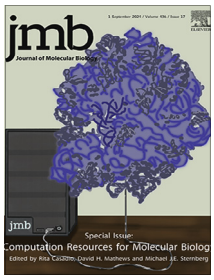
Review Article



The Enigma of Transcriptional Activation Domains

Alexandre M. Erkine * **Marcos A. Oliveira and Caleb A. Class**

Butler University, *Indianapolis, IN 46208, USA*

Correspondence to Alexandre M. Erkine: *aerkine@butler.edu (A.M. Erkine)*
https://doi.org/10.1016/j.jmb.2024.168766
Edited by Monika Fuxreiter

Abstract

Activation domains (ADs) of eukaryotic gene activators remain enigmatic for decades as short, extremely variable sequences which often are intrinsically disordered in structure and interact with an uncertain number of targets. The general absence of specificity increasingly complicates the utilization of the widely accepted mechanism of AD function by recruitment of coactivators. The long-standing enigma at the heart of molecular biology demands a fundamental rethinking of established concepts. Here, we review the experimental evidence supporting a novel mechanistic model of gene activation, based on ADs functioning via surfactant-like near-stochastic interactions with gene promoter nucleosomes. This new model is consistent with recent information-rich experimental data obtained using high-throughput synthetic biology and bioinformatics analysis methods, including machine learning. We clarify why the conventional biochemical principle of specificity for sequence, structures, and interactions fails to explain activation domain function. This perspective provides connections to the liquid–liquid phase separation model, signifies near-stochastic interactions as fundamental for the biochemical function, and can be generalized to other cellular functions.

Introduction

As an initial step in gene expression, the regulation of gene transcription is foundational for molecular biology and the life sciences generally. The first step in this process involves numerous gene-specific transcription factors, also called gene activators, that recognize a specific DNA sequence in a cognate gene promoter and, alone or in cooperation with other activators, initiate a chain of events necessary for gene activation. This process was and is a focus of intense study, with progress in the field reviewed on a regular basis. However, recent high-throughput experimental results and bioinformatics analyses increasingly indicate that the process of gene activation involves steps and players that defy the conventional logic that protein sequence determines protein structure, which in turn controls protein function via specific interactions. Contrary to this fundamental biochemistry paradigm, the short critical parts of gene activators, activation domains (ADs), are astronomically variable in sequence with some estimates of up to 10^{24} possible sequence variants for an individual activator,[1–3] are intrinsically disordered, and have "fuzzy" interactions with an uncertain number of targets. These interactions not only likely happen with low affinity and specificity but, contrary to fundamental biochemistry principles, are required to be at the near-stochastic level for successful gene activation. Here, we will briefly review the general mechanism of gene expression activation, formulate and interrogate the fundamental biochemical enigma of ADs, analyze the reason for the problem to be so hard and

long-standing, and provide a solution for the problem and perspectives for the future.

General mechanism of gene expression activation

The general mechanism driving gene expression was elucidated decades ago,[4,5] and our knowledge of this process is becoming increasingly detailed.[6] The main steps after activator binding to the promoter are the recruitment of coactivator enzymatic complexes that help remodel the gene promoter chromatin, enabling the assembly of a preinitiation transcription complex (PIC). Alternatively, gene activators may directly recruit PIC components, including the Mediator complex and significantly less-sequence-specific general transcription factors (GTFs). GTFs and the Mediator complex in turn carry the key transcriptional enzyme, RNA polymerase, which in the case of protein-encoding genes is RNA polymerase II.

Transcription initiation can be augmented via *cis*-regulatory gene elements (enhancers) that contain DNA sequences that are recognized by activator-like proteins, which cooperate with promoter activators and co-recruit promoter-binding proteins. It has been proposed that upon induction, enhancers and promoters organize themselves into larger condensates called super-enhancers, which create a transient liquid–liquid phase separation (LLPS) that facilitates transcription by concentrating all the necessary transcription enzymes and co-factors.[7,8] The assembly and disassembly of these condensates depends on multivalent interactions among proteins and nucleic acids and is regulated by intrinsically disordered regions of transcription factors, as well as Mediator subunits, and other coactivators.[9–11] The exact composition of super-enhancer condensates is likely defined by a specific pathway of gene regulation and is a subject of current investigations.[6]

Over the years, the core of important functional contacts has become clearer. The hub of interactions appears to be the large Mediator complex, which contains 21 to 26 core subunits, depending on the species.[12] Various interactions of transcription factors, including the ADs of activators, with Mediator have been demonstrated *in vitro* and *in vivo.* Several publications provide convincing data showing that AD mutations that decrease the specificity and the strength of these interactions *in vitro* also decrease functionality *in vivo*.[13–15] The Med15 subunit, with multiple activator-binding regions, seems to be especially important in yeast, although an additional 75 known and novel Mediator interacting targets were identified by mass spectrometry in higher eukaryotes.[16]

Among coactivators, the SAGA complex, which shares subunits with other coactivators such as NuA4, also appears to be an important target for ADs. Similar to the interactions of Mediator complex, the interactions between the critical SAGA/Nua4 complex subunit Tra1 and gene activators were correlated with functionality *in vivo*.[17–19] The critical binding partners of ADs seem to depend on the intrinsic inducibility of the gene being transcribed. Experiments in yeast demonstrated that for constitutively expressed genes (~85% of all genes[20]), essential interactions cluster around TFIID and other general transcription factors, while for inducible gene promoters (~15% of genes), the dominant AD interactions are with SAGA and other histone-modifying and chromatin-remodeling complexes, such as SWI/SNF.[6]

While the map of the interaction networks among these factors is becoming increasingly detailed, the fundamental mechanism formulated in 1990 s based on the initial promoter binding of gene-specific transcription activators followed by the physical recruitment of coactivators and necessary transcriptional enzymes and cofactors remains a key and overwhelmingly dominant concept for explaining transcription regulation in eukaryotes.[6] However, this model based on the biochemically fundamental specificity of sequence → structure → function relationships has increasingly become a subject for modifications, reflecting the growing body of experimental data demonstrating the intrinsic disorder of structures and low affinity of interactions involved. The observed inconsistencies are often framed as a long-standing enigma of ADs, frequently mentioned in recent information-rich publications[1,3,14,15,20] analyzing an increasingly large repertoire of AD sequences in experiments utilizing modern high-throughput synthetic biology advances.

The enigma of transcriptional activation domains

The relative clarity of the coactivator recruitment model begins to cloud once the attention switches to the composition of gene activator molecules and the function of their parts. These molecules perform two key functions in gene activation: (i) recognition of a specific DNA sequence within a cognate gene promoter or enhancer and (ii) recruitment of coactivators and transcription enzymes and cofactors. Each function is performed correspondingly by the DNA-binding domain (DBD) and AD. DBDs are clearly defined by structural motifs and the corresponding specific DNA-binding sequences[21–23] and follow the conventional **sequence → structure → target specificity paradigm**. ADs, however, do not follow this paradigm. The increasingly numerous controversies related to AD properties and characteristics of interactions with targets are outlined below.

The compositional analysis of lower (yeast)[2,3] and higher (human)[20] eukaryotes shows similarities in amino acid used – predominantly aromatic and

acidic. Thus the conventional logic suggests that to perform a recruitment function, ADs should have a **consensus sequence** that is responsible for the interaction with corresponding coactivators or cofactors. While ADs can be roughly divided into acidic, proline-rich, and glutamine-rich subtypes,[24,25] there is no consensus sequence within each category. The initial attempts to find a specific sequence within at least the most common acidic ADs led to the formulation of a loose 9 amino acid consensus.[26,27] However, subsequent investigations testing peptides, bearing these consensus sequences within the Gcn4 activator context, for the ability to activate reporter gene transcription *in vivo*[28] did not confirm their functionality. In addition, more powerful statistical approaches that analyzed large sets ($>10^6$) of sequences that had been functionally tested *in vivo* found instances with the 9-amino-acid proposed consensus sequence, but these sequences were present predominantly (92% of total instances identified) within the nonfunctional sub-pool.[3] Moreover, when the putative consensus sequence was used as a feature in machine learning (ML) predictive models, the consensus had no predictive power.[1,14,15] Together, these observations indicate that large ($>10^4$) sets of functional ADs analyzed *in vivo,* within the context of different activators and reporter genes, do not have a consensus sequence, which was confirmed recently.[20]

Realizing the absence of a solid consensus, researchers investigating individual eukaryotic activators by targeted mutagenesis formulated a number of potential **short linear motifs** (SLiMs) for ADs[29] that supposedly can recruit coactivators, for example, the WxxLF motif in yeast Gcn4 activator that interacts with the Med15 subunit of the Mediator complex.[28] However, again, testing with bioinformatics tools the presence of these motifs among $>10^6$ sequences showed that no SLiMs were significantly enriched in the pool of functional AD sequences and, in fact, were found predominantly (85–95% of the total) among nonfunctional sequences.[3] The same study also showed that using individual SLiMs or their combinations as machine learning features did not produce a predictive AD functionality algorithm. Why different transcription factor ADs vary so drastically in sequence and have neither a consensus sequence nor even common SLiMs has always been puzzling[15,25,30] and, as reiterated recently,[14,20] is a challenging and to-date unanswered question.

A structural approach to understanding ADs initially led to the formulation of the concept of an "acidic blob",[31] which later was developed into the idea that ADs form an amphipathic helix structure[4] in which the hydrophobic amino acid residues are clustered on one side of the α-helix and the acidic residues on the other side. This structural motif is considered to promote the formation of the recruitment bonds between ADs and coactivators, and was tested on several occasions; however, it has been shown not to be strictly required for AD function.[14,15] In fact, recent investigations involving thousands of variants of natural AD sequences established that the majority of AD sequences can function without α-helix formation.[15] Even more, increasing the amount of proline, a known "helix breaker", in the context of the amphipathic α-helix-forming AD sequences was shown to significantly increase transcription activation potential.[32] The very high *in vivo* reporter gene activation ability demonstrated for a WDWDWDWDWDWDWDWD WDWD sequence working as the AD of the HSF activator is also consistent with the absence of a requirement for a specific structure,[1] as the repulsion of negatively charged residues in this sequence likely prevents the formation of any canonical structural motif.

The absence of a specific sequence and/or structure in ADs suggests that each AD interacts with its targets only weakly and can alternate between a **large repertoire of targets**. Over the years, interactions with different ADs have been demonstrated for basal transcriptional machinery components such as TBP,[33–37] TFIIB,[38,39] TFIIH,[40,41] TFIIA,[42,43] RNA polymerase II,[44] variable TAFs,[45] and the Mediator subunits Med17 (Srb4), Srb10, Med15 (Gal11), Med2, and Med 25,[13,28,46–51] as well as with subunits of chromatin-remodeling and histone-modifying complexes such as Ada2, Taf17, Tra1 (SAGA and NuA4 complexes),[17,52–57] Swi1 and Snf2 (SWI/SNF complex),[57–59] and CBP.[60–62] The number of potential AD-interacting targets was estimated to be >200.[63] While early studies tried to postulate the sequence of co-activator recruitment events at gene promoter, it was proven to vary significantly between genes,[64] and it has been persistently challenging to establish this sequence of events without obvious specificity of AD–target interactions.

To investigate how the recruitment mechanism works, at least for a single target, recent investigations have focused primarily on the interactions of ADs with Mediator subunits and specifically with Med15.[14,15] The need to explain the recruitment of Mediator and other coactivators in the context of the uncertainty of AD sequence and structure led to the proposal of a "**fuzzy**" **interactions model**.[13,14,65] According to this model, intrinsically disordered ADs constantly alternate between several low-affinity AD-binding domains (ABDs) within the target, e.g., Med15, enabling binding in many different orientations to different ABDs, thus supposedly maintaining the relatively high overall affinity necessary for the recruitment and retention of the AD target. The possibility of "fuzzy" AD–Med15 interactions was proposed by results of NMR and plasmon resonance *in vitro* studies.[14] In addition, the same study showed a positive correlation between the AD–Med15 interaction constant and the functionality score of the

mutant ADs. However, *MED15* is a nonessential gene in yeast, and cells lacking this gene have only a mild phenotype,[66] making the relevance of focusing on Med15 unclear, and the extent to which this mechanism applies to the recruitment of other coactivators remains uncertain.

Another study proposed a mechanism in which specificity derives from the ability of the target (in this case, the Med25 subunit of Mediator) to rapidly adopt a certain structure upon binding to the intrinsically disordered AD (in this case, of the activator ETV/PEA3).[67] In this model, specific interactions with millimolar Kd values are established without the need for a specific AD sequence. A similar "fuzzy free-for-all", no-sequence-requirement mechanism was proposed for Med15.[14] While an intriguing proposal, this model suggests that the AD target, not AD itself, has a unique structural feature that can rapidly adopt a specific structure triggered by intrinsically disordered protein regions (IDRs). However, by various estimates, IDRs constitute 30–40% of the proteome,[68] and we must then assume that the majority of AD targets possess this IDR snapping feature and the ability to sort different IDRs and select ADs, which remains to be demonstrated.

A modern attempt to reconcile AD activity with its low-affinity and low-specificity interaction profile is based on the increasingly popular **LLPS** concept. According to LLPS, super-enhancers and gene promoters are assembled within cells into condensates with high concentrations of transcription-related enzymes and cofactors. In this case, the local concentration of AD targets is proposed to be high enough to enable even low-affinity, low-specificity interactions to occur.[9] The paradox is that ADs are supposed to be the trigger of the condensate formation, not *vice versa* when the condensate brings activators with their ADs together concentrating them. Using this logic, it is difficult to imagine how the AD, with its extremely low specificity of sequence, structure, and interactions, could trigger the recruitment of coactivators and Mediator complex and Pol II that is required for the condensate formation.

Near-stochastic interactions with extremely low Kd are fundamentally functional, contrary to the conventional wisdom

The enigma of AD function has persisted for decades and is becoming increasingly prominent. Such longstanding mysteries are highly unusual in the field of molecular biology, and the lack of resolution for the mechanism of AD function is especially surprising considering the importance of understanding the gene regulation mechanism for biology in general, and medicine in particular. The reason for this challenging situation seems to lie in the assumption that ADs recruit coactivators directly through sequence- or structure-specific interactions, which is increasingly inconsistent with the accumulating data on ADs. Since the inception of biochemistry as a field, its mindset and methodology have been based on the postulation that molecular structures and interactions are specific, while near-stochastic interactions are traditionally considered as detrimental and non-functional background. Perhaps it is time to reassess.

The low specificity and affinity of AD–target interactions become apparent from a closer look at the results of several independent *in vivo* screens for functional ADs in random libraries of short (20–30 amino acid) sequences.[1,2,69–71] Surprisingly, the results for different activator contexts consistently indicate that for each activator, 1–3% of all random sequences constitute *bona fide* ADs. Considering the 20 natural amino acids, the number of possible combinations for 20 positions is 20^{20}, which is $\sim 10^{26}$ and 1% of this number is $\sim 10^{24}$. This experimentally determined probability of functionality (1/100) is higher than the theoretical probability of finding a specific dipeptide ($P=1/20^2$ = 1/400) and certainly higher than the probability of finding a specific tripeptide ($P=1/20^3$ = 1/8000) at a certain position. Even considering an AD length of 20 amino acids, the experimental probability of finding a functional sequence is between the probability of finding a dipeptide and tripeptide anywhere in the sequence (Figure 1). The theoretical calculations and experimental data together suggest that functionality of ADs can be determined by just two to three amino acids and, thus the Kd of the AD–target interaction is extremely low, probably within the centimolar range. This extremely low level of affinity and specificity is considered meaningless by the standards of conventional biochemistry, and thus artifactual, and such interpretation of experimental results from screens of large pools of AD sequences are often ignored. We suggest that the conflict of experimental data with the conventional model of specific sequence/structure/interaction standards is likely the roadblock to our understanding of ADs.

The **struggle to adhere to the specificity principle** despite experimental contradictions becomes obvious with each attempt to define a consensus sequence,[26,72] or a SLiM,[28] or a specific structural motif for AD. Even the concept of "fuzzy" interactions seems to be an adaptation of the specificity principle to explain recruitment of the Mediator complex, but it is unable to explain how Mediator is chosen among a large repertoire of potential targets (see above and Ref. 63) or why "fuzzy" interactions between AD and Mediator components are not overwhelmed by equally possible but irrelevant or even functionally harmful interactions in the karyoplasm.

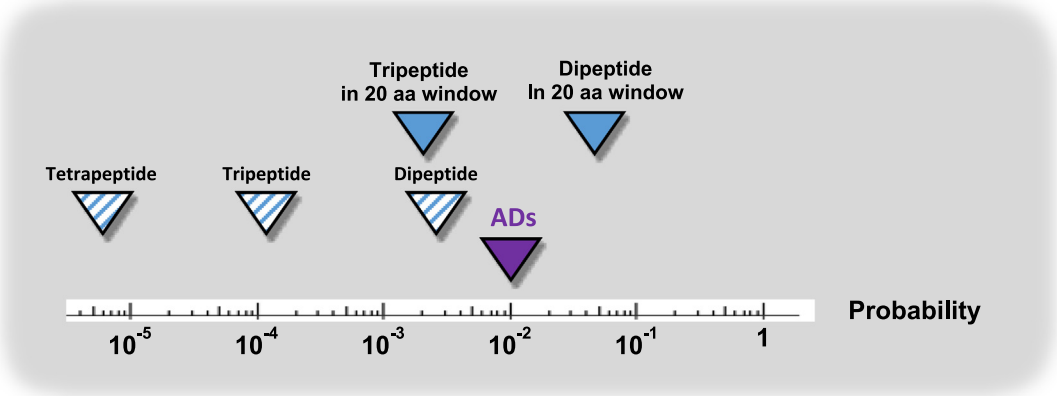


Figure 1. The probability of finding a functional AD sequence within a random sequence pool is similar to the probability of finding a di- or tri-peptide within a 20-amino-acid stretch. Striped triangles are probabilities of a specific di-, tri-, or tetra-amino acid sequence at a specific position, estimated by the formula $P = 1/20^N$, where 20 signifies the number of natural amino acids and N is the number of amino acids in the peptide (di-, tri-, tetra-). To estimate the probability of finding a specific di- or tri-peptide within a 20-amino-acid stretch (solid blue triangles), the previous probability P is multiplied by the number of possible starting positions within the stretch (19 for dipeptide, 18 for tripeptide, etc.), which is a reasonable estimation for low P. The purple triangle is the experimentally determined probability (~1–3%) of finding a functional AD sequence in a completely random pool.

The accumulation of longstanding and more recent contradictions related to the function of ADs, indicating that these critical protein parts do not follow the specificity paradigm, signals the need to reassess this paradigm. **Unless we fundamentally change how we consider the near-stochastic interactions of ADs, we will continue to be stymied by the AD paradox, wondering why ADs are so astronomically variable in sequence,[14] and struggling to understand how gene transcription is initiated**.

Specificity and stochasticity: Two biochemical worlds in one activator molecule

Separating from the biochemical specificity principle, especially for functional interactions of molecules, is certainly a challenge. However, considering AD functionality as nonspecific, intrinsically disordered hydrophobic-acidic amphiphilic surfactants does exactly this.[3,73] The working hypothesis for amphiphilic surfactant action of ADs is the intercalation of the aromatic extremity of the AD between the aromatic bases of DNA, while acidic amino acids of the AD interfere with histone-DNA salt bridges or directly interact with amino groups of lysine or arginine residues of histones. Such interactions could affect the surface of the local promoter nucleosome(s), triggering the chromatin remodeling and/or histone modifications that are associated with and necessary for gene activation.[73] In a similar fashion the nucleosome displacement was demonstrated recently by the action of DNA groove binders, which have an aromatic-hydrophilic chemical nature similar to Ads.[74] The surfactant model requires very low affinity (near-centimolar Kd values) for the AD–target interactions because ADs with high affinities to nucleosomes would likely function as promoter nucleosome-stabilizing repressors instead of activators. The surfactant model relies on the fact that ADs do not function alone but are always a part of an activator molecule containing a DBD. The DBD, by targeting the protein to a specific region (s) of the genome, increases the local AD concentration at the cognate gene promoter by many orders of magnitude (in inverse proportion to the Kd of the DBD), enabling “fuzzy” interactions with the nucleosomes, but only locally (Figure 2).

The problem of low-Kd AD–target interactions, arising from the extremely low specificity of ADs, is recognized, and as mentioned above, an attempt has been made to solve it within the framework of the LLPS model[9] by suggesting an increased coactivator concentration within the newly formed coacervate. However, in addition to the lack of a plausible explanation for how ADs with extremely low substrate specificity can trigger condensate formation by recruiting other factors, the condensates within the nucleus increase the AD–target interactions within only approximately an order or two of magnitude of the local concentration.[75] By contrast, in the surfactant model, targeting of the AD to the promoter by the DBD increases its local concentration by many orders of magnitude (nine orders of magnitude in the case of a DBD with a nanomolar Kd). The initial triggering of local chromatin remodeling and the creation of

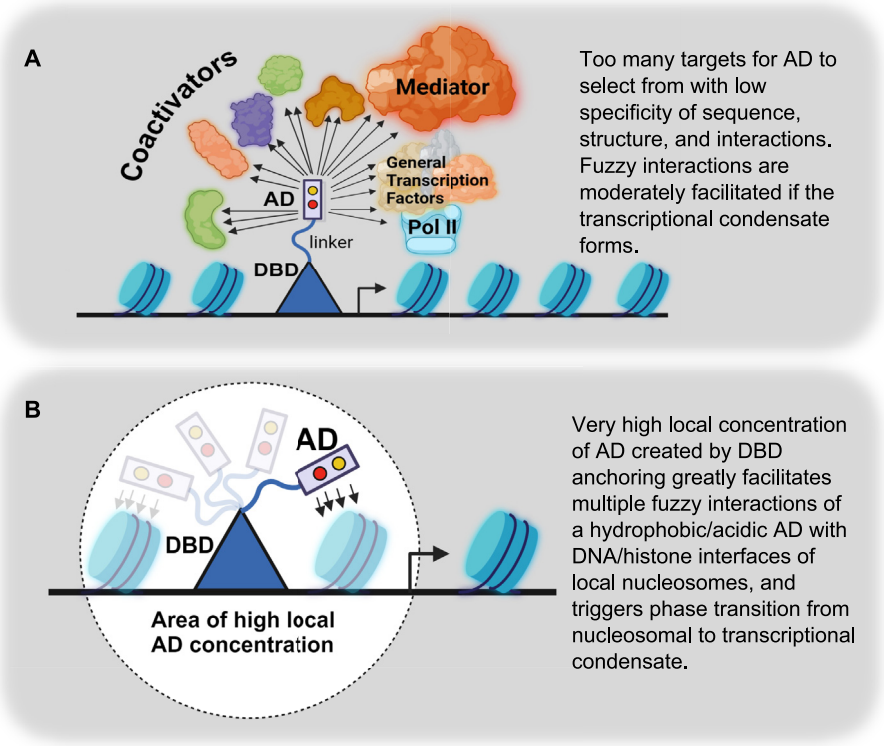


Figure 2. The comparison of the recruitment and surfactant models for AD function. A – The direct physical recruitment of co-activators model. Arrows indicate numerous possible interactions between AD and its possible targets. The red and yellow dots in the AD represent acidic and aromatic residues, respectively. B – The surfactant model. Due to anchoring by DBD, the AD has very high local concentration, thus is able to distort the structure of local nucleosomes and trigger chromatin remodeling leading to a phase conversion.

naked DNA stretches by ADs can promote the condensation of active promoters into a transcriptionally active phase. Thus, condensate formation could be a result of AD action and a subject for regulation by ADs.

The uncertainty with what initiates the LLPS condensate formation is solved within the surfactant model because triggering of the chromatin remodeling by AD and potential exposure of free DNA at gene promoters likely promotes binding to these regions of general transcription factors that have affinity to core promoter sequences, such as TATA box and initiator elements. General transcription factors with Pol II are an integral part of the PIC complex and together with the Mediator complex were shown to form condensates.[10] In addition, the local exposure of naked DNA creates a distinct pH environment critical for the phase separation.[76]

Potential duality of AD function

The surfactant idea for AD function is conducive to the development of the LLPS concepts because the action of the AD surfactant can be considered within the framework of processes related to the transition between two liquid phases: (i) destabilization of chromatin phase leading to the loosening of chromatin and local nucleosome displacement, which was shown to occur for some genes within seconds or minutes after the stimulus[77–79] and (ii) formation and stabilization of the transcription initiation condensate.[10] During chromatin loosening the AD might work as a true surfactant triggering the destabilization of the chromatin phase similar to the DNA groove binders.[74] These also aromatic-hydrophilic molecules were shown to translocate the histone octamer, especially in presence of RSC chromatin remodeling complex or Nap1 histone chaperone. Conversely for the transcription condensate formation, the role of AD might be in “prewetting”[80] by exposing locally IDRs important for LLPS[81] and creating a local naked DNA region exposing multiple DNA negatively charged phosphate groups, thus creating a local pH gradient also important for phase separation.[76] This consideration also suggests that the surfactant and the recruitment models for the AD function might not necessarily be mutually exclusive. The AD might work as a surfactant during the chromatin phase dissociation and following this step as a stabilizer/recruiter of the transcriptional machinery components including Mediator subunits[13–15] during the transcription condensate for-

mation. The formation of the transcriptional condensate is consistent with the facilitation of low affinity and low selectivity interactions between ADs and Mediator subunits demonstrated *in vitro*.[13–15] As suggested previously,[9] the transcriptional condensate formation increases the otherwise unlikely interactions between ADs and transcriptional machinery components.

Perspectives

While theoretically resolving the enigma of AD function, the surfactant model is not yet supported by a body of literature as extensive as that dedicated to the conventional but increasingly insufficient model of direct physical recruitment of coactivators by ADs. The biggest obstacles seem to be the limited experimental toolset available for studying near-stochastic interactions and the lack of perception of such interactions as fundamentally functional by conventional biochemistry. A promising trend, however, is that the near-stochastic "fuzzy" surfactant-like interactions that are characteristic of intrinsically disordered ADs are increasingly accessible to study using modern high-throughput experimentation utilizing massive parallel synthesis, NGS sequencing, and AI technologies for analysis.[1,2,15,20] This approach makes it increasingly possible to see the features of the whole "forest" of 10^{24} potential ADs[3,15] rather than just the "trees" and "branches" of individual AD sequences and their derivatives. It also allows the creation and study of sets of sequences designed around the possibility of multiple low-affinity interactions with alternating targets.

Additional possibilities have been opened recently with the development of AlphaFold 2[82] and AlphaFold 3[83] structural modeling, based on machine learning applied to vast structural databases. Currently the application of these AI tools to ADs and other IDRs is severely restricted[84] by the fact that they are trained mostly on static specific structures of proteins and do not account for IDRs. However, future AI training on expanding IDR datasets including growing list of AD datasets[2,3,15,20] and development of prediction tools focused on IDR-target interactions, could eventually allow predictions of subtle structural features and fuzzy interactions important for gene expression activation by ADs.

The opposite side of the extremely high variability of AD sequences is the fact, noticed by many research groups,[1–4,15] that individual mutations often do not drastically affect the activity of an AD but instead lead to a mild modification of its functionality. Analyzing large pools of AD sequences also shows how minute the differences in the activity levels between individual functional ADs can be and how wide the spectrum of activities for AD sequences within the context of an individual activator is.[1–3,15] The situation is drastically different for DBDs, where the majority of point mutations are expected to cause change of DBD structure and thus significantly affect or eliminate the DNA binding affinity. Considering that the main difference between biological species is not the amount or the divergence of individual gene sequences but instead the fine tuning of gene regulation,[85,86] it is plausible that the plasticity and mildness of mutational effects for gene activator ADs can be the foundation and the driving force of speciation. This mutational plasticity likely allows a significant pool of AD mutants to coexist in a biological population and to be utilized for environmental adaptation during natural selection.

Finally, the functionality of near-stochastic interactions involving IDRs is unlikely to be unique to ADs and gene expression regulation. IDRs are also known to be required for mRNA processing, apoptosis, molecular transport within and between cells, glycolysis, and many other biological processes.[68,87] Appreciating the existence of an additional world of plasticity based on near-random interactions, which is distinct from the conventional biochemical perspective of strong structure–function determinants, may be the key not only to solving the AD enigma but also to better understanding many other biological functions.

CRediT authorship contribution statement

Alexandre M. Erkine: Writing – review & editing, Writing – original draft, Conceptualization. **Marcos A. Oliveira:** Writing – original draft. **Caleb A. Class:** Writing – original draft.

DATA AVAILABILITY

No data was used for the research described in the article.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Bradley K. Broyles for discussions and constructive suggestions. This research was supported by grant to A.M.E. from the National Science Foundation (MCB 1925646).

Received 6 June 2024;
Accepted 23 August 2024;
Available online 28 August 2024

Keywords:
transcription regulation;
chromatin;
gene activator;
fuzzy interactions;
activation domain;
IDR

References

1. Ravarani, C.N. et al, (2018). High-throughput discovery of functional disordered regions: investigation of transactivation domains. *Mol. Syst. Biol.* **14**, e8190. https://doi.org/10.15252/msb.20188190.
2. Erijman, A. et al, (2020). A high-throughput screen for transcription activation domains reveals their sequence features and permits prediction by deep learning. *Mol. Cell* **78**, 890–902.e896. https://doi.org/10.1016/j.molcel.2020.04.020.
3. Broyles, B.K. et al, (2021). Activation of gene expression by detergent-like protein domains. *iScience* **24**, https://doi.org/10.1016/j.isci.2021.103017 103017.
4. Ptashne, M., Gann, A., (1997). Transcriptional activation by recruitment. *Nature* **386**, 569–577. https://doi.org/10.1038/386569a0.
5. Green, M.R., (2005). Eukaryotic transcription activation: right on target. *Mol. Cell* **18**, 399–402. https://doi.org/10.1016/j.molcel.2005.04.017.
6. Chen, H., Pugh, B.F., (2021). What do transcription factors interact with? *J. Mol. Biol.* **433**, https://doi.org/10.1016/j.jmb.2021.166883 166883.
7. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., Sharp, P.A., (2017). A phase separation model for transcriptional control. *Cell* **169**, 13–23. https://doi.org/10.1016/j.cell.2017.02.007.
8. Banani, S.F., Lee, H.O., Hyman, A.A., Rosen, M.K., (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nature Rev. Mol. Cell Biol.* **18**, 285–298. https://doi.org/10.1038/nrm.2017.7.
9. Boija, A. et al, (2018). Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855.e1816. https://doi.org/10.1016/j.cell.2018.10.042.
10. Cho, W.K. et al, (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415. https://doi.org/10.1126/science.aar4199.
11. Sabari, B.R. et al, (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361** https://doi.org/10.1126/science.aar3958.
12. Allen, B.L., Taatjes, D.J., (2015). The Mediator complex: a central integrator of transcription. *Nature Rev. Mol. Cell Biol.* **16**, 155–166. https://doi.org/10.1038/nrm3951.
13. Tuttle, L.M. et al, (2018). Gcn4-Mediator specificity is mediated by a large and dynamic fuzzy protein-protein complex. *Cell Rep.* **22**, 3251–3264. https://doi.org/10.1016/j.celrep.2018.02.097.
14. Tuttle, L.M. et al, (2021). Mediator subunit Med15 dictates the conserved "fuzzy" binding mechanism of yeast transcription activators Gal4 and Gcn4. *Nature Commun.* **12**, 2220. https://doi.org/10.1038/s41467-021-22441-4.
15. Sanborn, A.L. et al, (2021). Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. *eLife* **10** https://doi.org/10.7554/eLife.68068.
16. Quevedo, M. et al, (2019). Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. *Nature Commun.* **10**, 2669. https://doi.org/10.1038/s41467-019-10502-8.
17. Bhaumik, S.R., Raha, T., Aiello, D.P., Green, M.R., (2004). In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer. *Genes Dev.* **18**, 333–343. https://doi.org/10.1101/gad.1148404.
18. Reeves, W.M., Hahn, S., (2005). Targets of the Gal4 transcription activator in functional transcription complexes. *Mol. Cell. Biol.* **25**, 9092–9102. https://doi.org/10.1128/MCB.25.20.9092-9102.2005.
19. Majmudar, C.Y., Labut, A.E., Mapp, A.K., (2009). Tra1 as a screening target for transcriptional activation domain discovery. *Bioorg. Med. Chem. Lett.* **19**, 3733–3735. https://doi.org/10.1016/j.bmcl.2009.05.045.
20. DelRosso, N. et al, (2023). Large-scale mapping and mutagenesis of human transcriptional effector domains. *Nature* **616**, 365–372. https://doi.org/10.1038/s41586-023-05906-y.
21. Garvie, C.W., Wolberger, C., (2001). Recognition of specific DNA sequences. *Mol. Cell* **8**, 937–946.
22. Luscombe, N.M., Austin, S.E., Berman, H.M., Thornton, J. M., (2000). An overview of the structures of protein-DNA complexes. *Genome Biol.* **1**, https://doi.org/10.1186/gb-2000-1-1-reviews001 REVIEWS001.
23. Hossain, M.A., Barrow, J.J., Shen, Y., Haq, M.I., Bungert, J., (2015). Artificial zinc finger DNA binding domains: versatile tools for genome engineering and modulation of gene expression. *J. Cell. Biochem.*. https://doi.org/10.1002/jcb.25226.
24. Mitchell, P.J., Tjian, R., (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371–378.
25. Mapp, A.K., Ansari, A.Z., (2007). A TAD further: exogenous control of gene activation. *ACS Chem. Biol.* **2**, 62–75. https://doi.org/10.1021/cb600463w.
26. Piskacek, S. et al, (2007). Nine-amino-acid transactivation domain: establishment and prediction utilities. *Genomics* **89**, 756–768. https://doi.org/10.1016/j.ygeno.2007.02.003.
27. Piskacek, M., Havelka, M., Rezacova, M., Knight, A., (2016). The 9aaTAD transactivation domains: from Gal4 to p53. *PLoS One* **11**, https://doi.org/10.1371/journal.pone.0162842 e0162842.
28. Warfield, L., Tuttle, L.M., Pacheco, D., Klevit, R.E., Hahn, S., (2014). A sequence-specific transcription activator motif and powerful synthetic variants that bind Mediator using a fuzzy protein interface. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E3506–E3513. https://doi.org/10.1073/pnas.1412088111.
29. Staby, L. et al, (2017). Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochem. J.* **474**, 2509–2532. https://doi.org/10.1042/BCJ20160631.
30. Hahn, S., Young, E.T., (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189**, 705–736. https://doi.org/10.1534/genetics.111.127019.
31. Sadowski, I., Ma, J., Triezenberg, S., Ptashne, M., (1988). GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**, 563–564. https://doi.org/10.1038/335563a0.

32. Broyles, B.K. et al, (2023). Sequence features of transcriptional activation domains are consistent with the surfactant mechanism of gene activation. *bioRxiv*. https://doi.org/10.1101/2023.06.18.545482.
33. Horikoshi, M., Hai, T., Lin, Y.S., Green, M.R., Roeder, R. G., (1988). Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. *Cell* **54**, 1033–1042.
34. Stringer, K.F., Ingles, C.J., Greenblatt, J., (1990). Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* **345**, 783–786. https://doi.org/10.1038/345783a0.
35. Hermann, S., Berndt, K.D., Wright, A.P., (2001). How transcriptional activators bind target proteins. *J. Biol. Chem.* **276**, 40127–40132. https://doi.org/10.1074/jbc.M103793200.
36. Capella, M., Re, D.A., Arce, A.L., Chan, R.L., (2014). Plant homeodomain-leucine zipper I transcription factors exhibit different functional AHA motifs that selectively interact with TBP or/and TFIIB. *Plant. Cell Rep.* **33**, 955–967. https://doi.org/10.1007/s00299-014-1576-9.
37. Khan, S.H., Ling, J., Kumar, R., (2011). TBP binding-induced folding of the glucocorticoid receptor AF1 domain facilitates its interaction with steroid receptor coactivator-1. *PLoS One* **6**, e21939.
38. Lin, Y.S., Ha, I., Maldonado, E., Reinberg, D., Green, M.R., (1991). Binding of general transcription factor TFIIB to an acidic activating region. *Nature* **353**, 569–571. https://doi.org/10.1038/353569a0.
39. Choy, B., Green, M.R., (1993). Eukaryotic activators function during multiple steps of preinitiation complex assembly. *Nature* **366**, 531–536. https://doi.org/10.1038/366531a0.
40. Xiao, H. et al, (1994). Binding of basal transcription factor TFIIH to the acidic activation domains of VP16 and p53. *Mol. Cell. Biol.* **14**, 7013–7024.
41. Chabot, P.R. et al, (2014). Structural and functional characterization of a complex between the acidic transactivation domain of EBNA2 and the Tfb1/p62 subunit of TFIIH. *PLoS Pathog.* **10**, https://doi.org/10.1371/journal.ppat.1004042 e1004042.
42. Stargell, L.A., Struhl, K., (1995). The TBP-TFIIA interaction in the response to acidic activators in vivo. *Science* **269**, 75–78.
43. Ozer, J., Bolden, A.H., Lieberman, P.M., (1996). Transcription factor IIA mutations show activator-specific defects and reveal a IIA function distinct from stimulation of TBP-DNA binding. *J. Biol. Chem.* **271**, 11182–11190.
44. Tan, Q., Linask, K.L., Ebright, R.H., Woychik, N.A., (2000). Activation mutants in yeast RNA polymerase II subunit RPB3 provide evidence for a structurally conserved surface required for activation in eukaryotes and bacteria. *Genes Dev.* **14**, 339–348.
45. Goodrich, J.A., Tjian, R., (1994). TBP-TAF complexes: selectivity factors for eukaryotic transcription. *Curr. Opin. Cell Biol.* **6**, 403–409.
46. Koh, S.S., Ansari, A.Z., Ptashne, M., Young, R.A., (1998). An activator target in the RNA polymerase II holoenzyme. *Mol. Cell* **1**, 895–904.
47. Myers, L.C., Gustafsson, C.M., Hayashibara, K.C., Brown, P.O., Kornberg, R.D., (1999). Mediator protein mutations that selectively abolish activated transcription. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 67–72.
48. Jeong, C.J. et al, (2001). Evidence that Gal11 protein is a target of the Gal4 activation domain in the mediator. *Biochemistry* **40**, 9421–9427.
49. Ansari, A.Z. et al, (2002). Transcriptional activating regions target a cyclin-dependent kinase. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14706–14709. https://doi.org/10.1073/pnas.232573899.
50. Qiu, H. et al, (2004). An array of coactivators is required for optimal recruitment of TATA binding protein and RNA polymerase II by promoter-bound Gcn4p. *Mol. Cell. Biol.* **24**, 4104–4117.
51. Aguilar, X. et al, (2014). Interaction studies of the human and Arabidopsis thaliana Med25-ACID proteins with the herpes simplex virus VP16- and plant-specific Dreb2a transcription factors. *PLoS One* **9**, https://doi.org/10.1371/journal.pone.0098575 e98575.
52. Knutson, B.A., Hahn, S., (2011). Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and NuA4 complexes. *Mol. Cell. Biol.* **31**, 818–831. https://doi.org/10.1128/MCB.00687-10.
53. Brown, C.E. et al, (2001). Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. *Science* **292**, 2333–2337. https://doi.org/10.1126/science.1060214.
54. Barlev, N.A. et al, (1995). Characterization of physical interactions of the putative transcriptional adaptor, ADA2, with acidic activation domains and TATA-binding protein. *J. Biol. Chem.* **270**, 19337–19344.
55. Thut, C.J., Chen, J.L., Klemm, R., Tjian, R., (1995). p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. *Science* **267**, 100–104.
56. Henriksson, A. et al, (1997). Role of the Ada adaptor complex in gene activation by the glucocorticoid receptor. *Mol. Cell. Biol.* **17**, 3065–3073.
57. Neely, K.E. et al, (1999). Activation domain-mediated targeting of the SWI/SNF complex to promoters stimulates transcription from nucleosome arrays. *Mol. Cell* **4**, 649–655.
58. Neely, K.E., Hassan, A.H., Brown, C.E., Howe, L., Workman, J.L., (2002). Transcription activator interactions with multiple SWI/SNF subunits. *Mol. Cell. Biol.* **22**, 1615–1625.
59. Prochasson, P., Neely, K.E., Hassan, A.H., Li, B., Workman, J.L., (2003). Targeting activity is required for SWI/SNF function in vivo and is accomplished through two partially redundant activator-interaction domains. *Mol. Cell* **12**, 983–990.
60. Kundu, T.K. et al, (2000). Activator-dependent transcription from chromatin in vitro involving targeted histone acetylation by p300. *Mol. Cell* **6**, 551–561.
61. Lau, O.D. et al, (2000). HATs off: selective synthetic inhibitors of the histone acetyltransferases p300 and PCAF. *Mol. Cell* **5**, 589–595.
62. Mukherjee, S.P. et al, (2013). Analysis of the RelA:CBP/p300 interaction reveals its involvement in NF-kappaB-driven transcription. *PLoS Biol.* **11**, https://doi.org/10.1371/journal.pbio.1001647 e1001647.
63. Keung, A.J., Bashor, C.J., Kiriakov, S., Collins, J.J., Khalil, A.S., (2014). Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell* **158**, 110–120. https://doi.org/10.1016/j.cell.2014.04.047.

64. Cosma, M.P., (2002). Ordered recruitment: gene-specific mechanism of transcription activation. *Mol. Cell* **10**, 227–236. https://doi.org/10.1016/s1097-2765(02)00604-4.
65. Fuxreiter, M., Tompa, P., (2012). Fuzzy complexes: a more stochastic view of protein function. *Adv. Exp. Med. Biol.* **725**, 1–14. https://doi.org/10.1007/978-1-4614-0659-4_1.
66. Larsson, M. et al, (2013). Functional studies of the yeast med5, med15 and med16 mediator tail subunits. *PLoS One* **8**, https://doi.org/10.1371/journal.pone.0073137 e73137.
67. Henley, M.J. et al, (2020). Unexpected specificity within dynamic transcriptional protein-protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 27346–27353. https://doi.org/10.1073/pnas.2013244117.
68. Oldfield, C.J., Dunker, A.K., (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584. https://doi.org/10.1146/annurev-biochem-072711-164947.
69. Ma, J., Ptashne, M., (1987). A new class of yeast transcriptional activators. *Cell* **51**, 113–119.
70. Abedi, M. et al, (2001). Transcriptional transactivation by selected short random peptides attached to lexA-GFP fusion proteins. *BMC Mol. Biol.* **2**, 10.
71. Erkine, A.M., Gross, D.S., (2003). Dynamic chromatin alterations triggered by natural and synthetic activation domains. *J. Biol. Chem.* **278**, 7755–7764.
72. Piskacek, M., Havelka, M., Jendruchova, K., Knight, A., Keegan, L.P., (2020). The evolution of the 9aaTAD domain in Sp2 proteins: inactivation with valines and intron reservoirs. *Cell. Mol. Life Sci.* **77**, 1793–1810. https://doi.org/10.1007/s00018-019-03251-w.
73. Erkine, A.M., (2018). 'Nonlinear' biochemistry of nucleosome detergents. *Trends Biochem. Sci.* **43**, 951–959. https://doi.org/10.1016/j.tibs.2018.09.006.
74. Lorch, Y., Kornberg, R.D., Maier-Davis, B., (2023). Disruption of nucleosomes by DNA groove binders of clinical significance and implications for chromatin remodeling. *Proc. Natl. Acad. Sci. U. S. A.* **120**, https://doi.org/10.1073/pnas.2216611120 e2216611120.
75. Morin, J.A. et al, (2022). Sequence-dependent surface condensation of a pioneer transcription factor on DNA. *Nature Phys.*. https://doi.org/10.1038/s41567-021-01462-2.
76. King, M.R. et al, (2024). Macromolecular condensation organizes nucleolar sub-phases to set up a pH gradient. *Cell*. https://doi.org/10.1016/j.cell.2024.02.029.
77. Erkina, T.Y., Erkine, A.M., (2006). Displacement of histones at promoters of Saccharomyces cerevisiae heat shock genes is differentially associated with histone H3 acetylation. *Mol. Cell. Biol.* **26**, 7587–7600.
78. Erkina, T.Y., Tschetter, P.A., Erkine, A.M., (2008). Different requirements of the SWI/SNF complex for robust nucleosome displacement at promoters of heat shock factor and Msn2- and Msn4-regulated heat shock genes. *Mol. Cell. Biol.* **28**, 1207–1217. https://doi.org/10.1128/MCB.01069-07.
79. Sekinger, E.A., Gross, D.S., (2001). Silenced chromatin is permissive to activator binding and PIC recruitment. *Cell* **105**, 403–414.
80. Morin, J.A. et al, (2022). Sequence-dependent surface condensation of a pioneer transcription factor on DNA. *Nature Physics* **18**, 271–276. https://doi.org/10.1038/s41567-021-01462-2.
81. Borcherds, W., Bremer, A., Borgia, M.B., Mittag, T., (2021). How do intrinsically disordered protein regions encode a driving force for liquid-liquid phase separation? *Curr. Opin. Struct. Biol.* **67**, 41–50. https://doi.org/10.1016/j.sbi.2020.09.004.
82. Jumper, J. et al, (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
83. Abramson, J. et al, (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. https://doi.org/10.1038/s41586-024-07487-w.
84. Ruff, K.M., Pappu, R.V., (2021). AlphaFold and Implications for Intrinsically Disordered Proteins. *J. Mol. Biol.* **433**, https://doi.org/10.1016/j.jmb.2021.167208 167208.
85. Haerty, W., Singh, R.S., (2006). Gene regulation divergence is a major contributor to the evolution of Dobzhansky-Muller incompatibilities between species of Drosophila. *Mol. Biol. Evol.* **23**, 1707–1714. https://doi.org/10.1093/molbev/msl033.
86. Jovanovic, V.M. et al, (2021). Positive selection in gene regulatory factors suggests adaptive pleiotropic changes during human evolution. *Front. Genet.* **12**, https://doi.org/10.3389/fgene.2021.662239 662239.
87. Bondos, S.E., Dunker, A.K., Uversky, V.N., (2021). On the roles of intrinsically disordered proteins and regions in cell communication and signaling. *Cell Commun. Signal.: CCS* **19**, 88. https://doi.org/10.1186/s12964-021-00774-3.