J|A|C|S ARTICLES

Published on Web 06/29/2010

Design of Single-Stranded Nucleic Acid Binding Peptides Based on Nucleocapsid CCHC-Box Zinc-Binding Domains

Anthony L. Guerrerio[†,§] and Jeremy M. Berg*[,†,‡]

Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, and National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892

Received January 4, 2010; E-mail: bergj@mail.nih.gov

Abstract: The solution structures of nucleocapsid (NC)-like CCHC zinc-binding domains bound to nucleic acid targets have revealed that these domains bind guanosine residues within single-stranded nucleic acids. Here, we have performed initial studies examining the potential use of NC-like CCHC zinc-binding domains as modules to construct single-stranded nucleic acid binding peptides. The affinity for guanosine-containing single-stranded deoxyribooligonucleotides increases with the number of CCHC domains in the peptide. The length of the linker between domains affects the spacing of guanosine residues in oligonucleotides that are preferentially bound. These studies provide a proof of principle that NC-like CCHC zinc-binding domains can be utilized as a basis for designing peptides that bind specific single-stranded nucleic acid sequences.

Introduction

The prospect of designing functional proteins based on naturally occurring motifs, such as the Cys_2His_2 zinc finger proteins typified by *Xenopus* transcription factor IIIA, has shown great promise in recent years. Cys_2His_2 zinc fingers are small, approximately 28−30 amino acid domains that coordinate zinc through two cysteines and two histidines to form a stable structure well suited for binding double-stranded DNA.[1−3] Crystal structures of complexes between arrays of Cys_2His_2 domains from a series of DNA-binding proteins have revealed that each domain typically inserts an α-helix into the major groove and utilizes sets of amino acids residues to make sequence-specific contacts with the DNA.[3−10] Each domain contacts three or four bases, and the binding sites for each domain are adjacent or slightly overlapping. The modular nature of these proteins suggested that specific DNA-binding proteins might be designed by preparing proteins containing tandem arrays of these Cys_2His_2 domains with previously characterized DNA-binding characteristics. This suggestion has been realized with highly specific DNA-binding proteins being prepared, structurally characterized, and utilized in vivo to activate or repress the expression of specific genes or to target nucleases to specific sites.[6,11−21]

Despite their success, Cys_2His_2 zinc finger proteins bind primarily double-stranded nucleic acids. Therefore, expansion of designed protein binding into the realm of single-stranded nucleic acids requires a new module. Ideally, such a module would be a small, stable structure that recognizes a small number of bases, relatively independently of other modules surrounding it. Retroviral nucleocapsid-like CCHC-box zinc-binding domains (sometimes referred to as "zinc knuckles") have properties suggesting they may prove useful as such a module.[22−25] CCHC-box domains were first identified as a recurring sequence motif in retroviral nucleocapsid proteins of the form Cys-X_2-

† Johns Hopkins School of Medicine.
‡ National Institute of Diabetes and Digestive and Kidney Diseases.
§ Present address: Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21205.

(1) Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 99–102.
(2) Lee, M. S.; Gippert, G. P.; Soman, K. V.; Case, D. A.; Wright, P. E. *Science* **1989**, *245*, 635–637.
(3) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809–817.
(4) Kim, C. A.; Berg, J. M. *Nat. Struct. Biol.* **1996**, *3*, 940–945.
(5) Wolfe, S. A.; Nekludova, L.; Pabo, C. O. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 183–212.
(6) Pabo, C. O.; Peisach, E.; Grant, R. A. *Annu. Rev. Biochem.* **2001**, *70*, 313–40.
(7) Wolfe, S. A.; Grant, R. A.; Elrod-Erickson, M.; Pabo, C. O. *Structure* **2001**, *9*, 717–723.
(8) Wolfe, S. A.; Grant, R. A.; Pabo, C. O. *Biochemistry* **2003**, *42*, 13401–13409.
(9) Segal, D. J.; Crotty, J. W.; Bhakta, M. S.; Barbas, C. F., III; Horton, N. C. *J. Mol. Biol.* **2006**, *363*, 405–421.
(10) Stoll, R.; Lee, B. M.; Debler, E. W.; Laity, J. H.; Wilson, I. A.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2007**, *372*, 1227–1245.
(11) Desjarlais, J. R.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7345–7349.
(12) Pomerantz, J. L.; Sharp, P. A.; Pabo, C. O. *Science* **1995**, *267*, 93–96.
(13) Beerli, R. R.; Barbas, C. F., III *Nat. Biotechnol.* **2002**, *20*, 135–141.
(14) Jantz, D.; Amann, B. T.; Gatto, G. J., Jr.; Berg, J. M. *Chem. Rev.* **2004**, *104*, 789–799.
(15) Mandell, J. G.; Barbas, C. F., III *Nucleic Acids Res.* **2006**, *34*, W516–W523.
(16) Nomura, W.; Sugiura, Y. *Methods Mol. Biol.* **2007**, *352*, 83–93.
(17) Carroll, D. *Gene Ther.* **2008**, *15*, 1463–1468.
(18) Negi, S.; Imanishi, M.; Matsumoto, M.; Sugiura, Y. *Chemistry* **2008**, *14*, 3236–3249.
(19) Porteus, M. *Methods Mol. Biol.* **2008**, *435*, 47–61.
(20) Gordley, R. M.; Gersbach, C. A.; Barbas, C. F., III *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5053–5058.
(21) Le Provost, F.; Lillico, S.; Passet, B.; Young, R.; Whitelaw, B.; Vilotte, J. L. *Trends Biotechnol.* **2010**, in press.
(22) Berg, J. M. *Science* **1986**, *232*, 485–487.
(23) Green, L. M.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4047–4051.
(24) Summers, M. F.; South, T. L.; Kim, B.; Hare, D. R. *Biochemistry* **1990**, *29*, 329–340.
(25) Summers, M. F. *J. Cell Biochem.* **1991**, *45*, 41–48.

10.1021/ja910942v

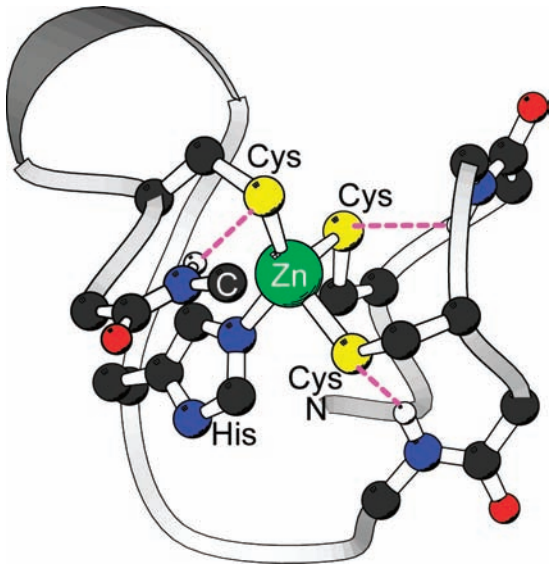

Figure 1. Representation of the structure typically formed by CCHC-box domains. The structure is centered around a zinc ion coordinated to the cysteinate sulfur atom of three cysteinate residues and the ϵ-nitrogen from a histidine. Three peptide backbone NH-to-S hydrogen bonds (shown as dotted lines) stabilize the structure.

Cys-X_4-His-X_4-Cys. These sequences were proposed[22] and confirmed[23,26] to form zinc-binding domains, with the Cys and His side chains providing ligands to a tetrahedrally coordinated zinc ion. The three-dimensional structures of a number of these domains and proteins containing these domains have been determined by NMR methods.[24,27−39] These domains form compact structures that are quite similar to one another (Figure 1). While some proteins contain single CCHC-box domains, many contain arrays of two or more such domains. In these proteins, the CCHC-box domains are connected, preceded, and followed by regions that are relatively unstructured in the absence of bound nucleic acid.

The structures of several complexes between CCHC-box domains and CCHC-box-containing proteins have been determined by NMR methods. The structure of one such complex, that between the nucleocapsid protein from HIV-1 bound to a 20-nucleotide fragment of RNA corresponding to stem loop 3 (SL3) from the RNA packaging signal,[29] is shown in Figure 2. In this structure, each of the CCHC-box domains interacts primarily with one guanosine residue, with the guanine base extending into a hydrophobic pocket within the domain. Guanine recognition is further facilitated by hydrogen bonds involving groups from the polypeptide backbone of the CCHC-box domain and NH1 and O6 of the base. In this structure, the two bound guanosines are part of a G-N-G sequence in a loop. This mode of interaction, with one CCHC-box domain binding a guanosine residue, has been observed for almost all other CCHC-box domains nucleic acid complexes that have been structurally characterized. The importance of guanosine residues is also supported by binding site selection studies of the HIV nucleocapsid protein.[40,41]

This recognition of guanosines by CCHC-box proteins is not limited to RNA. For example, the structure of the HIV NC amino-terminal CCHC zinc-binding domain complexed with a DNA pentanucleotide, d(ACGCC), revealed interactions that are quite similar to those observed for the same domain in the context of the intact HIV NC protein bound to RNA, including the same hydrogen-bonding interactions between the peptide backbone and the central deoxyguanosine.[27] Another protein containing CCHC-box domains is the universal minicircle sequence binding protein (UMSBP) from the trypanosome *Crithidia fasciculata*.[42−45] UMSBP contains five CCHC-box domains and functions to bind the guanosine-rich single-stranded DNA universal minicircle sequence (UMS), 5′-GGGGTTGGT-GTA-3′. UMSBP binds its single-stranded DNA target with a dissociation constant of approximately 0.04 nM. The protein can also bind to single-stranded RNA, but not to double- or quadruple-stranded DNA. Mutational analysis has shown that the guanine bases, specifically G4, G7, and G10 (underlined above), are essential for high-affinity UMSBP binding to UMS.[44]

Since CCHC-box domain-containing proteins bind single-stranded nucleic acids containing arrays of guanosines, we reasoned that these domains might be well-suited modules for the design of specific single-stranded nucleic acid binding proteins. The well-structured region of each CCHC-box domain extends approximately one amino acid before the first and three amino acids beyond the last zinc-coordinated cysteinate. The sequence of the HIV-1 nucleocapsid protein is MQKGNFR-NQRKTVK**C**FN**C**GKEG**H**IAKN**C**RAPRKKG**C**WK**C**GKEG-**H**QMKD**C**TE-RQAN, with the zinc-binding residues shown in bold and the linker between CCHC-box domains underlined. The linker, starting after the last cysteine in one CCHC-box domain and ending at the first cysteine of the next CCHC-box domain, is seven amino acids in length. This corresponds to a linker between the structured regions of the two CCHC-box domains of approximately three amino acids, and the binding site for this protein has the form 5′-GXG-3′. A potential

(26) South, T. L.; Kim, B.; Summers, M. F. *J. Am. Chem. Soc.* **1989**, *111*, 395–398.
(27) South, T. L.; Summers, M. F. *Protein Sci.* **1993**, *2*, 3–19.
(28) Demene, H.; Jullian, N.; Morellet, N.; de Rocquigny, H.; Cornille, F.; Maigret, B.; Roques, B. P. *J. Biomol. NMR* **1994**, *4*, 153–170.
(29) De Guzman, R. N.; Wu, Z. R.; Stalling, C. C.; Pappalardo, L.; Borer, P. N.; Summers, M. F. *Science* **1998**, *279*, 384–388.
(30) Gao, Y.; Kaluarachchi, K.; Giedroc, D. P. *Protein Sci.* **1998**, *7*, 2265–2280.
(31) Kodera, Y.; Sato, K.; Tsukahara, T.; Komatsu, H.; Maeda, T.; Kohno, T. *Biochemistry* **1998**, *37*, 17704–17713.
(32) Morellet, N.; Demene, H.; Teilleux, V.; Huynh-Dinh, T.; de Rocquigny, H.; Fournie-Zaluski, M. C.; Roques, B. P. *J. Mol. Biol.* **1998**, *283*, 419–434.
(33) Schuler, W.; Dong, C.; Wecker, K.; Roques, B. P. *Biochemistry* **1999**, *38*, 12984–12994.
(34) Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.; Chancellor, K. J.; Wu, Z. R.; Summers, M. F. *J. Mol. Biol.* **2000**, *301*, 491–511.
(35) Guo, J.; Wu, T.; Anderson, J.; Kane, B. F.; Johnson, D. G.; Gorelick, R. J.; Henderson, L. E.; Levin, J. G. *J. Virol.* **2000**, *74*, 8980–8988.
(36) Klein, D. J.; Johnson, P. E.; Zollars, E. S.; De Guzman, R. N.; Summers, M. F. *Biochemistry* **2000**, *39*, 1604–1612.
(37) Morellet, N.; Meudal, H.; Bouaziz, S.; Roques, B. P. *Biochem. J.* **2006**, *393*, 725–732.
(38) Matsui, T.; Kodera, Y.; Endoh, H.; Miyauchi, E.; Komatsu, H.; Sato, K.; Tanaka, T.; Kohno, T.; Maeda, T. *J. Biochem.* **2007**, *141*, 269–277.
(39) Matsui, T.; Kodera, Y.; Miyauchi, E.; Tanaka, H.; Endoh, H.; Komatsu, H.; Tanaka, T.; Kohno, T.; Maeda, T. *Biochem. Biophys. Res. Commun.* **2007**, *358*, 673–678.
(40) Berglund, J. A.; Charpentier, B.; Rosbash, M. *Nucleic Acids Res.* **1997**, *25*, 1042–1049.
(41) Kim, S. J.; Kim, M. Y.; Lee, J. H.; You, J. C.; Jeong, S. *Biochem. Biophys. Res. Commun.* **2002**, *291*, 925–931.
(42) Avrahami, D.; Tzfati, Y.; Shlomai, J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10511–10515.
(43) Tzfati, Y.; Abeliovich, H.; Avrahami, D.; Shlomai, J. *J. Biol. Chem.* **1995**, *270*, 21339–21345.
(44) Abu-Elneel, K.; Kapeller, I.; Shlomai, J. *J. Biol. Chem.* **1999**, *274*, 13419–13426.
(45) Onn, I.; Kapeller, I.; Abu-Elneel, K.; Shlomai, J. *J. Biol. Chem.* **2006**, *281*, 37468–37476.

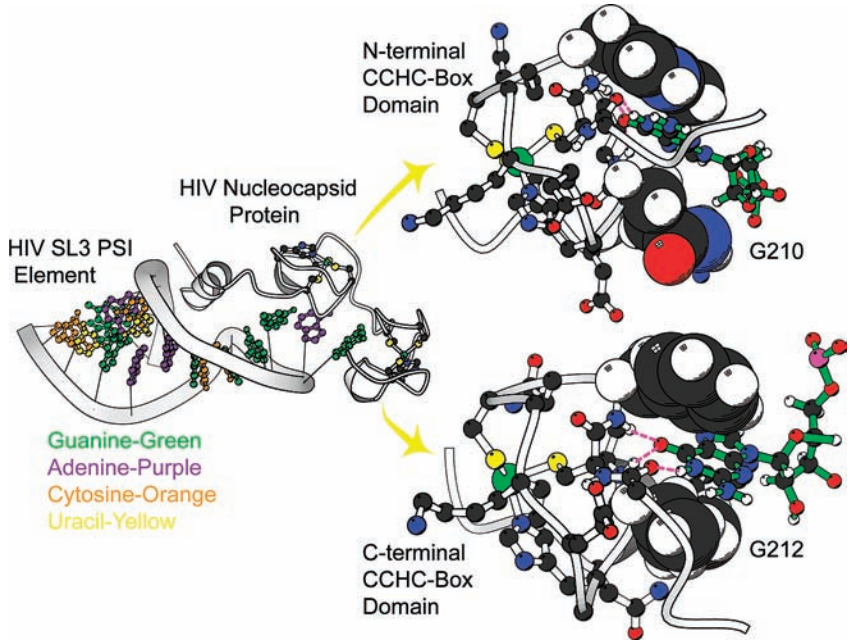

Figure 2. Structure of the HIV-1 nucleocapsid protein bound to a 20-base RNA target. Each of the two CCHC-box domains interacts primarily with a guanosine, with the guanine base lying in a pocket in the CCHC-box domain, making hydrogen bonds (shown as dotted lines) with the CCHC-box domain backbone. Note that the CCHC-box domains have been reoriented for clarity.

relationship between the length of the linker between CCHC-box domains and the spacing between key guanine residues within their binding site is revealed by comparing these results for HIV-1 nucleocapsid protein with those for UMSBP. The guanosine residues that are most crucial for binding in UMS are spaced with the pattern 5′-GXXGXXG-3′. The amino acid sequence of UMSBP, MSAAVT**C**YK**C**GEAG**H**MSRE**C**P-KAAASRT**C**YN**C**GQTG**H**LSRE**C**-PSERKPKA**C**YN**C**GSTE-**H**LSRE**C**PNEAKTGADSRT**C**YN**C**GQSG**H**LSRD**C**-PSERK-PKA**C**YN**C**GSTE**H**LSRE**C**PDRH, reveals that three of the four linkers between CCHC-box domains in UMSBP are one amino acid longer than that in HIV-1 nucleocapsid protein, while the other is five amino acids longer. This correlation between linker length and the spacing between key guanosine residues is highly suggestive and is supported by simple structural modeling.

On the basis of these observations, we hypothesized that proteins with arrays of CCHC-box domains recognize single-stranded nucleic acids with patterns of guanosine residues, with each CCHC-box domain recognizing a single guanosine and the lengths of the linkers between CCHC-box domains determining the spacing of the recognized guanosines. Here, we report results that provide support for this hypothesis and use this principle to design binding proteins that recognize specific DNA sequences with appropriately spaced guanosine residues.

Experimental Section

Preparation of Peptides. The following peptides were synthesized on a Milligen/Biosearch 9050 peptide synthesizer using 9-fluorenylmethoxycarbonyl chemistry:

UBPCON1:	SRT**C**YN**C**GQTG**H**LSRE**C**PSE
UBPCON2U:	SRT**C**YN**C**GQTG**H**LSRE**C**PSERKPKA**C**YN**C**GQTG**H**LSR E**C**PSE
UBPCON3UU:	SRT**C**YN**C**GQTG**H**LSRE**C**PSERKPKA**C**YN**C**GQTG**H**LSR E**C**PSERKPKA**C**YN**C**GQTG**H**LSRE**C**PSE
UBPCON2N:	SRT**C**YN**C**GQTG**H**LSRE**C**RAPRKKG**C**YN**C**GQTG**H**LSRE **C**PSE

Upon completion of each synthesis, the peptide was cleaved from the resin and deprotected by treatment with a solution of trifluoroacetic acid (80%), thioanisole (5%), ethanedithiol (3%), and anisole (2%) for 2 h. The crude peptides were filtered through glass wool, precipitated with ether, and washed with ether. The peptides were resuspended in 5% acetonitrile and 0.1% trifluoroacetic acid (TFA) and purified on a Vydac C18 reversed-phase HPLC column with a 0.5%/min acetonitrile gradient containing 0.1% TFA. The peptides were then resuspended in 400 mM Tris, pH 7.0, 100 mM tris(carboxyethyl)phosphine, 5% acetonitrile in order to reduce all cysteines. The peptides were then purified on a Vydac diphenyl reversed-phase HPLC column with a 0.5%/min acetonitrile gradient containing 0.1% TFA. Collected fractions were dried under a 95% nitrogen/5% hydrogen atmosphere in a Savant SpeedVac concentrator. All manipulations from this point on were performed in this atmosphere to prevent cysteine oxidation, and all solvents were degassed with helium prior to use. The mass of the peptides were confirmed using MALDI mass spectrometry:

	$M + H^+$ (Da)	
	calcd	obsd
UBPCON1	2229.4	2230.9
UBPCON2U	4676.2	4676.8
UBPCON3UU	7122.0	7122.3
UBPCON2N	4576.1	4576.4

Peptides were quantitated using $\epsilon_{275} = 1420\ M^{-1}\ cm^{-1}$ per tyrosine residue. The ability of the peptides to bind metal ions was monitored using cobalt(II) and methods that have been used for other CCHC-box peptides.[23] Absorption spectra very similar to those for other CCHC-box peptides were observed, with three peaks in the visible region having extinction coefficients greater than 400 $M^{-1}cm^{-1}$ (per metal binding site), consistent with tetrahedral metal coordination by three cysteinates and one histidine. Zinc(II) was found to readily displace cobalt(II). For the DNA-binding experiments, $ZnCl_2$ was added to provide 100 μM free Zn^{2+} when resuspended for the titration, taking into account the 1 equiv of Zn^{2+} bound per CCHC-box domain.

Preparation of Oligonucleotides. The following single-stranded deoxyribooligonucleotides, both unlabeled and 3′-fluorescein-labeled, were obtained in purified form from the Johns Hopkins School of Medicine biopolymer facility: d(TTGTT)-F, d(T_3GTTGT$_3$)-

F, d(T_3GTTGTTGT$_3$)-F, d(T_{10})-F, d(T_{13})-F, d(T_2GTTTGT$_3$), d(T_3GTTGT$_3$), d(T_4GTGT$_3$), and d(T_5GGT$_3$). Oligonucleotides were resuspended in 10 mM sodium phosphate buffer, pH 7.0, with 200 mM NaCl and then dialyzed against a 1000-fold excess of the same buffer plus 10 g of Chelex-100 beads (Sigma). The dialysis buffer was changed once. The oligonucleotides were dialyzed against a 1000-fold excess of double-distilled water plus 10 g of Chelex-100 beads and dried in a Savant SpeedVac concentrator. Light exposure for the fluorescein-labeled oligonucleotides was minimized. Oligonucleotides were resuspended under a 95% nitrogen/5% hydrogen atmosphere with buffers degassed with helium when used for a titration. Fluorescein-labeled oligonucleotides were quantified using $\epsilon_{490} = 70\,000\ M^{-1}\ cm^{-1}$. Unlabeled oligonucleotides were quantified using $\epsilon_{260} = 86\,800\ M^{-1}\ cm^{-1}$ for all oligonucleotides except d(T_5GGT$_3$), for which $\epsilon_{260} = 86\,200\ M^{-1}\ cm^{-1}$ was used.

Binding Studies. The binding of the peptides to fluorescently labeled oligonucleotides was monitored using a fluorescence anisotropy assay. The measurements were made on an ISA Fluorolog-3 spectrofluorometer configured in the L-format. Excitation and emission (band-pass/wavelength) were (2 nm/492 nm) and (6 nm/517 nm), respectively.

In a typical forward titration experiment, 220 μL of a 50−100 nM solution of the labeled oligonucleotide in 50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer, 150 mM NaCl, 100 μM $ZnCl_2$, pH 6.0, was added to a 0.4 mL Spectrosil stoppered fluorescence cuvette (Starna Cells, Inc.). The pH value of 6.0 was chosen on the basis of preliminary observations which indicated that several hours were required for the fluorescence anisotropy to stabilize at high peptide concentrations (greater than ~100 μM) at pH 7.0. The fluorescence anisotropy of the solution, corresponding to that of the free oligonucleotide, was recorded. The peptide was then titrated into the cuvette from a stock solution (50 mM MES, 150 mM NaCl, 100 μM free $ZnCl_2$, pH 6.0) in a stepwise fashion, and the resultant changes in anisotropy were recorded.

For most cases, the total change in fluorescence intensity over the course of an experiment was less than 10% and no correction was applied. In the cases where the total intensity did change more than 10%, the anisotropy, *r*, was converted to fraction bound, F_{bound}, using the equation

$$F_{bound} = \frac{r - r_{free}}{(r_{bound} - r)Q + (r - r_{free})}$$

where r_{free} is the anisotropy of the oligonucleotide, r_{bound} is the estimated anisotropy of the oligonucleotide−peptide complex at saturation, and *Q* is the ratio of the quantum yields for bound versus free oligonucleotide.

Competition titrations were begun as forward titrations, except the buffer used was 50 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES) buffer, 150 mM NaCl, 100 μM $ZnCl_2$, pH 7.0 (plus 0.05 mg/mL bovine serum albumin for the UBPCON2N titrations), and the peptide was added until binding was estimated to be ~90% saturated. At this point, unlabeled oligonucleotides were titrated into the cuvette from a stock solution (50 mM HEPES, 150 mM NaCl, 100 μM $ZnCl_2$, pH 7.0) in a stepwise fashion, and the resultant changes in anisotropy were recorded after 20 min, a time found to be sufficient for the anisotropy signal to stabilize.

Results and Discussion

In order to test the hypothesis that single-stranded nucleic acid binding proteins could be designed on the basis of arrays of CCHC-box domains, we prepared a series of peptides containing variable numbers of CCHC-box domains connected by different linkers. The ability of these peptides to binding appropriate oligonucleotides was monitored using a fluorescence anisotropy-based binding assay.

We designed and synthesized four peptides based on the consensus sequence of the five domains from UMSBP, which is identical in sequence to the second CCHC-box domain from UMSBP. The sequences of the peptides are

UBPCON1:	SRT**C**YN**C**GQTG**H**LSRE**C**PSE
UBPCON2U:	SRT**C**YN**C**GQTG**H**LSRE**C**PSERKPKA**C**YN**C**GQTG**H**LSR E**C**PSE
UBPCON3UU:	SRT**C**YN**C**GQTG**H**LSRE**C**PSERKPKA**C**YN**C**GQTG**H**LSR E**C**-PSERKPKA**C**YN**C**GQTG**H**LSRE**C**PSE
UBPCON2N:	SRT**C**YN**C**GQTG**H**LSRE**C**RAPRKKG**C**YN**C**GQTG**H**LSRE **C**PSE

The first peptide, UBPCON1, is identical in sequence to the second CCHC-box domain from UMSBP and the surrounding residues. The second, UBPCON2U, consists of two CCHC-box domains joined by the linker that occurs between domains 2 and 3 in UMSBP. Similarly, the UBPCON3UU consists of three CCHC-box domains, with each pair joined by the linker from UMSBP. The final protein, UBPCON2N, includes two CCHC-box domains identical to those in UBPCON2U, but the linker is based on the sequence that links the two CCHC-box domains in the HIV-1 nucleocapsid protein.

Fluorescence anisotropy was selected as the assay for studying nucleic acid binding.[46−48] This assay has many advantages, including the ability to study binding in solution under a wide range of conditions and the ability to monitor binding quantitatively for interactions with a wide range of peptide−nucleic acid binding affinities. Fluorescence anisotropy is generally applied to monitor protein−nucleic acid interactions through the use of nucleic acids labeled with fluorescent tags.

DNA targets for these peptides were designed to be as simple as possible, minimizing the likelihood of having alternative binding sites and of forming stable secondary structures. For UBPCON1, the pentadeoxyribonucleotide d(TT**G**TT)-F was used where the 3′-end of the oligonucleotide was labeled with fluorescein through the use of a fluorescein-based phosphoramidite during synthesis. For UBPCON2U and UBPCON2N, two labeled oligonucleotides were used: d(TTT**GTTG**TTT)-F and d(TTTTTTTTTT)-F. In addition, four unlabeled oligonucleotides were used in competition studies: d(TTT**GTTG**TTT), d(TTTT**GTG**TTT), d(TT**GTTTG**TTT), and d(TTTTT**GG**TTT). Finally, for UBPCON3UU, two labeled oligonucleotides were used: (TTT**GTTGTTG**TTT)-F and d(TTTTTTTTTTTTT)-F.

The results of initial studies with UBPCON1 being titrated into a solution d(TTGTT)-F are shown in Figure 3A. An essentially linear increase in anisotropy was observed up to a concentration of 600 μM UBPCON1. No significant change in fluorescence intensity was observed. The linear increase in anisotropy suggests that this single-domain peptide binds with this oligonucleotide with a dissociation constant greater than 300 μM. This binding is substantially weaker than that reported for a single domain from the HIV-1 nucleocapsid protein binding to a DNA pentanucleotide.[27] However, the present studies were performed in relatively high salt conditions (50 mM MES, 150

(46) Murakami, A.; Nakaura, M.; Nakatsuji, Y.; Nagahara, S.; Tran-Cong, Q.; Makino, K. *Nucleic Acids Res.* **1991**, *19*, 4097–4102.
(47) LeTilly, V.; Royer, C. A. *Biochemistry* **1993**, *32*, 7753–7758.
(48) Heyduk, T.; Ma, Y.; Tang, H.; Ebright, R. H. *Methods Enzymol.* **1996**, *274*, 492–503.

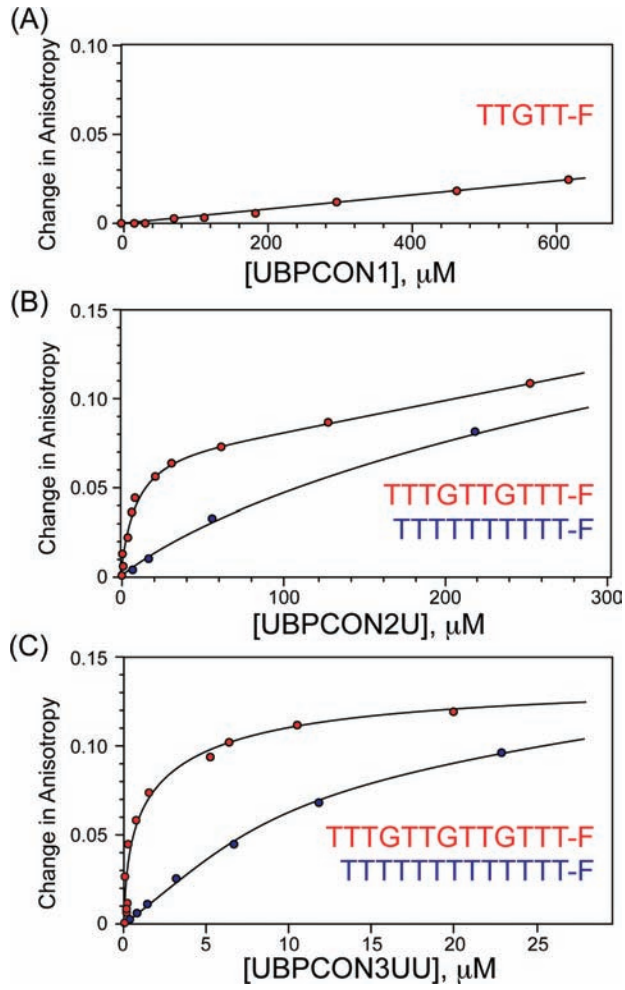

Figure 3. Plots of the changes in fluorescence anisotropy associated with oligonucleotide binding by peptides with one (A), two (B), or three (C) CCHC-box domains joined by linkers derived from UMSBP. Data are shown for oligonucleotides containing appropriately spaced guanosine residue (red) and the corresponding oligonucleotide containing only thymidine (blue), with fit curves in black.

mM NaCl), whereas previous studies were monitored by NMR in solutions with very low ionic strength.

UBPCON2U with two CCHC-box domains separated by the linker with sequence and length from UMSBP was titrated into d(T_3GTTGT$_3$)-F and d(T_{10})-F. The results are shown in Figure 3B. No significant changes in fluorescence intensity occurred over the course of the titrations. The binding curve for d(T_3GTTGT$_3$)-F showed a strong and weak binding phase and was fit to a rectangular hyperbola plus a line to take into account the weak-binding event. The tight phase was fit to yield a dissociation constant of 4.4 ± 0.2 μM (±95% confidence intervals). The binding curve for d(T_{10})-F was fit in a similar manner, with the maximum anisotropy change of the tight phase and the slope of the linear phase constrained to values from the d(T_3GTTGT$_3$)-F titration to yield a dissociation constant of 144 ± 12 μM.

The peptide UBPCON3UU, which contains three UMSBP CCHC-box domains with each pair separated by the linker with sequence and length from UMSBP, bound d(T_3GTTGTTGT$_3$)-F with a tight and weak phase in a manner similar to that for UBPCON2U. The results are shown in Figure 3C. The dissociation constant determined for the tight phase was 1.1 ± 0.1 μM. Similarly, UBPCON3UU bound d(T_{13})-F with a dissociation constant of 19.4 ± 1.1 μM.

The ability of the peptide UBPCON2N (with a seven amino acid linker derived from the HIV-1 nucleocapsid protein) to bind d(T_3**GTTG**T$_3$)-F was examined. This oligonucleotide was used so that a direct comparison with the results for UBPCON2U could be made. Studies described subsequently allowed examination of the effect of spacing between the guanosine residues on oligonucleotide binding by these two proteins. As shown in Figure 4A, the titration of UBPCON2N into a solution of d(T_3**GTTG**T$_3$)-F resulted in both a substantial increase in anisotropy and a significant decrease in fluorescence intensity. Because of the substantial change in fluorescence intensity over the course of the titration, the fraction bound (F_{bound}) was calculated from the anisotropy data as described in the Experimental Section. The intensity of the fluorescence continued to decrease even after the calculated fraction bound was essentially constant, suggesting that additional nonspecific binding also occurs as was observed for UBPCON2U and UBPCON3U. The fraction bound values were fit to yield a dissociation constant of 2.7 ± 0.4 μM, as shown in Figure 4B.

In order to address the influence of spacing between guanosine residues on the ability of the two-domain peptides to bind oligonucleotides, competition experiments with unlabeled oligonucleotides were performed. In these experiments, the peptides UBPCON2U and UBPCON2N were added to d(T_3**GTTG**T$_3$)-F until binding was ~90% saturated. At this point, an unlabeled oligonucleotide was added, and the loss of fluorescence anisotropy was monitored. For each protein, three unlabeled oligonucleotides were used. For UBPCON2U, these sequences were d(T_4**GTG**T$_3$), d(T_3**GTTG**T$_3$), and d(T_2**GTTTG**T$_3$). For UBPCON2N, these sequences were d(T_5**GG**T$_3$), d(T_4**GTG**T$_3$), and d(T_3**GTTG**T$_3$). The

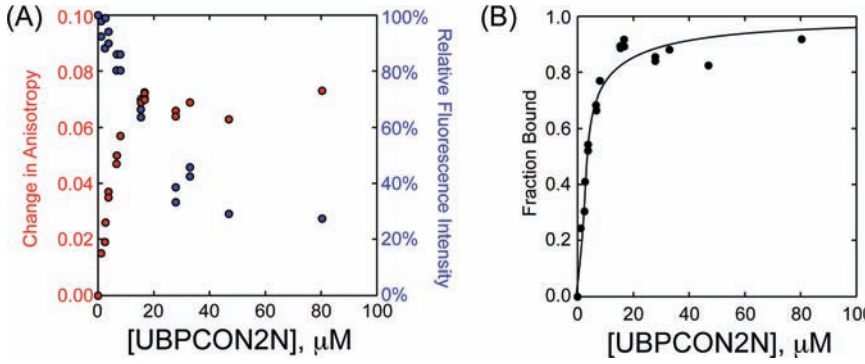

Figure 4. Binding studies with the peptide with two CCHC-box domains joined by the linker from the HIV-nucleocapsid protein with d(TTT**GTTG**TTT)-F. (A) Increase in fluorescence anisotropy (red) and decrease in fluorescence intensity. (B) Binding curve fit to the fraction bound, determined by the changes in anisotropy, corrected for the changes in fluorescence intensity.

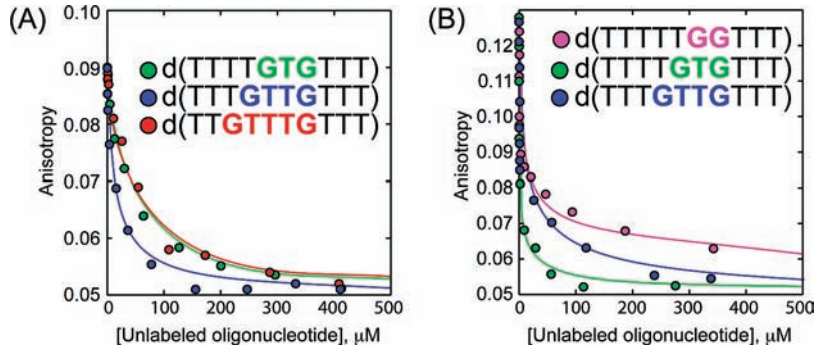

Figure 5. Effects of differences in spacing between guanosine residues on the binding of (A) UBPCON2U and (B) UBPCON2N determined by competition experiments. UBPCON2U preferentially bound the oligonucleotide with the guanosines separated by two thymidines, whereas UBPCON2N preferentially bound the oligonucleotide with the guanosines separated by a single thymidine.

results of these experiments are shown in Figure 5. There experiments demonstrate that UBPCON2U shows a preference for binding the oligonucleotide with two guanosine residues separated by two nucleotides over oligonucleotides with guanosine residues separated by either one or three nucleotides, in keeping with expectations. Fitting these data with a competitive binding model[49] suggests that the degree of preference is approximately 3-fold in each case. In contrast, UBPCON2N shows a preference for binding the oligonucleotide with guanosine residues separated by one nucleotide over oligonucleotides with guanosine residues separated by either none or two nucleotides. In this case, the level of discrimination is somewhat higher, with apparent preferences estimated to be 7- to 14-fold.

Summary and Conclusions

We have utilized the CCHC-box domain as a module for designing a series of peptides that were predicted to bind single-stranded oligonucleotides that contain appropriately spaced guanosine residues. Through the use of a binding assay based on fluorescence anisotropy measurements with fluorescein-labeled oligonucleotides, we demonstrated that these peptides preferentially bind guanosine-containing oligonucleotides over oligonucleotides lacking guanosine, that the affinity for guanosine-containing oligonucleotides increases as the number of CCHC-box domains is increased, and that the length of the linker between CCHC domains modulates the preference for oligonucleotides with different spacings between guanosine residues in a manner that is consistent with structural modeling.

These results represent only the first step in examining the properties of designed CCHC-box-based peptides for binding single-strand nucleic acids, as we have simplified the system as much as possible to facilitate these initial studies. We have used deoxyribonucleotides containing only guanosine and thymidine that have not been optimized and have not modified our original peptide designs in any way. Nonetheless, these results clearly demonstrate that our design principles generated peptides that have the nucleic acid binding properties anticipated in the generation of the peptides.

Acknowledgment. This work was supported by the intramural research program at the National Institute of Diabetes and Digestive and Kidney Diseases and by the National Institute of General Medical Sciences. We thank Drs. Gregory J. Gatto, Jr., and Barbara Amann for advice and assistance.

JA910942V

(49) Wang, Z. X. *FEBS Lett.* **1995**, *360*, 111–114.