Transcriptional Activator−Coactivator Recognition: Nascent Folding of a Kinase-Inducible Transactivation Domain Predicts Its Structure on Coactivator Binding†

Qing-Xin Hua,*,‡ Wen-hua Jia,‡ Bryant P. Bullock,§,|| Joel F. Habener,§ and Michael A. Weiss*,‡

Departments of Biochemistry & Molecular Biology and Chemistry, Center for Molecular Oncology, The University of Chicago, Chicago, Illinois 60637, and Howard Hughes Medical Institute, Laboratory of Molecular Endocrinology, Massachusetts General Hospital, Harvard Medical School, Fruit Street, Boston, Massachusetts 02114

Received January 12, 1998; Revised Manuscript Received February 23, 1998

ABSTRACT: A model of transcriptional activator−coactivator recognition is provided by the mammalian CREB activation domain and the KIX domain of coactivator CBP. The CREB kinase-inducible activation domain (pKID, 60 residues) is disordered in solution and undergoes an α-helical folding transition on binding to CBP [Radhakrishan, I., Perez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., and Wright, P. E. (1997) *Cell 91*, 741−752]. Binding requires phosphorylation of a conserved serine (RPpSYR) in pKID associated in vivo with the biological activation of CREB signaling pathways. The CBP-bound structure of CREB contains two α-helices (designated αA and αB) flanking the phosphoserine; the bound structure is stabilized by specific interactions with CBP. Here, the nascent structure of an unbound pKID domain is characterized by multidimensional NMR spectroscopy. The solubility of the phosphopeptide (46 residues) was enhanced by truncation of N- and C-terminal residues not involved in pKID−CBP interactions. Although disordered under physiologic conditions, the pKID fragment and its unphosphorylated parent peptide exhibit partial folding at low temperatures. One recognition helix (αA) is well-defined at 4 °C, whereas the other (αB) is disordered but inducible in 40% trifluoroethanol (TFE). Such nascent structure is independent of serine phosphorylation and correlates with the relative extent of engagement of the two α-helices in the pKID−KIX complex; whereas αA occupies a peripheral binding site with few intermolecular contacts, the TFE-inducible αB motif is deeply engaged in a hydrophobic groove. Our results support the use of TFE as an empirical probe of hidden structural propensities and define a correspondence between induced fit and the nascent structure of peptide fragments.

Regulation of eukaryotic gene expression is mediated in part through the assembly of promoter-specific preinitiation complexes (*1*). Such assembly requires not only protein−DNA recognition (*2*) but also a network of specific protein−protein interactions (*3*). Of particular interest are interactions between transcriptional activation domains and coactivators (*4*, *5*). A well-characterized example is provided by the mammalian CREB protein, a member of the bZIP[1] family of transcription factors (*6*−*10*) and a component of conserved signal-transduction pathways that can be induced by protein kinase A (*11*, *12*). Such signaling requires binding of CREB to a transcriptional coactivator, CREB-binding protein (CBP; *13*−*15*). Protein assembly (unlike DNA recognition) requires phosphorylation of a particular serine in the CREB kinase-inducible transactivation domain (designated pKID or P-Box; *8*, *11*, *13*, *16*). Although direct effects of phosphorylation on protein−DNA recognition are controversial (*17*−*22*), correlations are well-established between the strength of a biological signal, downstream CREB

† This work was supported in part by grants from the American Cancer Society of Illinois (Q.-X.H.) and National Institutes of Health (DK25532 to J.F.H. and HD33462 to M.A.W.). J.F.H. is an Investigator of the Howard Hughes Medical Institute. M.A.W. is an Established Investigator of the American Heart Association and a Bane and Lucille Markey Scholar at The University of Chicago.

* Address correspondence to either author at The University of Chicago.

‡ The University of Chicago.

§ Harvard Medical School.

|| Present address: Ciphergen Biosystems, Inc., 490 San Antonio Rd., Palo Alto, CA 94306.

[1] Abbreviations: 1D, one-dimensional; 2D, two-dimensional; 3D, three-dimensional; bZIP, basic region-leucine zipper motif; AMP, adenosine monophosphate; CBP, CREB-binding protein; CREB, cAMP-response-element binding protein; DG, distance geometry; DNA, deoxyribonucleic acid; DQF-COSY, double-quantum-filtered COSY; DTT, dithiothreitol; FPLC, fast-performance liquid chromatography; HMQC, heteronuclear multiple-quantum coherence; HMQC−NOESY−HMQC, 3D NMR experiment designed to resolve NOEs between amide protons with overlapping ^{1}H chemical shifts along two ^{15}N dimensions; HTH, helix−turn−helix; HTLV-1, human T-cell leukemia virus 1; KID, kinase-inducible domain; KIX, pKID-binding domain of CBP; MALDI, matrix-assisted laser desorption/ionization; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; NOESY−HMQC, heteronuclear 3D NOESY spectrum edited by $^{1}H-^{15}N$ correlations; PKA, protein kinase A; pKID, phosphorylated form of KID (S133); PE-COSY, primitive-exclusive correlation spectroscopy; ppm, parts per million; PTB, phosphotyrosine binding; RNA, ribonucleic acid; RNAP, RNA polymerase; rp-HPLC, reverse-phase high-performance liquid chromatography; P300, homologue of CBP; SA, simulated annealing; SH2, src-homology region 2; TBP, TATA-binding protein; t-BOC, *tert*-butoxycarbonyl; TOCSY, total correlation spectroscopy; TFE, trifluoroethanol; TFIID, general transcription factor II for RNAP-II; TOCSY−HMQC, hetereonuclear 3D spectrum edited by $^{1}H-^{15}N$ correlations; TOCSY−NOESY, homonuclear 3D NMR experiment combining successive TOCSY and NOESY mixing schemes; TOF, time-of-flight. Amino acids are designated by standard single-letter code.

S0006-2960(98)00080-4 CCC: $15.00
Published on Web 04/09/1998

phosphorylation, and transcriptional activation (*23*). The three-dimensional structure of a CREB pKID−CBP regulatory complex has recently been described (*24*) and provides a model of a coil → α-helix transition on assembly of a transcriptional activator−coactivator complex. The observed mode of phosphopeptide recognition, binding of an induced helix−turn−helix, differs from that of SH2 and PTB domains (*25*, *26*). Here, we demonstrate that the induced pKID HTH motif is predicted by the intrinsic structural propensities of the isolated phosphopeptide.

CREB−CBP-dependent transcriptional activation regulates diverse signaling pathways (such as those of peptide hormones, growth factors, calcium, and light; *8*, *27*−*29*). Such pathways have attracted broad interest as their deregulation is associated with impairment in memory, neoplasia (*30*), and developmental abnormalities (*31*). CREB contains two transcriptional activation regions, one kinase-inducible (designated pKID) and the other autonomous (Q2; *11*, *32*). Whereas the former domain interacts with CBP, the latter interacts with the $hTAF_{II}130$ subunit of basal transcriptional initiation factor TFIID (*33*, *34*). CBP (and its paralog P300; *35*) also participates in CREB-independent regulatory pathways, such as those involving the steroid-hormone receptor superfamily (*36*, *37*), nuclear proto-oncoproteins Jun (*14*) and Myb (*38*), signal-transduction factors Sap-1a (*39*), cubitus interruptus in hedgehog signaling in *Drosophila melanogaster* (*40*), and a viral virulence factor associated with human malignancy (HTLV-1 Tax; *41*). Structures of complexes between CBP and such CREB-independent regulatory factors have not to date been determined.

Biochemical studies of intact CREB have demontrated that the pKID domain is unusually susceptible to proteolysis and hence likely to be unstructured (*8*). Biophysical characterization of the isolated pKID domain (residues 101−160 of human CREB,[2] 60 residues) was limited in the original study of a pKID−CBP complex by the phosphopeptide's sparing solubility in the absence of guanidine hydrochloride (*24*). In accord with previous studies with circular dichroism (*16*), evidence was nevertheless obtained that showed that the domain is a random coil under experimental conditions employed with an extremely small propensity toward α-helix. In particular, ^1H and ^{15}N chemical shifts were observed to be not significantly different from random-coil values. In this study, we show that a central fragment of the pKID domain (residues 106−151, 46 residues) is by contrast highly soluble in aqueous solution, permitting its detailed characterization by multidimensional NMR spectroscopy. This fragment includes the critical phosphoserine and induced HTH recognition motif;[3] N- and C-terminal residues truncated in the present construction are disordered in the pKID−CBP complex and apparently not engaged in phosphopeptide−protein recognition[4] (*24*). At low temperatures, chemical shifts are observed to exhibit small but significant deviations from random-coil values, and these deviations are enhanced by addition of trifluoroethanol (TFE). Analysis of nuclear Overhauser effects (NOEs) demonstrates that the induced HTH motif is foreshadowed by the nascent structural propensities of the isolated phosphopeptide.[5] These observations support an analogy between induced fit in macromolecular assembly and early steps in the nucleation of protein folding (*42*−*46*).

MATERIALS AND METHODS

Materials. The human CREB pKID domain (residues 106−151; designated CREP-p in Figure 1B) was prepared by solid-phase peptide synthesis using t-BOC chemistry as described (*47*, *48*). The crude peptide was purified by size-exclusion chromatography (G25 Sephadex Superfine column) and reverse-phase (rp) HPLC (Waters C18 column). The fidelity of synthesis was verified in each case by N-terminal sequencing and mass laser-desorption time-of-flight (MALDI-TOF) mass spectrometry. Peptide purity was estimated to be >98% by analytical rp-HPLC. Whereas an extended CREB pKID peptide (60 residues) is sparingly soluble and aggregates in the absence of CBP or guanidine hydrochloride (*24*), a smaller fragment encompassing the CBP-binding motif (46 residues) was found to exhibit high solubility over a broad pH range (pH 2−8). Its elution position in size-exclusion chromatography is similar to that of monomeric and unfolded 34-residue fragments of parathyroid hormone (*49*).

Phosphorylation. The synthetic peptide (15 mg) was phosphorylated in vitro by cAMP-dependent protein kinase A (PKA) in the presence of 20 mM ATP in 50 mM MOPS (pH 7.0) and 10 mM $MgCl_2$ for 1 h at 37 °C. The phosphorylated peptide was repurified by ion-exchange FPLC (Mono-Q); the column was equilibrated in 20 mM Tris (pH 8.0). The phosphopeptide was eluted in a 0 to 1 M NaCl gradient. An aliquot of the recovered peak was mixed with an equal amount of the unphosphorylated peptide and reloaded on the column. Application of the same NaCl concentration gradient demonstrated that the phosphorylated and unphosphorylated peptides were well resolved. Analogous FPLC analysis of the original reaction mixture demonstrated that >95% of the original peptide was phosphorylated. The FPLC fraction containing the phosphopeptide was made 25 mM in ammonium/trifluoroacetic acid (TFA) and applied to a C18 Sep-Pak column. The column was washed with 25 mM ammonium/TFA containing 0 and 20% acetonitrile. The phosphopeptide was eluted with 25 mM ammonium/TFA in 50% acetonitrile and lyophilized. The final yield was 10 mg.

[2] Residue numbers refer to the intact CREB protein sequence and not to the synthetic peptide. The protein numbering scheme (*8*) includes residues encoded by exon D (14 residues; designated the CREB 341 isoform); an alternative numbering exists in the literature derived from a splicing isoform lacking this exon (CREB 327; *7*) (for example, see ref *22*). S133 in the present numbering scheme corresponds to S119 in the CREB 327 isoform.

[3] The use of "helix−turn−helix" is meant as a description of the phosphopeptide structure induced in the pKID−KIX structure (*24*) and is not meant to suggest any structural or evolutionary relationship to the HTH class of DNA-binding proteins (*2*). In particular, we note that the classical HTH contains characteristic helix−helix contacts not maintained in the orthogonal HTH in the pKID−KIX structure (Figure 1A).

[4] The phosphopeptide employed (46 residues) is smaller than that employed by Radhakrishan et al. (*24*) (60 residues). The N- and C-terminal residues (14 residues in total) not included in this present peptide (i.e., residues 101−106 and 152−160) are not engaged in CREB−KIX interactions and disordered in the solution structure of the complex. These 14 residues are shared between the CREB 341 and CREB 327 isoforms and are hence are unrelated to the 14 exon-D-encoded residues discussed in footnote 2.

[5] The sequence of αA suggests that a nascent helix may be stabilized by an aspartic acid N-cap (*82*) and possible (i,i+3) and (i,i+4) salt bridges (*63*); αB is amphipathic.

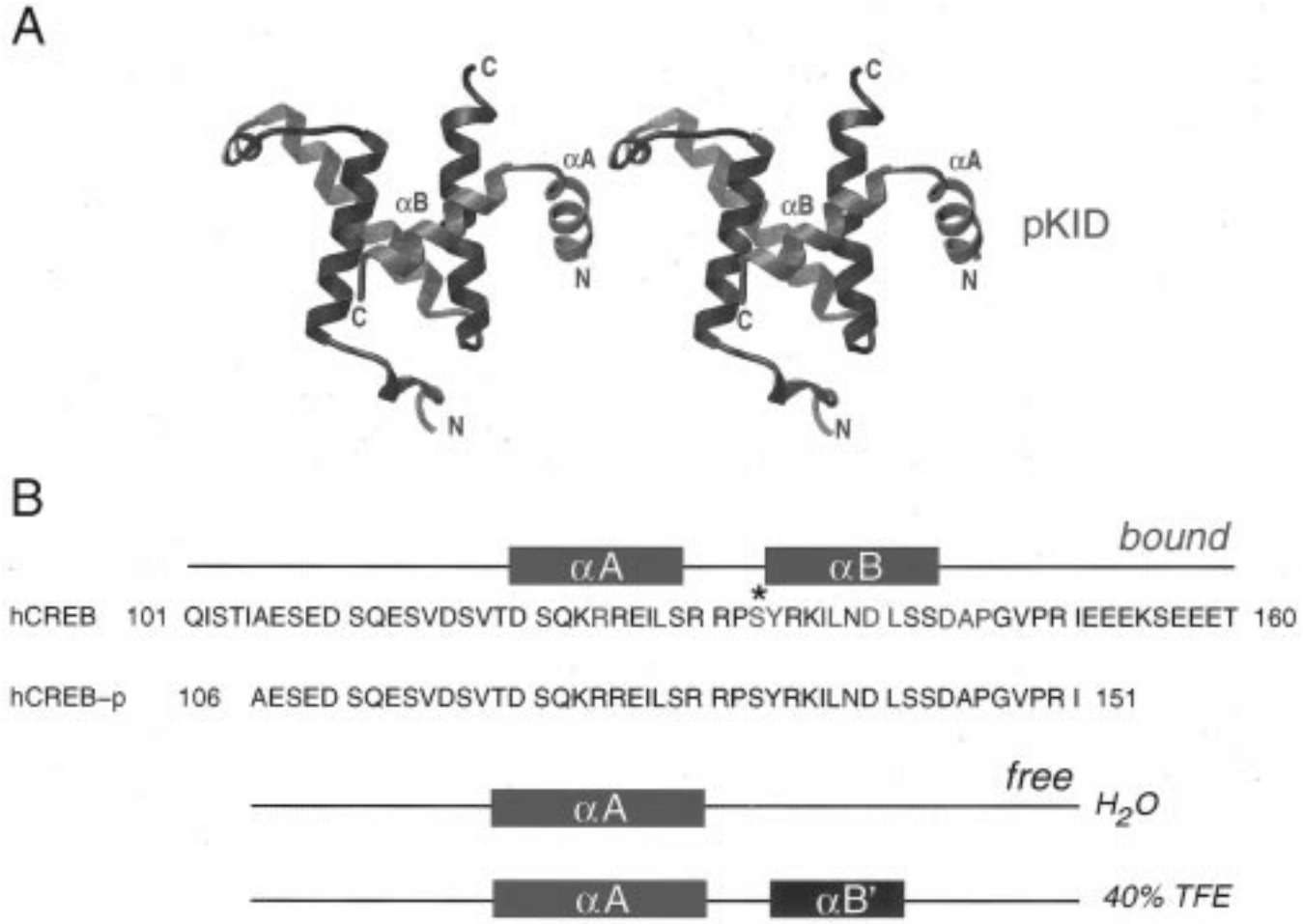

FIGURE 1: (A) Ribbon model (stereo representation) of the CREB pKID domain (red) and the CBP KIX domain (blue) as described by Radhakrishan et al. (*24*; figure reproduced with permission of the authors). Two recognition α-helices (designated αA and αB) are induced in CREB pKID on binding CPB. (B, upper panel) Sequence and position of induced α-helices (red rectangles) in the phosphopeptide fragment of CREB (residues 101−160, 60 residues) investigated by Radhakrishan et al. (*24*). The asterisk indicates the site of serine phosphorylation (S133). Residues shown in red are engaged in intermolecular side chain contacts at the pKID−KIX interface; contacts with αA are less extensive than are contacts with αB. (Lower panel) Sequence and position of α-helices in free CREB fragment in this study (designated CREB-p, residues 106−151). In aqueous solution at 4 °C, nascent folding of αA (red rectangle) is observed whereas αB′ (green) is induced on addition of 40% TFE (bottom panel). pKID residues not included in CREB-p (residues 101−105 and 152−160) are disordered in the solution structure of the KID−KIX complex (*24*). Their truncation was serendipitously found to enhance the solubility of the pKID fragment.

Isotopic Labeling. t-BOC protected amino acids containing ^{15}N were purchased from Cambridge Isotopes, Inc. (Woburn, MA), and used in solid-phase peptide synthesis (*47*). Two labeled peptides were prepared containing clusters of labeled residues at the N- or C-terminal portions of the peptide (set A, residues 106, 107, 109, 110, 113, 115, 116, 118, and 120; and set B, residues 137, 138, 140, 141, 144, 145, 147, and 148). The labeled peptides were purified and phosphorylated as described for the unlabeled peptide. The fidelity of incorporation was verified by 2D $^{15}N-^{1}H$ NMR HMQC spectroscopy (see Results; Figure 3A).

Circular Dichroism. CD spectra were obtained using an Aviv CD spectropolarimeter equipped with a thermister temperature control. A path length of 1 mm was used. Spectra were normalized to the mean-residue ellipticity on the basis of peptide concentrations as determined by optical density at 280 nm (estimated extinction coefficient based on an amino acid composition of one tyrosine). Spectra were obtained at successive temperatures (4−80 °C), percentages of TFE (0−80%), and peptide concentrations (10−100 mM). Normalized CD spectra and inferred helix contents are independent of peptide concentration in the range tested.

NMR Spectroscopy. Spectra were obtained at 500 MHz using a Varian Unity-plus spectrometer equipped with a 5 mm $^{1}H-^{15}N$ indirect detection probe. 1D spectra were obtained at successive temperatures (4−40 °C), percentages of TFE (0−40%), and peptide concentrations (0.5−5.0 mM). Chemical shifts were found to be independent of peptide concentration at the concentrations tested; resonance line widths are consistent with a monomeric fragment. 2D NMR studies focused on the phosphorylated peptide in aqueous solution and on the unmodified peptide under two conditions: aqueous solution at 4 °C and in 40% TFE (v/v) at 25 °C. 2D NOESY (mixing times of 200 and 400 ms), TOCSY (mixing time of 55 ms), DQF-COSY, and $^{15}N-^{1}H$ HMQC spectra were obtained in each case. Because of the similarity between spectra of the phosphorylated and unmodified peptides, 3D NMR experiments (*50*) focused on the more abundant unmodified sample. 3D $^{1}H-^{1}H$ TOCSY−NOESY spectra were obtained under the two solvent conditions with

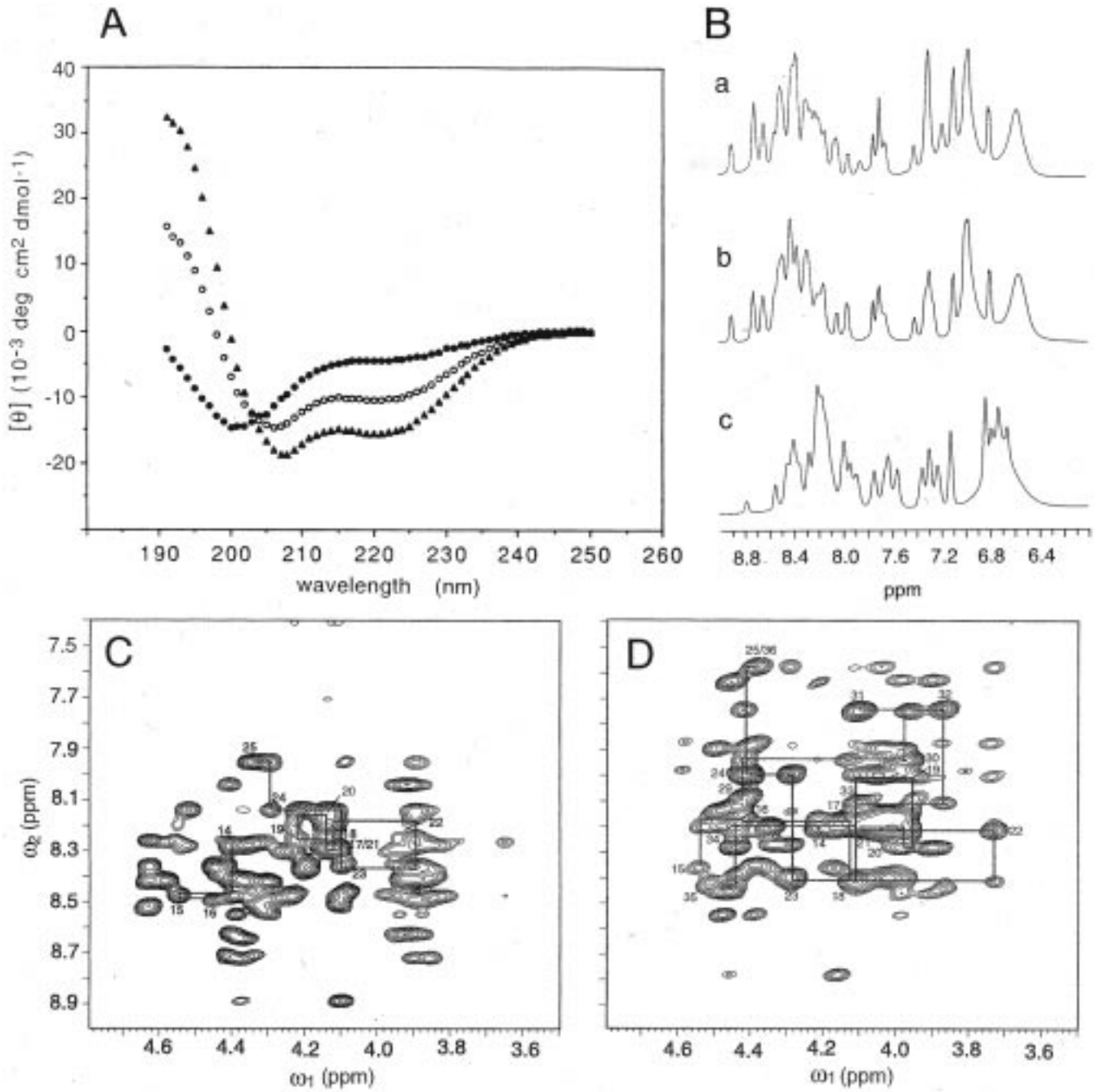

FIGURE 2: (A) Circular dichroism spectra of the CREB KID domain at 4 °C at successive concentrations of TFE: 0% (●), 20% (○), and 40% (▲). Spectra are unaffected by the presence or absence of phosphorylation at S133 and are independent of peptide concentration in the range of 10−100 μM. (B) 1D ^{1}H NMR spectra of the CREB KID domain: spectra a and b, phosphopeptide (pKID) and unmodified peptide (KID) at 4 °C in H_2O solution (50 mM KCl at pH 4.5); and spectrum c, unmodified peptide at 25 °C in 40% TFE. ^{1}H NMR spectra in either solvent are unaffected by the presence or absence of phosphorylation at S133 (except in the immediate neighborhood of the modification; see Supporting Information) and are independent of peptide concentration in the range of 0.5−5 mM. (C and D) Fingerprint $d_{\alpha N}$ regions of homonuclear 2D NOESY spectra of unmodified peptide in (C) aqueous solution at 4 °C and (D) 40% TFE at 25 °C. Helical segments of sequential assignment are outlined in each case (red, αA; and green, αB′ in 40% TFE). The limited resolution in these regions made necessary editing by homonuclear and ^{15}N heteronuclear NMR methods (Figure 3) for complete assignment.

respective mixing times of 55 and 400 ms. ^{15}N NOESY−HMQC (mixing time of 300 ms), ^{15}N TOCSY−HMQC (55 ms), and ^{15}N HMQC−NOESY−HMQC (200 ms) spectra were also obtained. 3D spectra were processed using Varian software with linear prediction. For the ^{1}H−^{1}H TOCSY−NOESY spectra, the observed 3D data matrix was 64 × 128 × 1024 and extended by linear prediction to 256 × 512 × 1024. Use of such data in sequential assignment was carried out as described (Figure 3C,D; *51*). For ^{15}N NOESY−HMQC and ^{15}N TOCSY−HMQC spectra, the observed 3D data matrices were 96 × 48 × 1024 and 96 × 24 × 1024, respectively, and extended to 512 × 256 × 1024 in each case. Use of such data in sequential assignment was carried out as described (*52*). Random-coil values of ^{1}H NMR chemical shifts were obtained from Wuthrich (*53*).

Buffers. CD and NMR spectra were obtained in aqueous solution containing 50 mM KCl at pH 4.5 as adjusted with dilute concentrations of HCl; aliquots of neat TFE were added to this solution to obtain successive concentrations of TFE (0−40% v/v).

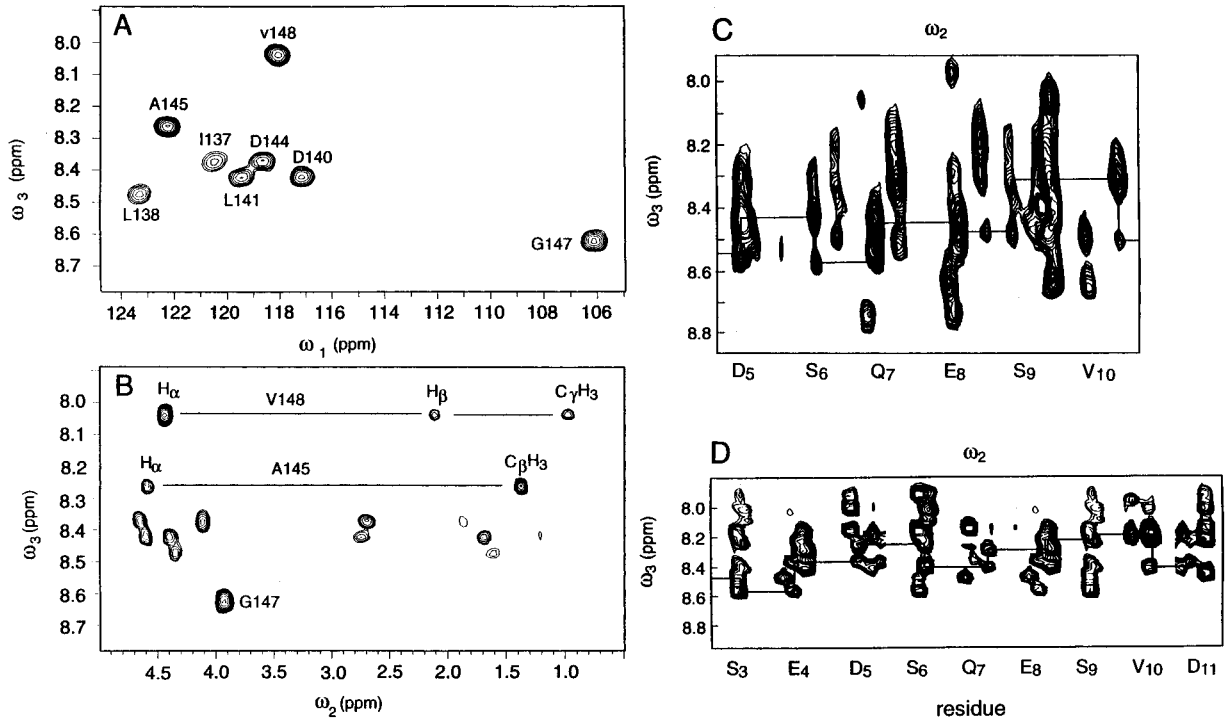

FIGURE 3: (A and B) Selective ^{15}N labeling of C-terminal residues permits observation of well-resolved $^{15}N-^{1}H$ HMQC correlations (A; obtained as a $\omega_1-\omega_3$ projection of the 3D TOCSY−HMQC experiment) and ^{15}N-edited ^{1}H TOCSY spin systems (the $\omega_2-\omega_3$ projection shown in panel B). Assignments in A and B are as indicated; the TOCSY mixing time was 55 ms at 4 °C. (C and D) Homonuclear 3D TOCSY−NOESY spectra of the unmodified peptide in (C) aqueous solution at 4 °C and (D) 40% TFE at 25 °C. Cross-peaks in individual $\omega_2-\omega_3$ planes are shown by residue. In panel C, sequential assignment of the peptide segment from D111 to V116 is illustrated at respective TOCSY editing frequencies in the ω_1 dimension of 2.75, 3.97, 2.09, 1.90, 3.97, and 2.19 ppm. In panel D, sequential assignment of the peptide segment from S109 to D117 is illustrated at respective TOCSY editing frequencies in the ω_1 dimension of 3.89, 2.10, 2.86, 3.98, 2.20, 2.10, 3.89, 2.20, and 2.86 ppm. $C_{\beta\alpha N}$ 3D TOCSY−NOESY connectivities were used in each case as defined by Gao et al. (*51*). Assignments are provided as Supporting Information.

RESULTS

Circular Dichroism. In accord with previous studies (*16*), the CD spectrum of the pKID fragment at room temperature lacks evidence of significant ordered structure. As the temperature is lowered to 4 °C, however, a partial and progressive α-helical folding transition is observed in the far ultraviolet. The magnitude of this transition is independent of peptide concentration (in the range of 10−100 μM) and of serine phosphorylation as previously inferred in studies of intact CREB (*21*). The helical signature of the phosphopeptide is accentuated by addition of 0−40% TFE (Figure 2A). On the basis of the values of the mean residue ellipticity ([θ]) at 222 nm, helix contents in aqueous solution and in 40% TFE are estimated to be 10−15% and 30−40%, respectively. As expected in an isolated peptide, thermal unfolding is not cooperative in either solvent (Supporting Information).

NMR Analysis. ^{1}H NMR spectra of KID and pKID fragments in aqueous solution (50 mM KCl at pH 4.5) exhibit temperature-dependent dispersion. ^{1}H NMR chemical shifts at 4 °C exhibit small but significant nonrandom values. These are independent of peptide concentration (in the range of 0.5−5 mM) and of serine phosphorylation (except in the immediate vicinity of S133). Dispersion is enhanced on addition of 0−40% TFE (Figure 2B). Complete sequential assignment at 4 °C and in 40% TFE at 25 °C was obtained using a combination of 2D NMR (NOESY, TOCSY, and DQF-COSY), homonuclear 3D NMR (3D TOCSY−NOESY), and heteronuclear 3D NMR experiments (^{15}N NOESY−HMQC, HMQC−NOESY−HMQC, and TOCSY−HMQC); the latter employed peptides containing multiple ^{15}N isotopic labels (see Materials and Methods). Resolution is incomplete in standard 2D spectra in either solvent (Figure 2C,D) but is extended in 3D spectra; shown in Figure 3 are representative sections of heteronuclear 3D NMR spectra (panels A and B) and homonuclear 3D NMR spectra (C and D). The former illustrates the essentially complete resolution of ^{15}N-edited spin systems. The latter are organized by residue to illustrate the use of homonuclear 3D TOCSY−NOESY cross-peaks in sequential assignment (*51*). Main chain ^{1}H chemical shifts are shown in Figure 4 in comparison with random-coil values; tables of assignments are provided as Supporting Information. In general, deviations from random-coil values (*53*) are largest for main chain resonances and decrease from H_β to distal side chain resonances.

Intrinsic Helical Propensities. Diagnostic short- and medium-range NOEs (*53*) provide a footprint of regions of nascent α-helix (Figure 5). This footprint is independent of serine phosphorylation. In aqueous solution at 4 °C, a contiguous series of $d_{\alpha N(i,i+3)}$ and $d_{\alpha\beta(i,i+3)}$ NOEs distinguishes residues 119−130 (NOEs shown in red in Figure 5). These and a second contiguous series (residues 134−141; NOEs shown in green in Figure 5) are observed in 40% TFE at 25 °C; NOEs observed only in 40% TFE are shown in black. We imagine that these regions exhibit fluctuating (and not rigid) structure as described for peptide fragments of native proteins (*42*, *44*). Contiguous TFE-independent and TFE-dependent (i,i+3) contacts between residues 134 and 141

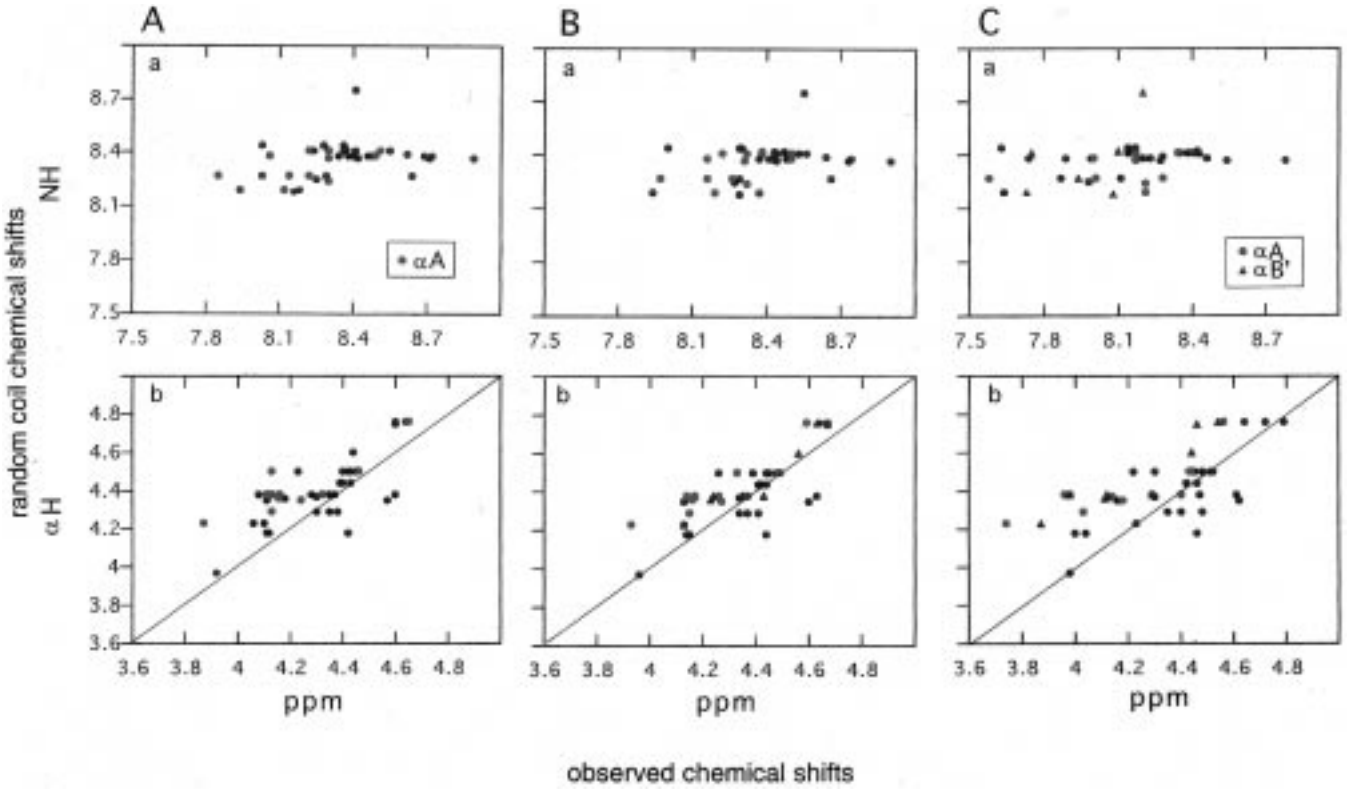

FIGURE 4: Comparison of observed main chain ^{1}H NMR chemical shifts and random-coil chemical shifts (*53*): (A) phosphorylated peptide in aqueous solution at 4 °C, (B) unphosphorylated peptide in aqueous solution at 4 °C, and (C) unphosphorylated peptide in 40% TFE at 25 °C. Each symbol indicates a residue. Residues in αA are shown in red, whereas in selected panels, residues in TFE-inducuble helix αB′ are shown as green triangles. Solid lines at the diagonal of the lower panels indicate the expected correlation in an ideal random coil. Residues in αA (red circles) exhibit systematic upfield deviations (left of diagonal). Because no correction to amide random-coil values was applied for temperature in the upper panels, in an ideal random coil, a linear correlation would be expected as an offset parallel to the diagonal of the plot.

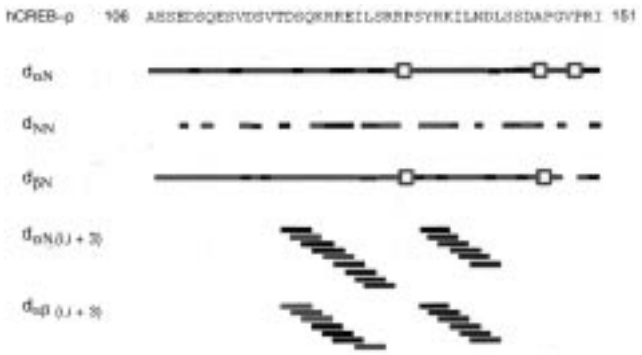

FIGURE 5: Summary of sequential assignment in the Wuthrich format. Red segments indicate NOEs observed both in aqueous solution at 4 °C and in 40% TFE at 25 °C; green segments indicate NOEs observed only in 40% TFE and black segments NOEs observed only in aqueous solution. Because results are identical for the phosphopeptide and unmodified peptide, only a single summary is provided.

further suggest that the cosolvent enhances a latent structural propensity (*54*). The two nascent helical segments correspond to well-defined recognition helices αA and αB in the structure of the pKID−CBP complex (Figure 1B; *24*). The induced helix αB′ is shorter than αB in the bound state (residues 133−144). As expected (*55*), H_α resonances in regions of nascent helix (shown in red and green in the lower panels of Figure 4) exhibit systematic upfield secondary shifts. The site of phosphorylation (RPpS) is not engaged in helical structure in either solvent. Non-native propensity for α-helix is not readily induced by 40% TFE elsewhere in the fragment.

An apparent helical stop (*56*) occurs following αA in aqueous solution and persists in 40% TFE (*57*). The helical stop includes the phosphoserine. In aqueous solution, residues RPpSY in fact exhibit a strong $d_{\alpha N(i,i+2)}$ contact between P132 and Y134 and a strong d_{NN} contact between pS133 and Y134. These contacts, characteristic of a β-turn, are consistent with the structure of the pKID−CBP complex (*24*). Although the 133−134 d_{NN} NOE is also observed in the unmodified peptide, it is possible that a propensity for a β-turn at this site is enhanced by the negative charge of phosphoserine as the P132−Y134 $d_{\alpha N(i,i+2)}$ contact is not observed in the absence of phosphorylation (its presence or absence in 40% TFE is obscured by resonance overlap). No long-range NOEs are observed in either solvent. In the pKID−CBP complex (*24*), a single interhelical contact is observed between the side chains of L128 and Y134 as the aromatic ring packs between the side chains of L128 and L138. Although the presence or absence of this contact in the isolated phosphopeptide could not be established due to overlap in the methyl resonances of the two leucines, we note that each of these resonances exhibits negligible secondary shifts. Because the orientation of Y134 in the structure of the pKID−CBP complex predicts large upfield ring-current shifts, the unremarkable chemical shifts of L128 and L138 in the isolated phosphopeptide suggest that this local hydrophobic cluster (if present) is not well organized. This suggests in turn that the relative orientation of the proto-

αA and αB segments is not well defined in the free phosphopeptide.

DISCUSSION

Cyclic AMP provides a "second messenger" in conserved signal-transduction pathways from membrane receptors to the nucleus (*11*, *12*). Transmission of the signal to responsive genes is mediated by families of specific transcription factors (*2*). This study focuses on one such factor, the cAMP-response-element binding protein (CREB; *7*−*9*). The functional system in the nucleus includes both CREB, an associated transcriptional coactivator CBP (*13*−*15*), and basal initiation factors including TFIID (*3*). Phosphorylation of a particular serine in (S133) is required for CBP−pCREB recognition (*8*, *13*, *16*) and regulation of target genes (*23*). This serine lies in a conserved kinase-inducible domain, designated pKID. The region of CBP that is involved in CREB binding and transcriptional activation (designated KIX) is likewise conserved among eukaryotic coactivators. In addition to binding CREB, CBP functions in diverse regulatory pathways (*14*, *36*−*40*), contains histone acetylase activity (*58*), and recruits other enzymatic activities (such as an RNA helicase; *59*) to the preinitiation complex (*5*, *60*). A human genetic disorder (Rubinstein-Taybi syndrome) is associated with mutations in CBP (*61*). The biological importance of the specific kinase-inducible CREB−CBP signaling pathway is highlighted by the observation of somatotrophic hypoplasia and dwarfism in transgenic mice expressing a nonphosphorylatable CREB mutant (*31*).

The structure of a specific complex between the pKID domain of CREB and the KIX domain of CBP has recently been determined (*24*) and provides the first model of a transcriptional activator−coactivator complex. The structure is remarkable for the induced fit of an HTH element in pKID (shown in red in Figure 1A); the two induced α-helices are nearly orthogonal and flank the site of serine phosphorylation (S133; asterisk in Figure 1B). Whereas an isolated KIX domain (residues 586−672 or 586−679 of mouse CBP) exhibits an autonomous structure, an unbound pKID phosphopeptide is essentially a random coil as previously inferred on the basis of CD studies at room temperature (*16*). In the original study (*24*), characterization of this phosphopeptide (residues 101−160 of the human CREB isoform 341) was limited by its sparing solubility in the absence of guanidine hydrochloride. In the course of screening pKID fragments of different length, we fortuitously found that a smaller fragment (residues 106−151; lower panel of Figure 1B) is by contrast highly soluble and monomeric in aqueous solution (pH 4.5). This fragment contains the site of serine phosphorylation and flanking HTH motif; the truncated residues are disordered in the pKID−KIX complex (*24*), and so their absence is unlikely to influence these results. The smaller phosphopeptide (46 residues) thus provides a model in which the folding propensities of the unbound pKID motif could be informatively investigated by NMR spectroscopy.

Our results delineate a partial folding transition at low temperatures. This transition is apparent in the CD spectrum at 4 °C (Figure 2A) and nonrandom pattern of ^{1}H NMR chemical shifts (Figures 2B and 4). Sites of nascent structure were localized by observation of residue-specific NOEs: The N-terminal recognition helix αA is observed as a nascent α-helix at 4 °C, whereas the amphipathic recognition helix αB is in part uncovered in 40% TFE at 25 °C (Figure 5). The mode of binding of pKID is thus foreshadowed by the intrinsic structural propensities of the phosphopeptide as summarized in Figure 1B. Phosphorylation of S133 does not affect the helical propensity of αA in aqueous solution but does enhance an adjoining turn-related NOE (*62*, *63*), perhaps via electrostatic interactions between the negative charge of the phosphoserine and the positive charges of neighboring arginine side chains. We speculate that the stronger propensity of αA than of αB to adopt a helical structure (in the absence of either KIX or cosolvent) is due to the former segment's favorable N-cap and possible (*i*,*i*+3) and (*i*,*i*+4) electrostatic interactions (*64*). It is intriguing that these relative propensities correlate with the extent of engagement of αA and αB in the pKID−KIX complex: Whereas αA occupies a peripheral binding site, αB lies within a hydrophobic groove and makes more extensive contributions to the protein−phosphopeptide interface (Figure 1A; *24*). Side chains in pKID in contact with CBP are highlighted in red in Figure 1B.

The utility of TFE as an empirical probe of hidden structural propensities (*64*, *65*) is of interest in light of its widespread use in diverse studies of peptide hormones and protein fragments (see, for example, refs *42*, *62*, and *66*−*71*). Because structures of receptor−ligand complexes are unavailable in many cases (for example, for glucagon, secretin, and parathyroid hormone), the relevance of TFE-induced structure in isolated peptides remains controversial. The pKID−CBP complex may be regarded as a model receptor−ligand complex in which the utility of TFE has been validated herein. Foreshadowing of the active structure of pKID in the nascent structure of the phosphopeptide does not imply, however, that such propensities limit its range of induced structures in unrelated protein complexes. Indeed, the structure of protein kinase A (PKA) and its mode of substrate binding (*72*, *73*) imply that KID binds to the kinase active site as an extended peptide. Critical recognition sites in the PKA−KID complex are likely to occur N-terminally with respect to the site of phosphorylation, i.e., in or amino-proximal to the region of the nascent αA helix. Whether the intrinsic structural propensities of the peptide sequence are realized or overridden is likely to be a function of the overall structural context and the available free energy of binding. The peripheral binding site of αA in the KIX complex thus exploits this segment's intrinsic helical propensity, whereas the deeply engaged catalytic site of PKA provides sufficient binding free energy to override this propensity. These considerations also underlie design of proteins containing "chameleon" sequences (*74*). Indeed, Radhakrishan et al. (*24*) speculate that the sequence of pKID represents an evolutionary compromise between the competing constraints of a helical binding mode in one case (KIX) and an extended binding mode in another (PKA). Analogous competition between constraints has been proposed to underlie the evolution of insulin sequences: a helix in the A-chain is required for receptor binding, whereas an extended strand is employed in prohormone processing (*75*, *76*).

The nascent structure of isolated peptides and the effects of TFE have been extensively investigated in relation to mechanisms of protein folding. Transient formation of

secondary structure is proposed to precede formation of compact intermediates (*43*−*45*). Whereas classic studies emphasized the role of such intermediates in a putative pathway (*77*), a recent perspective emphasizes an ensemble of nascent structures in rapid dynamic equilibrium (*78*). This ensemble coalesces to define a folding funnel en route to the folded state (*79*). Peptide fragments of sperm whale myoglobin, for example, have been studied by CD and NMR as models of possible folding initiation sites (*80*). Some peptides exhibit intrinsic conformational propensities in aqueous solution that correspond to their structures in the native state (the GH region), whereas others (helices B−E) are unstructured in water but exhibit a marked propensity to populate helical configurations on addition of cosolvents. Although the mechanism by which TFE affects peptide structure is unclear (*64*), it is thought that TFE biases a preexisting helix−coil equilibrium (in part by enhancing the strength of peptide hydrogen bonds) and so provides an empirical probe for uncovering propensities that would otherwise be inaccessible in experiments conducted in aqueous solution (*54*, *81*). This study suggests an analogy between the mechanisms of protein folding and assembly. The induced fit of the pKID HTH from an unstructured phosphopeptide at physiological temperature is likely to recapitulate physical events regulating the coalescence of partially folded conformations into a native state. In particular, just as CBP provides an external template for folding of pKID, we imagine that a folded subdomain in a compact intermediate can provide an internal template. An example of such an internal template has been proposed in a model of an oxidative protein-folding intermediate of human proinsulin (*76*). As in the KIX domain, the partial fold of the proinsulin intermediate provides a hydrophobic groove between helices (residues A12−A19 and B9−B19) in which docks an otherwise disordered strand (residues A1−A8) with an intrinsic helical propensity. Protein folding and protein recognition may thus be viewed as complementary problems. In particular, order−disorder transitions are likely to be of central importance in regulating the assembly of transcriptional regulatory complexes.

ACKNOWLEDGMENT

We thank H. T. Keutmann and Ashok Katri for assistance with peptide synthesis, characterization, and purification and J. P. Lee, K. Hallenga, and D. Jones for assistance with NMR spectroscopy.

SUPPORTING INFORMATION AVAILABLE

Five figures providing CD data as a function of temperature and concentration of TFE, secondary shifts of side chain ^{1}H NMR resonances, summaries of ^{1}H chemical-shift indices and short- and medium-range NOEs in the Wüthrich format, 1D ^{1}H NMR spectrum in 40% TFE at 4 °C, and the d_{NN} region of the NOESY spectra under the two conditions of study; seven tables providing ^{1}H and ^{15}N NMR resonance assignments for the phosphorylated and unphosphorylated peptides in aqueous solution and in 40% TFE; tables providing corresponding secondary shifts relative to random-coil values; and tables providing differences in chemical shifts between samples, including effects of S133 phosphorylation (28 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Tjian, R. (1996) *Philos. Trans. R. Soc. London 351*, 491−9.
2. Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−95.
3. Patikoglou, G., and Burley, S. K. (1997) *Annu. Rev. Biophys. Biomol. Struct. 26*, 289−325.
4. Uesugi, M., Nyanguile, O., Lu, H., Levine, A. J., and Verdine, G. L. (1997) *Science 277*, 1310−3.
5. Ptashne, M., and Gann, A. (1997) *Nature 386*, 569−77.
6. Montminy, M. R., and Bilezikjian, L. M. (1987) *Nature 328*, 175−8.
7. Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L., and Habener, J. F. (1988) *Science 242*, 1430−3.
8. Gonzalez, G. A., Menzel, P., Leonard, J., Fischer, W. H., and Montminy, M. R. (1991) *Mol. Cell. Biol. 11*, 1306−12.
9. Yun, Y. D., Dumoulin, M., and Habener, J. F. (1990) *Mol. Endocrinol. 4*, 931−9.
10. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell 71*, 1223−37.
11. Brindle, P. K., and Montminy, M. R. (1992) *Curr. Opin. Genet. Dev. 2*, 199−204.
12. Habener, J. F., Miller, C. P., and Vallejo, M. (1995) *Vitam. and Horm. (San Diego) 51*, 1−57.
13. Chrivia, J. C., Kwok, R. P., Lamb, N., Hagiwara, M., Montminy, M. R., and Goodman, R. H. (1993) *Nature 365*, 855−9.
14. Arias, J., Alberts, A. S., Brindle, P., Claret, F. X., Smeal, T., Karin, M., Feramisco, J., and Montminy, M. (1994) *Nature 370*, 226−9.
15. Kwok, R. P., Lundblad, J. R., Chrivia, J. C., Richards, J. P., Bachinger, H. P., Brennan, R. G., Roberts, S. G., Green, M. R., and Goodman, R. H. (1994) *Nature 370*, 223−6.
16. Parker, D., Ferreri, K., Nakajima, T., LaMorte, V. J., Evans, R., Koerber, S. C., Hoeger, C., and Montminy, M. (1996) *Mol. Cell. Biol. 16*, 694−703.
17. Merino, A., Buckbinder, L., Mermelstein, F. H., and Reinberg, D. (1989) *J. Biol. Chem. 264*, 21266−76.
18. Weih, F., Stewart, A. F., Boshart, M., Nitsch, D., and Schutz, G. (1990) *Genes Dev. 4*, 1437−49.
19. Nichols, M., Weih, F., Schmid, W., DeVack, C., Kowenz-Leutz, E., Luckow, B., Boshart, M., and Schutz, G. (1992) *EMBO J. 11*, 3337−46.
20. Hagiwara, M., Brindle, P., Harootunian, A., Armstrong, R., Rivier, J., Vale, W., Tsien, R., and Montminy, M. R. (1993) *Mol. Cell. Biol. 13*, 4852−9.
21. Richards, J. P., Bachinger, H. P., Goodman, R. H., and Brennan, R. G. (1996) *J. Biol. Chem. 271*, 13716−23.
22. Bullock, B. P., and Habener, J. F. (1998) *Biochemistry 37*, 3795−3809.
23. Hagiwara, M., Alberts, A., Brindle, P., Meinkoth, J., Feramisco, J., Deng, T., Karin, M., Shenolikar, S., and Montminy, M. (1992) *Cell 70*, 105−13.
24. Radhakrishan, I., Perez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., and Wright, P. E. (1997) *Cell 91*, 741−52.
25. Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M. D., Rios, C. B., Silverman, L., and Kuriyan, J. (1992) *Nature 358*, 646−53.
26. Zhou, M.-M., Ravichandran, K. S., Olejniczak, E. T., Petros, A. M., Meadows, R. P., Sattler, M., Harlan, J. E., Wade, W. S., Burakoff, S. J., and Fesik, S. W. (1995) *Nature 378*, 584−92.
27. Ginty, D. D., Kornhauser, J. M., Thompson, M. A., Bading, H., Mayo, K. E., Takahashi, J. S., and Greenberg, M. E. (1993) *Science 260*, 238−41.
28. Ginty, D. D., Bonni, A., and Greenberg, M. E. (1994) *Cell 77*, 713−25.
29. Bito, H., Deisseroth, K., and Tsien, R. W. (1996) *Cell 87*, 1203−14.
30. Arany, Z., Newsome, D., Oldread, E., Livingston, D. M., and Eckner, R. (1995) *Nature 374*, 81−4.
31. Struthers, R. S., Vale, W. W., Arias, C., Sawchenko, P. E., and Montminy, M. R. (1991) *Nature 350*, 622−4.

32. Yamamoto, K. K., Gonzalez, G. A., Menzel, P., Rivier, J., and Montminy, M. R. (1990) *Cell 60*, 611−7.
33. Ferreri, K., Gill, G., and Montminy, M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1210−3.
34. Nakajima, T., Uchida, C., Anderson, S., Parvin, J., and Montminy, M. (1997) *Genes Dev. 11*, 738−47.
35. Lundblad, J. R., Kwok, R. P., Laurance, M. E., Harter, M. L., and Goodman, R. H. (1995) *Nature 374*, 85−8.
36. Chakravarti, D., LaMorte, V. J., Nelson, M. C., Nakajima, T., Schulman, I. G., Juguilon, H., Montminy, M., and Evans, R. M. (1996) *Nature 383*, 99−103.
37. Kamei, Y., Xu, L., Heinzel, T., Torchia, J., Kurokawa, R., Gloss, B., Lin, S. C., Heyman, R. A., Rose, D. W., Glass, C. K., and Rosenfeld, M. G. (1996) *Cell 85*, 403−14.
38. Dai, P., Akimaru, H., Tanaka, Y., Hou, D. X., Yasukawa, T., Kaneilshii, C., Takahashi, T., and Ishii, S. (1996) *Genes Dev. 10*, 528−40.
39. Janknecht, R., and Nordheim, A. (1994) *Oncogene 12*, 1961−9.
40. Akimaru, H., Chen, Y., Dai, P., Hou, D. X., Nonaka, M., Smolik, S. M., Armstrong, S., Goodman, R. H., and Ishii, S. (1997) *Nature 386*, 735−8.
41. Kwok, R. P., Laurance, M. E., Lundblad, J. R., Goldman, P. S., Shih, H., Connor, L. M., Marriott, S. J., and Goodman, R. H. (1996) *Nature 380*, 642−6.
42. Dyson, H. J., and Wright, P. E. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 519−38.
43. Sancho, J., Neira, J. L., and Fersht, A. R. (1992) *J. Mol. Biol. 224*, 749−58.
44. Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., and Wright, P. E. (1992) *J. Mol. Biol. 226*, 795−817.
45. Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H. C., Lerner, R. A., and Wright, P. E. (1992) *J. Mol. Biol. 226*, 819−35.
46. Reymond, M. T., Huo, S., Duggan, B., Wright, P. E., and Dyson, H. J. (1997) *Biochemistry 36*, 5234−44.
47. Barany, G., Kneib-Cordonier, N., and Mullen, D. G. (1987) *Int. J. Pept. Protein Res. 30*, 705−39.
48. Kochoyan, M., Havel, T. F., Nguyen, D., Dahl, C. E., Keutmann, H. T., and Weiss, M. A. (1991) *Biochemistry 30*, 3371−86.
49. Gardella, T. J., Axelrod, D., Rubin, D., Keutmann, H. T., Potts, J. T., Jr., Kronenberg, H. M., and Nussbaum, S. R. (1991) *J. Biol. Chem. 266*, 13141−6.
50. Oschkinat, H., Griesinger, C., Kraulis, P. J., Sorensen, O. W., Ernst, R. R., Gronenborn, A. M., and Clore, G. M. (1988) *Nature 332*, 374−6.
51. Gao, X., and Burkhart, W. (1991) *Biochemistry 30*, 7730−9.
52. Clore, G. M., and Gronenborn, A. M. (1994) *Methods Enzymol. 239*, 349−63.
53. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
54. Jasanoff, A., and Fersht, A. R. (1994) *Biochemistry 33*, 2129−35.
55. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) *Biochemistry 31*, 1647−51.
56. Kim, P. S., and Baldwin, R. L. (1984) *Nature 307*, 329−34.
57. Nelson, J. W., and Kallenbach, N. R. (1989) *Biochemistry 28*, 5256−61.
58. Bannister, A. J., and Kouzarides, T. (1996) *Nature 384*, 641−3.
59. Nakajima, T., Uchida, C., Anderson, S., Parvin, J., and Montminy, M. (1997) *Cell 90*, 1107−12.
60. Greenblatt, J. (1997) *Curr. Opin. Cell Biol. 9*, 310−9.
61. Petrij, F., Giles, R. H., Dauwerse, H. G., Saris, J. J., Hennekam, R. C., Masuno, M., Tommerup, N., van Ommen, G. J., Goodman, R. H., Peters, D. J., et al. (1995) *Nature 376*, 348−51.
62. Yao, J., Feher, V. A., Espejo, B. F., Reymond, M. T., Wright, P. E., and Dyson, H. J. (1994) *J. Mol. Biol. 243*, 736−53.
63. Neira, J. L., and Fersht, A. R. (1996) *Folding Des. 1*, 231−41.
64. Lyu, P. C., Gans, P. J., and Kallenbach, N. R. (1992) *J. Mol. Biol. 223*, 343−50.
65. Storrs, R. W., Truckses, D., and Wemmer, D. E. (1992) *Biopolymers 32*, 1695−702.
66. Reed, J., Hull, W. E., Ponstingl, H., and Himes, R. H. (1992) *Biochemistry 31*, 11888−95.
67. Sonnichsen, F. D., Van Eyk, J. E., Hodges, R S., and Sykes, B. D. (1992) *Biochemistry 31*, 8790−8.
68. Zagorski, M. G., and Barrow, C. J. (1992) *Biochemistry 31*, 5621−31.
69. Neugebauer, W., Surewicz, W. K., Gordon, H. L., Somorjai, R. L., Sung, W., and Willick, G. E. (1992) *Biochemistry 31*, 2056−63.
70. Bruch, M. D., and Hoyt, D. W. (1992) *Biochim. Biophys. Acta 1159*, 81−93.
71. Laczko-Hollosi, I., Hollosi, M., Lee, V. M., and Mantsch, H. H. (1992) *Eur. Biophys. J. 21*, 345−8.
72. Knighton, D. R., Zheng, J., Eyck, L. F. T., Ashford, V. A., Xuong, N.-H., Taylor, S. S., and Sowadski, J. M. (1991) *Science 253*, 407−20.
73. Knighton, D. R., Zheng, J., Ten Eyck, L. F., Xuong, N.-H., Taylor, S. S., and Sowadski, J. M. (1991) *Science 253*, 414−20.
74. Minor, D. L, Jr., and Kim, P. S. (1996) *Nature 380*, 730−4.
75. Steiner, D. F., Rouille, Y., Gong, Q., Martin, S., Carroll, R., and Chan, S. J. (1996) *Diabetes Metab. 22*, 94−104.
76. Hua, Q. X., Hu, S. Q., Frank, B. H., Jia, W., Chu, Y. C., Wang, S. H., Burke, G. T., Katsoyannis, P. G., and Weiss, M. A. (1996) *J. Mol. Biol. 264*, 390−403.
77. Creighton, T. E. (1995) *Curr. Biol. 5*, 353−6.
78. Dill, K. A., and Chan, H. S. (1997) *Nat. Struct. Biol. 4*, 10−9.
79. Wolynes, P., Luthey-Schulten, Z., and Onuchic, J. (1996) *Chem. Biol. 3*, 425−32.
80. Reymond, M. T., Merutka, G., Dyson, H. J., and Wright, P. E. (1997) *Protein Sci. 6*, 706−16.
81. Luo, P., and Baldwin, R. L. (1997) *Biochemistry 36*, 8413−21.
82. Serrano, L., and Fersht, A. R. (1989) *Nature 342*, 296−9.

BI9800808