Intricate coupling between the transactivation and basic-leucine zipper domains governs phosphorylation of transcription factor ATF4 by casein kinase 2

Steven Siang, Eric S. Underbakke, and Julien Roche*

From the Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa, USA

Edited by Wolfgang Peti

Most transcription factors possess at least one long intrinsically disordered transactivation domain that binds to a variety of coactivators and corepressors and plays a key role in modulating the transcriptional activity. Despite the crucial importance of these domains, the structural and functional basis of transactivation remains poorly understood. Here, we focused on activating transcription factor 4 (ATF4)/cAMP response element-binding protein-2, an essential transcription factor for cellular stress adaptation. Bioinformatic sequence analysis of the ATF4 transactivation domain sequence revealed that the first 125 amino acids have noticeably less propensity for structural disorder than the rest of the domain. Using solution nuclear magnetic resonance spectroscopy complemented by a range of biophysical methods, we found that the isolated transactivation domain is predominantly yet not fully disordered in solution. We also observed that a short motif at the N-terminus of the transactivation domain has a high helical propensity. Importantly, we found that the N-terminal region of the transactivation domain is involved in transient long-range interactions with the basic-leucine zipper domain involved in DNA binding. Finally, *in vitro* phosphorylation assays with the casein kinase 2 show that the presence of the basic-leucine zipper domain is required for phosphorylation of the transactivation domain. This study uncovers the intricate coupling existing between the transactivation and basic-leucine zipper domains of ATF4, highlighting its potential regulatory significance.

The activating transcription factor 4 (ATF4), also called cAMP response element-binding protein-2 (CREB-2), is a transcription factor that mediates cellular gene expression changes in response to different types of cellular stress (1–3). While constitutively expressed at low concentrations, ATF4 can be rapidly induced under particular cell-stress conditions such as endoplasmic reticulum stress (4), amino acid deprivation (5), and oxidative stress (6). ATF4 expression is also found to be upregulated in response to cancer progression, making it an interesting therapeutic target for blocking angiogenesis and adaptation of cancer cells to hypoxia/anoxia

(7). Activity of ATF4 as a transcriptional repressor includes repression of genes encoding encephalin, chromogranin (8), and the phosphodiesterase PDE4D9 (9). ATF4 is known to form heterodimers with other transcription factors such as ATF3 (10), C/EBP homologous protein (CHOP) (11), and HTLV1 transactivator Tax (12). The numerous dimerization and interaction partners of ATF4 determine its wide range of transcriptional activities.

ATF4 general domain architecture

Human ATF4 is a 351 amino acid long protein organized into two functional domains, an N-terminal transactivation domain (TAD, residues 1–275) and a C-terminal basic-leucine zipper domain (bZip, residues 276–351). The C-terminal bZip domain is composed of a basic region essential for DNA binding (residues 280–301) and leucine zipper region (residues 306–334) essential for homo/heterodimerization (Fig. 1A). The bZip domain forms an extended α-helix encompassing both the basic and leucine-zipper regions in a continuous helical structure (13). Like many other transactivation domains in transcription factors, the TAD is predicted to be structurally disordered but plays a key role in modulating the activation and degradation of ATF4 (14). The N-terminal TAD contains a known binding site for several important binding partners, including p300/CBP-associated factor (PCAF) (1) and βTrCP (14–16). The TAD is also the target of multiple post-translational modifications. The primary phosphorylation sites of protein kinase A (PKA) (17), ribosomal S6 kinase 2 (RSK2) (18), casein kinase 2 (CK2) (19), and RET tyrosine kinase (20), among others, are all known to be localized within the TAD of ATF4.

Previous NMR studies have shown that a region of TAD encompassing the β-TrCP-binding motif (amino acids 208–230) is disordered in the unphosphorylated state but adopts a hairpin loop structure upon phosphorylation of serines 219 and 224 (15). This phosphorylation-induced conformational change appears to be critical for formation of the β-TrCP-ATF4 complex (16). However, very little is known about the structural features of the full TAD sequence, as an isolated domain or in the context of the full-length ATF4. Furthermore, the mechanisms of interdomain communication and potential coupling with the bZip domain remain unclear.

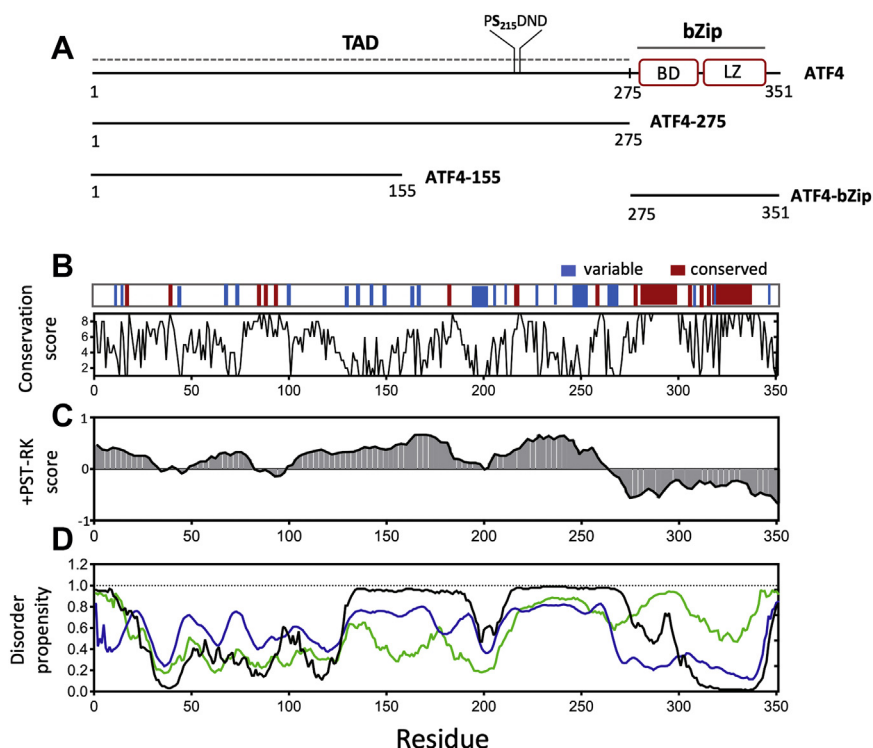* For correspondence: Julien Roche, jroche@iastate.edu.

Figure 1. Sequence analysis of ATF4. *A*, schematic representation of ATF4 sequence showing the N-terminal transactivation domain (TAD, residue 1–275) and C-terminal bZip region encompassing the basic domain (BD, residues 276–301) and leucine zipper domain (LZ, residues 302–351). The primary phosphorylation site of CK2, S215, is also indicated. The three additional constructs used in this study are shown below: ATF4-275 (isolated TAD), ATF4-155 (N-terminal region of the TAD), and ATF4-bZip (isolated bZip domain). *B*, conservation scores established from a Clustal W comparison of human ATF4 with five other species. Highly conserved amino acids (score > 8) and poorly conserved amino acids (score < 2) are shown in *red* and *blue*, respectively, in the *top panel*. *C*, analysis of amino acid composition with IDDomainSpotter shows regions depleted (+PST-RK score < 0) and enriched (+PST-RK score > 0) in disorder promoting amino acids. *D*, disorder propensity predicted with DisEMBL (*blue*), PONDR-FIT (*green*), and DISOPRED (*black*). ATF4, activating transcription factor 4; CK2, casein kinase 2.

In the present study, we used a combination of bioinformatic sequence analysis, circular dichroism (CD), analytical size-exclusion chromatography coupled with multiangle light scattering (SEC-MALS) and small-angle X-ray scattering (SEC-SAXS), and solution nuclear magnetic resonance (NMR) spectroscopy to determine the structural features of TAD. We found that the transactivation domain is predominantly disordered in solution with the exception of a short stretch of amino acids with a higher helical propensity at the N-terminus of the TAD. In addition, our results show that the TAD and bZip domains are not structurally independent but are coupled through long-range interactions involving the N-terminal region of the TAD. Finally, we found that the presence of the bZip is required to achieve complete phosphorylation of the TAD by the casein kinase CK2, suggesting that coupling between the TAD and bZip domains is essential for optimal posttranslational modification, a crucial mechanism controlling ATF4 proteasomal turnover.

Results

ATF4 sequence analysis

We first examined the degree of sequence conservation of ATF4 by comparing the human protein sequence (Uniprot P18848) with those of frog (*Xenopus laevis*, Uniprot Q3V796),

chicken (*Gallus gallus*, Uniprot Q9W610), bovine (*Bos taurus*, Uniprot F1MXZ1), goat (*Capra hircus*, Uniprot A0A452EN02), and mouse (*Mus musculus*, Uniprot Q06507). These sequences were aligned using Clustal Omega (21). Figure 1B shows the conservation score obtained throughout the sequence and highlights the regions that are highly conserved (score >8 in red) and those that are poorly conserved (score <2 in blue). As expected, the bZip domain, which is essential for DNA binding and dimerization of ATF4, is highly conserved among the compared species (Fig. 1B). In contrast, amino acids of the TAD are generally poorly conserved, especially at C-terminus of the transactivation domain. A few notable exceptions include several residues at the N-terminus of the TAD (residues 18, 39, 85, 87, 90, and 95). Analysis of ATF4 amino acid sequence with IDDomainSpotter (22) reveals that the TAD sequence is predominantly enriched in hydrophilic and disorder promoting amino acids (Pro, Ser, and Thr) and depleted in Arg and Lys (represented by positive +PST-RK scores in Fig. 1C). Yet, the amino acid composition of two short regions at the N-terminal of the TAD (residues 34–47 and 82–98), which contain conserved amino acids, noticeably differs from the rest of the TAD. Indeed, these regions have negative +PST-RK scores, which indicate a lower propensity for structural disorder (Fig. 1C). These differences in amino acid composition within

the TAD sequence were confirmed by disorder propensity predictions from DisEMBL (23), PONDR-FIT (24), and DIS-OPRED (25). Altogether, these predictions show a clear difference in propensity for structural disorder between the N-terminal and C-terminal regions of the TAD. As shown in Figure 1D, the first 120 amino acids are indeed predicted to have a significantly lower propensity for structural disorder than amino acids 130 to 275, which are predicted to be almost fully disordered in solution.

Biophysical characterization of isolated ATF4 TAD

Next, we conducted a range of biophysical experiments to characterize the properties of the isolated transactivation domain of ATF4. We recombinantly expressed and purified a truncated construct of ATF4 encompassing the isolated TAD (ATF4-275, Fig. 1A). The CD spectra of ATF4-275 collected at 293 K at a protein concentration of 15 μM, with pH ranging from 6.0 to 8.0, show large negative ellipticity between 200 and 205 nm indicating that the isolated TAD domain is predominantly disordered in solution in all tested conditions (Fig. 2A). As shown in Figure 2B, the elution profiles of ATF4-275 obtained by SEC-MALS were identical at all three sample concentrations tested (3 mg/ml, 6 mg/ml, and 9 mg/ml)

suggesting the absence of oligomerization at higher concentrations. The estimated molecular weight of 32.3 ± 0.3 kDa at 9 mg/mol confirms that the isolated TAD is monomeric in solution (theoretical molecular weight = 29.6 kDa). Additional investigation by SEC-SAXS revealed that ATF4-275 is largely extended in solution, with an Rg of 42.6 ± 4 Å, close to the theoretical dimensions of a fully extended polymeric chain (44 Å) (26) (Fig. S1). We then expressed and purified ATF4-275 in a minimal medium and collected a series of 2D ^{1}H-^{15}N TROSY-HSQC NMR spectra over a wide range of buffer and pH conditions. We observed in all tested conditions the same narrow ^{1}H chemical shift dispersion confirming that ATF4-275 is predominantly disordered in solution (Fig. 2C).

ATF4 TAD possesses a short N-terminal helical motif

Deviations of ^{13}C backbone chemical shifts from idealized random coil chemical shifts constitute a powerful method to precisely evaluate the degree of transient secondary structures in intrinsically disordered proteins (27). Yet, assigning the backbone chemical shifts of intrinsically disordered proteins is a challenging task due to the extent of crosspeak overlap. ATF4-275 presents the additional challenges of being a long domain (275 amino acids) with relatively poor sequence
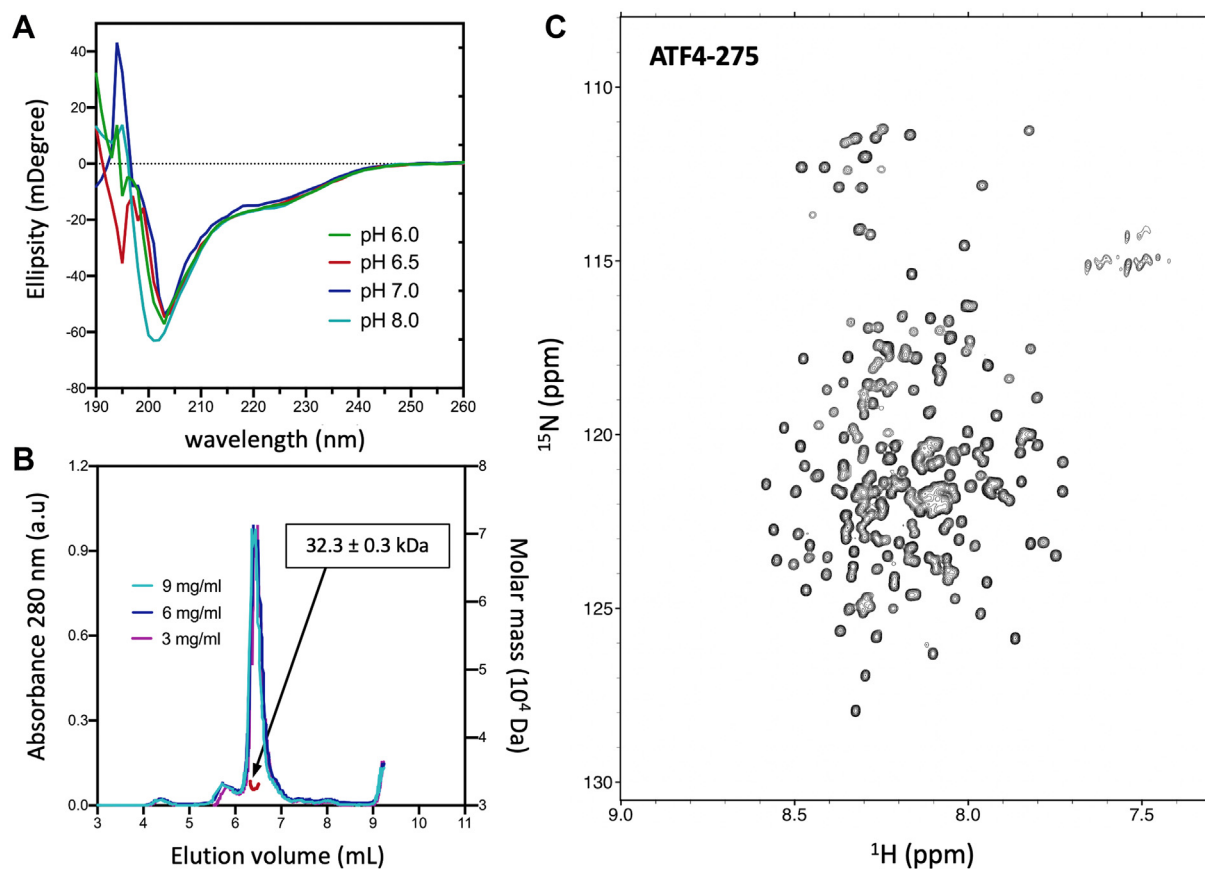

Figure 2. Biophysical characterization of ATF4 isolated transactivation domain. *A*, CD spectra of ATF4-275 collected at 293 K in four different buffer conditions, with pH ranging from 6.0 to 8.0. *B*, SEC-MALS data showing the elution profiles of ATF4-275 collected at three different protein concentrations, from 3 to 9 mg/ml (*left y-axis*). The MALS profile collected at 9 mg/ml is shown in *red* (*right y-axis*). *C*, 2D ^{1}H-^{15}N TROSY-HSQC spectrum of ATF4-275 collected in 20 mM MES pH 6.5, 50 mM NaCl, at 293 K. The spectrum was recorded at a ^{1}H frequency of 700 MHz. ATF4, activating transcription factor 4; CD circular dichroism; MALS, multiangle light scattering; SEC-MALS, size-exclusion chromatography coupled with multiangle light scattering.

complexity (*e.g.*, 32 serines, 32 leucines, 28 prolines). Among all conditions tested, we found that the 2D ^{1}H-^{15}N spectrum collected in 20 mM MES pH 6.5, 50 mM NaCl presented the most homogeneous crosspeak intensities and minimal cross-peak overlap (Fig. 2C). We therefore collected set of conventional 3D triple resonance experiments (HNCA, HNCB, HNCO, HN(CO)CA, and HN(COCA)CB) in this buffer condition. In addition to ATF4-275, we also collected the same set of triple-resonance experiments for a shorter construct encompassing the first 155 amino acids of the TAD (ATF4–155, Fig. 1A). This construct was primarily used as a control to check the consistency of our backbone chemical shifts assignment and to resolve potentially ambiguous sequential correlations. Comparison of 2D ^{1}H-^{15}H spectra collected for ATF4-275 and ATF4-155 shows an almost perfect overlap of the amide crosspeaks, which indicates the absence of long-range contacts between the N- and C-termini of ATF4-275 (Fig. S2). Overall, the set of triple-resonance experiments collected for ATF4-275 and ATF4-155 allowed us to assign the backbone chemical shifts of 254 amino acids (92% of ATF4–275 sequence), with no stretch of unassigned amino acids longer than five consecutive residues (BMRB entry 51062).

In good agreement with the CD data collected in the same conditions, the overall small ^{13}C chemical shift deviations (displayed here as ΔCα- ΔCβ), all within +2 ppm and -1 ppm, confirm that ATF4-275 is predominantly disordered in solution (black in Fig. 3A). Nevertheless, upon close examination, a systematic positive chemical shift deviation, indicative of a residual helical motif, is apparent at the N-terminal of the domain (residues 37–44). Strip plots of the HNCA and HN(CO)CA spectra used to solve the sequential assignment of this stretch of amino acids are shown in Fig. S3. The same positive ^{13}C chemical shift deviations were observed for these amino acids when measured in the shorter construct ATF4-155 (red in Fig. 3A). Beside residues 37 to 44, chemical shift deviations larger than +1 ppm were only found for five residues across TAD sequence. The unique nature of this short helical motif at the N-termini of the TAD is also apparent from temperature coefficients measurement. Here we define these temperature coefficients as ratios between the individual amide crosspeak intensities measured at 293 K and at 283 K (Fig. S4). Because amide protons of intrinsically disordered proteins are largely exposed to solvent, the intensity of amide crosspeaks tends to be very sensitive to changes that would affect the rate of exchange between amide protons and protons from the solvent. Since low temperature limits the rate of exchange with the solvent, amide crosspeaks of disordered proteins are expected to be more intense as the temperature is lowered (temperature coefficients <1). On the other hand, structured regions are expected to be less sensitive to chemical exchange (temperature coefficients ≥1) (28). Figure 3B shows that the temperature coefficients measured for ATF4-275 and ATF4-155 are mostly smaller than 1, confirming again that the TAD is predominantly disordered in solution. A notable exception to this trend is found for residues 38 to 44, which exhibit temperature coefficients larger than 1, suggesting that this short motif has a higher structural propensity than the rest of the TAD (Fig. 3B).

Overall, these results demonstrate that ATF4 TAD is predominantly disordered in solution but not fully devoid of secondary structure. It encompasses at least one stretch of amino acids at its N-terminus with a higher degree of helical propensity.

ATF4 TAD domain forms long-range interactions with the bZip domain

To further study the role played by the transactivation domain in the context of the full-length protein, we compared
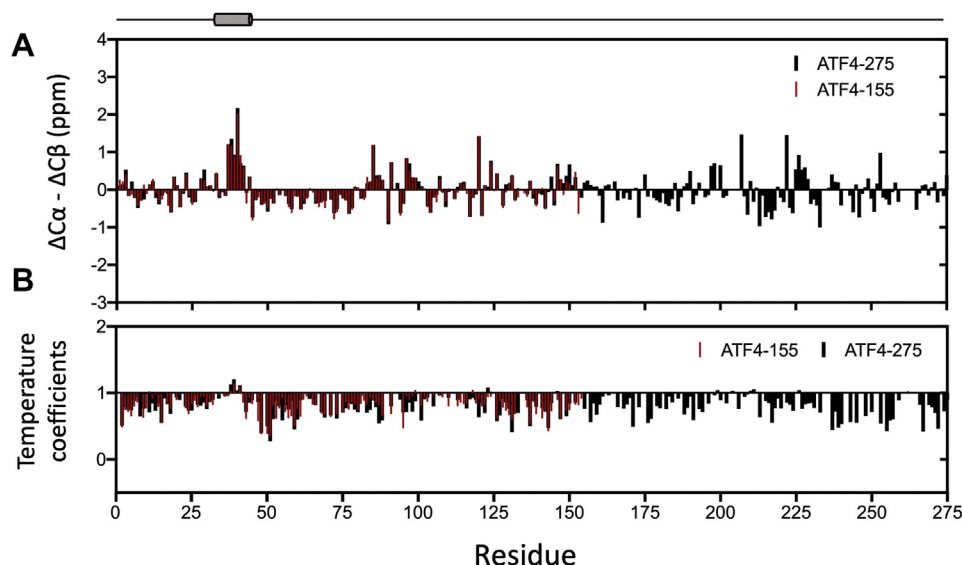

Figure 3. Solution NMR analysis of ATF4-275 and ATF4-155. *A*, ΔCα-ΔCβ chemical shift deviations shown for ATF4-275 (*black*) and ATF4-155 (*red*) (*B*) Temperature coefficients shown for ATF4-275 (*black*) and ATF4-155 (*red*), calculated as ratios of individual amide crosspeak intensities measured at 293 K and at 283 K (=I$_{293K}$/I$_{283K}$). ATF4, activating transcription factor 4; NMR, nuclear magnetic resonance.

the NMR spectra of ATF4-275 with that of the full-length ATF4. Besides small chemical shift changes and differences in peak line widths, we observed no additional amide crosspeak that could be assigned to the bZip domain (Figs. 4*A* and S5). This indicates that residues of the bZip domain experience conformational exchange on an intermediate timescale (relative to NMR timescale) leading to excessive line-width broadening. We probed a broad range of buffer and pH conditions, as well as external perturbations (*i.e.*, addition of TMAO and application of high-hydrostatic pressure, data not shown), but despite our efforts, we were unable to shift the timescale of the exchange and/or the relative population of the states in exchange to sufficiently reduce the line-width broadening of bZip resonances.

Another remarkable observation emerging from this comparison is that a significant number of crosspeaks corresponding to TAD residues are drastically attenuated in the full-length ATF4 spectrum (Fig. 4*A*). When plotted in terms of ratios of individual crosspeak intensities between ATF4 and ATF4-275, it clearly appears that theses residues are mostly localized within the N-terminus of the TAD (*e.g.*, residues 13–21, 32–48, and 80–126) (Fig. 4*B*). To assess whether the line-width broadening of these crosspeaks is due to transient interactions between the two domains, we performed a titration of ^{15}N-labeled ATF4-275 with unlabeled bZip domain (ATF4-bZip, Fig. 1*A*). We observed upon titration a similar pattern of amide crosspeak attenuation corresponding to residues localized at the N-terminal of the transactivation domain (Figs. 4*C* and S6). Taken together, these NMR results indicate that the TAD and bZip domains are not structurally

independent. The excessive line-width broadening appears to result from transient long-range contacts on an intermediate timescale.

ATF4 bZip domain is required for phosphorylation of the TAD

Finally, we investigated whether the structural coupling between the TAD and bZip domains could impact the ability of the TAD to be posttransitionally modified, which is a key mechanism regulating ATF4 function. The TAD is known to be the target of multiple posttranslational modifications, including phosphorylation by the casein kinase CK2, which controls the proteasomal turnover of ATF4 (14, 19). To test if the presence of the bZip domain influences the ability of TAD to be phosphorylated, we designed a set of *in vitro* phosphorylation assays with CK2 and compared the phosphorylation states of ATF4-275 with that of the full-length ATF4 using solution NMR and mass spectrometry (see Experimental procedures).

Comparison of 2D [1]H-^{15}N spectra collected for ATF4 subjected to a 3-h incubation with or without CK2 shows a drastic peak intensity loss for S215 and neighboring residues (red bar in Fig. 5*A*, and red spectrum in Fig. 5*B*), indicating that this serine is almost fully phosphorylated in such conditions by CK2. We also noticed that a second serine (S184) appears to be partially phosphorylated upon incubation with CK2 (Fig. 5*A*). Two new amide crosspeaks are indeed observed in the 2D [1]H-^{15}N spectrum of ATF4 after incubation with CK2: one intense resonance at 8.89 ppm (attributed to the phosphorylated form of S215) and a significantly less intense
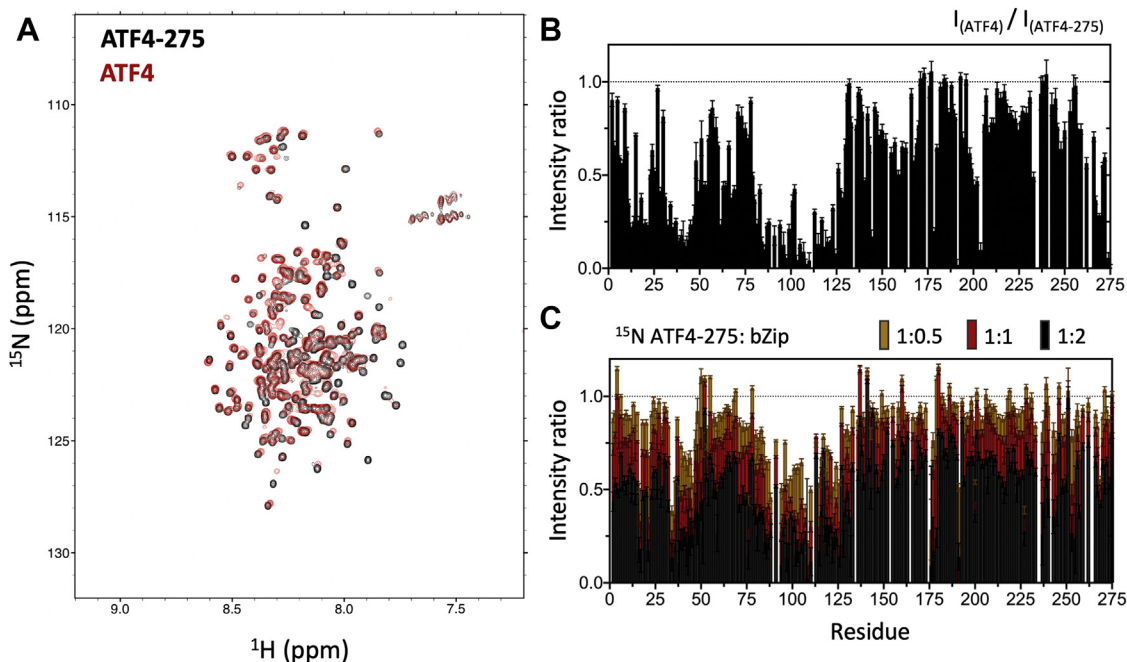

Figure 4. Comparison of ATF4-275 with the full-length ATF4. *A*, overlay of 2D ^{1}H-^{15}N TROSY-HSQC spectra collected for ATF4 (*black*) and ATF4-275 (*red*). The spectra were recorded at a ^{1}H frequency of 700 MHz. *B*, ratios of individual amide crosspeak intensities measured between ATF4 and ATF4-275 (=$I_{ATF4}/I_{ATF4-275}$) (*blue*), and ATF4 with 4 M equivalents of double stranded DNA (*cyan*). *C*, ratios of individual amide crosspeak intensities measured for 2D ^{1}H-^{15}N TROSY-HSQC spectra of ^{15}N-ATF4-275 collected at increasing concentrations of unlabeled ATF4-bZip. ATF4, activating transcription factor 4.
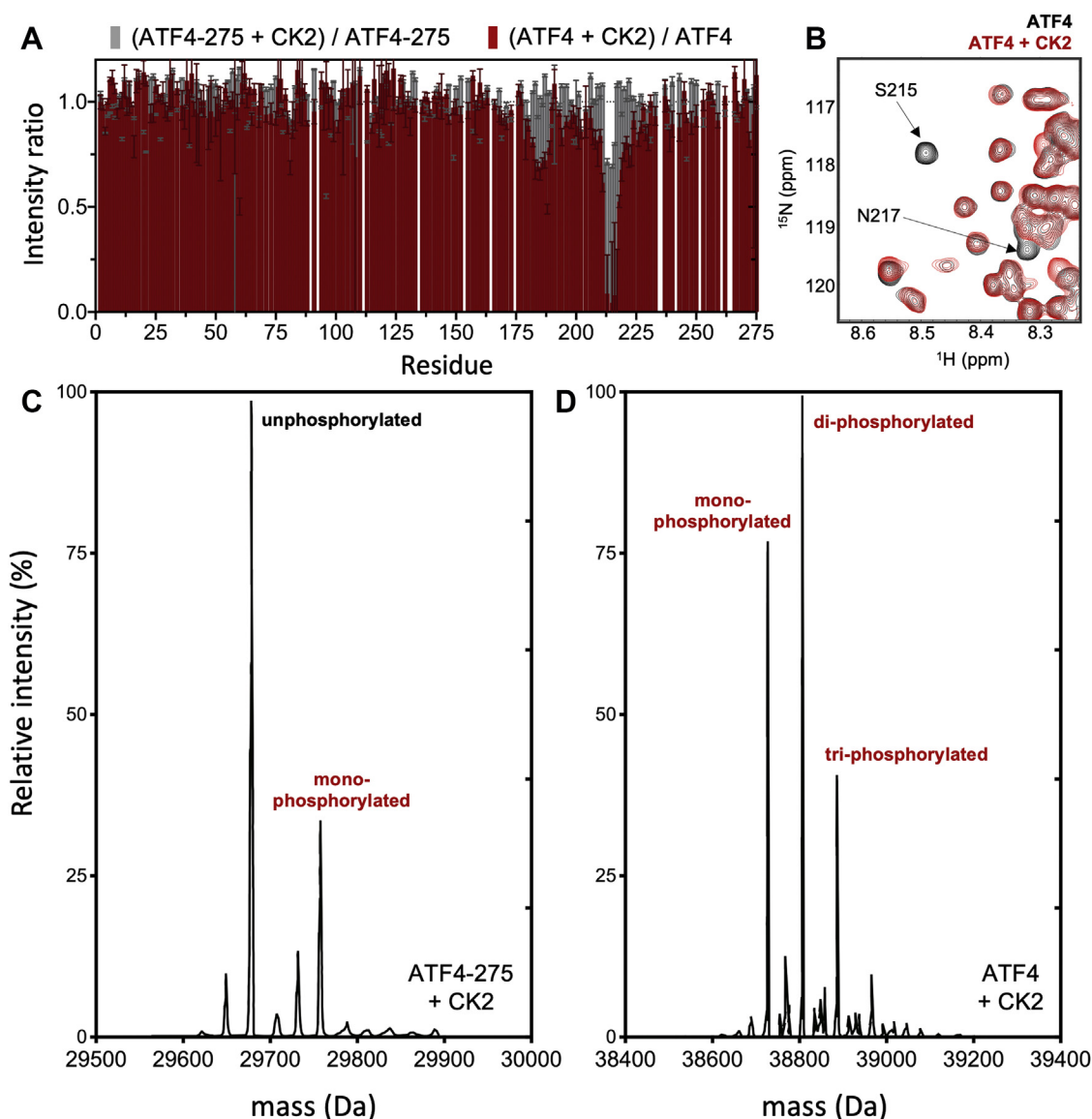
Figure 5. Comparison of ATF4 and ATF4-275 phosphorylation efficiency by CK2. *A*, ratios calculated as individual amide crosspeak intensities measured between samples incubated with and without the protein kinase CK2 ($=I_{(+)CK2}/I_{(−)CK2}$) for ATF4 (*red*) and ATF4-275 (*gray*). *B*, close-up of 2D ^{1}H-^{15}N TROSY-HSQC spectra of ATF4 non-incubated with CK2 (*black*) and ATF4 incubated with CK2 (*red*). This region of the spectrum shows the amide crosspeaks of S215 and neighboring residue N217. The spectra were recorded at a ^{1}H frequency of 700 MHz. *C*, deconvolution spectrum of intact ATF4-275 after 3 h of incubation with CK2 analyzed by high-performance liquid chromatography coupled to electrospray ionization mass spectrometry (LC-ESI-MS). *D*, deconvolution spectrum of intact ATF4 after 3 h of incubation with CK2. ATF4, activating transcription factor 4; CK2, casein kinase 2.

resonance 8.73 ppm (attributed to the phosphorylated form of S184) (Fig. S7). Interestingly, only minor losses of peak intensity are observed when ATF4-275 is incubated in the same conditions with CK2 (gray bar in Fig. 5*A*), suggesting that phosphorylation by CK2 is significantly less efficient when incubated with the isolated TAD. These results were confirmed by intact protein mass spectrometry showing only partial phosphorylation when CK2 is incubated with ATF4-275 (Fig. 5*C*), while complete phosphorylation is observed with the full-length ATF4, resulting in a distribution of mono, di-, and tri-phosphorylated species (Fig. 5*D* and Table S1). The specific phosphorylation of S215 was confirmed by liquid chromatography with tandem mass spectrometry (LC-MS/MS) (Table S2). Overall, these results suggest that the presence

of the bZip domain is essential for *in vitro* phosphorylation of TAD by the kinase CK2.

Discussion

Nearly 90% of eukaryotic transcription factors contain large transactivation domains that are intrinsically disordered (29). These TADs can form long continuous domains up to several hundred residues or be dispersed through different regions of the protein, as in the case of p53 (30). As regulatory hubs for transcriptional coactivators and repressors, TADs play an essential role in modulating the activity of transcription factors (31, 32). TADs are also usually heavily posttranscriptionally modified, which adds an extra layer of

complexity and modularity to their ability to control transcriptional activity (33, 34). Being predominantly disordered in solution, TADs are believed to exert their functions through relativity weak, highly dynamics, and short-lived interactions (35). While critically important, the structural and mechanistic basis of TAD functions is still poorly understood, in part due to the major challenges inherent to characterizing disordered domains. Biophysical approaches able to identify and characterize structured motifs at a residue level within a predominantly disordered conformational ensemble are essential to fully understand the mechanisms of regulation of transcription factors.

In the present study, we used a range of biophysical techniques to characterize the residual structural propensity of ATF4 TAD, a transcription factor that plays a key role in mediating the expression of a variety of genes in response to cellular stresses. Analysis of ATF4 sequence highlights the differences between the N-and C-terminal regions of the TAD. The first 125 amino acids of the TAD appear to have characteristics closer to that of a folded-like region rather than of a typical intrinsically disordered region. Indeed, we found that residues 1 to 125 are overall more conserved, less enriched in disorder-promoting amino acids, and have a relatively low propensity for structural disorder compared with residues 126 to 275 (Fig. 1, *B–D*). Differences between the N- and C-terminal regions of ATF4 TAD have been previously proposed by Liang and Hai (1), who identified residues 91 to 124 as a potential additional leucine zipper motif.

Here, a combination of CD, SEC-MALS, SEC-SAXS, and NMR experiments unambiguously show that the isolated TAD is predominantly disordered in solution (Fig. 2), suggesting that the more foldable characteristics of the first 125 amino acids are not sufficient to significantly shift the conformational ensemble of the TAD toward a folded-like ensemble with stable secondary structures. Yet, backbone chemical shift analysis, complemented by temperature coefficients, revealed that the N-terminal region is not fully disordered in solution but contains at least one short motif (residues 37–44) with high helical propensity (Fig. 3). Importantly, we found that most of the N-terminal region (*i.e.*, residues 13–21, 32–48, and 80–126) appears to be involved in transient long-range interactions with the bZip domain on an intermediate timescale (Fig. 4). While this type of coupling between the TAD and bZip domains have, to the best of our knowledge, not been identified before for ATF4, it is strongly reminiscent of other transcription factors such as p53 (36) and FOXO4 (37). In both cases, regions of the transactivation domains are known to interact directly with the DNA-binding domain (36, 37). Intrinsically disordered domains can mediate interactions through a wide range of mechanisms, from induced fit (38) to conformational selection (39), and various mixed mechanisms (40, 41). In the case of ATF4, the higher helical propensity of residues at the N-terminus of TAD suggests that interaction between the TAD and bZip domains takes place through conformational selection. However, precise kinetic information will be required to unequivocally determine the exact binding mechanism.

Finally, using a combination of NMR and mass spectrometry experiments, we found that the presence of the bZip domain is required for full phosphorylation of TAD by the kinase CK2. Among all 12 putative sites in ATF4 with the consensus sequence S/TXXD/E, S215 has been shown previously to be primary phosphorylation site of CK2, both *in vitro* and *in vivo* (19). Here, we found that incubation with CK2 led to complete phosphorylation of S215 in the full-length ATF4 but not in the isolated TAD (Fig. 5). This observation can be explained in several ways: (i) access of the phosphorylation motif to CK2 active site is facilitated by the disorder-mediated coupling between the two domains, (ii) CK2 binds to bZip, triggering a conformational change that brings the phosphorylation motif within reach, or (iii) CK2 optimal binding site covers regions of both TAD and bZip domains. Characterization of single point variants of the bZip domain is currently ongoing in our group in an effort to determine the precise mechanism of phosphorylation of ATF4 by CK2.

While it is well documented that TADs can modulate the accessibility of DNA-binding domains to limit or enhance DNA binding (36, 37, 42), our results suggest that this type of fine regulation may be reciprocal, as in the case of ATF4, the bZip domain modulating the accessibility of the TAD to a protein kinase. Altogether, these findings shed light on the complexity and intricate nature of these disorder-driven activation mechanisms that may be shared by many eukaryotic transcription factors.

Experimental procedures

Sequence analysis

Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) was used with default parameters. +PST-RK scores were calculated using IDDomainSpotter (http://www2.bio.ku.dk/IDR/) with 20 amino acids window. Disorder propensity calculated with DisEMBL (http://dis.embl.de/) is reported here for the "coils" values with coils threshold set to 1.20. The predictor VL3 was used with *via* PONDR-FIT (http://original.disprot.org/pondr-fit.php). The algorithm Disopred3 was used *via* PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/).

Sample preparation

cDNA fragments encoding ATF4, ATF4-155, ATF4-275, and ATF4-bZip sequences were amplified by PCR and inserted into pGEX-4T1 plasmid using EcoRI and XhoI restriction sites. All constructs contained an N-terminal glutathione S-transferase (GST) followed by TEV protease cleavage site. Expression was performed in *E. coli* BL21(DE3) cells. Cells were grown at 37 °C overnight. The next day, 10 ml of overnight culture was transferred to 1 L of fresh medium and grown at 37 °C until OD_{600nm}= 0.8. Protein overexpression was induced by adding IPTG to a final concentration of 400 μM and shaking at 16 °C overnight. Cells were harvested by centrifugation at 6000 rpm. Cells overexpressing ATF4, ATF4-275, and ATF4-155 were resuspended in Lysing Buffer A (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM DTT, 1 mM

EDTA). Cells overexpressing ATF4-bZip were resuspended in Lysing Buffer B (50 mM CHES 9.0, 100 mM NaCl, 5 mM DTT, 1 mM EDTA). Both lysing buffers were supplemented with PMSF, Leupeptin, DNase I, and Protease Inhibitor Cocktails (Alpha Aesar). Cells were lysed by sonicating at 4 °C. The insoluble fractions were removed by centrifugation at 16,000 rpm for 1 h at 4 °C. GST-tagged proteins were purified by passing the supernatant through Glutathione-Sepharose beads (UBPBio). Unbound proteins were removed by washing with lysing buffer, and GST-tagged proteins were eluted with lysing buffers containing 5 mM reduced gluta-thione. Additionally, ATF4 and ATF4-bZip were washed with 50 mM Tris-HCl 8.0, 500 mM NaCl, 5 mM DTT, 1 mM EDTA to remove nonspecific bound nucleic acids. After incubating with 0.01 mg of TEV protease per mg of GST-tagged ATF4, ATF4-155, and ATF4-275 for 1 h, the cleaved GST-tags were removed by running through two subtractive GST affinity columns. The GST-tagged ATF4-bZip was incubated with 0.01 mg of TEV protease overnight (∼16 h), and the cleaved GST-tag and ATF4-bZip were then separated by running through a HiLoad 16/600 SuperdexTM 200 pg gel filtration column in SEC Buffer (20 mM MES 6.5, 50 mM NaCl). The samples were maintained at 4 °C throughout the purification process.

Circular dichroism

Circular dichroism (CD) spectra were recorded using an MOS-500 (Biologic) spectropolarimeter controlled with Bio-Kine32 at room temperature in a 1 mm UV quartz cuvette (FireflySci). Samples were prepared at 15 μM in the following buffers: (i) 20 mM MES pH 6.0, 50 mM NaCl, (ii) 20 mM MES pH 6.5, 50 mM NaCl, (iii) 20 mM HEPES pH 7.0, 50 mM NaCl, and (iv) 20 mM Tris-HCl pH 8.0, 50 mM NaCl. The mean spectra were derived from three acquisitions collected between 190 −and 260 nm for 0.2 s with 2 nm slit width. Spectra were background-corrected with buffers.

SEC-MALS and SEC-SAXS

Multiangle light scattering (MALS) and small-angle X-ray scattering (SAXS) coupled in-line with size-exclusion chromatography (SEC) experiments were performed at SIBYLS ALS beamline 12.3.1 LBNL. SEC-MALS and SEC-SAXS experiments were performed with 70 μl samples of ATF4-275 prepared at 3 mg/ml, 6 mg/ml, and 9 mg/ml (∼100 μM, 200 μM, 300 μM protein concentration). The X-ray wavelength was set at λ = 1.03 Å incident light at a sample to detector distances of 1.5 m resulting in scattering vectors (q) ranging from 0.013 to 0.5 Å^{-1}. The scattering vector is defined as q = $4\pi\sin\theta/\lambda$, where 2θ is the scattering angle. All experiments were performed at 20 °C, and the data were processed as described previously (43). Briefly, an SAXS flow cell was directly coupled with an online 1260 Infinity HPLC system (Agilent) using a Shodex 803 analytical column (Showa Denko) for SEC separation. The column was equilibrated with a running buffer (20 mM MES pH 6.5, 50 mM NaCl) with a

flow rate of 0.5 ml/min. Sixty-five microliters of each sample was run through the SEC, and 3-s X-ray exposures were collected continuously during a 40-min elution. The SAXS frames recorded prior to the protein elution peak were used to subtract all other frames. Final merged SAXS profiles, derived by integrating multiple frames across the elution peak, were used for further analysis using PRIMUS (44) from the ATSAS suite of programs (45).

Eluent was subsequently split into 2:1 between the SAXS line and a series of UV at 280 and 260 nm detectors, quasi-elastic light scattering (QELS) detectors, and MALS detectors. MALS experiments were performed using a Wyatt Dawn Heleos MALS detector connected in tandem to an Optilab rEX refractive index concentration detector (Wyatt Technology). The light-scattering experiments were used to perform analytical scale chromatographic separations for molecular weight determination of the principal peaks in the SEC analysis.

Solution NMR spectroscopy

NMR spectra were collected on a Bruker 700 MHz and 800 MHz spectrometers, equipped with z-shielded gradient triple resonance 5 mm TCI cryoprobe. NMR samples were prepared at protein concentrations of 180 to 220 μM in 20 mM MES pH 6.5, 50 mM NaCl, and 90% H$_2$O/10% D$_2$O (v/v). 2D ^{1}H-^{15}N TROSY-HSQC experiments were recorded with a time domain matrix consisting of 100* (t$_1$, ^{15}N) × 1024* (t$_2$, ^{1}H) complex points with acquisition time of 50 ms (t$_1$) and 91.8 ms (t$_2$) using eight scans per FID and 1.5 s interscan delay. Spectral widths for 1H and 15N dimensions were set to 15.9 and 28.2 ppm, respectively, with carriers set at 4.821 ppm (^{1}H) and 119.138 ppm (^{15}N). Sequential ^{1}H/^{15}N/^{13}C backbone assignments of ATF4-275 and ATF4-155 were achieved using TROSY versions of conventional 3D triple resonance correlation experiments [HNCO, HNCA, HNCACB, HN(CO)CA, and HN(COCA)CB] (46) at 700 MHz. Secondary ^{13}C chemical shifts were calculated using Poulsen's random coil database with corrections for pH, temperature, and neighbor amino acid correction (47, 48). All spectra were processed using NMRPipe (49) and displayed with SPARKY (50).

In vitro phosphorylation

ATF4 and ATF4-275 constructs were phosphorylated in kinase buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 10 mM MgCl$_2$, and 1 mM DTT). The phosphorylation reactions were started by adding 2.5 mM ATP and 25,000 units of CK2 (two alpha and two beta subunits, NEB) per mg of protein sample. The phosphorylation reactions were performed for 3 h at 25 °C. The reactions were quenched by buffer exchange (*i.e.*, three rounds of centrifugal concentration (3K MWCO) with intervening buffer dilutions) to remove ATP and MgCl$_2$.

Intact protein mass spectrometry

Liquid chromatography–mass spectrometry (LC-MS) on intact protein samples was performed using an ACQUITY

UPLC H-Class coupled in line to a ESI-Q-TOF Synapt G2-Si (Waters) using a Restek Ultra C4 5 μm (#9102551) reverse-phase column. The column was operated at 25 °C with the flow rate of 0.4 ml/min. Chromatography was performed with a linear gradient between mobile phases A and B, with Buffer A consisting of aqueous 0.1% formic acid and Buffer B consisting of 100% acetonitrile, 0.1% formic acid. The elution was carried out using a linear gradient from 5% to 100% B at a flow rate of 0.4 ml/min for 7 min. Protein samples were injected at a concentration of 0.1 mg/ml. Mass spectrometry was performed using a LockSpray ESI (Waters) ionization source with [Glu1]-fibrinopeptide B as internal lock mass calibrant. The instrument was operated using MassLynx 4.1 (Waters). MS data were collected in positive ion, MS continuum, resolution mode with an m/z range of 300 to 5000. The intact protein molecular weight of each sample was deconvoluted from multiple charge states using the MaxEnt3 function of MassLynx 4.2 (Waters).

Liquid chromatography–tandem mass spectrometry

Phosphorylated samples were alkylated with 15 mM of iodoacetamide at 25 °C for 30 min before digesting with chymotrypsin at 25 °C for 20 h followed by trypsin at 37 °C for 10 h. Digested peptides were injected into EASY nLC-1200 coupled to a Nanospray FlexIon source for in-line analysis with a Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer equipped with an HCD fragmentation cell (Thermo Scientific). Peptides were searched with a Mascot database containing the GST-tagged ATF4 sequence. Peptides were identified with Proteome Discoverer Software 2.4 using the following search parameters: false discovery rate (FDR) 1%, precursor mass tolerance 20 ppm, and fragment mass tolerance ± 0.02 Da. Oxidation (M), deamidation (NQ), and phosphorylation (ST) were searched as dynamic modifications, and carbamidomethylation (C) was set as a static modification.

Data availability

The backbone chemical resonance assignment of ATF4-275 has been deposited at the Biological Magnetic Resonance Data Bank (BMRB) under accession number 51062. The mass spectrometry raw data and results are deposited to PRIDE Submission under accession number PXD029704.

Abbreviations—The abbreviations used are: ATF4, activating transcription factor 4; bZip, basic-leucine zipper domain; CD, circular dichroism; CREB-2, cAMP response element-binding protein-2; HSQC, heteronuclear single quantum correlation; LC-ESI-MS, high-performance liquid chromatography coupled to electrospray ionization mass spectrometry; LC-MS-MS, liquid chromatography with tandem mass spectrometry; MALS, multiangle light scattering; NMR, nuclear magnetic resonance; SAXS, small-angle X-ray scattering; SEC, size-exclusion chromatography; TAD, transactivation domain; TROSY, transverse relaxation optimized spectroscopy.

References

1. Liang, G., and Hai, T. (1997) Characterization of human activating transcription factor 4, a transcriptional activator that interacts with multiple domains of cAMP-responsive element-binding protein (CREB)-binding protein (CBP). *J. Biol. Chem.* **272**, 24088–24095
2. Ameri, K., and Harris, A. L. (2008) Activating transcription factor 4. *Int. J. Biochem. Cell Biol.* **40**, 14–21
3. Pakos-Zebrucka, K., Koryga, I., Mnich, K., Ljujic, M., Samali, A., and Gorman, A. M. (2016) The integrated stress response. *EMBO Rep.* **17**, 1374–1395
4. Armstrong, J. L., Flockhart, R., Veal, G. J., Lovat, P. E., and Redfern, C. P. (2010) Regulation of Endoplasmic reticulum stress-induced cell death by ATF4 in neuroectodermal tumor cells. *J. Biol. Chem.* **285**, 6091–6100
5. Kilberg, M. S., Shan, J., and Su, N. (2009) ATF4-dependent transcription mediates signaling of amino acid limitation. *Trends Endocrinol. Metab.* **20**, 436–443
6. Roybal, C. N., Hunsaker, L. A., Barbash, O., Vander Jagt, D. L., and Abcouwer, S. F. (2005) The oxidative stressor arsenite activates vascular endothelial growth factor mRNA transcription by an ATF4-dependent mechanism. *J. Biol. Chem.* **280**, 20331–20339
7. Singleton, D. C., and Harris, A. L. (2012) Targeting the ATF4 pathway in cancer therapy. *Expert Opin. Ther. Targets* **16**, 1189–1202
8. Schoch, S., Cibelli, G., Magin, A., Steinmüller, L., and Thiel, G. (2001) Modular structure of cAMP response element binding protein 2 (CREB2). *Neurochem. Int.* **38**, 601–608
9. Soda, T., Frank, C., Ishizuka, K., Baccarella, A., Park, Y.-U., Flood, Z., Park, S. K., Sawa, A., and Tsai, L.-H. (2013) DISC1-ATF4 transcriptional repression complex: Dual regulation of the cAMP-PDE4 cascade by DISC. *Mol. Psychiatry* **18**, 898–908
10. Wang, Q., Mora-Jensen, H., Weniger, M. A., Perz-Galan, P., Wolford, C., Hai, T., Ron, D., Chen, W., Trenkle, W., Wiestner, A., and Ye, Y. (2009) ERAD inhibitors integrate ER stress with an epigenetic mechanism to activate BH3-only protein NOXA in cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2200–2205
11. B'chir, W., Maurin, A.-C., Carraro, V., Vaerous, J., Jousse, C., Muranishi, Y., Parry, L., Stepien, G., Fafournoux, P., and Bruhat, A. (2013) The eIF2α/ATF4 pathway is essential for stress-induced autophagy gene expression. *Nucleic Acids Res.* **41**, 7683–7699
12. Reddy, T. R., Tang, H., Li, X., and Wong-Staal, F. (1997) Functional interaction of the HTLV-1 transactivator tax with activating transcription factor-4 (ATF4). *Oncogene* **14**, 2785–2792

13. Podust, L. M., Krezel, A. M., and Kim, Y. (2001) Crystal structure of the CCAAT box/enhancer-binding protein beta activating transcription factor-4 basic leucine zipper heterodimer in the absence of DNA. *J. Biol. Chem.* **276**, 505–513

14. Lassot, I., Ségéral, E., Berlioz-Torrent, X., Durand, H., Goussin, L., Hai, T., Benarous, R., and Margottin-Goguet, F. (2001) ATF4 degradation relies on a phosphorylation-dependent interaction with the SCF(betaTrCP) ubiquitin ligase. *Mol. Cell. Biol.* **21**, 2192–2202

15. Evrard-Todeshi, N., Bertho, G., Gharbi-Benarous, J., Benarous, R., and Girault, J.-P. (2007) Phosphorylation-dependent structure of ATF4 peptides derived from a human ATF4 protein, a member of the family of transcription factors. *Peptides* **28**, 2253–2267

16. Pons, J., Evrard-Todeshi, N., Bertho, G., Gharbi-Benarous, J., Tanchou, V., Benarous, R., and Girault, J.-P. (2008) Transfer-NMR and docking studies identify the binding of the peptide derived from activating transcription factor 4 to protein ubiquitin ligase beta-TrCP. Competition STD-NMR with beta-catenin. *Biochemistry* **47**, 14–29

17. Elefteriou, F., Ahn, F. J. D., Takeda, S., Starbuck, M., Yang, Z., Liu, Z., Kondo, H., Richards, W. G., Bannon, T. W., Noda, M., Clement, K., Vaisse, C., and Karsenty, G. (2005) Leptin regulation of bone resorption by the sympathetic nervous system and CART. *Nature* **434**, 514–520

18. Yang, X., Matsuda, K., Bialek, P., Jacquot, S., Masuoka, H. C., Schinke, T., Li, L., Brancorsini, S., Sassone-Corsi, P., Townes, T. M., Hanauer, A., and Karsenty, G. (2004) ATF4 is a substrate of RSK2 and an essential regulator of osteoblast biology; implication for Coffin-Lowry syndrome. *Cell* **117**, 387–398

19. Ampofo, E., Sokolowski, T., Götz, C., and Montenarh, M. (2013) Functional interaction of protein kinase CK2 and activating transcription factor 4 (ATF4), a key player in the cellular stress response. *Biochim. Biophys. Acta* **1833**, 439–451

20. Bagheri-Yarmand, R., Sinha, K. M., Gururaj, A. E., Ahmed, Z., Rizvi, Y. Q., Huang, S.-C., Ladbury, J. E., Bogler, O., Williams, M. D., Cote, G. J., and Gagel, R. F. (2015) A novel dual kinase function of the RET proto-oncogene negatively regulates activating transcription factor 4-mediated apoptosis. *J. Biol. Chem.* **290**, 11749–11761

21. Sievers, F., and Higgins, D. G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145

22. Millard, P. S., Bugge, K., Marabini, R., Boomsma, W., Burow, M., and Kragelund, B. B. (2020) IDDomainSpotter: Compositional bias reveals domains in long disordered protein regions-Insights from transcription factors. *Protein Sci.* **29**, 169–183

23. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., and Russell, R. B. (2003) Protein disorder prediction: Implications for structural proteomics. *Structure* **11**, 1453–1459

24. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., and Uversky, V. N. (2010) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* **1804**, 996–1010

25. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138–2139

26. Nygaard, M., Kragelund, B. B., Papaleo, E., and Lindorff-Larsen, K. (2017) An efficient method for estimating the hydrodynamic radius of disordered protein conformations. *Biophys. J.* **113**, 550–557

27. Kosol, S., Contreras-Martos, S., Cedeño, C., and Tompa, P. (2013) Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **18**, 10802–10828

28. Piai, A., Calçada, E. O., Tarenzi, T., del Grande, A., Varadi, M., Tompa, P., Felli, I. C., and Pierattelli, R. (2016) Just a flexible linker? The structural and dynamic properties of CBP-ID4 revealed by NMR spectroscopy. *Biophys. J.* **110**, 372–381

29. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873–6888

30. Raj, N., and Attardi, L. D. (2017) The transactivation domains of the p53 protein. *Cold Spring Harb. Perspect. Med.* **7**, a026047

31. Shammas, S. L. (2017) Mechanistic roles of protein disorder within transcription. *Curr. Opin. Struct. Biol.* **42**, 155–161

32. Clark, S., Myers, J. B., King, A., Fiala, R., Novacek, J., Pearce, G., Heierhorst, J., Reichow, S. L., and Barbar, E. J. (2018) Multivalency regulates activity in an intrinsically disordered transcription factor. *Elife* **7**, e36258

33. Conforti, F., Sayan, A. E., Sreekumar, R., and Sayan, B. S. (2012) Regulation of p73 activity by post-translational modifications. *Cell Death Dis.* **3**, e285

34. Ferraris, J. D., Williams, C. K., Persaud, P., Zhang, Z., Chen, Y., and Burg, M. B. (2002) Activity of the TonEBP/OREBP transactivation domain varies directly with extracellular NaCl concentration. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 739–744

35. Berlow, R. B., Dyson, H. J., and Wright, P. E. (2015) Functional advantages of dynamic protein disorder. *FEBS Lett.* **589**, 2433–2440

36. Krois, A. S., Dyson, H. J., and Wright, P. E. (2018) Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E11302–E11310

37. Kim, J., Ahn, D., and Park, C.-J. (2021) FOXO4 transactivation domain interaction with forkhead DNA binding domain and effect on selective DNA recognition for transcription initiation. *J. Mol. Biol.* **433**, 166808

38. Johnson, K. A. (2008) Role of induced fit in enzyme specificity: A molecular forward/reverse switch. *J. Biol. Chem.* **283**, 26297–26301

39. Kjaergaard, M., Teilum, K., and Poulsen, F. M. (2010) Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12535–12540

40. Arai, M., Sugase, K., Dyson, H. J., and Wright, P. E. (2015) Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9614–9619

41. Mollica, L., Bessa, M. L., Hanoulle, X., Jensen, M. R., Blackledge, M., and Schneider, R. (2016) Binding mechanisms of intrinsically disordered proteins: Theory, simulation, and experiments. *Front. Mol. Biosci.* **3**, 52

42. Desjardins, G., Meeker, C. A., Bhachech, N., Currie, S. L., Okon, M., Graves, B. J., and McIntosh, L. P. (2014) Synergy of aromatic residues and phosphoserines within the intrinsically disordered DNA-binding inhibitory elements of the Ets-1 transcription factor. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11019–11024

43. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., 2nd, Tsutakawa, S. E., Jenney, F. E., Jr., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S.-J., Scott, J. W., Dillard, B. D., Adams, M. W. W., and Tainer, J. A. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* **6**, 606–612

44. Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I., and Franke, D. (2021) Atsas 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **54**, 343–355

45. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) Primus - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36**, 1277–1282

46. Clore, G. M., and Gronenborn, A. M. (1998) Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol.* **16**, 22–34

47. Kjaergaard, M., and Poulsen, F. M. (2011) Sequence correction of random coil chemical shifts: Correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR* **50**, 157–165

48. Kjaergaard, M., Brander, S., and Poulsen, F. M. (2011) Random coil chemical shifts for intrinsically disordered proteins: Effects of temperature and pH. *J. Biomol. NMR* **49**, 139–149

49. Delaglio, F., Grzesiel, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **5**, 277–293

50. Goddard, T. D., and Kneller, D. G. (2010) *Sparky 3*, University of California San Francisco, San Francisco, CA

ASBMB