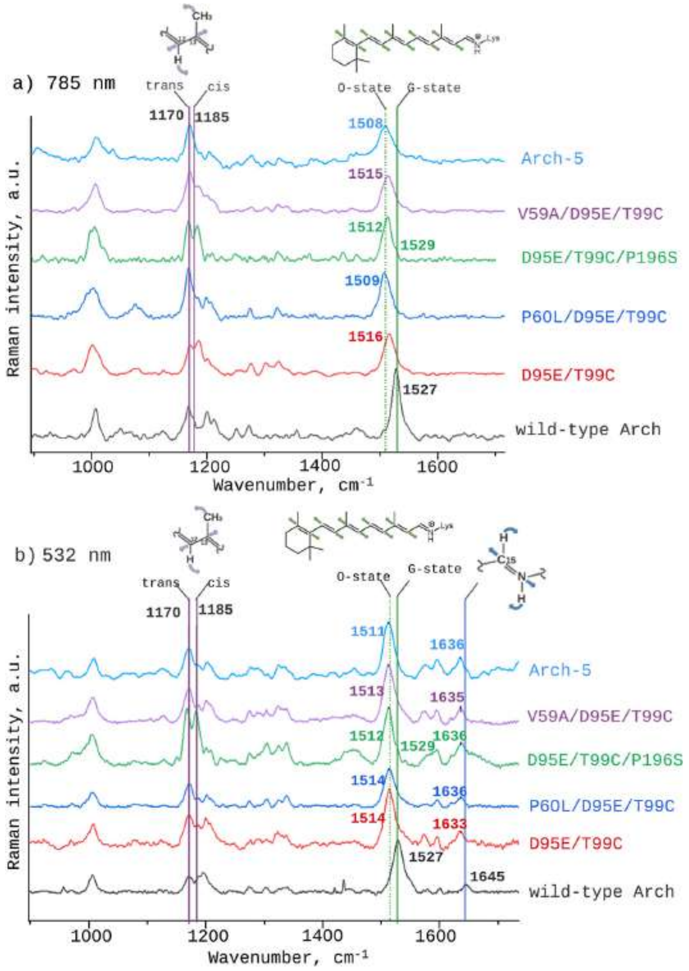


Figure 3.2 — Raman spectra of wild-type archaerhodopsin-3 and its mutants from the Archers group with increased fluorescence signal. Measurements were performed for purified protein samples dissolved in buffer (10 mM Tris, 200 mM NaCl, 0.15% DM, pH=6.5). Raman spectra were recorded under excitation at 785 nm [dark-adapted spectrum, (a)] and 532 nm [light-adapted spectrum, (b)]. The laser power was set at 100 mW. Arch-5 denotes Arch V59A/P60L/D95E/T99C/P196S.

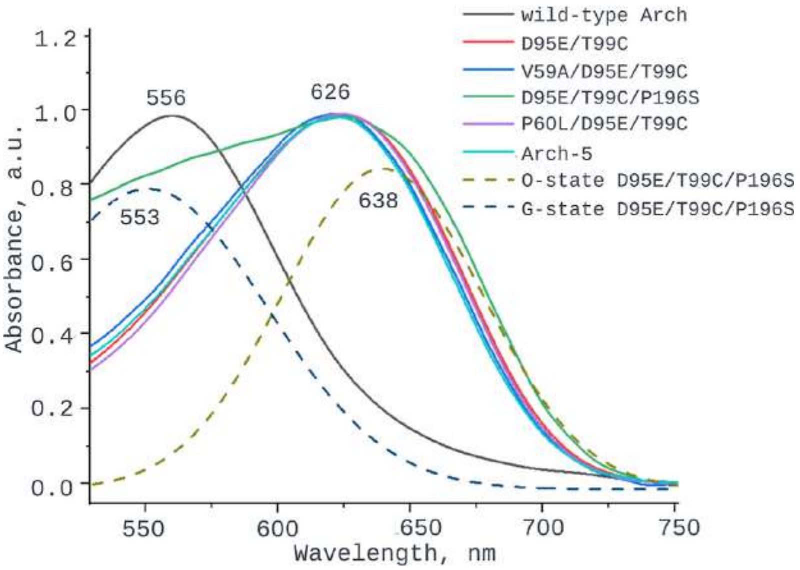


Figure 3.3 — UV-visible absorption spectra of wild-type archaerhodopsin-3 and its mutants from the Archers group with increased fluorescence signal. Measurements were performed for purified protein samples dissolved in buffer (10 mM Tris, 200 mM NaCl, 0.15% DM, pH=6.5). The result of decomposition of the Arch D95E/T99C/P196S absorption band into two Gaussian functions is depicted by dashed lines.

Table 2 — UV-Vis absorption maxima (λ_{max}), fluorescence excitation maxima (λ_{exc}), fluorescence emission maxima (λ_{em}) and fluorescence quantum yields (Φ) of the wild-type archaerhodopsin-3 (Arch) and its mutants discussed in this chapter. Arch-5 – Arch V59A/P60L/D95E/T99C/P196S

Protein	λ_{max}	λ_{exc}	λ_{em}	Φ,%
Arch	556	553 [17]	687 [16]	0.01 [17]
D95E/T99C	626	627	729	0.33
P60L/D95E/T99C	624	625	725	0.40
D95E/T99C/P196S	628	639	733	0.56
V59A/D95E/T99C	622	623	724	0.62
Arch-5	622	627	745	0.87

two clusters. The cluster corresponding to wild-type microbial rhodopsins is in the upper left corner; it contains wild-type archaerhodopsin-3. In the lower right corner is a cluster corresponding to forms of microbial rhodopsins that lack negatively charged residues near the C=NH+ portion of the chromophore. This cluster includes bacteriorhodopsin at pH=2.6 [113], as well as the O-intermediates of the photocycle of *Krokinobacter* rhodopsin 2 [114], halorhodopsin [115], and bacteriorhodopsin from archaea *Halobacterium salinarum.* [112]. The cluster also includes mutants of archaerhodopsin-3 from the Archers group. Thus, all the obtained experimental data and their comparison with known literature data indicate neutralization of charges of the amino acid environment of the C=NH+ part of the chromophore in fluorescent proteins from the Archers group.

It was also necessary to determine whether fluorescence occurs after photoexcitation of the main photocycle state or the photocycle intermediate. According to the published data, the voltage-dependent fluorescence of archaerhodopsin-3 occurs after photoexcitation of the Q-intermediate of the photocycle, which in turn occurs after photoexcitation of the N-intermediate of the photocycle. On the other hand, for the archaerhodopsin-3-based sensors, QuasAr1 and QuasAr2 fluorescence was shown to occur after photoexcitation of the ground state of the protein photocycle, without the involvement of photocycle intermediates [13]. We showed that for each protein from the Archer group, the maximum of the absorption spectrum and the maximum of the fluorescence excitation spectrum coincide, i.e., fluorescence arises from the photon-absorbing form of the protein – the ground state, most thermodynamically stable form of the protein (Figure 3.5). The only exception was Arch D95E/T99C/P196S, for which the absorption band maximum (628 nm) was shifted 11 nm relative to the fluorescence excitation band maximum (639 nm). Analysis of the absorption band of Arch D95E/T99C/P196S showed that its absorption band is two-component. Two Gaussian functions with maxima at 553 nm and 638 nm should be used to approximate the band (Figure 3.3). On the other hand, the absorption bands of the other proteins of the Archer group studied here were well approximated by a single Gaussian function. It was hypothesized that for Arch D95E/T99C/P196S there is a thermodynamic equilibrium of two forms with absorption maxima at 553 nm and 638 nm. Photoexcitation of a form of the protein with a maximum at 638 nm results in fluorescence, consistent with the proximity of the absorption maximum of this

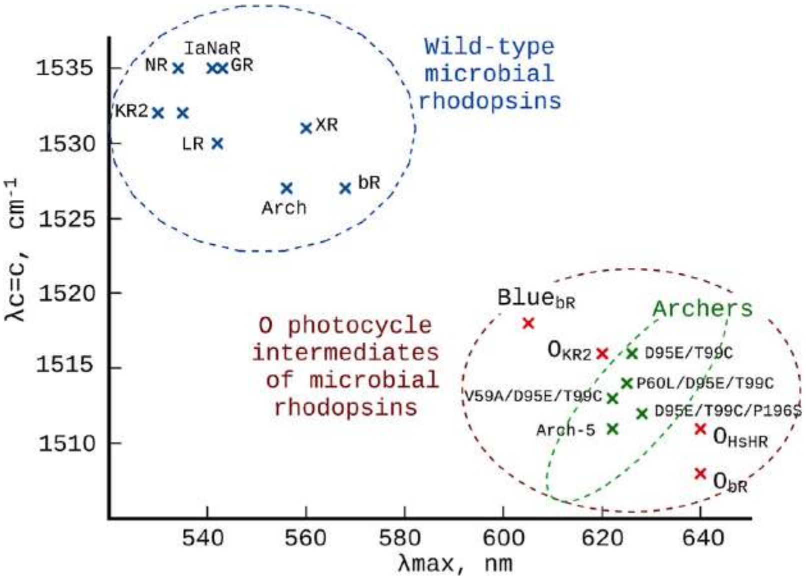


Figure 3.4 — Spectral properties of microbial rhodopsins. The band maxima of Raman spectra corresponding to the symmetric ethylenic C=C stretching vibrations of the chromophore ($\lambda_{C=C}$) are plotted on the y-axis, and the absorption maxima (λ_{max}) are plotted on the x-axis. Wild-type microbial rhodopsins are shown as blue dots, O-intermediates of the photocycle of microbial rhodopsins and the bacteriorhodopsin form at pH=2.6 are shown as red dots. Proteins from the Archers group are shown as green dots. Abbreviations: KR2, *Krokinobacter* rhodopsin 2 [114]; NR, *Neurospora* rhodopsin [116]; PspR, *Pseudomonas putida* rhodopsin [117]; IaNaR, *Indibacter alkaliphilus* rhodopsin [118]; GR, *Gloeobacter* rhodopsin [119]; LR, *Leptosphaeria* rhodopsin [120]; Arch, archaerhodopsin-3; XR, xanthorhodopsin [121]; bR, light-adapted bacteriorhodopsin from Archaea *Halobacterium salinarum* [113]; Blue$_{bR}$, form of bacteriorhodopsin at pH = 2.6 [113]; O$_{KR2}$, O-intermediate of the KR2 photocycle [122]; O$_{HsHR}$, O-intermediate of the halorhodopsin photocycle from Archaea *Halobacterium salinarum* [112]; O$_{bR}$, O-intermediate of the bacteriorhodopsin photocycle [112]; Arch-5, Arch V59A/P60L/D95E/T99C/P196S.

form to the maximum of the fluorescence excitation band of the protein (639 nm). On the other hand, the presence of a form with an absorption maximum at 553 nm causes the overall maximum of the absorption band of the protein to be shifted 11 nm into the blue region relative to the form of the protein whose photoexcitation results in fluorescence.

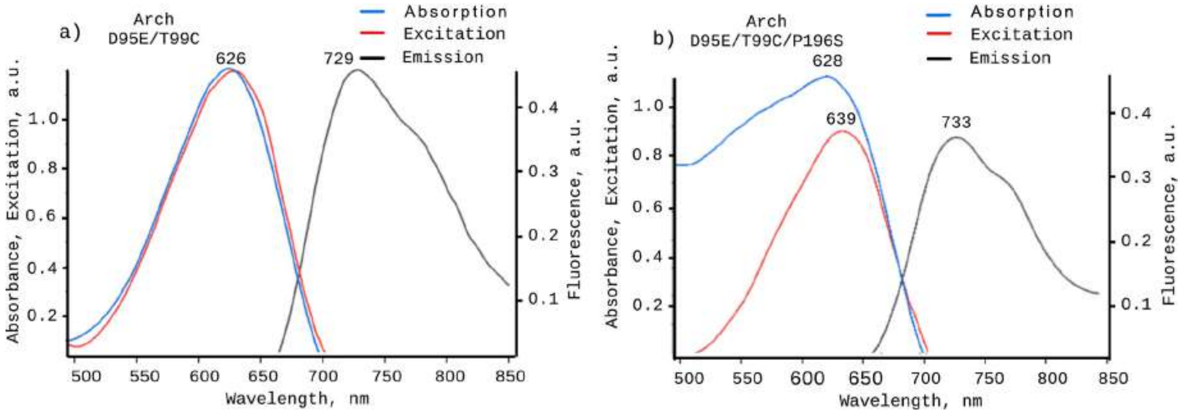


Figure 3.5 — Absorption, fluorescence excitation and fluorescence emission spectra of Arch D95E/T99C (a) and Arch D95E/T99C/P196S (b). Measurements were performed for purified protein samples dissolved in buffer (10 mM Tris, 200 mM NaCl, 0.15% DM, pH=6.5). Fluorescence excitation spectra were obtained by recording fluorescence at 740 nm. Fluorescence spectra were obtained by excitation at 630 nm.

3.3 Determination of the structural feature responsible for increased fluorescence in archaerhodopsin-3 mutants from the Archers group compared to the wild-type protein.

Computational models of all the proteins studied in this study were constructed to determine the structural feature whose alteration leads to increased fluorescence in mutants of archaerhodopsin-3 from the Archers group. The results of the experimental studies showed that the spectral properties of the Archers group proteins are similar to the spectral properties of the O-intermediates of the microbial rhodopsin photocycle. Based on this information, it was concluded that the electrostatic field near the chromophore in these proteins is also similar to the electrostatic field in the indicated region in the O-intermediates of the photocycle. Therefore, two models corresponding to the two forms of the protein, the most thermodynamically stable ground state of the wild-type protein photocycle (G-state) and the O-intermediate of the wild-type protein photocycle (O-state), were constructed for each protein. Initial models of the three-dimensional structures of the G-state models were constructed by homology modeling using the crystallographic structure of the wild-type protein as a template. Initial three-dimensional models of the O-state models were constructed by homology modeling using the crystallographic structure of the O-intermediate of bacteriorhodopsin with the L93A amino acid substitution as a template. Quantum mechanics/molecular mechanics (QM/MM) ASEC-FEG models were constructed for all initial three-dimensional protein models and their absorption band maxima were calculated. The methods described in Section 2.2 were used to build the models.

The results of the computational modeling are presented in Table 3. It is shown that the λ_{max} value calculated for the G-state model of the wild-type protein is close to the experimental one. Analysis of the model revealed the following protonation states of the counterions - counterion D95 is protonated and counterion D222 is in non-protonated form (Figure 3.6). Thus, the total charge generated by the counterions is -1. A cluster of three hydrogen-bonded water molecules, the counterions D95 and D222, and the NH group of the chromophore is observed in the counterion region. The proton release group contains two glutamic acids, one of which (E214) is protonated and forms a hydrogen bond with E207 (Figure 3.6). The orientation of the residues in the QM/MM model and the resulting hydrogen bonding system is consistent with that observed in the crystallographic structure of the protein.

A model of the O-state of wild-type archaerhodopsin-3 was constructed for comparison. The calculated absorption band maximum for the resulting model O-state of archaerhodopsin-3 is 634 nm, which is close to the absorption band maximum of the O-intermediate of bacteriorhodopsin (640 nm) estimated from time-resolved spectroscopy data [112]. In the O-state model, the counterions D95 and D222 are protonated (Figure 3.6). The cluster of water molecules observed in the G-state model of the wild-type protein is absent in the counterion region. Glutamic acids E207 and E214 are not protonated and cannot form a hydrogen bond. In contrast, E207 forms a hydrogen bond with tyrosine residue Y93 and E214 forms a hydrogen bond with threonine residue T203. The observed orientation of the residues and the hydrogen bonding system is consistent with that observed in the crystallographic structure of the O-intermediate of bacteriorhodopsin with amino acid substitution L93A.

Table 3 — Experimental and calculated absorption maxima of the wild-type archaerhodopsin-3 and the protein variants with amino acid substitutions considered in this study.

Protein	λ_{max}, exp.	λ^G_{max}, calc.	λ^O_{max}, calc.
Arch	558	557	634
D95E/T99C	626	548	621
P60L/D95E/T99C	624	548	623
D95E/T99C/P196S	628	564	639
V59A/D95E/T99C	622	542	625
Arch-5	622	558	626

The λ_{max} values calculated for the O-state models of archaerhodopsin-3 mutants from the Archers group reproduce the experimental values with an error of no more than 8 nm. The only exception is Arch D95E/T99C/P196S, for which the calculated value of λ_{max} (638 nm) differs by 10 nm from the experimental value (628 nm). However, the calculated λ_{max} value of this protein is consistent with the maximum of the fluorescence excitation band of the protein (639 nm) and the maximum of the second component in the decomposition of the absorption band into two Gaussian functions (638 nm). Thus, based on experimental data and computational models, it can be concluded that for Arch D95E/T99C/P196S an equilibrium between two forms is experimentally observed, one of which has $\lambda_{max} =$ 639 nm.

Analysis of computational models showed that the structures of the regions near the counterions and the proton release group are similar for all mutants with increased fluorescence, so detailed analysis was performed only for Arch D95E/T99C (Figure 3.7). According to the O-state model of this protein, the counterions D95 and D222 in these proteins are protonated, with only one water molecule in the counterion region. The glutamic acid residues of the proton release group are in the non-protonated state. E207 forms a hydrogen bond with tyrosine Y93 and E214 forms a hydrogen bond with threonine T203. The observed orientation of the residues and the hydrogen bonding system is consistent with that observed in a computer

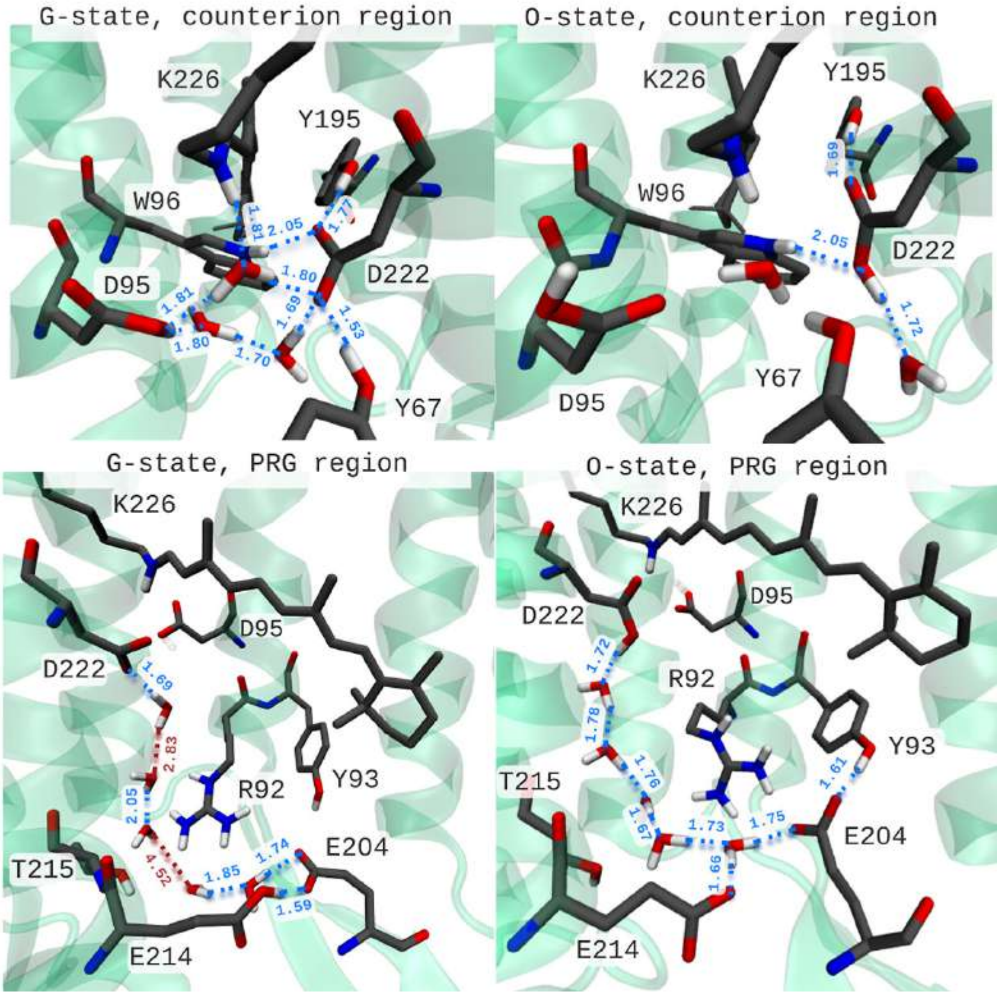
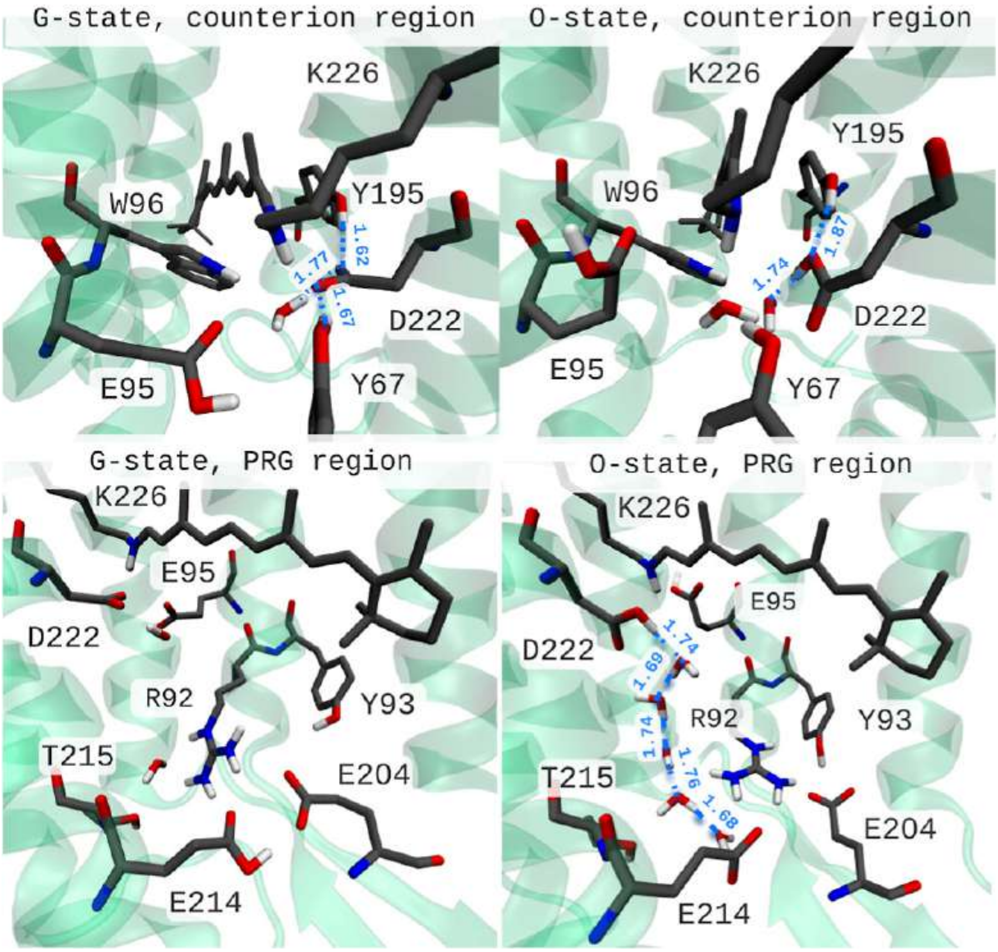


Figure 3.6 — Three-dimensional structures of archaerhodopsin-3 in the region of counterions and in the region of the proton release group in the G- and O-states of the protein obtained from the corresponding QM/MM models. In the G-state, three water molecules are observed near the counterions D95 and D222. The counterion D222 is not protonated and forms hydrogen bonds with Y67, Y195, W96 and two water molecules (upper left panel). Residue E214 from the proton release group is protonated and forms a hydrogen bond with E207. The positively charged residue R92 is oriented toward T215 (lower left panel). In the O-state, only one water molecule is observed near the counterions. D222 forms three hydrogen bonds, with Y195, W96 and one water molecule located outside the counterion region (upper right panel). E214 is in the non-protonated state, E207 forms a hydrogen bond with Y93. The positively charged residue R92 is oriented toward the negatively charged E214 and E207, making room for the formation of a chain of water molecules between E214 and D222 (lower right panel).

model of the O-state of archaerhodopsin-3 and in the crystallographic structure of the O-intermediate of bacteriorhodopsin with the L93A amino acid substitution. It should be noted that the λ_{max} values calculated for computational models of the G-states of the studied mutants of archaerhodopsin-3 differ significantly from the experimental values and are close to the λ_{max} of the wild-type protein. It can be assumed that in mutants with increased fluorescence intensity, the O-state of the protein is the most thermodynamically stable form.

Alterations in the protonation state of the counterion D222 in the investigated archaerhodopsin-3 mutants result from the following modifications. In the wild-type protein, the low value of pKa(D222) is determined by the formation of hydrogen bonds with two water molecules, tyrosine residues Y67 and Y195, and tryptophan residue W96 - a total of 5 hydrogen bonds. On the other hand, Arch D95E/T99C shows one hydrogen bond with water molecules, a weakened hydrogen bond with tyrosine Y195, no hydrogen bond with tyrosine Y67, and an increased distance to the counterion D95, which forms a hydrogen bond with cysteine C99 in Arch D95E/T99C. Thus, decreasing the number of hydrogen bonds and moving away from D95 leads to an increase in pKa(D222). Similar structural changes are observed in the other four proteins of the Archer group. An increase in pKa(D222) and the corresponding transition to the protonated state leads to a decrease in pKa(E214), a member of the proton release group. The relationship between the pKa of the counterion and the pKa of the glutamic acid of the proton release group has been studied experimentally and the existence of a direct correlation between these two values has been described in the literature [70]. It should be noted that in the O-state model of the wild-type archaerhodopsin-3 and in the Archers models considered, the presence of a chain of hydrogen-bonded water molecules is observed between E214 and D222. The structural changes observed in the Archers models in the proton release group region are a consequence of the E214 deprotonation.

The experimental data, namely the results of spectrophotometric titration of the proteins, are in agreement with the result of the computational model analysis. The experimental pKa(D222) value ~ 2.5 for the wild-type protein is in agreement with previously published data [123]. On the other hand, pKa(D222) values $>$ 7.8 were obtained for all considered proteins from the Archers group (Table 4). It should be noted that the lowest value of pKa(D222) $= 7.8$ was found for the Arch D95E/T99C/P196S protein, for which a two-component absorption spectrum band



G-state, counterion region

O-state, counterion region

G-state, PRG region

O-state, PRG region

Figure 3.7 — The three-dimensional structures of Arch D95E/T99C in the counterion region and the proton release group region in the G- and O-states of the protein obtained from the corresponding computational models. In the G-state, only one water molecule is observed near the counterions E95 and D222. The counterion D222 is not protonated and forms hydrogen bonds with Y67, Y195 and a water molecule from the counterion region (upper left panel). E214 from the proton release group is protonated and forms a hydrogen bond with E207. The positively charged residue R92 is oriented toward T215 (lower left panel). The O-state also shows only one water molecule near the counterions. D222 forms two hydrogen bonds with Y195 and a water molecule outside the counterion region (upper right panel). E214 is not protonated and E207 forms one hydrogen bond with Y93. The positively charged residue R92 is oriented toward the negatively charged E214 and E207, leaving room for a water chain to form between E214 and D222 (lower right panel).

and a blue shift of the absorption band relative to the fluorescence excitation band were observed. It was hypothesized above that the observed effects may be due to the presence of two forms of the protein coexisting in thermodynamic equilibrium. The titration results show that for this protein, at physiological pH values, there should be an equilibrium between a state with protonated D222, whose absorption band maximum is at 639 nm, and a state with D222 in the non-protonated state, whose absorption band maximum is at about 552 nm.

Table 4 — The pKa values of counterion D222 and residue E214 of the proton release group of wild-type archaerhodopsin-3 and the protein variants with amino acid substitutions considered in this study.

Protein	pKa(D222)	pKa(D222)
Arch	2.5	-
D95E/T99C	8.9	-
P60L/D95E/T99C	>9	-
D95E/T99C/P196S	7.8	7.0
V59A/D95E/T99C	>9	4.5
Arch-5	9.1	-

To perform the direct verification of our assumption that protonation of the counterion is a prerequisite for obtaining a protein with increased fluorescence intensity, new amino acid substitutions were proposed based on computational models. These substitutions should lead to an increase in pKa(D222) due to the breaking of hydrogen bonds of the counterion with surrounding residues and water molecules. Amino acid substitutions W96F and Y195F were proposed (Figure 3.6). No increase in fluorescence or red-shift of the absorption band maximum was observed for Arch W96F. On the other hand, an increase in absorbance in the >600 nm spectral region was observed for Arch Y195F. The measured fluorescence quantum yield of the protein was 0.23%, which is significantly higher than that of the wild-type protein (0.01%). The band maximum of the fluorescence excitation spectrum of Arch Y195F was 639 nm (Figure 3.8), which is consistent with the value of the absorption band maximum of the O-intermediate of bacteriorhodopsin. To further stabilize the O-state state, a T215A amino acid substitution was introduced into Arch W96F. For bacteriorhodopsin, it was previously shown that a similar substitution of T205A results in a 100-fold [124] increase in the lifetime of the

O-intermediate. Model analysis showed that the introduction of the T215A mutation leads to a change in the orientation of the arginine residue R92, which in the wild-type protein forms a water molecule-mediated hydrogen bond with T215. Breaking this bond results in the rotation of R92 toward the glutamic acids of the proton release group, which contributes to the decrease in pKa(E214) – a feature of the O-state. Besides, the rotation of R92 results in a free space for the formation of a chain of water molecules between D222 and E214. For Arch W96F/T215A, a complex absorption band consisting of two components was obtained, apparently corresponding to the O-state and G-state of the protein (Figure 3.9a). The protein is characterized by a fluorescence quantum yield of 0.25%, which is significantly higher than the fluorescence quantum yield of the wild-type protein (0.01%) (Figure 3.9b). Thus, the data show that charge neutralization in the C=NH+ environment of the chromophore group, achieved in the proteins from the Archer group by protonation of the counterion complex consisting of D95 and D222, is a prerequisite for the enhanced fluorescence of microbial rhodopsins.

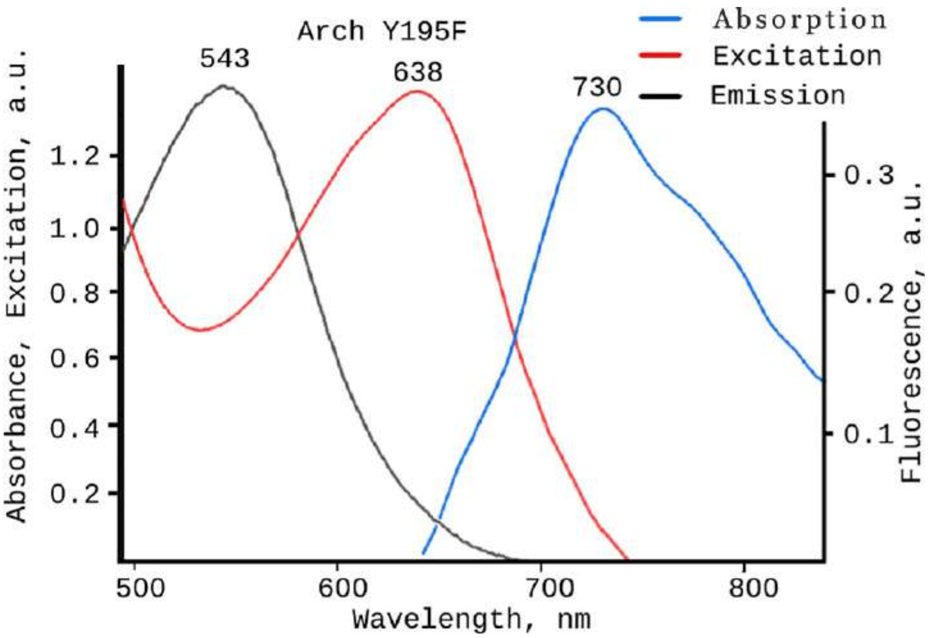


Figure 3.8 — Absorption, fluorescence excitation and fluorescence spectra of Arch Y195F. Measurements were performed for purified protein samples dissolved in buffer (10 mM Tris, 200 mM NaCl, 0.15% DM, pH=6.5). Fluorescence excitation spectra were obtained by recording fluorescence at 740 nm. Fluorescence spectra were obtained by excitation at 630 nm.

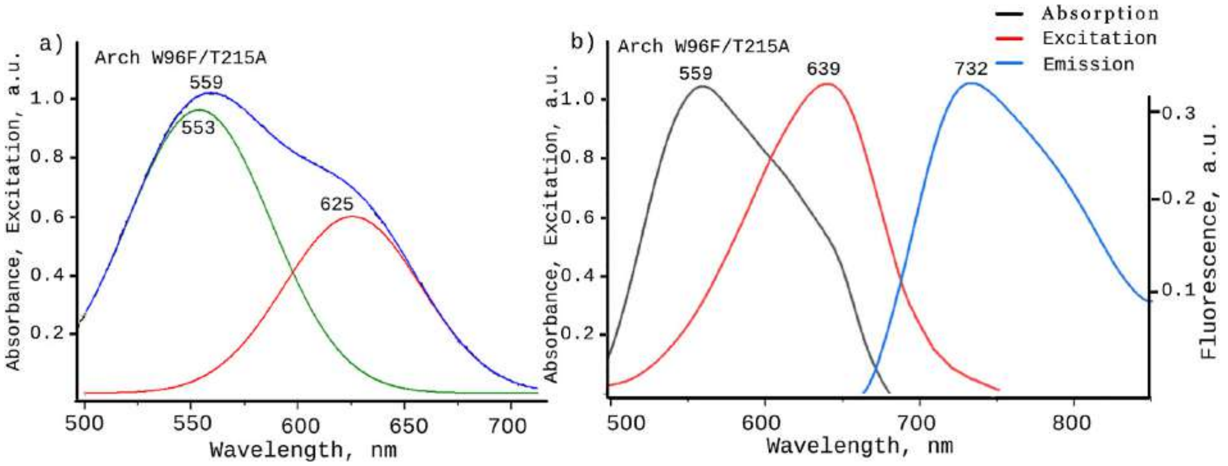


Figure 3.9 — a) Absorption spectrum of Arch W96F/T215A. b) Absorption, fluorescence excitation, and fluorescence emission spectra of Arch W96F/T215A. Measurements were performed for purified protein samples dissolved in buffer (10 mM Tris, 200 mM NaCl, 0.15% DM, pH=6.5). Fluorescence excitation spectra were obtained by recording fluorescence at 740 nm. Fluorescence spectra were obtained by excitation at 630 nm.

Chapter 4. Rational design of new archaerhodopsin-3-based fluorescent genetically encoded voltage indicators with enhanced fluorescence

This chapter describes the results of the rational design of novel archaerhodopsin-3 mutants from the Archers group with increased fluorescence quantum yields and absorption band maxima in the 638-643 nm range.

4.1 Structural feature determining voltage sensitivity of fluorescence in archaerhodopsin-3 and its mutants

The results described in Chapter 3 show that the introduction of the amino acid substitutions Y195F and W96F/T215A into the wild-type protein leads to an increase in the fluorescence quantum yield of the protein. In *E. coli* cells, the voltage dependence of the fluorescence of the obtained proteins was investigated. For Arch Y195F and Arch W96F/T215A, there was no detectable increase in the integrated fluorescence recorded from protein-expressing *E. coli* cells after their death due to the addition of sodium azide to the cell medium. This indicates either the absence of voltage dependence of fluorescence or the low sensitivity of fluorescence to voltage changes. On the other hand, a significant increase in fluorescence after cell death was observed for proteins from the Archer group.

In order to determine the structural feature of the protein whose value regulates the sensitivity of fluorescence to changes in voltage in archaerhdopsin-3 and its variants with amino acid substitutions, the pKa values of the chromophore of the studied proteins, the Schiff base, were measured in this study. It is shown that for all proteins of the Archer group, the pKa values of the Schiff base (pKa(SB)) were found to be close to the physiological pH, while the pKa(SB) values for Arch Y195F and Arch W96F/T215A were significantly higher (Table 5). In Figure 4.1 the values of fluorescence voltage sensitivity estimated as the relative change in fluorescence intensity with a 100 mV voltage increase are plotted on the y-axis and the pKa(SB) values are plotted on the x-axis. The graph shows an increase in sensitivity as pKa(SB) approaches the physiological value of pH=7.4 (Figure

4.1). Thus, the voltage sensitivity of fluorescence may be related to the equilibrium between forms with protonated and non-protonated chromophores.

Table 5 — The values of Schiff base pKa (pKa(SB)) and the presence of the reported change in fluorescence intensity after changing the membrane potential of protein-expressing *E. coli* cells from the initial value to 0 mV ($dF/dV \neq 0$), for the wild-type archaerhodopsin-3 and the considered mutants.

Protein	pKa(SB)	$dF/dV \neq 0$
D95E/T99C	7.5	Yes
P60L/D95E/T99C	7.7	Yes
D95E/T99C/P196S	7.4	Yes
V59A/D95E/T99C	7.9	Yes
Arch-5	7.6	Yes
Y196F	>10	No
W96F/T215A	9.4	No

4.2 The results of the rational design of new proteins.

The results presented in Chapters 3 and 4.1 showed that the rational design of novel fluorescent membrane potential sensors based on archaerhodopsin-3 must take into account two factors – the absence of negatively charged residues near the C=NH+ group of the chromophore and the proximity of the Schiff base pKa to the physiological value of pH=7.4. Since it was beyond the scope of this thesis study to investigate the mechanism that determines the pKa of the Schiff base, here the design was performed for the proteins from the Archers group with the D95E/T99C amino acid substitutions retained.

An analysis of the literature data presented in Chapter 1.3 shows that for microbial rhodopsins, an increase in the quantum fluorescence yields is observed with a shift of the maximum of the absorption band to the long-wavelength region. According to this correlation, the introduction of amino acid substitutions leading to a red shift in the absorption spectrum should lead to an increase in the quantum yield of fluorescence. According to the literature data and the results of research conducted

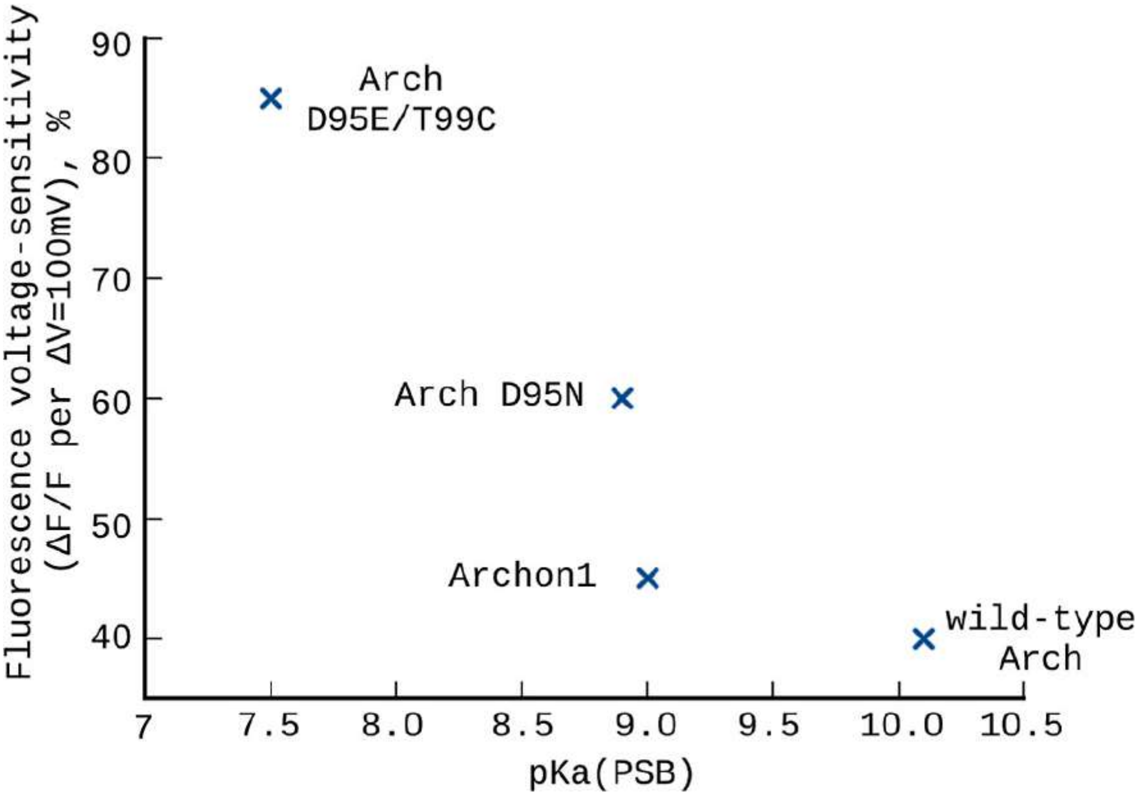


Figure 4.1 — Dependence of the fluorescence voltage sensitivity of archaerhodopsin-3-based genetically encoded voltage indicators, defined as the percentage change in fluorescence intensity for a 100 mV voltage change, on the pKa of the Schiff base.

during the thesis study, the introduction of polar amino acids in the region near the beta-ionone ring of the chromophore leads to a shift in the absorption spectrum of microbial rhodopsins to the long-wavelength region [21; 51; 101; 125]. This principle of introducing polar amino acids into the region near the beta-ionone ring was used for the rational design.

Nine sites were selected for the introduction of polar amino acids located near the beta-ionone ring of the chromophore. Since the number of possible combinations of polar amino acids at the 9 positions is too large, a preliminary analysis of databases of amino acid sequences of rhodopsins from bacteria, archaea, and fungi was performed. Combinations of polar amino acids at these nine positions found in natural rhodopsins were found. Computational screening was performed for the combinations found, and as a result, a final set of combinations of amino acid substitutions was proposed, which were introduced into the Archer1 (D95E/T99C) and Archer2 – archorhodopsin-3 with D95E/T99C/A225M amino acid substitutions, which is characterized by absorption maximum at 633 nm.

All proteins were synthesized and their absorption spectra and fluorescence quantum yields were measured, the results are shown in Figure 4.2. As a result, proteins with absorption maxima up to 643 nm and fluorescence quantum yields up to 1.18% were obtained (Table 6). For the obtained archaerhodopsin-3 variants, there is a tendency for the fluorescence quantum yield to increase with a shift of the absorption maximum to the long wavelength region ($R^2 = 0.48$, Figure 4.2). Fitting the bands of the absorption spectra of the proteins by Gaussian functions showed that for all proteins there is an equilibrium between the O-state with protonated counterion and an absorption maximum >620 nm and the G-state of the protein with non-protonated counterion and an absorption maximum <580 nm. The presence of this mixture causes a blue-shift of the absorption maxima relative to the band maxima of the corresponding O-states (Table 6). Note that the tendency for the fluorescence quantum yield to increase with increasing absorption band maxima persists for the O-state absorption band maxima, and the coefficient of determination for this dependence is greater than 0.5 ($R^2 = 0.62$, Figure 4.2).

It should be noted that a similar destabilization of the O-state was previously observed in the Arch D95E/T99C/P196S, in which the absorption band consisted of two components and decomposed into two Gaussian functions with maxima at 552 nm and 639 nm, corresponding to the forms with non-protonated and protonated

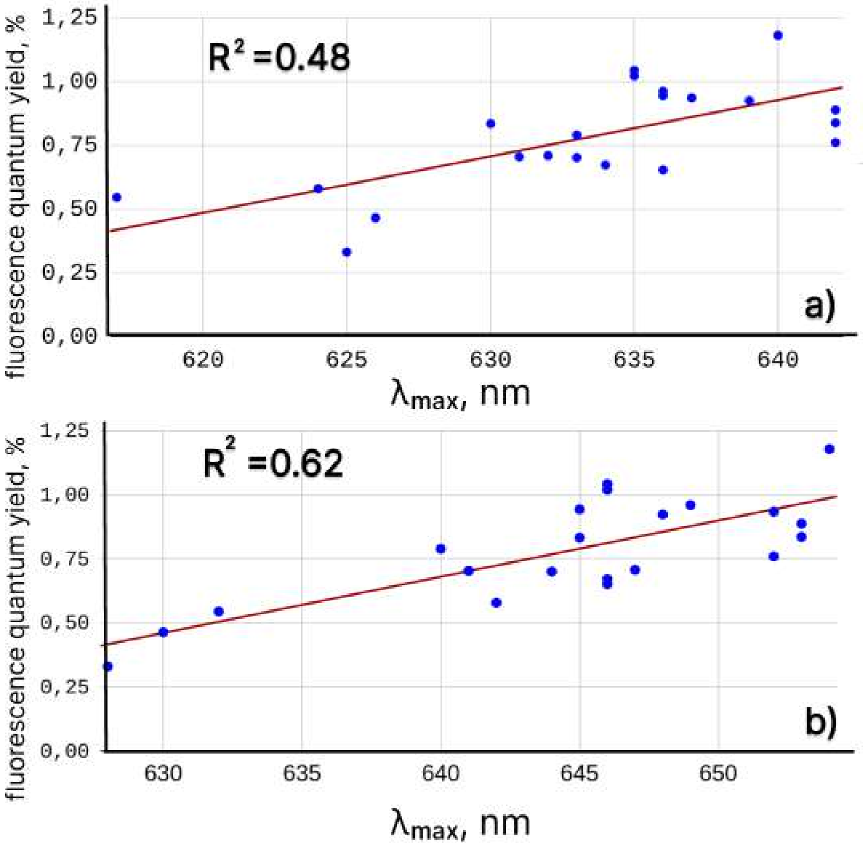


Figure 4.2 — a) Correlation between fluorescence quantum yields and absorption maxima for the new mutants of archaerhodopsin-3. b) Correlation between fluorescence quantum yields and O-state absorption maxima for the new mutants of archaerhodopsin-3.

Table 6 — Absorption band maxima (λ_{max}), absorption band maxima of protein O-states obtained from decomposition of absorption spectra into two Gaussian functions (λ_{max} [O]), and fluorescence quantum yields (Φ) of Archer2 and the best performing mutants of archaerhodopsin-3 from the Archers group proposed in this study.

Protein	λ_{max}, nm	λ_{max} [O], nm	Φ, %	pKa(SB)
Archer2	633	643	0.66	8.1
Archer2 + T215A	640	648	0.77	7.5
Archer1 + P196S/T215A	638	640	0.97	7.4
Archer2 + F218Y	642	653	0.89	7.8
Archer2 + P196S	640	654	1.18	7.3
Archer2 + P196S/F218Y	643	654	0.76	8.0

counterions, respectively. As part of the study, we performed an additional stabilization of the O-state in this protein. This was done by introducing a T215A amino acid substitution into the protein, which had previously stabilized the O-state in Arch W96F as described in Chapter 3. The Arch D95E/T99C/P196S/T215A protein was synthesized and its properties were studied experimentally. Consistent with the hypothesis of stabilization of the O-state after introducing the amino acid substitution T215A, λ_{max} had shifted to 638 nm and became in a close proximity to the fluorescence excitation band maximum $\lambda_{exc} = 640$ nm. The fluorescence quantum yield of the protein measured upon excitation at the λ_{max} increased from 0.56% to 0.97% after the introduction of the T215A substitution (Figure 4.3a). Importantly, the resulting protein retained the voltage dependence of the fluorescence signal, and it was also shown that its pKa(SB) = 7.4 was in the range characteristic of high fluorescence sensitivity to changes in membrane potential.

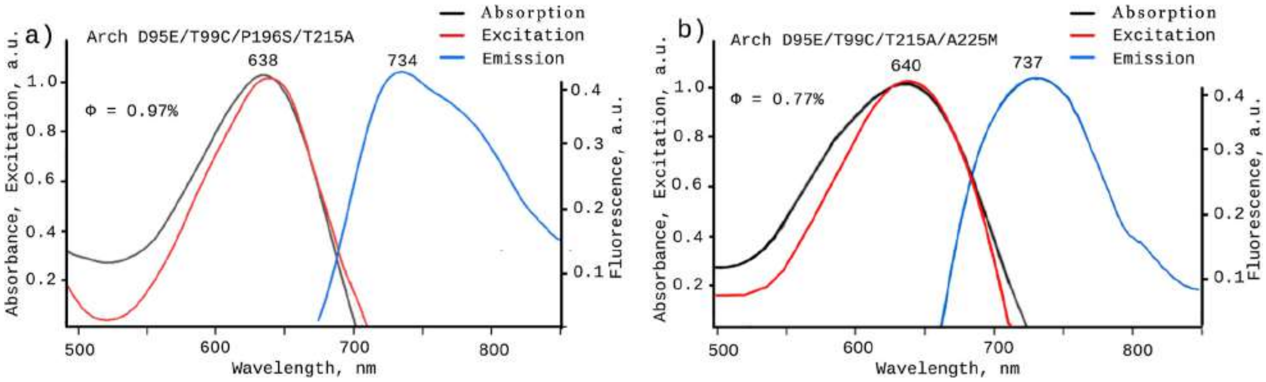


Figure 4.3 — Absorption, fluorescence excitation and fluorescence spectra of new proteins - Arch D95E/T99C/P196S/T215A (a) and Arch D95E/T99C/T215A/A225M (b). Measurements were performed for purified protein samples dissolved in buffer (10 mM Tris, 200 mM NaCl, 0.15% DM, pH=6.5). Fluorescence excitation spectra were obtained by recording fluorescence at 740 nm. Fluorescence spectra were obtained by excitation at 630 nm.

To further validate the found approach for the rational design of archaerhodopsin-3-based genetically encoded voltage indicators by stabilizing the O-state of the protein, Arch D95E/T99C/A225M (Archer2) with $\lambda_{max} = 633$ nm was considered. The absorption band of this protein can be decomposed into two Gaussian functions with maxima at 573 nm and 649 nm, which can be assigned to the non-protonated and protonated counterion states of the protein, respectively. The calculated value of the absorption band maximum for the O-state model of the protein is 643 nm. Apparently, the A225M substitution results in a red shift of the absorption band maximum relative to Arch D95E/T99C, but also destabilizes the O-state. Computational modeling analysis showed that replacing the alanine above the C13-Me chromophore group with a bulkier methionine has an indirect effect on the absorption band maximum. Three effects were observed: 1) deformation of the chromophore, the value of the dihedral angle C15-C14-C13-C12 increased from 172° to 165°; 2) increase of the distance between the counterions D95 and D222 and the NH group of the chromophore; 3) replacement of the water molecule located above the C13-Me group of the chromophore in D95E/T99C by the side chain M225.

Introduction of the T215A mutation into Archer2 to stabilize the O-state of the protein resulted in a shift of λ_{max} from 634 nm to 640 nm (Figure 4.3b). For Arch D95E/T99C/T215A/A225M, we observed an increase in fluorescence quantum yield upon excitation at the wavelength of the absorption band maximum relative to the protein without the T215A mutation, from 0.66% to 0.77%. It was also confirmed that the dependence of fluorescence on membrane potential is preserved for the obtained protein, and it was shown that its pKa(SB)=7.5 is in the region characteristic of high sensitivity of fluorescence to membrane potential.

To further evaluate the sensitivity of fluorescence to changes in membrane potential for the new Arch mutants, the increase in integrated fluorescence from *E. coli* cells expressing the protein was assessed for each protein after cell death by the action of sodium azide. For the Arch D95E/T99C protein, there are literature data on the sensitivity of fluorescence to changes in potential - the percentage change in fluorescence intensity with a voltage change of 100 mV is 85% (pH = 7.4). In this work it was found that for some of the new proteins - Arch D95E/T99C/T215A/A225M, D95E/T99C/P196S/T215A, Arch D95E/T99C/F218Y/A225M, Arch D95E/T99C/P196S/A225M - the sensitivity may be higher than for Arch D95E/T99C.

Thus, the thesis study resulted in the rational design of novel fluorescent membrane potential sensors based on archaerhodopsin-3. The obtained proteins are characterized by fluorescence quantum yields that exceed or are comparable to those of the brightest published variants of archaerhodopsin-3-based sensors. The absorption band maxima of the new proteins are closest to the optical transparency window of biological tissues compared to all published archaerhdopsin-3-based fluorescent genetically encoded voltage indicators. The proposed sensors and the formulated design principles of new fluorescent genetically encoded voltage indicators can be used in further studies for the rational design of new sensors of membrane potential based on microbial rhodopsins.

Conclusions

The main results of the thesis study:

1. Improved computational models for predicting the three-dimensional structure and calculating the spectral properties of rhodopsins starting the amino acid sequence. In particular, a new adaptation of the method of constructing quantum mechanics/molecular mechanics protein models, in which the amino acid environment of a cofactor described at the quantum mechanical level of theory is constructed by taking into account averaging over protein configurations existing in thermodynamic equilibrium.
2. Evidence has been obtained that a sufficient condition for obtaining mutants of microbial rhodopsins with increased intensity of the fluorescent signal is thermodynamic stabilization of the protein form with the absence of charged residues near the retinal Schiff base. Evidence has been obtained that in proteins from the Archers group, introduced amino acid substitutions lead to thermodynamic stabilization of the protein form with protonated counterion D222, which causes a significant increase in Archers fluorescence compared to the wild-type protein.
3. New mutants of archaerhodopsin-3 were obtained. The proteins demonstrate increased brightness compared to known variants of fluorescent genetically encoded voltage indicators from the Archers group. The fluorescence intensity of proteins depends on the magnitude of the cell membrane potential. The absorption maxima of the new proteins are closest to the window of optical transparency of biological tissues in comparison with all archaerhodopsin-3-based fluorescent genetically encoded voltage indicators published at the time of the thesis study.

References

1. *Xu Y.*, *Zou P.*, *Cohen A. E.* Voltage imaging with genetically encoded indicators // Curr. Opin. Chem. Biol. — 2017. — Vol. 39. — P. 1–10.
2. *Binggeli R.*, *Weinstein R. C.* Membrane potentials and sodium channels: hypotheses for growth regulation and cancer formation based on changes in sodium channels and gap junctions // Journal of theoretical biology. — 1986. — Vol. 123, no. 4. — P. 377–401.
3. *Yang M.*, *Brackenbury W. J.* Membrane potential and cancer progression // Frontiers in physiology. — 2013. — Vol. 4. — P. 185.
4. *Abdul Kadir L.*, *Stacey M.*, *Barrett-Jolley R.* Emerging roles of the membrane potential: action beyond the action potential // Frontiers in physiology. — 2018. — P. 1661.
5. Single action potentials and subthreshold electrical events imaged in neurons with a fluorescent protein voltage probe / L. Jin [et al.] // Neuron. — 2012. — Vol. 75, no. 5. — P. 779–785.
6. *Gong Y.*, *Li J. Z.*, *Schnitzer M. J.* Enhanced archaerhodopsin fluorescent protein voltage indicators // PLoS One. — 2013. — Vol. 8, no. 6. — e66959.
7. A bright and fast red fluorescent protein voltage indicator that reports neuronal activity in organotypic brain slices / A. S. Abdelfattah [et al.] // Journal of Neuroscience. — 2016. — Vol. 36, no. 8. — P. 2458–2472.
8. In vivo imaging of the coupling between neuronal and CREB activity in the mouse brain / T. Laviv [et al.] // Neuron. — 2020. — Vol. 105, no. 5. — P. 799–812.
9. Video-based pooled screening yields improved far-red genetically encoded voltage indicators / H. Tian [et al.] // Nat. Methods. — 2023. — Vol. 20, no. 7. — P. 1082–1094.
10. *Piatkevich K. D.*, *Boyden E. S.* Optogenetic Control of Neural Activity: the Biophysics of Microbial rhodopsins in Neuroscience // Q. Rev. Biophys. — 2023. — P. 1–81.

11. *Yang H. H.*, *St-Pierre F.* Genetically encoded voltage indicators: opportunities and challenges // Journal of Neuroscience. — 2016. — Vol. 36, no. 39. — P. 9977–9989.
12. Electrical spiking in Escherichia coli probed with a fluorescent voltage-indicating protein / J. M. Kralj [et al.] // Science. — 2011. — Vol. 333, no. 6040. — P. 345–348.
13. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins / D. R. Hochbaum [et al.] // Nat. Methods. — 2014. — Vol. 11, no. 8. — P. 825–833.
14. Directed evolution of a far-red fluorescent rhodopsin / R. S. McIsaac [et al.] // Proc. Natl. Acad. Sci. U.S.A. — 2014. — Vol. 111, no. 36. — P. 13034–13039.
15. Photoactivated voltage imaging in tissue with an archaerhodopsin-derived reporter / M.-P. Chien [et al.] // Sci. Adv. — 2021. — Vol. 7, no. 19. — eabe3216.
16. Optical recording of action potentials in mammalian neurons using a microbial rhodopsin / J. M. Kralj [et al.] // Nat. Methods. — 2012. — Vol. 9, no. 1. — P. 90–95.
17. Comparative studies of the fluorescence properties of microbial rhodopsins: spontaneous emission versus photointermediate fluorescence / K. Kojima [et al.] // J. Phys. Chem. B. — 2020. — Vol. 124, no. 34. — P. 7361–7367.
18. A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters / K. D. Piatkevich [et al.] // Nat. Chem. Biol. — 2018. — Vol. 14, no. 4. — P. 352–360.
19. Rational Design of Far-Red Archaerhodopsin-3-Based Fluorescent Genetically Encoded Voltage Indicators: from Elucidation of the Fluorescence Mechanism in Archers to Novel Red-Shifted Variants / D. M. Nikolaev [et al.] // ACS Phys. Chem. Au. — 2024. — Vol. 4, no. 4. — P. 347–362.

20. An assessment of water placement algorithms in quantum mechanics/molecular mechanics modeling: The case of rhodopsins' first spectral absorption band maxima / D. M. Nikolaev [et al.] // Phys. Chem. Chem. Phys. — 2020. — Vol. 22, no. 32. — P. 18114–18123.
21. Simple models to study spectral properties of microbial and animal rhodopsins: evaluation of the electrostatic effect of charged and polar residues on the first absorption band maxima / A. A. Shtyrov [et al.] // Int. J. Mol. Sci. — 2021. — Vol. 22, no. 6. — P. 3029.
22. Free Energy Computation for an Isomerizing Chromophore in a Molecular Cavity via the Average Solvent Electrostatic Configuration Model: Applications in Rhodopsin and Rhodopsin-Mimicking Systems / D. M. Nikolaev [et al.] // J. Chem. Theory Comput. — 2021. — Vol. 17, no. 9. — P. 5885–5895.
23. Fluorescence of the retinal chromophore in microbial and animal rhodopsins / D. M. Nikolaev [et al.] // Int. J. Mol. Sci. — 2023. — Vol. 24, no. 24. — P. 17269.
24. Fluorescence Imaging of Cell Membrane Potential: From Relative Changes to Absolute Values / D. M. Nikolaev [et al.] // Int. J. Mol. Sci. — 2023. — Vol. 24, no. 3. — P. 2435.
25. Voltage-and space-clamp errors associated with the measurement of electrotonically remote synaptic events / N. Spruston [et al.] // Journal of neurophysiology. — 1993. — Vol. 70, no. 2. — P. 781–802.
26. *Spira M. E.*, *Hai A.* Multi-electrode array technologies for neuroscience and cardiology // Nature nanotechnology. — 2013. — Vol. 8, no. 2. — P. 83–94.
27. *Scanziani M.*, *Häusser M.* Electrophysiology in the age of light // Nature. — 2009. — Vol. 461, no. 7266. — P. 930–939.
28. *Miller E. W.* Small molecule fluorescent voltage indicators for studying membrane potential // Current Opinion in Chemical Biology. — 2016. — Vol. 33. — P. 74–80.
29. *Loew L. M.* Design and use of organic voltage sensitive dyes // Membrane Potential Imaging in the Nervous System and Heart. — 2015. — P. 27–53.

30. *Kuhn B.*, *Roome C. J.* Primer to voltage imaging with ANNINE dyes and two-photon microscopy // Frontiers in Cellular Neuroscience. — 2019. — Vol. 13. — P. 321.

31. *Lin M. Z.*, *Schnitzer M. J.* Genetically encoded indicators of neuronal activity // Nat. Neurosci. — 2016. — Vol. 19, no. 9. — P. 1142–1153.

32. High-fidelity optical reporting of neuronal electrical activity with an ultrafast fluorescent voltage sensor / F. St-Pierre [et al.] // Nature neuroscience. — 2014. — Vol. 17, no. 6. — P. 884–889.

33. A fast and responsive voltage indicator with enhanced sensitivity for unitary synaptic events / Y. A. Hao [et al.] // Neuron. — 2024. — Vol. 112, no. 22. — P. 3680–3696.

34. Improving membrane voltage measurements using FRET with new fluorescent proteins / H. Tsutsui [et al.] // Nature methods. — 2008. — Vol. 5, no. 8. — P. 683–685.

35. High-speed recording of neural spikes in awake mice and flies with a fluorescent voltage sensor / Y. Gong [et al.] // Science. — 2015. — Vol. 350, no. 6266. — P. 1361–1366.

36. *St-Pierre F.*, *Chavarha M.*, *Lin M. Z.* Designs and sensing mechanisms of genetically encoded fluorescent voltage indicators // Current opinion in chemical biology. — 2015. — Vol. 27. — P. 31–38.

37. Comparative evaluation of genetically encoded voltage indicators / Y. Bando [et al.] // Cell reports. — 2019. — Vol. 26, no. 3. — P. 802–813.

38. Mechanism of voltage-sensitive fluorescence in a microbial rhodopsin / D. Maclaurin [et al.] // Proc. Natl. Acad. Sci. U.S.A. — 2013. — Vol. 110, no. 15. — P. 5939–5944.

39. Microbial and animal rhodopsins: structures, functions, and molecular mechanisms / O. P. Ernst [et al.] // Chem. Rev. — 2014. — Vol. 114, no. 1. — P. 126–163.

40. Microbial rhodopsins: diversity, mechanisms, and optogenetic applications / E. G. Govorunova [et al.] // Annual review of biochemistry. — 2017. — Vol. 86. — P. 845.

41. *Grip W. J. de*, *Ganapathy S.* Rhodopsins: an excitingly versatile protein species for research, development and creative engineering // Front. Chem. — 2022. — Vol. 10. — P. 879609.

42. *Inoue K.* Diversity, mechanism, and optogenetic application of light-driven ion pump rhodopsins // Optogenetics: Light-Sensing Proteins and Their Applications in Neuroscience and Beyond. — 2021. — P. 89–126.

43. Microbial rhodopsins: the last two decades / A. Rozenberg [et al.] // Annu. Rev. Microbiol. — 2021. — Vol. 75. — P. 427–447.

44. Similarities and Differences in Photochemistry of Type I and Type II Rhodopsins / M. A. Ostrovsky [et al.] // Biochemistry. — 2023. — Vol. 88, no. 10. — P. 1528–1543.

45. *Inoue K.* Photochemistry of the Retinal Chromophore in Microbial Rhodopsins // J. Phys. Chem. B. — 2023.

46. NeoR, a near-infrared absorbing rhodopsin / M. Broser [et al.] // Nat. Commun. — 2020. — Vol. 11, no. 1. — P. 5682.

47. *Lanyi J. K.* Proton transfers in the bacteriorhodopsin photocycle // Biochimica et Biophysica Acta (BBA)-Bioenergetics. — 2006. — Vol. 1757, no. 8. — P. 1012–1018.

48. Voltage Imaging with Engineered Proton-Pumping Rhodopsins: Insights from the Proton Transfer Pathway / X. Meng [et al.] // ACS Phys. Chem. Au. — 2023. — Vol. 3, no. 4. — P. 320–333.

49. O to bR transition in bacteriorhodopsin occurs through a proton hole mechanism / D. Maag [et al.] // Proc. Natl. Acad. Sci. U.S.A. — 2021. — Vol. 118, no. 39. — e2024803118.

50. Conformational changes in the archaerhodopsin-3 proton pump: detection of conserved strongly hydrogen bonded water networks / E. C. S. Clair [et al.] // J. Biol. Phys. — 2012. — Vol. 38, no. 1. — P. 153–168.

51. Directed evolution of Gloeobacter violaceus rhodopsin spectral properties / M. K. Engqvist [et al.] // J. Mol. Biol. — 2015. — Vol. 427, no. 1. — P. 205–220.

52. Archaerhodopsin variants with enhanced voltage-sensitive fluorescence in mammalian and Caenorhabditis elegans neurons / N. C. Flytzanis [et al.] // Nat. Commun. — 2014. — Vol. 5, no. 1. — P. 1–9.

53. Voltage imaging and optogenetics reveal behaviour-dependent changes in hippocampal dynamics / Y. Adam [et al.] // Nature. — 2019. — Vol. 569, no. 7756. — P. 413–417.

54. QuasAr Odyssey: the origin of fluorescence and its voltage sensitivity in microbial rhodopsins / A. Silapetere [et al.] // Nat. Commun. — 2022. — Vol. 13, no. 1. — P. 5501.

55. *Fujimoto K. J.* Electronic Couplings and Electrostatic Interactions Behind the Light Absorption of Retinal Proteins // Front. Mol. Biosci. — 2021. — Vol. 8. — P. 752700.

56. *Tsujimura M.*, *Ishikita H.* Insights into the protein functions and absorption wavelengths of microbial rhodopsins // J. Phys. Chem. B. — 2020. — Vol. 124, no. 52. — P. 11819–11826.

57. *Ryazantsev M. N.*, *Altun A.*, *Morokuma K.* Color tuning in rhodopsins: the origin of the spectral shift between the chloride-bound and anion-free forms of halorhodopsin // J. Am. Chem. Soc. — 2012. — Vol. 134, no. 12. — P. 5520–5523.

58. Ultrafast protein dynamics of bacteriorhodopsin probed by photon echo and transient absorption spectroscopy / J. T. Kennis [et al.] // J. Phys. Chem. B. — 2002. — Vol. 106, no. 23. — P. 6067–6080.

59. *Song L.*, *El-Sayed M.*, *Lanyi J.* Protein catalysis of the retinal subpicosecond photoisomerization in the primary process of bacteriorhodopsin photosynthesis // Science. — 1993. — Vol. 261, no. 5123. — P. 891–894.

60. Picosecond and nanosecond spectroscopies of the photochemical cycles of acidified bacteriorhodopsin / H. Ohtani [et al.] // Biochemistry. — 1986. — Vol. 25, no. 11. — P. 3356–3363.

61. Primary reactions of sensory rhodopsins / I. Lutz [et al.] // Proc. Natl. Acad. Sci. U.S.A. — 2001. — Vol. 98, no. 3. — P. 962–967.

62. Estimated acid dissociation constants of the Schiff base, Asp-85, and Arg-82 during the bacteriorhodopsin photocycle / L. Brown [et al.] // Biophys. J. — 1993. — Vol. 65, no. 1. — P. 124–130.

63. A unified view on varied ultrafast dynamics of the primary process in microbial rhodopsins / C.-F. Chang [et al.] // Angew. Chem. — 2022. — Vol. 134, no. 2. — e202111930.

64. Reconstitution of Gloeobacter violaceus rhodopsin with a light-harvesting carotenoid antenna / E. S. Imasheva [et al.] // Biochemistry. — 2009. — Vol. 48, no. 46. — P. 10948–10955.

65. Excitation energy-transfer and the relative orientation of retinal and carotenoid in xanthorhodopsin / S. P. Balashov [et al.] // Biophys. J. — 2008. — Vol. 95, no. 5. — P. 2402–2414.

66. Aspartate–Histidine interaction in the retinal Schiff base counterion of the light-driven proton pump of Exiguobacterium sibiricum / S. Balashov [et al.] // Biochemistry. — 2012. — Vol. 51, no. 29. — P. 5748–5762.

67. Acid–base equilibrium of the chromophore counterion results in distinct photoisomerization reactivity in the primary event of proteorhodopsin / C.-F. Chang [et al.] // Phys. Chem. Chem. Phys. — 2019. — Vol. 21, no. 46. — P. 25728–25734.

68. pH-dependent photoisomerization of retinal in proteorhodopsin / R. Huber [et al.] // Biochemistry. — 2005. — Vol. 44, no. 6. — P. 1800–1806.

69. Origin of the reactive and nonreactive excited states in the primary reaction of rhodopsins: pH dependence of femtosecond absorption of light-driven sodium ion pump rhodopsin KR2 / S. Tahara [et al.] // J. Phys. Chem. B. — 2018. — Vol. 122, no. 18. — P. 4784–4792.

70. Titration of aspartate-85 in bacteriorhodopsin: what it says about chromophore isomerization and proton release / S. P. Balashov [et al.] // Biophys. J. — 1996. — Vol. 70, no. 1. — P. 473–481.

71. Picosecond-millisecond dual-time-base spectroscopy of fluorescent photointermediates formed in the purple membrane of Halobacterium halobium / H. Ohtani [et al.] // Chem. Phys. Lett. — 1999. — Vol. 299, no. 6. — P. 571–575.

72. Femtosecond spectroscopy of acidified and neutral bacteriorhodopsin / T. Kobayashi [et al.] // Laser Applications in Life Sciences. Vol. 1403. — SPIE. 1991. — P. 407–416.

73. Picosecond fluorescence spectroscopy of the purple membrane of Halobacterium halobium in alkaline suspension / N. Kamiya [et al.] // Chem. Phys. Lett. — 1997. — Vol. 265, no. 6. — P. 595–599.

74. Directed evolution: methodologies and applications / Y. Wang [et al.] // Chem. Rev. — 2021. — Vol. 121, no. 20. — P. 12384–12444.

75. *Arnold F. H.* Directed evolution: bringing new chemistry to life // Angew. Chem. Int. Ed. — 2018. — Vol. 57, no. 16. — P. 4143–4148.

76. Structures of the archaerhodopsin-3 transporter reveal that disordering of internal water networks underpins receptor sensitization / J. F. B. Juarez [et al.] // Nat. Commun. — 2021. — Vol. 12, no. 1. — P. 1–10.

77. X-ray structure analysis of bacteriorhodopsin at 1.3 Å resolution / N. Hasegawa [et al.] // Sci. Rep. — 2018. — Vol. 8, no. 1. — P. 13123.

78. Crystal structure of the O intermediate of the Leu93 - Ala mutant of bacteriorhodopsin / J. Zhang [et al.] // Proteins. — 2012. — Vol. 80, no. 10. — P. 2384–2396.

79. *Muhammed M. T.*, *Aki-Yalcin E.* Homology modeling in drug discovery: Overview, current applications, and future perspectives // Chemical biology & drug design. — 2019. — Vol. 93, no. 1. — P. 12–20.

80. *Kelm S.*, *Shi J.*, *Deane C. M.* MEDELLER: homology-based coordinate generation for membrane proteins // Bioinformatics. — 2010. — Vol. 26, no. 22. — P. 2833–2840.

81. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules / A. Leaver-Fay [et al.] // Methods in enzymology. Vol. 487. — Elsevier, 2011. — P. 545–574.

82. *Yang J.*, *Zhang Y.* Protein structure and function prediction using I-TASSER // Curr. Protoc. — 2015. — Vol. 52, no. 1. — P. 5–8.

83. *Hill J. R.*, *Deane C. M.* MP-T: improving membrane protein alignment for structure prediction // Bioinformatics. — 2013. — Vol. 29, no. 1. — P. 54–61.

84. AlignMe—a membrane protein sequence alignment web server / M. Stamm [et al.] // Nucleic Acids Res. — 2014. — Vol. 42, W1. — W246–W251.

85. *Wu S.*, *Zhang Y.* MUSTER: improving protein sequence profile–profile alignments by using multiple sources of structure information // Proteins: Structure, Function, and Bioinformatics. — 2008. — Vol. 72, no. 2. — P. 547–556.

86. A comparative study of modern homology modeling algorithms for rhodopsin structure prediction / D. M. Nikolaev [et al.] // ACS omega. — 2018. — Vol. 3, no. 7. — P. 7555–7566.

87. *Bianco V.*, *Iskrov S.*, *Franzese G.* Understanding the role of hydrogen bonds in water dynamics and protein stability // Journal of Biological Physics. — 2012. — Vol. 38. — P. 27–48.

88. The roles of water in the protein matrix: a largely untapped resource for drug discovery / F. Spyrakis [et al.] // Journal of medicinal chemistry. — 2017. — Vol. 60, no. 16. — P. 6781–6827.

89. *Carugo O.*, *Bordo D.* How many water molecules can be detected by protein crystallography? // Acta Crystallographica Section D: Biological Crystallography. — 1999. — Vol. 55, no. 2. — P. 479–483.

90. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR / J. A. Ernst [et al.] // Science. — 1995. — Vol. 267, no. 5205. — P. 1813–1817.

91. *Morozenko A.*, *Stuchebrukhov A.* Dowser++, a new method of hydrating protein structures // Proteins. — 2016. — Vol. 84, no. 10. — P. 1347–1357.

92. *Michel J.*, *Tirado-Rives J.*, *Jorgensen W. L.* Prediction of the water content in protein binding sites // The journal of physical chemistry B. — 2009. — Vol. 113, no. 40. — P. 13337–13346.

93. *Yoon H.*, *Kolev V.*, *Warshel A.* Validating the water flooding approach by comparing it to grand canonical Monte Carlo simulations // The Journal of Physical Chemistry B. — 2017. — Vol. 121, no. 40. — P. 9358–9365.

94. *Zhang L.*, *Hermans J.* Hydrophilicity of cavities in proteins // Proteins: Structure, Function, and Bioinformatics. — 1996. — Vol. 24, no. 4. — P. 433–438.

95. *Ross G. A.*, *Morris G. M.*, *Biggin P. C.* Rapid and accurate prediction and scoring of water molecules in protein binding sites // PloS one. — 2012. — Vol. 7, no. 3. — e32036.

96. *Ross W.*, *Reichardt L.* Species-specific effects on the optical signals of voltage-sensitive dyes // The Journal of Membrane Biology. — 1979. — Vol. 48, no. 4. — P. 343–356.

97. *Sridhar A.*, *Ross G. A.*, *Biggin P. C.* Waterdock 2.0: Water placement prediction for Holo-structures with a pymol plugin // PloS one. — 2017. — Vol. 12, no. 2. — e0172743.

98. The ONIOM method and its applications / L. W. Chung [et al.] // Chem. Rev. — 2015. — Vol. 115, no. 12. — P. 5678–5796.

99. Gaussian 09, Revision D. 01, Gaussian / M. J. Frisch [et al.] // Inc.: Wallingford, CT. — 2009.

100. *Neese F.* The ORCA program system // Wiley Interdiscip. Rev. Comput. Mol. Sci. — 2012. — Vol. 2, no. 1. — P. 73–78.

101. *Altun A.*, *Yokoyama S.*, *Morokuma K.* Color tuning in short wavelength-sensitive human and mouse visual pigments: ab initio quantum mechanics/molecular mechanics studies // J. Phys. Chem. A. — 2009. — Vol. 113, no. 43. — P. 11685–11692.

102. An average solvent electrostatic configuration protocol for QM/MM free energy optimization: Implementation and application to rhodopsin systems / Y. Orozco-Gonzalez [et al.] // J. Chem. Theory Comput. — 2017. — Vol. 13, no. 12. — P. 6391–6404.

103. Comparison of multiple Amber force fields and development of improved protein backbone parameters / V. Hornak [et al.] // Proteins. — 2006. — Vol. 65, no. 3. — P. 712–725.

104. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers / M. J. Abraham [et al.] // SoftwareX. — 2015. — Vol. 1. — P. 19–25.

105. MOLCAS 7: the next generation / F. Aquilante [et al.] // J. Comput. Chem. — 2010. — Vol. 31, no. 1. — P. 224–247.

106. TINKER: Software tools for molecular design / J. W. Ponder [et al.] // Washington University School of Medicine, Saint Louis, MO. — 2004. — Vol. 3.

107. Structure optimization via free energy gradient method: Application to glycine zwitterion in aqueous solution / N. Okuyama-Yoshida [et al.] // J. Chem. Phys. — 2000. — Vol. 113, no. 9. — P. 3519–3524.

108. Progress in free energy perturbation: Options for evolving fragments / L. Zara [et al.] // Drug Discovery Today: Technologies. — 2021. — Vol. 40. — P. 36–42.

109. *Kochendoerfer G. G.*, *Mathies R. A.* Spontaneous emission study of the femtosecond isomerization dynamics of rhodopsin // J. Phys. Chem. — 1996. — Vol. 100, no. 34. — P. 14526–14532.

110. Chromophore structure in bacteriorhodopsin's O640 photointermediate / S. O. Smith [et al.] // Biochemistry. — 1983. — Vol. 22, no. 26. — P. 6141–6148.

111. *Lohrmann R.*, *Stockburger M.* Time-resolved resonance Raman studies of bacteriorhodopsin and its intermediates K590 and L550: Biological implications // J. Raman Spectrosc. — 1992. — Vol. 23, no. 10. — P. 575–583.

112. *Ames J. B.*, *Mathies R. A.* The role of back-reactions and proton uptake during the N - O transition in bacteriorhodopsin's photocycle: a kinetic resonance Raman study // Biochemistry. — 1990. — Vol. 29, no. 31. — P. 7181–7190.

113. *Smith S. O.*, *Mathies R. A.* Resonance Raman spectra of the acidified and deionized forms of bacteriorhodopsin // Biophys. J. — 1985. — Vol. 47, no. 2. — P. 251–254.

114. The photochemistry of sodium ion pump rhodopsin observed by watermarked femto-to submillisecond stimulated Raman spectroscopy / Y. Hontani [et al.] // Phys. Chem. Chem. Phys. — 2016. — Vol. 18, no. 35. — P. 24729–24736.

115. Resonance Raman study of halorhodopsin photocycle kinetics, chromophore structure, and chloride-pumping mechanism / J. B. Ames [et al.] // Biochemistry. — 1992. — Vol. 31, no. 50. — P. 12546–12554.

116. Photochemical Reaction Cycle and Proton Transfers in Neurospora Rhodopsin / L. S. Brown [et al.] // J. Biol. Chem. — 2001. — Vol. 276, no. 35. — P. 32495–32505.

117. A new group of eubacterial light-driven retinal-binding proton pumps with an unusual cytoplasmic proton donor / A. Harris [et al.] // Biochim. Biophys. Acta - Bioenerg. — 2015. — Vol. 1847, no. 12. — P. 1518–1529.

118. Low-temperature Raman spectroscopy of sodium-pump rhodopsin from Indibacter alkaliphilus: Insight of Na+ binding for active Na+ transport / Y. Nakamizo [et al.] // Phys. Chem. Chem. Phys. — 2021. — Vol. 23, no. 3. — P. 2072–2079.

119. Functional importance of the oligomer formation of the cyanobacterial H+ pump Gloeobacter rhodopsin / A. Iizuka [et al.] // Sci. Rep. — 2019. — Vol. 9, no. 1. — P. 10711.

120. Leptosphaeria rhodopsin: bacteriorhodopsin-like proton pump from a eukaryote / S. A. Waschuk [et al.] // Proc. Natl. Acad. Sci. U.S.A. — 2005. — Vol. 102, no. 19. — P. 6879–6883.

121. *Dioumaev A. K.*, *Wang J. M.*, *Lanyi J. K.* Low-temperature FTIR study of multiple K intermediates in the photocycles of bacteriorhodopsin and xanthorhodopsin // J. Phys. Chem. B. — 2010. — Vol. 114, no. 8. — P. 2920–2931.

122. Time-resolved IR spectroscopy reveals mechanistic details of ion transport in the sodium pump Krokinobacter eikastus rhodopsin 2 / M. Asido [et al.] // Phys. Chem. Chem. Phys. — 2019. — Vol. 21, no. 8. — P. 4461–4471.

123. Near-IR resonance Raman spectroscopy of archaerhodopsin 3: effects of transmembrane potential / E. C. Saint Clair [et al.] // J. Phys. Chem. B. — 2012. — Vol. 116, no. 50. — P. 14592–14601.

124. Directed evolution of bacteriorhodopsin for device applications / J. R. Hillebrecht [et al.] // Meth. Enzymol. — 2004. — Vol. 388. — P. 333–347.

125. *Altun A.*, *Yokoyama S.*, *Morokuma K.* Mechanism of spectral tuning going from retinal in vacuo to bovine rhodopsin and its mutants: multireference ab initio quantum mechanics/molecular mechanics studies // J. Phys. Chem. B. — 2008. — Vol. 112, no. 51. — P. 16883–16890.