A sequence-specific transcription activator motif and powerful synthetic variants that bind Mediator using a fuzzy protein interface

Linda Warfield^a, Lisa M. Tuttle^{a,b}, Derek Pacheco^a, Rachel E. Klevit^{b,1}, and Steven Hahn^{a,1}

^aDivision of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; and ^bDepartment of Biochemistry, University of Washington, Seattle, WA 98195

Edited* by Mark Ptashne, Memorial Sloan Kettering Cancer Center, New York, NY, and approved July 23, 2014 (received for review June 26, 2014)

Although many transcription activators contact the same set of coactivator complexes, the mechanism and specificity of these interactions have been unclear. For example, do intrinsically disordered transcription activation domains (ADs) use sequence-specific motifs, or do ADs of seemingly different sequence have common properties that encode activation function? We find that the central activation domain (cAD) of the yeast activator Gcn4 functions through a short, conserved sequence-specific motif. Optimizing the residues surrounding this short motif by inserting additional hydrophobic residues creates very powerful ADs that bind the Mediator subunit Gal11/Med15 with high affinity via a “fuzzy” protein interface. In contrast to Gcn4, the activity of these synthetic ADs is not strongly dependent on any one residue of the AD, and this redundancy is similar to that of some natural ADs in which few if any sequence-specific residues have been identified. The additional hydrophobic residues in the synthetic ADs likely allow multiple faces of the AD helix to interact with the Gal11 activator-binding domain, effectively forming a fuzzier interface than that of the wild-type cAD.

Mediator complex | protein NMR

Transcription activators are regulators of cell identity, cell growth, and the response to environmental conditions. These highly regulated factors contain one or more activation domains (ADs) that typically bind coactivators—the complexes that contact the transcription machinery and/or have chromatin modifying activity (1–5). AD–coactivator binding initiates a cascade of events leading to productive transcription including targeted chromatin remodeling and stimulation of both RNA polymerase II preinitiation complex formation and transcription elongation (6). Many broadly acting ADs bind several coactivators, allowing them to function at a wide range of promoters with different coactivator requirements (7–14). The function of most tested ADs is conserved among eukaryotes (15, 16), even though some key activator targets are not conserved. For example, the herpes virus protein VP16 strongly activates transcription in both yeast and mammalian cells, although human Med25, a critical target of VP16 in humans, is not found in yeast (17, 18). Thus, the ability to adapt to different coactivator targets is seemingly a key property of ADs.

Defining what constitutes a functional AD has been difficult, because there is little apparent sequence similarity among different ADs. All structurally characterized eukaryotic ADs lack a stable 3D structure and are disordered in the absence of a coactivator target (19–25). Initial studies classified ADs based on enriched residues: acidic proline-, glutamine-, or serine-rich activators (26). However, these residue types later were found to be generally enriched in intrinsically disordered proteins and were termed “disorder-promoting residues” (27, 28).

Attempts to define functional sequence motifs within ADs have led to ambiguous results. For many ADs, mutagenesis showed that one to three hydrophobic residues are important for activity, although often the bulky hydrophobic property of these

side chains is more important than the specific residue identity (29–34). VP16, containing tandem ADs, is a well-characterized and representative example. In the VP16 N-terminal AD (residues 412–456), only residue F442 is critical for function and cannot be replaced by any other side chain, whereas the flanking residues L439 and L444 are functionally important but can be replaced by other bulky hydrophobic side chains with little loss of activity (35, 36). In the C-terminal AD (residues 456–490), three phenylalanine residues are important for function, but they all can be individually replaced by other bulky hydrophobic residues with little effect on activity (37). Individual acidic residues within acidic ADs generally are not critical for function as long as a sufficient number of acidic residues are retained (19, 35, 38). These and other results raise the question of how so few sequence-specific residues can generate specificity in AD function and their interactions with coactivators.

In systems other than transcription, interactions between disordered peptides and structured proteins often are mediated although short linear motifs (SLMs) (39–42). These short motifs, ~3–10 residues in length, are found in otherwise unrelated proteins where they mediate molecular interactions, e.g., peptide motifs mediating specific binding to the SH2, SH3, and 14-3-3 domains, as well as motifs such as kinase, acetylase, or methylase recognition sites. Because of their fairly low information content, SLMs favor dynamic, low-affinity, and regulated interactions, and these short motifs can be generated readily through evolution to create new molecular interactions. Although some AD

Significance

How transcription activators recognize their coactivator targets is a longstanding question and is important for understanding activator specificity and synergy. Most activators are not obviously related in sequence, but they recognize a common set of coactivators, raising the question of whether these interactions are sequence-specific. We show that the yeast transcription factor Gcn4 central activation domain works via a short sequence-specific motif that can be optimized to generate powerful synthetic activators. Like many natural activators, the synthetic derivatives have redundant sequence and bind the Mediator subunit Gal11 with high affinity using a “fuzzy” protein interface. Our results suggest a mechanism to explain how a subset of natural activators use redundant sequence motifs and great flexibility in the binding interface to target unrelated coactivators.

Author contributions: L.W., L.M.T., R.E.K., and S.H. designed research; L.W., L.M.T., D.P., and S.H. performed research; L.W., L.M.T., D.P., R.E.K., and S.H. analyzed data; and L.W., L.M.T., D.P., R.E.K., and S.H. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

¹To whom correspondence may be addressed. Email: klevit@u.washington.edu or shahn@fhcr.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1412088111/-DCSupplemental.

properties are seemingly consistent with those of SLMs, a major difference is that ADs within different transcription factors are not obviously related in sequence, whereas SLMs are readily recognized in otherwise unrelated proteins.

Yeast Gcn4, an acidic activator with tandem ADs, regulates at least 70 genes in response to metabolic stress. Gcn4 is broadly acting, regulating both TATA-containing and TATA-less genes dependent on the coactivators Mediator, SAGA, and/or TFIID (7, 43, 44). The best-characterized Gcn4 coactivator target is the Mediator subunit Gal11/Med15 that contains at least three Gcn4 activator-binding domains (ABDs), ABD1, 2, and 3 (7, 10). NMR analysis showed that the Gcn4 central AD (cAD)–ABD1 complex is dynamic, with a short half-life, mediated entirely by hydrophobic interactions between three Gcn4 residues and a shallow hydrophobic groove in Gal11 ABD1 (19). The Gcn4 cAD is intrinsically disordered but binds Gal11 ABD1 as a helix in a “fuzzy” complex in which the purely hydrophobic protein–protein interface allows the Gcn4 helix to bind Gal11 in multiple orientations. Although this fuzzy interface provides an explanation for how some ADs bind unrelated coactivators, it raises the question of whether this type of AD ever could have a specific sequence motif.

Here we show that the Gcn4 cAD does work through a short, conserved, sequence-specific motif. However, the function of this natural motif is not optimized, and its activity can be enhanced by insertion of hydrophobic residues surrounding the motif to create very powerful synthetic activators. These synthetic ADs bind Gal11 with high affinity while still using a fuzzy interface, and, as in many natural ADs, their activation function is not strongly dependent on any one residue. We present a model to explain the action of these synthetic ADs, the nature of their redundant sequence motifs, and the mechanism of fuzzy coactivator binding.

Results

Hydrophobic Residues Cannot Substitute for the Specific Gcn4 cAD Motif. NMR analysis of the Gcn4 cAD–Gal11 ABD1 complex showed that Gcn4 residues W120, L123, and F124 directly contact multiple residues within a shallow hydrophobic groove on ABD1 (19). Consistent with the structural analysis, alanine substitution showed that these three residues are essential for AD function, whereas none of the 10 acidic residues within the cAD are required. Given the simple and fuzzy nature of the protein–protein interface, it was not obvious that there is a specific sequence requirement for cAD function other than the spacing pattern of three hydrophobic residues, which allows alignment of these side chains on one face of a helix.

As an initial test of whether the cAD has a specific sequence motif, we looked for another naturally occurring disordered polypeptide with the same pattern of hydrophobic residues but with a different sequence. The Gcn4 N-terminal AD (nAD; residues 1–100) is intrinsically disordered (19) and contains four clusters of hydrophobic residues important for function (19, 45). One of these clusters, found within residues 81–100 [sequence 93-VVESFF-98 (key hydrophobic residues underlined)], contains hydrophobic residues with the cAD spacing of *i*, *i*+3, *i*+4 embedded within an acidic disordered polypeptide (Fig. 1A). To test whether this motif has AD function, we fused Gcn4 residues 81–100 to the Gcn4 DNA-binding domain (DBD) via a flexible Gly–Ser linker and assayed AD function in vivo at the Gcn4- and Gal11-dependent genes *ARG3* and *HIS4*, respectively (Fig. 1B). Gcn4 mRNA was expressed under control of its natural promoter, and induction of Gcn4 translation was induced by the addition of sulfometuron methyl (SM), an inhibitor of amino acid synthesis, for 90 min before mRNA quantitation. In contrast to the WT Gcn4 cAD and full-length nAD (Table S1), Gcn4 residues 81–100 have no in vivo AD function, showing that AD activity involves more than just a simple pattern of bulky hydrophobic residues. However,

Fig. 1. The Gcn4 cAD is encoded by a specific sequence motif. (A) Sequences of the Gcn4 cAD (residues 101–134), the fourth hydrophobic cluster within the Gcn4 nAD (residues 81–100), and this hydrophobic cluster mutated to a cAD-like motif (Gcn4 cAD-like). Polypeptides were fused to the Gcn4 DBD using a glycine–serine linker. Acidic residues are colored blue, and key Gcn4 cAD residues and corresponding residues in the nAD hydrophobic cluster are colored red. (B) In vivo function of the AD derivatives. mRNA levels from the *ARG3* and *HIS4* genes were measured in yeast strains containing either vector (no Gcn4) or the Gcn4 derivatives shown in A. Strains were induced with 0.5 μ g/mL SM for 90 min. The mRNA levels were normalized to *ACT1* and are expressed as fold induction over Gcn4 cAD-like without SM.

changing the sequence of this hydrophobic pattern to that of the cAD [VVESFF→VWESLF (key hydrophobic residues underlined)] created a functional cAD-like AD with activity equivalent to that of the Gcn4 cAD (81–100 cAD-like) (Fig. 1B). Therefore, the activity of the cAD is encoded by a specific sequence motif rather than by a pattern of generic hydrophobic residues, and the motif can function when embedded in a disordered polypeptide of different sequence.

A Short AD Sequence Motif with a Preference for Surrounding Hydrophobic Residues. To characterize the cAD sequence motif, we used the Gcn4 81–100 cAD-like polypeptide as the template for extensive mutagenesis. As a first step, all side chains within a nine-residue segment centered on the three key hydrophobic residues were individually changed to all other side chains. The resulting derivatives were assayed in vivo for AD function at *ARG3*. Fig. 2 graphically illustrates the effects of mutations on AD function, and quantitative results are shown in Table S1. Surprisingly, the W15, L18, and F19 residues in the starting sequence were the optimal residues for function of the motif. Replacement of W15 with nearly any other residue completely abolished AD function. Similar results for the specificity of residues L18 and F19 were observed, except that substitution of tryptophan at these positions gave ~70–90% of the starting AD function. In sum, the three sequence-specific natural residues in the cAD motif are optimal, but the aromatic residue tryptophan

Fig. 2. Sequence specificity of the Gcn4 cAD motif. Residues within a nine-residue segment surrounding the cAD-like motif shown in the bracket were individually changed to all other residues, and these derivatives were tested for in vivo function at *ARG3*. In the chart, the starting cAD-like motif is in bold, and the key hydrophobic residues are colored red. The effects of single amino acid changes are displayed as the ability of the Gcn4 motif containing the mutation to activate transcription at *ARG3*, expressed as fold induction over the uninduced cAD-like AD.

can be substituted at two positions to generate an AD with significant activity.

In contrast to this sequence specificity, the four residues surrounding W15 are much more permissive in tolerated substitutions. About half of the substitutions at these positions produced AD function that was at least as good or was better than that of the starting cAD-like sequence, with less-optimal substitutions reducing, but not eliminating, AD function. Inspection of the tolerated residues shows that in nearly all cases substitutions of hydrophobic side chains lead to greater activity than the starting cAD-like motif. For residues Ser-20 and Ser-21, at the junction between the AD and the flexible linker, nearly all substitutions worked at least as well as serine, with hydrophobic residues generally being the most active for AD function. Taken together, our results show that the optimal Gcn4 cAD-like motif is $\Phi\Phi W\Phi\Phi LF$, embedded within a disordered polypeptide, where Φ represents any hydrophobic residue.

Creation of Short, Very Powerful ADs Based on the cAD Motif. We next asked whether the individual substitutions within the cAD-like motif that increase activity are additive for function. Based on the results of Fig. 2, synthetic ADs were created that had optimal (or nearly optimal) residues at many positions within the AD (Fig. 3 and Table S2). In most cases, multiple substitutions were additive or nearly additive for activation. For example, the starting cAD-like peptide activated *ARG3* sevenfold upon SM addition, and single-residue changes showed up to threefold increased activity over the cAD-like motif. However, multiple substitutions gave a wide range of more active ADs, with several showing 30- to 90-fold *ARG3* mRNA induction. None of these strong ADs bypassed the requirement for the coactivator subunit Gal11, because activated *ARG3* transcription was very low in the absence of Gal11 (Table S2). Remarkably, several of these strong synthetic ADs are more active than the full-length WT Gcn4 with its tandem ADs that activate *ARG3* transcription ~40-fold. The most active AD derivatives contained multiple tryptophan substitutions, with the strongest synthetic AD having the

sequence motif AWWWLFWFS. Because tryptophan is the least frequently used amino acid, such tryptophan-rich sequences are rare in nature. However, it was noted previously in a one-hybrid experiment that a short tryptophan-rich peptide derived from a tRNA synthetase has a transcription activation function when fused to a DBD (46).

To test if these powerful synthetic ADs can function within a different sequence context, we inserted selected synthetic AD motifs within the Gcn4 cAD peptide, residues 101–134, and assayed in vivo function (Table S3). The activity of the synthetic motifs was unaffected by this different context, showing that the short motif, and not the surrounding disordered sequence context, is the main determinant of function.

Highly redundant nine-residue AD signatures for yeast, mammalian, and viral activators were proposed based on sequence comparison of known activators and related factors (Fig. S1A) (47). However, we found that this motif does not correlate with AD function in our system. Using the nine-residue AD search algorithm (www.med.muni.cz/9aaTAD/index.php), we found that the Gcn4 cAD did not match the proposed motif signature, whereas Gcn4 residues 81–100, which has no AD function in our system, did match (Fig. S1B and C). Similarly, Gal4 residues 862–870 match the proposed signature, but a segment of Gal4 containing this peptide (857–881) has little activity, demonstrating that the proposed nine-residue motif does not predict AD function in our experimental system.

The Synthetic Activators Are Unusually Stable Compared with Gcn4.

It is known that the cellular levels of some transcription activators are inversely correlated with their transcription AD function; the strongest activators often are the least stable (48). Further, it has been proposed that destruction of some activators after a small number of transcription initiation events is integral to the transcription cycle, allowing expression to be shut off after the initial activation signal has ended (48–50). To test whether the activity-vs.-stability relationship holds for these Gcn4 derivatives, steady-state levels of all of the ADs were measured by Western blot, probing for SM-induced Gcn4 protein via a triple Flag tag at the

Fig. 3. In vivo activity of the synthetic ADs. The starting cAD-like motif is displayed in bold with the key motif residues in red. Motifs are displayed as fold induction over the uninduced cAD-like AD at *ARG3*. Horizontal lines between mutated residues indicate multiple mutations within each individual motif. Position indicates residue number in the peptide fused to the Gcn4 DBD shown above the chart. The bracket indicates the mutagenized region.

C terminus (Fig. S2). The inverse relationship between AD function and steady-state levels generally held for most single-residue substitutions shown in Fig. 2. However, most of the synthetic ADs showed clearly distinct behavior, with much higher protein levels than expected based on their strong AD function (Fig. S2B). If activator instability is an important part of biological regulation, then perhaps this unexpected stability is one reason why Gcn4 and other natural ADs are not optimized for AD function.

AD Function Correlates with Affinity for Gal11 ABD1. To investigate why the synthetic ADs strongly activate transcription, we tested whether they function by having higher coactivator-binding affinity than the natural cAD. As a representative assay for coactivator binding, we measured the affinity of the synthetic ADs for Gal11 ABD1, a well-characterized AD-binding target. Polypeptides with the synthetic motifs in the context of Gcn4 residues 101–134 were expressed as recombinant proteins and purified. We were unable to purify many of the highly tryptophan-rich peptides, because these recombinant proteins are insoluble. However, two of the synthetic ADs, M118 (AVWWSLFAS) and M120 (AFWMWLFAT), are soluble at high concentrations, allowing us to measure binding to ABD1 using isothermal titration calorimetry (ITC) (Table 1 and Fig. S3). WT Gcn4 cAD binds ABD1 with a K_d of 10 μ M (19). In contrast, a peptide with the Gcn4 81–100 motif (AVVESFFSS) lacking *in vivo* function shows no detectable ABD1 binding. As expected, the cAD-like motif used as the template for mutagenesis in Fig. 2 efficiently bound ABD1 with a K_d of 5 μ M. The AD derivatives M118 and M120, which respectively activate *ARG3* transcription 4.3- and 5.5-fold more strongly than the cAD motif, bind ABD1 with K_d s of 1.6 and 0.9 μ M, respectively. Thus, the strong activator function of these two synthetic ADs correlates with a higher affinity for the Mediator subunit Gal11.

The Strong Synthetic ADs Bind Gal11 via a Fuzzy Complex. The WT Gcn4 cAD binds Gal11 ABD1 via a fuzzy protein interface, but whether a fuzzy interface could account for the higher-affinity binding of the synthetic ADs was not known. One possible explanation for higher affinity is that the synthetic ADs “lock down” on ABD1 in a specific, higher-affinity, and less dynamic complex. An alternative possibility is that the additional hydrophobic residues in the synthetic motif allow a greater number of possible binding modes—effectively, a fuzzier, more redundant, but higher affinity complex. To distinguish these and other possibilities, we used NMR to examine the binding of ABD1 to the synthetic ADs M118 and M120.

As a first step we assessed the chemical-shift perturbations in 15 N-labeled ABD1 upon the addition of synthetic ADs M118 or M120. Localized chemical-shift changes on ABD1 would be more consistent with the synthetic activators adopting a constrained conformation upon binding, rather than the fuzzy binding of the WT cAD. Instead, we see widespread chemical-shift changes spanning the same shallow groove of ABD1, similar

to that observed when WT cAD binds to ABD1 (Fig. 4A). The same broad surface of ABD1 is affected by binding M118, M120, and WT cAD, with the synthetic activators inducing larger chemical-shift perturbations in ABD1, as is consistent with the increase in the number of aromatic residues in the AD and tighter ABD1 binding. The widespread chemical-shift changes on the surface of ABD1 are consistent with a fuzzy binding motif but could also be explained if the synthetic activators form longer helical regions upon binding than is the case for WT cAD.

To determine directly if the synthetic ADs bind ABD1 with a fuzzy interface, we performed NMR spin-label experiments. Analogous to the spin-label experiments previously performed with WT cAD (19), the paramagnetic spin-label TEMPO was incorporated at position 117 in the synthetic AD M120. A spin label induces loss of NMR peak intensity at sites proximal to the label. Intensity perturbations in the 15 N-heteronuclear single-quantum coherence spectroscopy (HSQC) of ABD1 upon binding to spin-labeled cAD or M120 are shown in Fig. 4B plotted on the structure of ABD1. We found that the spin-labeled M120 induces perturbations over an even larger area of ABD1 than does spin-labeled cAD. The large perturbations at distal sites of ABD1 are consistent with a fuzzy binding interface for the synthetic AD M120, wherein the helical binding region adopts many orientations with respect to the binding surface of ABD1. Therefore, although the addition of large hydrophobic residues to the cAD motif generally increases both activator activity and the binding affinity to ABD1, the mutations do not result in a single, preferred bound orientation of activator and ABD1.

Unbound WT cAD was determined to be intrinsically disordered, as supported by the very narrow range of 1 H chemical shifts in the 15 N-HSQC (Fig. S4A) (19). In contrast, the more dispersed 1 H chemical shifts of unbound M120 suggest the presence of some secondary structure even in the absence of a binding partner. Chemical-shift assignments of M120 were completed to gain structural information about the unbound synthetic AD. Comparison of measured α and β chemical shifts with those predicted for a random coil protein reveals information about secondary structure (51). As expected, the α and β chemical shifts of unbound WT cAD reveal no significant adoption of helical structure. However, the shifts for M120 show significant helical character for the region spanning residues 117–125 (Fig. 4C). We confirmed the increased helical propensity of M120 relative to cAD, at even lower concentrations than used for the NMR, by circular dichroism (CD) spectroscopy (Fig. S5). Far-UV CD is highly sensitive to protein secondary structure. CD scans of cAD and M120 were collected with increasing concentrations of trifluoroethanol (TFE), which stabilizes any intrinsic helical secondary structure (52). The minima at 208 and 222 nm characteristic of helical secondary structure are more prominent for M120 than for the cAD (Fig. S5C), as is consistent with the NMR analysis showing the helical character of the unbound M120 peptide.

The region with detectable helical character in unbound M120 (residues 117–125) is analogous to the region in the cAD that becomes helical upon binding to ABD1. The average amide chemical-shift difference upon binding to ABD1 generally is larger for the cAD than for M120 (Fig. 4D). This finding is consistent with the notion from the CD and NMR chemical shifts that the structure of unbound M120 resembles its bound structure more closely than is the case for the WT cAD. The preformed helical structure of M120 likely contributes to its increased binding affinity.

Like Many Natural ADs, the Synthetic AD Motifs Are Redundant. Because the synthetic AD motifs still bind Gal11 via fuzzy interfaces, one possibility is that the additional hydrophobic residues within the short motifs allow greater flexibility in binding ABD1

Table 1. Affinity of AD derivatives for Gal11 ABD1 measured by ITC

AD peptide	ABD1 K_d , μ M
Gcn4 cAD (101–134)	10.1 $\pm$ 1.4
Gcn4 81–100	ND
Gcn4 cAD-like	5.1 $\pm$ 0.35
Gcn4 M118	1.6 $\pm$ 0.11
Gcn4 M120	0.9 $\pm$ 0.09

Shown are the Gcn4 derivatives and their measured affinities for Gal11 ABD1. ND, not determined, too weak to measure. Data on WT cAD are from ref. 19.

Fig. 4. (A) The synthetic activators M118 and M120 bind to the same surface of ABD1 as does the cAD. The average change in amide chemical shift between free and bound ABD1 is plotted on the structure of ABD1 [Protein Data Bank (PDB) ID code 2LPB] with a red-to-white gradient indicating most to least affected: $\Delta\delta_{\text{avgNH}} = \sqrt{(\Delta\delta_{\text{H}})^2 + (\Delta\delta_{\text{N}}/5)^2}$. The widespread changes in chemical shift are suggestive of fuzzy binding. Data on WT cAD are from ref. 19. (B) M120 forms a fuzzy complex with ABD1. A paramagnetic spin label was incorporated at position S117 (mutated to cysteine in M120). The loss of signal intensity in NMR ^{15}N -HSQC experiments demonstrate the proximity of residue 117 to residues of ABD1. Loss of intensity is plotted on the structure of ABD1 (PDB ID code 2LPB) with a red-to-yellow gradient indicating most-to-least affected; prolines and residues retaining >50% signal intensity are gray. Results for M120 (Lower) resemble those previously published (19) for WT cAD (Upper). (C) Free M120 shows helicity in solution, whereas the cAD is unstructured. The difference between the measured α and β chemical shifts to random coil expected values (51) is indicative of a secondary structure. Larger positive values are indicative of a helical secondary structure, as seen for M120 residues ~117–125. (D) Average chemical-shift change in cAD and M120 upon binding to ABD1. For M120, the peaks of residues 121 and 123 broaden beyond detection (Fig. S4B). Generally, cAD shows larger chemical-shift changes in the binding region, as is consistent with a greater structural rearrangement upon binding ABD1. Unbound M120, with its preformed helical structure, more closely resembles its bound structure.

via alternative interactions. To test this possibility, we queried whether the starting tryptophan, leucine, and phenylalanine residues within the synthetic motifs are critical for function of the synthetic ADs. Fig. 5 shows that although replacing the tryptophan, leucine, and phenylalanine residues by alanine abolished the function of the WT cAD motif, these same alanine substitutions gave at most a twofold decrease in AD function in the context of the synthetic ADs. For example, in the synthetic AD M120, the tryptophan-to-alanine substitution led to a twofold decrease in function, whereas AD function was reduced by only 15–20% by alanine replacement of the leucine and phenylalanine residues. In the M118 and M121 derivatives, replacement of the tryptophan, leucine, or phenylalanine residues by alanine led to an approximately twofold decrease in activity but had little or no effect on function of the M122 motif. Such behavior is similar to that seen in some natural ADs where it has been difficult to assign important residues within the AD by mutagenesis of single residues (35, 53). Together, the results suggest that many ADs, including our synthetic ADs, contain redundant sequence motifs that allow alternative modes of binding to ABDs.

Discussion

To understand the mechanisms of transcription activation, one important goal is to determine how coactivators such as Mediator integrate signals from multiple factors to modulate transcription precisely. Although many transcription activators and coactivators are known, relatively little is understood about the mechanism and specificity of these interactions. Although a few ADs bind coactivators in complexes with relatively high affinity and specific geometry, many show dynamic binding and modest affinity in coactivator interactions. ADs often bind to several structurally unrelated coactivators, and it has been difficult to understand how ADs can adapt to different partners. Also confusing are the findings that unrelated ADs are not obviously related in sequence. What properties, if any, do ADs targeting the same coactivators have in common? All known ADs are intrinsically disordered polypeptides, and mutagenesis generally has revealed that a small number of hydrophobic residues are important for activity. However, no sequence-specific motif-encoding AD function has been identified to date.

The Gcn4 cAD binds the Mediator subunit Gal11/Med15 using a fuzzy protein–protein interface, potentially explaining how

Fig. 5. Synthetic AD motifs have redundant sequence. Analysis of *ARG3* mRNA from yeast strains containing either the Gcn4 cAD-like AD or the synthetic ADs and with single alanine substitutions within each of these motifs. *ARG3* mRNA levels were normalized to *ACT1* mRNA and are expressed as fold induction over uninduced Gcn4 cAD-like, with each starting motif set to 100% so that the effects of the single alanine mutations are compared with the starting motif.

many ADs bind to seemingly unrelated coactivators (19). If this binding mechanism is representative of a large class of ADs, then we expect that ABDs will not be very specific, allowing them to bind ADs of different sequences. Similarly, ADs of this type are predicted to have fairly low information content, allowing new AD functions to develop readily during evolution. Although peptide interactions with structured proteins in other systems often are mediated through short linear motifs, no short linear motifs have yet been recognized in ADs. However, AD sequence motifs that bind via a fuzzy interface may differ from those in complexes with specific geometry.

Structural studies found that the Gcn4 cAD uses three hydrophobic residues on one face of a helix to bind Gal11 ABD1. In accordance with the conservation of these three Gcn4 residues in closely related yeasts (Fig. S6), we show here that this short motif is sufficient to activate transcription when transferred to another disordered polypeptide. Our extensive mutagenesis also showed that the sequence of the three WT cAD hydrophobic residues is optimal for AD function. Furthermore, surrounding this motif with additional hydrophobic residues generates short and very powerful synthetic ADs. These synthetic ADs bind ABD1 with up to 10-fold higher affinity, providing an explanation for their powerful AD function.

We found that two of the synthetic ADs function via a fuzzy protein interface. Part of the reason for the higher activity of the synthetic ADs is that the additional hydrophobic residues within the motif can provide redundant interactions with coactivator targets. This redundancy was revealed by mutations within the tryptophan, leucine, and phenylalanine residues that are essential for function of the WT cAD but are only modestly important in the context of the synthetic ADs. In the WT cAD-ABD1 complex, a single side of the helix contains the tryptophan, leucine, and phenylalanine residues that face the ABD1 hydrophobic groove in multiple orientations (Movie S1). For the synthetic ADs, the extra hydrophobic residues in the motif yield a more extensive hydrophobic surface on the helix, likely allowing the peptide to rotate along its helical axis to present multiple faces of the helix to the ABD1 hydrophobic groove (Movie S2) in

addition to binding ABD1 in multiple orientations. This more extensive hydrophobic surface decreases the dependence on the original WxxLF motif. In other words, we predict that the higher affinity of the synthetic activators is generated in part by a fuzzier interface.

Our analysis also showed that the synthetic AD M120 has some helical character in the absence of a binding partner. In contrast to the disordered WT cAD, M120 AD shows a propensity for helix formation in the absence of Gal11. Thus, the strong activity of the synthetic ADs is likely a combination of the extra hydrophobic residues that make redundant contacts with ABDs and their stronger intrinsic helical character.

Our results raise the question of why the Gcn4 AD (and presumably other ADs) is not optimized for activity. The optimized synthetic motifs show that activity of the natural cAD can be increased >13-fold, with activity greater than that of full-length Gcn4. Overexpression of a strong AD is known to be detrimental and squelches transcription of other genes (54). Therefore, constitutive expression of a very strong AD may outcompete other natural ADs for their coactivator targets. In accordance with this notion, cell growth rate is ~30% slower when Gcn4 is constitutively expressed. Also, many natural genes are synergistically activated by combinations of factors, and a very powerful activator will likely override the need for multiple inputs in gene activation. Given these considerations, there may be no biological advantage to optimizing the function of most ADs apart from viral activators.

Finally, our results suggest that multiple sequence motifs function as transcription ADs. Although the Gcn4 cAD is one specific motif with AD function, it apparently is not used in most other yeast activators. A search of the yeast proteome for the WxxLF motif finds that it occurs in 180 yeast proteins, but only four of these (Gcn4, Gal4, Stb5, and Hir3) are predicted by sequence analysis to be sequence-specific transcription factors. Apart from Gcn4, none of the WxxLF motifs in the other transcription factors appear in known ADs. For example, the sequence WQILF is found in a helical segment of Gal4 that lies between the two known ADs and likely is involved in regulating AD function (55, 56). Therefore, we propose that multiple sequence-specific motifs embedded within disordered peptides can target the same set of ABDs within coactivators. It remains to be determined how often ADs are comprised of these short motifs and whether there are other classes of ADs that function by different mechanisms.

Materials and Methods

Yeast Strains and Plasmids. Yeast strain SHY823 ($\Delta gcn4$, $leu2$) was transformed with the following *LEU2*- and *GCN4*-derivative plasmids for protein expression and mRNA analysis: pRS315 (vector, no *GCN4*), pSH940 (WT *GCN4*), pSH942 (Gcn4 101–281; cAD), pSH943 (Gcn4 1–100; nAD), pSH1021 (Gcn4 81–100), or pSH1068 (Gcn4 81–100; cAD-like). Gcn4 motif mutations were generated in pSH1068 by site-directed mutagenesis. All Gcn4 derivatives except pSH942 contained the flexible linker GSGSGS linking the AD to Gcn4 125–281 containing the DBD. All Gcn4 derivatives contained a C-terminal 3x-Flag tag. Additional Gcn4 motif mutations were made in pSH942 as noted in the text to be in the context of the cAD (Table S3). Yeast strain SHY822 ($\Delta gcn4$, $\Delta gal11$, $leu2$) was used to test the Gal11 dependence of the synthetic ADs (Table S2).

mRNA Analysis. Yeast strains were grown in duplicate to an OD_{600} of 0.5–0.8 in 2% (wt/vol) dextrose synthetic complete Ile-Val-Leu medium at 30 °C. Cells were induced with 0.5 μ g/mL SM for 90 min to induce amino acid starvation (7), RNA was extracted and assayed in triplicate by RT-quantitative PCR, and the results were analyzed as described (7).

Quantitation of in Vivo Gcn4 Levels. Cells (1.5 mL) from the cultures used for the above mRNA analysis were pelleted and washed in water. Cells were resuspended in 1x lithium dodecyl sulfate sample buffer (Life Technologies) containing 50 mM DTT and then were lysed at 95 °C for 5 min, spun for 5 min, separated on a 4–12% Bis-Tris gel with 3-(N-morpholino)propanesulfonic acid

buffer, and then analyzed by Western blot. Gcn4-3×Flag was visualized with mouse anti-FLAG (Sigma) and Tfg2 with rabbit polyclonal anti-Tfg2 (S.H. laboratory). Western blots were quantitated with the Odyssey scanner and software (LI-COR Bioscience) using a titration of Gcn4 81–100 to standardize protein amounts. All motifs were normalized to the Tfg2 loading control.

Recombinant Protein Expression and Purification. Gal11 and Gcn4 derivatives were expressed as small ubiquitin-like modifier (SUMO) fusion proteins (Invitrogen), and the SUMO tag was removed during purification. Expression constructs and purification of Gal11 158–238 and Gcn4 derivatives were as described (7). Bacterial cells for expression of isotopically labeled proteins used in NMR were grown and peptides were purified as described (19).

ITC and Fluorescence Polarization. ITC experiments were carried out as described (19) using Gcn4 101–134 containing the AD motif sequences as indicated below. Gcn4 101–134 WT was added over the course of 21 injections as previously described (19). All other titrations were carried out with 16 injections at 0.5 μ L/s (injection 1 = 0.3 μ L; injections 2–16 = 2.55 μ L). The following concentrations were used: Gal11 158–238 (0.27 mM) vs. Gcn4 101–134 (2.37 mM); Gal11 158–238 (1.5 mM) vs. Gcn4 81–100 motif (0.15 mM); Gal11 158–238 (1.7 mM) vs. Gcn4 cAD-like motif (0.15 mM); Gal11 158–238 (2.2 mM) vs. Gcn4 M118 motif (0.22 mM); Gal11 158–238 (0.73 mM) vs. Gcn4 M120 motif (0.07 mM).

CD Measurements. CD spectra were recorded using a Jasco J-810 Circular Dichroism Spectropolarimeter. Spectra were recorded at room temperature using a 1-mm quartz cell, a measurement range of 250–200 nm, a scanning speed of 10 nm/min, 1-nm bandwidth, 8-s response time, and 0.1-nm data pitch. The initial peptide concentration was 100 μ M in 20 mM phosphate

buffer, pH 7.0, and TFE was added in increasing volumes to the final concentration of 70% (vol/vol). Peptide concentration was determined by measuring absorbance at 280 nm for Gcn4 101–134 WT (ϵ_{280} = 6,990 M/cm) and M120 (ϵ_{280} = 12,490 M/cm) using a Nanodrop ND-1000 Spectrophotometer.

NMR Experiments and Resonance Assignments. NMR HSQC titration and spin-label experiments were completed on a Bruker 500 MHz AVANCE spectrometer. All experiments were collected at 25 °C in NMR buffer (20 mM sodium phosphate, 150 mM NaCl, 0.1 mM EDTA, 0.1 mM PMSF, and 5 mM DTT) with 5% (vol/vol) D₂O. Chemical-shift assignments were obtained on a 1 mM [¹³C, ¹⁵N]-M120 sample using standard triple-resonance experiments (HNCA, HNCOC, HNCACB, and HNCO spectra) on a Bruker 800 MHz spectrometer. ¹⁵N-HSQC titration experiments were completed by adding unlabeled M118 or M120 to a 200 μ M ¹⁵N-ABD1 sample to a final ratio of 5:1 synthetic AD:ABD1, maintaining a constant concentration of ¹⁵N-ABD1. The reverse titration, adding unlabeled ABD1 to ¹⁵N-M120, was performed similarly. Paramagnetic relaxation enhancement experiments were performed as for cAD (19), with the spin-label 4-(2-iodoacetamido)-TEMPO incorporated at S117C of M120. HSQC spectra of 200 μ M ¹⁵N-ABD1 and 400 μ M of spin-labeled M120 were collected in the presence (reference intensity) and absence of ascorbic acid.

ACKNOWLEDGMENTS. We thank P. Rajagopal, P. Brzovic, and other members of the R.E.K. and S.H. laboratories for their suggestions and comments during this work; S. Petesch for sequence analysis of AD motifs and communication of unpublished results; and J. Söding for many insightful discussions on the nature of transcription ADs and for comments on the manuscript. This work was funded by National Institutes of Health (NIH) Grant R01GM075114 (to S.H.). R.E.K. is supported by NIH Grant R01GM088055.

- Näär AM, Lemon BD, Tjian R (2001) Transcriptional coactivator complexes. *Annu Rev Biochem* 70:475–501.
- Green MR (2005) Eukaryotic transcription activation: Right on target. *Mol Cell* 18(4):399–402.
- Ptashe M, Gann A (2002) *Genes and Signals* (Cold Spring Harbor Laboratory Press, Woodbury, NY).
- Poss ZC, Ebmeier CC, Taatjes DJ (2013) The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol* 48(6):575–608.
- Hahn S, Young ET (2011) Transcriptional regulation in *Saccharomyces cerevisiae*: Transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* 189(3):705–736.
- Weake VM, Workman JL (2010) Inducible gene expression: Diverse regulatory mechanisms. *Nat Rev Genet* 11(6):426–437.
- Herbig E, et al. (2010) Mechanism of Mediator recruitment by tandem Gcn4 activation domains and three Gal11 activator-binding domains. *Mol Cell Biol* 30(10):2376–2390.
- Reeves WM, Hahn S (2005) Targets of the Gal4 transcription activator in functional transcription complexes. *Mol Cell Biol* 25(20):9092–9102.
- Fishburn J, Mohibullah N, Hahn S (2005) Function of a eukaryotic transcription activator during the transcription cycle. *Mol Cell* 18(3):369–378.
- Jedidi I, et al. (2010) Activator Gcn4 employs multiple segments of Med15/Gal11, including the KIX domain, to recruit mediator to target genes in vivo. *J Biol Chem* 285(4):2438–2455.
- Langlois C, et al. (2008) NMR structure of the complex between the Tfb1 subunit of TFIID and the activation domain of VP16: Structural similarities between VP16 and p53. *J Am Chem Soc* 130(32):10596–10604.
- Mittler G, et al. (2003) A novel docking site on Mediator is critical for activation by VP16 in mammalian cells. *EMBO J* 22(24):6494–6504.
- Yang F, DeBeaumont R, Zhou S, Näär AM (2004) The activator-recruited cofactor/Mediator coactivator subunit ARC92 is a functionally important target of the VP16 transcriptional activator. *Proc Natl Acad Sci USA* 101(8):2339–2344.
- Liu W-L, et al. (2009) Structures of three distinct activator-TFIID complexes. *Genes Dev* 23(13):1510–1521.
- Fischer JA, Giniger E, Maniatis T, Ptashne M (1988) GAL4 activates transcription in *Drosophila*. *Nature* 332(6167):853–856.
- Fields S, Jang SK (1990) Presence of a potent transcription activating sequence in the p53 protein. *Science* 249(4972):1046–1049.
- Milbradt AG, et al. (2011) Structure of the VP16 transactivator target in the Mediator. *Nat Struct Mol Biol* 18(4):410–415.
- Vojnic E, et al. (2011) Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nat Struct Mol Biol* 18(4):404–409.
- Brzovic PS, et al. (2011) The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. *Mol Cell* 44(6):942–953.
- Di Lello P, et al. (2006) Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIID and the activation domain of p53. *Mol Cell* 22(6):731–740.
- Feng H, et al. (2009) Structural basis for p300 Taz2-p53 TAD1 binding and modulation by phosphorylation. *Structure* 17(2):202–210.
- Kussie PH, et al. (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274(5289):948–953.
- Sigler PB (1988) Transcriptional activation. Acid blobs and negative noodles. *Nature* 333(6170):210–212.
- Uesugi M, Nyanguile O, Lu H, Levine AJ, Verdine GL (1997) Induced alpha helix in the VP16 activation domain upon binding to a human TAF. *Science* 277(5330):1310–1313.
- Wells M, et al. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci USA* 105(15):5762–5767.
- Mitchell PJ, Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245(4916):371–378.
- Theillet F-X, et al. (2013) The alphabet of intrinsic disorder I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disordered Proteins* 1:e24360.
- Tomba P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27(10):527–533.
- Lin J, Chen J, Elenbaas B, Levine AJ (1994) Several hydrophobic amino acids in the p53 amino-terminal domain are required for transcriptional activation, binding to mdm-2 and the adenovirus 5 E1B 55-kD protein. *Genes Dev* 8(10):1235–1246.
- Chang J, Kim D-H, Lee SW, Choi KY, Sung YC (1995) Transactivation ability of p53 transcriptional activation domain is directly related to the binding affinity to TATA-binding protein. *J Biol Chem* 270(42):25014–25019.
- Iñiguez-Lluhi JA, Lou DY, Yamamoto KR (1997) Three amino acid substitutions selectively disrupt the activation but not the repression function of the glucocorticoid receptor N terminus. *J Biol Chem* 272(7):4149–4156.
- Almlöf T, Gustafsson JA, Wright AP (1997) Role of hydrophobic amino acid clusters in the transactivation activity of the human glucocorticoid receptor. *Mol Cell Biol* 17(2):934–945.
- Gill G, Pascal E, Tseng ZH, Tjian R (1994) A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTFII110 component of the *Drosophila* TFIID complex and mediates transcriptional activation. *Proc Natl Acad Sci USA* 91(1):192–196.
- Sainz MB, Goff SA, Chandler VL (1997) Extensive mutagenesis of a transcriptional activation domain identifies single hydrophobic and acidic amino acids important for activation in vivo. *Mol Cell Biol* 17(1):115–122.
- Cress WD, Triezenberg SJ (1991) Critical structural elements of the VP16 transcriptional activation domain. *Science* 251(4989):87–90.
- Regier JL, Shen F, Triezenberg SJ (1993) Pattern of aromatic and hydrophobic amino acids critical for one of two subdomains of the VP16 transcriptional activator. *Proc Natl Acad Sci USA* 90(3):883–887.
- Sullivan SM, et al. (1998) Mutational analysis of a transcriptional activation region of the VP16 protein of herpes simplex virus. *Nucleic Acids Res* 26(19):4487–4496.
- Leuther KK, Salmeron JM, Johnston SA (1993) Genetic evidence that an activation domain of GAL4 does not require acidity and may form a beta sheet. *Cell* 72(4):575–585.
- Stein A, Pache RA, Bernadó P, Pons M, Aloy P (2009) Dynamic interactions of proteins in complex networks: A more structured view. *FEBS J* 276(19):5390–5405.
- Nguyen Ba AN, et al. (2012) Proteome-wide discovery of evolutionary conserved sequences in disordered regions. *Sci Signal* 5(215):rs1.
- Das RK, Mao AH, Pappu RV (2012) Unmasking functional motifs within disordered regions of proteins. *Sci Signal* 5(220):pe17.
- Gould CM, et al. (2010) ELM: The status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res* 38(Database issue):D167–D180.

43. Brown CE, et al. (2001) Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. *Science* 292(5525):2333–2337.
44. Swanson MJ, et al. (2003) A multiplicity of coactivators is required by Gcn4p at individual promoters in vivo. *Mol Cell Biol* 23(8):2800–2820.
45. Jackson BM, Drysdale CM, Natarajan K, Hinnebusch AG (1996) Identification of seven hydrophobic clusters in GCN4 making redundant contributions to transcriptional activation. *Mol Cell Biol* 16(10):5557–5571.
46. Lin C-H, Lin G, Chang C-P, Wang C-C (2010) A tryptophan-rich peptide acts as a transcription activation domain. *BMC Mol Biol* 11:85.
47. Piskacek S, et al. (2007) Nine-amino-acid transactivation domain: Establishment and prediction utilities. *Genomics* 89(6):756–768.
48. Geng F, Wenzel S, Tansey WP (2012) Ubiquitin and proteasomes in transcription. *Annu Rev Biochem* 81:177–201.
49. Rosonina E, Duncan SM, Manley JL (2010) SUMO functions in constitutive transcription and during activation of inducible genes in yeast. *Genes Dev* 24(12):1242–1252.
50. Chi Y, et al. (2001) Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. *Genes Dev* 15(9):1078–1092.
51. Kjaergaard M, Poulsen FM (2011) Sequence correction of random coil chemical shifts: Correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol NMR* 50(2):157–165.
52. Luo P, Baldwin RL (1997) Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* 36(27):8413–8421.
53. Ansari AZ, Reece RJ, Ptashne M (1998) A transcriptional activating region with two contrasting modes of protein interaction. *Proc Natl Acad Sci USA* 95(23):13543–13548.
54. Gill G, Ptashne M (1988) Negative effect of the transcriptional activator GAL4. *Nature* 334(6184):721–724.
55. Ma J, Ptashne M (1987) Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* 48(5):847–853.
56. Stone G, Sadowski I (1993) GAL4 is regulated by a glucose-responsive functional domain. *EMBO J* 12(4):1375–1385.