# Basic Leucine Zippers: Aureochromes Versus the Rest

Madhurima Khamaru[a], Anwesha Deb[a] and Devrani Mitra[a*]

[a]Department of Life Sciences, Presidency University, 86/1 College Street, Kolkata – 700073, WB, India

[*]Address of correspondence - devrani.dbs@presiuniv.ac.in

Running title – bZIP-DNA interactions: An *in silico* analysis

## Key words

Aureochrome; basic leucine zipper; protein-DNA interaction; phylogeny and (co)evolution; all-atom network

## Acknowledgement

MK and AD acknowledge University Grants Commission (UGC) and Department of Biotechnology (DBT), Government of India respectively, for their doctoral research fellowships. DM acknowledges funding from DBT, Government of India (BT/PR26435/BRB/10/1627/2017). The authors thank Soumen Roy, Bose Institute - Kolkata for important discussions on network analysis.

## Abstract

The photoreceptor cum transcription factors, Aureochromes, is present exclusively in marine photosynthetic algae. Co-existence of Light-Oxygen-Voltage (LOV) sensor and basic leucine zipper (bZIP) effector in aureochromes is unique -- no other photoreceptors barring aureochromes are known to possess bZIP as effector. The ability of aureochromes to participate in diverse biological activities, inverse topological arrangement make them a model protein not only to study light-dependent gene expression or transcriptional regulation but also as potential optogenetic scaffold. While detailed structural information is available for the LOV sensor, 3D structural information is absent for bZIP/bZIP+linker+LOV module. This study therefore aims for a thorough investigation of the bZIP domains from aureochromes and others, and their interaction with substrate DNA using tools from sequence/structural bioinformatics and network theory. With an in-depth comparison between 26 aureochrome subtypes and 147 plant/ophistokont bZIPs, we explore important residues at basic region and zipper - necessary for dimerization stability and DNA binding specificity. Perhaps the most notable finding is the unique histidine substitution at the DNA binding signature sequence of aureochromes, completely absent in any other bZIPs. Not only is this residue important for DNA binding, this can serve as a potential switch point in aureochrome/bZIP evolution. An all-atom network analysis on representative bZIP-DNA co-crystal structures, especially the measurement of eigenvector centrality, further adds importance to hydrophobic interactions in the zipper region to stabilize bZIP dimer and facilitate DNA binding.

## Introduction

Regulation of gene expression is essential for the survival of every life form. Transcription factors (TFs), upon binding with the cognate DNA, turn genes on/off at precise spatio-temporal resolution. Different classes of TFs exist. They differ on the basis of their DNA binding structural motifs and activity. Examples include basic leucine zipper (bZIP), basic helix-loop-helix (bHLH), Zinc-coordinating DNA-binding motifs (Zinc finger), helix-turn-helix (HTH), beta scaffold factors like STAT, MADS box, HMG box, TBP, RUNT, Beta barrel alpha helix etc. Among these, bZIPs comprise one of the largest group of eukaryotic transcription factors (Wang et al., 2018). The uniqueness of bZIPs lies in their conserved nature of structure and sequence specificity. They possess N-X7-R/K-X9 signature motif and binds mostly to the *cis* acting DNA elements of a promoter having an 'ACGT' core sequence (Wang et al., 2018) (Rodríguez-Martínez et al., 2017a). Interdigitation between the partner helices leads to the dimerization of bZIP — resulting in a superimposed coiled coil structure or 'zipper' (Bader & Vogt, 2006) (Sinden, 1994). This zipper-like structure is extended by the N-terminal helical segment, which can bind to the DNA major groove via basic amino acid residues (Glover & Harrison, 1995). Dimerization is essential for the recognition of substrate DNA and transcriptional regulation. The C-terminal zipper motif consists of heptads, which can be represented by a helical wheel. The residues are conventionally marked from 'a' to 'g' (Grigoryan & Keating, 2006), where leucines generally occupy every 'd' position, whereas the hydrophobic residues including leucine/isoleucine dominate 'a' (Bader & Vogt, 2006). However, presence of a conserved asparagine at 'a' position in the middle of zipper could result in destabilization of the oligomer (Harbury et al., 1993). The 'b', 'c' and 'f' positions are generally filled by polar residues. The charged residues at the 'e' and 'g' positions also aid in the zipper dimer stability (Deppmann et al., 2004) (Podust et al., 2001). Leucine residues, owing to their hydrophobic side chains, create 'knobs' which can fit into the 'holes' generated in between the side chains of amino acids present in adjacent helices (Bader & Vogt, 2006). At the other end, the basic amino acid residues complement well with the negatively charged phosphate backbone of DNA — this results in favorable electrostatic interactions. Besides, they can also participate in hydrogen bonding/polar interactions and can even engage in non-polar interactions with the portions of extended side chains. Besides phosphates, the nitrogenous base and deoxyribose sugar are also involved in hydrogen bonding, van der Waal's as well as stacking interactions. The affinity of bZIP TFs towards the cognate target DNA sequences is imparted by the sequence composition of its basic region as well as the length of the leucine zipper (Hakoshima, 2005). The dimerization partner is also crucial for its sequence specific DNA binding (Pogenberg et al., 2014). Their ability to bring in change in the binding specificity upon hetero-dimerization is indicative of this (Busch & Sassone-Corsi, 1990). The electrostatic attraction and repulsion forces, acting between the polar residues that flank the hydrophobic interaction surface of the zippers, are major deciding factors behind the formation of homo/heterodimers (Vinson et al., 1993). The oppositely charged residues located at 'g' and the following e' positions across the hydrophobic interface, form the inter-helical ion bridges. The interactions between d-d', a-a', g-e' are vital for the zipper formation and stabilization (Podust et al., 2001).

Widespread among the regulatory proteins of gene expression, the bZIPs essay crucial role in cellular proliferation and differentiation in eukaryotes. It is known to be the ancient DNA binding module found across three kingdoms of life (Hakoshima, 2005). Several groups of animal and plant bZIPs are well defined and characterized. Few homologs of bZIPs are also evident in fungi, algae, protists and even in bacteria. However, a unique blue light (BL) responsive photoreceptor with a bZIP transcription factor like activity has been recently

discovered (Takahashi et al., 2007) in the photosynthetic stramenopiles (Rayko et al., 2010). It was named 'aureochrome' after 'aureus' (meaning 'gold') — the place of its origin being the golden alga/xanthophyte *Vaucheria frigida* (Takahashi et al., 2007). The flavin binding Light-Oxygen-Voltage (LOV) BL sensor at the C-terminus, makes aureochromes different from any other conventional bZIPs. Given maximum penetration in the sea water, BL is predominant in the marine environment. Therefore, in photosynthetic marine algae, most of the physiological and developmental responses are BL dependent. Aureochromes are found exclusively in the members of algal phylum – heterokontophyta (**SI- 1.a, b**). This phylum consists of 12 families and is present under group-4 of the algal classification proposed by Robert Lee (Lee, 2018). Heterokontophyta typically lacks phototropins (Rayko et al., 2010) - one of the major LOV photoreceptors in land plants. However, aureochrome is known to aid in the optimization of photosynthesis (Mann et al., 2020) under varied light conditions. The broad spectrum functions of aureochromes include promotion of high light acclimation, photo protection in diatoms (Costa et al., 2013). Additionally, aureochromes perform organism specific diverse functionalities (**SI-1.c**) like - branching in *Vaucheria sp* (Takahashi et al., 2007), cell cycle progression in *Phaeodactylum sp* (Kroth et al., 2017), lipid accumulation in *Nannochloropsis sp.* (Huang et al., 2014) etc. Aureochromes' relevance in carrying out BL-mediated gene expression is not only important to study BL signaling in photosynthetic marine algae, they are potentially relevant for optogenetics as well. Dual presence of a LOV photoreceptor and a bZIP TF domain, albeit in a reverse arrangement contrary to commonly-sighted topology in LOV photoreceptors, empowers aureochrome to serve as a natural optogenetic tool (Hepp et. al., 2020) (Duan et al., 2018) (Jerng et al., 2021). In fact, the inverse domain topology is similar to effector-sensor arrangement in several eukaryotic TFs (Matiiv & Chekunova, 2018) — a feature suitable for developing synthetic sensory transcription factors (Boral et al., 2022) as well as optogenetic designs (Banerjee & Mitra, 2020). Using tools from structural bioinformatics, phylogenetic studies, and network theory, we therefore attempt a thorough sequence-structure-function characterization of this unique class of LOV photoreceptor-TF. We include all available varieties of aureochromes and compare them with other bZIP TFs.

## Materials and Methods

***Retrieval of bZIP sequences from different organisms*:** The amino acid sequences of Aureochromes and other representative bZIP proteins from plants and animals were collected from the UniProt (Apweiler et al., 2018) protein sequence database. BLASTP (Altschul et al., 1990) search was also performed against the database of non-redundant protein sequences using available bZIP sequences as queries. A few sequences, especially those of aureochromes from *Phaeodactylum sp*. were obtained from the available literature (Costa et al., 2013). Amino acid sequences of the 78 bZIP proteins from plants, 69 bZIPs from opisthokont species and 26 aureochrome bZIP sequences were shortlisted for further analyses. The respective domain boundaries of the bZIPs were finalized either from 'conserved domain' analysis of NCBI and Pfam, or from PSI-BLAST or domain BLAST of NCBI.

***Sequence comparative and molecular evolutionary analysis*:** A total of 173 bZIP sequences thus retrieved, were subjected to multiple sequence alignment using Clustal omega (Sievers et al., 2011) algorithm — first group-wise and then collectively. This gave us a preliminary understanding on the subtle differences in bZIP sequences that could lead to differences in DNA binding activity. The sequences were further considered for the construction of phylogenetic tree using MEGA X programme (Kumar et al., 2018). The corresponding

accession numbers of all the database entries of these bZIP proteins along with the names and source organisms are listed in the table (**SI-2**). For molecular evolutionary analysis, all 173 bZIP sequences were first aligned using MUSCLE algorithm, inbuilt in MEGAX. The most rational tree of all the selected bZIP proteins was drawn using neighbor joining (NJ) method, which is simple and fast. Poisson model, which considers equality of substitution rates among all the sites and assumes equal frequencies of amino acids at the time of correction of the multiple substitutions at the same sites, was followed as the substitution method. The phylogeny was tested through bootstrap analysis where the number of performances of bootstrap replications for each node was kept at 1000. The redundant and partial sequences were discarded. Only the reviewed and annotated sequences were considered. The guide tree was then represented and labeled by the online phylogenetic tree manipulation tool iTOLv5. The clades in red and green represent the maximum and minimum values of bootstrap respectively. Covariance study on aureochrome bZIPs were performed using GREMLIN (Kamisetty et al., 2013) server. Default parameters like 4 iterations, >75% query coverage and $10^{-10}$ /$10^{-4}$ E-value cut-off were chosen as appropriate for this study.

***Modeling and docking of aureochrome bZIPs - comparison with other bZIPs*:** With the availability of complete genome sequence and to cover the maximum range of aureochrome sub-types, we chose representatives mostly from *Ectocarpus*. The bZIP sequences, representative of all the five sub-types of aureochromes, were individually modeled using SWISS-MODEL (Waterhouse et al., 2018). bZIPs of *Ectocarpus siliculosus* aureochrome 1, 2, 3, 4 and *Saccharina japonica* aureochrome 5 were chosen as representatives. The models, selected on the basis of best Ramachandran parameters, were next energy minimized using Chimera. The energy minimized structures were next subjected to molecular docking analysis using NPDock (Tuszynska et al., 2015). The CRE (cAMP responsive element) DNA substrate containing 'aureo-box', TGACGT, (Banerjee et al., 2016) from CREB structure (PDB ID: 1DH3, (Schumacher et al., 2000) was used as the ligand in all five cases. Preliminary knowledge on protein-DNA interaction interface, as acquired from sequence-structure superposition with solved bZIP-DNA co-crystal structures, was used as input in docking. The energy parameters, interface interactions as well as buried surface area on the best scoring structures of aureochrome_bZIP-DNA complexes were calculated using PDBePISA (Evgeny, 2010) and Pymol. In order to perform a detailed comparative study between aureochromes and the rest, available bZIP proteins having co-crystal structures in PDB, were selected. We analyzed the hydrogen bonding and polar interactions at the basic region of bZIP-DNA interface using Chimera and Pymol. *In silico* mutagenesis was necessary in few cases, to understand the structure-function relations as well as evolutionary significance of the N-terminal basic region residues.

***Residue Interaction Graphs to understand bZIP-DNA interactions*:** The protein-DNA complex structure can be represented as residue interaction graph (RIG) or an all-atom network, where nodes (amino acids/nucleotide residues) are linked by edges within a given distance cut off. To investigate further on bZIP-DNA interactions, RIGs were first generated from representative bZIP co-crystal structures using the Biographs module of Biopython (Cock et al., 2009) and NetworkX (Hagberg et al., 2008). Eight structures representing different bZIP subgroups are - 2WTY (MAF), 1JNM (JUN with CRE), 1GD2 (PAP), 1YSA (GCN4), 1FOS (JUN-FOS) 1NWQ (CEBP), 1DH3 (CREB) and 2H7H (JUN with AP1). The RIGs were generated considering all atom-pair contacts within a given distance of 4.5Å. Centrality measures as simple as degree centrality, geodesic/shortest

path based closeness and betweenness centrality have been extremely helpful to identify influential nodes from residue-level interaction data (Deb et al., 2020). However, to capture all communication at the residue level, we calculated eigenvector centrality (EC) of the constructed RIGs. How well connected a node is, in terms of its connectivity with other well-connected nodes of a network, can be measured by EC. The weighted sum of the centralities of all nodes, which are connected to node *'i'* by an edge *'$A_{ij}$'*, is defined as EC of node *'i'* ($C_i$) (Negre et al., 2018)- $C_i = \varepsilon^{-1}\sum_{j=1}^{n} A_{ij}c_j$ , where '$c$' denotes the eigenvector associated with eigenvalue $\varepsilon$ of '$A$'.

The distribution of EC was plotted next. We studied the value of Z (X) = (X-<X>)/σ, where <X> and σ represent mean and standard deviation respectively. For subsequent analysis, we thence considered only those values of X which are associated with Z (X) > 1.5.

## Results

***Sequence retrieval, phylogeny estimation and course of bZIP evolution*:**

A total number of 173 sequences inclusive of plant, ophistokont and aureochrome bZIPs are considered for this study. The evolutionary position of all the selected representatives of the bZIP proteins from plants, ophistokonts and their phylogenetic relationship with aureochrome bZIPs could be inferred from the constructed phylogenetic tree. However, the divergence rooted deeply within the ancient nodes might not have been observed due to the relatively shorter length of bZIP domains. Nevertheless, three distinct lineages with a common ancestral origin are observed. In one lineage, BATF, PAR and E4BP4 are present. While the second lineage comprises animal bZIP families like CREB2, ATF2, CEBP, ATF3, FRA, FOS along with the viral bZIP family Zta. The remaining plant and animal bZIPs are found to cluster together with aureochrome bZIPs in the third lineage. However, the last common ancestor of the 3rd lineage further diverges into two distinct lines of evolution -- where one is solely represented by all the aureochromes. And, the other lineage further divides into two -- the ophistokont bZIPs e.g. PAP, GCN4, JUN, MAF, CREB3, CREB and ATF6, and all plant bZIPs like A, B, C, D, E, F, G, H, I and S. Close observations reveal that most aureochrome bZIPs are close to S-type plant bZIPs. The fungal bZIPs like PAP, GCN4 are also found to be closer to aureochrome bZIPs. However, the most appropriate sister clade or the immediate ancestor of aureochrome bZIPs could not be ascertained. In all-aureochrome lineage, most of aureochrome1 and 2, barring few exceptions, tend to align together in separate sub-lineages, while the remaining aureochrome members lose distinctiveness and cluster randomly [**Figure 1**]. To summarize, the phylogeny estimation of bZIPs from multiple genomes not only identifies the ancestor-descendant relationship among different members, but also highlights the evolutionary relationship between aureochromes and the rest.

***Decoding bZIP-DNA interactions - insights from modeled and docked aureochrome subtypes*:** Following *EsAureo1*, *EsAureo2*, *EsAureo3*, *EsAureo4* and *SjAureo5* homodimer modeling and energy minimization, each one was subjected to *in silico* docking experiments with the substrate CRE DNA sequence. The best scoring complex structures were selected and bZIP-DNA interfaces were characterized based on hydrogen bonding/salt bridge/van der Waals' interactions, interface area and energy. Complex structures of all five aureochrome representatives are depicted in **Figure 2**. The N-terminal basic regions being the key determinants behind DNA binding specificity of bZIPs, they were first screened for polar including hydrogen bonding interactions with the

DNA substrate. All interactions are color coded and presented in tabular format [**SI-3**]. In *Es*Aureo1-DNA complex, Arg'355 (NH2) forms hydrogen bond with dC'14 (OP2), while Arg'360 (NH1) interacts with dG'12 (OP1). Asn'359 interacts with N4 of dC'14 via hydrogen bonding, whereas Lys'371 (NZ) engages with dA'10 (O5). His'362 (NE2) interacts with dT'8 (O4). In *Es*Aureo2, Arg'190 (NH1), Arg'193 (NH2), Asn'194 (ND), and Arg'199 (NH1) form polar contacts with dC'14 (OP2), dG'5 (OP2), dT'8 (O4) and dG'12 (OP1) respectively. In *EsAureo3*, His'51 and Asn'48 are in close proximity to multiple deoxy-ribonucleotides liked dG'6, dA'15 and dT'13. Additionally, Arg'47 (NH1), Ser'55 (OG), Arg'54 (NE) and Arg'56 (NH2) make polar contacts with dC'7 (OP2), dA'10 (N7), dG'9 (OP2) and dG'9 (N7) respectively. The residues from the basic region of *Es*Aureo4 like Arg'55 (NH2), Asn'47 (ND2), Arg'53 (NH1), Arg'46 (NH1) and His'50 (NE2) also show similar interaction pattern like in other aureochromes. The respective interacting partner residues from the substrate are dG'9 (OP2), dG'16 (O6), dG'6 (OP2), dC'7 (OP2) and dT'8 (OP2). Surprisingly, *Es*Aureo5 does not contain bZIP domain – instead a LOV with huge N-terminal extension is present. Therefore, as a representative of Aureo5, we chose *Sj*Aureo5. In *Sj*Aureo5, His'153 (NE2) and Lys'161 (NZ) interact respectively with OP2 of dG'6 and dG'9. The basic region Arginines, Arg'162, Arg'151and Arg'158 interact with dC'11 (OP2), dT'13 (OP2) and dA'10 (OP2) respectively. *Perhaps the most important observation that we made in this study is the unique presence of Histidine in all aureochrome bZIPs, except Aureo2 from several stramenopiles*. Additionally, Histidine is found to be interacting with substrate DNA, albeit at *in silico* docking studies. Absence of Histidine at the homologous position of plant/animal bZIPs leads to further speculation about its possible evolutionary impact on the origin and evolution of aureochromes in the bZIP superfamily. We investigate this further at the latter part of our manuscript. To test the reliability of our docked structures, we conducted PDBePISA (Protein Interfaces, Surfaces and Assemblies) analysis and compared them with experimental data on bZIP-DNA complex [1DH3 and 1FOS representing homo- and heterodimer-DNA interactions respectively]. The emphasis was given on three parameters – the interface area, solvation energy and number of non-covalent interactions, precisely hydrogen bonding and salt bridge interactions. While the interface area between monomeric units of protein chains range from ~900Å$^2$-1200Å$^2$ (homodimer to heterodimer respectively), that between protein-DNA range from ~300-400Å$^2$. A negative value of gain in solvation free energy ($\Delta^i$G) upon interface formation in both protein-protein/DNA complex formations indicates positive affinity between them and hydrophobic interface. In case of all homodimeric aureochromes [*Es*Aureo1-4, *Sj*Aureo5], interface area is very similar to homodimeric 1DH3/1JNM, with an average of 915Å$^2$. Aureochrome-DNA interface area ranges in between ~350-450Å$^2$ with very similar number of hydrogen bonding/salt bridge interactions like in experimental structures. Further, the $\Delta^i$G for both protein-protein/DNA in aureochrome-DNA complexes are not only negative but the values are very similar to those of native co-crystal complexes. Characteristic of all bZIPs, in aureochrome-DNA complexes too, the first and last in signature sequence Asparagine and Arginine/Lysine-nucleotide interactions are conserved. Additionally, interactions from basic amino acids towards substrate binding are also prominent. Perhaps the most notable exception is the presence of DNA-binding Histidine in most aureochromes.

***Screening into the N-terminal basic region of the bZIPs*:** N(X)$_7$R/Kis the signature basic region motif, which is present in all bZIPs [**SI-4.a, b**]. While NRESA(X)$_2$SR constitutes the signature plant bZIP sequence, N(X)$_2$AA(X)$_3$R is the characteristic basic region sequence from most animal bZIPs, excluding CREB, CREB3 and ATF6. Throughout the plant kingdom a serine at the 4$^{th}$ position is conserved. And, in all aureochromes

barring Aureo2, residue at the 4th position is histidine. In Aureo2, histidine is substituted with a leucine. An in-depth investigation into all the aureochromes further revealed that the *Sj*Aureo2 despite its designation as type-2 aureochrome, possesses a histidine instead of leucine. This observation can be justified from the phyletic pattern [**Figure 1**], where *Sj*Aureo2 parts with Aureo1/3 separately from the Aureo2 cluster [**SI-4.c**]. Since neither histidine nor leucine is found at the 4th position in any other bZIP across three kingdoms of life, their conservation in aureochromes must bear an evolutionary significance. Further, the presence of threonine in Aureo2 bZIPs [$N(X)_2LA(X)_2TR$] is also striking considering the presence of serine at the homologous position in other aureochrome bZIPs. Perhaps this is worth mentioning that Aureo2 also possesses non-functional LOV (Takahashi et al., 2007). Out of eleven conserved amino acids in the flavin-binding pocket, nine are conserved in Aureo2. The reason behind such differences in Aureo2 is elusive – it is possible that Aureo2 LOV domain serves as an intermediate between PAS and functional LOV. *Further, considering its functional differences, sequence composition as well as distribution in phylogenetic tree with respect to the effector bZIP, it is quite possible that Aureo2 emerged first in the evolutionary timescale. As gain of function variation to acquire light-dependent DNA binding, Aureo1 might have emerged next. And, perhaps to preserve this very important light-dependent transcription, several aureochrome paralogs originated through gene duplication events.*

**Exploring the importance of Histidine substitution by *in silico* mutageneis, phylogeny and co-evolution analysis:** To further understand the role of histidine at the N-terminal basic region of the aureochrome bZIPs, *in-silico* mutagenesis in sequence and structure was conducted. Histidine was mutated to serine (as in plant bZIPs), alanine (as in animal bZIPs) and leucine (as in Aureo2) to find out the consequences in terms of evolution as well as DNA binding respectively. When leucine replaces His'362/His'153 in the docked *Es*Aureo1 / *Sj*Aureo5-DNA complex, no interactions with any of the nucleotides can be seen. His'51 in *Es*Aureo3, upon substitution by leucine, remains in close contact with dG'6 of the substrate DNA. Similar situation arises in *Es*Aureo4, where His'50 upon mutation to leucine remains in close contact with dG'6. However, no polar contacts could be observed in both these cases. His'362 of *Es*Aureo1 on being mutated to serine, comes in close contact with dC`7. While *Es*Aureo2's Leu'197 to serine mutation maintains dT'8 contact, His'51➔Ser in *Es*Aureo3, shows possible interaction with the N7 of DG'6. Incorporation of serine at the homologous histidine positions in *Es*Aureo4 and *Sj*Aureo5, fails to show any significant interaction. We must acknowledge that experimental data on 3D bZIP-DNA complex structures could be most reliable. However considering modeled structures, we could not find the indispensability of histidine at this specific position of the basic region of bZIPs.

Next, we wanted to check the evolutionary significance of histidine at this specific position. Thence, we re-constructed phylogenetic tree after altering the histidine/leucine position in aureochrome by serine (similar to plant bZIPs) and alanine (similar to animal bZIPs). Very interestingly, upon serine substitution, aureochromes shifted closer to the plant bZIPs sharing a common evolutionary lineage, while all animal bZIPs clustered in a different lineage [**Figure 3**]. Upon mutation to alanine, the bZIPs of aureochrome departed from the plant specific group and instead aligned with several opisthokont bZIPs (CEBP, PAP, JUN) and the plant bZIP subgroup-D [**Figure 3**]. All through we considered a bootstrap value of 1000 to obtain statistically confident data. And indeed, presence of histidine in aureochromes does make a difference in their evolutionary origin. Co-evolution studies conducted with EsAureo1 bZIP with an E-value cut-off $10^{-10}$ followed by Weblogo

analysis shows distinct presence of histidine in aureochrome subtypes [**Figure 3**]; whereas increasing the E-value to $10^{-4}$ ([**Figure 3**] thus accommodating more bZIP sequences for calculations) leads to complete disappearance of histidine and appearance of serine/alanine at the homologous position. *Histidine in aureochromes can therefore be designated as key class switching residue and is significant not only for aureochromes but in the history of bZIP evolution*.

Considering histidine's presence in all aureochromes except Aureo2 and complete absence in other bZIPs, we thought of exploring co-evolution status of histidine using GREMLIN (Kamisetty et al., 2013). Histidine does not co-evolve with any other residues present in LOV and this is expected given the fact that co-evolving residues are generally nearby in sequence. However, histidine co-evolves with serine, located four residues downstream in the bZIP signature sequence. And indeed in most Aureo2s, both histidine and serine are replaced by leucine and threonine respectively [**SI-4b**].

**Network analysis and bZIP dimer stability towards substrate DNA interaction:** Aureochrome bZIPs are unique in the sense that they are BL-regulated. Starting with photon absorption to signal generation at the C-terminal LOV domain to signal transmission via linker and zipper regions to DNA binding activity at the N-terminal basic region, the intervening amino acids play essential roles. And, as already discussed in the earlier sections, bZIP dimerization is the essential pre-requisite of DNA binding activity. We therefore constructed all-atom networks from eight co-crystallized bZIP-DNA complexes. While centrality measures are frequently used for network analysis, in this study, we particularly focused on Eigenvector Centrality (EC). EC that considers both the number of connections made by a node as well as the flow of information associated with them, EC has been successfully used to identify residues responsible for allosteric communication in enzymes (Negre et al., 2018). We did not consider aureochrome model structures for RIG analysis because structural artifacts could affect the result. We therefore calculated EC from networks constructed upon eight co-crystal structures of homo/heterodimeric bZIP-DNA complexes [PDB IDs: 1DH3, 1FOS, 1GD2, 1JNM, 1NWQ, 1YSA, 2H7H and 2WTY] [**SI-5**]. Residues with high EC values with $Z(X) > 1.5$ are only considered and marked in green [**Figure 4**]. In all cases, they are found to congregate at the zipper region, albeit at variable distances from the basic region. To show their relative locations with respect to DNA binding, the residues participating in the latter are highlighted in blue. Two-three residues from monomeric units showing highest EC values are further color coded in red to check any further significance. And in all cases, they occupy 'a' and 'd' positions of a heptad, highlighted in yellow [**SI-5**]. We found this result extremely interesting considering well-known hydrophobic interaction between residues at positions 'a' and 'd' towards leucine zipper dimerization and its stability.

## Discussion

The global network of TFs comprising interactions with specific substrate DNA or partner protein molecules regulates one of the most fundamental processes of life - gene expression. Besides the possibility of simple/auto regulation, a TF can participate in feed forward loops/feedback loops (FFLs/FBLs) (Alon, 2007) as homodimer or heterodimer, in complex and dynamic gene expression regulatory units. BL-responsive sensory TFs like aureochromes, the pivot of this study, come with an additional advantage. Using light to control TFs provides an unprecedented way to control/interrogate cellular behavior spatio-temporally. In view of numerous possibilities by which a sensory TF like aureochrome can bring about substantial changes into the gene expression profile,

both naturally and synthetically, we present **Figure 5.** The central part of this figure is inspired by the works of Alon, 2007, although on network motif analysis from a different perspective. **Figure 5** immediately tells us about the possible global impact of a sensory TF interaction network, how easy it would be to regulate gene expression with light – perhaps in one of the most non-invasive ways! Co-existence of LOV sensor and bZIP effector [**Figure 5**], distribution of effector-sensor similar to eukaryotic TFs, makes aureochrome a potential contributor in optogenetics. Other than aureochromes, LOV does not co-exist with bZIP in any other proteins known till date. Even from a natural perspective, aureochromes are unique given their exclusive presence in photosynthetic marine algae (Takahashi et al., 2007). Absence of phototropins in photosynthetic heterokontophyta makes aureochrome vital BL-responsive photoreceptors besides cryptochromes. The direct BL-dependent DNA binding ability of aureochrome without the need of an elaborate signaling cascade enables aureochrome to bring about faster changes in gene expression under fluctuating light conditions (Mann et al., 2020). While a substantial amount of literature including multiple papers on structural information (Mitra et al., 2012) (Banerjee et al., 2016) (Hepp et al., 2020) (Heintz & Schlichting, 2016), is found against the LOV sensors of aureochromes, structure based mechanistic insights is scarce for the effector bZIPs. Time-resolved solution scattering experiments (Heintz & Schlichting, 2016) along with H/D-exchange measurements did show BL-dependent conformational changes promoting LOV dimerization and enhancement of DNA binding activity. Dynamic light scattering combined with Fluorescence Resonance Energy Transfer and analytical size exclusion chromatography did establish intermolecular LOV-LOV and bZIP-bZIP interactions (Nakatani & Hisatomi, 2015) towards dimer stability and subsequent DNA binding activity. Upon BL-illumination, the light signal possibly transmits via the hydrophobic residues of core β-sheet and activates bZIP (Nakajima et al., 2021). This study aims for the holistic characterization of bZIP domains and compares aureochromes with the rest.

bZIPs constitute one of the major groups of TFs, found mostly in eukaryotes. The evolution of eukaryotes is positively correlated with the expansion and diversification of the bZIP superfamily (Jindrich & Degnan, 2016). Both the number of bZIP homologs as well as the complexity of bZIPs has increased over time in the process of evolution. Initially bZIPs were supposed to evolve independently in 4 eukaryotic lineages - opisthokonta, amoebozoa, planta, heterokonta (Jindrich & Degnan, 2016). There are 17 bZIP subfamilies within opisthokont, namely, YAP, MAF, XBP1, CNC, BATF, CREB, CREB-2, OASIS, PAR, E4BP4, ATF-2, ATF-4, ATF-6, FOS, JUN, C/EBP, PAP (Amoutzias et al., 2007) (Zhu et al., 2018) (Fujii et al., 2000). In contrast, bZIPs are found only in very few species of protists or amoebozoa. There are 10 distinct classes of bZIPs in plants (Jakoby et al., 2002) (Corrêa et al., 2008) whereas in heterokonta, the most abundant bZIP transcription factor is Aureochrome (Rayko et al., 2010). The BATF, JUN as well as FOS and ATF have originated from the primogenitor present in viruses. C/EBP and CNC-bZIP appear to be stemmed from marine bacteria (Zhu et al., 2018). Rest of the bZIP subfamilies are considered as the derivatives of the primogenitors present in viruses. Although bZIPs are supposed to have far reaching origin, it is still elusive when and where the ancestral gene has actually been incorporated. Although bZIPs are ubiquitous among eukaryotes, the number of the existing bZIP transcription factors in the different organisms varies greatly. For example, 17 bZIPs have been reported in *Saccharomyces cerevisiae*, 27 are there in *Drosophila* and 56 bZIPs have been observed in humans. In plants, a relatively large number of bZIPs are found to be present. Approximately 75 are found in *Arabidopsis*, 64 in cucumber, 49 in castor bean, 89 in rice, 92 in sorghum, 125 in maize, 89 in barley, 131 in soybean, 55 in grapevine and 96 in *Brachypodium*. The substrate DNA sequences differ subtly and bZIPs can bind with

specific substrates with great specificity. The bZIPs of the metazoans recognize 6 different consensus DNA-binding sites- c-AMP responsive element (CRE), TPA responsive element (TRE), CAAT box, PAR binding sites, Maf recognition element (MARE) and CRE-like elements (Amoutzias et al., 2007) (Rodríguez-Martínez et al., 2017b). Again, CRE, TRE and YAP (yeast activator protein) sites can be recognized by fungal bZIPs. Plant bZIPs recognize G-box (CACGTG), A-box (TACGTA), C-box (GACGTC), S-box (TGACGTG), D-box (TGACGTG) (Wang et al., 2018) etc. Hydrogen bonding interactions involving basic region and substrate DNA were analyzed for all available co-crystal structures of bZIP-DNA complexes [**SI-6**]. It is indeed interesting to note the multifarious specificities of bZIP proteins despite overall structure-sequence conservation, especially at the basic region. *This study therefore investigated how subtle changes at the residue level are manifested in the phyletic distribution of bZIP members, including the evolution of aureochromes and their subtypes*. Though close to few ophistokont and plant bZIPs, all aureochrome bZIPs cluster in a distinct lineage. Perhaps the major difference is the substitution of histidine (leucine in type-2 aureochromes) as the 4th amino acid in the $N(X)_7R/K$ signature motif. This single amino acid variation is striking and brings substantial alteration to the phyletic pattern. The presence of serine in replacement of this histidine/leucine results into an assembly of aureochrome bZIPs with plant bZIPs sharing a common evolutionary lineage. Again, mutation from histidine➔alanine aligns them with several animal bZIPs along with plant subgroup D. *In silico* docking studies reveal the potential of histidine in DNA binding activity, although this awaits experimental verification. This unique histidine further co-evolves with serine at the 8th position of the signature bZIP sequence. And indeed this is the case in Aureo2, which differs from the rest of the aureochrome sub-types. Despite the presence of leucine and threonine respectively at the 4th and 8th position of signature sequence, Aureo2 clubs with rest of the aureochrome sub-types. However, upon single substitution from leucine to alanine, Aureo2 shifts lineage to share with CEBPs, while retaining proximity with other aureochromes. Undoubtedly, histidine/leucine is the hallmark of aureochrome bZIPs. Considering the presence of non-functional sensor domain and close association with ophistokont bZIPs just upon single amino acid substitution, we speculate that Aureo2 might serve as the middle point of aureochrome evolution. During the course of evolution, possibly with the gain of function mutation, Aureo1 has acquired the ability of performing light dependent transcription and following gene duplication events several paralogs have emerged.

While the orderliness and sequence composition of the basic and fork regions play a crucial role in DNA binding specificity, sequence organization and stability of heptad repeats in zipper region is equally important. Even a subtle change in heptad sequence can impact dimerization specificity and stability, and the resulting manifestation could be significant. Choosing partner from the same or different group of bZIPs, respectively for homo- and hetero-dimerization can generate substantial functional diversity. As we know, the 3D structure of a protein can be considered as a network of nodes (amino acid residues) connected via edges (non-covalent interactions). And, the relative importance of the individual amino acid residues can be obtained by calculating different topological parameters of RIG. Therefore, apart from characterizing protein-protein/DNA interface in terms of bonding interactions, surface area and energy, we conducted all-atom network analysis. Centrality measures like degree, closeness and betweenness centrality effectively capture influential nodes in a network. In this study, we emphasized on measuring EC values considering its potential to detect nodes important for making connections and information flow. Interestingly, *residues with highest EC values are always at 'a' and 'd' positions of a heptad in all eight representative co-crystal structures*. Importance of hydrophobic residues at

the strategic positions necessary for dimerization and consequent enhancement of DNA binding activity has been shown for c-Fos mutants (Porte et al., 1997). Incorporation of isoleucine at the 2nd heptad not only promoted homodimerization but also conferred DNA binding specificity towards TRE substrate in c-Fos mutants. While EC calculations for aureochromes await determination of experim ental structures, *undoubtedly this measurement reiterates the importance of hydrophobic residues towards maintaining bZIP dimerization, stability and promoting DNA binding activity*.

## Figure Legends:

**Figure 1**: Phylogenetic tree of the bZIPs from plants, opisthokonts and aureochromes, where the latter makes distinct lineage. Different subgroups of bZIPs are represented by different colours. Bootstrap values are denoted by the circles of various sizes.

**Figure 2**: a) Docked bZIP structures of *Ectocarpus siliculosus* (*Es*) aureochrome1 (gray), aureochrome2 (pale green), aureochrome3 (yellow), aureochrome4 (pink) and *Saccharina japonica* (*Sj*) aureochrome5 (teal) bZIPs in complex with substrate DNA sequences containing 'aureo-box', TGACGT motif. This shows overall structural similarity in aureochrome-DNA complexes. Histidines, unique to aureochromes are involved in DNA binding activity. Pairwise interactions between an amino acid residue and the corresponding nucleotide are elaborated further in SI.

**Figure 3**: a) Re-construction of phylogenetic tree upon mutation of aureochromes' histidine to either serine (plant like) or alanine (animal like), showing class switching potential of histidine in aureochrome bZIP evolution; b) Weblogo results followed by co-evolution studies of *Es*Aureo1 at different E-value cut-offs show distinct presence of histidine in aureochrome bZIP homologues unlike other bZIPs.

**Figure 4**: Distinct location of bZIP residues involved in DNA binding (blue) and having high (green) eigenvector centrality (EC) values. Residues with highest EC values are highlighted in red. Incidentally they are all present at 'a' and 'd' positions of heptad repeats (detailed in SI), indicating importance of hydrophobic residues in bZIP dimerization and stability.

**Figure 5**: Impact of light-regulated transcription factor (TF) like blue-light responsive aureochromes on global TF interaction network. This figure shows several gene expression modules where aureochrome like sensory TF can participate as homo/heterodimer.

## Author contribution

DM conceptualized and supervised the work; MK carried out sequence/structure characterization (including curation of data and construction of tables, phylogeny analysis, structure analysis); AD and DM carried out modeling/docking studies; DM performed network analysis; MK and DM generated figures and wrote the manuscript.

## References

Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat Rev Genet, 2007 8:6*, *8*(6), 450–461. https://doi.org/10.1038/nrg2102

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amoutzias, G. D., Veron, A. S., Weiner, J., Robinson-Rechavi, M., Bornberg-Bauer, E., Oliver, S. G., & Robertson, D. L. (2007). One billion years of bZIP transcription factor evolution: Conservation and change in dimerization and DNA-binding site specificity. *Mol Biol Evol*, *24*(3), 827–835. https://doi.org/10.1093/molbev/msl211

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., & Yeh, L. S. L. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Res*, *46*(5), 2699–2699. https://doi.org/10.1093/NAR/GKY092

Bader, A. G., & Vogt, P. K. (2006a). Leucine Zipper Transcription Factors: bZIP Proteins. In *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*. https://doi.org/10.1007/3-540-29623-9_2180

Bader, A. G., & Vogt, P. K. (2006b). Leucine Zipper Transcription Factors: bZIP Proteins. In *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine* (pp. 964–967). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-29623-9_2180

Banerjee, A., Herman, E., Serif, M., Maestre-Reyna, M., Hepp, S., Pokorny, R., Kroth, P. G., Essen, L.-O., & Kottke, T. (2016). Allosteric communication between DNA-binding and light-responsive domains of diatom class I aureochromes. *Nucleic Acids Res*, *44*(12), 5957–5970. https://doi.org/10.1093/nar/gkw420

Banerjee, S. & Mitra, D. (2020) Structural Basis of Design and Engineering for Advanced Plant Optogenetics. *Trends Plant Sci*. 25(1): 35-65. https://doi.org/10.1016/j.tplants.2019.10.002

Boral, A., Khamaru, M., & Mitra, D. (2022). Designing synthetic transcription factors: A structural perspective. *Avd Protein Chem Struct Biol*. 130: 245-287. https://doi.org/10.1016/bs.apcsb.2021.12.003

Busch, S. J., & Sassone-Corsi, P. (1990). Dimers, leucine zippers and DNA-binding domains. *Trends Genet*. https://doi.org/10.1016/0168-9525(90)90071-D

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & De Hoon, M. J. L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics (Oxford, England)*, *25*(11), 1422–1423. https://doi.org/10.1093/BIOINFORMATICS/BTP163

Corrêa, L. G. G., Riaño-Pachón, D. M., Schrago, C. G., Vicentini dos Santos, R., Mueller-Roeber, B., & Vincentz, M. (2008). The Role of bZIP Transcription Factors in Green Plant Evolution: Adaptive Features Emerging from Four Founder Genes. *PLoS ONE*, *3*(8), e2944. https://doi.org/10.1371/journal.pone.0002944

Costa, B., Sachse, M., Jungandreas, A., Bartulos, C. R., & Gruber, A. (2013). Aureochrome 1a Is Involved in the Photoacclimation of the Diatom Phaeodactylum tricornutum. *PLoS ONE*, *8*(9). https://doi.org/10.1371/journal.pone.0074451

Deb, A., Grewal, R. K., Roy, S., & Mitra, D. (2020). Residue interaction dynamics in Vaucheria aureochrome1 light-oxygen-voltage: Bridging theory and experiments. *Proteins*, *88*(12), 1660–1674. https://doi.org/10.1002/PROT.25984

Deppmann, C. D., Acharya, A., Rishi, V., Wobbes, B., Smeekens, S., Taparowsky, E. J., & Vinson, C. (2004). Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana : a comparison to Homo sapiens B-ZIP motifs. *Nucleic Acids Res*, *32*(11), 3435–3445. https://doi.org/10.1093/NAR/GKH653

Duan, L., Hope, J. M., Guo, S., Ong, Q., François, A., Kaplan, L., Scherrer, G., & Cui, B. (2018). Optical Activation of TrkA Signaling. *ACS Synth Biol*, *7*(7), 1685–1693. https://doi.org/10.1021/ACSSYNBIO.8B00126

Evgeny, K. (2010). Crystal contacts as nature's docking solutions. *J Comp Chem*, *31*(1), 133–143. https://doi.org/10.1002/JCC.21303

Fujii, Y., Shimizu, T., Toda, T., Yanagida, M., & Hakoshima, T. (2000). Structural basis for the diversity of DNA recognition by bZIP transcription factors. *Nat Struct Mol Biol.,* 7: 889-893. https://doi.org/10.1038/82822

Glover, J. N. M., & Harrison, S. C. (1995). Crystal structure of the heterodimeric bZIP transcription factor c-Fos–c-Jun bound to DNA. *Nature 373*(6511), 257–261. https://doi.org/10.1038/373257a0

Grigoryan, G., & Keating, A. E. (2006). Structure-based Prediction of bZIP Partnering Specificity. *J Mol Biol.*, *355*(5), 1125–1142. https://doi.org/10.1016/j.jmb.2005.11.036

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. *7th Python in Science Conference (SciPy 2008)*, 11–15. https://www.bibsonomy.org/bibtex/272096c6553d7057409fe78ca698eb332/mschuber

Hakoshima, T. (2005). Leucine Zippers. In *Encyclopedia of Life Sciences*. https://doi.org/10.1038/npg.els.0005049

Harbury, P. B., Zhang, T., Kim, P. S., & Alber, T. (1993). A switch between two-, three-, and four-stranded

coiled coils in GCN4 leucine zipper mutants. *Science (New York, N.Y.)*, *262*(5138), 1401–1407. https://doi.org/10.1126/SCIENCE.8248779

Hepp, S., Trauth, J., Hasenjäger, S., Bezold, F., Essen, L. O., & Taxis, C. (2020). An Optogenetic Tool for Induced Protein Stabilization Based on the Phaeodactylum tricornutum Aureochrome 1a Light–Oxygen–Voltage Domain. *J Mol Biol*. https://doi.org/10.1016/j.jmb.2020.02.019

Heintz, U. & Schlichting, I. (2016). Blue light-induced LOV domain dimerization enhances the affinity of Aureochrome 1a for its target DNA sequence. *eLife*, 5: e11860. https://doi.org/10.7554/eLife.11860

Huang, Y., Wang, L., Zheng, M., Zheng, L., Tong, Y., & Li, Y. (2014). Overexpression of NgAUREO1, the gene coding for aurechrome 1 from Nannochloropsis gaditana, into *Saccharomyces cerevisiae* leads to a 1.6-fold increase in lipid accumulation. *Biotechnol. Lett.*, *36*(3), 575–579. https://doi.org/10.1007/s10529-013-1386-0

Ishikawa, M., Takahashi, F., Nozaki, H., Nagasato, C., Motomura, T., & Kataoka, H. (2009). Distribution and phylogeny of the blue light receptors aureochromes in eukaryotes. *Planta*, *230*, 543–552. https://doi.org/10.1007/s00425-009-0967-6

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., & Parcy, F. (2002). bZIP transcription factors in Arabidopsis. *Trends Plant Sci*, *7*(3), 106–111. https://doi.org/10.1016/S1360-1385(01)02223-3

Jerng, H. H., Patel, J. M., Khan, T. A., Arenkiel, B. R., & Pfaffinger, P. J. (2021). Light-regulated voltage-gated potassium channels for acute interrogation of channel function in neurons and behavior. *PLOS ONE*, *16*(3), e0248688. https://doi.org/10.1371/JOURNAL.PONE.0248688

Jindrich, K., & Degnan, B. M. (2016). The diversification of the basic leucine zipper family in eukaryotes correlates with the evolution of multicellularity Genome evolution and evolutionary systems biology. *BMC Evol Biol*, *16*(1), 1–12. https://doi.org/10.1186/s12862-016-0598-z

Kamisetty, H., Ovchinnikov, S., & Baker, D. (2013). Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proc Natl Acad Sci., USA*, *110*(39), 15674–15679. https://doi.org/10.1073/PNAS.1314045110/SUPPL_FILE/SD05.XLSX

Kroth, P. G., Wilhelm, C., & Kottke, T. (2017). An update on aureochromes_ Phylogeny – mechanism – function. *J Plant Physiol.*, *217*, 20–26. https://doi.org/10.1016/j.jplph.2017.06.010

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*, *35*(6), 1547–1549. https://doi.org/10.1093/MOLBEV/MSY096

Lee, R. E. (2018). Phycology. *Phycology*. https://doi.org/10.1017/9781316407219

Mann, M., Serif, M., Wrobel, T., Eisenhut, M., Madhuri, S., Flachbart, S., Weber, A. P. M., Lepetit, B., Wilhelm, C., & Kroth, P. G. (2020). The Aureochrome Photoreceptor PtAUREO1a Is a Highly Effective Blue Light Switch in Diatoms. *IScience*, *23*(11), 101730. https://doi.org/10.1016/j.isci.2020.101730

Matiiv, A. B., & Chekunova, E. M. (2018). Aureochromes – Blue Light Receptors. *Biochemistry (Moscow)*, *83*(6), 662–673. https://doi.org/10.1134/S0006297918060044

Mitra, D., Yang, X., & Moffat, K. (2012) Crystal structures of Aureochrome1 LOV suggest new design strategies for optogenetics. *Structure*. 20(4): 698-706. https://doi.org/ 10.1016/j.str.2012.02.016

Nakajima, H., Kobayashi, I., Adachi, Y., & Hisatomi, O. (2021) Transmission of light signals from the light-oxygen-voltage core via the hydrophobic region of the β-sheet surface in aureochrome-1. *Sci Rep*., 11:11995. https://doi.org/ 10.1038/s41598-021-91497-5

Nakatani, Y., & Hisatomi, O. (2015). Molecular Mechanism of Photozipper, a Light-Regulated Dimerizing Module Consisting of the bZIP and LOV Domains of Aureochrome-1. Biochemistry, 54(21): 3302-2213. https://doi.org/10.1021/acs.biochem.5b00320

Negre, C. F. A., Morzan, U. N., Hendrickson, H. P., Pal, R., Lisi, G. P., Patrick Loria, J., Rivalta, I., Ho, J., & Batista, V. S. (2018). Eigenvector centrality for characterization of protein allosteric pathways. *Proc Natl Acad Sci., USA*, *115*(52), E12201–E12208. https://doi.org/10.1073/PNAS.1810452115

Podust, L. M., Krezel, A. M., & Kim, Y. (2001). Crystal structure of the CCAAT box/enhancer-binding protein beta activating transcription factor-4 basic leucine zipper heterodimer in the absence of DNA. *J Biol Chem*, *276*(1), 505–513. https://doi.org/10.1074/JBC.M005594200

Pogenberg, V., Consani Textor, L., Vanhille, L., Holton, S. J., Sieweke, M. H., & Wilmanns, M. (2014). Design of a bZip Transcription Factor with Homo/Heterodimer-Induced DNA-Binding Preference. *Structure*, *22*(3), 466–477. https://doi.org/10.1016/j.str.2013.12.017

Porte, D., Oertel-Buchheit, P., John, M., Granger-Schnarr, M., & Schnarr, M. (1997). DNA binding and transactivation properties of Fos variants with homodimerization capacity. *Nucleic Acids Res*., 25(15): 3026-3033

Rayko, E., Maumus, F., Maheswari, U., Jabbari, K., & Bowler, C. (2010). Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol*, *188*(1), 52–66. https://doi.org/10.1111/j.1469-8137.2010.03371.x

Rodríguez-Martínez, J. A., Reinke, A. W., Bhimsaria, D., Keating, A. E., & Ansari, A. Z. (2017). Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. *ELife*. https://doi.org/10.7554/eLife.19272

Schumacher, M. A., Goodman, R. H., & Brennan, R. G. (2000). *The Structure of a CREB bZIPSomatostatin CRE Complex Reveals the Basis for Selective Dimerization and Divalent Cation-enhanced DNA Binding*. *J Biol Chem*, 275(45): 35242-35247. https://doi.org/10.1074/jbc.M007293200

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, *7*(1), 539. https://doi.org/10.1038/MSB.2011.75

Sinden, R. R. (1994). DNA–Protein Interactions. *DNA Structure and Function*, 287–325. https://doi.org/10.1016/B978-0-08-057173-7.50013-4

Takahashi, F., Yamagata, D., Ishikawa, M., Fukamatsu, Y., Ogura, Y., Kasahara, M., Kiyosue, T., Kikuyama, M., Wada, M., & Kataoka, H. (2007). AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles. *Proc Natl Acad Sci*, *USA, 104*(49), 19625–19630. https://doi.org/10.1073/pnas.0707692104

Tuszynska, I., Magnus, M., Jonak, K., Dawson, W., & Bujnicki, J. M. (2015). NPDock: a web server for protein–nucleic acid docking. *Nucleic Acids Res*, *43*(Web Server issue), W425. https://doi.org/10.1093/NAR/GKV493

Vinson, C. R., Hai, T., & Boyd, S. M. (1993). Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: Prediction and rational design. *Genes Dev.*, *7*(6), 1047–1058. https://doi.org/10.1101/gad.7.6.1047

Wang, Y., Zhang, Y., Zhou, R., Dossa, K., Yu, J., Li, D., Liu, A., Mmadi, M. A., Zhang, X., & You, J. (2018). Identification and characterization of the bZIP transcription factor family and its expression in response to abiotic stresses in sesame. *PLOS ONE*, *13*(7), e0200850. https://doi.org/10.1371/journal.pone.0200850

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*, *46*(W1), W296–W303. https://doi.org/10.1093/NAR/GKY427

Zhu, Y.-P., Wang, M., Xiang, Y., Qiu, L., Hu, S., Zhang, Z., Mattjus, P., Zhu, X., & Zhang, Y. (2018). Nach Is a Novel Subgroup at an Early Evolutionary Stage of the CNC-bZIP Subfamily Transcription Factors from the Marine Bacteria to Humans. *Int J Mol Sci*., 19(10): 2927. https://doi.org/10.3390/ijms19102927

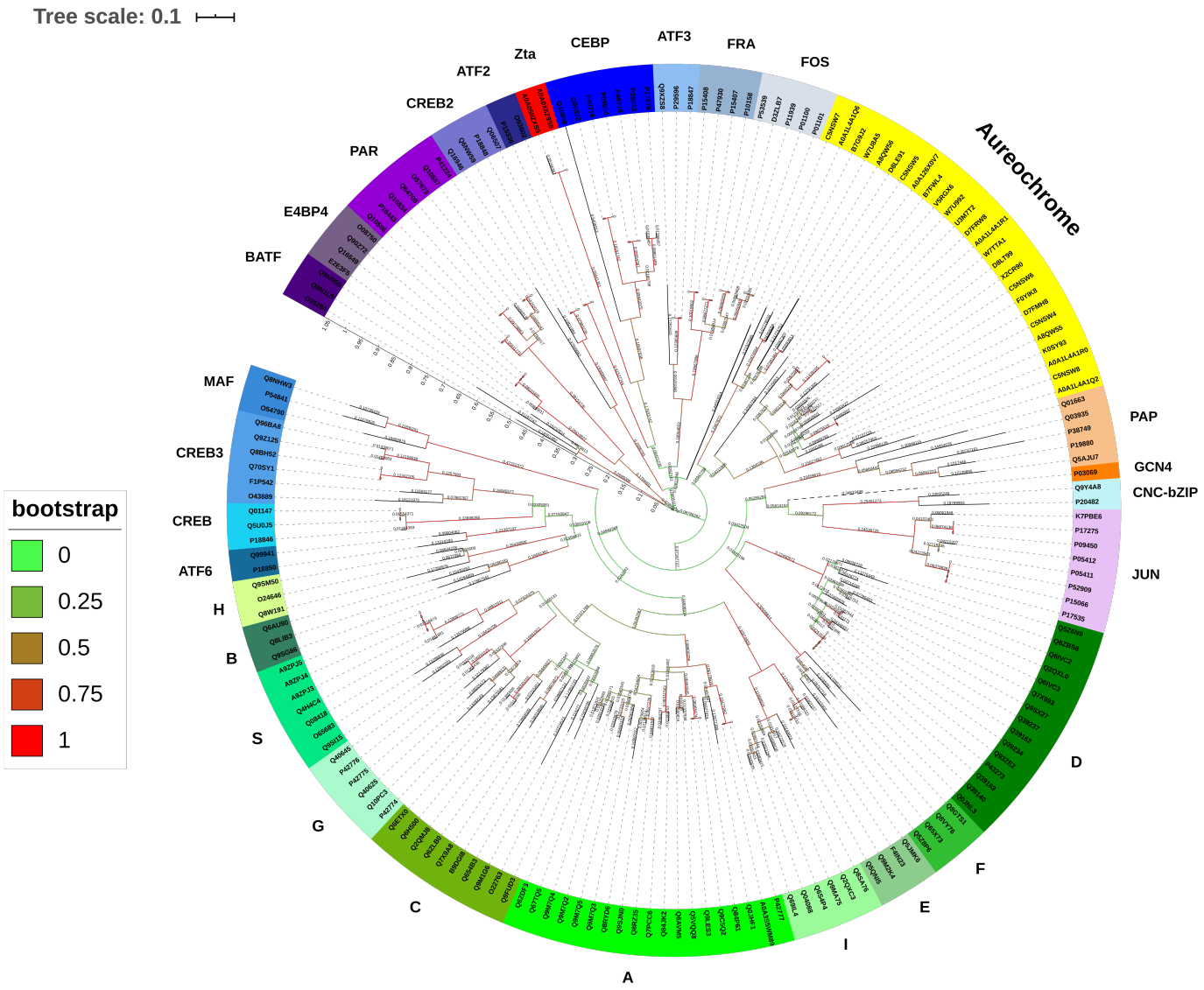


Figure 1

**Colored ranges**

- Aureochrome bZIP
- GCN4
- JUN
- PAP
- FOS
- FRA
- ATF3
- CEBP
- ATF2
- CREB2
- PAR
- BATF
- E4BP4
- ATF6
- CREB3
- CREB
- MAF
- CNC-bZIP
- D-type of Plants
- F-type of Plants
- E-type of Plants
- I-type of Plants
- A-type of Plants
- C-type of Plants
- S-type of Plants
- B-type of Plants
- H-type of Plants
- G-type of Plants
- Viral Zta

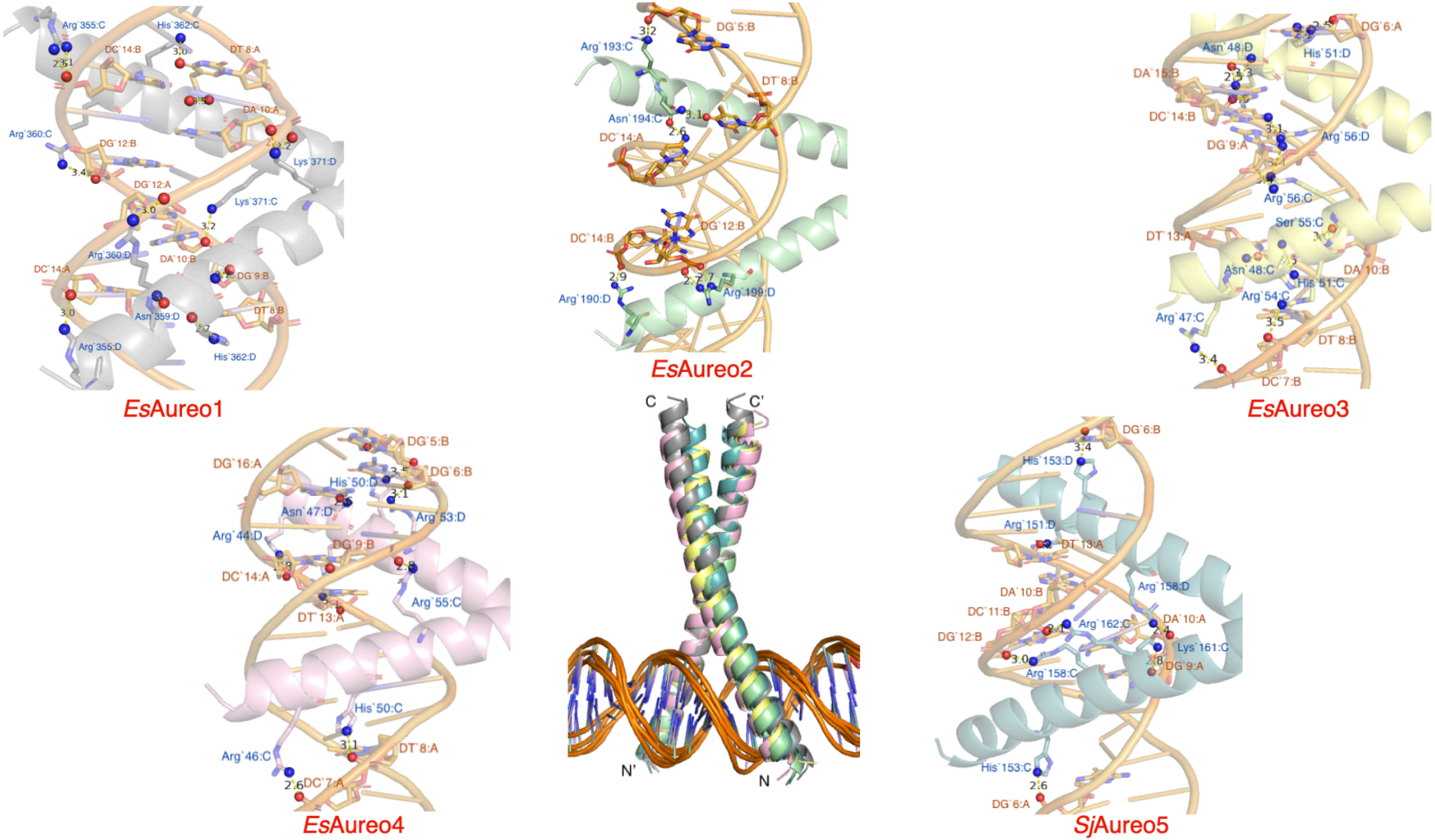
Arg`355:C
His`362:C
DC`14:B
DT`8:A
DA`10:A
Arg`360:C
DG`12:B
Lys`371:D
DG`12:A
Lys`371:C
Arg`360:D
DA`10:B
DC`14:A
DG`9:B
DT`8:B
Asn`359:D
Arg`355:D
His`362:D
EsAureo1
DG`5:B
Arg`193:C
DT`8:B
Asn`194:C
DC`14:A
DC`14:B
DG`12:B
Arg`190:D
Arg`199:D
EsAureo2
DG`6:A
Asn`48:D
His`51:D
DA`15:B
DC`14:B
DG`9:A
Arg`56:D
Arg`56:C
Ser`55:C
DT`13:A
Asn`48:C
DA`10:B
His`51:C
Arg`54:C
Arg`47:C
DT`8:B
DC`7:B
EsAureo3
DG`16:A
DG`5:B
His`50:D
DG`6:B
Asn`47:D
Arg`53:D
Arg`44:D
DG`9:B
DC`14:A
Arg`55:C
DT`13:A
His`50:C
Arg`46:C
DT`8:A
DC`7:A
EsAureo4
C
C'
N'
N
DG`6:B
His`153:D
Arg`151:D
DT`13:A
Arg`158:D
DA`10:B
DC`11:B
Arg`162:C
DA`10:A
DG`12:B
Lys`161:C
Arg`158:C
DG`9:A
His`153:C
DG`6:A
SjAureo5

Figure 2

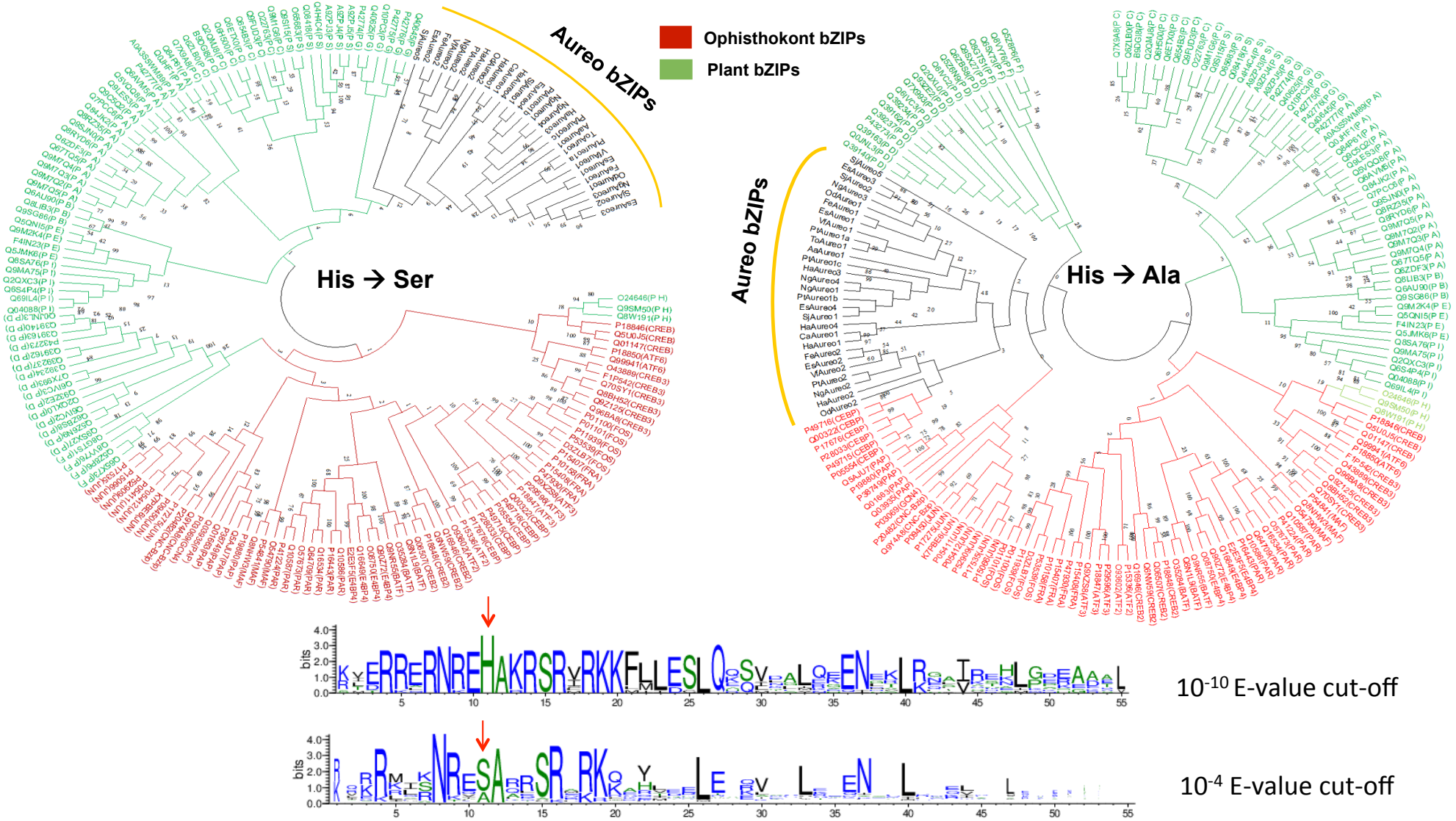
Ophisthokont bZIPs
Plant bZIPs
Aureo bZIPs
Aureo bZIPs
His → Ser
His → Ala
bits
10-10 E-value cut-off
10-4 E-value cut-off

Figure 3

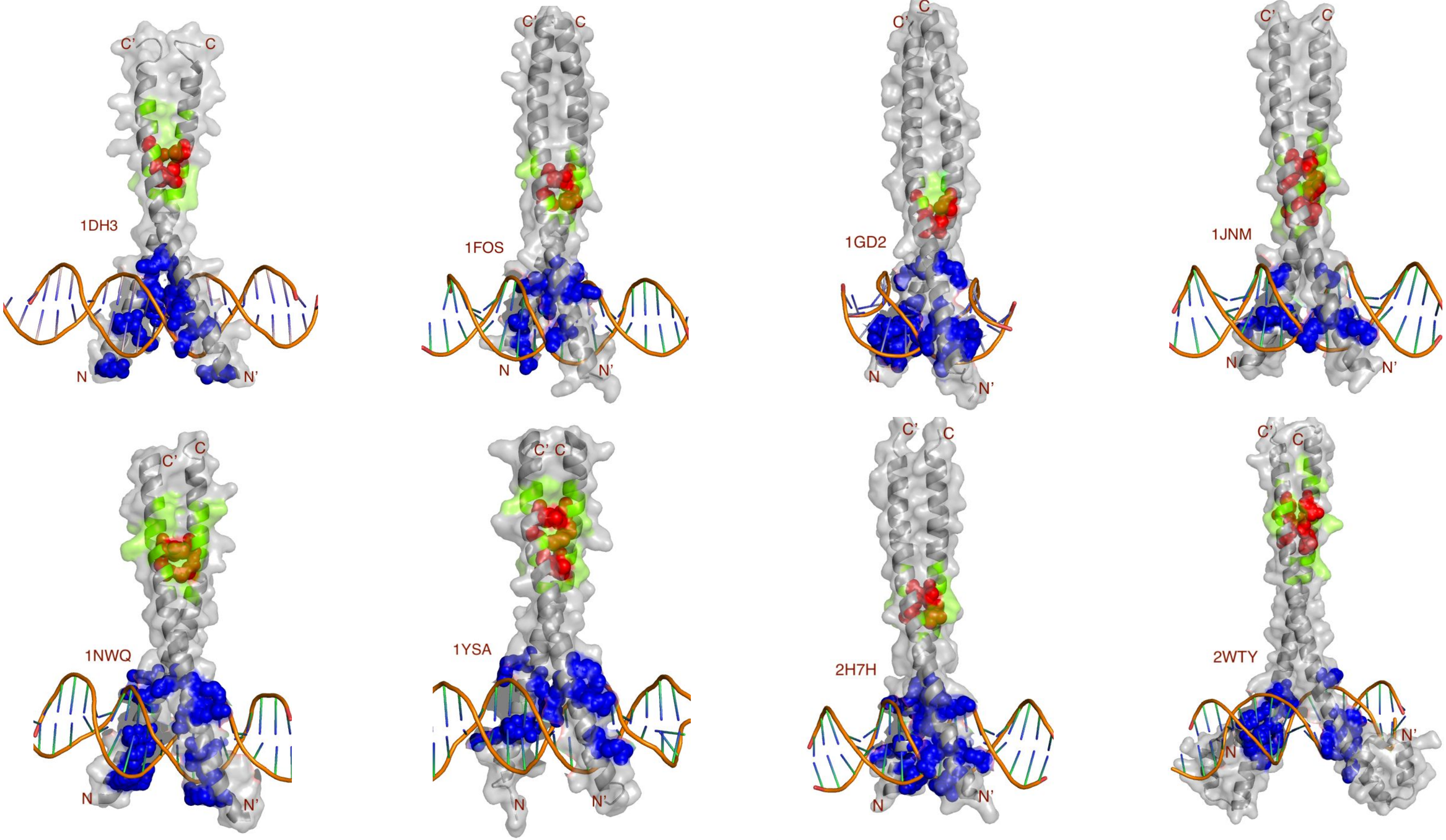
C'
C
1DH3
N
N'
C'
C
1FOS
N
N'
C
C'
1GD2
N
N'
C'
C
1JNM
N
N'
C'
C
1NWQ
N
N'
C' C
1YSA
N
N'
C'
C
2H7H
N
N'
C'
C
2WTY
N
N'

Figure 4

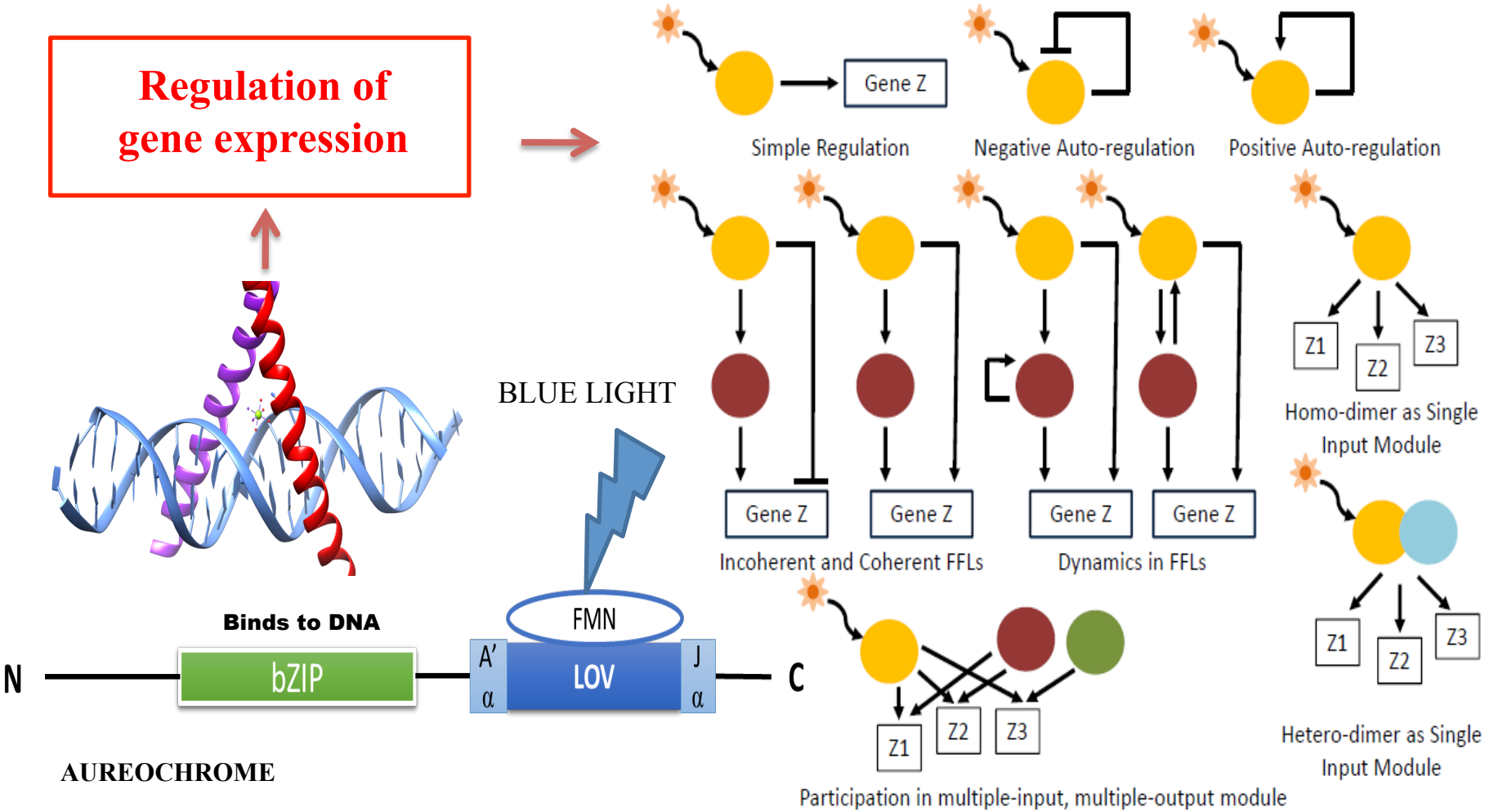
Regulation of gene expression
Gene Z
Simple Regulation
Negative Auto-regulation
Positive Auto-regulation
BLUE LIGHT
Z1
Z2
Z3
Homo-dimer as Single Input Module
Gene Z
Gene Z
Gene Z
Gene Z
Incoherent and Coherent FFLs
Dynamics in FFLs
Binds to DNA
FMN
N
bZIP
A'
α
LOV
J
α
C
AUREOCHROME
Z1
Z2
Z3
Hetero-dimer as Single Input Module
Z1
Z2
Z3
Participation in multiple-input, multiple-output module

Figure 5