PROTEIN-DNA INTERACTIONS OF OXIDATIVE-STRESS TRANSCRIPTION FACTORS OXYR1 AND OXYR2 IN *ALIIVIBRIO FISCHERI*

By

Alexander Castro Martínez

A thesis submitted to the

DEPARTMENT OF BIOLOGY

FACULTY OF NATURAL SCIENCES

UNIVERSITY OF PUERTO RICO

RIO PIEDRAS CAMPUS

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN BIOLOGY

May 11, 2022

Río Piedras, Puerto Rico

Alexander Castro Martínez

This thesis has been accepted by the:

DEPARTMENT OF BIOLOGY
FACULTY OF NATURAL SCIENCES
UNIVERSITY OF PUERTO RICO
RIO PIEDRAS CAMPUS

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN BIOLOGY

Thesis committee:

_____Advisor
José A. Rodríguez-Martínez, Ph.D.

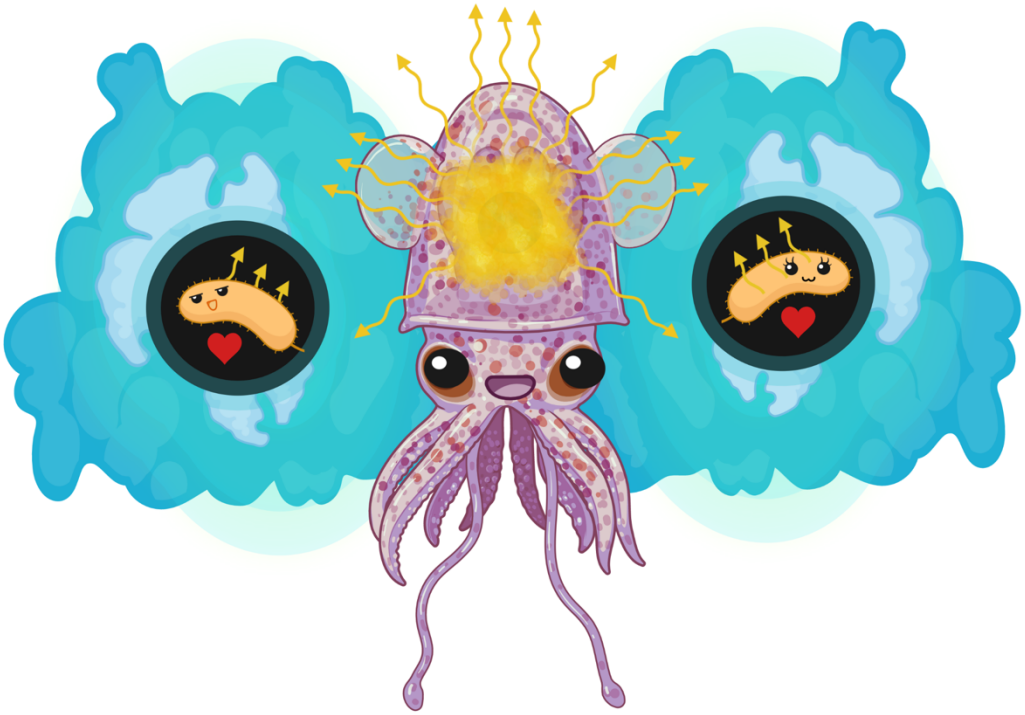
_____
Esther Peterson-Peguero, Ph.D.

_____
Abel Baerga-Ortiz, Ph.D.

_____
Gary A. Toranzos, Ph.D.

I would like to dedicate this work to the beautiful
people in my life, including those no longer present.
May each one of us
arrive at our heart's chosen destination.

Be present each day, and
let the warmth of the experience fuel your days.
Walk with gratitude, and step into your power.



ACKNOWLEDGEMENTS

Foremost, please accept my deepest gratitude to Dr. José A. Rodriguez-Martínez. Up to this day, I am blown away by your kindness, scientific prowess, and the sum of all your parts as a human being. Thank you for your patience, enthusiasm, guidance, and support throughout this journey. Your contribution to my formation as a better scientist is incommensurable. I am grateful that I was allowed to become part of the JARM-Lab. Gifting the amazing people I can dare to call family at heart. Thank you, Dr. Esther Peterson-Peguero, Dr. Abel Baerga, and Dr. Gary Toranzos, for providing insight as my thesis committee. To Dr. Peterson, thank you for our conversations, your advice, and for always helping me see through adversity before undertaking an important decision. I would like to include my expression of gratitude to Dr. Zomary Flores for providing the OxyR sequences, and her assistantship during the course of this research project.

Shout out to my friend, roommate, and lab partner Jessica M. Rodriguez-Rios. Forever grateful to have met such a fantastic human being. No words can begin to describe the positive impact you and your brother have made in my life. Changed for good. To Anthony Barreto and Emmanuel Carrasquillo, thank you for accepting me when I was freshly new at the lab and for teaching me skills and techniques I will continue to use during my career as a scientist. Your contributions to my experience during these past few years at the lab, were top tier. To Rosalba Velazquez Roig for your words of encouragement and self-esteem boosters always needed during tremulous times; your candor and words of truth will always be appreciated. To Paola, Sebastian, Laura, and Victoria for providing technical support and standing with me during the duration of this research and for allowing me to be part of your formation as arising scientists, I owe every step of the road to you, nothing would be possible without your contribution. I also want to thank my mother, family, and friends for their patience and love. Last but not least, blessings to the people with whom I started this process and have taken roots towards new spaces.

ABBREVIATIONS

TFs	Transcription factors
SELEX	Systematic Evolution Ligands by Exponential Enrichment
ROS	Reactive Oxygen Species
H_2O_2	Hydrogen peroxide
mRNA	Messenger RNA
(σ-factor)	Sigma factor
CAP	Catabolite activator protein
RD	Regulatory domain
DBD	DNA binding domain
PWMs	Position Weight Matrixes
LTTRs	LysR-type transcriptional regulators
GA	Gibson Assembly

ABSTRACT

Transcription factors (TFs) are regulatory proteins that bind tightly to specific DNA sequences 15-20 base pairs long. Being sequence-specific DNA-binding proteins, TFs bear the key to the cellular state and control how organisms respond to different environmental stresses. OxyR, a LysR-type transcription factor, binds strongly to H_2O_2 and activates a set of genes whose main purpose is to protect bacteria against oxidative stress. The genome of *Aliivibrio fischeri* bacteria codes for two different OxyR proteins, OxyR1 and OxyR2. However, the collection of target genes of OxyR1 and OxyR2 in *A. fischeri* remains to be determined. This study aims to clone, overexpress, and purify OxyR1 and OxyR2 TFs to determine their DNA binding functionality throughout the assessment of their intrinsic DNA-binding preferences. Purified OxyR1 and OxyR2 TFs were used to determine their DNA binding specificity using Systematic Evolution of Ligands by Exponential Enrichment (SELEX-seq).

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: LITERATURE REVISION

CHAPTER 2: PROTEIN CLONING AND PURIFICATION OF OXYR1 AND OXYR2 FROM *ALIIVIBRIO FISCHERI*

CHAPTER 3: EMSA FUNCTIONALITY ASSESSMENT OF *ALIIVIBRIO FISCHERI* OXYR1 AND OXYR2 TFS AND DETERMINING THEIR INTRINSIC BINDING SITES THROUGH SELEX-SEQ

LIST OF TABLES

CHAPTER 1: LITERATURE REVISION

CHAPTER 2: PROTEIN CLONING AND PURIFICATION OF OXYR1 AND OXYR2 FROM *ALIIVIBRIO FISCHERI*

CHAPTER 3: EMSA FUNCTIONALITY ASSESSMENT OF *ALIIVIBRIO FISCHERI* OXYR1 AND OXYR2 TFS AND DETERMINING THEIR INTRINSIC BINDING SITES THROUGH SELEX-SEQ

CHAPTER 3: EMSA FUNCTIONALITY ASSESSMENT OF *ALIIVIBRIO FISCHERI* OXYR1 AND OXYR2 TFS AND DETERMINING THEIR INTRINSIC BINDING SITES THROUGH SELEX-SEQ

CHAPTER 4: *ALLIVIBRIO FISCHERI* OXYR1 AND OXYR2 PROTEIN STRUCTURE PREDICTIONS AND MODEL BUILDING

1 CHAPTER 1

LITERATURE REVIEW

1.1 INTRODUCTION

Transcription factors (TFs) can recognize and bind to DNA in a specific manner, occasionally near their target genes in the promoter or enhancer regions (Browning et al., 2019). TFs have a unique affinity toward preferred nucleotide sequences. Each TF regulates a set of genes in response to specific environmental or intracellular triggers. TFs integrate multiple environmental inputs and translate them into coordinated responses (Young, 2013). For example, the OxyR transcription factor interacts with hydrogen peroxide and controls transcriptional expression activation in multiple bacteria. In other words, TFs serve as central nodes in intracellular signaling, playing an essential role in the organism's overall expression.

Reactive oxygen species (ROS) is an umbrella term for various molecular oxygen derivatives that occur as an ordinary attribution of aerobic life: oxygen-containing reactive species. The collective ROS umbrella term includes hydrogen peroxide (H_2O_2), hydroxyl radical ($^{\bullet}OH$), hypochlorous acid (HOCl), lipid hydroperoxide (LOOH), ozone (O_3), , superoxide (O_2^-), among others (Halliwell, 1996). ROS can lead to oxidative damage of macromolecules, such as DNA, protein, and lipids (Dwyer et al., 2015). For example, since hydrogen peroxide (H_2O_2) was identified in a living cell, it has been considered a toxic byproduct of aerobic metabolism, something that cells had to remove (Imlay, 2008).

The adaptive response of bacterial antioxidant systems toward oxidative stress is coordinated mainly by one well-studied system: the *oxyR* regulon (Seth et al., 2020). OxyR is a TF that acts as a sensor for hydrogen peroxide (H_2O_2) levels in bacteria. The sensing step involves the oxidation of two highly conserved cysteines that lead to the formation of a disulfide bond. Furthermore, this sensing step allows for the activation of OxyR and RNA polymerase recruitment to the *oxyR* operon promoter, starting transcription (Dubbs & Mongkolsuk, 2012).

In *Aliivibrio fischeri* there are two different known OxyR TFs: OxyR1 and OxyR2, unlike *Escherichia coli* which only possess a single copy *oxyR*. As seen before, modulation of oxidative stress becomes a critical step in confronting possible threats presented to the biological activity of an organism. Mainly, the OxyR protein is a modulator of oxidative stress works as a biosensor capable of recognizing hydrogen peroxide. OxyR is well known to activate the expression of a regulon of hydrogen peroxide-inducible genes such as *katG*, *gor*, *ahpC*, *ahpF*, *oxyS* (a regulatory RNA), *dps*, *fur*, and *grxA* in order to defend bacteria against ROS (James A. Imlay, 2015a). For example, during the initial association stages of animal-bacterial colonization of the bobtail squid *Euprymna scolopes* by the bioluminescent bacterium *A. fischeri* must sense and resist ROS. Therefore, the bacterial defense against oxidative stress ensures its colonization of the crypts within the mantle cavity of the bobtail squid, where the light organ is localized (Norsworthy & Visick, 2013). Therefore, examining the DNA binding preferences of OxyR1 and OxyR2 will provide insight into the value they possess as members of the oxidative stress biological defense pathway. In addition, it could bring insight into the overall genes involved in the oxidative pathways involved during the early stages of symbiotic colonization.

1.2 PROKARYOTIC TRANSCRIPTION

Transcription is the synthesis of a complementary strand of RNA directly from a DNA template. In general, transcription allows for the transferring of the genetic information stored in the sequence of nitrogenous bases of DNA into a messenger RNA (mRNA). The newly formed mRNA strand, in essence, is the carrier of coded information that allows specific proteins and functional RNA to be synthesized by an organism. The transcription process is modulated by an assortment of molecules, like RNA polymerase (RNAP), a supply of RNA nucleotides, cis-acting DNA sequence, structural DNA elements, and trans-acting proteins (Browning & Busby, 2016).

Trans-acting proteins like TFs recognize specific DNA sequences at target promoters up-or down-regulating gene expression. In bacteria, transcription is a tightly regulated process that facilitates bacterial adaptation to different environments in which transcription factors play a crucial role.

In bacteria, transcription initiation occurs at the transcription site with the recruitment of an RNAP by its association with a sigma factor (σ-factor) to the promoter in a specific locus upstream of the gene to be transcribed (Seshasayee et al., 2011). However, bacterial TFs that function as repressors bind to DNA targets that overlap elements at their promoters, occluding access to RNAP. On the other hand, TFs that function as activators bind to DNA targets located upstream of the essential elements at their target promoters, recruiting the RNAP. Gene expression in bacteria relies upon promoter recognition by the DNA-dependent RNAP and subsequent transcription initiation (Browning & Busby, 2016). Therefore, transcriptional initiation represents one of the most critical control points for gene expression.

Transcription, including translation in bacteria, occurs in the cytosol due to the lack of compartments, allowing quick usage of resources. The mechanism by which transcription occurs is a step-by-step process governed by RNAP, as shown in **Figure 1**. A recount of the important events would be: (1) RNAP binds to DNA at a site called the promoter 'open complex', allowing only one of the two strands to serve as the template for RNA synthesis in the 5' to 3' direction; (2) Once the open complex is formed, the RNAP assembles free nucleotides into the newly synthesized chain, using complementary base pairing as a guide; (3) RNAP continues to synthesize the mRNA until reaching a site called terminator; (4) the RNAP reaches the transcription termination site and the newly synthesized single-strand RNA is released from the DNA. The primary promoter elements that facilitate specific transcript initiation by RNAP are the UP (upstream) element, the −35 element, the extended −10 element, and the −10 element, with other

elements located in the spacer region between the −10 and −35 element (Lee et al., 2012). In addition, the C-terminal domains involving the polymerase α subunits (αCTD) make specific contacts with UP elements located upstream of the −35 element (Gourse et al., 2000).

σ-factors allow the transcription of all genes, therefore contributing to the sustainability of bacterial biological functions. For example, σ-factor S (RpoS) plays a vital role as a subunit of the RNAP holoenzyme by activating genes for protection when it encounters environmental stress conditions such as DNA damage, carbon starvation, temperature, pH, oxidative stress, among others (Battesti et al., 2011). For example, RpoS is a crucial response regulator to stress conditions during an oxidative stress event provided by phagocytes during an *E. coli*'s infection (Hryckowian & Welch, 2013). In addition, the association of appropriate alternative sigma factors with core RNA polymerase provides a mechanism for cellular responses mediated through redirection of transcription initiation (Kazmierczak et al., 2005). Some σ-factors can enhance the survival of bacteria by aiding against oxidative stress generated within macrophages that could lead to harmful oxidative bursts. (RpoS, a master regulator of general stress response). For example, *E. coli* and *Salmonella typhimurium* derive multiple general stress-responsive alternatives such as *rpoE*, *rpoH*, and *rpoS* to protect themselves from different environmental stress conditions, including hyperosmolarity, oxidative damage, and reduced pH (Kazmierczak et al., 2005).

The synthesis of proteins requires the expenditure of tremendous amounts of energy. The cell must optimize the spent energy by ensuring proteins are made when needed. Consequently, bacteria use regulation of protein synthesis and gene expression to optimize the cell's metabolic energy, improving the cell's machinery economy. TFs regulate transcription. These *trans*-acting factors regulate gene expression by binding directly to a *cis*-regulatory region in a specific manner. Direct recognition occurs when protein amino acid chains interact with specific bases in the DNA

sequences. For example, the catabolite activator protein (CAP) response element contains a specific affinity for the nucleotide consensus sequence (TGACGTCA). CAP activation depends widely on its interactions with the cAMP response element (Siegel, Aggranoff, 1999).

On the other hand, indirect recognition by a transcription factor is guided by the structural features in the DNA major and minor groove, backbone features, intrinsic curvature, and flexibility (Steffen et al., 2002). Possibly, TFs twist the DNA helix, altering the binding sites presentation to regulatory proteins that rely on indirect readout (DNA shape) to bind DNA (Dorman, 2019). Transcription factors that are members of the LysR-type transcription regulators (LTTRs) use helix-turn-helix binding motifs that engage both the major and minor grooves in DNA. The binding of the *trans*-acting factor to the *cis*-regulatory region alters transcription initiation, probably through a direct interaction of the *trans*-acting factor with the RNAP complex (Siegel, Aggranoff, 1999).

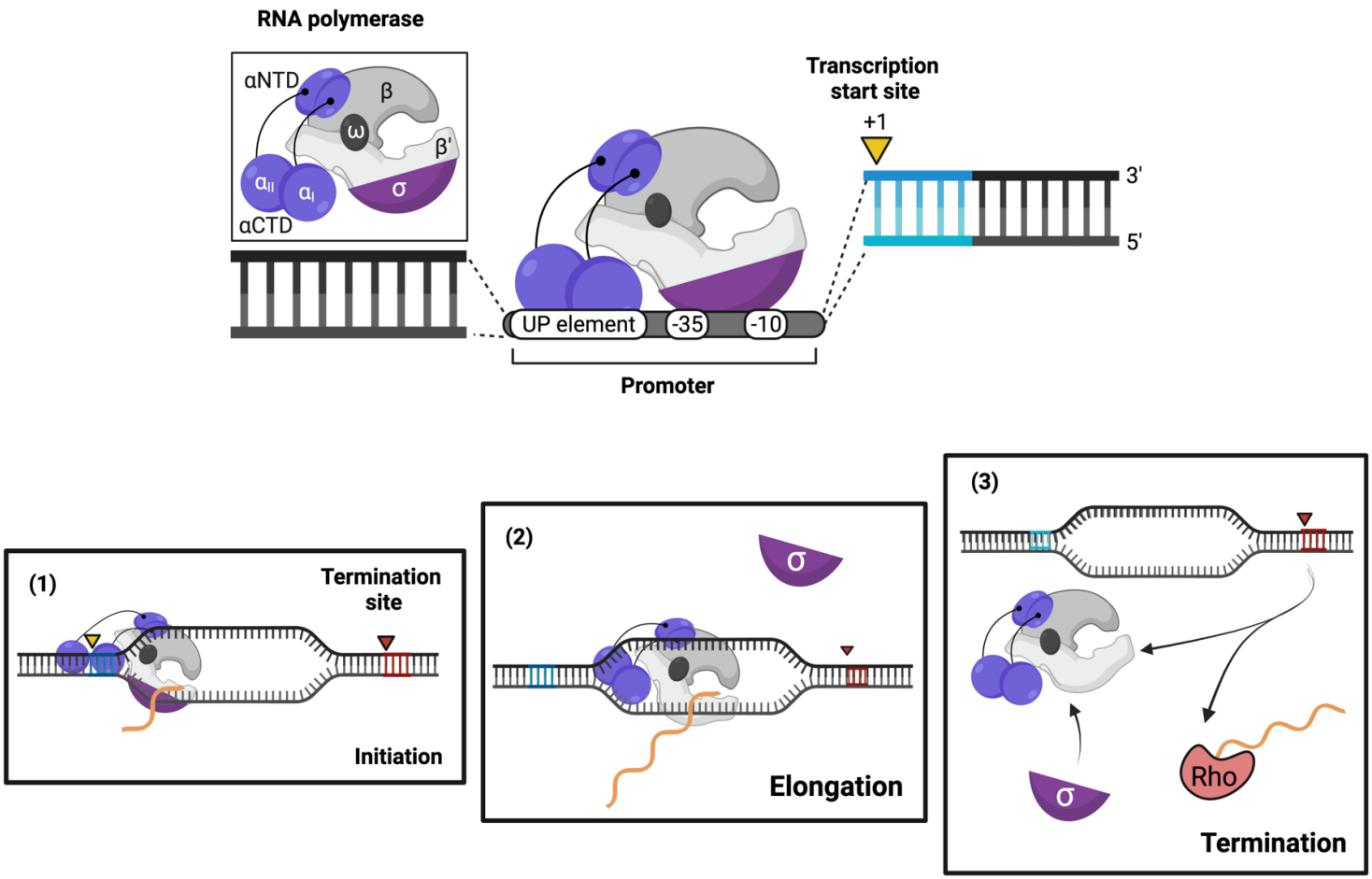


Figure 1. Transcription: Initiation, Elongation, and Termination steps. (1) Initiation: RNAP binds to DNA at a site called the promoter, allowing only one of the two strands to serve as the template for RNA synthesis in the 5' to 3' direction. RNA polymerase forms a transcription bubble by melting DNA near the transcription start site, and a short strand of RNA is made. (2) Elongation: RNAP advances 3' to 5' direction down the template strand, melting duplex DNA and adding rNTPs to the mRNA. Sigma factor dissociates from the RNAP complex. (3) Termination: At the transcription termination site, polymerase releases RNA and dissociates from DNA. In Rho-dependent termination, the RNA contains a binding site for Rho. Rho binds to this sequence; it catches up with RNAP at the transcription bubble and pulls the RNA transcript and the DNA template apart, releasing the mRNA and ending transcription. Created with Biorender.com

1.3 TRANSCRIPTION REGULATION

Gene regulation is the global component of mechanisms that act to induce or repress the expression of a gene. Likewise, gene regulation confers structural and chemical changes to the DNA by binding proteins to specific DNA elements (Range, 2012). Furthermore, gene regulation involves mechanisms that modulate the posterior translation of mRNA. Finally, gene regulation determines which genes will be active and not active and define their level of expression. Combining multiple gene regulation systems will allow for different pathways of differential gene expression. For example, two known gene control mechanisms are repression and activation. Activation allows for the appropriate mRNA transcription and consequently the production of the operon's enzymes. In contrast, repression is the regulatory mechanism that inhibits gene expression and, consequently, decrease the synthesis of operon enzymes. Repression is mediated by proteins called repressors, and it is often a response to the overabundance of an end-product. Some repressors are known to inhibit the ability of RNA polymerase to initiate transcription. In contrast to repressors, inducers are proteins that act to activate gene transcription.

A classic example of gene regulation is the lactose (*lac*) operon. This operon is a genetic unit that specifies and produces necessary enzymes important for the metabolism of lactose by *E. coli*. The *lac* operon contains three cistrons or DNA fragments that encode a functional protein (*lacZ*, *lacY*, and *lacA*). *E. coli*'s lactose operon encodes three enzymes: b-galactosidase (*lacZ)*, which digests lactose into its two constituent sugars: glucose and galactose, a permease (*lacY*) that helps transfer lactose into the cell, and finally, a trans-acetylase (*lacA*) the relevance of which is not entirely clear in lactose metabolism. Transcription of the lac operon normally occurs only when lactose is available for it to digest. When lactose is unavailable, the lac repressor inhibits transcription by binding to the operator, which partially overlaps the promoter. When the lac

repressor is bound to the promoter, RNA polymerase entry to the DNA becomes hindered, and there is no transcription of the operon. However, if lactose is present in high concentrations, the lac repressor cannot bind to DNA.

In contrast, the catabolite activator protein (CAP) is only active when glucose levels become low and cAMP levels are high. Therefore, the lac operon can only be transcribed at high levels when glucose is absent, and cAMP is present. This strategy ensures that bacteria only use lactose after depleting glucose, their preferred energy source. Altogether, these two events of gene regulation (repression and induction) allow RNAP to bind strongly to the promoter to give it a clear path for transcription. Diagrams for the lac operon are shown in **Figures 2- 4**. CAP is a transcription factor, and TFs have a global impact on the overall expression of its regulated genes during metabolic activity.

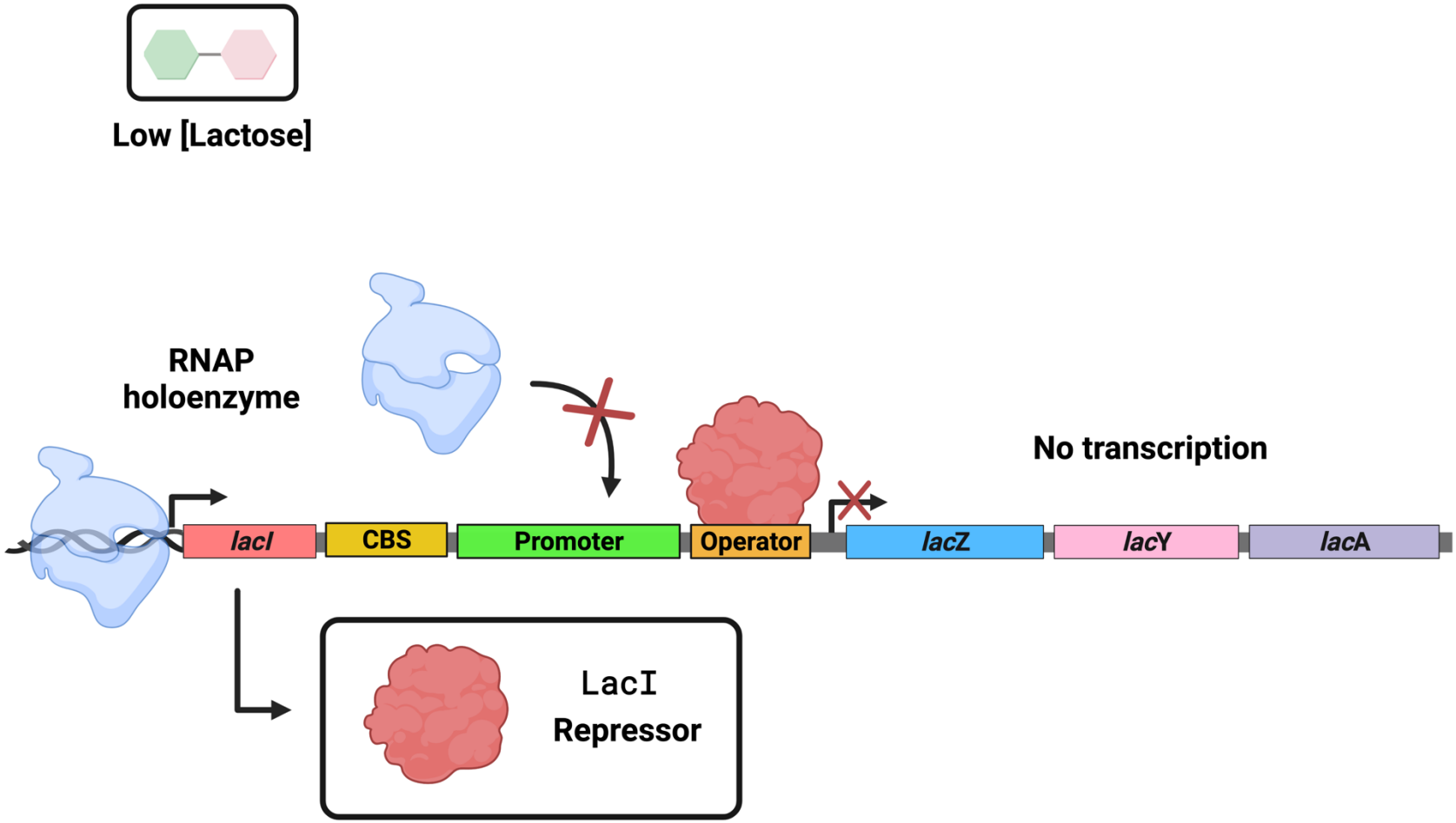


Figure 2. The *lac* operon: active LacI repression. In the presence of low lactose concentration, the LacI repressor is able to block the expression of the lac operon by forming a loop between the operators. The binding of LacI occurs as a homotetramer to the at the operator sites; located downstream from the promoter and upstream of the transcriptional initiation site in the DNA. Once the LacI homotetramers is at the operator sites, it hinders the accessibility of the RNAP into the promoter. Therefore, stopping the transcription of the *lacZ*, *lacY*, and *lacA* genes. Created with Biorender.com

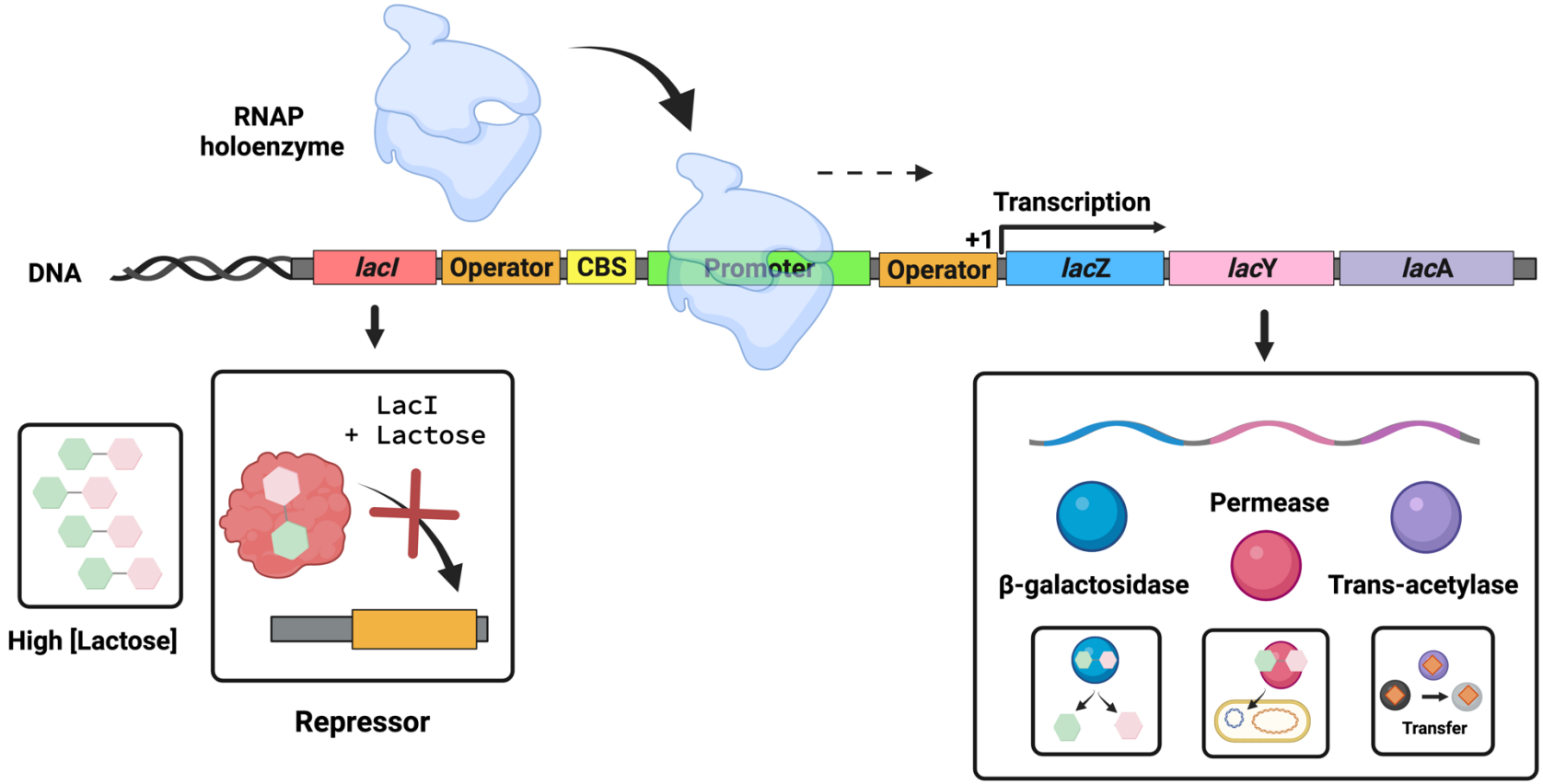


Figure 3. The *lac* operon: inactive LacI activation. The lac repressor has a high affinity for lactose. When a high amount of lactose is present the lac repressor will bind it causing dissociation from the DNA operator thus freeing the operon for gene expression. When the LacI is no present in the operator sites, RNAP recruitment into the promoter. Therefore, activating the transcription of the *lacZ*, *lacY*, and *lacA* genes. Created with Biorender.com

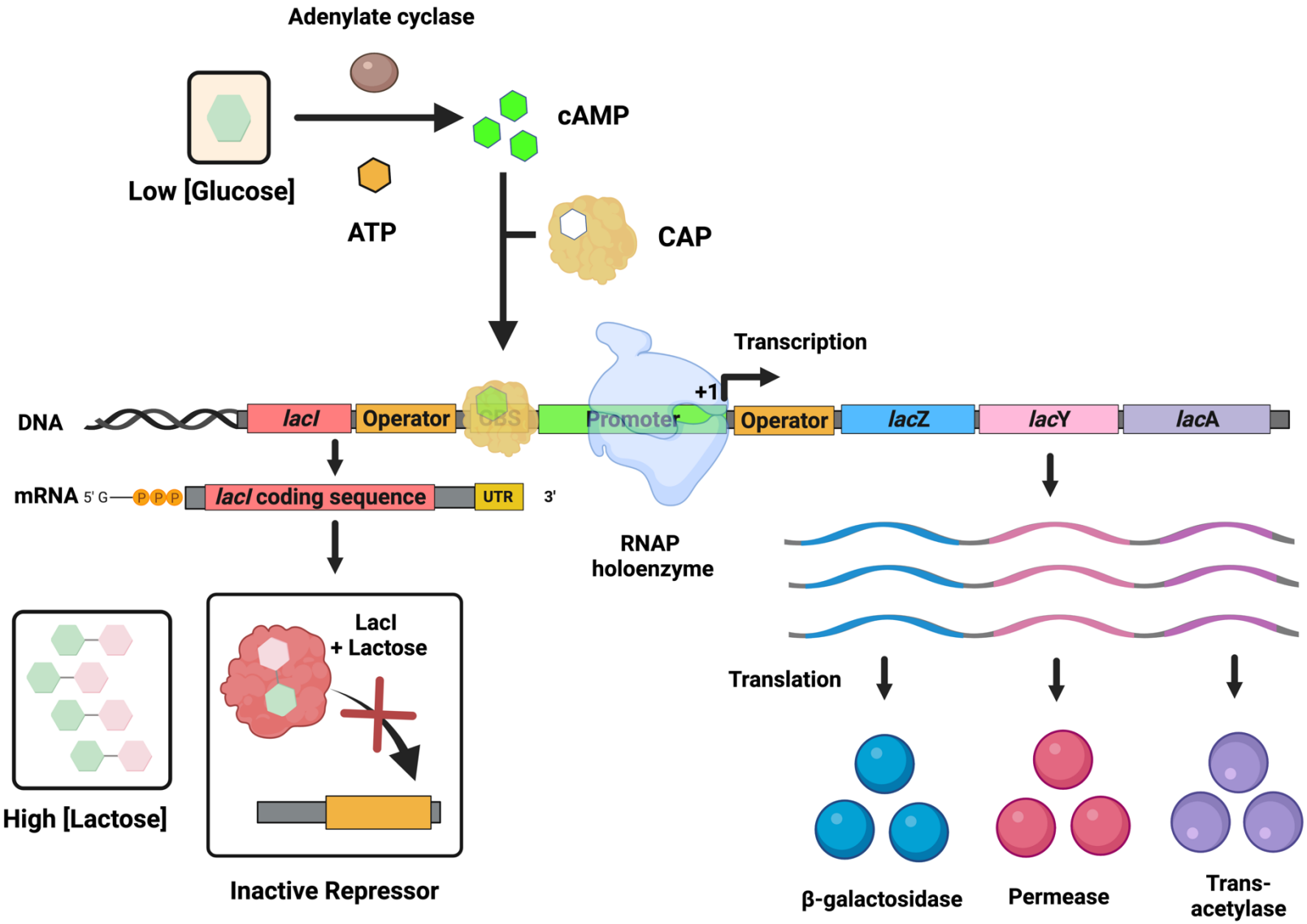


Figure 4. The lac operon: CAP mediated activation. In the presence of low concentrations of glucose, CAP is able to binds CAP binding site (CBS). CAP recruits the RNA polymerase, thus initiating and enhancing the transcription of the lac operon. In order for CAP to bind DNA it must first bind cyclic AMP (cAMP), and the promoter must not be hindered by LacI. LacI is inactive during high concentrations of lactose. Created with Biorender.com

1.4 TRANSCRIPTION FACTORS (TFs)

TFs are proteins defined by their ability to bind regulatory DNA sequences and regulate transcription. The binding of TFs to DNA is a bimolecular process governed by two rates: an on-rate (k_{on}) for the formation of the DNA-protein complex and an off-rate (k_{off}) for its dissociation (Stormo & Zhao, 2010). TFs interact with a target DNA sequence of 4 - 20 bp. TFs can bind DNA as monomers, homodimers, and tetramers. TFs can activate or repress transcription initiation depending on where they bind relative to the gene transcriptional target site. TFs are classified into families. Often this classification is based on their regulatory function involving their regulatory domain (RD), sequences homology (structural similarity) presented by their DNA binding domain (DBD), and their mechanism of action (Brivanlou & Darnell, 2002). The RD functions as a signal sensor by ligand-binding or protein-protein interaction. The DBD is an independently folded protein domain that contains at least one structural motif that recognizes DNA. Both the DBD and RD are essential components necessary for DNA binding and gene expression, respectively.

Both prokaryotic and eukaryotic TFs are identified by the presence of a DNA-binding domain using sequence searches against protein family databases such as PFAM and by BLAST-based detection of homologs. TFs can also be described by their DNA binding affinity and DNA binding specificity. The affinity of the TF for a specific DNA sequence is defined as the dissociation constant (Kd), the concentration of free TF for which the DNA is half bound (Stormo & Zhao, 2010). On the other hand, specificity can be defined as how a protein distinguishes between two different DNA sequences; the specificity of TFs entails the list of Kds to all possible binding sites. DNA-binding specificities can be represented as consensus sequences (e.g., using IUPAC codes), Position Weight Matrices (or PWMs) (which are visualized as “sequence logos”), or by a table of relative affinities to individual sequences (Sai et al., 2011); **Figure 5**.

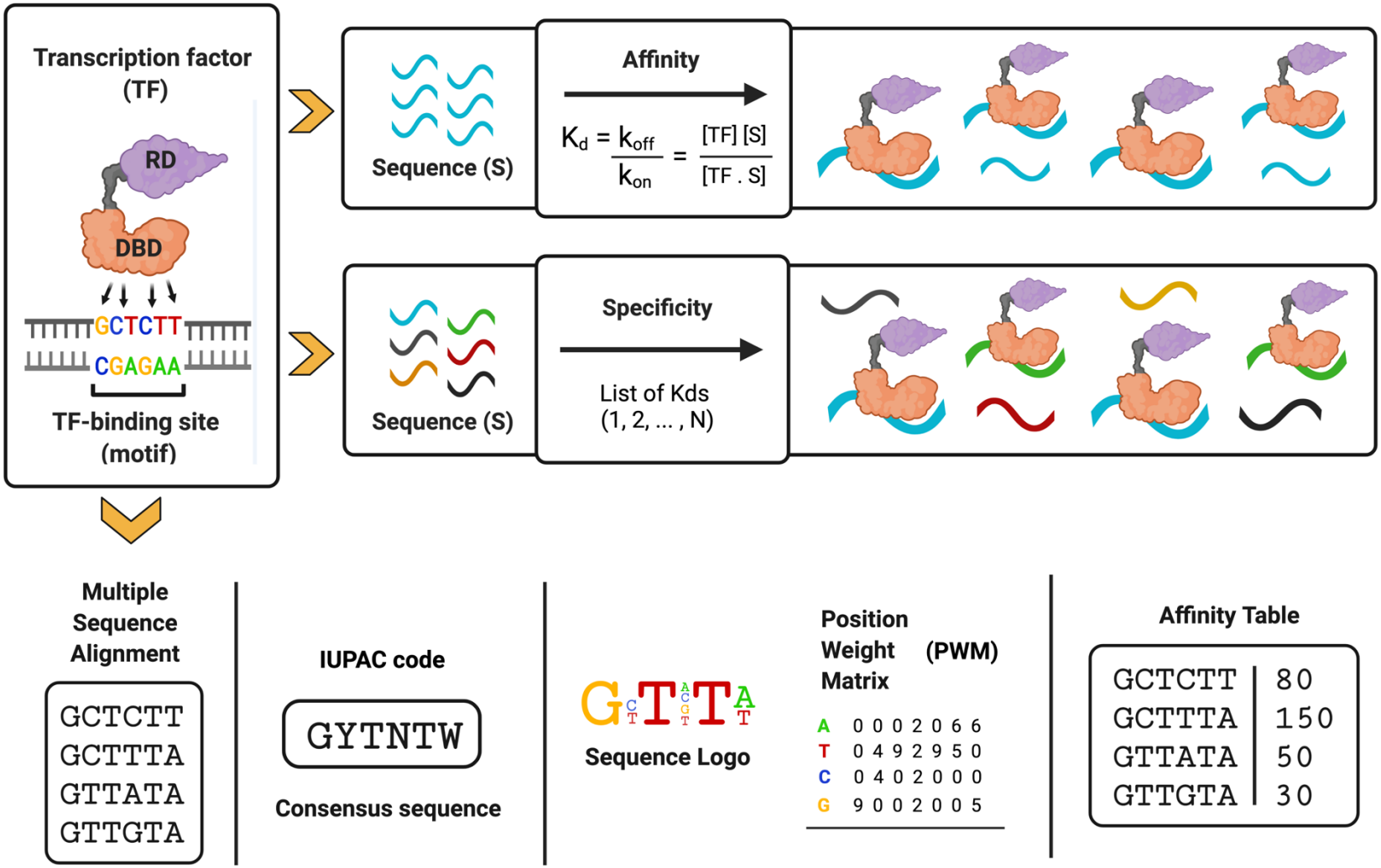


Figure 5. Transcription factors: overview. The binding of protein to DNA is a bimolecular process governed by two rates: an on-rate for the formation of the complex and an off-rate for its dissociation, this value is the Kd. Affinity is the strength of binding of a single molecule to its ligand. Whereas specificity refers to how well a protein can distinguish between different sequences, value is given by the complete Kd list of all selected sequences. Bioinformatically speaking, DNA-binding specificities can be represented as consensus sequences (derived from a multiple sequence alignment), PWMs which are visualized as "sequence logos" or by a table of relative affinities to individual sequences. Created with Biorender.com

1.5 THE LYSR TRANSCRIPTION FACTORS

The LysR family comprises similar-sized, LTTRs respond to a diversity of co-inducers during transcription regulation. LTTRs usually have a composition of approximately 330 amino acids (Schell, 1993). The name LysR comes from the transcriptional activation of the *lysA* gene, which codes for an enzyme that produces lysine. Most family members share the following characteristics: encoding of an inducer-responsive transcriptional activator protein (276 to 324 residues in length); binding at DNA sequences that have similar position and structural motif and is independent of the co-inducer presence; divergently transcribed from a promoter that is nearby to or that often overlaps a promoter of a regulated gene; repression of their own transcription; can bind as a dimer or tetramer (Schell, 1993). A schematic representation model for LTTR-dependent transcriptional regulation can be observed in **Figure 6**. Multiple studies at the molecular level of the LysR TFs itself has made it the best characterized member of the group, providing in turn the family namesake (Maddocks & Oyston, 2008).

LTTRs have a ligand-binding domain C-terminal to the DNA-binding domain. The C-terminal of the LTTRs often acts as a co-factor binding domain, diagram in **Figure 7**. The C-terminal comprises an α-helix and β-strand traversing both the RD and DBD domains. For example, residues 236-246 in the β-strand contain part of the β-turn-β ligand binding crevice predicted for many family members such as NodD (Györgypa & Kondorosi, 1991). The C-terminal region includes RD1 and RD2 domains joined by hinge regions found at residues 167–170 and 270–273. The co-inducer-binding cleft resides between the RD1 and RD2 regions, with a depth of 10 Å and a diameter of 7 Å. Mutational analysis experiments support the importance of the co-inducer-binding cleft and its flanking sites in multiple functions of LTTRs. Mutational changes in the protein's sequence can lead to loss of transcriptional function and DNA binding activity. For

example, substitution mutations at positions 231 and 252 in NahR (Huang & Schell, 1991) or position 234 of OxyR (Morgan et al., 1986), members of the LTTRs, lead to a co-inducer independent phenotype.

The N-terminus is often composed of a helix-turn-helix (HTH) motif, which provides a means of binding to DNA. Uniquely, >20% amino acid sequence identity or the presence of a consensus sequence for the highly conserved amino terminus with another LTTR family member indicates the presence of an LTTR (Schell, 1993). The central portion of this highly conserved region (residues 23 - 42) is nearly 40% identical to all LTTRs, containing an HTH at the N-terminal. The HTH motif is present in all LTTRs and 95% of all prokaryotic DNA binding proteins (Maddocks & Oyston, 2008). The first full-length LTTR crystal structure resolved was CbnR, **Figure 8**. CbnR is divergently transcribed from the *cbnABCD* operon, the products involved in the degradation of chlorocatechol (Maddocks & Oyston, 2008). Each dimer of CbnR comprises one short-form subunit and one extended form subunit, giving the tetrameric molecule an asymmetrical ellipsoidal shape (130 Å x 670 Å x 660 Å). Each subunit has two domains, a DBD (residues 1 – 58) and a regulatory domain (residues 88 – 294) joined by a linker region (residues 59 – 87). The subunits dimerize through an anti-parallel helix-helix interaction, and the dimers interact along a twofold axis. The resulting ellipsoid has a cavity of (30 Å x 615 Å x 610 Å) that accommodates the co-inducer, which can be 3-chlorobenzoate or benzoate. The DBD lies in a V-shape at the base of the tetramer that consists of three α-helices and two β-strands that form a winged-HTH (Ogawa et al., 1999). This structure is similar to related LysR TF ModE of *E. coli* and has been used to model numerous HTH regions, including OxyR.

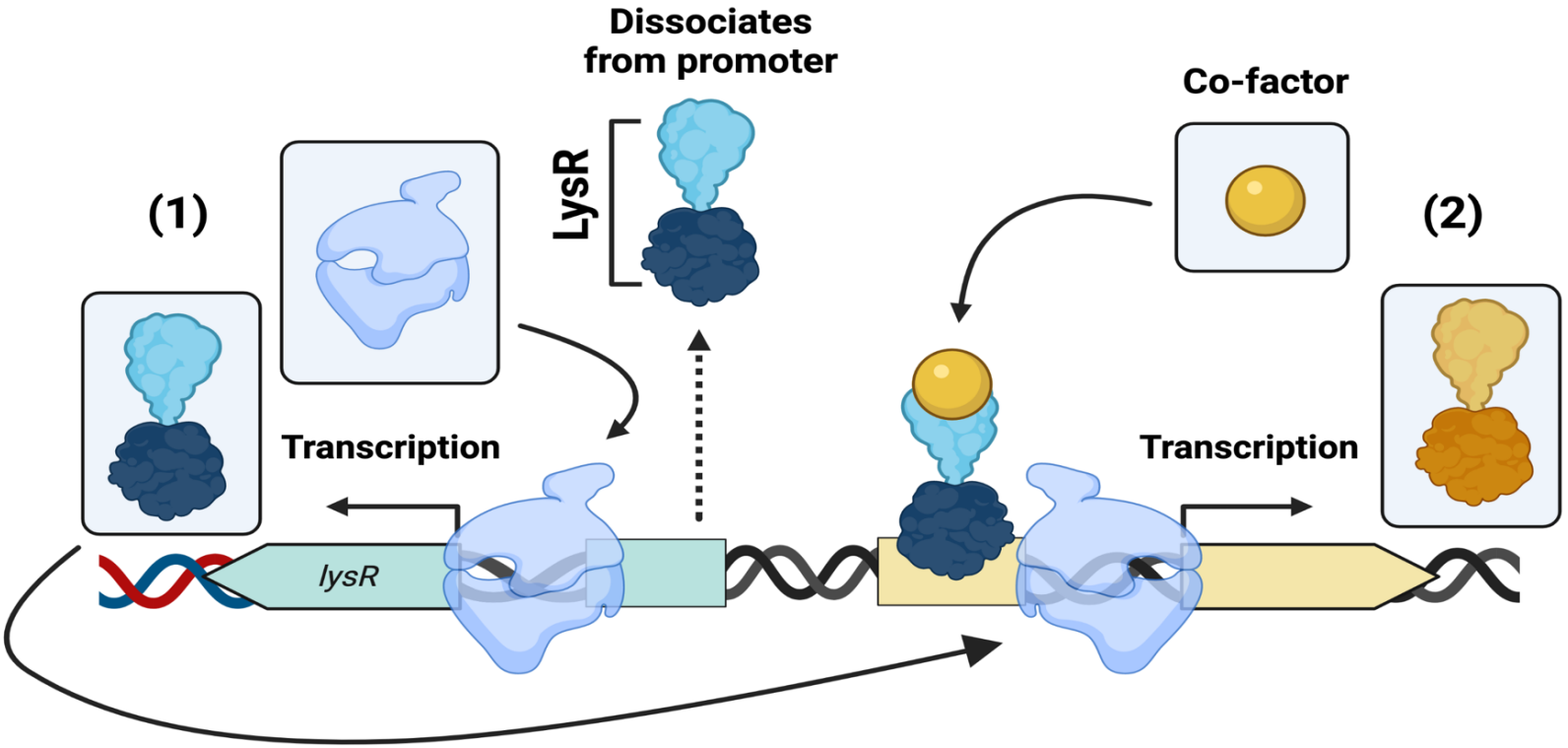


Figure 6. Schematic representation of the classical model for LTTR-dependent transcriptional regulation. (1) The *lysR* gene is transcribed when the LysR protein is dissociated from its promoter. (2) The LysR protein product binds upstream of the promoter to the divergently transcribed target gene; when the co-factor interacts with the LTTR, transcription of this gene is activated. Created with Biorender.com

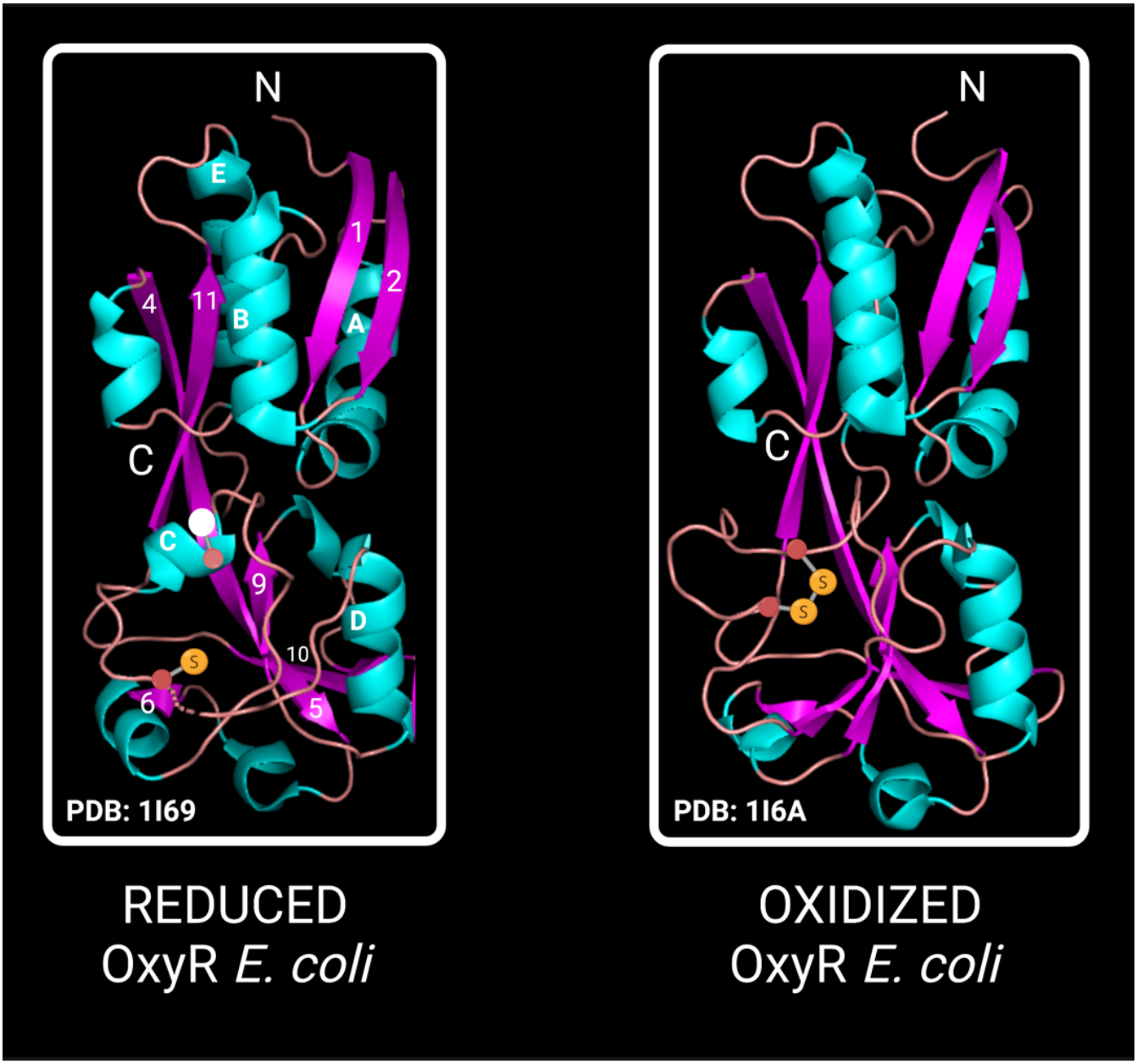


Figure 7. The schematic ribbon diagrams of the OxyR monomers in the reduced (A) and oxidized (B) forms. They are shown with the redox-active cysteines Cys-199 (C199S in the reduced form) and Cys-208 in a ball-and-stick representation. The β strands and α helices are shown in magenta and cyan blue, respectively. Secondary structural elements are labeled in the reduced form (A); the boundaries are β1 (92–98), β2 (120–126), β3 (143–145), β4 (156–160), β5 (162–169), β6 (180–182), β7 (190–193), β8 (219–222), β9 (238–241), β10 (254–260), β11 (264–272), αA (102–116), αB (129–138), αC (199–203), αD (224–233), and αE (279–293).

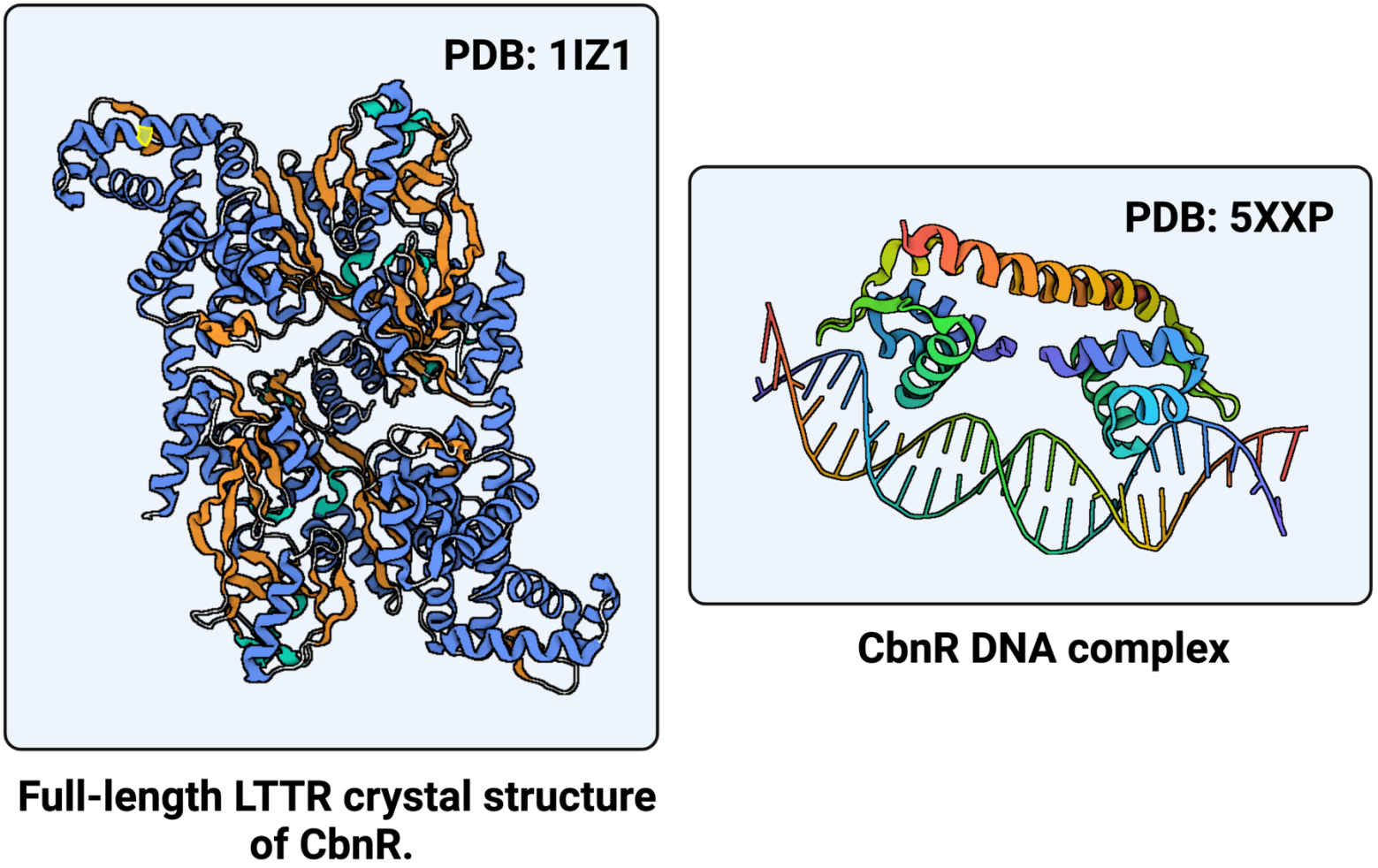


Figure 8. Full-length LTTR crystal structure of CbnR, and the CbnR DNA complex. Created with Biorender.com

1.6 TRANSCRIPTION FACTOR OXYR

OxyR is a LTTR master peroxide sensor TF. OxyR acts primarily as a global regulator of the peroxide stress response by activating the expression of a range of antioxidant defense genes. OxyR also functions as a transcriptional repressor for some genes under normal growth conditions by binding to a more extended region of the target promoters occluding RNA polymerase binding. OxyR consists of an N-terminal DBD with an HTH motif and a C-terminal Regulatory Domain can be seen in **Figure 9**. OxyR senses H_2O_2 with conserved cysteine residues in the RD. By sensing H_2O_2, OyxR is activated, forming an intramolecular disulfide bond via rapid kinetic reaction. The disulfide bond is formed between two highly conserved cysteine residues (Cys-199 and Cys-208) via the oxidation of the Cys-199 to a sulfenic acid intermediate (Zheng & Storz, 1998). The oxidized state of the protein results in a structural change in the RD that accounts for the redox-dependent switch. Hence, only oxidized OxyR can induce the cooperative binding of RNAP and, in turn, activate transcription. OxyR is considered a novel example of protein regulation by fold editing through a reversible disulfide bond formation.

The OxyR protein consists of 305 amino acid residues in *E. coli*. The N-terminal domain contains a helix-turn-helix DNA binding motif, connected to the C-terminal regulatory and oligomerization domain (residues 80 – 305) by a flexible linker that is sensitive to proteolytic digestion (Zheng & Storz, 1998). The C-terminal domain contains the redox-active cysteines (Cys-199 and Cys-208) that mediate the redox-dependent conformational switch. Biochemical analysis and DNase I and hydroxyl radical footprinting studies showed that the interaction between the OxyR occurs in the form of a tetramer. The interaction between the OxyR tetramer and the DNA is significantly different between the oxidized and reduced forms (Toledano et al., 1994). Reduced

OxyR binds DNA at two pairs of major grooves separated by one helical turn, while oxidized OxyR occupies four consecutive major grooves.

OxyR, being a member of the LTTRs, has been seen to control a regulon that mediates the oxidative-stress response in multiple bacterial members such as *E. coli*, *S. typhimurium, Vibrio vulnificus*, among others For instance, exposure of *E. coli* to 0.05 mM H_2O_2 leads to the production of 30 proteins that protect it from subsequent exposure to 200-fold higher levels of H_2O_2 (Morgan et al., 1986). Several of these H_2O_2-induced proteins and their genes characterized are [e.g., *katG* (catalase), *ahpCF* (alkyl peroxide reductase), and *gorA* (glutathione reductase)]. OxyR positively regulates these genes in *E. coli* (**Table 1**). Unlike other LTTRs, OxyR activates transcription after being oxidized by hydrogen peroxide rather than in response to the binding of a specific co-inducer. Thus, the oxidation state of OxyR affects its ability to activate transcription but not the binding of the protein to its target sites.

The binding sites and mechanism of action of OxyR appear somewhat different from those of other LTTRs, perhaps so OxyR can recognize many different promoters in response to oxidative stress (Schell, 1993). The protein functions as a dimer and tetramer, displaying some cooperativity that may avoid activation when H_2O_2 is below a threshold. For example, the facultative bacterium *V. vulnificus* employs two OxyR proteins similar to *A. fischeri*. OxyR1 and OxyR2 in *V. vulnificus* can sense distinct levels of H_2O_2. In *V. vulnificus*, OxyR1 is activated by the low levels of H_2O_2 endogenously generated upon aeration, whereas OxyR2 responds to severe influxes of H_2O_2 from the external environment (Kim et al., 2014a). These combined results demonstrate that OxyR1 and OxyR2 function as a gradient-state redox switch to regulate expression tightly.

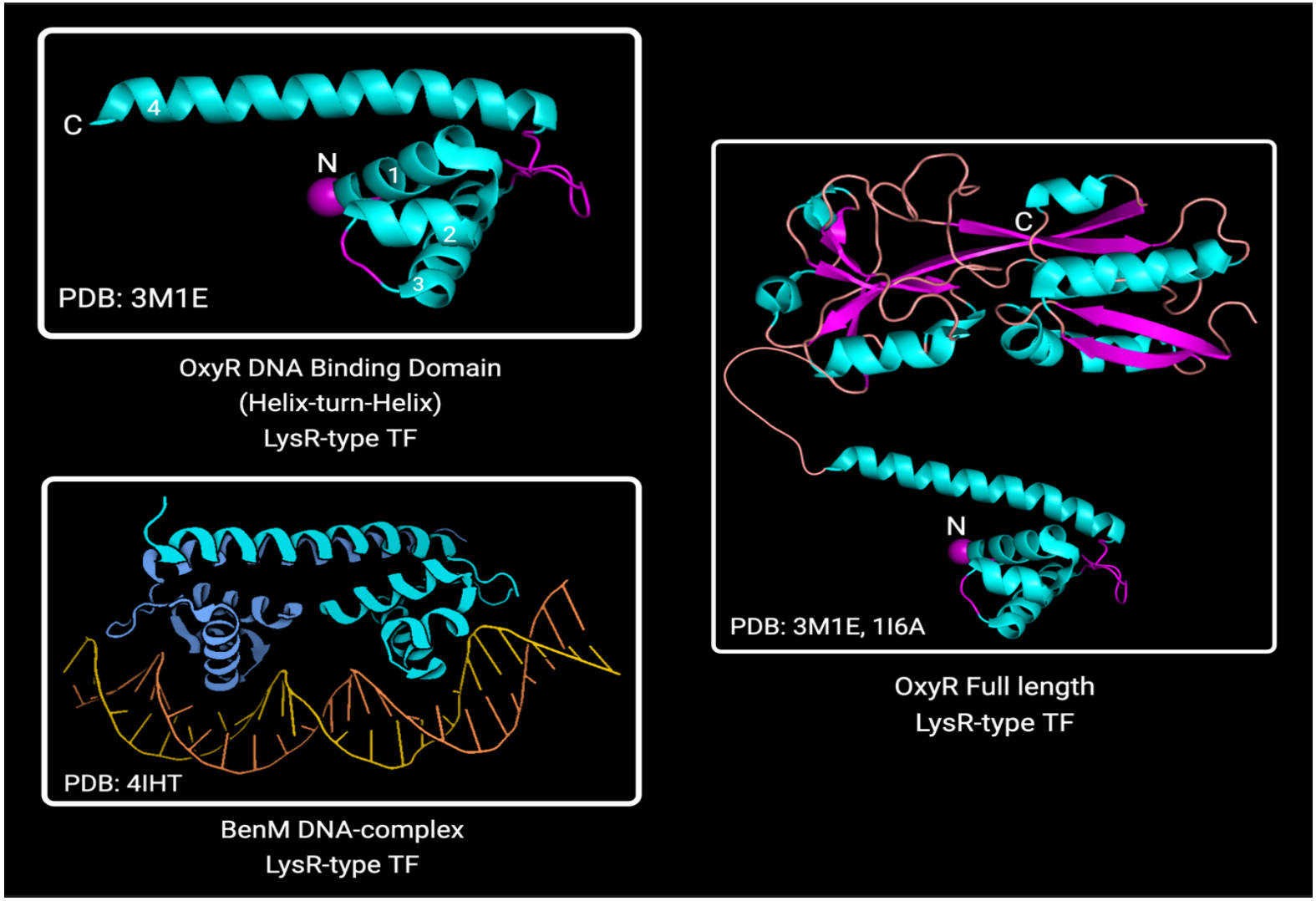


Figure 9. Schematic ribbon diagrams of the structure of the DNA binding domain, full-length OxyR transcription factor, and BenM DNA-complex. Homology model of the DBD. The HTH is marked in blue. The full-length monomer of the OxyR oxidized form.

Table 1. Summary of important genes that control the oxidative stress response in *Escherichia coli*

Genes	Oxidative stress defense activity	Impact in the stress response	Reference
katG	catalase	Scavenge H_2O_2	Jacobson et.al., 1989
ahpCF	NADH peroxidase		
dps	Mini-ferrtin	Minimize free iron	Grant et.al., 1998
fur	iron-import repressor		Varghese et.al., 2007
mntH	manganese importer	activates mononuclear enzymes	Anjem et.al., 2009
sufA-E	iron-sulfur assembly	activated Fe/S enzymes	Outten et.al., 2004
hemH	ferrochelatase	sustain heme synthesis	Mancine et.al., 2015
gor	gluthathione reductase	thiol maintenance	Aslund et.al., 1999
trxC	thioredoxin		
grxA	glutaredoxin		
dsbg	protein sulfenate reductase		

1.7 REACTIVE OXYGEN SPECIES (ROS) AND DNA DAMAGE

Among the factors that can mutate an organism DNA, seen in **Figure 10,** two main classes involve endogenous and exogenous agents. Exogenous DNA damage arises when environmental, physical, and chemical agents damage the DNA (e.g., UV, ionizing radiation, alkylating agents, and crosslinking agents). On the other hand, endogenous DNA damage arises from chemically active DNA engaging in hydrolytic and oxidative reactions with water and ROS (Kim et al., 2014b).

ROS is a collective term that includes oxygen radicals like superoxide, hydroxyl, peroxyl, and hydroperoxyl. ROS also includes nonradical oxidizing agents, such as H_2O_2, HOCl, and O_3. ROS can be converted easily into radicals or species that contain one or more unpaired electrons (Nimrat Chatterjee, 2017). A variety of natural circumstances can elevate ROS beyond homeostatic levels, that can lead to oxidative burst damage. How do bacteria defend themselves against ROS and oxidative burst damage? Often microbes maintain transcription factor inducible defense systems to prevent any detrimental shifts in cellular homeostasis. Three transcription factors extensively studied are SoxR, which defrays the toxic effects of ROS, along with SoxS, and OxyR, which responds to H_2O_2 stress. For example, *E. coli* 's exposure to hydrogen peroxide induces the activation of *oxyR* gene, which in turn minimizes DNA damage by activating the active scavenging of ROS (Bayir, 2005).

Hydrogen peroxide and superoxide are two of the most harmful reduced forms of oxygen, leading to cellular death. For example, the addition of multicopy plasmids carrying the *oxyR*-regulated genes, catalase (*katg*), and alkyl hydroperoxide reductase (*ahp*) in strains with *oxyR* deletions result in the overexpression of enzymatic activity. This activity causes reversion of spontaneous mutations to wild type levels, by the protective effect of these enzymes that capable

of scavenging active oxygen species (Imlay & Linn, 1987). H_2O_2 is a small and uncharged molecule that passively crosses membranes and is difficult to be excluded from the cell. Once inside the cell, it disrupts multiple aspects of iron metabolism, thereby attacking homeostasis of the system, a universal feature of life (G. Storz et al., 1987). Often hydrogen peroxide arises in natural habitats through reactions between sulfur and oxygen at oxic/anoxic interfaces. In addition, several organisms, such as plants, animals, and certain bacteria, excrete hydrogen peroxide to poison local microbes (Imlay, 2015b). Therefore, *oxyR* and *oxyR*-regulated genes have a clear role in the defense against oxidative DNA damage and regulation of transcription.

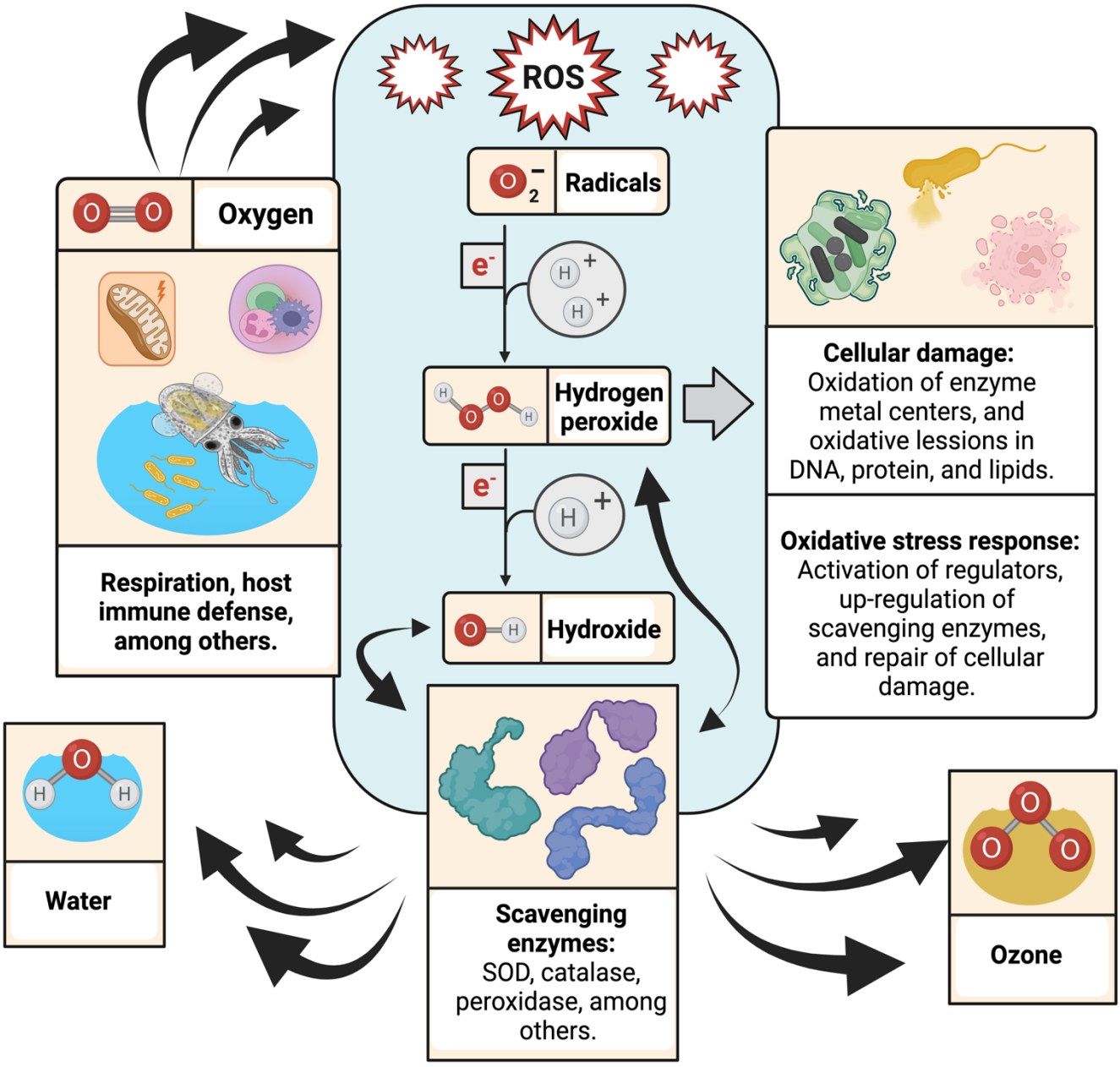


Figure 10. Source and consequences of ROS in bacteria. Summary of the sources and consequences of deleterious ROS such as oxygen radical, hydrogen peroxide, and hydroxyl radical. These ROS are mainly involved in the immune defense by both animals and plants. Although ROS can be beneficial, they are generally detrimental to organisms due to their interaction with DNA, proteins, lipids and other biomolecules, leading to dysfunction, genetic mutation, and peroxidation. Scavenging enzymes are responsible for cleansing the damaging effects of ROS. Created with Biorender.com

1.8 TUNE INTO THE TIMELINE: DISCOVERY AND KEY STUDIES OF OXYR

Pre-treatment of *E. coli* and *S. typhimurium* with small doses (10 and 30 μM) of H_2O_2 induces protection against the toxic effects of subsequent treatment of H_2O_2 (Imlay, 2015b). *E. coli* and *S. Typhimurium* are introduced as good experimental models for assessing ROS stress adaptation. Notably, *E. coli* and *S. typhimurium* use an array of enzymes highly involved in the detoxification of ROS and repairing ROS-induced damage . The induced protection against H_2O_2 mutations is proportional to catalase production (Winquist et al., 1984). Catalase accelerates the breakdown of H_2O_2, lowering its effective dose and, in turn, reducing damaging effects (Torraca et al., 2014).

S. typhimurium, a member of the *Enterobacteriaceae* family, acts as a facultative intracellular pathogen causing intestinal and invasive diseases in animals (Winquist et al., 1984) and humans (Hur et al., 2012). For example, in mice, *S. typhimurium* has to survive inside macrophages to cause disease (Caradonna et al., 2000). Inside the macrophages, the excessive production of reactive ROS can lead to oxidative burst, a challenge bacteria have to overcome for survival (Fields et al., 1986). For example, once the macrophages have engulfed *S. typhimurium,* it is exposed to ROS; one gateway to survival is activating its oxidative defense mechanisms.

The Ames group began to study *S. typhimurium*, to understand the regulation of bacterial defenses against H_2O_2 generated oxidative stress. It was determined that during the H_2O_2 pre-adaptive step, thirty proteins were induced and subsequently analyzed using two-dimensional gel electrophoresis. Nine out of the thirty proteins constitutively overexpressed are dominant H_2O_2-resistant (Minakami & Sumimoto, 2006). H_2O_2-resistant mutants were isolated following mutagenesis with diethyl sulfate and were assigned the name of *oxyR1* mutants.

Strains carrying a Tn5 insertion element linked to oxyR1 in regions near *argE* on *S. typhimurium* were isolated. One H_2O_2-resistant transductant (oxyR11argH1823:Tn10) overexpress the nine *oxyR1* proteins constitutively. A single genetic locus coordinately controls the proteins produced by *oxyR1* (Christman et al., 1985). Deletions of the oxyR gene are recessive and uninducible by H_2O_2 for the nine proteins overexpressed in the oxyR1 H_2O_2-resistant mutants, proposing *oxyR* as a positive regulatory element. Three of them: KatG (catalase), Ahp (alkyl hydroperoxide reductase), and DnaK (DNA biosynthesis), have activities involved in the defense against oxidative damage and are regulated by *oxyR* at the transcriptional level (Christman et al., 1985). This discovery allowed for a previously uncharacterized global regulatory system, the *oxyR* regulatory network.

In 1989 the Christman, Storz, and Ames groups revealed *E. coli* 's *oxyR* gene cloning and sequencing, suggesting an opening reading frame (305 amino acids) that encodes a 34.4 kDa protein named OxyR (Morgan et al., 1986). Early studies of amino acid sequence comparison using the Atlas of Protein Sequence and Structure (Dayhoff data bank) found homology of the OxyR protein with the LysR *E. coli* 's protein (member of the LTTRs). Storz, Tartaglia, and Ames group (1990) successfully purified and characterized the first OxyR protein from *E. coli*. The article was published in Science under "*Transcriptional Regulator of Oxidative Stress-Inducible Genes: Direct Activation by Oxidation*."

Footprinting studies suggest that OxyR recognizes different functional binding sites with high and nearly equal affinity and specificity, despite the lack of sequence similarity. Differences in the footprints between reduced and oxidized OxyR at the *katg*, *ahpC*, and *oxyR* promoters suggest that a conformational change in OxyR transduces an oxidative stress signal to RNA polymerase (Tao et al., 1989). Specific recognition of non-homologous sites is possible due to

contacts made over a large area (~ 43 bp) in which degenerate DNA recognition occurs at many positions (Storz et al., 1990). Based on DNA footprinting studies, Toledano, Kullik, and Trinh (1994) reported that the oxidized and reduced OxyR contact a DNA motif comprised of four 5' - ATAGnT- 3' elements. The reduced forms bind the ATAGnT elements spaced at 10 bp intervals but oxidized OxyR contact these elements in four adjacent major grooves at the DNA helix (Tartaglia et al., 1992).

Different transcription factors such as OxyR, SoxR, and SoxS have played a key role in the transcriptional regulation of the defense system against oxidative stress in multiple bacteria species. OxyR, SoxR, and SoxS have around 18, 3, and 21 target binding sites; identified by studying *E. coli* through *in vitro* DNA-binding experiments, mutational analysis, and microarray-based comparative transcriptomics (Toledano et al., 1994). Seo, Kim, Szubin, and Palsson (2015) reconstructed OxyR, SoxR, and SoxS transcriptional regulatory networks under oxidative stress in *E. coli in vivo*. Techniques such as chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Exo) are useful for *in vivo* binding site identification in the *E. coli* K-12 MG1655 genome for each transcription factor. In addition, strand-specific massively parallel cDNA sequencing (RNA-seq) assesses transcription level measurements of both wild-type and knockout mutants for each TF to identify causal relationships. Integrative data from both experiments allow for reconstructing the oxidative stress metabolic network.

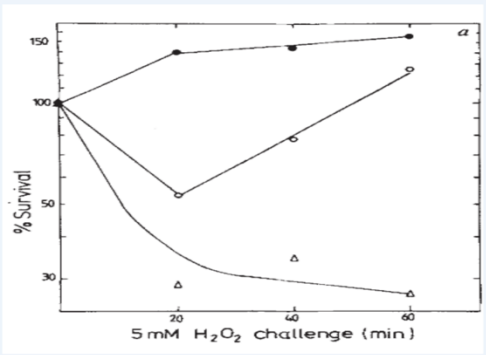
Analysis revealed that OxyR, SoxR, and SoxS regulons, with this new data, expands to 68 target genes in 51 transcriptional units (OxyR, 38 genes in 28 TUs; SoxR, 11 genes in 10 TUs; and SoxS, 34 genes in 25 TUs). Thirty-eight target genes in 29 TUs were previously identified (Seo et al., 2015). ChIP-exo experiments identified 28, 10, and 25 reproducible binding sites for OxyR,

SoxR, and SoxS, respectively, under oxidative stress. They also detected 76% (32 of 42) of OxyR, SoxR, and SoxS binding sites reported from *in vitro* experiments (Seo et al., 2015).

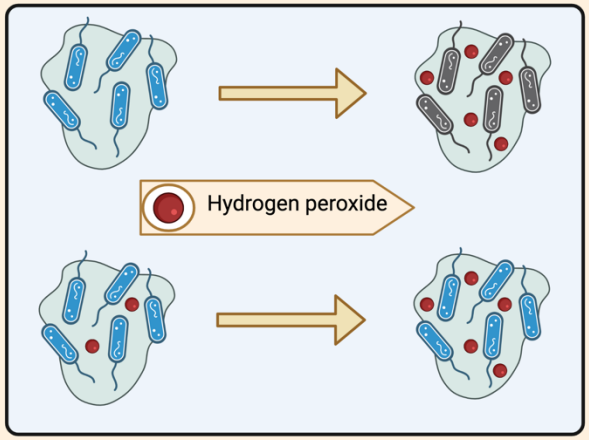
Reconstruction of the TFs genome-wide role includes but is not limited to the activation of genes related to amino acid and bacterial cell wall biosynthesis and divalent metal ion transportation. They suggest that ROS responses in microbes hold an important role in novel antibiotic treatments. High-throughput experiments allowed for a more comprehensive understanding of the coordinated genome-wide regulatory roles of the TFs in cellular responses specific to oxidative stress. An Overview of the OxyR timeline is presented in **Figure 11**.

START

Pre-treatment of *E. coli* and *S. typhimurium* with small doses of H_2O_2 induces protection against toxic effects of subsequent treatment of H_2O_2.



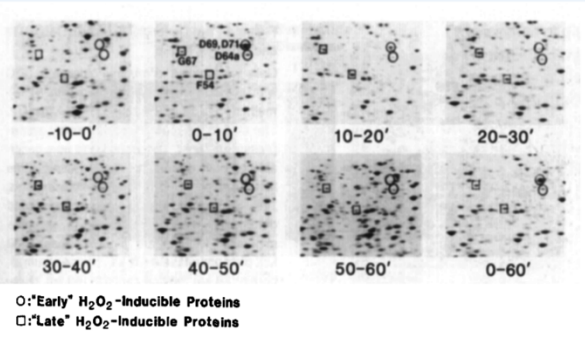
1983 - Demple and Halbrook



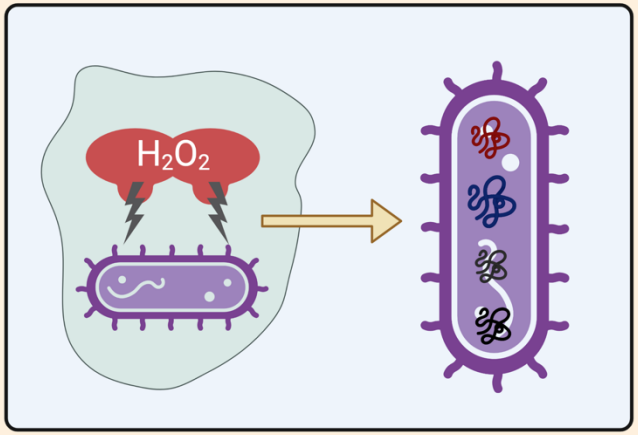
Demple B, Halbrook J. Inducible repair of oxidative DNA damage in Escherichia coli. Nature. 1983 Aug 4-10;304(5925):466-8. doi: 10.1038/304466a0. PMID: 6348554.

A total of thirty proteins are induces during the H_2O_2 pre-adaptative step.

H_2O_2 resistant mutants were isolated and were assigned the name of oxyR1 mutants.

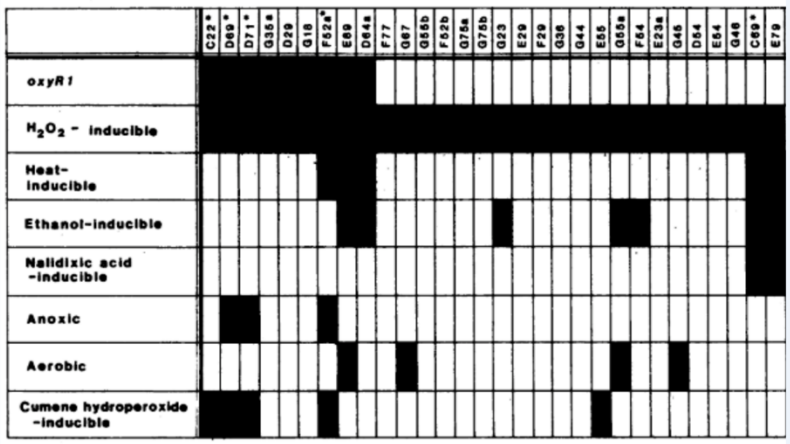


1985 - Christman, et. al

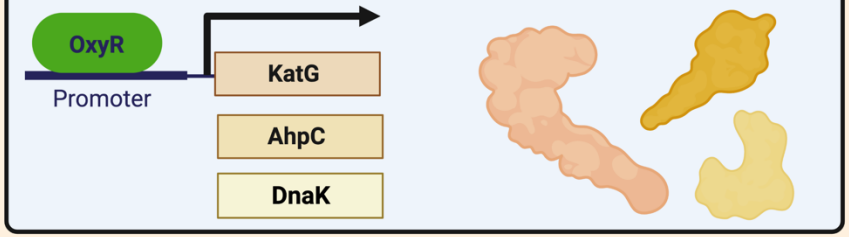


Christman MF, Morgan RW, Jacobson FS, Ames BN. Positive control of a regulon for defenses against oxidative stress and some heat-shock proteins in Salmonella typhimurium. Cell. 1985 Jul;41(3):753-62. doi: 10.1016/s0092-8674(85)80056-8. PMID: 2988786.

oxyR is proposed to be a positive regulator element for anti-oxidative stress genes

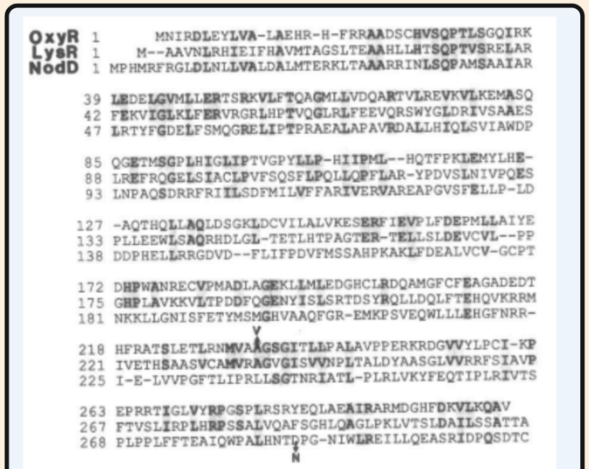


1986 - Morgan, et. al

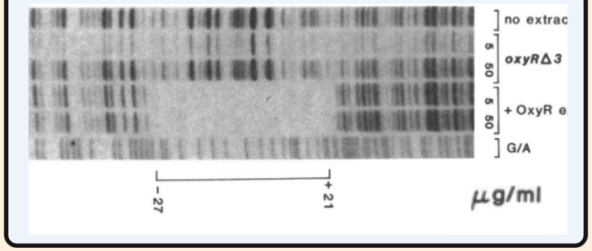


Morgan, R. W., Christman, M. F., Jacobson, F. S., Storz, G., & Ames, B. N. (1986). Hydrogen peroxide-inducible proteins in Salmonella typhimurium overlap with heat shock and other stress proteins. Proceedings of the National Academy of Sciences of the United States of America, 83(21), 8059–8063. https://doi.org/10.1073/pnas.83.21.8059

*E. coli*s oxyR gene is cloned and sequenced; 34.4 kDa protein named OxyR1. A footprint made for OxyR reveals DNA interaction.



1989 - Christman, Storz, and Ames



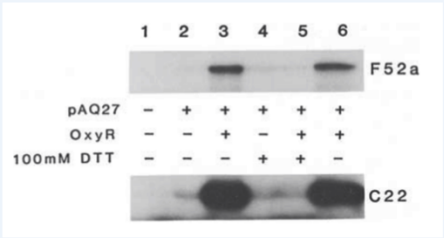
Christman MF, Storz G, Ames BN. OxyR, a positive regulator of hydrogen peroxide-inducible genes in Escherichia coli and Salmonella typhimurium, is homologous to a family of bacterial regulatory proteins. Proc Natl Acad Sci U S A. 1989 May;86(10):3484-8. doi: 10.1073/pnas.86.10.3484. PMID: 2471187; PMCID: PMC287162.

Purification and characterization of the OxyR protein from *Escherichia coli*

Table 1. The *oxyR* Regulon

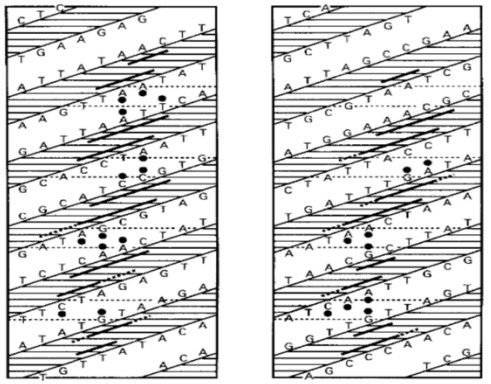
Activity		Gene	*E. coli* map position
OxyR		*oxyR*	89.6
Catalase	HPI	*katG*	89.2
Alkyl hydroperoxide reductase	C22, F52	*ahpC, ahpF*	13
Glutathione reductase		*gorA*	77–78

1990 - Storz, Tartaglia, and Ames

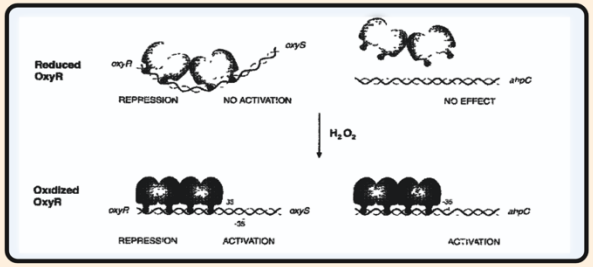


Storz G, Tartaglia LA, Ames BN. The OxyR regulon. Antonie Van Leeuwenhoek. 1990 Oct;58(3):157-61. doi: 10.1007/BF00548927. PMID: 2256675.

Oxidized and reduced OxyR versions contact a DNA motif comprised of four ATAGnT elements.



1994 - Toledano, Kulik, and Trinh



Toledano MB, Kullik I, Trinh F, Baird PT, Schneider TD, Storz G. Redox-dependent shift of OxyR-DNA contacts along an extended DNA-binding site: a mechanism for differential promoter selection. Cell. 1994 Sep 9;78(5):897-909. doi: 10.1016/s0092-8674(94)90702-1. PMID: 8087856.

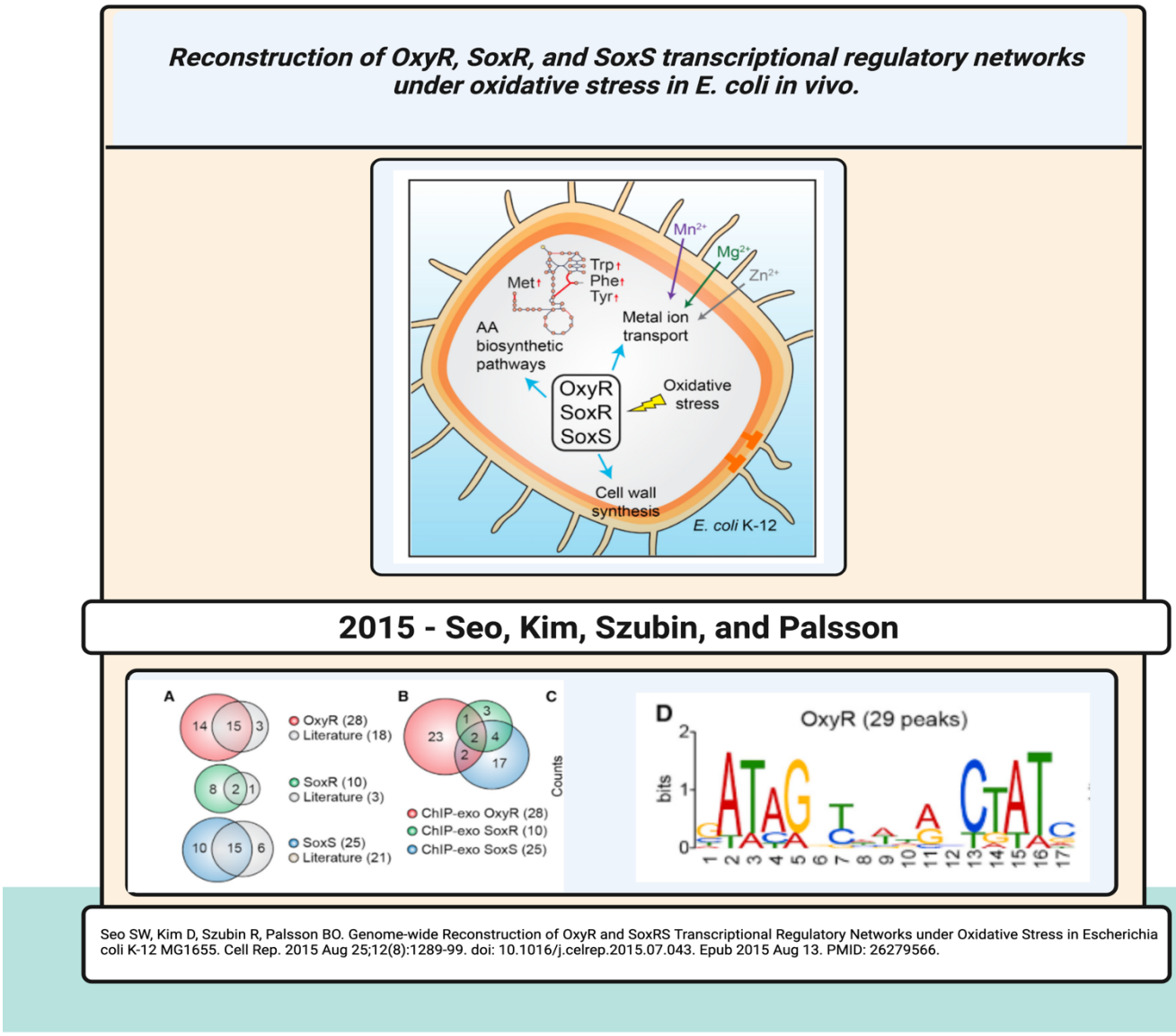


Figure 11. Discovery and key studies of OxyR.

1.9 SYMBIOTIC RELATIONSHIP: *A. FISCHERI - E. SCOLOPES*

A. fischeri is a model system for studying the symbiotic relationships between bacteria and animals. An animal associated with *A. fischeri* is the bobtail squid *Euprymna scolopes*. *E. scolopes* is a small, nocturnal, sepiolid squid that is bioluminescent owing to the presence of bacterial symbionts (*A. fischeri*) contained within the squid's light-emitting organ, **Figure 12**. The genome of *A. fischeri* encodes for two different OxyR transcription factors: OxyR1 and OxyR2. Understanding OxyR1 and OxyR2 would lead to comprehending their overall function and how they could be related to oxidative stress regulation; oxidative stress response by *E. scolopes* during the initial stages of *A. fischeri* colonization.

The *A. fischeri* symbiont is acquired soon after an *E. scolopes* juvenile is born; due to the ubiquitous presence of *A. fischeri* in the vicinity of the newly born bobtail squid. Analysis of the *E. scolopes* juvenile host has revealed biomechanical (e.g., ciliary currents along the ducts leading to the crypt spaces (Nyholm & McFall-Ngai, 2004) and biochemical (e.g., oxidative stress) (Nyholm et al., 2000) mechanisms by which colonization specificity is ensured. Macrophage-like hemocytes isolated from *E. scolopes* during light organ colonization recognized and phagocytosed *A. fischeri* less than other closely related bacterial species (Nyholm et al., 2009). *E. scolopes* innate immune system role during *A. fischeri's* light organ colonization suggests specificity.

Bacteria can harness the deleterious effects of various oxidants that exploit oxidative damage by using TFs or proteins that act as physiological signals that trigger global antioxidant responses (Pomposiello & Demple, 2001). Producing the purified protein will allow for the examination of OxyR1 and OxyR2 DNA-binding specificity, leading to the discovery of crucial target genes; identifying the DNA sequences that TFs bind requires the protein of interest to be purified in enough quantities to perform downstream processes.

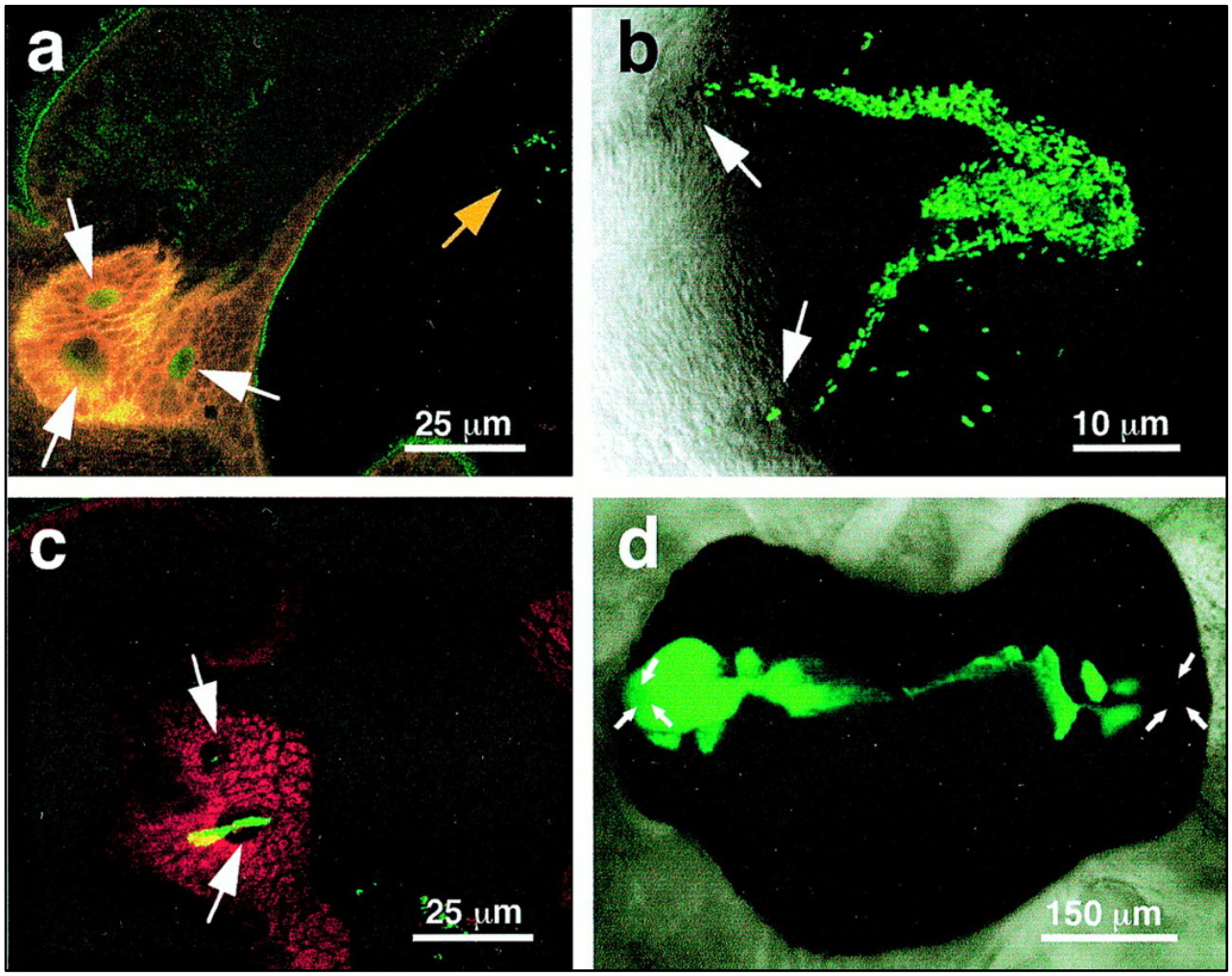


Figure 12. Stages in the process of symbiotic colonization of the bobtail squid's light organ by *A. fischeri*. (a) After 1 hour of exposure to GFP-labeled *A. fischeri*, a laser-scanning confocal micrograph (LSM) revealed a small aggregate (orange arrow) forming above a pore of the light organ. (b) Between 2 and 4 hours after inoculation, bacteria were seen as streams migrating from the aggregate pores. (c) Between 4 and 6 hours after injection, a mass of GFP-labeled *A. fischeri* cells appeared to be migrating through a pore and into a duct of the light organ. (d) Differential interference contrast image of the fully colonized light organ of *E. scolopes*, showing the population of GFP-labeled symbionts (green). White arrows indicate the location of the pore in all panels. https://doi.org/10.1073/pnas.97.18.10231

2 CHAPTER 2

PROTEIN CLONING AND PURIFICATION OF OXYR1 AND OXYR2 FROM *ALIIVIBRIO FISCHERI*

2.1 INTRODUCTION

The concept of plasmid was introduced by Lederberg in 1952 and defined as an extrachromosomal genetic element in bacterial hosts (Lederberg, 1952). Plasmids were used during the first recombinant DNA experiments, and forty years later, they remain as the carriage horses of molecular cloning (Green MR, 2021). Today plasmid vectors are available and optimized for particular purposes. Frequently, plasmids contain genes coding for enzymes advantageous to the bacterial hosts. These genes specify a remarkably diverse set of traits, many of which are of great medical and commercial significance, such as the production of antibiotics, degradation of organic compounds, and enterotoxins, among others (Green, Sambrook, 2012). DNA cloning is a molecular biology technique that allows the synthesis of many identical copies of a DNA piece, such as a gene. A target gene is inserted into a plasmid and then introduced into a bacterium via transformation during a cloning experiment. Although multiple cloning techniques exist and their historical value holds significance in many scientific discoveries, the innovative approach published by Daniel Gibson in 2008 revolutionized the way we perform cloning (Green, Sambrook, 2012).

The Gibson Assembly Cloning method is a procedure that allows the cloning of two or more fragments without the need for restriction enzyme digestion or compatible restriction sites; diagram in **Figure 13**. Instead, user-defined overlapping ends are incorporated into the fragments to allow the seamless joining of adjacent fragments. Due to its many advantages over traditional restriction enzyme cloning, Gibson assembly has rapidly become a preferred method for cloning DNA into plasmids. Some advantages include the rapid cloning of multiple DNA fragments into any vector without restriction enzymes in less than an hour. In addition, the exonuclease contained in the Gibson master mix mediates DNA strand chew back, exposing a single strand which allows

for annealing of the terminal homologous overlap sequences. Annealing the homologous overlap sequences is followed by extension by the DNA polymerase and ligation by a ligase to yield an assembled product. This seamless assembly can be applied to routine ORF cloning and large-size and complex cloning projects.

The purification of proteins is an essential step toward understanding their function. However, the attainment of high quantities of a specific protein derived from endogenous sources is often challenging. To address this situation, researchers take advantage of heterologous expression systems. Heterologous expression systems consist mainly of cloning target genes into artificial vectors designed to operate within easily cultured cells, such as *E. coli*. This system, to some degree, allows the incorporation of modifications to optimize the expression of proteins and facilitate their purification with engineered affinity tags. Purification to near homogeneity is essential for characterizing protein structure, biochemical and biophysical properties, and ligand binding assays (Green, Sambrook, 2012).

Fusing the protein of interest with carrier proteins or tags can increase solubility, proteolysis protection, protein detection by immunological methods, and affinity purification. (Costa et al., 2014). For example, a variety of more than 20 different fusion systems like polyhistidine (6xHis) and S-fusion tags are intrinsically designed to facilitate the purification and detection of targeted proteins. Affinity chromatography is a separation technique that allows for a 'lock and key' fit between the target protein and a specific ligand. For example, immobilized metal affinity chromatography (IMAC), portrayed in **Figure 14**, uses the affinity of 6xHis for divalent metal cations immobilized as transition metal chelate complexes on a chromatography resin.

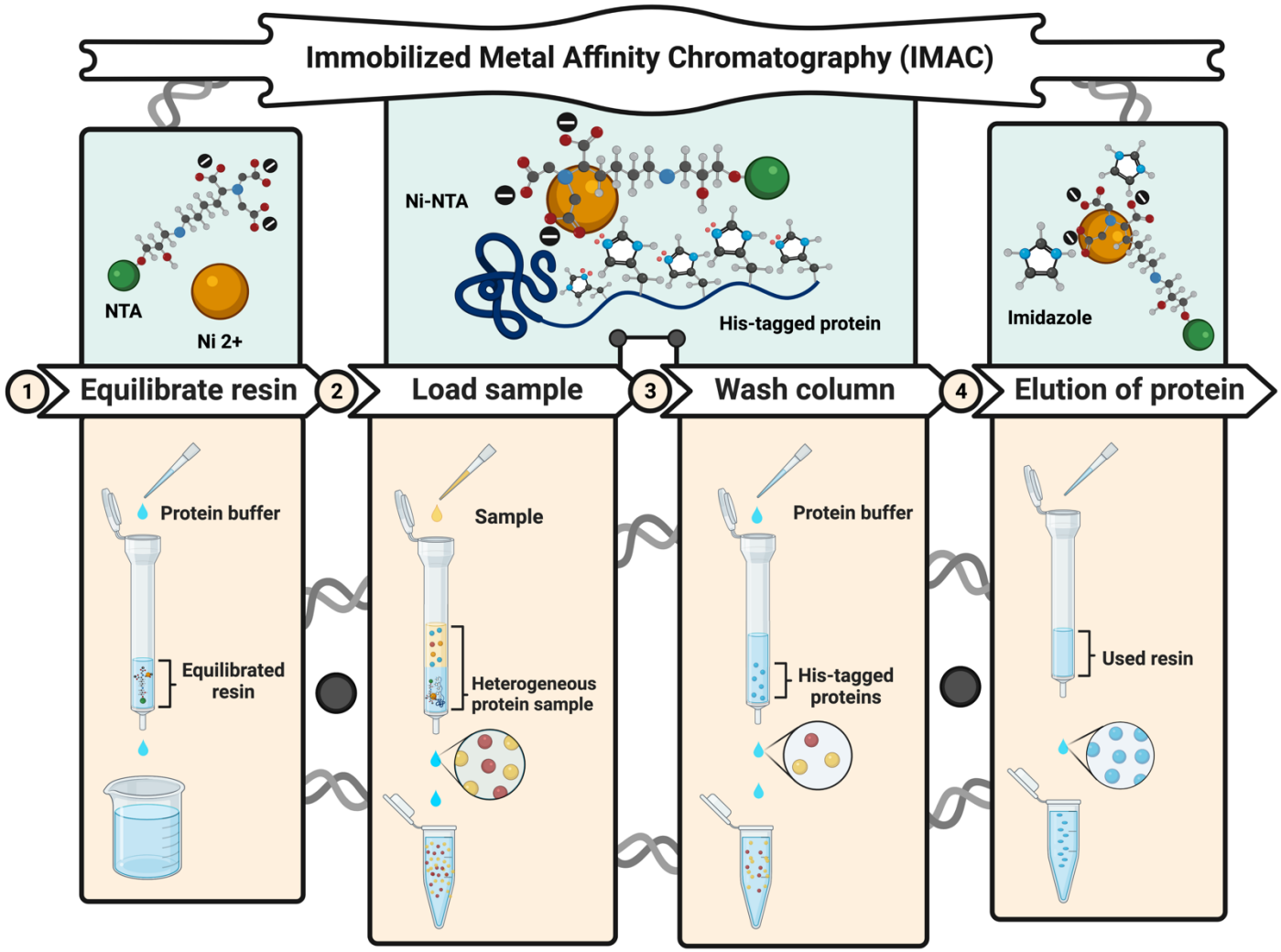


Figure 13. Proteins and peptides with an affinity for metal ions can be separated using metal chelate affinity chromatography. (1) The metals are immobilized onto a chromatographic medium by chelation. (2) The resin composed of the Ni-NTA complex interacts with His-tagged protein. As a result, other proteins are washed away. (3) The sample is further washed in order to remove impurities. (4) Elution and recovery of captured His-tagged protein from an IMAC column are accomplished by using a high imidazole concentration. Metal chelate affinity chromatography is excellent for purifying recombinant (His) fusion proteins; his tagged protein. Created with Biorender.com

2.2 MATERIALS AND METHODS

2.2.1 CLONING

The full-length (FL) and DNA binding domain (DBD) coding sequences of *A. fischeri*'s oxyR1 and oxyR2 genes were obtained from Dr. Zomary Flores' laboratory. *E.coli* genomic DNA was extracted from a BL21(DE3) strain (Millipore Sigma, EU) using the Wizard® Genomic DNA Purification Kit. *E. coli* oxyR gene (b3961) FL and DBD coding sequences of *E. coli* were PCR-amplified from BL21(DE3) the extracted genomic DNA; primers in **Table 2**. Information KEGG database (oxyR1: VF_1974; oxyR2: VF_2299; oxyRE: b3961). Each coding sequences was base-modified with a 15 bp overhang with homology to either, pET-32a(+): (p32a) or pET-51b(+): (p51b) plasmids; primers in **Tables 2, 3**. These customized, synthetic primers were procured from Integrated DNA Technologies, Iowa. Primers were designed using ApE and IDT Oligo Analyzer tools. p32a and p51b were obtained from Novagen, Merck Millipore. Cloning vector maps for p32a and p51b are shown in **Figure 15**.A Phusion High Fidelity PCR amplification was employed for the plasmid linearization of p32a and p51b. PCR-linearized plasmids were gel purified (Qiagen kit), and insert PCR products were column purified (Qiagen kit). DNA concentrations were measured by absorbance at 260 nm using the Nanodrop One Spectrophotometer from Thermo Fisher Scientific Inc. (Madison, WI).

These customized, synthetic gene inserts were cloned into p32a and p51b expression vectors by Gibson Assembly (GA). Judith Kimble's lab (UW Madison) GA master mix protocol was adapted for GA. GA master mix was prepared by combining 36.56 µL water, 133.33 µL 2x isothermal reaction buffer (200 mM TrisCl, pH 7.5, 10% PEG-8000, 20 mM MgCl2, 20 mM DTT, 0.4 mM each of four dNTPs, 2 mM beta-NAD), 0.2 µL T5 Exonuclease + Buffer (0.005 U/ ul), 3.3 µL Phusion DNA polymerase (NEB, 0.033 U/ µL) and 26.6 µL Taq DNA ligase (NEB, 5.3 U/

μL); stocks of 15 μL aliquots were made and stored at −20 ◦C for use. Conversely, a 2X Gibson Assembly Master Mix from New England Biolabs, Massachusetts was incorporated too. In the GA reaction 30 ng of the linearized vector was incorporated into the mixture with a 3-fold excess of the insert. The total volume for each reaction: 20 μL = 5 μL GA master mix + 5 μL water-insert. The GA reaction is incubated for 1 h at 50 °C and after incubation 5 μL is used to transform Dh5-alpha Competent *E. coli* cells. Transformants were plated on LB agar plates containing 100 mg/ml of ampicillin. Transformants were analyzed by colony PCR and DNA sequencing using standard T7 promoter (5′-TAATACGACTCACTATAGGG-3′) and terminator (5′-GCTAGTTATTGCTCAGCGG-3′) primers. The samples obtained were named: *oxyR1 Aliivibrio fischeri* FL (OX1F), *oxyR1 A. fischeri* DBD (OX1D), *oxyR2 A. fischeri* FL (OX2F), *oxyR2 A. fischeri* DBD (OX2D), *oxyR E. coli* FL (OXEF), and *oxyR E. coli* DBD (OXED). Overview of cloning protocols in **Figures 16, 17**.

Table 2. p32a cloning primers.

Primers	Sequence	Information (length, Tm, GC%)
pET32a_linear_Rv	cttgtcgtcgtcgtc	15 bp, 49 °C, 60
pET32a_linear_Fw	caccaccaccaccac	15bp, 52 °C, 67
pET32a_OxyR1_FD_Fw	gacgacgacgacaagATGAATAAATGGCCCTCATTA	21bp, 51 °C, 33
pET32a_OxyR1_F_Rv	gtggtggtggtggtgTAAAATATCTTCGACAATATCAGC	24 bp, 50 °C, 29
pET32a_OxyR1_D_Rv	gtggtggtggtggtgCAACAATTCTCGTCCTTGCTC	21 bp, 55 °C, 48
pET32a_OxyR2_FD_Fw	gacgacgacgacaagATGAATATCCGAGATTTTGAG	21bp, 49 °C, 33
pET32a_OxyR2_D_Rv	gtggtggtggtggtgTTTTACTTCAAGTAAGATCTTTTTC	25 bp, 50 °C, 24
pET32a_OxyR2_F_Rv	gtggtggtggtggtgGCTACGAATAACATTAGAC	19 bp, 46 °C, 37
pET32a_OxyR3_FD_Fw	gacgacgacgacaagATGAATATTCGTGATCTTGAG	21 bp, 49 °C, 33
pET32a_OxyR3_D_Rv	gtggtggtggtggtgCAGCACGGTACGCG	14 bp, 53 °C, 71
pET32a_OxyR3_F_Rv	gtggtggtggtggtgAACCGCCTGTTTTAAAAC	18 bp, 50 °C, 39

Table 3. p51b cloning primers.

Primers	Sequence	Information (length, Tm, GC)
pET51_Fw_new	catcaccaccatcatcaccatc	15bp, 51°C, 60
pET51_Rv_new	cttgtcgtcgtcatctgc	18bp, 54°C, 56
pET51b(+)_OxyR1_Full_Fw	gatgacgacgacaagAATAAATGGCCCTCATTAAAAC	22bp, 51C, 32
pET51b(+)_OxyR1_Full_Rv	gatgatggtggtgatgTAAAATATCTTCGACAATATCAG	21 bp, 48C, 26
pET51b(+)_OxyR1_DBD_Rv	gatgatggtggtgatgCAACAATTCTCGTCCTTG	18 bp, 49C, 44
pET51b(+)_OxyR2_Full_Fw	gatgacgacgacaagAATATCCGAGATTTTGAGTAC	21bp, 49C, 33
pET51b(+)_OxyR2_Full_Rv	gatgatggtggtgatgGCTACGAATAACATTAGAC	19 bp, 46C, 37
pET51b(+)_OxyR2_DBD_Rv	gatgatggtggtgatgTTTTACTTCAAGTAAGATCTTTTTC	25 bp, 50C, 24
pET51b(+)_OxyRE_Full_Fw	gatgacgacgacaagAATATTCGTGATCTTGAGTAC	21bp, 49C, 33
pET51b(+)_OxyRE_Full_Rv	gatgatggtggtgatgAACCGCCTGTTTTAAAAC	18 bp, 50C, 39
pET51(+)_OxyR3_DBD_Rv	gatggtgatgatggtggtgatgAACCGCCTGTTTTAAAACTTTATC	18 bp, 50 °C, 39

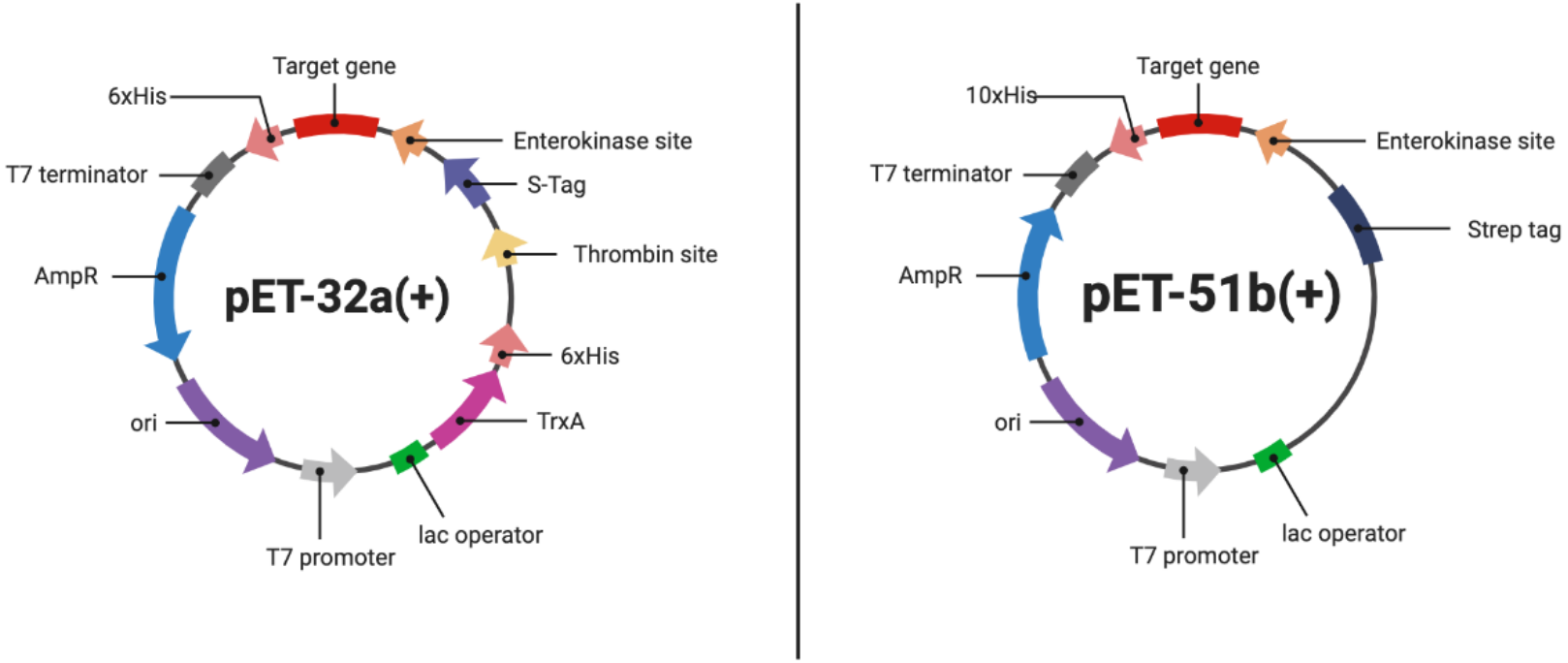


Figure 14. p32a series is designed for cloning and high-level expression of peptide sequences fused with the 109 aa Trx Tag (thioredoxin protein) and S Tag for detection and purification. Fusion proteins also containing cleavable His Tag. Unique sites are shown on the circle map. p51b vector is designed for cloning and high-level expression of target proteins fused with the 8 aa Strep Tag II coding sequence that is cleavable with enterokinase (Ek) protease. Created with Biorender.com

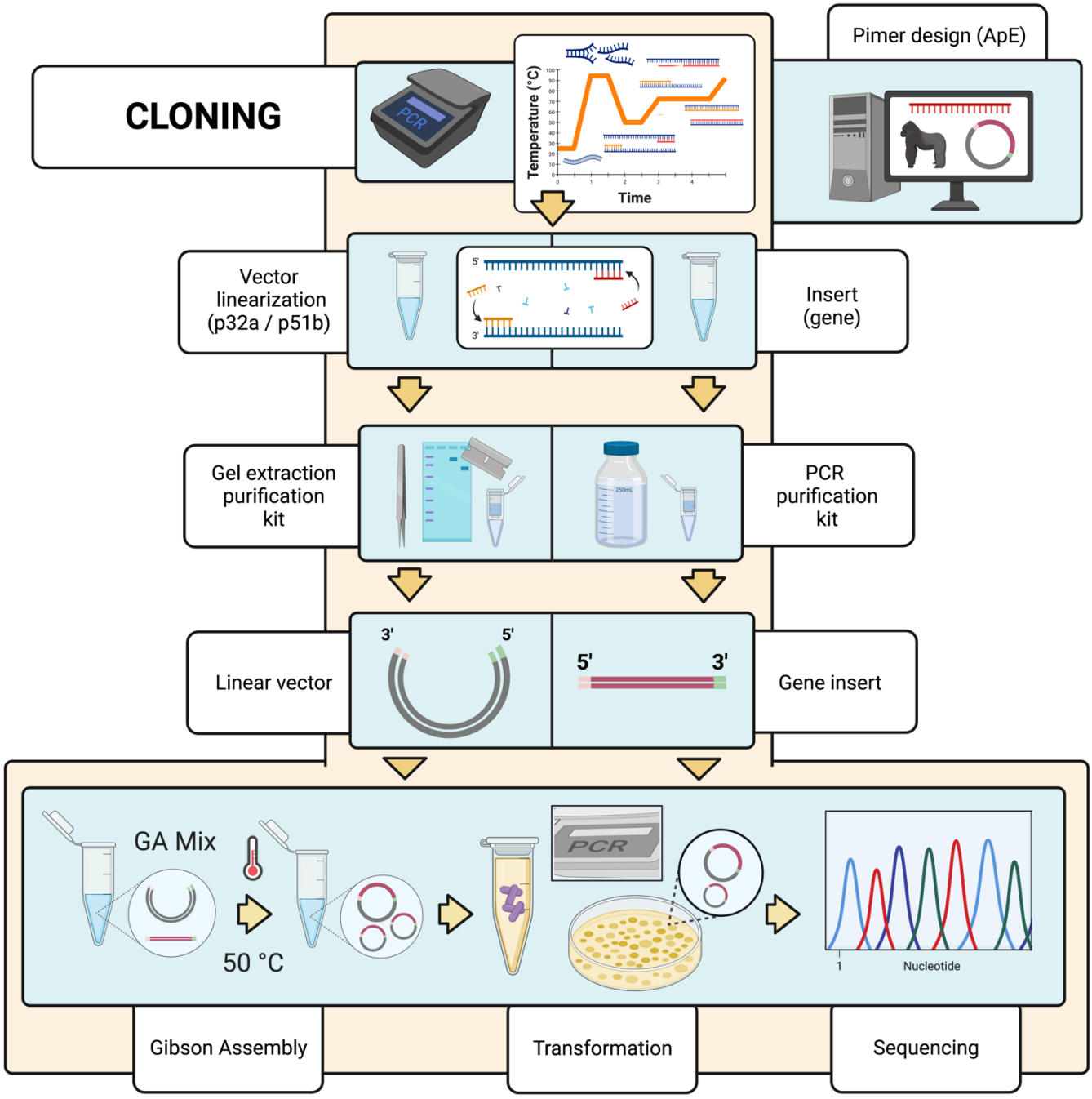


Figure 15. Overview of the cloning protocol. Design your plasmid (think about what DNA segments you will need to join) and order the primers. Amplify the linearized vector and insert each gene of interest. Purify these samples by using a kit according to product size and characteristics. Gibson Assembly is performed. An hour-long reaction produces multicopy plasmids of each gene of interest in the desired vector. The sample is transformed into bacteria, analyzed through colony PCR, and Sanger sequenced. Created with Biorender.com

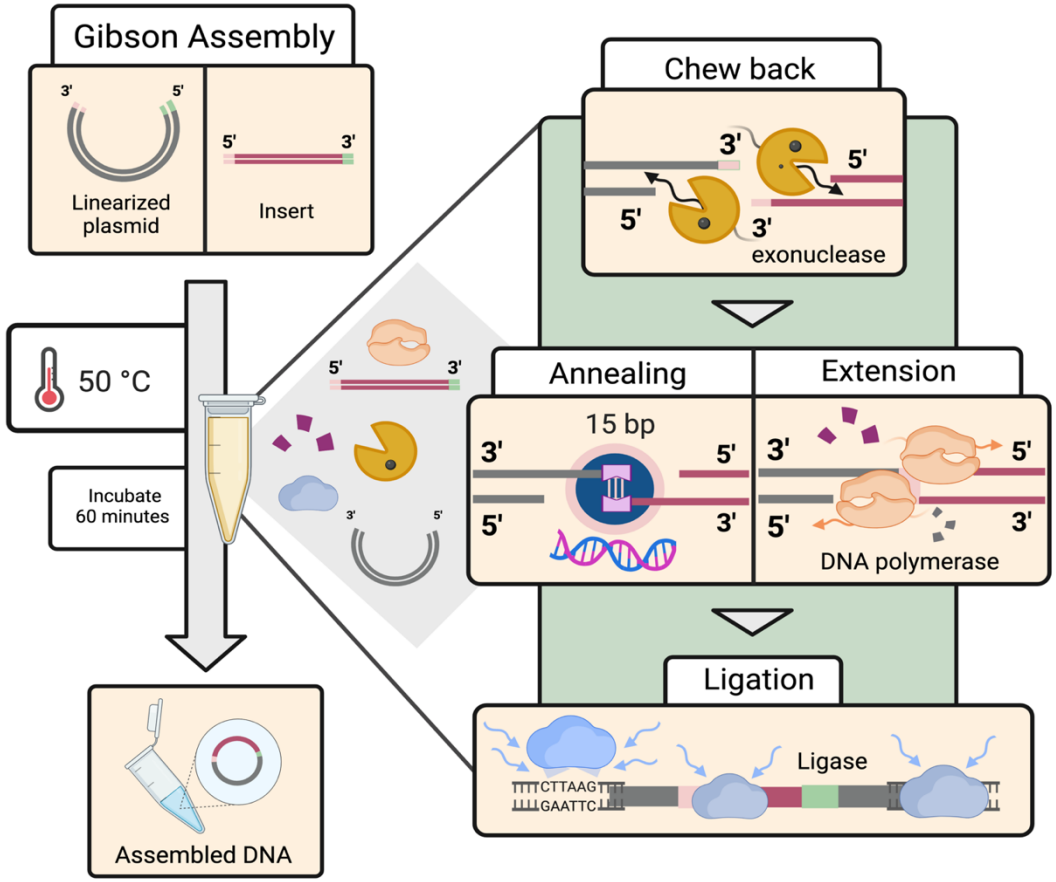


Figure 16. Gibson Assembly overview**.** In a Gibson Assembly reaction in equimolar concentrations, insert and linearized plasmid are combined. In the GA reaction, during the 1 h at 50°C, the T5 Exonuclease, an enzyme that creates a single-strand DNA in the 3' by chewing from the DNA 5' end (chew back); allows for DNA fragments to anneal by using the 15 bp overhang (annealing). Phusion DNA Polymerase incorporates nucleotides to "fill in" the gaps in the annealed DNA fragments (extension). Taq DNA Ligase joins the annealed complementary DNA fragments (ligation). Once the DNA is assembled, it is transformed into a competent cell. Created with Biorender.com

2.2.2 PROTEIN EXPRESSION

For the expression analysis experiments, BL21(DE3) competent *E. coli* cells transformed with p32a and p51b vectors harboring OX1F, OX2F, OXEF, OX1D, OX2D, and OXED fusion gene, were grown overnight in 5 mL of Terrific Broth (TB) supplemented with ampicillin (100 μg/mL) and chloramphenicol (50 μg/mL). An additional sample, known as thioredoxin (THX) was also overexpressed as an experimental control. In all these experiments, uninduced cultures were used as a control. The secondary cultures were prepared with the 5 mL inoculum in a 500 mL TB also supplemented with ampicillin (100 μg/mL) and chloramphenicol (50 μg/mL). Optimal bacterial concentration was when OD600 reached ~0.6 and then it was induced with 0.4 mM of Isopropyl β-D-1- thiogalactopyranoside (IPTG) for optimal time and temperature of 16 h at 23 °C, continuously shaking at 150 rpm.

After overnight induction cells were pelleted by centrifugation at 10,000 rpm for 8 x minutes, and frozen overnight at -80 °C. Overview of the induction protocol is shown in **Figure 17**. Subsequently, cells were harvested in prechilled lysis buffer (20 mM Tris-HCl, 2 M NaCl, 30 mM imidazole, 10% glycerol, protease inhibitor cocktail EDTA free; pH 8.0) and lysed by ultrasonication (on ice) with a Sonicator Q500 from Terra Universal, Inc., at 40% Amplitude, in four rounds of 30 secs ON/OFF, and prepped for purification, by centrifuging at 15,000 g for 40 min at 4 °C. Overview of the sonication protocol is shown in **Figure 18**. Protein timepoint samples were then resolved and analyzed using sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting.

To purify each of the recombinant OxyR fusion proteins, an IMAC was carried out under native conditions. The OxyR fusion proteins were induced with the identified optimal expression parameters in 500 mL culture volumes. Cells were then harvested, lysed, and centrifuged to obtain supernatant/soluble/insoluble cell fractions. Further, the soluble cell fraction is diluted with an

equal volume of lysis buffer and loaded onto the purification column (charged with nickel-nitrilotriacetic acid) followed by incubation with continuous shaking for 1 h at 4 °C. After the binding of each OxyR fusion protein to the nickel, the unbound/bacterial proteins were discharged, and the column was washed sequentially with 60 mL of wash buffer (20 mM Tris-HCl, 2 M NaCl, 30 mM Imidazole, 10% glycerol; pH 8.0). The bound OxyR fusion protein was then eluted with an elution buffer 20 mM Tris-HCl, 2 M NaCl, 500 mM Imidazole, 10% glycerol; pH 8.0). The samples were the dialyzed against a storage buffer (50 mm Tris-HCl, pH 8.0, 150 mM NaCl and 10% glycerol) overnight and concentrated using Amicon Ultra-0.5 Centrifugal Filter Unit. Ni-NTA chromatography overview at **Figure 19**.

For each SDSPAGE and Western Blot, overview in **Figure 20**, 5 µL of protein were mixed with 2 µL of Beta-mercaptoethanol loading dye; the samples were heated at 95 °C for 5 minutes and then loaded into a 12% polyacrylamide gel; 15 mins at 80 V, 1 h at 150 V. For the western blot, a polyvinylidene fluoride (PVDF) western blot membrane was submerged in methanol for 60 seconds, and 2 paper stacks were soaked in the transfer buffer (50 ml of 5X transfer buffer, 50 ml of ethanol and 150 ml of purified water) for 60 seconds. A transfer sandwich (paper stack, membrane, gel, and paper stack) was prepared and placed in the Trans-blot turbo transfer system from Bio-Rad Laboratories, Inc. (Hercules, CA). Specification for the run 1 x 1.5mm gel = 1.3A, 25V, 15min. After transference, the membrane was blocked with 0.25 g of milk resuspended in 10 ml of 1X TBST buffer (450 ml of purified water, 50 ml of 10X TBS, 500 µL of 20% Tween; pH 7.4) for 1 hr. The membrane is incubated overnight at 4°C with a desired antibody; [anti-His (1:10000) and anti-S (1:10000)] were used in the Western blotting analysis. Visibility of the results Chemiluminescent properties, and imagery possible by the Azure Sapphire Biomolecular Imager, Azure Biosystems (Dublin, CA).

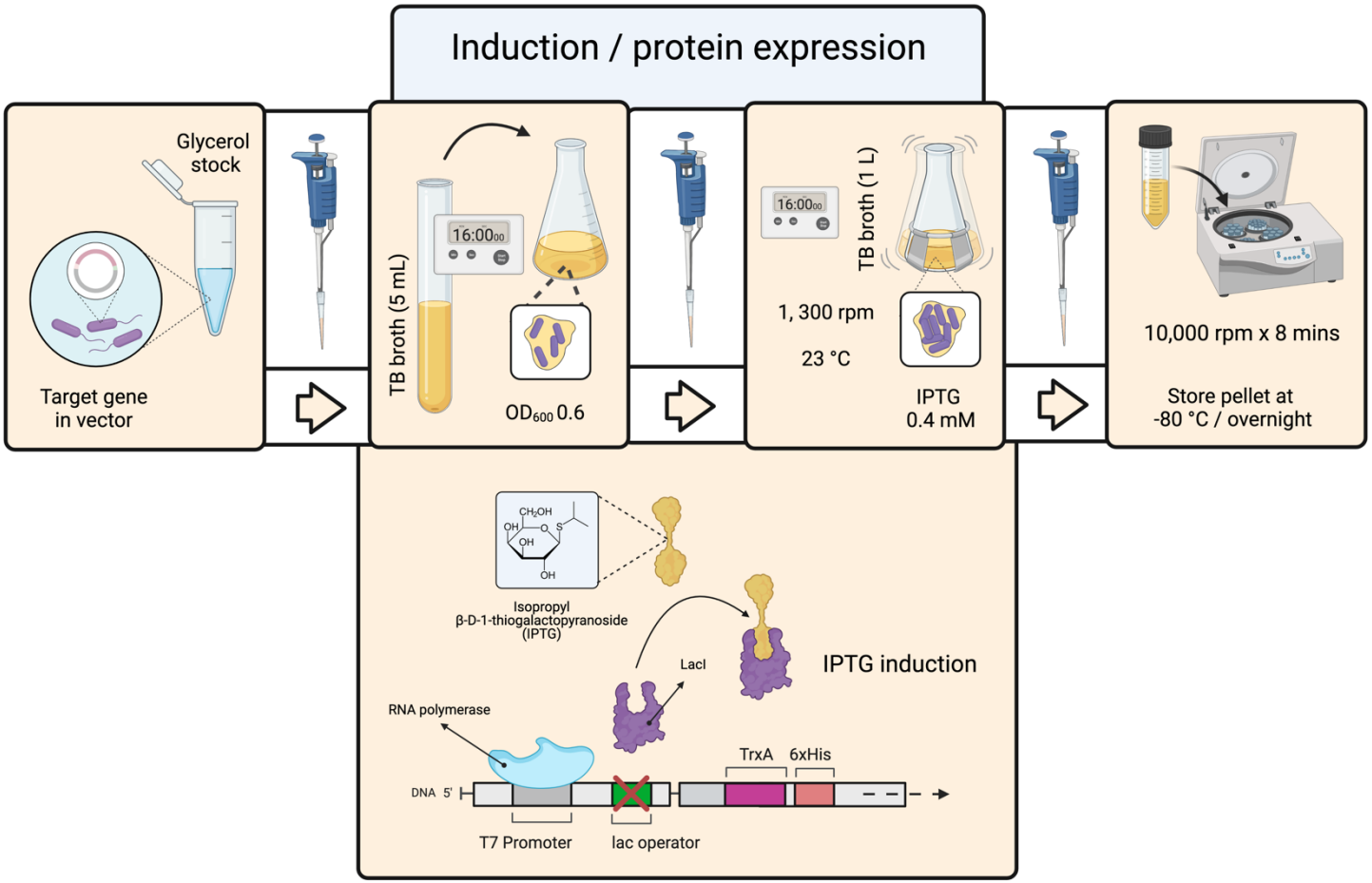


Figure 17. Induction and protein expression overview. IPTG is a compound that molecularly mimic of allolactose, a lactose metabolite that triggers transcription of the lac operon, and it is therefore used to induce *E. coli* protein expression where the gene is under the control of the lac operator. IPTG binds to the lac repressor and releases the tetrameric repressor from the lac operator in an allosteric manner, thereby allowing the transcription of genes in the lac operon. Glycerol stocks are cultured in 5mL TB test tube and grown overnight (16 h). Cells are re-cultured in 500 mL of TB until OD600 reaches 0.6. Samples are induced with 0.4 mM of IPTG and kept growing overnight at a temperature of 23 °C. Samples are centrifuged at 10, 000 rpm for 8 mins and stored at -80 °C overnight. Created with Biorender.com

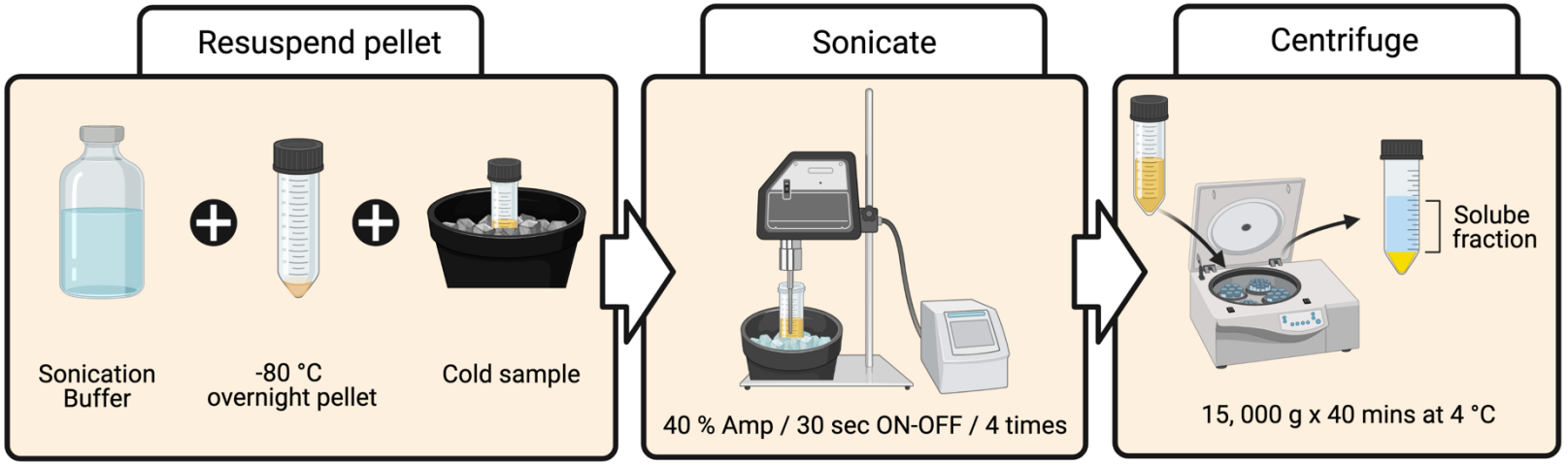


Figure 18. Sonication overview. Sonication is a cell lysis method commonly used, it is the exposure of cells to frequencies of sound that can disrupt their membrane. Initially, cells are resuspended in a sonication or lysis buffer; crucial to keep samples cold throughout the sonication process. Use an ice bath so that the heat created by pulsing doesn't compromise sample. Put the sonication probe inside the tube with the chilled resuspended cells. Frequency is adjusted depending on how difficult it is to lyse the cell and the sensitivity of the contents you want to purify. The final step is centrifuging the cell debris that accumulated in solution during sonication. Centrifugation is done at 15, 000 g at 4 °C degrees for approximately 30 mins.

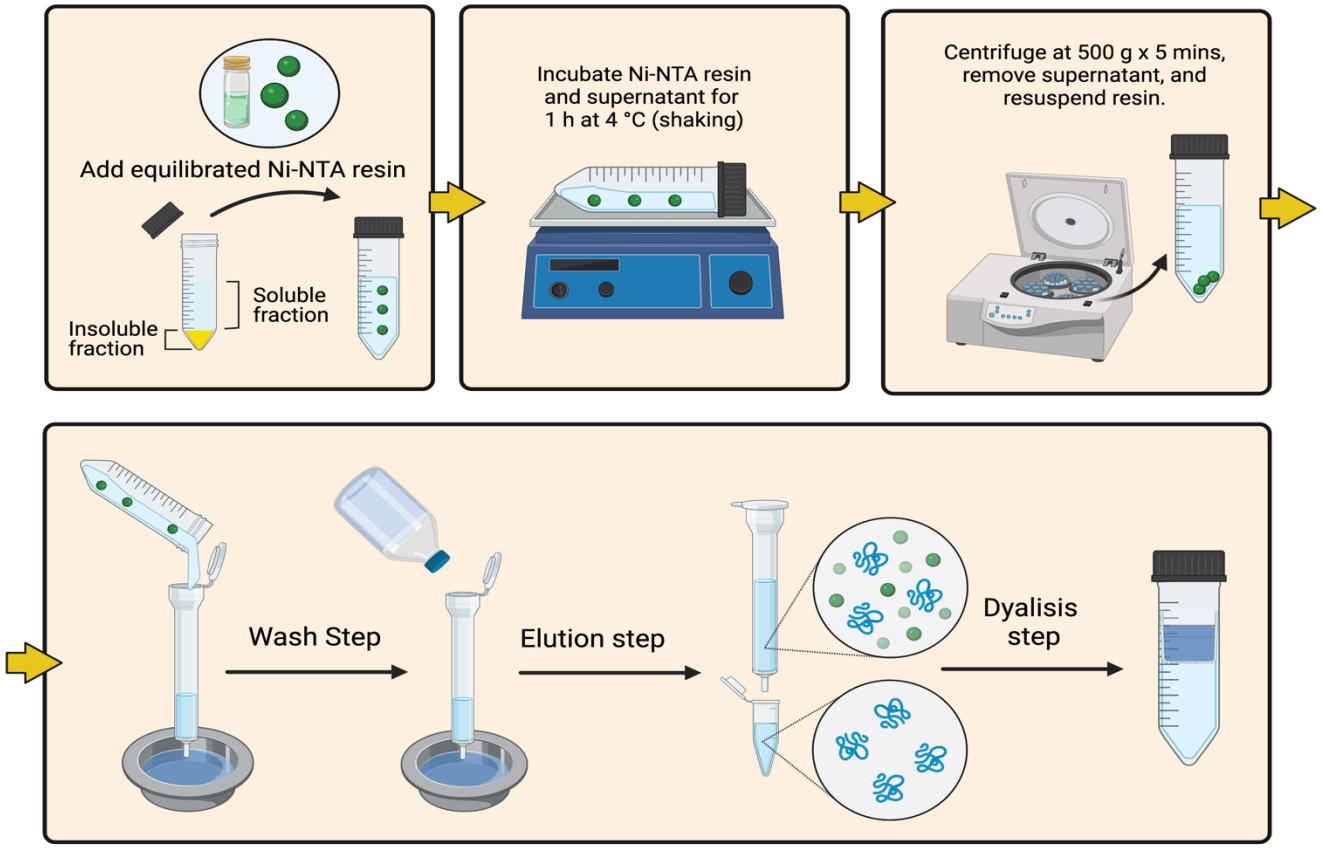


Figure 19. Ni-NTA purification overview. His-tag purification requires the equilibrated Ni-NTA resin to be equilibrated and resuspended with the soluble fraction of the purified target protein. A tailored equilibration buffer is required to properly equilibrate the resin. The samples are then mixed with the equilibrated resin and shaken for mixture for an hour in cold. Samples are then centrifuged and ready for column purification. Mixture of the resin and samples are washed with column buffer or wash buffer and eluted with an elution buffer containing imidazole. Finally, samples are dialyzed to transfer the purified protein to a storage buffer without imidazole or other contaminants. Created with Biorender.com

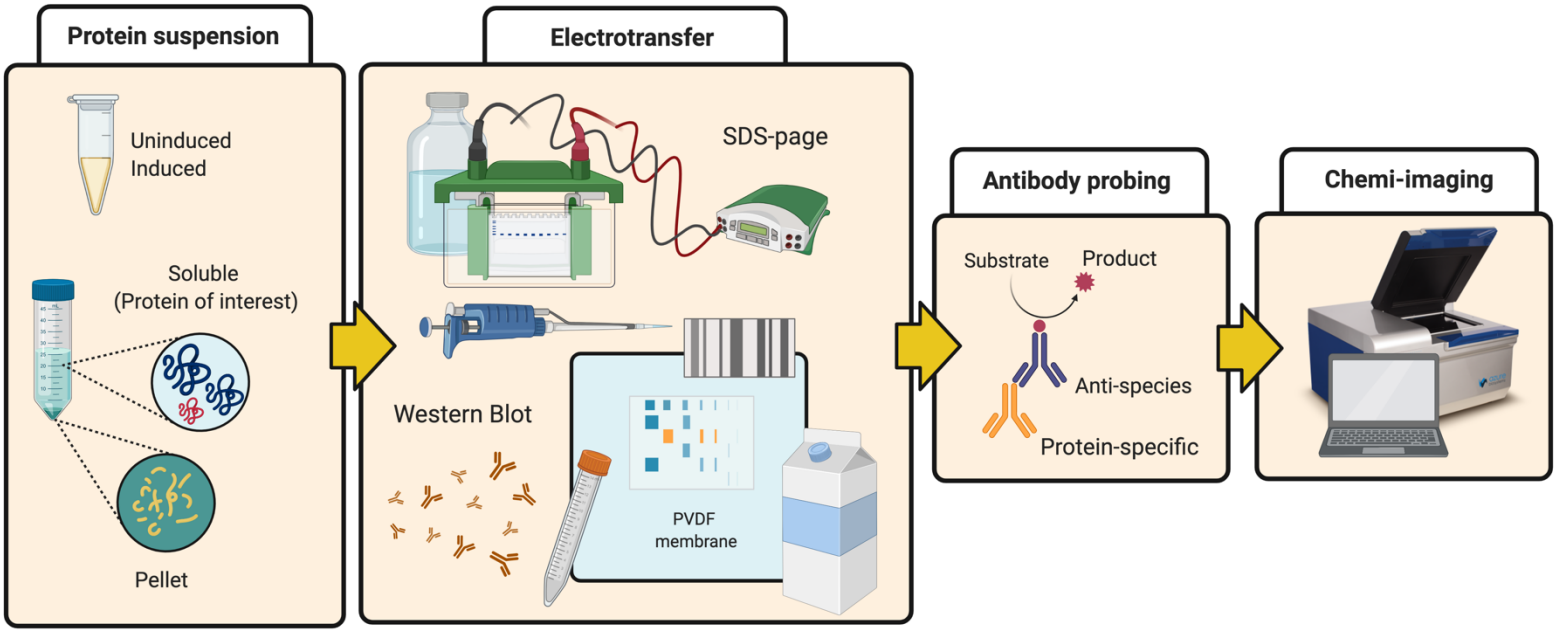


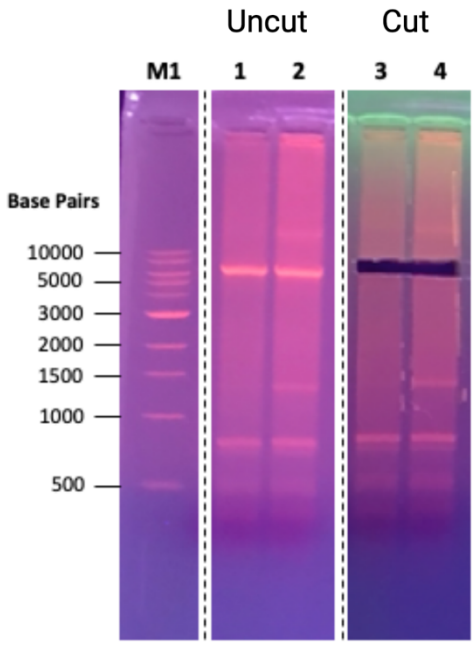
Figure 20. Overview of the SDS-PAGE and Western Blot techniques. Created with Biorender.com

2.3 RESULTS

2.3.1 CLONING

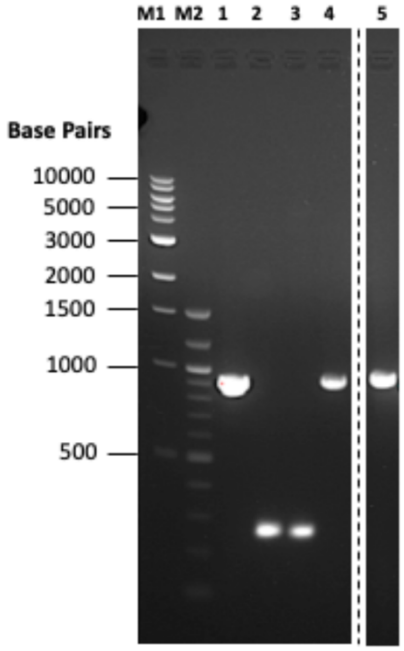
The validated oxyr fusion plasmids for p32a were coupled with a set of tags (His) at the start of the sequence and before the stop codon. The reason to have fusion tags at either end is that its position can influence expression level, solubility, and stability of human proteins expressed in *E. coli* (Haridhasapavalan et al., 2021). The customized, synthetic oxyR inserts were then cloned under the control of a tightly regulated strong inducible T7 promoter of p32a expression vector. Results from PCR amplification of linearized vector and inserts are in **Figures 21 - 23**. Gibson assembly cloning technique was used to clone the inserts into the p32a and p51b plasmids, to generate minipreps. A Full length (FL) and DNA Binding Domain (DBD) version of each oxyR was constructed for both plasmids, results in **Figures 24, 25**.

After a successful Gibson assembly protocol, each gene was inserted in p32a and p51b plasmids, and the reaction product was transformed in *E. coli* DH-5 alpha. Subsequently, after transformation and overnight incubation was carried, a colony PCR was conducted to screen and confirm the appropriate construct of the inserts in each of the plasmids. DNA concentrations were assessed with a Nanodrop, results in **Table 4**. Samples, successfully identified in their appropriate theoretical weights using an agarose gel electrophoresis, were transformed BL21 (DE3) for protein overexpression. Plasmids were verified by Sanger sequencing using T7 forward and T7 reverse primers. NCBI-BLAST tool alignment was used for the construct of every OxyR protein generated.



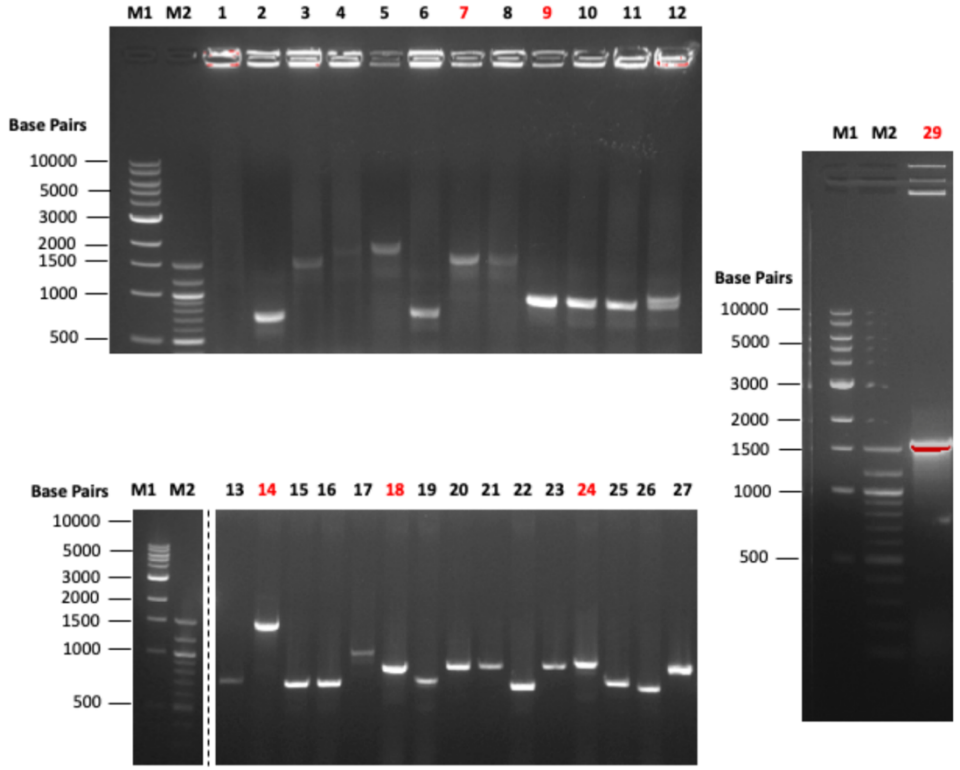
L	Sample	Expected length
1	pET32a(+)_linearized	5839
2	pET32a(+)_linearized	5839
3	pET32a(+)_linearized	5839
4	pET32a(+)_linearized	5839
M1	1 kb ladder	

Figure 22. Linearized plasmid PCR product of the p32a plasmid. The gel contains information on the linearized plasmid of p32a before and after incision from agarose gel 1%. The expected theoretical length of p32a linearized is 5,839 bp.



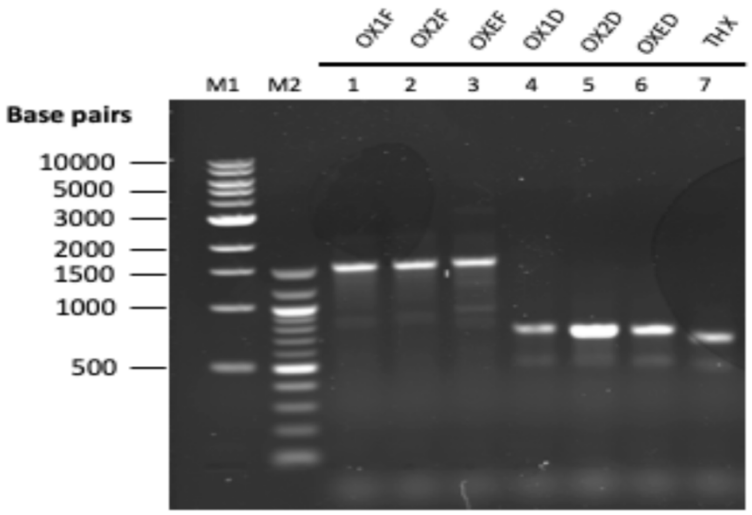
L	pET-32a(+) inserts	Expected length
1	OX1F	927
2	OX1D	218
3	OX2D	218
4	OX2F	927
5	OXEF	945
M1	1 kb ladder	
M2	100 bp ladder	
	TAATACGACTCACTATAGGG T7 promoter, forward primer	
	GCTAGTTATTGCTCAGCGG T7 terminator, reverse primer	

Figure 21. Insert PCR products of the oxyR genes to be inserted in p32a through GA cloning.



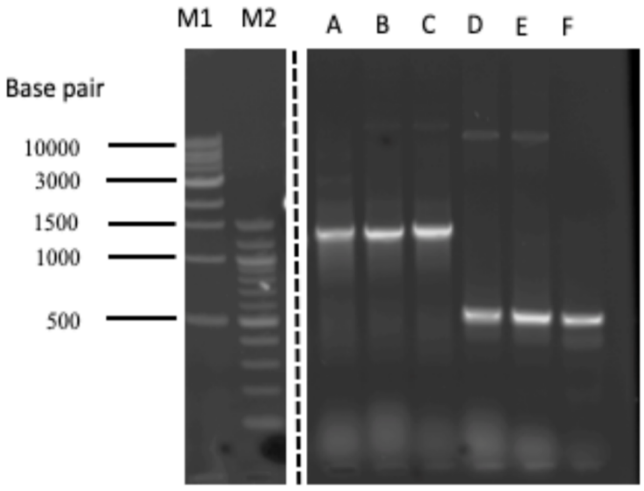
L	pET-32a(+) inserts	Expected length
1 - 8	OXEF	927
9 - 12	OXED	218
13 -17	OX1F	218
18 -22	OX2D	927
23 -27	OX1D	945
29	OX2F	246
M1	1 kb ladder	
M2	100 bp ladder	
TAATACGACTCACTATAGGG T7 promoter, forward primer		
GCTAGTTATTGCTCAGCGG T7 terminator, reverse primer		
*Selected samples after colony screening		

Figure 23. Colony screening after GA. Several colonies were selected for each of the successfully transformed oxyR genes to identify appropriate cloning.



L	pET-32a(+) Miniprep	Expected length
1	OX1F	1548
2	OX2F	1548
3	OXEF	1566
4	OX1D	878
5	OX2D	878
6	OXED	867
7	THX	712
M1	1 kb ladder	
M2	100 bp ladder	
	TAATACGACTCACTATAGGG T7 promoter, forward primer	
	GCTAGTTATTGCTCAGCGG T7 terminator, reverse primer	

Figure 24. p32a miniprep analysis.



Samples	Expected Mass (bp)	Gene	Plasmid
A	1162	OX1F	pET51b(+)
B		OX2F	
C		OXEF	
D	480	OX1D	
E		OX2D	
F		OXED	
M1	1 kb ladder		
M2	100 bp ladder		

Figure 25. p51b miniprep analysis with T7 promoter/ terminator.

Table 4. p32a and p51b oxyR miniprep concentrations

Primers	Concentration [ng/ µL]
pET51b(+)_OX1F	129.7
pET51b(+)_OX2F	126.5
pET51b(+)_OXEF	139.1
pET51b(+)_OX1D	132.8
pET51b(+)_OX2D	112.3
pET51b(+)_OXED	112.4
pET32a(+)_OX1Ft	132.7
pET32a(+_OX2F	138.9
pET32a(+)_OXEF	124.5
pET32a(+)_OX1D	140
pET32a(+)_OX2D	101.4
pET32a(+)_THX	154.3

2.3.2 PROTEIN EXPRESSION AND PURIFICATION

We next sought to identify the optimal expression conditions such as inducer concentration, pre-induction cell density, induction temperature, and postinduction incubation time to obtain soluble heterologous expression of recombinant oxyR fusion proteins in *E. coli* strain BL21(DE3). Since numerous studies have demonstrated that identifying optimal expression parameters is key to achieve soluble recombinant proteins, this aspect was emphasized. The concentration of inducer is the first and foremost vital parameter for the overexpression of recombinant proteins. Therefore, to identify the optimal IPTG concentration, BL21(DE3) competent *E. coli* cells transformed with p32a and p51b for each oxyR were induced with various concentrations of IPTG for 2 h. Maximum expression was observed when cells were induced with 0.4 mM IPTG.

Protein expression timepoints of induction were taken to corroborate the appropriate synthesis of the OxyR protein; timepoints: uninduced (before the addition of IPTG), induced (after the addition of IPTG and incubation), supernatant (soluble proteins), and pellet (insoluble proteins). Molecular weight was assessed through SDS-PAGE for each of the oxyR fusion protein timepoints, **Figures 26 - 29**. The samples are in appropriate length as interpreted from the SDS-PAGE and the theoretical values established for the fusion proteins. The dentification of optimal conditions facilitated expression of recombinant OxyR proteins in soluble form, we next purified this fusion protein using the most commonly used protein purification technique, IMAC. Importantly, purification under native conditions often resulted in generating biologically active protein molecules with native-like secondary structure conformation (Wingfield, 2015). Results of the Ni-NTA purification in **Figures 30 - 36**. Further, the fusion proteins produced by p32a and p51b were detected by Western blotting using an anti-His antibody and anti-S antibody, **Figures 37, 38**. A band corresponding to human OxyR fusion proteins was observed; the calculated molecular weight of the fusion protein are present in the images.

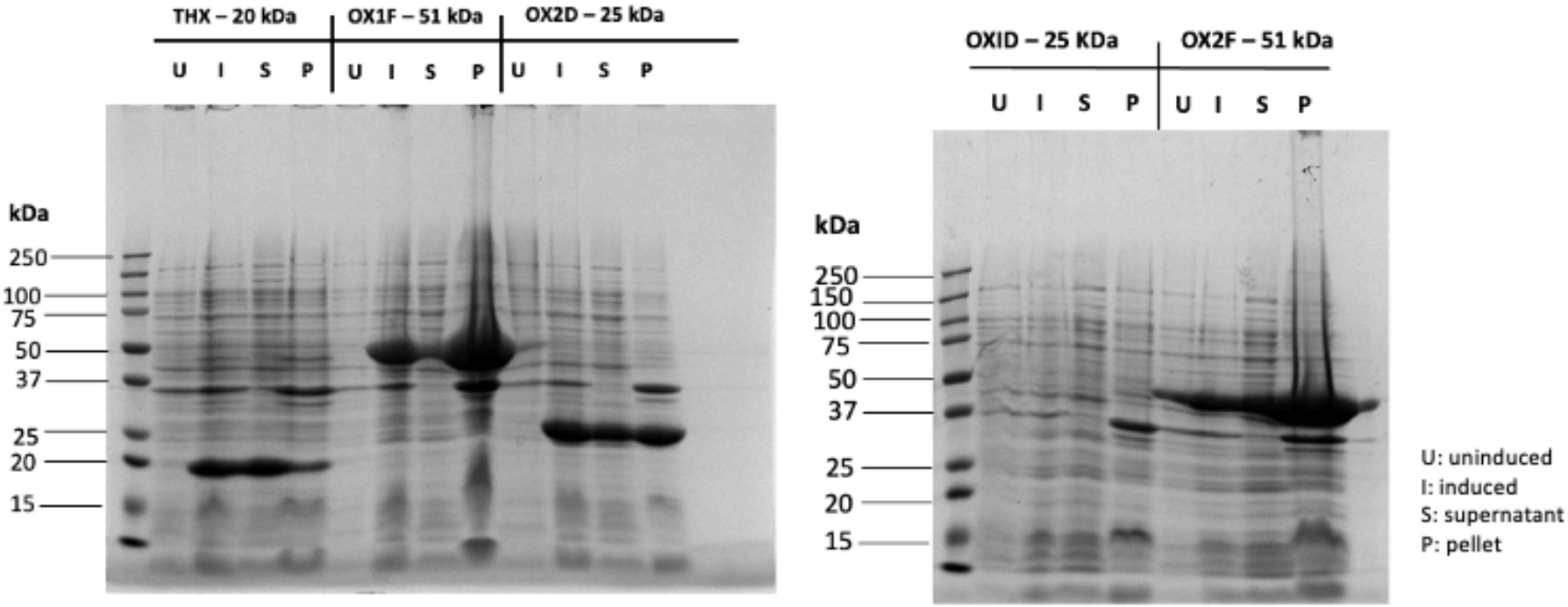


Figure 26. Induction timepoints: THX, OX1F, OX2D, OX1D, OX2F.

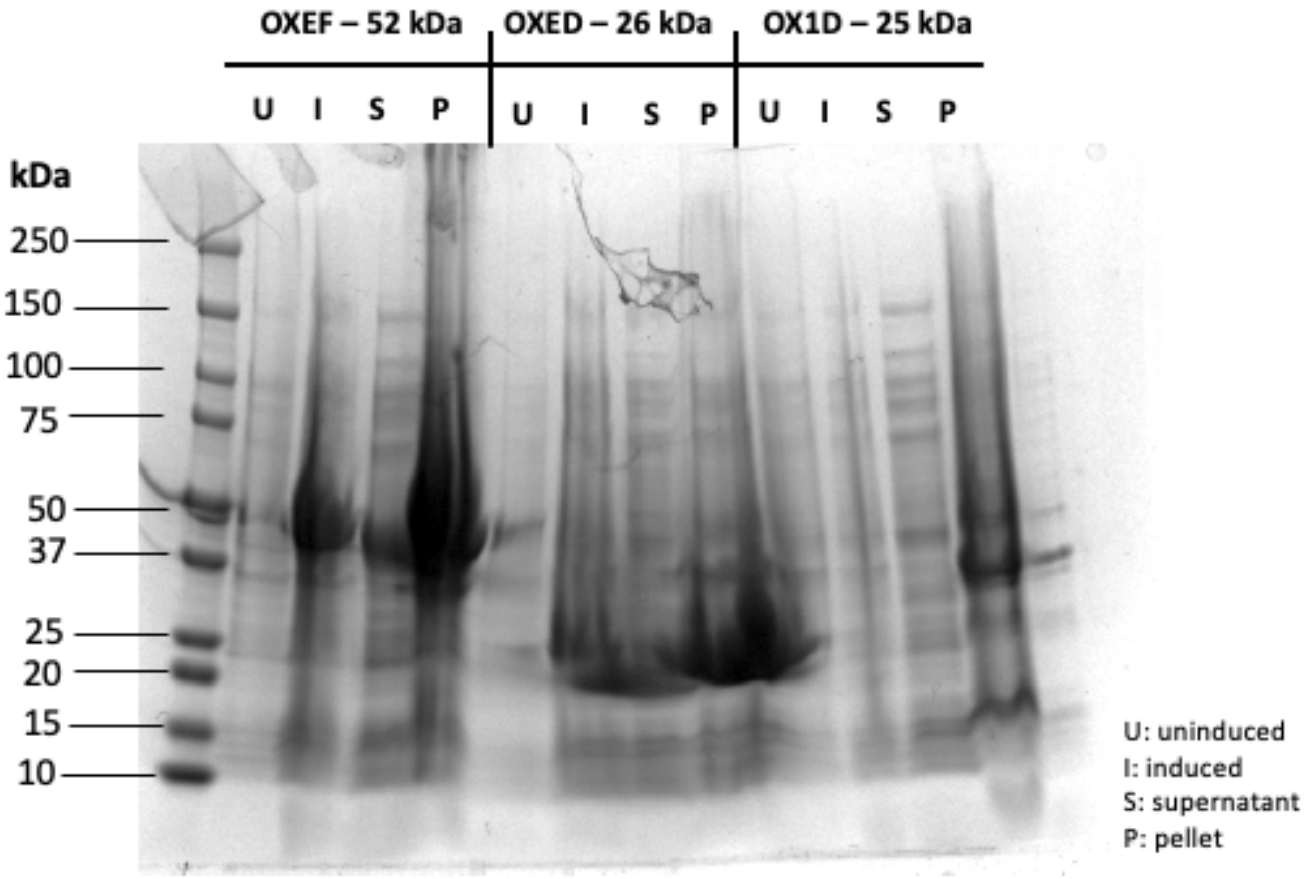


Figure 27. Induction timepoints: OXEF, OXED, OX1D.

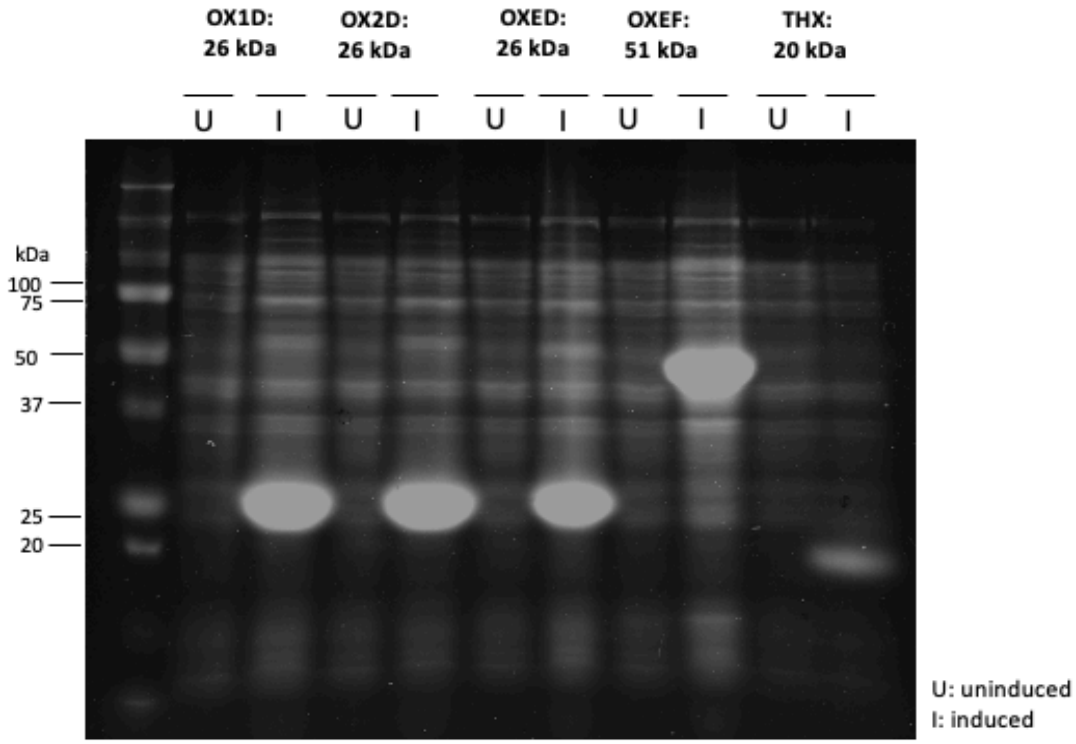


Figure 28. Induction timepoints: OX1D, OX2D, OXED, OXEF, THX.

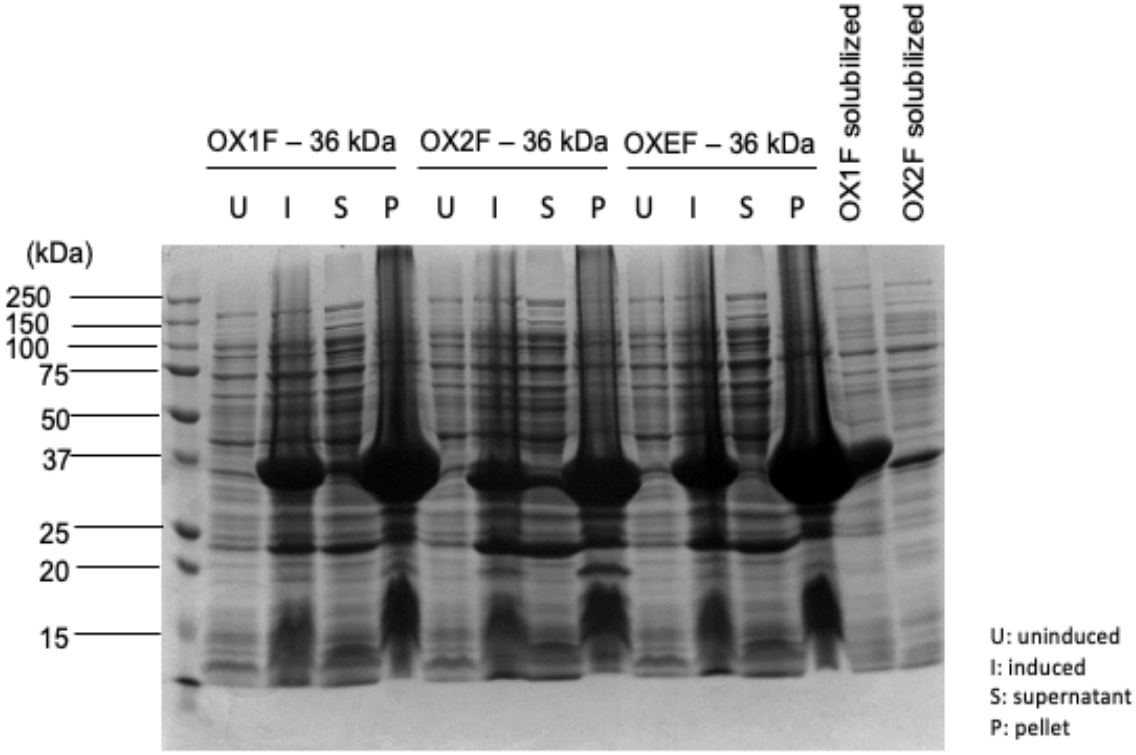


Figure 29. Induction timepoints: OX1F, OX2F, and OXEF.

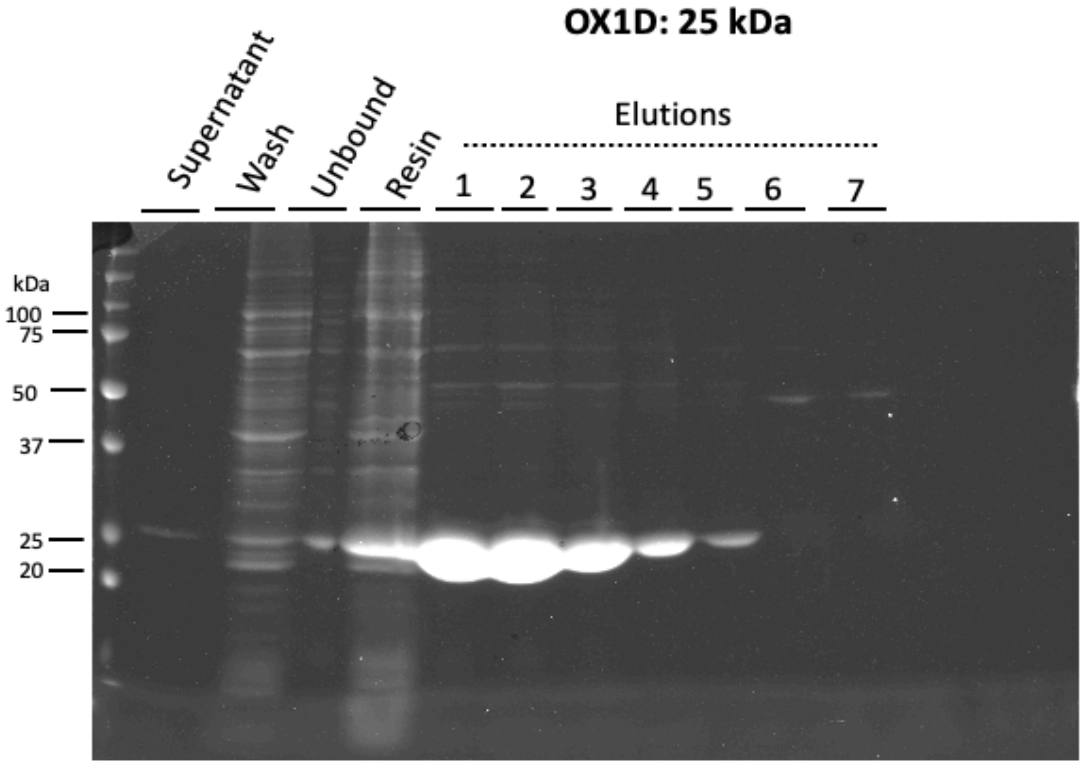


Figure 30. Purification timepoints: OX1D.

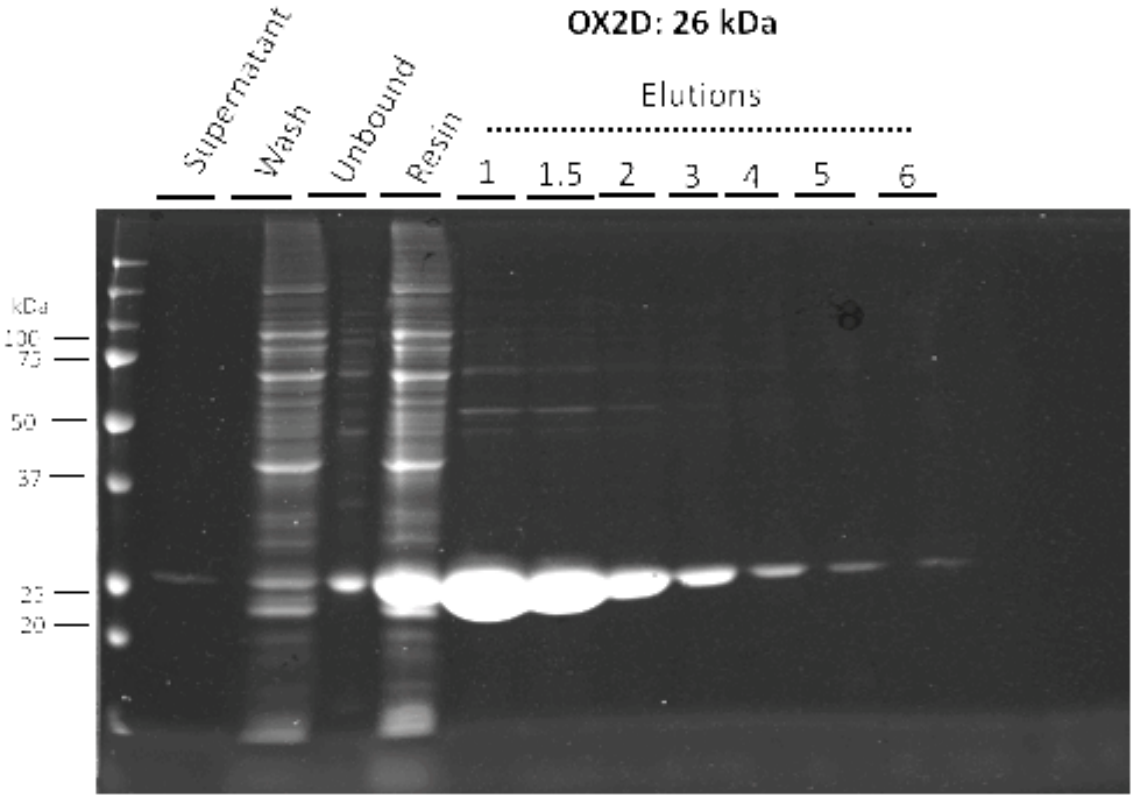


Figure 31. Purification timepoints: OX2D.

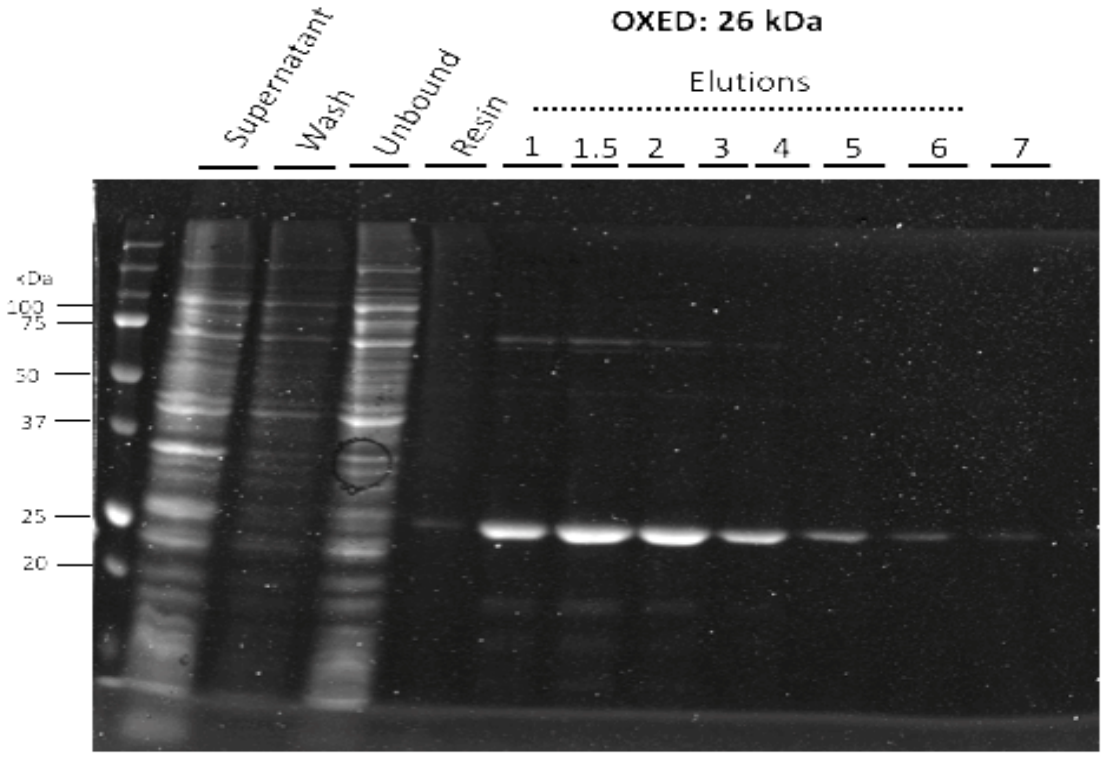


Figure 32. Purification timepoints: OXED.

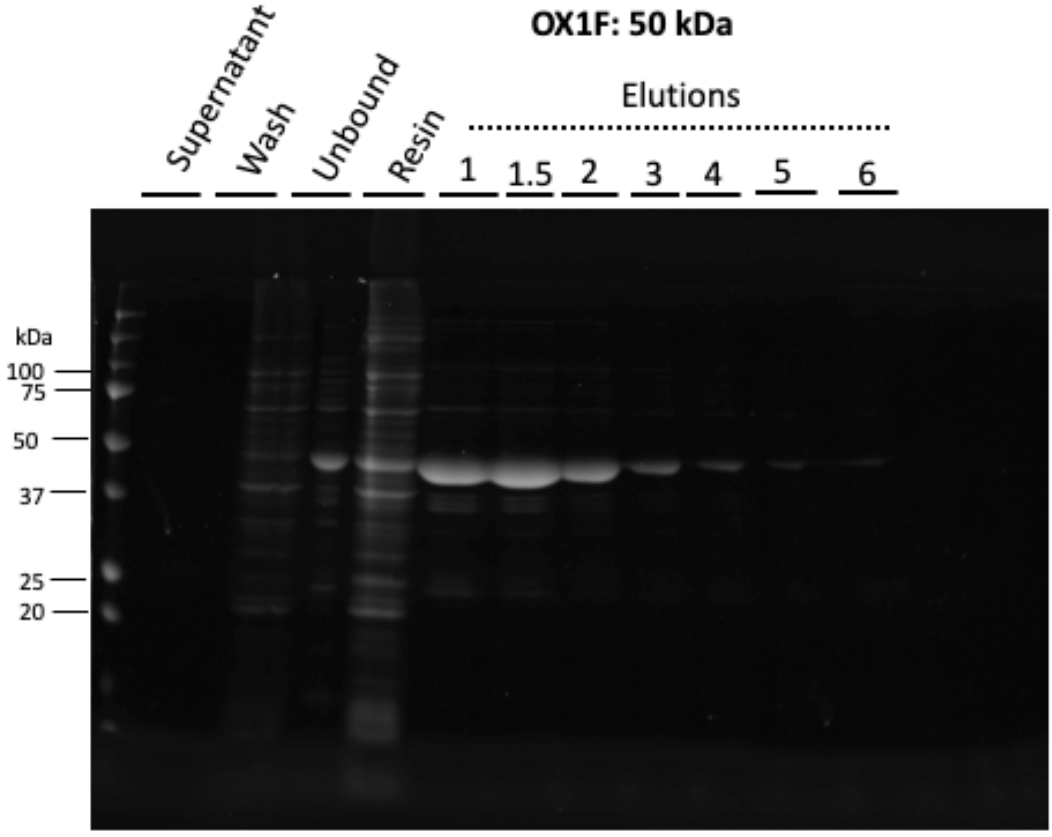


Figure 33. Purification timepoints: OX1F.

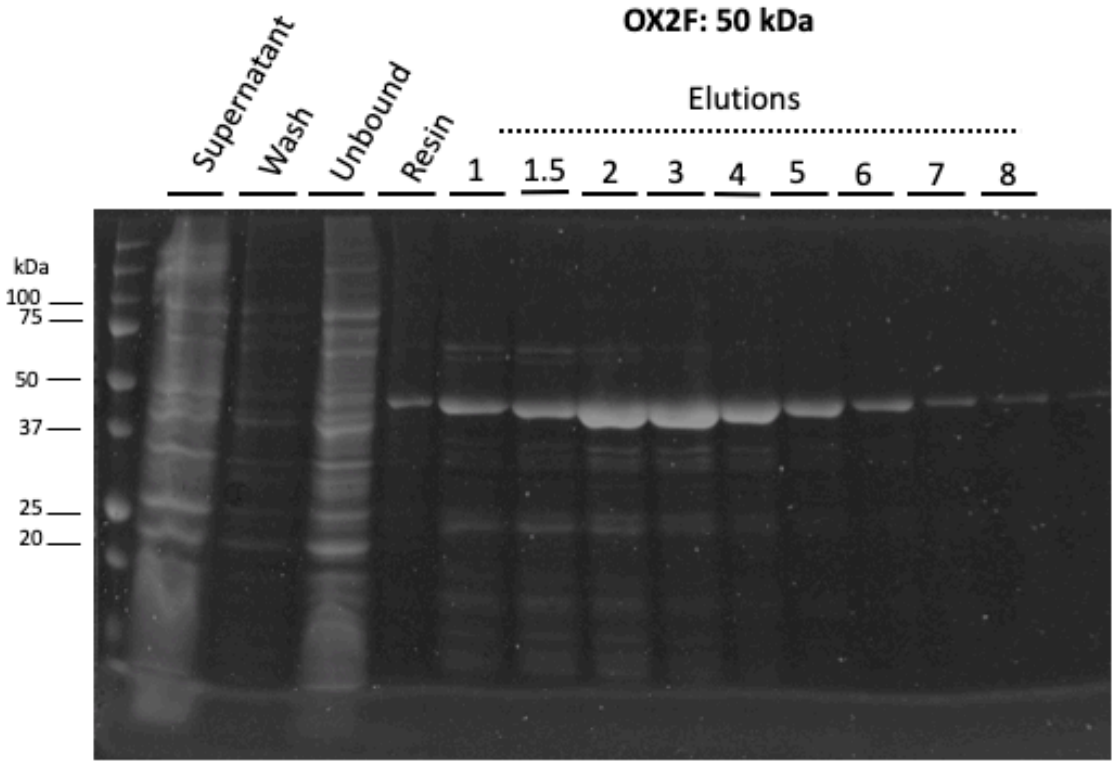


Figure 34. Purification timepoints: OX2F.

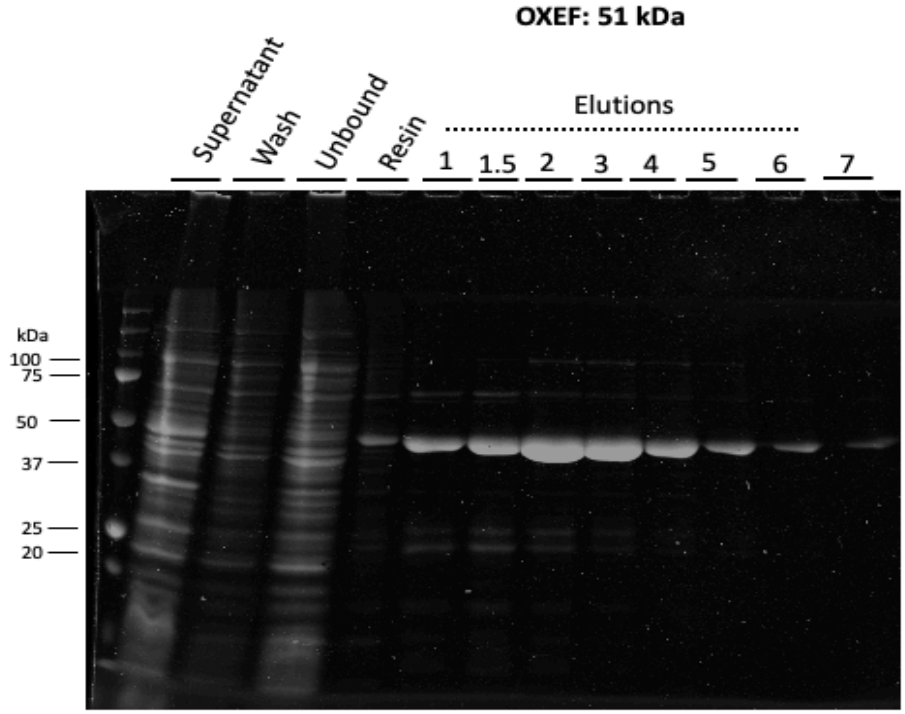


Figure 35. Purification timepoints: OXEF.

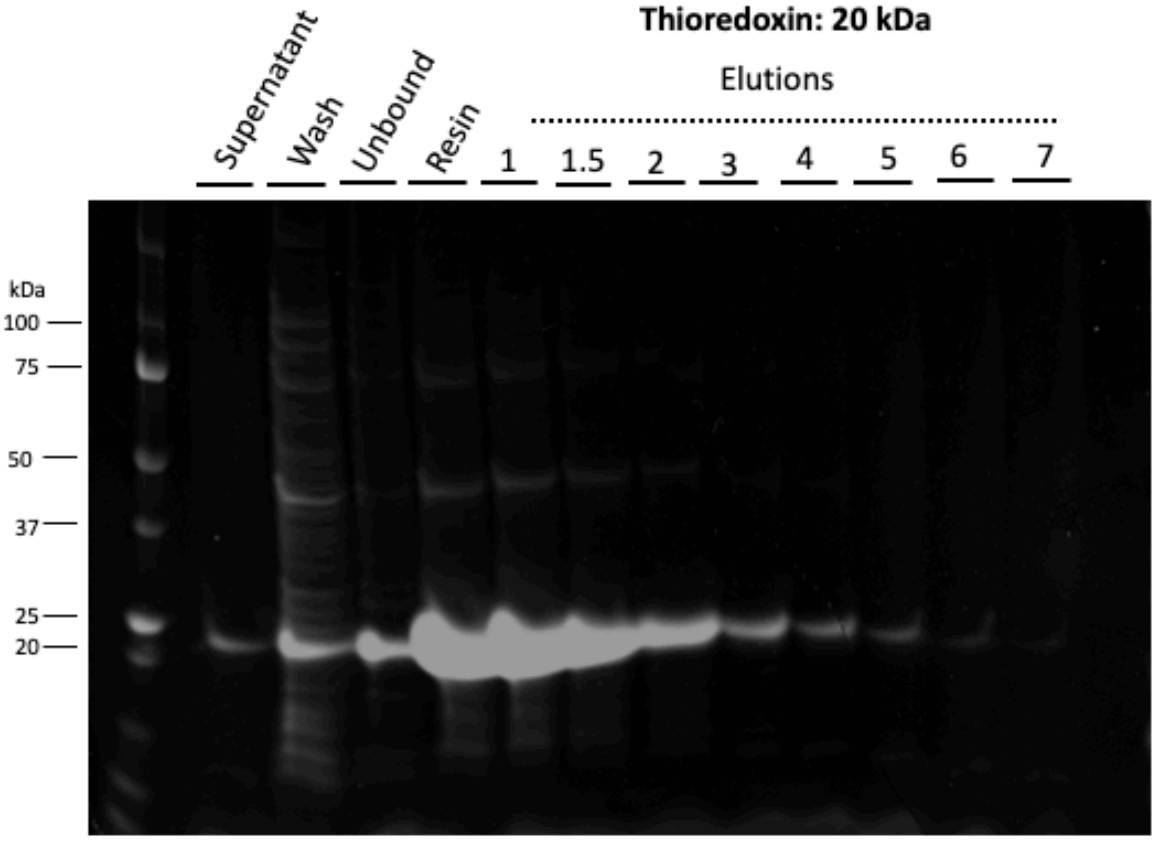


Figure 36. Purification timepoints: THX.

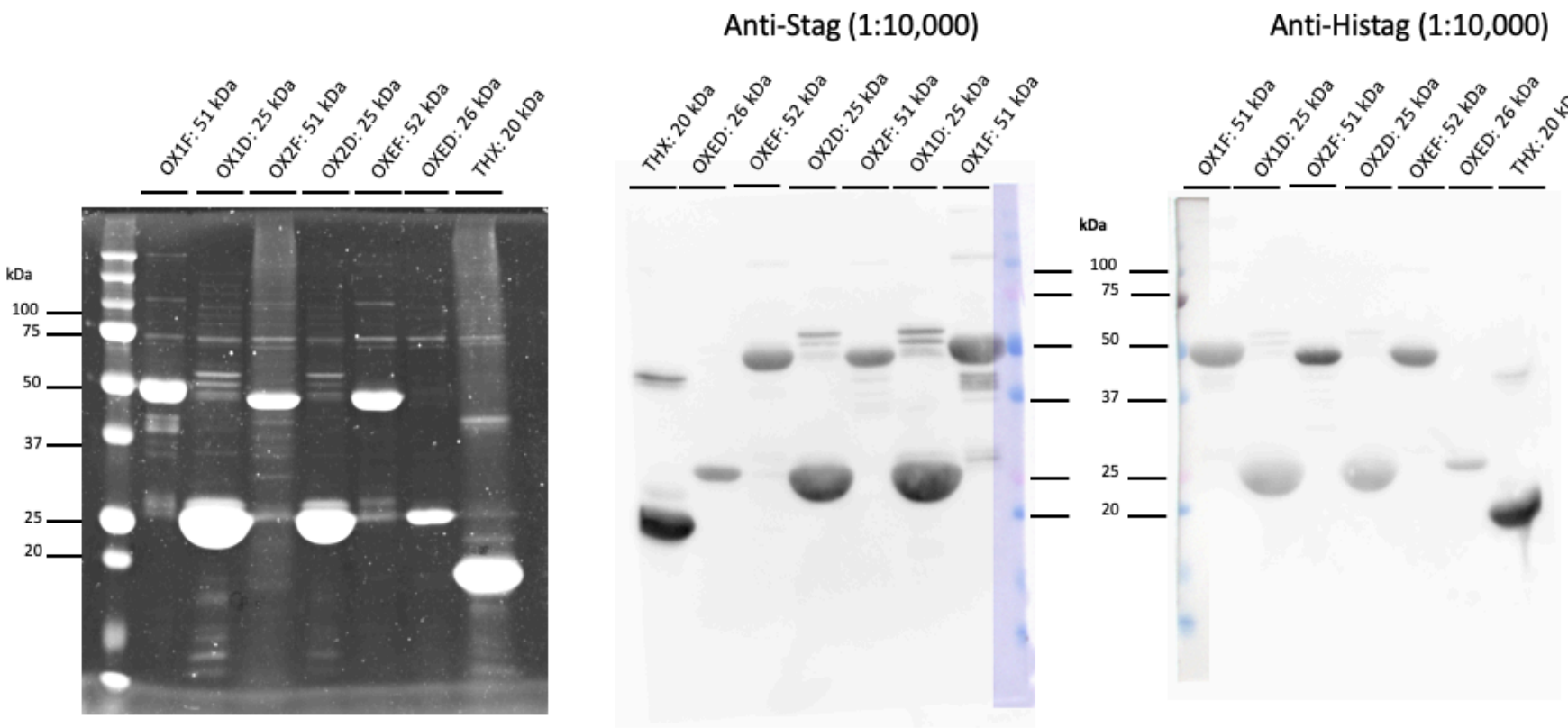
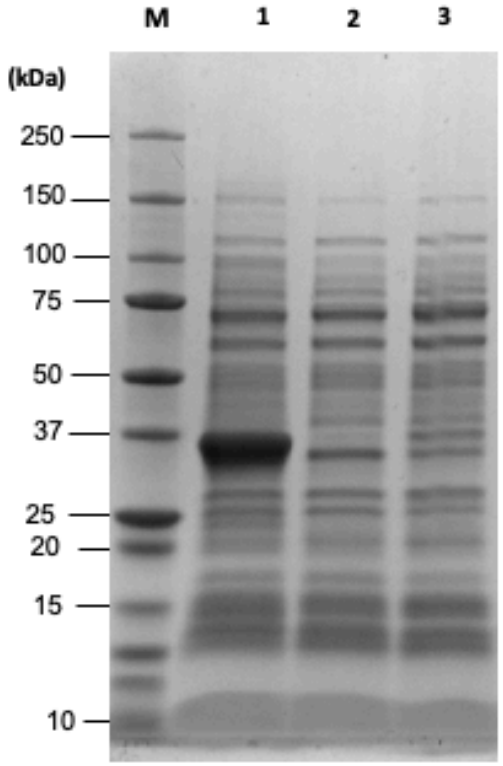
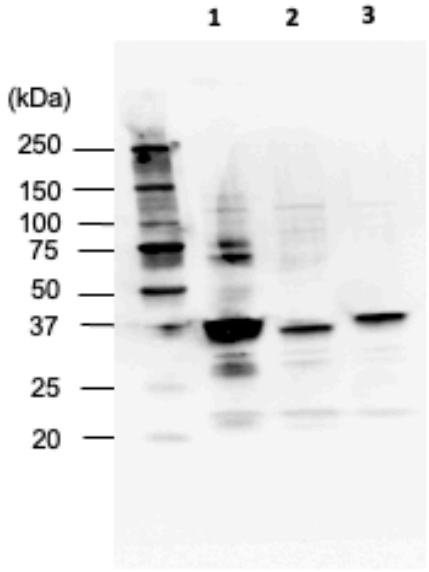


Figure 37. SDS-PAGE and Western blot of Ni-NTA purified fusion proteins: OX1F, OX1D, OX2F, OX2D, OXEF, OXED, THX; derived from p32a. Anti S-tag (1:10000) and Anti His-tag (1:10000) were used for the Western blot.





	pET-51b(+)	Expected size (kDa)
1	OX1F	36
2	OX2F	
3	OXEF	37
M	Protein ladder	

Figure 38. SDS-PAGE and Western blot of Ni-NTA purified fusion proteins: OX1F, OX2F, OXEF; derived from p51b. Anti His-tag (1:10000) was used for the Western blot.

2.4 DISCUSSION

This study reports the production of a homogeneous recombinant OxyR1, OxyR2, and OxyRE fusion protein in their FL and DBD versions., which is one of the critical transcription factors in regulating the defense against oxidative stress. Notably, based on its defensive role, several studies included OxyR in the activation of antioxidative genes. Generating a recombinant version of this crucial transcription factor is key to circumvent the limitations associated with DNA binding and oxidative state of the protein. Cloning protocol functional plasmids, (1) A total of 14 plasmids were generated for both p32a and p51, (2) The average length of each of the constructs is appropriate to the theoretical value calculated from the Orf at ApE, (3) Overexpression of the protein was successful, theoretical value calculated from the Orf once translated into kDa matches the obtained results. In the current study, we selected *E. coli* as an expression host due to its fast growth rate, high transformation efficiency, well-understood genetics, and cost-effective production of recombinant proteins in large quantities. In this study we specifically used the BL21(DE3) strain of *E. coli* due to its various advantages: (1) compatible with p32a and p51b vectors, (2) engineered to produce T7 RNA polymerase, (3) allows high-level of stable expression of the protein of interest.

Fusion tag optimization considers various features of DNA sequence into account that can influence protein expression levels: the secondary structure, stability, cis-regulatory elements, among others. BL21(DE3) competent *E. coli* transformed with p32a vector harboring oxyR1, oxyR2, and oxyRE fusion gene was induced under optimal expression conditions (temperature, 23 °C; cell density, OD600 = ~0.6; IPTG concentration, 0.04 mM; induction time, 16 h), and the expressed proteins were purified using affinity chromatography Based on these studies, as well as our *in vitro* analysis, we have optimized an induction strategy for the OxyR1, OxyR2, and OxyRE fusion proteins to achieve its successful heterologous expression in *E. coli*. The purification

samples were run on SDS-PAGE gels and then either stained using Coomassie Brilliant Blue or performed Western blotting with anti-His antibody and anti-S antibody. Purification of proteins successful, although is not optimal due to the use of only one step of purification through Ni-NTA. Despite the impurity in the sample there is enough purified target protein to perform downstream processes such as the Electrophoretic Mobility Shift Assay (protein's functionality) and SELEX-seq (protein's relative affinity).

In summary, we demonstrated the generation of pure, full-length and DNA binding domain recombinant OxyR1, OxyR2, and OxyRE transcription factors. Screening and identifying the optimal expression conditions were determined. Fusion tags like the thioredoxin (THX) tag play a vital role in obtaining the soluble expression of the recombinant fusion proteins. OxyR fusion proteins were expressed under optimum conditions: OD_{600} = 0.6 at 23°C and then induced by 0.4 mM IPTG for 16 h. His-tagged proteins were purified using Ni-NTA resin and detected by SDS-PAGE and Western blot. Out of all the samples the purest ones were obtained from the p32a plasmids. Given the multiplicity of bacterial TFs regulators that have been discovered in the past years, it is believed that defining and determining a TFs overall specific binding landscape provides a useful step towards understanding the overall role of these proteins in the regulation of transcription.

3 CHAPTER 3

EMSA FUNCTIONALITY ASSESSMENT OF *ALIIVIBRIO FISCHERI* OXYR1 AND OXYR2 TFS AND DETERMINING THEIR INTRINSIC BINDING SITES THROUGH SELEX-SEQ

3.1 INTRODUCTION

The interaction of transcription factors with specific DNA sites is a significant step in controlling many cellular processes, including DNA replication, recombination and repair, transcription, and viral assembly (Savinkova et al., 2021). An EMSA is a rapid and sensitive method to detect protein-nucleic acid interactions. Mobility-shift assays are often used for qualitative purposes, although under appropriate conditions, they can provide quantitative data to determine binding stoichiometries, affinities, and kinetics (Hellman & Fried, 2007). In an EMSA the observation typically is the electrophoretic mobility of a DNA-protein complex that is less than that of the free nucleic acid. During electrophoresis, the protein-DNA complexes are quickly resolved from free DNA, providing a 'snapshot' of the original sample's equilibrium between bound and free DNA. In this manner, the EMSA can provide information about the DNA-protein complex formation and detect the presence or the absence of binding.

Notably, no single set of binding and electrophoresis conditions works well for all molecular systems. However, several variables like the design of the nucleic acid target, binding reaction conditions, and electrophoresis conditions can be optimized to study a particular interaction. For example, the small number of non-specific protein binding sites in a small DNA can be advantageous when the binding protein has low sequence-specificity. In addition, electrophoretic resolution of complexes from free nucleic acid is highest with small nucleic acids, making possible short electrophoresis times (Hellman & Fried, 2007). One way to identify a DNA-protein complex is using fluorescent probes that can detect a shift of the DNA with a fluorescent imaging system. Protein-nucleic acid interactions are sensitive to mono- and divalent salt concentrations and pH (Hellman & Fried, 2007). Therefore, buffers that approximate physiological salt concentrations and pH and provide needed co-factors at appropriate concentrations are used for EMSA

experiments. The EMSA technique has been reviewed from theoretical and experimental perspectives by studying the binding of purified prokaryotic repressor/activator proteins and those studying eukaryotic nuclear transcription factors (Hellman & Fried, 2007). Apart from identifying DNA-protein complexes and characterizing these proteins, the EMSA can also be useful to distinguish the complexes formed according to the equilibrium constants of each binding reaction, enabling to compare relative binding affinities and discriminate the interaction between a given protein with more than one nucleic acid sequence (Ferraz et al., 2021).

The SELEX-seq technology is a combinatorial chemistry technique for producing oligonucleotides of either single-stranded DNA or RNA that binds to a target ligand (Slattery et al., 2011); diagram of the technique is in **Figure 39**. SELEX-seq differs from traditional SELEX in two ways: (1) The number of selected (bound) DNA oligos characterized and (2) The number of rounds of selection performed (Riley, et.al., 2014). Unlike traditional SELEX, where the order of 102 selected DNA oligos is identified at the very end of the reiterative selection process, SELEX-seq leverages the depth of next generation sequencing to characterize 107 or more selected DNA molecules at each round of selection (Riley, et.al., 2014). The technique is widely applied as an *in vitro* selection method to isolate high-affinity DNAs bound by a transcription factor from an extensive library with random sequences. Therefore, SELEX-seq is ideal for exploring the DNA binding preferences of single proteins or multi-protein complexes.

SELEX-seq allows identifying relevant binding sites after only one to two rounds of selection (Riley, et.al., 2014)). Using a biophysical model of the SELEX-seq procedure, relative affinities for selected sequences are obtained by comparing the sequence composition of later rounds to that of the unselected DNA library. The success of a SELEX-seq experiment is primarily driven by two components: the DNA-binding protein(s) of interest and the randomized DNA

SELEX library. Beyond protein preparation, multiple variables must be considered when generating the randomized SELEX library (Riley, et.al., 2014). The randomized region must be large enough to encompass the expected core DNA binding motif and also allow for capturing information regarding sequences immediately flanking the core motif.

We have already established that DNA-protein interactions are essential for molecular and cellular mechanisms, such as transcription, transcriptional regulation, and DNA modifications. Considering the characterization of the DNA-binding protein, EMSA is an easy *in vitro* method that rapidly identifies DNA-protein binding interactions. For this chapter, we studied the overall binding interactions of the different purified OxyR and assessed their overall affinity using the SELEX-seq technique.

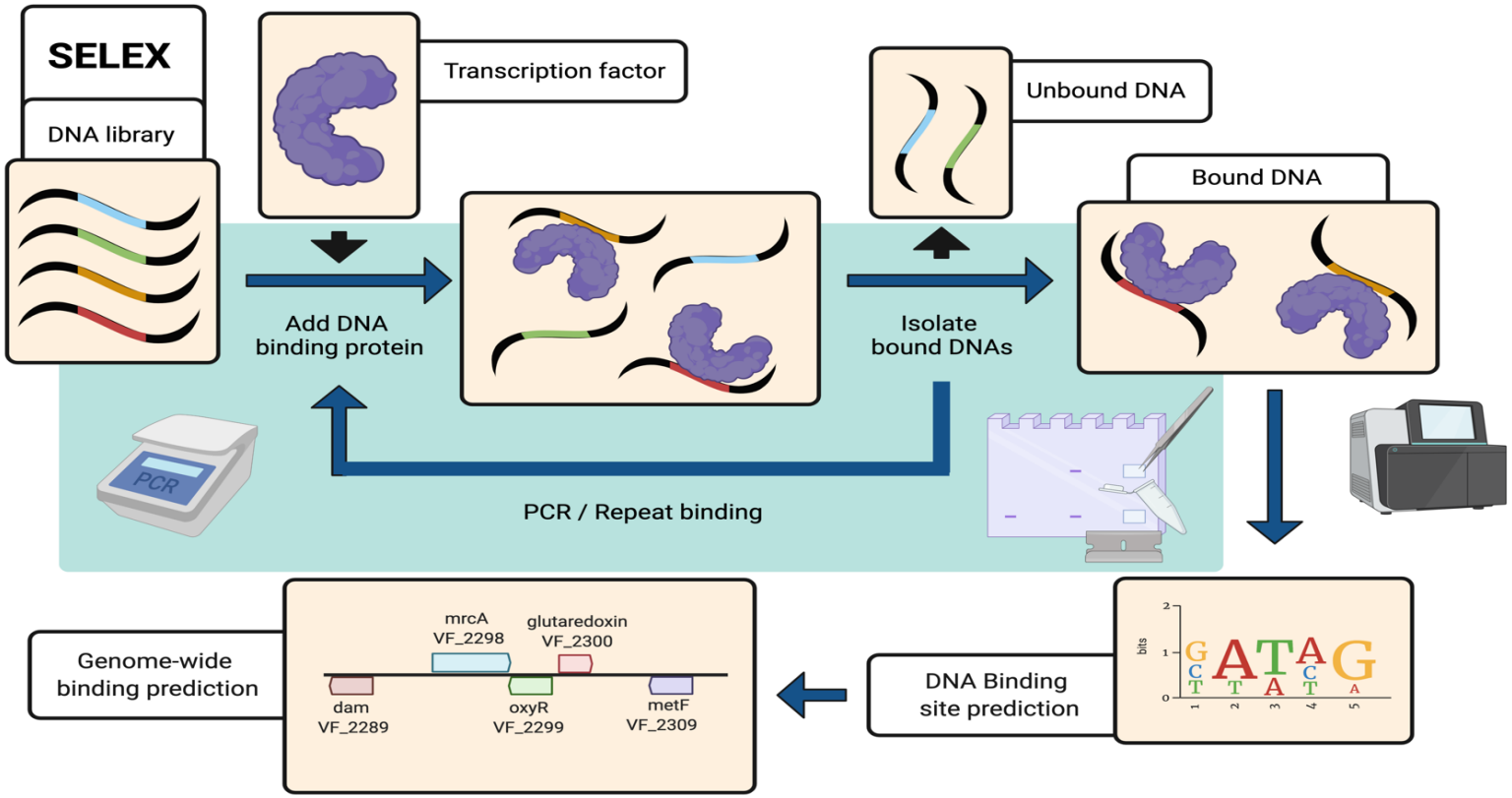


Figure 39. Overview of the SELEX-seq method. They start with a library generated with millions of sequences (20N). The binding reaction occurs when the selected transcription factor is incubated with the library. Once selection through affinity and specificity occurs, only the sequences bound by the TF are selected for PCR amplification, and the others are discarded. Next, the selected samples are excised from a gel by following a sample of the TF with a fluorescent probe. Finally, the samples are then processed through PCR. The SELEX cycle is repeated for three rounds (R1, R2, R3). The results lead to the attainment of only specific sequences enriched through the SELEX for each of the studied TF. These sequences are barcoded, and Ilumina is sequenced posteriorly to properly identify the selected binding sites for which the TF holds affinity. Created with Biorender.com

3.2 MATERIALS AND METHODS:

3.2.1 ELECTROPHORETIC MOBILITY SHIFT ASSAY (EMSA)

An EMSA was used to test the DNA-binding activity of the Ni-NTA purified OX1F, OX2F, OXEF, OX1D, OX2D, OXED, and THX proteins. This experiment depends on fluorescence measures from the designed DNA probe and the use of a native gel. A 6% polyacrylamide gel electrophoresis is run to assess the binding properties of the TF to the probe; **Figure 40**. The DNA probe sequence consists of 60 nucleotides, containing 20 bases of a previously characterized *E. coli* OxyR binding sites, probe called OXGWCELL (Keseler et al., 2011; Salgado et al., 2013; Seo et al., 2015). The probe is [5-CTCGCCTGGGCAGAAGTGTC**CCTCAAGTTAACTTGAGG**GACACTTCTGCCCAGGCGAG-3′] flanked by 20 constant bases on the 5' end and 20 constant bases on the 3' end. The 5' end of the probe was modified with the dye IR700 from Integrated DNA Technologies, Inc. (Coralville, IA). The double-stranded DNA probe was synthesized with a primer-extension reaction containing 2 µL of the probe, 6 µL of 50 µM IR700, 25 µL of EconoTaq Master Mix 2X from Lucigen Corp (Middleton, WI), and 17 µL of nuclease free water. The protocol for the thermocycler was: 95 °C for 2 min., 55 °C for 1 min., 72 °C for 5min., 10 °C on hold followed by PCR column purification.

Binding reactions (15 µL contained 10 nM of the labeled DNA probe in a 1X OxyR binding buffer (20 mM Tris pH 8.0, 50 mM NaCl, 2.5 mM $MgCl_2$, 0.45 mM EDTA, 10 % glycerol), 33.3 ng pdl-dC, 33.3 ng BSA, 0.07% Tween-20, and nuclease free water. For each separate binding reaction, a total of 5 ul of the proteins were added. The samples were incubated at room temperature for 1 h. Native 6% polyacrylamide gels were pre-ran at 60 V for 15 min in the 1x OxyR binding buffer. We loaded 15 µL of each sample at 30 V and the gel ran at 120 V for 3

hours. The gels were imaged using the Azure Sapphire Biomolecular Imager (Azure Biosystems, , CA).

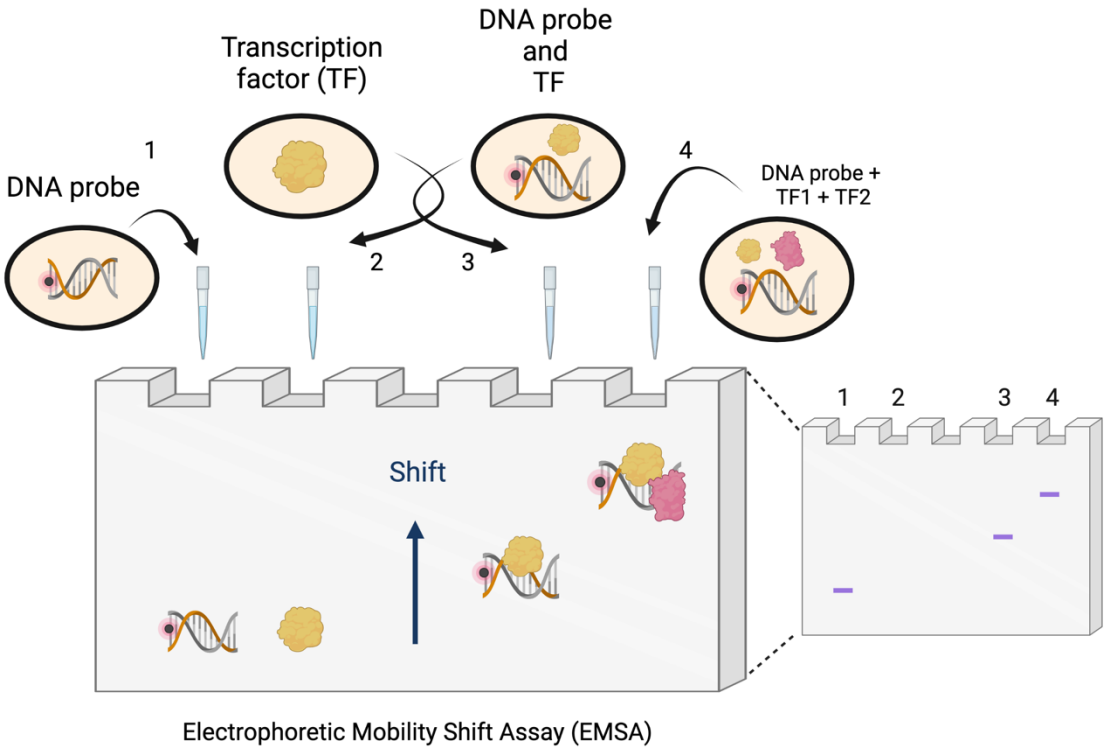


Figure 40. Overview of the Electrophoretic Mobility Shift Assay (EMSA). The illustration depicts the basic steps for an EMSA to assess the functionality of a purified transcription factor. The DNA probe usually contains a canonical region known to be bound by the TF. For example: (1) Contains the DNA probe with a region known to be bound by the TF (Positive control). (2) Contains only the TF to be studied (Negative control). (3) DNA probe and TF are incubated in a binding reaction and added to the lane to examine binding activity. (4) In the case that multiple TF interactions want to be assessed during DNA binding, the binding reaction can be prepared with both TFs and introduced as a sample. Created with Biorender.com

3.2.2 SYSTEMATIC EVOLUTION OF LIGANDS BY EXPONENTIAL ENRICHMENT (SELEX-SEQ)

The DNA-binding sites of which *A. fischeri* OxyR1, OxyR2, and *E. coli* OxyR (FL and DBD) were determined by SELEX-seq. A 200 nM biotinylated DNA library from Integrated DNA Technologies, Inc. (Coralville, IA) with a central randomized 20 bp sequence flanked by 20 constant nucleotides on each side for PCR. Binding reactions contained 25 nM of the OXGWCELL and 200 nM of the DNA library) and 5 μL of the Ni-NTA purified protein. The reaction master mixes also contained: OxyR binding buffer (20 mM Tris pH 8.0, 50 mM NaCl, 2.5 mM $MgCl_2$, 0.45 mM EDTA, 10 % glycerol), 33.3 ng pdl-dC, 33.3 ng BSA, .07% Tween-20, and nuclease-free water. Incubation of samples was at room temperature for 1 hr. Native 6% polyacrylamide gels were pre-ran at 120 V for 15 min. We loaded 20 μL of each sample at 30 V, and the gel ran at 120 V for 3 hours. Using the Azure Sapphire Imager every gel was imaged. Bands corresponding to the TF-DNA library were excised. The DNA was eluted in 500 μL of the elution buffer from Qiagen (Germantown, MD) overnight at 30°C and shaking at 1,200 rpm (Thermoshaker BIOGRANT). Bound DNA was enriched with Dynabeads M280 Streptavidin magnetic beads using Invitrogen (Carlsbad, CA) protocol. After washing the beads with the elution buffer 3 times, the pulled-down DNA was resuspended in a PCR master mix with EconoTaq and amplified for 18 cycles. The DNA was purified using the Qiagen kit, and their concentration was measured with the Nanodrop and was used for subsequent SELEX rounds. After the three selection rounds were performed, unique 6-bp barcodes and Illumina sequencing adapters were incorporated into the DNA sequences of all the rounds by PCR, including the original library. Samples were sequenced using Illumina technology at Novogene (California, USA). SELEX-seq overview in **Figure 41**.

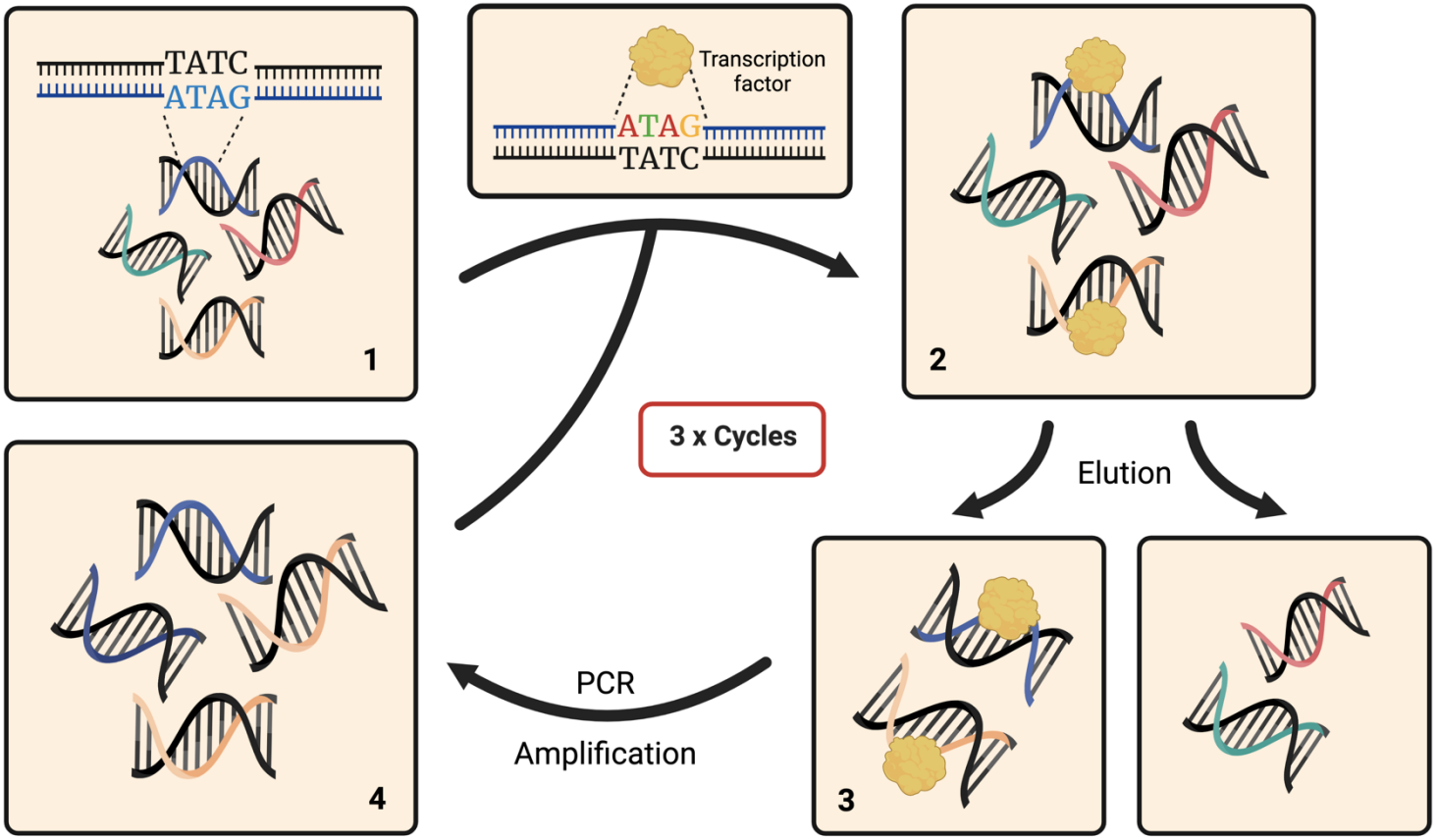


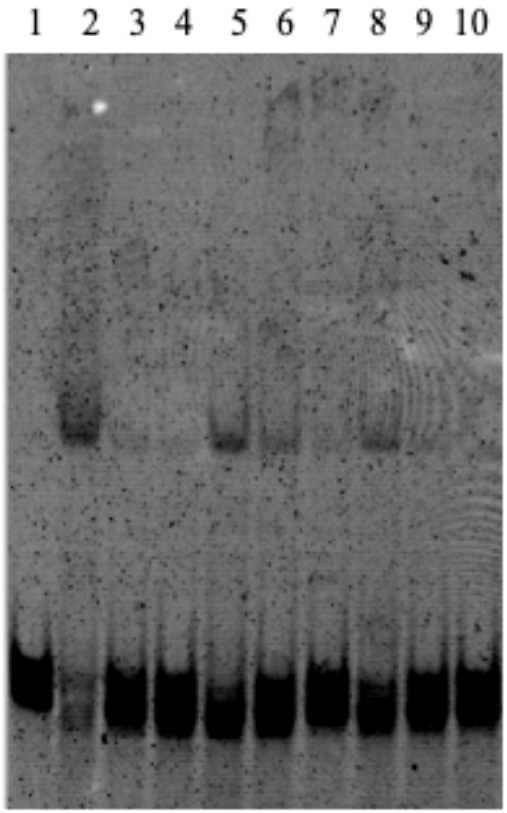
Figure 41. SELEX-seq process overview. (1) Incubation of target molecules with the library pool, (2) the subsequent separation of unbound oligonucleotides and the elution of bound oligonucleotides, and (3) then PCR amplification of bound sequences for three rounds. Created with Biorender.com

3.3 RESULTS

3.3.1 EMSA

In a typical EMSA experiment, the protein-DNA complex migrates more slowly than free nucleic acid. The essence of this technique is to layer a DNA-protein solution onto a gel and rapidly separate unbound DNA from DNA complex with proteins and hence has diminished electrophoretic mobility. The level of 'free' DNA seen in the gel must correspond to the amount free in the initial DNA-protein solution of interest, changes in free and complex DNA levels during the electrophoresis experiment is informative of binding (Garner & Revzin, 1986). The binding of OX1F, OX2F, OXEF, OX1D, OX2D, OXED, and THX was confirmed by EMSAs using a known binding site for OxyR in *E. coli (*OXGWCELL). After applying the adequate protein purification protocol of the p32a purified and dialyzed samples, these were probed in an EMSA; the appropriate binding functionality of each TFs can be seen in **Figures 42 - 46**.

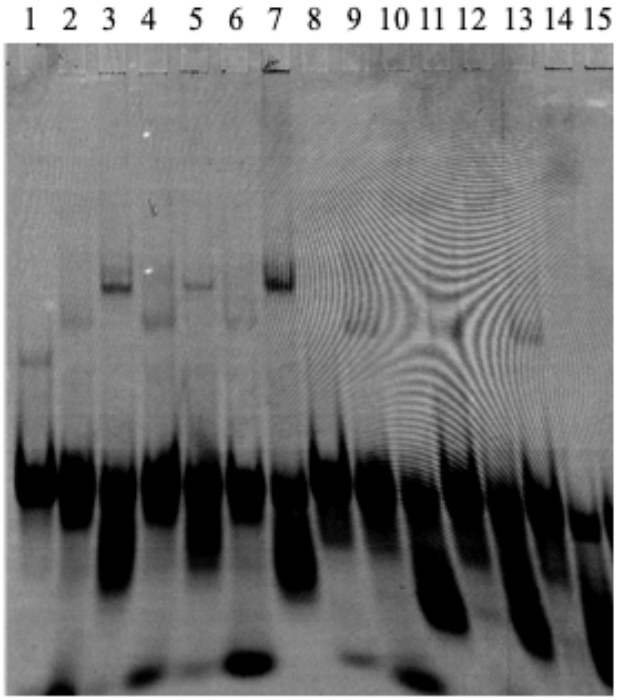
Initially, throughout the experiments, a volume of 2.5 µL of the purified protein shifted the DNA. However, a more intense shift was observed by using 5 µL. Specificity of the OxyR samples was assessed by incorporating a sequence-scrambled probe and maintaining a constant protein volume at 5 µL. Again, a shift was observed for each sample. A shift can be observed for the OxyR DBDs produced in some cases. The results demonstrate how easily this technique can characterize the DNA-binding proteins, requiring only prior knowledge of transcription factor consensus sequences. There are limitations to using EMSA as a technique for representing DNA–protein interactions *in vivo*, as physiological conditions will not be identical *in vitro*. Nevertheless, a shift can be observed for all the purified proteins under examination.



	Sample
1	Free DNA probe: OXGWCELL
2 - 4	OX1F
5 - 7	OX2F
8 - 10	OXEF
volume of protein added: 2.5 µL (lanes: 2, 5, 8, 11) , 1 µL (lanes: 3, 6, 9, 12) , and 1/10 dilution (lanes: 4, 7, 10, 13)	

OXGCELL- aGATAGGTTnAACCTATCgc

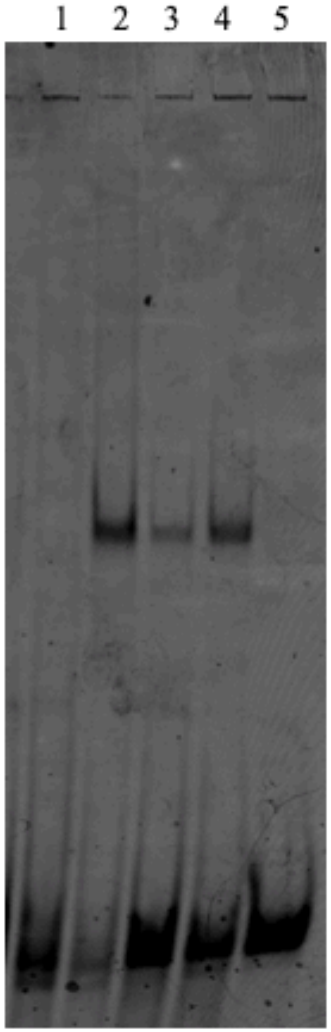
Figure 42. EMSA of OX1F, OX2F, and OXEF.



	Sample
1	Free DNA probe: OXGCELL
2	OX1F
3	OX1D
4	OX2F
5	OX2D
6	OXEF
7	OXED
8	Free DNA probe: OXGCELL_SCRAM
9	OX1F
10	OX1D
11	OX2F
12	OX2D
13	OXEF
14	OXED
15	THX
volume of protein added: 5 µL	

OXGCELL- aGATAGGTTnAACCTATCgc

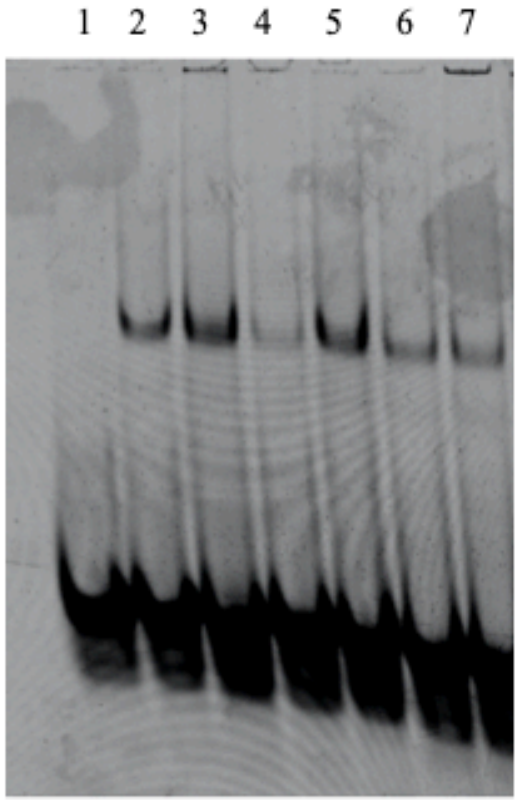
Figure 43. EMSA of OX1F, OX1D, OX2F, OX2D, OXEF, OXED, and THX.



	Sample
1	Free DNA probe: OXGWCELL
2	OX1D
3	OX2D
4	OXED
5	THX
volume of protein added: 5 µL	

OXGCELL- aGATAGGTTnAACCTATCgc

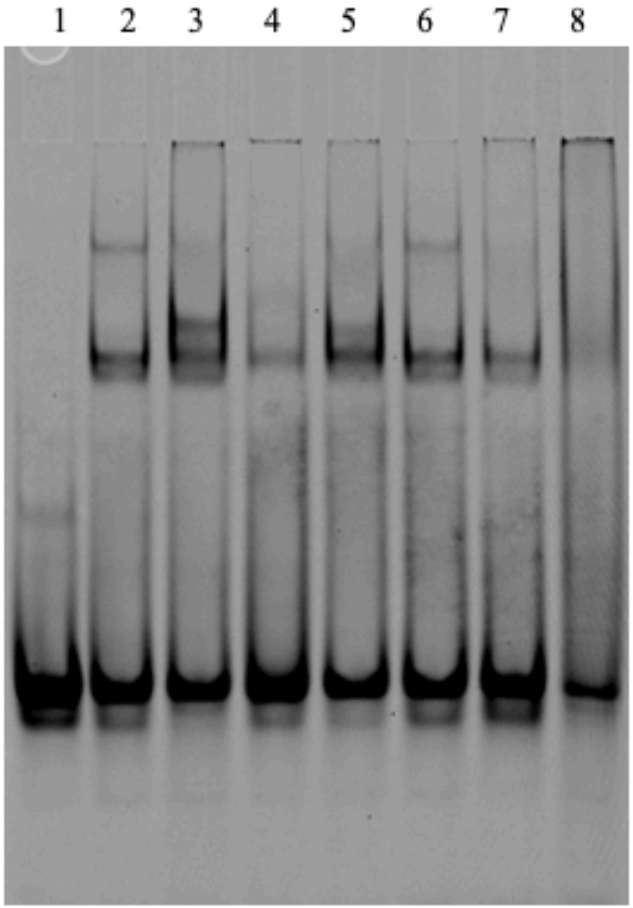
Figure 44. EMSA of OX1D, OX2D, OXED.



	Sample
1	Free DNA probe: OXGCELL
2	OX1F
3	OX2F
4	OXEF
5	OX1F
6	OX2F
7	OXEF
volume of protein added: 5 µL	

OXGCELL- aGATAGGTTnAACCTATCgc

Figure 45. EMSA of OX1F, OX1D, OX2F, OX2D, OXEF, OXED, THX.



	Sample
1	Free DNA probe: OXGCELL
2	OX1F
3	OX1D
4	OX2F
5	OX2D
6	OXEF
7	OXED
8	THX
volume of protein added: 5 µL	

OXGCELL- aGATAGGTTnAACCTATCgc

Figure 46. EMSA of OX1F, OX1D, OX2F, OX2D, OXEF, OXED, THX.

3.3.2 SELEX-SEQ

Starting with a synthetic oligonucleotide library with 20 randomized bases, we performed three SELEX-seq selection rounds for the purified OX1F, OX2F, OXEF, OX1D, OX2D, OXED, THX samples. For the SELEX-seq, a DNA probe was included to assess binding and it also serves as a tracking lane to monitor the mobility of the protein-DNA complex. In addition to the tracking lane, a control probe with a TF was included to monitor the mobility of the complex, and finally, one lane, contained the TF with the prepared library. Both the probe and the library are 60 bp long, which makes them migrate similarly in the gel. SELEX-seq gels for round 1, 2, and 3 are shown in **Figures 47, 48, 50, 51, 53, 54**. An agarose gel was used to monitor the DNA products after PCR and purification. Sample concentrations were obtained with Nanodrop UV/vis absorption, **Figures 49, 52** and **Tables 5, 6, 7**.

3.3.3 SELEX-SEQ ANALYSIS

The algorithm Autoseed, identifies all subsequences that represent local maxima, that is, are enriched more than any closely related sequence (Nitta et al., 2015). The method is based on a novel distance measure between two subsequences, the Huddinge distance. The Autoseed method is capable of identifying any specificity that results in an enrichment of subsequences longer than 2 bp. In this method, a subsequence is aligned to all other subsequences, and its count is compared to counts of subsequences that contain n − 1 perfectly matching bases, where n is the maximum number of defined bases in the aligned subsequences. The method compares also gapped subsequences to ungapped ones, and can thus identify diverse motifs with widely spaced recognition sites, and differentiate between monomers and dimers of the same subsequence (Nitta et al., 2015).

In general, longer subsequences are expected to be rarer and thus more easily identified. SELEX-seq combined with Autoseed may not be able to identify specificity for relatively non-

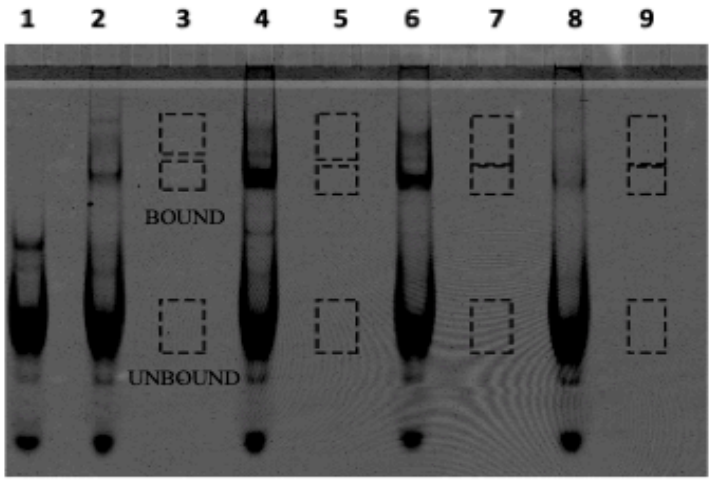
specific or accessory DNA-binding proteins that display weak mononucleotide or dinucleotide preferences. We tracked eight purified TFs over the selection rounds. The barcoded sequences with proper insert length were finally sorted by copy number and compiled into local max enrichment sequences, from which the TF logos of the most enriched sequences were generated.

The binding profile of the -expressed OxyR FL and DBDs logos, and local max enrichment sequences of the R3 of the SELEX-seq EMSA are indicated in **Figures 55 - 62.** Analysis of the OxyR-TFs SELEX-seq dataset using Autoseed identified a previously identified motif registered as consensus sequence for the *E. coli* OxyR, ATAGnT. The possible motifs found, based on their local max enrichment, for each OxyR are OXIF: CTAT (1,039), ATAG (1,039), ATGT (1,162), ACAT (1,162); OX1D: GTAT (2,392), ATAC (2,392); OX2FA: ATAG (916), CTAT (916), ATTA (1,121), TAAT (1,121); OX2FB: ATAG (1,588), CTAT (1,588), TCAT (1,172), ATGA (1,172), GATC (648); OX2D: ATAG (1,563), CTAT (1,563), ATAA (1,701), TTAT (1,701); OXEF: CTAC(T) (1,319), (A)TAG (1,319), CTTA (1,031), TAAG (1,031); OXED: ATAG (908), CTAT (908); THX: TATC (1,305), AAGG (1,305). The binding profile of the expressed OxyR FL and DBDs logos, and local max enrichment sequences of the R3 of the SELEX-seq EMSA are indicated in **Figures 55 - 62.** Despite possible logos corresponding to enriched sequences, most enrichment was suppressed by a low quantity of fold change in the present sequences. These results lead to low confidence in the data obtained through some bound sequences.

Table 5. UV-Vis Spectrophotometry OxyR sample Absorbance and estimated concentration.

Protein	Abs	Ext. Coe.	MW	[M]	[uM]	[nM]	PI	AA
OX1F	0.582	31440	51183.89	0.00001851145038	18.51145038	18511.45038	5.6	463
OX2F	0.549	37275	51398.02	0.00001472837022	14.72837022	14728.37022	5.99	463
OXEF	0.502	30410	52157.34	0.00001650772772	16.50772772	16507.72772	5.96	469
OX1D	2.007	22710	26703.83	0.00008837516513	88.37516513	88375.16513	6.11	240
OX2D	1.35	15595	26537.83	0.00008656620712	86.56620712	86566.20712	6.17	240
OXED	0.221	15595	26129.29	0.00001417120872	14.17120872	14171.20872	6.2	236
THX	1.207	14230	20398.09	0.00008482080112	84.82080112	84820.80112	5.81	189

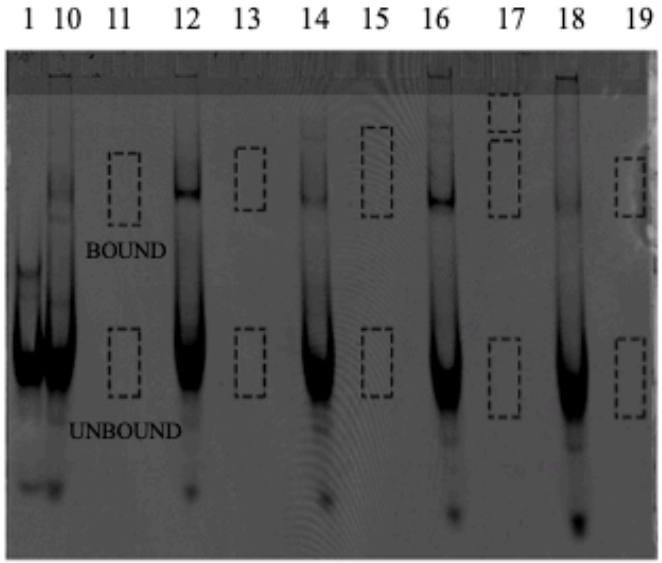
SELEX ROUND 1 (R1); GEL (1):



	Sample	DNA probe or DNA library
1		Free DNA probe: OXGCELL
2	OX1F	Free DNA probe: OXGCELL
3	OX1F	DNA library (20 N)
4	OX1D	Free DNA probe: OXGCELL
5	OX1D	DNA library (20 N)
6	OX2D	Free DNA probe: OXGCELL
7	OX2D	DNA library (20 N)
8	THX	Free DNA probe: OXGCELL
9	THX	DNA library (20 N)
volume of protein added: 5 µL		

Figure 47. EMSA gel (1); SELEX-seq round 1 (R1).

SELEX ROUND 1 (R1); GEL (2):



	Sample	DNA probe or DNA library
1		Free DNA probe: OXGCELL
10	OX2F_A	Free DNA probe: OXGCELL
11	OX2F_A	DNA library (20 N)
12	OX2F_B	Free DNA probe: OXGCELL
13	OX2F_B	DNA library (20 N)
14	OXEF	Free DNA probe: OXGCELL
15	OXEF	DNA library (20 N)
16	OXED	Free DNA probe: OXGCELL
17	OXED	DNA library (20 N)
18	THX	Free DNA probe: OXGCELL
19	THX	DNA library (20 N)
volume of protein added: 5 µL		

Figure 48. EMSA gel (2); SELEX-seq round 1 (R1).

Table 6. SELEX-seq round 1 (*R1*); Nanodrop concentrations after PCR purification; p32a purified TFs SELEX-seq samples.

#	Sample	Concentration [ng/uL]	Bound/ Unbound
1	OX1F	20.4	Bound
2	OX1F	27.2	Unbound
3	OX1D	29.4	Bound
4	OX1D	22.6	Unbound
5	OX2D	22.5	Bound
6	OX2D	23.9	Unbound
7	THX	18.2	Bound
8	THX	37.5	Unbound
9	GEL	16.4	Bound
10	OX2F_A	21.1	Unbound
11	OX2F_A	29.6	Bound
12	OX2F_B	21.8	Unbound
13	OX2F_B	23.8	Bound
14	OXEF	21.8	Bound
15	OXEF	23.8	Unbound

SELEX ROUND # 1:

	Sample	Concentration [ng/uL]	Bound/ Unbound
16	OXED	12.4	Bound 1
17	OXED	15.6	Bound 2
18	OXED	25.0	Unbound
19	THX	20.7	Bound
20	THX	27.9	Unbound
21	Negative control	9.9	
22	PCR + Beads	9.3	

SELEX ROUND 1 (R1); Purified samples (Bound / Unbound):

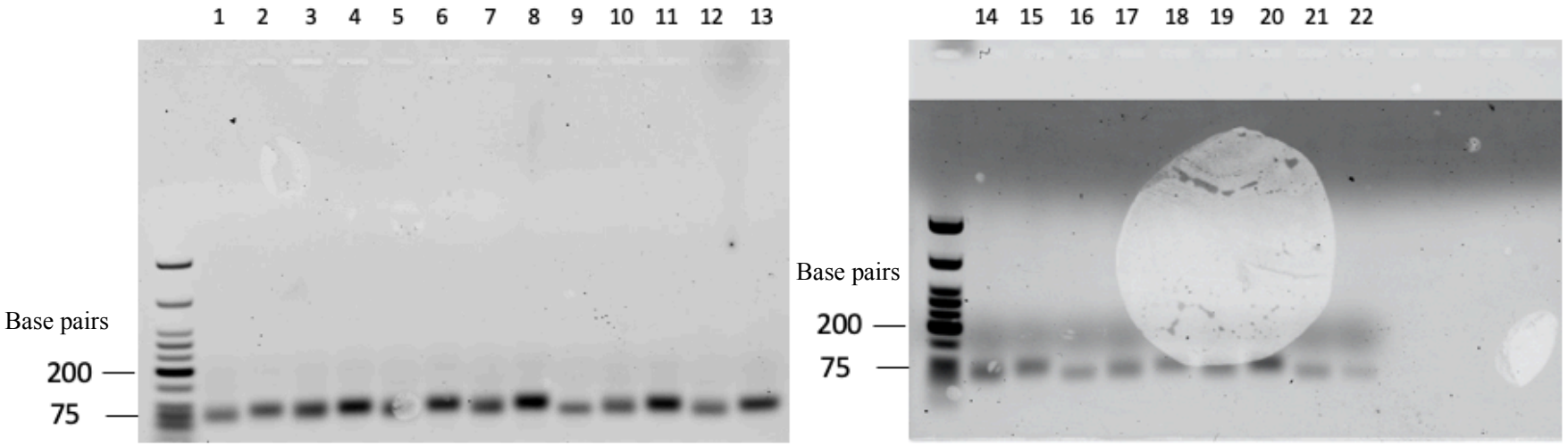
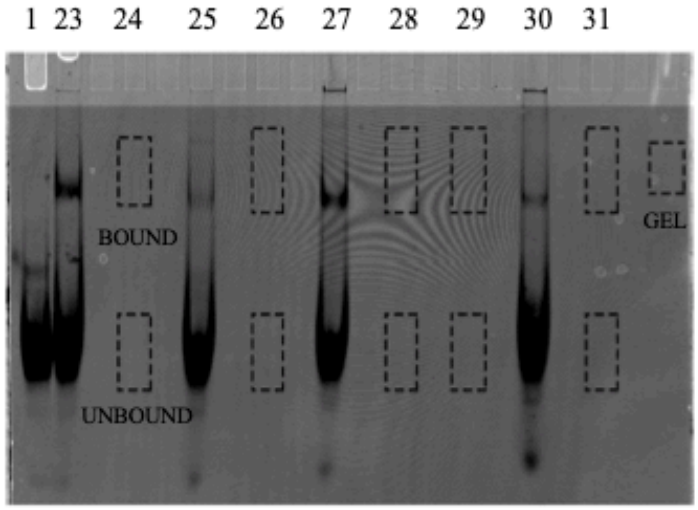


Figure 49. SELEX-seq round 1 (R1); Agarose gels after PCR purification of p32a purified TFs; Samples are labeled according to Table 6.

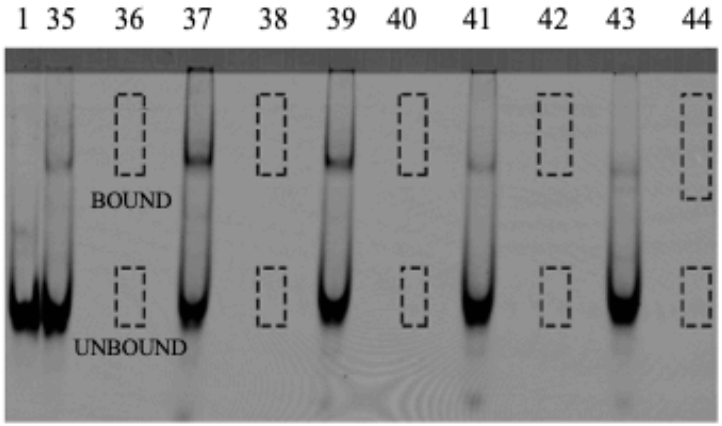
SELEX ROUND 2 (R2); GEL (3):



	Sample	DNA probe or DNA library
1		Free DNA probe: OXGCELL
23	OX2F_B	Free DNA probe: OXGCELL
24	OX2F_B	DNA library (20 N)
25	OXEF	Free DNA probe: OXGCELL
26	OXEF	DNA library (20 N)
27	OXED	Free DNA probe: OXGCELL
28	OXED_B1	DNA library (20 N)
29	OXED_B2	DNA library (20 N)
30	THX	Free DNA probe: OXGCELL
31	THX	DNA library (20 N)
volume of protein added: 5 µL		

Figure 50. EMSA gel (3); SELEX-seq round 2 (R2).

SELEX ROUND 2 (R2); GEL (4):



	Sample	DNA probe or DNA library
1		Free DNA probe: OXGCELL
35	OX1F	Free DNA probe: OXGCELL
36	OX1F	DNA library (20 N)
37	OX1D	Free DNA probe: OXGCELL
38	OX1D	DNA library (20 N)
39	OX2D	Free DNA probe: OXGCELL
40	OX2D	DNA library (20 N)
41	THX	Free DNA probe: OXGCELL
42	THX	DNA library (20 N)
43	OX2F_A	Free DNA probe: OXGCELL
44	OX2F_A	DNA library (20 N)
volume of protein added: 5 µL		

Figure 51. EMSA gel (4); SELEX-seq round 2 (R2).

Table 7. SELEX-seq round 1 (R2); Nanodrop concentrations after PCR purification; purified TFs SELEX-seq samples.

	Sample	Concentration [ng/uL]	Bound/ Unbound
23	OX2F_B	21.1	Bound
24	OX2F_B	26.8	Unbound
25	OXEF	27.4	Bound
26	OXEF	28.0	Unbound
27	OXED BAND1	24.9	Bound
28	OXED BAND 1	24.0	Unbound
29	OXED BAND 2	18.6	Bound
30	OXED BAND2	24.3	Unbound
31	THX	25.5	Bound
32	THX	21.7	Unbound
33	GEL (-)	12.2	
34	PCR + BEADS	2.6	
35	OX1F	16.0	Bound
36	OX1F	21.8	Unbound

SELEX ROUND # 2:

	Sample	Concentration [ng/uL]	Bound/ Unbound
37	OX1D	20.7	Bound
38	OX1D	20.6	Unbound
39	OX2D	17.0	Bound
40	OX2D	23.7	Unbound
41	THX	16.6	Bound
42	THX	18.0	Unbound
43	OX2F_D1	19.5	Bound
44	OX2F_D1	18.5	Unbound
45	GEL NEGATIVE	9.0	
46	PCR + BEADS	4.1	

SELEX ROUND 2 (R2); Purified samples (Bound / Unbound):

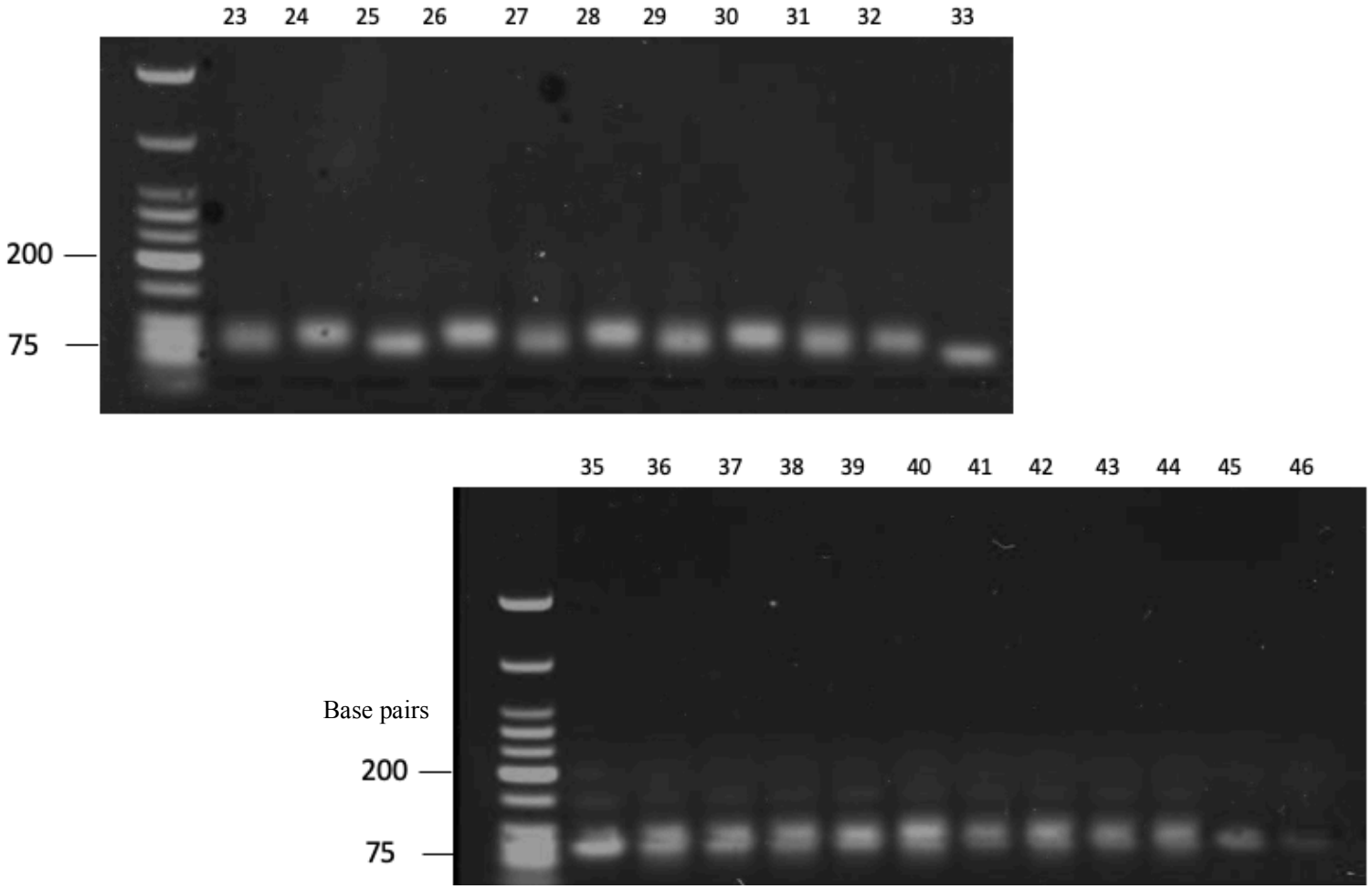
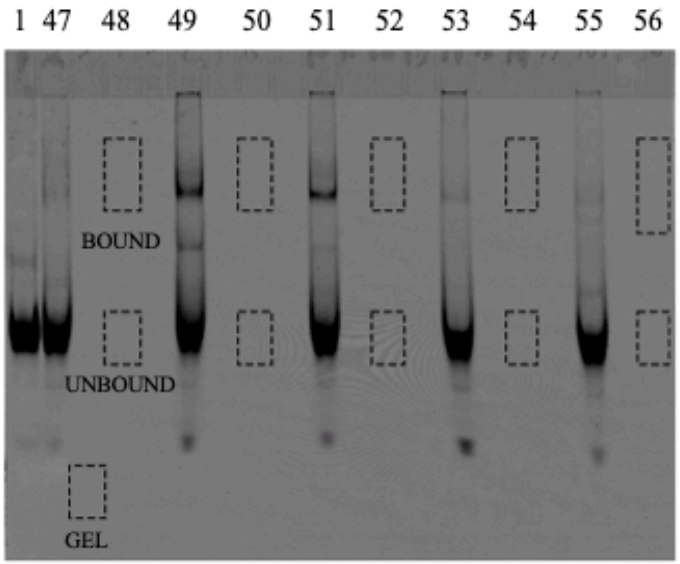


Figure 52. SELEX-seq round 2 (R2); Agarose gels after PCR purification of purified TFs; Samples are labeled according to Table 6.

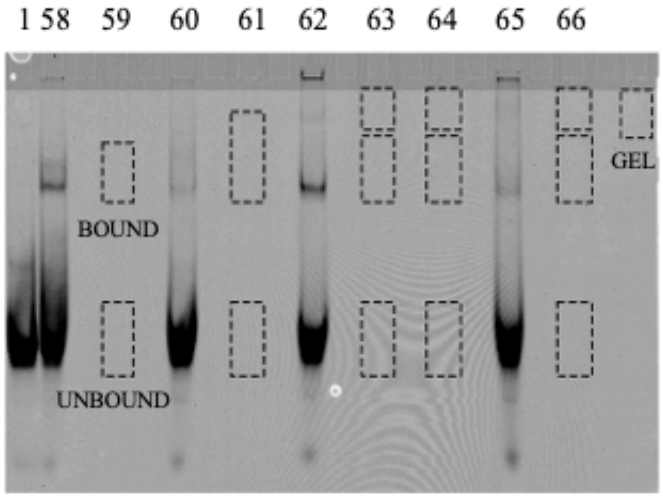
SELEX ROUND 3 (R3); GEL (5):



	Sample	DNA probe or DNA library
1		Free DNA probe: OXGCELL
47	OX1F	Free DNA probe: OXGCELL
48	OX1F	DNA library (20 N)
49	OX1D	Free DNA probe: OXGCELL
50	OX1D	DNA library (20 N)
51	OXED	Free DNA probe: OXGCELL
52	OXED	DNA library (20 N)
53	THX	Free DNA probe: OXGCELL
54	THX	DNA library (20 N)
55	OX2F_A	Free DNA probe: OXGCELL
56	OX2F_A	DNA library (20 N)
volume of protein added: 5 µL		

Figure 53. EMSA gel (5); SELEX-seq round 3 (R3); purified TFs.

SELEX ROUND 3 (R3); GEL (6):



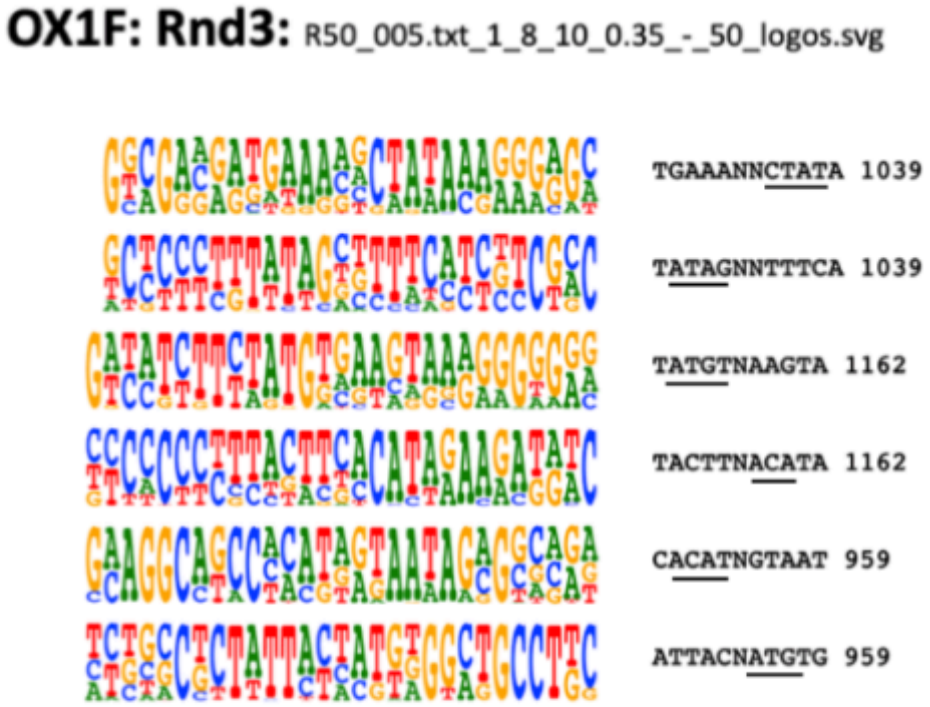
	Sample	DNA probe or DNA library
1		Free DNA probe: OXGCELL
58	OX2F_B	Free DNA probe: OXGCELL
59	OX2F_B	DNA library (20 N)
60	OXEF	Free DNA probe: OXGCELL
61	OXEF	DNA library (20 N)
62	OXED	Free DNA probe: OXGCELL
63	OXED_B1	DNA library (20 N)
64	OXED_B2	DNA library (20 N)
65	THX	Free DNA probe: OXGCELL
66	THX	DNA library (20 N)
volume of protein added: 5 µL		

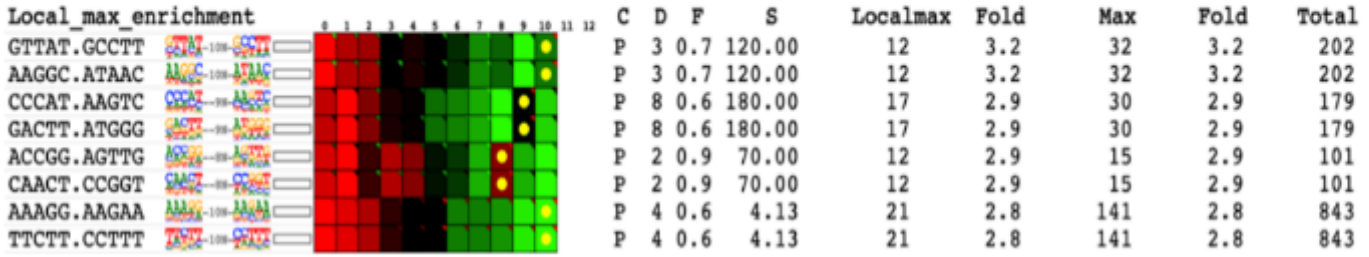
Figure 54. EMSA gel (6); SELEX-seq round 3 (R3); p32a purified TFs.

Table 8. SELEX-seq round 3 (R3); Nanodrop concentrations after PCR purification; p32a purified TFs SELEX-seq samples.

	Sample	Concentration [ng/uL]	Bound/ Unbound
47	OX1F	20.0	Bound
48	OX1F	25.3	Unbound
49	OX1D	22.0	Bound
50	OX1D	24.0	Unbound
51	OX2D	21.9	Bound
52	OX2D	20.7	Unbound
53	THX	22.6	Bound
54	THX	21.6	Unbound
55	OX2F_D1	32.5	Bound
56	OX2F_D1	22.7	Unbound
57	GEL	5.5	

	Sample	Concentration [ng/uL]	Bound/ Unbound
58	OX2F_D2	22.2	Bound
59	OX2F_D2	19.7	Unbound
60	OXEF	16.2	Bound
61	OXEF	26.4	Unbound
62	OXED BAND 1 - 1	9.8	Bound
63	OXED BAND 1 - 2	21.8	Bound
64	OXED	21.8	Unbound
65	OXED BAND 2 – 1	14.5	Bound
66	OXED BAND 2- 2	16.2	Bound
67	OXED	18.2	Unbound
68	THX BAND 1	11.6	Bound
69	THX BAND 2	15.8	Bound
70	THX	20.2	Unbound
71	Gel Negative	11.8	
72	PCR + Beads	14.7	





Local_max_enrichment																		
GTTAT.GCCTT												C	D	F	S	Localmax	Fold	Max	Fold	Total
AAGGC.ATAAC												P	3	0.7	120.00	12	3.2	32	3.2	202
CCCAT.AAGTC												P	3	0.7	120.00	12	3.2	32	3.2	202
GACTT.ATGGG												P	8	0.6	180.00	17	2.9	30	2.9	179
ACCGG.AGTTG												P	8	0.6	180.00	17	2.9	30	2.9	179
CAACT.CCGGT												P	2	0.9	70.00	12	2.9	15	2.9	101
AAAGG.AAGAA												P	2	0.9	70.00	12	2.9	15	2.9	101
TTCTT.CCTTT												P	4	0.6	4.13	21	2.8	141	2.8	843
												P	4	0.6	4.13	21	2.8	141	2.8	843

Figure 55. SELEX-seq round 3 (R3); OX1F; Autoseed analysis identifying all subsequences that represent local max enrichment.

Local_max_enrichment	C	D	F	S	Localmax	Fold	Max	Fold	Total
TATTC.GTATT	-	0	0.0	0.69	224	5.4	224	5.4	508
AATAC.GAATA	-	0	0.0	0.69	224	5.4	224	5.4	508
ATAAT.ACGAA	P	7	0.3	4.23	173	5.0	173	5.0	427
TTCGT.ATTAT	P	7	0.3	4.23	173	5.0	173	5.0	427
GAACG.TTCCC	P	4	0.5	2.58	68	4.8	68	4.8	101
GGGAA.CGTTC	P	4	0.5	2.58	68	4.8	68	4.8	101
TATTC.CTTCT	P	2	0.8	180.00	15	4.6	40	4.6	257
AGAAG.GAATA	P	2	0.8	180.00	15	4.6	40	4.6	257
CCTCT.AGTTA	P	3	0.9	160.00	17	4.1	24	4.1	125
TAACT.AGAGG	P	3	0.9	160.00	17	4.1	24	4.1	125
GGAAC.GTTCC	P	0	0.0	480.00	50	4.0	56	4.5	120
CATTC.TTGCA	P	4	0.8	130.00	14	4.0	27	4.0	157
TGCAA.GAATG	P	4	0.8	130.00	14	4.0	27	4.0	157
TAAGG.TGATA	-	2	0.7	1.08	117	4.0	117	4.0	239
TATCA.CCTTA	-	2	0.7	1.08	117	4.0	117	4.0	239
AATAC.CACCA	-	7	0.6	0.99	15	4.0	25	4.0	139
TGGTG.GTATT	-	7	0.6	0.99	15	4.0	25	4.0	139

Figure 56. SELEX-seq round 3 (R3); OX1D; Autoseed analysis identifying all subsequences that represent local max enrichment.

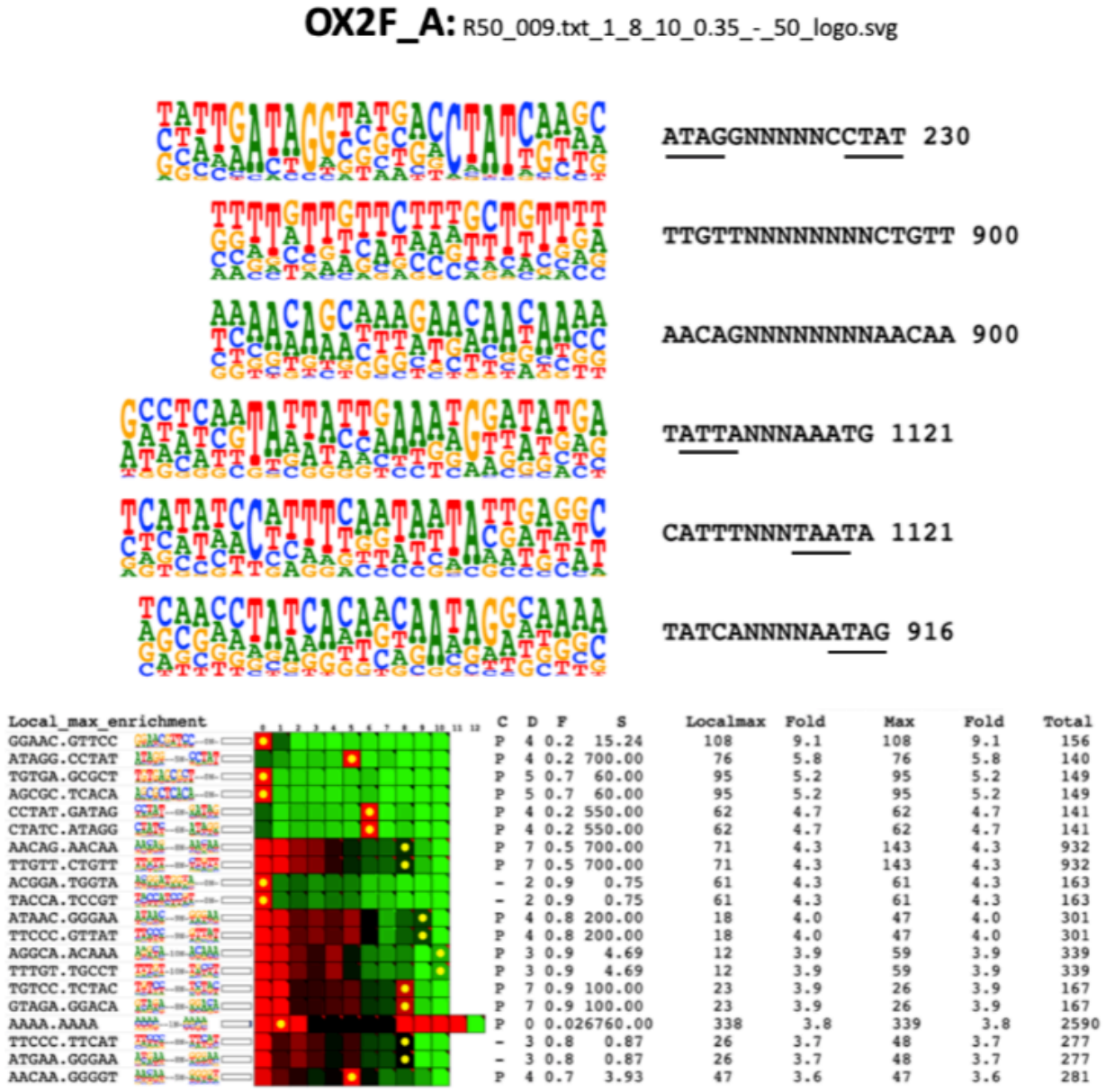


Figure 57. SELEX-seq round 3 (R3); OX2F; Autoseed analysis identifying all subsequences that represent local max enrichment.

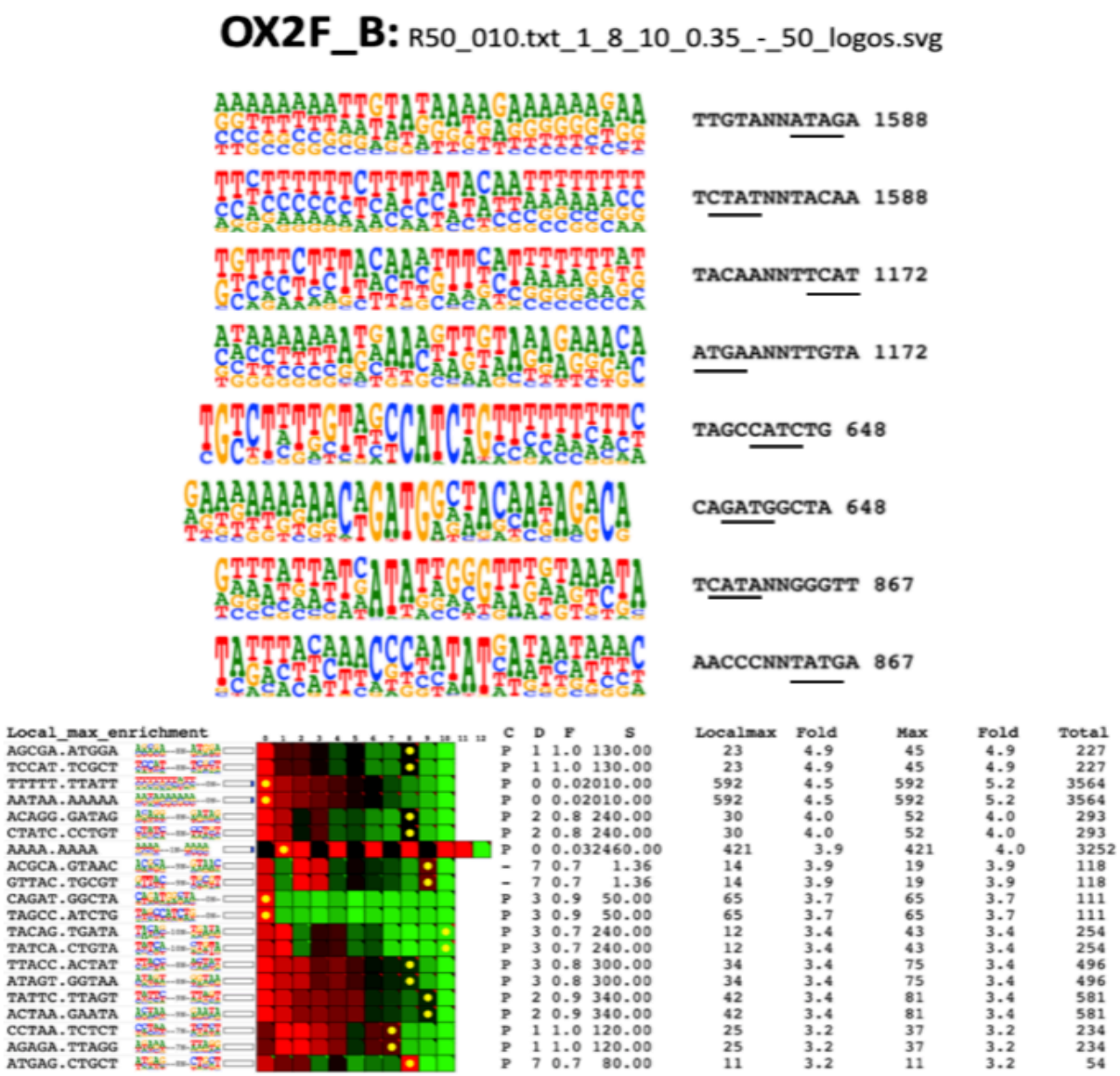
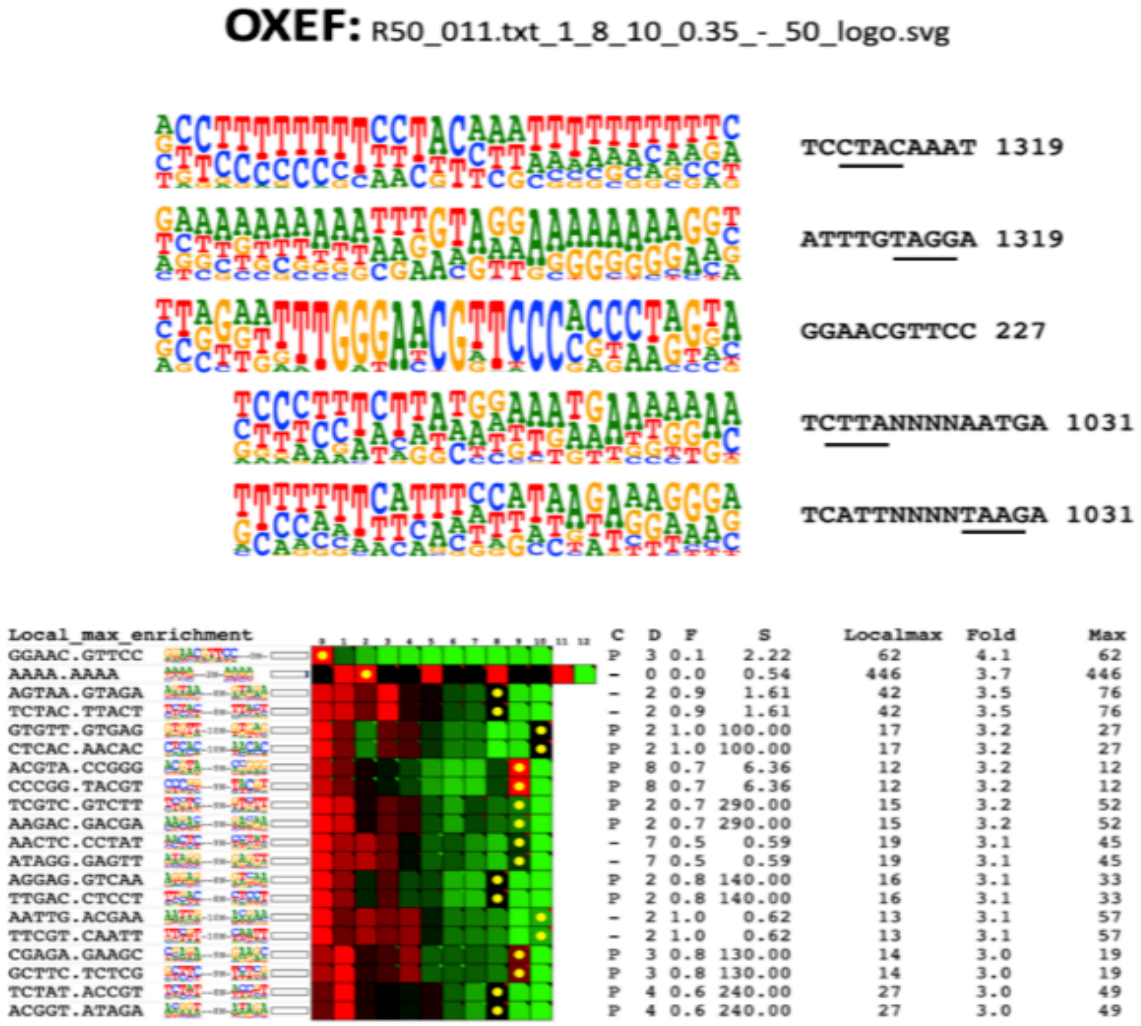


Figure 58. SELEX-seq round 3 (R3); OX2F; Autoseed analysis identifying all subsequences that represent local max enrichment.

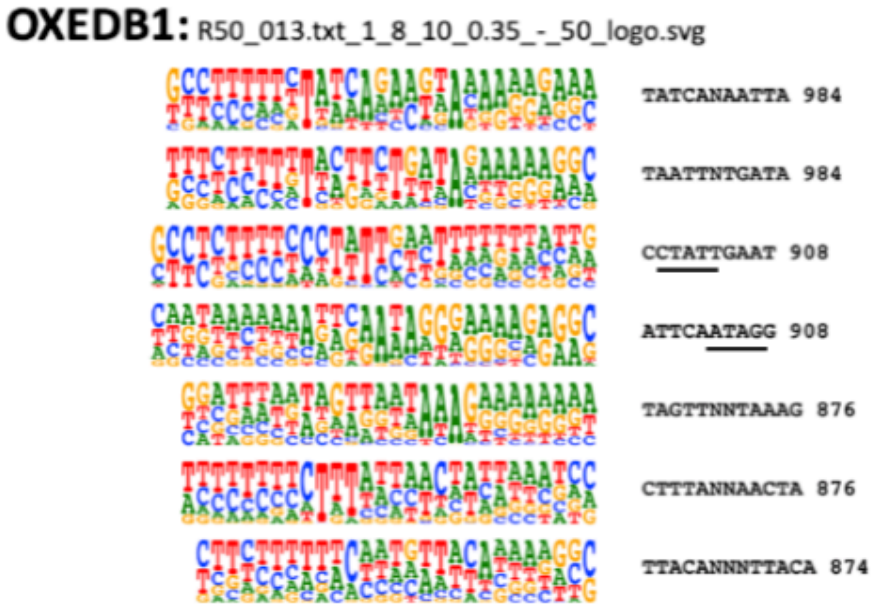
Local_max_enrichment	C	D	F	S	Localmax	Fold	Max	Fold	Total
AATAC.GAATA	P	3	0.9	240.00	305	5.5	305	5.5	639
TATTC.GTATT	P	3	0.9	240.00	305	5.5	305	5.5	639
GAACG.TTCCC	-	1	1.0	1.48	102	5.3	102	5.3	154
GGGAA.CGTTC	-	1	1.0	1.48	102	5.3	102	5.3	154
AATAA.GGTAA	-	2	0.9	0.86	228	5.0	228	5.0	707
TTACC.TTATT	-	2	0.9	0.86	228	5.0	228	5.0	707
TATCA.CCTTA	-	2	0.8	0.73	192	4.9	192	4.9	347
TAAGG.TGATA	-	2	0.8	0.73	192	4.9	192	4.9	347
TATTA.CTATT	-	3	0.7	1.00	275	4.7	275	4.7	660
AATAG.TAATA	-	3	0.7	1.00	275	4.7	275	4.7	660
GTGTT.TGCTT	P	2	0.8	260.00	20	3.7	49	3.7	305
AAGCA.AACAC	P	2	0.8	260.00	20	3.7	49	3.7	305
GGAAC.GTTCC	P	0	0.1	26.67	58	3.5	100	5.9	176
AAGGG.ACAGC	P	2	0.8	6.36	11	3.5	20	3.5	134
GCTGT.CCCTT	P	2	0.8	6.36	11	3.5	20	3.5	134
AGCAC.GGATG	P	3	0.9	2.38	51	3.4	51	3.4	103
CATCC.GTGCT	P	3	0.9	2.38	51	3.4	51	3.4	103
ACGGA.TGGTA	P	3	0.8	15.45	50	3.3	50	3.3	179
TACCA.TCCGT	P	3	0.8	15.45	50	3.3	50	3.3	179
TTCCC.CGTTT	-	2	0.8	0.88	17	3.2	66	3.2	302

Figure 59. SELEX-seq round 3 (R3); OX2D; Autoseed analysis identifying all subsequences that represent local max enrichment.



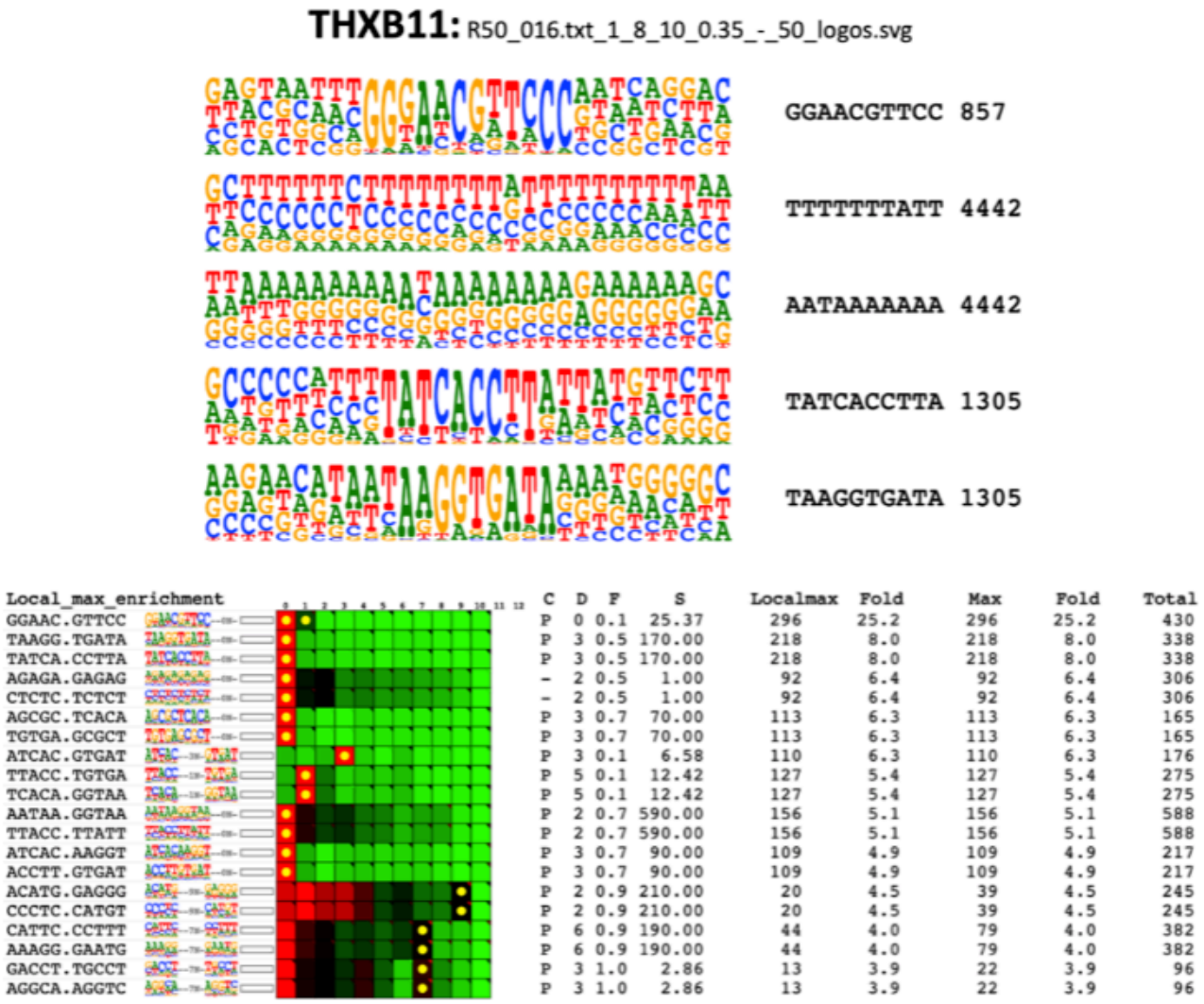
Local_max_enrichment	C	D	F	S	Localmax	Fold	Max	Fold	Total
GGAAC.GTTCC	P	3	0.1	2.22	62	4.1	62	4.1	94
AAAA.AAAA	-	0	0.0	0.54	446	3.7	446	3.9	3497
AGTAA.GTAGA	-	2	0.9	1.61	42	3.5	76	3.5	498
TCTAC.TTACT	-	2	0.9	1.61	42	3.5	76	3.5	498
GTGTT.GTGAG	P	2	1.0	100.00	17	3.2	27	3.2	173
CTCAC.AACAC	P	2	1.0	100.00	17	3.2	27	3.2	173
ACGTA.CCGGG	P	8	0.7	6.36	12	3.2	12	3.2	63
CCCGG.TACGT	P	8	0.7	6.36	12	3.2	12	3.2	63
TCGTC.GTCTT	P	2	0.7	290.00	15	3.2	52	3.2	281
AAGAC.GACGA	P	2	0.7	290.00	15	3.2	52	3.2	281
AACTC.CCTAT	-	7	0.5	0.59	19	3.1	45	3.1	253
ATAGG.GAGTT	-	7	0.5	0.59	19	3.1	45	3.1	253
AGGAG.GTCAA	P	2	0.8	140.00	16	3.1	33	3.1	145
TTGAC.CTCCT	P	2	0.8	140.00	16	3.1	33	3.1	145
AATTG.ACGAA	-	2	1.0	0.62	13	3.1	57	3.1	335
TTCGT.CAATT	-	2	1.0	0.62	13	3.1	57	3.1	335
CGAGA.GAAGC	P	3	0.8	130.00	14	3.0	19	3.0	110
GCTTC.TCTCG	P	3	0.8	130.00	14	3.0	19	3.0	110
TCTAT.ACCGT	P	4	0.6	240.00	27	3.0	49	3.0	261
ACGGT.ATAGA	P	4	0.6	240.00	27	3.0	49	3.0	261

Figure 60. SELEX-seq round 3 (R3); OXEF; Autoseed analysis identifying all subsequences that represent local max enrichment.



Local_max_enrichment	C	D	F	S	Localmax	Fold	Max	Fold	Total
CACTC.TCTGT	P	1	1.0	120.00	22	4.3	33	4.3	221
ACAGA.GAGTG	P	1	1.0	120.00	22	4.3	33	4.3	221
AGTTG.GGAAA	P	3	0.9	110.00	14	4.3	32	4.3	231
TTTCC.CAACT	P	3	0.9	110.00	14	4.3	32	4.3	231
TGAGA.GGAAG	P	6	1.0	160.00	35	4.3	36	4.3	232
CTTCC.TCTCA	P	6	1.0	160.00	35	4.3	36	4.3	232
CCTCT.ACTAT	-	5	0.5	1.52	52	4.3	52	4.3	281
ATAGT.AGAGG	-	5	0.5	1.52	52	4.3	52	4.3	281
AACTG.AATAG	-	4	0.9	1.04	40	4.2	44	4.2	262
CTATT.CAGTT	-	4	0.9	1.04	40	4.2	44	4.2	262
TCACA.CGAAA	P	7	0.8	160.00	23	4.2	33	4.2	186
TTTCG.TGTGA	P	7	0.8	160.00	23	4.2	33	4.2	186
TTACC.TTTTT	P	6	0.6	660.00	87	4.2	157	4.2	1002
AAAAA.GGTAA	P	6	0.6	660.00	87	4.2	157	4.2	1002
AAGCA.AACGG	P	2	0.7	2.23	47	4.1	47	4.1	240
CCGTT.TGCTT	P	2	0.7	2.23	47	4.1	47	4.1	240
CCTGG.GTTCT	P	6	0.9	90.00	14	4.0	27	4.0	109
AGAAC.CCAGG	P	6	0.9	90.00	14	4.0	27	4.0	109

Figure 61. SELEX-seq round 3 (R3); OXED; Autoseed analysis identifying all subsequences that represent local max enrichment.



	C	D	F	S	Localmax	Fold	Max	Fold	Total
GGAAC.GTTCC	P	0	0.1	25.37	296	25.2	296	25.2	430
TAAGG.TGATA	P	3	0.5	170.00	218	8.0	218	8.0	338
TATCA.CCTTA	P	3	0.5	170.00	218	8.0	218	8.0	338
AGAGA.GAGAG	-	2	0.5	1.00	92	6.4	92	6.4	306
CTCTC.TCTCT	-	2	0.5	1.00	92	6.4	92	6.4	306
AGCGC.TCACA	P	3	0.7	70.00	113	6.3	113	6.3	165
TGTGA.GCGCT	P	3	0.7	70.00	113	6.3	113	6.3	165
ATCAC.GTGAT	P	3	0.1	6.58	110	6.3	110	6.3	176
TTACC.TGTGA	P	5	0.1	12.42	127	5.4	127	5.4	275
TCACA.GGTAA	P	5	0.1	12.42	127	5.4	127	5.4	275
AATAA.GGTAA	P	2	0.7	590.00	156	5.1	156	5.1	588
TTACC.TTATT	P	2	0.7	590.00	156	5.1	156	5.1	588
ATCAC.AAGGT	P	3	0.7	90.00	109	4.9	109	4.9	217
ACCTT.GTGAT	P	3	0.7	90.00	109	4.9	109	4.9	217
ACATG.GAGGG	P	2	0.9	210.00	20	4.5	39	4.5	245
CCCTC.CATGT	P	2	0.9	210.00	20	4.5	39	4.5	245
CATTC.CCTTT	P	6	0.9	190.00	44	4.0	79	4.0	382
AAAGG.GAATG	P	6	0.9	190.00	44	4.0	79	4.0	382
GACCT.TGCCT	P	3	1.0	2.86	13	3.9	22	3.9	96
AGGCA.AGGTC	P	3	1.0	2.86	13	3.9	22	3.9	96

Figure 62. SELEX-seq round 3 (R3); THX; Autoseed analysis identifying all subsequences that represent local max enrichment.

3.4 DISCUSSION

The OXGWCELL probe presented a shift for all the OxyRs under study, meaning the proteins were purified successfully and are functional. Although the shifts sometimes are dim, the presence of binding is corroborated multiple times, and can also be seen during the SELEX-seq experiments. In the EMSA approach, the protein-bound DNA was examined by locating the DNA/protein binding band with native polyacrylamide gels.

In this study, the SELEX-seq strategy was also incorporated. The specificity of PCR amplification was further protected by low-cycle amplification (15 cycles) with Taq polymerase. Also, barcoding was split to fit coding to the library from all the rounds. The specificity and the enrichment of the SELEX-seq selection rounds were monitored by PCR amplification sequences obtained from each of the rounds. According to the data obtained, some of the TFs shared a similar motif in certain cases. Although the reads for each motif is low, it is recognized that the motifs are represented in previous literature.

The R3 for OX1F, OX2F_A, OX2F_B, and OXED presented an affinity consensus sequence previously characterized for OxyR, but just a low number of fold enrichment was observed throughout the SELEX rounds. Extensive research of the different rounds may provide insight into patterns that may identify DNA-binding motifs for each TFs. In this study, the transcription factor OxyR was used as the target protein, because studies on its DNA-binding specificity have been done by other groups, in different bacterial strains. Results obtained by SELEX-seq strategy can be compared with those obtained by other previous studies.

4 CHAPTER 4

ALLIVIBRIO FISCHERI OXYR1 AND OXYR2 PROTEIN STRUCTURE PREDICTIONS AND MODEL BUILDING

In collaboration with Laura M. Rodriguez-Bonilla and Victoria Pinto as part of their undergraduate research

4.1 INTRODUCTION

Around the year 2014, The UniProtKB/TrEMBL protein database contained over 80 million protein sequences. However, the Protein Data Bank contains just over 100,000 experimentally determined 3D structures (Kelley et al., 2015). This ever-widening gap between our knowledge of sequence space and structure space poses serious challenges when examining the structure and function of a protein sequence of interest. Advances in computational techniques to predict protein structure and function allow, on average, to structurally model 50–70% of a typical genome by solely computational modeling techniques (Kelley et al., 2015). Protein structure is more conserved in evolution than protein sequence and evidences a finite and relatively small (1,000–10,000) number of unique protein folds in nature.

Some computational methods include simulated folding using physics-based or empirically derived energy functions construction of models from small fragments of known structures. Threading, where the compatibility of a sequence with an experimentally derived fold, is determined using similar energy functions and template-based modeling (TBM), in which a sequence is aligned to a sequence of known structures based on patterns of evolutionary variation (Kelley et al., 2015). Usually, protein structure prediction relies on a method that compares a protein sequence of interest with a large database of sequences; this allows for constructing an evolutionary or statistical profile of that sequence and subsequently scans this profile against a database of profiles for known structures. Some of the most widely used web servers for protein modeling are Phyre2, i-TASER, Swiss-Model, HHpred, PSI-BLAST–based secondary structure prediction (PSIPRED), Robetta, and Raptor. Phyre2 is one of the most widely used protein structure prediction servers, and it serves ∼40,000 unique users per year, processing ∼1,000 user-submitted proteins per day (Kelley et al., 2015).

4.2 MATERIALS AND METHODS

4.2.1 PROTEIN STRUCTURE PREDICTIONS

To predict *A. fischeri* OxyR1, OxyR2, and *E.coli* OxyR (EcOxyR) transcription factor structures, we used the web-based Protein Homology/analogy Recognition Engine V 2.0 server (Phyre2)and the molecular visualization system PyMol. We first downloaded the *A. fischeri* OxyR1, OxyR2, and OxyRE full-length and DNA-binding domain (DBD) gene sequences from the Kyoto Encyclopedia of Genes and Genomes (KEGG). Then, Blastx, a subprogram of the Basic Local Alignment Search Tool (BLAST), was used to translate the nucleotide gene sequence query resulting in the protein sequences (Wheeler D., 2007). These sequences were uploaded separately into the Phyre2 server, which uses a library of known protein structures taken from the Structural Classification of Proteins (SCOP) database and augmented with newer depositions in the Protein Data Bank (Kelley & Sternberg, 2009). Our protein sequence queries were then scanned against a non-redundant sequence (PDB) database, and a protein model profile was constructed and deposited in the 'fold library' (Kelley & Sternberg, 2009). The secondary query structure was predicted and scanned against the fold library using a profile-profile alignment algorithm following profile construction. The ten highest-scoring alignments were used to construct the full 3D protein model (Kelley & Sternberg, 2009). Finally, the Phyre database was searched for proteins with known structures aligned with the sequence of full-length and OxyR DBD (Zaim & Kierzek, 2003). For our final protein predictions, we chose to model and download our PDB files based on the *Corynebacterium glutamicum* (PDB code ID 6g4r) OxyR1 full-length, which had a confidence value and percent identity (% i.d.) of 100 and 34, respectively. OxyR1 DBD 99.9 and 38, OxyR2 full-length 100 and 34, OxyR2 DBD 99.9 and 39, EcOxyR full-length 100.0 and 35, and EcOxyR DBD 99.9 and 42.

4.2.2 MODEL BUILDING

The molecular visualization system PyMol was used to model the protein structure predictions. We uploaded each PDB file downloaded from Phyre2 and followed the next steps. To model the protein prediction structures, we worked with each protein individually (OxyR1 full-length, OxyR1 DBD, OxyR2 full-length, OxyR2 DBD, OxyRE full-length, and OxyRE DBD). We started with OxyR1 full-length and worked only with the DNA-binding domain residues in our alignments. First, we identified the DNA-binding domain (DBD) and the regulatory domain (RD) in the protein with different colors. Then we uploaded the PDB file of a well-known LysR-type transcription factor known as BenM (PDB code ID 4IHT) into the same PyMol window. This was done to align our protein to a DNA model and identify DNA interacting residues using BenM as a guide. BenM is a tetramer with two DBDs and two RDs. Using the 'align' tool in PyMol, we aligned one of the DBD chains of the BenM protein to the DBD residues of OxyR1. Since the DNA binding residues of BenM have been identified in previous studies, aligning the DBD of OxyR1 to BenM would yield accurate predictions of the DNA-binding residues in our OxyR proteins. Next, we looked at each amino acid in the BenM protein known to interact with DNA and identified the amino acid residue in OxyR1 that had been aligned at that position.

To view DNA-interacting residues more easily, each residue had a different color and was shown as a 'sticks' model, using the 'show' tool in PyMol. To create the mirrored protein, we uploaded a second identical copy of the OxyR1 full-length protein PDB file (obtained as before, from Phyre2) in the same PyMol window. This second OxyR1full-length molecule was aligned to the remaining DBD chain of the BenM protein. Next, the DNA-interacting residues of the remaining DBD chain were identified following the steps described before. Then, the BenM protein was removed from the PyMol window, and the remaining molecule and DNA strands were uploaded to Chiron (Ramachandran et al., 2011). This automated web server evaluates and

resolves clashes in protein structures. After this final step, the 3D model of OxyR1 with our predicted DNA-interacting residues identified was ready for publication. To model the protein structures of the OxyR2 and OxyRE proteins, all the previous steps were repeated without modification for both full-length and DBD proteins.

4.3 RESULTS

4.3.1 PROTEIN STRUCTURE PREDICTIONS & MODEL BUILDING

Table 9. Prediction of DNA-binding residues based on OxyR protein model predictions

Protein	Residues that interact with DNA
A. fischeri OxyR1	K56, Q32, S33, Q40, S31, T34, G38
A. fischeri OxyR2	R53, Q29, P30, K37, S28, T31, Q35
E.coli OxyR	R53, Q29, P30, K37, S28, T31, Q35

Table 10. Importance of DNA-binding residues in OxyR protein model predictions

Amino	Amino Acid Charge at pH 7	Role in DNA-binding
Lysine (K)	+ 1	Positive charges interact favorably with negatively charged DNA
Glutamine (Q)	- 1	Provides hydrogen bonding to adenines
Serine (S)		Provides methyl-group recognition of thymidine and hydrogen bonding
Arginine (R)	+1	Positive charges interact favorably with negatively charged DNA
Proline (P)		Easily bends when in an □-helix conformation and provides structure.

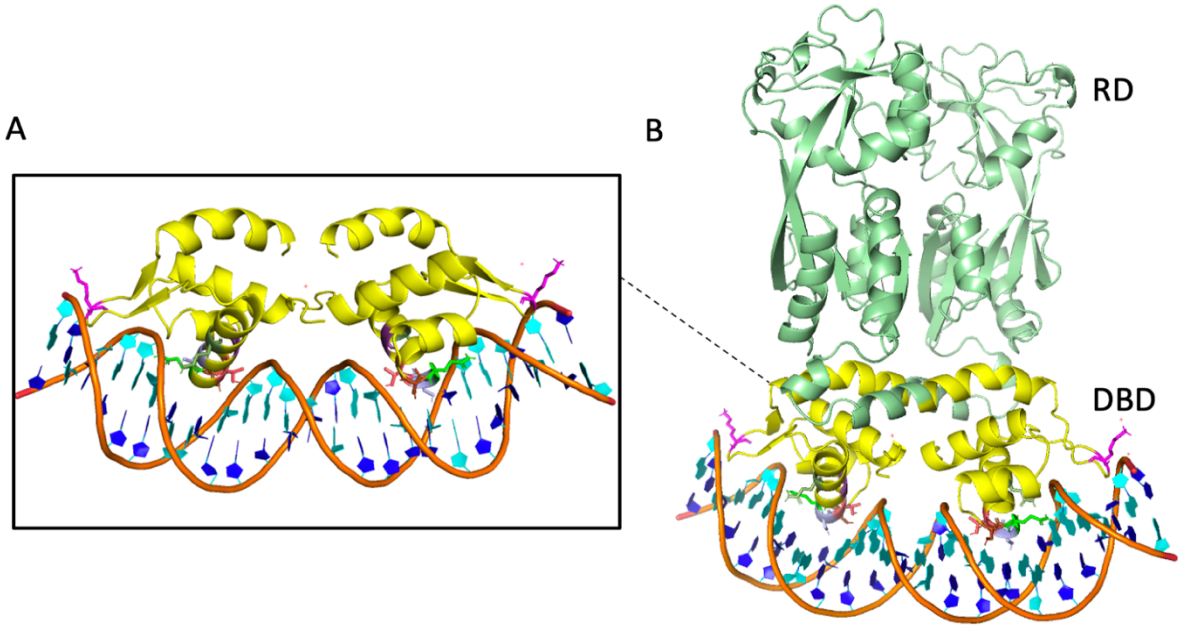


Figure 63. Predicted models for *A. fischeri* OxyR1 DBD and OxyR1 full-length. (A) *A. fischeri* OxyR1 DBD dimer. (B) *A. fischeri* OxyR1 full-length dimers. DNA interacting residues are shown as sticks.

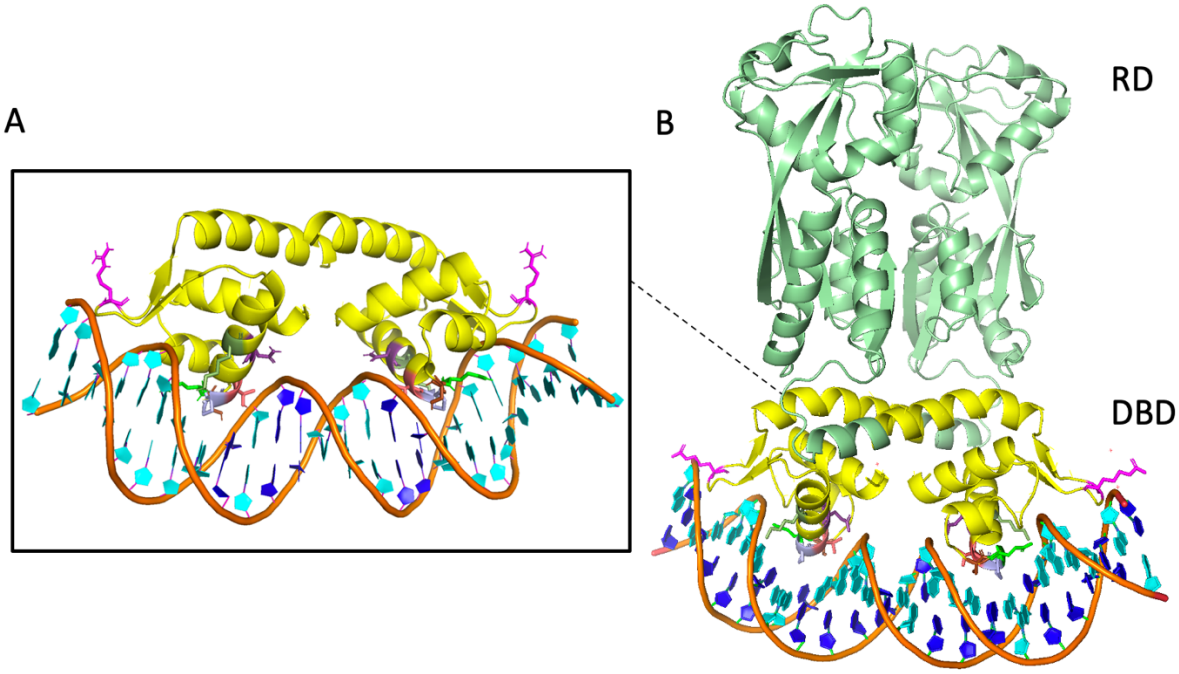


Figure 64. Predicted models of *A. fischeri* OxyR2 DBD and OxyR2 full-length. (A) *A. fischeri* OxyR2 DBD dimer. (B) *A. fischeri* OxyR2 full-length dimers. DNA interacting residues are shown as sticks.

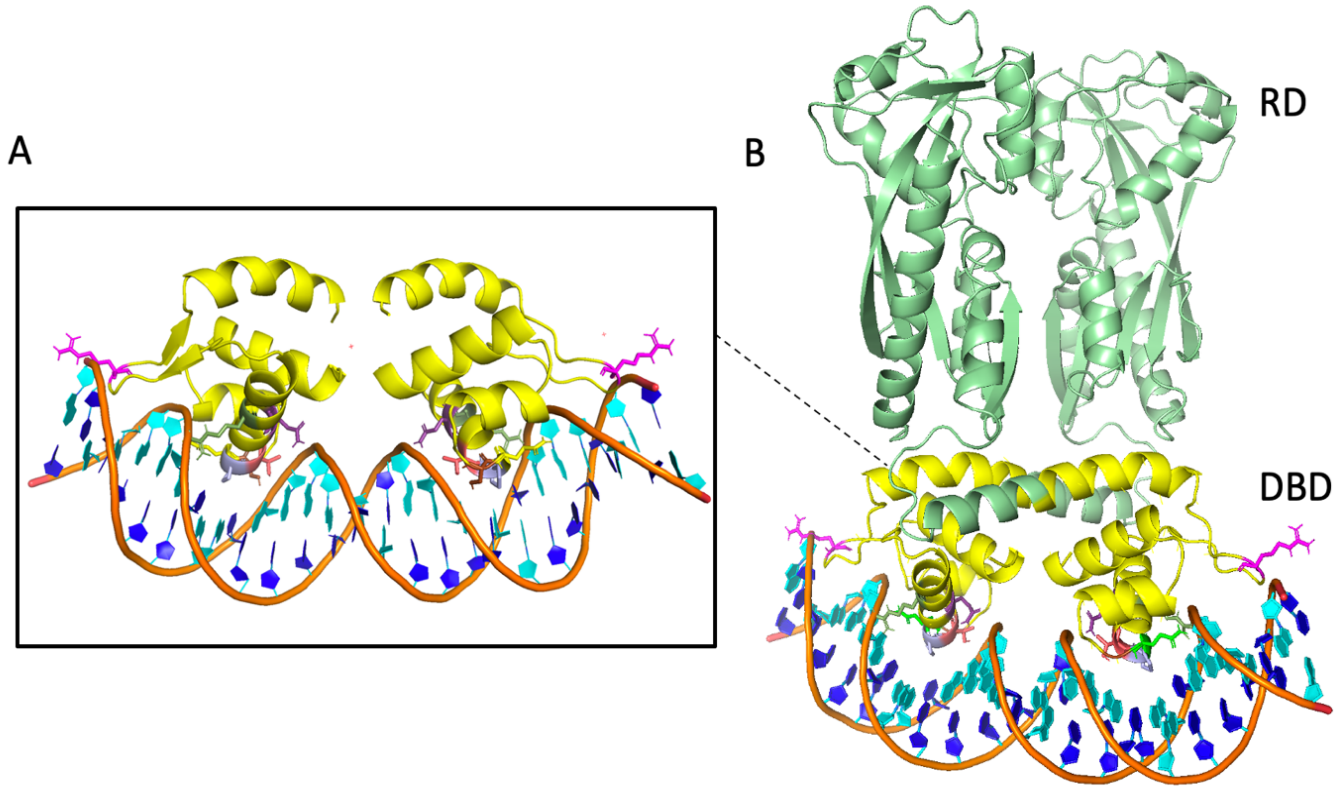


Figure 65. Predicted models of *E. coli* OxyR DBD and OxyR full-length. (A) *E. coli* OxyR DBD dimer. (B) *E. coli* OxyR full-length dimes. DNA interacting residues are shown as sticks.

4.4 DISCUSSION

The structure prediction of *A. fischeri* full-length and DBD OxyR1 and OxyR2, and *E. coli* OxyR were successfully modeled based on the *C. glutamicum* OxyR. Moreover, their modeled predictions were complexed with the cognate DNA of another transcriptional regulator commonly known as BenM (Alanazi et al., 2013). As expected by the conservation of sequence and structural similarities among LysR-type transcriptional regulators (LTTRs), critical structures were observed among all modeled predictions, such as the effector-binding regulatory domain, linker helix (LH), DBD turn-helix-turn motif, and DNA-binding interactions in the alpha-helix 3 (α3). LTTRS usually function as tetramers, with each subunit consists of an N-terminal DBD containing a helix-turn-helix motif connected by LH to an effector-binding RD at the C-terminus.

As an LTTR, *C. glutamicum* full-length OxyR has a common structural fold to that of its family, which includes an N-terminal domain comprised of β-strands sandwiched between two α-helices (subdomain 1), a bundle of β-strands and α-helices (subdomain 2), and a C-terminal α-helix and β-strand traversing both domains (Pedre et al., 2018). Although the conformation of *A. fischeri* OxyR1, OxyR2, and OxyRE differ significantly from that of *C. glutamicum* full-length OxyR published structure, all β-strand α-helix structures compromising the DBD and hydrogen peroxide (H_20_2) binding RD are showed to be conserved. As expected, based on the *C. glutamicum* full-length OxyR published crystal structure (Pedre et al., 2018), DBD α3-helix is shown to be conserved. It has been shown that this specific alpha helix takes part in major DNA-binding interactions between LTTRs and major grooves in the DNA. Based on our predictions, we were able to identify critically conserved DNA-interacting residues in the α3-helix of all of our protein models. In *A. fischeri* OxyR 1, full-length amino acids K56, Q32, S33, Q40, S31, T34, and G38 (Fig. 63) were interacting with DNA when complexed with the cognate DNA of a

BenM (Alanazi et al., 2013)**.** For *A. fischeri* OxyR2, full-length amino acids R53, Q29, P30, K37, S28, T31, Q35 (Fig. 64) were identified as DNA-binding residues, while in OxyRE, the identified residues were R53, Q29, P30, K37, S28, T31, Q35 (Fig. 65). In terms of the functionality of these residues, many have been shown to play critical roles in other LTTRs' ability to bind to DNA. For example, biochemical studies on mutated LTTRs have confirmed the importance of residues equivalent to T31 in *E. coli* OxyR and P30 in *P. putida* NahR DNA-binding (Kullik et al., 1995; M. A. Schell et al., 1990; Zaim & Kierzek, 2003). These OxyR protein prediction models provided structural and functional insight into the OxyR regulators expressed by two distinct organisms and the potential role of their DNA-binding residues in their activation.

5 CHAPTER 5

CONCLUSIONS

5.1 CONCLUSIONS

Data presented here indicate that OxyR1, OxyR2, and OxyRE fusion proteins were successfully cloned in both their full length and DNA binding domain versions. pET-32a(+) and pET-51b(+) clones are available for protein expression. pTXB1and pEU are available for alternate methods of protein expression. A catalog of the available plasmid minipreps were generated for the JARM-lab. Overexpression of OxyR1, OxyR2, and OxyRE fusion proteins in both their Full length and DNA binding domain version was achieved. In turn, the expression methodology for the pET-32a(+) constructs was effectively optimized and reproduced. The DNA complex interaction was examined with a known OxyRE consensus sequence, determined by *in vivo* ChIP-exo studies. Protein functionality for OxyR1, OxyR2, and OxyRE fusion protein in their FL and DBD versions were assessed and verified significant times through EMSA. All OxyR proteins examined to date, regardless of their regulatory effects, are capable of binding to promoter regions of their target genes in both reduced and oxidized forms (Wan et al., 2018).

DNA binding was visualized in a shift in the gel corresponding to the interaction of the DNA-complex. The DNA binding preferences of OxyR1, OxyR2, and OxyRE FL and DBD were examined through the use of SELEX-seq. In literature, the alignment of previously identified OxyR-regulated promoters from *Pseudomonas aeruginosa* has revealed four putative OxyR-binding tetranucleotide sequences (ATAG) spaced by heptanucleotides (Ochsner et al., 2000). The binding profile of the -expressed OxyR FL and DBDs logos, and local max enrichment sequences of the R3 of the SELEX-seq EMSA indicate the presence of similar motifs for OxyR1_FL: CTAT (1,039), ATAG (1,039); OxyR2_FL: ATAG (916), CTAT (916); OxyR2_DBD: ATAG (1,563), CTAT (1,563); OxyRE_DBD: ATAG (908), CTAT (908). Different from the previous list, OX1D: GTAT (2,392), ATAC (2,392) and OXEF: CTAC(T) (1,319), (A)TAG (1,319) have a similar motif but vary by just one position in the tetranucleotide sequence.

It is noteworthy that a portion of the palindrome-like sequence (ATAGATTNAATCTAT) was presented in OxyR1_FL, OxyR2_FL, OxyR2_DBD, OxyRE_DBD logos as a DNA motif enriched sequence. The motifs were enriched with A+T. It is common for OxyR to have extremely enriched A+T regions in their binding motifs (Wei et al., 2012). However, a more accurate determination of the DNA-binding motif of OxyR based on a genome-wide analysis should be performed to corroborate the identified motifs. A palindromic DNA sequence has been identified to which LTTRs are known to bind; this is often found to form part of an imperfect, dyadic region. The LTTR box was identified first in Rhizobium spp. as an interrupted palindrome with the sequence ATC-N9-GAT(Maddocks & Oyston, 2008).

Possible experiments to further the knowledge obtained from this research would be: (1) Validation of the obtained SELEX-seq enriched motifs will provide a solid foundation to build the consensus sequence of the transcription factor. (2) Determining Genome wide binding predictions using SELEX-seq data, to develop OxyR pathway. (3) RNA-seq transcriptional regulation under oxidative stress conditions in order to examine the overall expression caused by OxyR activation.

6 REFERENCES

Alanazi, A. M., Neidle, E. L., & Momany, C. (2013). The DNA-binding domain of BenM reveals the structural basis for the recognition of a T-N11-A sequence motif by LysR-type transcriptional regulators. *Acta Crystallographica Section D: Biological Crystallography*, *69*(10), 1995–2007. https://doi.org/10.1107/S0907444913017320

Battesti, A., Majdalani, N., & Gottesman, S. (2011). The RpoS-Mediated General Stress Response in Escherichia coli. *Annual Review of Microbiology*, *65*, 189–213. https://doi.org/10.1146/annurev-micro-090110-102946.The

Bayir, H. (2005). Reactive oxygen species. *Critical Care Medicine*, *33*(12), 498–501. https://doi.org/10.1097/01.CCM.0000186787.64500.12

Brivanlou, A. H., & Darnell, J. E. (2002). Transcription: Signal transduction and the control of gene expression. *Science*, *295*(5556), 813–818. https://doi.org/10.1126/science.1066355

Browning, D. F., & Busby, S. J. W. (2016). Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology*, *14*(10), 638–650. https://doi.org/10.1038/nrmicro.2016.103

Browning, D. F., Butala, M., & Busby, S. J. W. (2019). Bacterial Transcription Factors: Regulation by Pick "N" Mix. *Journal of Molecular Biology*, *431*(20), 4067–4077. https://doi.org/10.1016/j.jmb.2019.04.011

Caradonna, L., Amati, L., Magrone, T., Pellegrino, N. M., Jirillo, E., & Caccavo, D. (2000). Enteric bacteria, lipopolysaccharides and related cytokines in inflammatory bowel disease: Biological and clinical significance. *Journal of Endotoxin Research*, *6*(3), 205–214. https://doi.org/10.1177/09680519000060030101

Christman, M. F., Morgan, R. W., Jacobson, F. S., & Ames, B. N. (1985). Positive Control of a Regulon for Defenses against Oxidative Stress and Some Heat-Shock Proteins in Salmonella typhimurium. *Cell*, *41*(3), 753–762.

Costa, S., Almeida, A., Castro, A., & Domingues, L. (2014). Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: The novel Fh8 system. *Frontiers in Microbiology*, *5*(FEB), 1–20. https://doi.org/10.3389/fmicb.2014.00063

Dorman, C. J. (2019). DNA supercoiling and transcription in bacteria: a two-way street. *BMC Molecular and Cell Biology*, *20*(1), 26. https://doi.org/10.1186/s12860-019-0211-6

Dubbs, J. M., & Mongkolsuk, S. (2012). Peroxide-sensing transcriptional regulators in bacteria. *Journal of Bacteriology*, *194*(20), 5495–5503. https://doi.org/10.1128/JB.00304-12

Dwyer, D. J., Collins, J. J., & Walker, G. C. (2015). Unraveling the physiological complexities of antibiotic lethality. *Annual Review of Pharmacology and Toxicology*, *55*, 313–332.

https://doi.org/10.1146/annurev-pharmtox-010814-124712

Ferraz, R. A. C., Lopes, A. L. G., da Silva, J. A. F., Moreira, D. F. V., Ferreira, M. J. N., & de Almeida Coimbra, S. V. (2021). DNA–protein interaction studies: a historical and comparative analysis. *Plant Methods*, *17*(1), 1–21. https://doi.org/10.1186/s13007-021-00780-z

Fields, P. I., Swanson, R. V., Haidaris, C. G., & Heffron, F. (1986). Mutants of Salmonella typhimurium that cannot survive within the macrophage are avirulent. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(14), 5189–5193. https://doi.org/10.1073/pnas.83.14.5189

Garner, M. M., & Revzin, A. (1986). The use of gel electrophoresis to detect and study nucleic acid- protein interactions. *Trends in Biochemical Sciences*, *11*(10), 395–396. https://doi.org/10.1016/0968-0004(86)90149-0

Gourse, R. L., Ross, W., & Gaal, T. (2000). UPs and downs in bacterial transcription initiation: The role of the alpha subunit of RNA polymerase in promoter recognition. *Molecular Microbiology*, *37*(4), 687–695. https://doi.org/10.1046/j.1365-2958.2000.01972.x

Green, M. R., & Sambrook, J. (2012). *Molecular cloning: A laboratory manual*. *Cold Spring Harbor Laboratory Press*.

Green MR, S. J. (2021). Cloning and Transformation with Plasmid Vectors. *Cold Spring Harb Protoc*, *11*.

Györgypa, Z., & Kondorosi, A. (1991). Homology of the ligand-binding regions of Rhizobium symbiotic regulatory protein NodD and vertebrate nuclear receptors. *MGG Molecular & General Genetics*, *226*(1–2), 337–340. https://doi.org/10.1007/BF00273624

Halliwell, B. (1996). *Antioxidants in Human*. 33–50.

Haridhasapavalan, K. K., Ranjan, S. H., Bhattacharyya, S., & Thummer, R. P. (2021). Soluble expression, purification, and secondary structure determination of human MESP1 transcription factor. *Applied Microbiology and Biotechnology*, *105*(6), 2363–2376. https://doi.org/10.1007/s00253-021-11194-1

Hryckowian, A. J., & Welch, R. A. (2013). RpoS contributes to phagocyte oxidase-mediated stress resistance during urinary tract infection by Escherichia coli CFT073. *MBio*, *4*(1). https://doi.org/10.1128/mBio.00023-13

Huang, J., & Schell, M. A. (1991). In vivo interactions of the NahR transcriptional activator with its target sequences: Inducer-mediated changes resulting in transcription activation. *Journal of Biological Chemistry*, *266*(17), 10830–10838. https://doi.org/10.1016/s0021-9258(18)99094-0

Hur, J., Jawale, C., & Lee, J. H. (2012). Antimicrobial resistance of Salmonella isolated from

food animals: A review. *Food Research International*, *45*(2), 819–830. https://doi.org/10.1016/j.foodres.2011.05.014

Imlay, J. A., & Linn, S. (1987). Mutagenesis and stress responses induced in Escherichia coli by hydrogen peroxide. *Journal of Bacteriology*, *169*(7), 2967–2976. https://doi.org/10.1128/jb.169.7.2967-2976.1987

Imlay, James A. (2008). Cellular defenses against superoxide and hydrogen peroxide. *Revista Brasileira de Ortopedia*, *77*(1), 755–776. https://doi.org/10.1146/annurev.biochem.77.061606.161055.Cellular

Imlay, James A. (2015a). Diagnosing oxidative stress in bacteria: not as easy as you might think. *Current Opinion in Microbiology*, *24*, 124–131. https://doi.org/10.1016/j.mib.2015.01.004

Imlay, James A. (2015b). Transcription factors that defend bacteria against reactive oxygen species. *Annual Review of Microbiology*, *69*, 93–108. https://doi.org/10.1146/annurev-micro-091014-104322.Transcription

Kazmierczak, M. J., Wiedmann, M., & Boor, K. J. (2005). Alternative Sigma Factors and Their Roles in Bacterial Virulence. *Microbiology and Molecular Biology Reviews*, *69*(4), 527–543. https://doi.org/10.1128/mmbr.69.4.527-543.2005

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858. https://doi.org/10.1038/nprot.2015.053

Kelley, L. A., & Sternberg, M. J. E. (2009). Protein structure prediction on the web: A case study using the phyre server. *Nature Protocols*, *4*(3), 363–373. https://doi.org/10.1038/nprot.2009.2

Keseler, I. M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muniz-Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T., Kaipa, P., Spaulding, A., Pacheco, J., Latendresse, M., Fulcher, C., Sarker, M., Shearer, A. G., Mackie, A., Paulsen, I., … Karp, P. D. (2011). EcoCyc: A comprehensive database of Escherichia coli biology. *Nucleic Acids Research*, *39*(SUPPL. 1), 583–590. https://doi.org/10.1093/nar/gkq1143

Kevin Range, and D. M. Y. A. M. (2012). The structure, function, and evolution of proteins that bind DNA and RNA. *Bone*, *23*(1), 1–7. https://doi.org/10.1038/nrm3884.The

Kim, S., Bang, Y. J., Kim, D., Lim, J. G., Oh, M. H., & Choi, S. H. (2014a). Distinct characteristics of OxyR2, a new OxyR-type regulator, ensuring expression of Peroxiredoxin 2 detoxifying low levels of hydrogen peroxide in Vibrio vulnificus. *Molecular Microbiology*, *93*(5), 992–1009. https://doi.org/10.1111/mmi.12712

Kim, S., Bang, Y. J., Kim, D., Lim, J. G., Oh, M. H., & Choi, S. H. (2014b). Distinct characteristics of OxyR2, a new OxyR-type regulator, ensuring expression of Peroxiredoxin

2 detoxifying low levels of hydrogen peroxide in Vibrio vulnificus. *Molecular Microbiology*, *93*(5), 992–1009. https://doi.org/10.1111/mmi.12712

Kullik, I., Toledano, M. B., Tartaglia, L. A., & Storz, G. (1995). Mutational analysis of the redox-sensitive transcriptional regulator OxyR: Regions important for oxidation and transcriptional activation. *Journal of Bacteriology*, *177*(5), 1275–1284. https://doi.org/10.1128/jb.177.5.1275-1284.1995

Lance M. Hellman, M. G. F. (2007). Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein- Nucleic Acid Interactions. *Nature Protocols*, *2*(8), 1849–1861. https://doi.org/10.3969/j.issn.1672-7347.2012.03.002

Lederberg, J. (1952). Cell genetics and hereditary symbiosis. *Physiological Reviews*, *32*(4), 403–430. https://doi.org/10.1152/physrev.1952.32.4.403

Lee, D. J., Minchin, S. D., & Busby, S. J. W. (2012). Activating transcription in bacteria. *Annual Review of Microbiology*, *66*(June), 125–152. https://doi.org/10.1146/annurev-micro-092611-150012

Maddocks, S. E., & Oyston, P. C. F. (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology*, *154*(12), 3609–3623. https://doi.org/10.1099/mic.0.2008/022772-0

Minakami, R., & Sumimoto, H. (2006). Phagocytosis-coupled activation of the superoxide-producing phagocyte oxidase, a member of the NADPH oxidase (Nox) family. *International Journal of Hematology*, *84*(3), 193–198. https://doi.org/10.1532/IJH97.06133

Morgan, R. W., Christman, M. F., Jacobson, F. S., Storz, G., & Ames, B. N. (1986). Hydrogen peroxide-inducible proteins in Salmonella typhimurium overlap with heat shock and other stress proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(21), 8059–8063. https://doi.org/10.1073/pnas.83.21.8059

Nimrat Chatterjee, G. C. W. (2017). Mechanisms of DNA damage, repair and mutagenesis. *Environ Mol Mutagen*, *58*(5), 235–263. https://doi.org/10.1002/em.22087.Mechanisms

Nitta, K. R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J., Hens, K., Toivonen, J., Deplancke, B., Furlong, E. E. M., & Taipale, J. (2015). Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *ELife*, *2015*(4), 1–20. https://doi.org/10.7554/eLife.04837

Norsworthy, A. N., & Visick, K. L. (2013). Gimme shelter: how Vibrio fischeri successfully navigates an animal's multiple environments. *Frontiers in Microbiology*, *4*(November), 1–14. https://doi.org/10.3389/fmicb.2013.00356

Nyholm, S. V., & McFall-Ngai, M. J. (2004). The winnowing: Establishing the squid - Vibrios symbiosis. *Nature Reviews Microbiology*, *2*(8), 632–642.

https://doi.org/10.1038/nrmicro957

Nyholm, S. V., Stabb, E. V., Ruby, E. G., & McFall-Ngai, M. J. (2000). Establishment of an animal-bacterial association: Recruiting symbiotic vibrios from the environment. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(18), 10231–10235. https://doi.org/10.1073/pnas.97.18.10231

Nyholm, S. V., Stewart, J. J., Ruby, E. G., & McFall-Ngai, M. J. (2009). Recognition between symbiotic Vibrio fischeri and the haemocytes of Euprymna scolopes. *Environmental Microbiology*, *11*(2), 483–493. https://doi.org/10.1111/j.1462-2920.2008.01788.x

Ochsner, U. A., Vasil, M. L., Alsabbagh, E., Parvatiyar, K., & Hassett, D. J. (2000). Role of the Pseudomonas aeruginosa oxyR-recG operon in oxidative stress defense and DNA repair: OxyR-dependent regulation of katB-ankB, ahpB, and ahpC-ahpF. *Journal of Bacteriology*, *182*(16), 4533–4544. https://doi.org/10.1128/JB.182.16.4533-4544.2000

Ogawa, N., McFall, S. M., Klem, T. J., Miyashita, K., & Chakrabarty, A. M. (1999). Transcriptional activation of the chlorocatechol degradative genes of Ralstonia eutropha NH9. *Journal of Bacteriology*, *181*(21), 6697–6705. https://doi.org/10.1128/jb.181.21.6697-6705.1999

Pedre, B., Young, D., Charlier, D., Mourenza, Á., Rosado, L. A., Marcos-Pascual, L., Wahni, K., Martens, E., de la Rubia, A. G., Belousov, V. V., Mateos, L. M., & Messens, J. (2018). Structural snapshots of OxyR reveal the peroxidatic mechanism of H2O2 sensing. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(50), E11623–E11632. https://doi.org/10.1073/pnas.1807954115

Pomposiello, P. J., & Demple, B. (2001). Redox-operated genetic switches: The SoxR and OxyR transcription factors. *Trends in Biotechnology*, *19*(3), 109–114. https://doi.org/10.1016/S0167-7799(00)01542-0

Ramachandran, S., Kota, P., Ding, F., & Dokholyan, N. V. (2011). Automated minimization of steric clashes in protein structures. *Proteins: Structure, Function and Bioinformatics*, *79*(1), 261–270. https://doi.org/10.1002/prot.22879

Sai, A., Seshasayee, N., Sivaraman, K., & Luscombe, N. M. (2011). *A Handbook of Transcription Factors*. *52*. https://doi.org/10.1007/978-90-481-9069-0

Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñiz-Rascado, L., García-Sotelo, J. S., Weiss, V., Solano-Lira, H., Martínez-Flores, I., Medina-Rivera, A., Salgado-Osorio, G., Alquicira-Hernández, S., Alquicira-Hernández, K., López-Fuentes, A., Porrón-Sotelo, L., Huerta, A. M., Bonavides-Martínez, C., Balderas-Martínez, Y. I., Pannier, L., … Collado-Vides, J. (2013). RegulonDB v8.0: Omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Research*, *41*(D1), 203–213. https://doi.org/10.1093/nar/gks1201

Savinkova, L. K., Sharypova, E. B., & Kolchanov, N. A. (2021). In Vitro Methods Used to Study DNA–Protein Interactions. *Biology Bulletin Reviews*, *11*(4), 344–357. https://doi.org/10.1134/S2079086421040071

Schell, M. A., Brown, P. H., & Raju, S. (1990). Use of saturation mutagenesis to localize probable functional domains in the NahR protein, a LysR-type transcription activator. *Journal of Biological Chemistry*, *265*(7), 3844–3850. https://doi.org/10.1016/s0021-9258(19)39671-1

Schell, Mark A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annual Reviews Microbiolgy*, *47*(1), 597–626.

Seo, S. W., Kim, D., Szubin, R., & Palsson, B. O. (2015). Genome-wide Reconstruction of OxyR and SoxRS Transcriptional Regulatory Networks under Oxidative Stress in Escherichia coli K-12 MG1655. *Cell Reports*, *12*(8), 1289–1299. https://doi.org/10.1016/j.celrep.2015.07.043

Seshasayee, A. S. N., Sivaraman, K., & Luscombe, N. M. (2011). An overview of prokaryotic transcription factors: A summary of function and occurrence in bacterial genomes. *Sub-Cellular Biochemistry*, *52*, 7–23. https://doi.org/10.1007/978-90-481-9069-0_2

Seth, D., Hausladen, A., & Stamler, J. S. (2020). Anaerobic transcription by OxyR: A novel paradigm for nitrosative stress. *Antioxidants and Redox Signaling*, *32*(12), 803–816. https://doi.org/10.1089/ars.2019.7921

SiegelAggranoff, B. W. (1999). *Basic neurochemistry: Molecular, cellular, and medical aspects*.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H. J., & Mann, R. S. (2011). Cofactor binding evokes latent differences in DNA binding specificity between hox proteins. *Cell*, *147*(6), 1270–1282. https://doi.org/10.1016/j.cell.2011.10.053

Steffen, N. R., Murphy, S. D., Tolleri, L., Hatfield, G. W., & Lathrop, R. H. (2002). DNA sequence and structure: Direct and indirect recognition in protein-DNA binding. *Bioinformatics*, *18*(SUPPL. 1). https://doi.org/10.1093/bioinformatics/18.suppl_1.S22

Stormo, G. D., & Zhao, Y. (2010). Determining the specificity of protein-DNA interactions. *Nature Reviews Genetics*, *11*(11), 751–760. https://doi.org/10.1038/nrg2845

Storz, G., Christman, M. F., Sies, H., & Ames, B. N. (1987). Spontaneous mutagenesis and oxidative damage to DNA in Salmonella typhimurium. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(24), 8917–8921. https://doi.org/10.1073/pnas.84.24.8917

Storz, Gisela, Tartaglia, L. A., Farr, S. B., & Ames, B. N. (1990). Bacterial defenses against oxidative stress. *Trends in Genetics*, *6*(C), 363–368. https://doi.org/10.1016/0168-9525(90)90278-E

Tao, K., Makino, K., Yonei, S., Nakata, A., & Shinagawa, H. (1989). Molecular cloning and nucleotide sequencing of oxyR, the positive regulatory gene of a regulon for an adaptive response to oxidative stress in Escherichia coli: Homologies between OxyR protein and a family of bacterial activator proteins. *MGG Molecular & General Genetics*, *218*(3), 371–376. https://doi.org/10.1007/BF00332397

Tartaglia, L. A., Gimeno, C. J., Storz, G., & Ames, B. N. (1992). Multidegenerate DNA recognition by the OxyR transcriptional regulator. *Journal of Biological Chemistry*, *267*(3), 2038–2045. https://doi.org/10.1016/s0021-9258(18)46050-4

Todd R. Riley, Matthew Slattery, Namiko Abe, Chaitanya Rastogi, Richard Mann, and H. B. (2014). *SELEX-seq, a method for characterizing the complete repertoire of binding site preferences for transcription factor complexes*. *1196*, 255–278. https://doi.org/10.1007/978-1-4939-1242-1

Toledano, M. B., Kullik, I., Trinh, F., Baird, P. T., Schneider, T. D., & Storz, G. (1994). Redox-dependent shift of OxyR-DNA contacts along an extended DNA-binding site: A mechanism for differential promoter selection. *Cell*, *78*(5), 897–909. https://doi.org/10.1016/S0092-8674(94)90702-1

Tong Ihn, L., & Young, R. A. (2013). Transcriptional Regulation and its Misregulation in Disease Tong. *Cell*, *1152*(6), 1237–1251. https://doi.org/10.1016/j.cell.2013.02.014.Transcriptional

Torraca, V., Masud, S., Spaink, H. P., & Meijer, A. H. (2014). Macrophage-pathogen interactions in infectious diseases: New therapeutic insights from the zebrafish host model. *DMM Disease Models and Mechanisms*, *7*(7), 785–797. https://doi.org/10.1242/dmm.015594

Wan, F., Kong, L., & Gao, H. (2018). Defining the binding determinants of Shewanella oneidensis OxyR: Implications for the link between the contracted OxyR regulon and adaptation. *Journal of Biological Chemistry*, *293*(11), 4085–4096. https://doi.org/10.1074/jbc.RA117.001530

Wei, Q., Le Minh, P. N., Dötsch, A., Hildebrand, F., Panmanee, W., Elfarash, A., Schulz, S., Plaisance, S., Charlier, D., Hassett, D., Häussler, S., & Cornelis, P. (2012). Global regulation of gene expression by OxyR in an important human opportunistic pathogen. *Nucleic Acids Research*, *40*(10), 4320–4333. https://doi.org/10.1093/nar/gks017

Wheeler D., B. M. (2007). *Comparative Genomics* (Bergman NH (ed.)). Humana Press.

Wingfield, P. T. (2015). Overview of the purification of recombinant proteins. In *Current Protocols in Protein Science* (Vol. 2015, Issue April). https://doi.org/10.1002/0471140864.ps0601s80

Winquist, L., Rannug, U., Rannug, A., & Ramel, C. (1984). Protection from toxic and mutagenic

effects of H_2O_2 by catalase induction in Salmonella typhimurium. *Mutation Research Letters*, *141*(3–4), 145–147. https://doi.org/10.1016/0165-7992(84)90087-3

Zaim, J., & Kierzek, A. M. (2003). The structure of full-length LysR-type transcriptional regulators. Modeling of the full-length OxyR transcription factor dimer. *Nucleic Acids Research*, *31*(5), 1444–1454. https://doi.org/10.1093/nar/gkg234

Zheng, M., & Storz, G. (1998). *Activation of the OxyR Transcription Factor by Reversible Disulfide Bond Formation*. *279*(March), 11–14.