Pergamon

Bioorganic & Medicinal Chemistry 9 (2001) 2335–2339

BIOORGANIC & MEDICINAL CHEMISTRY

Selection of a High-Affinity DNA Pool for a bZip Protein with an Out-of-Phase Alignment of the Basic Region Relative to the Leucine Zipper

Young-sam Lee, Daniel G. Gurnon, Jessica J. Hollenbeck and Martha G. Oakley*

Department of Chemistry, Indiana University, 800 East Kirkwood Avenue, Bloomington, IN 47405-7102, USA

Received 16 February 2001; accepted 12 April 2001

Dedicated to Professor Peter B. Dervan in recognition of his many seminal contributions in the fields of bioorganic chemistry and chemical biology

Abstract—bZip transcription factors contain two regions that are required for DNA binding: a leucine zipper dimerization domain and a highly charged basic region that directly contacts DNA. The spacing between these subdomains is strictly conserved, and changes in this spacing result in a loss of function. Using an in vitro selection strategy, we have investigated the ability of a bZip protein with incorrect spacing between these two regions to bind specifically to DNA. Surprisingly, we find that although such a protein does not bind to its predicted site, it is possible to isolate a pool of DNAs that bind with very similar affinity to that of GCN4 for its optimum DNA site.

Introduction

The basic-region leucine-zipper (bZip) group of transcriptional activators and protooncogenes recognize DNA through a conserved bipartite structural motif. This motif consists of a coiled-coil leucine zipper dimerization domain and a highly charged basic region that directly contacts DNA.[1–3] Although the leucine zipper region does not contact the DNA directly, it serves to position the two basic regions appropriately for DNA binding. The spacing of the basic region relative to the leucine zipper is strictly conserved across this class of proteins. Indeed, Struhl and co-workers have shown for the yeast transcriptional activator GCN4 that insertions or deletions that change the rotational orientation of the basic regions relative to the leucine zipper lead to a loss of function.[4] In contrast, insertions of seven amino acids, corresponding to two complete helical turns within the coiled-coil domain, are tolerated.[4] Using an in vitro selection strategy, we have investigated the ability of a bZip protein with incorrect spacing between these two regions to bind to DNA specifically. We show here that, although such a protein does not bind to its predicted site, it is possible to isolate a pool of DNAs that bind with very similar affinity to that of GCN4 for its optimum DNA site.

Results and Discussion

In naturally occurring bZip proteins, the basic region is N-terminal to the leucine zipper. We have recently designed a series of model bZip peptides in which the GCN4 basic region is placed C-terminal to the leucine zipper.[5] Because the correct spacing between an N-terminal leucine zipper and a C-terminal basic region was unknown, we tested seven 'reverse' GCN4 (RGCN4) peptides, each with a different number of amino acids between the leucine zipper and basic regions, to span the full 360° of rotational space. We defined the leucine zipper as residues 251–281 of GCN4 and the basic region as residues 231–246. The four residues between these regions were considered the 'linker' region for the parent protein, RGCN4(0). Other members of the series contain insertions or deletions in this region (Fig. 1A). Finally, we considered residues 226–230, which are also required to maintain full DNA binding affinity,[6] to be important for the basic region to assume a helical conformation. These residues have

*Corresponding author. Tel.: +1-812-855-4843; fax: +1-812-855-8300; e-mail: oakley@indiana.edu

0968-0896/01/$ - see front matter
PII: S0968-0896(01)00158-4

been replaced in the RGCN4 peptides by residues commonly found at the C-termini of helices.[7]

GCN4 binds with high affinity to the inverted repeat sites 5′-TGACTCA-3′[8–10] and 5′-TGACGTCA-3′,[11] with each monomer of the basic region recognizing the half-site 5′-TGAC-3′.[12,13] Because the basic regions in the RGCN4•DNA complex lie in the opposite orientation with respect to the DNA, we expected the RGCN4 binding site to be a palindromic repeat of the inverted half-site, 5′-GTCA-3′ (Fig. 1B). It has been shown previously that two basic regions covalently linked with a flexible disulfide bond at their N-termini recognize the INV4 site.[14] We assayed all seven RGCN4 proteins for binding to five 24-bp oligonucleotide duplexes (INV0-4; Fig. 1B) containing DNA sites that differed only in the spacing between the 5′-GTCA-3′ and 5′-TGAC-3′ half-sites. As expected, we found that only one of these seven RGCN4 proteins is able to bind specifically and with high affinity to one of these rationally designed DNA sites.[5] This peptide, RGCN4(+3), binds to the INV2 site (Fig. 1B) with an affinity as high as that of GCN4 for its optimal site.[5]

Of the seven RGCN4 peptides tested, only RGCN4(+1) (Fig. 1) failed to bind to any of the five INV sites, even at a concentration of 6 μM.[5] RGCN4(+1) has two fewer residues in the 'linker' region than does RGCN4(+3). As the residues that make specific contacts in the major groove of DNA are appropriately positioned in RGCN4(+3), we expect them to be approximately 160° out-of-phase in RGCN4(+1), assuming an α-helical conformation of the RGCN4(+1) basic regions in the presence of DNA. Because the GCN4 basic region is largely disordered in the absence of DNA and becomes helical only upon DNA binding,[15–22] other conformations of the RGCN4(+1) basic regions must be considered. Similarly, as there is conformational flexibility in the side chains involved in DNA binding, the positions of the functional groups that make specific contacts with DNA cannot be predicted precisely. Thus, although it is clear that RGCN4(+1) cannot bind to the expected recognition sites, it is possible that a different spatial arrangement of the basic-region residues would lead to altered DNA-binding specificity.

To determine whether high-affinity DNA binding can occur at any DNA site for a bZip protein with incorrect spacing between the leucine zipper and basic regions, we carried out an in vitro selection assay[23,24] to search for high-affinity ligands for RGCN4(+1). Several groups have carried out selection experiments for GCN4. The optimum binding site for intact GCN4, 5′-TGACTCA-3′ was identified in an early selection experiment that did not include amplification steps.[9] GCN4-binding duplexes were isolated from a pool containing 23 randomized nucleotides through four cycles of affinity chromatography. Forty of the 43 selected DNA duplexes contained all seven nucleotides in this consensus sequence. More recently, an in vitro selection assay was carried out for the bZip domain of GCN4. In this assay, GCN4-binding DNAs were isolated from a pool containing 10 randomized base pairs through four rounds of gel retardation and PCR amplification.[10] Although the same optimum binding site was identified, a greater variety of sequences were obtained, with only 13/37 sequences matching the consensus 7-bp binding site. Indeed, some of the isolated DNAs bound to GCN4 with an affinity 10–20 times lower than that of the consensus site.[10] Nonetheless, these results clearly demonstrate that the bZip domain is sufficient to select the optimal DNA binding site for GCN4. We carried out a similar selection using the GCN4 bZip domain and a DNA pool with 12 randomized base pairs and obtained similar results.[25]

We used a gel retardation selection strategy to probe for high-affinity DNA ligands for RGCN4(+1). One pmol of oligonucleotide 66N18, which contains an 18-nucleotide randomized region (Fig. 2), was amplified by PCR and used for the first round of selection. Roughly 10 pmol (500 nM) of the double-stranded DNA pool were incubated with RGCN4(+1) (3 μM) in phosphate

A.

GCN4-56	basic region	linker	leucine zipper
	KDPAAL KRARNTEAARRSRARK	LQRV	KQLEDKVEELLSKNYHLENEVARLKKLVGER

RGCN4	leucine zipper	linker	basic region
(0)	M KQLEDKVEELLSKNYHLENEVARLKKLVGE	**LQRV**	KRARNTEAARRSRARK AALKG
(+1)	M KQLEDKVEELLSKNYHLENEVARLKKLVGE	**LQRRV**	KRARNTEAARRSRARK AALKG
(+3)	M KQLEDKVEELLSKNYHLENEVARLKKLVGE	**LQKLQRV**	KRARNTEAARRSRARK AALKG

B.

5′-ATAGCAG**GTCA**TA**TGAC**AGCGGAT-3′ 5′-ATTATA**GTCA**TCGA**TGAC**AAGGAT-3
3′-TATCGTC**CAGT**AT**ACTG**TCGCCTA-5′ 3′-TAATAT**CAGT**AGCT**ACTG**TTCCTA-5

INV2 **INV4**

Figure 1. (A) Comparison of the protein sequences of GCN4 and three RGCN4 variants. Standard single-letter amino acid abbreviations are used. (B) Sequences of the INV2 and INV4 oligonucleotides. The expected consensus half-sites for RGCN4, 5′-GTCA-3′ and 5′-TGAC-3′ are underlined in bold. INV0, INV1, and INV3 (not shown) are also 24 bp in length with 0, 1, and 3 base pairs between the half sites, respectively.

buffer at 4 °C for 30 min. The buffer also contained NaCl (350 mM) and polydeoxyinosinic-deoxycytidylic acid (poly[dI-dC]•poly[dI-dC]; 100 μM in base pairs) to minimize non-specific protein–DNA binding. Protein-bound DNAs were separated from free DNAs by gel retardation[26,27] and were then extracted from the gel and amplified by PCR. The isolated DNAs were subjected to further rounds of selection, and the protein concentration was reduced gradually to 100 nM. The progress of the selection was monitored by determining an apparent composite affinity constant (K_a) for RGCN4(+1) to intermediate and final pools of DNAs in the presence of excess non-specific DNA. Although it is clearly not possible to assign a single apparent affinity constant to a heterogeneous mixture of DNAs, determining an apparent composite K_a provides a convenient measure of the overall affinity of the DNA pool. Both the apparent composite K_a and the maximum percentage of bound DNAs increased as the stringency of the selection increased. At 12 rounds of selection, no further increase in the apparent composite K_a was observed.

The pool resulting from the 12th round of selection was amplified independently by three different workers, and the apparent composite dissociation constant (K_d) was determined to be $(3 \pm 1) \times 10^{-7}$ M (Fig. 3). This number is surprisingly close to the apparent K_d of $2–3 \times 10^{-7}$ M observed for our final GCN4-selected pool.[25] Half of the sequences obtained for the GCN4-selected DNAs contained the consensus seven-base pair site, which binds to the GCN4 bZip domain with a K_d of approximately 9×10^{-8} M under these conditions.[28] We therefore predict that the RGCN4(+1)-selected pool contains a high proportion of sequences that bind to RGCN4(+1) with an affinity similar to that of GCN4 for its consensus site.

66N18	5'-GCACCATATGGTCGCGTGGATCCT(N)$_{18}$CCGCACAGATCTCCAGGTACCCGC-3
66AP1	5'-GCACCATATGGTCGCGTGGATCCTGATGACTCATCCCGCACAGATCTCCAGGTACCCGC-3'
24P3	5'-GCACCATATGGTCGCGTGGATCCT-3'
24P4	5'-GCGGGTACCTGGAGATCTGTGCGG -3'

Figure 2. Sequences of oligonucleotides used for the RGCN4(+1) in vitro selection experiment. 66N18 is the pool used for the selection; 66AP1 contains the optimal binding site for GCN4 (underlined). 24P3 and 24P4 are the primers used for PCR amplification.

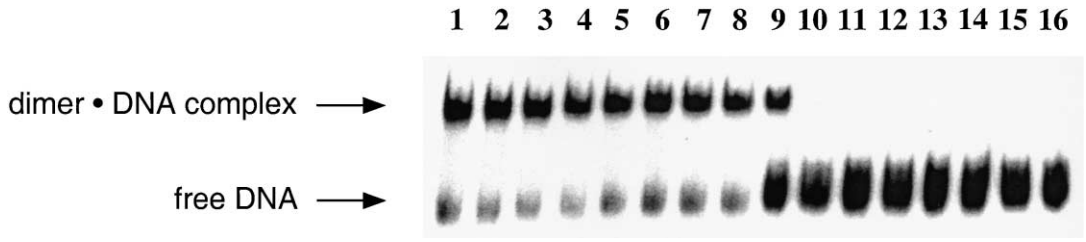

Figure 3. Gel mobility-shift assay of RGCN4(+1) binding to an oligonucleotide pool after 12 rounds of selection: 137 mM NaCl, 2.7 mM KCl, 4.3 mM sodium phosphate, 1.4 mM potassium phosphate, 1 mM EDTA, 1 mM DTT, 0.1% Igepal CA-630, 0.4 mg mL^{-1} BSA, 100 μM bp calf thymus DNA, pH 7.4. Lanes 1–15 contain ^{32}P-end-labeled DNA and decreasing amounts of peptide (8 μM, 6 μM, 4 μM, 2 μM, 1 μM, 800 nM, 600 nM, 400 nM, 200 nM, 100 nM, 80 nM, 60 nM, 40 nM, 20 nM, and 10 nM, respectively). Lane 16 contains end-labeled DNA only. Approximately 70% of the DNAs in the pool are bound at high concentrations of protein.

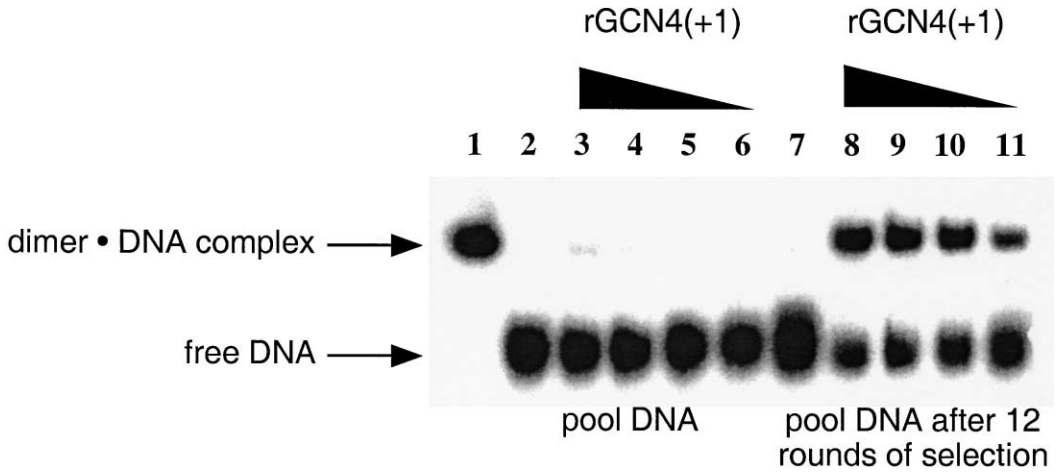

Figure 4. Equilibrium binding of RGCN4(+1) to oligonucleotide pool DNA under the conditions described for Figure 3. Lane 1, GCN4-56 (600 nM) binding to double-stranded 66AP1; lane 2, oligonucleotide pool DNA, lanes 3–6, RGCN4(+1) binding to oligonucleotide pool DNA at peptide concentrations of 1 μM, 500 nM, 250 nM, and 125 nM, respectively; lane 7, oligonucleotide pool DNA after 12 rounds of selection, lanes 8–11, RGCN4(+1) binding to oligonucleotide pool DNA after 12 rounds of selection at peptide concentrations of 1 μM, 500 nM, 250 nM, and 125 nM, respectively.

The binding by most of the DNAs in this pool appears to be specific. The selection was carried out in the presence of high concentrations of NaCl and in the presence of non-specific competitor DNA. These conditions have been shown previously to reduce binding by GCN4 to non-specific, but not to specific DNA sites.[9,28] In addition, little or no binding is observed by RGCN4(+1) to the randomized DNA pool at micromolar concentrations (Fig. 4). In contrast, a significant fraction of 12th-round-selected DNAs bind to RGCN4(+1) in the presence of excess non-specific DNA at concentrations as low as 125 nM. Finally, it should be noted that the mobility of the RGCN4(+1) complex is very similar to that of GCN4 bound to 66AP-1 (Fig. 4), strongly suggesting that RGCN4(+1) binds as a dimer to the selected DNAs.

Conclusions

We have isolated high-affinity, specific DNA ligands for a bZip protein with incorrect spacing between the basic region and leucine zipper. This result is particularly surprising as this protein, RGCN4(+1), fails completely to bind to five rationally designed sites, including one that can be bound by two basic regions with their N-termini connected by a flexible disulfide linker.[14] Changes in protein or DNA structure may be required for high-affinity protein·DNA complex formation, or unanticipated primary DNA sequences may be recognized. Further investigation is required to address these issues.

Experimental

Materials

Deionized water was further purified with a Corning MegaPure Filtration System. (Poly[dI-dC]·poly[dI-dC]) was purchased from ICN Pharmaceuticals, Inc. (Costa Mesa, CA) and dissolved in water to a concentration of 2 mM in base pairs. Enzymes were obtained from New England Biolabs (Beverly, MA) and used with the buffers supplied. Adenosine 5′-[γ-^{32}P]-triphosphate was purchased from ICN Pharmaceuticals, Inc. Oligonucleotides were supplied by Genosys Biotechnologies, Inc. (The Woodlands, TX). Cerenkov radioactivity was measured with a Packard Tri-Carb® 1600TR Liquid Scintillation Analyzer (Meriden, CT). A Molecular Dynamics 400E PhosphorImager and ImageQuant Software v. 3.3 were used to obtain data from storage phosphor screens purchased from Molecular Dynamics (Sunnyvale, CA).

In vitro selection

Construction of an expression vector for RGCN4(+1) will be described elsewhere.[5] RGCN4(+1) was expressed and purified from *Escherichia coli* as described previously.[28] Double-stranded 66N18 was obtained by standard PCR procedures, using ~1 pmol template, and ~160 pmol each of the 5′-^{32}P-labeled primers 24P3 and 24P4 (Fig. 2). The radioactive product was purified by non-denaturing polyacrylamide gel electrophoresis (8% polyacrylamide, 79:1 acrylamide/bisacrylamide) before use in the first round of selection. Complex formation was initiated by adding template DNA to RGCN4(+1) in phosphate-buffered saline (350 mM NaCl, 2.7 M KCl, 4.3 mM sodium phosphate, 1.4 mM potassium phosphate, 1 mM EDTA, 1 mM DTT, 0.1% Igepal CA-360, 0.4 mg/mL BSA, pH 7.4). This buffer was supplemented with poly[dI-dC]·poly[dI-dC] to a final concentration of 100 μM in base pairs. After a 30 min equilibration period at 4 °C, samples were subjected to polyacrylamide gel electrophoresis (8% polyacrylamide, 19:1 acrylamide/bisacrylamide) at 4 °C. A positive control lane containing 66AP-1 and GCN4-56 or similar was used to determine the location of protein-bound DNAs. The protein-bound DNAs were eluted from the gel and subjected to PCR amplification and purification as described above. 25–30% of the amplified DNA pool was used in the next round. Protein concentrations were: 3 μM for rounds 1 and 2, 750 nM for rounds 3–5, 400 nM for rounds 6 and 7, 200 nM for rounds 8–10, and 100 nM for rounds 11 and 12.

Quantitative electrophoretic mobility-shift assays

These assays were performed as described previously.[28,29] Protein concentrations ranged from 8 μM to 10 nM. The labeled DNA concentration was ~ 2 nM. After electrophoresis, gels were dried and analyzed using storage phosphor technology, and the amounts of free and bound DNA were quantified using ImageQuant v. 3.3 software. The fraction DNA bound (θ_{app}) was calculated as the volume contained within the band corresponding to the bound DNA divided by the sum of the volume present in the bands corresponding to the bound and free DNAs within each lane. Data were fit to the modified Hill eq 1 describing formation of a 2:1 protein/DNA complex:

$$\theta_{app} = \theta_{min} + (\theta_{max} - \theta_{min})\frac{K_a^2[L]_{tot}^2}{1 + K_a^2[L]_{tot}^2} \quad (1)$$

$[L]_{tot}$ corresponds to the total protein monomer concentration, K_a corresponds to the apparent composite monomeric equilibrium association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. Data were fit using the nonlinear least-squares fitting procedure of KaleidaGraph software (version 3.0.5, Abelbeck software) with K_a, θ_{max}, and θ_{min} as the adjustable parameters.

Acknowledgements

This work was supported by Grant GM57571 from the National Institutes of Health and Grant 32029-G4 from the American Chemical Society Petroleum Research Fund.

References and Notes

1. Kerppola, T. K.; Curran, T. *Curr. Opin. Struct. Biol.* **1991**, *1*, 71.
2. Hu, J. C.; Sauer, R. T. *Nucl. Acids Mol. Biol.* **1992**, *6*, 82.
3. Hurst, H. C. *Protein Profile* **1995**, *2*, 105.
4. Pu, W. T.; Struhl, K. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 6901.
5. Hollenbeck, J. J.; Gurnon, D. G.; Fazio, G. C.; Carlson, J. J.; Oakley, M. G. Submitted for publication.
6. Oakley, M. G.; Dervan, P. B. *Science* **1990**, *248*, 847.
7. Richardson, J. S.; Richardson, D. C. *Science* **1988**, *240*, 1648.
8. Hill, D. E.; Hope, I. A.; Macke, J. P.; Struhl, K. *Science* **1986**, *234*, 451.
9. Oliphant, A. R.; Brandl, C. J.; Struhl, K. *Mol. Cell. Biol.* **1989**, *9*, 2944.
10. Mavrothalassitis, G.; Beal, G.; Papas, T. S. *DNA Cell Biol.* **1990**, *9*, 783.
11. Sellers, J. W.; Vincent, A. C.; Struhl, K. *Mol. Cell. Biol.* **1990**, *10*, 5077.
12. Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223.
13. Keller, W.; Köönig, P.; Richmond, T. J. *J. Mol. Biol.* **1995**, *254*, 657.
14. Park, C.; Campbell, J. L.; Goddard, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9094.
15. Talanian, R. V.; McKnight, C. J.; Kim, P. S. *Science* **1990**, *249*, 769.
16. Weiss, M. A. *Biochemistry* **1990**, *29*, 8020.
17. Weiss, M. A.; Ellenberger, T.; Wobbe, C. R.; Lee, J. P.; Harrison, S. C.; Struhl, K. *Nature* **1990**, *347*, 575.
18. O'Neil, K. T.; Hoess, R. H.; DeGrado, W. F. *Science* **1990**, *249*, 774.
19. Saudek, V.; Pasley, H. S.; Gibson, T.; Gausepohl, H.; Frank, R.; Pastore, A. *Biochemistry* **1991**, *30*, 1310.
20. Cuenoud, B.; Schepartz, A. *Science* **1993**, *259*, 510.
21. O'Neil, K. T.; Shuman, J. D.; Ampe, C.; DeGrado, W. F. *Biochemistry* **1991**, *30*, 9030.
22. Talanian, R. V.; McKnight, C. J.; Rutkowski, R.; Kim, P. S. *Biochemistry* **1992**, *31*, 6871.
23. Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505.
24. Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818.
25. Lee, Y-s.; Oakley, M. G. Unpublished results.
26. Fried, M.; Crothers, D. M. *Nucleic Acids Res.* **1981**, *9*, 6505.
27. Garner, M. M.; Revzin, A. *Nucleic Acids Res.* **1981**, *9*, 3047.
28. Hollenbeck, J. J.; Oakley, M. G. *Biochemistry* **2000**, *39*, 6380.
29. Metallo, S. J.; Schepartz, A. *Chem. Biol.* **1994**, *1*, 143.