Computational Modeling of Structural Changes in Selected Optically Regulated Proteins and Ligands

Sylwia Czach

Computational Modeling of Structural Changes in Selected Optically Regulated Proteins and Ligands

Sylwia Czach

Under the supervision of:

main supervisor
prof. dr hab. Wiesław Nowak

auxilairy supervior
dr inż. Jakub Rydzewski

Department of Biophysics, Institute of Physics
Faculty of Physics, Astronomy and Informatics

Nicolaus Copernicus University in Toruń

Toruń, 2025

Acknowledgements

I would like to express my sincere appreciation to my supervisors Wiesław Nowak and Jakub Rydzewski whose expertise, mentorship and constructive feedback have been essential to the completion of this dissertation. Their patience, understanding, dedication and encouragement have been a source of great motivation and I am deeply grateful for their support.

To my beloved, words cannot fully express my gratitude for your great and constant support throughout this process. Thank you for believing in me even when I doubted myself, for always lifting me up with your encouragement, and for your incredible smile that made even the most difficult moments easier to bear. Every moment we have shared has been a source of strength and inspiration. This thesis would not have been possible without you. From the bottom of my heart, thank you.

I would like to express my sincere gratitude to Katarzyna Walczewska-Szewc, Beata Niklas, Julia Duda, Karolina Mikulska-Rumińska, Łukasz Pepłowski and James Mattock for their support, collaboration and constructive discussions throughout this research project. Their professionalism and willingness to share ideas have been invaluable in driving this thesis forward, making the research process both productive and enjoyable.

I would like to thank my family and friends for their endless support, love and encouragement. To my family, your belief in me has been my greatest source of strength. To my friends, thank you for always being there, for your words of motivation and for bringing positivity to this experience. Your presence in my life has been invaluable and I am truly lucky to have you by my side.

I acknowledge Polish high-performance computing infrastructure PLGrid for awarding this project access to the LUMI supercomputer, owned by the EuroHPC Joint Undertaking, hosted by CSC (Finland) and the LUMI consortium through PLL/2024/07/017528. The support from IDUB/NCU Emerging Fields Nanoscale Biophysics is also acknowledge. PYP work was supported by the National Science Center in Poland (Sonata 2021/43/D/ST4/00920, "Statistical Learning of Slow Collective Variables from Atomistic Simulations"). PhD project has been conducted within Doctoral School of Exact and Natural Sciences (Academia Scientiarum Thoruniensis).

Computational Modeling of Structural Changes in Selected Optically Regulated Proteins and Ligands

Abstract

Although electromagnetic radiation is a harmful factor that degrades most organic macromolecules, some proteins can profit from interaction with light. Furthermore, these proteins have essential functions in all living organisms. Direct interaction with light in these proteins is possible because they contain a small light-sensitive molecule called a chromophore. After absorbing a photon, this molecule changes its spatial conformation, triggering a series of structural changes in the protein and influencing its function and the course of metabolic processes in cells.

The dissertation explores the application of molecular modeling methods to examine the mechanisms underlying light-induced conformational changes in two photoactive systems and to investigate the potential for optical regulation of the receptor P2X7, which plays a crucial role in glioma development. In the first photoactive system – bacterial phytochrome – advanced molecular dynamics (MD) methods were used to overcome high-energy barriers and study conformational changes induced by biliverdin photoisomerization. Previously unknown metastable states were revealed, and two signal transduction pathways from the chromophore-binding pocket to the remote protein domains were identified. The second photoactive system – photoactive yellow protein (PYP), which contains the chromophore *para*-coumaric acid, was studied using a new machine learning technique called spectral map. Data from the MD simulation were used to train neural networks and to construct slow collective variables that describe conformational changes upon light absorption. In the purinergic P2X7 receptor, the binding affinities of light-sensitive ligands to the orthosteric pocket were studied using molecular docking. The results showed that the compounds fit the receptor pocket sufficiently well, demonstrating that using photoactive ligands to regulate chromophore-free proteins is a promising novel drug development approach.

Komputerowe modelowanie zmian strukturalnych w wybranych białkach i ligandach regulowanych optycznie

Abstrakt

Chociaż promieniowanie elektromagnetyczne jest czynnikiem szkodliwym, który degraduje większość makrocząsteczek organicznych, niektóre białka mogą czerpać korzyści z interakcji ze światłem. Ponadto białka te pełnią istotne funkcje we wszystkich organizmach żywych. Bezpośrednia interakcja tych białek ze światłem jest możliwa, ponieważ zawierają one małą cząsteczkę wrażliwą na światło, zwaną chromoforem. Po absorpcji fotonu cząsteczka ta zmienia swoją konformację przestrzenną, wywołując szereg zmian strukturalnych w białku i wpływając na jego funkcję oraz przebieg procesów metabolicznych w komórkach.

Rozprawa przedstawia wyniki zastosowania metod komputerowych do badania zmian konformacyjnych indukowanych światłem w dwóch białkowych układach fotoaktywnych oraz badania możliwości optycznej regulacji receptora P2X7 istotnego w rozwoju glejaka. Zmiany konformacyjne zachodzące w fitochromie bakteryjnym powodowane fotoizomeryzacją biliwerdyny badane były przy wykorzystaniu metod wzmocnionego próbkowania. Odkryto nieznane wcześniej stany metastabilne i zidentyfikowano dwa szlaki przekazywania sygnału z kieszeni wiążącej chromofor do pozostałych domen białka. Fotoaktywne białko żółte (PYP), posiadające jako chromofor kwas *para*-kumarowy, jako układ modelowy, zbadane zostało przy zastosowaniu nowej metody, opartej o uczenie maszynowe, zwanej mapą spektralną. Dane uzyskane z symulacji dynamiki molekularnej wykorzystano do trenowania sieci neuronowych i konstruowania wolnych zmiennych kolektywnych. Dokowanie molekularne zastosowano do badania powinowactwa ligandów fotoizomeryzujących do kieszeni ortosterycznej receptora purynergicznego P2X7. Wyniki pokazały, że związki fotoaktywne wiążą się do kieszeni tego receptora wystarczająco dobrze. W efekcie stwierdzono, że stosowanie fotoaktywnych ligandów do regulacji kanałów jonowych i receptorów stanowi obiecującą metodę opracowywania nowych leków.

List of abbreviations

MD	Molecular Dynamics
OM/MM	Quantum Mechanics/Molecular Mechanics
ML	Machine Learning
BLUF	Blue Light Using Flavin
GFP	Green Fluorescent Protein
pCA	*para*-Coumaric Acid
PYP	Photoactive Yellow Protein
PAS	Per-Arnt-Sim
BphP	Bacterial Phytochrome
BV	Biliverdin
GPCR	G-Protein-Coupled Receptor
ChR	Channelrhodopsin
PHR	Photolyase Homology Region
CCE	Cryptochrome C-terminal Extension
FAD	Flavin Adenine Dinucleotide
MTHF	Levomefolic Acid, L-5-methyltetrahydrofolate
LOV	Light, Oxygen, Voltage
FMN	Flavin Mononucleotide
FPs	Fluorescent Proteins
BLAST	Basic Local Alignment Search Tool
AI	Artificial Intelligence
FF	Force Fields
REMD	Replica Exchange Molecular Dynamics
US	Umbrella Sampling
CVs	Collective Variables
OCP	Orange Carotenoid Protein

VES	Variational Enhanced Sampling
TD-DFT	Time-Dependent Density Functional Theory
CASSCF	Complete Active Space Self-Consistent Field
CC	Coupled Cluster
ADC	Algebraic Diagrammatic Construction
TDSE	Time-Dependent Schrödinger Equation
NAMD	Nonadiabatic Dynamics Molecular Dynamics
PES	Potential Energy Surfaces
NACR	Non-Adiabatic Coupling Vector R
FSSH	Fewest Switches Surface Hopping
NNP	Neural Network Potential
GNN	Graph Neural Networks
PCM	Photosensory Core Module
cGMP	Cyclic Guanosine Monophosphate
CBD	Chromophore Binding Domain
HO	Heme Oxygenase
LINCS	Linear Constraint Solver
FE	Free Energy
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
CNS	Central Nervous System
PNS	Peripheral Nervous System
ANS	Autonomic Nervous System
ATP	Adenosine Triphosphate
BBB	Blood Brain Barrier
CSF	Cerebrospinal Fluid
WHO	World Health Organization
TMZ	Temozolomide
MTIC	Methyl-Triazeno-Imidazole-Carboxamide
MMR	Mismatch Repair
TM	Transmembrane
BzATP	2',3'-O-(4-Benzoilobenzoilo)Adenozyno-5'-Trifosforan
ADP	Adenosine Diphosphate
AP4	Adenosine Tetraphosphate
AP5	P1,P5-Di(adenosine-5'-)Pentaphosphate
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine

pLDDT	Predicted Local Distance Difference Test
PAE	Predicted Aligned Error
MSA	Multiple Sequence Alignment

Publications *

Sylwia Czach, Jakub Rydzewski, Understanding Mechanistic Transitions in Photoactive Yellow Protein with Machine Learning, (*in preparation*)
PhD candidate's contribution: Searching and reviewing the literature. Force field parameterization for non-standard system elements. Conducting molecular dynamics simulations. Analysis and visualization of results. Writing.
(CRediT: data curation, formal analysis, investigation, methodology, validation, visualization, writing – original draft, writing – review and editing.)

Sylwia Czach, Jakub Rydzewski, Wiesław Nowak, Illuminating Protein Dynamics: A Review of Computational Methods for Studying Photoactive Proteins, 2025, (*to be published*)
PhD candidate's contribution: Developing a structure, identifying issues and questions of the work. Searching, reviewing and analyzing the literature. Preparing figures, writing.
(CRediT: conceptualization, investigation, visualization, writing – original draft, writing – review and editing.)

Sylwia Czach, Katarzyna Walczewska-Szewc, Investigating substrate binding mechanism in prolyl oligopeptidase through molecular dynamics, 2025, *Physical Biology*, 22, 5, https://doi.org/10.1088/1478-3975/adf429
PhD candidate's contribution: Searching and reviewing the literature. Identifying issues and questions of the work, developing research questions and hypotheses. Preparation of ligand structures, molecular docking, and molecular dynamics simulations. Analysis and visualization of results. Writing.
(CRediT: conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing – original draft, writing – review and editing.)

* PhD candidate's contribution according to the Contributor Role Taxonomy (CRediT)

Bartosz Szymczak, Joanna Czarnecka, **Sylwia Czach**, Wiesław Nowak, Katarzyna Roszek, Purinergic approach to effective glioma treatment with temozolomide reveals enhanced anti-cancer effects mediated by P2X7 receptor, 2023 *Cellular Signalling*, 106, 110641, https://doi.org/10.1016/j.cellsig.2023.110641
PhD candidate's contribution: Searching and reviewing the literature. Bioinformatics analysis of amino acid sequences. Generation and validation of a homology model. Ligand selection and preparation. Molecular docking. Analysis and visualization of results. Writing.
(CRediT: data curation, formal analysis, investigation, methodology, validation, visualization, writing – review and editing.)

Jakub Rydzewski, Katarzyna Walczewska-Szewc, **Sylwia Czach**, Wiesław Nowak, Krzysztof Kuczera, Enhancing the inhomogeneous photodynamics of canonical bacteriophytochrome, 2022 *The Journal of Physical Chemistry B*, 126 (14), 2647-2657, https://doi.org/10.1021/acs.jpcb.2c00131
PhD candidate's contribution: Searching and reviewing the literature. Analysis and discusing of the results. Writing.
(CRediT: investigation, formal analysis, validation, writing – review and editing.)

Contents

"Ah, not in knowledge is happiness, but in the acquisition of knowledge! In for ever knowing, we are for ever blessed; but to know all were the curse of a fiend."

"The Power of Words",
Edgar Allan Poe

1

General introduction

Chapter based on:

Authors: Sylwia Czach, Jakub Rydzewski, Wiesław Nowak

Original title: Illuminating Protein Dynamics: A Review of Computational Methods for Studying Photoactive Proteins,

Published in: *to be published in Springer*, 2025

https://doi.org/10.48550/arXiv.2503.21418

1.1 Purpose and Scope of the Thesis

Each chapter of the thesis focuses on a different research problem related to the study of optically regulated proteins.

1. Introduction
   Based on the publication by Czach et al. [1] and presents a comprehensive review of current knowledge on the biochemistry and biophysics of photoactive proteins, as well as computer methods used to study these systems. The aim of this chapter is to describe and explain the structure and mechanisms that are responsible for the function of photoactive proteins. The text outline the challenges associated with studying such systems and review computational methods that address these challenges, making it possible to study the mechanisms of conformational changes in photoactive proteins.

2. Bacteriophytochrome and biliverdin IXα.
   The aim of this chapter is to examine, using enhanced sampling molecular dynamics simulation, the initial stages of mechanical signal transduction in the bacteriophytochrome-biliverdin complex, as well as the metastable states of the photoproducts Pr and Pfr. This chapter presents evidence of significant differences in the free energy landscape between the Pr and Pfr conformers. It also emphasises the importance of the protein environment in forming and expanding metastable states.

3. Photoactive Yellow Protein and *p*-coumaric acid.
   The main goal of this chapter was to investigate a well-known system of photoactive yellow protein using new methods based on machine learning, called spectral map. Data obtained from the series of molecular dynamics simulation were used to train neural networks, construct slow collective variables and the corresponding free-energy landscapes, and explain the relation between fluctuations in the chromophore pocket and large structural rearrangements observed in PYP.

4. Light-induced regulation of proteins, photoactivated ligands of P2X7 receptor.
   In this chapter, the potential use of photoactive compounds for regulating the purinergic ion channel P2X7 was investigated through a series of molecular docking experiments. This chapter presents a homology model of the human receptor, based on the human sequence and the structure of the rat receptor. The binding affinity of photoactive compounds for the ATP-binding pocket of the homology P2X7 receptor was evaluated in relation to the docking results for the native ligand, ATP.

ALTHOUGH there are over 500 naturally occurring amino acids, only 22 α-amino acids play a central role in the formation of proteins, making them the most biologically important. These α-amino acids are the basic building units of proteins, and their unique sequences and structural arrangements determine the wide range of biological functions they perform. For example, the human proteome contains at least 20,000 most well-studied human protein structures, each of which is involved in essential molecular and cellular processes such as enzymatic catalysis, signal transduction, immune response, and structural support [2–5].

The spatial organization of amino acids in proteins is not fixed or rigid; rather, proteins are highly dynamic structures that constantly adapt to their environment. This adaptability is primarily due to conformational changes, which are structural rearrangements that occur in response to various external and internal factors. These changes are fundamental to the regulation of metabolic processes and the overall functionality of living organisms. Several external factors can affect the conformation of proteins, including changes in environmental pH, interactions with other biomolecules (such as substrates, inhibitors, toxins, and drugs), and exposure to physical agents such as electromagnetic radiation. These structural changes can enhance or inhibit protein activity, ultimately affecting biochemical pathways, cells condition, and body homeostasis [2, 6–8].

Understanding the dynamic properties of proteins is crucial to fields such as biochemistry, molecular biology, and pharmacology, as it provides insights into disease mechanisms, drug design, and the development of new therapeutic treatments. Biophysics research groups pay a lot of attention to protein dynamics.

Electromagnetic radiation is widely recognized for its ability to damage and irreversibly denature most proteins and nucleic acids. According to the principles of molecular biology, such damage can lead to changes in protein and nucleic acids structure, potentially resulting in mutations. However, a unique class of proteins exhibits remarkable resistance to light-induced damage. Rather than being harmed by exposure to light, these proteins have evolved to detect and process light signals, thereby triggering specific biological responses.

At the core of this ability is the presence of a chromophore, an organic molecule embedded in the protein structure that enables the absorption of electromagnetic radiation. This absorption often initiates a process known as photoisomerization, which forms the basis of light-activated molecular switches. When light is absorbed, the chromophore undergoes conformational changes, which in turn trigger a cascade of structural transitions both within the protein itself and in its immediate environment. These structural changes can be transmitted to other macromolecules, affecting entire metabolic pathways within

the cell [9–13].

1

A key feature of this light-responsive system is its ability to revert to its original state, allowing the cycle to be repeated. This dynamic transition between active and inactive states, known as photocycling, which is the foundation of the function of these molecular switches, allows organisms to harness light for various biological processes.

A large number of metabolic processes critical to the proper functioning of all living organisms are pathways that depend on the interaction of biological macromolecules with light. The most important of these processes is photosynthesis. However, they also include pigmentation and phototaxis in prokaryotes and plants, vision and circadian regulation, and light-dependent gene expression and DNA repair. Light-regulated proteins are also involved in cell aggregation and colony formation in bacteria. Because of their predictable responses to light, they have applications in fields such as optogenetics, bioimaging, biotechnology, and medicine [13–18].

Current knowledge about optically regulated proteins is primarily based on structural and spectroscopic data. High-resolution crystallography and cryo-electron microscopy data provide access to the detailed structures of many photoreceptors. Moreover, the monitoring of light-induced processes is made possible by time-resolved spectroscopy [19–21]. However, a full microscopic characterization of the photoactivation process of multiple photoreceptors remains out of reach. This is due to the difficulty of experimentally reconstructing photoactivation at different timescales. These range from the ultrafast timescales of femtoseconds and picoseconds, which are characteristic of photochemical events, to the slower millisecond timescale of protein conformational changes. Capturing all ultrafast intermediate states and understanding the interactions between chromophore dynamics and protein conformational changes is a significant challenge. The complexity of researching light-induced reactions stems primarily from the fact that they occur in excited states [22]. The gap in knowledge about the mechanisms of photoactive systems resulting from the limitations of experimental methods can be overcome by using atomistic computational approach [13]. Methods that have been developed to explore organic photoactive systems include bioinformatics [23–25], molecular dynamics (MD) [17], enhanced sampling MD [26–29], and hybrid quantum-classical (QM/MM) MD [30, 31]. The current trend in computational studies is directed towards the use of machine learning (ML) methods to analyze organic systems, including those that are light-regulated. The limitations of current computational techniques when it comes to predicting dynamics across multiple time scales, coupled with the need for a deeper understanding of photoactive proteins in their excited states, are driving the rapid development of new computational methods [32–35].

1.2 Overview of Photoactive Proteins

Photoactive proteins have the unique ability to absorb light at a specific wavelength, triggering precise biological activation. Absorption of light provides the energy necessary to overcome high activation barriers, allowing proteins to undergo conformational changes. These changes can be reversible. Proteins can return to their original structural state after fulfilling their role, allowing them to respond repeatedly to light stimuli. This cyclic process ensures that these proteins can function effectively through multiple cycles of activation, allowing them to periodically interact with other molecules and regulate various cellular processes, as shown in Figure 1.1.

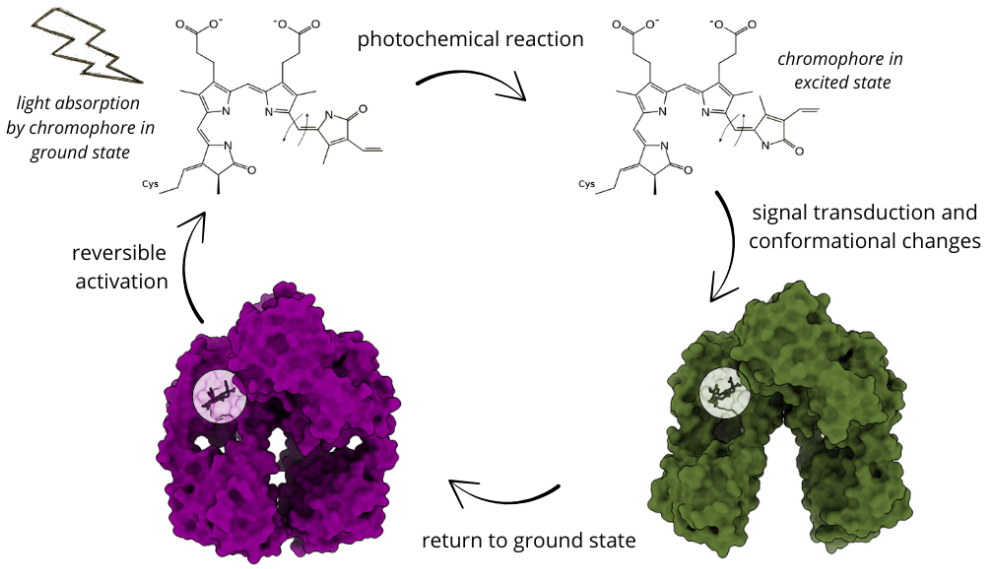
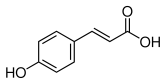
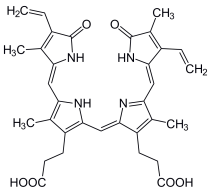
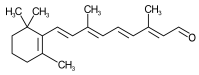
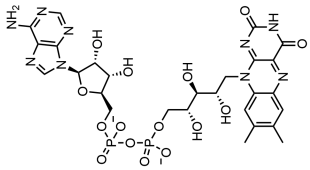
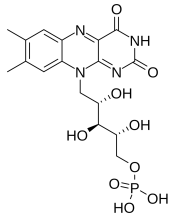
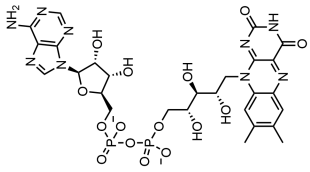
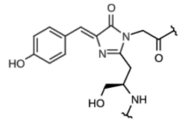


Figure 1.1: Schematic representation of the action of photoactive proteins (protein in the dark state with chromophore in ground state PDB: 4O0P, protein in illuminated state with chromophore in excited state PDB: 4O01).

Photosensitive proteins can be classified into several groups based on the chemical structure of the chromophore and their mechanism of action [12]. Table 1.1 and Figure 1.2 show the emerging classification of the proteins reported in the literature into seven families: xanthopsins [36, 37], phytochromes [38], rhodopsins [39, 40], cryptochromes [41], phototropins [42], blue light using flavin (BLUF) proteins [43], and fluorescent proteins [44].

1

Table 1.1: Classification and overview of photoactive protein families: presentation of protein families, chromophore types, representative examples of chromophores, and photochemical mechanisms underlying light-driven biological processes.

Protein family	Chromophore	Chromophore example	Mechanism
Xantophins	Polyenes	 *p*-Coumaric acid	*Trans–cis* photoisomerization
Phytochromes	Tetrapyroles	 Biliverdin	*Trans–cis* photoisomerization
Rhodopsins	Polyenes	 Retinal	*Trans–cis* photoisomerization
Cryptochromes	Aromatic	 Flavin adenine dinucleotide	Electron transfer
Phototropins	Aromatic	 Flavin adenine mononucleotide	Cysteinyl adduct formation
BLUF proteins	Aromatic	 Flavin adenine dinucleotide	Proton transfer
Fluorescent proteins	Amino acids	 Serine, tyrosine, glycine	Fluorophore formation

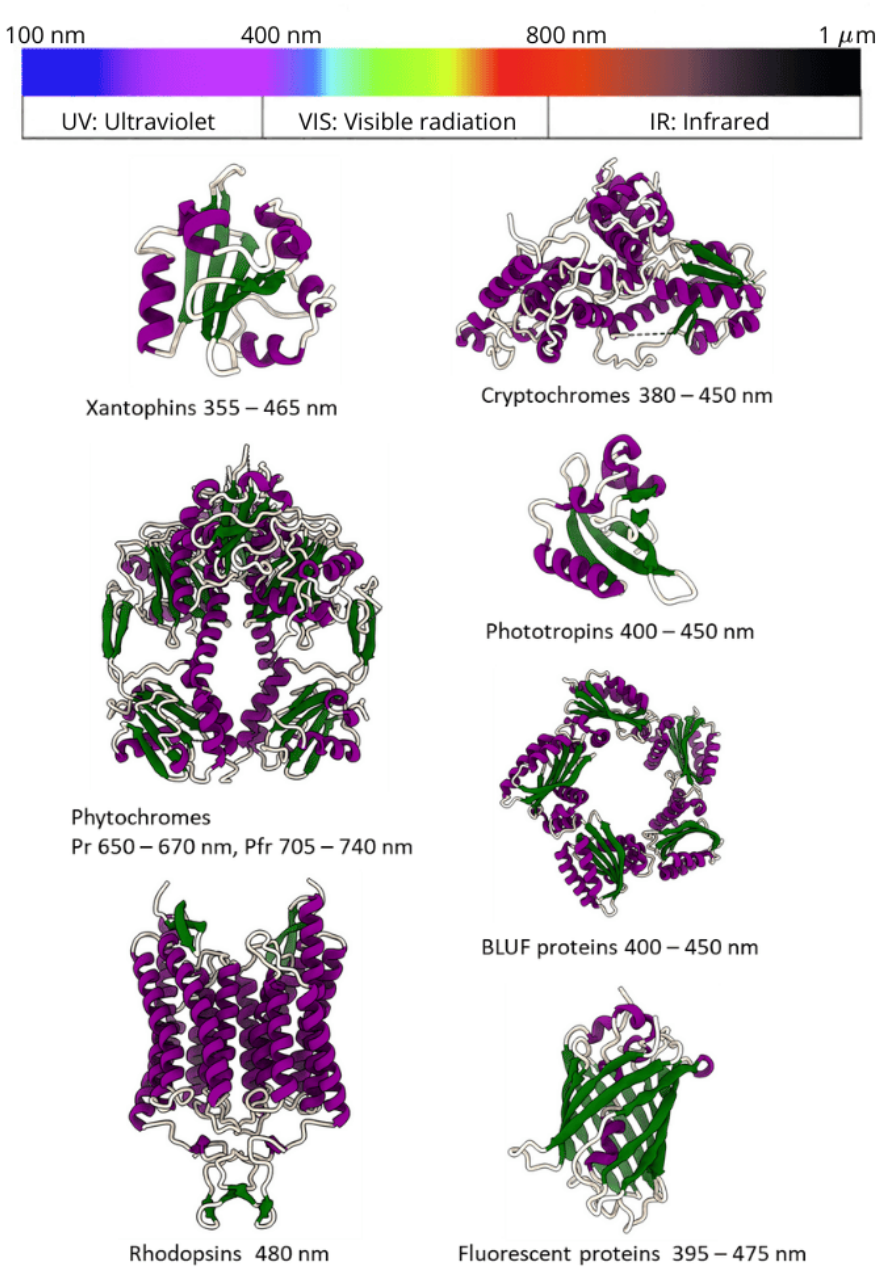


Figure 1.2: Visible light spectrum and visualization of the structures of representative proteins from each family with information on the wavelength to which the chromophore reacts (xantophins PDB:1TS6 [45], cryptochromes PDB:5T5X [46], phytochromes PDB:4O0P [47], phototropins PDB:8QI8 [48], BLUF proteins PDB:1X0P [49], rhodopsins PDB:3UG9 [50], fluorescent proteins PDB:1GFL [51]).

1

Mechanism of actions can differ within each protein family. Xanthopsins, phytochromes, and rhodopsins act by photoisomerization of the chromophore. Cryptochromes are activated by electron transfer, BLUF proteins by proton transfer, and phototropins require the formation of cysteinyl adducts. Unlike the families mentioned above, fluorescent proteins have a different mechanism of responding to light. A fluorophore is located within the protein structure. In green fluorescent protein (GFP), it consists of three amino acids (serine, tyrosine, glycine) that undergo transformations in the environment of the protein: cyclization and dehydration, leading to the formation of a fluorescent system of conjugated double bonds [52, 53].

Xanthopsins are a group of photosensitive proteins characterized by the presence of *para*-coumaric acid (*p*CA) a derivative of hydroxycinnamic acid as a chromophore. This molecule is covalently bound to the conserved cysteine residue through a thiol ester bond, forming a Schiff base. Photoactive yellow protein (PYP), which was initially discovered in the halophilic purple bacteria *Halorhodospira halophila*, is a well-studied example of the xanthopsin family. PYP has a characteristic α/β fold, as shown in the Figure 1.3, which was used as the structural basis for the definition of the Per-Arnt-Sim (PAS) domain. This domain, characterized by a central five-stranded β-sheet flanked by α-helical regions, plays a crucial role in biological signaling in all life forms [11, 36].

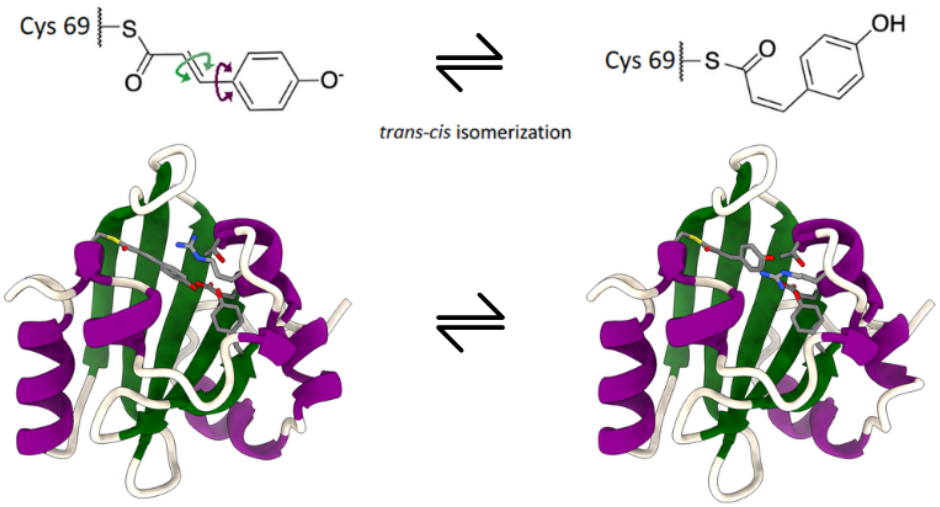


Figure 1.3: PYP in pG state on the left and pB state on the right, above both structures, the chromophore *p*-coumaric acid is shown in states corresponding to the structures (PDB:1TS6.) [45].

The chromophore is located within a well-defined hydrophobic binding pocket, which

is formed by residues that stabilize *p*CA through hydrogen bonding and hydrophobic interactions. The thiol-ester linkage between *p*CA and cysteine and the hydrogen bonds involving tyrosine and glutamine in the case of PYP are key aspects that stabilize the deprotonated phenolate form of *p*CA in the dark state. The surrounding protein environment controls the protonation state of *p*CA and its ability to respond to light [11, 54, 55]. In the dark state, the absorption maximum typically takes place around 446 nm (in the blue light region). When light is absorbed, the chromophore undergoes *trans–cis* isomerization, initiating the photocycle. The photocycle involves a transition from the ground state, through an intermediate state to a signal state, and finally the recovery of the system. This transition is associated with changes to the hydrogen bond network and the structure of the protein [55, 56].

This process enables xanthopsins to act as photoreceptors, allowing microorganisms to respond to light signals from their environment. They are involved in regulating phototaxis, protecting against photooxidative stress and regulating signaling pathways that affect metabolism and gene expression [12, 57].

Phytochromes are a family of photoreceptor proteins that act as molecular switches and enable organisms to detect and respond to light in the red and far-red regions of the electromagnetic radiation spectrum (typically covering wavelengths from ~600 to 750 nm). They play a crucial role in regulating cellular processes in response to environmental light signals [47, 58, 59]. In plant cells, proteins from the phytochrome family regulate fundamental physiological processes, including photosynthesis, seed germination, flowering, and phototaxis. In bacteria, these proteins regulate various intracellular processes, including gene expression, protein phosphorylation, and protein degradation. They are also thought to play a role in the transport of calcium and other ions [60–64]. Phytochromes are classified according to their origin, structure, spectral properties, and photochemical behavior. Although they all play a similar role as red/far-red light photoreceptors using a linear tetrapyrrole chromophore, phytochromes differ significantly in terms of their architecture and biological function. Tetrapyrroles are a class of compounds that contain four pyrrole or pyrrole-like rings. These rings can be connected either linearly or cyclically; in either case, a double bond pattern is formed that allows for the absorption of electromagnetic radiation. They exhibit a response to light within the red and far-red range of the visible spectrum. As molecular switches, they occur in two forms: Pfr, which is triggered by far-red light (~705–740 nm), and Pr, which is responsive to red light (~650–670 nm)(Figure 1.2) [65–68].

Figure 1.4 shows a representative protein from the phytochrome family: a bacterial phy-

1

tochrome (BphP), which contains a biliverdin IXα (BV) molecule as a chromophore. The BphP photosensor's basic module has a three-part region architecture consisting of well-conserved domains: PAS, GAF, and PHY [59, 68].

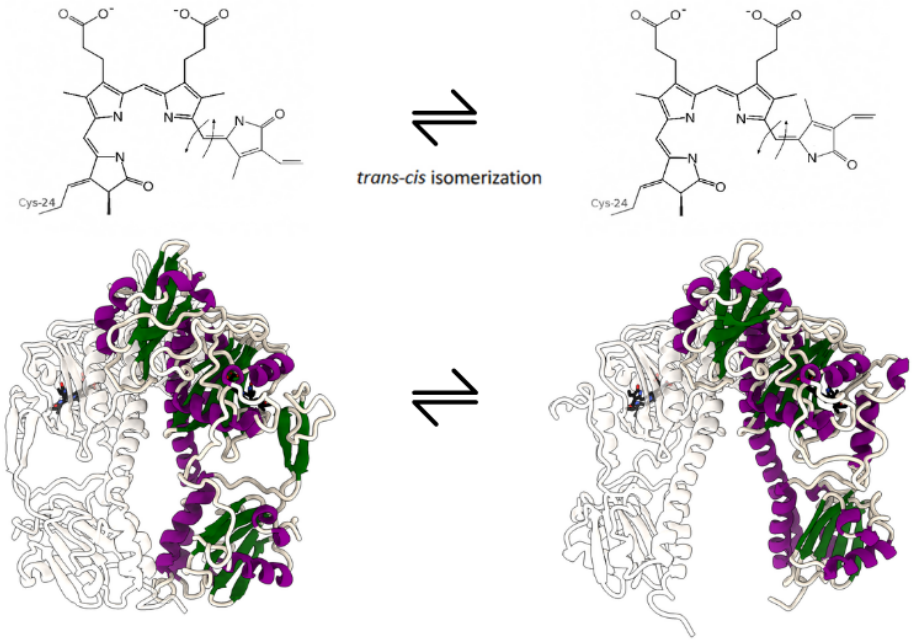


Figure 1.4: Structural features of the BphP-BV complex. The photoisomerization of biliverdin from the *ZZZ* state to the *ZZE* state (top). Bacterial phytochrome in the Pr state on the left and in the Pfr state on the right [59] (protein in the dark state with chromophore in ground state PDB: 4O0P, protein in illuminated state with chromophore in excited state PDB: 4O01). Structures colored by secondary structures with chromophore in stick representation in black.

Rhodopsins are a group of proteins that include visual rhodopsins, found in *Eukaryotes*, as well as ion-transporting rhodopsins and proton pumps, found in *Prokaryotes* and *Archaea*. Furthermore, some rhodopsins function as sensory proteins involved in signaling pathways, which can be found in all organisms [69–73].

Eukaryotic rhodopsin is a membrane protein and a member of the G-protein-coupled receptor (GPCR) class, characterized by seven transmembrane helices that form a ligand-binding region. It has two main components: the opsin protein and the chromophore 11-*cis*-retinal (in ground state), which is covalently linked to Lys296, form of a protonated Schiff base, located in the seventh transmembrane domain. The chromophore absorbs radiation in the visible light range, at a wavelength of approximately 498 nm. Following the absorption of a photon, the chromophore undergoes photoisomerization from 11-*cis*-retinal to all-*trans*-retinal, resulting in conformational changes in the protein and

activation of the photoreceptor [74, 75]. Due to their role in vision, eukaryotic rhodopsins are particularly important and well-studied photoactive proteins.

In contrast, rhodopsins derived from prokaryotic organisms and algae have found application in biotechnological processes as they lack coupling with G proteins and have a simpler activation mechanism than their eukaryotic equivalents. One such rhodopsin is the channelrhodopsin (ChR) subfamily, which acts as a light-gated ion channel. These proteins perform the role of photoreceptors in unicellular green algae, controlling their phototaxis responses [70].

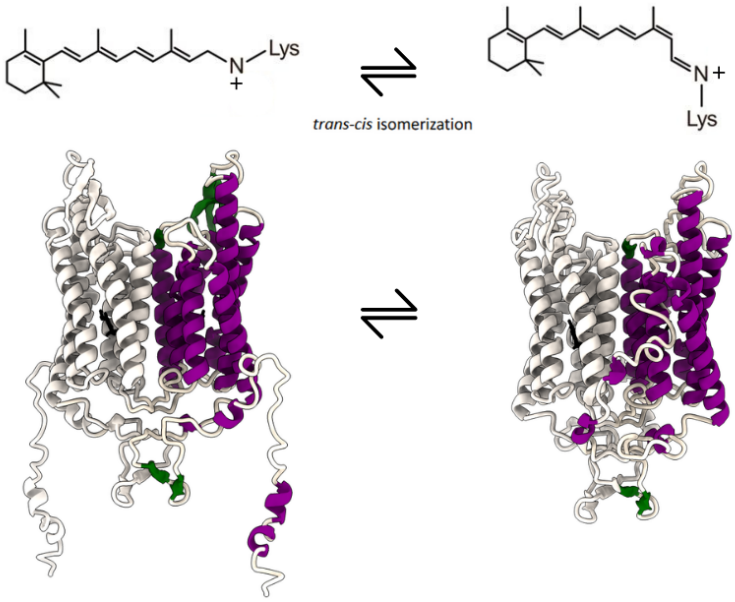


Figure 1.5: At the top, photoisomerization of the chromophore from *trans*-retinal to 13-*cis*-retinal. Below are the structural changes in the protein. On the left, the ground state protein with the chromophore in the *trans* state (PDB: 7C86) and on the right, the protein with 13-*cis*-retinal (PDB: 7E71) from *Chlamydomonas reinhardtii*. Structures colored by secondary structures with chromophore in stick representation in black.

The first proteins belonging to this subfamily to be discovered were the Channelrhodopsin-1 (ChR1) and Channelrhodopsin-2 (ChR2) proteins derived from the model organism *Chlamydomonas reinhardtii*. A chimera between ChR1 and ChR2, the recombinant hybrid protein is also used in research. Non-eukaryotic variants are also composed of seven transmembrane α-helices. They contain a chromophore that is covalently linked to the rest of the protein, forming a protonated Schiff base. In the ground state, this chromophore is all-*trans*-retinal (Figure 1.5) (unlike in eukaryotic rhodopsins, where it is 11-*cis*-retinal) [50, 76–78]. ChRs, when expressed in cells of other organisms, enable the light control of electrical excitability, ion flow, and other cellular processes using light, a

method known as optogenetics [79–82].

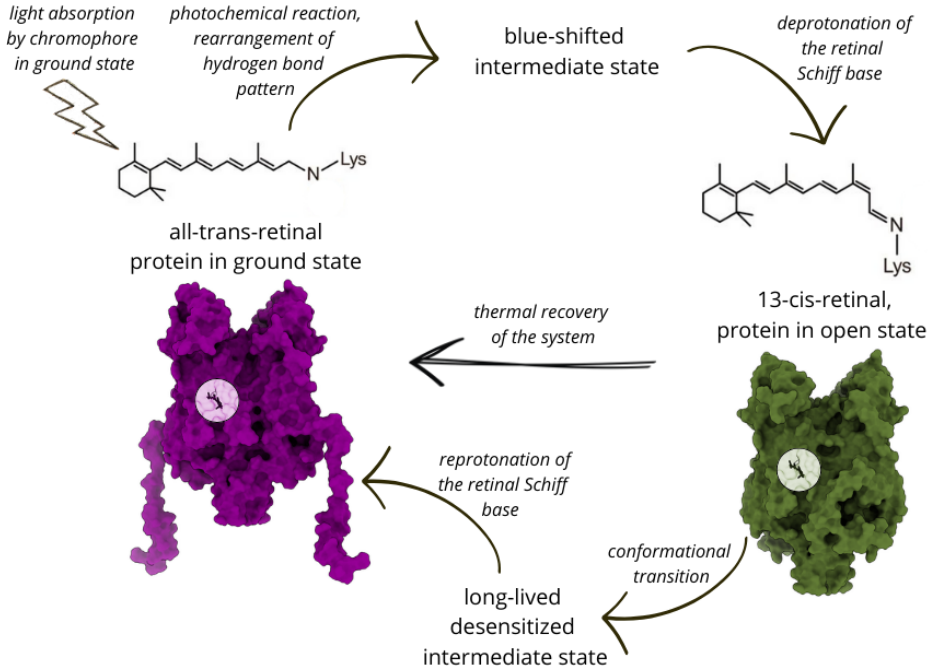


Figure 1.6: Schematic representation of the photocycle by which channelrhodopsin acts. The protein is shown in purple in its ground state with the all-trans chromophore. When radiation is absorbed, the hydrogen bond network rearranges and the protein transitions to an open form with the 13-cis chromophore via an intermediate state. Return to the ground state can occur in two ways: directly, via thermal reversion, or via an intermediate state. [83]

ChRs form ion channels directly, and, since they do not require coupling with a G protein for activation, cell depolarization involving them occurs very efficiently. Photocycle is initiated by the absorption of blue light with a maximum wavelength of 480 nm by the chromophore molecule in the ground state (Figure 1.6). This leads to a conformational change from all-*trans*-retinal to 13-*cis*-retinal. After isomerization, the chromophore and local side chains relax. As the system progresses towards the intermediate state, small shifts in nearby helices and rearrangement of hydrogen bonding occur. These initial relaxations create the conditions necessary for proton transfer [84, 85]. From a chemical perspective, the key step is the deprotonation of the retinal Schiff base, which results in the formation of a blue-shifted intermediate. The main pore opening in ChR2 is usually triggered by Schiff base deprotonation [86, 87]. Signal propagation leads to larger conformational changes in transmembrane helices, with the opening of internal gates allowing more water to penetrate the internal region. This process results in the formation of the red-shifted intermediate, which correlates with the open state. These structural changes

then propagate through the protein, leading to the opening of the channel and allowing ions to flow. [83, 88]. The conducting red-shifted state is transient and can decay in two ways: it can return to the dark ground state through thermal recovery, or it can pass into a long-lived desensitized intermediate state that is spectrally distinct and less conducting [83,89]. The dark chromophore can be regenerated through reprotonation of the Schiff base. Within a few milliseconds, the retinal relaxes and returns to its ground all-*trans* form. Structural changes then cause the channel to close again, blocking the ion flow [50,76–78].

Cryptochromes are proteins that can be categorized as flavoproteins (pigment proteins), which mostly operate as blue light receptors. The term 'cryptochrome' comes from the Greek words *kryptos*, meaning 'hidden', and *chroma*, meaning 'color'. This reflects their initially 'hidden' role as photoreceptors. Although they exhibit structural similarities to DNA photolyases (bacterial enzymes activated by light and involved in repairing UV-induced DNA damage), they do not demonstrate effective DNA repair activity. Cryptochromes are highly conserved proteins that can be found in all kingdoms of life [90–93]. In plants, blue light photoreception is a mechanism by which developmental signals, such as hypocotyl elongation, seed germination, and pigment accumulation, can be transmitted [94, 95]. In animals, cryptochromes participate in processes involving circadian clock regulation and may play a role in navigation and magnetoreception [93, 96].

Cryptochromes typically consist of two domains: an N-terminal photolyase homology region (PHR) of around 500 amino acids and a cryptochrome C-terminal extension (CCE) domain of varying size. The PHR domain is highly conserved in plants, animals, and insects. It is involved in binding the chromophore flavin adenine dinucleotide (FAD) and, in some cases, a secondary chromophore site binds 5,10-methenyltetrahydrofolate (pterin or MTHF). The CCE domain plays a key role in protein-protein interactions and signal transduction. While it often exhibits a disordered structure in solution, it undergoes conformational rearrangements upon light activation in plant cells. Differences in this domain between species determine functional diversity [92, 94, 97, 98]. Light in the 320–500 nm range is primarily absorbed by the main chromophore: FAD. In some organisms, particularly plants, MTHF or other pterins can act as antennas, harvesting photons and transferring excitation energy to FAD. In *Arabidopsis thaliana*, the flavin resting state in cryptochromes is the oxidized (FAD) form. Light absorption induces the accumulation of a mixture of radical ($FADH^{\circ}$) and reduced ($FADH^{-}$) flavin. This triggers electron transfer through a conserved chain of tryptophan residues, known as the 'Trp triad' electron transport pathway. Light-induced redox changes in FAD cause conformational changes in the protein structure, converting cryptochromes from an inactive to a signaling state. This

is followed by the reorganization of the hydrogen bond network near the chromophore binding pocket. The PHR and CCE domains then reorient and expose protein-protein interaction motifs [93, 99–103].

Phototropins are blue light photoreceptor proteins (flavoproteins) that exhibit kinase activity. They are primarily found in plants and algae, where they regulate a variety of light-dependent physiological processes. They are not found in animal cells. The name comes from their key role in phototropism. They are also responsible for opening the stomata, moving the chloroplasts, and positioning the leaves [104, 105]. Evolution occurred through proteins containing the LOV (Light, Oxygen, Voltage) domain, which are widespread among bacteria, fungi, and plants. In plants, they have probably adapted to specialized functions related to light perception. They share a similar structure to proteins from the PAS sensory domain family [106, 107].

The overall protein organization includes a photosensitive N-terminal region containing two LOV domains, LOV1 and LOV2. It also features a C-terminal signaling region, comprising a serine-threonine protein kinase domain. The composition of each LOV domain is approximately 100–120 amino acids, and there is structural similarity with PAS domains. A conserved chromophore-binding pocket in each LOV domain binds one flavin mononucleotide (FMN) molecule non-covalently. The LOV1 domain serves as both a stabilizer and a regulatory element. It participates in dimerization, thereby increasing the stability of photoreceptors and light sensitivity. The LOV2 domain is the main photoreceptor and functions as a molecular switch. Light absorption triggers conformational changes, leading to kinase activation [108–111]. In the phototropin structure, FMN acts as a light-absorbing co-factor that absorbs blue light (at around 450 nm). When excited, FMN forms a covalent adduct with the conserved cysteine residue in the LOV domains. This covalent bond between the cysteine residue and FMN represents the signal state. The return to the dark state occurs spontaneously within a few seconds. In the dark state, the kinase is auto-inhibited. In the signaling state, structural changes in LOV2 (includes the unfolding of one of the helices) release the inhibition and trigger kinase activity [111–115].

Blue Light Using Flavin (BLUF) Proteins are a diverse family of photoreceptors that facilitate light perception and signaling in a variety of organisms [116]. The BLUF domain is a conserved light-sensing module composed of 100–150 amino acids. It is usually linked to an effector domain that determines the function of the protein. Proteins containing this domain act as light-regulated switches, converting photon energy into biochemical or structural signals that influence metabolic pathway function [117, 118]. In the early 21st

century, BLUF domains were identified independently by different research groups in two different organisms: the single-celled flagellate *Euglena gracilis* [119] and the purple bacterium *Rhodobacter sphaeroides* [120]. These discoveries followed research into other flavin-dependent photoreceptors, such as cryptochromes and LOV proteins, which demonstrated that flavin molecules can mediate a variety of light responses in different organisms. Early spectroscopic and mutagenesis studies confirmed that BLUF domains use FAD as a chromophore and exhibit unique, reversible photochemistry that differs from that found in LOV systems and cryptochromes [118, 121]. BLUF sensors regulate various processes in the cells of bacteria, cyanobacteria, algae, and some fungi, including behavioral and metabolic processes that are often related to adapting to changing light conditions. The outcome of BLUF activation depends on the effector domain linked to the BLUF sensor and the organism's ecological environment [12, 118].

Due to their ability to easily bind to a wide range of effector domains, proteins containing the BLUF domain can perform a variety of functions [122, 123]. They may be involved in regulating phototaxis [124] and influencing the circadian rhythm [49]. They can act as light-regulated transcription factors, controlling gene expression in response to blue light [125]. They often act as light-controlled molecular switches, modulating interactions with other proteins [126] and enzymatic reactions [127]. In non-photosynthetic bacteria such as *E. coli*, BLUF photoreceptors contribute to the regulation of biofilm formation and adaptation to stress [128–130]. In photosynthetic bacteria, however, they control processes related to photosynthetic efficiency [131]. Due to their small size, fast response, and reversibility, BLUF domains are increasingly used as optogenetic tools. In biotechnology, they are used to create synthetic light-activated proteins by combining them with various signaling or enzymatic domains [132, 133].

A BLUF protein is a modular flavoprotein that consists of a BLUF sensor domain linked to an effector domain. The BLUF domain has a compact α/β structure consisting of five β-sheets surrounded by α-helices on both sides. The N- and C-terminal regions of the BLUF domain often contribute to structural stability and may facilitate communication between domains. Blue light absorption and signaling in BLUF proteins is carried out by the chromophore FAD. This chromophore is situated in a conserved hydrophobic pocket near the β-sheet and is stabilized by multiple hydrogen bonds and van der Waals interactions. The amino acid residues forming the chromophore-binding pocket in the BLUF domain include the following: Tyr, Gln, Trp, Asn, and Met [117, 118, 123, 124, 131]. Photoactivation causes changes in the network of hydrogen bonds surrounding FAD. Electrons and protons are transferred between the conserved Tyr, Trp, and Gln residues. This results in tautomerization and rotation of the glutamine residue, altering its hydro-

1

gen bond pattern [134–136]. These changes lead to the formation of a light-activated signaling state, which is spectrally separate from the dark state (typically exhibiting a red shift in absorption). This signaling state persists for milliseconds to seconds, after which it returns thermally to the dark state, enabling reversible light detection. Small local changes propagate throughout the protein, leading to global structural changes. The C-terminal cap region typically experiences the greatest movement, transmitting the signal to the effector domain [123, 137–139].

Fluorescent Proteins (FPs) have the ability to emit light after absorbing radiation. As most known proteins can be attached to FP, the FPs can be used as markers to observe biochemical events in living cells. Various organisms can be genetically encoded with green fluorescent protein (GFP). This enables real-time monitoring of metabolic pathways [140].

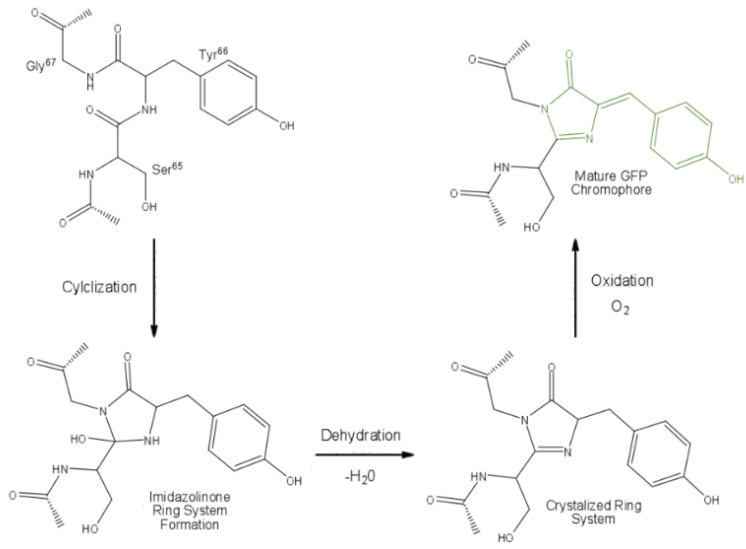


Figure 1.7: Schematic representation of the formation of the GFP fluorophore [51, 141].

The example of this type of macromolecule that is best studied and most widely used is the GFP found in the jellyfish *Aequorea victoria* [44]. GFP has a typical β-barrel structure, consisting of several antiparallel β-sheets formed into a cylindrical shape (see Figure 1.2). It has a fluorophore made up of three amino acids (serine, tyrosine, and glycine), which can undergo cyclization and dehydration to form a system of conjugated double bonds that causes fluorescence, as shown at the Figure 1.7. The fluorophore is excited at 395 nm and 475 nm, with the emitted light having a wavelength of 509 nm [44, 51, 142–145].

Many quantum effects have been investigated in GFP, such as coherent dynamics,

where electronic and nuclear motions synchronize after photoexcitation [146], and photonic entanglement, which occurs during fluorescence emission [147]. Several mutations of GFP have been developed due to their potential for various applications, including color mutants such as blue, cyan, and yellow fluorescent protein. These modified proteins exhibit broad radiation absorption and emission allowing the study of the activity of several different proteins and metabolic pathways simultaneously [51, 148, 149]. Additionally, fluorescent proteins may exhibit unique properties when they contain other chromophores. For example, UnaG is a ligand-activated fluorescent protein that interacts with bilirubin. These features include red-shifted emissions occurring above 600 nm, and photoconversion from green to red emission [150].

Optical Regulation of Non-Photoactive Proteins is an exciting and rapidly developing field within modern photobiology and bioengineering. Non-photoactive proteins are proteins that do not respond directly to light due to the absence of an integral chromophore or light-sensitive domain. Optical regulation involves linking light-responsive components to these proteins in order to control their structure, activity or interactions. Optical regulation can be accomplished through genetic, chemical, or hybrid bioengineering approaches [151–153].

The process of fusion with natural photoactive domains involves attaching a light-sensitive domain to a non-photoactive target protein. When exposed to light, the photoactive domain undergoes a conformational change which is transmitted to the target and modulates its activity [154–157].

Light-controlled protein regulation can be achieved by introducing photoswitches or chromophores (e.g. azobenzene or resveratrol) via covalent bonding to specific amino acid residues in the protein, or by inserting them into native ligand-binding pockets without forming chemical bonds [158–162].

1.3 Computational Methods for Studying Photoactive Proteins

1

Photoactive proteins are crucial for the functioning of living organisms. However, understanding the mechanisms underlying the activation process of light-sensitive proteins remains challenging, as achieving the required temporal and spatial resolution is highly difficult in an experimental setting. Therefore, computational studies are the main source of knowledge about photoreceptor activation mechanisms. The following section provides an overview of the computational research methods employed in the study of biological systems that interact with light.

1.3.1 Bioinformatics

The analysis of protein sequences and the obtaining of valuable information about their structure, evolution, and interactions is made possible by bioinformatics. These tools enable sequence alignment, structure prediction, and evolutionary analysis. They provide key insights into the properties and mechanisms of action of a wide variety of proteins [23–25].

Analyzing the order in which sequence parts are arranged and how they are connected can help to identify regions that are similar across different sequences. This process also enables the identification of key sites, such as where the chromophore is attached. This process can also reveal functional motifs and characteristic domains that are common to other photoactive proteins, highlighting evolutionary links between protein families [163]. Useful tools for comparing protein sequences with known sequences available in databases include BLAST (Basic Local Alignment Search Tool) [164] and Clustal Omega [165]. In bioinformatics, functional annotation is a crucial method. It helps to identify sites of post-translational modification and interaction regions by detecting functional domains. This can be achieved using tools such as InterPro [166] and UniProt [167, 168]. The application of functional annotation to photoactive proteins can reveal the potential impact of mutations in the protein sequence on functionality. Moreover, it makes it possible to predict particular features associated with photoactivation, such as light absorption and signaling [163, 169]. Bioinformatics tools can be used to track the evolution of light-sensitive proteins and identify adaptations that enable them to interact with light [170]. Tools such as MEGA [171] and PhyML [172] are used for evolutionary research. These tools enable phylogenetic analysis and the creation of family trees, facilitating the determination of evolutionary relationships between members of photoactive protein families [173].

In recent years, significant progress has been made in predicting the three-dimensional

structure of proteins based on their amino acid sequence. The importance of these methods is evidenced by the fact that researchers were awarded the 2024 Nobel Prize in Chemistry [174, 175]. This has been made possible by tools such as AlphaFold [176], Rosetta [177] and the SWISS-MODEL server [178, 179]. The primary limitation of these methods is their current inability to predict light-induced structural changes in proteins. However, they do allow the ground state structure to be obtained. In combination with MD methods, this enables the study of conformational changes induced by light absorption, such as chromophore conformational changes and signaling pathways to other parts of the protein. Methods based on ML and artificial intelligence (AI) are continuously being developed to overcome these limitations, with an increasing number of techniques emerging that aim to predict the structures and dynamics of biological systems [180–184].

The full complexity of protein responses to light cannot be captured by a single layer of data (whether that data is from a gene, a transcript, or a protein). Bioinformatic integration brings together multiple 'omic' datasets to provide insight into how photoreceptor-encoding genes are regulated and expressed, how photoactive proteins fold and assemble, how they interact with signaling partners and influence downstream metabolic networks and physiological responses [163].

Kyndt et al. [169] have published an article demonstrating how bioinformatics can be used to analyze photoactive systems. In this study, the authors used BLAST to perform sequence alignment and analysis, comparing protein sequences (e.g., aligning bacterial rhodopsins and PYPs to detect conserved regions and functional residues). They also conducted phylogenetic analysis to infer evolutionary relations between phytochrome-related proteins in different bacterial species. A new family of carotenoid proteins was identified and characterized using bioinformatics tools [185,186]. Furthermore, sequence alignment allowed the construction of a phylogenetic tree, the classification of protein sequences, and the prediction of the functional roles of bacteriophytochrome-like proteins, as demonstrated by Nadella et al. [187]. Hasegawa et al. [188] have also used bioinformatics to show that the genome of cyanobacteria may contain genes that encode rhodopsins, suggesting that these proteins may have evolved in photoautotrophic organisms.

1.3.2 Molecular Dynamics (MD)

MD simulations have become an effective way of complementing experimental results and gaining new insights into complex systems at a microscopic level. MD is fundamentally a computational method used to analyze the physical motion of atoms and molecules, as well as their deformation and interaction, over timescales that are often inaccessible in experiments [189–191].

Table 1.2: The major categories of MD computational approaches used to study biological systems [1].

Approach	Physics Level	Time Scale	Key Application in Photoactive Proteins
Classical MD	Classical	ns–μs	Structure stability and dynamics
Enhanced Sampling	Classical	μs–ms	Rare events, conformational transitions
Quantum Calculations	Quantum	fs–ps	Charge transfer, absorption spectra
Hybrid QM/MM Methods	Quantum + Classical	ps–ns	Light-induced reactions in proteins
Nonadiabatic Dynamics	Quantum + Classical	fs–ps	Ultrafast photochemistry

This section describes the main categories of computational approaches (shown in Table 1.2 used in MD to study biological systems, including photoactive proteins.

Classical MD Simulations can be used to simulate the behavior of a molecular system as a function of time. Integrating a set of differential equations derived from classical dynamics enables the trajectory in phase space (i.e., the space of positions and momenta) to be obtained. MD simulations are based on classical Newtonian mechanics in order to describe the motion of atoms. Each atom is treated as a point mass characterized by its position, velocity, and acceleration [192–194].

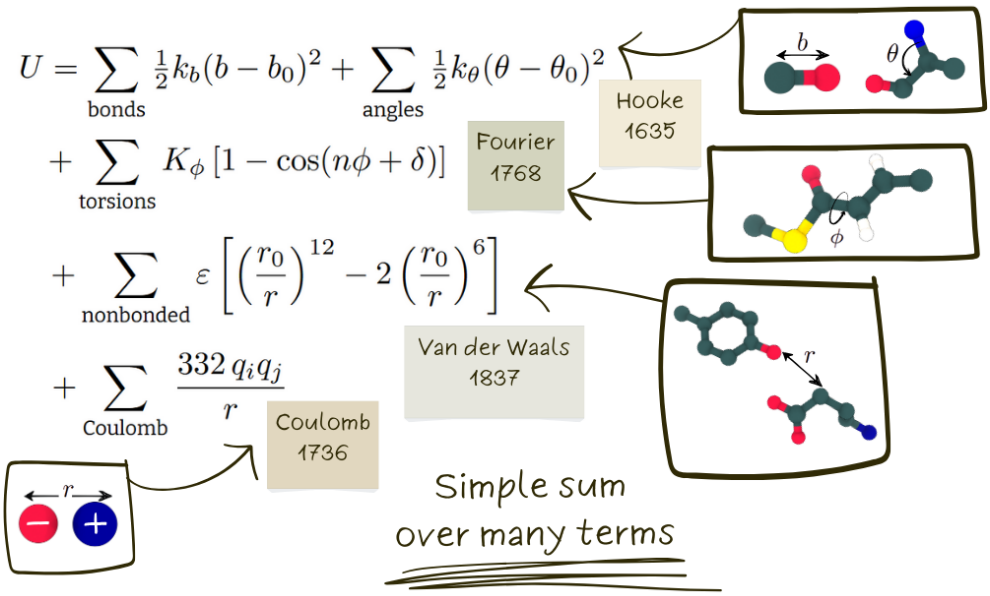


Figure 1.8: The total potential energy U can be expressed as a sum of bonded and nonbonded interaction terms, including harmonic bond stretching, harmonic angle bending, periodic torsional (dihedral) rotations, Lennard–Jones, van der Waals interactions, and Coulombic electrostatic interactions. Molecular cartoons depict the corresponding structural degrees of freedom associated with each term. Based on 2013 Nobel Lecture [195, 196].

In MD simulations, the potential energy function of a system is derived from empirical force fields (FF). In the case of MD of proteins, the term 'force field' is used to describe the mathematical formulas and parameters that are employed to calculate the energy of a protein as a function of its atomic coordinates [197]. The elements of typical FF include various parameters (shown in the Figure 1.8), such as bond strain (a harmonic potential that approximates Hooke's law), angular deflection (a harmonic potential for angles between bonds), dihedral torsion (periodic functions for rotation around bonds) and non-bonded interactions (van der Waals: Lennard-Jones potential and electrostatics: Coulomb's law) [198–201]. Examples of popular force fields include CHARMM [202], AMBER [203], OPLS [204], and GROMOS [205].

Popular software that uses the above-mentioned force fields to study protein stability, conformational dynamics of biological systems [206], and protein-protein or protein-ligand interactions [207, 208] include GROMACS [209], NAMD [210], and AMBER [211].

As analytical solutions are impossible for large systems, numerical integration algorithms such as Verlet, leapfrog and velocity-Verlet advance the system in discrete time steps. These algorithms provide updated positions and velocities, producing a trajectory of atomic motion. The aim is to achieve stability, accuracy and energy conservation with minimal computational cost [212].

- Verlet integration is one of the longest-standing and most widely used schemes in MD. The algorithm updates positions using those from previous time steps [213]:

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \mathbf{a}(t)\Delta t^2 \tag{1.1}$$

  The advantages of this algorithm are its simplicity, reliability and reversibility over time. However, it requires additional reconstruction of the velocities [213].

- In Leapfrog, integration velocities are defined at half-step increments [214]:

$$\mathbf{v}(t + \frac{\Delta t}{2}) = \mathbf{v}(t - \frac{\Delta t}{2}) + \mathbf{a}(t)\Delta t \tag{1.2}$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t + \frac{\Delta t}{2})\Delta t \tag{1.3}$$

  This algorithm provides high numerical accuracy. Velocities 'leapfrog' over positions to ensure optimal energy management. This is a common option in many MD software packages. Interpolation is necessary because velocities are not available at the same time point as positions [214].

- Velocity Verlet integration method combines the advantages of the Verlet method

with the use of explicit velocity updates [213]:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 \tag{1.4}$$

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t) + \frac{1}{2}[\mathbf{a}(t) + \mathbf{a}(t+\Delta t)]\Delta t \tag{1.5}$$

The algorithm provides high levels of accuracy and stability. Velocities and positions are defined within the same time step. It also facilitates coupling with thermostats and barostats [213].

Free energy landscapes illustrate the thermodynamic stability and accessibility of molecular conformations. They are based on the idea that biomolecular function is regulated by transitions between different energy minima. They provide a multidimensional representation of folding pathways, binding/unbinding transitions, and conformational changes (escaping metastable states) [215].

The whole procedure is more complicated in terms of photoactive systems, which, in addition to the protein component modeled via force fields, also include non-standard chromophore molecules requiring additional parametrization. This involves additional quantum chemical calculations to provide essential information about the molecule, such as bond lengths and angles, dihedral angles, and charge, as well as the connection between the chromophore molecule and the protein component [216–219].

Enhanced Sampling Techniques are used in situations where classical MD exhibits performance issues related to the efficient sampling of conformational space. This problem is caused by rough energy landscapes with many local minima that are separated by energy barriers higher than the thermal energy. These limitations may result in an inaccurate representation of the system's conformational states, which can hinder the analysis and understanding of its functional properties. In order to correctly characterize the dynamics and function of a system, MD simulations must cover all of its relevant states [27–29]. Such challenges are often overcome using enhanced sampling techniques, such as replica exchange [220], umbrella sampling [221], and metadynamics [222]. These methods are implemented in the PLUMED library, which can interface with many MD software packages [223–225].

The conformational space of biomolecular systems can be efficiently explored using a classical enhanced sampling method known as replica exchange molecular dynamics (REMD). This is achieved by performing a series of simulations (replicas) of the system under different conditions, periodically exchanging configurations between them. This helps

to avoid local minima and sample a broad area of the energy landscape [226, 227]. For instance, REMD has been used to efficiently sample PYP conformations in work by Vreede et al. [228]. Moreover, REMD simulations of the LOV protein successfully reproduced essential structural changes associated with the light activation process observed experimentally [229].

Umbrella sampling (US) is the first enhanced sampling method to apply an additional bias potential to enhance the fluctuations of significant degrees of freedom [221]. In the work of Bondanza et al. [230], the US-REMD approach was used to investigate the molecular mechanisms of activation in orange carotenoid proteins. Furthermore, Nottoli et al. [231] have combined a polarizable quantum mechanical approach with US to demonstrate excited state intramolecular proton transfer in solvated 3-hydroxyflavone.

To accelerate the sampling of high-energy or poorly visited states, time-dependent bias potentials along a function of a few selected degrees of freedom, called collective variables (CVs), are used [232–234]. CVs are selected parameters (e.g., torsion angles, distances, or coordinates) that capture conformational transitions particularly relevant for the system dynamics. In metadynamics, the bias potential is constructed by accumulating Gaussians in CV space to improve the sampling and unravel the free-energy landscape (schematic representation shown in Figure 1.9).

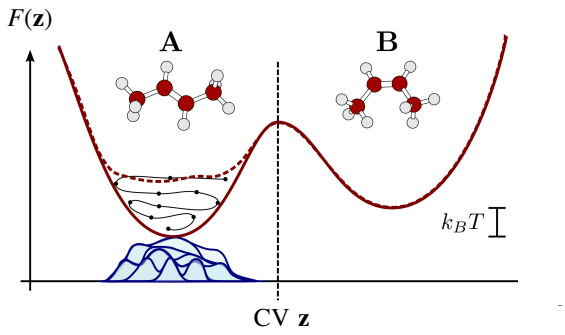


Figure 1.9: Schematic representation of the conformational flooding scheme for computing the kinetics for the transition from state A to state B [234]. Modified from `https://atomistic-cookbook.org/`.

CVs $\mathbf{z}$ are given by a set of d functions of the system coordinates $\mathbf{R}$:

$$\mathbf{z}(\mathbf{R}) = (\mathbf{z}_1(\mathbf{R}), \mathbf{z}_2(\mathbf{R}), ..., \mathbf{z}_d(\mathbf{R})). \tag{1.6}$$

The bias potential at time t can be given by the following formula (based on the Alessan-

dro Barducci paper [232]):

$$V_G(\mathbf{z},t) = \int_0^t dt'\, \omega \exp\left(-\sum_{i=1}^{d} \frac{(\mathbf{z}_i(\mathbf{R}) - \mathbf{z}_i(\mathbf{R}(t')))^2}{2\sigma_i^2}\right), \tag{1.7}$$

where ω is a rate of energy (expressed as Gaussian height W divided by a deposition stride τ_G) and σ_i is the width of the Gaussian for the ith CV.

Gaussians effectively 'fill in' local minima in the free energy landscape, making it easier for the system to escape and explore new regions as illustrated in Figure 1.9. As the bias accumulates, the system is pushed away from previously visited states, allowing it to overcome energy barriers.

After the biased passage of the system through various minima and energy barriers in the free energy landscape, the free energy surface of the system must be reconstructed. The bias potential is used as an approximation of the free energy of the system [232]:

$$V_G(\mathbf{z}, t \to \infty) = -F(\mathbf{z}) + C, \tag{1.8}$$

where C is an additive constant and the free energy $F(\mathbf{z})$ can be defined as following [232]:

$$F(\mathbf{z}) = -\frac{1}{\beta} \ln\left(\int d\mathbf{R}\, \delta(\mathbf{z} - \mathbf{z}(\mathbf{R}))e^{-\beta U(\mathbf{R})}\right), \tag{1.9}$$

where $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant, T is the temperature of the system and $U(\mathbf{R})$ is the potential energy function.

This methodology was employed to investigate the photoregulation mechanism of orange carotenoid protein (OCP), which is linked to the carotenoid chromophore. This study examined the conformational changes in OCP that facilitate carotenoid translocation and interaction with the light-harvesting antenna [235]. Metadynamics has also been used to design the molecular structures of photoswitches by efficiently exploring their conformational space, as demonstrated by Raucci et al. [236].

Another enhanced sampling technique used for the study of photoactive proteins is variational enhanced sampling (VES) [237]. This technique has been employed to study the mechanism of phytochrome photoisomerization, with a particular focus on the biliverdin IXα (BV)-bacteriophytochrome system and signal transduction pathways, as well as thermal transitions [59]. This has been described in detail in Chapter 2.

Quantum Chemical Calculations significantly improve our understanding of photoactive systems by providing detailed insights into the electronic structure and photochemical

properties of their chromophores [17]. Characterizing light absorption, excited-state dynamics, and photoinduced reactions at the atomic level is possible with methods such as time-dependent density functional theory (TD-DFT) [238,239] and multi-configuration approaches (e.g., CASSCF) [240]. Quantum calculations are crucial for understanding the interactions between chromophores and their protein environments, including charge transfer and polarization effects. These methods also support classical techniques by providing a mechanistic understanding of processes such as isomerization, fluorescence, and energy transfer in photoactive proteins [241,242].

TD-DFT is a method that can be used to simulate how the electronic density of a system responds to external factors, such as light absorption. Valuable insights into absorption spectra and excited-state properties are offered by its accurate description of electron transitions between ground and excited states [243,244]. The primary use of TD-DFT is in the analysis of both absorption spectra and processes such as fluorescence and energy transfer. Understanding how protein environments affect the absorption spectrum is also made easier by predicting shifts in the absorption maximum. The reliability of the application of TD-DFT is dependent on the appropriate choice of detailed methodology. In particular, DFT functionals should be benchmarked on molecules similar to the chromophore [245].

Another approach, complete active space self-consistent field (CASSCF), is a wave function-based method. It considers part of the electron correlation within a defined active space, such as the static correlation energy, which goes beyond the single Slater determinant approximation [246]. The active space for photoactive systems encompasses the crucial orbitals and electrons that participate in photochemical reactions, such as the π and π^* orbitals of the chromophore. The CASSCF wave function is obtained by initially dividing the space of molecular orbitals into three separate categories: inactive orbitals, active orbitals, and secondary orbitals. The active orbitals can be occupied in any way, as long as the overall spin and spatial symmetry of the wave function are maintained [247,248]. When optimizing the wave function, it is assumed that inactive orbitals are doubly occupied and that secondary orbitals are empty. Post-Hartree-Fock methods that include both static and dynamic electron correlation effects, such as CASSCF/CASPT2 and coupled cluster (CC) methods, are frequently applied to study photoactive phenomena [249–252]. It is particularly important to include dynamic correlation effects if accurate vertical excitation energies are to be correctly predicted for a chromophore embedded in a protein environment. Methods for accurately modeling excited states in small molecules, such as CC2 [253], and those related to the second-order algebraic diagrammatic construction ADC(2) [254], are impractical for studying complex systems,

including large molecules of chromophores or proteins [255].

1

Due to the computational cost of quantum calculations for complex systems, the study has mainly focused on chromophores that can undergo photochemical reactions, such as biliverdin [256–258], p-coumaric acid [259–261], and retinal [262, 263].

Hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) Methods of simulations are an essential tool for understanding photoreceptor proteins, since it is not possible to study such large systems effectively using quantum calculations as the only method. QM/MM enables the modeling of chromophores and their immediate surroundings. This is achieved by applying more expensive quantum methods to describe the chemically active parts of chromophores and their local environments. Classical MD methods can be used to study the rest of the system, which is the protein macromolecule [31, 195, 264, 265]. In order to apply such hybrid methods, the system under consideration must be divided into two main regions: the QM region, which contains the atoms that are directly involved in the molecular process (e.g. the chromophore and the neighboring amino acids that stabilize the chromophore's position and form critical hydrogen bonds), and the MM region, which covers the environment surrounding of the QM region (e.g. the remaining amino acids that comprise the protein and the solvent). The QM part is treated using quantum mechanical techniques (e.g., DFT and post-Hartree-Fock), while the MM part is calculated using classical molecular mechanics techniques based on predefined FF. The software for QM/MM simulations is available in various packages, such as ORCA [266], Gaussian [267], or ChemShell [268]. Programs such as AMBER-Gaussian [269], GROMACS-CP2K [270, 271], AMBER/NAMD/GROMACS-TeraChem [272, 273] provide interfaces between classical and quantum calculations.

The interface between the quantum and molecular regions is the primary challenge when implementing these methods [248, 274]. Interaction between the two mentioned regions is usually direct, so the total energy of the system cannot be defined as the sum of the contributions of its component energies [275, 276]. The coupling between the QM and MM regions determines the extent to which the two subsystems interact and contribute to the total energy of the hybrid system. This coupling can involve non-bonded interactions (such as electrostatic and van der Waals) and bonded interactions (such as covalent bonds crossing the QM-MM boundary). The total energy of the QM/MM system as a whole is expressed as [271, 277]:

$$E_{\text{total}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}}, \tag{1.10}$$

where E_{QM} is the energy of the QM region, calculated using a quantum mechanical method, E_{MM} is the energy of the MM region, calculated using a classical force field, and $E_{\text{QM/MM}}$ is

the interaction energy between the QM and MM regions [271,278]. Several types of QM/MM coupling schemes can be distinguished, such as mechanical, electrostatic, and polarizable embedding [248,279].

In QM/MM systems, particularly those involving covalent bonds crossing the interface, the boundary between the QM and MM regions is crucial. Failing to handle this boundary properly can lead to nonphysical results. Non-covalent boundaries are usually treated using the electrostatic embedding coupling scheme, while van der Waals interactions are handled classically using the MM force field. However, the treatment of boundaries passing through a covalent bond is more complicated and can be achieved using several strategies, such as hydrogen capping, frozen orbitals, localized molecular orbitals, and effective core potentials [280–283].

QM/MM methods can be used to study light-induced processes in biomolecules. This is an area of research that has grown in popularity in recent years [284–288]. The use of QM/MM has been particularly widespread in the study of PYP, covering aspects such as protonation, hydrogen bonds, and the absorption spectrum [273,289–292]. Hybrid approach has also been used to study other photoactive systems, such as channelrhodopsin [293], rhodopsin [294–296], retinal [297,298], and biliverdin [258,299–302].

Nonadiabatic Dynamics (NAMD) methods allow the study of light-induced processes, including excited-state dynamics, internal conversion, and photochemical reactions [303, 304]. Photoactive proteins convert light energy into structural changes by exciting embedded chromophores. Photochemical events, especially those near conical intersections (CIs), involve strong coupling between electronic and nuclear motions, resulting in nonadiabatic transitions. When a photon is absorbed, the system transitions from the ground electronic state (S_0) to one or more excited states (S_1, S_2, ...). Relaxation then occurs via nonadiabatic transitions, i.e. transitions between electronic states that are coupled by nuclear motion and electronic–nuclear interactions. However, the Born–Oppenheimer molecular dynamics approach, which assumes that electrons adjust instantly to nuclear motion, is not applicable in such cases because the electronic and nuclear degrees of freedom are strongly coupled [305–307]. The goal of nonadiabatic dynamics is to simulate the time evolution of the nuclei and electrons simultaneously. This involves tracking the population transfer between the electronic states and the structural relaxation that leads to photochemical outcomes [308,309].

The mathematical framework of NAMD is based on solving the time-dependent Schrödinger equation (TDSE). This is usually achieved by using the Born–Oppenheimer approximation and other methods for modeling complex molecular systems [307].

1

As mentioned, NAMD's description begins with TDSE:

$$\hbar \frac{\partial \Psi}{\partial t} = \hat{H}(\mathbf{r}, \mathbf{R}) \Psi(\mathbf{r}, \mathbf{R}, t), \tag{1.11}$$

where: $\hbar$ is the reduced Planck constant, $\hat{H}(r, \mathbf{R})$ is Hamiltonian operator describing the system's total energy, and $\Psi(\mathbf{r}, \mathbf{R}, t)$ is the total wavefunction (electronic $\mathbf{r}$ and nuclear $\mathbf{R}$ coordinates) [307, 310].

In terms of Born–Oppenheimer Expansion the total wavefunction $\Psi(\mathbf{r}, \mathbf{R}, t)$ is expanded in terms of a basis set, which involves electronic and nuclear wavefunctions in either the adiabatic (ψ_a) or the diabatic (ϕ_a) representation [307, 310]:

$$\Psi(\mathbf{r}, \mathbf{R}, t) = \sum_a \psi_a(\mathbf{r}, \mathbf{R}) \chi_a(\mathbf{R}, t) = \sum_a \phi_a(r) \chi'_a(\mathbf{R}, t), \tag{1.12}$$

where: $\psi_a(\mathbf{r}, \mathbf{R})$ is the electronic wavefunction in the adiabatic representation, $\phi_a(\mathbf{r})$ is the electronic wavefunction in the diabatic representation, $\chi_a(\mathbf{R}, t)$, and $\chi'_a(\mathbf{R}, t)$ are the corresponding nuclear wavefunctions [307, 310].

The adiabatic electronic wavefunctions $\psi_a(\mathbf{r}, \mathbf{R})$ and their corresponding eigenvalues, $E_a(\mathbf{R})$ (the Potential Energy Surfaces, PESs), are obtained by solving the static electronic Schrödinger equation. Using this expansion in the TDSE provides coupled differential equations for the nuclear wavefunctions $\chi_a(\mathbf{R}, t)$ [311].

The coupling between different electronic states is mathematically described by the non-adiabatic coupling vector $d_{ab}(\mathbf{R})$ (NACR).

$$d_{ab}(\mathbf{R}) = \langle \psi_a(\mathbf{r}, \mathbf{R}) \,|\, \nabla_{\mathbf{R}} \psi_b(\mathbf{r}, \mathbf{R}) \rangle_r. \tag{1.13}$$

This coupling is the result of the nuclear kinetic energy operator acting on the electronic wavefunctions. When the adiabatic energy gap $|E_b(\mathbf{R}) - E_a(\mathbf{R})|$ becomes marginal (near avoided crossings or conical intersections), the NACR terms couple the PESs, indicating a breakdown of the Born–Oppenheimer approximation. The NACR can be explained using the Hellmann–Feynman theorem [312, 313].

As the exact solution to the nuclear TDSE is computationally challenging for large systems due to its exponential scaling with dimension, NAMD relies on approximations, primarily mixed quantum-classical methods. This methodology employs a classical approach to the treatment of nuclei, while electrons are handled within the framework of quantum mechanics. The two most popular mixed quantum-classical methods are: Ehrenfest dynamics [314, 315] and fewest switches surface hopping (FSSH) [316, 317].

1.3.3 Machine Learning Base Methods

ML have become a revolutionary approach in the computational study of photoactive proteins (see Table 1.1 and Figure 1.2), which undergo complex photoinduced processes such as electronic excitation, isomerization and proton or electron transfer. In order to model these events with high accuracy, it is essential to calculate electronic and nuclear degrees of freedom. This is usually achieved through QM and MD simulations. However, such approaches remain computationally demanding for large protein–chromophore systems and for capturing the broad range of relevant timescales. To overcome these limitations, ML offers a data-driven approach that complements and enhances traditional simulation techniques [318, 319].

ML models can learn potential energy surfaces (PESs) and associated forces directly from QM reference data [318, 320]. They can accelerate molecular simulations by replacing or augmenting costly electronic-structure calculations [321], predict photophysical properties such as excitation energies, fluorescence lifetimes and spectral shifts with near-quantum accuracy [322,323], and automate the analysis of large-scale simulations to reveal structure–function relationships [324, 325]. By combining statistical learning with theoretical principles, ML-based approaches offer an effective fusion of computational efficiency and physical accuracy. This synergy is transforming the field of computational photobiology by enabling the investigation of photoactive systems on a scale that was previously beyond the scope of conventional theoretical frameworks [318, 319, 326].

ML is a broad field of computer science that includes a variety of algorithms which enable computers to extract patterns and make predictions from data. In photoactive protein modeling, ML approaches can be categorized into three main types: supervised, unsupervised and reinforcement learning. Each type offers unique advantages for analyzing structure–function relationships and light-driven dynamics [327–329]. Supervised learning involves training a model using labeled datasets, where the desired output is known for each input. The algorithm learns the relationship between molecular descriptors and target properties by minimizing prediction errors in the training data [330–332]. Unsupervised learning aims to identify hidden patterns, clusters or reduced-dimensional representations within unlabeled data. As these methods do not require explicit target outputs, they are especially effective at discovering novel features in high-dimensional molecular simulations [328, 333]. Reinforcement learning is a concept that describes how an algorithm learns to make choices through interaction with an environment. Rather than relying on labeled data, the model receives feedback in the form of rewards or penalties, optimizing strategies that achieve long-term objectives [334–337].

A variety of ML frameworks have been developed and adapted for the study of different

types of protein. These frameworks differ in terms of their learning objectives, data requirements and integration with physical modeling approaches [338]. Neural network potentials (NNPs) are one of the most commonly implemented ML approaches for accurately reproducing potential energy surfaces (PESs). When trained on datasets obtained from ab initio calculations, NNPs learn the mapping between atomic configurations and their associated energies or forces [339–341]. Frameworks such as Behler–Parrinello neural networks [342], and ANI [343] have demonstrated great versatility across different chromophore environments. Graph neural networks (GNNs) treat molecules and proteins as networks of nodes and edges, where the nodes represent atoms and the edges represent bonds or spatial relationships. This representation intuitively captures both chemical topology and the impact of the local environment [344]. Kernel-Based and Gaussian Process Models offer a mathematically transparent alternative to deep neural networks. These models interpolate between known quantum chemical data points to predict energies, forces, and spectroscopic observables while also providing uncertainty estimates [345, 346]. Generative models have become increasingly popular in inverse molecular design. By learning the relationships between sequences, structures and properties, these models can suggest new chromophores, predict the effects of mutations, and design synthetic variants with specific photoresponses. This can support the development of new optogenetic tools and photoactive drug compounds [347–349].

2

Bacteriophytochrome and biliverdin IXα

Chapter based on:

Authors: Jakub Rydzewski, Katarzyna Walczewska-Szewc, **Sylwia Czach**, Wiesław Nowak, Krzysztof Kuczera

Original title: Enhancing the inhomogeneous photodynamics of canonical bacteriophytochrome

Published in: The Journal of Physical Chemistry B, 126 (14), 2647-2657, 2022

https://doi.org/10.1021/acs.jpcb.2c00131

2.1 Introduction

Bacteriophytochromes (BphPs) belong to the phytochromes family (Table 1.1). BphPs have been found in both photosynthetic and non-photosynthetic bacteria. This indicates that their function is not restricted to photosynthesis, but also involves general environmental light sensing. They are present in purple phototrophic bacteria, cyanobacteria, and green sulfur bacteria [350–352]. Their function is to optimize photosynthetic and metabolic processes by perceiving light. They also help with chromatic acclimation, which involves adjusting the light-harvesting complexes in response to environmental light conditions. They can often regulate light-dependent behavioral or pathogenic responses, such as adhesion to surfaces, migration, colony formation, and gene expression [353–356]. BphPs are light-sensitive proteins that act as a molecular switch. They transition between two forms: the dark state (Pr, which absorbs red light at around 660 nm) and the illuminated state (Pfr, which absorbs far-red light at around 700 nm). They have covalently bound a linear tetrapyrrole chromophore, which is typically biliverdin IXα (BV) [47, 58, 59].

2.1.1 Structure of Bacteriophytochromes

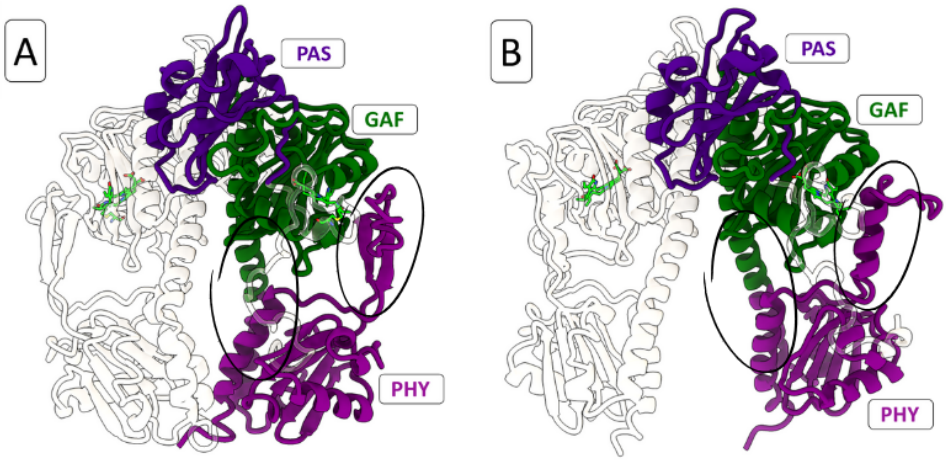


Figure 2.1: Structure of bacteriophytochrome A) in the Pr state, B) in the Pfr state with PAS domains in indigo, GAF in dark green, and PHY in purple, biliverdin in stick representation colored according to atom type with carbons in lime. The sites of the greatest conformational changes in both structures are marked with circles.

Bacteriophytochromes (BphPs) are sensory proteins that exist as dimers, composed of a highly conserved photosensory core and variable output domains. Their structure allows

them to detect red and far-red light via a biliverdin chromophore, and to convert this information into biological responses through structural changes and enzymatic activity. The Photosensory Core Module (PCM) comprises three highly conserved domains arranged in a linear order: PAS–GAF–PHY (as shown in Figure 2.1) [351, 357–359].

The PAS Domain is present in all living organisms [360]. The PAS domain name refers to the three proteins in which the domain was first discovered: Per (period circadian protein), Arnt (aryl hydrocarbon receptor nuclear translocator protein), and Sim (single-minded protein) [19]. The core architecture of the PAS domain was first characterized in the structure of the photoactive yellow protein (PYP) in 1998 (see Chapter 3 for more details). The PAS domain functions as a molecular sensor, binding small compounds and other proteins. In BphP, it stabilizes the orientation of the chromophore-binding domain (GAF) by acting as a structural scaffold (Figure 2.1 illustrates the spatial orientation of both domains).

The GAF Domain, the same as in the case of the PAS domain, it was named after the proteins it is found in: cyclic guanosine monophosphate (cGMP), phosphodiesterase–Adenylate, and cyclase–FhlA [361]. The GAF domain is a highly conserved site among all phytochromes. It plays a crucial role in determining the spectral properties and kinetics of photoswitching. It is the location of the primary chromophore binding site, where biliverdin IXα is covalently linked with a conserved cysteine residue (the Chromophore Binding Domain (CBD) as illustrated in Figure 2.3. The pocket formed by the amino acids stabilizes the chromophore and plays a role in photoisomerization [362–365].

The PHY Domain is the phytochrome-specific part of the protein. It contains a flexible 'tongue' region that undergoes a major conformational change during photoconversion. This improves the efficiency of the photocycle's reversibility and thermal reversion [366–368].

The Output Domains which are located C-terminal to the PCM, exhibit species-specific variation and define the functional role of BphP in cellular signaling. Examples of common domains include histidine kinases [369,370], diguanylate cyclases [367], guanylate cyclases [371], phosphatase domains [366], and DNA-binding transcription factors [372,373]. The ability of phytochromes to combine different domains enables them to perform various functions in response to light.

2.1.2 Chromophore - Biliverdin IXα

Biliverdin IXα ($C_{33}H_{34}N_4O_6$, 582.65 g/mol) (BV) is the main light-absorbing chromophore found in most bacteriophytochromes. It plays a key role in the perception of red and far-red light, converting photon energy into a structural change that initiates signaling path-

ways [59,367]. It consists of four pyrrole rings (A–D) in an open-chain conformation, which are connected by methine bridges (illustrated in Figure 2.2). Of the four possible isomers (IXα, IXβ, IXγ, and IXδ), only IXα is biologically relevant in phytochromes. Biliverdin IXα is formed as a result of enzymatic heme cleavage by heme oxygenase (HO). The IXα isomer is formed by cutting the α-methine bridge between pyrrole rings A and D, as shown in Figure 2.2 [352,374,375].

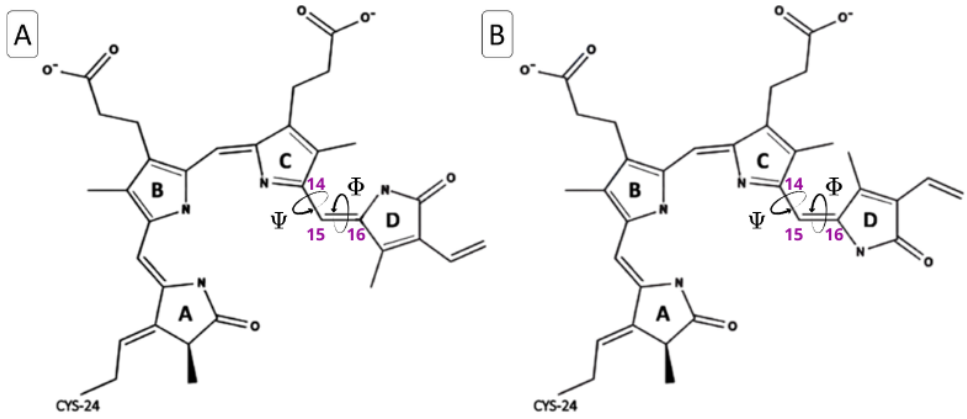


Figure 2.2: The structure of biliverdin A) in the Pr state in the *ZZZ* form, B) in the Pfr state in the *ZZE* form. The angles (Ψ and Φ) around which rotations resulting from photoisomerization occur are marked with arrows on the structures.

BV's strong light-absorbing characteristics are a result of extensive π-electron delocalization along the methine bridges. The action of BphP is based on the unique properties of BV related to its reaction to light. BV exists in a ground state (Pr) that absorbs radiation in the red spectrum (~660 nm). Upon absorbing red light, BV undergoes *ZZZ* → *ZZE* (*trans* → *cis*) isomerization in the C15=C16 double bond (between rings C and D). This results in D-ring rotation and the formation of an illuminated state (Pfr), which absorbs radiation in the far red range (~730 nm). The absorption peaks differ depending on the protein environment in which the chromophore is located, especially the GAF domains, as well as on specific interactions with amino acid residues present in the chromophore binding pocket. Over time, the light-induced Pfr state can revert to the Pr state through a process of thermal reversion (the kinetics of which depend on the protein environment). This reversibility enables cells to detect and respond to light dynamically [59,351,352,367,374].

2.1.3 Chromophore-Protein Interaction

The function of phytochromes is determined by the interaction between the BV chromophore and the amino acids surrounding the CBD. This determines the chromophore's position, photoconversion efficiency, and signal transmission to the output domains [366].

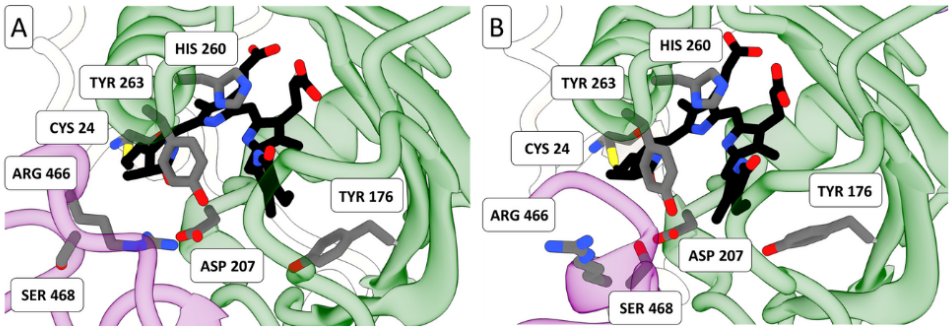


Figure 2.3: Clope-up view on chromophore binding pocket A) of BphP in the Pr state with biliverdin in the *ZZZ* form, B) of BphP in the Pfr state with biliverdin in the *ZZE* form. The semi-transparent protein structure is colored according to domains, while the structures of amino acids forming the CBD are colored according to atom type, amino acid carbons in gray, and biliverdin carbons in black.

In the GAF domain of BphP proteins, BV is located in the CBD and is bound via a thioether bond between ring A of BV and the thiol group of conserved cysteine Cys24. This can be seen in Figure 2.3 [59]. Additionally, the chromophore is stabilized by a network of hydrogen bonds and electrostatic interactions. Stabilizing hydrogen bonds typically form between the side chains of histidine, aspartate, tyrosine, and serine, and the side chains of propionate (rings A and D), pyrrole nitrogen atoms, and methine bridge atoms. Other stabilizing interactions are also identified, including $\pi - \pi$ stacking and hydrophobic interactions. Aromatic residues (e.g., phenylalanine, tyrosine, and tryptophan) interact with the flat arrangement of tetrapyrrole rings via $\pi - \pi$ interactions [351, 358, 362, 376–378].

2.1.4 Photocycle and Mechanism of Signal Transduction

The ground state of the protein before radiation absorption is the Pr state with a covalently bound chromophore in the *ZZZ* conformation. This complex is sensitive to radiation in the red light range (~660 nm). Following the absorption of radiation by BV, photoisomerization occurs, changing the conformation from *ZZZ* in the Pr state to *ZZE* in the Pfr state. The photoconversion of BV from *Z* to *E* takes place around the double bond between C15 and C16, which results in the rotation of the D [301, 367]. Additionally, there are indications that a single bond between C14 and C15 may be able to rotate [379]. Conformational changes in

the chromophore induced by exposure to red light induce a rearrangement of the hydrogen bond network pattern in CBD and changes in the position of the remaining amino acids of the GAF domain [59, 350, 357]. The signal of conformational changes then propagates through the helical spine according to Isaksson et al., 2021 [380] and through the tongue region according to Takala et al., 2014 [47]. The chromophore undergoes a conformational change upon photon absorption, resulting in significant alterations to the protein's secondary structure. Specifically, the secondary structure of the PHY 'tongue' region shifts from a β-sheet (Pr) to an α-helix (Pfr). This changes the orientation of the PHY domain in relation to the PAS–GAF region [47, 59, 380]. Changes caused by the photoisomerization of BV result in the formation of Pfr states that are sensitive to far-red radiation (~700 nm). The reversible transition between phytochrome forms is essential for the proper functioning of the protein and the occurrence of metabolic processes. After absorbing the far-red photon, a transition from the Pfr state to the Pr state occurs. In the absence of far-red light, Pfr can thermally revert to Pr over time. Protein conformational switches, triggered by signal transmission, modulate biological processes such as gene expression, enzyme activity, and motility [58, 351, 381]. A diagram in Figure 2.4 illustrates the described photo cycle.

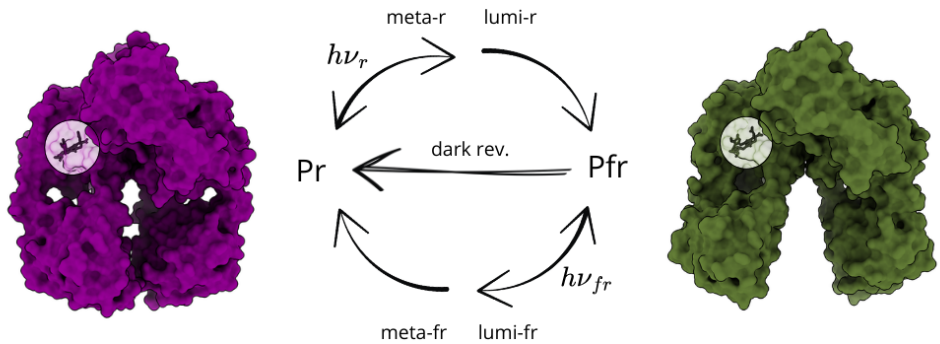


Figure 2.4: The reversible photocycle of phytochrome. Illuminating Pr phytochrome with red light produces the primary photoproducts: lumi-r and meta-r. These are subsequently converted to Pfr. Pfr can be converted into Pr by illumination with far-red light to produce lumi-fr and meta-fr, which then convert to Pr, or via a subsequent thermal process known as dark thermal reversion, which is known to occur in canonical phytochromes.

2.1.5 Purpose and Scope of the Chapter

This chapter aims to investigate the early stages of mechanical signal transduction in the canonical BphP-BV complex and the metastable states of the photoproducts Pr and Pfr. MD simulations with enhanced sampling were conducted to overcome the high free energy (FE) barrier of the initial photoisomerization event. This allowed the non-equilibrium configurations visited by the BphP-BV complex in the Pr and Pfr conformations to be sampled

effectively. The simulations focused on the Pr and Pfr conformers, beginning with their equilibrium configurations, i.e., the state of Pr before photoisomerization and the state of Pfr after the initial photochemical event.

This chapter presents evidence that the Pr and Pfr conformers have notably different free energy (FE) landscapes. Metastable states within these landscapes allow for temperature-dependent transitions associated with dark thermal reversion. These transitions proceed via an intermediate state (Pfr), which is caused by the rotation of Arg466, His467, and Ser468 from the tongue region.

This study also highlights the key thermodynamic role of the protein environment in the formation and expansion of important metastable states in Pfr.

Additionally, it was observed that enhanced local fluctuations of the BV chromophore contribute to nanoscale conformational movements in BphP, consistent with two experimentally determined signal transduction pathways (Takala et al., 2014 [47] and Isaksson et al., 2021 [380]).

2.2 Materials and methods

2.2.1 Models

Each MD simulation was performed using the GROMACS 2018.1 code [209] with the PLUMED 2.8 plug-in [224, 382]. X-ray structures of Pr and Pfr were used as the starting models of BphP-BV complexes (Figure 2.1): Pr (PDB 4O0P [47], dark state; ~660 nm red light absorption) and Pfr (PDB 4O01 [47], illuminated state; ~700 nm far-red light absorption). The models consist of PAS-GAF-PHY domains (linking the PAS-GAF to the PHY domain via two ways: a long α-helical spine and the so-called tongue region [47, 380]). They also have a chromophore covalently attached to CYS24 in the form of a biliverdin IXα molecule (the mentioned structures shown in the Figure 2.2).

The GROMACS topology for BV originates from Modi et al. work [301], where the authors performed *ab initio* quantum chemical calculations to parameterize BV in the *ZZZ* and *ZZE* states. The chromophore BV is modeled with protonation of all nitrogen atoms, as suggested in the reference work [301]. The complexes were solvated with TIP3P water molecules model [383] in a box with water margin of the 0.7 nm along all axes. The systems were neutralized with a concentration of Na^+ and Cl^- ions of 0.15 M. The simulations were performed using the AMBER03 force field [384]. In each simulation, an integration time step of 2 fs was used. Periodic boundary conditions were applied. LINCS [385] was used to constrain hydrogen bonds, including those between heavy and hydrogen atoms in the protein and chromophore, as well as O–H and artificial H–H bonds in the rigid TIP3P water model. Long-range electrostatic interactions are handled using the particle-mesh Ewald method [386] with cutoffs of 1.2 nm for both electrostatic and nonbonded van der Waals interactions. The complexes were minimized for 5000 steps until convergence was achieved. Then, the systems were equilibrated using a velocity rescaling thermostat [387] at 300 K with a relaxation time of 1 ps to simulate the canonical ensemble (*NVT*). Next, 200 ns of unbiased MD simulations were performed for each protein-chromophore system to check the level of thermal fluctuation in the equilibrium metastable state.

2.2.2 Enhanced Sampling Simulations

Enhanced MD simulations were performed using variationally enhanced sampling [237] (VES), which is part of a module VES implemented in PLUMED 2.8 [224, 382].

Variationally Enhanced Sampling (VES)

In the canonical ensemble, the microscopic coordinates $\mathbf{R}$ are sampled from the Boltzmann distribution $P(\mathbf{R}) \propto \mathrm{e}^{-\beta U(\mathbf{R})}$, where β is the inverse temperature and $U(\mathbf{R})$ is the potential energy function. Since molecular systems are multidimensional ($3N$, where N is the number of atoms), a small set of collective variables (CVs), $\mathbf{z}(\mathbf{R})$, has been identified to reduce their dimensionality. The marginal equilibrium of the distribution of $\mathbf{z}$ can to be expressed as follows:

$$P(\mathbf{z}) = \int d\mathbf{R}\delta[\mathbf{z} - \mathbf{z}(\mathbf{R})]P(\mathbf{R}) = \langle\delta[\mathbf{z} - \mathbf{z}(\mathbf{R})]\rangle_U, \tag{2.1}$$

where $\delta[\mathbf{z} - \mathbf{z}(\mathbf{R})]$ is the Dirac delta function and $\langle\delta[\mathbf{z} - \mathbf{z}(\mathbf{R})]\rangle_U$ is the ensemble average under the potential $U(\mathbf{R})$.

In systems with sampling issues, (i.e., systems characterized by the presence of metastable states), the free energy (FE) landscape, $F(\mathbf{z}) = -\frac{1}{\beta}\log P(\mathbf{z})$, consists of many metastable states separated by energy barriers higher than thermal energy. The system remains kinetically trapped and unable to explore the CV space on a timescale that can be simulated. In such cases, it is extremely difficult to achieve convergence using standard MD methods.

Enhanced sampling methods based on CV can be used to alleviate the sampling problem. These approaches introduce a bias potential $V(\mathbf{z})$ that acts in the $\mathbf{z}$ space. This changes the sampling probability from $P(\mathbf{z})$ to the following:

$$P_V(\mathbf{z}) = \Big\langle \delta[\mathbf{z} - \mathbf{z}(\mathbf{R})] \Big\rangle_{U+V}. \tag{2.2}$$

Therefore, this approach is referred to as non-Boltzmann sampling.

In the VES method developed by Valsson and Parrinello [237], a functional dependent on the bias embedded in the CV space $\mathbf{z}$, is optimized:

$$\Omega[V] = \frac{1}{\beta} log\left(\frac{\int d\mathbf{z}\,\mathrm{e}^{-\beta[F(\mathbf{z})+V(\mathbf{z})]}}{\int d\mathbf{z}\,\mathrm{e}^{-\beta F(\mathbf{z})}}\right) + \int d\mathbf{z} p_T(\mathbf{z})V(\mathbf{z}), \tag{2.3}$$

where $p_T(\mathbf{z})$ is the target distribution to which the sampling is biased. Equation 2.3 shows that, up to an immaterial constant, F can be given as follows:

$$F(\mathbf{z}) = -V(\mathbf{z}) - \frac{1}{\beta}\log p_T(\mathbf{z}), \tag{2.4}$$

when V is such that $\Omega[V]$ is stationary. If we assume that p_T is uniform, we simply obtain $F = V$. In this case, finding $F(\mathbf{z})$ is possible by minimizing $\Omega[V]$. Another option to calcu-

late $F(\mathbf{z})$ is the standard umbrella-sampling reweighting [221]:

$$F(\mathbf{z}) = -\frac{1}{\beta}\log\left\langle \mathrm{e}^{\beta V(\mathbf{z})}\,\delta[\mathbf{z}-\mathbf{z}(\mathbf{R})]\right\rangle, \tag{2.5}$$

where the statistical weights associated with $\mathbf{z}$ are given by $\mathrm{e}^{\beta V(\mathbf{z})}$. The VES approach typically involves representing $V(\mathbf{z})$ as a linear combination of basis functions [237].

Simulations

The parameters used to run the simulation remained unchanged (the same as in the case of unbiased simulations in the previous subsection 2.2.1). Enhanced sampling simulations were carried out for 500 ns for the Pr and Pfr forms of the BphP-BV complex. While simulations of BV chromophores, without a protein environment in *ZZZ* and *ZZE* conformations, were performed for 200 ns. The bias in the simulations performed was presented as a Fourier series:

$$V_\alpha(\mathbf{z}) = \sum_k \alpha_k\,\mathrm{e}^{i\mathbf{k}\mathbf{z}} \tag{2.6}$$

where $\alpha = [\alpha_{-N}, ..., \alpha_N]$ are real variational coefficients and $\mathbf{z} \equiv (\Phi, \Psi)$ and represents the set of the biased variables (Figure 2.2). The expansion order was used as N = 10, the periodic domain of the basis set as $[-\pi, \pi]$, and $\mathbf{k} \equiv [k_1, k_2, ...]$, where each term goes from $-N$ to N [17, 237, 301]. A uniform p_T distribution was used as the target probability distribution, which should be obtained after simulation convergence. The functional $[V\alpha]$ (a detailed definition and description can be found in the section on Variationally Enhanced Sampling (VES) 2.2.2) was optimized using stochastic gradient descent (SGD) [388] with a learning rate of μ = 0.1 and a step size of 1000.

2.2.3 Data Analysis and Visualization

GROMACS 2018.1 [209] with the PLUMED 2.8 plug-in [224,382] were used to analyze all atomistic trajectories. To avoid statistical weights $\mathrm{e}^{\beta V}$ imbalance caused by the initial rapid changes in deviation values, the first 10 ns of the simulation were discarded from the analysis. The remaining samples are used to create free energy landscapes. The weights are scaled to a range from 0 to 1 to avoid numerical problems.

The quantitative differences between FE Pr and Pfr landscapes were calculated by using integrating over the metastable aggregate probabilities:

$$\Delta F = -\frac{1}{\beta} log\left(\frac{\int_k d\mathbf{z}\,\mathrm{e}^{-\beta F(\mathbf{z})}}{\int_l d\mathbf{z}\,\mathrm{e}^{-\beta F(\mathbf{z})}}\right), \tag{2.7}$$

where integration domains correspond to the regions in the CV space that represent

the metastable states k and l, respectively. Calculating FE differences required selecting CV samples from metastable states. Therefore, only CV samples belonging mainly to metastable states were initially selected, using a threshold value to filter out CV samples with high statistical weights. The rest of the samples are not important because their weight decreases significantly with energy, according to the Boltzmann factor. The important CV samples are then grouped into metastable states using clustering. The calculation of FE differences between the corresponding metastable states in Pr and Pfr was performed based on the following scheme.

1. A histogram of CV samples was created and transformed into an FE landscape using $F(\mathbf{z}) = -\beta^{-1} \log P(\mathbf{z})$. The histogram was reweighted with statistical weights $w(\mathbf{z}) = \mathrm{e}^{\beta V(\mathbf{z})}$.

2. CV samples above the threshold value (70 kJ/mol) were rejected in order to prepare a training data set for clustering. CV samples had negligible value of $w(\mathbf{z}) = \mathrm{e}^{\beta V(\mathbf{z})}$.

3. The remaining CV samples were grouped for different numbers of clusters, and silhouette analysis was used to find the optimal number of clusters. Silhouette analysis is a method used to evaluate the quality of clustering by measuring how well each data point fits within its assigned cluster compared to other clusters.

4. The FE differences between pairs of metastable states were calculated using the equation 2.7.

2.3 Results

2.3.1 Collective Variables for Enhanced Chromophore Dynamics

The aim was to investigate how the mechanical signal propagates within the protein structure in Pr and Pfr conformations. Therefore, only the chromophore dynamics were enhanced. The first step was selecting the relevant degrees of freedom for the chromophore embedded in the CBD region of the BphP protein. For this reason, a set of generalized variables was created to describe the chromophore's dynamics quantitatively in the CBD region of the BphP protein. CVs are the term used to describe these generalized variables. The main change in the BV chromophore upon the transition from *ZZE* to *ZZZ* occurs between the C- and D-ring pyrroles (shown at Figure 2.2). Therefore, the first CV considered was the dihedral angle Φ (highlighted in the Figure 2.2), which is involved in the rotation of the D-pyrrole ring around the double bond between the two carbons (C15=C16) [47,65]. Due to its potential involvement in the photoisomerization of the chromophore, the dihedral angle Ψ (Figure 2.2) describing the rotation of the D-pyrrole ring around a single carbon bond (C14-C15) was chosen as the next CV [376,389].

Once the CVs were identified, a 200 ns of unbiased MD simulation was performed at a temperature of 300 K to verify the stability of the Pr and Pfr forms. The equilibrium fluctuations of the CVs in the Pr and Pfr X-ray conformations were also observed during the simulation. Within 200 ns, it was found that the systems are kinetically trapped in X-ray structure basins, and thermal fluctuations of Pr and Pfr do not allow for the exploration of other metastable states. The enhanced sampling simulations were the subject of the next part of the study. These simulations were conducted at a temperature of 300 K for 500 ns for both of the protein-chromophore complex conformations and for 200 ns for the chromophore in solution, excluding the protein environment.

2.3.2 Thermodynamical Landscapes from Variational Sampling

Figure 2.8 illustrates the FE landscapes estimated using the umbrella reweighting method. FE profiles of the systems were calculated by integrating out unused CVs along the dihedral angles Φ and Ψ. The deepest FE minimum was placed at 0 kJ/mol in each landscape. This was done by analyzing the difference between global FE minima and maxima in both Pr and Pfr.

To evaluate the convergence of the enhanced sampling simulations, it was necessary to verify whether the simulations effectively explore the CV space. For BphP-BV, both simulations repeatedly intersected the CV space after 500 ns, as shown in Figure 2.5. The simu-

lation's convergence is achieved earlier in the case of the BV chromophore when it is in a solvent without a protein environment (Figure 2.6).

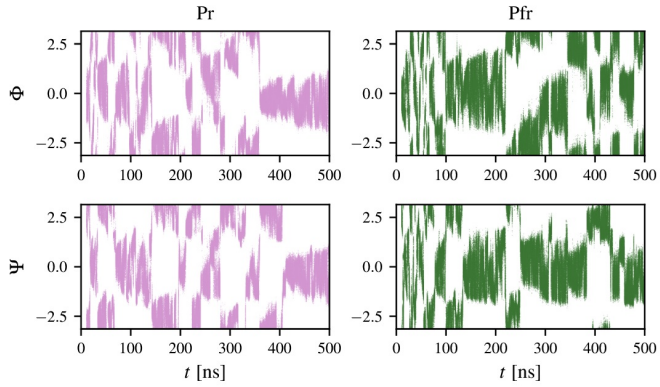


Figure 2.5: Time-series of the Φ and Ψ CVs for the Pr and Pfr conformers for the chromophore in protein environment. The first 10 ns are not shown because the statistical weights were not equilibrated (from Rydzewski et al. [59]).

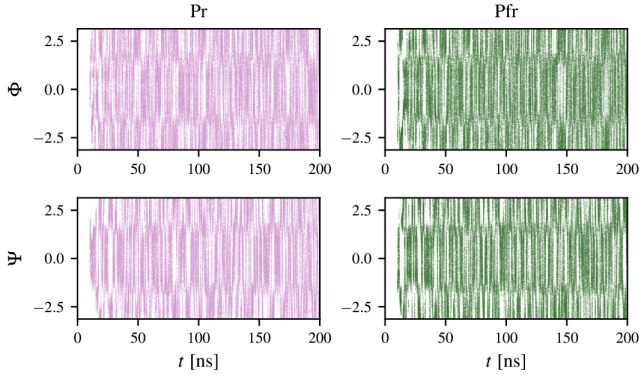


Figure 2.6: Time-series of the Φ and Ψ CVs for the Pr and Pfr conformers for the chromophore in solvent. The first 10 ns are not shown because the statistical weights were not equilibrated (from Rydzewski et al. [59]).

In the final stage of the simulation, transitions in CV space occur much more frequently because the deviation potential causes the dynamics to behave similarly to diffusion.

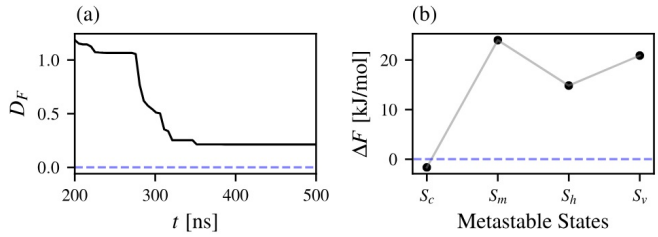


Figure 2.7: (a) Statistical distance, $D_F[F(\mathbf{z}), F(\mathbf{z})]$, used to compare FE landscapes on a global scale. Here, the FE landscapes Pr and Pfr are used as $F(\mathbf{z})$ and $F'(\mathbf{z})$, respectively. FE landscapes are calculated based on samples up to time t. (b) FE differences between the corresponding metastable states in Pr and Pfr are calculated using Equation 2.7. The names of the metastable states are shown on the x-axis (from Rydzewski et al. [59]).

Figure 2.8 shows the FE landscapes associated with CV deviation for both states of the system: Pr and Pfr. The white dots visible in the diagrams correspond to X-ray structures from the original PDB files (Pfr, 4O0P, and Pr, 4O01). In addition, Figure 2.8 shows the FE landscapes of the *ZZZ* and *ZZE* conformations of the BV chromophore without the protein environment. The FE profiles along each CV are shown in Figure 2.9.

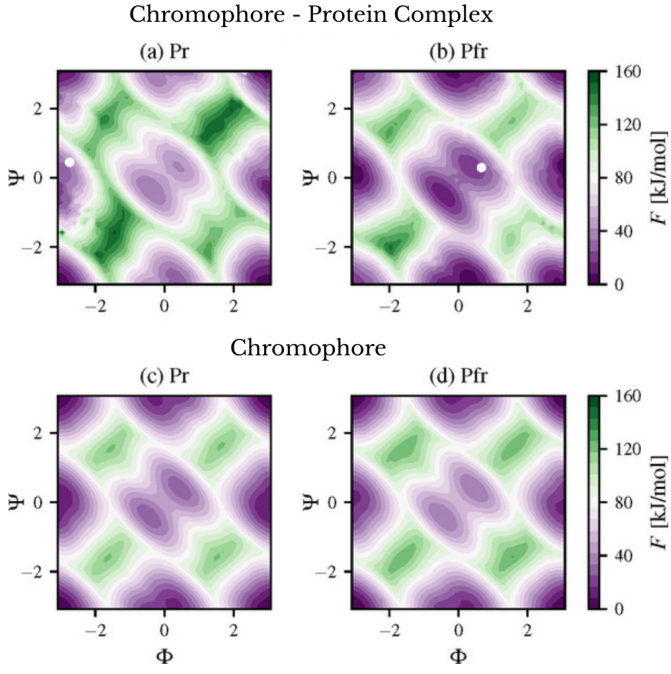


Figure 2.8: Free energy calculated by biasing the dihedral angles Φ and Ψ (in radians) using variationally enhanced sampling (VES) for different photoactive conformers of BphP-BV: (a) the Pr form (PDB:4O01 [47]) and (b) the Pfr form (PDB:4O0P [47]). The X-ray structures from the PDB are represented by white dots. In the lower part, FE landscapes for the chromophore in solvent are shown in (c) *ZZZ* and (d) *ZZE* (from Rydzewski et al. [59]).

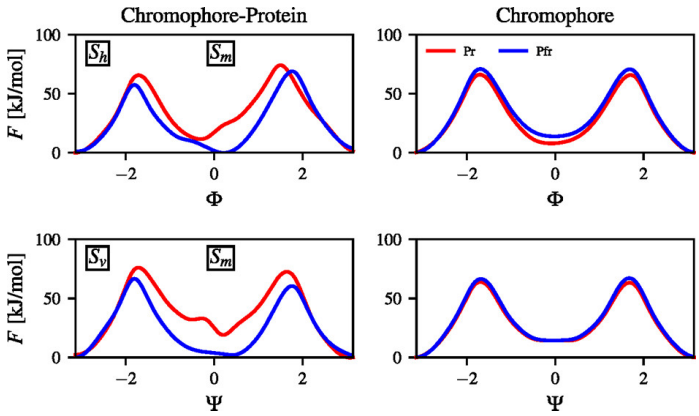


Figure 2.9: The first column shows the free energy profiles for the Pr and Pfr conformers of the BphP-BV complex, and the second column shows the free energy profiles for the chromophore without the protein environment. These profiles were calculated from the FE landscapes of the Pr and Pfr conformers in dihedral angle space (Φ, Ψ) in radians (from Rydzewski et al. [59]).

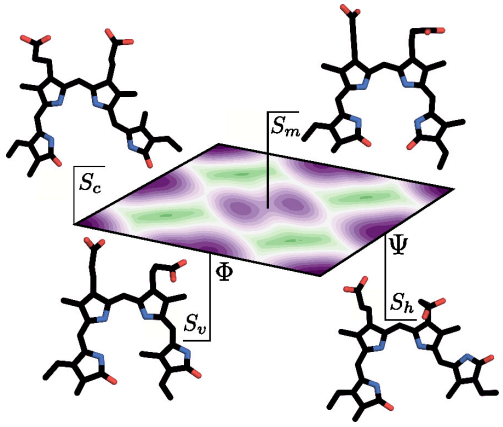


Figure 2.10: Naming convention: the labels visible next to the structures of BV correspond to the names of the metastable states used for Pr and Pfr and the respective conformations: corner S_c, vertical S_v, horizontal S_h, and middle S_m (from Rydzewski et al. [59]).

Figure 2.10 shows the naming convention used in this work, which illustrates BV conformations corresponding to locations on the FE landscape of BV in solution. Metastable states are designated in the following order: S_m (middle), S_v (vertical), S_h (horizontal), and S_c (corner). The S_h and S_m metastable states have also been identified as the X-ray structures of the Pr and Pfr states, respectively. The metastable state of S_m is defined as a single superstate, rather than two separate states. However, it consists of two substates with a relatively low energy barrier and frequent conformational transitions (as illustrated in Figure 2.8 and Figure 2.10).

As shown in Figure 2.8 and Figure 2.10, the EF landscapes for the Pr and Pfr states consist of four metastable states located in similar CV regions, separated by high energy barriers. Upon closer analysis of Figure 2.8, it becomes evident that there are noticeable differences between the FE Pr and Pfr states when a chromophore is bound to the protein. The change in the FE landscapes of the chromophore-protein complex is caused by the interactions between the BV chromophore and the amino acids that form the CBD, as shown in the Pr and Pfr crystallographic structures. Furthermore, as shown in Figure 2.8, the FE landscapes of the chromophore itself in the solvent do not demonstrate any differences between the *ZZZ* and *ZZE* forms. Despite the initial sampling of S_m, it is clear that the *ZZE* chromophore form transitions at 0 kJ/mol to the more preferred metastable state, S_h. It can also be seen that the *ZZZ* and *ZZE* chromophore forms have the highest FE barrier in the solvent, at approximately 125 kJ/mol (as shown in Figure 2.8). The calculated FE barriers are higher for BV in a protein environment than for the chromophore in solution alone. The presence of amino acids in the protein environment appears to provide interactions that restrict the rotation of the D ring, thereby increasing the energy barriers. The FE barrier is highest for Pr at approximately 140 kJ/mol, and for Pfr at approximately 125 kJ/mol (see Figure 2.8). The highest free energy (FE) barriers of the Pfr and Pr forms of the chromophore in the solvent are similar to the Pfr state barriers of the chromophore-protein complex. This observation suggests that the chromophore may exhibit a wider range of conformational freedom in the Pfr form of CBD compared to the Pr form. Figure 2.11 shows that the BphP-BV complex in the Pfr form has lower FE barriers than the isolated BV chromophore. Based on the FE landscapes (Figure 2.8) and FE profiles (Figure 2.9), it is clear that both CVs are necessary to distinguish all distinct metastable states.

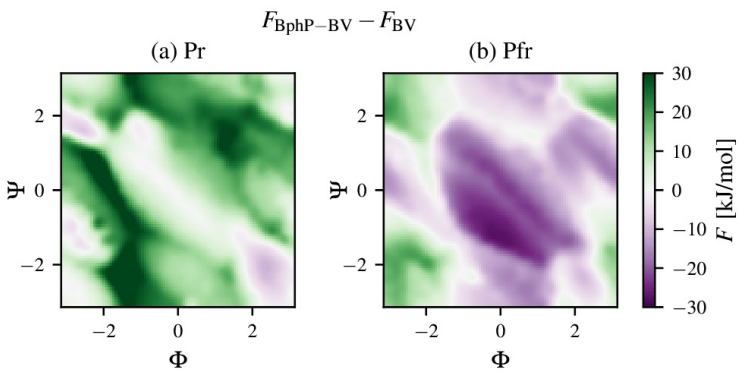


Figure 2.11: Protein constraint effect on FE landscapes of the Pr and Pfr conformers estimated by calculating differences between FE landscapes of the protein-chromophore complex and the isolated chromophore (from Rydzewski et al. [59]).

2.3.3 Metastable States

Despite the fact that the metastable states containing the X-ray structures of the Pr and Pfr conformations are located in the same CV regions (Figure 2.8), they differ significantly from each other. As Figure 2.8 shows, this difference is indicated by a change in the metastable X-ray states, i.e., states that are filled with X-ray chromophore conformations. It is important to note that the Pr X-ray state is slightly shifted relative to the FE basin minimum. For the Pr conformer, the S_m state is characterized by low FE values. On the other hand, the crystallographic state for Pfr is the metastable S_m state.

The change in X-ray states can be reduced to a rotation of the dihedral angle, denoted by Φ, of approximately 180 degrees. The linear approximation of this transition shows that there is almost no change in the dihedral angle Ψ.

Upon analyzing the Figure 2.7b, significant FE differences were found between Pr and Pfr. Differences are calculated by comparing Pr with Pfr. A negative value indicates that the state is more predominant in Pr, and a positive value indicates that the state is more common in Pfr. The largest FE difference happens between states S_m and is about 25 kJ/mol. This metastable state has a lower FE basin in Pfr than in Pr. This can be justified by the fact that the X-ray structure of Pfr belongs to the S_m state. Minor differences (15–20 kJ/mol) are observed in the case of S_h and S_c. This is especially interesting since S_h is the X-ray state of Pfr. The only state more favorable to Pr is S_c, which is defined by a lower FE than S_c in Pfr (Figure 2.7b). These findings support previous analyses (Figure 2.8) and suggest that the FE landscapes of Pr and Pfr are substantially different.

2.3.4 Crossing Free Energy Barriers and Dark Thermal Reversion

In canonical phytochromes, in addition to photoconversion occurring between the Pr and Pfr states, a one-way thermal pathway is often seen that makes it possible to return from the Pfr state to the Pr state [65, 367]. Only BV-binding BphPs with a resting Pr state exhibit Pfr structural heterogeneity of the chromophore, which is determined by a temperature-dependent equilibrium. This may be related to the chromophore's ability to undergo thermal isomerization and revert to its original state, Pr. This dark reversion is a rare event. It occurs over a specific timescale that depends on the type of phytochrome. This process can take anywhere from minutes to hours [390, 391].

As mentioned earlier, the Pfr form (S_m) crystallographic state has two substates that frequently transition between each other. The first substate includes the X-ray conformation of Pfr, while the second substate is divided from the X-ray substate by a low FE barrier (as shown in Figure 2.8). This is in opposition to the S_m state of Pr, which is unpopulated

compared to its X-ray state (S_h). One possible pathway involves the X-ray substate of Pfr, which goes through the intermediate state and the S_h metastable state. This metastable state has a similar FE to that of the X-ray state of Pr. This kind of shift relates to the dark thermal reversion from Pfr to Pr. It is more likely than the conversion from the X-ray state of Pr to the S_m metastable state, since the FE barrier separating these states is considerably higher for Pr.

Furthermore, the time scales for the thermal reversions are in the range of seconds for the dark thermal reversion of Pfr to Pr and hours for the thermal reversion from Pr to Pfr. These time scales indicate that the probability of observing the thermal reversion in canonical phytochromes is higher for the Pfr to Pr transition. For bathy phytochromes, though, this trend may reverse, as the Pfr form is the main ground state conformation for such phytochromes (Figure 2.4) [381, 392].

Figure 2.12 shows the intermediate state of Pfr compared to the X-ray state of Pfr. Based on the visible conformations, it can be concluded that a hydrogen bond with Ser468 stabilizes the D ring of BV in the X-ray state of Pfr. However, in the intermediate state of Pfr, the stabilizing hydrogen bond is broken. A change in the hydrogen bond pattern is followed by a shift of the tongue region away from the chromophore position (Figure 2.12b) and a clockwise rotation of Arg466, His467, and Ser468. These movements suggest that the conserved amino acid triad in the tongue region significantly impacts the BV in the CBD and likely influences the structural change from an α-helix to a β-sheet.

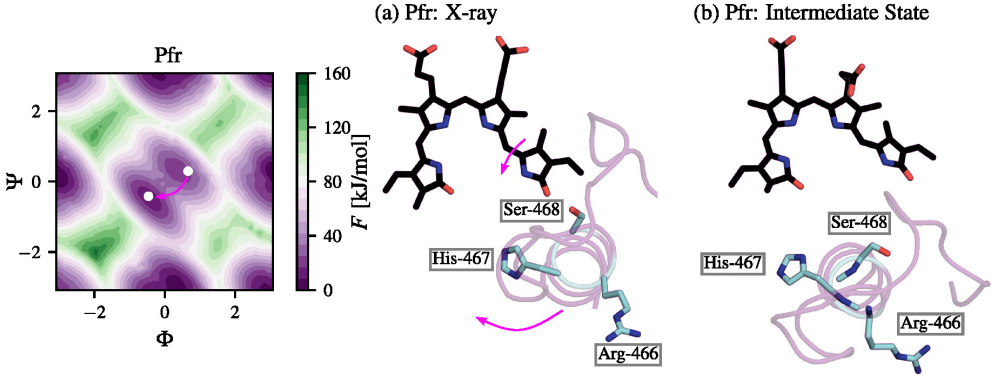


Figure 2.12: Representative BphP-BV structures of the Pfr conformer in (a) the X-ray FE basin and (b) the intermediate state crucial for the dark thermal reversion toward Pr. The tongue region of the protein is shown in magenta and the chromophore structure is illustrated in black. Amino acids important for the transition are labeled. The dihedral angles Φ and Ψ are shown in radians (from Rydzewski et al. [59]).

This intermediate state may be an important step in the thermal dark reversion of Pfr to Pr. These results also suggest that clockwise triad movement is more likely to occur in

the Pr form. The next section analyzes how the movement of the catalytic triad influences the secondary structure change of the tongue region and the conformational constraints and freedom of the BV chromophore in the CBD.

The following transition model was adopted in the FE landscape to explain the transition between the Pr state and the Pfr state, and vice versa. In the Pr-Pfr transition, dynamics begin in the X-ray Pr state (S_h), propagate upward (Figure 2.8), and overcome the FE barrier of approximately 100 kJ/mol. Then, the transition continues downwards and reaches the FE minimum of 30 kJ/mol. The system undergoes a series of intermediate transitional states [62, 393–396] (Figure 2.4) that cannot be detected due to the limitations of the methods used. Then, it reaches the FE Pfr basin (0 kJ/mol). When the transition occurs in the reverse direction (from Pfr to Pr), the system begins to exhibit dynamics in the X-ray FE Pfr basin (0 kJ/mol). It then overcomes the FE barrier at approximately 75 kJ/mol and reaches the metastable S_h FE state via Pfr transitional conformations at approximately 0 kJ/mol. In this case, the FE basin has comparable depths (as can be seen in Figure 2.8).

Based on the previous information, the FE barriers estimated in this work are comparable to the absorption spectra of Pr and Pfr. Upon analyzing the results, it was found that the transition from Pr to Pfr requires higher energy than the opposite transition from Pfr to Pr. This is demonstrated by the FE landscape of Pr in the chromophore-protein case along Ψ (Figure 2.8), where the barrier between the S_h and S_m states of Pr is wider and higher than the barrier between the S_m and S_h states of Pfr. Thus, the energy required to change the conformation of Pfr is lower. This outcome aligns with the behavior observed in the transition between Pr and Pfr. As mentioned in the introduction to this chapter, the Pr-Pfr transition occurs when a red photon is absorbed. A far-red photon, which carries lower energy, is required to trigger the reverse transition. The Q absorption band of phytochromes ranges from 660 to 700 nm, equivalent to an energy range of around 155 to 181 kJ/mol. Figure 2.8 shows that the FE barrier required for the Pr transition from S_h to S_m is approximately 100 kJ/mol, compared to approximately 75 kJ/mol for the Pfr transition from S_m to S_h. These observations align with the expectation that a higher excitation energy is necessary to reach the excited state, which leads to the transformation of Pr into Pfr via conical intersection [397, 398]. The dissipation of energy within the protein may be a contributing factor to the observed variations in transition energies, particularly in instances where heat is transferred to the surroundings [399].

2.3.5 Mechanical Signal Transduction

Various conformational changes in protein environment are caused by their ligand fluctuation and diffusion [208, 400–408]. In the simulations analyzed in this study, the FE land-

scape for chromophore photoisomerization was determined by the local enhancement of chromophore fluctuations, primarily in the D-ring pyrrole of the BV bound with the protein. This enabled evaluation of the hypothesis that mechanical signal propagation across the BphP-BV complex could be initiated by local structural changes in the chromophore. These changes could trigger a series of conformational shifts in the amino acids of the CBD, which would eventually propagate to the GAF and PHY domains. During propagation from the CBD to the PHY domain, the signal amplifies, leading to significant conformational changes in the PHY domain, but on timescales longer than those examined in this study [47, 380, 409, 410].

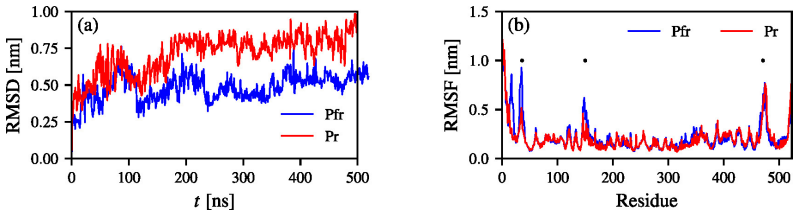


Figure 2.13: RMS descriptors of the enhanced sampling trajectories. (a) RMSD values calculated by first fitting the Cα of the PAS-GAF region to a reference structure. (b) RMSF values calculated for each residue of Pr state and Pfr state of the protein. Dots indicate RMSF values for most fluctuating residue regions (from Rydzewski et al. [59]).

Figure 2.13a shows the RMSD values calculated for the Pr(Pfr) trajectory obtained using VES, plotted against the Pr(Pfr) X-ray state. The average RMSD values are around 0.5 nm for Pfr and 0.7 nm for Pr. Initially, the fluctuations of the BV D ring are low, up to roughly 150 ns. After this time, the RMSD values begin to rise. The conformational change in the PHY regions occurs simultaneously for the Pr and Pfr conformers. The changes in PHY occur with a considerable delay because, at this stage, the chromophores in the Pr and Pfr conformers pass through the CV space multiple times (as shown in Figure 2.5). Figure 2.13b shows the RMSF values for each BphP amino acid. Three notable peaks are visible, indicating increased fluctuations in three regions: the PAS region (residues 30–70), the small α-helix at the PAS-GAF domain interface (residues 120–170), and the tongue region (residues 450–490) close to the chromophore. As suggested in the work of Isaksson et al., 2021 [380], the main part in signal transduction in phytochromes is the PAS-GAF spine helix. The results presented here indicate that the Pr form is more stable.

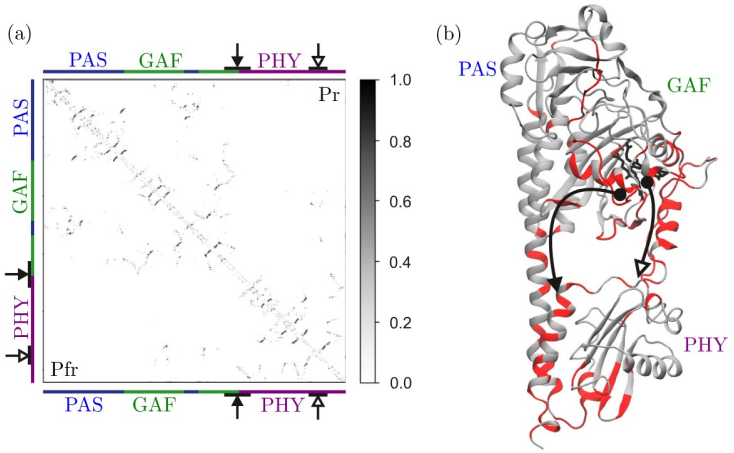


Figure 2.14: Occurrence of pairwise close contact during the enhanced sampling simulations for the Pr and Pfr conformers. a) Black and white-headed arrows indicate regions of the tongue region and the helix spine. Putative signal transduction pathways are shown in red as the regions with the greatest difference (> 0.7) in pairwise contact frequency between the Pr and Pfr models (from Rydzewski et al. [59]).

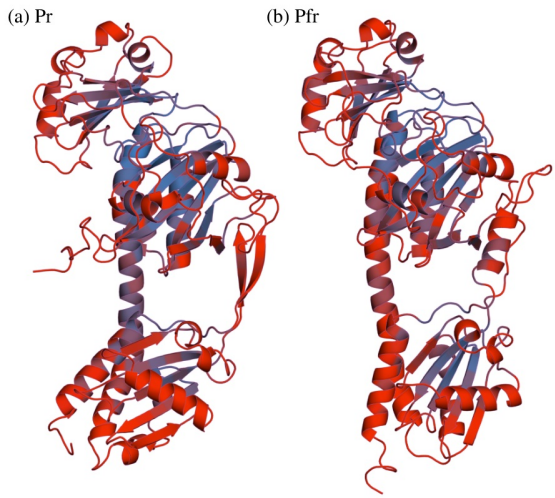


Figure 2.15: Structures of (a) Pr and (b) Pfr with residues colored by B-factor values calculated from RMSF. The main difference lies in the PAS-GAF helix spine that is more flexible in the Pfr conformer than in the Pr conformer (from Rydzewski et al. [59]).

Furthermore, the frequency of close residue-residue contacts (i.e., within a distance of 0.35 nm) during simulations with enhanced sampling was analyzed (the results are illustrated in Figure 2.14). By investigating the most frequent residue-residue contacts, we were able to identify where structural changes took place and whether these conforma-

tional changes could be related to enhanced chromophore dynamics. Figure 2.15 shows that, based on these results, there are two distinct structural pathways: one via the partially disordered tongue region, and the other associated with spine helix motions.

The results of this study indicate that by enhancing the fluctuations of the BV chromophore in the CBD, the mechanical signal may be partly transmitted via the tongue region (as suggested in work of Takala et al., 2014 [47]) and the spine helix connecting the PAS-GAF domains to the terminal part of the PHY domain (as demonstrated by NMR in the work of Isaksson et al., 2021 [380]). These subtle reorganizations in the CBD may contribute to the propagation of mechanical signals through both pathways. This suggests that both paths are biologically relevant and can operate together.

2.3.6 Interactions in the Chromophore-Binding Domain

To examine the interactions between the chromophore and the protein, the number of hydrogen bonds between BV and the amino acids that compose the CBD were calculated in the Pr and Pfr states (see Figure 2.16).

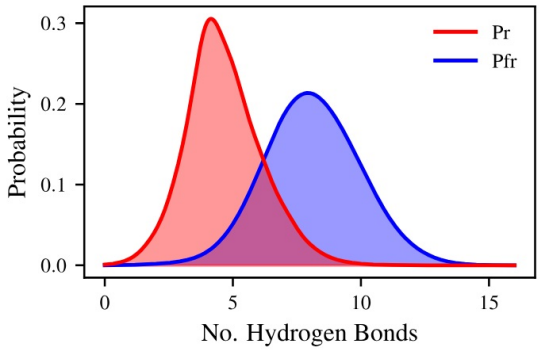


Figure 2.16: The histogram of the number of hydrogen bonds formed during the VES simulations for the Pr and Pfr conformers (from Rydzewski et al. [59]).

The obtained results suggest that for Pr, the average highest number of hydrogen bonds between the protein and the chromophore is approximately twice that of Pfr. This suggests that, in the Pfr state, the chromophore is more likely to fluctuate. However, in the Pr form, the amino acid environment of the CBD restricts the chromophore's movements more (see Figure 2.11). This is supported by the results of the FE landscape calculations, which show that the Pfr conformer has wider FE basins than Pr (Figure 2.8). The distribution of H-bonds is significantly different, suggesting that the arrangement of these bonds between the chromophore and the photosensory core may be a crucial factor. This ar-

rangement could influence the protein's structure, affecting the Pr and Pfr conformers in distinct ways.

Analysis of the hydrogen bonding pattern shows that the most stabilizing hydrogen bonds for Pr occur between rings C and B as well as by bonds with Ser274, Arg254, Tyr216, and Ser257. This is also reflected in the frequency of close contact between residues. For Pfr, the stabilizing hydrogen bonds are nearly identical. These propionate side-chain interactions are unaffected by the enhanced simulations, suggesting that the chromophore photoisomerization has a small impact on the CBD. This is consistent with the findings of related studies [301].

Figure 2.17 illustrates the noticeable rearrangement of amino acids in the CBD that interact frequently with the BV chromophore. The BV-BphP structures shown are taken from X-ray metastable basins: S_h for the Pr conformer and S_m for the Pfr conformer. As can be seen, some amino acids are relocated to the edge of the CBD in the Pfr conformer compared to the Pr conformer; for example, Tyr176, Asp221, and His260. Significant structural changes to the tongue region can be observed in the Pr (β-sheet) and Pfr (α-helix) conformers near the BV chromophore. These structural changes create more conformational space for the BV chromophore in the Pfr conformer.

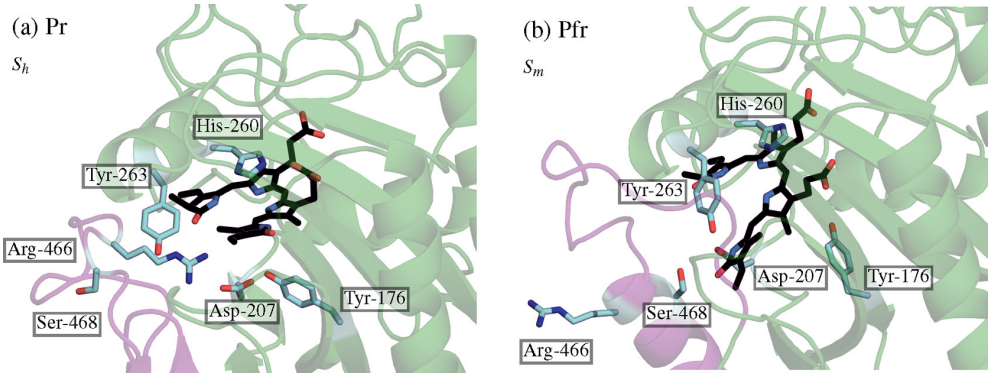


Figure 2.17: Representative structures of the BV-BphP complex taken from the corresponding X-ray metastable basins: (a) S_h for Pr state, and (b) S_m for Pfr state. The BV chromophore in the CBD is shown in black, and the labeled amino acids that interact with the D-ring of the chromophore are shown in cyan. The tongue region (residues 450–490) is depicted in magenta, and the CBD is shown in green (from Rydzewski et al. [59]).

In both Pr and Pfr, the OD1 and OD2 atoms of Asp207 accept a hydrogen from the nitrogen of the D ring of the chromophore, and the nitrogen of His260 accepts a hydrogen from the D ring of the BV. The Asp-207 side chain is part of a ligand pocket in the CBD. The Tyr176 and Tyr263 side chains also belong to this triad, which is located in proximity to the D ring. The side chains of Arg466, His467, and Ser468 in the tongue region behave differently in Pr and Pfr. In the Pr conformer, Arg466 donates two hydrogen atoms to the D-ring nitrogen

due to highly dynamic hydrogen-bonding patterns for C=O [411]. His467 then accepts an H-bond from the D-ring nitrogen. These amino acids stabilize the D-ring, which restricts its rotation. These interactions are disrupted in Pfr, however, due to conformational changes in the tongue region that move Arg466 and His467 outside the pocket (Figure 2.17). Ser468 then directs towards the CBD and donates a hydrogen atom to the D-ring oxygen.

The results obtained from the simulations suggest that the primary photoisomerization event, which causes the BV to change the rotation angle of the D-ring, leads to the destabilization of the CBD around the D-ring. This replaces stable interactions with the formation of a loose cavity, in which H-bonds between the CBD and BV are less likely to occur. The D-ring rotation induces changes in the disordered region of the tongue, which leads to a secondary structural change from an antiparallel β-sheet to an α-helix. However, it is the interactions between the BV and the disordered region of the tongue that allow for this conformational change. Further study is needed to confirm whether the in-plane rotation of the BV relative to the protein matrix is an important factor in the rearrangement of the CBD, despite the clear differences in the angle distribution of BV between the Pr and Pfr forms.

2.4 Conclusion

It should be noted that, by design, enhanced sampling MD simulations cannot observe the ultrafast photochemical steps during the photoisomerization process. Nevertheless, in any transition through short-lived intermediaries [62, 393–396], these stages result in the complex populating the Pfr conformer. The behavior of the chromophore in terms of its conformation is different in the protein matrix for the Pr and Pfr conformers, as demonstrated. Using enhanced sampling MD pushes these conformers beyond the equilibrium states that are inaccessible thermally. Recent applications of enhanced sampling MD to photoactive proteins have shown that simulating these processes using unbiased MD only provides part of the picture [230, 407].

This work reveals that the Pr and Pfr conformers are characterized by multiple metastable states. This was revealed by monitoring and biasing both the Φ and Ψ dihedral angles. However, not all metastable states can be observed only by examining the Φ dihedral angle. This highlights the significance of selecting a set of CVs that can describe the studied systems while preserving as much information as possible [412]. It is not possible to answer the question of whether the Φ dihedral is the only degree of freedom necessary to describe BV-BphP photoisomerization quantitatively in experimental conditions. But a detailed thermodynamic characterization of Pr and Pfr can be provided by biasing both dihedral angles. Photoisomerization events, which were once thought to only affect double bond rotation, have recently been found to modify single bonds as well [17, 379, 413]. It can be seen from the FE landscapes that the transition pathways involving both dihedral angles are possible.

The most advanced experimental techniques have not been able to answer one key question: which structural changes in phytochromes are responsible for transmitting the mechanical signal from the CBD to the PHY domain? Although IR studies can provide general knowledge about the nature of structural rearrangements, the resulting insights are limited in terms of atomic-level resolution. However, time-resolved data at the required accuracy level cannot be provided by XFMS experiments (X-ray hydroxy radical footprinting (XF) with mass spectroscopy (XFMS)) [414], which have a single-residue spatial resolution. The model presented in this study provides a thermodynamic and atomistic description of the driving force that enhances local BV chromophore fluctuations, leading to multiscale structural rearrangements that affect the Pr and Pfr conformers of the protein. It has been found that the mechanical signal is potentially propagated through the tongue region [47] (found by the X-ray structures) and via the PAS-GAF helical spine [380] (found by using NMR) to the terminal part of the PHY domain.

The higher energy required for the protein-chromophore to change from Pr to Pfr compared to the reverse transition, as shown in this work. This is in agreement with the absorption spectra of Pr and Pfr. The Q-band absorption of phytochromes is within the range of 660–700 nm [415], which correlates to an energy range of approximately 155–181 kJ/mol. The results of this study indicate that the estimated FE barriers are consistent with the requirement that the excitation energy must be higher to reach an excited state, which converts Pr via the conical intersection to Pfr. It has been suggested that energy may be dissipated within the protein, which could explain the significant difference observed between the Q-band energies and the FE barriers estimated using VES. This is because energy is dissipated in the form of heat to a thermal bath under experimental conditions [399].

Our significant finding is the potential for dark thermal reversion from Pfr to Pr. Experiments have proven that the BV-like chromophore may have alternate conformations. This is demonstrated by the presence of temperature-dependent chromophore conformations in many phytochromes [390]. In some cases, the Pfr state of canonical phytochromes is heterogeneous, with a temperature-dependent equilibrium crucial for the thermal reversion [416]. The existence of an interplay between two different Pfr chromophore conformations is demonstrated in this study. These metastable states are separated by a low free energy (FE) barrier, resulting in transitions between them occurring on the microsecond timescale. This aligns with the work of Salewski et al. [392], where it was suggested that the Pfr conformer is more likely to introduce heterogeneous transient chromophore conformations into the CBD.

It has been demonstrated that the disordered part of the tongue region is affected by the rotation of the D-ring pyrrole of BV. During dark thermal reversion, a clockwise rotation of the amino acid triad Arg466, His467, and Ser468 is observed, which leads to the breaking of the hydrogen bond between Ser468 and the D-ring pyrrole. This clockwise rotation appears to be of biological significance in stabilizing the Pr conformation through H-bond with Arg466 and a closing motion of the CBD around the BV chromophore. This demonstrates how the protein environment can influence the heterogeneous photodynamics of the chromophore, as well as the potential conversion from a β-sheet to an α-helix within the conserved tongue region of BphP.

3

Photoactive Yellow Protein and *p*-coumaric acid

Chapter based on:

Authors: Sylwia Czach, Jakub Rydzewski

Original title: Understanding Mechanistic Transitions in Photoactive Yellow Protein with Machine Learning

Published in: *in preparation*

3.1 Introduction

Photoactive Yellow Protein (PYP) was first discovered in 1985 in the species *Halorhodospira halophila* [417], but can be found in other halophilic purple sulfur bacteria. *H. halophila* belongs to a group of obligatory photosynthetic, anaerobic, purple, sulfur-oxidizing bacteria. It thrives in salty environments with a salinity of 15–20% and uses sulfur compounds to produce energy. Unlike plants, algae, and cyanobacteria, it does not produce oxygen during photosynthesis [54, 418–420]. PYP is a specialized photoreceptor protein that enables bacterial cells to move away from potentially harmful blue light through negative phototaxis [421, 422].

PYP is a highly soluble globular protein. It exhibits the characteristic fold of the Per-ARNT-Sim (PAS) domain, which is a structural motif commonly found in proteins that detect and respond to light. This fold typically consists of a five-stranded antiparallel β-sheet surrounded by several α-helices. Due to its small size, stability, and solubility, PYP is experimentally accessible and widely used as a reference model in photoreceptor research. Furthermore, the reversible, fast, and well-defined photocyclic transition it undergoes under light makes it an ideal model system. For these reasons, PYP is one of the simplest and best characterized systems for studying light perception at the molecular level [11, 423, 424].

3.1.1 Structural Overview of Photoactive Yellow Protein

PYP is a small protein consisting of around 125 amino acids (14 kDa) that form a rigid structure around the chromophore *para*-coumaric acid (*p*CA). PYP's secondary structure consists of five main antiparallel β-sheets (βI–βV), which form a central β-sheet scaffold structure, as well as four α-helices (illustrated in Figure 3.1). The α1 helix, which is located at the N-terminus region of the protein (residues 1–25), is more flexible and undergoes conformational changes when signaling occurs. The α2, α3 and α3 helices surround the β-sheet, stabilizing the structure [11, 45, 425].

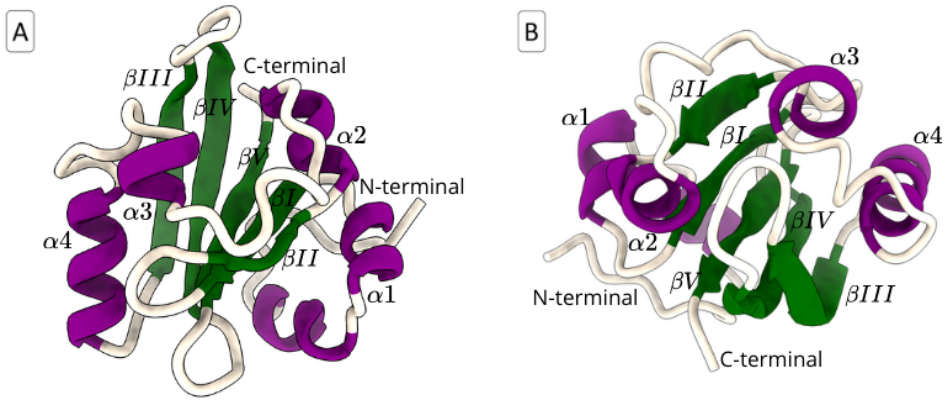


Figure 3.1: Structure of the photoactive yellow protein A) in front view and B) top view, colored according to secondary structures. Numbering of secondary structures βI–βV and α1–α4 applied. Protein in ground state, based on PDB: 1TS6

The chromophore is covalently bound to Cys69 via a thioester bond (as shown in Figure 3.2). *p*CA is positioned deep within the hydrophobic cavity of the protein core, and its placement is strictly limited by non-covalent interactions (e.g., hydrogen bonds and van der Waals forces) with the surrounding residues [11,173,423].

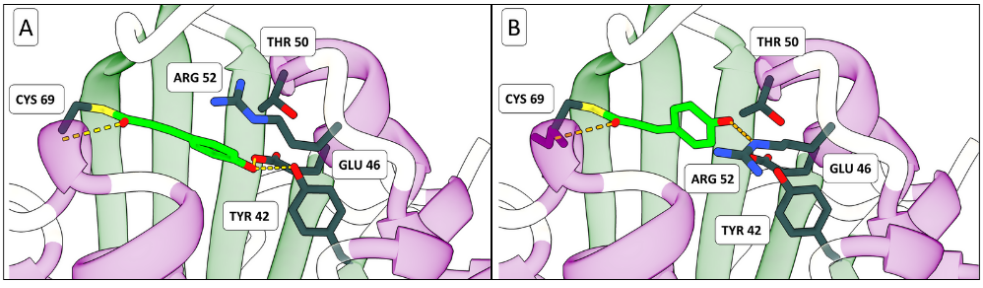


Figure 3.2: Binding site of the PYP chromophore A) in the ground state with the chromophore in the *trans* form (pG), and B) in the illuminated state with the chromophore in the *cis* state (pB). The binding of *p*CA (in lime) to Cys69 and labeled amino acids forming part of the hydrogen bond network (highlighted in yellow) are shown in both panels.

Figure 3.2 shows that several conserved amino acid residues are important for the structural stability and proper photochemistry of the PYP system:

- **Cys69** forms a covalent thioester bond with the *p*CA chromophore, which is essential for its function [426], additionally, *p*CA is stabilized by the interaction between the oxygen of the chromophore and the nitrogen of cysteine [427].

- **Glu46** affects the pKa and photoreactivity of the chromophore by acting as a hydro-

gen bond donor for its phenol oxygen, which stabilizes the chromophore in its deprotonated, light-absorbing form [428].

- **Tyr42** forms a network of hydrogen bonds with Glu46 and the chromophore, which provides additional stabilization of the active site [429].
- **Arg52** interacts electrostatically with the chromophore's carboxylate group, contributing to its orientation in ground state and stabilization via H-bond in excited state [430].
- **Thr50** it may also indirectly participate in the positioning and stabilization of the chromophore by maintaining the stability of amino acids that are directly involved in forming hydrogen bonds with the chromophore (e.g. Tyr42).

The hydrogen bonds formed between *p*CA and Glu46 and Tyr42 in the ground state of PYP are extremely short, 2.47 Å and 2.52 Å, respectively [428, 431]. This bond is a non-standard hydrogen bond, which can be considered to have properties similar to a covalent bond, significantly stabilizing the transition state and facilitating enzymatic reactions [431–434].

3.1.2 The Chromophore: *p*-Coumaric Acid

The chromophore present in PYP is *p*-coumaric acid ($C_9H_8O_3$, chemically defined as 4-hydroxycinnamic acid), a naturally occurring phenolic compound derived from the phenylpropanoid pathway as one of the three isomers of coumaric acid. Its chemical structure (shown in the Figure) consists of a phenolic ring with a hydroxyl group (-OH) in the para position and a conjugated system of double bonds (C=C–C=O) forming a propenoic acid side chain [435–438].

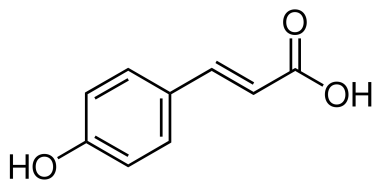


Figure 3.3: The structure of *para*-coumaric acid.

In a solution without a protein environment, the maximum absorption (λ_{max}) of neutral *p*CA is typically around 310–330 nm in the UV range. Upon deprotonation to the phenolate ion, the absorption spectrum shifts towards the near visible range (around 370–390

nm) due to increased electron delocalization. However, the exact value of the maximum absorption depends on the polarity of the solvent and its pH, as these factors affect the protonation state of *p*CA [259,435,439,440]. In the protein environment of PYP, after binding to the hydrophobic pocket of PYP and deprotonation, *p*CA exhibits a strong absorption peak at a wavelength of approximately 446 nm in the blue part of the visible spectrum. This shift in the absorption maximum is the result of a reduction in the influence of the solvent through the protein environment, as well as deprotonation of the phenolic OH group, the formation of a network of hydrogen bonds with Glu46 and Tyr42, and electrostatic stabilization by Arg52 [440–442].

3.1.3 Photocycle and Mechanism of Signal Transduction

PYP is a blue-light photoreceptor that undergoes a well-defined photocycle involving a series of light-induced conformational and chemical changes to its chromophore and protein structure. This reversible process begins in the ground state (pG). It involves multiple intermediate states, each with distinct spectroscopic and structural properties, which ultimately lead to the formation of a long-lived state (pB) [443,444].

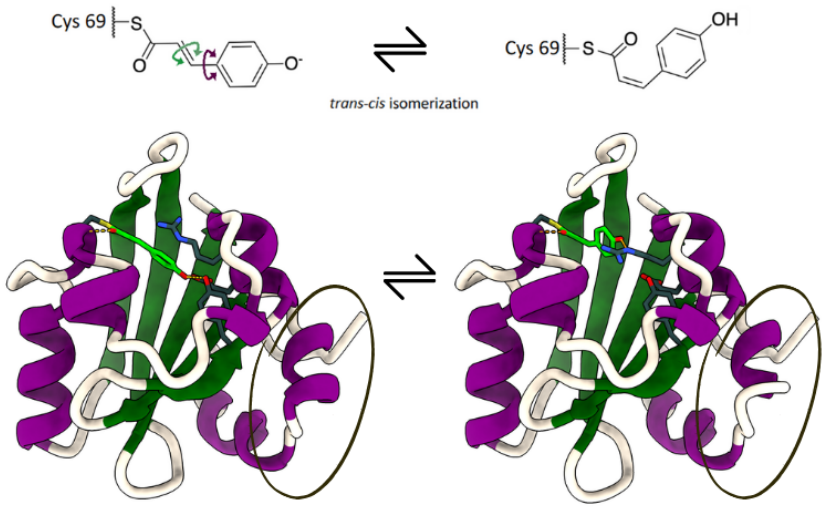


Figure 3.4: Structural changes occurring as a result of photoisomerization of the chromophore from the *trans* to the *cis* state on top. The protein undergoes structural changes as the chromophore transitions from the ground state to the illuminated state. The location of the greatest structural changes is marked with a circle.

The photochemical cycle is initiated when the chromophore, which is covalently linked to Cys69 via a thioester bond, absorbs blue light with a wavelength of around 446 nm [445]. This photon absorption excites the chromophore from its ground electronic state to

an excited state, leading to ultrafast photoisomerization (femtosecond to picosecond timescale) [443]. *Trans-cis* isomerization occurs around the C7=C8 double bond in the propenoic acid side chain of the chromophore [426]. The rotation of the single C6-C7 bond may also play an important role in the primary photochemistry of the PYP (as demonstrated in Figure 3.4) [446]. An interesting fact is that *trans-cis* photoisomerization occurs monoexponentially and without intermediate states in *p*CA in solution, but approximately 10 times more slowly than the process observed for PYP. Isomerization triggers a series of structural changes, resulting in the formation of a series of intermediate products [443, 447]. The basic photocycle scheme can be summarized as follows: pG → pR → pB → pG [11].

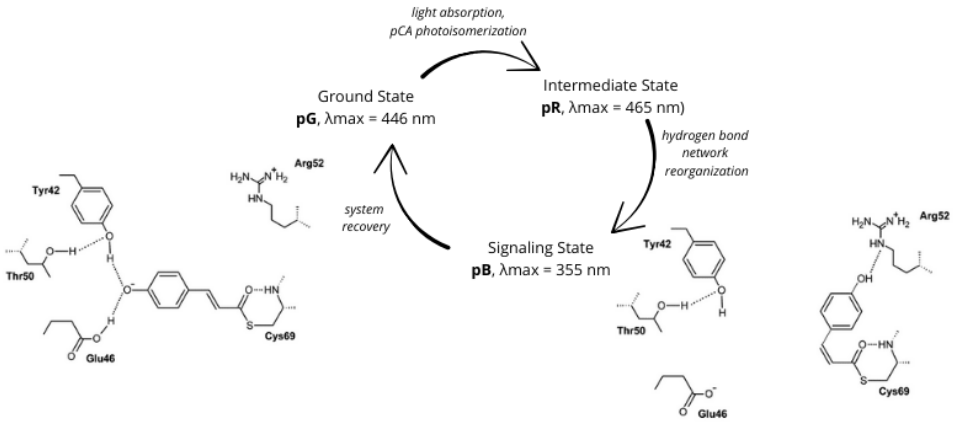


Figure 3.5: Schematic representation of the photocycle, illustrating changes in the hydrogen bonding pattern.

Ground State (pG, λ_{max} = 446 nm) is the thermodynamically stable, dark-adapted form of the protein before it is exposed to light. This represents the initial and final stages of the photocycle, indicating the state to which PYP reverts after signaling state (pB) decay. In this state, the chromophore is completely planar and deprotonated, and the protein exhibits its native conformation. The *p*CA anionic form is stabilized by a hydrogen bond with Glu46 (which acts as a proton acceptor) and a hydrogen bond with Tyr42 (hydrogen bond network pattern shown in Figure 3.5) [11, 425, 426].

Intermediate State (pR, λ_{max} = 465 nm) forms after the absorption of blue light (446 nm) by the *p*CA. The chromophore remains covalently bound to Cys69 and transitions from its ground state to an excited singlet state. Ultrafast isomerization then occurs around the C7=C8 double bond, transforming the chromophore from *trans* to *cis*. While the *p*CA remains anionic and deprotonated at the phenolic oxygen atom, the hydrogen bond network begins to reorganize (there is a partial breakage of hydrogen bonds between the chro-

mophore and Glu46 and Tyr42). The chromophore pocket becomes more flexible. All of these events destabilize the electronic environment around the chromophore, preparing the protein for major conformational changes. These modifications result in a shift in maximum absorption towards the red spectrum, typically around 465 nm [45, 55, 448, 449].

Signaling State (pB, λ_{max} = 355 nm) called bleached or blue-shifted, is the biologically active, signaling conformation of PYP. The pB state forms after the pR intermediate as the photocycle progresses. This state involves the protonation of the *p*CA (changing it from anionic to neutral), a significant reorganization of the hydrogen-bond network, and major conformational changes to the protein structure (in particular, the N-terminal region). The phenolic oxygen of the chromophore no longer forms strong hydrogen bonds with either Glu46 or Tyr42; however, the chromophore is now stabilized by a hydrogen bond with Arg52. The N-terminus undergoes partial unfolding and displacement from the protein core, thereby increasing its solvent accessibility. This likely exposes new surfaces for potential protein-protein interactions. It is thought that these conformational changes modulate protein-protein interactions, enabling PYP to interact with additional signaling partners or cellular components. The maximum absorption shifts to around 355 nm due to the mentioned structural rearrangements [426, 430, 441, 450, 451].

Reversion to Ground State this can occur through the absorption of radiation by the signaling state, or through thermal relaxation over time. The pB state is long-lived (up to milliseconds) but thermally unstable, eventually decaying back to the pG state. The process of thermal reversion is significantly slower than photochemical-induced and is more sensitive to environmental conditions. In both cases, returning to the ground state involves steps such as the deprotonation of the chromophore, the isomerization of the double bond from *cis* to *trans*, the regeneration of the hydrogen bond network, and the re-folding of the N-terminal region [11, 452, 453].

3.1.4 Purpose and Scope of the Chapter

The main objective of this chapter was to examine a well-known photoactive yellow protein system using novel machine-learning-based methods, namely Spectral Map (SM) developed by J. Rydzewski [454]. After parameterizing the CHARMM36 force field for the systems and performing MD simulations, data from the trajectories were used to train neural networks, construct slow collective variables, and understand how local changes in the vicinity of the chromophore affect larger structural changes in the protein.

3.2 Materials and methods

3.2.1 Structures

The study was based on the structure of the photoactive yellow protein (PYP), PBD number: 1TS6 (structure of the pG and pB from time-resolved Laue crystallography) [45]. The downloaded file was split into two files containing the structures of the two PYP states: the ground-state pG and the signaling-state pB. Hydrogen atoms were added to the structures in both states. The difference between the states was the presence of hydrogen in the pB state at oxygen O4' and its absence in the pG state (as can be seen in Figure 3.5). It was important to consider the protonation state of Glu46 (as shown in Figure 3.2 and Figure 3.5) by keeping the hydrogen atom from the carboxyl group in the pG state and in the case of the pB state, loss of hydrogen from the carboxyl group.

The structures of the chromophore in two states and three amino acids (Pro68, Cys69, and Thr70) with optimized geometries using Schrödinger's Maestro package [455] were saved, and a set of force-field parameters were generated using the SwissParam server [456]. The structure of the chromophore-free protein for the pG and pB states was prepared (pKa values, protonation) using Schrödinger's Maestro package [455]. Next, the structures were used as input for the CHARMM-GUI Input Generator [457, 458] server to obtain CHARMM36 force field parameters for MD simulations in GROMACS. The protein structure was merged with the chromophore structure generated by SwissParam.

3.2.2 Molecular Dynamics

The protein-chromophore complexes in the two states, pG and pB, were solvated, and ions were added. MD simulations were performed using GROMACS (version 2019) [209] and the CHARMM36 force field [459] modified with parameters for the ligand. Two-state PYP circuits prepared in the previous stage were used as starting structures. Five repetitions were created for each state. The TIP3P water model was used to solvate the system (60 Å in all directions). The complexes were neutralized by adding 7 Na^+ ions in pG state and 6 Na^+ ions in pB state. In all the simulations, an integration time step of 2 fs was used, and all hydrogen bonds were constrained using the LINCS algorithm [460]. Long-range electrostatic interactions were treated using particle mesh Ewald [386] with a 1.2 nm cutoff and a 1.2 nm cutoff for nonbonded van der Waals interactions.

The complexes were minimized and subsequently equilibrated at 300 K using the velocity rescaling thermostat. Short MD simulations (0.5 ns each) were then run in the NVT ensemble at 300 K with positional restraints of 1000 and 100 kJ mol^{-1} nm^{-2} on protein

Cα atoms and chromophore. Subsequently, short MD simulations in the NPT ensemble were run at a constant pressure of 1 atm using the Berendsen barostat [461]. Harmonic restraints of 10 kJ mol^{-1} nm^{-2} were applied to protein Cα and chromophore for 1 ns. A production MD simulation of 100 ns was run five times in the NPT ensemble at constant pressure, which was maintained using the Parrinello-Rahman barostat [462]. The velocity rescaling thermostat [387] was used to maintain a constant temperature of 300 K.

3.2.3 Mechanistic Descriptors

To investigate the mechanisms underlying the system's behavior, it was necessary to identify mechanical descriptors that could be physically interpreted and serve as inputs to a neural network used to learn CVs using SM. The key atoms involved in forming the hydrogen-bond network that stabilizes the chromophore position in both the pG and pB states were identified (hydrogen atoms were not included). Initially, ten atoms were identified from the following amino acids in the local environment of the chromophore (1 Å radius): Tyr42(O), Glu46(O), Thr50(O), Arg52(N), Cys69 (N, C, S), and from p-coumaric acid (C, O, O). These atoms were used to define mechanistic descriptors. Subsequent groups of atoms were then selected at distances of 2 Å and 3 Å from each of the initial atoms. The pairwise distances between the aforementioned atoms were used as mechanistic descriptors to describe structural changes in the system. Therefore, for each pG and pB, three datasets were obtained, denote as 1 Å, 2 Å, and 3 Å. The increasing neighborhood around the chromophore served as an indicator of how parts of the protein located at greater distances from the chromophore affected its local environment and dynamics. Finally, for each dataset, 45 distances were selected that explained most of the variance and used them as inputs to SM to learn slow CVs.

3.2.4 Spectral Map

SM simplifies the multidimensional configuration space of the system represented with the n mechanistic descriptors $\mathbf{x} = (x_1, \ldots, x_n)$ by mapping it into reduced space $\mathbf{z} = (z_1, \ldots, z_d)$ given by a set of d functions of the descriptors, commonly referred to as CVs ($d \ll n$). In SM, CVs are represented using a neural network (NN):

$$\mathbf{z} = \xi_w(\mathbf{x}) \equiv \left\{\xi_k(\mathbf{x}, w)\right\}_{k=1}^{d}, \tag{3.1}$$

where w are learnable parameters ensuring that the NN describes slow CVs. By sampling the system in CV space, its dynamics propagates in a free-energy landscape:

$$F(\mathbf{z}) = -\frac{1}{\beta} \log \int d\mathbf{x}\, \delta(\mathbf{z} - \xi_w(\mathbf{x}))\, \mathrm{e}^{-\beta U(\mathbf{x})} \tag{3.2}$$

where $\beta = 1/(k_{\mathrm{B}}T)$ is the inverse temperature with k_{B} denoting the Boltzmann constant, and T temperature.

SM constructs slow CVs by increasing the effective timescales characteristic of the system. To this aim, its reduced dynamics was first modeled as a Markov chain using kernel functions. To measure the similarity between CV samples $\mathbf{z}_k$ and $\mathbf{z}_l$, Gaussian kernel was used:

$$G(\mathbf{z}_k, \mathbf{z}_l) = \exp\left(-\frac{1}{\varepsilon}\|\mathbf{z}_k - \mathbf{z}_l\|^2\right) \tag{3.3}$$

where $\|\mathbf{z}_k - \mathbf{z}_l\|$ denotes pairwise Euclidean distances between every pair of CV samples $k, l = 1, \ldots, N$ and N is the number of samples. The Gaussian kernel exhibits a notion of locality by defining a neighborhood around each sample of radius ε, which was estimated as the median of pairwise distances.

As the marginal equilibrium density in CV space is often far from uniform for dynamical systems with complex free-energy landscapes, a density-preserving kernel is required for data sampled from any underlying probability distribution. For this purpose, an anisotropic diffusion kernel was used:

$$K(\mathbf{z}_k, \mathbf{z}_l) = \frac{G(\mathbf{z}_k, \mathbf{z}_l)}{\sqrt{\varrho(\mathbf{z}_k)\varrho(\mathbf{z}_l)}}, \tag{3.4}$$

where $\varrho(\mathbf{z}_k) = \sum_l G(\mathbf{z}_k, \mathbf{z}_l)$ is a kernel density estimate. A Markov transition matrix was then constructed by row-normalizing K:

$$m_{kl} \sim M(\mathbf{z}_k, \mathbf{z}_l) = \frac{K(\mathbf{z}_k, \mathbf{z}_l)}{\sum_n K(\mathbf{z}_k, \mathbf{z}_n)} \tag{3.5}$$

which models a discrete Markov chain in the CV space $m_{kl} = \Pr\{\mathbf{z}_{\tau+1} = \mathbf{z}_l \mid \mathbf{z}_\tau = \mathbf{z}_k\}$ expressing a probability of transition between CV samples $\mathbf{z}_k$. The Markov chain approximates the long-time asymptotics of the system by describing the dynamics by the Fokker–Planck anisotropic diffusion.

Finally, to estimate the dominant time scales encoded in the system, a spectral decomposition of the Markov transition matrix $M\Psi_k = \lambda_k \Psi_k$ was performed, where Ψ_k and λ_k are the k-th right eigenfunctions and eigenvalues of M, respectively. The real-valued

eigenvalues of M are (sorted in non-ascending order):

$$\lambda_0 = 1 > \lambda_1 \cdots \geq \lambda_N, \tag{3.6}$$

where the eigenvalue λ_0 corresponds to the equilibrium distribution of the Markov chain (3.5) given by the eigenfunction Ψ_0. The dominant eigenvalues related to the slowest relaxation timescales in the system can be found by associating each eigenvalue with an effective timescale $t_k = -1/\log \lambda_k$. The loss function was minimize by the negative variance explained by the first k eigenvalues:

$$L = -\frac{\sum_{i=0}^{k} \lambda_i}{\sum_{i=0}^{N} \lambda_i}, \tag{3.7}$$

where N indicates the total number of eigenvalues. By iteratively minimizing (3.7), the NN learns CVs associated with the slowest modes in the system, while keeping any fast variables negligible.

A schematic illustration of the learning algorithm of SM is shown in Figure 3.6.

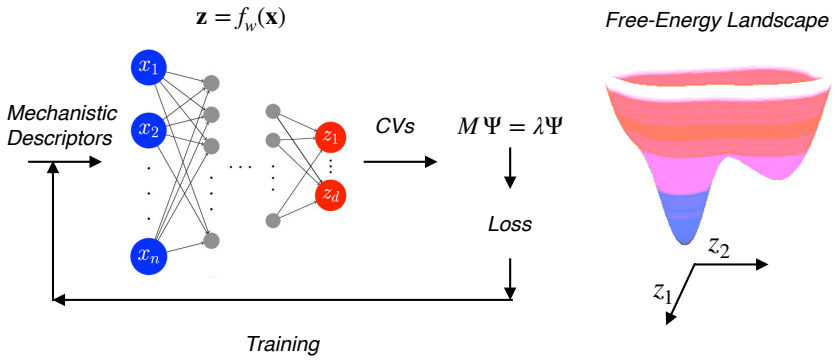


Figure 3.6: Schematic illustration of the learning algorithm of SM. After training, the NN returns the slowest CVs and the associated free-energy landscape can be constructed.

The package implementing SM is written in Python 3. It uses the PyTorch library for the training of NNs. It can be run on a GPU for faster computations. It also employs the PyTorch Lightning framework, which serves as a high-level wrapper for PyTorch to simplify the training process for the users.

3.3 Results

The analysis study mainly on the analysis of CVs learned using SM, the free-energy landscapes calculated based on these CVs, and structural analysis of the metastable states. The first step was to define mechanical descriptors that capture conformational changes in PYP captured during the system dynamics.

Figure 3.7 shows a visualization of the mechanistic descriptors of the system. Figure 3.7:A and Figure 3.7:E present the chromophore and CBD for the pG and pB states, respectively. This figure also highlights the atoms that are important for forming the hydrogen bond network in PYP. By gradually expanding the neighborhood around the chromophore, three datasets containing pairwise distances between the selected atoms for each system: pG (Figure 3.7:B, C, and D) and pB (Figure 3.7:F, G, and H) were defined. The first dataset contains the pairwise distances between atoms in a neighborhood of radius 1 Å around the chromophore; the second of 2 Å, and the third of 3 Å. By learning CVs from these datasets, the influence of structural parts of PYP on the local dynamics of the chromophore could be checked and vice versa.

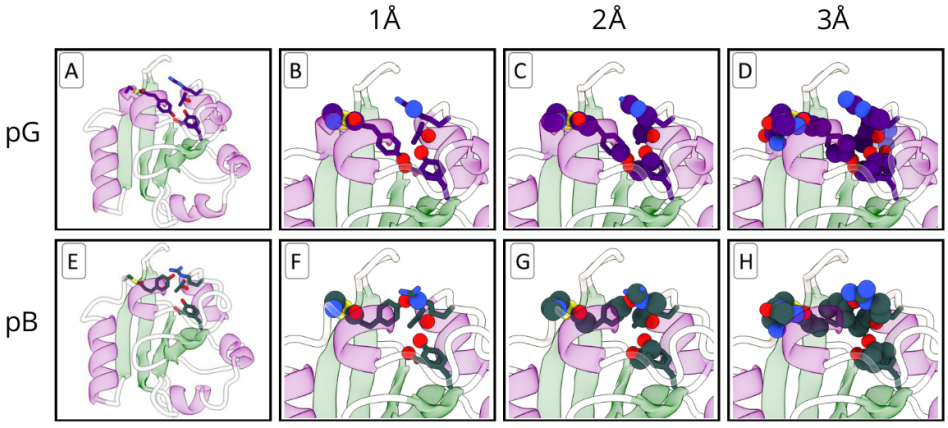


Figure 3.7: A) The p-coumaric acid and the CBD amino acids in the ground state in stick representation, colored by atom type. B) The crucial atoms involved in forming the hydrogen bond network in pG state shown as spheres. C) Initial atoms and atoms located at a distance of 2 Å form them in pG state shown as spheres. D) Initial atoms and atoms located in 3 Å area in pG state shown as spheres. E) The p-coumaric acid and the CBD amino acids in the signaling state in stick representation, colored by atom type. F) The key atoms involved in forming the hydrogen bond network in pB state shown as spheres. C) Initial atoms and atoms located at a distance of 2 Å form them in pB state shown as spheres. D) Initial atoms and atoms located in 3 Å area in pB state shown as spheres.

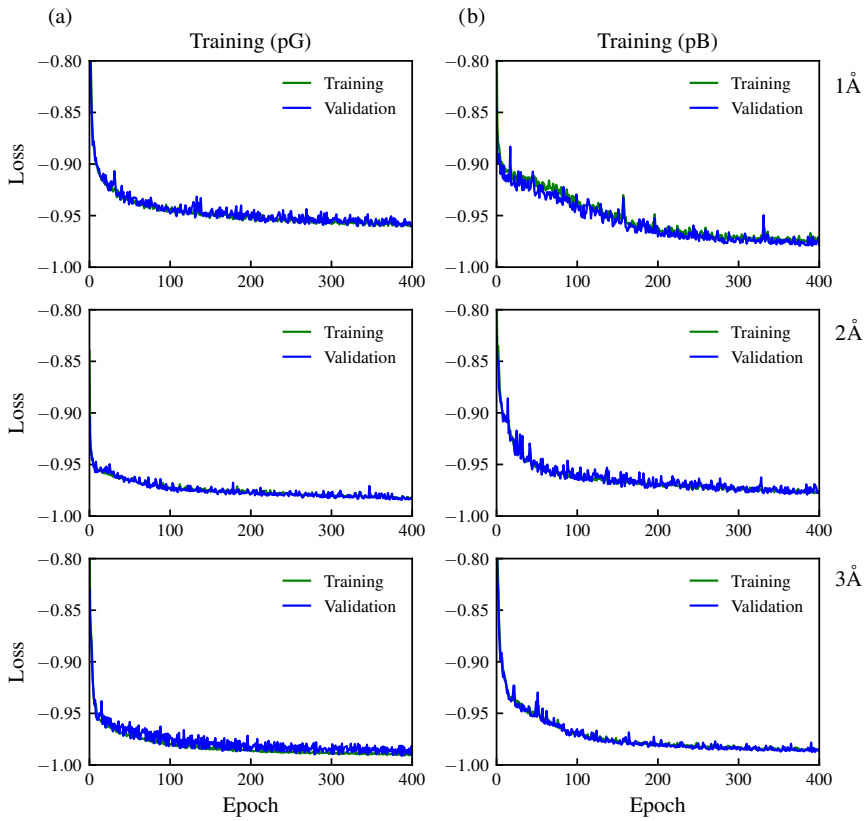


Figure 3.8: Learning results (training and validation losses) for (a) pG and (b) pB represented by three datasets. Training is run on datasets that consist of high-variance pairwise distances defined by the neighborhood of p-coumaric acid of radius 1 Å, 2 Å, and 3 Å.

From each dataset, 45 mechanistic descriptors that had the highest variance to use NNs with the same size were selected. Therefore, all NNs had 45 input nodes and 2 output CVs. In all cases, NNs were employed with two hidden layers and ELU activation functions. Each layer contained 75% of the input size in nodes. Batches of 1000 were used. The data sets were divided into training and validation sets, with the validation set representing 20% of the full data set. The learning over 400 epochs were conducted, but used early stopping to terminate the process when the validation loss reached a precision of 10^{-6}. As an optimizer, Adam with a learning rate of 10^{-3} and default parameters was used. For the loss function was explain by variance in the first $k = 3$ eigenvalues (Eq. 3.7).

As shown in Figure 3.8, both the training and validation losses converged around 400 epochs. Throughout the training process, the validation losses (20 % of each dataset) closely follow the training losses, indicating very stable learning and the ability of the NNs

in each case to generalize the estimation of CVs from data not included in the training dataset.

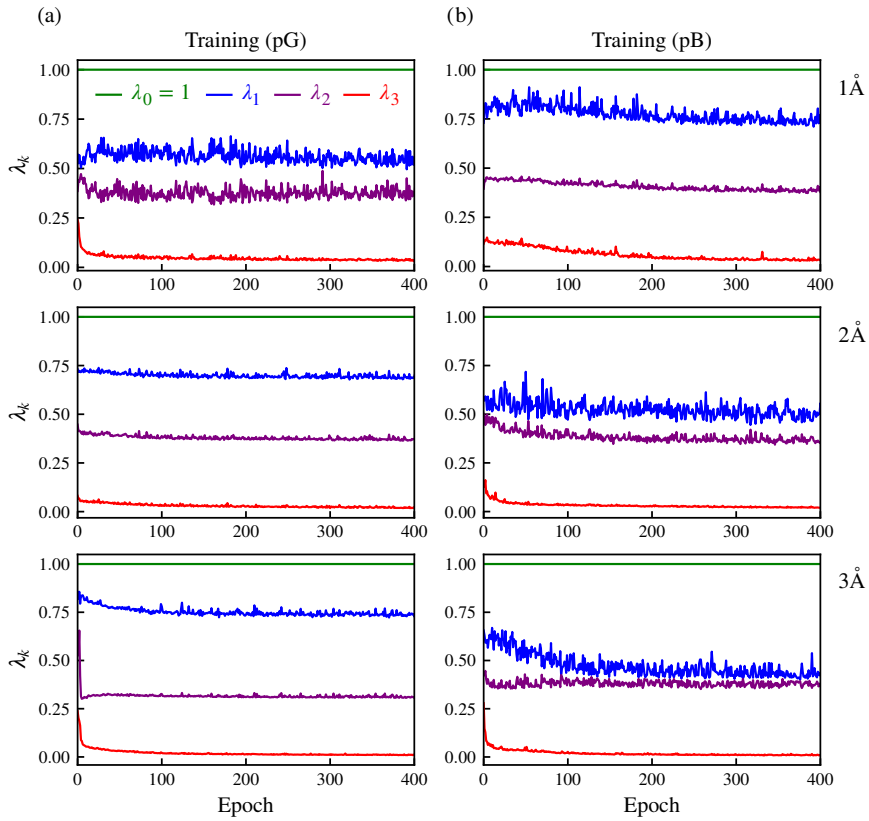


Figure 3.9: Learning results (first four dominant eigenvalues) for (a) pG and (b) pB represented by three datasets. The loss function maximizes the first three eigenvalues ($k = 3$) and minimizes the rest. Training is run on datasets that consist of high-variance pairwise distances defined by the neighborhood of p-coumaric acid of radius 1 Å, 2 Å, and 3 Å.

For both pB and pG, the dominant eigenvalues indicate three main metastable states (Figure 3.9). For pB, increasing the neighborhood from 1 Å to 3 Å results in λ_1 and λ_2 converging to a similar value. This indicates that, at 1 Å, there is a single dominant timescale associated with the highest barrier in the free-energy landscape. As the neighborhood increases, all the barriers in the system converge to approximately the same value. Interestingly, an inverse effect for pG was observe: increasing the neighborhood around p-coumaric acid results in the emergence of a single dominant free-energy barrier. This leads to a very important observation: PYP structural elements, even those separated by 3 Å , affect the local environment around p-coumaric acid. For pB, these elements affect

the dynamics, so it is spread evenly between three metastable states. Conversely, for pG, this leads to the dynamics focusing on two dominant states, while the third state is less populated.

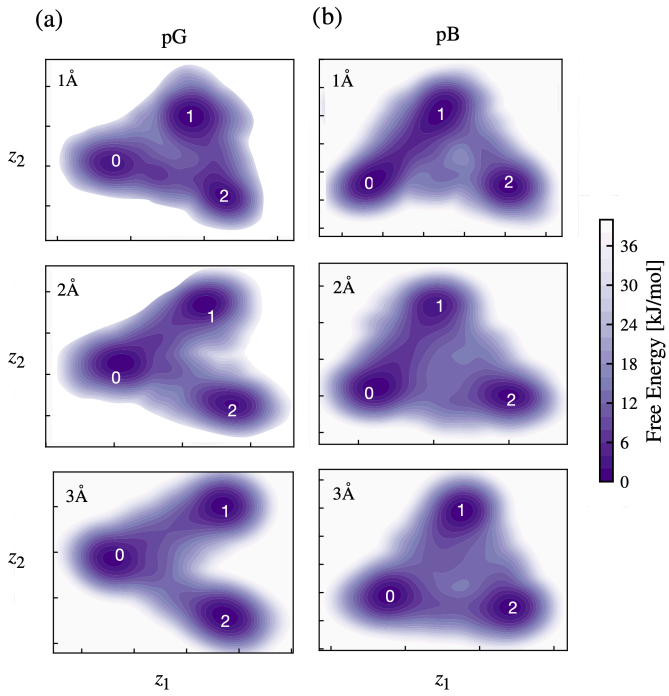


Figure 3.10: Free-energy landscape in CV space calculated for (a) pG and (b) pB corresponding to the datasets of 1 Å, 2 Å, and 3 Åneighborhoods around the chromophore.

The results from training SM are consistent with the free-energy landscapes calculated for each dataset. As indicated by the eigenvalues of the transition matrices, the free-energy landscapes consists of three dominant metastable states with high energy barriers reaching approximately 20 kJ/mol on average. As can be seen in Figure 3.10a, CVs from SM are very consistent between different datasets, showing that the learning algorithm accurately captured the slow dynamics encoded in the MD trajectories. As suggested by the eigenvalues, when the neighborhood around the chromophore increases from 1 Å to 3 Å, this results in the isolation of the metastable state 2; see Figure 3.10(a). In other words, the transition from state 1 to 2 is no longer accessible due to an energy barrier higher than 40 kJ/mol. The opposite behavior can be seen for pB (Figure 3.10b), where state 2, initially relatively isolated (1 Å), starts to be more populated by the increase in the neighborhood radius, which suggests that structural parts of PYP quite distanced from the chromophore

are noticeably affected by the local dynamics.

Based on the estimated free-energy landscapes, conformations of the complex from three metastable states were extracted to analyze its structural dynamics. The results are shown in Figure 3.11 and are labeled as on the free-energy landscapes (Figure 3.10).

In the state 0 of pG (Figure 3.11a), the chromophore is positioned close to Glu46 (in green) and Tyr42 (in magenta), at distances that allow for the formation of hydrogen bonds with these amino acids. In state 1, two dominant conformations of the chromophore were observed: one is near Arg52 (in grey), while the other is farther from this stabilizing triad. State 2 in pG bears structural similarities to state 0, with the chromophore being stabilized by Glu46 and Tyr42. Additionally, the PYP structures in states 0 and 2 exhibit global similarities, indicating that state 1 is likely an intermediate transient state. Interestingly, for the dataset with a neighborhood of 1 Å, the energy barrier between states 0 and 2 is relatively low (approximately 15 kJ/mol). However, when the neighborhood was increased to 3 Å, the barrier height rises significantly, suggesting that distant rearrangements in the PYP structure are affected by the fluctuations and dynamics around the chromophore.

For pB (Figure 3.11b), significant variations in the distances between the chromophore and its immediate surroundings were observed. In state 0, the chromophore is closely positioned near Glu42 and Tyr42, which play a crucial role in stabilizing the chromophore in pG. However, in the 3 Å dataset, the chromophore often interacts with Arg52. States 1 and 2 resemble state 1 in pG, with conformations splitting into two groups: one directed toward Glu46 and the other toward Arg52. In contrast, for the 1 Å dataset, the transitions between states 0 and 2 pass through state 1. In this case, including a larger neighborhood reduces the energy barrier between states 0 and 2, resulting in similar barriers for transitions between all states.

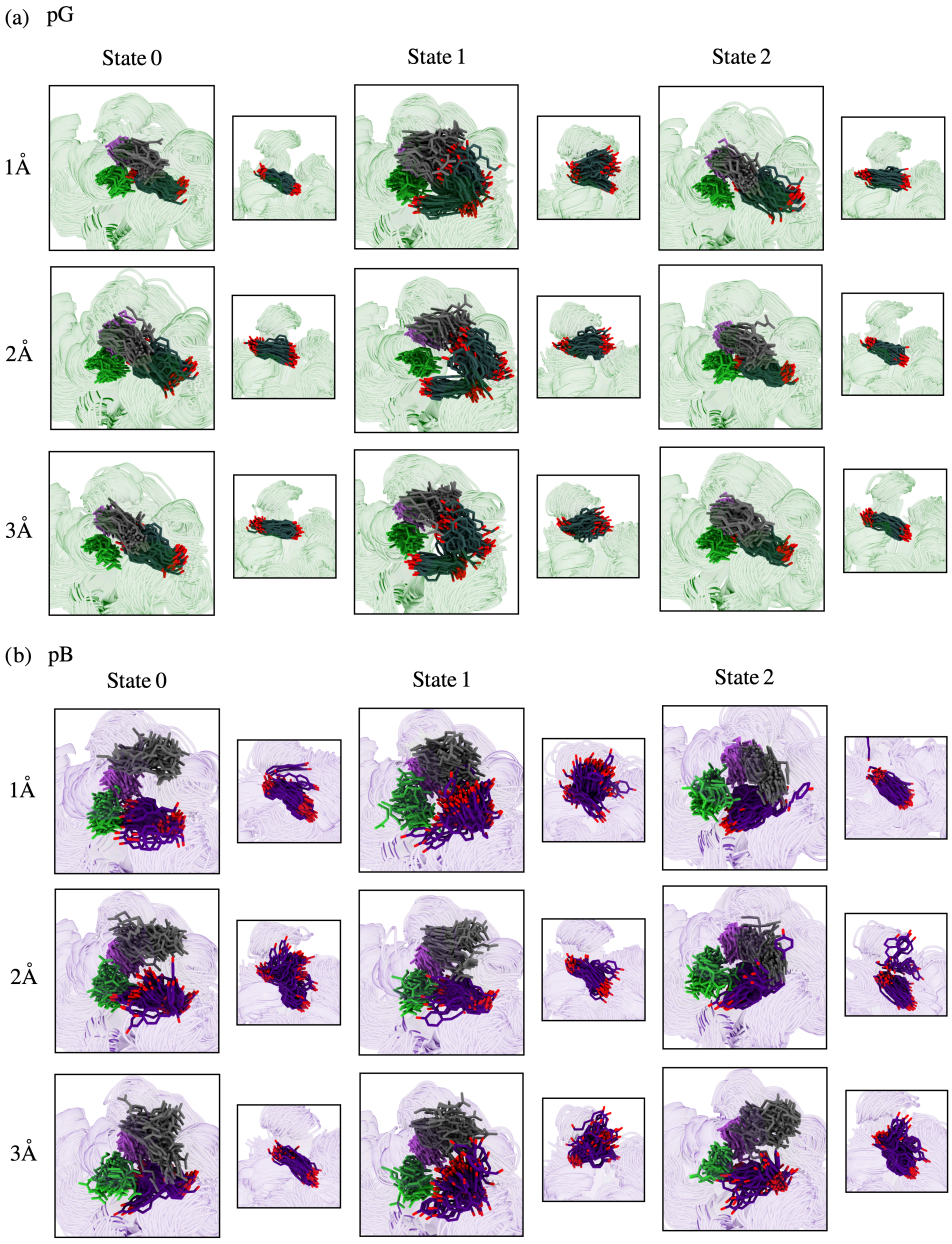


Figure 3.11: Metastable state analysis for (a) pG and (b) pB. PYP conformations from the three identified states (indexed 0, 1, and 2) are numbered as in Figure 3.10. The bigger panel of the figure shows a top view of the chromophore surrounded by amino acids in a stick representation (Tyr42 in pink, Glu46 in green and Arg52 in gray). The smaller panel next to it shows the chromophore in a stick representation from the front (the orientation as in the Figure 3.10) in dark green for pG, and purple for pB.

3.4 Conclusion

This chapter investigates a well-characterized photoactive yellow protein (PYP) system using the Spectral Map (SM) machine learning approach. Following chromophore parameterization in the CHARMM36 force field and subsequent MD simulations, the resulting trajectories were used to train neural networks and identify slow CVs and the associated free-energy landscapes and metastable states. The results demonstrate the applicability of machine learning techniques for extracting mechanistically relevant dynamical features from molecular simulation data.

Although it is not possible to simulate the reversible transitions between pG and pB triggered by light, it is possible to characterize pG and pB and their behavior under equilibrium conditions. The analysis revealed that the structural rearrangements in pB occur in the opposite direction to those observed in pG, indicating that this findings are physically consistent with the PYP photocycle mechanics. It is important to note that used approach and analysis, while promising, are still in the preliminary stages. A more detailed investigation is currently underway and will be published shortly after the PhD defense.

4

Light-induced regulation of proteins, photoactivated ligands of P2X7 receptor

Chapter based on:

Authors: Bartosz Szymczak, Joanna Czarnecka, **Sylwia Czach**, Wiesław Nowak, Katarzyna Roszek

Original title: Purinergic approach to effective glioma treatment with temozolomide reveals enhanced anti-cancer effects mediated by P2X7 receptor

Published in: Cellular Signalling, 106, 110641, 2023

https://doi.org/10.1016/j.cellsig.2023.110641

4.1 Introduction

4.1.1 Overview of the Nervous System

The nervous system is anatomically divided into two parts: the central nervous system (CNS) and the peripheral nervous system (PNS). However, functionally, the two systems are closely linked and operate as an integrated, coordinated whole. The CNS is the brain and spinal cord, both surrounded by skeletal bones: the brain in the cranial cavity of the skull and the spinal cord in the spinal canal. CNS is the control center for the processing and integration of sensory information. The PNS controls the transmission of motor commands and skeletal muscle movement and activity, and manages involuntary functions such as heart rate, digestion, and glandular activity. The PNS consists of 12 cranial nerve pairs, 31 spinal nerve pairs, right and left sympathetic chains, and pelvic parasympathetic chains and associated ganglia [463–466].

From a functional perspective, the nervous system is divided into the somatic and autonomic nervous systems, and this classification is not limited to topographical separation. The autonomic nervous system (ANS) can be divided into sympathetic, parasympathetic, and enteric, and is mostly visceral and, in contrast to the somatic nervous system (SNS), is not under voluntary regulation [463, 466].

In terms of the microanatomy of the nervous system, it consists of two main cell types: neurons and neuroglia. The basic structural and functional components of the nervous system are neurons, which are highly specialized and capable of forming extremely complex networks that receive and integrate signals and generate adaptive responses. Each neuron is made up of an input domain: multiple dendrites along with the cell body, and an output domain: the axon, a single extension of the cell body that propagates an action potential to the synapse, where signals are transmitted from one cell to another. However, in addition to the structural schema, neurons are highly heterogeneous in size and shape, as shown in Figure 4.1 [467–469]. As mentioned, another cell type in the nervous system besides neurons is glial cells. These cells are the most numerous in the human brain, and have long been considered passive cells that support neurons. Compared with the extensive studies of various neuronal functions in the nervous system, our knowledge of glial cells remains limited. However, this knowledge gap is being systematically and dynamically filled by many researchers. We now know that the function of glial cells is to support, protect, and nourish neurons, maintain homeostasis of the extracellular fluid that surrounds them, and participate in synaptogenesis. In contrast to neurons, neuroglial cells can replicate. They multiply to fill gaps caused by neuron loss. There are six glial cell types in the nervous system, divided into two topographically distinct groups.

Those that are found in the CNS are astrocytes, oligodendrocytes, microglia, and lining cells, and the Schwann cells and satellite cells that are only found in the PNS [467,470–472]. The above-mentioned division of cells is illustrated in Figure 4.1.

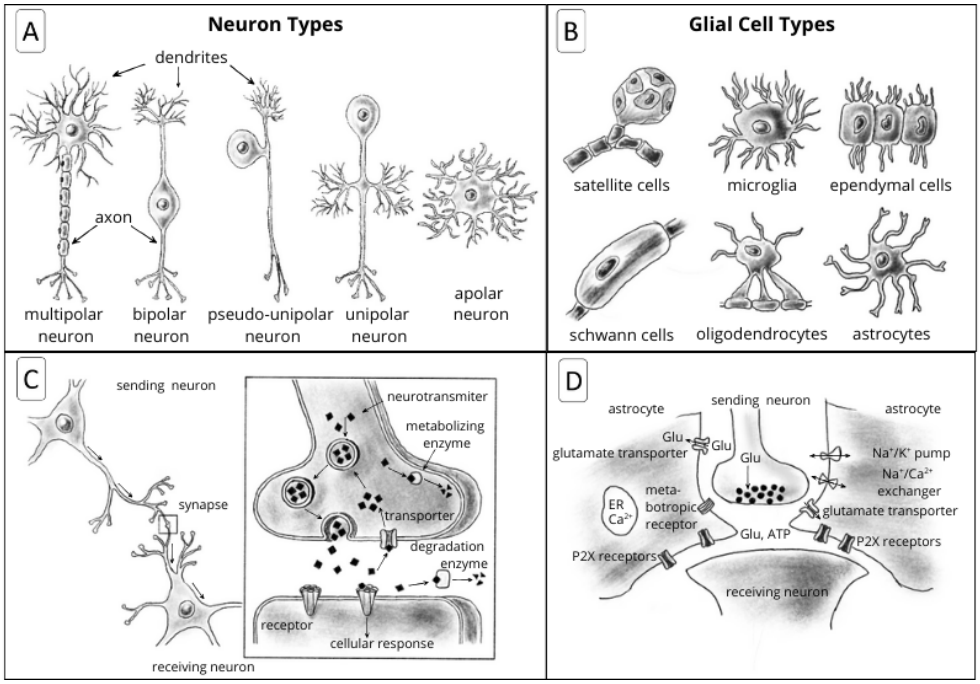


Figure 4.1: A) The types of nerve cell are shown, with their key functional elements — the axons and dendrites — indicated. B) Types of glial cells. C) The schematic representation of a synapse shows the site of communication between neurons by sending and receiving neurotransmitters. D) The schematic representation of the site of interaction between neurons and glial cells.

Nerve cells communicate with each other and with other types of cells through synaptic connections, which are shown in Figure 4.1. The synaptic junction is the place where the nerve impulse is transmitted from one cell to another (depending on the type of synapse) with the participation of a substance with the nature of a neurotransmitter in chemical synapses or by means of an electrical impulse in electrical synapses. Correct communication of neurons through signal transmission at synapses ensures the maintenance of the body's homeostasis, enables adequate perception and response to stimuli, and is responsible for memorization and learning [473–475].

In addition to being a highly specialized form of signal transmission within the nervous system between neurons or between a neuron and an effector (such as a muscle or gland), synaptic communication is one of many interactions between nerve cells, glial cells, and other cells in the body. Other forms of cell-to-cell communication involve neurotransmit-

ters, ion channels, and receptors that transmit, modulate, and interpret signals [476–479].

Glial cells were once thought to be passive support cells that did not produce electrical impulses. However, they are now known to be active communicators. Astrocytes are the most common type of glial cell. They work with neurons in the body to help them work well. They also help form, maintain, and strengthen connections between neurons. They detect synaptic activity and release gliotransmitters, including glutamate, adenosine triphosphate (ATP), and D-serine. These compounds affect the synaptic strength, plasticity, and excitability of neurons [477, 480, 481].

Other glial cells (shown in the Figure 4.1) can also modulate neuronal activity. Oligodendrocytes insulate and myelinate axons in the CNS, facilitating the faster propagation of action potentials. Similarly, Schwann cells insulate and myelinate axons in the PNS. Microglia are phagocytic cells that circulate in the CNS. They protect the nervous system by destroying microbial invaders and removing debris from nerve cell death. Microglia also release cytokines that modulate neuronal activity and synaptic remodeling. Lining cells form the linings of the ventricles of the brain and the central canal of the spinal cord. These cells play a role in forming the blood-brain barrier (BBB) and producing cerebrospinal fluid (CSF). Satellite cells provide structural support and electrical insulation for neuronal cell bodies in the PNS. Nutrient exchange and waste removal between cells is also regulated by them [466, 467, 470–472, 482].

4.1.2 Glioma Cancer

Neurodegenerative diseases and cancer are associated with dysfunctional glial activity. Gliomas originate from glial cells or their precursors and account for the vast majority of malignant brain tumors [472, 483]. There are two classifications of gliomas in use. The first is by cell type. Gliomas are classified based on which cell type they most closely resemble histologically. Examples include astrocytomas, oligodendrogliomas, and glioblastomas [484, 485]. The second classification is by grade, according to the World Health Organization (WHO) Classification of Tumors of the Central Nervous System. This classification includes grades I, II, III, and IV [486].

Glioblastomas are the most predominant CNS tumor, contributing to 81% of malignant brain tumors. Glioblastoma is the most common histology type of glioma (about 45% of all gliomas) with a five-year survival rate of about 5% [487].

The therapy of gliomas is very challenging due to the location of the tumors, the complex tumor microenvironment, and the presence of the BBB. The standard treatment for glioblastoma involves complete resection with a margin of healthy tissue, radiation therapy to the focal tumor area, and temozolomide (TMZ) chemotherapy [488–490].

TMZ is a prodrug and is the preferred chemotherapeutic agent for treating glioblastoma. It is a small, orally administered compound (shown in Figure 4.2) that is quickly and fully absorbed. It can also penetrate the BBB. Spontaneous conversion to an active metabolite (methyl-triazeno-imidazole-carboxamide - MTIC) occurs without the need for enzymatic demethylation in the liver. MTIC methylates DNA and induces single-strand breaks and mismatches during replication. This results in the activation of the mismatch repair (MMR) system. Consequently, cell cycle arrest and apoptosis occur [491,492].

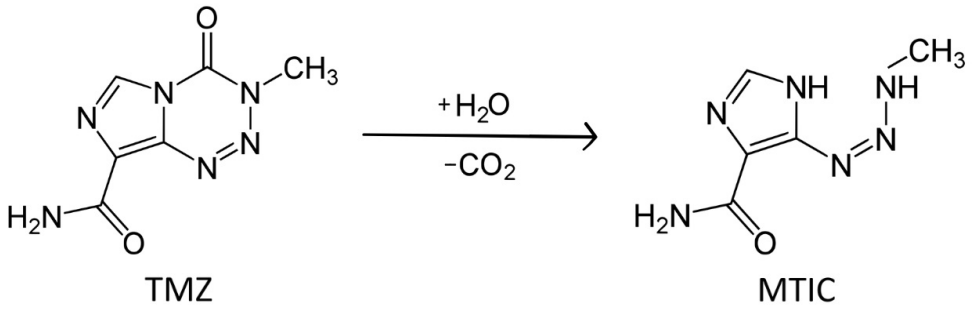


Figure 4.2: The structure of temozolomide, the pro-drug, and the structure of MTIC, the activated metabolite, after undergoing a series of metabolic reactions shown schematically in the image.

Even though conventional methods have improved, patient survival remains limited, underscoring the need for new approaches. Some of them are based on the search for drugs that target purinergic receptors, mainly P2X7, because its increased expression is observed in differentiated glioma cells [493]. Other innovative therapies for treating brain tumors focus on immunotherapy [494,495], using non-coding RNA [496], and light involving therapies [497].

4.1.3 P2X7 Receptor

The P2X7 receptor, which belongs to the P2X purinoreceptor family, is a trimeric, nonselective ion channel gated by extracellular ATP (adenosine triphosphate, structure shown at Figute 4.3). Its long intracellular C-terminal domain sets it apart from other P2X receptors. Like the other P2X family receptors, the P2X7 receptor mediates the flow of Na^+, K^+, Ca^{2+}, and other small cations [498,499]. However, the P2X7 receptor has a feature: it can open a large cellular pore that allows organic ions, such as N-methyl-D-glucamine ($NMDG^+$) and $choline^+$, to pass through. It can also transmit signals through the ATP-dependent pathway without creating a channel. A high concentration of ATP is required to activate the P2X7 receptor, which is higher than that required for the other receptors in the family

and is approximately 100 μM. It is present in many cell types, including stem cells, blood cells, glia, neurons, bone cells, muscle cells, and skin cells. Due to its presence on the surface of many cell types, P2X7 is involved in a variety of physiological and disease-related processes [500–503]. These procedures involve participation in the initiation of inflammatory reactions by stimulating the inflammasome and secreting interleukins [504]. The production of reactive oxygen and nitrogen species can be affected by the activation of the receptor to mediate the destruction of intracellular pathogens [505]. The proliferation and death of various cell types can also be triggered by P2X7 receptor [506]. The metabolism of glucose and lipids is also influenced by P2X7 receptor activities [507, 508].

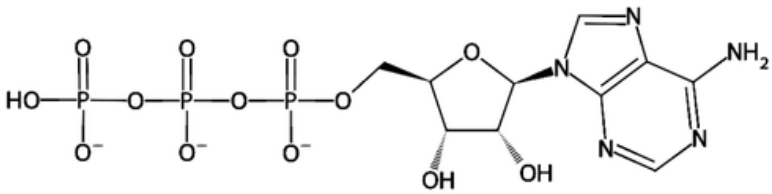


Figure 4.3: The structure of ATP (adenosine triphosphate) PubChem CID:5957, ATP has three phosphate groups, each with a negative charge at physiological pH.

In addition to humans (*Homo sapiens*), six other mammalian species have been found to have the P2X7 receptor: macaques (*Macaca*), dogs (*Canis lupus familiaris*), pandas (*Ailuropoda melanoleuca*), mice (*Mus musculus*), rats (*Rattus norvegicus*), and guinea pigs (*Cavia porcellus*). These species exhibit a degree of similarity to the human receptor that approaches 75% identity. Among non-mammalian species, such as zebrafish (*Danio rerio)*, the similarity decreases to less than 40%. The P2X7 receptor is a protein composed of three usually identical subunits and contains 1785 amino acids (in rare cases, P2X7 subunits can combine with P2X4 subunits to form a receptor in heterotrimer form).

Each subunit of the human receptor is composed of 595 amino acids. The receptor has two alpha-helical transmembrane domains (TM1 and TM2), intracellular N-terminal and C-terminal domains, and a large ectodomain with multiple anti-parallel β-strands, containing two ATP-binding sites. The structure of the monomer with domains highlighted, as well as the entire receptor with a visible ATP binding pocket, is shown in detail in Figure 4.5 [500, 502, 509, 510].

The P2X7 subunit is distinguished by its larger C-terminus, which contains an additional 200 amino acid residues and forms two domains: a cysteine-rich domain and a lipopolysaccharide-binding domain [511]. The N-terminus has been found to have a specific site for protein kinase C, which is thought to regulate P2X7 desensitization and

control receptor gating. This site also enables activation and regulates Ca^{2+} flow through the channel [512, 513].

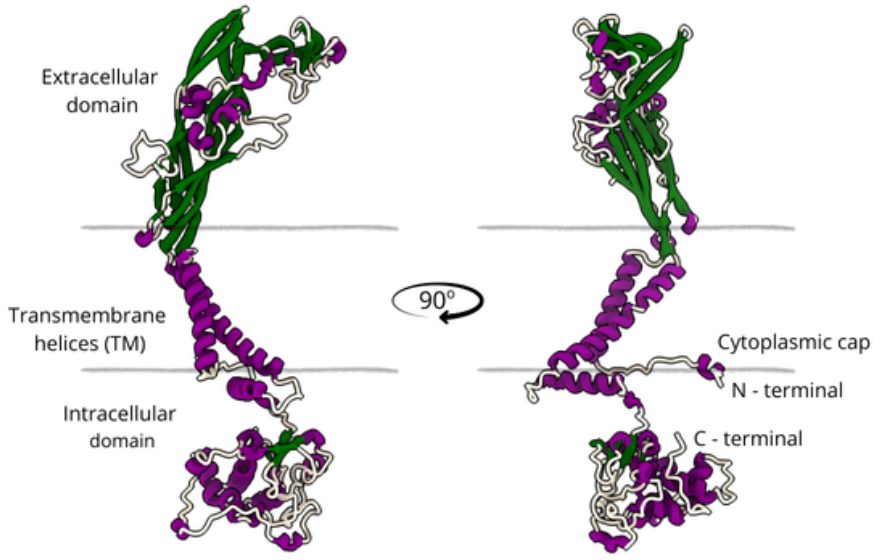


Figure 4.4: Structure of the p2x7 receptor monomer with visible extracellular and intracellular domains and transmembrane α-helixes. Structure colored according to secondary structures, with the cell membrane shown as a gray lines. Monomer of the closed rat receptor PDB: 6U9V.

There are three ATP-binding pockets (also called orthosteric binding pockets) in the extracellular part of the receptor, formed by the ATP-binding sites of each pair of adjacent monomers. The pocket is formed by six highly conserved amino acid residues: Lys64, Lys66, and Thr189 from one monomer, and Asn292, Arg294, and Lys311 from an adjacent monomer. The amino acids form a narrow (around 11 Å), shallow (around 8 Å), and U-shaped cavity that is exposed to the solvent. Three important amino acids in the ATP-binding process are the hydrophobic residues Leu191, Ile214, and Ile228. These residues are oriented toward the ligand and generate a local hydrophobic environment that supports ATP binding [514, 515].

Figure 4.6 shows the amino acids that make up the pocket and its surface area, other essential and supporting ATP-binding residues, as well as a representation of how ATP is located inside the cavity. As previously mentioned, the P2X7 receptor is sensitive to high concentrations of ATP. The presence of this molecule in at least two pockets results in channel opening. Short activation opens a standard ligand-gated ion channel that is permeable to small cations. Prolonged, repeated activation of the receptor results in the formation of

large pores across the plasma membrane that allow molecules up to 900 Da to pass freely. This allows fluorescent dyes to enter the cell. However, prolonged opening of the pore contributes to cell death [502, 514, 515].

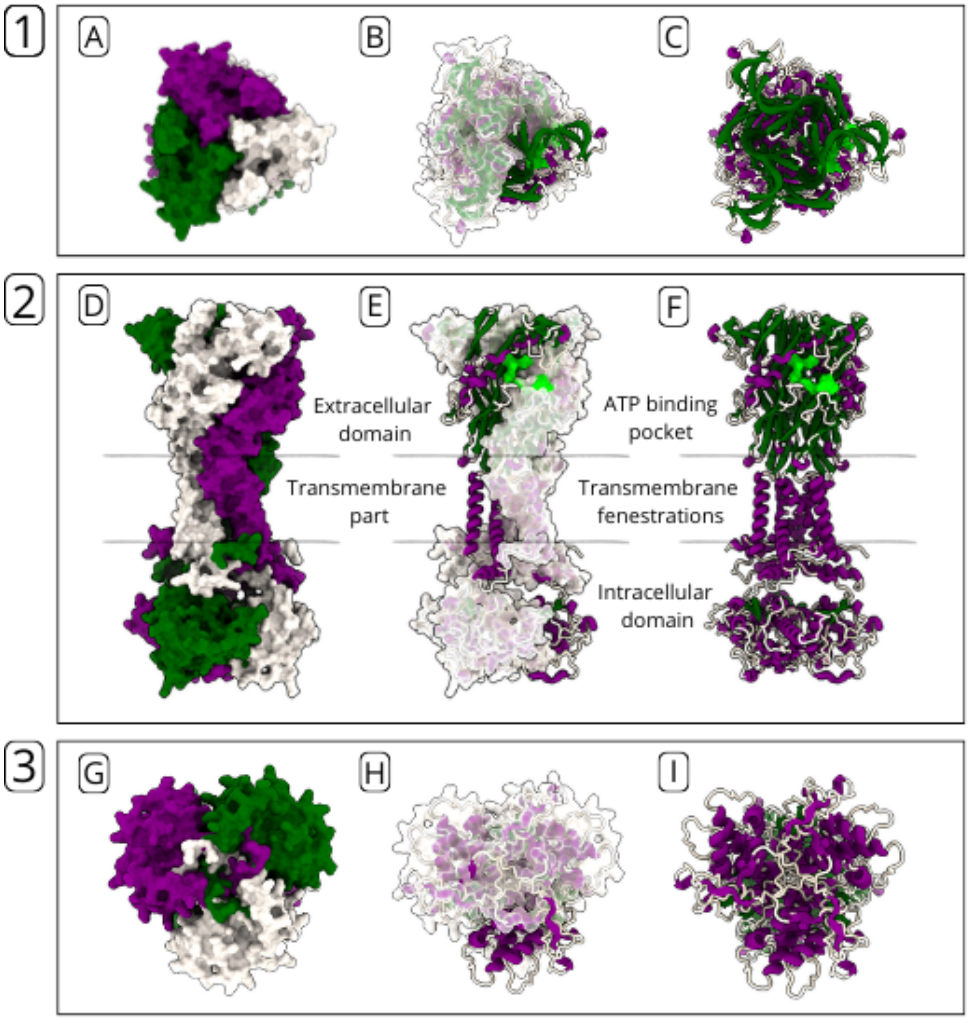


Figure 4.5: Overall architecture of p2x7 with visible extracellular and intracellular domains, transmembrane part and membrane shown as a gray lines. Structure of the closed rat receptor PDB: 6U9V. **1.** Top view **A)** Organization of three monomers in relation to each other, each shown as a surface in different color. **B)** The arrangement of the monomer relative to the other chains. Conformation of the monomer in color relative to the secondary structures (α-helix in purple, β-strand in dark green and coil in white), the other chains forming the channel shown as a semitransparent surface with ATP-binding pocket visible as surface in lime color. **C)** Overall architecture of the receptor visualize as secondary structure with ATP-binding pocket visible in the extracellular part shown as surface in lime color. **2.** Front view **D)** Organization of three monomers in relation to each other, each shown as a surface in different color. **E)** The arrangement of the monomer relative to the other chains. Conformation of the monomer in color relative to the secondary structures (α-helix in purple, β-strand in dark green and coil in white), the other chains forming the channel shown as a semitransparent surface with ATP-binding pocket visible as surface in lime color. **F)** Overall architecture of the receptor visualize as secondary structure with ATP-binding pocket visible in the extracellular part shown as surface in lime color. **3.** Bottom view **G)** Organization of three monomers in relation to each other, each shown as a surface in different color. **H)** The arrangement of the monomer relative to the other chains. **I)** Overall architecture of the receptor visualize as secondary structure.

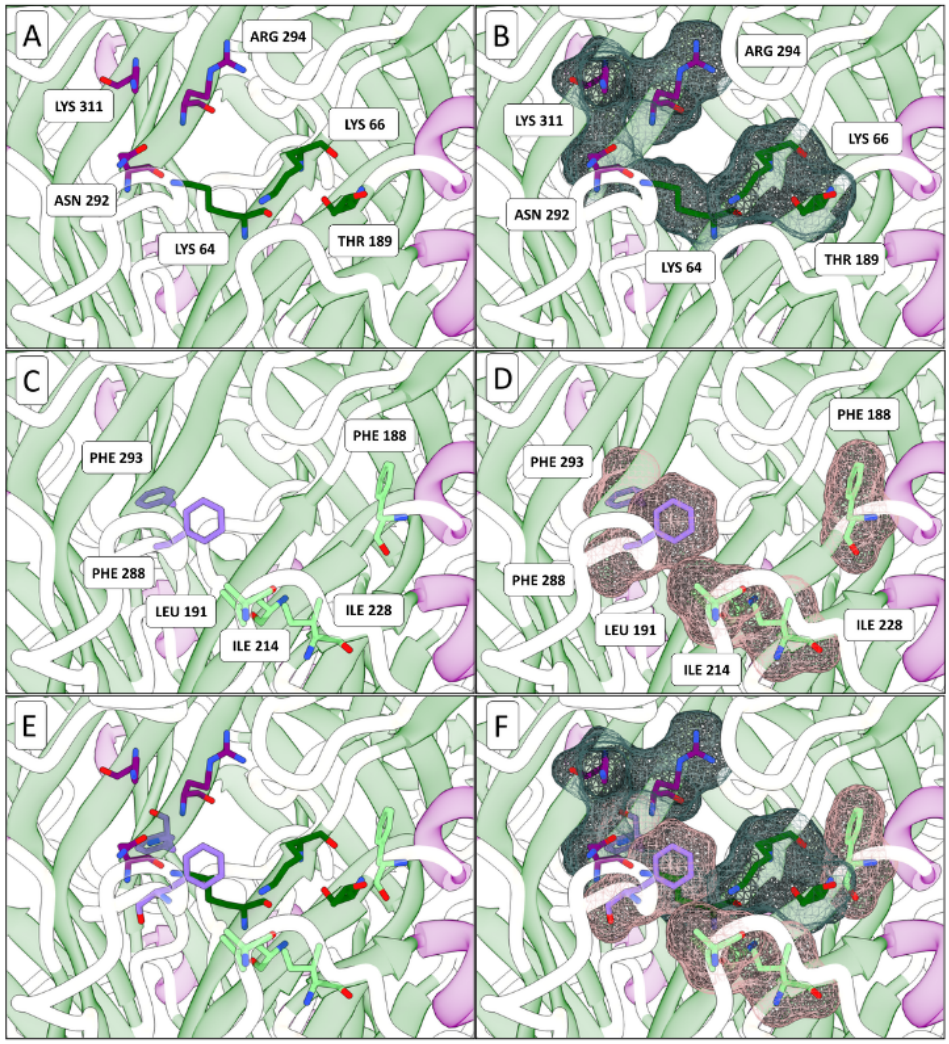


Figure 4.6: Close-up view of the amino acids forming the ATP binding pocket and the associated amino acids. (A) The amino acids forming the ATP binding pocket are shown in dark green (Lys64, Lys66 and Thr189) from monomer A, and in purple (Gln292, Arg294 and Lys311) from an adjacent monomer B. B) Representation of the surface of the amino acids forming the pocket. C) Supporting amino acids shown in light green from monomer A (Phe188, Leu191, Ile214 and Ile228), and in light purple from an adjacent monomer B (Phe293 and Phe288). D) Representation of the surface of the supporting amino acids. E) Amino acids forming the ATP binding pocket and related amino acids in stick form, colored as in the previous panels. F) Surface of amino acids forming the ATP binding pocket and related amino acids.

In addition to the ATP-binding pocket, another pocket can be identified in the receptor structure to which inhibitors and small drug molecules bind. The drug-binding pocket (also called the allosteric binding pocket) is located between neighboring subunits, adjacent to the ATP-binding site. The following amino acids are involved in forming the drug-

binding cavity: Phe88, Phe95, Phe103, Met105, Phe108, Phe295, and Val312 [516, 517]. The site is also shown in the Figure 4.6.

The P2X7 receptor is highly expressed in glioma cells. In the tumor microenvironment, elevated extracellular ATP levels activate it. Stimulation of the P2X7 receptor enhances glioma cell proliferation and increases cell survival. Activation of the receptor may influence the development of adaptive mechanisms that condition cancer cells to resist apoptosis (independent of the P2X7 receptor). The inhibition of P2X7 has been confirmed to decrease glioma proliferation [518–521]. However, the receptor expressed on glioma cells can still form cell pores. Under certain conditions, prolonged activation and long-lasting opening of the P2X7 receptor can lead to the apoptosis of cancer cells. This phenomenon is tied to the P2X7's dual functionality. Depending on the intensity and duration of its activation, as well as the surrounding cellular and molecular context, it can mediate either cell survival or cell death [493, 510, 522]. Its role in tumor growth has made P2X7 a potential therapeutic target. These therapies could block its activity or selectively induce apoptosis in glioma cells by prolonging its activation. The second approach presents two challenges. First, long-term activation of P2X7 receptors can damage healthy cells on whose surfaces these receptors are located. Second, glioma cells can adapt by reducing P2X7 expression [510, 523, 524].

Long-term P2X7 activation, leading to pore opening, is a promising strategy for inducing apoptosis in glioma cells. However, careful modulation is required to exploit its apoptotic effects while minimizing damage to normal tissues. A combined approach may be the solution. This chapter explores photoactive ligands that could reduce the ATP concentration required for receptor activation, allowing for selective anti-tumor activity while mitigating off-target toxicity.

4.1.4 Light-Induced Protein Regulation

Light-induced protein regulation uses a photosensitive ligand and light as external stimuli to precisely and controllably modulate protein activity. Early research on this topic included optogenetics, a technique that uses genetically encoded light-sensitive proteins. More recent developments, however, focus on external factors, such as small light-sensitive molecules. Light is a unique tool because it is minimally invasive (short exposures do not cause tissue damage). Light can be precisely controlled for localization, wavelength, intensity, and exposure duration, enabling real-time control of protein activity.

However, applying these methods to ligand-gated ion channels, such as P2X7, requires the use of well-designed, light-responsive compounds [525–530]. This part of the study

focuses on molecules that specifically bind to the ATP-binding pocket of the P2X7 receptor, modulating its activity in response to light. This could lead to the selective apoptosis of glioma cells by forming cell pores. The studies presented used naturally occurring compounds and commercially available drugs with photoactive potential. Retinal derivatives were also tested because these compounds had previously been used by the research team for cell differentiation [531]. Artificial compounds, which are a combination of natural photoactive compounds and BzATP, were also tested (see Table 4.A0.9 for the structures of the compounds).

4.1.5 Purpose and Scope of the Chapter

This chapter integrates fundamental knowledge of the nervous system and glial biology with the emerging field of light-regulated proteins studied using computational methods. The focus is on the P2X7 receptor as a potential therapeutic target and its possible photoactive ligands. This section primarily focuses on exploring the potential of light-induced regulation of the P2X7 receptor for treating glioma. To address this knowledge gap, the affinity of photoactive ligands for the receptor was investigated by comparing it with that of native P2X7 receptor ligands.

4.2 Materials and methods

4.2.1 Homology Modeling

As no structure of the human P2X7 receptor was available in the Protein Data Bank (PDB), a homology modeling approach was used. Complete structures of the rat (*Ratus norvegicus*) P2X7 receptor in the form of a closed and an open channel, determined by the cryo-EM technique, were available in the PDB [514]. However, the amino acid sequences of the dominant A isoform of the human P2X7 receptor (UniProtKB ID: Q99572, GeneBank: CAA73360) and the rat P2X7 receptor (UniProtKB ID: Q64663, GeneBank: CAA65131) are available.

A comparison of the amino acid sequences of the two receptors was made using the UniProt Align tool [532], and the Percent Identity of the alignment was 80.34%. The results of the comparison are shown in Figure 4.7 below (comparison of sequences in the text version made using the TEXshade package for LaTeX [533]). Due to the high level of similarity between the two sequences, the sequence of the human receptor and the rat structures (PDB: 6U9V, the closed state of the receptor, and PDB: 6U9W, the open state of the receptor [514]) were used as templates to build the hP2X7R homology model. For this purpose, the SWISS-MODEL [179] automated server for the homology modeling of protein structures was used. Hydrogen atoms were added to the resulting structures using Maestro [534, 535], a Schrödinger modeling environment, to set the system pH to 7. The quality of the generated protein structure was checked with the Procheck [536] server. Only a few residues from loop regions far from the important binding sites in this study were found to be slightly beyond a standard deviation.

Due to the growing popularity of AlphaFold2 [176], it was decided to also check the quality of the receptor model generated using this tool. For this purpose, ColabFold [537] implemented in UCSF ChimeraX [538] was used to generate the structure of each of the three monomers based on its sequence from UniProtKB with ID: Q99572. Additionally, models were generated from amino acid sequences and PDB structures of the rat receptor in open and closed conformations. The structures generated using AlphaFold2 were aligned and compared with those available in the PDB and with the obtained homologous model. In further stages of the work, a homologous model was used.

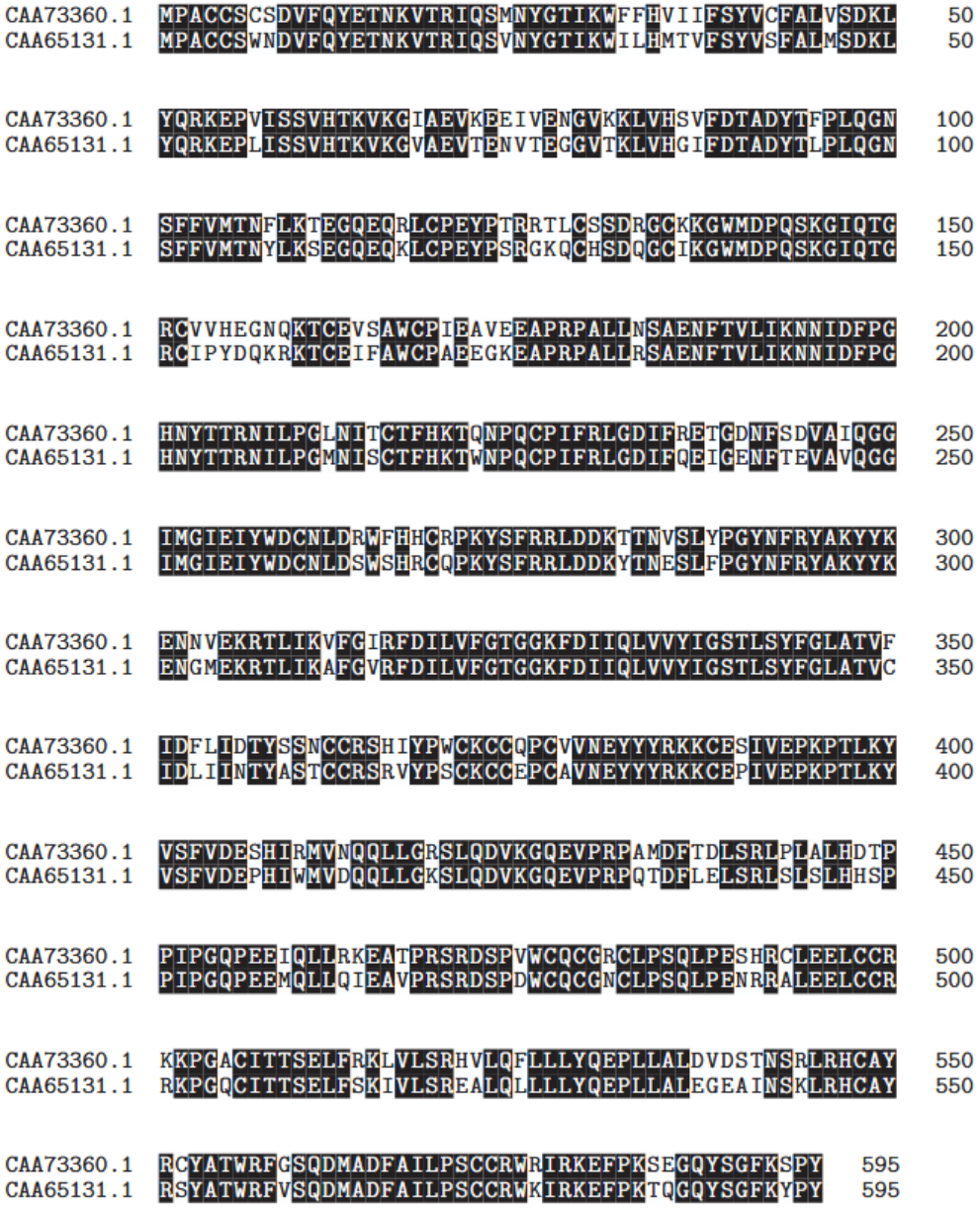


Figure 4.7: Comparison of the amino acid sequences of the dominant isoform A of human P2X7R (GeneBank: CAA73360, on top) and rat P2X7R (GeneBank: CAA65131, bottom). Identical amino acids are highlighted in black frames. Sequence comparisons were made using the TEXshade package for LaTeX [533].

4.2.2 Ligand Preparation

Table 4.1: List of ligands with PubChem CID numbers docked to P2X7 receptor divided into five groups.

Group	Ligand	PubChem CID with hyperlinks
1. known ligands	ATP	5957
	TMZ	5394
	MTIC	76953
	BzATP	115205
	AZ10606120	10310632
	ADP	6022
	AP4	14003
	AP5	56777378
2. natural photoactive compounds	Azobenzene	2272
	Stilbene	638088
	Pterostilbene	5281727
	Resveratrol	445154
3. commercial drugs	Agelasine	73440802
	Combretastatin A1	5458993
	Combretastatin A1 Phosphate	6918546
	Combretastatin A4	5351344
	Combretastatin A4 Phosphate	5351387
	Ombrabulin	6918405
4. retinoids	Retinal	638015
	Retinoic Acid	444795
5. photoactive derivatives of BzATP	Azobenzene + BzATP	
	Stilbene + BzATP	
	Pterostilbene + BzATP	
	Resveratrol + BzATP	

The compounds listed in Table 4.1 were prepared for molecular docking and divided into five groups based on structural similarity and properties. For this purpose, the geometry of the three-dimensional structure of the ligands from Group 1: ATP (adenosine triphosphate), TMZ (temozolomide), MTIC (3-methyl-(triazen-1-yl)imidazole-4-carboxamide), BzATP (2',3'-O-(4-benzoilobenzoilo)adenozyno-5'-trifosforan), AZ10606120 and derivatives of ATP, with are ADP (adenosine diphosphate), AP4 (adenosine tetraphosphate) and AP5 (P1,P5-Di(adenosine-5'-)pentaphosphate) (structures obtained from PubChem,

CID numbers given in the Table 4.1 all CID numbers are hyperlinked to their respective PubChem pages) was optimized using the Schrödinger modeling environment Maestro [534, 535].

The PubChem database had structures only in the trans form for the Group 2 ligand, Group 3 agelasine, and Group 4, while for the other Group 3 compounds, only structures in the cis state were available. As it was important to have the structures of these compounds in both forms, the available Simplified Molecular Input Line Entry System (SMILES) was used. The missing states were written in SMILES notation, and then the three-dimensional structures of the mentioned compounds were posed using the webpage www.cheminfo.org. Compounds 5 are hybrids modified by photoactive parts (ligands from group 2) of BzATP molecules. The molecules were designed using Maestro software tools in two states: trans and cis. The two-dimensional structures and SMILES for these compounds are provided in the Appendix. All prepared molecular structures of groups 2-5 in trans and cis states were optimized using the Schrödinger modeling environment Maestro [534, 535].

4.2.3 Molecular Docking

The prepared ligand structures were minimized using Schrödinger Maestro LigPrep. A total of 43 ligands were subjected to molecular docking using the SMINA code [539], a fork of AutoDock Vina [540] (which provides support for minimization and scoring of docked ligand positions), to the resulting homology model structure. These ligands included group 1 ligands in one state and ligands of other groups in trans and cis states. In the case of retinal, two cis forms were considered: 11-cis and 13-cis, and for retinoic acid, three cis forms were considered: 9-cis, 11-cis, and 13-cis. Since the exact binding site of the native ligand of this receptor is known, the docking site was restricted to the ATP-binding pocket, which is shown in detail in the Figure 4.6 (docking space created by: Lys64, Lys66, Thr189, Gln292, Arg294, Lys311). A total of 50 independent docking runs were performed for each ligand using the default settings, with up to 100 poses generated per run. The ligand with the highest score within the pocket was selected and prepared in Schrödinger Maestro [534, 535] by adding hydrogen atoms, then minimized to obtain the optimal conformation.

4.2.4 Data Analysis and Visualization

The bioinformatic analysis was performed using the UniProt Align tool [532]. The AlphaFold Error Plot tool [176, 541] implemented in USCF ChimeraX [538, 542] was used to analyze the structures generated with AlphaFold2. The USCF ChimeraX program was

also used to visually analyze and compare the structures generated by AlphaFold2 with homologous structures in the PDB. The UCSF ChimeraX *coulombic* command was used to calculate the Coulomb electrostatic potential in real time and color the molecular surfaces. Coulombic electrostatic potential was calculated using atomic partial charges and coordinates according to Coulomb's law: $\varphi = \sum[\frac{q_i}{\varepsilon d_i}]$, where φ is the potential, q are the atomic partial charges, ε is is the dielectric constant, and d denotes the distances from the atoms (by default 1.4 Å). The `mlp` command was also used to calculate molecular lipophilicity potential maps of the homologous model. The tables and graphs prepared for the analysis of docking results were created using Microsoft Excel [543], and the UCSF ChimeraX program was used to visualize the docking results.

4.3 Results

4.3.1 Homology Modeling

Three separate models were generated using AlphaFold2, based on their amino acid sequence (in the case of the homologous model Figure 4.8:A) and based on their amino acid sequence and structural seed information available in the PDB (in the case of the closed Figure 4.8:B and open Figure 4.8:C forms of the rat receptor structure). For each generated model, a sequence coverage plot, a predicted local distance difference test (pLDDT, AlphaFold's per-residue confidence score), and a predicted aligned error (PAE) graph were obtained.

On the sequence coverage plot, each horizontal line represents a sequence from the multiple sequence alignment (MSA), colored by sequence identity (blue for high similarity and red for low). The vertical axis shows the number of sequences investigated, and the horizontal axis shows the positions within the generated protein structure. The black line shows the MSA coverage for each residue. All three graphs in Figure 4.8 display the same key features: substantial coverage (orange) from ~1–450, a significant drop in coverage from ~450–520, and enhanced coverage from ~520–595. The black line (MSA depth per position) indicates that there are no significant differences between the Figure 4.8 panels. Large peaks occur between residues ~50-450, meaning that this region is well conserved and that many sequences cover this position. Dips around residues ~450–520 correspond to regions of lower confidence in the structure and/or flexible regions in the 3D model. This may be explained by the fact that some of the P2X7 receptor structures deposited in the PDB lack the intracellular region.

pLDDT is a per-residue confidence score produced by AlphaFold that estimates the reliability of the predicted atomic positions in a protein structure. All three 3D structures colored by pLDDT in Figure 4.8 show blue helical elements, which means high confidence folded cores, yellow/green parts mostly localized in the region at the edge of the membrane on the intracellular side, and a few red loops with low confidence due to high flexibility. There are no visible differences between models B and C, which were seeded with information from two different PDB structures. The overlapped structures generated using AlphaFold2 are shown in Figure 4.A0.28 in the Appendix. In contrast, model A (the human receptor) shows less confidence in the extracellular loops, as well as structural differences in the membrane edge on the intracellular side, compared to models B and C.

The overall pattern of the PAE heatmaps for the three models is very similar. It is identical for models B and C, and exhibits minor changes for model A: the red areas intersecting the blue blocks are narrower. In summary, according to AlphaFold2, all three models show

high consistency in terms of domain localization.

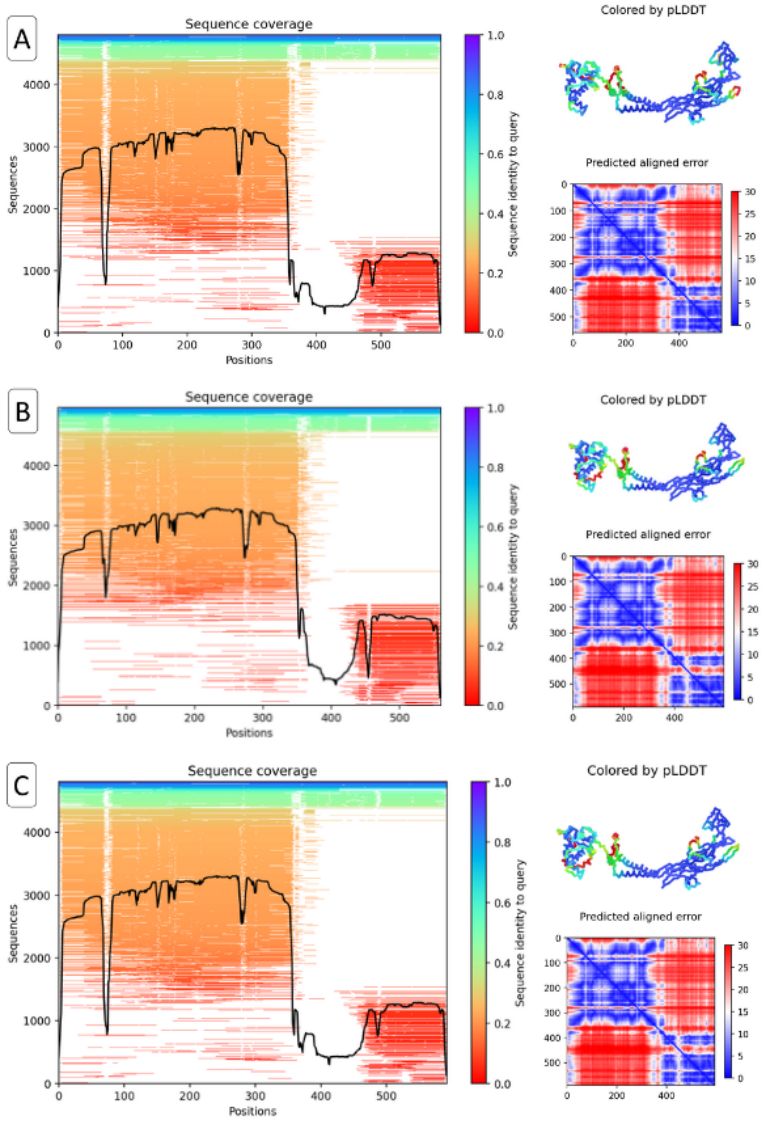


Figure 4.8: AlphaFold2 prediction results pLDDT (Predicted Local Distance Difference Test) color scheme is blue for very high (>90) confidence, cyan/green for medium (70–90) confidence, and yellow/orange/red for poor (<70) confidence. A) AlphaFold2 prediction results for monomer of homology model developed based on the human amino acid sequence and the structure of the rat receptor. B) AlphaFold2 prediction results monomer of closed rat model PDB: 6U9V, and C) for monomer of open rat model PDB: 6U9W

Figure 4.9 shows the alignment structures. Figure 4.9:A shows the structure generated by AlphaFold2 in purple and the homologous model obtained in this work in dark green (RMSD between all 592 pairs: 14.74 Å). A visual comparison of the two models reveals a significant inconsistency, namely a partial shift in the intracellular portion of the receptor. The shift is also visible in the lower part of the transmembrane α-helices. The consistency in the extracellular part is much higher, with differences appearing in the highly flexible loops.

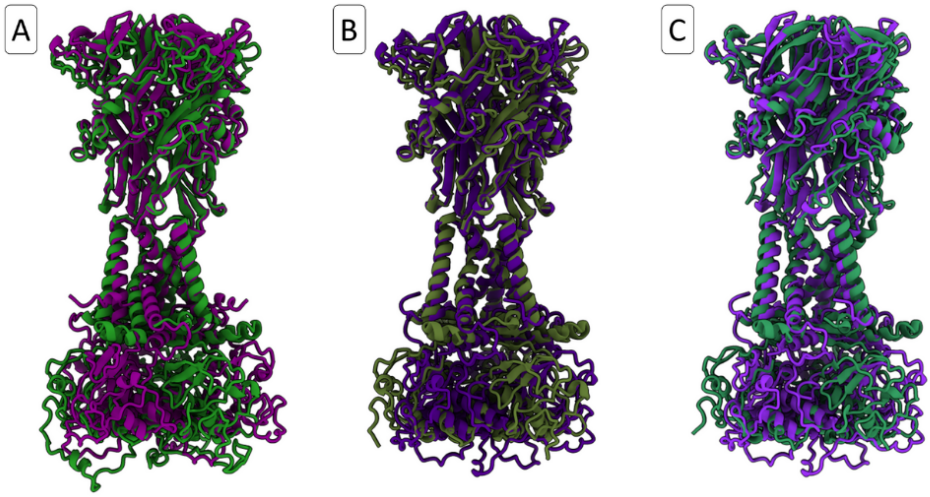


Figure 4.9: Comparison of models generated using AlphaFold2 with the homologous model and models available in PDB. A) Superimposed structure generated using Alphafold2 in purple and the structure of the homologous human receptor model developed based on the human amino acid sequence and the structure of the rat receptor in green, B) Superimposed structure generated using Alphafold2 in indigo and the structure of the closed rat receptor PDB: 6U9V in olive green. C) Superimposed structure generated using Alphafold2 in dark magenta and the structure of the open rat receptor PDB: 6U9W in see green.

Figure 4.9:B and Figure 4.9:C show the alignment of the structure with AlphaFold2 (in the shades of purple) and the closed (RMSD 13.35 Å) and open rat receptor (RMSD 14.57 Å) from PDB (in the green shades), respectively. A visual comparison of the two green structures reveals significant differences between the closed and open structures. The transmembrane fenestrations of the open receptor are significantly larger than in the closed state, and conformational changes are also visible in the extracellular region, resulting from the presence of the ligand in the ATP-binding pockets of the open receptor. In contrast, the structures generated by AlphaFold2 show no differences between these two states, based on the initial information from the available PDB structures. When comparing AlphaFold2 models with PDB models, the largest shifts are also observed in the

intracellular region. The high uncertainty in AlphaFold2's prediction of this region may be due to the lack of complete structures of P2X family receptors in databases, as most available structures lack the intracellular region. The greatest similarity to the structure generated by AlphaFold2 is visible in the PDB structure of the closed rat receptor, which can be explained by the high availability of closed P2X family receptor structures in databases.

All results presented in the following sections of this paper were obtained using a homology model, as the work began before the rapid development of AlphaFold2. The choice of a homology model is further justified by the fact that the reliability of the generated structures may be questionable.

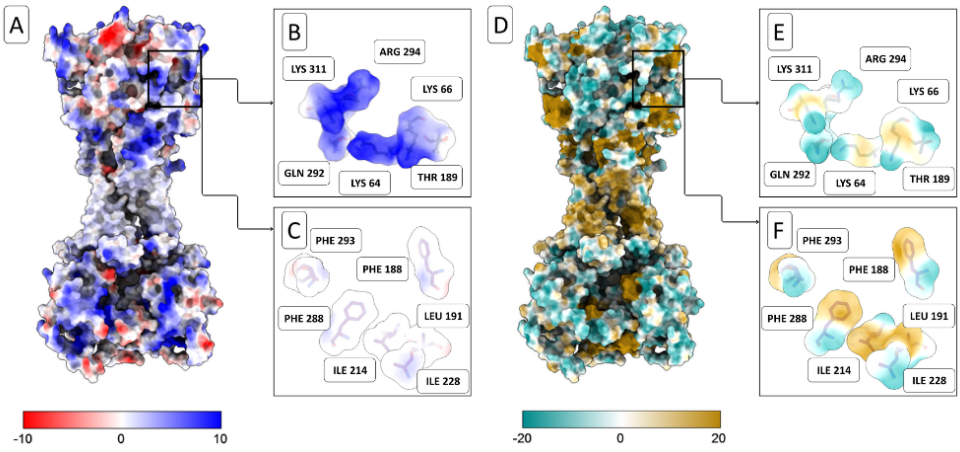


Figure 4.10: Analysis of the features of the human P2X7 receptor homologous model. A) Coulombic electrostatic potential with coloring ranging from red for negative potential through white to blue for positive potential. The result of Coulombic electrostatic potential is in units of kcal/(mol·e) at 298 K. B) Close-up view of Coulomb electrostatic potential map of the surface of ATP binding pocket. C) Close-up view of Coulomb electrostatic potential map of the surface of the supporting amino acids. D) Lipophilicity potential maps the results are shown with coloring on the molecular surface of the protein ranging from cyan (most hydrophilic) to white to amber (most lipophilic). E) Close-up view of lipophilicity potential map of the surface of ATP binding pocket. C) Close-up view of lipophilicity potential map of the surface of the supporting amino acids.

An analysis of the characteristics of the homologous model of the human P2X7 receptor, including Coulombic electrostatic potential and lipophilicity, was performed using ChimeraX software tools and is presented in Figure 4.10. Figure 4.10:A shows a map of the Coulombic electrostatic potential of the entire receptor structure, with a neutral region visible in the transmembrane α-helices, while positive regions can be observed at the contact point between monomers in the extracellular region. This area contains the

ATP binding pocket, which is shown in Figure 4.10:B. The blue area represents a positive potential, which enables the binding of negatively charged ATP molecules or other ligands. The amino acid residues surrounding the binding site shown in Figure 4.10:C are neutral, suggesting that their role in stabilizing the ligand is not related to the presence of charge.

Maps of the lipophilic potential of the entire receptor are shown in Figure 4.10:D, which shows a large, strongly hydrophobic region in the part where transmembrane helices are present. Figure 4.10:E,F presents a close-up of the lipophilic potential map of the amino acids comprising the ATP binding pocket and the accompanying amino acids, respectively. The ATP-binding pocket exhibits higher hydrophilicity, while the surrounding amino acids are strongly hydrophobic, suggesting that the difference in lipophilicity potential between the pocket amino acids and the supporting amino acids may affect the stabilization of ligands during binding.

In the final stages of this dissertation, the authors published a paper presenting the cryo-EM structure of the human P2X7 receptor in the apo-closed state [544].

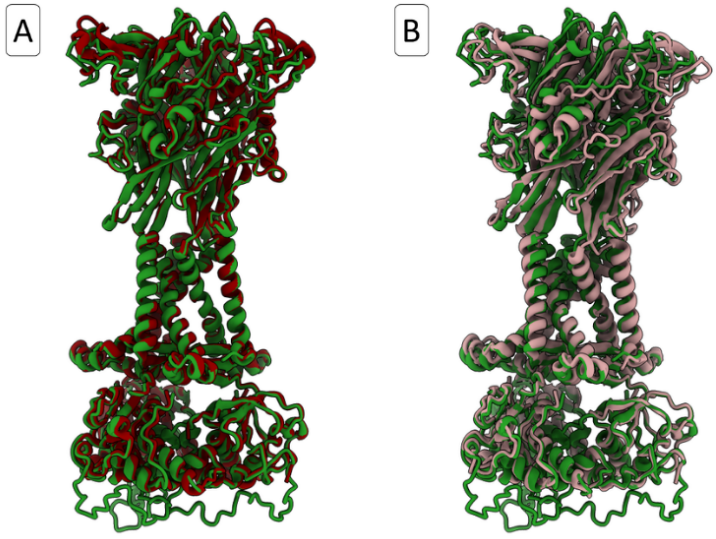


Figure 4.11: A) A comparison of the homology model generated for the thesis, shown in green, with the receptor in closed state PDB: 9E3M model [544], shown in maroon. B) A comparison of the homology model, shown in green, with the open structure of the receptor PDB: 9E3N model [544], shown in beige.

Figure 4.11 shows a comparison of the homology model generated for this thesis with the PDB: 9E3M model of the receptor in the closed state and with PDB: 9E3N open structure of the receptor. The homology model and the PDB structure of the closed receptor demonstrate a high level of similarity, as indicated by an RMSD of 0.819 Å for 483 atom pairs in the single chain. The PDB structure contains missing residues; therefore, the protein's amino

acid count does not match that of the homologous model. The missing residues are located mainly in the intracellular part of the protein. The structural differences between the homologous model and the open-state receptor are larger, with an RMSD of 0.980 angstroms over 304 atom pairs. Most of the changes occur in the transmembrane helix region and the intracellular part. This is because the homologous model is based on the closed receptor. The cryo-EM structure of the P2X7 receptor in the open state is missing residues, resulting in RMSDs for only 304 chain atoms and the largest changes occurring in the intracellular part of the protein.

4.3.2 Molecular Docking

For each of the 43 ligands, which are divided into five groups (see Table 4.1 for the list of ligands and the aforementioned division), 50 docking series were performed using the smina program. Each docking session generated up to 100 ligand positions in the ATP-binding pocket, to which the program assigned scores to evaluate them. The results presented in this section maintain the ligand grouping shown in Table 4.1. Docking results provide a static approximation of binding. The lowest score (the most negative) indicates the most accurate predicted ligand-protein docking position. Therefore, a high negative score corresponds to the strongest ligand binding in the binding pocket of the protein.

Table 4.2 shows the average results from 50 rounds of docking into the ATP binding pocket for the five ligand positions with the best results in Group 1. The appendix includes results for all positions from 50 docking rounds for each ligand. Figure 4.12 shows the score of each best position from 50 docking rounds in the form of dots, and the average score is shown as a dotted line (which is the value from the first row of Table 4.2). The docking results for ATP will serve as a reference point for further analyses, and the other results will be interpreted in relation to them. This reference has been chosen because ATP is a native ligand of the P2X7 receptor. BzATP exhibits a binding affinity approximately 1.0 kcal/mol higher than that of the native ATP receptor ligand, which has been confirmed experimentally [545]. In each of the 50 docking rounds, the positions of the best BzATP show better consistency and less deviation from the mean value (black in Figure 4.12) than those of the ATP (grey in Figure 4.12). However, both compounds are characterized by low smina scores, which are significantly better than those of the other ligands presented in the Table 4.2. This can be explained by the presence of a negatively charged phosphate chain in their structure that interacts strongly with the positively charged amino acids in the pocket. The smallest deviations from the mean value in Figure 4.12 are shown by TMZ (in purple). Despite the fact that the values oscillate around -6.5 kcal/mol — approximately 2.0 kcal/mol lower than the values for ATP — it can be concluded that this compound exhibits some affinity for the ATP binding pocket of the P2X7 receptor. This has been confirmed by the experimental team. The combined action of ATP and TMZ results in a 70% reduction in viability and reduced migration capacity of differentiated glioblastoma cells. This allowed the use of a low, non-toxic concentration of ATP and a reduced effective concentration of TMZ [531]. MTIC is an active metabolite and breakdown product of TMZ that does not exhibit high affinity for the ATP-binding pocket of the P2X7 receptor. These values do not exceed -6.0 kcal/mol and are significantly worse than ATP. The affinity of AZ10606120 (AZ), an experimentally confirmed P2X7 receptor antagonist [545], was tested to better determine the binding strength of TMZ and photoactive ligands. AZ shows values oscillating

around -7.6 kcal/mol, which are worse than those obtained with ATP but better than those obtained with TMZ.

Table 4.2: Average score of the 50 docking rounds to the ATP binding pocket for the top five positions for Group 1 ligands, smina score (predicted binding energy in kcal/mol)

ATP	BzATP	TMZ	MTIC	AZ
-8.32	-9.66	-6.46	-5.42	-7.59
-8.20	-9.58	-5.69	-5.14	-7.54
-8.11	-9.50	-5.64	-5.05	-7.50
-8.06	-9.44	-5.58	-4.99	-7.48
-8.01	-9.40	-5.49	-4.93	-7.44

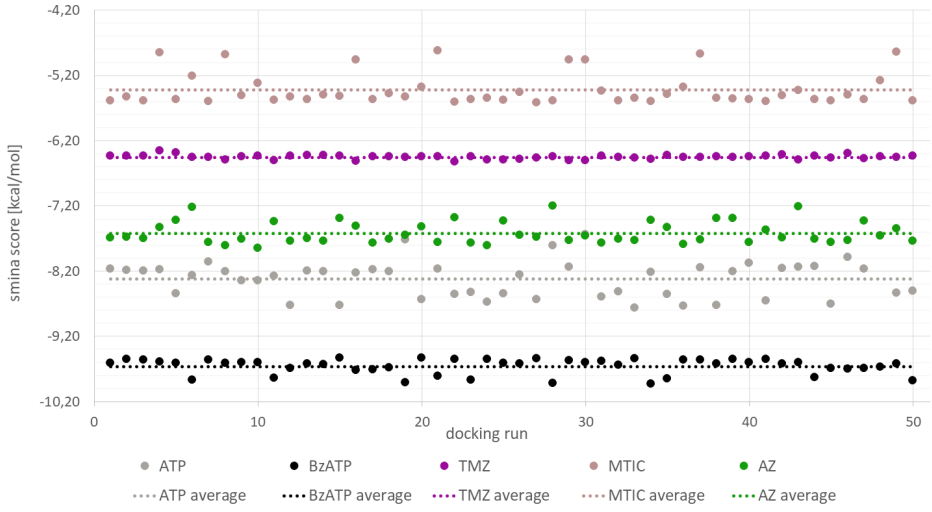


Figure 4.12: The graph shows the results for the best positions from each of the 50 docking rounds for ligands from Group 1 in the form of points, and the average value (shown in Table 4.2) as a dotted line in the color corresponding to the points. ATP in gray, BzATP in black, TMZ in purple, MTIC in beige and AZ in forest green.

Since ATP is treated as a reference, the docking results were compared with ATP present in the PDB model 6U9W, visualizing both positions in Figure 4.13. The difference in ATP positions within the receptor pocket is due to the conformational difference between the two models. It can therefore be concluded that the conformational change within the pocket, caused by ligand binding, allows movement of the ligand and changes to the network of hydrogen bonds that stabilize its position.

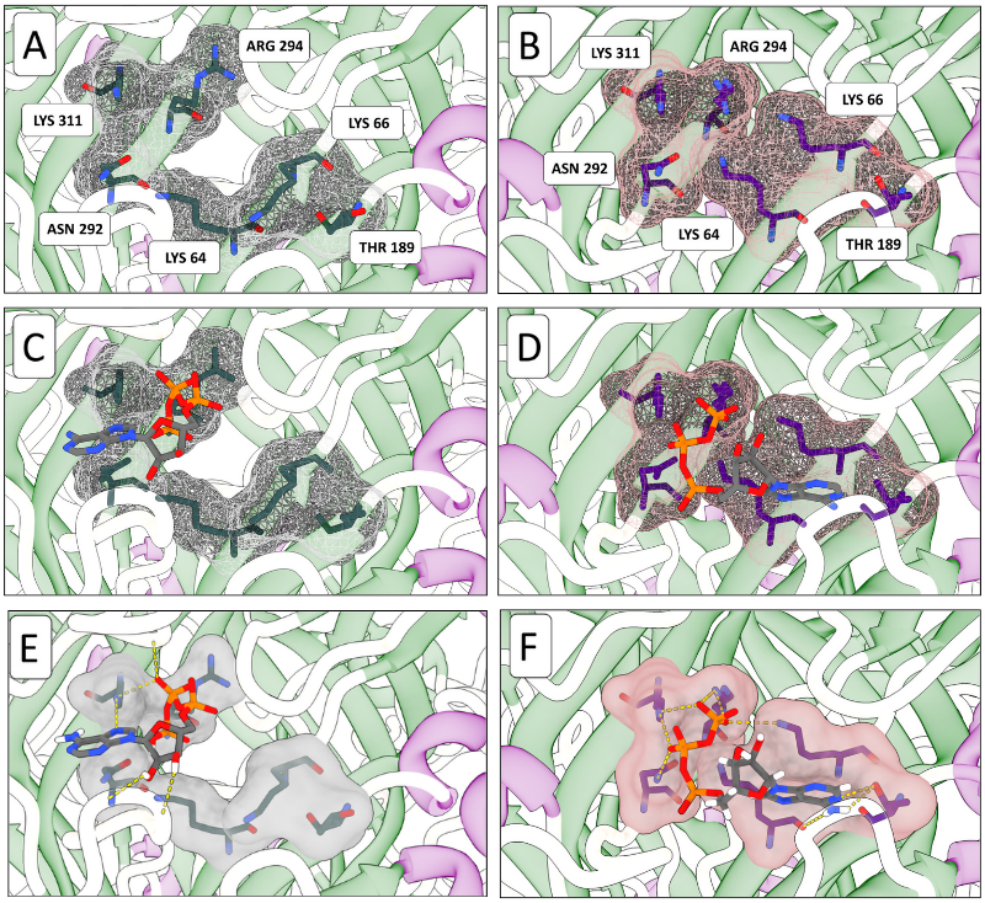


Figure 4.13: Visualization of A) the ATP binding pocket with labeled amino acids and C) the best ATP docking position in the homologous closed P2X7 receptor model structure. B) ATP binding pocket D) ATP position in the pocket in the open rat receptor structure PDB: 6U9W. E) Hydrogen bond pattern between ATP and amino acids in homology model shown as yellow dash-line. F) Hydrogen bond pattern between ATP and amino acids in the open rat receptor structure PDB: 6U9W shown as yellow dash-line.

Figure 4.14 shows the comparison of the ATP-binding pocket architecture and surface of the ATP-binding pocket of the homology model and ATP-binding pocket of the Cryo-EM structure of the human P2X7 receptor in the ATP-bound open state PDB: 9E3N. Comparing the human receptor pocket with the rat receptor ATP binding pocket shown above in Figure 4.13:D shows that these places in the structure of the receptors in the open state are very similar. As in the previous comparison, the conformational change within the pocket may be caused by ligand binding.

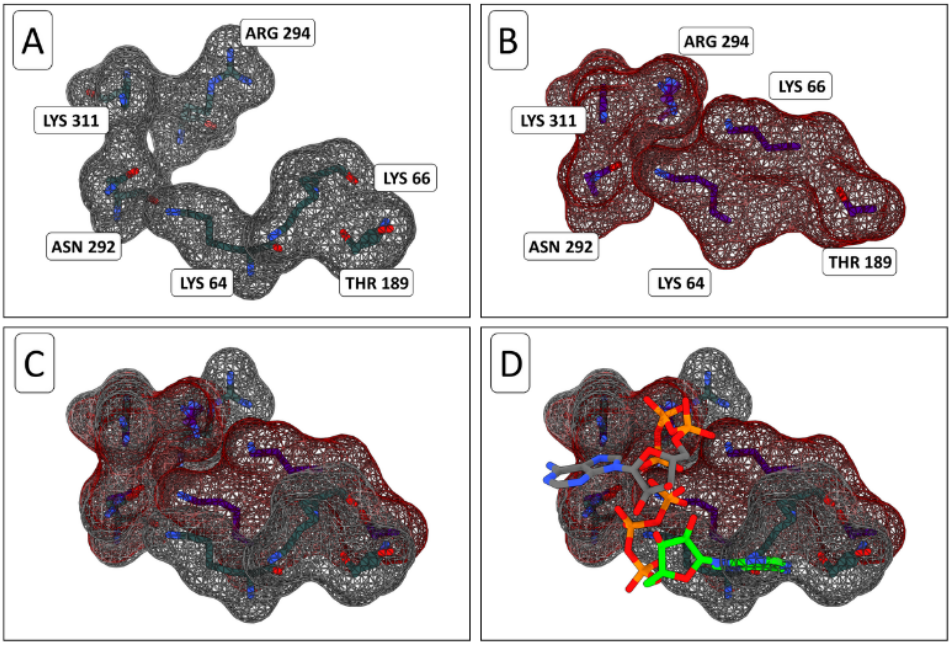


Figure 4.14: A) Isolated ATP-binding pocket of the homology model, mesh surface in gray. B) Isolated ATP-binding pocket of the Cryo-EM structure of the human P2X7 receptor in the ATP-bound open state PDB: 9E3N, mesh surface in maroon. C) Overlap of the two pockets. D) The ATP binding pockets of both models are shown, with the ATP visible. The ATP resulting from docking is shown in grey and the ATP from the cryo-EM structure is shown in lime.

Figure 4.15 shows the visualization of the best positions (positions with the lowest smina score) for each ligand from Group 1. For clearer visualization and easier interpretation, the structures of amino acids and ligands have been visualized without hydrogen atoms. Additionally, Figure 4.15:A represents an empty binding pocket with labels of the amino acids that form it. The colors of carbon atoms in individual ligands correspond to their color representation in Figure 4.12. All compounds fit well into the binding pocket. ATP (Figure 4.15:B in gray) and BzATP (Figure 4.15:C in black) interact with the negatively charged phosphate chain and the positively charged surface of the pocket. Ligands such as TMZ (Figure 4.15:D

in purple) and MTIC (Figure 4.15:E in beige) penetrate deep into the pocket cavity and potentially have a high degree of freedom of movement there due to their small size and lack of charge. In the case of AZ (Figure 4.15:F in green), the adamantane part of the compound is located outside the binding pocket area. Interaction with the pocket's amino acids occurs via the rings present in the AZ structure. The adamantane moiety is likely involved in hydrogen bonding with the pocket environment.

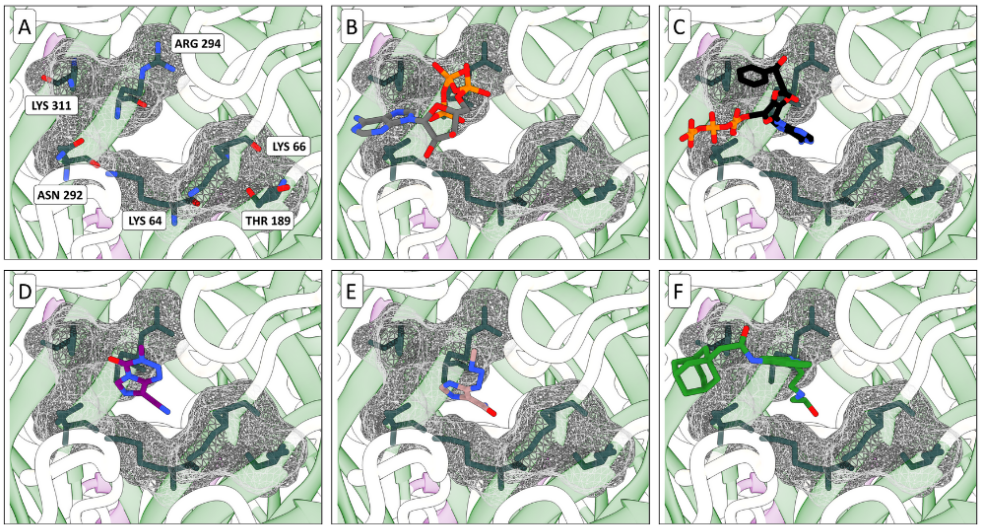


Figure 4.15: Visualization of the docking results for group 1 ligands with the highest smina score. To improve clarity, hydrogen atoms have been omitted from both the amino acid structures and the ligands with structures are colored by heteroatoms. A) The structure of the ATP binding pocket is shown again for convenience of analysis and interpretation. Surface of the pocket shown as a mesh. B) ATP in gray. C) BzATP in black. D) TMZ in purple. E) MTIC in beige. F) AZ in forest green.

Docking of ATP derivatives with varying numbers of phosphate residues was performed to examine their effect on the binding affinity of the compound in the receptor pocket. The results are presented in the Table 4.3 and in the Figure 4.16. Adenosine diphosphate (ADP), a derivative with one less phosphate group than ATP, exhibits a slightly weaker binding force, represented by a higher value of approx. 0.4 kcal/mol. In the case of adenosine tetraphosphate (AP4), a derivative with one more phosphate group than ATP, the smina score results show no significant difference between AP4 and ATP. P1,P5-Di(adenosine-5'-)pentaphosphate (AP5), an ATP derivative containing five phosphate groups and two adenine rings, exhibits high affinity for the binding pocket, represented by a smina score approximately 1 kcal/mol lower than that of ATP. However, the results for AP5 show large deviations from the mean (Figure 4.16 in light green). Based on the presented results, it can

be concluded that the number of phosphate residues may affect the binding strength of compounds in the ATP-binding pocket of the P2X7 receptor. This can also be explained by the negative charge of the ATP-binding pocket surface.

Table 4.3: Average score of the 50 docking rounds to the ATP binding pocket for the top five positions for ATP, BzATP, and additional ligands that are ATP derivatives, with varying numbers of phosphate groups, smina score in kcal/mol

ATP	BzATP	ADP	AP4	AP5
-8.32	-9.66	-7.88	-8.25	-9.38
-8.20	-9.58	-7.78	-8.06	-9.13
-8.11	-9.50	-7.73	-7.99	-8.95
-8.06	-9.44	-7.70	-7.92	-8.78
-8.01	-9.40	-7.67	-7.83	-8.66

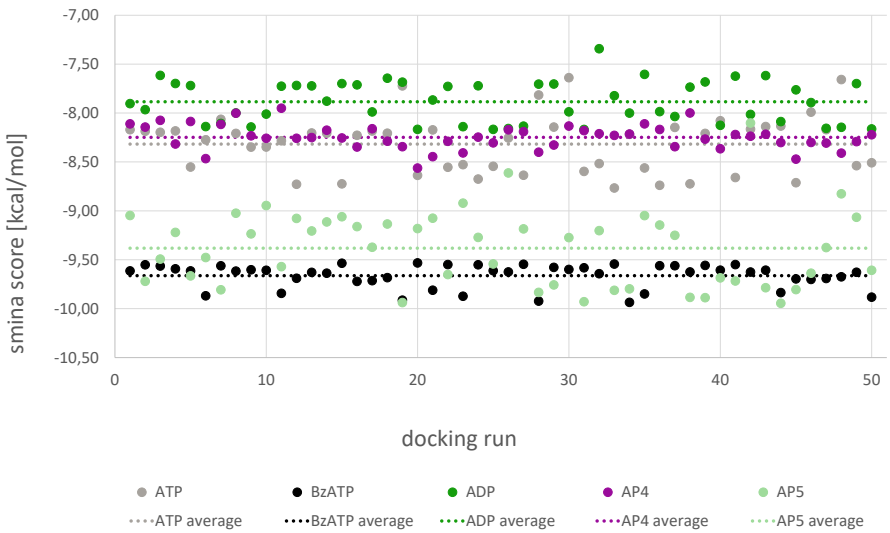


Figure 4.16: The graph shows the results for the best positions from each of the 50 docking rounds for ATP, BzATP, and additional ligands that are ATP derivatives, with varying numbers of phosphate groups in the form of points, and the average value (shown in Table 4.3) as a dotted line in the color corresponding to the points. ATP in gray, BzATP in black, ADP in forest green, AP4 in purple and AP5 in light green.

The first panel of the Figure 4.17:A shows the structure of the pocket and the amino acid labels that make up the pocket. Figure 4.17:B and Figure 4.17:C show the previously presented positions of ATP (in gray) and BzATP (in black), respectively, to facilitate comparison of the positions of these compounds in the receptor pocket with other ATP derivatives. The ADP Figure 4.17:D (in forest green) position is similar to the way ATP is arranged in the recep-

tor pocket. By contrast, despite having a greater number of phosphate residues, AP4 Figure 4.17:E (in purple) and AP5 Figure 4.17:F (in light green) arrange themselves in the pocket cavity in a manner similar to BzATP. This is counterintuitive because strongly negatively charged phosphate residues should be drawn toward the positively charged pocket. This may be caused by the presence of a network of hydrogen bonds holding the ligands in this position. This topic requires further investigation.

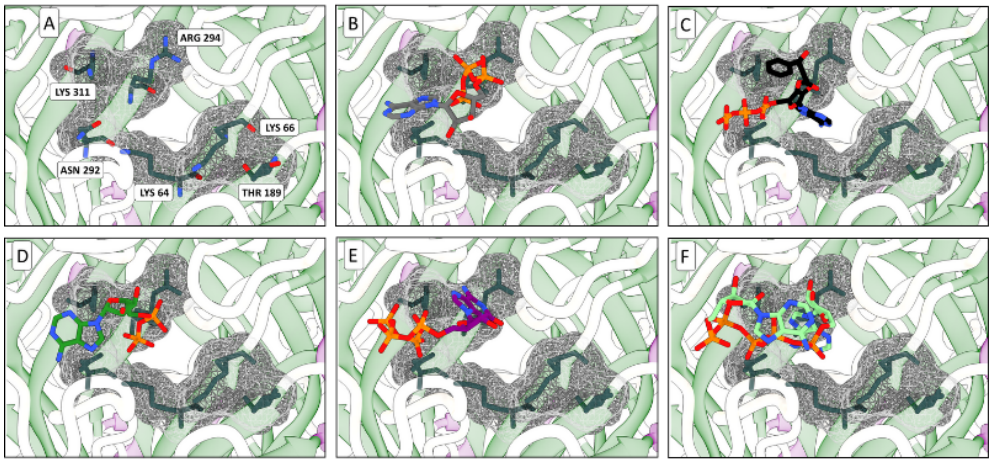


Figure 4.17: Visualization of the docking results for ATP, BzATP, and additional ligands that are ATP derivatives, with the highest smina score. To improve clarity, hydrogen atoms have been omitted from both the amino acid structures and the ligands with structures are colored by heteroatoms. A) The structure of the ATP binding pocket is shown again for convenience of analysis and interpretation. Surface of the pocket shown as a mesh. B) ATP in gray. C) BzATP in black. D) ADP in forest green. E) AP4 in purple. F) AP5 in light green.

Table 4.4: Average score of the 50 docking rounds to the ATP binding pocket for the top five positions for Group 2 ligands

Azobenzene		Stilbene		Pterostilbene		Resveratrol	
trans	cis	trans	cis	trans	cis	trans	cis
-5.65	-5.76	-5.70	-5.81	-6.08	-6.58	-6.95	-6.72
-5.63	-5.75	-5.68	-5.80	-6.04	-6.54	-6.82	-6.69
-5.61	-5.69	-5.54	-5.73	-6.00	-6.29	-6.78	-6.62
-5.58	-5.52	-5.37	-5.56	-5.97	-6.22	-6.74	-6.56
-5.53	-5.31	-5.30	-5.39	-5.94	-6.14	-6.69	-6.50

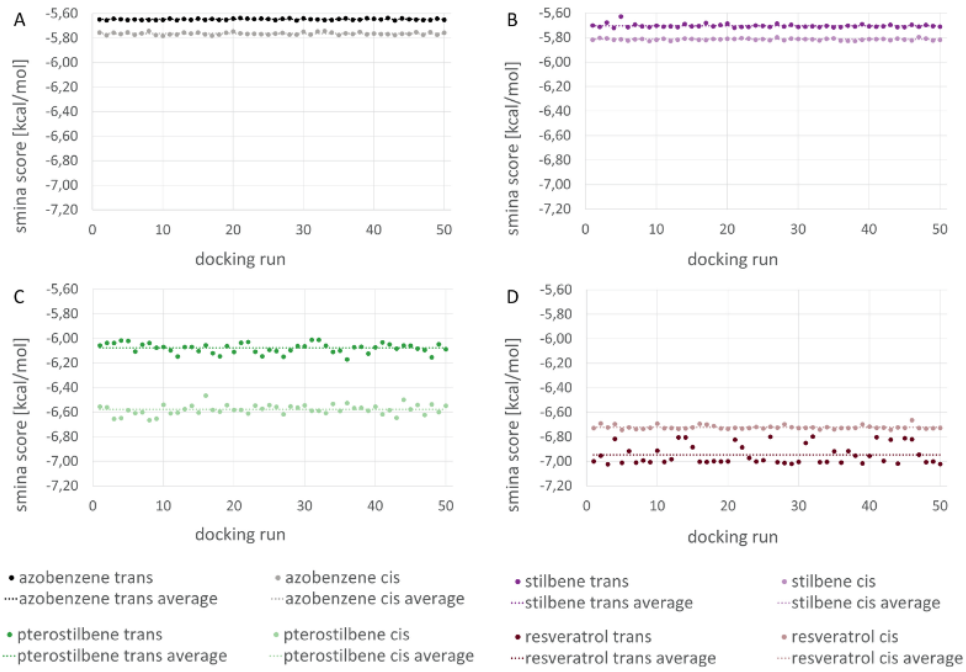


Figure 4.18: The graphs show the results for the best positions from each of the 50 docking rounds for ligands from Group 2 in the form of points, and the average value (shown in Table 4.4) as a dotted line in the color corresponding to the points. A: azobenzene in *trans* in black, in *cis* in gray. B: stilbene in *trans* in purple, in *cis* in lavender. C: pterostilbene in *trans* in forest green, in *cis* in light green. D: resveratrol in *trans* in maroon, in *cis* beige.

A series of docking studies was performed using ATP as a reference (-8.32 kcal/mol average score for the best docking pose across 50 docking rounds) to test the potential use of photoactive compounds as P2X7 receptor ligands. Table 4.4 and Figure 4.18 show the results for two states (*trans* and *cis*) of photoactive ligands of Group 2, commonly known and used compounds that exhibit photoisomerization. Azobenzene and stilbene show minor deviations from the average (as shown in Figure 4.18:A and Figure 4.18:B) and a poor result in terms of ATP score (the difference is approximately 3 kcal/mol). Additionally, in these cases, there is no significant difference between the *trans* and *cis* states. In both cases, however, the *cis* state shows a slightly better score of approximately 0.1 kcal/mol than the *trans* state. Better results were observed in the case of pterostilbene (shown in Figure 4.18:C). Although the score is weaker than that of ATP, the result for the *cis* state is comparable to that obtained for TMZ docking. Pterostilbene exhibits the most significant differences among the Group 2 ligands between the *trans* and *cis* states, with an approximate difference of 0.5 kcal/mol. For this reason, this compound was selected for further study. Although resveratrol showed slightly worse results than pterostilbene, it was also selected for further in-depth studies because it is the only ligand in Group 2 that exhibits better results for *trans* state (as shown in Figure 4.18:D).

Figure 4.19 shows visualisations of the best docking positions for group 2 ligands. One of the *trans* azobenzene rings is located in the proximity of Arg294 (Figure 4.19:A), while the other is positioned between Asn292 and Lys311. Conformational change to the *cis* form causes the ring to move away from Asn292 and Lys311 (see Figure 4.19:B). The *trans* position of azobenzene differs from that of the other ligands, which consistently have one ring located inside the pocket cavity and the other positioned between Arg294 and Lys311 (Figure 4.19:C, Figure 4.19:E, and Figure 4.19:G). For ligands in the *cis* state, one of the rings remains in the pocket cavity. In contrast, the ligand that was located near the amino acids Arg294 and Lys311 in the *trans* state is moved away from them towards the outside (Figure 4.19:D, Figure 4.19:F, and Figure 4.19:H).

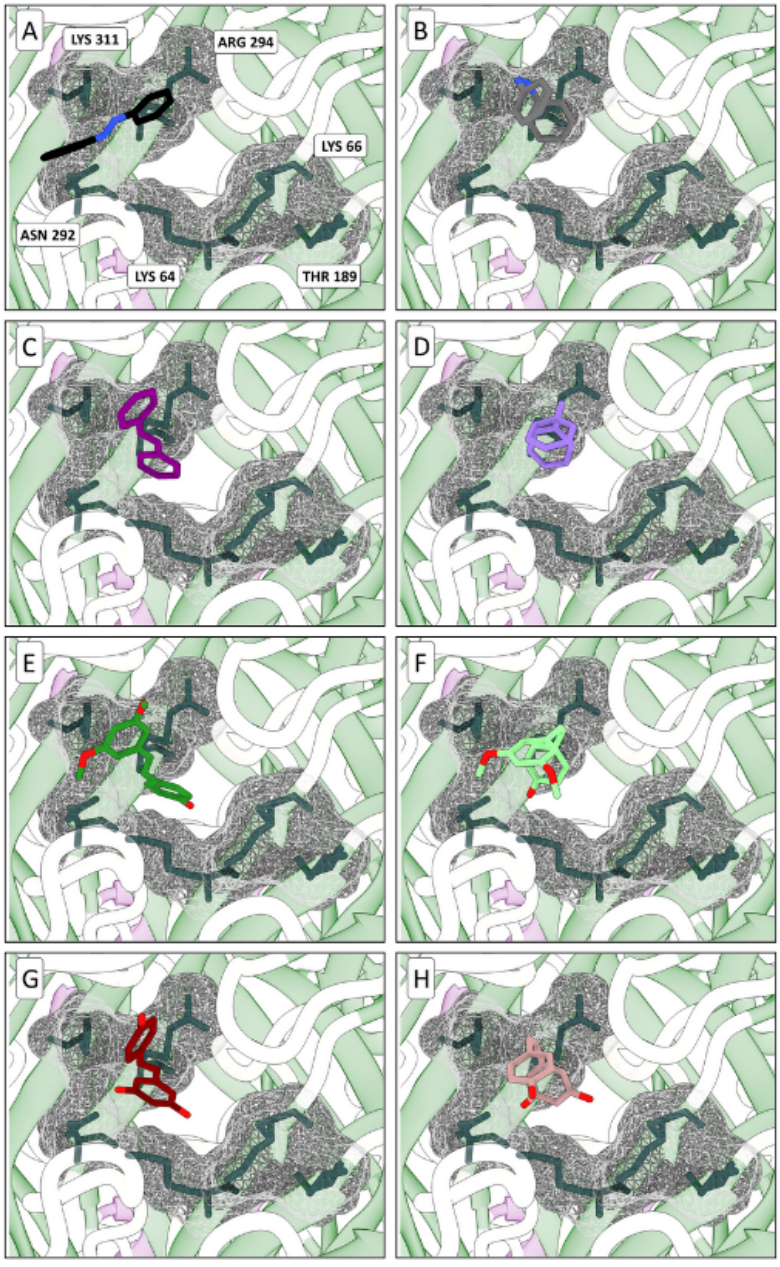


Figure 4.19: Visualization of the docking results for group 2 ligands with the highest smina score. To improve clarity, hydrogen atoms have been omitted from both the amino acid structures and the ligands with structures are colored by heteroatoms. A) Azobenzene in *trans* in black, B) in *cis* in gray. C) Stilbene in *trans* in purple, D) in *cis* in lavender. E) Pterostilbene in *trans* in forest green, F) in *cis* in light green. G) Resveratrol in *trans* in maroon, H) in *cis* in beige.

Table 4.5: Average score of the 50 docking rounds to the ATP binding pocket for the top five positions for Group 3 ligands

Agelasine		Combretastatin A1		Combretastatin A4		Ombrabulin	
trans	cis	trans	cis	trans	cis	trans	cis
-8.25	-8.16	-6.56	-6.89	-6.20	-6.65	-6.92	-6.94
-7.97	-8.01	-6.50	-6.85	-6.16	-6.57	-6.85	-6.87
-7.79	-7.91	-6.46	-6.76	-6.15	-6.48	-6.79	-6.73
-7.74	-7.80	-6.43	-6.73	-6.13	-6.44	-6.74	-6.68
-7.67	-7.71	-6.40	-6.69	-6.11	-6.42	-6.68	-6.62

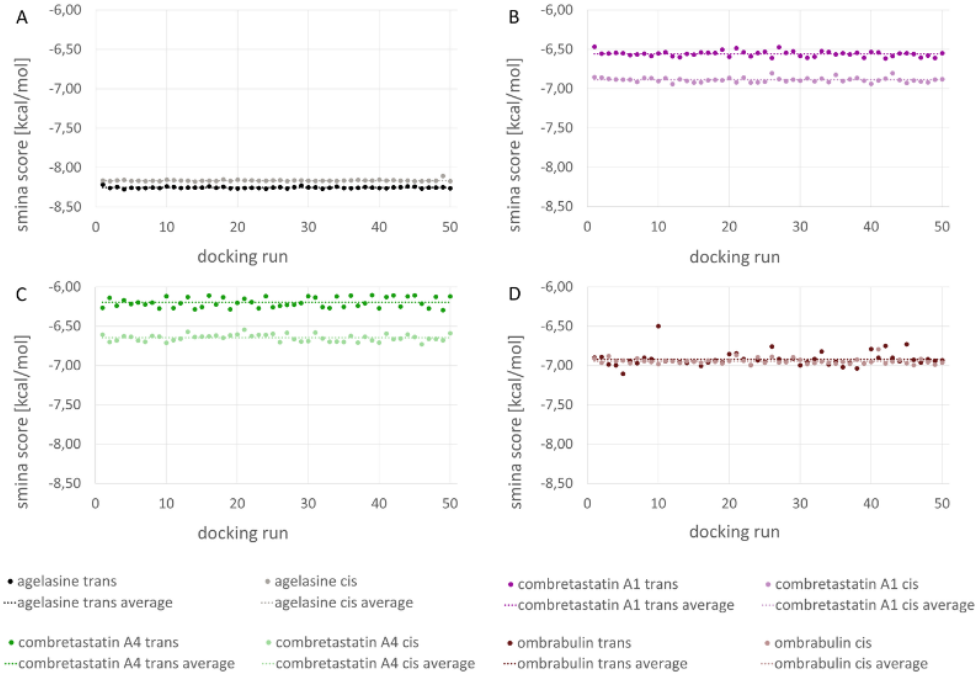


Figure 4.20: The graphs show the results for the best positions from each of the 50 docking rounds for ligands from Group 3 in the form of points, and the average value (shown in Table 4.5) as a dotted line in the color corresponding to the points. A: agelasine in *trans* in black, in *cis* in gray. B: combretastatin A1 in *trans* in purple, in *cis* in lavender. C: combretastatin A4 in *trans* in forest green, in *cis* in light green. D: ombrabulin in *trans* in maroon, in *cis* beige.

The results of the commercial docking of compounds with photoactive potential to the ATP-binding pocket of the P2X7 receptor are presented in Table 4.5 and Figure 4.20. The results for agelasine are very consistent, with negligible deviation from the mean; they are also closest to those obtained from ATP docking. Ombrabulin shows poorer results (approximately -7.0 kcal/mol), so its binding strength in the pocket can be compared with that of TMZ. It also has the largest deviation from the mean value among the ligands of Group 3. These results show no difference between the *trans* and *cis* states for agelasine (Figure 4.20:A, *trans* in black, in *cis* in gray) and ombrabulin (Figure 4.20:D, *trans* in maroon, in *cis* beige). The average scores for the top 50 docking positions of combretastatin A1 and combretastatin A4 are comparable to those of TMZ. Slight differences of approximately 0.4 kcal/mol are noticeable in the case of combretastatin A1 (Figure 4.20:B, *trans* in purple, in *cis* in lavender) and combretastatin A4 (Figure 4.20:C, *trans* in forest green, in *cis* in light green).

Figure 4.21 shows a visualization of the best docking positions for ligands from group 3. In the *trans* form (shown in Figure 4.21:A in black), both agelasine rings are located in proximity to Asn292, Arg294, and Lys311. The *cis* form (shown in Figure 4.21:B in gray) exhibits similar positioning to the *trans* form and does not demonstrate any significant differences. The trans states of combretastatin A1 (Figure 4.21:C), combretastatin A4 (Figure 4.21:E) and ombrabulin (Figure 4.21:G) exhibit consistent ring positioning near the amino acids Asn292, Arg294 and Lys311 and the second ring next to Lys66. After the ligand conformational change to the *cis* state in the case of combretastatin A1 (Figure 4.21:D), the position of the ring that was near Lys66 changes to one near Arg294, and the ring interacting in the *trans* state with Asn292, Arg294, and Lys311 is moved away from them towards the outside. Both combretastatin A4 (Figure 4.21:F) and ombrabulin (Figure 4.21:H) exhibit a shift of the ring from the Lys66 region towards Arg294 in the *cis* state, while the position of the second ring simultaneously changes slightly, remaining in the region of the amino acids Asn292 and Lys311.

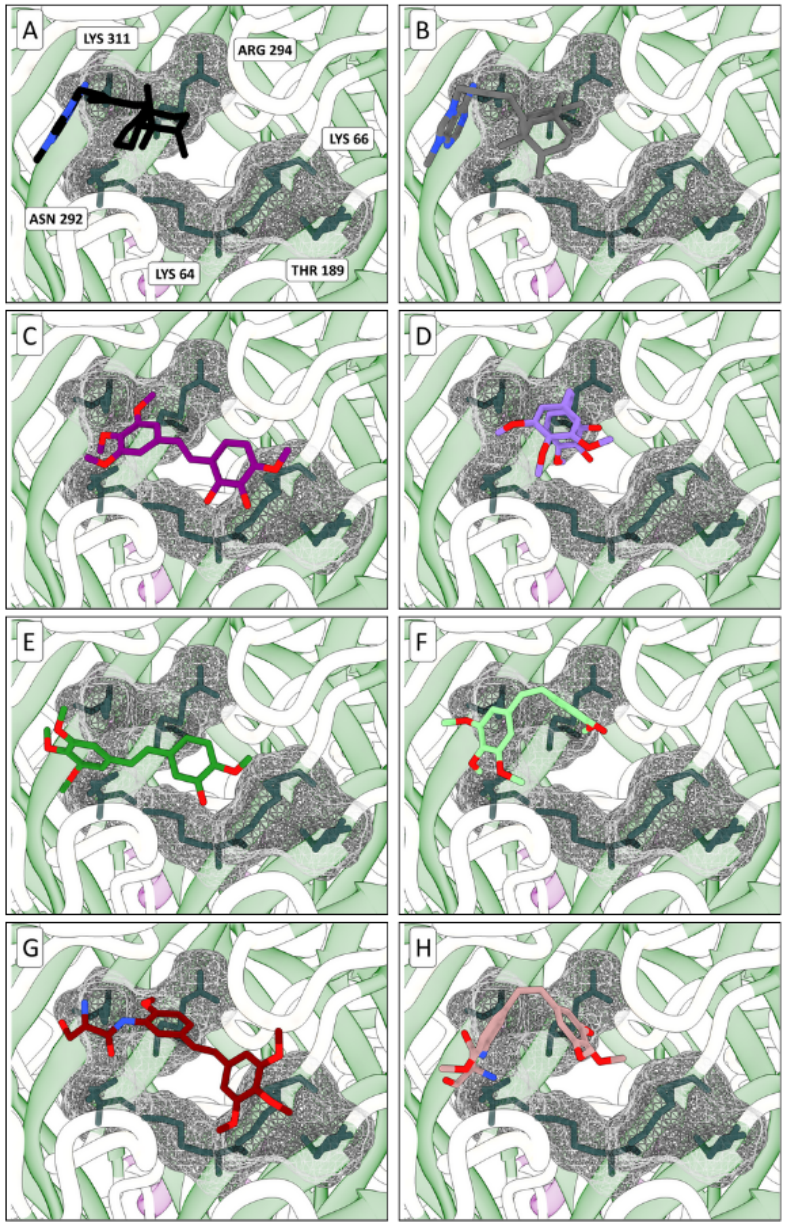


Figure 4.21: Visualization of the docking results for group 3 ligands with the highest smina score. To improve clarity, hydrogen atoms have been omitted from both the amino acid structures and the ligands with structures are colored by heteroatoms. A) Agelasine in *trans* in black, B) in *cis* in gray. C) Combretastatin A1 in *trans* in purple, D) in *cis* in lavender. E) Combretastatin A4 in *trans* in forest green, F) in *cis* in light green. G) Ombrabulin in *trans* in maroon, H) in *cis* beige.

As the native ligand P2X7-ATP and its derivative, BzATP, both contain phosphate residues, it was decided to test the phosphate derivatives, combretastatin A1 and A4. Table 4.6 and Figure 4.22 present the docking results for the ligands mentioned above.

Table 4.6: Average score of the 50 docking rounds to the ATP binding pocket for the top five positions for combretastatin derivatives

Combretastatin A1		Combretastatin A1 Phosphate		Combretastatin A4		Combretastatin A4 Phosphate	
trans	cis	trans	cis	trans	cis	trans	cis
-6.56	-6.89	-6.60	-6.73	-6.20	-6.65	-6.73	-6.68
-6.50	-6.85	-6.56	-6.68	-6.16	-6.57	-6.69	-6.63
-6.46	-6.76	-6.38	-6.63	-6.15	-6.48	-6.66	-6.60
-6.43	-6.73	-6.34	-6.58	-6.13	-6.44	-6.62	-6.59
-6.40	-6.69	-6.31	-6.50	-6.11	-6.42	-6.57	-6.48

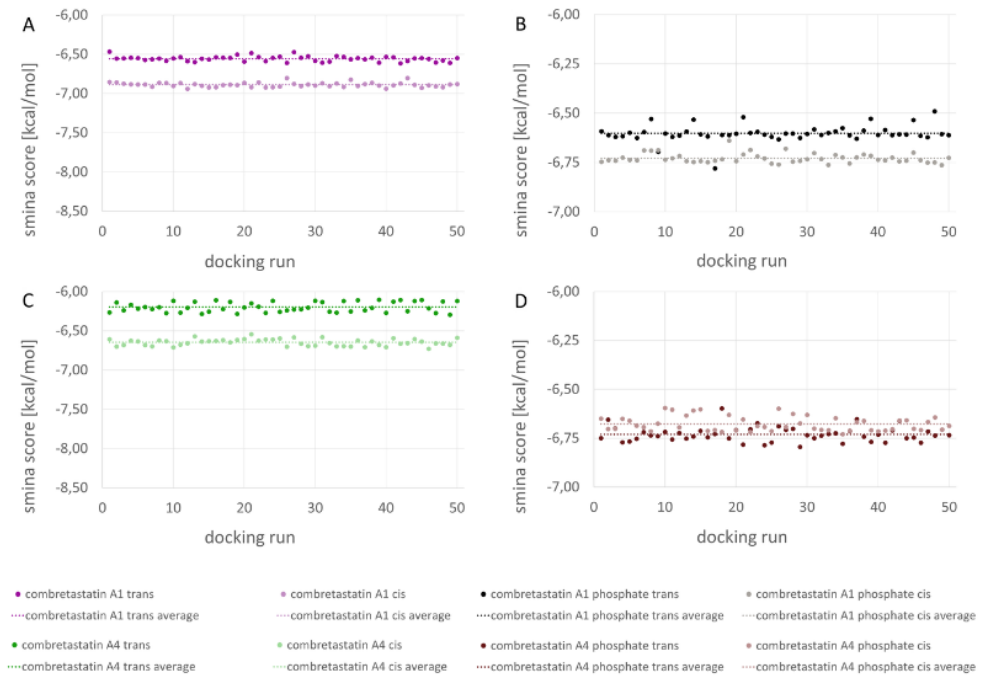


Figure 4.22: The graphs show the results for the best positions from each of the 50 docking rounds for ligands from Group 3 in the form of points, and the average value (shown in Table 4.6) as a dotted line in the color corresponding to the points. A: combretastatin A1 in *trans* in purple, in *cis* in lavender. B: combretastatin A1 phosphate in *trans* in black and *cis* in gray. C: combretastatin A4 in *trans* in forest green, in *cis* in light green. D: combretastatin A4 phosphate in *trans* in maroon, in *cis* beige.

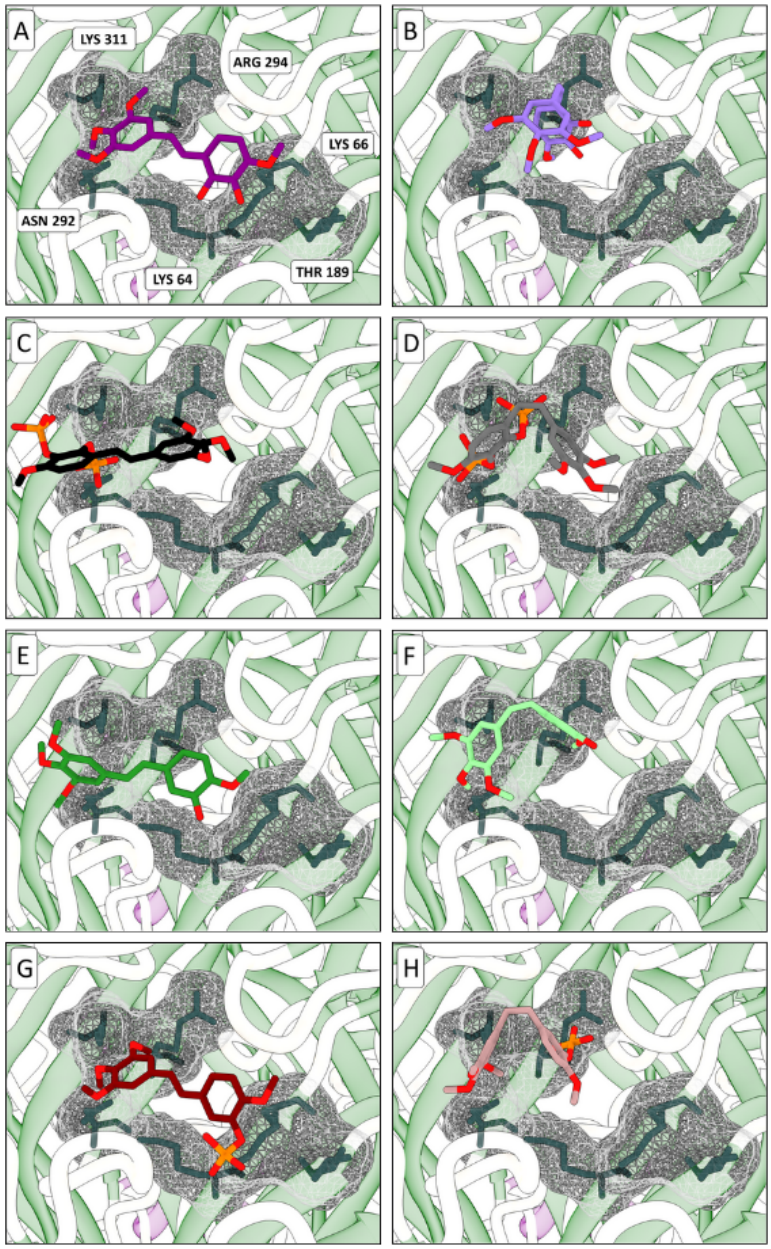


Figure 4.23: Visualization of the docking results for group 3 ligands with the highest smina score. To improve clarity, hydrogen atoms have been omitted from both the amino acid structures and the ligands with structures are colored by heteroatoms. A) Combretastatin A1 in *trans* in purple, B) in *cis* in lavender. C) Combretastatin A1 phosphate in *trans* in black, D) in *cis* in gray. E) Combretastatin A4 in *trans* in forest green, F) in *cis* in light green. G) Combretastatin A4 phosphate in *trans* in maroon, H) in *cis* beige.

The results for phosphate derivatives of combretastatins are comparable to those obtained for combretastatins without phosphate groups. The scores obtained also show no significant differences between the trans and cis states.

As the differentiation of glioblastoma cells in laboratory experiments was carried out using retinoic acid, which exhibits photoisomerisation, it was decided to test the affinity of this compound and its derivative, retinal, for the ATP-binding pocket of the P2X7 receptor. The results are presented in Table 4.7 and Figure 4.24.

Table 4.7: Average score of the 50 docking rounds to the ATP binding pocket for the top five positions for Group 4 ligands

Retinal			Retinoic Acid			
trans	11-cis	13-cis	trans	9-cis	11-cis	13-cis
-6.56	-6.60	-6.40	-7.12	-6.87	-6.82	-6.75
-6.44	-6.55	-6.34	-6.98	-6.75	-6.77	-6.72
-6.37	-6.44	-6.30	-6.93	-6.65	-6.73	-6.68
-6.33	-6.40	-6.25	-6.89	-6.57	-6.68	-6.66
-6.30	-6.34	-6.19	-6.86	-6.50	-6.65	-6.62

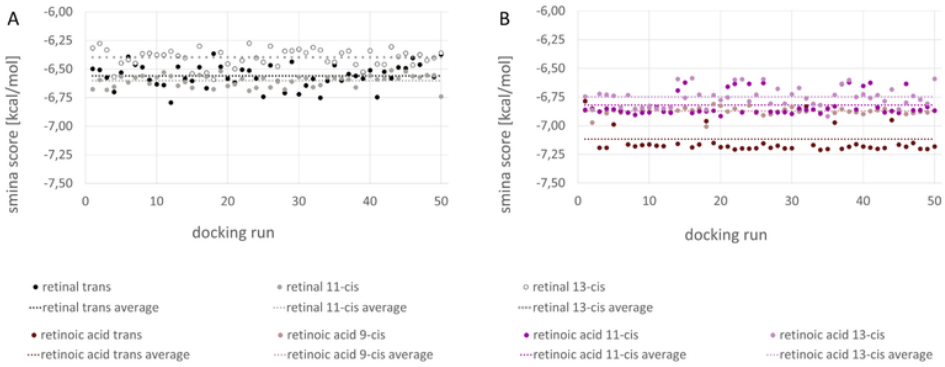


Figure 4.24: The graphs show the results for the best positions from each of the 50 docking rounds for ligands from Group 4 in the form of points, and the average value (shown in Table 4.7) as a dotted line in the color corresponding to the points. A: retinal *trans* in black, retinal 11-*cis* in gray, and retinal 13-*cis* in white. B: retinoic acid *trans* in maroon, retinoic acid 9-*cis* in beige, reticoic acid 11-*cis* in purple, and retinoic acid 13-*cis* in lavender.

Retinal in each state (Figure 4.24:A, trans in black, 11-cis in gray, and 13-cis in white) shows a score comparable to the results for TMZ, while the results show large deviations from the average value. No significant differences in scores are observed between the above-mentioned states. Better docking results were obtained for retinoic acid. The scores are lower than those for retinal, and a separation between the trans state of retinoic acid

(Figure 4.24:B in maroon), which exhibits higher binding affinity to the receptor pocket, and the cis states can be observed (Figure 4.24:B, retinoic acid 9-*cis* in beige, reticoic acid 11-*cis* in purple, and retinoic acid 13-*cis* in lavender).

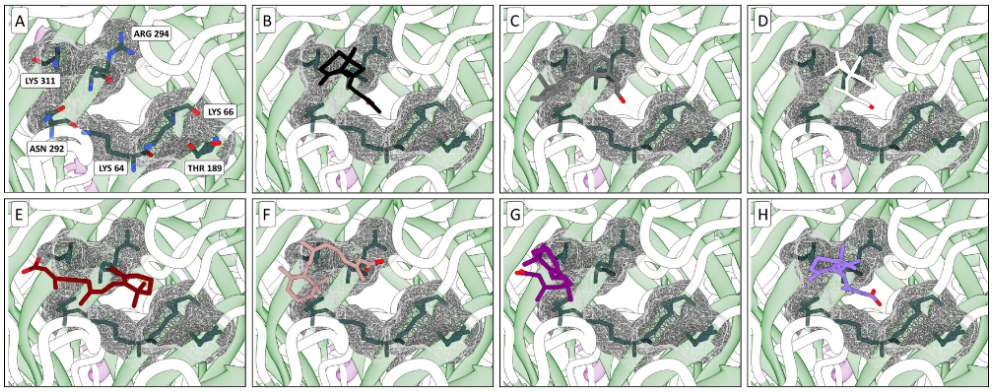


Figure 4.25: Visualization of the docking results for group 4 ligands with the highest smina score. To improve clarity, hydrogen atoms have been omitted from both the amino acid structures and the ligands with structures are colored by heteroatoms. A) retinal *trans* in black, B) retinal 11-*cis* in gray, C) and retinal 13-*cis* in white. D) retinoic acid *trans* in maroon, E) retinoic acid 9-*cis* in beige, F) reticoic acid 11-*cis* in purple, G) and retinoic acid 13-*cis* in lavender.

As shown in Figure 4.25, the visualisations of the best ligand positions in the ATP binding pocket of the P2X7 receptor do not reveal a consistent pattern of change between the trans and cis states. However, it is visible that all ligands position themselves within the amino acids Asn292, Arg294, and Lys311.

Based on the obtained scores and analysis of ligand docking positions in the ATP-binding pocket of the P2X7 receptor, it can be concluded that retinoic acid and its derivatives do not directly affect P2X7 receptor activity or channel formation.

The last group of ligands studied using molecular docking was BzATP molecules modified with photoactive fragments. The results obtained from the dockings are presented in Table 4.8 and Figure 4.26.

Table 4.8: Average score of the 50 docking rounds to the ATP binding pocket for the top five positions for Group 5 ligands

BzATP + Azobenzene		BzATP + Stilbene		BzATP + Pterostilbene		BzATP + Resveratrol	
trans	cis	trans	cis	trans	cis	trans	cis
-9.02	-9.27	-9.08	-9.30	-9.45	-9.75	-9.59	-9.32
-8.89	-9.09	-8.98	-9.18	-9.34	-9.50	-9.45	-9.14
-8.79	-8.99	-8.84	-9.08	-9.27	-9.35	-9.32	-9.03
-8.66	-8.82	-8.74	-8.99	-9.16	-9.26	-9.19	-8.96
-8.57	-8.74	-8.65	-8.91	-9.08	-9.18	-9.09	-8.91

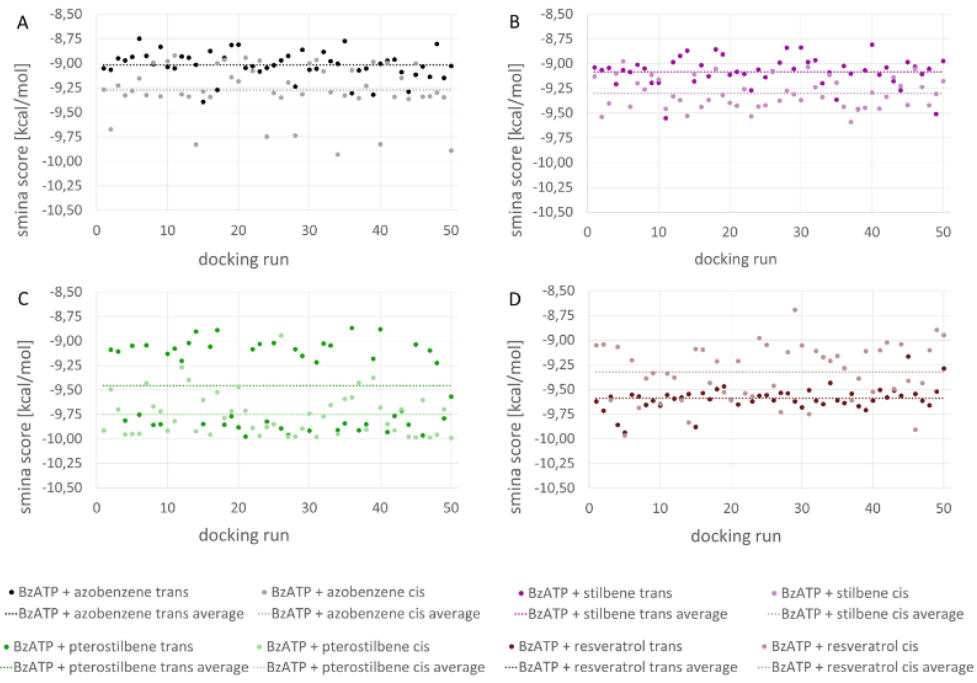


Figure 4.26: The graphs show the results for the best positions from each of the 50 docking rounds for ligands from Group 5 in the form of points, and the average value (shown in Table 4.8) as a dotted line in the color corresponding to the points. A: BzATP + azobenzene in *trans* in black, in *cis* in gray. B: BzATP + stilbene in *trans* in purple, in *cis* in lavender. C: BzATP + pterostilbene in *trans* in forest green, in *cis* in light green. D: BzATP + resveratrol in *trans* in maroon, in *cis* beige.

The docking scores obtained with modified BzATP containing photoactive parts are only slightly worse than those with BzATP without modification (-9.66 kcal/mol). However, they show a large deviation from the average value (visible in Figure 4.26). Each of the four compounds studied shows a slight difference of approximately 0.2 kcal/mol in the average value between the *trans* and *cis* states.

Figure 4.27 shows a visual analysis of the optimal docking positions for group 5 ligands. The positioning of ligands in the ATP binding pocket is highly consistent in *trans* state. BzATP modified with azobenzene (Figure 4.27:A in black), stilbene (Figure 4.27:C in purple), and resveratrol (Figure 4.27:G in maroon) in the *trans* state position phosphate residues in the pocket cavity near Arg294. The adenosine rings are located between the Asn292 and Lys311 residues, while the ring of the photoactive group is located near Lys66 and Thr189. These compounds in the *cis* position (azobenzene in Figure 4.27:B in gray, stilbene in Figure 4.27:D in lavender, and resveratrol in Figure 4.27:H in beige) are characterized by a conformational change resulting in a different arrangement, with the phosphate residues located in the vicinity of the Asn292 and Lys311 residues. The position of BzATP modified with a pretostilbene part in the *trans* state (Figure 4.27:E in forest green) shows a different orientation compared to the other compounds. In this case, the phosphate residue chain is arranged in the region of Asn292 and Lys311 (as in the *cis* states of other ligands), while the adenosine rings are located in the proximity of Arg294. The ring of the photoactive residue behaves in the same way as in the other compounds in the *trans* state and is located close to the amino acids Lys66 and Thr189. In the *cis* state of BzATP modified with a part of pterostilbene (Figure 4.27:F in light green), the phosphate residues and adenosine rings do not change their position and remain located in the region of the same amino acid residues. On the photoactive side, however, the ring moves away from Lys66 and Thr189, leaving the pocket. The difference in the behavior of BzATP modified with a pterostilbene moiety can be explained by the presence of two methoxy groups in the photoactive ring, which can stabilize the ligand in the position visualized in Figure 4.27:E via hydrogen bonds.

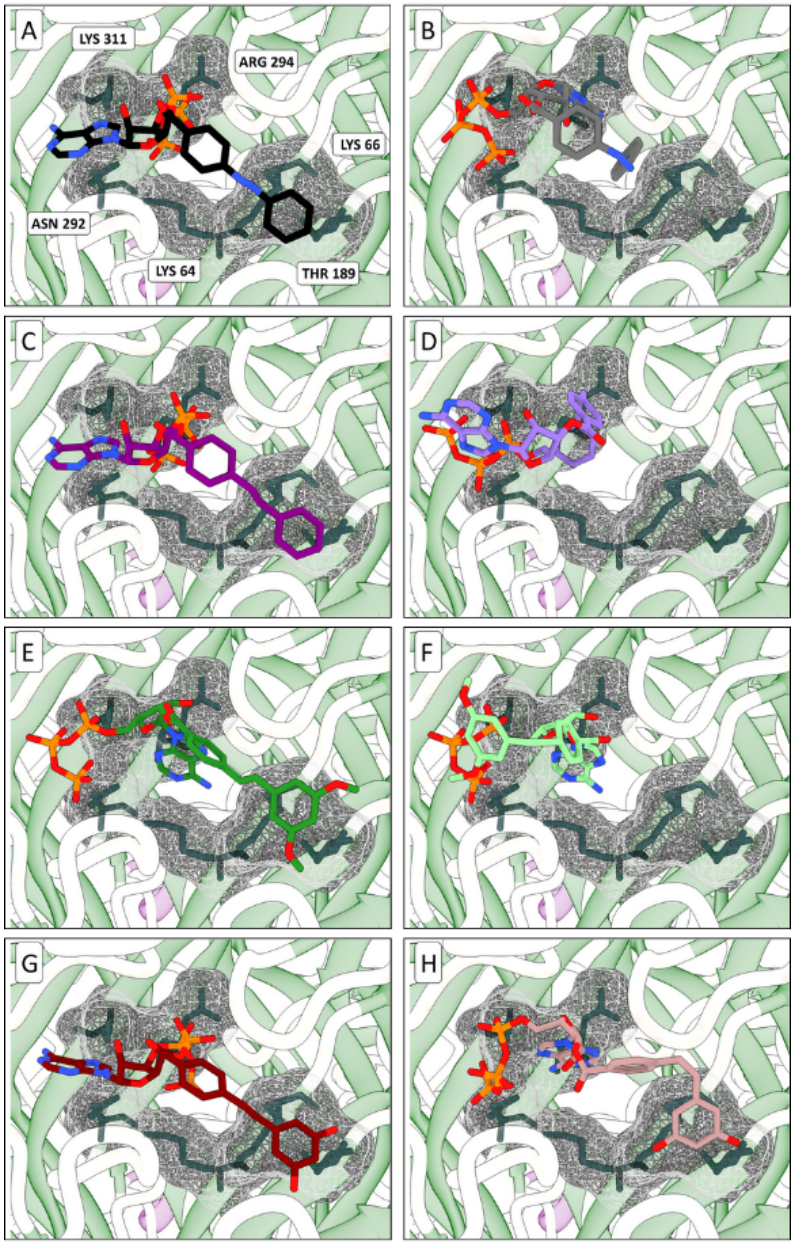


Figure 4.27: Visualization of the docking results for group 5 ligands with the highest smina score. To improve clarity, hydrogen atoms have been omitted from both the amino acid structures and the ligands with structures are colored by heteroatoms. A) BzATP + azobenzene in *trans* in black, B) in *cis* in gray. C) BzATP + stilbene in *trans* in purple, D) in *cis* in lavender. E) BzATP + pterostilbene in *trans* in forest green, F) in *cis* in light green. G) BzATP + resveratrol in *trans* in maroon, H) in *cis* in beige.

4.4 Conclusion

Interest in the P2X7 receptor has grown in recent years with the discovery of new metabolic pathways and diseases, such as depression [546, 547] Parkinson's disease [548], Alzheimer's disease [549–551], and various types of cancer [552, 553] (including those outside the central nervous system [554–556]), in which this receptor plays a role. However, complete channel inhibition is not a desirable approach in drug discovery, as the receptor also plays a crucial role in the normal functioning of many cells. While the concept of optically regulating proteins is not novel, it is an innovative approach for the P2X7 receptor.

All 43 of the docked compounds exhibit a reasonable binding affinity for the ATP-binding pocket of the P2X7 receptor. ATP derivatives have a higher affinity for the pocket as the number of negative phosphate groups increases, reflecting the pocket's strongly positive surface charge. Pterostilbene exhibited the most significant differences in binding energy between *trans* and *cis* state, making it the compound with the greatest potential for further study. The next compound selected for further research was resveratrol. The difference between the states is comparatively small, but resveratrol binds more strongly in the *trans* state than in the *cis* state. In the docking results for group 3 and 4 ligands, conformational changes from *trans* to *cis* do not show significant differences. In the case of BzATP modified with a partially photoactive component, the affinity for the binding pocket is exceptionally high. However, there are no significant differences in this case either. Conformational changes do not affect binding strength, probably because the cavity is large and the photoactive groups have considerable freedom of movement during these shifts.

Further research into suitable compounds is necessary. Research on this topic, conducted in collaboration with the experimental team, is ongoing. Future studies will continue to use BzATP due to its high affinity for the ATP-binding pocket. Follow-up studies will examine larger photoactive compounds that better fill the pocket space. This could potentially lead to conformational changes resulting in differences in binding strength. A molecular docking study of the inhibitor binding pockets would be a valid next step. It is also planned to study more receptor-photoactive ligand systems in the future using MD simulations.

Appendix

Comparison of the obtained homologous model with the structure generated by AlphaFold.

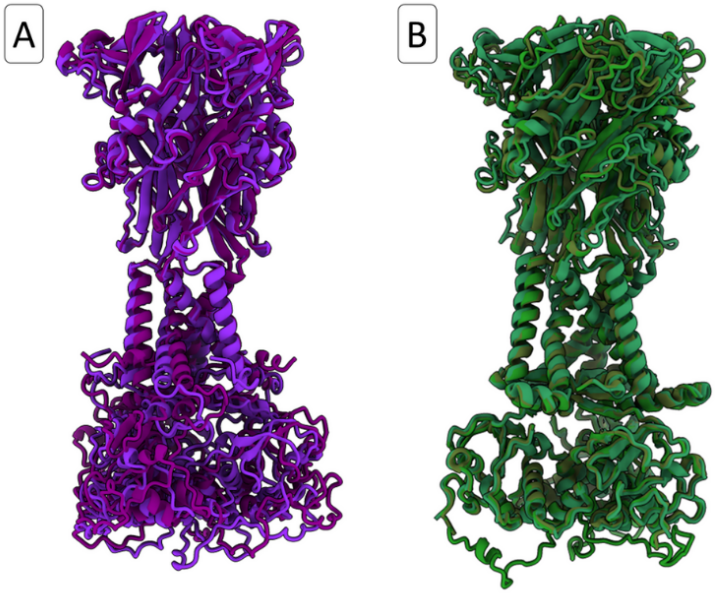
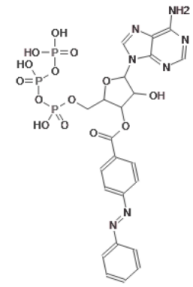
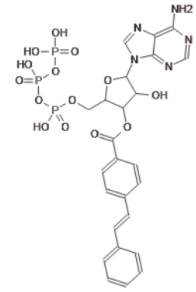
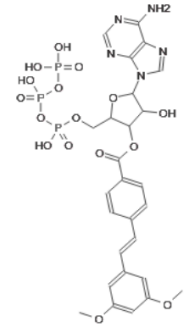
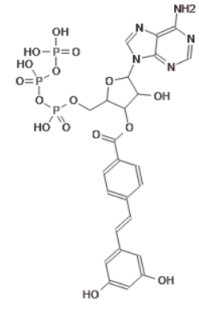


Figure 4.A0.28: Comparison of A) overlapping models generated using AlphaFold2 in shades of purple with B) overlapping structures of the homologous model and models available in PDB (the structure of the closed rat receptor PDB: 6U9V and the structure of the open rat receptor PDB: 6U9W) in shades of green.

Structures and SMLIES codes for modified compounds.

Table 4.A0.9: Structures and SMLIES codes for modified compounds.

Compound	SMILES code	
BzATP + Azobenzene	NC1=C2N=C[N](C3OC(CO[P](O)(=O)O[P](O)(=O)O[P](O)(O)=O)C(OC(=O)C4=CC=C(C=C4)N=NC5=CC=CC=C5)C3O)C2=NC=N1	
BzATP + Stilbene	NC1=C2N=C[N](C3OC(CO[P](O)(=O)O[P](O)(=O)O[P](O)(O)=O)C(OC(=O)C4=CC=C(\C=C\C5=CC=CC=C5)C=C4)C3O)C2=NC=N1	
BzATP + Pterostilbene	COC1=CC(=CC(=C1)\C=C\C2=CC=C(C=C2)C(=O)OC3C(O)C(OC3CO[P](O)(=O)O[P](O)(=O)O[P](O)(O)=O)[N]4C=NC5=C4N=CN=C5N)OC	
BzATP + Resveratrol	NC1=C2N=C[N](C3OC(CO[P](O)(=O)O[P](O)(=O)O[P](O)(O)=O)C(OC(=O)C4=CC=C(\C=C\C5=CC(=CC(=C5)O)O)C=C4)C3O)C2=NC=N1	

Docking Results

The results of docking compounds into the ATP binding pocket of the P2X7 receptor are available in the repository at: https://github.com/s-czach/PhD-Thesis.

Figure 4.A0.29: A QR code that redirects to a page containing the docking results and other relevant dissertation data.

5 Summary

Each chapter of the thesis aimed to provide a comprehensive overview of the field of optically regulated proteins by addressing a distinct research problem.

- Bacteriophytochrome and biliverdin IXα.
  This chapter explores the mechanical signal transduction steps in bacteriophytochrome and biliverdin through the use of enhanced sampling molecular dynamics simulations. Particular focus was given to characterizing the metastable conformational states associated with the Pr and Pfr photoproducts. The simulations reveal significant differences in the free energy landscapes of the two states, reflecting their distinct conformational preferences and potential transition pathways. Analysis of the trajectories indicates that the way chromophores interact with the proteins they are surrounded by is crucial in determining the shapes of these landscapes. These findings provide molecular insight into how light-induced structural changes translate into mechanical signals in bacteriophytochromes. These findings are consistent with experimental data for this system.

- Photoactive Yellow Protein and *p*-coumaric acid.
  This chapter uses new machine learning-based methods called spectral map to study the PYP model system. The slow collective variables were constructed by training neural networks on data from MD simulations. The analysis showed that the structural rearrangements in pB take place in the opposite direction to those observed in pG. These findings are physically consistent with the mechanics of

the PYP photocycle. This allows for the conclusion that machine learning methods enable a detailed study of photoactive systems and observation of their behavior at a level that cannot be achieved using conventional methods.

- Light-induced regulation of proteins, photoactivated ligands of P2X7 receptor.
  This chapter describes the process of creating a homologous model of a human receptor based on human sequences and rat receptor structures. A series of molecular docking experiments was conducted to investigate the potential use of photoactive compounds to regulate the purinergic ion channel P2X7. The binding affinity of these compounds for the ATP-binding pocket of the homology receptor was then evaluated in relation to the docking results for the native ligand, ATP. The encouraging results of ligand docking suggest that the use of photoactive ligands could be a promising future area of research for developing new therapeutic compounds, motivating continued work in this area.

Bibliography

[1] S. Czach, J. Rydzewski, W. Nowak *et al.*, "Illuminating protein dynamics: A review of computational methods for studying photoactive proteins," *arXiv preprint arXiv:2503.21418*, 2025.

[2] J. M. Berg, G. J. Gatto, J. K. Hines, J. L. Tymoczko *et al.*, *Biochemistry (-2024)*. Macmillan Learning, 2024.

[3] E. A. Ponomarenko, E. V. Poverennaya, E. V. Ilgisonis, M. A. Pyatnitskiy, A. T. Kopylov, V. G. Zgoda, A. V. Lisitsa, and A. I. Archakov, "The size of the human proteome: the width and depth," *International journal of analytical chemistry*, vol. 2016, no. 1, p. 7436849, 2016.

[4] P. Sinitcyn, A. L. Richards, R. J. Weatheritt, D. R. Brademan, H. Marx, E. Shishkova, J. G. Meyer, A. S. Hebert, M. S. Westphall, B. J. Blencowe *et al.*, "Global detection of human variants and isoforms by deep proteome sequencing," *Nature biotechnology*, vol. 41, no. 12, pp. 1776–1786, 2023.

[5] M. J. Lopez and S. S. Mohiuddin, "Biochemistry, essential amino acids," in *StatPearls [Internet]*. StatPearls Publishing, 2024.

[6] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "The shape and structure of proteins," in *Molecular Biology of the Cell. 4th edition*. Garland Science, 2002.

[7] R. N. McLaughlin Jr, F. J. Poelwijk, A. Raman, W. S. Gosal, and R. Ranganathan, "The spatial architecture of protein function and adaptation," *Nature*, vol. 491, no. 7422, pp. 138–142, 2012.

[8] H.-X. Zhou and X. Pang, "Electrostatic interactions in protein structure, folding, binding, and condensation," *Chemical reviews*, vol. 118, no. 4, pp. 1691–1741, 2018.

[9] J. Hughes and T. Lamparter, "A prokaryotic phytochrome." *Nature*, vol. 386, no. 6626, 1997.

[10] Z. Jiang, L. R. Swem, B. G. Rushing, S. Devanathan, G. Tollin, and C. E. Bauer, "Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes," *Science*, vol. 285, no. 5426, pp. 406–409, 1999.

[11] K. J. Hellingwerf, J. Hendriks, and T. Gensch, "Photoactive yellow protein, a new type of photoreceptor protein: will this "yellow lab" bring us where we want to go?" *The Journal of Physical Chemistry A*, vol. 107, no. 8, pp. 1082–1094, 2003.

[12] M. A. Van Der Horst and K. J. Hellingwerf, "Photoreceptor proteins,"star actors of modern times": a review of the functional dynamics in the structure of representative members of six different photoreceptor families," *Accounts of chemical research*, vol. 37, no. 1, pp. 13–20, 2004.

[13] M.-a. Mroginski, S. Adam, G. S. Amoyal, A. Barnoy, A.-n. Bondar, V. A. Borin, J. R. Church, T. Domratcheva, B. Ensing, F. Fanelli *et al.*, "Frontiers in multiscale modeling of photoreceptor proteins," *Photochemistry and photobiology*, vol. 97, no. 2, pp. 243–269, 2021.

[14] J. Jamieson, "The influence of light on the development of bacteria," *Nature*, vol. 26, no. 663, pp. 244–245, 1882.

[15] E. B. Purcell, D. Siegal-Gaskins, D. C. Rawling, A. Fiebig, and S. Crosson, "A photosensory two-component system regulates bacterial cell attachment," *Proceedings of the National Academy of Sciences*, vol. 104, no. 46, pp. 18 241–18 246, 2007.

[16] L. Capece, L. Boechi, L. L. Perissinotti, P. Arroyo-Mañez, D. E. Bikiel, G. Smulevich, M. A. Marti, and D. A. Estrin, "Small ligand–globin interactions: reviewing lessons derived from computer simulation," *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, vol. 1834, no. 9, pp. 1722–1738, 2013.

[17] S. Gozem, H. L. Luk, I. Schapiro, and M. Olivucci, "Theory and Simulation of the Ultrafast Double-bond Isomerization of Biological Chromophores," *Chem. Rev.*, vol. 117, no. 22, pp. 13 502–13 565, 2017.

[18] M. Gomelsky and W. D. Hoff, "Light helps bacteria make important lifestyle decisions," *Trends in microbiology*, vol. 19, no. 9, pp. 441–448, 2011.

[19] A. Möglich, R. A. Ayers, and K. Moffat, "Structure and Signaling Mechanism of Per-Arnt-Sim Domains," *Structure*, vol. 17, no. 10, pp. 1282–1294, 2009.

[20] J. T. Kennis and T. Mathes, "Molecular Eyes: Proteins that Transform Light into Biological Information," *Interface Focus*, vol. 3, no. 5, p. 20130005, 2013.

[21] G. Salvadori, P. Mazzeo, D. Accomasso, L. Cupellini, and B. Mennucci, "Deciphering Photoreceptors through Atomistic Modeling from Light Absorption to Conformational Response," *J. Mol. Biol.*, vol. 436, no. 5, p. 168358, 2024.

[22] G. Groenhof, V. Modi, and D. Morozov, "Observe while it Happens: Catching Photoactive Proteins in the Act with Non-Adiabatic Molecular Dynamics Simulations," *Curr. Opin. Struct. Biol.*, vol. 61, pp. 106–112, 2020.

[23] J. Gauthier, A. T. Vincent, S. J. Charette, and N. Derome, "A Brief History of Bioinformatics," *Brief. Bioinform.*, vol. 20, no. 6, pp. 1981–1996, 2019.

[24] Y. J. Edwards and A. Cottage, "Bioinformatics Methods to Predict Protein Structure and Function: A Practical Approach," *Mol. Biotechnol.*, vol. 23, pp. 139–166, 2003.

[25] A. Pavlopoulou and I. Michalopoulos, "State-of-the-Art Bioinformatics Protein Structure Prediction Tools," *Int. J. Mol. Med.*, vol. 28, no. 3, pp. 295–310, 2011.

[26] R. C. Bernardi, M. C. Melo, and K. Schulten, "Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems," *Biochim. Biophys. Acta*, vol. 1850, no. 5, pp. 872–877, 2015.

[27] O. Valsson, P. Tiwary, and M. Parrinello, "Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint," *Annu. Rev. Phys. Chem.*, vol. 67, no. 1, pp. 159–184, 2016.

[28] G. Bussi and A. Laio, "Using Metadynamics to Explore Complex Free-Energy Landscapes," *Nat. Rev. Phys.*, vol. 2, pp. 200–2012, 2020.

[29] J. Hénin, T. Lelièvre, M. R. Shirts, O. Valsson, and L. Delemotte, "Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]," *Living Journal of Computational Molecular Science*, vol. 4, no. 1, p. 1583, 2022.

[30] E. Brunk and U. Rothlisberger, "Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States," *Chem. Rev.*, vol. 115, no. 12, pp. 6217–6263, 2015.

[31] T. Andruniów and M. Olivucci, *QM/MM Studies of Light-responsive Biological Systems*. Springer, 2021.

[32] Y. Wang, J. M. L. Ribeiro, and P. Tiwary, "Machine Learning Approaches for Analyzing and Enhancing Molecular Dynamics Simulations," *Curr. Opin. Struct. Biol.*, vol. 61, pp. 139–145, 2020.

[33] J. Rydzewski, M. Chen, and O. Valsson, "Manifold Learning in Atomistic Simulations: A Conceptual Review," *Mach. Learn.: Sci. Technol.*, vol. 4, p. 031001, 2023.

[34] S. Mehdi, Z. Smith, L. Herron, Z. Zou, and P. Tiwary, "Enhanced Sampling with Machine Learning," *Annu. Rev. Phys. Chem.*, vol. 75, no. 1, p. 347–370, 2024.

[35] T. Gökdemir and J. Rydzewski, "Machine Learning of Slow Collective Variables and Enhanced Sampling via Spatial Techniques," *Chem. Phys. Rev.*, vol. 6, no. 1, 2025.

[36] R. Kort, W. Hoff, M. Van West, A. Kroon, S. Hoffer, K. Vlieg, W. Crielaand, J. Van Beeumen, and K. J. Hellingwerf, "The xanthopsins: a new family of eubacterial blue-light photoreceptors." *The EMBO journal*, vol. 15, no. 13, pp. 3209–3218, 1996.

[37] W. D. Hoff, A. Xie, I. H. Van Stokkum, X.-j. Tang, J. Gural, A. R. Kroon, and K. J. Hellingwerf, "Global conformational changes upon receptor stimulation in photoactive yellow protein," *Biochemistry*, vol. 38, no. 3, pp. 1009–1017, 1999.

[38] P. H. Quail, "The phytochrome family: dissection of functional roles and signalling pathways among family members," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 353, no. 1374, pp. 1399–1403, 1998.

[39] W. D. Hoff, K.-H. Jung, and J. L. Spudich, "Molecular mechanism of photosignaling by archaeal sensory rhodopsins," *Annual review of biophysics and biomolecular structure*, vol. 26, no. 1, pp. 223–258, 1997.

[40] J. L. Spudich, C.-S. Yang, K.-H. Jung, and E. N. Spudich, "Retinylidene proteins: structures and functions from archaea to humans," *Annual review of cell and developmental biology*, vol. 16, no. 1, pp. 365–392, 2000.

[41] M. Ahmad and A. R. Cashmore, "Hy4 gene of a. thaliana encodes a protein with characteristics of a blue-light photoreceptor," *Nature*, vol. 366, no. 6451, pp. 162–166, 1993.

[42] E. Huala, P. W. Oeller, E. Liscum, I.-S. Han, E. Larsen, and W. R. Briggs, "Arabidopsis nph1: a protein kinase with a putative redox-sensing domain," *Science*, vol. 278, no. 5346, pp. 2120–2123, 1997.

[43] M. Gomelsky and G. Klug, "Bluf: a novel fad-binding domain involved in sensory transduction in microorganisms," *Trends in biochemical sciences*, vol. 27, no. 10, pp. 497–500, 2002.

[44] R. Y. Tsien, "The Green Fluorescent Protein," *Annu. Rev. Biochem.*, vol. 67, no. 1, pp. 509–544, 1998.

[45] H. Ihee, S. Rajagopal, V. Šrajer, R. Pahl, S. Anderson, M. Schmidt, F. Schotte, P. A. Anfinrud, M. Wulff, and K. Moffat, "Visualizing Reaction Pathways in Photoactive Yellow Protein From Nanoseconds to Seconds," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 20, pp. 7145–7150, 2005.

[46] A. K. Michael, J. L. Fribourgh, Y. Chelliah, C. R. Sandate, G. L. Hura, D. Schneidman-Duhovny, S. M. Tripathi, J. S. Takahashi, and C. L. Partch, "Formation of a Repressive Complex in the Mammalian Circadian Clock is Mediated by the Secondary Pocket of CRY1," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 7, pp. 1560–1565, 2017.

[47] H. Takala, A. Björling, O. Berntsson, H. Lehtivuori, S. Niebling, M. Hoernke, I. Kosheleva, R. Henning, A. Menzel, J. A. Ihalainen *et al.*, "Signal Amplification and Transduction in Phytochrome Photosensors," *Nature*, vol. 509, no. 7499, pp. 245–248, 2014.

[48] G. Gotthard, S. Mous, T. Weinert, R. N. A. Maia, D. James, F. Dworkowski, D. Gashi, A. Furrer, D. Ozerov, E. Panepucci *et al.*, "Capturing the Blue-Light Activated State of the Phot-LOV1 Domain from Chlamydomonas Reinhardtii using Time-Resolved Serial Synchrotron Crystallography," *IUCrJ*, vol. 11, no. 5, 2024.

[49] A. Kita, K. Okajima, Y. Morimoto, M. Ikeuchi, and K. Miki, "Structure of a Cyanobacterial BLUF Protein, Tll0078, Containing a Novel FAD-binding Blue Light Sensor Domain," *J. Mol. Biol.*, vol. 349, no. 1, pp. 1–9, 2005.

[50] H. E. Kato, F. Zhang, O. Yizhar, C. Ramakrishnan, T. Nishizawa, K. Hirata, J. Ito, Y. Aita, T. Tsukazaki, S. Hayashi *et al.*, "Crystal Structure of the Channelrhodopsin Light-Gated Cation Channel," *Nature*, vol. 482, no. 7385, pp. 369–374, 2012.

[51] F. Yang, L. G. Moss, and G. N. Phillips Jr, "The Molecular Structure of Green Fluorescent Protein," *Nat. Biotechnol.*, vol. 14, no. 10, pp. 1246–1251, 1996.

[52] D. P. Barondeau, J. A. Tainer, and E. D. Getzoff, "Structural Evidence for an Enolate Intermediate in Gfp Fluorophore Biosynthesis," *J. Am. Chem. Soc.*, vol. 128, no. 10, pp. 3166–3168, 2006.

[53] L. Zhang, H. N. Patel, J. W. Lappe, and R. M. Wachter, "Reaction Progress of Chromophore Biogenesis in Green Fluorescent Protein," *J. Am. Chem. Soc.*, vol. 128, no. 14, pp. 4766–4772, 2006.

[54] T. E. Meyer, J. A. Kyndt, S. Memmi, T. Moser, B. Colón-Acevedo, B. Devreese, and J. Van Beeumen, "The growing family of photoactive yellow proteins and their presumed functional roles," *Photochemical & Photobiological Sciences*, vol. 11, no. 10, pp. 1495–1514, 2012.

[55] Y. Imamoto and M. Kataoka, "Structure and photoreaction of photoactive yellow protein, a structural prototype of the pas domain superfamily," *Photochemistry and photobiology*, vol. 83, no. 1, pp. 40–49, 2007.

[56] F. Schotte, H. S. Cho, F. Dyda, and P. Anfinrud, "Watching a signaling protein function: What has been learned over four decades of time-resolved studies of photoactive yellow protein," *Structural Dynamics*, vol. 11, no. 2, 2024.

[57] K. J. Hellingwerf, "The molecular basis of sensing and responding to light in microorganisms," *Antonie Van Leeuwenhoek*, vol. 81, no. 1, pp. 51–59, 2002.

[58] P. H. Quail, M. T. Boylan, B. M. Parks, T. W. Short, Y. Xu, and D. Wagner, "Phytochromes: photosensory perception and signal transduction," *Science*, vol. 268, no. 5211, pp. 675–680, 1995.

[59] J. Rydzewski, K. Walczewska-Szewc, S. Czach, W. Nowak, and K. Kuczera, "Enhancing the inhomogeneous photodynamics of canonical bacteriophytochrome," *The Journal of Physical Chemistry B*, vol. 126, no. 14, pp. 2647–2657, 2022.

[60] M. Legris, Y. Ç. Ince, and C. Fankhauser, "Molecular Mechanisms Underlying Phytochrome-controlled Morphogenesis in Plants," *Nat. Commun.*, vol. 10, no. 1, p. 5219, 2019.

[61] O. S. Oliinyk, K. G. Chernov, and V. V. Verkhusha, "Bacterial Phytochromes, Cyanobacteriochromes and Allophycocyanins as a Source of Near-Infrared Fluorescent Probes," *Int. J. Mol. Sci.*, vol. 18, no. 8, p. 1691, 2017.

[62] A. Schmidt, L. Sauthof, M. Szczepek, M. F. Lopez, F. V. Escobar, B. M. Qureshi, N. Michael, D. Buhrke, T. Stevens, D. Kwiatkowski *et al.*, "Structural Snapshot of a Bacterial Phytochrome in Its Functional Intermediate State," *Nat. Commun.*, vol. 9, no. 1, p. 4912, 2018.

[63] Y. Kuwasaki, K. Suzuki, G. Yu, S. Yamamoto, T. Otabe, Y. Kakihara, M. Nishiwaki, K. Miyake, K. Fushimi, R. Bekdash *et al.*, "A Red Light–Responsive Photoswitch for Deep Tissue Optogenetics," *Nat. Biotechnol.*, vol. 40, no. 11, pp. 1672–1679, 2022.

[64] L. Qiao, L. Niu, M. Wang, Z. Wang, D. Kong, G. Yu, and H. Ye, "A Sensitive Red/Far-Red Photoswitch for Controllable Gene Therapy in Mouse Models of Metabolic Diseases," *Nat. Commun.*, vol. 15, no. 1, p. 10310, 2024.

[65] N. C. Rockwell, Y.-S. Su, and J. C. Lagarias, "Phytochrome Structure and Signaling Mechanisms," *Annu. Rev. Plant Biol.*, vol. 57, no. 1, pp. 837–858, 2006.

[66] J. Li, G. Li, H. Wang, and X. W. Deng, *Phytochrome Signaling Mechanisms*. American Society of Plant Biologists, 2011, vol. 9.

[67] K. J. Halliday and S. J. Davis, "Light-Sensing Phytochromes Feel the Heat," *Science*, vol. 354, no. 6314, pp. 832–833, 2016.

[68] M. Mroginski, D. Von Stetten, S. Kaminski, F. V. Escobar, N. Michael, G. Daminelli-Widany, and P. Hildebrandt, "Elucidating Photoinduced Structural Changes in Phytochromes by the Combined Application of Resonance Raman Spectroscopy and Theoretical Methods," *J. Mol. Struct.*, vol. 993, no. 1-3, pp. 15–25, 2011.

[69] O. Béja, L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich *et al.*, "Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea," *Science*, vol. 289, no. 5486, pp. 1902–1906, 2000.

[70] O. A. Sineshchekov, K.-H. Jung, and J. L. Spudich, "Two Rhodopsins Mediate Phototaxis to Low-and High-intensity Light in Chlamydomonas Reinhardtii," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, no. 13, pp. 8689–8694, 2002.

[71] T. Friedrich, S. Geibel, R. Kalmbach, I. Chizhov, K. Ataka, J. Heberle, M. Engelhard, and E. Bamberg, "Proteorhodopsin is a Light-driven Proton Pump With Variable Vectoriality," *J. Mol. Biol.*, vol. 321, no. 5, pp. 821–838, 2002.

[72] K. Palczewski, "G Protein–Coupled Receptor Rhodopsin," *Annu. Rev. Biochem.*, vol. 75, no. 1, pp. 743–767, 2006.

[73] S. O. Smith, "Structure and Activation of the Visual Pigment Rhodopsin," *Annu. Rev. Biophys.*, vol. 39, no. 1, pp. 309–328, 2010.

[74] J. M. Berg, G. J. Gatto, J. Hines, J. L. Tymoczko, and L. Stryer, *Biochemistry*, 10th ed. Macmillan, 2023.

[75] L. N. Caro, Z. Li, A. R. Balo, N. Van Eps, J. M. Rini, and O. P. Ernst, "Rapid and Facile Recombinant Expression of Bovine Rhodopsin in HEK293S Gnti-Cells Using a Piggybac Inducible System," in *Meth. Enzymol.* Elsevier, 2015, vol. 556, pp. 307–330.

[76] G. Nagel, T. Szellas, W. Huhn, S. Kateriya, N. Adeishvili, P. Berthold, D. Ollig, P. Hegemann, and E. Bamberg, "Channelrhodopsin-2, a Directly Light-gated Cation-selective Membrane Channel," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, no. 24, pp. 13 940–13 945, 2003.

[77] C. Bamann, T. Kirsch, G. Nagel, and E. Bamberg, "Spectral Characteristics of the Photocycle of Channelrhodopsin-2 and Its Implication for Channel Function," *J. Mol. Biol.*, vol. 375, no. 3, pp. 686–694, 2008.

[78] E. G. Govorunova, Y. Gou, O. A. Sineshchekov, H. Li, X. Lu, Y. Wang, L. S. Brown, F. St-Pierre, M. Xue, and J. L. Spudich, "Kalium Channelrhodopsins Are Natural Light-Gated Potassium Channels That Mediate Optogenetic Inhibition," *Nat. Neurosci.*, vol. 25, no. 7, pp. 967–974, 2022.

[79] Y. Zhou, Y. Wei, L. Li, T. Yan, and H. Ye, "Optogenetics in Medicine: Innovations and Therapeutic Applications," *Curr. Opin. Biotechnol.*, vol. 92, p. 103262, 2025.

[80] V. S. Ovechkina, S. K. Andrianova, I. O. Shimanskaia, P. S. Suvorova, A. Y. Ryabinina, M. L. Blagonravov, V. V. Belousov, and A. A. Mozhaev, "Advances in Optogenetics and Thermogenetics for Control of Non-Neuronal Cells and Tissues in Biomedical Research," *ACS Chem. Biol.*, 2025.

[81] M. Charette, C. Rosenblum, O. Shade, and A. Deiters, "Optogenetics with Atomic Precision - A Comprehensive Review of Optical Control of Protein Function through Genetic Code Expansion," *Chem. Rev.*, 2025.

[82] W. Zhou, L. Jia, L. Yue, and L. Hu, "Advances and Applications of Peripheral Optogenetics in Animal Models," *Neuroscience*, 2025.

[83] J. Kuhne, J. Vierock, S. A. Tennigkeit, M.-A. Dreier, J. Wietek, D. Petersen, K. Gavriljuk, S. F. El-Mashtoly, P. Hegemann, and K. Gerwert, "Unifying photocycle model for light adaptation and temporal evolution of cation conductance in channelrhodopsin-2," *Proceedings of the National Academy of Sciences*, vol. 116, no. 19, pp. 9380–9389, 2019.

[84] J. Becker-Baldus, C. Bamann, K. Saxena, H. Gustmann, L. J. Brown, R. C. Brown, C. Reiter, E. Bamberg, J. Wachtveitl, H. Schwalbe *et al.*, "Enlightening the photoactive site of channelrhodopsin-2 by dnp-enhanced solid-state nmr spectroscopy," *Proceedings of the National Academy of Sciences*, vol. 112, no. 32, pp. 9896–9901, 2015.

[85] C. Bamann, R. Gueta, S. Kleinlogel, G. Nagel, and E. Bamberg, "Structural guidance of the photocycle of channelrhodopsin-2 by an interhelical hydrogen bond," *Biochemistry*, vol. 49, no. 2, pp. 267–278, 2010.

[86] K. Eisenhauer, J. Kuhne, E. Ritter, A. Berndt, S. Wolf, E. Freier, F. Bartl, P. Hegemann, and K. Gerwert, "In channelrhodopsin-2 glu-90 is crucial for ion selectivity and is deprotonated during the photocycle," *Journal of Biological Chemistry*, vol. 287, no. 9, pp. 6904–6911, 2012.

[87] V. A. Lórenz-Fonfría, T. Resler, N. Krause, M. Nack, M. Gossing, G. Fischer von Mollard, C. Bamann, E. Bamberg, R. Schlesinger, and J. Heberle, "Transient protonation changes in channelrhodopsin-2 and their relevance to channel gating," *Proceedings of the National Academy of Sciences*, vol. 110, no. 14, pp. E1273–E1281, 2013.

[88] Q. Xin, W. Zhang, and S. Yuan, "The mechanism of the channel opening in channelrhodopsin-2: A molecular dynamics simulation," *International Journal of Molecular Sciences*, vol. 24, no. 6, p. 5667, 2023.

[89] P. Hegemann, S. Ehlenbeck, and D. Gradmann, "Multiple photocycles of channelrhodopsin," *Biophysical journal*, vol. 89, no. 6, pp. 3911–3918, 2005.

[90] J. Gressel, "Blue light photoreception," *Photochemistry and photobiology*, vol. 30, no. 6, pp. 749–754, 1979.

[91] A. Sancar, "Structure and function of dna photolyase and cryptochrome blue-light photoreceptors," *Chemical reviews*, vol. 103, no. 6, pp. 2203–2238, 2003.

[92] C. Lin and T. Todo, "The cryptochromes," *Genome biology*, vol. 6, no. 5, p. 220, 2005.

[93] I. Chaves, R. Pokorny, M. Byrdin, N. Hoang, T. Ritz, K. Brettel, L.-O. Essen, G. T. Van Der Horst, A. Batschauer, and M. Ahmad, "The cryptochromes: blue light photoreceptors in plants and animals," *Annual review of plant biology*, vol. 62, no. 1, pp. 335–364, 2011.

[94] Q.-H. Li and H.-Q. Yang, "Cryptochrome signaling in plants," *Photochemistry and photobiology*, vol. 83, no. 1, pp. 94–101, 2007.

[95] Q. Wang and C. Lin, "Mechanisms of cryptochrome-mediated photoresponses in plants," *Annual Review of Plant Biology*, vol. 71, no. 1, pp. 103–129, 2020.

[96] M. Liedvogel and H. Mouritsen, "Cryptochromes—a potential magnetoreceptor: what do we know and what do we want to know?" *Journal of the Royal Society Interface*, vol. 7, no. suppl_2, pp. S147–S162, 2010.

[97] X. Yu, H. Liu, J. Klejnot, and C. Lin, "The cryptochrome blue light receptors," *The Arabidopsis Book/American Society of Plant Biologists*, vol. 8, p. e0135, 2010.

[98] C. C. DeOliveira and B. R. Crane, "A structural decryption of cryptochromes," *Frontiers in chemistry*, vol. 12, p. 1436322, 2024.

[99] C. Lin and D. Shalitin, "Cryptochrome structure and signal transduction," *Annual review of plant biology*, vol. 54, no. 1, pp. 469–496, 2003.

[100] C. L. Partch and A. Sancar, "Photochemistry and photobiology of cryptochrome blue-light photopigments: the search for a photocycle," *Photochemistry and photobiology*, vol. 81, no. 6, pp. 1291–1304, 2005.

[101] B. Liu, H. Liu, D. Zhong, and C. Lin, "Searching for a photocycle of the cryptochrome photoreceptors," *Current opinion in plant biology*, vol. 13, no. 5, pp. 578–586, 2010.

[102] M. Ahmad, "Photocycle and signaling mechanisms of plant cryptochromes," *Current Opinion in Plant Biology*, vol. 33, pp. 108–115, 2016.

[103] B. Aguida, J. Babo, S. Baouz, N. Jourdan, M. Procopio, M. A. El-Esawi, D. Engle, S. Mills, S. Wenkel, A. Huck *et al.*, "'seeing'the electromagnetic spectrum: spotlight on the cryptochrome photocycle," *Frontiers in Plant Science*, vol. 15, p. 1340304, 2024.

[104] R. B. Celaya and E. Liscum, "Phototropins and associated signaling: providing the power of movement in higher plants," *Photochemistry and photobiology*, vol. 81, no. 1, pp. 73–80, 2005.

[105] W. R. Briggs, J. M. Christie, and T. E. Swartz, "Phototropins," in *Photomorphogenesis in plants and bacteria*. Springer, 2006, pp. 223–252.

[106] W. R. Briggs, "Phototropin overview," in *Light sensing in plants*. Springer, 2005, pp. 139–146.

[107] F.-W. Li, C. J. Rothfels, M. Melkonian, J. C. Villarreal, D. W. Stevenson, S. W. Graham, G. K.-S. Wong, S. Mathews, and K. M. Pryer, "The origin and evolution of phototropins," *Frontiers in plant science*, vol. 6, p. 637, 2015.

[108] S. Tokutomi, D. Matsuoka, and K. Zikihara, "Molecular structure and regulation of phototropin kinase by blue light," *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, vol. 1784, no. 1, pp. 133–142, 2008.

[109] H. Katsura, K. Zikihara, K. Okajima, S. Yoshihara, and S. Tokutomi, "Oligomeric structure of lov domains in arabidopsis phototropin," *FEBS letters*, vol. 583, no. 3, pp. 526–530, 2009.

[110] J. Łabuz, O. Sztatelman, and P. Hermanowicz, "Molecular insights into the phototropin control of chloroplast movements," *Journal of Experimental Botany*, vol. 73, no. 18, pp. 6034–6051, 2022.

[111] J. E. Hart and K. H. Gardner, "Lighting the way: Recent insights into the structure and regulation of phototropin blue light receptors," *Journal of Biological Chemistry*, vol. 296, 2021.

[112] M. T. Alexandre, R. van Grondelle, K. J. Hellingwerf, B. Robert, and J. T. Kennis, "Perturbation of the ground-state electronic structure of fmn by the conserved cysteine in phototropin lov2 domains," *Physical Chemistry Chemical Physics*, vol. 10, no. 44, pp. 6693–6702, 2008.

[113] T. Koyama, T. Iwata, A. Yamamoto, Y. Sato, D. Matsuoka, S. Tokutomi, and H. Kandori, "Different role of the jα helix in the light-induced activation of the lov2 domains in various phototropins," *Biochemistry*, vol. 48, no. 32, pp. 7621–7628, 2009.

[114] K. Okajima, "Molecular mechanism of phototropin light signaling," *Journal of plant research*, vol. 129, no. 2, pp. 149–157, 2016.

[115] L. Henry, O. Berntsson, M. Panman, A. Cellini, A. Hughes, I. Kosheleva, R. Henning, and S. Westenhoff, "New light on the mechanism of phototransduction in phototropin," *Biochemistry*, vol. 59, no. 35, pp. 3206–3215, 2020.

[116] Y. Zhou, S. Tang, Z. Chen, Z. Zhou, J. Huang, X.-W. Kang, S. Zou, B. Wang, T. Zhang, B. Ding *et al.*, "Origin of the multi-phasic quenching dynamics in the bluf domains across the species," *Nature Communications*, vol. 15, no. 1, p. 623, 2024.

[117] S. Masuda, "Light detection and signal transduction in the bluf photoreceptors," *Plant and Cell Physiology*, vol. 54, no. 2, pp. 171–179, 2013.

[118] S.-Y. Park and J. R. Tame, "Seeing the light with bluf proteins," *Biophysical reviews*, vol. 9, no. 2, pp. 169–176, 2017.

[119] M. Iseki, S. Matsunaga, A. Murakami, K. Ohno, K. Shiga, K. Yoshida, M. Sugai, T. Takahashi, T. Hori, and M. Watanabe, "A blue-light-activated adenylyl cyclase mediates photoavoidance in euglena gracilis," *Nature*, vol. 415, no. 6875, pp. 1047–1051, 2002.

[120] S. Masuda and C. E. Bauer, "Appa is a blue light photoreceptor that antirepresses photosynthesis gene expression in rhodobacter sphaeroides," *Cell*, vol. 110, no. 5, pp. 613–623, 2002.

[121] A. Losi and W. Gärtner, "The evolution of flavin-binding photoreceptors: an ancient chromophore serving trendy blue-light sensors," *Annual review of plant biology*, vol. 63, pp. 49–72, 2012.

[122] A. Falciatore and C. Bowler, "The evolution and function of blue and red light photoreceptors," *Current Topics in Developmental Biology*, vol. 68, pp. 317–350, 2005.

[123] A. Lukacs, P. J. Tonge, and S. R. Meech, "Photophysics of the blue light using flavin domain," *Accounts of chemical research*, vol. 55, no. 3, pp. 402–414, 2022.

[124] T. Fujisawa and S. Masuda, "Light-induced chromophore and protein responses and mechanical signal transduction of bluf proteins," *Biophysical reviews*, vol. 10, no. 2, pp. 327–337, 2018.

[125] T. Drepper, U. Krauss, S. Meyer zu Berstenhorst, J. Pietruszka, and K.-E. Jaeger, "Lights on and action! controlling microbial gene expression by light," *Applied microbiology and biotechnology*, vol. 90, no. 1, pp. 23–40, 2011.

[126] B. D. Zoltowski and K. H. Gardner, "Tripping the light fantastic: blue-light photoreceptors as examples of environmentally modulated protein- protein interactions," *Biochemistry*, vol. 50, no. 1, pp. 4–16, 2011.

[127] T. Kanazawa, S. Ren, M. Maekawa, K. Hasegawa, F. Arisaka, M. Hyodo, Y. Hayakawa, H. Ohta, and S. Masuda, "Biochemical and physiological characterization of a bluf protein- eal protein complex involved in blue light-dependent degradation of cyclic diguanylate in the purple bacterium rhodopseudomonas palustris," *Biochemistry*, vol. 49, no. 50, pp. 10 647–10 655, 2010.

[128] N. Tschowri, S. Lindenberg, and R. Hengge, "Molecular function and potential evolution of the biofilm-modulating blue light-signalling pathway of escherichia coli," *Molecular microbiology*, vol. 85, no. 5, pp. 893–906, 2012.

[129] I. Chitrakar, J. N. Iuliano, Y. He, H. A. Woroniecka, J. Tolentino Collado, J. M. Wint, S. G. Walker, P. J. Tonge, and J. B. French, "Structural basis for the regulation of biofilm formation and iron uptake in a. baumannii by the blue-light-using photoreceptor, blsa," *ACS infectious diseases*, vol. 6, no. 10, pp. 2592–2603, 2020.

[130] S. Shi, W. Qi, J. Zhang, C. Liang, W. Liu, H. Han, W. Zhuang, T. Chen, W. Sun, and Y. Chen, "Proteotranscriptomic analysis reveals the mechanisms underlying escherichia coli phenotypic shifts under blue light," *Biotechnology and Bioengineering*, vol. 122, no. 5, pp. 1258–1271, 2025.

[131] A. Jung, T. Domratcheva, M. Tarutina, Q. Wu, W.-h. Ko, R. L. Shoeman, M. Gomelsky, K. H. Gardner, and I. Schlichting, "Structure of a bacterial bluf photoreceptor: insights into blue light-mediated signal transduction," *Proceedings of the National Academy of Sciences*, vol. 102, no. 35, pp. 12 350–12 355, 2005.

[132] G. Y. Fraikin, M. Strakhovskaya, N. Belenikina, and A. Rubin, "Lov and bluf flavoproteins' regulatory photoreceptors of microorganisms and photosensory actuators in optogenetic systems," *Moscow University Biological Sciences Bulletin*, vol. 71, no. 1, pp. 50–57, 2016.

[133] M. S. Kaushik, R. Sharma, S. K. Veetil, S. K. Srivastava, and S. Kateriya, "Modular diversity of the bluf proteins and their potential for the development of diverse optogenetic tools," *Applied Sciences*, vol. 9, no. 18, p. 3924, 2019.

[134] T. Domratcheva, E. Hartmann, I. Schlichting, and T. Kottke, "Evidence for tautomerisation of glutamine in bluf blue light receptors by vibrational spectroscopy and computational chemistry," *Scientific reports*, vol. 6, no. 1, p. 22669, 2016.

[135] K. Karadi, S. M. Kapetanaki, K. Raics, I. Pecsi, R. Kapronczai, Z. Fekete, J. N. Iuliano, J. T. Collado, A. A. Gil, J. Orban *et al.*, "Functional dynamics of a single tryptophan residue in a bluf protein revealed by fluorescence spectroscopy," *Scientific reports*, vol. 10, no. 1, p. 2061, 2020.

[136] Y. Hontani, J. Mehlhorn, T. Domratcheva, S. Beck, M. Kloz, P. Hegemann, T. Mathes, and J. T. Kennis, "Spectroscopic and computational observation of glutamine tautomerization in the blue light sensing using flavin domain photoreaction," *Journal of the American Chemical Society*, vol. 145, no. 2, pp. 1040–1052, 2023.

[137] A. Losi, "Flavin-based blue-light photosensors: a photobiophysics update," *Photochemistry and photobiology*, vol. 83, no. 6, pp. 1283–1300, 2007.

[138] Y. Zhou, X.-W. Kang, Z. Zhou, Z. Chen, S. Zou, S. Tang, B. Wang, K. Wang, D. Zhong, and B. Ding, "Unified mechanism of light-state bluf domain photocycles by capturing proton relay intermediates," *Ultrafast Science*, vol. 4, p. 0072, 2024.

[139] X.-W. Kang, K. Wang, X. Zhang, D. Zhong, and B. Ding, "Elementary reactions in the functional triads of the blue-light photoreceptor bluf domain," *The Journal of Physical Chemistry B*, vol. 128, no. 9, pp. 2065–2075, 2024.

[140] A. M. Gest, A. Z. Sahan, Y. Zhong, W. Lin, S. Mehta, and J. Zhang, "Molecular Spies in Action: Genetically Encoded Fluorescent Biosensors Light up Cellular Signals," *Chem. Rev.*, vol. 124, no. 22, pp. 12 573–12 660, 2024.

[141] B. G. Reid and G. C. Flynn, "Chromophore formation in green fluorescent protein," *Biochemistry*, vol. 36, no. 22, pp. 6786–6791, 1997.

[142] F. G. Prendergast and K. G. Mann, "Chemical and Physical Properties of Aequorin and the Green Fluorescent Protein Isolated From Aequorea Forskalea," *Biochemistry*, vol. 17, no. 17, pp. 3448–3453, 1978.

[143] M. Chalfie, "Green Fluorescent Protein," *Photochem. Photobiol.*, vol. 62, no. 4, pp. 651–656, 1995.

[144] S. J. Remington, "Green Fluorescent Protein: A Perspective," *Protein Sci.*, vol. 20, no. 9, pp. 1509–1519, 2011.

[145] M. Chalfie and S. R. Kain, *Green Fluorescent Protein: Properties, Applications and Protocols*. John Wiley & Sons, 2005.

[146] R. A. Cinelli, V. Tozzini, V. Pellegrini, F. Beltram, G. Cerullo, M. Zavelani-Rossi, S. De Silvestri, M. Tyagi, and M. Giacca, "Coherent Dynamics of Photoexcited Green Fluorescent Proteins," *Phys. Rev. Lett.*, vol. 86, no. 15, p. 3439, 2001.

[147] S. Shi, P. Kumar, and K. F. Lee, "Generation of Photonic Entanglement in Green Fluorescent Proteins," *Nat. Commun.*, vol. 8, no. 1, p. 1934, 2017.

[148] K. Brejc, T. K. Sixma, P. A. Kitts, S. R. Kain, R. Y. Tsien, M. Ormö, and S. J. Remington, "Structural Basis for Dual Excitation and Photoisomerization of the Aequorea Victoria Green Fluorescent Protein," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 94, no. 6, pp. 2306–2311, 1997.

[149] P. Gaytán and A. Roldán-Salgado, "Photoactivatable Blue Fluorescent Protein," *ACS Omega*, vol. 9, no. 26, pp. 28 577–28 582, 2024.

[150] J. Kwon, J.-S. Park, M. Kang, S. Choi, J. Park, G. T. Kim, C. Lee, S. Cha, H.-W. Rhee, and S.-H. Shim, "Bright Ligand-Activatable Fluorescent Protein for High-quality Multicolor Live-cell Super-Resolution Microscopy," *Nat. Commun.*, vol. 11, no. 1, p. 273, 2020.

[151] M. W. Hoorens and W. Szymanski, "Reversible, spatial and temporal control over protein activity using light," *Trends in Biochemical Sciences*, vol. 43, no. 8, pp. 567–575, 2018.

[152] J. M. Reis, D. C. Burns, and G. A. Woolley, "Optical control of protein–protein interactions via blue light-induced domain swapping," *Biochemistry*, vol. 53, no. 30, pp. 5008–5016, 2014.

[153] K. E. Brechun, K. M. Arndt, and G. A. Woolley, "Strategies for the photo-control of endogenous protein activity," *Current opinion in structural biology*, vol. 45, pp. 53–58, 2017.

[154] X. X. Zhou, H. K. Chung, A. J. Lam, and M. Z. Lin, "Optical control of protein activity by fluorescent protein domains," *Science*, vol. 338, no. 6108, pp. 810–814, 2012.

[155] L. Benedetti, A. E. Barentine, M. Messa, H. Wheeler, J. Bewersdorf, and P. De Camilli, "Light-activated protein interaction with high spatial subcellular confinement," *Proceedings of the National Academy of Sciences*, vol. 115, no. 10, pp. E2238–E2245, 2018.

[156] U. Galindo-García, M. Vanegas-Reza, R. Arreguín-Espinosa, K. S. Pérez, R. Pérez-Solis, M. E. Mendoza, K. Y. Cervantes-Quintero, S. R. Islas, M. Cuéllar-Cruz, and A. Moreno, "A short and practical overview on light-sensing proteins, optogenetics, and fluorescent biomolecules inside biomorphs used as optical sensors," *Crystals*, vol. 13, no. 9, p. 1343, 2023.

[157] L. Z. Fan and M. Z. Lin, "Optical control of biological processes by light-switchable proteins," *Wiley Interdisciplinary Reviews: Developmental Biology*, vol. 4, no. 5, pp. 545–554, 2015.

[158] O. Bozovic, B. Jankovic, and P. Hamm, "Using azobenzene photocontrol to set proteins in motion," *Nature Reviews Chemistry*, vol. 6, no. 2, pp. 112–124, 2022.

[159] R. J. Mart and R. K. Allemann, "Azobenzene photocontrol of peptides and proteins," *Chemical Communications*, vol. 52, no. 83, pp. 12 262–12 277, 2016.

[160] A. A. Beharry and G. A. Woolley, "Azobenzene photoswitches for biomolecules," *Chemical Society Reviews*, vol. 40, no. 8, pp. 4422–4437, 2011.

[161] M. Zhu and H. Zhou, "Azobenzene-based small molecular photoswitches for protein modulation," *Organic & biomolecular chemistry*, vol. 16, no. 44, pp. 8434–8445, 2018.

[162] M. Pantusa, R. Bartucci, and B. Rizzuti, "Stability of trans-resveratrol associated with transport proteins," *Journal of agricultural and food chemistry*, vol. 62, no. 19, pp. 4384–4391, 2014.

[163] M. A. v. d. Horst, T. P. Stalcup, S. Kaledhonkar, M. Kumauchi, M. Hara, A. Xie, K. J. Hellingwerf, and W. D. Hoff, "Locked Chromophore Analogs Reveal That Photoactive Yellow Protein Regulates Biofilm Formation in the Deep Sea Bacterium Idiomarina Loihiensis," *J. Am. Chem. Soc.*, vol. 131, no. 47, pp. 17 443–17 451, 2009.

[164] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden, "BLAST+: Architecture and Applications," *BMC Bioinf.*, vol. 10, pp. 1–9, 2009.

[165] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding *et al.*, "Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega," *Mol. Syst. Biol.*, vol. 7, no. 1, p. 539, 2011.

[166] M. Blum, A. Andreeva, L. C. Florentino, S. R. Chuguransky, T. Grego, E. Hobbs, B. L. Pinto, A. Orr, T. Paysan-Lafosse, I. Ponamareva *et al.*, "Interpro: the Protein Sequence Classification Resource in 2025," *Nucleic Acids Res.*, vol. 53, no. D1, pp. D444–D456, 2025.

[167] E. Coudert, S. Gehant, E. De Castro, M. Pozzato, D. Baratin, T. Neto, C. J. Sigrist, N. Redaschi, and A. Bridge, "Annotation of Biologically Relevant Ligands in UniProtKB using ChEBI," *Bioinformatics*, vol. 39, no. 1, p. btac793, 2023.

[168] A. Bateman, M.-J. Martin, S. Orchard, M. Magrane, A. Adesina, S. Ahmad, E. Bowler-Barnett, H. Bye-A-Jee, D. Carpentier, P. Denny *et al.*, "UniProt: the Universal Protein Knowledgebase in 2025," *Nucleic Acids Res.*, 2024.

[169] J. A. Kyndt, T. E. Meyer, and M. A. Cusanovich, "Photoactive Yellow Protein, Bacteriophytochrome, and Sensory Rhodopsin in Purple Phototrophic Bacteria," *Photochem. Photobiol. Sci.*, vol. 3, pp. 519–530, 2004.

[170] M. Khatun, W. Shoombuatong, M. M. Hasan, H. Kurata *et al.*, "Evolution of Sequence-Based Bioinformatics Tools for Protein-protein Interaction Prediction," *Curr. Genomics*, vol. 21, no. 6, pp. 454–463, 2020.

[171] K. Tamura, G. Stecher, and S. Kumar, "MEGA11: Molecular Evolutionary Genetics Analysis (version 11)," *Mol. Biol. Evol.*, vol. 38, no. 7, pp. 3022–3027, 2021.

[172] S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel, "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0," *Syst. Biol.*, vol. 59, no. 3, pp. 307–321, 2010.

[173] J. Xing, V. M. Gumerov, and I. B. Zhulin, "Photoactive yellow protein represents a distinct, evolutionarily novel family of pas domains," *Journal of bacteriology*, vol. 204, no. 11, pp. e00 300–22, 2022.

[174] L. A. Abriata, "The nobel prize in chemistry: past, present, and future of ai in biology," *Communications Biology*, vol. 7, no. 1, p. 1409, 2024.

[175] E. Callaway, "Developers of alphafold win chemistry nobel," *Nature*, vol. 634, pp. 525–526, 2024.

[176] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko *et al.*, "Highly Accurate Protein Structure Prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.

[177] J. Abbass and J.-C. Nebel, "Rosetta and the journey to predict proteins' structures, 20 years on," *Current Bioinformatics*, vol. 15, no. 6, pp. 611–628, 2020.

[178] M. Bertoni, F. Kiefer, M. Biasini, L. Bordoli, and T. Schwede, "Modeling Protein Quaternary Structure of Homo-and Hetero-oligomers Beyond Binary Interactions by Homology," *Sci. Rep.*, vol. 7, no. 1, p. 10480, 2017.

[179] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli *et al.*, "SWISS-MODEL: Homology Modelling of Protein Structures and Complexes," *Nucleic Acids Res.*, vol. 46, no. W1, pp. W296–W303, 2018.

[180] M. T. Degiacomi, "Coupling molecular dynamics and deep learning to mine protein conformational space," *Structure*, vol. 27, no. 6, pp. 1034–1040, 2019.

[181] F. Noé, G. De Fabritiis, and C. Clementi, "Machine learning for protein folding and dynamics," *Current opinion in structural biology*, vol. 60, pp. 77–84, 2020.

[182] S. Matsumoto, S. Ishida, M. Araki, T. Kato, K. Terayama, and Y. Okuno, "Extraction of protein dynamics information from cryo-em maps using deep learning," *Nature Machine Intelligence*, vol. 3, no. 2, pp. 153–160, 2021.

[183] M. Audagnotto, W. Czechtizky, L. De Maria, H. Käck, G. Papoian, L. Tornberg, C. Tyrchan, and J. Ulander, "Machine learning/molecular dynamic protein structure prediction approach to investigate the protein conformational ensemble," *Scientific Reports*, vol. 12, no. 1, p. 10018, 2022.

[184] G. Janson and M. Feig, "Generation of protein dynamics by machine learning," *Current Opinion in Structural Biology*, vol. 93, p. 103115, 2025.

[185] H. Bao, M. R. Melnicki, E. G. Pawlowski, M. Sutter, M. Agostoni, S. Lechno-Yossef, F. Cai, B. L. Montgomery, and C. A. Kerfeld, "Additional Families of Orange Carotenoid Proteins in the Photoprotective System of Cyanobacteria," *Nat. Plants*, vol. 3, no. 8, pp. 1–11, 2017.

[186] F. Muzzopappa, A. Wilson, V. Yogarajah, S. Cot, F. Perreau, C. Montigny, C. Bourcier de Carbon, and D. Kirilovsky, "Paralogs of the C-terminal Domain of the Cyanobacterial Orange Carotenoid Protein are Carotenoid Donors to Helical Carotenoid Proteins," *Plant Physiol.*, vol. 175, no. 3, pp. 1283–1303, 2017.

[187] R. Nadella and D. Hernandez-Baltazar, "Identification of New Protein Sources for Renewable Energy Storage Systems: A Bioinformatic Study," in *AIP Conf. Proc.*, vol. 1992, no. 1. AIP Publishing, 2018.

[188] M. Hasegawa, T. Hosaka, K. Kojima, Y. Nishimura, Y. Nakajima, T. Kimura-Someya, M. Shirouzu, Y. Sudo, and S. Yoshizawa, "A Unique Clade of Light-Driven Proton-Pumping Rhodopsins Evolved in the Cyanobacterial Lineage," *Sci. Rep.*, vol. 10, no. 1, p. 16752, 2020.

[189] M. Karplus and G. A. Petsko, "Molecular Dynamics Simulations in Biology," *Nature*, vol. 347, no. 6294, pp. 631–639, 1990.

[190] T. Hansson, C. Oostenbrink, and W. van Gunsteren, "Molecular Dynamics Simulations," *Curr. Opin. Struct. Biol.*, vol. 12, no. 2, pp. 190–196, 2002.

[191] B. Leimkuhler and C. Matthews, "Molecular Dynamics," *Interdiscip. Appl. Math.*, vol. 39, no. 1, 2015.

[192] E. Shchukin and V. Yushchenko, "Molecular Dynamics Simulation of Mechanical Behaviour," *J. Mat. Sci.*, vol. 16, pp. 313–330, 1981.

[193] T.-x. Xiang and B. D. Anderson, "Molecular Distributions in Interphases: Statistical Mechanical Theory Combined With Molecular Dynamics Simulation of a Model Lipid Bilayer," *Biophys. J.*, vol. 66, no. 3, pp. 561–572, 1994.

[194] M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, 2nd ed. Oxford University Press, 2023.

[195] M. Levitt, "Birth and Future of Multiscale Modeling for Macromolecular Systems (Nobel Lecture)," *Angew. Chem. Int. Ed.*, vol. 53, no. 38, pp. 10 006–10 018, 2014.

[196] M. Karplus, "Development of Multiscale Models for Complex Chemical Systems: From H+$H_2$2 to Biomolecules (Nobel Lecture)," *Angew. Chem. Int. Ed.*, vol. 53, no. 38, 2014.

[197] O. Guvench and A. D. MacKerell Jr, "Comparison of protein force fields for molecular dynamics simulations," *Molecular modeling of proteins*, pp. 63–88, 2008.

[198] M. A. González, "Force Fields and Molecular Dynamics Simulations," *École thématique de la Société Française de la Neutronique*, vol. 12, pp. 169–200, 2011.

[199] L. Monticelli and D. P. Tieleman, "Force fields for classical molecular dynamics," *Biomolecular simulations: Methods and protocols*, pp. 197–213, 2012.

[200] S. Hug, "Classical molecular dynamics in a nutshell," *Biomolecular simulations: Methods and protocols*, pp. 127–152, 2012.

[201] S. A. Hollingsworth and R. O. Dror, "Molecular Dynamics Simulation for All," *Neuron*, vol. 99, no. 6, pp. 1129–1143, 2018.

[202] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha *et al.*, "All-atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins," *J. Phys. Chem. B*, vol. 102, no. 18, pp. 3586–3616, 1998.

[203] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, 1995.

[204] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-atom Force Field on Conformational Energetics and Properties of Organic Liquids," *J. Am. Chem. Soc.*, vol. 118, no. 45, pp. 11 225–11 236, 1996.

[205] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, "A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: the Gromos Force-field Parameter Sets 53a5 and 53a6," *J. Comput. Chem.*, vol. 25, no. 13, pp. 1656–1676, 2004.

[206] Y. Jin, L. O. Johannissen, and S. Hay, "Predicting New Protein Conformations from Molecular Dynamics Simulation Conformational Landscapes and Machine Learning," *Proteins: Struct., Funct., Bioinf.*, vol. 89, no. 8, pp. 915–921, 2021.

[207] R. Baron and J. A. McCammon, "Molecular Recognition and Ligand Association," *Annu. Rev. Phys. Chem.*, vol. 64, pp. 151–175, 2013.

[208] J. Rydzewski and W. Nowak, "Ligand diffusion in proteins via enhanced sampling in molecular dynamics," *Physics of Life Reviews*, vol. 22, pp. 58–74, 2017.

[209] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1, pp. 19–25, 2015.

[210] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable Molecular Dynamics with NAMD," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, 2005.

[211] D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzesh, T. J. Giese, A. W. Götz, H. Gohlke *et al.*, "AmberTools," *J. Chem. Inf. Model.*, vol. 63, no. 20, pp. 6183–6191, 2023.

[212] D. Janezic and F. Merzel, "An efficient symplectic integration algorithm for molecular dynamics simulations," *Journal of chemical information and computer sciences*, vol. 35, no. 2, pp. 321–326, 1995.

[213] M. G. Paterlini and D. M. Ferguson, "Constant temperature simulations using the langevin equation with velocity verlet integration," *Chemical Physics*, vol. 236, no. 1-3, pp. 243–252, 1998.

[214] B. J. Leimkuhler, S. Reich, and R. D. Skeel, "Integration methods for molecular dynamics," in *Mathematical Approaches to biomolecular structure and dynamics*. Springer, 1996, pp. 161–185.

[215] F. Pietrucci, "Strategies for the exploration of free energy landscapes: Unity in diversity and challenges ahead," *Reviews in Physics*, vol. 2, pp. 32–45, 2017.

[216] K. L. Breyfogle, D. L. Blood, A. M. Rosnik, and B. P. Krueger, "Molecular Dynamics Force Field Parameters for the EGFP Chromophore and Some of Its Analogues," *J. Phys. Chem. B*, vol. 127, no. 26, pp. 5772–5788, 2023.

[217] K. Claridge and A. Troisi, "Developing Consistent Molecular Dynamics Force Fields for Biological Chromophores via Force Matching," *J. Phys. Chem. B*, vol. 123, no. 2, pp. 428–438, 2018.

[218] F.-Y. Lin and A. D. MacKerell, *Force Fields for Small Molecules*. Springer, 2019, pp. 21–54.

[219] X. He, B. Walker, V. H. Man, P. Ren, and J. Wang, "Recent Progress in General Force Fields of Small Molecules," *Curr. Opin. Struct. Biol.*, vol. 72, pp. 187–193, 2022.

[220] Y. Sugita and Y. Okamoto, "Replica-Exchange Molecular Dynamics Method for Protein Folding," *Chemi. Phys. Lett.*, vol. 314, no. 1-2, pp. 141–151, 1999.

[221] G. M. Torrie and J. P. Valleau, "Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling," *J. Comp. Phys.*, vol. 23, no. 2, pp. 187–199, 1977.

[222] A. Laio and M. Parrinello, "Escaping Free-Energy Minima," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, no. 20, pp. 12 562–12 566, 2002.

[223] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "PLUMED 2: New Feathers for an Old Bird," *Comp. Phys. Commun.*, vol. 185, no. 2, pp. 604–613, 2014.

[224] T. P. consortium, "Promoting transparency and reproducibility in enhanced molecular simulations," *Nature methods*, vol. 16, no. 8, pp. 670–673, 2019.

[225] G. A. Tribello, M. Bonomi, G. Bussi, C. Camilloni *et al.*, "PLUMED Tutorials: A Collaborative, Community-Driven Learning Ecosystem," *J. Chem. Phys.*, vol. 162, no. 9, 2025.

[226] E. Rosta and G. Hummer, "Error and efficiency of replica exchange molecular dynamics simulations," *The Journal of chemical physics*, vol. 131, no. 16, 2009.

[227] D. J. Sindhikara, D. J. Emerson, and A. E. Roitberg, "Exchange often and properly in replica exchange molecular dynamics," *Journal of Chemical Theory and Computation*, vol. 6, no. 9, pp. 2804–2808, 2010.

[228] J. Vreede, K. J. Hellingwerf, and P. G. Bolhuis, "Helix Formation is a Dynamical Bottleneck in the Recovery Reaction of Photoactive Yellow Protein," *Proteins: Struct., Funct., Bioinf.*, vol. 72, no. 1, pp. 136–149, 2008.

[229] A. Ganguly, W. Thiel, and B. R. Crane, "Glutamine Amide Flip Elicits Long Distance Allosteric Responses in the LOV Protein Vivid," *J. Am. Chem. Soc.*, vol. 139, no. 8, pp. 2972–2980, 2017.

[230] M. Bondanza, L. Cupellini, P. Faccioli, and B. Mennucci, "Molecular Mechanisms of Activation in the Orange Carotenoid Protein Revealed by Molecular Dynamics," *J. Am. Chem. Soc.*, vol. 142, no. 52, pp. 21 829–21 841, 2020.

[231] M. Nottoli, M. Bondanza, F. Lipparini, and B. Mennucci, "An Enhanced Sampling QM/AMOEBA Approach: The Case of the Excited State Intramolecular Proton Transfer in Solvated 3-Hydroxyflavone," *J. Chem. Phys.*, vol. 154, no. 18, 2021.

[232] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 1, no. 5, pp. 826–843, 2011.

[233] L. Sutto, S. Marsili, and F. L. Gervasio, "New Advances in Metadynamics," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 2, no. 5, pp. 771–779, 2012.

[234] D. Ray and M. Parrinello, "Kinetics From Metadynamics: Principles, Applications, and Outlook," *J. Chem. Theory Comput.*, vol. 19, no. 17, pp. 5649–5670, 2023.

[235] M. Bondanza, L. Cupellini, F. Lipparini, and B. Mennucci, "The Multiple Roles of the Protein in the Photoactivation of Orange Carotenoid Protein," *Chem*, vol. 6, no. 1, pp. 187–203, 2020.

[236] U. Raucci, D. M. Sanchez, T. J. Martínez, and M. Parrinello, "Enhanced Sampling Aided Design of Molecular Photoswitches," *J. Am. Chem. Soc.*, vol. 144, no. 42, pp. 19 265–19 271, 2022.

[237] O. Valsson and M. Parrinello, "Variational approach to enhanced sampling and free energy calculations," *Physical review letters*, vol. 113, no. 9, p. 090601, 2014.

[238] E. Runge and E. K. Gross, "Density-functional theory for time-dependent systems," *Physical review letters*, vol. 52, no. 12, p. 997, 1984.

[239] D. Jacquemin, B. Mennucci, and C. Adamo, "Excited-state calculations with td-dft: from benchmarks to simulations in complex environments," *Physical chemistry chemical physics*, vol. 13, no. 38, pp. 16 987–16 998, 2011.

[240] S. Larsson, "Applications of casscf," *International Journal of Quantum Chemistry*, vol. 111, no. 13, pp. 3424–3430, 2011.

[241] T. Kubař, M. Elstner, and Q. Cui, "Hybrid Quantum Mechanical/Molecular Mechanical Methods for Studying Energy Transduction in Biomolecular Machines," *Annu. Rev. Biophys.*, vol. 52, no. 1, pp. 525–551, 2023.

[242] M. Manathunga, A. W. Götz, and K. M. Merz Jr, "Computer-Aided Drug Design, Quantum-Mechanical Methods for Biological Problems," *Curr. Opin. Struct. Biol.*, vol. 75, p. 102417, 2022.

[243] H. Tandon, T. Chakraborty, and V. Suhag, "A Brief Review on Importance of DFT in Drug Design," *Res. Med. Eng. Sci*, vol. 7, no. 4, pp. 791–795, 2019.

[244] M. Elstner, T. Frauenheim, and S. Suhai, "An Approximate DFT Method for QM/MM Simulations of Biological Structures and Processes," *J. Mol. Struct.*, vol. 632, no. 1-3, pp. 29–41, 2003.

[245] R. Sarkar, M. Boggio-Pasqua, P.-F. Loos, and D. Jacquemin, "Benchmarking td-dft and wave function methods for oscillator strengths and excited-state dipole moments," *Journal of Chemical Theory and Computation*, vol. 17, no. 2, pp. 1117–1132, 2021.

[246] S. J. Bintrim and K. Carter-Fenk, "Optimal-reference excited state methods: Static correlation at polynomial cost with single-reference coupled-cluster approaches," *Journal of Chemical Theory and Computation*, vol. 21, no. 8, pp. 4080–4094, 2025.

[247] G. Cardenas and J. J. Nogueira, "An Algorithm to Correct for the CASSCF Active Space in Multiscale QM/MM Calculations Based on Geometry Ensembles," *Int. J. Quantum Chem.*, vol. 121, no. 6, p. e26533, 2021.

[248] H. M. Senn and W. Thiel, "QM/MM Methods for Biomolecular Systems," *Angew. Chem. Int. Ed.*, vol. 48, no. 7, pp. 1198–1229, 2009.

[249] J. W. Park and T. Shiozaki, "On-the-fly CASPT2 Surface-Hopping Dynamics," *J. Chem. Theory Comput.*, vol. 13, no. 8, pp. 3676–3683, 2017.

[250] A. G. Rao and I. Schapiro, "Photoisomerization of Phytochrome Chromophore Models: an XMS-CASPT2 Study," *Phys. Chem. Chem. Phys.*, vol. 24, no. 48, pp. 29 393–29 405, 2022.

[251] L. Barneschi, D. Kaliakin, M. Huix-Rotllant, N. Ferré, M. Filatov, and M. Olivucci, "Assessment of the Electron Correlation Treatment on the Quantum-Classical Dynamics of Retinal Protonated Schiff Base Models: XMS-CASPT2, RMS-CASPT2, and REKS Methods," *J. Chem. Theory Comput.*, vol. 19, no. 22, pp. 8189–8200, 2023.

[252] S. Sirimatayanant and T. Andruniów, "Tuning Two-Photon Absorption in Rhodopsin Chromophore via Backbone Modification: The Story Told by CC2 and TD-DFT," *J. Chem. Theory Comput.*, vol. 20, no. 18, pp. 8118–8126, 2024.

[253] O. Christiansen, H. Koch, and P. Jørgensen, "The second-order approximate coupled cluster singles and doubles model cc2," *Chemical Physics Letters*, vol. 243, no. 5-6, pp. 409–418, 1995.

[254] J. Schirmer, "Beyond the random-phase approximation: A new approximation scheme for the polarization propagator," *Physical Review A*, vol. 26, no. 5, p. 2395, 1982.

[255] A. Dreuw, "Quantum chemical methods for the investigation of photoinitiated processes in biological systems: Theory and applications," *ChemPhysChem*, vol. 7, no. 11, pp. 2259–2274, 2006.

[256] K. Jóźwiak, P. Ogrin, T. Urbic, and A. Filarowski, "Molecular Dynamics and Density Functional Theory Studies of Conformational Stability of Bilirubin and Biliverdin," *J. Mol. Liquids*, vol. 391, p. 123287, 2023.

[257] S. Stoll, A. Gunn, M. Brynda, W. Sughrue, A. C. Kohler, A. Ozarowski, A. J. Fisher, J. C. Lagarias, and R. D. Britt, "Structure of the Biliverdin Radical Intermediate in Phycocyanobilin: Ferredoxin Oxidoreductase Identified by High-Field EPR and DFT," *J. Am. Chem. Soc.*, vol. 131, no. 5, pp. 1986–1995, 2009.

[258] I. V. Polyakov, B. L. Grigorenko, V. A. Mironov, and A. V. Nemukhin, "Modeling Structure and Excitation of Biliverdin-Binding Domains in Infrared Fluorescent Proteins," *Chem. Phys. Lett.*, vol. 710, pp. 59–63, 2018.

[259] E. C. Semidalas and C. E. Semidalas, "Structure and Vibrational Spectra of p-Coumaric Acid Dimers by DFT Methods," *Vib. Spectrosc.*, vol. 101, pp. 100–108, 2019.

[260] A. Garzon, I. Bravo, A. J. Barbero, and J. Albaladejo, "Mechanistic and Kinetic Study on the Reactions of Coumaric Acids with Reactive Oxygen Species: A DFT Approach," *J. Agric. Food Chem.*, vol. 62, no. 40, pp. 9705–9710, 2014.

[261] N. Kumar, V. Pruthi, and N. Goel, "Structural, Thermal and Quantum Chemical Studies of p-Coumaric and Caffeic Acids," *J. Molec. Struct.*, vol. 1085, pp. 242–248, 2015.

[262] M. Huix-Rotllant, M. Filatov, S. Gozem, I. Schapiro, M. Olivucci, and N. Ferré, "Assessment of Density Functional Theory for Describing the Correlation Effects on the Ground and excited state potential energy surfaces of a retinal chromophore model," *J. Chem. Theory Comput.*, vol. 9, no. 9, pp. 3917–3932, 2013.

[263] F. Blomgren and S. Larsson, "Exploring the Potential Energy Surface of Retinal, a Comparison of the Performance of Different Methods," *J. Comput. Chem.*, vol. 26, no. 7, pp. 738–742, 2005.

[264] A. Warshel and M. Levitt, "Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme," *J. Mol. Biol.*, vol. 103, no. 2, pp. 5227–249, 1976.

[265] A. Altun, S. Yokoyama, and K. Morokuma, "Quantum Mechanical/Molecular Mechanical Studies on Spectral Tuning Mechanisms of Visual Pigments and Other Photoactive Proteins," *Photochem. Photobiol.*, vol. 84, no. 4, pp. 845–854, 2008.

[266] F. Neese, "The ORCA Program System," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 2, no. 1, pp. 73–78, 2012.

[267] R. Dennington, T. A. Keith, and J. M. Millam, "GaussView Version 6," 2019, semichem Inc. Shawnee Mission KS.

[268] Y. Lu, M. R. Farrow, P. Fayon, A. J. Logsdail, A. A. Sokol, C. R. A. Catlow, P. Sherwood, and T. W. Keal, "Open-Source, Python-Based Redevelopment of the ChemShell Multiscale QM/MM Environment," *J. Chem. Theory Comput.*, vol. 15, no. 2, pp. 1317–1328, 2018.

[269] T. Okamoto, K. Yamada, Y. Koyano, T. Asada, N. Koga, and M. Nagaoka, "A Minimal Implementation of the AMBER–GAUSSIAN Interface for Ab Initio QM/MM-MD Simulation," *J. Comput. Chem.*, vol. 32, no. 5, pp. 932–942, 2011.

[270] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann *et al.*, "CP2K: An Electronic Structure and Molecular Dynamics Software Package-Quickstep: Efficient and Accurate Electronic Structure Calculations," *J. Chem. Phys.*, vol. 152, no. 19, 2020.

[271] T. Laino, F. Mohamed, A. Laio, and M. Parrinello, "An Efficient Real Space Multigrid QM/MM Electrostatic Coupling," *J. Chem. Theory Comput.*, vol. 1, no. 6, pp. 1176–1184, 2005.

[272] V. W. D. Cruzeiro, Y. Wang, E. Pieri, E. G. Hohenstein, and T. J. Martínez, "TeraChem Protocol Buffers (TCPB): Accelerating QM and QM/MM Simulations with a Client–Server Model," *J. Chem. Phys.*, vol. 158, no. 4, 2023.

[273] C. M. Isborn, A. W. Gotz, M. A. Clark, R. C. Walker, and T. J. Martínez, "Electronic Absorption Spectra From MM and Ab Initio QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein," *J. Chem. Theory Comput.*, vol. 8, no. 12, pp. 5092–5106, 2012.

[274] G. Groenhof, "Introduction to QM/MM Simulations," *Biomolecular Simulations: Methods and Protocols*, pp. 43–66, 2013.

[275] H. Lin and D. G. Truhlar, "Redistributed Charge and Dipole Schemes for Combined Quantum Mechanical and Molecular Mechanical Calculations," *J. Phys. Chem. A*, vol. 109, no. 17, pp. 3991–4004, 2005.

[276] H. M. Senn and W. Thiel, "QM/MM Studies of Enzymes," *Curr. Opin. Chem. Biol.*, vol. 11, no. 2, pp. 182–187, 2007.

[277] S. Roßbach and C. Ochsenfeld, "Influence of Coupling and Embedding Schemes on QM Size Convergence in QM/MM Approaches for the Example of a Proton Transfer in DNA," *J. Chem. Theory Comput.*, vol. 13, no. 3, pp. 1102–1107, 2017.

[278] R. Khare, S. L. Mielke, G. C. Schatz, and T. Belytschko, "Multiscale Coupling Schemes Spanning the Quantum Mechanical, Atomistic Forcefield, and Continuum Regimes," *Comput. Methods Appl. Mech. Eng.*, vol. 197, no. 41-42, pp. 3190–3202, 2008.

[279] A. Humeniuk and W. J. Glover, "Multistate, Polarizable QM/MM Embedding Scheme Based on the Direct Reaction Field Method: Solvatochromic Shifts, Analytical Gradients and Optimizations of Conical Intersections in Solution," *J. Chem. Theory Comput.*, vol. 20, no. 5, pp. 2111–2126, 2024.

[280] R. B. Murphy, D. M. Philipp, and R. A. Friesner, "Frozen Orbital QM/MM Methods for Density Functional Theory," *Chem. Phys. Lett.*, vol. 321, no. 1-2, pp. 113–120, 2000.

[281] G. A. Bramley, O. T. Beynon, P. V. Stishenko, and A. J. Logsdail, "The Application of QM/MM Simulations in Heterogeneous Catalysis," *Phys. Chem. Chem. Phys.*, vol. 25, no. 9, pp. 6562–6585, 2023.

[282] Q. Sun and G. K.-L. Chan, "Exact and Optimal Quantum Mechanics/Molecular Mechanics Boundaries," *J. Chem. Theory Comput.*, vol. 10, no. 9, pp. 3784–3790, 2014.

[283] H. C. Watanabe and Q. Cui, "Quantitative Analysis of QM/MM Boundary Artifacts and Correction in Adaptive QM/MM Simulations," *J. Chem. Theory Comput.*, vol. 15, no. 7, pp. 3917–3928, 2019.

[284] E. Boulanger and J. N. Harvey, "QM/MM Methods for Free Energies and Photochemistry," *Curr. Opin. Struct. Biol.*, vol. 49, pp. 72–76, 2018.

[285] R. Liang, F. Liu, and T. J. Martínez, "Nonadiabatic Photodynamics of Retinal Protonated Schiff Base in Channelrhodopsin 2," *J. Phys. Chem. Lett.*, vol. 10, no. 11, pp. 2862–2868, 2019.

[286] J. Wen, S. Mai, and L. González, "Excited-state Dynamics Simulations of a Light-driven Molecular Motor in Solution," *J. Phys. Chem. A*, vol. 127, no. 45, pp. 9520–9529, 2023.

[287] A. R. Murphy, M. A. Hix, and A. R. Walker, "Exploring the Effects of Mutagenesis on FusionRed by Using Excited-State QM/MM Dynamics and Classical Force Field Simulations," *ChemBioChem*, vol. 24, no. 12, p. e202200799, 2023.

[288] D. A. Tracy, S. Fernandez-Alberti, J. F. Galindo, S. Tretiak, and A. E. Roitberg, "Nonadiabatic Excited-State Molecular Dynamics with an Explicit Solvent: NEXMD-SANDER Implementation," *J. Phys. Chem. B*, vol. 128, no. 46, pp. 11 426–11 434, 2024.

[289] P. Campomanes and S. Vanni, "Protonation Equilibrium in the Active Site of the Photoactive Yellow Protein," *Molecules*, vol. 26, no. 7, p. 2025, 2021.

[290] V. W. D. Cruzeiro, M. Manathunga, K. M. Merz Jr, and A. W. Götz, "Open-Source Multi-GPU-Accelerated QM/MM Simulations with AMBER and QUICK," *J. Chem. Inf. Model.*, vol. 61, no. 5, pp. 2109–2115, 2021.

[291] M. Tsujimura, H. Tamura, K. Saito, and H. Ishikita, "Absorption Wavelength along Chromophore Low-Barrier Hydrogen Bonds," *Iscience*, vol. 25, no. 5, 2022.

[292] M. Manathunga, H. M. Aktulga, A. W. Götz, and K. M. Merz Jr, "Quantum Mechanics/Molecular Mechanics Simulations on NVIDIA and AMD Graphics Processing Units," *J. Chem. Inf. Model.*, vol. 63, no. 3, pp. 711–717, 2023.

[293] I. Dokukina, A. Nenov, M. Garavelli, C. M. Marian, and O. Weingart, "QM/MM Photodynamics of Retinal in the Channelrhodopsin Chimera C1C2 with OM3/MRCI," *ChemPhotoChem*, vol. 3, no. 2, pp. 107–116, 2019.

[294] O. Valsson, P. Campomanes, I. Tavernelli, U. Rothlisberger, and C. Filippi, "Rhodopsin Absorption from First Principles: Bypassing Common Pitfalls," *J. Chem. Theory Comput.*, vol. 9, no. 5, pp. 2441–2454, 2013.

[295] X. Yang, M. Manathunga, S. Gozem, J. Léonard, T. Andruniów, and M. Olivucci, "Quantum–Classical Simulations of Rhodopsin Reveal Excited-State Population Splitting and its Effects on Quantum Efficiency," *Nat. Chem.*, vol. 14, no. 4, pp. 441–449, 2022.

[296] D. Di Prima, L. Pedraza-González, P. Reinholdt, J. Kongsted, and B. Mennucci, "Fluorescent Rhodopsins: A Challenging Test for Cost-Effective QM/MM Approaches," *J. Phys. Chem. A*, 2024.

[297] C. Punwong, J. Owens, and T. J. Martínez, "Direct QM/MM Excited-State Dynamics of Retinal Protonated Schiff Base in Isolation and Methanol Solution," *J. Phys. Chem. B*, vol. 119, no. 3, pp. 704–714, 2015.

[298] P. Malakar, S. Gholami, M. Aarabi, I. Rivalta, M. Sheves, M. Garavelli, and S. Ruhman, "Retinal Photoisomerization versus Counterion Protonation in Light and Dark-Adapted Bacteriorhodopsin and its Primary Photoproduct," *Nat. Commun.*, vol. 15, no. 1, p. 2136, 2024.

[299] F. S. Alavi, M. Gheidi, M. Zahedi, N. Safari, and U. Ryde, "A Novel Mechanism of Heme Degradation to Biliverdin Studied by QM/MM and QM Calculations," *Dalton Trans.*, vol. 47, no. 25, pp. 8283–8291, 2018.

[300] E. Iijima, M. P. Gleeson, M. Unno, and S. Mori, "QM/MM Investigation for Protonation States in a Bilin Reductase PcyA-Biliverdin IXα Complex," *ChemPhysChem*, vol. 19, no. 15, pp. 1809–1813, 2018.

[301] V. Modi, S. Donnini, G. Groenhof, and D. Morozov, "Protonation of the Biliverdin IXα Chromophore in the Red and Far-Red Photoactive States of a Bacteriophytochrome," *J. Phys. Chem. B*, vol. 123, no. 10, pp. 2325–2334, 2019.

[302] S. Santra, R. N. Manna, S. Chakrabarty, and D. Ghosh, "Conformational Effects on the Absorption Spectra of Phytochromes," *J. Phys. Chem. B*, vol. 128, no. 15, pp. 3614–3620, 2024.

[303] J. C. Tully, "Nonadiabatic dynamics," in *Modern methods for multidimensional dynamics computations in chemistry*. World Scientific, 1998, pp. 34–72.

[304] M. Vazdar, M. Eckert-Maksić, M. Barbatti, and H. Lischka, "Excited-state non-adiabatic dynamics simulations of pyrrole," *Molecular physics*, vol. 107, no. 8-12, pp. 845–854, 2009.

[305] H. Nakamura, *Nonadiabatic transition: concepts, basic theories and applications*. World Scientific, 2012.

[306] ——, *Introduction to nonadiabatic dynamics*. World Scientific, 2019.

[307] T. R. Nelson, A. J. White, J. A. Bjorgaard, A. E. Sifain, Y. Zhang, B. Nebgen, S. Fernandez-Alberti, D. Mozyrsky, A. E. Roitberg, and S. Tretiak, "Non-adiabatic excited-state molecular dynamics: Theory and applications for modeling photophysics in extended molecular materials," *Chemical reviews*, vol. 120, no. 4, pp. 2215–2287, 2020.

[308] I. Tavernelli, "Nonadiabatic molecular dynamics simulations: Synergies between theory and experiments," *Accounts of chemical research*, vol. 48, no. 3, pp. 792–800, 2015.

[309] E. Tapavicza, G. D. Bellchambers, J. C. Vincent, and F. Furche, "Ab initio non-adiabatic molecular dynamics," *Physical Chemistry Chemical Physics*, vol. 15, no. 42, pp. 18 336–18 348, 2013.

[310] M. Born and K. Huang, *Dynamical theory of crystal lattices*. Oxford university press, 1996.

[311] J. E. Subotnik, E. C. Alguire, Q. Ou, B. R. Landry, and S. Fatehi, "The requisite electronic structure theory to describe photoexcited nonadiabatic dynamics: Nonadiabatic derivative couplings and diabatic electronic couplings," *Accounts of chemical research*, vol. 48, no. 5, pp. 1340–1350, 2015.

[312] S. Tretiak and S. Mukamel, "Density matrix analysis and simulation of electronic excitations in conjugated and aggregated molecules," *Chemical reviews*, vol. 102, no. 9, pp. 3171–3212, 2002.

[313] A. Piryatinski, M. Stepanov, S. Tretiak, and V. Chernyak, "Semiclassical scattering on conical intersections," *Physical review letters*, vol. 95, no. 22, p. 223001, 2005.

[314] S.-I. Sawada, A. Nitzan, and H. Metiu, "Mean-trajectory approximation for charge-and energy-transfer processes at surfaces," *Physical Review B*, vol. 32, no. 2, p. 851, 1985.

[315] F. Ding, J. J. Goings, H. Liu, D. B. Lingerfelt, and X. Li, "Ab initio two-component ehrenfest dynamics," *The Journal of chemical physics*, vol. 143, no. 11, 2015.

[316] A. W. Jasper and D. G. Truhlar, "Improved treatment of momentum at classically forbidden electronic transitions in trajectory surface hopping calculations," *Chemical physics letters*, vol. 369, no. 1-2, pp. 60–67, 2003.

[317] J. E. Subotnik, "Fewest-switches surface hopping and decoherence in multiple dimensions," *The Journal of Physical Chemistry A*, vol. 115, no. 44, pp. 12 083–12 096, 2011.

[318] G. W. Richings and S. Habershon, "Predicting molecular photochemistry using machine-learning-enhanced quantum dynamics simulations," *Accounts of Chemical Research*, vol. 55, no. 2, pp. 209–220, 2022.

[319] J. Li, M. Vacher, P. O. Dral, and S. A. Lopez, "Machine learning methods in photochemistry and photophysics," in *Theoretical and Computational Photochemistry*. Elsevier, 2023, pp. 163–189.

[320] P. O. Dral and M. Barbatti, "Molecular excited states through a machine learning lens," *Nature Reviews Chemistry*, vol. 5, no. 6, pp. 388–405, 2021.

[321] Y. Xu, D. Verma, R. P. Sheridan, A. Liaw, J. Ma, N. M. Marshall, J. McIntosh, E. C. Sherer, V. Svetnik, and J. M. Johnston, "Deep dive into machine learning models for protein engineering," *Journal of chemical information and modeling*, vol. 60, no. 6, pp. 2773–2790, 2020.

[322] J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. E. Roitberg, "Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning," *Nature communications*, vol. 10, no. 1, p. 2903, 2019.

[323] C.-W. Ju, H. Bai, B. Li, and R. Liu, "Machine learning enables highly accurate predictions of photophysical properties of organic fluorescent materials: Emission wavelengths and quantum yields," *Journal of Chemical Information and Modeling*, vol. 61, no. 3, pp. 1053–1065, 2021.

[324] F. Liu, L. Du, D. Zhang, and J. Gao, "Direct learning hidden excited state interaction patterns from ab initio dynamics and its implication as alternative molecular mechanism models," *Scientific Reports*, vol. 7, no. 1, p. 8737, 2017.

[325] P. Szymaszek, M. Tyszka-Czochara, and J. Ortyl, "Application of photoactive compounds in cancer theranostics: review on recent trends from photoactive chemistry to artificial intelligence," *Molecules*, vol. 29, no. 13, p. 3164, 2024.

[326] P. Kouba, P. Kohout, F. Haddadi, A. Bushuiev, R. Samusevich, J. Sedlar, J. Damborsky, T. Pluskal, J. Sivic, and S. Mazurenko, "Machine learning-guided protein engineering," *ACS catalysis*, vol. 13, no. 21, pp. 13 863–13 895, 2023.

[327] E. F. Morales and H. J. Escalante, "A brief introduction to supervised, unsupervised, and reinforcement learning," in *Biosignal processing and classification using computational learning and intelligence*. Elsevier, 2022, pp. 111–129.

[328] A. Karthikeyan and U. D. Priyakumar, "Artificial intelligence: machine learning for chemical sciences," *Journal of Chemical Sciences*, vol. 134, no. 1, p. 2, 2022.

[329] C. Qu and H. Liu, *Machine Learning in Molecular Sciences*. Springer, 2023.

[330] T. Jiang, J. L. Gradus, and A. J. Rosellini, "Supervised machine learning: a brief primer," *Behavior therapy*, vol. 51, no. 5, pp. 675–687, 2020.

[331] D. S. Wigh, J. M. Goodman, and A. A. Lapkin, "A review of molecular representation in the age of machine learning," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 12, no. 5, p. e1603, 2022.

[332] P. Geurts, A. Irrthum, and L. Wehenkel, "Supervised learning with decision tree-based methods in computational and systems biology," *Molecular Biosystems*, vol. 5, no. 12, pp. 1593–1605, 2009.

[333] A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé, and A. Laio, "Unsupervised learning methods for molecular simulation data," *Chemical Reviews*, vol. 121, no. 16, pp. 9722–9758, 2021.

[334] M. Dodds, J. Guo, T. Löhr, A. Tibo, O. Engkvist, and J. P. Janet, "Sample efficient reinforcement learning with active learning for molecular design," *Chemical Science*, vol. 15, no. 11, pp. 4146–4160, 2024.

[335] X. Liu, J. Zhang, Z. Hou, Y. I. Yang, and Y. Q. Gao, "From predicting to decision making: Reinforcement learning in biomedicine," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 14, no. 4, p. e1723, 2024.

[336] M. Popova, O. Isayev, and A. Tropsha, "Deep reinforcement learning for de novo drug design," *Science advances*, vol. 4, no. 7, p. eaap7885, 2018.

[337] R. Haddad, E. E. Litsa, Z. Liu, X. Yu, D. Burkhardt, and G. Bhisetti, "Targeted molecular generation with latent reinforcement learning," *Scientific Reports*, vol. 15, no. 1, p. 15202, 2025.

[338] A. Dosajh, P. Agrawal, P. Chatterjee, and U. D. Priyakumar, "Modern machine learning methods for protein property prediction," *Current Opinion in Structural Biology*, vol. 90, p. 102990, 2025.

[339] J. Behler, "Four generations of high-dimensional neural network potentials," *Chemical Reviews*, vol. 121, no. 16, pp. 10 037–10 072, 2021.

[340] T. T. Duignan, "The potential of neural network potentials," *ACS Physical Chemistry Au*, vol. 4, no. 3, pp. 232–241, 2024.

[341] R. Martin-Barrios, E. Navas-Conyedo, X. Zhang, Y. Chen, and J. Gulín-González, "An overview about neural networks potentials in molecular dynamics simulation," *International Journal of Quantum Chemistry*, vol. 124, no. 11, p. e27389, 2024.

[342] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Physical review letters*, vol. 98, no. 14, p. 146401, 2007.

[343] J. S. Smith, O. Isayev, and A. E. Roitberg, "Ani-1: an extensible neural network potential with dft accuracy at force field computational cost," *Chemical science*, vol. 8, no. 4, pp. 3192–3203, 2017.

[344] Z. Li, K. Meidani, P. Yadav, and A. Barati Farimani, "Graph neural networks accelerated molecular dynamics," *The Journal of Chemical Physics*, vol. 156, no. 14, 2022.

[345] K. Hansen, D. Baehrens, T. Schroeter, M. Rupp, and K.-R. Müller, "Visual interpretation of kernel-based prediction models," *Molecular Informatics*, vol. 30, no. 9, pp. 817–826, 2011.

[346] Y. Guan, D. Li, S. Xue, and Y. Xi, "Feature-fusion-kernel-based gaussian process model for probabilistic long-term load forecasting," *Neurocomputing*, vol. 426, pp. 174–184, 2021.

[347] A. Strokach and P. M. Kim, "Deep generative modeling for protein design," *Current opinion in structural biology*, vol. 72, pp. 226–236, 2022.

[348] R. L. van den Broek, S. Patel, G. J. van Westen, W. Jespers, and W. Sherman, "In search of beautiful molecules: a perspective on generative modeling for drug design," *Journal of chemical information and modeling*, vol. 65, no. 18, pp. 9383–9397, 2025.

[349] J. B. Ingraham, M. Baranov, Z. Costello, K. W. Barber, W. Wang, A. Ismail, V. Frappier, D. M. Lord, C. Ng-Thow-Hing, E. R. Van Vlack *et al.*, "Illuminating protein space with a programmable generative model," *Nature*, vol. 623, no. 7989, pp. 1070–1078, 2023.

[350] P. Scheerer, N. Michael, J. H. Park, S. Nagano, H.-W. Choe, K. Inomata, B. Borucki, N. Krauß, and T. Lamparter, "Light-induced conformational changes of the chromophore and the protein in phytochromes: Bacterial phytochromes as model systems," *ChemPhysChem*, vol. 11, no. 6, pp. 1090–1105, 2010.

[351] M. E. Auldridge and K. T. Forest, "Bacterial phytochromes: more than meets the light," *Critical reviews in biochemistry and molecular biology*, vol. 46, no. 1, pp. 67–88, 2011.

[352] B. Karniol and R. D. Vierstra, "Structure, function, and evolution of microbial phytochromes," in *Photomorphogenesis in plants and bacteria*. Springer, 2006, pp. 65–98.

[353] A. A. Kaberniuk, A. A. Shemetov, and V. V. Verkhusha, "A bacterial phytochrome-based optogenetic system controllable with near-infrared light," *Nature methods*, vol. 13, no. 7, pp. 591–597, 2016.

[354] I. Kraiselburd, L. Moyano, A. Carrau, J. Tano, and E. G. Orellano, "Bacterial photosensory proteins and their role in plant–pathogen interactions," *Photochemistry and Photobiology*, vol. 93, no. 3, pp. 666–674, 2017.

[355] L. B. Wiltbank and D. M. Kehoe, "Diverse light responses of cyanobacteria mediated by phytochrome superfamily photoreceptors," *Nature Reviews Microbiology*, vol. 17, no. 1, pp. 37–50, 2019.

[356] R. Ohlendorf and A. Möglich, "Light-regulated gene expression in bacteria: Fundamentals, advances, and perspectives," *Frontiers in Bioengineering and Biotechnology*, vol. 10, p. 1029403, 2022.

[357] H. A. Schneider-Poetsch, "Signal transduction by phytochrome: phytochromes have a module related to the transmitter modules of bacterial sensor proteins," *Photochemistry and photobiology*, vol. 56, no. 5, pp. 839–846, 1992.

[358] A. Nagatani, "Phytochrome: structural basis for its functions," *Current opinion in plant biology*, vol. 13, no. 5, pp. 565–570, 2010.

[359] S. G. Sokolovski, E. A. Zherebtsov, R. K. Kar, D. Golonka, R. Stabel, N. B. Chichkov, A. Gorodetsky, I. Schapiro, A. Möglich, and E. U. Rafailov, "Two-photon conversion of a bacterial phytochrome," *Biophysical journal*, vol. 120, no. 5, pp. 964–974, 2021.

[360] J. T. Henry and S. Crosson, "Ligand-binding pas domains in a genomic, cellular, and structural context," *Annual review of microbiology*, vol. 65, no. 1, pp. 261–286, 2011.

[361] L. Aravind and C. P. Ponting, "The gaf domain: an evolutionary link between diverse phototransducing proteins." *Trends in biochemical sciences*, vol. 22, no. 12, pp. 458–459, 1997.

[362] J. Hughes, "Phytochrome three-dimensional structures and functions," *Biochemical Society Transactions*, vol. 38, no. 2, pp. 710–716, 2010.

[363] J. E. Schultz, "Structural and biochemical aspects of tandem gaf domains," *Cgmp: generators, effectors and therapeutic implications*, pp. 93–109, 2009.

[364] X. Yang, J. Kuk, and K. Moffat, "Crystal structure of pseudomonas aeruginosa bacteriophytochrome: photoconversion and signal transduction," *Proceedings of the National Academy of Sciences*, vol. 105, no. 38, pp. 14 715–14 720, 2008.

[365] G. Battocchio, R. González, A. G. Rao, I. Schapiro, and M. A. Mroginski, "Dynamic properties of the photosensory domain of deinococcus radiodurans bacteriophytochrome," *The Journal of Physical Chemistry B*, vol. 124, no. 9, pp. 1740–1750, 2020.

[366] G. Gourinchas, S. Etzl, and A. Winkler, "Bacteriophytochromes–from informative model systems of phytochrome function to powerful tools in cell biology," *Current opinion in structural biology*, vol. 57, pp. 72–83, 2019.

[367] H. Takala, P. Edlund, J. A. Ihalainen, and S. Westenhoff, "Tips and turns of bacteriophytochrome photoactivation," *Photochemical & Photobiological Sciences*, vol. 19, no. 11, pp. 1488–1510, 2020.

[368] C. Böhm, N. Todorović, M. Balasso, G. Gourinchas, and A. Winkler, "The phy domain dimer interface of bacteriophytochromes mediates cross-talk between photosensory modules and output domains," *Journal of molecular biology*, vol. 433, no. 15, p. 167092, 2021.

[369] A. H. West and A. M. Stock, "Histidine kinases and response regulator proteins in two-component signaling systems," *Trends in biochemical sciences*, vol. 26, no. 6, pp. 369–376, 2001.

[370] P. Casino, V. Rubio, and A. Marina, "Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction," *Cell*, vol. 139, no. 2, pp. 325–336, 2009.

[371] S. Etzl, R. Lindner, M. D. Nelson, and A. Winkler, "Structure-guided design and functional characterization of an artificial red light–regulated guanylate/adenylate cyclase for optogenetic applications," *Journal of Biological Chemistry*, vol. 293, no. 23, pp. 9078–9089, 2018.

[372] P. H. Quail, "Phytochrome-interacting factors," in *Seminars in cell & developmental biology*, vol. 11, no. 6. Elsevier, 2000, pp. 457–466.

[373] V. N. Pham, P. K. Kathare, and E. Huq, "Phytochromes and phytochrome interacting factors," *Plant physiology*, vol. 176, no. 2, pp. 1025–1038, 2018.

[374] S. Sano, T. Sano, I. Morishima, Y. Shiro, and Y. Maeda, "On the mechanism of the chemical and enzymic oxygenations of alpha-oxyprotohemin ix to fe. biliverdin ix alpha." *Proceedings of the National Academy of Sciences*, vol. 83, no. 3, pp. 531–535, 1986.

[375] M. D. Maines, "Heme oxygenase: function, multiplicity, regulatory mechanisms, and clinical applications," *The FASEB journal*, vol. 2, no. 10, pp. 2557–2568, 1988.

[376] J. R. Wagner, J. S. Brunzelle, K. T. Forest, and R. D. Vierstra, "A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome," *Nature*, vol. 438, no. 7066, pp. 325–331, 2005.

[377] H. Lehtivuori, I. Rissanen, H. Takala, J. Bamford, N. V. Tkachenko, and J. A. Ihalainen, "Fluorescence properties of the chromophore-binding domain of bacteriophytochrome from deinococcus radiodurans," *The Journal of Physical Chemistry B*, vol. 117, no. 38, pp. 11 049–11 057, 2013.

[378] S. J. Davis, A. V. Vener, and R. D. Vierstra, "Bacteriophytochromes: phytochrome-like photoreceptors from nonphotosynthetic eubacteria," *Science*, vol. 286, no. 5449, pp. 2517–2520, 1999.

[379] J.-y. Hasegawa, M. Isshiki, K. Fujimoto, and H. Nakatsuji, "Structure of phytochromobilin in the pr and pfr forms: Sac-ci theoretical study," *Chemical physics letters*, vol. 410, no. 1-3, pp. 90–93, 2005.

[380] L. Isaksson, E. Gustavsson, C. Persson, U. Brath, L. Vrhovac, G. Karlsson, V. Orekhov, and S. Westenhoff, "Signaling mechanism of phytochromes in solution," *Structure*, vol. 29, no. 2, pp. 151–160, 2021.

[381] G. Merga, M. F. Lopez, P. Fischer, P. Piwowarski, Ż. Nogacz, A. Kraskov, D. Buhrke, F. V. Escobar, N. Michael, F. Siebert *et al.*, "Light-and temperature-dependent dynamics of chromophore and protein structural changes in bathy phytochrome agp2," *Physical Chemistry Chemical Physics*, vol. 23, no. 33, pp. 18 197–18 205, 2021.

[382] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "Plumed 2: New feathers for an old bird," *Computer physics communications*, vol. 185, no. 2, pp. 604–613, 2014.

[383] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of chemical physics*, vol. 79, no. 2, pp. 926–935, 1983.

[384] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee *et al.*, "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations," *Journal of computational chemistry*, vol. 24, no. 16, pp. 1999–2012, 2003.

[385] B. H. P-LINCS, "A parallel linear constraint solver for molecular simulation," *Journal of Chemical Theory and Computation*, vol. 4, no. 1, pp. 116–122, 2008.

[386] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh ewald method," *The Journal of chemical physics*, vol. 103, no. 19, pp. 8577–8593, 1995.

[387] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *The Journal of chemical physics*, vol. 126, no. 1, 2007.

[388] F. Bach and E. Moulines, "Non-strongly-convex smooth stochastic approximation with convergence rate o (1/n)," *Advances in neural information processing systems*, vol. 26, 2013.

[389] K. Inomata, M. A. Hammam, H. Kinoshita, Y. Murata, H. Khawn, S. Noack, N. Michael, and T. Lamparter, "Sterically locked synthetic bilin derivatives and phytochrome agp1 from agrobacterium tumefaciens form photoinsensitive pr-and pfr-like adducts," *Journal of Biological Chemistry*, vol. 280, no. 26, pp. 24 491–24 497, 2005.

[390] P. W. Kim, N. C. Rockwell, S. S. Martin, J. C. Lagarias, and D. S. Larsen, "Dynamic inhomogeneity in the photodynamics of cyanobacterial phytochrome cph1," *Biochemistry*, vol. 53, no. 17, pp. 2818–2826, 2014.

[391] C. Klose, F. Nagy, and E. Schäfer, "Thermal reversion of plant phytochromes," *Molecular Plant*, vol. 13, no. 3, pp. 386–397, 2020.

[392] J. Salewski, F. V. Escobar, S. Kaminski, D. von Stetten, A. Keidel, Y. Rippers, N. Michael, P. Scheerer, P. Piwowarski, F. Bartl *et al.*, "Structure of the biliverdin cofactor in the pfr state of bathy and prototypical phytochromes," *Journal of Biological Chemistry*, vol. 288, no. 23, pp. 16 800–16 814, 2013.

[393] J. J. Van Thor, K. L. Ronayne, and M. Towrie, "Formation of the early photoproduct lumi-r of cyanobacterial phytochrome cph1 observed by ultrafast mid-infrared spectroscopy," *Journal of the American Chemical Society*, vol. 129, no. 1, pp. 126–132, 2007.

[394] T. Rohmer, C. Lang, C. Bongards, K. B. S. S. Gupta, J. Neugebauer, J. Hughes, W. Gärtner, and J. Matysik, "Phytochrome as molecular machine: revealing chromophore action during the pfr→ pr photoconversion by magic-angle spinning nmr spectroscopy," *Journal of the American Chemical Society*, vol. 132, no. 12, pp. 4431–4437, 2010.

[395] M. Carrillo, S. Pandey, J. Sanchez, M. Noda, I. Poudyal, L. Aldama, T. N. Malla, E. Claesson, W. Y. Wahlgren, D. Feliz *et al.*, "High-resolution crystal structures of transient intermediates in the phytochrome photocycle," *Structure*, vol. 29, no. 7, pp. 743–754, 2021.

[396] D. Wang, Y. Qin, M. Zhang, X. Li, L. Wang, X. Yang, and D. Zhong, "The origin of ultrafast multiphasic dynamics in photoisomerization of bacteriophytochrome," *The journal of physical chemistry letters*, vol. 11, no. 15, pp. 5913–5919, 2020.

[397] D. M. Shcherbakova and V. V. Verkhusha, "Chromophore chemistry of fluorescent proteins controlled by light," *Current opinion in chemical biology*, vol. 20, pp. 60–68, 2014.

[398] V. A. Sineshchekov and O. D. Bekasova, "Two distinct photoprocesses in cyanobacterial bilin pigments: Energy migration in light-harvesting phycobiliproteins versus photoisomerization in phytochromes," *Photochemistry and Photobiology*, vol. 96, no. 4, pp. 750–767, 2020.

[399] J. K. Weber, D. Shukla, and V. S. Pande, "Heat dissipation guides activation in signaling proteins," *Proceedings of the National Academy of Sciences*, vol. 112, no. 33, pp. 10 377–10 382, 2015.

[400] J. Rydzewski and W. Nowak, "Memetic algorithms for ligand expulsion from protein cavities," *The Journal of Chemical Physics*, vol. 143, no. 12, 2015.

[401] ——, "Machine learning based dimensionality reduction facilitates ligand diffusion paths assessment: A case of cytochrome p450cam," *Journal of Chemical Theory and Computation*, vol. 12, no. 4, pp. 2110–2120, 2016.

[402] ——, "Rare-event sampling in ligand diffusion: Reply to comments on" ligand diffusion in proteins via enhanced sampling in molecular dynamics"." *Physics of Life Reviews*, vol. 22, pp. 85–87, 2017.

[403] ——, "Thermodynamics of camphor migration in cytochrome p450cam by atomistic simulations," *Scientific reports*, vol. 7, no. 1, p. 7736, 2017.

[404] J. Rydzewski, R. Jakubowski, W. Nowak, and H. Grubmuller, "Kinetics of huperzine a dissociation from acetylcholinesterase via multiple unbinding pathways," *Journal of Chemical Theory and Computation*, vol. 14, no. 6, pp. 2843–2851, 2018.

[405] J. Rydzewski and O. Valsson, "Finding multiple reaction pathways of ligand unbinding," *The Journal of Chemical Physics*, vol. 150, no. 22, 2019.

[406] J. Rydzewski, "maze: Heterogeneous ligand unbinding along transient protein tunnels," *Computer Physics Communications*, vol. 247, p. 106865, 2020.

[407] K. Walczewska-Szewc and W. Nowak, "Photo-switchable sulfonylureas binding to atp-sensitive potassium channel reveal the mechanism of light-controlled insulin release," *The Journal of Physical Chemistry B*, vol. 125, no. 48, pp. 13 111–13 121, 2021.

[408] S. Czach and K. Walczewska-Szewc, "Investigating substrate binding mechanism in prolyl oligopeptidase through molecular dynamics," *Physical Biology*, vol. 22, no. 5, p. 056001, 2025.

[409] V. Macaluso, L. Cupellini, G. Salvadori, F. Lipparini, and B. Mennucci, "Elucidating the role of structural fluctuations, and intermolecular and vibronic interactions in the spectroscopic response of a bacteriophytochrome," *Physical Chemistry Chemical Physics*, vol. 22, no. 16, pp. 8585–8594, 2020.

[410] V. Macaluso, G. Salvadori, L. Cupellini, and B. Mennucci, "The structural changes in the signaling mechanism of bacteriophytochromes in solution revealed by a multiscale computational investigation," *Chemical Science*, vol. 12, no. 15, pp. 5555–5565, 2021.

[411] J. A. Ihalainen, E. Gustavsson, L. Schroeder, S. Donnini, H. Lehtivuori, L. Isaksson, C. Thoing, V. Modi, O. Berntsson, B. Stucki-Buchli *et al.*, "Chromophore–protein interplay during the phytochrome photocycle revealed by step-scan ftir spectroscopy," *Journal of the American Chemical Society*, vol. 140, no. 39, pp. 12 396–12 404, 2018.

[412] J. Rydzewski and O. Valsson, "Multiscale reweighted stochastic embedding: Deep learning of collective variables for enhanced sampling," *The Journal of Physical Chemistry A*, vol. 125, no. 28, pp. 6286–6302, 2021.

[413] S. Mustalahti, D. Morozov, H. L. Luk, R. R. Pallerla, P. Myllyperkio, M. Pettersson, P. M. Pihko, and G. Groenhof, "Photoactive yellow protein chromophore photoisomerizes around a single bond if the double bond is locked," *The Journal of Physical Chemistry Letters*, vol. 11, no. 6, pp. 2177–2181, 2020.

[414] S. Gupta, "Using x-ray footprinting and mass spectrometry to study the structure and function of membrane proteins," *Protein and peptide letters*, vol. 26, no. 1, pp. 44–54, 2019.

[415] A. V. Nemukhin, B. L. Grigorenko, M. G. Khrenova, and A. I. Krylov, "Computational challenges in modeling of representative bioimaging proteins: Gfp-like proteins, flavoproteins, and phytochromes," *The Journal of Physical Chemistry B*, vol. 123, no. 29, pp. 6133–6149, 2019.

[416] F. Velazquez Escobar, D. Von Stetten, M. Günther-Lütkens, A. Keidel, N. Michael, T. Lamparter, L.-O. Essen, J. Hughes, W. Gärtner, Y. Yang *et al.*, "Conformational heterogeneity of the pfr chromophore in plant and cyanobacterial phytochromes," *Frontiers in molecular biosciences*, vol. 2, p. 37, 2015.

[417] T. E. Meyer, "Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium ectothiorhodospira halophila," *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 806, no. 1, pp. 175–183, 1985.

[418] J. A. Kyndt, F. Vanrobaeys, J. C. Fitch, B. V. Devreese, T. E. Meyer, M. A. Cusanovich, and J. J. Van Beeumen, "Heterologous production of halorhodospira halophila holo-photoactive yellow protein through tandem expression of the postulated biosynthetic genes," *Biochemistry*, vol. 42, no. 4, pp. 965–970, 2003.

[419] T. Meyer, S. Devanathan, T. Woo, E. Getzoff, G. Tollin, and M. Cusanovich, "Site-specific mutations provide new insights into the origin of ph effects and alternative spectral forms in the photoactive yellow protein from halorhodospira halophila," *Biochemistry*, vol. 42, no. 11, pp. 3319–3325, 2003.

[420] R. G. Saer and R. E. Blankenship, "Light harvesting in phototrophic bacteria: structure and function," *Biochemical Journal*, vol. 474, no. 13, pp. 2107–2131, 2017.

[421] W. D. Hoff, M. A. van der Horst, C. B. Nudel, and K. J. Hellingwerf, "Prokaryotic phototaxis," in *Chemotaxis: Methods and Protocols*. Springer, 2009, pp. 25–49.

[422] J. P. Armitage and K. J. Hellingwerf, "Light-induced behavioral responses (phototaxis') in prokaryotes," *Photosynthesis Research*, vol. 76, no. 1, pp. 145–155, 2003.

[423] W. Criellard, R. Kort, and K. J. Hellingwerf, "Photoactive yellow protein, a photoreceptor from purple bacteria," in *Comprehensive Series in Photosciences*. Elsevier, 2001, vol. 1, pp. 179–192.

[424] M. A. Cusanovich and T. E. Meyer, "Photoactive yellow protein: a prototypic pas domain sensory protein and development of a common signaling mechanism," *Biochemistry*, vol. 42, no. 17, pp. 4759–4770, 2003.

[425] P. Düx, G. Rubinstenn, G. W. Vuister, R. Boelens, F. A. Mulder, K. Hård, W. D. Hoff, A. R. Kroon, W. Crielaard, K. J. Hellingwerf *et al.*, "Solution structure and backbone dynamics of the photoactive yellow protein," *Biochemistry*, vol. 37, no. 37, pp. 12 689–12 699, 1998.

[426] J. Vreede, W. Crielaard, K. J. Hellingwerf, and P. G. Bolhuis, "Predicting the signaling state of photoactive yellow protein," *Biophysical journal*, vol. 88, no. 5, pp. 3525–3535, 2005.

[427] K. Okamoto, N. Hamada, T.-a. Okamura, N. Ueyama, and H. Yamamoto, "Color regulation and stabilization of chromophore by cys69 in photoactive yellow protein active center," *Organic & Biomolecular Chemistry*, vol. 7, no. 18, pp. 3782–3791, 2009.

[428] S. Anderson, S. Crosson, and K. Moffat, "Short hydrogen bonds in photoactive yellow protein," *Biological Crystallography*, vol. 60, no. 6, pp. 1008–1016, 2004.

[429] E. J. Leenders, L. Guidoni, U. Röthlisberger, J. Vreede, P. G. Bolhuis, and E. J. Meijer, "Protonation of the chromophore in the photoactive yellow protein," *The Journal of Physical Chemistry B*, vol. 111, no. 14, pp. 3765–3773, 2007.

[430] G. Groenhof, M. Bouxin-Cademartory, B. Hess, S. P. De Visser, H. J. Berendsen, M. Olivucci, A. E. Mark, and M. A. Robb, "Photoactivation of the photoactive yellow protein: why photon absorption triggers a trans-to-cis isomerization of the chromophore in the protein," *Journal of the American Chemical Society*, vol. 126, no. 13, pp. 4228–4233, 2004.

[431] K. Saito and H. Ishikita, "Formation of an unusually short hydrogen bond in photoactive yellow protein," *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1827, no. 3, pp. 387–394, 2013.

[432] S. Yamaguchi, H. Kamikubo, K. Kurihara, R. Kuroki, N. Niimura, N. Shimizu, Y. Yamazaki, and M. Kataoka, "Low-barrier hydrogen bond in photoactive yellow protein," *Proceedings of the National Academy of Sciences*, vol. 106, no. 2, pp. 440–444, 2009.

[433] W. Cleland and M. M. Kreevoy, "Low-barrier hydrogen bonds and enzymic catalysis," *Science*, vol. 264, no. 5167, pp. 1887–1890, 1994.

[434] P. A. Sigala, M. A. Tsuchida, and D. Herschlag, "Hydrogen bond dynamics in the active site of photoactive yellow protein," *Proceedings of the National Academy of Sciences*, vol. 106, no. 23, pp. 9232–9237, 2009.

[435] I. Kiliç and Y. Yeşiloğlu, "Spectroscopic studies on the antioxidant activity of p-coumaric acid," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 115, pp. 719–724, 2013.

[436] J. Kaur and R. Kaur, "p-coumaric acid: A naturally occurring chemical with potential therapeutic applications," *Current Organic Chemistry*, vol. 26, no. 14, pp. 1333–1349, 2022.

[437] Z. Lou, H. Wang, S. Rao, J. Sun, C. Ma, and J. Li, "p-coumaric acid kills bacteria through dual damage mechanisms," *Food control*, vol. 25, no. 2, pp. 550–554, 2012.

[438] D. R. Klein and L. S. Starkey, *Organic chemistry*. John Wiley & Sons, 2025.

[439] M. Putschögl, P. Zirak, and A. Penzkofer, "Absorption and emission behaviour of trans-p-coumaric acid in aqueous solutions and some organic solvents," *Chemical Physics*, vol. 343, no. 1, pp. 107–120, 2008.

[440] L. L. Premvardhan, F. Buda, M. A. van der Horst, D. C. Lührs, K. J. Hellingwerf, and R. van Grondelle, "Impact of photon absorption on the electronic properties of p-coumaric acid derivatives of the photoactive yellow protein chromophore," *The Journal of Physical Chemistry B*, vol. 108, no. 16, pp. 5138–5148, 2004.

[441] P. Changenet-Barret, A. Espagne, P. Plaza, K. Hellingwerf, and M. M. Martin, "Investigations of the primary events in a bacterial photoreceptor for photomotility: photoactive yellow protein (pyp)," *New Journal of Chemistry*, vol. 29, no. 4, pp. 527–534, 2005.

[442] V. Molina and M. Merchán, "On the absorbance changes in the photocycle of the photoactive yellow protein: a quantum-chemical analysis," *Proceedings of the National Academy of Sciences*, vol. 98, no. 8, pp. 4299–4304, 2001.

[443] T. Gensch, C. Gradinaru, I. Van Stokkum, J. Hendriks, K. Hellingwerf, and R. Van Grondelle, "The primary photoreaction of photoactive yellow protein (pyp): anisotropy changes and excitation wavelength dependence," *Chemical physics letters*, vol. 356, no. 3-4, pp. 347–354, 2002.

[444] S. Ohishi, N. Shimizu, K. Mihara, Y. Imamoto, and M. Kataoka, "Light induces destabilization of photoactive yellow protein," *Biochemistry*, vol. 40, no. 9, pp. 2854–2859, 2001.

[445] L. T. Mix, E. C. Carroll, D. Morozov, J. Pan, W. R. Gordon, A. Philip, J. Fuzell, M. Kumauchi, I. Van Stokkum, G. Groenhof *et al.*, "Excitation-wavelength-dependent photocycle initiation dynamics resolve heterogeneity in the photoactive yellow protein from halorhodospira halophila," *Biochemistry*, vol. 57, no. 11, pp. 1733–1747, 2018.

[446] A. D. Stahl, M. Hospes, K. Singhal, I. van Stokkum, R. van Grondelle, M. L. Groot, and K. J. Hellingwerf, "On the involvement of single-bond rotation in the primary photochemistry of photoactive yellow protein," *Biophysical Journal*, vol. 101, no. 5, pp. 1184–1192, 2011.

[447] P. Changenet-Barret, P. Plaza, and M. M. Martin, "Primary events in the photoactive yellow protein chromophore in solution," *Chemical physics letters*, vol. 336, no. 5-6, pp. 439–444, 2001.

[448] K. Takeshita, Y. Imamoto, M. Kataoka, F. Tokunaga, and M. Terazima, "Themodynamic and transport properties of intermediate states of the photocyclic reaction of photoactive yellow protein," *Biochemistry*, vol. 41, no. 9, pp. 3037–3048, 2002.

[449] S. Yeremenko, I. H. van Stokkum, K. Moffat, and K. J. Hellingwerf, "Influence of the crystalline state on photoinduced dynamics of photoactive yellow protein studied by ultraviolet-visible transient absorption spectroscopy," *Biophysical journal*, vol. 90, no. 11, pp. 4224–4235, 2006.

[450] M. Kamiya and I. Ohmine, "A molecular dynamics study for the structure determination of the signaling state in the photocycle of photoactive yellow protein," *The Journal of Physical Chemistry B*, vol. 114, no. 19, pp. 6594–6600, 2010.

[451] Y.-W. Hsiao and W. Thiel, "pb2 intermediate of the photoactive yellow protein: Structure and excitation energies," *The Journal of Physical Chemistry B*, vol. 115, no. 9, pp. 2097–2106, 2011.

[452] G. Rubinstenn, G. W. Vuister, F. A. Mulder, P. E. Düx, R. Boelens, K. J. Hellingwerf, and R. Kaptein, "Structural and dynamic changes of photoactive yellow protein during its photocycle in solution," *nature structural biology*, vol. 5, no. 7, pp. 568–570, 1998.

[453] J. Hendriks and K. J. Hellingwerf, "ph dependence of the photoactive yellow protein photocycle recovery reaction reveals a new late photocycle intermediate with a deprotonated chromophore," *Journal of Biological Chemistry*, vol. 284, no. 8, pp. 5277–5288, 2009.

[454] J. Rydzewski, "Spectral map for slow collective variables, markovian dynamics, and transition state ensembles," *Journal of Chemical Theory and Computation*, vol. 20, no. 18, pp. 7775–7784, 2024.

[455] Schrödinger, "Schrödinger release 2024-3: Maestro," *LLC*, 2024.

[456] M. Bugnon, M. Goullieux, U. F. Röhrig, M. A. Perez, A. Daina, O. Michielin, and V. Zoete, "Swissparam 2023: a modern web-based tool for efficient small molecule parametrization," *Journal of Chemical Information and Modeling*, vol. 63, no. 21, pp. 6469–6475, 2023.

[457] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch *et al.*, "Charmm: the biomolecular simulation program," *Journal of computational chemistry*, vol. 30, no. 10, pp. 1545–1614, 2009.

[458] J. Lee, X. Cheng, S. Jo, A. D. MacKerell, J. B. Klauda, and W. Im, "Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field," *Biophysical journal*, vol. 110, no. 3, p. 641a, 2016.

[459] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov *et al.*, "Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields," *Journal of computational chemistry*, vol. 31, no. 4, pp. 671–690, 2010.

[460] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, "Lincs: a linear constraint solver for molecular simulations," *Journal of computational chemistry*, vol. 18, no. 12, pp. 1463–1472, 1997.

[461] H. J. Berendsen, J. v. Postma, W. F. Van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of chemical physics*, vol. 81, no. 8, pp. 3684–3690, 1984.

[462] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied physics*, vol. 52, no. 12, pp. 7182–7190, 1981.

[463] R. Nieuwenhuys, H. J. Ten Donkelaar, and C. Nicholson, *The central nervous system of vertebrates*. Springer, 2014.

[464] G. J. Tortora and B. H. Derrickson, *Principles of anatomy and physiology*. John wiley & sons, 2018.

[465] M. Ayub and A. Mallamaci, "An introduction: Overview of nervous system and brain disorders," *The role of natural antioxidants in brain disorders*, pp. 1–24, 2023.

[466] P. J. Bazira, "An overview of the nervous system," *Surgery (Oxford)*, 2024.

[467] J. W. Bigbee, "Cells of the central nervous system: an overview of their structure and function," *Glycobiology of the Nervous System*, pp. 41–64, 2022.

[468] J. Ma and J. Tang, "A review for dynamics of collective behaviors of network of neurons," *Science China Technological Sciences*, vol. 58, pp. 2038–2045, 2015.

[469] R. S. Snell, *Clinical neuroanatomy*. Lippincott Williams & Wilkins, 2010.

[470] F. He and Y. E. Sun, "Glial cells more than support cells?" *The international journal of biochemistry & cell biology*, vol. 39, no. 4, pp. 661–665, 2007.

[471] M. D. Ehlers and F. Polleux, "Neuronal and glial cell biology," *Current Opinion in Neurobiology*, vol. 20, no. 5, pp. 529–530, 2010.

[472] V. Parpura, M. T. Heneka, V. Montana, S. H. Oliet, A. Schousboe, P. G. Haydon, R. F. Stout Jr, D. C. Spray, A. Reichenbach, T. Pannicke *et al.*, "Glial cells in (patho) physiology," *Journal of neurochemistry*, vol. 121, no. 1, pp. 4–27, 2012.

[473] D. Choquet and A. Triller, "The dynamic synapse," *Neuron*, vol. 80, no. 3, pp. 691–703, 2013.

[474] T. C. Südhof, "Towards an understanding of synapse formation," *Neuron*, vol. 100, no. 2, pp. 276–293, 2018.

[475] T. C. Sudhof, "The cell biology of synapse formation," *Journal of Cell Biology*, vol. 220, no. 7, 2021.

[476] S. Koizumi, K. Fujishita, and K. Inoue, "Regulation of cell-to-cell communication mediated by astrocytic atp in the cns," *Purinergic signalling*, vol. 1, pp. 211–217, 2005.

[477] G. Schiera, C. M. Di Liegro, and I. Di Liegro, "Cell-to-cell communication in learning and memory: from neuro-and glio-transmission to information exchange mediated by extracellular vesicles," *International Journal of Molecular Sciences*, vol. 21, no. 1, p. 266, 2019.

[478] M. Portela, T. Mitchell, and S. Casas-Tintó, "Cell-to-cell communication mediates glioblastoma progression in drosophila," *Biology open*, vol. 9, no. 9, p. bio053405, 2020.

[479] S. Ahmad, R. K. Srivastava, P. Singh, U. P. Naik, and A. K. Srivastava, "Role of extracellular vesicles in glia-neuron intercellular communication," *Frontiers in Molecular Neuroscience*, vol. 15, p. 844194, 2022.

[480] A. Adamsky, A. Kol, T. Kreisel, A. Doron, N. Ozeri-Engelhard, T. Melcer, R. Refaeli, H. Horn, L. Regev, M. Groysman *et al.*, "Astrocytic activation generates de novo neuronal potentiation and memory enhancement," *Cell*, vol. 174, no. 1, pp. 59–71, 2018.

[481] V. Naranjo, A. Contreras, B. Merino, A. Plaza, M. P. Lorenzo, C. García-Cáceres, A. García, J. A. Chowen, M. Ruiz-Gayo, N. Del Olmo *et al.*, "Specific deletion of the astrocyte leptin receptor induces changes in hippocampus glutamate metabolism, synaptic transmission and plasticity," *Neuroscience*, vol. 447, pp. 182–190, 2020.

[482] K. R. Jessen, "Glial cells," *The international journal of biochemistry & cell biology*, vol. 36, no. 10, pp. 1861–1867, 2004.

[483] L. P. Deleyrolle, A. Harding, K. Cato, F. A. Siebzehnrubl, M. Rahman, H. Azari, S. Olson, B. Gabrielli, G. Osborne, A. Vescovi *et al.*, "Evidence for label-retaining tumour-initiating cells in human glioblastoma," *Brain*, vol. 134, no. 5, pp. 1331–1343, 2011.

[484] M. Weller, W. Wick, K. Aldape, M. Brada, M. Berger, S. M. Pfister, R. Nishikawa, M. Rosenthal, P. Y. Wen, R. Stupp *et al.*, "Glioma," *Nature reviews Disease primers*, vol. 1, no. 1, pp. 1–18, 2015.

[485] A. Perry and P. Wesseling, "Histologic classification of gliomas," *Handbook of clinical neurology*, vol. 134, pp. 71–95, 2016.

[486] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. Ng, S. M. Pfister, G. Reifenberger *et al.*, "The 2021 who classification of tumors of the central nervous system: a summary," *Neuro-oncology*, vol. 23, no. 8, pp. 1231–1251, 2021.

[487] Q. T. Ostrom, L. Bauchet, F. G. Davis, I. Deltour, J. L. Fisher, C. E. Langer, M. Pekmezci, J. A. Schwartzbaum, M. C. Turner, K. M. Walsh *et al.*, "The epidemiology of glioma in adults: a "state of the science" review," *Neuro-oncology*, vol. 16, no. 7, pp. 896–913, 2014.

[488] K. Yang, Z. Wu, H. Zhang, N. Zhang, W. Wu, Z. Wang, Z. Dai, X. Zhang, L. Zhang, Y. Peng *et al.*, "Glioma targeted therapy: insight into future of molecular approaches," *Molecular cancer*, vol. 21, no. 1, p. 39, 2022.

[489] T. Li, J. Li, Z. Chen, S. Zhang, S. Li, S. Wageh, O. A. Al-Hartomy, A. G. Al-Sehemi, Z. Xie, R. K. Kankala *et al.*, "Glioma diagnosis and therapy: current challenges and nanomaterial-based solutions," *Journal of controlled release*, vol. 352, pp. 338–370, 2022.

[490] R. S. Angom, N. M. R. Nakka, and S. Bhattacharya, "Advances in glioblastoma therapy: an update on current approaches," *Brain sciences*, vol. 13, no. 11, p. 1536, 2023.

[491] R. Stupp, M. Gander, S. Leyvraz, and E. Newlands, "Current and future developments in the use of temozolomide for the treatment of brain tumours," *The lancet oncology*, vol. 2, no. 9, pp. 552–560, 2001.

[492] N. Singh, A. Miner, L. Hennis, and S. Mittal, "Mechanisms of temozolomide resistance in glioblastoma-a comprehensive review," *Cancer drug resistance*, vol. 4, no. 1, p. 17, 2021.

[493] M. Monif, T. J. O'Brien, K. J. Drummond, C. A. Reid, S. V. Liubinas, and D. A. Williams, "P2x7 receptors are a potential novel target for anti-glioma therapies," *Journal of Inflammation*, vol. 11, pp. 1–10, 2014.

[494] L. Rong, N. Li, and Z. Zhang, "Emerging therapies for glioblastoma: current state and future directions," *Journal of Experimental & Clinical Cancer Research*, vol. 41, no. 1, p. 142, 2022.

[495] F. Yasinjan, Y. Xing, H. Geng, R. Guo, L. Yang, Z. Liu, and H. Wang, "Immunotherapy: a promising approach for glioma treatment," *Frontiers in immunology*, vol. 14, p. 1255611, 2023.

[496] A. S. Doghish, A. Mahmoud, M. A. Abd-Elmawla, M. B. Zaki, N. M. Aborehab, A. Hatawsh, A. F. Radwan, G. A. Sayed, R. Moussa, M. A. Abdel-Reheim *et al.*, "Innovative perspectives on glioblastoma: the emerging role of long non-coding rnas," *Functional & Integrative Genomics*, vol. 25, no. 1, p. 43, 2025.

[497] B. A. Cesca, K. P. San Martin, M. D. Caverzan, P. M. Oliveda, and L. E. Ibarra, "State-of-the-art photodynamic therapy for malignant gliomas: innovations in photosensitizers and combined therapeutic approaches," *Exploration of Targeted Anti-tumor Therapy*, vol. 6, p. 1002303, 2025.

[498] R. A. North, "Molecular physiology of p2x receptors," *Physiological reviews*, vol. 82, no. 4, pp. 1013–1067, 2002.

[499] G. Burnstock, "Purine and purinergic receptors," *Brain and neuroscience advances*, vol. 2, p. 2398212818817494, 2018.

[500] R. Sluyter, "The p2x7 receptor," *Protein Reviews: Volume 19*, pp. 17–53, 2017.

[501] G. Collo, S. Neidhart, E. Kawashima, M. Kosco-Vilbois, R. North, and G. Buell, "Tissue distribution of the p2x7 receptor," *Neuropharmacology*, vol. 36, no. 9, pp. 1277–1283, 1997.

[502] M. Á. Martínez-Cuesta, M. A. Blanch-Ruiz, R. Ortega-Luna, A. Sánchez-López, and Á. Álvarez, "Structural and functional basis for understanding the biological significance of p2x7 receptor," *International journal of molecular sciences*, vol. 21, no. 22, p. 8454, 2020.

[503] E. M. Jimenez-Mateos, J. Smith, A. Nicke, and T. Engel, "Regulation of p2x7 receptor expression and function in the brain," *Brain research bulletin*, vol. 151, pp. 153–163, 2019.

[504] G. R. Dubyak, "P2x7 receptor regulation of non-classical secretion from immune effector cells," *Cellular microbiology*, vol. 14, no. 11, pp. 1697–1706, 2012.

[505] C. M. Miller, N. R. Boulter, S. J. Fuller, A. M. Zakrzewski, M. P. Lees, B. M. Saunders, J. S. Wiley, and N. C. Smith, "The role of the p2x7 receptor in infectious diseases," *PLoS pathogens*, vol. 7, no. 11, p. e1002212, 2011.

[506] E. Adinolfi, C. Pizzirani, M. Idzko, E. Panther, J. Norgauer, F. Di Virgilio, and D. Ferrari, "P2x 7 receptor: Death or life?" *Purinergic signalling*, vol. 1, pp. 219–227, 2005.

[507] M. W. Grol, I. Zelner, and S. J. Dixon, "P2x7-mediated calcium influx triggers a sustained, pi3k-dependent increase in metabolic acid production by osteoblast-like cells," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 302, no. 5, pp. E561–E575, 2012.

[508] K. L. Beaucage, A. Xiao, S. I. Pollmann, M. W. Grol, R. J. Beach, D. W. Holdsworth, S. M. Sims, M. R. Darling, and S. J. Dixon, "Loss of p2x7 nucleotide receptor function leads to abnormal fat distribution in mice," *Purinergic signalling*, vol. 10, pp. 291–304, 2014.

[509] L.-H. Jiang, J. M. Baldwin, S. Roger, and S. A. Baldwin, "Insights into the molecular mechanisms underlying mammalian p2x7 receptor functions and contributions in diseases, revealed by structural modeling and single nucleotide polymorphisms," *Frontiers in pharmacology*, vol. 4, p. 55, 2013.

[510] M. Tewari and P. Seth, "Emerging role of p2x7 receptors in cns health and disease," *Ageing research reviews*, vol. 24, pp. 328–342, 2015.

[511] H. M. Costa-Junior, F. Sarmento Vieira, and R. Coutinho-Silva, "C terminus of the p2x7 receptor: treasure hunting," *Purinergic signalling*, vol. 7, pp. 7–19, 2011.

[512] X. Liang, D. S. Samways, K. Wolf, E. A. Bowles, J. P. Richards, J. Bruno, S. Dutertre, R. J. DiPaolo, and T. M. Egan, "Quantifying ca2+ current and permeability in atp-gated p2x7 receptors," *Journal of Biological Chemistry*, vol. 290, no. 12, pp. 7930–7942, 2015.

[513] J. Amstrup and I. Novak, "P2x7 receptor activates extracellular signal-regulated kinases erk1 and erk2 independently of ca2+ influx," *Biochemical Journal*, vol. 374, no. 1, pp. 51–61, 2003.

[514] A. E. McCarthy, C. Yoshioka, and S. E. Mansoor, "Full-length p2x7 structures reveal how palmitoylation prevents channel desensitization," *Cell*, vol. 179, no. 3, pp. 659–670, 2019.

[515] L.-H. Jiang, E. A. Caseley, S. P. Muench, and S. Roger, "Structural basis for the functional properties of the p2x7 receptor for extracellular atp," *Purinergic Signalling*, vol. 17, no. 3, pp. 331–344, 2021.

[516] A. Karasawa and T. Kawate, "Structural basis for subtype-specific inhibition of the p2x7 receptor," *elife*, vol. 5, p. e22153, 2016.

[517] T. Kawate, "P2x receptor activation," *Protein Reviews: Volume 19*, pp. 55–69, 2017.

[518] L. K. Kan, D. Williams, K. Drummond, T. O'Brien, and M. Monif, "The role of microglia and p2x7 receptors in gliomas," *Journal of Neuroimmunology*, vol. 332, pp. 138–146, 2019.

[519] J. G. McLarnon, "Roles of purinergic p2x7 receptor in glioma and microglia in brain tumors," *Cancer Letters*, vol. 402, pp. 93–99, 2017.

[520] Z. Ji, Y. Xie, Y. Guan, Y. Zhang, K.-S. Cho, M. Ji, and Y. You, "Involvement of p2x7 receptor in proliferation and migration of human glioma cells," *BioMed research international*, vol. 2018, no. 1, p. 8591397, 2018.

[521] L. K. Kan, S. Seneviratne, K. J. Drummond, D. A. Williams, T. J. O'Brien, and M. Monif, "P2x7 receptor antagonism inhibits tumour growth in human high-grade gliomas," *Purinergic Signalling*, vol. 16, pp. 327–336, 2020.

[522] D. Matyśniak, V. Chumak, N. Nowak, A. Kukla, L. Lehka, M. Oslislok, and P. Pomorski, "P2x7 receptor: the regulator of glioma tumor development and survival," *Purinergic Signalling*, vol. 18, no. 1, pp. 135–154, 2022.

[523] C. Volonte, S. Apolloni, S. D. Skaper, and G. Burnstock, "P2x7 receptors: channels, pores and more," *CNS & Neurological Disorders-Drug Targets-CNS & Neurological Disorders)*, vol. 11, no. 6, pp. 705–721, 2012.

[524] A. Morelli, D. Ferrari, G. Bolognesi, R. Rizzuto, and F. D. Virgilio, "Proapoptotic plasma membrane pore: P2x7 receptor," *Drug development research*, vol. 52, no. 4, pp. 571–578, 2001.

[525] P. Bregestovski, G. Maleeva, and P. Gorostiza, "Light-induced regulation of ligand-gated channel activity," *British journal of pharmacology*, vol. 175, no. 11, pp. 1892–1902, 2018.

[526] R. H. Kramer, J. J. Chambers, and D. Trauner, "Photochemical tools for remote control of ion channels in excitable cells," *Nature chemical biology*, vol. 1, no. 7, pp. 360–365, 2005.

[527] P. Paoletti, G. C. Ellis-Davies, and A. Mourot, "Optical control of neuronal ion channels and receptors," *Nature Reviews Neuroscience*, vol. 20, no. 9, pp. 514–532, 2019.

[528] D. L. Fortin, M. R. Banghart, T. W. Dunn, K. Borges, D. A. Wagenaar, Q. Gaudry, M. H. Karakossian, T. S. Otis, W. B. Kristan, D. Trauner *et al.*, "Photochemical control of endogenous ion channels and cellular excitability," *Nature methods*, vol. 5, no. 4, pp. 331–338, 2008.

[529] A. C. Kneuttinger, "A guide to designing photocontrol in proteins: methods, strategies and applications," *Biological Chemistry*, vol. 403, no. 5-6, pp. 573–613, 2022.

[530] C. W. Riggsbee and A. Deiters, "Recent advances in the photochemical control of protein function," *Trends in biotechnology*, vol. 28, no. 9, pp. 468–475, 2010.

[531] B. Szymczak, J. Czarnecka, S. Czach, W. Nowak, and K. Roszek, "Purinergic approach to effective glioma treatment with temozolomide reveals enhanced anti-cancer effects mediated by p2x7 receptor," *Cellular Signalling*, vol. 106, p. 110641, 2023.

[532] T. U. Consortium, "Uniprot: the universal protein knowledgebase in 2025," *Nucleic Acids Research*, vol. 53, no. D1, pp. D609–D617, 2025.

[533] E. Beitz, "Texshade: shading and labeling of multiple sequence alignments using latex2e," *Bioinformatics*, vol. 16, no. 2, pp. 135–139, 2000.

[534] Schrödinger, "Schrödinger release 2025-2: Maestro," 2025.

[535] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments," *Journal of computer-aided molecular design*, vol. 27, pp. 221–234, 2013.

[536] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, "Procheck: a program to check the stereochemical quality of protein structures," *Applied Crystallography*, vol. 26, no. 2, pp. 283–291, 1993.

[537] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, "Colabfold: making protein folding accessible to all," *Nature methods*, vol. 19, no. 6, pp. 679–682, 2022.

[538] E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris, and T. E. Ferrin, "Ucsf chimerax: Tools for structure building and analysis," *Protein Science*, vol. 32, no. 11, p. e4792, 2023.

[539] D. R. Koes, M. P. Baumgartner, and C. J. Camacho, "Lessons learned in empirical scoring with smina from the csar 2011 benchmarking exercise," *Journal of chemical information and modeling*, vol. 53, no. 8, pp. 1893–1904, 2013.

[540] O. Trott and A. J. Olson, "Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of computational chemistry*, vol. 31, no. 2, pp. 455–461, 2010.

[541] R. Evans, M. O'Neill, A. Pritzel, N. Antropova, A. Senior, T. Green, A. Žídek, R. Bates, S. Blackwell, J. Yim *et al.*, "Protein complex prediction with alphafold-multimer," *biorxiv*, pp. 2021–10, 2021.

[542] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, and T. E. Ferrin, "Ucsf chimerax: Structure visualization for researchers, educators, and developers," *Protein science*, vol. 30, no. 1, pp. 70–82, 2021.

[543] Microsoft Corporation, "Microsoft excel." [Online]. Available: https://office.microsoft.com/excel

[544] A. C. Oken, A. L. Turcu, E. Tzortzini, K. Georgiou, J. Nagel, F. G. Westermann, M. Barniol-Xicota, J. Seidler, G.-R. Kim, S.-D. Lee *et al.*, "A polycyclic scaffold identified by structure-based drug design effectively inhibits the human p2x7 receptor," *Nature communications*, vol. 16, no. 1, p. 8283, 2025.

[545] A. C. Oken, N. E. Lisi, I. Krishnamurthy, A. E. McCarthy, M. H. Godsey, A. Glasfeld, and S. E. Mansoor, "High-affinity agonism at the p2x7 receptor is mediated by three residues outside the orthosteric pocket," *Nature communications*, vol. 15, no. 1, p. 6662, 2024.

[546] J. Liu, T.-T. Liu, L. Mou, Y. Zhang, X. Chen, Q. Wang, B.-L. Deng, and J. Liu, "P2x7 receptor: A potential target for treating comorbid anxiety and depression," *Purinergic Signalling*, vol. 21, no. 4, pp. 829–839, 2025.

[547] J. Xu, Y. Zhang, X. Wang, W. Xu, S. Chen, L. Pan, and J. Gao, "A p2x7 receptor antagonist alleviates ptsd-like behaviors in adolescent rats through gut microbiota modulation and hippocampal transcriptomic remodeling," *Journal of Affective Disorders*, p. 119859, 2025.

[548] E. Pislyagin, S. Kozlovskiy, I. Agafonova, E. Menchinskaya, E. Chingizova, T. Gorpenchenko, A. Mirochnik, E. Fedorenko, Y. Sabutski, S. Polonik *et al.*, "Sea urchin pigment ethylspinazarin (u-573): A novel p2x7 receptor antagonist with neuroprotective and antiparkinsonian effects," *International Journal of Molecular Sciences*, vol. 26, no. 17, p. 8639, 2025.

[549] C. A. Maschio, O. Moreno, J. Wang, U. Maheshwari, A. Keller, U. Konietzko, D. Razansky, A. Nordberg, A. Rominger, C. Hock *et al.*, "Increased levels of gfap and purinergic p2x7 receptor in alzheimer's disease brain are associated with aβ, tau pathologies and synaptic loss," *Alzheimer's Research & Therapy*, 2025.

[550] S.-Y. Mei, N. Zhang, M.-j. Wang, P.-r. Lv, and Q. Liu, "Microglial purinergic signaling in alzheimer's disease," *Purinergic Signalling*, vol. 21, no. 4, pp. 815–827, 2025.

[551] R. Maurya, A. Sharma, and S. Naqvi, "Decoding nlrp3 inflammasome activation in alzheimer's disease: A focus on receptor dynamics," *Molecular Neurobiology*, pp. 1–21, 2025.

[552] P. Jiang, C. Wang, F. Jia, H. Wu, H. Hao, and S. Jing, "P2x7 receptor as a key player in pathological pain: Insights into neuropathic, inflammatory, and cancer pain," *Frontiers in Pharmacology*, vol. 16, p. 1585545, 2025.

[553] S. N. Mousavikia, M. M. Matin, M. T. B. Tossi, and H. Azimian, "Unraveling the role of the p2x7 receptor in cancer radioresistance: Molecular insights and therapeutic implications," *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, p. 119910, 2025.

[554] Y. Du, Y. Cao, W. Song, X. Wang, Q. Yu, X. Peng, and R. Zhao, "Role of the p2x7 receptor in breast cancer progression," *Purinergic Signalling*, vol. 21, no. 4, pp. 791–799, 2025.

[555] Q. Yu, X. Peng, G. Xu, X. Bai, Y. Cao, Y. Du, X. Wang, and R. Zhao, "Overexpression or knockdown of the p2x7 receptor regulates the progression of non-small cell lung cancer, involving gsk-3β and jnk signaling pathways," *European Journal of Pharmacology*, vol. 995, p. 177421, 2025.

[556] Y. Xue, X. Peng, M. Yan, Y. Du, Y. Cao, and R. Zhao, "The p2x7 receptor in leukemia: pathological mechanisms and therapeutic potential," *Purinergic Signalling*, pp. 1–12, 2025.