



University of Tennessee, Knoxville  
**TRACE: Tennessee Research and Creative  
Exchange**

---

Chancellor's Honors Program Projects

Supervised Undergraduate Student Research  
and Creative Work


---

5-2012

## Investigating the Flexibility of Intrinsically Disordered Proteins in Folding and Binding

Amanda Leilah Debuhr  
adebuhr@utk.edu

Follow this and additional works at: [https://trace.tennessee.edu/utk\\_chanhonoproj](https://trace.tennessee.edu/utk_chanhonoproj)

 Part of the [Biochemistry Commons](#), [Biophysics Commons](#), [Molecular Biology Commons](#), and the [Structural Biology Commons](#)

---

### Recommended Citation

Debuhr, Amanda Leilah, "Investigating the Flexibility of Intrinsically Disordered Proteins in Folding and Binding" (2012). *Chancellor's Honors Program Projects*.  
[https://trace.tennessee.edu/utk\\_chanhonoproj/1509](https://trace.tennessee.edu/utk_chanhonoproj/1509)

This Dissertation/Thesis is brought to you for free and open access by the Supervised Undergraduate Student Research and Creative Work at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Chancellor's Honors Program Projects by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact [trace@utk.edu](mailto:trace@utk.edu).

# **Investigating the Flexibility of Intrinsically Disordered Proteins in Folding and Binding**

**Amanda Leilah DeBuhr**

Biological Sciences-BCMB

An undergraduate honors thesis submitted in partial fulfillment of the requirements for the  
chancellors honors program

**Faculty Advisor:** Dr. Valerie Berthelier ([vberthel@mc.utmck.edu](mailto:vberthel@mc.utmck.edu))

Conformational Diseases & Therapeutics

The University of Tennessee Health Science Center - Knoxville

Graduate School of Medicine

## **ABSTRACT**

Intrinsically disordered proteins are involved in many important cellular processes such as the regulation of gene expression at the levels of transcription and translation, and signal transduction. They are also involved in the regulation of cancerous and neurodegenerative disease processes. Investigating the mechanism by which intrinsically disordered proteins (IDPs) acquire structure upon binding to their specific partners is technically challenging as the structural flexibility of IDPs limits the number of suitable techniques. Circular dichroism (CD) spectroscopy coupled with osmotic stress was used to measure the conformational changes of two IDPs separate from their mutual binding. The disordered region of the CREB binding protein called NCBD and its binding partner ACTR were used in this study. The osmotic stress applied by the presence of osmolytes revealed information on the hydration changes and energetics that accompany folding. Overall, this allows us to more closely discern the folding properties of these IDPs to uncover how folding contributes toward binding.

## **BACKGROUND AND INTRODUCTION**

### **Characteristics of Intrinsically Disordered Proteins and Their Biological Functions**

Intrinsically disordered proteins (IDPs) lack a tertiary structure and exist in an ensemble of states both in solution and when unbound to a ligand *in vivo*. They usually have some local elements of secondary structure and some clusters of hydrophobic amino acids (1). Despite the high degree of disorder, these proteins are far from randomly structured and form transient, yet specific secondary and tertiary structural elements (2).

Intrinsically disordered proteins have prompted researchers to question the structure-function paradigm, which states that the three-dimensional structure of a protein determines its specific function (3,1). In fact, many proteins and domains within proteins have been found to lack ordered globular structure *in vivo* and in solution. More than 15,000 proteins in the Swiss Protein Database were found to contain disordered regions of at least 40 consecutive amino acid residues (3). Furthermore, the occurrence of unstructured regions of significant size (>50 residues) is common in functional proteins (1). Predictions on 29 genomes have established that intrinsic disorder is more common in proteins from eukaryotes than in those from bacteria and archaea, with more than 30% of eukaryotic proteins having disordered regions greater than 50 consecutive residues (3).

The amino acid composition of intrinsically disordered proteins is distinguished from those of structured proteins by certain features. These features include the presence of numerous uncompensated charged groups, resulting in a large net charge at neutral pH and a low content of hydrophobic amino acid residues (3). Specifically, amino acid sequences particular to IDPs are characterized by the presence of low sequence complexity and amino-acid compositional bias, a low content of bulky hydrophobic amino acids which

would normally form the core of a folded globular protein (Val, Leu, Ile, Met, Phe, Trp, and Tyr), and a high proportion of particular polar and charged amino acids (Gln, Ser, Pro, Glu, Lys, and on occasion, Gly, and Ala) (1). This combination of the low hydrophobicity and relatively high net charge is the characteristic of these proteins that contribute to their lack of compact structure under physiological conditions (3).

Intrinsically disordered proteins are structurally adaptable and bind ligands promiscuously, which means that they can adopt different conformations when they meet different ligands (2). For example, the structure of NCBD in complex with IRF-3 has an entirely different topology from its unbound conformation, both of which are different from its structure in complex with ACTR. NMR data suggests that NCBD may sample conformations other than its ACTR binding structure. The ability to sample different conformations is facilitated by the low structural stability which allows the domain to have sizeable populations of unfolded or partially folded states (2).

Many biological functions require the participation of intrinsically disordered proteins. They function in the regulation of transcription and translation, regulation of the cell cycle, cellular signal transduction, protein phosphorylation, the storage of small molecules, and the regulation of the self-assembly of large multiprotein complexes such as the bacterial flagellum and the ribosome (1,2). Furthermore, IDPs that are involved in signaling and transcription are relatively unstable and can be targeted for proteolytic degradation through the ubiquitin-proteasome system (1,2). They also function in protein-nucleic acid recognition as DNA-binding proteins and RNA-binding proteins. Intrinsically disordered regions are also found in chaperones for other proteins and for RNA molecules,

in which capacity they bind to the misfolded proteins and RNA molecules and help to unfold kinetically-trapped folding intermediates (2). It is hypothesized that their structural flexibility is what allows them to be suited for involvement in so many different regulatory and signaling processes.

### **Characteristics of ACTR and NCBD and Their Biological Functions**

The activator for thyroid hormone and retinoid receptors protein (ACTR) of the p160 nuclear receptor co-activator is disordered in the free state, is a coactivator for nuclear hormone receptors, and has multiple interaction domains (1,4). It directly binds nuclear receptors and stimulates their transcriptional activities in a hormone-dependent fashion (4). It recruits two other nuclear factors, CBP and P/CAF, and thus serves as a scaffold for creating a multisubunit coactivator complex. ACTR also is a histone acetyltransferase (HAT) protein and plays an important role in hormonal signaling by acetylating both free histones and nucleosome histones with substrate preference on histones H3 and H4 (4).

Northern analysis showed that the mRNA level of ACTR was found in several different human tissues and varied in its expression among them. Longer transcripts were abundant in heart, skeletal muscle, pancreas, and placenta, relatively low in brain, very low in lung and liver, and barely detectable in kidney (4). Shorter transcripts were also found in different tissues at lower abundances (4). Human cancer cell lines examined also showed restricted expression of ACTR mRNA (4).

The steroid receptor coactivator-3 (SRC-3) is the homologue of ACTR in mouse. It is a coactivator of nuclear receptors and is expressed in a tissue-specific manner throughout

the oocytes, mammary glands, hippocampus, olfactory bulb, smooth muscle, hepatocytes, and vaginal epithelium (5). Genetic disruption of SRC-3 in mice results in a pleiotropic phenotype showing dwarfism, delayed puberty, reduced female reproductive function, and blunted mammary gland development (5). Hormonal analysis indicates that SRC-3 plays a role in both the growth hormone regulatory pathway and the production of estrogen, which explains the mutant phenotypes observed when it is disrupted (5).

The cyclic-AMP-response-element-binding protein, (CREB)-binding protein, or (CBP), which contains the intrinsically disordered region NCBD, is a large nuclear protein that enhances gene expression through the modification of histones in the process of chromatin remodeling (1,6,7). Studies of steroid-regulated gene promoters have revealed that p160s and HAT proteins like CBP are among the first cofactors recruited in response to ligand (6). Recruitment of coactivators to promoters/enhancers is crucial for the sequential chromatin modification and remodeling events preceding transcription (6). Changes in the chromatin structure are one of the ways in which gene expression is controlled. Hyperacetylation of core histones is associated with gene activation at the level of transcription while deacetylation is involved in gene repression (7). The recruitment of the CREB-binding protein (CBP) to target gene promoters/enhancers facilitates acetylation of histone N-terminal tails, leading to chromatin remodeling and enhanced gene expression (6). This has been demonstrated for nuclear receptors, which activate transcription of their target genes in response to ligand binding (6). Ligand bound receptors undergo a conformational change that stimulates their interaction with cofactors that contain functional LXXLL motifs, such as the p160 coactivators SRC1, TIF2, and ACTR (6).

The CBP system exemplifies how the binding of intrinsically disordered proteins to their targets is often regulated by covalent modifications that lead to simple biological switches (1). CBP is a bridging protein that transmits the signal of ligand-induced conformation change to the basal transcriptional machinery (4). It modifies both chromatin and transcription factors through its intrinsic acetyltransferase activity, and also contributes to the scaffold for the recruitment and assembly of the transcriptional machinery (1). It transmits the signal for the activation of transcription by nuclear receptors, including CREB, AP-1, and bHLH (7). Nuclear receptors are a large family of ligand-dependent transcription factors that bind as homodimers or heterodimers to their cognate DNA elements and regulate genes involved in critical aspects of cell proliferation, differentiation, and homeostasis (7). CBP plays a role in the control of cell growth and differentiation by integrating signals for gene expression through interacting with the transcription machinery and chromatin structure modification (7).

NCBD is the C-terminal domain comprising residues 2058-2116 of human CBP that is responsible for the interaction of CBP with a number of other important proteins such as the steroid receptor coactivators, interferon regulatory proteins, and the viral protein Tax (1,2). In the ligand free state NCBD is compact and has a high degree of helicity, but apparently it lacks sigmoidicity in the unfolding curve. These observations suggest that NCBD is a molten globule (2).

While NCBD forms a molten globule with extensive tertiary interactions both in the free and ligand bound states, ACTR is a random-coil like IDP that folds into an extended structure upon binding to its cognate ligands and has very few intramolecular, tertiary



interactions in the bound state (2). While an extended random coil structure is needed to wrap around the binding partner, a molten globule ensemble of interconverting folded structures provides a target for conformational selection by diverse ligands (2).

NCBD and ACTR fold upon mutually binding into a complex in a process called mutual synergistic folding (1,2). Unlike the well ordered TAZ and KIX domains of CBP, the NCBD resembles a highly helical molten globule more than a stably folded protein (1). The NCBD has an amino-acid composition that would be expected for an intrinsically disordered protein. It is highly enriched in Gln, Pro, Ser, and Leu residues, and cannot bury enough of its hydrophobic surface to form a stable globular structure in the absence of its binding partner (1). This is reinforced by the denaturation behavior of isolated NCBD, which shows neither the sigmoidal unfolding curve nor the characteristic heat capacity change that accompany the unfolding of a globular protein (1).

### **Studying the IDPs ACTR and NCBD**

The discovery of so many unstructured proteins and regions within proteins is recent and can be ascribed to the availability of large amounts of sequence data coupled with gene based functional analysis, and to spectroscopic techniques that have been developed to analyze their structural properties in solution (1). Crystal-structure analysis is useful for indicating the presence of unstructured states in a protein, but cannot provide structural information about these unstructured states (1). Instead, spectroscopic methods including NMR and CD spectroscopy, among others, are commonly used to study the dynamics of disordered proteins (1).

In order to investigate the structural changes of NCBD and ACTR before and upon binding, the CD signal was measured as increasing osmotic stress was applied to the peptides in solution. It was hypothesized that increased osmotic pressure would induce protein folding by effecting a transition from random coil to  $\alpha$ -helix in unbound ACTR and NCBD as measured by CD spectroscopy. The CD-signal at 200nm was used as an indicator of random coil, while the signal at 222nm was used as an indicator of  $\alpha$ -helix. Increasing negativity in the signal at 222nm means that the protein contains more  $\alpha$ -helix, while increasing positivity in the signal at 200nm means that the protein contains less random coil.

This technique is relevant to the environment inside living cells, as small cosolutes called osmolytes are used to counteract external osmotic pressure that would otherwise cause proteins to misfold. Osmolytes protect proteins by stabilizing their native conformations, and are excluded from protein-water interfaces (8). However, the exact mechanism by which osmolytes stabilize macromolecules is not fully understood (8).

One possible mechanism by which osmolytes stabilize proteins is molecular crowding due to excluded volume interactions. It has been demonstrated that a high fractional volume occupancy of crowding agents, such as soluble polymers, significantly shifts the non-native to folded thermodynamic equilibrium ( $D \leftarrow \rightarrow N$ ) toward the more compact native states (8). This effect is associated with the restriction of protein conformations to allow larger free volume for osmolytes, thereby destabilizing the unfolded state with respect to the native conformation (8).

## **MATERIALS AND METHODS**

## **Materials**

The nuclear receptor coactivator binding domain (NCBD) of mouse cAMP response element binding (CREB) protein (CBP, accession: NP\_001020603), CBP(2059-2117) (59 residues, 6545 Da), and the interaction domain of mouse activator for thyroid hormone and retinoid receptor (ACTR, accession: o09000), ACTR(1025-1098) (74 residues, 7989 Da), were synthesized by solid-phase Fmoc chemistry (Keck-Yale facility). The crude peptides were purified by high performance liquid chromatography using a Zorbax SB-C3 column (Agilent Technologies) with a reverse phase water (+ 0.05% TFA) / acetonitrile (+ 0.05% TFA) gradient and lyophilized. Fractions were collected in 100 x 13 mm disposable glass tubes. Mass spectrometry confirmed the molecular mass of the peptides. Phosphate buffered saline (PBS, 20X), TFA, and ethylene glycol, triethylene glycol polyethylene glycol 400, betaine, and xylitol were used without further purification.

## **Purification and Preparation of Peptide Samples**

The peptide samples (ACTR and NCBD) were prepared for CD spectroscopy by HPLC purification. The frozen peptides were allowed to come to room temperature and then re-suspended in 1%(wt) TFA until dissolved completely, and then centrifuged for 8 minutes at high speed to collect insoluble impurities in a pellet at the bottom. The supernatant was removed and the pellet containing resins, impurities, and contaminants was allowed to remain behind. The supernatant was injected onto the HPLC column.

The peptide fractions corresponding to the 215 nm absorbance peak on the graph produced by the BioLogic DuoFlow software were collected and combined. A 35  $\mu$ L sample was taken from the combined fractions for analysis by the analytic HPLC. The remainder was lyophilized and frozen for use in the CD experiments.

The yields of the peptides after purification were calculated according to the linear fit from a standard curve made for the ACTR peptide. These linear fits of these standard curves were:

$$\text{Area} = (866.16 \pm 4.81) \times (\mu\text{g ACTR})$$

$$\text{Area} = (1440) \times (\mu\text{g NCBD})$$

A 35  $\mu\text{L}$  sample was taken from each tube of combined fractions and analyzed by analytical HPLC. The HPLC produced a graph which had mAU on the y axis and minutes on the x axis. The area under the curve on a graph of mAU vs. time (minutes) corresponding to the peptide was taken and used in the following calculation:

$$M_{\text{purified ACTR}} \cdot \text{mg} = \frac{\text{Area } \mu\text{g} \cdot \text{mL}^{-1}}{866.16} \cdot 10^{-3} \text{mg} \cdot \mu\text{g}^{-1} \cdot \frac{(V_{\text{collected}} - V_{\text{HPLC sample}})}{V_{\text{collected}}} \cdot \frac{\text{mL}}{\text{mL}}$$

$$M_{\text{purified NCBD}} \cdot \text{mg} = \frac{\text{Area } \mu\text{g} \cdot \text{mL}^{-1}}{1440} \cdot 10^{-3} \text{mg} \cdot \mu\text{g}^{-1} \cdot \frac{(V_{\text{collected}} - V_{\text{HPLC sample}})}{V_{\text{collected}}} \cdot \frac{\text{mL}}{\text{mL}}$$

$$M_{\text{crude ACTR or NCBD}} \cdot \text{mg} = \frac{\frac{m_{\text{dissolved ACTR or NCBD}}}{2} \cdot \text{mg}}{1.1 \text{ mL}} \cdot 0.9 \text{ mL}$$

$$\text{Percent Yield} = \frac{M_{\text{purified ACTR or NCBD}}}{M_{\text{crude ACTR or NCBD}}} \cdot 100$$

The amount of crude peptide used in the calculation was adjusted for the volume of the peptide that was actually injected onto the HPLC. Only 0.9 mL of the 1.1 mL volume of the peptide was dissolved in were actually injected. The yields were very low and ranged for NCBD from 2.57 to 4.03%, and for ACTR from 3.89 to 6.62%. The solvent that was chosen to dissolve the crude peptide may in part explain these low yields. Indeed, the 1% TFA solution might not have been strong enough to fully dissolve the peptide, and so the pellet that was observed in the bottom of the tube after centrifugation likely contained not

only impurities like resins but also insolubilized peptide. One possible option would have been to use formic acid instead, with the caution to not allow it to come into contact with the peptide for more than 30 minutes in order to avoid formylation of the methionine residues.

### **Circular Dichroism Spectroscopy**

CD spectra were collected with an Olis RSM 1000 circular dichroism spectrophotometer using the Olis spectralWorks software (Bogart, GA). Wavelength scans from 260 to 190 nm were performed with 1 nm steps and integration time per data point determined automatically as a function of PMT voltage. All CD experiments were conducted at 22°C.

Two vials of lyophilized peptide were combined and dissolved into 0.75 mL of 2X PBS so that final concentrations in 1X PBS would be in the range of 0.2-0.4 mg/mL, the concentration needed to elicit a CD signal. Exact peptide concentrations were determined by HPLC using a standard curve generated from amino acid analysis (AlbioTech, Richmond, VA). Two solutions were then made: one was a solution without any osmolyte, which consisted of 0.35 mL of the peptide 2X PBS solution + 0.35 mL of H<sub>2</sub>O. The second solution was the highest concentration osmolyte solution, which had 0.35 mL peptide 2X PBS solution + 0.35 mL 80 wt% osmolyte in water (for EG, TEG, and PEG 400). Serial dilutions were made from these two solutions. To measure CD of each sample, 320 µL was put into a 1mm quartz cell. The weight percents that were tested for EG, TEG, and PEG400 were 0%, 5%, 10%, 17.5%, 25%, 30%, and 40%. For xylitol, the molalities that were tested included

1.0m, 1.75m 2.5m, 3.0m, and 4.0m. For betaine, the molalities were 0m, 0.25m, 0.5m, 1.25m, 1.5m, and 2.0m.

First, background scans of all the solutions of varying osmolyte concentration in percent by weight were taken so their CD signals could be subtracted from the corresponding CD signals of the solutions that included the peptides.

Next, a 1:1 molar ratio of NCBD:ACTR in 2X PBS was prepared from the individual peptide solutions to yield NCBD/ACTR dimer. To do this, first the molarities of a sample each of ACTR and of NCBD were calculated.

$$\text{For ACTR: } 0.2255 \text{ mg} \cdot \frac{\text{mol}}{7989\text{g}} = 1.534 \cdot 10^{-8} \text{ mol}$$

$$\text{For NCBD: } 0.077698 \text{ mg} \cdot \frac{\text{mol}}{6545\text{g}} = 1.187 \cdot 10^{-8} \text{ mol}$$

The tubes containing the ACTR and the NCBD were each resuspended in 0.360 mL of 2X PBS. Since the two molecules bind in a 1:1 ratio, NCBD in this case is kind of like the limiting reagent in a chemical reaction. To obtain  $1.187 \cdot 10^{-8} \text{ mol}$  of ACTR to react with an equivalent amount of NCBD,  $1.187 \cdot 10^{-8} \cdot \frac{\text{L}}{4.261 \cdot 10^{-5} \text{ mol}} \cdot \frac{10^3 \text{ mL}}{1 \text{ L}} = 0.279 \text{ mL}$

Then 0.279 mL of the ACTR solution made as described above were added into the tube containing the 0.36 mL of NCBD to produce a solution that was 0.639 mL in volume. This 0.630 mL of solution was diluted with 0.111 mL of 2X PBS to get 0.750 mL of peptide-containing solution that was needed for the dilution scheme.

The concentration of the solution was calculated. At this point, the solution contained  $1.187 \cdot 10^{-8} \text{ mol}$  of the ACTR/NCBD complex. The molar mass of the complex is

$14534 \frac{g}{mol}$ . So the mass of complex present in solution was  $1.187 \cdot 10^{-8} mol \cdot \frac{14534g}{mol} = 0.172519 mg$

The concentration of the solution containing NCBD/ACTR was  $\frac{0.172519 mg}{0.750 mL} = 0.230 mg mL^{-1}$ , which was in the 0.2-0.4 mg mL<sup>-1</sup> range required in the experiments in which ACTR and NCBD were studied separately.

Samples were loaded into a 1 mm pathlength quartz cuvette. The CD spectrum of the 2X PBS buffer solution was subtracted from the sample spectra. These background subtracted CD spectra were converted to units of mean residue ellipticity,  $[\theta] = [deg cm^2 dmol^{-1}]$ . The additive molecular mass and number of residues for NCBD and ACTR were used to normalize the NCBD/ACTR signal to  $[\theta]$ .

The mean residue ellipticity adjustment of the CD signal is based on the fact that the light source in the spectrometer traces out an ellipse, and as a result the light is elliptically polarized (9). Elliptically polarized light can be resolved into left and right-circularly polarized components, which are absorbed unequally through an optically active or chiroptical absorption band (9). Thus CD can be expressed in two ways, one of which is to measure the ellipticity. The ratio of the minor axis to the major axis of the ellipse defines the tangent of ellipticity (9). Because the difference in absorbances of the two components is only a small fraction of the average absorbance, the ellipse is extremely elongated. Thus, tangent (ellipticity) can be approximated as ellipticity. For macromolecules the CD data are expressed in mean residue ellipticities  $[\theta]$ , rather than molar absorption coefficients,  $\epsilon_L - \epsilon_R$ . These quantities are independent of the relative molecular mass (or molecular

weight) of the biopolymer, unless the conformation of, say, oligomers of  $\alpha$ -helical polypeptides depends on the degree of polymerization. Traditionally, CD can be calculated from  $\epsilon_L - \epsilon_R = (A_L - A_R)/lC$ , And

$$[\theta] = \frac{M_0}{100} \phi(l'c)$$

$$\text{or } [\theta] = 100\phi(lm),$$

where the As are absorbances, l and l' are the light path in cm and dm, respectively,  $M_0$  is the mean residue weight (about 115 for proteins), c is the concentration in g/cm<sup>3</sup>, and m is the mean residue molar concentration. The two quantities are related by a conversion factor of 3300 (Yang, 1969):  $[\theta] = 3000(\epsilon_L - \epsilon_R)$ . The dimensions of  $[\theta]$  and  $(\epsilon_L - \epsilon_R)$  are deg cm<sup>2</sup> dmole<sup>-1</sup> and M<sup>-1</sup>cm<sup>-1</sup>, respectively (9).

A plot of  $[\theta]$  (deg cm<sup>2</sup> dmole<sup>-1</sup>) vs. wavelength (nm) was made for all the different osmolytes. Then, a plot of  $[\theta]_{222}$  (deg cm<sup>2</sup> dmole<sup>-1</sup>) vs. OsmP (mmol kg<sup>-1</sup>) was also made for all the different osmolytes. It was necessary to find the fraction of helix in the peptide under each osmolyte condition. The equation for fraction of  $\alpha$ -helix is  $f_h = \frac{-[\theta]_{222} + 3000}{3900}$

(10). However in NCBD and ACTR, not all of the residues can adopt an  $\alpha$ -helix structure, so

the following modified equation is used:  $f_h = \frac{MRE_{measured} - MRE_{min}}{MRE_{max} - MRE_{min}}$ . As calculated from the

NMR structure of NCBD/ACTR complex (1KBH.pbd), in the bound state NCBD has 61%  $\alpha$ -helix and ACTR has 38%  $\alpha$ -helix. Thus the maximum helical content for each will yield a CD signal, MRE of the following:

$$[\theta]_{222}(NCBD, \text{max helix}) = -(0.61 * 39000 - 3000) = 20790$$

$$[\theta]_{222}(ACTR, \text{max helix}) = -(0.38 * 39000 - 3000) = 11820$$



From  $[\theta]_{222}$  deg cm<sup>2</sup> dmole<sup>-1</sup>, the fraction of  $\alpha$ -helix for each peptide and osmolyte condition was calculated according to  $f_h = \frac{[\theta]_{222}-3000}{-11820-3000}$  for ACTR and  $f_h = \frac{[\theta]_{222}-3000}{-20790-3000}$  for NCBD, and for ACTR:NCBD  $f_h = \frac{[\theta]_{222}-3000}{-15720-3000}$ .

Next, the natural logarithm of the equilibrium constant for folding for each peptide in each osmolyte condition was calculated, as in  $\ln K_{eq} = \ln \left( \frac{f_h}{1-f_h} \right)$ , and substituted into the expression for  $\Delta\Delta G$  for each peptide/osmolyte condition, which is  $\Delta\Delta G = -8.3145 * 293.15 * \ln K_{eq} - \ln K_{eq,0}$ .

Using  $\Delta\Delta G$ , the number of waters released from upon folding was calculated according to  $\Delta N_w = -\frac{\partial\Delta\Delta G}{\partial\mu_w} = \frac{55.6}{RT} \frac{\partial\Delta\Delta G}{\partial[Osm]}$  where  $\mu_w$  is the chemical potential of water, 55.6 is the number of moles of water in 1 kg, and [Osm] represents the solute osmolal concentration, a measure of water chemical potential (8).

Vapor pressure osmometry was used to measure the vapor pressure of the samples, however, for many of the osmolytes, including TEG and PEG 400, the instrument wouldn't measure past 3500 mmol/kg, and the vapor pressure of the 30 wt% TEG exceeded the capacity of the osmometer, so calculated osmotic pressures were used instead.

## RESULTS AND DISCUSSION

### Describing ACTR and NCBD

The peptides that were used in this study were ACTR (activator for thyroid hormone and retinoid receptor), 74 residues, and NCBD (nuclear receptor coactivator binding domain), 59 residues. Table 1 shows the percentages of secondary structure elements in ACTR and NCBD in the bound and unbound states. NCBD has twice the percentage of  $\alpha$ -helix in the unbound state as does ACTR. The actual CD signal at 222 is more negative for the complex than the prediction represented by the gray curve in Figure 1. When these two peptides bind, an increase occurs in the total  $\alpha$ -helical content over what would be expected in the absence of structural changes, which indicates that they experience secondary structure changes upon binding.

Table 1. This chart displays the percentages of the different types of secondary structure elements in the unbound and bound forms of NCBD and ACTR, as well as the percentages found in the NCBD/ACTR complex as calculated from CD and NMR data.

structure	NCBD		ACTR		NCBD/ACTR	
	unbound*	bound†	unbound*	bound†	CD*	NMR†
$\alpha$ -helix (%)	36 (40)	61	18 (19)	38	36 (38)	48
$\beta$ -sheet (%)	1	0	5	0	2	0
turn (%)	15	31	11	5	14	17
unfolded (%)	48	9	66	57	49	35

\* Calculated from fitting the CD data using the CONTINLL method with the CLUSTER basis set. Values in parentheses are %  $\alpha$ -helix calculated using the equation:

$$\% \alpha\text{-helix} = (-[\theta]_{222\text{ nm}} + 3000)/39000.$$

† Calculated from the first NMR structure in 1KBH.pdb and assuming the first 27 N-terminal ACTR peptide residues are unfolded.

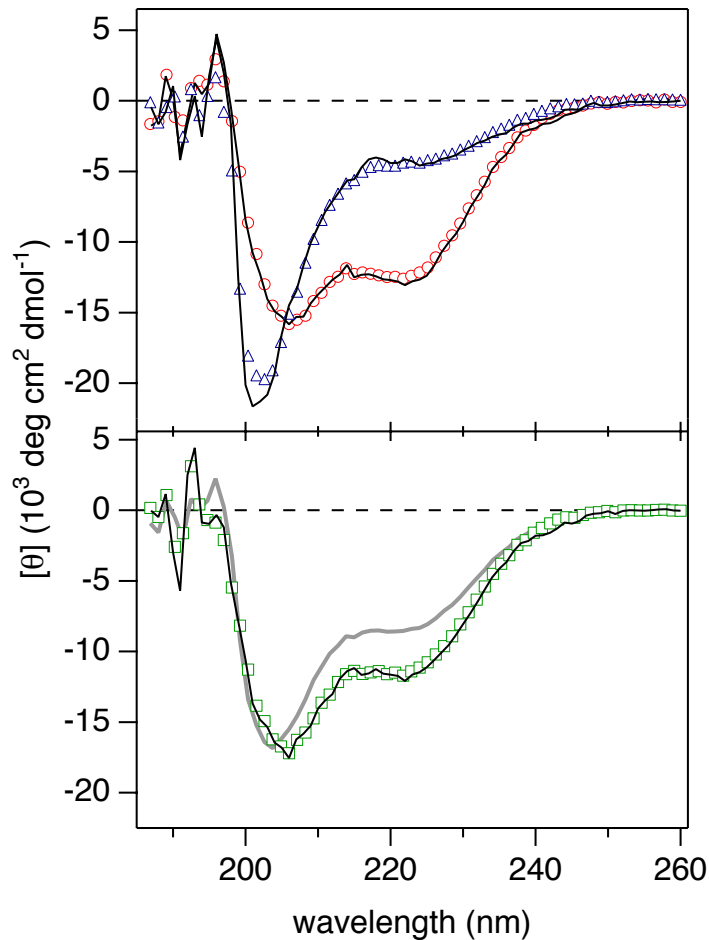


Figure 1. Circular dichroism (CD) spectra of NCBD (○), ACTR (△), and NCBD/ACTR complex (□) with the corresponding secondary structure fits (solid black lines). NCBD (58.1 mM), ACTR (51.3 mM), and NCBD/ACTR (27.2 mM) were in 1X TBS buffer. The calculated CD spectrum for a 50:50 mixture of NCBD and ACTR from the respective individual spectra, which assumes no interactions or changes in their secondary structure, is shown (solid grey line) for comparison. The more pronounced negative peak around 222nm for the experimentally measured NCBD/ACTR is indicative of increased  $\alpha$ -helical content from the binding interaction between NCBD and ACTR.

### CD on NCBD, ACTR, and NCBD/ACTR with Osmolytes

As the weight percent of PEG400 increases,  $[\theta]_{222}$  becomes more negative while  $[\theta]_{200}$  becomes more positive. The signal at 222nm is an indicator of  $\alpha$ -helix while the signal at 200 nm is an indicator of random coil. This effect is seen for ACTR (Fig.2), NCBD (Fig. 3), and NCBD/ACTR (Fig. 4), indicating that the content of  $\alpha$ -helical structures present

in the protein in solutions containing higher concentrations of osmolyte than in solutions containing less. At 40% (v/v) of PEG 400, the percentage of  $\alpha$ -helix is significantly closer to the percentage of  $\alpha$ -helix in the folded protein than at 0%. The percentage of  $\alpha$ -helix in ACTR relative to the percentage in the bound state increases from 19% to 71% between 0% (v/v) of PEG 400 and 40%. For NCBD, the increase is from 40% to 85%, and for the complex, from 100% to 114%. Thus, increased osmolyte concentration effected a change in secondary structure in NCBD and ACTR separately and when bound. It is significant that an increase in folding was observed in the complex, in which ACTR and NCBD are represented as already having reached 100% of their  $\alpha$ -helical content.

Figures 5, 6, and 7 display the relationship between the signal at 222nm and the osmotic pressure in osmolality for ACTR, NCBD, and NCBD/ACTR in PEG400. Osmotic pressure is proportional to the weight percent, and as expected, the signal becomes more negative as the osmotic pressure increases, indicating that more  $\alpha$ -helix is present in the peptide at higher osmotic pressures. Figures 5, 6, and 7 are like taking a slice through figures 2, 3, and 4 at the wavelength of 222 nm. A similar slice taken at 200 nm would show an increase in the in the CD signal with weight percent, indicating a corresponding decrease in the random coil content of the peptide.

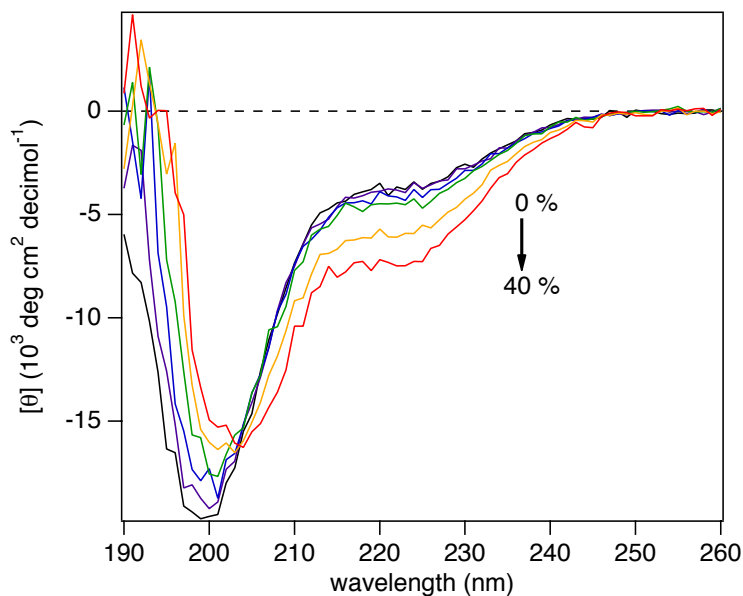


Figure 2. ACTR in PEG 400. Mean residue ellipticity as a function of the wavelength of circularly polarized light for solutions containing ACTR and PEG 400. CD indicates that ACTR initially contains 18%  $\alpha$ -helix without added osmolyte. As the osmolyte concentration in solution increases from 0-40% (v/v), the CD signal at 222 nm becomes increasingly negative, indicating that the peptide adopts more  $\alpha$ -helical structure. Calculations from the NMR structure show that ACTR contains 38%  $\alpha$ -helix in the bound state. At 40% (v/v) of PEG 400, ACTR achieved 71% of the bound state  $\alpha$ -helicity.

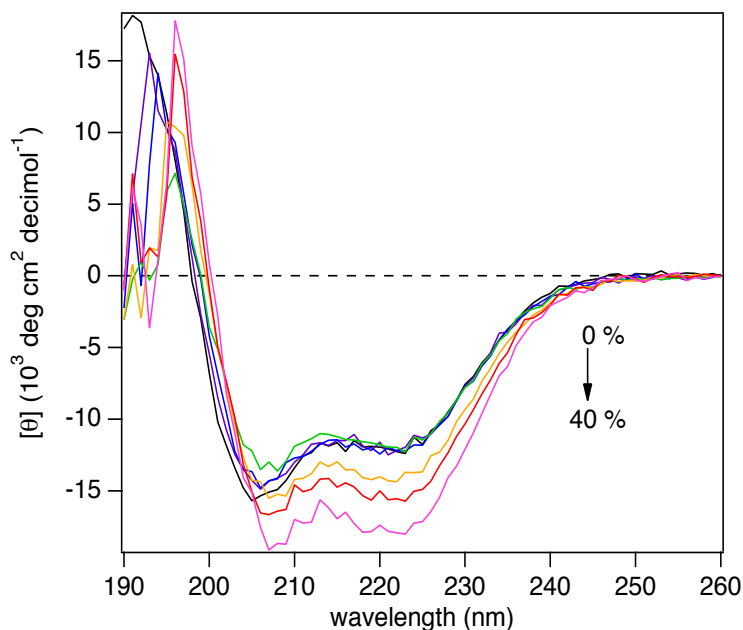


Figure 3. NCBD in PEG 400. Mean residue ellipticity as a function of the wavelength of circularly polarized light for solutions containing NCBD and PEG 400. CD indicates that NCBD initially contains 40%  $\alpha$ -helix without added osmolyte. As the osmolyte concentration in solution increases from 0-40% (v/v), the CD signal at 222 nm becomes increasingly negative, indicating that the peptide adopts more  $\alpha$ -helical structure. Calculations from the NMR structure show that NCBD contains 61%  $\alpha$ -helix in the bound state. At 40% (v/v) of PEG 400, NCBD achieved 88% of the bound state  $\alpha$ -helicity.

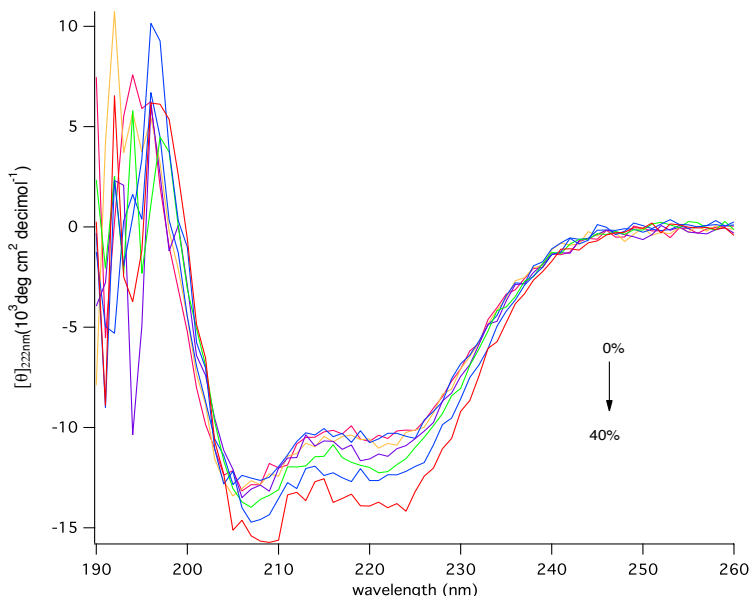


Figure 4. NCBD/ACTR in PEG400. Mean residue ellipticity as a function of the wavelength of circularly polarized light for solutions containing NCBD/ACTR and PEG 400. CD indicates that NCBD/ACTR initially contains 38%  $\alpha$ -helix without added osmolyte. As the osmolyte concentration in solution increases from 0-40% (v/v), the CD signal at 222 nm becomes increasingly negative, indicating that the peptide adopts more  $\alpha$ -helical structure. Calculations from the NMR structure show that NCBD/ACTR contains 48%  $\alpha$ -helix in the bound state. At 40% (v/v) of TEG, NCBD/ACTR achieved 115% of the bound state  $\alpha$ -helicity.

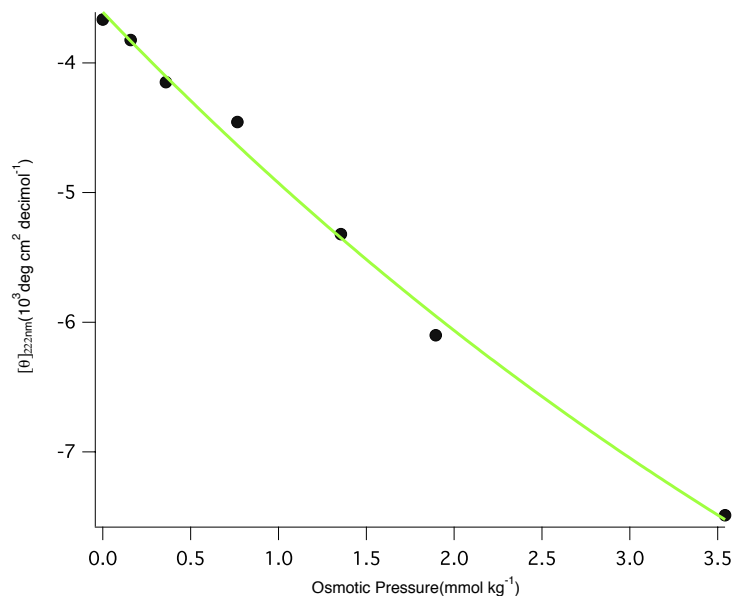


Figure 5. This graph shows mean residue ellipticity  $[\theta]_{222}$  (deg cm<sup>2</sup> dmole<sup>-1</sup>) as a function of osmotic pressure for ACTR in PEG400. The increasing negativity of the signal at 222 nm indicates that there is more  $\alpha$ -helix present as the osmotic pressure increases.

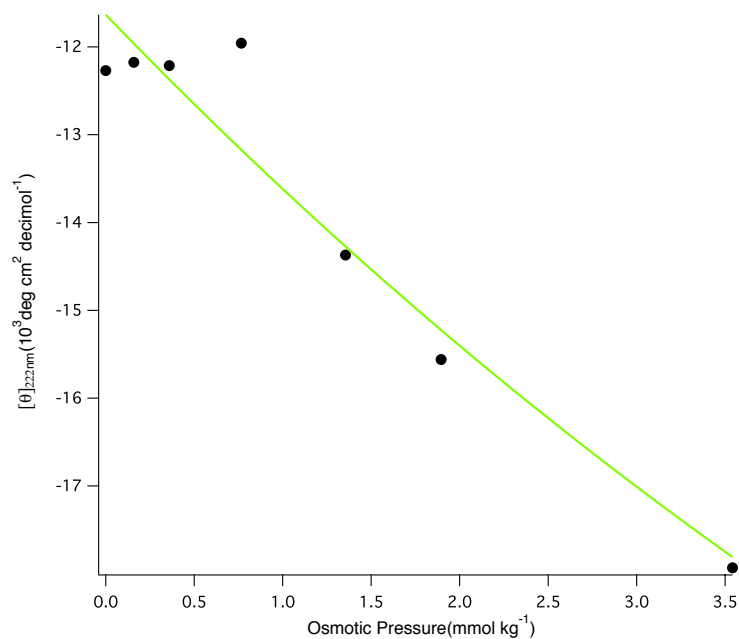


Figure 6. This graph shows mean residue ellipticity  $[\theta]_{222}$  (deg cm<sup>2</sup> dmole<sup>-1</sup>) as a function of osmotic pressure for NCBD in PEG400. The increasing negativity of the signal at 222 nm indicates that there is more  $\alpha$ -helix present as the osmotic pressure increases.

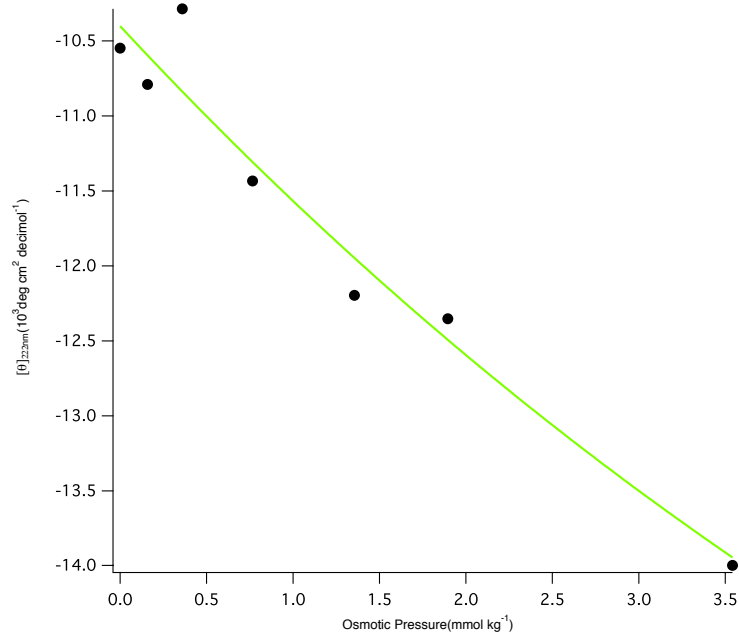


Figure 7. This graph shows mean residue ellipticity  $[\theta]_{222}$  ( $\text{deg cm}^2 \text{ dmole}^{-1}$ ) as a function of osmotic pressure for NCBD/ACTR in PEG400. The increasing negativity of the signal at 222 nm indicates that there is more  $\alpha$ -helix present as the osmotic pressure increases.

### Thermodynamics of osmolyte-induced folding of ACTR, NCBD, and NCBD/ACTR

The osmotic stress technique used to induce folding is relevant to the *in vivo* environment in which IDPs are found. The intracellular environment is crowded with salts, sugars, amino acids, macromolecules like DNA and proteins. Thus, this intracellular environment differs from the dilute aqueous conditions that are usually used in enzymatic, molecular recognition, and protein biomolecular assembly reactions (11). Protein stability and activity vary according to variances including hydration levels, ion concentrations and pH, among other solvent conditions. Cells strive to maintain proteins' shapes and functions in the face of environmental stress by accumulating osmolytes which most commonly include polyols, amino acids, and combinations of methylamines with urea (12). These protective osmolytes shift the thermodynamic equilibrium of folding toward more compact native states by being preferentially excluded from the protein-water interfaces (12).



One explanation for how osmolytes shift the folding equilibrium is the mechanism of molecular crowding, according to which, the restriction of protein conformations to allow larger free volume for the added osmolytes destabilizes the unfolded state with respect to the folded, functional conformation (12). In these CD experiments, the response to osmotic stress was considered to be related to the hydration of the molecule. Osmotic pressure causes some of the solute-excluding waters to be released from the surface of the protein. Information about the nature of physical interactions between macromolecules can be gained from studying the dependence of  $\Delta N_w$  on the chemical nature of the osmolytes and the macromolecule's surface (11). Many salts and polar solutes are preferentially excluded from hydrophobic surfaces, and the number of preferentially bound waters depends on the chemical nature and size of the osmolyte as well as on the nature of the macromolecular surface (11). Changes in the number of preferentially bound water accompanying folding and binding reactions were measured from the change in  $\Delta G$ , the free energy of binding with osmolyte concentration. For a constant difference in the number of included waters, it was expected that  $\Delta\Delta G$  would vary linearly with solute osmotic pressure. (11).

For ACTR, NCBD, and NCBD/ACTR the  $\Delta\Delta G$  becomes more negative as the osmotic pressure increases. As  $\Delta\Delta G$  is a measure of stability (8), the negative increase in its value indicates that the peptide is more stable at higher osmotic pressures. Larger sized osmolytes cause a leftward shift of the lines towards more negative  $\Delta\Delta G$  values, indicating that larger osmolytes are better at stabilizing peptides. Figure 8 displays these relationships for ACTR and NCBD/ACTR. Likewise,  $\Delta N_w$  scales with the osmolyte molecular volume,  $V_m$ . It was expected that,  $\Delta N_w$  would vary exponentially with partial molar volume, and level off at the point where  $\Delta N_w$  continues to increase even as molecular volume keeps

increasing. For small changes in molecular volume, it was expected to vary linearly. The results are fit with an exponential curve, and the results do confirm the expected general trend of an increase in the number of waters released as the molecular volume of the osmolyte increases. It is notable that at the highest weight percent of PEG400, the complex released 4 water molecules, comparable with the effect of EG on ACTR.

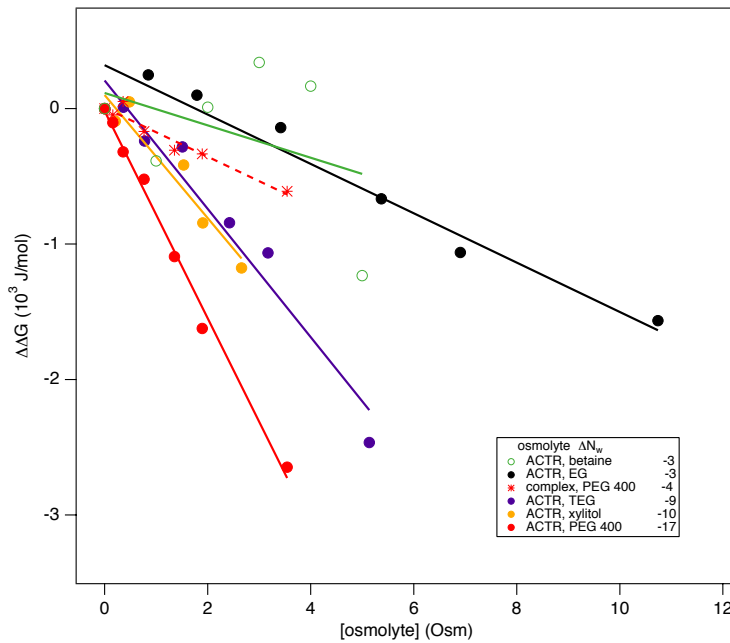


Fig 8. The free energy of folding  $\Delta G = -RT \ln K_{eq}$ , where  $K_{eq} = \frac{[\text{folded peptide}]}{[\text{unfolded peptide}]}$ .  $\Delta\Delta G$  is defined as the difference in the energy of the peptide's folding between the osmolyte solution and water:  $\Delta\Delta G = \Delta G(\text{osmolyte}) - \Delta G(\text{water})$ . The numbers of waters released from ACTR,  $\Delta N_w$ , depends on the osmolyte used.  $\Delta N_w$  scales with osmolyte molecular volume.  $\Delta N_w = -\frac{\partial \Delta\Delta G}{\partial \mu_w} = \frac{55.6}{RT} \frac{\partial \Delta\Delta G}{\partial [Osm]}$ .

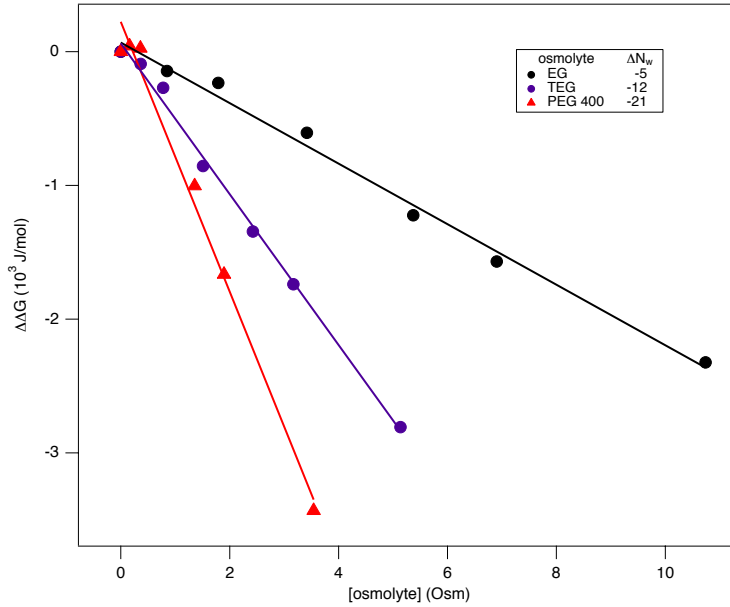


Figure 9. Similar to ACTR,  $\Delta\Delta G = \Delta G(\text{osmolyte}) - \Delta G(\text{water})$ .  $\Delta N_w$ , depends on the osmolyte used and scales with osmolyte molecular volume.

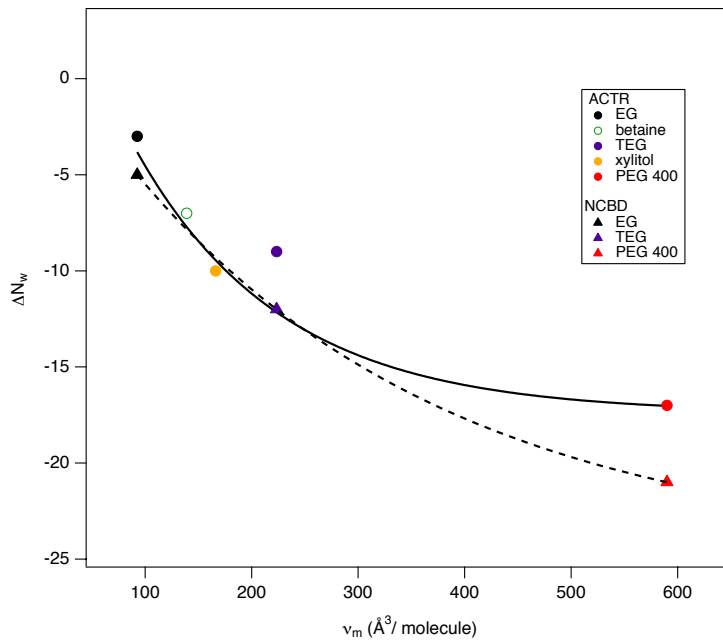


Figure 10. The number of cosolute-excluding waters released upon folding,  $\Delta N_w$  vs.  $V_m$  is shown for both ACTR and NCBD.  $\Delta N_w$ , depends on the osmolyte used and scales with osmolyte molecular volume.

Although this experiment provided insightful results, there are some limitations. Due to time limitations the effects of varying pH were not fully explored and all CD experiments were carried out in solutions of neutral pH. In future experiments, the CD measurements could be repeated in solutions of varying pH. In fact, one experiment had to be repeated because using TEG that had acidified due to improper storage produced an interesting outlier in the  $\Delta N_w$  data; it was calculated that 54 waters had been released upon folding in this experiment. Thus, changing the pH in addition to varying the osmotic pressure has a strong effect on the secondary structure of these peptides. Another limitation was the fact that the yields of the purified peptide were low. As a result, usually the product from two purification rounds had to be combined to obtain one set of CD measurements. However these limitations did not impact the results drastically, as the results generally agree with the literature.

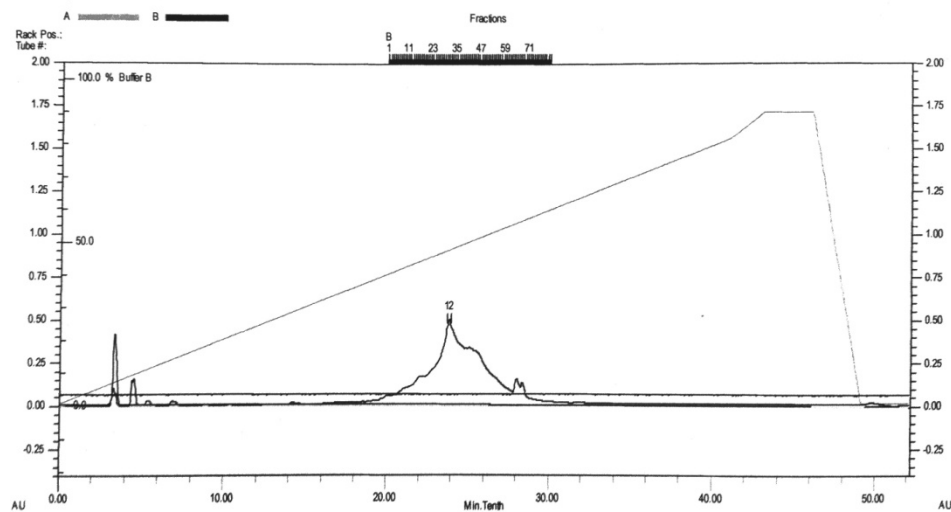


Figure 11. Purification run report. The fractions that were collected corresponded to the tip of the peak. However, a great portion of the area under the curve corresponds to impurities that likely occurred as a result of a messy peptide synthesis process.

## CONCLUSIONS

In order to accomplish their functions, most IDPs must undergo a disorder to order transition (3). A lack of intrinsic structure and the capacity to fold into an ordered conformation upon binding are advantageous because they confer upon IDPs high specificity coupled with low affinity, the ability to bind to several different targets (one to many signaling), and the capability to overcome steric restrictions (3).

Here it was found that osmotic stress caused the peptides to form more  $\alpha$ -helix structures and to release waters. The osmolyte molecular volume was proportional to the number of waters released. The osmolytes act by stabilizing the more compact, ordered conformations of the peptide and cause water molecules associated with the surface of the peptide to be released. Also, in most cases, the linear free energy change with osmolality includes the initial, zero osmolyte condition at  $\Delta\Delta G = 0$ . This is most easily observed for NCBD (Fig. 9) and indicates that the variation in secondary structure by the added osmolyte is due to osmotic effects and not by any preferential interactions between peptide and osmolyte. The one exception appears to be ACTR with EG, where repeat measurements confirmed the particular free energy variation observed. Here, it does seem that specific interactions by EG have altered the initial state of ACTR. Interestingly, EG does not have strong preferential interaction with NCBD. These subtle features of EG as an osmolyte for these IDPs deserve further study to fully understand EG action.

It is notable that even the complex was able to fold into a structure containing more  $\alpha$ -helix. Because the amino acids that have formed  $\alpha$ -helices in the bound form locked into their conformations, and thus cannot be induced to fold any further, the extra  $\alpha$ -helices that

formed were probably formed in the flexible tail of ACTR that was not bound to NCBD. This flexible tail is characterized as an additional N-terminal leucine-rich motif located between residues 1029 - 1,033 that does not participate in the hydrophobic core and remains disordered in the complex but rather is available to interact with other proteins (13). These hydrophobic leucine regions are significant because the molecular interface of the ACTR binding domains consist of clusters of hydrophobic residues separated by regions of polar amino acids rich in glycine, serine, proline, or glutamine. In ACTR and CBP, the interaction clusters form recognizable motifs with sequence patterns f-X-X-f-f, f-f-X-X-f or a combination of the two, where f is a bulky hydrophobic residue (13). Many of these hydrophobic motifs are rich in leucine, and they may function similarly to the L-X-X-L-L motifs that are responsible for the interactions between the p160 coactivators and the ligand-binding domains of nuclear receptors (13). However, while the L-X-X-L-L motifs form an isolated helix upon binding to a structured nuclear receptor ligand-binding domain, the disordered leucine-rich motifs in the ACTR and CBP binding domains undergo synergistic folding to form the binding interface (13).

There are two competing models that attempt to explain protein association reactions that involve binding and folding. One model is the induced fit model and the other is the conformational selection model (14). According to the induced fit model, the protein binds or associates with its ligand in a fully disordered state and then undergoes the conformational change (folding) (2,14). According to the conformational selection model, the bound-state conformation must exist in the free ensemble, previous to binding, and one partially prefolded ligand selectively associates with its partially prefolded binding partner. According to the conformational selection model, NCBD would adopt a conformation

similar to that in the complex with ACTR as it is associating with ACTR. NMR results confirm that conformational selection plays an important role in this interaction (2).

Literature results show that the conformation of NCBD in the complex with ACTR is present even in the absence of ACTR (2). Although they do not provide an answer to the debate between the induced fit model and the conformational selection model, these current experiments confirm the fact that NCBD and ACTR can be induced to fold in the absence of their binding partners and thus support the existence of structured NCBD molecules implies a conformational selection mechanism in which ACTR selectively binds to the fraction of the prefolded ACTR binding conformation of the NCBD ensemble (2). Disordered proteins and regions are required for transcriptional activators, signaling proteins, and regulatory proteins to function. However, they are also the sites of many chromosomal translocations result in disease (1). For example, translocations that fuse regions of CBP or p300 to segments of MOZ (the monocytic zinc-finger leukemia protein) or MLL are associated with leukemia in humans. Translocations within disordered regions leave the folded domains intact and produce fusion proteins with abnormal functions (2). In contrast, truncations or translocations of gene regions that include fully structured domains would lead to the production of a misfolded protein that would be degraded by the cellular machinery quickly enough to avoid producing disease (2).

The conformational flexibility of IDPs gives them the functional versatility needed to bind to and recognize a diverse array of other proteins, RNA, DNA, and other small ligands (15). However, disordered regions are also involved in the promotion of neurodegenerative protein folding diseases such as Parkinson's and Huntington's diseases and prion diseases (2). For example, human CBP, which contains a stretch of 19 Gln

residues, can potentially be sequestered in polyglutamine aggregates in the brains of patients with Huntington's disease and it has been proposed that the disease might result from interference with the normal transcriptional functions of CBP (2). Accordingly, it has been shown that the MET-16 peptide can fold into a stable monomeric  $\beta$ -hairpin conformation under some conditions, while under others it aggregates into amyloid fibers (16). In the presence of polyols and polyethylene glycols acting as excluded cosolutes, the monomeric  $\beta$ -hairpin conformation was stabilized with respect to the unfolded state (16). Just as this osmolyte study demonstrates, it was also found that stabilization free energy varied linearly with osmolyte concentration and grew with molecular volume (16). The reaction that was studied was related to amyloid aggregation, which is the form of protein self-oligomerization that is implicated in the formation of several neurodegenerative diseases (16). A recent survey found that human proteins that have numerous tracts of extremely low complexity and features characteristic of IDPs, most commonly consisting of Gln, Ser, or acidic residues, are often associated with neurological diseases and cancer (2). Thus, further study into structure-function relationship of IDPs may provide information about the mechanisms of neurodegenerative disease processes.

## **ACKNOWLEDGEMENTS**

I would like to thank all members of the Berthelie lab, including Valerie Berthelie, Erica Rowe, Chris Stanley, Penney McWilliams-Koeppen, Tatiana Perevozchikova, and Monique LeMieux, for their mentoring and their assistance with this project.

## **REFERENCES**

1. Dyson, H.J. and P.E. Wright. 2005. Intrinsically Unstructured Proteins and their Functions. *Nature Reviews*. USA. 6: 197-208.



2. Kjaergaard, M., Teilum, K., and F.M. Poulsen. 2010. Conformational selection in the molten globule states of the nuclear coactivator binding domain of CBP. *PNAS*. USA. 107(28): 12535-12540.
3. Uversky, V.N. 2002. Natively unfolded proteins: A point where biology waits for physics. *Protein Science*. USA. 11: 739-756.
4. Chen, H., Lin, R.J., Schlitz, R.L., Chakravarti, D. Nash, A., Nagy, L., Privalsky, M.L., Nakatani, Y., R.M. Evans. 1997. Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. *Cell*. 90: 569-580.
5. Xu, J., Liao, L., Ning, G., Yoshida-Komiya, H., Deng, C., and B.W. O'Malley. 2000. The steroid receptor coactivator SRC-3 (p/CIP/RAC3/AIB1/ACTR/TRAM-1) is required for normal growth, puberty, female reproductive function, and mammary gland development. *Proceedings of the National Academy of Sciences*. USA. 97(12): 6379-6384.
6. Waters, L., Yue, B., Veverka, V., Renshaw, P., Bramham, J., Matsuda, S., Frenkiel, T., Kelly, G., Muskett, F., Car, M., and D.M. Heery. 2006. Structural diversity in p/160 CREB-binding protein coactivator complexes. *Journal of Biological Chemistry*. USA. 281(21): 14787-14795.
7. Torchia, J., Rose, D.W., Inostroza, J., Kamei, Y., Westin, S., Glass, C.K., and M.G. Rosenfeld. 1997. The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. *Nature*. USA. 387: 677-684.
8. Politi, R. and D. Harries. 2010. Enthalpically driven peptide stabilization by protective osmolytes. *The Royal Society of Chemistry*. UK. 46: 6559-6451.
9. Venyaminov, S.Y., and J.T. Yang. Determination of Protein secondary structure. Fasman, G.D. Ed. 1996. *Circular dichroism and the conformational analysis of biomolecules*. 69-108.
10. <http://www.mrl.ucsb.edu/mrl/centralfacilities/polymer/cd.html>
11. Harries, D., Rau, D.C., and Parsegian, V.A. 2005. Solutes Probe Hydration in Specific Association of Cyclodextrin and Adamantane. *Journal of the American Chemical Society*. USA. 127: 2184-2190.
12. Gilman-Politi, R., and D. Harries. 2011. Unraveling the Molecular Mechanism of Enthalpy Driven Peptide Folding by Polyol Osmolytes. *Journal of Chemical Theory and Computation*. USA. 7: 3816-3828.
13. Demarest, S.J., M. Martinez-Yamout, J. Chung, H. Chen, W. Xu, H.J. Dyson, R.M. Evans, and P.E. Wright. 2002. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature*. 415:549-553.

14. Wright, P. E. and H. J. Dyson. 2009. Linking folding and binding. *Curr. Opin. Struct. Biol.* 19:31-38.
15. Fuxreiter, M., P. Tompa, I. Simon, V. N. Uversky, J. C. Hansen, and F. J. Asturias. 2008. Malleable machines take shape in eukaryotic transcriptional regulation. *Nat. Chem. Biol.* 4:728-737.
16. Sukenik, S., R. Politi, L. Ziserman, D. Danino, A. Friedler, and D. Harries. Crowding Alone Cannot Account for Cosolute Effect on Amyloid Aggregation. *PLoS ONE*. USA. 6(1): 1-11.