Specific DNA-RNA Hybrid Binding by Zinc Finger Proteins

Yigong Shi and Jeremy M. Berg

Zinc finger proteins of the Cys_2His_2 type represent a large class of proteins that have been assumed to function by means of specific interactions with DNA. Experiments motivated by structural characteristics of zinc finger protein–DNA complexes revealed that certain zinc finger proteins bound DNA-RNA hybrids with affinities comparable to or greater than those for DNA duplexes. The interactions between the zinc finger proteins and the DNA-RNA hybrids were dependent on which strand was RNA and were sequence-specific. Thus, interactions with DNA-RNA hybrids should be considered with regard to the biological roles of zinc finger proteins.

It has been estimated that up to 1000 proteins of the Cys_2His_2 zinc finger type are encoded by human and other higher eukaryotic genomes (*1*). For most of these proteins, no biological or biochemical role has yet been established. However, for nearly all zinc finger proteins that have been studied, sequence-specific DNA binding activity has been found in vitro. Other than for known RNA binding proteins, such as transcription factor IIIA (TFIIIA) (*2*) and another 5S RNA binding protein (*3*), the possibility that these proteins may act by binding other types of nucleic acids has not been previously explored. Here, we report that certain Cys_2His_2 zinc finger proteins can bind to DNA-RNA hybrids in a sequence-dependent manner.

Analyses of the crystal structures of zinc finger protein–DNA complexes (*4*, *5*) revealed that (i) the proteins have more contacts with one strand of the DNA than with the other, and (ii) the structures of the DNA in the complexes have characteristics of both A- and B-form DNA (*6*). This is illustrated by the Zif268-DNA cocrystal structure shown in Fig. 1. The major distinction between RNA and DNA is the presence of 2′ hydroxyl groups and the concomitant changes in conformation preferences. For more than half of the 2′ carbon atoms on the more contacted strand, distances to the nearest amino acid side chains are shorter than 4 Å, whereas no such close contacts are seen for the less contacted strand (*7*). On the basis of these observations, we examined the binding of a prototypical three–zinc finger DNA-binding unit from transcription factor Sp1 (*8*) to DNA-RNA hybrids. We reasoned that no unfavorable contacts should be made with the less contacted strand and that such hybrids should have structures distorted toward the A-form (*9*). The sequences of the oligonucleotides used are shown (Fig. 2) (*10*), and the results of the binding studies are indicated (Fig. 3) (*11*). The Sp1 unit bound to a 19-base pair DNA fragment (D1-D2) containing the site 5′-GGGGCGGGG-3′ with an apparent dissociation constant of 40 nM under our experimental conditions (*11*). Binding with very similar affinity was observed for the analogous DNA-RNA hybrid (D1-R2) with the DNA strand containing the guanine-rich site. This strand has been strongly implicated as the more contacted strand for Sp1 (*12*). In contrast, binding to the alternative DNA-RNA hybrid (R1-D2) was approximately one-tenth as strong, and that to the corresponding RNA-RNA duplex (R1-R2) was approximately one-hundredth as strong. The binding data are summarized in Table 1.

To further probe the specificity of these interactions, we examined the binding properties of ZF-QQR, a zinc finger protein with different site preferences (Fig. 2). This protein is a designed consensus sequence–based protein (*13*) and is known to bind preferentially to DNA sites of the form 5′-GGGGAAGAA-3′ (*14*). The purine-rich strand is almost certainly the more contacted strand (*15*). Binding studies (Fig. 3) revealed that ZF-QQR binds preferentially to the DNA-RNA hybrid (D3-R4) with the purine-rich site on the DNA strand. This hybrid was bound approximately five times as tightly as the DNA-DNA duplex (D3-D4). Moreover, it was bound approximately 30 times as tightly as the alternative hybrid (R3-D4) and more than 150 times as tightly as the RNA-RNA duplex (R3-R4) (Table 1). In addition, the binding of this protein to these sites was at least 100 times tighter than its binding to the corresponding Sp1 sites (*16*). This result directly demonstrates that the protein–DNA-RNA hybrid interactions are quite sequence-specific as well as strand-specific.

Two additional experiments were performed to explore the sequence-specific nature of the protein–DNA-RNA hybrid interactions (Fig. 4). First, a methylation interference assay (*17*) revealed that methylation of any of the guanines within the 5′-GGGGAAGAA-3′ site on the DNA strand strongly interfered with binding, whereas methylation of guanines outside this site did not. This finding suggests that the specificity-determining contacts in the complexes with the DNA-RNA hybrids are very similar to those observed in zinc finger protein–

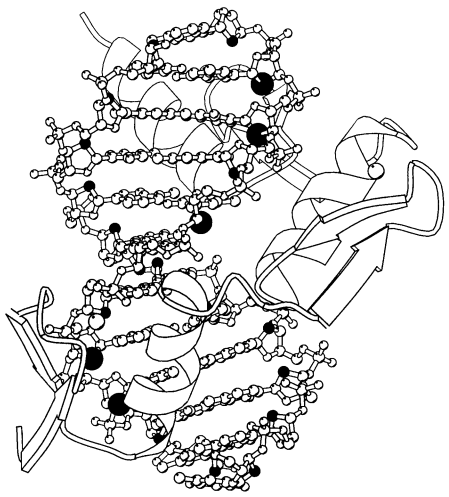

Fig. 1. A schematic view (*23*) of the structure of the three zinc finger–DNA complex from Zif268 (*4*). The 2′ carbon atoms are shown in black. The 2′ carbon atoms shown as larger spheres are less than 4 Å from the nearest protein atom. These atoms all lie on one strand of the DNA, which is also involved in specific protein-DNA hydrogen bonding interactions. Arrows represent β strands. The structures of complexes with DNA-RNA hybrids have not been determined and may be somewhat different.

A

```
D1   5'-TCACTGGGGCGGGGTCCTC-3'
R1      U   U          U  U

D2   5'-GAGGACCCCGCCCCAGTGA-3'
R2                     U

D3   5'-TCACTGGGGAAGAAGAATCCTC-3'
R3      U   U          U  U

D4   5'-GAGGATTCTTCTTCCCCAGTGA-3'
R4         UU UU UU       U
```

B

```
Sp1
                    M E K L R N G S G D P G K K K -
Q H I C H I Q G C G K V Y G K T S H L R A H L R W H T G E R -
P F M C T W S Y C G K R F T R S D E L Q R H K R T H T G E K -
K F A C P E - - C P K R F M R S D H L S K H I K T H Q N K K

ZF-QQR
                    M E K L R N G S G D P G K K K -
Q H A C P E C G K S F S Q S S N L Q K H Q R T H T G E K -
P Y K C P E C G K S F S Q S S N L Q K H Q R T H T G E K -
P Y K C P E C G K S F S R S D H L Q R H Q R T H Q N K K
```

Fig. 2. Oligonucleotide and zinc finger protein sequences. (**A**) The binding sequence is 5′-GGGGCGGGG-3′ for Sp1 and 5′-GGGGAAGAA-3′ for ZF-QQR. In our nomenclature, D denotes DNA, R denotes RNA, 1 and 3 correspond to the more contacted strand, and 2 and 4 correspond to the less contacted strand. (**B**) Amino acid sequences of the DNA binding domains of Sp1 and ZF-QQR. Metal-chelating residues are in boldface; presumed specificity-determining residues are underlined. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

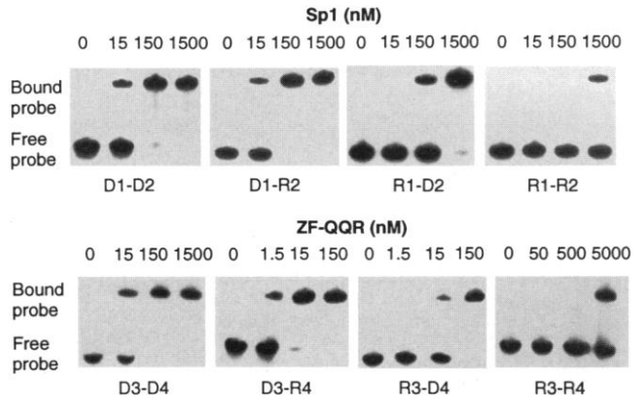

Fig. 3. Binding of Sp1 and ZF-QQR to DNA duplexes, to DNA-RNA hybrids, and to RNA duplexes (probes are named as described in Fig. 2). Upper and lower panels show the results of gel shift assays for Sp1 and ZF-QQR, respectively.

Table 1. Dissociation constants (K_d) of Sp1 and ZF-QQR binding to DNA duplexes, to DNA-RNA hybrids, and to RNA duplexes (estimated standard error, ±15%). The sequences of the nucleic acids used are given in Fig. 2.

Binding sequence	K_d (nM)
Sp1	
D1-D2	40
D1-R2	40
R1-D2	400
R1-R2	>3200
ZF-QQR	
D3-D4	15
D3-R4	2.8
R3-D4	76
R3-R4	>5000

DNA complexes. Second, a ribonuclease H (RNase H) footprinting experiment (*18*) was performed. This experiment revealed that 80 nM ZF-QQR completely and specifically protects the binding site region of the RNA strand of the hybrid from cleavage by *Escherichia coli* RNase H. In addition to providing data about binding specificity, this observation reveals that zinc finger binding is capable of blocking RNase H action in vitro. The sequences we used had been optimized for DNA duplex binding, and it is possible that the requirements for optimal DNA-RNA hybrid binding are somewhat different. With fully optimized sequences, even greater DNA-RNA hybrid specificity might be realized.

To determine whether DNA-RNA hybrid binding was a property common to other classes of DNA binding proteins, we examined the behavior of the homeodomain protein Ubx (*19*). This protein exhibited no detectable binding either to a DNA-RNA hybrid or to an RNA duplex under conditions where tight DNA binding was seen (*20*). This finding indicates that the preference of Ubx for DNA duplexes is more than 1000 times as great as for DNA-RNA hybrids or RNA duplexes, and it is consistent with the observation that nearly all classes of structurally characterized DNA binding proteins (except Cys_2His_2 zinc finger proteins) interact strongly with both nucleic acid strands. This property increases the likelihood that the introduction of 2′-hydroxyl groups, which are associated with converting either strand to RNA, will lead to significant steric clashes.

TFIIIA, the first member of the Cys_2His_2 zinc finger protein family to be discovered, binds both 5S RNA and the internal control regions of 5S RNA genes specifically (*2*). Subsequent studies (*21*), however, have revealed that TFIIIA is best regarded as a fusion protein between a specific DNA binding protein (zinc finger domains 1 to 3) and an RNA binding protein (domains 4 to 7). The DNA-RNA binding activity described here appears to be a general property common to many (although not necessarily all) members of this protein class. At present, we do not know of any specific biological system for such activity. Nonetheless, such heteroduplexes do occur during a number of biological processes, including gene transcription, DNA replication, and the reverse transcription of retroviruses and retrotransposons. In addition, our observations provide further experimental support for the suggestion that these proteins may prefer to bind to nucleic acids that can adopt conformations distinct from the canonical B-form (*6*). The structures of DNA-RNA hybrids are not well understood, although these hybrids appear to adopt structures along the A-B continuum (*9*). Finally, regardless of the occurrence of any natural zinc finger–heteroduplex interactions, our results suggest that such proteins could be used intentionally to target specific DNA-RNA hybrid structures in vivo.

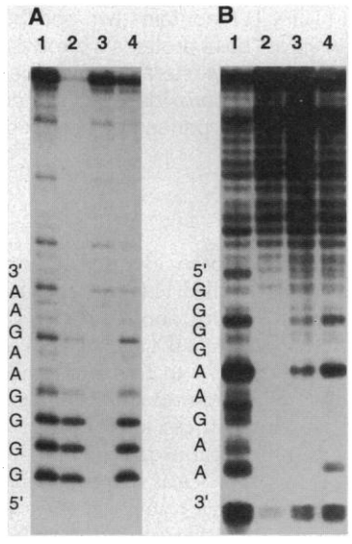

Fig. 4. (**A**) Methylation interference for ZF-QQR binding to a DNA-RNA hybrid. Lane 1, control, no ZF-QQR; lane 2, unshifted signal with 100 nM ZF-QQR; lane 3, shifted signal with 100 nM ZF-QQR; and lane 4, unshifted signal with 20 nM ZF-QQR. (**B**) RNase H footprinting for ZF-QQR binding to a DNA-RNA hybrid. Lane 1, control, no ZF-QQR; lane 2, 80 nM ZF-QQR; lane 3, 8 nM ZF-QQR; and lane 4, 0.8 nM ZF-QQR.

Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

REFERENCES AND NOTES

1. D. Rhodes and A. Klug, *Sci. Am.* **268**, 56 (February 1993).
2. H. R. Pelham and D. D. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4170 (1980).
3. K. E. Joho, M. K. Darby, E. T. Crawford, D. D. Brown, *Cell* **61**, 293 (1990).
4. N. P. Pavletich and C. O. Pabo, *Science* **252**, 809 (1991).
5. L. Fairall, J. W. R. Schwabe, L. Chapman, J. T. Finch, D. Rhodes, *Nature* **366**, 483 (1993).
6. L. Nekludova and C. O. Pabo, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6948 (1994).
7. Calculations were performed with the use of the coordinates of the Zif268-DNA complex (*4*).
8. J. T. Kadonaga, K. R. Carner, F. R. Masiarz, R. Tjian, *Cell* **51**, 1079 (1987).
9. M. Salazar, O. Y. Fedoroff, J. M. Miller, N. S. Ribeiro, B. R. Reid, *Biochemistry* **32**, 4207 (1993) (and references therein).
10. All DNA oligonucleotides were synthesized on an Applied Biosystems 392 synthesizer, deprotected at 55°C overnight, and purified through oligonucleotide purification cartridges (Applied Biosystems). RNA oligonucleotides were synthesized on Applied Biosystems 380B or 392 synthesizers and were processed as described in Applied Biosystems User Bulletin 69 (1992).
11. For gel mobility-shift experiments, 10 pmol of single-strand DNA or RNA were labeled at their 5′ ends with [γ-^{32}P]adenosine triphosphate (ATP) by T4 polynucleotide kinase (New England Biolabs), and 1.5 equivalents of the corresponding complementary strand were added to each reaction. The mixture was placed in a 90°C water bath and was then allowed to cool to room temperature over a period of 40 min. The annealed product was purified through an NAP-10 column (Pharmacia). The double-stranded species was precipitated in ethanol and further purified on a 10% native acrylamide gel (20:1 acrylamide:bisacrylamide) (*22*). The purified probe was resuspended in 3 ml of 10 mM tris-Cl (pH 7.6). The amount of probe used for each gel shift experiment was 0.2 μl (about 3000 cpm). The final concentration of the probe was estimated to be less than 0.1 nM. Various concentrations of a protein were incubated with the radioactive probe in a 20-μl gel shift buffer containing 25 mM tris (pH 8.0), 100 mM KCl, 2 mM dithiothreitol, 100 μM $ZnCl_2$, 10% glycerol, bovine serum albumin (BSA; 50 μg/ml), and polydeoxyinosinic-deoxycytidylic acid [poly(dI-dC)] (2 μg/ml) at 4°C for 30 min. The reaction mixtures were run on 10%

acrylamide gels (75:1 acrylamide:bisacrylamide) in 50 mM tris (pH 8.0) and 400 mM glycine buffer at 250 V for 1 hour. The proteins were expressed and purified as described (*13*). The protein concentrations were determined by absorbance at 280 nm with the use of extinction coefficients of 11,200 M^{-1} cm^{-1} for Sp1 and 2840 M^{-1} cm^{-1} for ZF-QQR. For each protein–nucleic acid pair, a pilot experiment was performed to estimate the dissociation constant. Then, five different concentrations of the protein, centered around the dissociation constant, were used to determine binding affinity. Four independent gel shift experiments were carried out for each protein–nucleic acid pair. The binding affinities determined from the independent experiments were found to be within 20% of one another. The radioactive signal was quantitated with PhosphorImager (Molecular Dynamics) analysis, and the data were fit with the use of the program KaleidaGraph (Synergy Software, Reading, PA).

12. J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11109 (1992).
13. J. R. Desjarlais and J. M. Berg, *ibid.* **90**, 2256 (1993).
14. The optimal binding site, 5′-GGGGAAGAA-3′, was derived from gel mobility-shift assay–based selection with the use of randomized DNA oligonucleotides.
15. The purine-rich strand is thought to be the more contacted strand because of the presence of five guanines, the first three of which, in analogy to the Zif268-DNA structure, are presumed to interact with the third zinc finger domain of ZF-QQR through arginine and histidine residues. This inference is also supported by the methylation interference studies, which indicated that methylation of any of these five guanines abolished DNA binding (Y. Shi and J. M. Berg, unpublished data).
16. Y. Shi and J. M. Berg, unpublished data.
17. A single-stranded DNA, 5′-CTCATGTCACTGGGGAAGAAGAAGAATCGATCTGATC-3′, was labeled at its 5′ end with [^{32}P]ATP, hybridized to its complementary RNA strand, and purified from 10% native acrylamide gel as described (*10*). The methylation interference assay was carried out as described [(*22*), pp. 12.3.1–12.3.4].
18. A single-stranded RNA, 5′-GAUCAGAUCGAUUCUUCUUCUUCCCCAGUGACAUGAG-3′, was labeled at its 5′ end with [^{32}P]ATP, hybridized to its complementary DNA strand, and purified as described (*10*). The RNase H protection assay was performed at 25°C for 2 min in a 100-μl reaction buffer containing 40 mM tris (pH 7.5), 100 mM KCl, 4 mM $MgCl_2$, 1 mM dithiothreitol, 4% glycerol, BSA (30 μg/ml), and poly(dI-dC) (2 μg/ml). One unit of *E. coli* RNase H (Pharmacia) was used for each reaction. The reaction was quenched with 400 μl of RNase H stop solution containing 20 mM EDTA, 92% ethanol (v/v), 750 mM ammonium acetate, and 4 μg of poly(dI-dC) at −70°C for 20 min. The precipitated probe was run on a 10% denaturing acrylamide gel.
19. S. C. Ekker, K. E. Young, D. P. von Kessler, P. A. Beachy, *EMBO J.* **10**, 1179 (1991).
20. Y. Shi and J. M. Berg, unpublished data.
21. K. R. Clemens *et al.*, *Science* **260**, 530 (1993).
22. A. S. Baldwin Jr., in *Current Protocols in Molecular Biology*, F. M. Ausubel *et al.*, Eds. (Wiley, New York, 1989), vol. 2, pp. 12.3.2–12.3.4.
23. P. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
24. Supported by NIH grant GM46257. We thank P. Beachy and S. Ekker for a sample of Ubx.

17 October 1994; accepted 3 February 1995

Killing of Trypanosomes by the Human Haptoglobin-Related Protein

Andrea B. Smith, Jeffrey D. Esko, Stephen L. Hajduk*

African trypanosomes cause disease in humans and animals. *Trypanosoma brucei brucei* affects cattle but not humans because of its sensitivity to a subclass of human high density lipoproteins (HDLs) called trypanosome lytic factor (TLF). TLF contains two apolipoproteins that are sufficient to cause lysis of *T. b. brucei* in vitro. These proteins were identified as the human haptoglobin-related protein and paraoxonase–arylesterase. An antibody to haptoglobin inhibited TLF activity. TLF was shown to exhibit peroxidase activity and to be inhibited by catalase. These results suggest that TLF kills trypanosomes by oxidative damage initiated by its peroxidase activity.

Trypanosomes are protozoan parasites of medical and veterinary importance. In humans, *T. brucei gambiense* and *T. brucei rhodesiense* cause African sleeping sickness. In animals, the disease *nagana* is caused by the parasite *T. b. brucei*. These three subspecies of trypanosomes are morphologically indistinguishable and differ only in their host range and pathologic presentation. The difference in host specificity is caused by a nonimmune killing factor in human serum; the African sleeping sickness trypanosomes are resistant to this factor (*1*).

The trypanocidal activity of normal human serum has been localized to a minor subset of HDLs called TLF, with the majority of serum HDL being nontoxic (*2*). TLF differs from nonlytic HDLs in that it has a higher density (1.21 to 1.24 g/ml) and contains two characteristic apolipoproteins, LI (94.5 kD) and LIII (45 kD) (*2*). LI is composed of three subunits that are 45, 36, and 13.5 kD in size, whereas LIII consists of two apolipoproteins that comigrate, one a 45-kD protein and the other composed of 36- and 13.5-kD subunits. Reconstitution studies have shown that LIII is required for lysis in micellar form in combination with either LI or AI, a major apolipoprotein found in HDLs (*3*).

To establish the identity of LI and LIII, we determined their NH_2-terminal sequences. The subunits of TLF were isolated by two-dimensional gel electrophoresis (*4*) and the individual bands were analyzed by Edman degradation. The LI and LIII 45-kD subunits yielded sequences that were identical to the uncleaved leader sequence of human paraoxonase–arylesterase. The NH_2-termini of the LI and LIII 36-kD subunits were identical to the β subunit of both haptoglobin and the haptoglobin-related protein. The LI and LIII 13.5-kD subunits yielded sequences that were identical (allowing for one sequencing ambiguity) to the uncleaved leader sequence of the α subunit of the haptoglobin-related protein (Fig. 1). Haptoglobin has a similar signal sequence that is normally cleaved at maturation. Haptoglobin is initially translated as a precursor containing the α and β subunits; cleavage at Arg^{84} occurs after a disulfide bond has formed between the two subunits (*5*). By analogy, it is likely that the 36-kD protein is the β subunit of the haptoglobin-related protein.

Human serum paraoxonase–arylesterase associates with HDL_2 and HDL_3 (*6*) and is known to hydrolyze aromatic esters, although its physiological role is unknown. It was an unlikely candidate for the TLF toxin because it is found in HDL_2, which has been shown to be nontoxic (*2*). Haptoglobins are serum glycoproteins that bind free hemoglobin and decrease the loss of iron from the body; however, overloading of the cell with haptoglobin-hemoglobin complexes can cause iron toxicity with the formation of reactive free radicals (*7*). Although it is expressed in amounts that are one-thousandth those of haptoglobin (*8*), the haptoglobin-related protein is predicted to have functional domains that are similar to those of haptoglobin (*9*). Because the haptoglobin-related protein appeared to be a potential toxin, we focused our studies on this protein.

To investigate whether the haptoglobin-related protein was a component of TLF, we measured TLF activity (*2*) in the presence of rabbit polyclonal antibodies to human haptoglobin that cross-react with the haptoglobin-related protein (Fig. 2A). These antibodies inhibited TLF-mediated lysis in a concentration-dependent fashion; this observation implies that the haptoglobin-related protein is part of the lytic molecule. In contrast, nonspecific rabbit polyclonal antibodies had no effect (*10*).

To determine whether the haptoglobin-related protein retained the ability to bind hemoglobin, we assayed for hemoglobin in purified TLF by means of size exclusion chromatography. Total protein absorbance at 280 nm (A_{280}), heme protein absorbance at 407 nm (A_{407}), and TLF lysis activity all coeluted with an apparent molecular mass of 500 kD (Fig. 2B). Hemoglobin was also detected in TLF fractions by immunoblot anal-

Department of Biochemistry and Molecular Genetics, School of Medicine and Dentistry, University of Alabama at Birmingham, Birmingham, AL 35294, USA.

*To whom correspondence should be addressed.