Accelerated Publications

Minimum Length of a Sequence-Specific DNA Binding Peptide†

Robert V. Talanian,‡ C. James McKnight, Rheba Rutkowski, and Peter S. Kim*

Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, Massachusetts 02142

Received March 24, 1992; Revised Manuscript Received May 28, 1992

ABSTRACT: NMR experiments show that a stable complex can be formed between a 14-base-pair oligonucleotide and a disulfide-bonded dimer of a peptide containing 27 residues of the basic region of the yeast transcriptional activator GCN4; the complex is in slow exchange on the NMR time scale. In contrast, a nonspecific complex is in fast exchange on the NMR time scale. DNase I footprinting experiments show that dimers of peptides containing as few as 20 residues of GCN4 bind DNA with sequence specificity similar to that of the intact protein. Circular dichroism experiments suggest that specific binding involves only 15 residues, corresponding to residues 231–245 of GCN4, in an α-helical conformation. These results limit substantially the region of GCN4 involved in sequence-specific DNA contacts and provide a uniquely simple model for studying protein–DNA interactions in detail.

The "bZip" is a structural motif responsible for DNA binding in several transcriptional activator proteins including yeast GCN4, mammalian C/EBP, and the nuclear oncogene products Fos and Jun (Vinson et al., 1989). Compared to other DNA binding motifs such as the zinc finger and helix–turn–helix, the bZip motif is simple, consisting of an α-helical coiled coil, the "leucine zipper" (Landschulz et al., 1988; O'Shea et al., 1989), that mediates dimerization, and an adjacent α-helical region rich in basic amino acids, the "basic region", that contacts DNA.

The DNA binding activity of GCN4 can be modeled with a short peptide (GCN4-br1) that contains 31 residues from the basic region of GCN4 plus the C-terminal linker Gly-Gly-Cys, which replaces functionally the leucine zipper (Talanian et al., 1990). The oxidized peptide dimer (GCN4-br1^{SS}), but not the reduced monomer, binds to the GCN4 recognition element 5′-ATGACTCAT-3′ (GCRE)[1] in a sequence-specific manner as judged by gel mobility shift and DNase I footprinting assays. Thus, the essential contribution of the leucine zipper to DNA binding appears to be dimerization, and the basic region contains sufficient information for specific DNA binding (Talanian et al., 1990).

A prediction of both the "scissors grip" (Vinson et al., 1989) and "induced helical fork" (O'Neil et al., 1990) models of the bZip motif is that the basic region binds DNA in an α-helical conformation. Circular dichroism (CD) spectroscopy reveals that GCN4-br1^{SS}, which does not contain a leucine zipper, assumes an α-helical structure on binding DNA (Talanian et al., 1990), consistent with other studies of bZip proteins (O'Neil et al., 1990, 1991; Weiss, 1990; Patel et al., 1990; Weiss et al., 1990).

Here we report efforts to minimize the size of the GCN4-br1 peptide while retaining sequence-specific DNA binding activity. Three criteria are used to evaluate the shorter peptides: DNase I footprinting, NMR analysis, and stability measurements using CD spectroscopy.

MATERIALS AND METHODS

GCN4-br1 was prepared as described (Talanian et al., 1990). Other peptides were synthesized by tBoc chemistry on an Applied Biosystems (ABI) Model 430A peptide synthesizer using standard HOBt/NMP cycles modified to include acetic anhydride capping. Peptides were cleaved by low–high HF or TFMSA cleavage and purified as described (Talanian et al., 1990). Peptide stock concentrations were determined in triplicate by the method of Zahler and Cleland (1968). Fast atom bombardment mass spectra (Mass Search, Modesto, CA, or M-Scan, West Chester, PA) gave molecular masses within ±1 Da of calculated values. Oligonucleotides were synthesized on an ABI Model 392 DNA synthesizer and purified as described in ABI User Bulletin No. 50.

DNase I Footprinting. DNase I footprint assays (Galas & Schmidt, 1978) were performed as described (Talanian et al., 1990). DNA probes were prepared from plasmids pUC9-Sc4251 (Hill et al., 1986) (GCRE probe) or pCRE (CRE probe) as described (Talanian et al., 1990). pCRE is a derivative of pUC9-Sc4251 with the GCRE site 5′-ATGACTCAT-3′ substituted with the cAMP response element (CRE) 5′-ATGACGTCAT-3′. Footprint gels were quantitated with a Betagen Betascope 603 blot analyzer.

Circular Dichroism Spectroscopy. CD experiments were performed on an AVIV Model 60DS or 62DS CD spectrometer with a 5- or 10-mm path-length cell. Samples in 10 mM phosphate (pH 7.0) and 100 mM NaCl contained 5 μM peptide dimers and 5 μM oligonucleotides (double stranded) when

† This research was supported by the National Institutes of Health (GM44162). R.V.T. was supported by a National Research Service Award (GM13665) from the National Institutes of Health. C.J.M. is a postdoctoral fellow of the American Cancer Society. P.S.K. is a Pew Scholar in the Biomedical Sciences and a Rita Allen Foundation Scholar.

‡ Present address: BASF Bioresearch Corp., 195 Albany St., Cambridge, MA 02139.

[1] Abbreviations: GCRE, GCN4 recognition element; CD, circular dichroism; CRE, cAMP response element; NMR, nuclear magnetic resonance. Peptide abbreviations are explained in Table I. Abbreviations for the amino acids are A, Ala; C, Cys; D, Asp; E, Glu; G, Gly; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; and T, Thr.

0006-2960/92/0431-6871$03.00/0

present. All melts were >90% reversible and were baseline corrected but not smoothed.

NMR Spectroscopy. NMR spectra were collected at 20 °C on a Bruker AMX 500 NMR spectrometer using a selective 1–1 pulse with a spin echo sequence for water peak suppression (Sklenár & Bax, 1987). Chemical shift values are relative to 2,2,3,3-tetradeuterio-3-(trimethylsilyl)propionic acid (De Marco, 1977). Samples contained 200 μM oligonucleotides (double stranded) and various peptide concentrations with 50 mM phosphate buffer (pH 6.5) in 90% H_2O/10% D_2O. Spectra were collected as 2048 scans of 8K complex data points.

RESULTS

The Minimum-Length GCN4 Basic Region Peptide That Binds DNA Sequence Specifically. Twelve truncated analogs of the peptide GCN4-br1 were prepared synthetically (Table I). Each contained the C-terminal flexible linker Gly-Gly-Cys (Talanian et al., 1990; O'Shea et al., 1989) to allow covalent dimer formation. In addition, the peptides were synthesized as C-terminal amides, and the N-termini were acetylated to avoid introduction of charges and possible unfavorable interactions with the α-helix dipole [e.g., Shoemaker et al. (1987)].

For each peptide dimer, DNase I footprinting was tested over a range of peptide concentrations, typically 0.05–20 μM. Intact GCN4 protein and GCN4 bZip peptides bind in vitro to two different DNA binding sites with similar affinity (Sellers et al., 1990; Weiss et al., 1990); probes corresponding to both sites were used here. One probe contains the optimized GCRE sequence 5′-ATGACTCAT-3′ (Hill et al., 1986). The other probe contains the sequence 5′-ATGACGTCAT-3′, corresponding to the ATF/CREB and cAMP response element (CRE) sites (Hai et al., 1988; Roesler et al., 1988). The results (Table I) indicate that CRE site binding is more tolerant to peptide truncation than is GCRE binding. As judged by DNase I footprinting, GCN4-br6^{SS}, containing 20 residues of GCN4, is the shortest peptide dimer that retains specific binding to the CRE site.

The greater affinity of some peptides for the CRE compared to the GCRE is illustrated in Figure 1. GCN4-br10^{SS} binds the CRE with 2–3-fold greater affinity than the GCRE. This result supports the hypothesis that the true half-site recognized by GCN4 is 5′-ATGAC-3′ (Sellers et al., 1990).

Analysis of Peptide–DNA Complex Stability by Circular Dichroism Spectroscopy. We measured peptide–DNA complex stabilities to assess the contributions of various basic region residues to DNA binding. Peptides were incubated with a self-complementary oligonucleotide (CRE14, 5′-GGATGACGTCATCC-3′) in a 1:1 molar ratio (peptide dimer:oligonucleotide dimer) at 5 μM, more than 100 times the K_D of a GCN4-br1^{SS}–GCRE complex (Talanian et al., 1990). The CD signal at 222 nm, a measure of α-helicity (Woody, 1985), was monitored as a function of temperature for each complex (Figure 2). Peptides containing the C-terminal sequence KLQRMKQ bind DNA with significantly greater thermal stability (Figure 2A) than peptides lacking this sequence (Figure 2B). Further C-terminal truncations result in dramatic decreases in the stabilities of the peptide–DNA complexes (Figure 2C).

The thermal stability of GCN4-br5^{SS} in the presence of CRE14 (Figure 2A) was surprising given the failure of that peptide to bind the CRE site under the conditions used for DNase I footprinting (Table I). The stability of a nonspecific DNA complex of GCN4-br5^{SS} with DNA was tested using

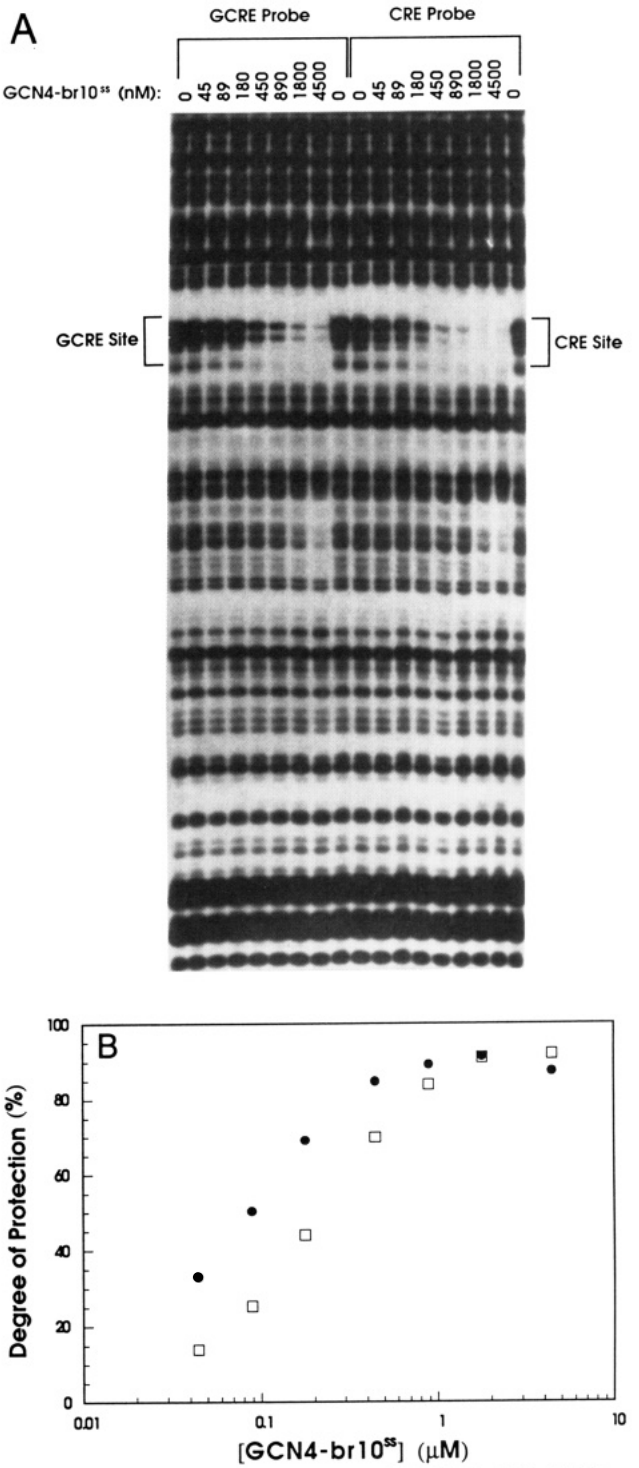

FIGURE 1: DNase I footprint titrations of GCN4-br10^{SS}. (A) Autoradiogram of footprint gel. At the highest peptide concentrations tested, a second site present in both probes and containing the sequence 5′-GTGACTGGG-3′ also is bound specifically. (B) Quantitation of DNase I protection. Each lane of the gel in (A) was quantitated with a beta scanner. The degree of protection was calculated as the ratio of the radioactivity within each site [as indicated on (A)] to that outside of the site divided by the value obtained for control (no peptide) lanes. Background was measured in the empty lanes adjacent to the outer lanes of the gel. Symbols: (□) GCRE probe; (●) CRE probe.

the oligonucleotide NON14 (5′-GGATCTCGAGATCC-3′), which has the same base composition as CRE14 but has two bases within each half-site inverted (underlined). NON14 induces a significant α-helical CD signal in GCN4-br5^{SS} at 0 °C (Figure 2D), but unlike the result obtained with CRE14, the shape of the thermal transition suggests noncooperative

Table I: DNase I Footprinting of the GCRE and CRE Sites by GCN4 Basic Region Peptides

peptide	sequence[a]	binding site[b] GCRE	CRE
N-terminal truncations			
GCN4-br1	PESSDPAALKRARNTEAARRSRARKLQRMKQ	+	+
GCN4-br2	DPAALKRARNTEAARRSRARKLQRMKQ	+	+
GCN4-br3	PAALKRARNTEAARRSRARKLQRMKQ	–	+
GCN4-br4	AALKRARNTEAARRSRARKLQRMKQ	–	+
GCN4-br5	KRARNTEAARRSRARKLQRMKQ	–	–[c]
GCN4-br6	DPAALKRARNTEAARRSRAR	–	+
GCN4-br7	PAALKRARNTEAARRSRAR	–	–
GCN4-br8	AALKRARNTEAARRSRAR	–	–
GCN4-br9	LKRARNTEAARRSRAR	–	–
C-terminal truncations			
GCN4-br2	DPAALKRARNTEAARRSRARKLQRMKQ	+	+
GCN4-br10	DPAALKRARNTEAARRSRARKLQ	+	+
GCN4-br6	DPAALKRARNTEAARRSRAR	–	+
GCN4-br11	DPAALKRARNTEAARRSRA	–	–
GCN4-br12	DPAALKRARNTEAARRSR	–	–
GCN4-br13	DPAALKRARNTEAARR	–	–

[a] Each peptide contained in addition the sequence GGC at its C-terminus. Each peptide was synthesized as the C-terminal amide and, except for GCN4-br1, was acetylated at the N-terminus. Note that some peptides fit into both the N- and C-terminal series of truncated peptides. The underlined residues of GCN4-br1^{SS} are those in common between GCN4-br5^{SS} and GCN4-br6^{SS} and probably include all of the residues that make sequence-specific contacts with DNA (see text). [b] DNase I footprinting assays were performed as described under Materials and Methods and scored positive if selective GCRE or CRE site protection was observed. [c] Although the footprinting assay scored negative, GCN4-br5^{SS} binds to the CRE site sequence specifically as judged by CD (Figure 2D).

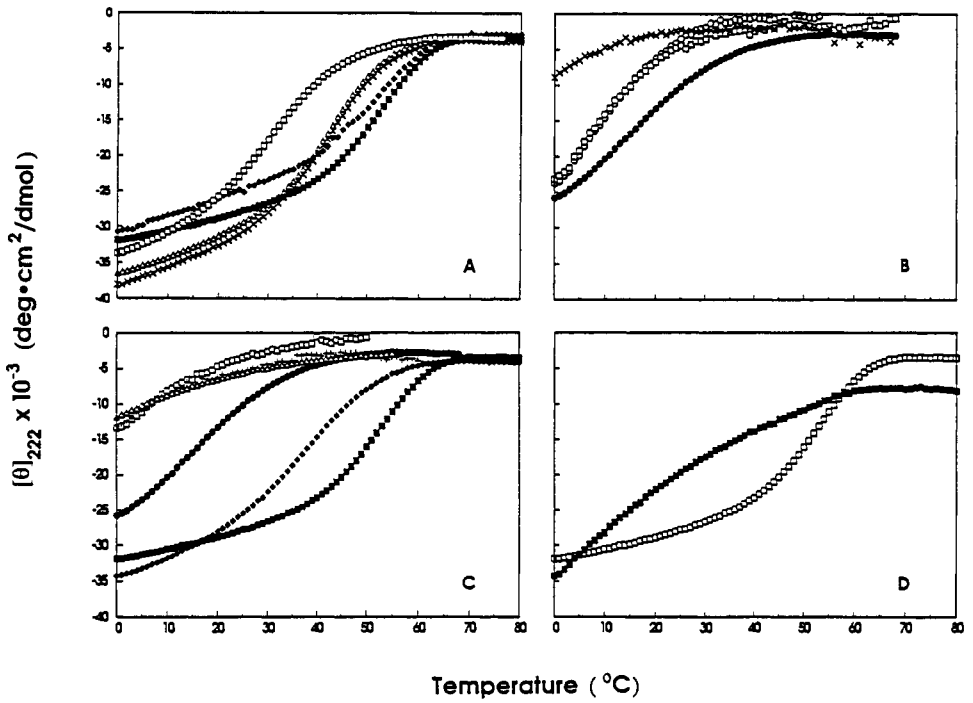

FIGURE 2: CD melts of peptides in complex with CRE14. (A) N-Terminal truncations: (◆) GCN4-br1^{SS}; (■) GCN4-br2^{SS}; (×) GCN4-br3^{SS}; (△) GCN4-br4^{SS}; (□) GCN4-br5^{SS}. (B) N-Terminal truncations lacking the C-terminal sequence KLQRMKQ: (●) GCN4-br6^{SS}; (○) GCN4-br7^{SS}; (□) GCN4-br8^{SS}; (×) GCN4-br9^{SS}. (C) C-Terminal truncations: (■) GCN4-br2^{SS}; (◆) GCN4-br10^{SS}; (●) GCN4-br6^{SS}; (□) GCN4-br11^{SS}; (+) GCN4-br12^{SS}; (△) GCN4-br13^{SS}. (D) GCN4-br5^{SS} in complex with (■) NON14 or (□) CRE14.

binding. The difference between the melted baselines of the two complexes can be accounted for entirely by the difference in CD signals of the isolated oligonucleotides at 222 nm (not shown). We conclude that GCN4-br5^{SS} binds to CRE14 in a sequence-specific manner.

One-Dimensional NMR Analysis of a Peptide–DNA Complex. Figure 3A shows the base imino proton NMR resonances of CRE14 at various GCN4-br2^{SS}:CRE14 molar ratios. When the peptide binds, there are changes in the chemical shifts for the imino protons of the oligonucleotide. As peptide is added, resonances corresponding to unbound imino protons decrease in intensity while new resonances corresponding to the bound state appear. This indicates that the peptide–DNA complex is in slow exchange on the NMR time scale (τ > msec) [e.g., Otting et al. (1990)]. In contrast, addition of GCN4-br2^{SS} to a solution containing NON14 (Figure 3B) results in gradual changes in the imino resonance chemical shifts, demonstrating that this complex is in fast exchange on the NMR time scale. Further, the decrease in intensity of the NON14 imino resonances observed on addition of peptide is a consequence of peptide:DNA aggregation at molar ratios above 0.4, resulting in broadening and decreased intensity of all peptide and DNA resonances (not shown). Aggregation is not observed with CRE14. Thus, there are substantial differences in the lifetimes of specific and non-specific complexes.

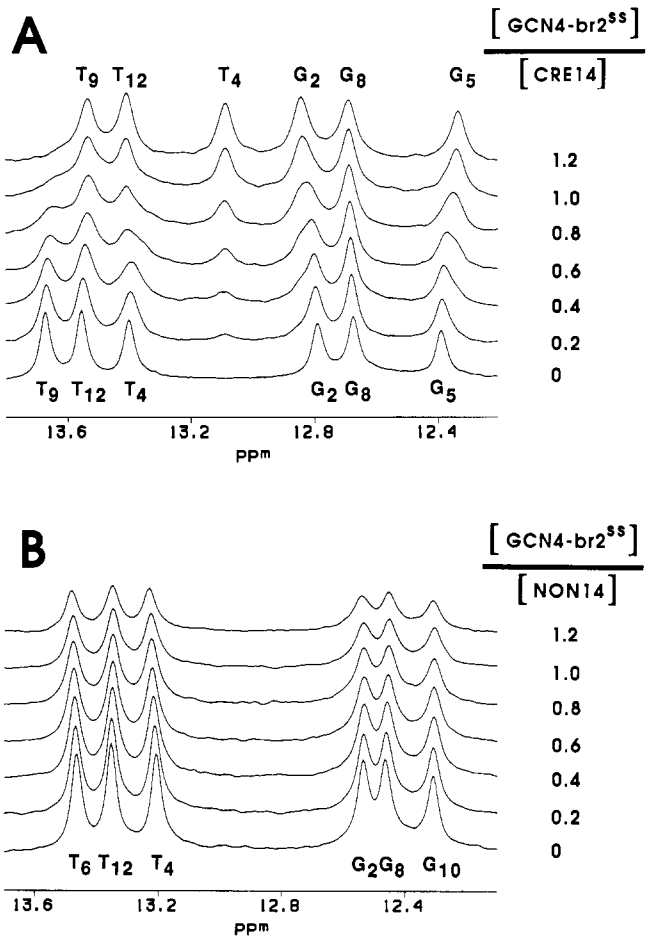

FIGURE 3: NMR titrations of CRE14 and NON14 by GCN4-br2^{SS}. Peptide:oligonucleotide ratios are as indicated. Imino proton resonance assignments were determined from two-dimensional NOESY spectra (Chou et al., 1983; Ulrich et al., 1983; Clore & Gronenborn, 1983; Boelens et al., 1985) of 740 μM CRE14, 1.65 mM NON14, or a stoichiometric mixture of approximately 1 mM GCN4-br2^{SS} and CRE14 (not shown). Key: (A) CRE14. (B) NON14.

DISCUSSION

The DNA binding activity of GCN4-br6^{SS}, containing only 20 residues from the basic region of GCN4, demonstrates that a remarkably short region of GCN4 is capable of binding DNA in a sequence-specific manner. Further, in the presence of seven additional residues at the C-terminus (GCN4-br2^{SS}), at least two of the N-terminal residues of GCN4-br6^{SS} are dispensable for footprinting (GCN4-br4^{SS}; Table I), and the five N-terminal residues are dispensable for sequence-specific DNA binding in CD experiments (GCN4-br5^{SS}; Figure 2D). Thus it appears that most or all of the GCN4 residues that make specific DNA contacts lie within the 15-residue region from GCN4 residues 231 to 245 (underlined for GCN4-br1 in Table I).

The present study also sheds light on the roles of the residues that flank the essential GCN4 basic region residues. In GCN4-br1^{SS} the four N-terminal residues (PESS) contribute little to DNA binding, since they are dispensable for DNase I footprinting (Table I) and do not increase significantly the thermal stability of the complex with CRE14 (Figure 2A). We hypothesize that these residues extend away from and do not contact the DNA. The seven C-terminal GCN4 residues of GCN4-br2^{SS} (KLQRMKQ) contribute significantly to complex stability; removal of these residues in GCN4-br6^{SS} decreases the thermal transition midpoint by >30 °C (Figure 2C). Removal of these residues, however, does not compromise DNase I footprinting (Table I), so it seems unlikely that they contact DNA directly. Rather, these residues may form the N-terminal portion of the leucine zipper coiled coil; the four C-terminal residues of GCN4-br2^{SS} (RMKQ) are the first four residues in a GCN4 leucine zipper peptide and are in a coiled-coil conformation in the crystal structure of this peptide (O'Shea et al., 1991). Additional leucine zipper residues should increase complex stability further, since a 58-residue GCN4 bZip peptide bound to a CRE site-containing oligonucleotide gave a CD melt with a transition midpoint of about 70 °C (Weiss, 1990).

Two structural models for bZip–DNA complexes have been proposed. The scissors grip model (Vinson et al., 1989) postulates that the two basic regions of a bZip dimer form α-helices and track from the base of the leucine zipper in opposite directions along a DNA major groove. The model also postulates a sharp bend in the basic region helix at a conserved asparagine (e.g., N235 of GCN4) which serves in the model as an α-helical "N-cap" residue (Richardson & Richardson, 1988). This bend would allow more extensive contacts between the basic region and DNA. The induced helical fork model (O'Neil et al., 1990) is similar but does not postulate a bend in the basic region helix. It is clear that the basic region forms an α-helix upon binding DNA (Talanian et al., 1990; O'Neil et al., 1990, 1991; Weiss, 1990; Patel et al., 1990; Weiss et al., 1990), but whether that helix is bent is not resolved. The present study limits the region of GCN4 that makes important sequence-specific contacts with DNA to 15 contiguous residues. The X-ray crystal structure of the MATα2 homeodomain–operator DNA complex (Wolberger et al., 1991) demonstrates that a helix of this length need not be kinked to make DNA contacts along its entire length, and thus it is unnecessary to postulate a bend in the basic region helix of GCN4 when bound to DNA. The simplicity of peptide models of the type studied here makes them attractive for studies of the structural basis for protein–DNA recognition and also for the design of DNA binding polypeptides and peptidomimetics.

ACKNOWLEDGMENT

We thank Michael Burgess for technical assistance in peptide synthesis.

REFERENCES

Boelens, R., Scheek, R. M., Dijkstra, K., & Kaptein, R. (1985) *J. Magn. Reson. 62*, 378–386.
Chen, Y.-S., Yang, J. T., & Chau, K. H. (1974) *Biochemistry 13*, 3350–3359.
Chou, S.-H., Hare, D. R., Wemmer, D. E., & Reid, B. R. (1983) *Biochemistry 22*, 3037–3041.
Clore, G. M., & Gronenborn, A. M. (1983) *EMBO J. 2*, 2109–2115.
De Marco, A. (1977) *J. Magn. Reson. 26*, 527–528.
Galas, D. J., & Schmidt, A. (1978) *Nucleic Acids Res. 5*, 3157–3170.
Hai, T., Liu, F., Allegretto, E. A., Karin, M., & Green, M. R. (1988) *Genes Dev. 3*, 2083–2090.
Hill, D. E., Hope, I. A., Macke, J. P., & Struhl, K. (1986) *Science 234*, 451–457.
Hinnebusch, A. G. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6442–6446.
Landschulz, W. H., Johnson, P. F., & McKnight, S. L. (1988) *Science 240*, 1759–1764.
O'Neil, K. T., Hoess, R. H., & DeGrado, W. F. (1990) *Science 249*, 774–778.
O'Neil, K. T., Shuman, J. D., Ampe, C., & DeGrado, W. F. (1991) *Biochemistry 30*, 9030–9034.

O'Shea, E. K., Rutkowski, R., & Kim, P. S. (1989) *Science 243*, 538–542.
O'Shea, E. K., Klemm, J. D., Kim, P. S., & Alber, T. (1991) *Science 254*, 539–544.
Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J., & Wüthrich, K. (1990) *EMBO J. 9*, 3085–3092.
Patel, L., Abate, C., & Curran, T. (1990) *Nature 347*, 572–575.
Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648–1652.
Roesler, W. J., Vandenbark, G. R., & Hanson, R. W. (1988) *J. Biol. Chem. 263*, 9063–9066.
Sellers, J. W., Vincent, A. C., & Struhl, K. (1990) *Mol. Cell. Biol. 10*, 5077–5986.
Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M. & Baldwin, R. L. (1987) *Nature 326*, 563–567.
Sklenár, V., & Bax, A. (1987) *J. Magn. Reson. 74*, 469–479.
Talanian, R. V., McKnight, C. J., & Kim, P. S. (1990) *Science 249*, 769–771.
Ulrich, E. L., John, E.-M. M., Gough, G. R., Brunden, M. J., Gilham, P. T., Westler, W. M., & Markley, J. L. (1983) *Biochemistry 22*, 4362–4365.
Vinson, C. R., Sigler, P. B., & McKnight, S. L. (1989) *Science 246*, 911–916.
Weiss, M. A. (1990) *Biochemistry 29*, 8020–8024.
Weiss, M. A., Ellenberger, T. E., Wobbe, C. R., Lee, J. P., Harrison, J. C., & Struhl, K. (1990) *Nature 347*, 575–578.
Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., & Pabo, C. O. (1991) *Cell 67*, 517–528.
Woody, R. W. (1985) in *The Peptides* (Undenfriend, S., Meienhofer, J., & Hruby, V. J., Eds.) pp 15–114, Academic Press, Inc., New York.
Zahler, W. L., & Cleland, W. W. (1968) *J. Biol. Chem. 243*, 716–719.