Nucleic Acids Research, 1997, Vol. 25, No. 15 **2967–2972**

The role of a basic amino acid cluster in target site selection and non-specific binding of bZIP peptides to DNA

Steven J. Metallo, David N. Paolella and Alanna Schepartz*

Department of Chemistry, PO Box 208107, Yale University, New Haven, CT 06520-8107, USA

Received May 6, 1997; Accepted June 2, 1997

ABSTRACT

The ability of a transcription factor to locate and bind its cognate DNA site in the presence of closely related sites and a vast array of non-specific DNA is crucial for cell survival. The CREB/ATF family of transcription factors is an important group of basic region leucine zipper (bZIP) proteins that display high affinity for the CRE site and low affinity for the closely related AP-1 site. Members of the CREB/ATF family share in common a cluster of basic amino acids at the N-terminus of their bZIP element. This basic cluster is necessary and sufficient to cause the CRE site to bend upon binding of a CREB/ATF protein. The possibility that DNA bending and CRE/AP-1 specificity were linked in CREB/ATF proteins was investigated using chimeric peptides derived from human CRE-BP1 (a member of the CREB/ATF family) and yeast GCN4, which lacks both a basic cluster and CRE/AP-1 specificity. Gain of function and loss of function experiments demonstrated that the basic cluster was not responsible for the CRE/AP-1 specificity displayed by all characterized CREB/ATF proteins. The basic cluster was, however, responsible for inducing very high affinity for non-specific DNA. It was further shown that basic cluster-containing peptides bind non-specific DNA in a random coil conformation. We postulate that the high non-specific DNA affinities of basic cluster-containing peptides result from cooperative electrostatic interactions with the phosphate backbone that do not require peptide organization.

INTRODUCTION

Eukaryotic transcription factors in the basic region–leucine zipper (1) (bZIP) family employ an extremely simple structural motif to accomplish the sequence-specific recognition of duplex DNA (2). The DNA binding activity of a bZIP protein is localized within 60 contiguous amino acids termed the bZIP element (3; Fig. 1). Each bZIP element contains three segments that each play a specific role in DNA recognition. A basic segment of ~20 amino acids near the N-terminus of the element contains residues that contact DNA directly (4–7), while a zipper segment of ~25 amino acids orchestrates the formation of a coiled coil dimer from two protein monomers (4–6,8). The basic and zipper segments are connected by a six residue spacer segment whose sequence is not conserved amongst members of the bZIP family (9).

Although bZIP proteins are structurally simple, they recognize a set of inverted half-sites of widely varying sequence. In addition to their ability to discriminate between target sites that differ in half-site sequence, bZIP proteins also discriminate between target sites of identical half-site sequence but different half-site spacing (half-site spacing specificity) (10). For example, bZIP proteins related to the oncogene products Fos and Jun (AP-1 family) prefer the pseudosymmetric 9 bp AP-1 target site (ATGACTCAT) (11,12), whereas those related to CREB and ATF-2 (CREB/ATF family) prefer the symmetric CRE target site (ATGACGTCAT) (13) in which the same inverted pair of half-sites is separated by 2 bp. The yeast bZIP protein GCN4 binds both sites with comparable affinity (14). In spite of a large body of research on bZIP–DNA interactions, the determinants of CRE/AP-1 selectivity displayed by CREB/ATF proteins are not well understood.

Many, if not all, CREB/ATF family members share in common the ability to bend their specific CRE target site (15–18). Because the CRE site bends intrinsically toward the major groove, interaction with a CREB/ATF protein results in an overall straightening of the DNA. It has been demonstrated that bending of the CRE site by the CREB/ATF protein CRE-BP1 requires two basic residues (the basic cluster) located at the N-terminus of the basic segment (18). Bending is proposed to result from neutralization by these two basic residues of two symmetry-related phosphate linkages on a single face of the DNA helix. The resulting asymmetric charge distribution bends the DNA spontaneously in the direction of the neutralized phosphates, as predicted by Mirzabekov and Rich (19) and demonstrated in model systems by Strauss and Maher (20). Because all members of the CREB/ATF bZIP family studied discriminate effectively between the CRE and AP-1 target sites and also bend DNA, we wondered whether CRE/AP-1 specificity and DNA bending were linked, i.e. whether DNA bending resulted in high CRE/AP-1 specificity. We demonstrate here that it does not. The presence of the basic cluster in native and chimeric bZIP peptides does not result in selective binding to the CRE site. The basic cluster does, however, have a dramatic effect on the ability of the bZIP peptide to distinguish specific DNA in the presence of competing non-specific sequences.

*To whom correspondence should be addressed. Tel: +1 203 432 5094; Fax: +1 203 432 6144; Email: alanna@milan.chem.yale.edu

MATERIALS AND METHODS

Peptides

The chimeric bZIP peptides used in these experiments (Fig. 1) were reported previously (10,18) and, with the exception of g5c, were purified as described. g5c was subjected to additional purification by reverse phase HPLC using a semipreparative Vydac C-18 column and a gradient of 98–20% water in acetonitrile containing 0.05% TFA over 40 min at a flow rate of 4 ml/min. The identities of all peptides were confirmed by amino acid analysis and either MALDI or electrospray mass spectrometry.

Electrophoretic mobility shift assay

The CRE and AP-1 affinities of peptides g5c, c5g and c10g were determined by use of an electrophoretic mobility shift assay employing binding buffer (1.4 mM KH_2PO_4, 4.3 mM Na_2HPO_4, 2.7 mM KCl, 137 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.1% NP-40, 400 µg/ml BSA and 5% glycerol, pH 7.4) and oligonucleotides, at a final concentration of <50 pM, containing the CRE or AP-1 site (CRE_{24} or $AP\text{-}1_{23}$) reported previously (10). The conditions and data analysis were as used previously to determine the specificity of the peptides ccc and ggg (10). c5g and c10g bound well to both CRE_{24} and $AP\text{-}1_{23}$ at 25°C, whereas g5c bound to both only at 4°C. Complexes of peptides with non-specific DNA were not stable under the described gel electrophoresis conditions; neither specific bands nor well shifts were detected. The affinities of the chimeric peptides (Fig. 1) for non-specific DNA were therefore measured by use of a competition electrophoretic mobility shift assay in which the fraction of $[^{32}P]CRE_{24}$ bound to a given peptide was monitored as a function of the concentration of either Non_{24} (AGTGGAGTAAGGCCT-ATCTCGTGC), calf thymus DNA (CTD) (Gibco BRL) or poly(dI-dC)·poly(dI-dC) (Pharmacia). Peptide concentrations were chosen such that maximal binding was achieved in the absence of competitor DNA. The concentrations used were as follows: ggg, 9 nM; ggc, 20 nM; gcc, 15 nM; c5g, 30 nM; c10g, 60 nM; ccc, 90 nM; ccg, 60 nM; cgg, 120 nM; g5c, 60 nM. Reactions were incubated at 25°C in binding buffer. Addition of 10 mM $MgCl_2$ to the binding buffer did not affect the affinity of c5g for Non_{24} (data not shown). Hence, there was no effect on specificity when a divalent cation was present.

Data analysis

Direct electrophoretic mobility shift data were analyzed as described previously (10). This assay monitors the equilibrium shown in Scheme 1 in which two unfolded bZIP peptide monomers (U) are converted into a single, dimeric DNA complex (A_2O).

$$2U + O \underset{}{\overset{K_{app}}{\leftrightarrows}} A_2O \qquad \text{(Scheme 1)}$$

Regardless of the pathway used to assemble the bZIP dimer·DNA complex (10), the fraction of DNA bound to peptide (θ) can be represented by the following equation where A_{tot} represents the total concentration of peptide monomer:

$$\theta = 1/(1 + K_{app}/[A_{tot}]^2) \qquad \mathbf{1}$$

A fit of the experimental binding data to equation **1** yields a value for the apparent dissociation constant (K_{app}) that accounts for binding of two bZIP monomers to DNA, with concomitant formation of a coiled coil, in a manner that is independent of binding pathway.

For competition electrophoretic mobility shift assays, the dissociation constants for formation of peptide·CRE_{24} and peptide·Non_{24} complexes, K_1 and K_2 respectively, may be represented by the following equilibrium and mass action expressions:

$$K_1 = ([R]^2[O])/[R_2O] \qquad \mathbf{2}$$
$$K_2 = ([R]^2[D])/[R_2D] \qquad \mathbf{3}$$
$$[D] = [D_t] - [R_2D] \qquad \mathbf{4}$$
$$[O] = [O_t] - [R_2O] \qquad \mathbf{5}$$
$$2[R_2D] = [R_t] - [R] - 2[R_2] - 2[R_2O] \qquad \mathbf{6}$$

where R, O and D represent the concentrations of peptide monomer, specific DNA (CRE_{24}) and non-specific DNA (Non_{24}) respectively. Combining equations **2** and **3** yields:

$$K_2 = (K_1[R_2O][D])/([O][R_2D]) \qquad \mathbf{7}$$

Under conditions where non-specific DNA is in excess ($[D_t] >> [R_2D]$) and the concentrations of peptide dimer and specific DNA are small ($[R_2]\rightarrow 0$ and $[R_2O]\rightarrow 0$), then equation **7** may be reduced and rearranged to:

$$K_2[R_t] = (2K_1[D_t][R_2O]/[O]) + K_2[R] \qquad \mathbf{8}$$

Substituting equations **5** and **2** into equation **8** leads to:

$$K_2[R_t] - K_2\{K_1[R_2O]/([O_t] - [R_2O])\}^{1/2} = (2K_1[D_t][R_2O])/([O_t] - [R_2O]) \qquad \mathbf{9}$$

Squaring the equations leads to equation **10**:

$$(K_2^2K_1^2[R_2O])/([O_t] - [R_2O]) = 4K_1^2[D_t]^2\{[R_2O]/([O_t] - [R_2O])\}^2 - 2K_2[R_t]K_1[D_t]\{[R_2O]/([O_t] - [R_2O])\} + K_2^2[R_t]^2 \qquad \mathbf{10}$$

Rearranging equation **10** and setting $[R_2O]/([O_t] - [R_2O]) = \theta/(1 - \theta)$, $a = 4K_1^2[D_t]^2$, $b = 2K_2[R_t]K_1[D_t] + K_2^2K_1$ and $c = K_2^2[R_t]^2$ yields expression **11**, which can be rearranged to **12**:

$$a(\theta/1 - \theta)^2 - b(\theta/1 - \theta) + c = 0 \qquad \mathbf{11}$$
$$(a + b + c)\theta^2 = (-b - 2c)\theta + c = 0 \qquad \mathbf{12}$$

Equation **12** may be converted using the quadratic formula into an expression **13** that yields the fraction of labeled DNA bound to peptide (θ) as a function of the concentration of competitor DNA. This expression was used in KaleidaGraph (v.3.0.2; Abelbeck Software, Reading, PA) to fit the competition data.

$$\theta = \langle K_1K_2^2 + 2K_1K_2[D_t][R_t] + K_2^2[R_t]^2 + \{K_1^2K_2^2(K_2^2 + 4K_2[D_t][R_t] - 12[D_t]^2[R_t]^2)\}^{1/2}\rangle \div \langle 2(K_1K_2^2 + 4K_1^2K_2^2 + 2K_1K_2[D_t][R_t] + K_2^2[R_t]^2)\rangle \qquad \mathbf{13}$$

Circular dichroism and fluorescence

Circular dichroism and fluorescence experiments were performed in CD buffer (1.4 mM KH_2PO_4, 4.3 mM Na_2HPO_4, 2.7 mM KCl, 137 mM NaCl, 1 mM EDTA and 1 mM DTT, pH 7.4), which contains the same salt concentration as binding buffer. Circular dichroism experiments were conducted at 25°C on an Aviv 62DS spectrometer using a 1 mm path length cell. Fluorescence experiments were conducted at 25°C on a SLM Aminco 4800s spectrofluorimeter using a 3 × 3 mm cell. The intrinsic fluorescence of the unique tryptophan residue in the spacer region of ccc was monitored between 300 and 400 nm upon excitation at 292 nm, a wavelength chosen to reduce inner filter effects from the DNA.

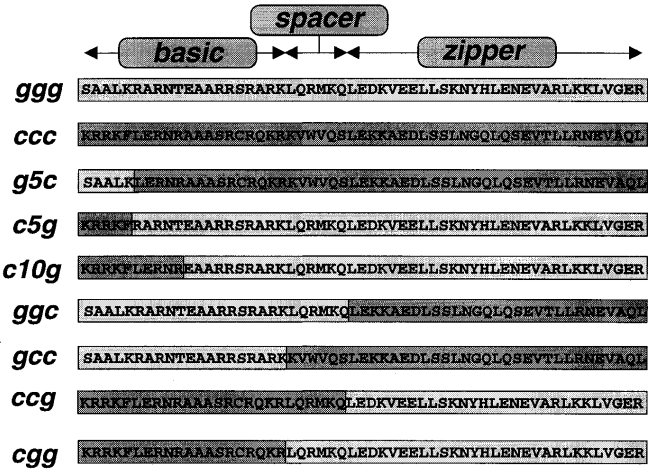

Figure 1. Chimeric bZIP peptides were assigned names depending on whether the sequence contained within the basic, spacer and zipper segments was derived from GCN4 (g) or CRE-BP1 (c). For example, ccg contained the basic and spacer segments from CRE-BP1 fused to the zipper segment from GCN4 and c5g contained five residues from the CRE-BP1 basic segment fused to the remaining basic, spacer and zipper of GCN4.

RESULTS

Basic cluster not involved in CRE/AP-1 selectivity

We made use of a series of chimeric bZIP element peptides containing sequence from CRE-BP1 and GCN4 (Fig. 1) to examine the role of the CRE-BP1 basic cluster in CRE/AP-1 selectivity. Peptide c5g contained the N-terminal basic cluster (KRRKF) of CRE-BP1 in the context of sequence from GCN4 (21); g5c contained the corresponding five residues from GCN4 (SAALK) in the context of sequence from CRE-BP1 (22). The GCN4 and CRE-BP1 bZIP element peptides are referred to as ggg and ccc respectively.

First we compared the CRE/AP-1 specificities of peptides ggg and c5g to determine whether the basic cluster of CRE-BP1 was sufficient to induce preferential recognition of the CRE site. The free energy of the c5g·CRE_{24} complex, ΔG_{cre} = –22.6 kcal/mol, was essentially equivalent to the free energy of the c5g·AP-1_{23} complex, $\Delta G_{ap\text{-}1}$ = –22.5 kcal/mol. The free energy of the ggg·CRE_{24} complex, ΔG_{cre} = –23.7 kcal/mol, was approximately equal to the free energy of the ggg·AP-1_{23} complex, $\Delta G_{ap\text{-}1}$ = –24.1 kcal/mol (10). Thus both ggg and c5g bound equally well to oligonucleotides containing the CRE and AP-1 sites CRE_{24} and AP-1_{23} (Fig. 2). The peptide c10g, in which 10 N-terminal residues of ggg were replaced with the corresponding sequence from CRE-BP1 (KRRKFLERNR), also exhibited little selectivity for one target site over the other; the free energy of the c10g·CRE_{24} complex, ΔG_{cre} = –20.6 kcal/mol, was almost identical to the free energy of the c10g·AP-1_{23} complex, $\Delta G_{ap\text{-}1}$ = –20.5 kcal/mol. These results indicate that the CRE-BP1 basic cluster is not sufficient to encode for preferential recognition of the CRE site.

Next we compared the CRE/AP-1 specificities of peptides ccc and g5c to determine whether the basic cluster was necessary for specific recognition of the CRE site by CRE-BP1. Both g5c and ccc bound with higher affinity to CRE_{24} than to AP-1_{23} (Fig. 2). At 4°C the free energy of the g5c·CRE_{24} complex, ΔG_{cre} = –19.7 kcal/mol, was 2.9 kcal/mol more favorable than the free energy of the g5c·AP-1_{23} complex, $\Delta G_{ap\text{-}1}$ = –16.8 kcal/mol. At the same temperature the free energy of the ccc·CRE_{24} complex, ΔG_{cre} = –20.9 kcal/mol, was 2.3 kcal/mol more favorable than the free energy of the ccc·AP-1_{23} complex, $\Delta G_{ap\text{-}1}$ = –18.6 kcal/mol (10). The absence of the basic cluster had no effect on the CRE/AP-1 selectivity of the intrinsically selective peptide ccc. Thus, the basic cluster is not a necessary component of CRE/AP-1 specificity. Although residues within the CRE-BP1 basic cluster specify the direction of DNA bending by CRE-BP1 peptides (18), they do not encode any determinants of CRE/AP-1 selectivity that are characteristic of the CREB/ATF class of bZIP proteins.

While addition or removal of the CRE-BP1 basic cluster did not modulate CRE/AP-1 selectivity, it did perturb binding affinity. c5g bound CRE_{24} and AP-1_{23} 1.1 and 1.6 kcal/mol more poorly respectively than did ggg; c10g bound these sequences 3.1 and 3.6 kcal/mol more poorly than did ggg. Thus, the presence of the basic cluster reduced the affinity of these peptides for CRE_{24} and AP-1_{23}. Interestingly, the presence of the basic cluster in the context of ccc enhances binding affinity; the stability of the g5c·CRE_{24} complex was 1.2 kcal/mol less than the stability of the ccc·CRE_{24} complex under identical conditions. Thus, while the basic cluster perturbs the affinities of peptides that do not normally contain a basic cluster, its presence is not inherently destabilizing to the bZIP·DNA complex.

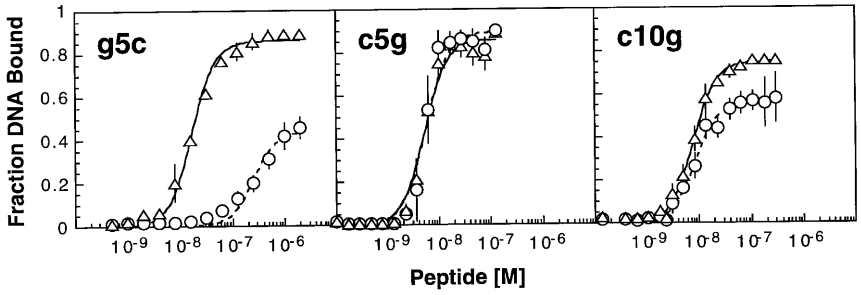

Figure 2. CRE/AP-1 selectivities of selected chimeric peptides as determined by direct electrophoretic mobility shift analysis. Plots illustrate the fraction of AP-1_{23} (○) or CRE_{24} (Δ) bound as a function of the concentration of g5c, c5g or c10g. Data were fitted to a Langmuir binding isotherm as described previously (10). Experiments using c5g and c10g were performed at 25°C; experiments using g5c were performed at 4°C because the g5c·AP-1_{23} complex could not be observed at 25°C.

Table 1. Non-specific DNA affinities of chimeric peptides as determined by competition electrophoretic mobility shift analysis

Peptide	$K_{cre} \times 10^{18}$ (M^2)	$K_{non} \times 10^{18}$ (M^2)	$\Delta\Delta G_{cre/non}$ (kcal/mol)	Induced bend?
ggg	5 ± 1.2	1986 ± 662	3.5	No
ccc	1128 ± 341	940 ± 112	–0.1	Yes
g5c	37 ± 11	4416 ± 1535	2.8	No
c5g	90 ± 23	274 ± 2	0.6	Yes
cgg	645 ± 328	1104 ± 61	0.3	Yes
gcc	6 ± 0.4	2128 ± 642	3.5	No
ccg	80 ± 39	552 ± 5	1.1	Yes
c10g	3 ± 0.8	83 ± 24	2.0	Yes
ggc	29 ± 12	623 ± 260	1.8	No

$\Delta\Delta G_{cre/non}$ was calculated from the relationship $\Delta\Delta G_{cre/non} = \Delta G_{non} - \Delta G_{cre}$, where $\Delta G_{non} = -RT \ln(1/K_{non})$ and $\Delta G_{cre} = -RT \ln(1/K_{cre})$. Bending data are from references (16,18).

Basic cluster involvment in non-specific binding

Next we asked whether the presence of the CRE-BP1 basic cluster influenced affinity for non-specific DNA. To measure non-specific DNA affinity, we made use of a competition assay in which the fraction of [^{32}P]CRE_{24} bound to a given peptide was monitored as a function of the concentration of non-specific DNA added to the reaction (23,24). The non-specific DNA used for this purpose, Non_{24}, contained a sequence identical to CRE_{24} except that the 10 bp of the CRE site were scrambled.

Both ccc and c5g displayed surprisingly high affinities for Non_{24} (Table 1). The stability of the ccc·Non_{24} complex, ΔG_{non} = –20.4 kcal/mol, was almost identical to the stability of the ccc·CRE_{24} complex, ΔG_{cre} = –20.5 kcal/mol, and the stability of the c5g·Non_{24} complex, ΔG_{non} = –21.8 kcal/mol, was only 0.6 kcal/mol less than the stability of the c5g·CRE_{24} complex. Thus, there is virtually no free energy difference between the specific and non-specific complexes for ccc and c5g. Although there are potentially multiple non-specific binding sites on Non_{24} and a commensurate statistical advantage for the composite site over a single specific site, we compare overall affinities of the peptides for two oligonucleotide sequences of the same length, one which contains a specific site and one which does not.

In contrast to results obtained with peptides ccc and c5g, the analogous peptides ggg and g5c, which do not contain the basic cluster, bound more poorly to non-specific DNA, Non_{24}, than to specific DNA, CRE_{24} (Table 1). The free energy of the ggg·Non_{24} complex, –20.1 kcal/mol, was 3.5 kcal/mol less favorable than the free energy of the ggg·CRE_{24} complex and the free energy of the g5c·Non_{24} complex, –19.6 kcal/mol, was 2.8 kcal/mol less favorable than that of the g5c·CRE_{24} complex. Thus, the presence of the basic cluster in ccc decreases the specificity for CRE_{24} over Non_{24} by 2.9 kcal/mol relative to g5c and the absence of the basic cluster in ggg increases the specificity for CRE_{24} over Non_{24} by 2.9 kcal/mol relative to c5g. This dramatic change in the specificity of these peptides for their target site relative to non-specific DNA requires only the five residues of the CRE-BP1 basic cluster.

The CRE_{24} and Non_{24} affinities of chimeric peptides in which whole segments of the GCN4 and CRE-BP1 bZIP elements were interchanged (Fig. 1) followed, with two exceptions, the same trend (Table 1). When the GCN4 basic region, which lacks a basic cluster, was fused to the CRE-BP1 spacer and zipper segments, the result was a peptide (gcc) that preferred the CRE_{24} sequence to the Non_{24} sequence by >3 kcal/mol. Conversely, when the CRE-BP1 basic segment was fused to the GCN4 spacer and zipper segments, the result was a peptide (cgg) that preferred the CRE_{24} sequence to the Non_{24} sequence by only $\Delta\Delta G_{non/cre}$ = 0.3 kcal/mol. Peptide ccg, which contains the basic and spacer segments of CRE-BP1 fused to the zipper segment of GCN4, as expected, displayed a high degree of non-specific binding ($\Delta\Delta G_{non/cre}$ = 1.1 kcal/mol). Peptides c10g and ggc behaved anomalously. While one contained a basic cluster and one did not, they displayed $\Delta\Delta G_{non/cre}$ = 2.0 and 1.8 kcal/mol respectively, directly between the two extreme cases.

Differences in specificity are not DNA dependent

To test the abilities of both large pieces of random DNA and also a simple repeating DNA polymer to compete with CRE_{24} for specific binding, we measured the affinities of ggg, ccc and c5g for both calf thymus DNA (CTD) and poly(dI-dC)·poly(dI-dC) (Fig. 3). The abilities of these large DNAs to compete with [^{32}P]CRE_{24} was compared with the ability of unlabeled CRE_{24} to compete with itself. The data show that ccc and c5g bound CTD and poly(dI-dC)·poly(dI-dC) approximately as well as they bound CRE_{24}, whereas ggg did not. There was a 3.3 kcal/mol difference between the stabilities of both the ggg·CTD and ggg·poly(dI·dC) complexes and the stability of the ggg·CRE_{24} complex. The corresponding complexes of c5g differed by 1.5 and –0.1 kcal/mol respectively from the stability of the c5g·CRE_{24} complex. The simple repeating polymer poly(dI·dC) competed as effectively for c5g as DNA containing a specific binding site. A similar difference between specific and non-specific binding was seen in the case of ccc, which, as expected from the earlier results, displayed low specificity. While the difference in stability between the ccc·CTD complex and the ccc·CRE_{24} complex was a moderate 2.0 kcal/mol, there was no difference between the stabilities of ccc·poly(dI·dC) and ccc·CRE_{24} complexes. The high affinity of a peptide for both a non-specific, heterogeneous DNA sequence, CTD, as well as a simple repeating polymer, poly(dI·dC), results directly from the presence of the basic cluster in the peptide sequence.

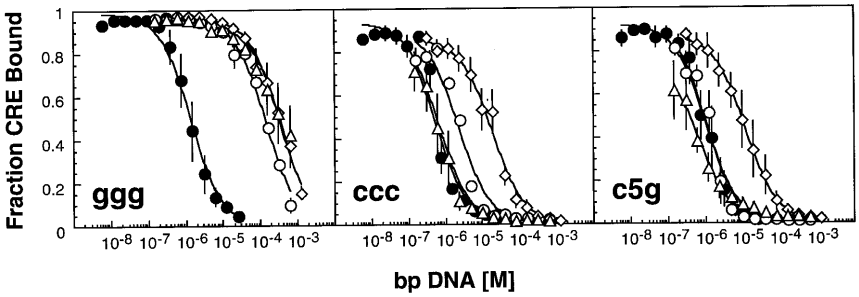

Figure 3. Non-specific binding affinities of selected chimeric peptides as determined by competition electrophoretic mobility shift analysis. Plots illustrate the fraction of [^{32}P]CRE$_{24}$ bound to ggg, ccc or c5g in the presence of CRE$_{24}$ (●), Non$_{24}$ (○), calf thymus DNA (◇) or poly(dI-dC)·poly(dI-dC) (Δ). DNA concentrations are listed in terms of moles of base pairs. Binding data were fitted to equation **12**. Error bars represents the standard error for an average of three trials.

Binding to non-specific DNA via a random coil conformation

Although peptides ccc and c5g displayed affinities for Non$_{24}$ that were comparable with those for CRE$_{24}$, the complexes with Non$_{24}$ were not stable during gel electrophoresis (data not shown). Fluorescence spectroscopy was therefore used to verify the existence of a direct peptide–Non$_{24}$ interaction. Experiments were performed with peptide ccc which contains a unique tryptophan within the spacer segment. The fluorescence of a 20 μM solution of ccc was monitored in the presence or absence of CRE$_{24}$ or Non$_{24}$ (Fig. 4A). Fluorescence intensity was greatest in the absence of DNA. Strong quenching of the fluorescence was observed upon addition of CRE$_{24}$ (48%) or Non$_{24}$ (63%). The fluorescence data indicate that ccc interacts with both DNA sequences and suggests that the interaction with the specific CRE$_{24}$ sequence differs from that with the non-specific Non$_{24}$ sequence.

Circular dichroism was used to determine the extent of α-helicity within the ccc·Non$_{24}$ complex. Experiments were performed at a ccc concentration of 20 μM, a concentration well below the dissociation constant of the ccc dimer (K_{dim} > 200 μM) (10). The circular dichroism spectrum between 200 and 270 nm was unchanged upon addition of 20 μM Non$_{24}$ to ccc (Fig. 4B). In both cases the spectrum corresponded to that of a random coil peptide, with θ_{MRW} = –6000 deg·cm^2/dmol at 222 nm. In contrast, addition of 20 μM CRE$_{24}$ to ccc resulted in a large change in the circular dichroism spectrum, with θ_{MRW} = –26 000 deg·cm^2/dmol at 222 nm, consistent with formation of considerable α-helical structure. It was not possible to conduct circular dichroism experiments on peptides containing a GCN4 zipper (e.g. ggg and c5g), because addition of Non$_{24}$ led to precipitate formation. The circular dichroism experiments indicate that although the ccc peptide interacted equally well with both the CRE$_{24}$ and Non$_{24}$ sequences, the structures of the resultant complexes differed considerably. The peptide in the Non$_{24}$·ccc complex did not undergo the α-helical folding transition characteristic of formation of a specific bZIP·DNA complex (25–29).

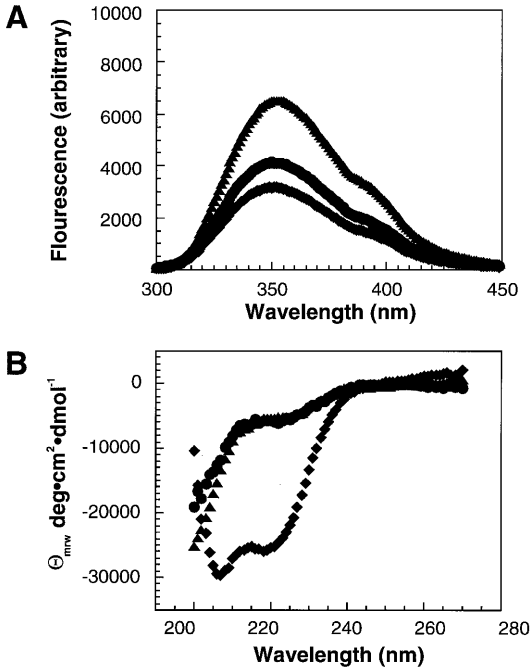

Figure 4. Fluorescence and circular dichroism spectra of ccc. (**A**) Fluorescence spectra of 20 μM ccc at 25°C in the absence (▲) or presence of Non$_{24}$ (●) or CRE$_{24}$ (◆). (**B**) Circular dichroism spectra of 20 μM ccc at 25°C in the absence (▲) or presence of 20 μM Non$_{24}$ (●) or CRE$_{24}$ (◆). Spectra are corrected for the small signals due to CRE$_{24}$ or Non$_{24}$.

DISCUSSION

Bending does not effect target site selectivity

We tested whether the basic cluster of CREB/ATF proteins could control CRE/AP-1 specificity. The basic cluster was removed from a CREB/ATF peptide or added to a GCN4 peptide and the CRE/AP-1 selectivities of the resulting chimeric peptides were determined. The results of both the loss and gain of function experiments were the same: basic cluster residues do not control CRE/AP-1 selectivity and therefore there is no causal relationship between DNA bending and CRE/AP-1 selectivity. In agreement with a study of the specificity of the ATF-1 bZIP element (30), our results point to residues in the C-terminal half of the basic segment and in the spacer segment as being important for target site selection: c10g, which contains sequence from CRE-BP1 in the N-terminal half of the basic segment, fails to discriminate between the CRE and AP-1 sites, while cgg, in which all 20 basic segment residues are from CRE-BP1, discriminates between the two sites almost as well as does ccc (10).

The basic cluster and non-specific binding

While the basic cluster and its accompanying ability to bend DNA have no influence on the relative affinity of a peptide for the

closely related CRE and AP-1 target sites, it has a dramatic influence on the relative affinity for specific and non-specific DNA. Removal of the basic cluster from peptide ccc increases $\Delta\Delta G_{non/cre}$ by 3 kcal/mol and addition of the basic cluster to peptide ggg decreases $\Delta\Delta G_{non/cre}$ by 3 kcal/mol. The fact that CTD and poly(dI-dC)·poly(dI-dC) compete well for peptides containing a basic cluster strengthens the assertion that non-specific binding is enhanced and that the basic cluster increases DNA affinity regardless of sequence.

Although our data reveal that the ccc·CRE_{24} and ccc·Non_{24} complexes possess similar stabilities, only the specific ccc·CRE_{24} complex contains significant α-helical structure. The ccc·Non_{24} complex, like the uncomplexed peptide, remains in a predominantly random coil conformation. Thus, formation of the ccc·Non_{24} complex requires little conformational reorganization of the peptide and, hence, neglecting solvation, the energy gained through DNA interactions translates directly into binding energy. In contrast, part of the energy gained by additional interactions with DNA in the specific ccc·CRE_{24} complex must be expended in driving the coil-to-helix transition of the basic region. Part of the effectiveness of the basic cluster in binding to random DNA could be its ability to form multiple, simultaneous salt bridges to backbone phosphates in a conformation (or set of conformations) that does not require protein folding and decreases the access of solute cations to the phosphate backbone. A similar situation has been noted in the interaction between arginine-rich peptides from HIV Rev and the Rev response element (RRE) RNA (31). In the specific Rev peptide·RRE complex, the peptide is helical, whereas in non-specific complexes the peptide is not. Increasing the helical propensity of the peptide enhances the specific but not the non-specific binding.

Several groups (18,20,34) have provided evidence that asymmetric neutralization of the anionic phosphodiester backbone via salt bridges caused DNA to bend. It has been suggested that appending a cationic surface to an adjacent sequence-specific DNA recognition motif should bend the DNA in a controlled, predictable manner (32). The peptide c5g exhibits this design; it bends DNA toward the neutralized surface. The result in terms of bending is as predicted. However, the large increase in non-specific binding was unexpected and urges caution in the design of similar molecules.

Summary

The 10^9 bp that comprise the human genome provide a vast background of non-specific DNA through which a transcription factor must sort to find its cognate site (33). A large fraction of this DNA differs considerably from the cognate sequence, but other sequences may differ by just 1 or 2 bp. Transcription factors must not only avoid activating from sequences that are similar to their target site, they must also remain free from the massive excess of completely non-specific DNA to bind their target sites at reasonable concentrations. Hence the ability of a transcription factor to avoid non-specific binding is a critical factor in its function of specific regulation. Here we have shown that a cluster of basic amino acids present at the N-terminus of all CREB/ATF bZIP elements is not responsible for their preference for the CRE site over the closely related AP-1 site. The presence of the basic cluster does, however, greatly increase the affinity of a peptide for non-specific versus specific DNA. It appears as though the function of the basic cluster, DNA bending, is paid for at the expense of specific binding. We surmise that the cost must be regained from some transcriptional advantage conferred on the complex containing an altered DNA architecture.

ACKNOWLEDGEMENTS

We are grateful to the National Institutes of Health (GM52544) for their financial support of this research. We also thank the Arthur Wayland Dox Foundation and the Organic Chemistry Division of the American Chemical Society for graduate fellowships to S.J.M.

REFERENCES

1 McKnight,S.L. (1991) *Scient. American*, **264**, 54–64.
2 Hurst,H.C. (1995) *Protein Profile*, **2**, 101–168.
3 Harrison,S.C. (1991) *Nature*, **353**, 715–719.
4 König,P. and Richmond,T. (1993) *J. Mol. Biol.*, **233**, 139–154.
5 Glover,J.N.M. and Harrison,S.C. (1995) *Nature*, **373**, 257–261.
6 Ellenberger,T.E., Brandl,C.J., Struhl,K. and Harrison,S.C. (1992) *Cell*, **71**, 1223–1237.
7 Keller,W., König,P. and Richmond,T.J. (1995) *J. Mol. Biol.*, **254**, 657–667.
8 O'Shea,E.K., Klemm,J.D., Kim,P.S. and Alber,T. (1991) *Science*, **254**, 539–544.
9 Pu,W.T. and Struhl,K. (1991) *Mol. Cell. Biol.*, **11**, 4918–4926.
10 Metallo,S.J. and Schepartz,A. (1994) *Chem. Biol.*, **1**, 143–151.
11 Short,N. (1987) *Nature*, **330**, 209–210.
12 Bos,T., Bohmann,D., Tsuchie,H., Tjian,R. and Vogt,P. (1988) *Cell*, **52**, 705–712.
13 Hai,T., Liu,F., Coukos,J. and Green,M.R. (1989) *Genes Dev.*, **3**, 2083–2090.
14 Sellers,J.W., Vincent,A.C. and Struhl,K. (1990) *Mol. Cell. Biol.*, **10**, 5077–5086.
15 Kerppola,T.K. and Curran,T. (1993) *Mol. Cell. Biol.*, **13**, 5479–5489.
16 Paolella,D.N., Palmer,C.R. and Schepartz,A. (1994) *Science*, **264**, 1130–1133.
17 Kerppola,T.K. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 10117–10122.
18 Paolella,D., Liu,Y. and Schepartz,A. (1997) *Biochemistry*, in press.
19 Mirzabekov,A.D. and Rich,A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 1118–1121.
20 Strauss,J.K. and Maher,L.J.M. (1994) *Science*, **266**, 1829–1834.
21 Hinnebusch,A.G. and Fink,G.R. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5374–5378.
22 Maekawa,T., Sakura,H., Kanei-Ishii,C., Sudo,T., Yoshimura,T., Fujisawa,J., Yoshida,M. and Ishii,S. (1989) *EMBO J.*, **8**, 2023–2028.
23 Lin,S.-Y. and Riggs,A.D. (1972) *J. Mol. Biol.*, **72**, 671–690.
24 Cuenoud,B. and Schepartz,A. (1993) *Science*, **259**, 510–513.
25 O'Neil,K.T., Hoess,R.H. and DeGrado,W.F. (1990) *Science*, **249**, 774–778.
26 Johnson,N.P., Lindstrom,J., Baase,W.A. and Hippel,P.H.v. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4840–4844.
27 Weiss,M.A., Ellenberger,T., Wobbe,C.R., Lee,J.P., Harrison,S.C. and Struhl,K. (1990) *Nature*, **347**, 575–578.
28 Weiss,M. (1990) *Biochemistry*, **29**, 8020–8024.
29 O'Neil,K.T., Shumann,J.D., Ampe,C. and DeGrado,W.F. (1991) *Biochemistry*, **30**, 9030–9034.
30 Kim,J. and Struhl,K. (1995) *Nucleic Acids Res.*, **23**, 2531–2537.
31 Tan,R., Chen,L., Buettner,J., Hudson,D. and Frankel,A. (1993) *Cell*, **73**, 1031–1040.
32 Strauss,J., Roberts,C., Nelson,M., Switzer,C. and Maher,L. (1996) *Chem. Biol.*, **3**, 671–8.
33 Kodadek,T. (1995) *Chem. Biol.*, **2**, 267–279.
34 Leonard,D.A., Rajaram,N. and Kerppola,T.K. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 4913–4918.