Title
Commonly asked questions about transcriptional activation domains.

Authors
Udupa, Aditya
Kotha, Sanjana
Staller, Max

Publication Date
2024-02-01

DOI
10.1016/j.sbi.2023.102732

HHS Public Access

Author manuscript

Curr Opin Struct Biol. Author manuscript; available in PMC 2024 June 22.

Published in final edited form as:

Curr Opin Struct Biol. 2024 February ; 84: 102732. doi:10.1016/j.sbi.2023.102732.

Commonly asked questions about transcriptional activation domains

Aditya Udupa[1], **Sanjana R. Kotha**[1,2], **Max V. Staller**[1,2,3,*]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, 94720

[2]Center for Computational Biology, University of California, Berkeley, 94720

[3]Chan Zuckerberg Biohub–San Francisco, San Francisco, CA 94158

Abstract

Eukaryotic transcription factors activate gene expression with their DNA binding domains and activation domains. DNA binding domains bind the genome by recognizing structurally related DNA sequences; they are structured, conserved, and predictable from protein sequence. Activation domains recruit chromatin modifiers, coactivator complexes, or basal transcriptional machinery via structurally diverse protein-protein interactions. Activation domains and DNA binding domains have been called independent, modular units, but there are many departures from modularity, including interactions between these regions and overlap in function. Compared to DNA binding domains, activation domains are poorly understood because they are poorly conserved, intrinsically disordered, and difficult to predict from protein sequence. This review, organized around commonly asked questions, describes recent progress that the field has made in understanding the sequence features that control activation domains and predicting them from sequence.

Keywords

Transcription; Transcription factor; Coactivator; RNA Polymerase II; Activation domain; Transactivation domain; Transcriptional activation domain; Protein-protein interactions (PPIs); intrinsically disordered protein; convolutional neural network; protein function prediction

What are transcriptional activation domains?

Transcriptional activation domains are the regions of transcription factors (TFs) that bind to coactivator complexes to activate transcription [1-3]. These regions are also called transactivation domains or activator domains, and all three terms have been applied both to minimized regions of high activity (10-80 AA) and the entirety of the TF outside the DNA binding domain (DBD; 100s of AA). We will use the term activation domain to refer to short regions that directly bind to coactivators. Activation domains are defined experimentally, most often in sufficiency assays, where candidate protein regions are fused to a heterologous DBD and activity is measured with a reporter gene. There are high-throughput sufficiency

*Correspondence: 16 Barker Hall, Berkeley CA, 94707, mstaller@berkeley.edu.

assays in yeast, fly, and human cells that use pooled oligo synthesis to study short TF fragments, ≤80AA [4-9], but some groups have queried longer regions [10,11]. The boundaries of nearly all annotated activation domains should be regarded as approximations because very little experimental effort has been devoted to defining boundaries precisely.

The primary known function of activation domains is to recruit coactivator complexes. These interactions are highly dynamic, with short dwell times that are hard to catch with pull-down assays *in vivo* [12-14]. There are now dozens of beautiful NMR structures of these interactions, as reviewed by Dyson and Wright [1]. It remains a goal in the field to map interactions between activation domains and coactivators. Genetically defining the coactivator dependence of activation domains yields complex results that have been hard to interpret [15]. Recent TURBO-ID experiments, which capture dynamic interactions *in vivo*, suggest that most activation domains preferentially bind to only 1-2 coactivators [16]. Importantly, different members of a single DBD family (e.g. FOXO) recruited different combinations of the TFIID, CBP/p300, NuA4, and BAF complexes [16]. Organizing TFs into families by DBD homology has been useful, but going forward we will require an orthogonal organization system for grouping TFs with functionally similar activation domains.

Are all activation domains acidic?

The first few dozen activation domains were all negatively charged [17,18], inspiring a seminal paper by Paul Sigler entitled, “Transcriptional activation. Acid blobs and negative noodles” [19]. Sigler leveraged his authority as a respected structural biologist to argue that activation domains did not need to fold in order to be functional. This seemingly simple idea was heretical at the time. It is now clear that transcription factors are highly enriched for intrinsically disordered protein regions (IDRs), which do not fold into a single 3D structure, comprise roughly a third of the residues in eukaryotic proteomes and are enriched for protein-protein interactions and post-translational modifications [20-22]. Sigler’s paper remains highly recommended reading.

Traditionally, activation domains are classified by their most common residue as acidic, glutamine-rich, proline-rich, or serine-rich. For this review, we collected lists of activation domains from recent surveys [9,16,23-25] (Table S1). After confirming the UniProt ID of each domain, we obtained the full-length sequences of all isoforms and used the published coordinates of the domain to find the sequence of each region. If the region matched the domain, we saved the UniProt ID of the isoform in the column of Table S1 titled “Matching Isoforms.” In cases where there were multiple matches, we selected one to record in the Uniprot ID column of Table S1, preferring the canonical isoform when it was among the matches. For a few members of the DelRosso library, we used the Ensembl ID to confirm which isoform to designate as the UniProt ID. Finally, we merged domains with the same UniProt ID with overlapping start and end coordinates, yielding the union of overlapping annotations. We used the updated start and end coordinates to find the region of the full-length sequence to use as the new domain sequence.

The combined list of activation domains revealed that glutamine-rich activation domains are rare (19/760, 2.50%) and that the three other classes overlap highly (Figure 1) [2,26,27]. Acidic activation domains were the first group described [17,18], remain the largest, and contain the strongest members [28]. In these domains, aromatic and leucine residues make the largest contributions to activation domain function [29]. Individual acidic residues are dispensable and poorly conserved but collectively essential for function [7]. Many acidic activation domains are well-described by our **acidic exposure model**, wherein the critical aromatic and leucine residues make contact with shallow hydrophobic grooves on coactivators (Figure 2) [4,30]. However, in the unbound state, the aromatic and leucine residues interact with each other and drive collapse into an inactive state. The acidic residues repel each other and favor solvation, keeping the hydrophobic residues exposed to solvent, where they are available to bind coactivators. Intermixing the positions and balancing the numbers of acidic and hydrophobic residues is important for full activity [4,5,7,26]. The exchange between collapsed and expanded states might be rapid, but the time scale has not yet been measured. The acidic exposure model is an instance of the stickers-and-spacers model with a very active role for the spaces [31]. This model is supported by work from many groups [5-8,27,32,33]. Surface plasmon resonance assays showed that acidic residues can also mediate fast, low-affinity electrostatic binding to coactivators and that hydrophobic residues mediate slow, high-affinity binding [7,34,35]. The overlap between acidic, serine-rich, proline-rich and glutamine (Q) rich activation domains have led Bintu et al. to playfully describe them all together as greasy acidic noodles sprinkled with salt (S), pepper (P), and queso (Q). Evidence that phosphorylation can modulate activation domain activity has led to speculation that some S-rich or P-rich activation domains are inducible acidic activation domains [36,37].

Within acidic activation domains, there is functional diversity. There are hints that L-rich activation domains bind to CBP/p300 and aromatic rich activation domains bind to Med25 [30,33,38]. Binding specificity arises from the structure of the coactivator binding interface imposing geometric constraints on the activation domain. For example, the deep hydrophobic canyon of Taz1 imposes more constraint than the shallow hydrophobic canyon of Med15 [9,39] A live-imaging study of transcriptional bursting found that 45/78 activation domains primarily regulate either transcriptional burst size or burst duration, but less often both (9/78) [40]. Activation domains that recruit Mediator or the general transcriptional machinery tended to modulate burst size, while activation domains that recruit SWI/SNF, histone acetyltransferases, or the super elongation complex tended to regulate burst intensity [16,40]

Why are so many active domains acidic?

In principle, the exposure of hydrophobic residues offered by acidic residues in the acidic exposure model could be achieved by basic residues, but acidic residues have several advantages. Most importantly, because DNA is acidic, it repels acidic activation domains, promotes exposure, and prevents non-specific DNA binding [41]. Acidic activation domains can have low-affinity, intramolecular, electrostatic interactions with DBD that can increase DNA specificity [42-45] and electrostatic interactions with basic coactivators [34,35]. When DNA repels activation domains, evolution can tune DNA affinity by acting only on the

DBD. A drawback of positively charged residues is that they can have cation-pi interactions with aromatic residues, which would increase collapse instead of exposure [46]. The electrostatic constraint posed by DNA can explain why many activation domains are acidic.

Acidic activation domain function is deeply conserved across eukaryotes. The Gal4/UAS system from yeast works beautifully in flies, mammals, and plants [47]. Acidic activation domains from animals, viruses, and plants work well in yeast [11,28,48]. This promiscuous species-crossing has fueled speculation that acidic activation domains existed in the ancestor of all eukaryotes.

Why are activation domains disordered?

Virtually all activation domains are intrinsically disordered, but many undergo coupled binding and folding, often into short alpha helices [1]. The first explanation for intrinsic disorder is that it allows activation domains to fold differently with each interaction partner [1,49]. In p53, varying the helical propensity trades off affinity for two partners, drastically modulating protein function [50]. Known activation domains are enriched for low-confidence secondary structural predictions in AlphaFold models [16]. So far, there are few clear examples where the activation domain remains disordered while bound to the coactivator, but we suspect this type of interaction is underreported due to ascertainment bias [51]. The second explanation comes from the acidic exposure model, where disorder reduces the entropic cost of keeping W,F,Y,L residues exposed to solvent because they need to be exposed for only a fraction of the time to allow coactivator binding. The third explanation is that intrinsically disordered sequences can use long–range, low–affinity electrostatic interactions to achieve diffusion-limited binding [52-56].

A controversial idea is the detergent model, which argues that activation domains loosen the interactions between nucleosomes and DNA to help create nucleosome-free regions for TF binding [8,57]. This idea contrasts with the standard model where activation domains recruit chromatin remodeling enzymes, including ISWI, SWI/SNF, CHD, and INO80, which use ATP-hydrolysis to slide or evict nucleosomes [58]. We do not endorse this model.

What is the molecular grammar in activation domains?

Molecular grammar describes how the arrangement or order of amino acids contributes to function. There is a spectrum ranging from an extreme of “no grammar,” where only the composition matters, to an extreme of “strict grammar,” where the exact order of residues is essential for function. The dominant model for activation domains is that they are short linear motifs (SLiMs) of hydrophobic residues surrounded by a permissive context [59]. For example, the ΦxxΦΦ motif, where Φ is a hydrophobic residue, is surrounded by acidic residues on many activation domains, often forming an amphipathic alpha helix that presents a continuous hydrophobic surface to the coactivator [1,60]. There are two problems with motif-centered models. First, individual motifs are conserved within a family of orthologs, but each motif is rarely present in many families, making each one too specific to be a useful predictor of activation domains [4,37,60]. Second, our mutagenesis has revealed that multiple motifs are necessary for full activity [4,30]. The reliance on motifs has served as

a robust set of training wheels for the field, but as our understanding of activation domain function matures, the fixation on motifs is holding us back because motifs imply a strict molecular grammar. We believe that it is time to focus on **clusters** of hydrophobic residues embedded in a permissive context, emphasizing a much more flexible grammar.

There is strong evidence against strict grammar: random peptides with activation domain activity do not have enriched motifs [6], shuffling activation domain sequence can increase activity in a third of examples [4,7], and searching for clusters of W,F,Y,L residues in acidic regions is a good predictor of activation domains [26,30]. Interesting work on Abf1 in yeast completely blurs the line between motifs and context [61]. There is also strong evidence against no grammar: shuffling sequence often has profound effects on activity, especially in helices [4,7,30], and interchanging aromatic residues can disrupt activity [30]. We are left with weak grammar, which we still do not fully understand and is at times disconcerting. Shuffling sequences can disrupt activity, or it can have little effect; breaking helices can disrupt activity or have little effect; interchanging similar residues (D>E or F>W) can disrupt activity or have little effect [7,29,30,39]. We refer the reader to Kotha and Staller 2023 for an extended discussion of the role of motifs and grammar in activation domains [26].

Can we predict activation domains from protein sequence?

Recently, high-throughput assays for measuring activation domain activity have powered convolutional neural network (CNN) models for predicting activation domains from protein sequence. The first computational model for activation domains, the 9aaTAD model, used regular expressions to find matches to a highly degenerate motif and context pattern [62]. However, in two high-throughput screens in yeast, this pattern was not enriched in activation domains and was not a useful predictor [6,7]. Rational mutagenesis of one activation domain and lasso regression models on random peptides revealed key amino acids [4,5,8]. The first CNN activation domain predictor was developed with a dataset of 3.6 million 30-AA random peptides tested in yeast [6]. The second CNN was developed with a dataset of 53-AA peptides from 180 *Saccharomyces cerevisiae* transcription factors (n = 7460 tiles) [7]. In our experience [63] and the work of others [11,26,27,63], both of these models do a good job predicting the general location of activation domains on human and plant TFs. We have found they do an excellent job of prioritizing a few regions of a TF that are likely to be activation domains. Both models struggle to find activation domain boundaries accurately, but these boundaries are poorly defined. To our surprise, scanning for clusters of W,F,L residues in acidic regions performs nearly as well as CNN models for human TFs, implying both that prediction is simpler than anticipated and that the grammar is highly degenerate [26]. Second-generation CNN models with more sophisticated architectures are already more accurate [64]. As more data becomes available, we anticipate the activation domain predictors will improve.

What is the link between activation domains and phase separation?

It is now clear that transcription occurs in dynamic clusters [65]. These clusters are non-stoichiometric assemblies with dozens of copies of each TF and coactivator complexes that

together recruit dozens of Pol II molecules, some of which successfully transcribe mRNAs [66-71]. It remains deeply contested whether these clusters of active transcription are phase-separated biomolecular condensates [72,73]. There is evidence that the same protein-protein interactions that enable activation domain function *in vivo* enable phase separation *in vitro* [72]. Modulation of phase separation *in vitro* can identify interaction binding partners or drugs [74]. There are examples from plants where a TF becomes inactive as it enters the condensate for long-term storage [75,76]. Careful studies in a synthetic system showed that phase separation can be completely separated from activation domain strength [77].

We speculate that the reason there has been so much confusion between transcriptional activation and phase separation is that both processes rely on multivalency. Multivalency is essential in many phase-separated systems and dynamic protein-protein interactions (See companion review by Berlow and colleagues). TFs and activation domains engage in multivalent binding with coactivators. Activation domains show multivalency on two length scales. First, many TFs have multiple activation domains that can bind the same coactivator: there are five patches of Gcn4 that each contact many (but not all) of the four activation domain-binding domains of Med15 [13]. For p53, the active form is a tetramer [78], and it can form four contacts with CBP/p300 [49]. Second, within activation domains, adding aromatic or leucine residues near key hydrophobic motifs boosts activity by lengthening the interaction surface [4,60,79].

Concluding remarks

Over the past 6 years, new methods have clarified the sequence determinants that control acidic activation domain function. We anticipate the next few years will expand these approaches to activation domains from other classes and investigate post-translational modifications. The major open questions are to define functional classes of activation domains, map interactions with coactivators, and build improved predictors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Nick Morffy, Thomas Graham, Darren Kahn, Vinson Fan, and John J. Ferrie for comments on the manuscript.

Funding

S.R.K. is an undergraduate at UC Berkeley supported by the STEM Excellence through Equity & Diversity (SEED) Scholars Honors Program. A.U. is supported by NIH grant 5T32GM007232-38. This work is supported by Simons Foundation grant 1018719, NIH grant R35GM15081, and NSF grant 2112057. M.V.S is a Chan Zuckerberg Biohub – San Francisco Investigator.

References

1. Dyson HJ, Wright PE: Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300. J Biol Chem 2016, 291:6714–6722. [PubMed: 26851278]

2. Soto LF, Li Z, Santoso CS, Berenson A, Ho I, Shen VX, Yuan S, Fuxman Bass JI: Compendium of human transcription factor effector domains. Molecular Cell 2022, 82:514–526. [PubMed: 34863368] After carefully curating a long list of published activation and repression domains, this paper analyzed conservation and genetic variation. DBDs are much more conserved than activation domains. Effector domains contain more neutral mutations and fewer clinical mutations than DBDs. An excellent starting point for people entering the field.

3. Már M, Nitsenko K, Heidarsson PO: Multifunctional Intrinsically Disordered Regions in Transcription Factors. Chemistry 2023,

4. Staller MV, Holehouse AS, Swain-Lenz D, Das RK, Pappu RV, Cohen BA: A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. Cell Syst 2018, 6:444–455.e6. [PubMed: 29525204]

5. Ravarani CN, Erkina TY, De Baets G, Dudman DC, Erkine AM, Babu MM: High-throughput discovery of functional disordered regions: investigation of transactivation domains. Mol Syst Biol 2018, 14:e8190. [PubMed: 29759983]

6. Erijman A, Kozlowski L, Sohrabi-Jahromi S, Fishburn J, Warfield L, Schreiber J, Noble WS, Söding J, Hahn S: A High-Throughput Screen for Transcription Activation Domains Reveals Their Sequence Features and Permits Prediction by Deep Learning. Mol Cell 2020, 78:890–902.e6. [PubMed: 32416068]

7. Sanborn AL, Yeh BT, Feigerle JT, Hao CV, Townshend RJ, Lieberman Aiden E, Dror RO, Kornberg RD: Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. Elife 2021, 10.

8. Broyles BK, Gutierrez AT, Maris TP, Coil DA, Wagner TM, Wang X, Kihara D, Class CA, Erkine AM: Activation of gene expression by detergent-like protein domains. iScience 2021, 24:103017. [PubMed: 34522860]

9. Staller MV, Ramirez E, Kotha SR, Holehouse AS, Pappu RV, Cohen BA: Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains. Cell Systems 2022, 13:334–345.e5. [PubMed: 35120642]

10. Arnold CD, Nemčko F, Woodfin AR, Wienerroither S, Vlasova A, Schleiffer A, Pagani M, Rath M, Stark A: A high-throughput method to identify trans-activation domains within transcription factor sequences. EMBO J 2018, 37:e98896. [PubMed: 30006452]

11. Hummel NFC, Zhou A, Li B, Markel K, Ornelas IJ, Shih PM: The trans-regulatory landscape of gene networks in plants. Cell Syst 2023, 14:501–511.e4. [PubMed: 37348464]

12. Herbig E, Warfield L, Fish L, Fishburn J, Knutson BA, Moorefield B, Pacheco D, Hahn S: Mechanism of Mediator recruitment by tandem Gcn4 activation domains and three Gal11 activator-binding domains. Mol Cell Biol 2010, 30:2376–2390. [PubMed: 20308326]

13. Tuttle LM, Pacheco D, Warfield L, Luo J, Ranish J, Hahn S, Klevit RE: Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex. Cell Rep 2018, 22:3251–3264. [PubMed: 29562181]

14. Nishikawa JL, Boeszoermenyi A, Vale-Silva LA, Torelli R, Posteraro B, Sohn Y-J, Ji F, Gelev V, Sanglard D, Sanguinetti M, et al. : Inhibiting fungal multidrug resistance by disrupting an activator-Mediator interaction. Nature 2016, 530:485–489. [PubMed: 26886795]

15. Swanson MJ, Qiu H, Sumibcay L, Krueger A, Kim S-J, Natarajan K, Yoon S, Hinnebusch AG: A multiplicity of coactivators is required by Gcn4p at individual promoters in vivo. Mol Cell Biol 2003, 23:2800–2820. [PubMed: 12665580]

16. Alerasool N, Leng H, Lin Z-Y, Gingras A-C, Taipale M: Identification and functional characterization of transcriptional activators in human cells. Mol Cell 2022, doi:10.1016/j.molcel.2021.12.008. This systematic screen of full-length cDNAs identified new activator TFs and dozens of activation domains. TubroID proteomics revealed that most activation domains predominantly recruit one coactivator complex. Importantly, paralogs from a DBD family could be activators or repressors. Strong activators were shown to be involved in stress response, reprogramming, and differentiation. Oncogenic fusion proteins were enriched for binding to specific coactivators.

17. Hope IA, Struhl K: Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell 09 1986, 46:885–894. [PubMed: 3530496]

18. Ma J, Ptashne M: A new class of yeast transcriptional activators. Cell 1987, 51:113–119. [PubMed: 3115591]

19. Sigler PB: Transcriptional activation. Acid blobs and negative noodles. Nature 1988, 333:210–212. [PubMed: 3367995]

20. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, et al. : Classification of intrinsically disordered regions and proteins. Chem Rev 2014, 114:6589–6631. [PubMed: 24773235]

21. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK: Intrinsic disorder in transcription factors. Biochemistry 2006, 45:6873–6888. [PubMed: 16734424]

22. Uversky VN, Kulkarni P: Intrinsically disordered proteins: Chronology of a discovery. Biophys Chem 2021, 279:106694. [PubMed: 34607199]

23. Soto LF, Li Z, Santoso CS, Berenson A, Ho I, Shen VX, Yuan S, Bass JIF: Compendium of human transcription factor effector domains. 2021,

24. Choi Y, Asada S, Uesugi M: Divergent hTAFII31-binding Motifs Hidden in Activation Domains ,. J Biol Chem 05 2000, 275:15912–15916. [PubMed: 10821850]

25. DelRosso N, Tycko J, Suzuki P, Andrews C, Aradhana, Mukund A, Liongson I, Ludwig C, Spees K, Fordyce P, et al. : Large-scale mapping and mutagenesis of human transcriptional effector domains. Nature 2023,

26. Kotha SR, Staller MV: Clusters of acidic and hydrophobic residues can predict acidic transcriptional activation domains from protein sequence. Genetics 2023, Described a simple mechanistic model for predicting activation domains from protein sequence. An intersection of this model with published convolutional neural network models led to more accurate predictions. Contains an extended discussion of motifs and sequence features of activation domains.

27. DelRosso N, Tycko J, Suzuki P, Andrews C, Aradhana, Mukund A, Liongson I, Ludwig C, Spees K, Fordyce P, et al. : Large-scale mapping and mutagenesis of human transcriptional effector domains. Nature 2023, doi:10.1038/s41586-023-05906-y. This systematic screen of 80AA fragments from 2000 human TFs and chromatin regulators uncovered hundreds of new activation domains and repression domains. Mutagenesis revealed all activation domains rely on W,F,Y,L residues for activity. Sumoylation sites were enriched in repression domains. Found many new serine and proline rich activation domains. Also found 72 bifunctional domains that both activate and repress.

28. Sadowski I, Ma J, Triezenberg S, Ptashne M: GAL4-VP16 is an unusually potent transcriptional activator. Nature 1988, 335:563–564. [PubMed: 3047590]

29. Cress WD, Triezenberg SJ: Critical structural elements of the VP16 transcriptional activation domain. Science 1991, 251:87–90. [PubMed: 1846049]

30. Staller MV, Ramirez E, Kotha SR, Holehouse AS, Pappu RV, Cohen BA: Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains. Cell Syst 2022, 13:334–345.e5. [PubMed: 35120642] This rational mutagenesis of multiple acidic activation domains yielded the original formulation of the acidic exposure model. Activation domains required both acidic and hydrophobic residues. The balance between acidic and hydrophobic residues predicts strong activation domains.

31. Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, Grace CR, Soranno A, Pappu RV, Mittag T: Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science 2020, 367:694–699. [PubMed: 32029630]

32. Bjarnason S, McIvor JAP, Prestel A, Demény KS, Bullerjahn JT, Kragelund BB, Mercadante D, Heidarsson PO: DNA binding redistributes activation domain ensemble and accessibility in pioneer factor Sox2. bioRxiv 2023, doi:10.1101/2023.06.16.545083. This study leveraged the power of SM-FRET, NMR, and simulations to demonstrate that the structural ensemble of SOX2's activation domains are compacted due to interactions between its charged residues and those of the DBD. When the DBD binds SOX2's cognate motif, this ensemble expands, exposing the activation domains to coactivators. A beautiful example of a seemingly-modular TF containing important intramolecular interactions.

33. Berlow RB, Jane Dyson H, Wright PE: Multivalency enables unidirectional switch-like competition between intrinsically disordered proteins. Proceedings of the National Academy of Sciences 2022, 119.

34. Hermann S, Berndt KD, Wright AP: How transcriptional activators bind target proteins. J Biol Chem 2001, 276:40127–40132. [PubMed: 11514548]
35. Ferreira ME, Hermann S, Prochasson P, Workman JL, Berndt KD, Wright APH: Mechanism of transcription factor recruitment by acidic activators. J Biol Chem 2005, 280:21779–21784. [PubMed: 15826952]
36. Conti MM, Li R, Narváez Ramos MA, Zhu LJ, Fazzio TG, Benanti JA: Phosphosite Scanning reveals a complex phosphorylation code underlying CDK-dependent activation of Hcm1. Nat Commun 2023, 14:310. [PubMed: 36658165]
37. Raj N, Attardi LD: The Transactivation Domains of the p53 Protein. Cold Spring Harb Perspect Med 2017, 7.
38. Henley MJ, Linhares BM, Morgan BS, Cierpicki T, Fierke CA, Mapp AK: Unexpected specificity within dynamic transcriptional protein-protein complexes. Proc Natl Acad Sci U S A 2020, 117:27346–27353. [PubMed: 33077600]
39. Brzovic PS, Heikaus CC, Kisselev L, Vernon R, Herbig E, Pacheco D, Warfield L, Littlefield P, Baker D, Klevit RE, et al. : The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. Mol Cell 12 2011, 44:942–953.
40. Mamrak NE, Alerasool N, Griffith D, Holehouse AS, Taipale M, Lionnet T: The kinetic landscape of human transcription factors. bioRxiv 2022, doi:10.1101/2022.06.01.494187. Imaging time courses of nascent transcription measures how activation domains modulate bursting kinetics. Activation domains regulate burst size, burst frequency, or, rarely, both. This paper presents a promising approach to classifying activation domains functionally. Activation domains that recruit Mediator, super elongation complex, or basal transcriptional machinery tended to increase burst amplitude. Activation domains that recruit histone acetyl transferases or SWI/SNF tended to increase burst duration (active fraction). Viral activation domains could increase both burst size and burst duration.
41. Xiang L, Chen K, Yan R, Li W, Xu K: Single-molecule displacement mapping unveils nanoscale heterogeneities in intracellular diffusivity. Nat Methods 2020, 17:524–530. [PubMed: 32203387]
42. He F, Borcherds W, Song T, Wei X, Das M, Chen L, Daughdrill GW, Chen J: Interaction between p53 N terminus and core domain regulates specific and nonspecific DNA binding. Proc Natl Acad Sci U S A 2019, 116:8859–8868. [PubMed: 30988205]
43. Krois AS, Dyson HJ, Wright PE: Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A 2018, 115:E11302–E11310. [PubMed: 30420502]
44. Gregory E, Daughdrill GW: Sequence Properties of An Intramolecular Interaction That Inhibits p53 DNA Binding. Biomolecules 2022, 12.
45. Schütz S, Bergsdorf C, Goretzki B, Lingel A, Renatus M, Gossert AD, Jahnke W: The Disordered MAX N-terminus Modulates DNA Binding of the Transcription Factor MYC:MAX. J Mol Biol 2022, 434:167833. [PubMed: 36174765]
46. Wang J, Choi J-M, Holehouse AS, Lee HO, Zhang X, Jahnel M, Maharana S, Lemaitre R, Pozniakovsky A, Drechsel D, et al. : A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell 2018, 174:688–699.e16. [PubMed: 29961577]
47. Brand AH, Perrimon N: Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 1993, 118:401–415. [PubMed: 8223268]
48. Burz DS, Hanes SD: Isolation of mutations that disrupt cooperative DNA binding by the Drosophila bicoid protein. J Mol Biol 2001, 305:219–230. [PubMed: 11124901]
49. Krois AS, Ferreon JC, Martinez-Yamout MA, Dyson HJ, Wright PE: Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein. Proc Natl Acad Sci U S A 2016, 113:E1853–62. [PubMed: 26976603]
50. Borcherds W, Theillet F-X, Katzer A, Finzel A, Mishall KM, Powell AT, Wu H, Manieri W, Dieterich C, Selenko P, et al. : Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. Nat Chem Biol 2014, 10:1000–1002. [PubMed: 25362358]

51. Risør MW, Jansma AL, Medici N, Thomas B, Dyson HJ, Wright PE: Characterization of the High-Affinity Fuzzy Complex between the Disordered Domain of the E7 Oncoprotein from High-Risk HPV and the TAZ2 Domain of CBP. Biochemistry 2021, 60:3887–3898. [PubMed: 34905914]
52. Kim J-Y, Chung HS: Disordered proteins follow diverse transition paths as they fold and bind to a partner. Science 2020, 368:1253–1257. [PubMed: 32527832]
53. Shoemaker BA, Portman JJ, Wolynes PG: Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci U S A 2000, 97:8868–8873. [PubMed: 10908673]
54. Kim J-Y, Meng F, Yoo J, Chung HS: Diffusion-limited association of disordered protein by non-native electrostatic interactions. Nat Commun 2018, 9:4707. [PubMed: 30413699]
55. Shammas SL, Travis AJ, Clarke J: Remarkably fast coupled folding and binding of the intrinsically disordered transactivation domain of cMyb to CBP KIX. J Phys Chem B 2013, 117:13346–13356. [PubMed: 23875714]
56. Wicky BIM, Shammas SL, Clarke J: Affinity of IDPs to their targets is modulated by ion-specific changes in kinetics and residual structure. Proc Natl Acad Sci U S A 2017, 114:9882–9887. [PubMed: 28847960]
57. Erkina TY, Erkine AM: Nucleosome distortion as a possible mechanism of transcription activation domain function. Epigenetics Chromatin 2016, 9:40. [PubMed: 27679670]
58. Reyes AA, Marcum RD, He Y: Structure and Function of Chromatin Remodelers. J Mol Biol 2021, 433:166929. [PubMed: 33711345]
59. Bugge K, Brakti I, Fernandes CB, Dreier JE, Lundsgaard JE, Olsen JG, Skriver K, Kragelund BB: Interactions by Disorder - A Matter of Context. Front Mol Biosci 2020, 7:110. [PubMed: 32613009]
60. Warfield L, Tuttle LM, Pacheco D, Klevit RE, Hahn S: 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 2014, 111:E3506–13. [PubMed: 25122681]
61. Langstein-Skora I, Schmid A, Emenecker RJ, Richardson MOG, Götz MJ, Payer SK, Korber P, Holehouse AS: Sequence- and chemical specificity define the functional landscape of intrinsically disordered regions. bioRxiv 2022, doi:10.1101/2022.02.10.480018.
62. Piskacek S, Gregor M, Nemethova M, Grabner M, Kovarik P, Piskacek M: Nine-amino-acid transactivation domain: establishment and prediction utilities. Genomics 2007, 89:756–768. [PubMed: 17467953]
63. Hummel N, Markel K, Stefani J, Staller MV, Shih P: Systematic identification of transcriptional activator domains from non-transcription factor proteins in plants and yeast. bioRxiv 2023, doi:10.1101/2023.09.12.557247.
64. Mahatma S, Van den Broeck L, Morffy N, Staller MV, Strader LC, Sozzani R: Prediction and functional characterization of transcriptional activation domains. In 2023 57th Annual Conference on Information Sciences and Systems (CISS). . 2023:1–6.
65. Lu F, Lionnet T: Transcription Factor Dynamics. Cold Spring Harb Perspect Biol 2021, 13.
66. Cho W-K, Jayanth N, English BP, Inoue T, Andrews JO, Conway W, Grimm JB, Spille J-H, Lavis LD, Lionnet T, et al. : RNA Polymerase II cluster dynamics predict mRNA output in living cells. Elife 2016, 5.
67. Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, Heckert A, Banala S, Lavis L, Darzacq X, et al. : Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science 2018, 361.
68. Cho W-K, Spille J-H, Hecht M, Lee C, Li C, Grube V, Cisse II: Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science 2018, 361:412–415. [PubMed: 29930094]
69. Kawasaki K, Fukaya T: Functional coordination between transcription factor clustering and gene activity. Mol Cell 2023, 83:1605–1622.e9. [PubMed: 37207625] This tour de force in high-resolution live-imaging for *Drosophila* embryos convincingly demonstrates how clustering of activator TFs precede transcriptional bursts. This work provides direct evidence of the standard model in live cells. The authors found that Mediator similarly clustered at these enhancer proximal sites and found feedback between active transcription and cluster size. Adding glutamine repeats

to the TFs increased cluster sizes and led to increased intensity and frequency of transcriptional bursts.

70. Ingersoll S, Brown K, Ma B, Ren X: Quantifying the Binding and Target-Search Kinetics of Transcriptional Regulatory Factors by Live-Cell Single-Molecule Tracking. In DNA-Protein Interactions: Methods and Protocols. Edited by Simoes-Costa M Springer US; 2023:141–162.
71. Pomp W, Meeuss en JV, Lenstra TL: Transcription factor exchange enables prolonged transcriptional bursts. biorxiv 2023, doi:10.1101/2023.05.15.540758.
72. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, Li CH, Shrinivas K, Manteiga JC, Hannett NM, et al. : Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell 2018, 175:1842–1855.e16. [PubMed: 30449618]
73. Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, et al. : Coactivator condensation at super-enhancers links phase separation and gene control. Science 2018, 361.
74. Zhu J, Salvatella X, Robustelli P: Small molecules targeting the disordered transactivation domain of the androgen receptor induce the formation of collapsed helical states. Nat Commun 2022, 13:6390. [PubMed: 36302916]
75. Powers SK, Holehouse AS, Korasick DA, Schreiber KH, Clark NM, Jing H, Emenecker R, Han S, Tycksen E, Hwang I, et al. : Nucleo-cytoplasmic Partitioning of ARF Proteins Controls Auxin Responses in Arabidopsis thaliana. Mol Cell 2019, 76:177–190.e5. [PubMed: 31421981]
76. Jung J-H, Barbosa AD, Hutin S, Kumita JR, Gao M, Derwort D, Silva CS, Lai X, Pierre E, Geng F, et al. : A prion-like domain in ELF3 functions as a thermosensor in Arabidopsis. Nature 2020, 585:256–260. [PubMed: 32848244]
77. Trojanowski J, Frank L, Rademacher A, Mücke N, Grigaitis P, Rippe K: Transcription activation is enhanced by multivalent interactions independent of phase separation. Mol Cell 2022, 82:1878–1893.e10. [PubMed: 35537448] High-resolution imaging of a synthetic system demonstrates how transcriptional activation can be completely uncoupled from phase separation. The combination of a dCas system with Optogenetics probed the relationship between phase separation and multivalency. In several systems, phase-separated droplets resulted in a decrease of total and nascent RNA levels at synthetic locus. VPR and FUSN fusion resulted in higher nascent and total RNA concomitant with phase separation; however, in cells with comparable VPR & FUSN concentration that did not form droplets, increased RNA levels were comparable.
78. Gaglia G, Lahav G: Constant rate of p53 tetramerization in response to DNA damage controls the p53 response. Mol Syst Biol 2014, 10:753. [PubMed: 25344068]
79. Scholes NS, Weinzierl ROJ: Molecular Dynamics of "Fuzzy" Transcriptional Activator-Coactivator Interactions. PLoS Comput Biol 2016, 12:e1004935. [PubMed: 27175900]
80. Ludwig CH, Thurm AR, Morgens DW, Yang KJ, Tycko J, Bassik MC, Glaunsinger BA, Bintu L: High-Throughput Discovery and Characterization of Viral Transcriptional Effectors in Human Cells. bioRxiv 2022, doi:10.1101/2022.12.16.520835. This systematic screen of viral proteins uncovered new activation domains and repression domains. Includes a careful comparison of mutations in minimal activation domains and full-length proteins. Activation domains are enriched for the ΦxxΦΦ motif surrounded by acidic residues. Repression domains are enriched for the ΦxxΦΦ motif surrounded by basic residues. This result emphasizes how motif context is as important as motif composition.

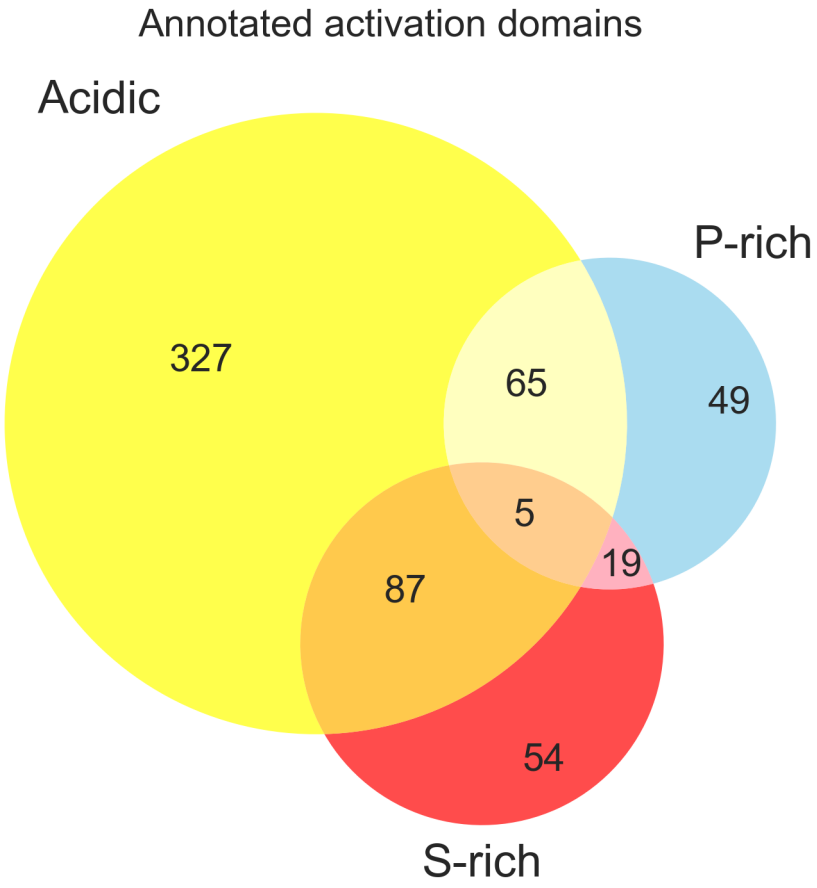

Figure 1:
Among annotated activation domains, the traditional classes are highly overlapping. Acidic activation domains have a net charge < −3, P-rich have >15% proline, and S-rich have >15% serine.

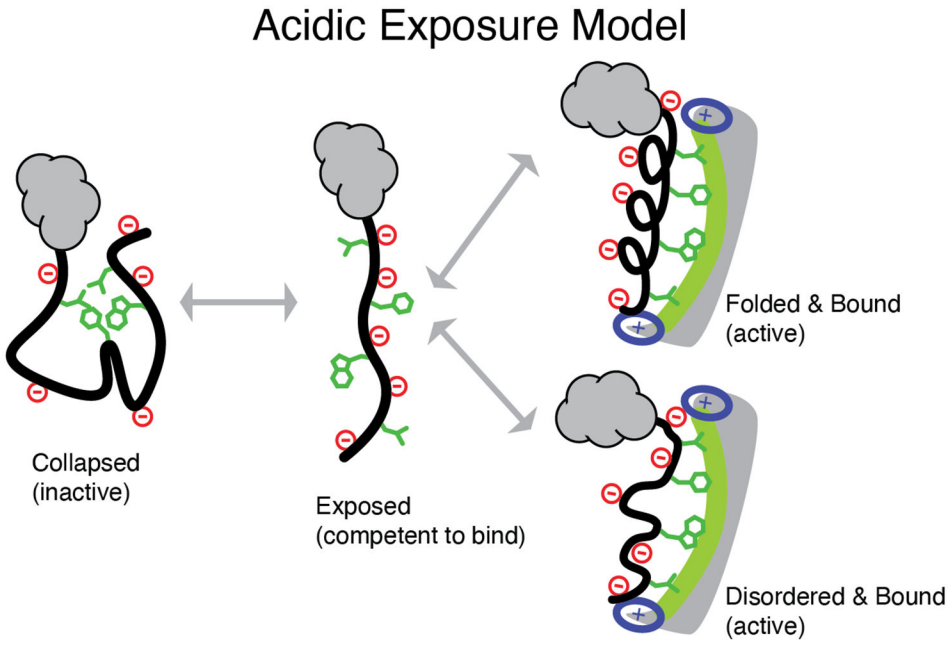

Figure 2:

In our acidic exposure model, disordered activation domains rapidly transition between collapsed and expanded states. The collapsed state is inactive. The expanded state is competent to bind coactivators because the W,F,Y,L residues are exposed to solvent. The W,F,Y,L residues make critical contacts with hydrophobic surfaces on the coactivator. Many activation domains experience coupled folding and binding, but folding is not essential. Electrostatic interactions between the activation domain and coactivator can contribute to binding or steering, but these interactions are of low affinity and not always necessary.