OXFORD SIMB Society for Industrial Microbiology and Biotechnology

Journal of Industrial Microbiology and Biotechnology, 2024, 51, kuae011

https://doi.org/10.1093/jimb/kuae011
Advance access publication date: 3 April 2024
Natural Products – Review

Harnessing regulatory networks in Actinobacteria for natural product discovery

Hannah E. Augustijn [1,2,†], Anna M. Roseboom [2,†], Marnix H. Medema [1,2], Gilles P. van Wezel [2,3]

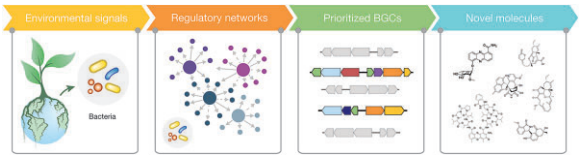
[1]Bioinformatics Group, Wageningen University, Wageningen, The Netherlands
[2]Molecular Biotechnology, Institute of Biology, Leiden University, Leiden, The Netherlands
[3]Netherlands Institute for Ecology (NIOO-KNAW), Wageningen, The Netherlands
Correspondence should be addressed to: Gilles P. van Wezel at g.wezel@biology.leidenuniv.nl and Marnix H. Medema at marnix.medema@wur.nl.
[†]These authors contributed equally.

Abstract: Microbes typically live in complex habitats where they need to rapidly adapt to continuously changing growth conditions. To do so, they produce an astonishing array of natural products with diverse structures and functions. Actinobacteria stand out for their prolific production of bioactive molecules, including antibiotics, anticancer agents, antifungals, and immunosuppressants. Attention has been directed especially towards the identification of the compounds they produce and the mining of the large diversity of biosynthetic gene clusters (BGCs) in their genomes. However, the current return on investment in random screening for bioactive compounds is low, while it is hard to predict which of the millions of BGCs should be prioritized. Moreover, many of the BGCs for yet undiscovered natural products are silent or cryptic under laboratory growth conditions. To identify ways to prioritize and activate these BGCs, knowledge regarding the way their expression is controlled is crucial. Intricate regulatory networks control global gene expression in Actinobacteria, governed by a staggering number of up to 1000 transcription factors per strain. This review highlights recent advances in experimental and computational methods for characterizing and predicting transcription factor binding sites and their applications to guide natural product discovery. We propose that regulation-guided genome mining approaches will open new avenues toward eliciting the expression of BGCs, as well as prioritizing subsets of BGCs for expression using synthetic biology approaches.

One-Sentence Summary: This review provides insights into advances in experimental and computational methods aimed at predicting transcription factor binding sites and their applications to guide natural product discovery.

Keywords: Actinobacteria, Regulatory networks, Natural product biosynthesis

Graphical abstract



Introduction

Actinobacteria are known as nature's medicine makers, producing a huge variety of natural products (NPs) (Barka et al., 2016; Bérdy, 2005; Hopwood, 2007). Activities of NPs include antibiotic, antifungal, anticancer, anthelmintic, herbicide, and immunosuppressant, and humans harness these properties for application in agriculture, biotechnology, and medicine. Identifying novel bioactive molecules via screening of strain collections is a costly process that, particularly for antibiotics, has rarely paid off in recent decades (Cooper & Shlaes, 2011; Payne et al., 2006). This discovery void coincides with the rapid spread of antimicrobial resistance. Together, these two issues create an urgent need for novel antibiotics.

A major revolution in the way we look at drug discovery was prompted by whole genome sequencing. When the genome of the model organism *Streptomyces coelicolor* was published some 20 years ago (Bentley et al., 2002), followed by several other well-studied Actinobacteria (Cruz-Morales et al., 2013; Ikeda et al., 2003; Ohnishi et al., 2008; Oliynyk et al., 2007), it became evident that these bacteria contain far more biosynthetic potential than originally anticipated. Genes encoding the biosynthesis of NPs are typically co-located on bacterial genomes in so-called biosynthetic gene clusters (BGCs), enabling the prediction of biosynthetic potential based on sequence data (Blin et al., 2019). Large-scale genome sequencing efforts brought to light that only a fraction of the biosynthetic potential that is potentially encoded by microbial genomes has been experimentally characterized (Bérdy, 2012; Gavriilidou et al., 2022; Newman & Cragg, 2020), and streptomycetes were shown to have a particularly diverse arsenal of BGCs (Gavriilidou et al., 2022). The vast majority of the chemical space of their specialized metabolites so far remains concealed, largely because we lack the understanding of how we should activate their biosynthesis (Ohnishi et al., 2008) or how we should prioritize BGCs that are likely to have biological activities of interest.

Received: January 22, 2024. **Accepted:** April 2, 2024.

The production of specialized metabolites is a resource-intensive process for the bacterial cell, usually taking place under conditions of resource limitation (Bruggeman et al., 2020; Kerkhoven, 2022). Therefore, the expression of BGCs is subject to extensive transcriptional control, governed by transcription factors (TFs) that mediate direct or indirect activation or repression of gene expression (Albert, 2005; Balleza et al., 2009). This regulatory network involves dynamic interplay between sigma factors for promoter recognition and specific and pleiotropic regulators that act as transcriptional repressors or activators, which together dictate the gene expression profiles. As of yet, the number of global regulators in streptomycetes for which the complete regulon has been described is still very limited, while approximately 12% of their genes are annotated as regulatory (Romero-Rodriguez et al., 2015). This underscores the current limitations in the understanding of the intricate regulatory networks controlling the biosynthesis of NPs. Understanding how BGCs are controlled may provide new leads to activate their gene expression or prioritize them for likely ecological functions (van Bergeijk et al., 2020; Yoon & Nodwell, 2014; Zhu et al., 2014). To do so, we need to characterize the TFs and their cis-regulatory elements (CREs) and identify the environmental triggers and cues that activate the expression of silent BGCs (Rigali et al., 2018). Recent advances in high-throughput technologies for the detection and prediction of transcription factor binding sites (TFBSs) hold great promise for extensive mapping of the regulatory networks of NP producers.

Here we will summarize the established principles of transcriptional control of BGCs gleaned from *Streptomyces coelicolor*, whereafter we will review experimental and computational methods employed in the pre- and post-genomic era for identifying and characterizing TFBSs in Actinobacteria. Subsequently, we will delve into the present and prospective applications of these methods within the genomics-based NP genome mining field, which promises many advances for NP discovery for both industrial and clinical applications.

A Short Overview of Transcriptional Control in the Model Organism *Streptomyces Coelicolor*

The intertwining regulatory networks translating environmental triggers to intracellular signals activating BGCs have been relatively well studied in the model organism *Streptomyces coelicolor* (Urem et al., 2016). The chemical differentiation of streptomycetes, marked by the production of secondary metabolites, also known as specialized metabolites or NPs, closely aligns with their morphological differentiation (Bibb, 2005; van der Heul et al., 2018). Streptomycetes undergo a complex life cycle in which stress conditions, such as nutrient depletion, initiate aerial hyphae formation, eventually leading to sporulation (Chater & Losick, 1997; Jakimowicz & van Wezel, 2012; Kelemen & Buttner, 1998). Following dispersion, spores germinate, giving rise to a branched, vegetative mycelium composed of long multinucleoid hyphae (Claessen et al., 2014; Elliot et al., 2008). At this stage of the developmental life cycle of *S. coelicolor*, the production of undecylprodigiosin (Red) is initiated. Subsequently, the vegetative mycelium lyses, and a programmed cell death provides nutrients for the newly formed aerial mycelium (Manteca et al., 2005; Tenconi et al., 2018). In *S. coelicolor*, the formation of this aerial mycelium coincides with the production of actinorhodin (Act). The aerial mycelium eventually produces chains of unigenomic spores (Jakimowicz & van Wezel, 2012; McCormick & Flardh, 2012). Due to this strict growth phase-dependent regulation of antibiotic production, developmental mutants that are halted in the life cycle generally fail to produce antibiotics.

The biosynthesis of Act, Red, and calcium-dependent antibiotic (CDA) is directly controlled by ActII-ORF4, RedD, and CdaR, respectively, which belong to the family of *Streptomyces* antibiotic regulatory proteins (SARPs) (Wietzorrek & Bibb, 1997). Transcription of *redD* is in turn controlled by the orphan response regulator RedZ, illustrating a hierarchical cascade of cluster-situated regulators (CSRs) governing BGC expression (Bibb, 2005). Interestingly, expression of the Red BGC is completely dictated by the expression of RedD, and expression of *redD* from a developmentally controlled promoter allows production of prodigionines in aerial hyphae (van Wezel et al., 2000). The expression of CSRs may be controlled by an intricate network of global regulators. As an example, transcription of the *actII*-orf4 is controlled by complex multilevel control that consists of some 20 different regulatory networks (Liu et al., 2013). These include several pleiotropic regulators such as Crp, DasR, and AtrA, each of which modulate primary and specialized metabolism as well as morphological differentiation (Gao et al., 2012; Świątek-Połatyńska et al., 2015). This intertwining network of regulatory proteins facilitates multilevel control of BGC expression, of vital importance for *Streptomyces* but challenging for researchers to completely decipher.

To delineate the organizational structure of the regulatory network, it is interesting to compare the regulons controlled by the pleiotropic regulators BldD (Den Hengst et al., 2010) and DasR (Świątek-Połatyńska et al., 2015). The regulon of the highly conserved BldD in *S. coelicolor* spans some 160 direct targets, with a particular focus on the control of developmental genes. In *S. coelicolor*, BldD only marginally and indirectly affects antibiotic production, for example, via its influence on the transcription of genes like *adpA* and *nsdA* (Den Hengst et al., 2010; Yan et al., 2020). While later studies showed that BldD may directly control the biosynthesis of avermectin in *Streptomyces avermilis* (Yan et al., 2020), daptomycin in *Streptomyces roseosporus* (Yan et al., 2020), and lincomycin in *Streptomyces lincolnensis* (Li et al., 2019), its core regulon clearly revolves around the control of development. Conversely, the regulon of DasR, which controls a similar number of genes as BldD, mainly revolves around the control of primary and specialized metabolism (Rigali et al., 2008; Świątek-Połatyńska et al., 2015; Urem et al., 2016). Considering that both DasR and BldD act during early development, and mutation of either gene locks streptomycetes in the vegetative growth phase, it is surprising to see how little overlap there is between these two master regulons (Rigali et al., 2018). This suggests an evolutionary strategy to separate the global metabolic and developmental control networks in streptomycetes. In *S. coelicolor*, DasR responsive elements were discovered within most BGCs that specify antibiotics (*act, red, cda, cpk*), while its control over siderophore biosynthesis is mediated via repression of the iron master regulator *dmdR1* (Craig et al., 2012). DasR was the first example of a pleiotropic regulator that controls a cryptic BGC, namely *cpk* for the cryptic type I polyketide Cpk. The production of Cpk (later called coelimycin P1) could be elicited via the addition of *N*-acetylglucosamine (GlcNAc), which is metabolized and subsequently interferes with DasR binding, thereby unlocking the expression of the *cpk* cluster (Rigali et al., 2008). This underlines the power of mapping regulons for NP discovery. For a more detailed description, including visualizations, of the regulatory networks described in this introduction, we refer to reviews elsewhere (Liu et al., 2013; van der Heul et al., 2018).

Gaining insights into the TF regulatory networks in model organisms such as *S. coelicolor* is important to improve our general

Table 1. Examples of Conserved Pleiotropic Regulators in *Streptomycetes* With Identified TFBSs

TF	Studied strain	Method	Reference
Primary metabolism			
ArgR	*S. coelicolor*	EMSA & footprinting	Botas et al. (2018), Perez-Redondo et al. (2012)
Crp	*S. coelicolor*	ChIP-chip	Gao et al. (2012)
GlnR	*S. coelicolor*	EMSA & footprinting	Liao et al. (2015), Tiffert et al. (2008), Wang & Zhao (2009)
LexA	*S. venezuelae*	ChIP-seq	Stratton et al. (2022)
NsrR	*S. venezuelae*	ChIP-seq	Crack et al. (2015), Munnoch et al. (2016)
HrdB	*S. coelicolor*	ChIP-seq	Šmídová et al. (2019)
Specialized metabolism			
AdpA	*S. griseus*	ChIP-seq, EMSA & footprinting	Hara et al. (2009), Higo et al. (2012)
AtrA	*S. coelicolor*	EMSA & footprinting	Uguru et al. (2005)
DasR	*S. coelicolor*	ChIP-chip	Swiatek-Polatynska et al. (2015)
ScbR	*S. coelicolor*	ChIP-seq, EMSA & footprinting	Li et al. (2015)
ScbR2	*S. coelicolor*	ChIP-seq, EMSA & footprinting	Li et al. (2015)
Development			
AbrC3	*S. coelicolor*	ChIP-chip	Rico et al. (2014)
BldC	*S. venezuelae*	ChIP-chip & ChIP-seq	Bush et al. (2019)
BldD	*S. coelicolor*	ChIP-chip	Den Hengst et al. (2010)
BldM	*S. venezuelae*	ChIP-seq	Al-Bassam et al. (2014)
OsdR	*S. coelicolor*	EMSA & footprinting	Urem et al. (2016)
PhoP	*S. coelicolor*	EMSA & footprinting	Santos-Beneit et al. (2009), Zheng et al. (2019)
WhiB	*S. venezuelae*	ChIP-seq	Bush et al. (2013)

Note. The regulators are subclassified in their main functional control group.

comprehension of regulons, and this information can then be translated to other strains. Table 1 provides an overview of well-studied conserved pleiotropic regulators in streptomycetes for which at least a part of the regulon has been mapped using experimental methods.

Experimental Techniques for the Detection of Regulatory Systems

In the research domain of Actinobacterial gene regulation, the sequencing of the complete genome of the model organism *Streptomyces coelicolor* A3(2) (Bentley et al., 2002) marked a transition from traditional pre-genomic methodologies, which focused mostly on single transcription units and simple *cis-trans* relationships, to post-genomic techniques, known for their high-throughput and genome-wide analysis capabilities. To study the intricate interactions between TFs and the DNA, numerous experimental methods have been developed, and for comprehensive details, we refer readers to in-depth reviews elsewhere (Dey et al., 2012; Ferraz et al., 2021; Guille & Kneale, 1997). In this section, we present the most relevant experimental techniques used to study TFs, their binding sites, and their regulons, illustrating their progression from the pre- to post-genomic era. These methodologies provide the basic toolbox to unravel the regulatory networks controlling BGCs in Actinobacteria and thus constitute the technological basis for regulation-guided genome mining efforts.

Detection of TFBSs in the Pre-genomic Era

Initial research on DNA-protein interactions primarily centered on globally acting TFs that regulate various aspects of Actinobacterial biology. In the pre-genomic era, DNA binding studies were typically performed using electrophoretic mobility shift assay or gel retardation assay (Fig. 1a) (Dey et al., 2012; Fried & Crothers, 1981; Garner & Revzin, 1981). Here, the protein-nucleic complex formation reduces migration speed compared to the electrophoretic run of a free nucleic acid fragment in a gel. DNase I footprinting, whereby enzymatic cleavage of the DNA is prevented by binding of a protein, can then be employed to decipher the exact nucleotides a TF binds to (Fig. 1a) (Ferraz et al., 2021; Galas & Schmitz, 1978). While these techniques are limited in throughput and focus, often addressing single genes at a time, they have been instrumental in unraveling complex regulatory systems (Bibb, 1996). To predict and identify binding sequences for novel regulators, the systematic evolution of ligands by the exponential enrichment (SELEX) method was developed (Fig. 1a) (Ferraz et al., 2021; Tuerk & Gold, 1990). This method has been applied successfully to localize TFBSs on a genome-wide scale in Actinobacteria (Qian et al., 2002). However, SELEX fails to decipher the complex binding motifs of TFs with variable binding sequences, a trait that is commonly seen for pleiotropic regulators, for example, for DasR (Swiatek-Polatynska et al., 2015).

Although these pre-genomic methods may appear somewhat outdated in the face of the high-throughput post-genomic techniques that will be explored in the following section, they remain widely applied (Fig. 1c). Presumably, because of the highly reliable information they provide on specific regulatory interactions. In *Streptomyces*, the genus with the most regulatory-focused papers (Fig. 1c), they were used to elucidate the CREs of the A-factor dependent transcriptional activator AdpA (Yamazaki et al., 2004), PhoP that is involved in the response to stress caused by phosphate limitation (Sola-Landa et al., 2005; Yang et al., 2015), and DasR that links the control of primary and secondary metabolism (Rigali et al., 2006, 2008). More recently, these techniques have been used to enhance the titer of the macrolide pesticide milbemycin (Wang et al., 2022a), demonstrating the ongoing relevance of these pre-genomic methods in current research.

The Post-genomic Era: Genome-wide Transcription Network Analysis

The advent of genome sequencing technologies led to major advances in the analysis of transcriptional networks and DNA binding studies. Around the turn of the century, a breakthrough was achieved with the development of ChIP-chip or ChIP-on-chip, a method that combined Chromatin Immuno Precipitation (ChIP)

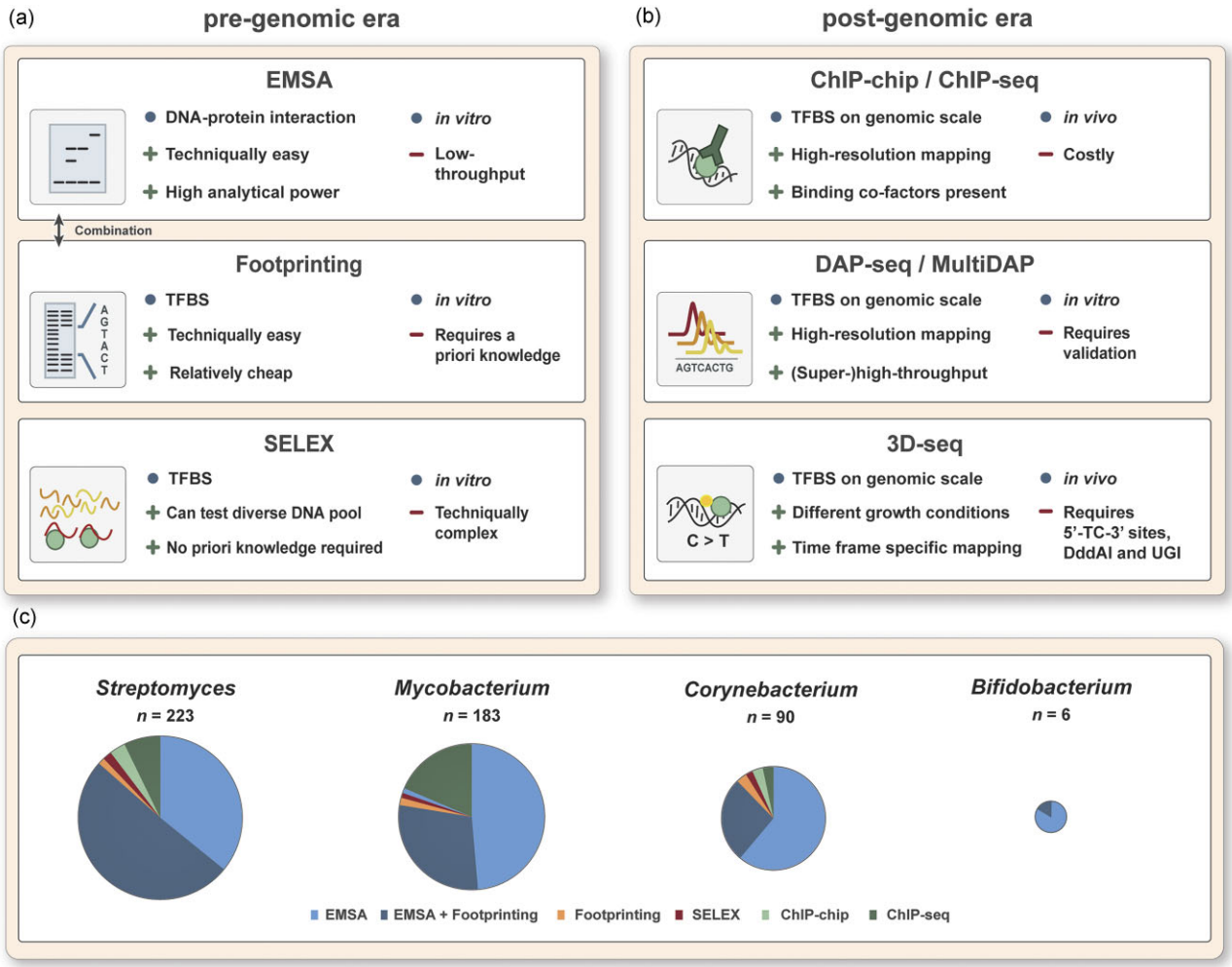


Fig. 1. The core toolbox of techniques for regulatory network analysis, including methods from the pre-genomic era (a) and the post-genomic era (b) to study DNA-protein interactions and transcription factor binding sites (TFBSs) in Actinobacteria, and their advantages and disadvantages. Circle charts illustrate how regularly these common methods have been applied to study transcription factors in four genera within the phylum of Actinobacteria; *Streptomyces*, *Mycobacterium*, *Corynebacterium*, and *Bifidobacterium* (c). These charts represent papers on transcription factors that were found using the keyword "EMSA," "electrophoretic mobility shift assay," "gel retardation assay," "band shift assay," "footprinting," "DNAse protection assay," "SELEX," "ChIP-chip," "ChIP-on-chip," or "ChIP-seq" in combination with the respective genus name on PubMed on October 5, 2023. The size of each circle plot corresponds to the dataset size, as denoted by the value "*n*," which reflects the actual number of papers represented by the size of the plots.

with DNA microarray analysis to study genome-wide TF binding under physiological conditions in living cells (Blat & Kleckner, 1999; Ren et al., 2000). With this technique, TFs bound to their target sites are cross-linked chemically or via UV light, followed by shearing of the genomic DNA (Gilmour & Lis, 1984; Solomon & Varshavsky, 1985). DNA-protein complexes are then retrieved using antibodies against a genetically engineered protein tag, such as Flag, HA, V5, or Myc (Kidder et al., 2011), whereafter DNA analysis is conducted using DNA microarrays. A subsequent major advance was the integration of ChIP with high-throughput DNA sequencing (ChIP-seq) (Fig. 1b) (Johnson et al., 2007). This combination revolutionized the process by enabling the quick and accurate analysis of large amounts of precipitated DNA fragments. Drawbacks of ChIP-seq include its complex data analysis and the fact that the use of antibodies, reagents, and the required sequencing depth make this a relatively expensive method (Dey et al., 2012; Ferraz et al., 2021). Despite this, ChIP-seq has been broadly used to analyze various regulatory networks in key Actinobacteria genera such as *Corynebacterium* (Jeong et al., 2021; Jungwirth et al., 2013; Zhang et al., 2022), *Mycobacterium* (Galagan et al., 2013; Minch et al., 2015), and *Streptomyces* (McLean et al., 2016; Munnoch et al., 2016; Stratton et al., 2022) (Fig. 1c).

More recently, DNA affinity purification sequencing (DAP-seq) was developed to achieve even greater throughput in the mapping of regulons (Bartlett et al., 2017). The method is similar to ChIP-seq, except that proteins are produced via *in vitro* transcription-translation and then bound to genomic DNA in the absence of their natural stimuli, allowing the generation of genome-wide DNA binding profiles of many TFs in an inexpensive and rapid way (Bartlett et al., 2017; O'Malley et al., 2016). Originally developed for *Arabidopsis* (Galli et al., 2018), DAP-seq has now been successfully applied to fungi (Fischer et al., 2018), insects (De Mendoza et al., 2019), and bacteria (Wang et al., 2022b; Zhang et al., 2023). Adaptations of the classical DAP-seq method include biotin-DAP-seq and MultiDAP, which are respectively capable of mapping the regulons even more rapidly or for multiple organisms and TFs simultaneously (Baumgart et al., 2021). The success rate of biotin-DAP-seq is comparable to other *in vitro* methods (Baumgart et al., 2021). The *in vitro* transcription-sequencing (RIViT-seq) technology that has recently been developed combines *in vitro* transcription by

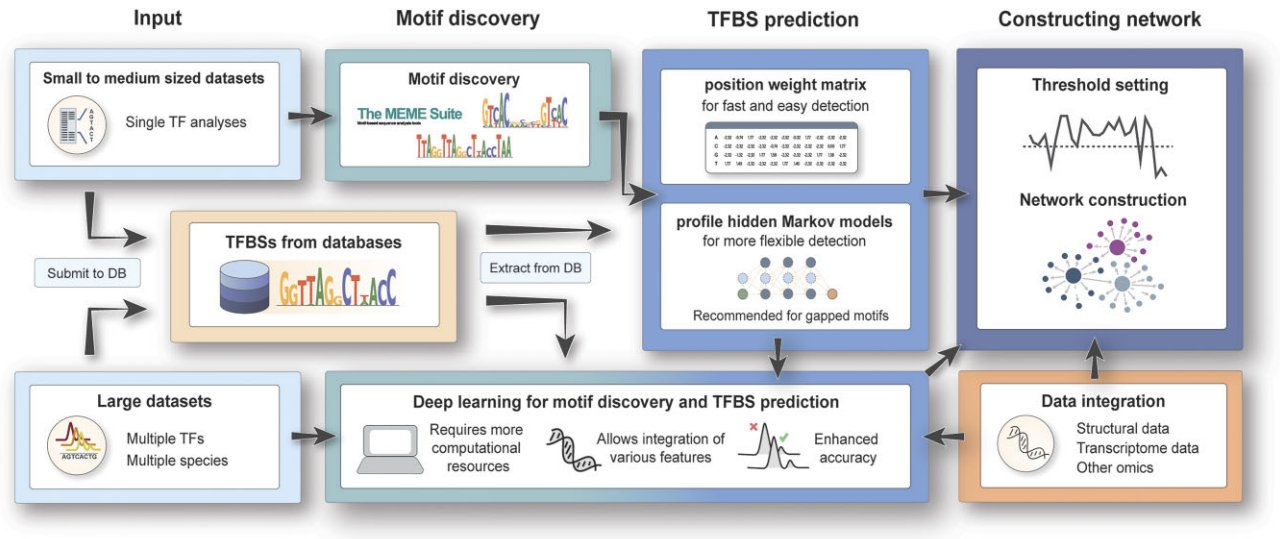


Fig. 2. Workflow for constructing regulatory networks from experimental data. The process can be roughly categorized into four stages. Importantly, the nature of the input data determines the subsequent procedural steps. For small to medium-sized datasets, such as footprinting results or single-TF ChIP-seq, a conventional motif discovery phase is employed, followed by TFBS predictions utilizing PWMs or pHMMs. In contrast, large datasets from high-throughput analyses, such as MultiDAP, could be suitable for advanced deep learning techniques. For all methodologies, threshold determination is essential to discern connections between regulators, resulting in the construction of a regulatory network.

reconstituted RNA polymerase with RNA-sequencing to identify regulons (Otani & Mouncey, 2022). However, like for all *in vitro* DNA binding studies, the disadvantage of these methods is that key biological information is missing, such as post-translational modifications, ligands or cofactors that influence the binding activity of a TF, and information on, for example, the growth-phase-dependent activity of the TFs. Combining various multi-omics datasets, for instance, *in vitro* with *in vivo* assays, compensates for the lack of this key biological information and allows for the systematic analysis of regulatory networks (Zhang et al., 2020). The effectiveness of this combinational approach is showcased in a study of cell growth phase regulation in *Streptomyces griseus* (Hwang et al., 2022). In this study, the use of four multi-omics datasets, including RNA-seq, dRNA-seq, Term-seq, and ribosome profiling, led to the discovery of several novel regulons. This highlights the potential of combining diverse datasets for a more complete understanding of regulatory interactions. Another recently developed high-throughput method providing this much-needed additional layer of information to study DNA-protein interactions *in vivo* is 3D-seq or DddA-seq (Fig. 1b) (Gallagher et al., 2022). Here, a regulator is fused to the active deaminase domain of the nucleic acid-targeting deaminase Ddda that targets 5'-TC-3' sites around the TFBS and introduces C-G to T-A mutations that localize the CRE of the protein. Moreover, DddA activity can be modified by adjusting the expression of $DddA_I$, whereby TF binding can be mapped within a specific timeframe, for instance, during host colonization. This would provide another extension to the biological understanding of the function of a BGC, which could contribute to BGC prioritization in the search for new NPs.

As sequencing costs continue to decrease, high-throughput methods such as ChIP-seq and DAP-seq will rapidly become the new standard for studying genome-scale *cis-trans* relationships. Processing these data involves statistical filtering that will inevitably lead to false negatives, and hence loss of important binding interactions. Therefore, data generated by these methods should always be validated with other experimental methods.

Computational Methods for TFBS Detection

The development of high-throughput screening methods, together with reduced experimental costs, has significantly increased the availability of experimental data in genomics. Despite these advancements in experimental methods, many researchers encounter challenges when working with their own custom strains or species that lack existing genomic information. Moreover, even within well-researched model organisms, the process of validating putative binding sites is often labor-intensive and time-consuming. Addressing this, the existing experimentally validated binding site data offers the opportunity to create sophisticated *in silico* TFBS prediction models. Such models not only identify potential TFBSs in previously unstudied species but also provide assurance of legitimacy when a putative binding site aligns with a recognized motif and thus facilitate high-confidence assignment of regulatory interactions with BGCs. In this section, we will describe several computational techniques and methodologies for TFBS detection, as well as their potential applications in genome mining, with a particular focus on Actinobacteria.

TFBS Detection

Numerous methods for detecting TFBSs have been developed to streamline and enhance regulatory predictions, each tailored to specific datasets and research objectives. Tools like MEME (Bailey et al., 2015) and Weeder (Pavesi et al., 2004) excel in identifying TFBS patterns or motifs from raw data, visualized as sequence logos and consensus sequences for convenient motif representation. However, consensus sequences do not fully capture the variation of nucleotide distribution in the motif, limiting their standalone effectiveness for TFBS detection (Eggeling, 2018; Ladunga, 2010; Stormo, 2000).

To accurately predict TFBSs, position weight matrices (PWMs) were introduced (Staden, 1984; Stormo, 2000). In contrast to consensus-based methods, these matrices are generated by quantifying the frequency of each nucleotide or amino acid at every position in a set of aligned sequences. With the use of PWMs,

it becomes possible to predict entire regulons with a minimal amount of experimentally validated binding sites. Bioinformatics tools, such as PREDetector (Hiard et al., 2007), RSAT (Santana-Garcia et al., 2022), or various tools from the MEME-suite, aid in the construction and utilization of custom PWMs or motifs of data obtained from databases or experimental results. Despite PWMs being the predominant method for predicting TFBS occurrences in DNA, they are not without limitations. The model operates under the assumption that each nucleotide functions independently, which might lead to an oversimplification of the complex interactions and dependencies between nucleotides (Bulyk et al., 2002; Hannenhalli, 2008; Nguyen & Androulakis, 2009). Consequently, using prediction models, true binding sites might be camouflaged among lower prediction scores and, hence, indistinguishable from false positives, making experimental validation essential to determine complete regulons (Bang et al., 2022; Ma et al., 2015; Wasserman & Sandelin, 2004). Although these challenges led to the development of numerous adaptations of this classical PWM approach (Boytsov et al., 2022; Hannenhalli & Wang, 2005; Jayaram et al., 2016; Nguyen & Androulakis, 2009), the classical PWM model continues to command greater attention and preference within the field. Its sustained popularity is likely due to its blend of simplicity, speed, and minimal computational requirements, all while maintaining commendable accuracy.

While the PWM remains the go-to method for TFBS detection, certain situations call for more advanced models. For example, its rigidity, particularly its fixed length, becomes problematic when addressing TFs with variable sequence lengths and spacer regions (Mathelier & Wasserman, 2013). Recognizing these challenges, there is a shift toward nucleotide profile hidden Markov model (HMM)-based frameworks as an alternative (Eddy, 2004; Maaskola & Rajewsky, 2014; Yoon, 2009). HMMs employ a state-based probabilistic system, granting them the capability to represent position interdependencies within TFBSs and accommodate variable motif lengths. For motif discovery, these HMMs demonstrate more optimal precision, particularly when identifying gapped motifs. GLAM2 (Frith et al., 2008), a component of the MEME suite, stands as a commendable tool for this purpose. Transitioning from motif discovery to TFBS detection, the adaptability of HMMs continues to be evident. For example, the tool nhmmer of HMMER (Finn et al., 2011; Wheeler & Eddy, 2013) has gained traction as it exploits the probabilistic nature of HMMs to accurately pinpoint binding sites in complex and often noisy datasets, like those from ChIP-seq or DAP-seq (Madera & Gough, 2002; Mathelier & Wasserman, 2013). Nevertheless, HMM approaches demand substantial input data to surpass the performance of PWMs (Nguyen & Androulakis, 2009). Within the Actinobacterial research domain, this has been a bottleneck due to the limited availability of vast datasets (Fig. 1c). However, as high-throughput techniques continue to advance and become more accessible, we anticipate a shift toward more complex methodologies.

Building on the methodologies previously explored, there is another layer of complexity that demands attention. The binding dynamics of TFs to their corresponding TFBSs rely heavily on structural nuances, such as the shape of the DNA, making structure-centric approaches more accurate compared to sequence-based methods (Cui et al., 2022; Mathelier et al., 2016). Predictive deep learning models for binding site detection increasingly incorporate these structure-based features in their training phase to refine predictive capabilities (Chiu et al., 2020; Schnepf et al., 2020; Wang et al., 2021a; Yang & Ramsey, 2015). Furthermore, several efforts have been made to utilize TF-DNA structural information, such as X-ray crystallography data, to obtain insights into the binding orientation of the TF protein itself (Trerotola et al., 2022; Wetzel et al., 2022). However, a significant bottleneck is the scarcity of these experimentally defined crystal structures, driving the need for innovative prediction models. Recently, the availability of these predicted protein structures has significantly increased with the development of novel prediction methods such as AlphaFold (Jumper et al., 2021). However, the effective application of these models in actinobacterial contexts hinges on high-quality predicted TFs and subsequent experimental validation.

Regulatory Network Reconstruction

The identification of individual binding sites through regulatory detection methods is a crucial first step in understanding gene regulation. The challenge then lies in constructing comprehensive TF regulatory networks. Here, we will focus on how to bridge the gap between available experimental data for TBFSs and the formation of a comprehensive network, which is essential for elucidating the complex regulatory cascades (Fig. 2). As a simple example, the activation of avermectin biosynthesis in *S. avermitilis* is governed by the CSR AveR, which is in turn regulated by the pleiotropic phosphate regulator PhoP (Martín et al., 2017). This then predicts that depleting phosphate should lead to the activation of the avermectin BGC.

Constructing a gene regulatory network based on TFBSs begins with the careful curation of binding interactions, whether experimentally determined or predicted. It is of critical importance to choose the proper thresholds to minimize false positives while at the same time preventing too many false negatives. For experimentally validated binding sites, confidence levels can be assigned based on the extent of experimental evidence supporting each site (Escorcia-Rodríguez et al., 2020). In the case of predictive models such as PWMs, thresholds are established by considering the information content of motifs or examining the ratio of hits in coding versus non-coding regions (Hiard et al., 2007). Once these curated interactions are compiled, it becomes possible to construct the network. In such a network, genes are represented as nodes, while the interactions between regulators are depicted as edges. Another prevalent method involves the derivation of networks from transcriptomic data, known as transcriptional regulatory networks. Contrasting with TFBS-based networks, TRNs are more focused on aspects of gene expression. A notable example of this approach can be seen in recent research on *S. coelicolor*, where they compared networks generated from both TFBS and transcriptomic data to construct a comprehensive gene regulatory network (Zorro-Aranda et al., 2022). This study revealed that the TFBS-based network appeared to be more comprehensive than the transcriptional regulatory network. However, integrating these two methodologies seems a logical progression. The mere presence of a TFBS does not necessarily result in transcription, and conversely, transcriptomic data can be constrained by the specificity of the culture conditions used. Therefore, a combined approach could effectively mitigate the inherent limitations of each method, providing a more complete understanding of gene regulation.

Regulatory Networks and Elicitation of Cryptic BGCs

When analyzing sequence data from even a single actinobacterial strain for the discovery of, for example, novel bioactive compounds, many target BGCs with potentially desired properties can be identified. This identification is usually achieved by utilizing genome mining tools such as antiSMASH (Blin et al., 2023)

or PRISM (Skinnider et al., 2020). However, strains frequently fail to express these BGCs under conventional experimental conditions, leading these gene clusters to be referred to as "silent" or "cryptic" BGCs. Given the complexity of the native regulatory networks that control BGC expression, a prevalent strategy in such instances is to omit these regulatory systems by refactoring and heterologous expression of these BGCs in alternative host organisms (Baral et al., 2018; Liu et al., 2021). Though this approach has been successfully applied for gene clusters derived from Actinobacteria (Ahmed et al., 2020; Gomez-Escribano & Bibb, 2011; Ikeda et al., 2014; Mevaere et al., 2018), it often mandates intricate manual design. Additionally, heterologous expression might lead to variations in the metabolites produced, raising questions about how closely these compounds mirror those generated by the original strain (Xu et al., 2022). Finally, compounds may be produced from more than one gene cluster, in which case heterologous expression becomes undoable at a significant scale (van Bergeijk et al., 2020; van Wezel et al., 1994). Therefore, in many cases, scientists will have to rely on the original host for the production of the compound, thus driving the need for alternative BGC expression methods.

One such method is through the identification and subsequential utilization of environmental signals that trigger the native regulatory networks that govern the production of specialized metabolites. These specialized metabolites play an important role in survival through mediating resource competition (Abrudan et al., 2015; Wright & Vetsigian, 2016), protection against oxidative stresses (Lee et al., 2020), and in the uptake of essential nutrients (Kramer et al., 2020). This requires careful timing of production, suggesting that environmental signals, indicative of specific stresses such as nutrient scarcity or the presence of competitors, have been incorporated in the regulation of BGC expression (Bibb, 2005; Hoskisson & Fernández-Martínez, 2018; Kramer et al., 2020; van der Heul et al., 2018). Indeed, nutrient availability and co-culturing of Actinobacteria with other microorganisms significantly influence their specialized metabolite production (Rigali et al., 2008; Traxler et al., 2013). To express cryptic or silent BGCs, it is essential to understand these environmental signals, or "elicitors," that influence metabolite production. Current methods to identify these elicitors for BGC activation include changing the growth media (Bode et al., 2002; Zhu et al., 2014), use of antibiotics to induce spontaneous antibiotic resistance (Hosaka et al., 2009; Tanaka et al., 2009), and co-cultivation approaches (Bertrand et al., 2014; Hoshino et al., 2015; Sugiyama et al., 2015). Additionally, screening for novel chemical elicitors can enhance the success rate in high-throughput screening of bacterial strain collections. Examples of such screens include GlcNAc, which works via interference with DasR (Rigali et al., 2008), γ-butyrolactones (Hsiao et al., 2009; Willey & Gaskell, 2011), and histone deacetylase inhibitors (Albright et al., 2015). The high-throughput elicitor screening technology (HiTES) has been developed for large-scale elicitor screens (Moon et al., 2019a, b), where bacteria are challenged with a library of small molecules, followed by metabolomics-based screening of the responses. This approach has successfully identified cryptic antibiotics, such as the lanthipeptide cebulantin (Moon et al., 2019a) and the naphtoquinone epoxide hiroshidine (Moon et al., 2019b). Alternatively, the identification of ligands inducing conformational changes in TFs, thereby eliciting the production of NPs via TF binding, can be predicted through *in silico* small molecule docking (Agu et al., 2023; Oladejo et al., 2023). This method allows the prediction of binding poses of ligands to receptor proteins and has become an emerging trend in drug discovery (Rudrappa et al., 2023). Although it relies on high-resolution 3D representations of target proteins, the method has been successfully applied for the identification of various ligands, such as the role of a Zn(II) ion in the functioning of UxuR in *E. coli* (Ferraz et al., 2021; Purtov et al., 2019). Additionally, we envision that we may also look at regulatory networks for pathway activation and to predict BGC function—not by the enzymes they encode or the molecules they produce, but by their regulatory controls and responsiveness to elicitors. This targeted TFBS detection approach, for example, as implemented in antiSMASH (Blin et al., 2023), will provide valuable information on when a certain BGC may be expressed and how its expression can be elicited. For this, we can examine the binding sites of known regulators as markers to discover possible novel pathways. By understanding the function of a specific regulator, the functionality of the genes it controls can be inferred. A possible methodology underlying this approach is versatile and can be applied to various TFs (Fig. 3), as illustrated by the application of this strategy to the INBEKT (Identification of Natural compound Biosynthesis pathways by Exploiting Knowledge of Transcriptional regulation) system (Spohn et al., 2016). This system successfully identified a novel zinc-associated gene cluster by targeting the zinc-regulon through the zinc-dependent regulator ZuR. With knowledge of the elicitor at hand, the novel gene cluster could be experimentally validated, underlining the potential of utilizing regulatory networks and elicitor screens as an integrated regulator-guided strategy for functional inference and elicitation of BGCs.

Clearly, these methods require prior knowledge of the TFBSs from which the regulon can be deduced. In other model organisms, such as *Escherichia coli*, *Bacillus subtilis*, or *Pseudomonas aeruginosa*, many *cis-trans* relationships between TFs and their cognate binding sites have been characterized (Gao et al., 2018; Moreno-Campuzano et al., 2006; Wang et al., 2021b). In contrast, for *Streptomyces*, it is estimated that only about 6% of the complete network has been experimentally elucidated (Zorro-Aranda et al., 2022). Hence, expanding the knowledge of TFBSs is a prerequisite if we are to make significant strides in the application of regulatory networks as prediction tools for regulation and elicitation of BGCs. To overcome this limitation, the tool AURTHO exploits the frequent autoregulatory nature of TFs to detect their corresponding sequence motifs (Anderssen et al., 2022). Such an approach suggests that if a shared motif is found upstream of both a TF and pathway-related components, like transporters or core biosynthesis genes, it could indicate the TF's role in regulating that specific regulon. Conversely, the function of a TF might be inferred by examining the roles of its target genes.

All in all, we believe that the concept of connecting regulatory network information and elicitor screens is a promising approach in genome mining-based NP discovery, which will aid in prioritizing BGCs and finding new ways to elicit their expression.

Concluding Remarks and Future Perspectives

In summary, this review has highlighted the crucial yet often neglected regulatory aspect of specialized metabolite production in the search for novel drug candidates. Understanding and manipulating BGC regulation has the potential to facilitate the activation of silent or prioritization of orphaned gene clusters, potentially unlocking a wealth of untapped NPs. Moreover, the adoption of regulation-guided strategies, alongside the advancement of high-throughput techniques and genome mining

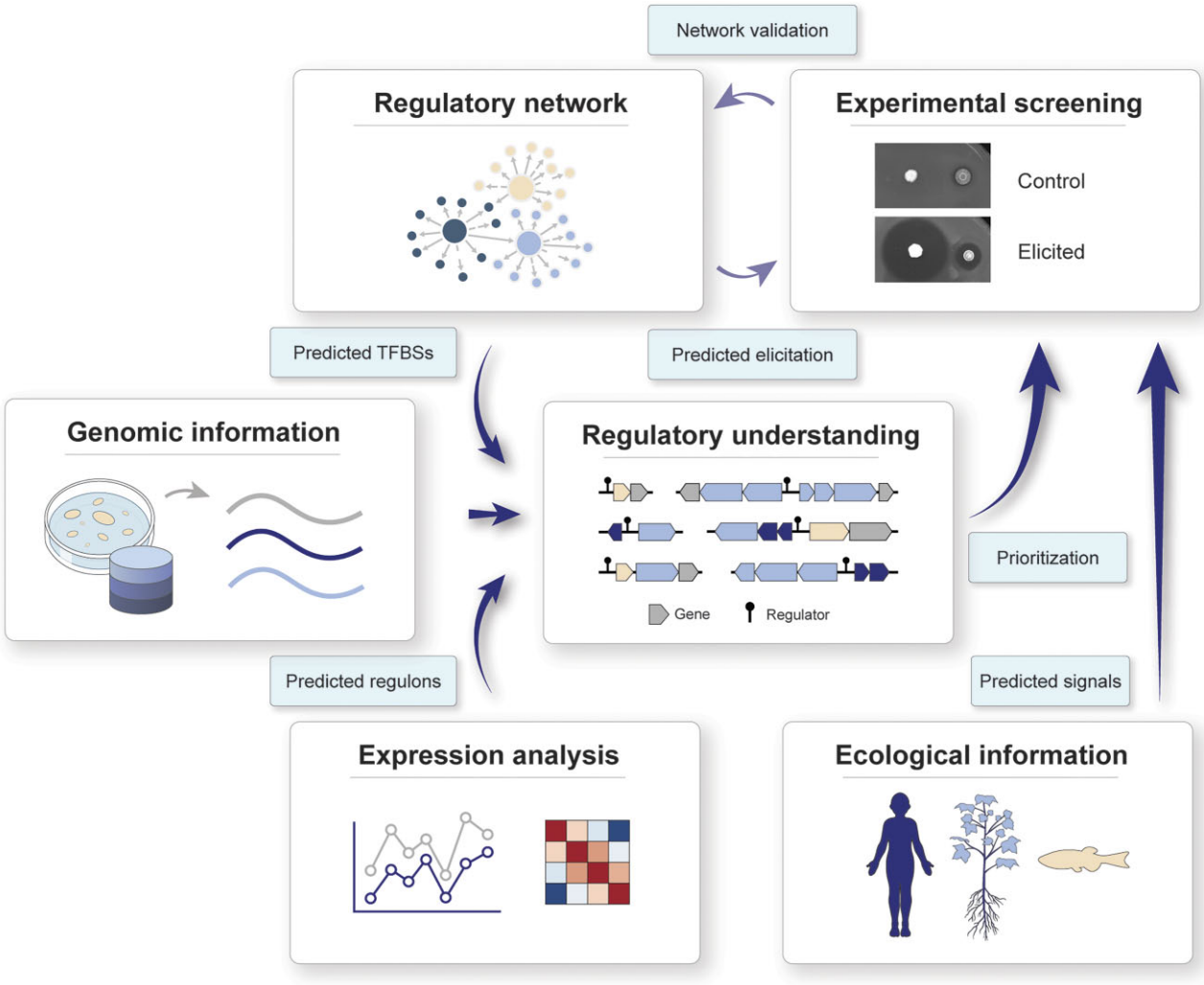


Fig. 3. Schematic workflow for combining genomic information with regulatory networks, and expression data for the functional understanding and prioritization of gene clusters for experimental validation. Genomic data obtained from databases or from sequencing strain collections are combined with experimentally validated or predicted regulatory interactions. These data connected to gene cluster and TFBS detection methods and their associated regulatory networks, aid in the prioritization of BGCs for further detailed experimental validation. Ecological information can be incorporated as an additional layer for the prediction of elicitation conditions to activate BGC expression *in vivo*. Improved understanding of the underlying regulatory networks also provides important clues for the elucidation and chemical characterization of the natural products derived from the BGCs. In turn, the experimental characterization feeds the regulatory network with new validated interactions. Created with BioRender.com.

methods, is paving the way for advances in how we prioritize and exploit these pathways.

The potential of binding site detection techniques is undeniable. Yet, the primary constraint facing regulation-guided genome mining strategies is not the lack of innovation but rather the scarcity of diverse, high-quality datasets, particularly for Actinobacteria. The advancement of affordable, high-throughput experimental methodologies could offer a solution, as it is set to rapidly increase our collection of large datasets dedicated to biosynthetic pathway exploration. Specifically, state-of-the-art global binding site screening methods, such as MultiDAP, will be of great value, especially when combined with complementary data types, such as large-scale transcriptome datasets from elicitor screenings. This combined approach can shed light on the connection between regulons and elicitors, subsequently unveiling the triggers needed for gene cluster expression. As the field continues to evolve, computational strategies will be indispensable in harnessing the full potential of these datasets. With the influx of more data, machine learning methods, renowned for their ability of handling complex datasets, are expected to take the lead in predicting functions or elicitors of gene clusters based on their regulatory characteristics (Mullowney et al., 2023).

Overall, regulator-guided genome mining strategies hold immense promise for decoding the complex regulatory networks that govern gene expression. By acquiring a deeper understanding of these networks, we can improve our ability to prioritize and efficiently utilize gene clusters, unlocking their potential for biomedical and biotechnological innovations and implementations.

Funding

This work was supported by the European Union via ERC Starting Grant 948770-DECIPHER to M.H.M. and ERC Advanced Grant 101055020-320 COMMUNITY to G.P.v.W.

Conflict of interest

M.H.M. is a member of the scientific advisory board of Hexagon Bio. The other authors declare no conflict of interests.

References

Abrudan, M. I., Smakman, F., Grimbergen, A. J., Westhoff, S., Miller, E. L., van Wezel, G. P., & Rozen, D. E. (2015). Socially mediated induction and suppression of antibiosis during bacterial coexistence. *Proceedings of the National Academy of Sciences*, 112(35), 11054–11059. https://doi.org/10.1073/pnas.1504076112

Agu, P. C., Afiukwa, C. A., Orji, O. U., Ezeh, E. M., Ofoke, I. H., Ogbu, C. O., Ugwuja, E. I., & Aja, P. M. (2023). Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management [Review]. *Scientific Reports*, 13(1), 13398. https://doi.org/10.1038/s41598-023-40160-2

Ahmed, Y., Rebets, Y., Estévez, M. R., Zapp, J., Myronovskyi, M., & Luzhetskyy, A. (2020). Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters. *Microbial Cell Factories*, 19(1), 5. https://doi.org/10.1186/s12934-020-1277-8

Al-Bassam, M. M., Bibb, M. J., Bush, M. J., Chandra, G., & Buttner, M. J. (2014). Response regulator heterodimer formation controls a key stage in *streptomyces* development. *PLos Genetics*, 10(8), e1004554. https://doi.org/10.1371/journal.pgen.1004554

Albert, R. (2005). Scale-free networks in cell biology. *Journal of Cell Science*, 118(Pt 21), 4947–4957. https://doi.org/10.1242/jcs.02714

Albright, J. C., Henke, M. T., Soukup, A. A., McClure, R. A., Thomson, R. J., Keller, N. P., & Kelleher, N. L. (2015). Large-scale metabolomics reveals a complex response of *Aspergillus nidulans* to epigenetic perturbation. *ACS Chemical Biology*, 10(6), 1535–1541. In press. https://doi.org/10.1021/acschembio.5b00025

Anderssen, S., Naômé, A., Jadot, C., Brans, A., Tocquin, P., & Rigali, S. (2022). AURTHO: Autoregulation of transcription factors as facilitator of cis-acting element discovery. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1865(5), 194847. https://doi.org/10.1016/j.bbagrm.2022.194847

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME suite. *Nucleic Acids Research*, 43(W1), W39–W49. https://doi.org/10.1093/nar/gkv416

Balleza, E., López-Bojorquez, L. N., Martínez-Antonio, A., Resendis-Antonio, O., Lozada-Chávez, I., Balderas-Martínez, Y. I., Encarnación, S., & Collado-Vides, J. (2009). Regulation by transcription factors in bacteria: Beyond description. *FEMS Microbiology Review*, 33(1), 133–151. https://doi.org/10.1111/j.1574-6976.2008.00145.x

Bang, S., Galli, M., Crisp, P. A., Gallavotti, A., & Schmitz, R. J. (2022). Identifying transcription factor–DNA interactions using machine learning. *in silico Plants*, 4(2), diac014. https://doi.org/10.1093/insilicoplants/diac014

Baral, B., Akhgari, A., & Metsä-Ketelä, M. (2018). Activation of microbial secondary metabolic pathways: Avenues and challenges. *Synthetic and Systems Biotechnology*, 3(3), 163–178. https://doi.org/10.1016/j.synbio.2018.09.001

Barka, E. A., Vatsa, P., Sanchez, L., Gavaut-Vaillant, N., Jacquard, C., Meier-Kolthoff, J., Klenk, H. P., Clément, C., Oudouch, Y., & van Wezel, G. P. (2016). Taxonomy, physiology, and natural products of the actinobacteria. *Microbiology and Molecular Biology Reviews*, 80(1), 1–43. https://doi.org/10.1128/MMBR.00019-15

Bartlett, A., O'Malley, R. C., Huang, S.-S. C., Galli, M., Nery, J. R., Gallavotti, A., & Ecker, J. R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nature Protocols*, 12(8), 1659–1672. https://doi.org/10.1038/nprot.2017.055

Baumgart, L. A., Lee, J. E., Salamov, A., Dilworth, D. J., Na, H., Mingay, M., Blow, M. J., Zhang, Y., Yoshinaga, Y., Daum, C. G., & O'Malley, R. C. (2021). Persistence and plasticity in bacterial gene regulation. *Nature Methods*, 18(12), 1499–1505. https://doi.org/10.1038/s41592-021-01312-2

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., … Hopwood, D. A. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 417(6885), 141–147. https://doi.org/10.1038/417141a

Bérdy, J. (2005). Bioactive microbial metabolites [Review]. *The Journal of Antibiotics*, 58(1), 1–26. https://doi.org/10.1038/ja.2005.1

Bérdy, J. (2012). Thoughts and facts about antibiotics: Where we are now and where we are heading. *Journal of Antibiotics*, 65(8), 441. https://doi.org/10.1038/ja.2012.54

Bertrand, S., Bohni, N., Schnee, S., Schumpp, O., Gindro, K., & Wolfender, J.-L. (2014). Metabolite induction via microorganism co-culture: A potential way to enhance chemical diversity for drug discovery. *Biotechnology Advances*, 32(6), 1180–1204. https://doi.org/10.1016/j.biotechadv.2014.03.001

Bibb, M. (1996). The regulation of antibiotic production in *Streptomyces coelicolor* A3(2). *Microbiology (Reading, England)*, 142(Pt 6), 1335–1344. https://doi.org/10.1099/13500872-142-6-1335

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes [Review]. *Current Opinion in Microbiology*, 8(2), 208–215. https://doi.org/10.1016/j.mib.2005.02.016

Blat, Y. & Kleckner, N. (1999). Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell*, 98(2), 249–259. https://doi.org/10.1016/s0092-8674(00)81019-3

Blin, K., Shaw, S., Augustijn, H. E., Reitz, Z. L., Biermann, F., Alanjary, M., Fetter, A., Terlouw, B. R., Metcalf, W. W., Helfrich, E. J. N., van Wezel, G. P., Medema, M. H., & Weber, T. (2023). antiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research*, 51(W1), W46–W50. https://doi.org/10.1093/nar/gkad344

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., & Weber, T. (2019). antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, 47(W1), W81–W87. https://doi.org/10.1093/nar/gkz310

Bode, H. B., Bethe, B., Höfs, R., & Zeeck, A. (2002). Big effects from small changes: Possible ways to explore nature's chemical diversity. *Chembiochem*, 3(7), 619–627. https://doi.org/10.1002/1439-7633(20020703)3:7%3C619::aid-cbic619%3E3.0.co;2-9

Botas, A., Pérez-Redondo, R., Rodríguez-García, A., Álvarez-Álvarez, R., Yagüe, P., Manteca, A., & Liras, P. (2018). ArgR of *Streptomyces coelicolor* is a pleiotropic transcriptional regulator: Effect on the transcriptome, antibiotic production, and differentiation in liquid cultures. *Frontiers in Microbiology*, 9, 361. https://doi.org/10.3389/fmicb.2018.00361

Boytsov, A., Abramov, S., Makeev, V. J., & Kulakovskiy, I. V. (2022). Positional weight matrices have sufficient prediction power for analysis of noncoding variants. *F1000Res*, 11, 33. https://doi.org/10.12688/f1000research.75471.3

Bruggeman, F. J., Planqué, R., Molenaar, D., & Teusink, B. (2020). Searching for principles of microbial physiology. *FEMS Microbiology Review*, 44(6), 821–844. https://doi.org/10.1093/femsre/fuaa034

Bulyk, M. L., Johnson, P. L. F., & Church, G. M. (2002). Nucleotides of transcription factor binding sites exert interdependent effects

on the binding affinities of transcription factors. *Nucleic Acids Research*, 30(5), 1255–1261. https://doi.org/10.1093/nar/30.5.1255

Bush, M. J., Bibb, M. J., Chandra, G., Findlay, K. C., & Buttner, M. J. (2013). Genes required for aerial growth, cell division, and chromosome segregation are targets of WhiA before sporulation in Streptomyces venezuelae [Research Support, Non-U.S. Gov't]. *MBio*, 4(5), e00684–00613. https://doi.org/10.1128/mBio.00684-13

Bush, M. J., Chandra, G., Al-Bassam, M. M., Findlay, K. C., & Buttner, M. J. (2019). BldC delays entry into development to produce a sustained period of vegetative growth in *Streptomyces venezuelae*. *MBio*, 10(1), e02812–e02818. https://doi.org/10.1128/mbio.02812-18

Chater, K. F. & Losick, R. (1997). Mycelial life style of *Streptomyces coelicolor* A3(2) and its relatives. In J. A. Shapiro & M. Dworkin (Eds.), *Bacteria as Multicellular Organisms* (pp. 149–182). Oxford University Press.

Chiu, T.-P., Xin, B., Markarian, N., Wang, Y., & Rohs, R. (2020). TFB-Sshape: An expanded motif database for DNA shape features of transcription factor binding sites. *Nucleic Acids Research*, 48(D1), D246–D255. https://doi.org/10.1093/nar/gkz970

Claessen, D., Rozen, D. E., Kuipers, O. P., Sogaard-Andersen, L., & van Wezel, G. P. (2014). Bacterial solutions to multicellularity: A tale of biofilms, filaments and fruiting bodies [Research Support, Non-U.S. Gov't Review]. *Nature Reviews Microbiology*, 12(2), 115–124. https://doi.org/10.1038/nrmicro3178

Cooper, M. A. & Shlaes, D. (2011). Fix the antibiotics pipeline. *Nature*, 472, 32. https://doi.org/10.1038/472032a

Crack, J. C., Munnoch, J., Dodd, E. L., Knowles, F., Al Bassam, M. M., Kamali, S., Holland, A. A., Cramer, S. P., Hamilton, C. J., Johnson, M. K., Thomson, A. J., Hutchings, M. I., & Le Brun, N. E. (2015). NsrR from *Streptomyces coelicolor* is a nitric oxide-sensing [4Fe-4S] cluster protein with a specialized regulatory function. *Journal of Biological Chemistry*, 290(20), 12689–12704. https://doi.org/10.1074/jbc.M115.643072

Craig, M., Lambert, S., Jourdan, S., Tenconi, E., Colson, S., Maciejewska, M., Ongena, M., Martin, J. F., van Wezel, G., & Rigali, S. (2012). Unsuspected control of siderophore production by N-acetylglucosamine in streptomycetes. *Environmental Microbiology Reports*, 4(5), 512–521. https://doi.org/10.1111/j.1758-2229.2012.00354.x

Cruz-Morales, P., Vijgenboom, E., Iruegas-Bocardo, F., Girard, G., Yáñez-Guerra, L. A., Ramos-Aboites, H. E., Pernodet, J.-L., Anné, J., van Wezel, G. P., & Barona-Gómez, F. (2013). The genome sequence of *Streptomyces lividans* 66 reveals a novel tRNA-dependent peptide biosynthetic system within a metal-related genomic island. *Genome Biology and Evolution*, 5(6), 1165–1175. https://doi.org/10.1093/gbe/evt082

Cui, F., Zhang, Z., Cao, C., Zou, Q., Chen, D., & Su, X. (2022). Protein-DNA/RNA interactions: Machine intelligence tools and approaches in the era of artificial intelligence and big data. *Proteomics*, 22(8), e2100197. https://doi.org/10.1002/pmic.202100197

De Mendoza, A., Pflueger, J., & Lister, R. (2019). Capture of a functionally active methyl-CpG binding domain by an arthropod retrotransposon family. *Genome Research*, 29(8), 1277–1286. https://doi.org/10.1101/gr.243774.118

Den Hengst, C. D., Tran, N. T., Bibb, M. J., Chandra, G., Leskiw, B. K., & Buttner, M. J. (2010). Genes essential for morphological development and antibiotic production in *Streptomyces coelicolor* are targets of BldD during vegetative growth. *Molecular Microbiology*, 78(2), 361–379. https://doi.org/10.1111/j.1365-2958.2010.07338.x

Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C., & Rani, V. (2012). DNA-protein interactions: Methods for detection and analysis. *Molecular and Cellular Biochemistry*, 365(1-2), 279–299. https://doi.org/10.1007/s11010-012-1269-z

Eddy, S. R. (2004). What is a hidden Markov model? *Nature Biotechnology*, 22(10), 1315–1316. https://doi.org/10.1038/nbt1004-1315

Eggeling, R. (2018). Disentangling transcription factor binding site complexity. *Nucleic Acids Research*, 46(20), e121. https://doi.org/10.1093/nar/gky683

Elliot, M. A., Buttner, M. J., & Nodwell, J. R. (2008). Multicellular development in *Streptomyces*. In D. E. Whitworth (Ed.), *Myxobacteria: Multicellularity and Differentiation* (pp. 419–438). ASM Press. https://doi.org/10.1128/9781555815677.ch24

Escorcia-Rodríguez, J. M., Tauch, A., & Freyre-González, J. A. (2020). Abasy Atlas v2.2: The most comprehensive and up-to-date inventory of meta-curated, historical, bacterial regulatory networks, their completeness and system-level characterization. *Computational and Structural Biotechnology Journal*, 18, 1228–1237. https://doi.org/10.1016/j.csbj.2020.05.015.

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 [Review]. *Plant Methods*, 17(1), 82. https://doi.org/10.1186/s13007-021-00780-z

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, 39(Web Server issue), W29–37. https://doi.org/10.1093/nar/gkr367

Fischer, M. S., Wu, V. W., Lee, J. E., O'Malley, R. C., & Glass, N. L. (2018). Regulation of cell-to-cell communication and cell wall integrity by a network of MAP kinase pathways and transcription factors in *Neurospora crassa*. *Genetics*, 209(2), 489–506. https://doi.org/10.1534/genetics.118.300904

Fried, M. & Crothers, D. M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Research*, 9(23), 6505–6525. https://doi.org/10.1093/nar/9.23.6505

Frith, M. C., Saunders, N. F. W., Kobe, B., & Bailey, T. L. (2008). Discovering sequence motifs with arbitrary insertions and deletions. *Plos Computational Biology*, 4(5), e1000071. https://doi.org/10.1371/journal.pcbi.1000071

Galagan, J. E., Minch, K., Peterson, M., Lyubetskaya, A., Azizi, E., Sweet, L., Gomes, A., Rustad, T., Dolganov, G., Glotova, I., Abeel, T., Mahwinney, C., Kennedy, A. D., Allard, R., Brabant, W., Krueger, A., Jaini, S., Honda, B., Yu, W.-H., … Schoolnik, G. K. (2013). The *Mycobacterium tuberculosis* regulatory network and hypoxia. *Nature*, 499(7457), 178–183. https://doi.org/10.1038/nature12337

Galas, D. J. & Schmitz, A. (1978). DNAse footprinting: A simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Research*, 5(9), 3157–3170. https://doi.org/10.1093/nar/5.9.3157

Gallagher, L. A., Velazquez, E., Peterson, S. B., Charity, J. C., Radey, M. C., Gebhardt, M. J., Hsu, F., Shull, L. M., Cutler, K. J., Macareno, K., de Moraes, M. H., Penewit, K. M., Kim, J., Andrade, P. A., LaFramboise, T., Salipante, S. J., Reniere, M. L., de Lorenzo, V., Wiggins, P. A., … Mougous, J. D. (2022). Genome-wide protein-DNA interaction site mapping in bacteria using a double-stranded DNA-specific cytosine deaminase. *Nature Microbiology*, 7(6), 844–855. https://doi.org/10.1038/s41564-022-01133-9

Galli, M., Khakhar, A., Lu, Z., Chen, Z., Sen, S., Joshi, T., Nemhauser, J. L., Schmitz, R. J., & Gallavotti, A. (2018). The DNA binding landscape of the maize auxin response factor family. *Nature Communications*, 9(1), 4526. https://doi.org/10.1038/s41467-018-06977-6

Gao, C., Hindra, Mulder, D., Yin, C., & Elliot, M. A. (2012). Crp is a global regulator of antibiotic production in streptomyces. *MBio*, 3(6), e00407–e00412. https://doi.org/10.1128/mBio.00407-12

Gao, Y., Yurkovich, J. T., Seo, S. W., Kabimoldayev, I., Dräger, A., Chen, K., Sastry, A. V., Fang, X., Mih, N., & Yang, L. (2018). Systematic discovery of uncharacterized transcription factors in *Escherichia coli* K-12 MG1655. *Nucleic Acids Research*, 46(20), 10682–10696. https://doi.org/10.1093/nar/gky752.

Garner, M. M. & Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: Application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Research*, 9(13), 3047–3060. https://doi.org/10.1093/nar/9.13.3047

Gavriilidou, A., Kautsar, S. A., Zaburannyi, N., Krug, D., Muller, R., Medema, M. H., & Ziemert, N. (2022). Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes [Research Support, Non-U.S. Gov't]. *Nature Microbiology*, 7(5), 726–735. https://doi.org/10.1038/s41564-022-01110-2

Gilmour, D. S. & Lis, J. T. (1984). Detecting protein-DNA interactions in vivo: Distribution of RNA polymerase on specific bacterial genes. *Proceedings of the National Academy of Sciences*, 81(14), 4275–4279. https://doi.org/10.1073/pnas.81.14.4275

Gomez-Escribano, J. P. & Bibb, M. J. (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology*, 4(2), 207–215. https://doi.org/10.1111/j.1751-7915.2010.00219.x

Guille, M. J. & Kneale, G. G. (1997). Methods for the analysis of DNA-protein interactions. *Molecular Biotechnology*, 8(1), 35–52. https://doi.org/10.1007/BF02762338

Hannenhalli, S. (2008). Eukaryotic transcription factor binding sites—Modeling and integrative search methods. *Bioinformatics*, 24(11), 1325–1331. https://doi.org/10.1093/bioinformatics/btn198

Hannenhalli, S. & Wang, L.-S. (2005). Enhanced position weight matrices using mixture models. *Bioinformatics*, 21(Suppl 1), i204–i212. https://doi.org/10.1093/bioinformatics/bti1001

Hara, H., Ohnishi, Y., & Horinouchi, S. (2009). DNA microarray analysis of global gene regulation by A-factor in *Streptomyces griseus* [Article]. *Microbiology (Reading, England)*, 155(7), 2197–2210. https://doi.org/10.1099/mic.0.027862-0

Hiard, S., Marée, R., Colson, S., Hoskisson, P. A., Titgemeyer, F., van Wezel, G. P., Joris, B., Wehenkel, L., & Rigali, S. (2007). PREDetector: A new tool to identify regulatory elements in bacterial genomes. *Biochemical and Biophysical Research Communications*, 357(4), 861–864. https://doi.org/10.1016/j.bbrc.2007.03.180

Higo, A., Hara, H., Horinouchi, S., & Ohnishi, Y. (2012). Genome-wide distribution of AdpA, a global regulator for secondary metabolism and morphological differentiation in *Streptomyces*, revealed the extent and complexity of the AdpA regulatory network. *DNA Research*, 19(3), 259–273. https://doi.org/10.1093/dnares/dss010

Hopwood, D. A. (2007). *Streptomyces in Nature and Medicine: the Antibiotic Makers*. Oxford University Press.

Hosaka, T., Ohnishi-Kameyama, M., Muramatsu, H., Murakami, K., Tsurumi, Y., Kodani, S., Yoshida, M., Fujie, A., & Ochi, K. (2009). Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nature Biotechnology*, 27(5), 462–464. https://doi.org/10.1038/nbt.1538

Hoshino, S., Wakimoto, T., Onaka, H., & Abe, I. (2015). Chojalactones A-C, cytotoxic butanolides isolated from *Streptomyces sp.* cultivated with mycolic acid containing bacterium. *Organic Letters*, 17(6), 1501–1504. https://doi.org/10.1021/acs.orglett.5b00385

Hoskisson, P. A. & Fernández-Martínez, L. T. (2018). Regulation of specialised metabolites in actinobacteria—expanding the paradigms. *Environmental Microbiology Reports*, 10(3), 231–238. https://doi.org/10.1111/1758-2229.12629

Hsiao, N.-H., Gottelt, M., & Takano, E. (2009). Chapter 6. Regulation of antibiotic production by bacterial hormones. *Methods in Enzymology*, 458, 143–157. https://doi.org/10.1016/S0076-6879(09)04806-X

Hwang, S., Lee, N., Choe, D., Lee, Y., Kim, W., Kim, J. H., Kim, G., Kim, H., Ahn, N. H., Lee, B. H., Palsson, B. O., & Cho, B. K. (2022). System-level analysis of transcriptional and translational regulatory elements in *Streptomyces griseus*. *Frontiers in Bioengineering and Biotechnology*, 10, 844200. https://doi.org/10.3389/fbioe.2022.844200

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., & Ōmura, S. (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nature Biotechnology*, 21(5), 526–531. https://doi.org/10.1038/nbt820

Ikeda, H., Kazuo, S.-Y., & Omura, S. (2014). Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *Journal of Industrial Microbiology & Biotechnology*, 41(2), 233–250. https://doi.org/10.1007/s10295-013-1327-x

Jakimowicz, D. & van Wezel, G. P. (2012). Cell division and DNA segregation in *Streptomyces*: How to build a septum in the middle of nowhere? *Molecular Microbiology*, 85(3), 393–404. https://doi.org/10.1111/j.1365-2958.2012.08107.x

Jayaram, N., Usvyat, D., Martin, R., & A. C. (2016). Evaluating tools for transcription factor binding site prediction. *BMC Bioinformatics*, 17(1), 1–12. https://doi.org/10.1186/s12859-016-1298-9

Jeong, H., Kim, Y., & Lee, H.-S. (2021). OsnR is an autoregulatory negative transcription factor controlling redox-dependent stress responses in *Corynebacterium glutamicum*. *Microbial Cell Factories*, 20(1), 203. https://doi.org/10.1186/s12934-021-01693-1

Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of *in vivo* protein-DNA interactions. *Science*, 316(5830), 1497–1502. https://doi.org/10.1126/science.1141319

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Jungwirth, B., Sala, C., Kohl, T. A., Uplekar, S., Baumbach, J., Cole, S. T., Pühler, A., & Tauch, A. (2013). High-resolution detection of DNA binding sites of the global transcriptional regulator GlxR in *Corynebacterium glutamicum*. *Microbiology (Reading, England)*, 159(Pt 1), 12–22. https://doi.org/10.1099/mic.0.062059-0

Kelemen, G. H. & Buttner, M. J. (1998). Initiation of aerial mycelium formation in *streptomyces*. *Current Opinion in Microbiology*, 1(6), 656–662. https://doi.org/10.1016/s1369-5274(98)80111-2 [pii]

Kerkhoven, E. J. (2022). Advances in constraint-based models: Methods for improved predictive power based on resource allocation constraints. *Current Opinion in Microbiology*, 68, 102168. https://doi.org/10.1016/j.mib.2022.102168

Kidder, B. L., Hu, G., & Zhao, K. (2011). ChIP-Seq: Technical considerations for obtaining high-quality data. *Nature Immunology*, 12(10), 918–922. https://doi.org/10.1038/ni.2117

Kramer, J., Özkaya, Ö., & Kümmerli, R. (2020). Bacterial siderophores in community and host interactions. *Nature Reviews Microbiology*, 18(3), 152–163. https://doi.org/10.1038/s41579-019-0284-4

Ladunga, I. (2010). An overview of the computational analyses and discovery of transcription factor binding sites. *Methods in Molecular Biology*, 674, 1–22. https://doi.org/10.1007/978-1-60761-854-6_1

Lee, L.-H., Goh, B.-H., & Chan, K.-G. (2020). Actinobacteria: Prolific producers of bioactive metabolites. *Frontiers in Microbiology*, 11, 138. https://doi.org/10.3389/fmicb.2020.01612

Li, J., Wang, N., Tang, Y., Cai, X., Xu, Y., Liu, R., Wu, H., & Zhang, B. (2019). Developmental regulator BldD directly regulates

lincomycin biosynthesis in *Streptomyces lincolnensis*. *Biochemical and Biophysical Research Communications*, 518(3), 548–553. https://doi.org/10.1016/j.bbrc.2019.08.079

Li, X., Wang, J., Li, S., Ji, J., Wang, W., & Yang, K. (2015). ScbR- and ScbR2-mediated signal transduction networks coordinate complex physiological responses in *Streptomyces coelicolor* [Research Support, Non-U.S. Gov't]. *Scientific Reports*, 5(1), 14831. https://doi.org/10.1038/srep14831

Liao, C. H., Yao, L., Xu, Y., Liu, W. B., Zhou, Y., & Ye, B. C. (2015). Nitrogen regulator GlnR controls uptake and utilization of non-phosphotransferase-system carbon sources in Actinomycetes. *Proceedings of the National Academy of Sciences*, 112(51), 15630–15635. https://doi.org/10.1073/pnas.1508465112

Liu, G., Chater, K. F., Chandra, G., Niu, G., & Tan, H. (2013). Molecular regulation of antibiotic biosynthesis in *Streptomyces* [Research Support, Non-U.S. Gov't]. *Microbiology and Molecular Biology Reviews*, 77(1), 112–143. https://doi.org/10.1128/MMBR.00054-12

Liu, Z., Zhao, Y., Huang, C., & Luo, Y. (2021). Recent advances in silent gene cluster activation in *streptomyces*. *Frontiers in Bioengineering and Biotechnology*, 9, 632230. https://doi.org/10.3389/fbioe.2021.632230

Ma, X., Ezer, D., Navarro, C., & Adryan, B. (2015). Reliable scaling of position weight matrices for binding strength comparisons between transcription factors. *BMC Bioinformatics*, 16(1), 1–13. https://doi.org/10.1186/s12859-015-0666-1

Maaskola, J. & Rajewsky, N. (2014). Binding site discovery from nucleic acid sequences by discriminative learning of hidden Markov models. *Nucleic Acids Research*, 42(21), 12995–13011. https://doi.org/10.1093/nar/gku1083

Madera, M. & Gough, J. (2002). A comparison of profile hidden Markov model procedures for remote homology detection. *Nucleic Acids Research*, 30(19), 4321–4328. https://doi.org/10.1093/nar/gkf544

Manteca, A., Fernandez, M., & Sanchez, J. (2005). Mycelium development in *Streptomyces antibioticus* ATCC11891 occurs in an orderly pattern which determines multiphase growth curves [Research Support, Non-U.S. Gov't]. *BMC Microbiology*, 5(1), 51. https://doi.org/10.1186/1471-2180-5-51

Martín, J. F., Rodríguez-García, A., & Liras, P. (2017). The master regulator PhoP coordinates phosphate and nitrogen metabolism, respiration, cell differentiation and antibiotic biosynthesis: Comparison in *Streptomyces coelicolor* and *Streptomyces avermitilis*. *The Journal of Antibiotics*, 70(5), 534–541. https://doi.org/10.1038/ja.2017.19

Mathelier, A. & Wasserman, W. W. (2013). The next generation of transcription factor binding site prediction. *Plos Computational Biology*, 9(9), e1003214. https://doi.org/10.1371/journal.pcbi.1003214

Mathelier, A., Xin, B., Chiu, T. P., Yang, L., Rohs, R., & Wasserman, W. W. (2016). DNA shape features improve transcription factor binding site predictions *in vivo*. *Cell Systems*, 3(3), 278–286 e274. https://doi.org/10.1016/j.cels.2016.07.001

McCormick, J. R. & Flardh, K. (2012). Signals and regulators that govern *Streptomyces* development [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review]. *FEMS Microbiology Review*, 36(1), 206–231. https://doi.org/10.1111/j.1574-6976.2011.00317.x

McLean, T. C., Hoskisson, P. A., & Seipke, R. F. (2016). Coordinate regulation of antimycin and candicidin biosynthesis. *mSphere*, 1(6), e00305–e00316. https://doi.org/10.1128/mSphere.00305-16

Mevaere, J., Goulard, C., Schneider, O., Sekurova, O. N., Ma, H., Zirah, S., Afonso, C., Rebuffat, S., Zotchev, S. B., & Li, Y. (2018). An orthogonal system for heterologous expression of actinobacterial lasso peptides in *Streptomyces* hosts. *Scientific Reports*, 8(1), 8232. https://doi.org/10.1038/s41598-018-26620-0

Minch, K. J., Rustad, T. R., Peterson, E. J. R., Winkler, J., Reiss, D. J., Ma, S., Hickey, M., Brabant, W., Morrison, B., Turkarslan, S., Mawhinney, C., Galagan, J. E., Price, N. D., Baliga, N. S., & Sherman, D. R. (2015). The DNA-binding network of *mycobacterium tuberculosis*. *Nature Communications*, 6(1), 5829. https://doi.org/10.1038/ncomms6829

Moon, K., Xu, F., & Seyedsayamdost, M. R. (2019a). Cebulantin, a cryptic lanthipeptide antibiotic uncovered using bioactivity-coupled HiTES. *Angewandte Chemie International Edition*, 58(18), 5973–5977. https://doi.org/10.1002/anie.201901342

Moon, K., Xu, F., Zhang, C., & Seyedsayamdost, M. R. (2019b). Bioactivity-HiTES unveils cryptic antibiotics encoded in actinomycete bacteria. *ACS Chemical Biology*, 14(4), 767–774. https://doi.org/10.1021/acschembio.9b00049

Moreno-Campuzano, S., Janga, S. C., & Pérez-Rueda, E. (2006). Identification and analysis of DNA-binding transcription factors in Bacillus subtilis and other firmicutes-a genomic approach. *BMC Genomics*, 7(1), 1–10. https://doi.org/10.1186/1471-2164-7-147

Mullowney, M. W., Duncan, K. R., Elsayed, S. S., Garg, N., van der Hooft, J. J., Martin, N. I., Meijer, D., Terlouw, B. R., Biermann, F., & Blin, K. (2023). Artificial intelligence for natural product drug discovery. *Nature Reviews Drug Discovery*, 22(11), 895–916. https://doi.org/10.1038/s41573-023-00774-7

Munnoch, J. T., Martinez, M. T. P., Svistunenko, D. A., Crack, J. C., Le Brun, N. E., & Hutchings, M. I. (2016). Characterization of a putative NsrR homologue in *Streptomyces venezuelae* reveals a new member of the Rrf2 superfamily. *Scientific Reports*, 6(1), 31597. https://doi.org/10.1038/srep31597

Newman, D. J. & Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products*, 83(3), 770–803. https://doi.org/10.1021/acs.jnatprod.9b01285

Nguyen, T. T. & Androulakis, I. P. (2009). Recent advances in the computational discovery of transcription factor binding sites. *Algorithms*, 2(1), 582–605. https://doi.org/10.3390/a2010582

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M., & Horinouchi, S. (2008). Genome sequence of the streptomycin-producing microorganism streptomyces griseus IFO 13350. *Journal of Bacteriology*, 190(11), 4050–4060. https://doi.org/10.1128/jb.00204-08

Oladejo, D. O., Duselu, G. O., Dokunmu, T. M., Isewon, I., Oyelade, J., Okafor, E., Iweala, E. E., & Adebiyi, E. (2023). In silico structure prediction, molecular docking, and dynamic simulation of *plasmodium falciparum* AP2-I transcription factor. *Bioinformatics and Biology Insights*, 17, 11779322221149616. https://doi.org/10.1177/11779322221149616

Oliynyk, M., Samborskyy, M., Lester, J. B., Mironenko, T., Scott, N., Dickens, S., Haydock, S. F., & Leadlay, P. F. (2007). Complete genome sequence of the erythromycin-producing bacterium saccharopolyspora erythraea NRRL23338. *Nature Biotechnology*, 25(4), 447–453. https://doi.org/10.1038/nbt1297

O'Malley, R. C., Huang, S.-S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., Galli, M., Gallavotti, A., & Ecker, J. R. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*, 166(6), 1598. https://doi.org/10.1016/j.cell.2016.08.063

Otani, H. & Mouncey, N. J. (2022). RIVIT-seq enables systematic identification of regulons of transcriptional machineries. *Nature Communications*, 13(1), 3502. https://doi.org/10.1038/s41467-022-31191-w

Pavesi, G., Mereghetti, P., Mauri, G., & Pesole, G. (2004). Weeder Web: Discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Research*, 32(Web Server issue), W199–203. https://doi.org/10.1093/nar/gkh465

Payne, D. J., Gwynn, M. N., Holmes, D. J., & Pompliano, D. L. (2006). Drugs for bad bugs: Confronting the challenges of antibacterial discovery. *Nature Reviews Drug Discovery*, 6(1), 29–40. https://doi.org/10.1038/nrd2201

Perez-Redondo, R., Rodriguez-Garcia, A., Botas, A., Santamarta, I., Martin, J. F., & Liras, P. (2012). ArgR of *Streptomyces coelicolor* is a versatile regulator. *PLoS One*, 7(3), e32697. https://doi.org/10.1371/journal.pone.0032697

Purtov, Y. A., Tutukina, M. N., Nikulin, A. D., & Ozoline, O. N. (2019). The topology of the contacts of potential ligands for the UxuR transcription factor of *Escherichia coli* as revealed by flexible molecular docking. *Molecular Biophysics*, 64(1), 49–56. https://doi.org/10.1134/S0006350919010160

Qian, Y., Lee, J. H., & Holmes, R. K. (2002). Identification of a DtxR-regulated operon that is essential for siderophore-dependent iron uptake in *Corynebacterium diphtheriae*. *Journal of Bacteriology*, 184(17), 4846–4856. https://doi.org/10.1128/JB.184.17.4846-4856.2002

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P., & Young, R. A. (2000). Genome-wide location and function of DNA binding proteins. *Science*, 290(5500), 2306–2309. https://doi.org/10.1126/science.290.5500.2306

Rico, S., Santamaría, R. I., Yepes, A., Rodríguez, H., Laing, E., Bucca, G., Smith, C. P., & Díaz, M. (2014). Deciphering the regulon of *streptomyces coelicolor* AbrC3, a positive response regulator of antibiotic production. *Applied and Environmental Microbiology*, 80(8), 2417–2428. https://doi.org/10.1128/aem.03378-13

Rigali, S., Anderssen, S., Naômé, A., & van Wezel, G. P. (2018). Cracking the regulatory code of biosynthetic gene clusters as a strategy for natural product discovery. *Biochemical Pharmacology*, 153, 24–34. https://doi.org/10.1016/j.bcp.2018.01.007

Rigali, S., Nothaft, H., Noens, E. E., Schlicht, M., Colson, S., Muller, M., Joris, B., Koerten, H. K., Hopwood, D. A., Titgemeyer, F., & van Wezel, G. P. (2006). The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. *Molecular Microbiology*, 61(5), 1237–1251. https://doi.org/10.1111/j.1365-2958.2006.05319.x.

Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A. W., Hopwood, D. A., & van Wezel, G. P. (2008). Feast or famine: The global regulator DasR links nutrient stress to antibiotic production by *streptomyces* [Research Support, Non-U.S. Gov't]. *Embo Reports*, 9(7), 670–675. https://doi.org/10.1038/embor.2008.83

Romero-Rodriguez, A., Robledo-Casados, I., & Sanchez, S. (2015). An overview on transcriptional regulators in *streptomyces*. *Biochimica Et Biophysica Acta*, 1849(8), 1017–1039. https://doi.org/10.1016/j.bbagrm.2015.06.007

Rudrappa, M., Nayaka, S., & Kumar, R. S. (2023). *In silico* molecular docking approach of melanin against melanoma causing MITF proteins and anticancer, oxidation-reduction, photoprotection, and drug-binding affinity properties of extracted melanin from *Streptomyces* sp. strain MR28. *Applied Biochemistry and Biotechnology*, 195(7), 4368–4386. https://doi.org/10.1007/s12010-023-04358-4

Santana-Garcia, W., Castro-Mondragon, J. A., Padilla-Gálvez, M., Nguyen, N. T. T., Elizondo-Salas, A., Ksouri, N., Gerbes, F., Thieffry, D., Vincens, P., Contreras-Moreira, B., van Helden, J., Thomas-Chollier, M., & Medina-Rivera, A. (2022). RSAT 2022: Regulatory sequence analysis tools. *Nucleic Acids Research*, 50(W1), W670–W676. https://doi.org/10.1093/nar/gkac312

Santos-Beneit, F., Rodriguez-Garcia, A., Sola-Landa, A., & Martin, J. F. (2009). Cross-talk between two global regulators in *streptomyces*: PhoP and AfsR interact in the control of afsS, pstS and phoRP transcription. *Molecular Microbiology*, 72(1), 53–68. https://doi.org/10.1111/j.1365-2958.2009.06624.x

Schnepf, M., von Reutern, M., Ludwig, C., Jung, C., & Gaul, U. (2020). Transcription factor binding affinities and DNA shape readout. *Iscience*, 23(11), 101694. https://doi.org/10.1016/j.isci.2020.101694

Skinnider, M. A., Johnston, C. W., Gunabalasingam, M., Merwin, N. J., Kieliszek, A. M., MacLellan, R. J., Li, H., Ranieri, M. R. M., Webster, A. L. H., Cao, M. P. T., Pfeifle, A., Spencer, N., To, Q. H., Wallace, D. P., Dejong, C. A., & Magarvey, N. A. (2020). Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nature Communications*, 11(1), 1–9. https://doi.org/10.1038/s41467-020-19986-1

Šmídová, K., Ziková, A., Pospíšil, J., Schwarz, M., Bobek, J., & Vohradsky, J. (2019). DNA mapping and kinetic modeling of the HrdB regulon in *Streptomyces coelicolor*. *Nucleic Acids Research*, 47(2), 621–633. https://doi.org/10.1093/nar/gky1018

Sola-Landa, A., Rodríguez-García, A., Franco-Domínguez, E., & Martín, J. F. (2005). Binding of PhoP to promoters of phosphate-regulated genes in *streptomyces coelicolor*: Identification of PHO boxes. *Molecular Microbiology*, 56(5), 1373–1385. https://doi.org/10.1111/j.1365-2958.2005.04631.x

Solomon, M. J. & Varshavsky, A. (1985). Formaldehyde-mediated DNA-protein crosslinking: A probe for in vivo chromatin structures. *Proceedings of the National Academy of Sciences*, 82(19), 6470–6474. https://doi.org/10.1073/pnas.82.19.6470

Spohn, M., Wohlleben, W., & Stegmann, E. (2016). Elucidation of the zinc-dependent regulation in *Amycolatopsis japonicum* enabled the identification of the ethylenediamine-disuccinate ([S,S]-EDDS) genes. *Environmental Microbiology*, 18(4), 1249–1263. https://doi.org/10.1111/1462-2920.13159

Staden, R. (1984). Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Research*, 12(1Part2), 505–519. https://doi.org/10.1093/nar/12.1Part2.505

Stormo, G. D. (2000). DNA binding sites: Representation and discovery. *Bioinformatics*, 16(1), 16–23. https://doi.org/10.1093/bioinformatics/16.1.16

Stratton, K. J., Bush, M. J., Chandra, G., Stevenson, C. E. M., Findlay, K. C., & Schlimpert, S. (2022). Genome-wide identification of the LexA-mediated DNA damage response in *streptomyces venezuelae*. *Journal of Bacteriology*, 204(8), e0010822. https://doi.org/10.1128/jb.00108-22

Sugiyama, R., Nishimura, S., Ozaki, T., Asamizu, S., Onaka, H., & Kakeya, H. (2015). 5-Alkyl-1,2,3,4-tetrahydroquinolines, new membrane-interacting lipophilic metabolites produced by combined culture of *Streptomyces nigrescens* and *Tsukamurella pulmonis*. *Organic Letters*, 17(8), 1918–1921. https://doi.org/10.1021/acs.orglett.5b00607

Swiatek-Polatynska, M. A., Bucca, G., Laing, E., Gubbens, J., Titgemeyer, F., Smith, C. P., Rigali, S., & van Wezel, G. P. (2015). Genome-wide analysis of In vivo binding of the master regulator DasR in Streptomyces coelicolor identifies novel non-anonical targets. *PLoS One*, 10(4), e0122479. https://doi.org/10.1371/journal.pone.0122479

Świątek-Połatyńska, M. A., Bucca, G., Laing, E., Gubbens, J., Titgemeyer, F., Smith, C. P., Rigali, S., & van Wezel, G. P. (2015). Genome-wide analysis of *in vivo* binding of the master regulator DasR in *Streptomyces coelicolor* identifies novel non-canonical targets. *PLoS One*, 10(4), e0122479. https://doi.org/10.1371/journal.pone.0122479

Tanaka, Y., Komatsu, M., Okamoto, S., Tokuyama, S., Kaji, A., Ikeda, H., & Ochi, K. (2009). Antibiotic overproduction by *rpsL* and *rsmG* mutants of various actinomycetes. *Applied and Environmental Microbiology*, 75(14), 4919–4922. https://doi.org/10.1128/AEM.00681-09

Tenconi, E., Traxler, M. F., Hoebreck, C., van Wezel, G. P., & Rigali, S. (2018). Production of prodiginines is part of a programmed cell death process in *Streptomyces coelicolor*. *Frontiers in Microbiology*, 9, 1742. https://doi.org/10.3389/fmicb.2018.01742

Tiffert, Y., Supra, P., Wurm, R., Wohlleben, W., Wagner, R., & Reuther, J. (2008). The *streptomyces coelicolor* GlnR regulon: Identification of new GlnR targets and evidence for a central role of GlnR in nitrogen metabolism in actinomycetes. *Molecular Microbiology*, 67(4), 861–880. https://doi.org/10.1111/j.1365-2958.2007.06092.x

Traxler, M. F., Watrous, J. D., Alexandrov, T., Dorrestein, P. C., & Kolter, R. (2013). Interspecies interactions stimulate diversification of the streptomyces coelicolor secreted metabolome. *MBio*, 4(4), e00459–e00513. https://doi.org/10.1128/mBio.00459-13

Trerotola, M., Antolini, L., Beni, L., Guerra, E., Spadaccini, M., Verzulli, D., Moschella, A., & Alberti, S. (2022). A deterministic code for transcription factor-DNA recognition through computation of binding interfaces. *NAR Genomics and Bioinformatics*, 4(1), lqac008. https://doi.org/10.1093/nargab/lqac008

Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249(4968), 505–510. https://doi.org/10.1126/science.2200121

Uguru, G. C., Stephens, K. E., Stead, J. A., Towle, J. E., Baumberg, S., & McDowall, K. J. (2005). Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor* [Research Support, Non-U.S. Gov't]. *Molecular Microbiology*, 58(1), 131–150. https://doi.org/10.1111/j.1365-2958.2005.04817.x

Urem, M., Swiatek-Polatynska, M. A., Rigali, S., & van Wezel, G. P. (2016). Intertwining nutrient-sensory networks and the control of antibiotic production in *streptomyces*. *Molecular Microbiology*, 102(2), 183–195. https://doi.org/10.1111/mmi.13464

van Bergeijk, D. A., Terlouw, B. R., Medema, M. H., & van Wezel, G. P. (2020). Ecology and genomics of actinobacteria: New concepts for natural product discovery [Review]. *Nature Reviews Microbiology*, 18(10), 546–558. https://doi.org/10.1038/s41579-020-0379-y

van der Heul, H. U., Bilyk, B. L., McDowall, K. J., Seipke, R. F., & van Wezel, G. P. (2018). Regulation of antibiotic production in actinobacteria: New perspectives from the post-genomic era [Review]. *Natural Product Reports*, 35(6), 575–604. https://doi.org/10.1039/c8np00012c

van Wezel, G. P., White, J., Hoogvliet, G., & Bibb, M. J. (2000). Application of *redD*, the transcriptional activator gene of the undecylprodigiosin biosynthetic pathway, as a reporter for transcriptional activity in *Streptomyces coelicolor* A3(2) and *Streptomyces lividans*. *Journal of Molecular Microbiology and Biotechnology*, 2(4), 551–556. https://doi.org/11075931

van Wezel, G. P., Woudt, L. P., Vervenne, R., Verdurmen, M. L., Vijgenboom, E., & Bosch, L. (1994). Cloning and sequencing of the *tuf* genes of *Streptomyces coelicolor* A3(2). *Biochimica Et Biophysica Acta*, 1219(2), 543–547. https://doi.org/7918656.

Wang, H., Liu, Y., Cheng, X., Zhang, Y., Li, S., Wang, X., & Xiang, W. (2022a). Titer improvement of milbemycins via coordinating metabolic competition and transcriptional co-activation controlled by *Streptomyces* antibiotic regulatory protein family regulator in *Streptomyces bingchenggensis*. *Biotechnology and Bioengineering*, 119(5), 1252–1263. https://doi.org/10.1002/bit.28044

Wang, J. & Zhao, G. P. (2009). GlnR positively regulates *nasA* transcription in *Streptomyces coelicolor*. *Biochemical and Biophysical Research Communications*, 386(1), 77–81. https://doi.org/10.1016/j.bbrc.2009.05.147

Wang, M., Tian, Y., Xu, L., Zhang, F., Lu, H., Li, M., & Li, B. (2022b). High osmotic stress increases OmpK36 expression through the regulation of KbvR to decrease the antimicrobial resistance of *Klebsiella pneumoniae*. *Microbiology Spectrum*, 10(3), e0050722. https://doi.org/10.1128/spectrum.00507-22

Wang, S., Zhang, Q., Shen, Z., He, Y., Chen, Z.-H., Li, J., & Huang, D.-S. (2021a). Predicting transcription factor binding sites using DNA shape features based on shared hybrid deep learning architecture. *Molecular Therapy-Nucleic Acids*, 24, 154–163. https://doi.org/10.1016/j.omtn.2021.02.014

Wang, T., Sun, W., Fan, L., Hua, C., Wu, N., Fan, S., Zhang, J., Deng, X., & Yan, J. (2021b). An atlas of the binding specificities of transcription factors in Pseudomonas aeruginosa directs prediction of novel regulators in virulence. *Elife*, 10, e61885. https://doi.org/10.7554/eLife.61885.

Wasserman, W. W. & Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. *Nature Reviews Genetics*, 5(4), 276–287. https://doi.org/10.1038/nrg1315

Wetzel, J. L., Zhang, K., & Singh, M. (2022). Learning probabilistic protein-DNA recognition codes from DNA-binding specificities using structural mappings. *Genome Research*, 32(9), 1776–1786. https://doi.org/10.1101/gr.276606.122

Wheeler, T. J. & Eddy, S. R. (2013). nhmmer: DNA homology search with profile HMMs. *Bioinformatics*, 29(19), 2487–2489. https://doi.org/10.1093/bioinformatics/btt403

Wietzorrek, A. & Bibb, M. (1997). A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. *Molecular Microbiology*, 25(6), 1181–1184. https://doi.org/10.1046/j.1365-2958.1997.5421903.x.

Willey, J. M. & Gaskell, A. A. (2011). Morphogenetic signaling molecules of the streptomycetes. *Chemical Reviews*, 111(1), 174–187. https://doi.org/10.1021/cr1000404

Wright, E. S. & Vetsigian, K. H. (2016). Inhibitory interactions promote frequent bistability among competing bacteria. *Nature Communications*, 7(1), 11274. https://doi.org/10.1038/ncomms11274

Xu, Y., Du, X., Yu, X., Jiang, Q., Zheng, K., Xu, J., & Wang, P. (2022). Recent advances in the heterologous expression of biosynthetic gene clusters for marine natural products. *Marine Drugs*, 20(6), 341. https://doi.org/10.3390/md20060341

Yamazaki, H., Tomono, A., Ohnishi, Y., & Horinouchi, S. (2004). DNA-binding specificity of AdpA, a transcriptional activator in the A-factor regulatory cascade in *Streptomyces griseus*. *Molecular Microbiology*, 53(2), 555–572. https://doi.org/10.1111/j.1365-2958.2004.04153.x

Yan, H., Lu, X., Sun, D., Zhuang, S., Chen, Q., Chen, Z., Li, J., & Wen, Y. (2020). BldD, a master developmental repressor, activates antibiotic production in two *Streptomyces* species. *Molecular Microbiology*, 113(1), 123–142. https://doi.org/10.1111/mmi.14405

Yang, J. & Ramsey, S. A. (2015). A DNA shape-based regulatory score improves position-weight matrix-based recognition of transcription factor binding sites. *Bioinformatics*, 31(21), 3445–3450. https://doi.org/10.1093/bioinformatics/btv391

Yang, R., Liu, X., Wen, Y., Song, Y., Chen, Z., & Li, J. (2015). The PhoP transcription factor negatively regulates avermectin biosynthesis in *Streptomyces avermitilis*. *Applied Microbiology and Biotechnology*, 99(24), 10547–10557. https://doi.org/10.1007/s00253-015-6921-6

Yoon, B.-J. (2009). Hidden markov models and their applications in biological sequence analysis. *Current Genomics*, 10(6), 402–415. https://doi.org/10.2174/138920209789177575

Yoon, V. & Nodwell, J. R. (2014). Activating secondary metabolism with stress and chemicals. *Journal of Industrial Microbiology & Biotechnology*, 41(2), 415–424. https://doi.org/10.1007/s10295-013-1387-y

Zhang, H., Ouyang, Z., Zhao, N., Han, S., & Zheng, S. (2022). Transcriptional regulation of the creatine utilization genes of

Corynebacterium glutamicum ATCC 14067 by AmtR, a Central nitrogen regulator. *Frontiers in Bioengineering and Biotechnology*, 10, 816628. https://doi.org/10.3389/fbioe.2022.816628

Zhang, H., Sun, T., Cao, X., Wang, Y., Ma, Z., Wang, Y., Yang, N., Xu, M., Deng, X., Li, H., Wang, B., Yi, J., Wang, Z., Zhang, Q., & Chen, C. (2023). Scanning iron response regulator binding sites using Dap-seq in the *Brucella* genome. *PLOS Neglected Tropical Diseases*, 17(7), e0011481. https://doi.org/10.1371/journal.pntd.0011481

Zhang, Q., Huang, Q., Fang, Q., Li, H., Tang, H., Zou, G., Wang, D., Li, S., Bei, W., Chen, H., Li, L., & Zhou, R. (2020). Identification of genes regulated by the two-component system response regulator NarP of *Actinobacillus pleuropneumoniae* via DNA-affinity-purified sequencing [Evaluation Study]. *Microbiological Research*, 230, 126343. https://doi.org/10.1016/j.micres.2019.126343

Zheng, Y., Sun, C. F., Fu, Y., Chen, X. A., Li, Y. Q., & Mao, X. M. (2019). Dual regulation between the two-component system PhoRP and AdpA regulates antibiotic production in *Streptomyces*. *Journal of Industrial Microbiology & Biotechnology*, 46(5), 725–737. https://doi.org/10.1007/s10295-018-02127-5

Zhu, H., Swierstra, J., Wu, C., Girard, G., Choi, Y. H., van Wamel, W., Sandiford, S. K., & van Wezel, G. P. (2014). Eliciting antibiotics active against the ESKAPE pathogens in a collection of actinomycetes isolated from mountain soils. *Microbiology (Reading, England)*, 160(8), 1714. https://doi.org/10.1099/mic.0.078295-0

Zorro-Aranda, A., Escorcia-Rodríguez, J. M., González-Kise, J. K., & Freyre-González, J. A. (2022). Curation, inference, and assessment of a globally reconstructed gene regulatory network for *Streptomyces coelicolor*. *Scientific Reports*, 12(1), 2840. https://doi.org/10.1038/s41598-022-06658-x

Received: January 22, 2024. **Accepted:** April 2, 2024