The recognition of a noncanonical RNA base pair by a zinc finger protein

Pilar Blancafort[1,2*], Sergey V Steinberg[1], Bruno Paquin[1], Roscoe Klinck[1†], Jamie K Scott[2] and Robert Cedergren[1‡]

Background: The zinc finger (ZF) is the most abundant nucleic-acid-interacting protein motif. Although the interaction of ZFs with DNA is reasonably well understood, little is known about the RNA-binding mechanism. We investigated RNA binding to ZFs using the Zif268–DNA complex as a model system. Zif268 contains three DNA-binding ZFs; each independently binds a 3 base pair (bp) subsite within a 9 bp recognition sequence.

Results: We constructed a library of phage-displayed ZFs by randomizing the α helix of the Zif268 central finger. Successful selection of an RNA binder required a noncanonical base pair in the middle of the RNA triplet. Binding of the Zif268 variant to an RNA duplex containing a G·A mismatch (rG·A) is specific for RNA and is dependent on the conformation of the mismatched middle base pair. Modeling and NMR analyses revealed that the rG·A pair adopts a head-to-head configuration that counterbalances the effect of S-puckered riboses in the backbone. We propose that the structure of the rG·A duplex is similar to the DNA in the original Zif268–DNA complex.

Conclusions: It is possible to change the specificity of a ZF from DNA to RNA. The ZF motif can use similar mechanisms in binding both types of nucleic acids. Our strategy allowed us to rationalize the interactions that are possible between a ZF and its RNA substrate. This same strategy can be used to assess the binding specificity of ZFs or other protein motifs for noncanconical RNA base pairs, and should permit the design of proteins that bind specific RNA structures.

Addresses: [1]Département de Biochimie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, QC, Canada, H3C 3J7. [2]Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6.

Present addresses: *Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. †Ribotargets, Ltd, Kett House, Station Road, Cambridge, CB1 2IP, UK.

‡Deceased

Correspondence: Pilar Blancafort and Jamie K Scott
E-mail: pilar@scripps.edu
jkscott@sfu.ca

Key words: double-helix structure, major-groove geometry, phage display, protein–nucleic-acid interaction, zinc finger

Received: **18 March 1999**
Revisions requested: **16 April 1999**
Revisions received: **12 May 1999**
Accepted: **17 May 1999**

Published: **21 July 1999**

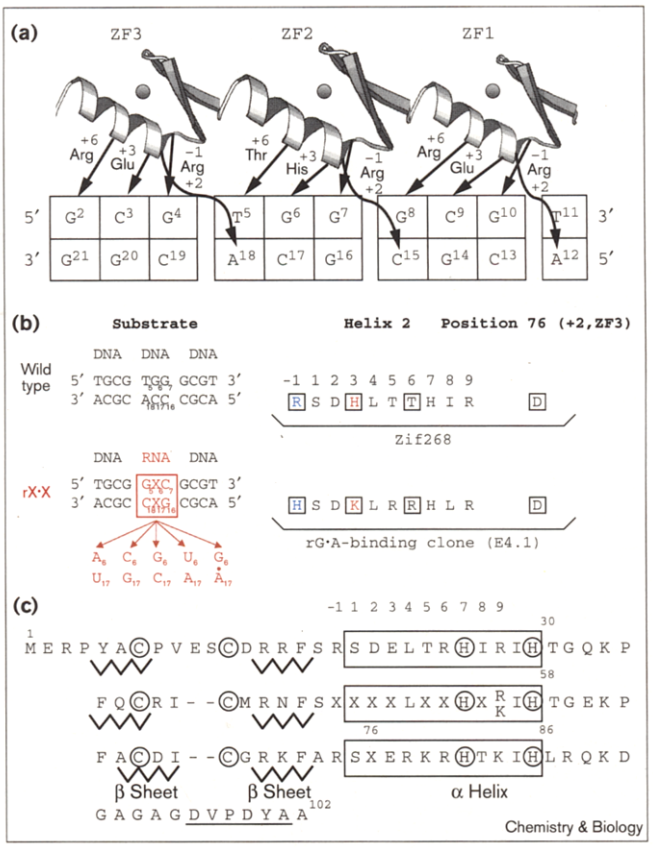
Chemistry & Biology August 1999, **6**:585–597
http://biomednet.com/elecref/1074552100600585

Introduction

RNA–protein interactions play a pivotal role in many cellular processes, including RNA synthesis, processing, transport and translation [1]. The elucidation of RNA–protein structures in complex should allow new insights into these mechanisms. Further, the design of artificial proteins capable of binding to specific RNA structures *in vivo* should allow the modulation of gene expression at the RNA level.

Because of the paucity of data on structures of RNA–protein complexes, our knowledge of RNA–protein interactions is more limited than that of DNA–protein interactions, for which a great body of structural information is available [2]. DNA–protein interactions, however, can also be useful for understanding those between RNA and proteins. The major difference between DNA and RNA in their association with proteins is expected to come from the presence of the 2′ OH groups in RNA, which restrict its conformation as compared to that of DNA. Due to this restriction, the RNA Watson–Crick (WC) double helix, unlike its DNA counterpart, is entirely of the A-type conformation, which is characterized by a deep and poorly accessible major groove [3]. As a result, even if the base pairs (bp) of RNA in the major groove were chemically identical to those of DNA, they might not be as accessible to an interacting protein as their DNA counterpart. Accordingly, it has been shown for a number of DNA–protein complexes that the direct replacement of DNA by RNA in the strands contacted by a protein leads to a dramatic decrease in affinity [4]. To obtain a stable association between a DNA-binding protein and an RNA double helix, therefore, special strategies are needed. Here we describe a strategy by which it was possible to use a DNA template to model protein–RNA interactions. A discrete DNA domain known to bind to a protein was substituted by specific combinations of ribonucleotides. We used molecular modeling to formulate hypotheses about the combinations of ribonucleotides that can optimally interact with

Figure 1



General structure of mouse transcription factor Zif268. **(a)** A summary of the base contacts made by Zif268 with the cognate DNA sequence. **(b)** The design of the nucleic-acid substrates for the ZF library selections (rX•X), compared with the wild-type dsDNA duplex substrate of Zif268. The identity of the helix 2 and residue 76 in the original Zif268 and in the rG•A-selected ZF clone are indicated. The contact positions for the middle triplet of the corresponding substrate are indicated by squares. **(c)** The ZF library design, which is based on the sequence of Zif268. The randomized positions are indicated by X, and are specified by the degenerate codon NNK (N = A,C,G,T; K = G,T). Position +9 of helix 2 is specified as an arginine or lysine residue. The zinc-coordination sites are surrounded by circles. The underlined sequence DVPDYA is a linear epitope for the 17/9 monoclonal antibody [33]. The ZF library was displayed as a fusion with the major coat protein (pVIII) of the filamentous bacteriophage vector, f88-4 [19].

the model protein. We then constructed a phage-display library, based on the structure of the DNA-binding domain, that varied specific amino-acid residues expected to interact with these modeled RNA combinations. Proteins specific for particular RNA combinations were isolated in selection experiments, and finally the complexes were physically characterized using direct binding studies and subjected to modeling.

We chose as our model system for this study the complex of a three zinc finger (ZF) protein, the mouse transcription factor Zif268, with WC double-helical DNA. The crystal structure of this complex is available [5], as well as extensive information on its nucleic-acid–protein interactions [6–13]; this makes it possible to interpret the structural role of particular nucleotides and amino-acid residues involved in the complex. Several aspects of the structure of this complex were important for the design of a successful selection experiment. Firstly, Zif268 has a multidomain structure consisting of three ZF modules, ZF1, ZF2 and ZF3, each of which binds almost independently to a different 3 bp sequence on the DNA ligand. The binding of each ZF module proceeds by the same general mechanism (Figure 1a), in which contacts with nucleotide bases in the major groove of the DNA are made by amino-acid residues at positions –1, +3 and +6. The modular nature of the ZF–DNA interaction allowed us to test potential ZF–RNA interactions, by replacing with RNA the part of the DNA that interacts with the middle ZF2. After such a replacement, the normal interaction between the flanking DNA triplets and their corresponding flanking ZFs (ZF1 and ZF3) should guarantee that the general landscape for the nucleic-acid–protein interaction had not changed, yet binding affinity would be low unless ZF2 interacted with the RNA triplet. Second, the DNA in the Zif268–DNA complex has a specific conformation called $B_{enlarged\ groove}$-DNA (B_{eg}-DNA) that shares properties of both the B and A forms [14]. The replacement of a portion of the DNA by RNA, which is presumably in the A conformation, was not expected to lead to major conformational perturbations. Third, we introduced different irregularities into the WC double helix that are known to open the major groove of the RNA molecule, and make it accessible to protein sidechains [15,16]. Such interactions have been observed for the Rev-binding element of HIV-1; this sequence has two purine–purine base pairs and a bulged pyrimidine in the otherwise WC double helix, which facilitate the binding of the Rev protein to the major groove of the double-stranded (ds) viral RNA [15].

In the work described here, we replaced the middle triplet in the DNA-binding site of Zif268 with a variety of RNA triplets, including those composed entirely of WC base pairs, and those containing non-WC base pairs, and using a phage-displayed ZF library, selected a Zif268 variant that specifically interacts with the nonWC duplex. For the substrates analyzed here, substantial affinity was achieved only after the introduction of a G·A mismatch in the middle position of the RNA triplet. Based on the results of computer modeling and nuclear magnetic resonance (NMR) studies of the G·A-containing duplex, we propose a detailed scheme for ZF–RNA binding, and discuss the structural elements of the RNA groove responsible for the shape-selective interaction with the middle ZF.

Results

Experimental design

The primary structure of the DNA/RNA hybrid used in the selection experiments is shown in Figure 1b. The central

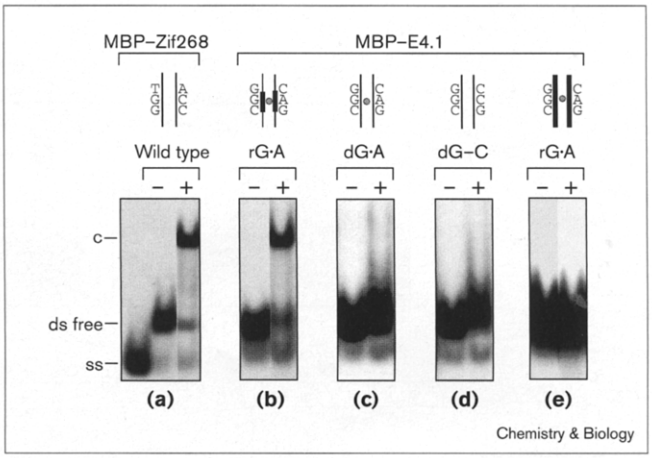
3 bp consist entirely of ribonucleotides, whereas the flanking regions are composed of deoxyribonucleotides. The nucleotide sequences of the first and third triplets remain unchanged. To ensure good base pairing in the RNA triplet, we incorporated G–C pairs in the first (G5–C18) and third (C7–G16) positions. For the middle pair, 6–17, each of the four WC combinations and the G6•A17 pair were used to make a total of five different duplexed molecules. The G•A pair was chosen for this study because, among the noncanonical RNA base pairs, it is known to disturb the structure of the double helix, while adopting a pairing scheme that resembles those of WC pairs.

The design of the phage-displayed ZF library was based on the framework of interactions described for the Zif268–DNA complex (Figure 1a). We fully randomized all positions in the region between –1 and +8 of the middle finger, except for the highly conserved residues Leu+4 and His+7, which are not involved in direct interaction with DNA (Figure 1c). In addition, only limited variation was permitted for position +9, allowing either lysine or arginine, the two residues that most frequently occur at this position. Finally, position 76 (position +2 of ZF3) was fully randomized, as in the Zif268–DNA complex it has cross-strand contact with nucleotide 18 [17,18]. The eight fully randomized positions and one binary position produce a library with a theoretical nucleotide sequence complexity of 2.2×10^{12} that encodes 5.1×10^{10} different ZF sequences. The number of phage clones comprising the library is 2.5×10^{7}. This library was displayed as an amino-terminal fusion with the major coat protein (pVIII) of the filamentous phage vector f88-4 [19–21]. The level of wild-type Zif268 expression by phage, as well as that of several library variants, was determined by enrichment studies that measured the binding of phage-bearing ZFs to a variety of control RNA and DNA/RNA ligands (data not shown). ZF expression was also deduced from ELISA, in which antibody binding was measured to a linear epitope tag that was carboxy-terminal to a given ZF (Figure 1c).

Phage selection

We undertook the selection of the ZF library against duplexes containing one of the four possible WC nucleotide combinations in the middle position 6•17 (duplexes rG–C, rC–G, rA–U and rU–A, where r denotes an RNA central triplet containing the indicated base pair). After eight rounds of panning, none of the duplexes captured ZF phage with a significant yield. When selection was performed using a RNA duplex whose middle pair was G6•A17 (rG•A), however, ZF phages were selected with a yield of about 1% of the phage screened in the fourth round of panning. To verify the specificity of the association of the G•A pair with the phage that it selected, we tested the binding of phage from the rG•A-enriched pool to another duplex containing a U6•U17 pair (rU•U). In this experiment, the yield did not exceed the background level of 10^{-4}%, which clearly indicated that the detected phage–nucleic acid interaction was specific and dependent on the G6•A17 pair. In addition, no interaction was observed between ZF variants from this library and WC RNA bp substitutions in this particular duplex.

Figure 2



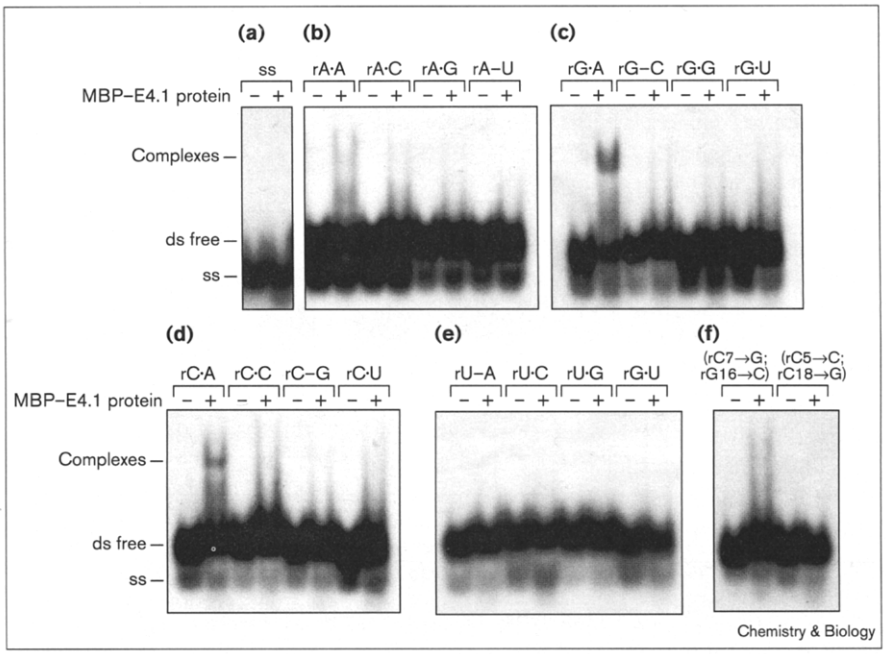
Gel mobility shift experiments assess the specificity of the selected ZF protein as a carboxy-terminal fusion with MBP with different DNA or RNA-containing duplexes. **(a)** The radiolabeled wild-type dsDNA sequence (Figure 1b) was incubated in the absence (–) or presence (+) of saturating amounts (100 nM) of Zif268 expressed as a fusion with MBP (MBP–Zif268). **(b–e)** Different radiolabeled duplexes were incubated in the absence (–) or presence (+) of saturating amounts (600 nM) of the selected ZF protein (MBP–E4.1). DNA is indicated by a thin line, RNA by a thick line. Only the sequence of the middle triplet of the duplex is indicated. c, nucleic acid–ZF complex; ss, single-stranded oligonucleotide; ds free, uncomplexed ds oligonucleotide.

Ten clones picked from the pool of phage selected after the fourth round of panning with the rG•A duplex had identical sequences (clone E4.1, Figure 1b). Like wild-type Zif268, this clone bore aspartate, serine, aspartate and arginine residues at positions 76, +1, +2 and +9, respectively, whereas the positions that had changed were: Arg–1→His, His+3→Lys, Thr+5→Arg, Thr+6→Arg and Ile+8→Leu.

Specificity of the rG•A-binding peptide (E4.1)

The dependence of the E4.1–rG•A association on the RNA region of the duplex was analyzed by gel mobility shift experiments (Figure 2). The rG•A-binding peptide was transferred from pVIII to the carboxyl terminus of the maltose-binding protein (MBP) of *Escherichia coli*, and expressed; purified MBP–E4.1 fusion was used in the gel shift assay. Figure 2b shows that the selected ZF protein (MBP–E4.1) was able to bind the rG•A duplex. The binding was markedly reduced when full-length DNA substrates were used (Figure 2c,d), however, and was completely eliminated in the case of the full-length RNA substrate (Figure 2e). The fact that neither dsDNA or dsRNA was recognized by the protein strongly suggests

Figure 3

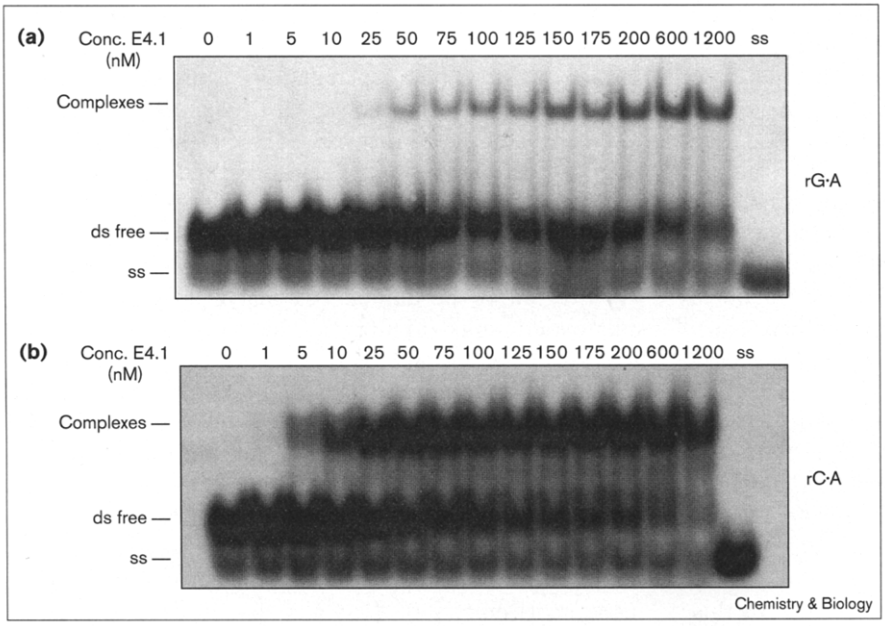


Gel mobility shift experiments probe the specificity of the selected ZF protein (MBP–E4.1) for each of the three positions in the RNA triplet. **(a)** The ^{32}P-labeled ss oligonucleotides were incubated in the absence (–) or presence (+) of saturating amounts of MBP–E4.1 (600nM). **(b–e)** Gel mobility shift experiments assess the specificity for the middle pair of the RNA triplet, indicated on the top of the panels. Sixteen duplexes containing each of the possible changes in the middle pair of the RNA triplet were incubated in the absence (–) or presence (+) of saturating amounts of MBP–E4.1 fusion (600nM). **(f)** A duplex changing the 3′ position of the triplet (rC7→G;rG16→C), and a duplex changing the 5′ position of the triplet (rG5→C; rC18→G) were incubated in the absence (–) or presence (+) of 600nM of MBP–E4.1.

that the rG·A duplex possesses a distinctive conformation that is different from both WC duplexes.

The degree to which the association of E4.1 with rG·A is dependent on the G·A base pair was tested by systematically replacing this base pair with the 15 other dinucleotide combinations possible. The results presented in Figure 3 show that in addition to the G·A pair, only the C·A pair allowed significant binding; in fact, the affinity of rC·A for E4.1 was even higher than that of rG·A (Figure 4 and Table 1). None of the other dinucleotide combinations, including the four bearing WC base pairs, showed detectable association with E4.1. Next, we tested the dependence of the association on the nature of the G–C

Figure 4



Gel mobility shift analysis determines the relative dissociation constant (K_d) of MBP–E4.1 for its substrates. **(a)** Gel shifts at various MBP–E4.1 concentrations using a constant concentration of ^{32}P-labeled rG·A duplex. **(b)** Gel shifts at various MBP–E4.1 concentrations using a constant concentration of ^{32}P-labeled rC·A duplex.

Table 1

Relative dissociation constants (K_d) of the ZF fusions.

	ZF fusion	
Substrate	MBP–E4.1	MBP–E4.1 (His–1→Arg)
rG•A	79.76 + 1.04 nM	65.92 + 0.088 nM
rC•A	8.84 + 1.16 nM	32.82 + 0.035 nM
(C7→G; G16→C)	nd	0.53 + 0.10 nM

Affinities of the ZF proteins for the target duplexes were determined using a gel shift assay. K_d values are averaged from 3–6 experiments. Standard errors are indicated. nd, not determined (because the complex was too weak).

base pairs flanking the central G•A pair. We made the most conservative changes possible by simply flipping these G–C base pairs one at a time. As shown in Figure 3f, mutation (G5→C; C18→G) eliminated binding entirely, whereas mutation (C7→G; G16→C) decreased binding significantly.

Modeling of the rG•A duplex

The specificity of the interaction between E4.1 and rG•A makes it reasonable to suggest that the structure of this complex resembles that of Zif268–DNA in its essential aspects. This permits the crystal conformation of the latter to be used as a basis for modeling of the E4.1–rG•A complex.

The two complementary trideoxyribonucleotide regions T5-G6-G7 and C16-C17-A18 that form the three WC base pairs interacting with ZF2 of Zif268 were replaced with triribonucleotides G5-G6-C7 and G16-A17-C18, respectively. The flanking DNA triplets preserved their positions as in the original DNA molecule. Because the conformation of the original DNA within the complex with Zif268 has some similarities to A form [14], the replacement of the middle DNA portion by RNA was not expected to dramatically distort the overall structure of the molecule. In modeling the conformation of the middle G•A pair, four different nucleotide arrangements were considered: head-to-head, sheared, G-*syn*•A-*anti* and G-*anti*•A-*syn*. All these possibilities are known to exist in nucleic-acid double helices [22]. We found a preference for a head-to-head arrangement in rG•A, because it involves a stacking pattern that is characteristic of the A form, and because all the other nucleotide arrangements had essential hydrophobic surfaces exposed to the solvent, rather than interacting with other bases. The 3′-*endo* and 2′-*endo* conformations were assigned for the riboses and deoxyriboses, respectively. On the basis of these deductions, a model of rG•A was built and refined by energy minimization (Table 2 and Figures 5,6). As mentioned, pair C6•A17 can substitute for pair G6•A17 in the modeled rG•A interaction with E4.1. For this C•A pair, a possible nucleotide arrangement that is isosteric to that of the head-to-head G•A pair is shown in Figure 6d.

The replacement of six deoxyriboses by riboses and the introduction of a head-to-head purine•purine G•A base pair into the double helix resulted in rearrangement of the backbone in all ribonucleotide, as well as in the flanking deoxyribonucleotide base pairs. As for the positions of the bases, the main differences between the modeled rG•A conformation and the crystal structure of the original DNA occur around nucleotides G5 and C16, where a deoxyribonucleotide is followed by a ribonucleotide. The nucleotides experiencing the strongest displacement in comparison to the corresponding nucleotide positions in the original DNA structure are G5 and G16, which shift 2.5–3.0 Å in the direction of the minor groove, and nucleotides G6 and A17, which shift 1.5–2.0 Å in the same direction. The bases of all six ribonucleotides are tilted by about 10° with respect to their corresponding bases in the original DNA molecule. Despite these differences, the overall arrangement of base pairs in rG•A resembles that of the original DNA, making it possible that both nucleic acid molecules interact with their

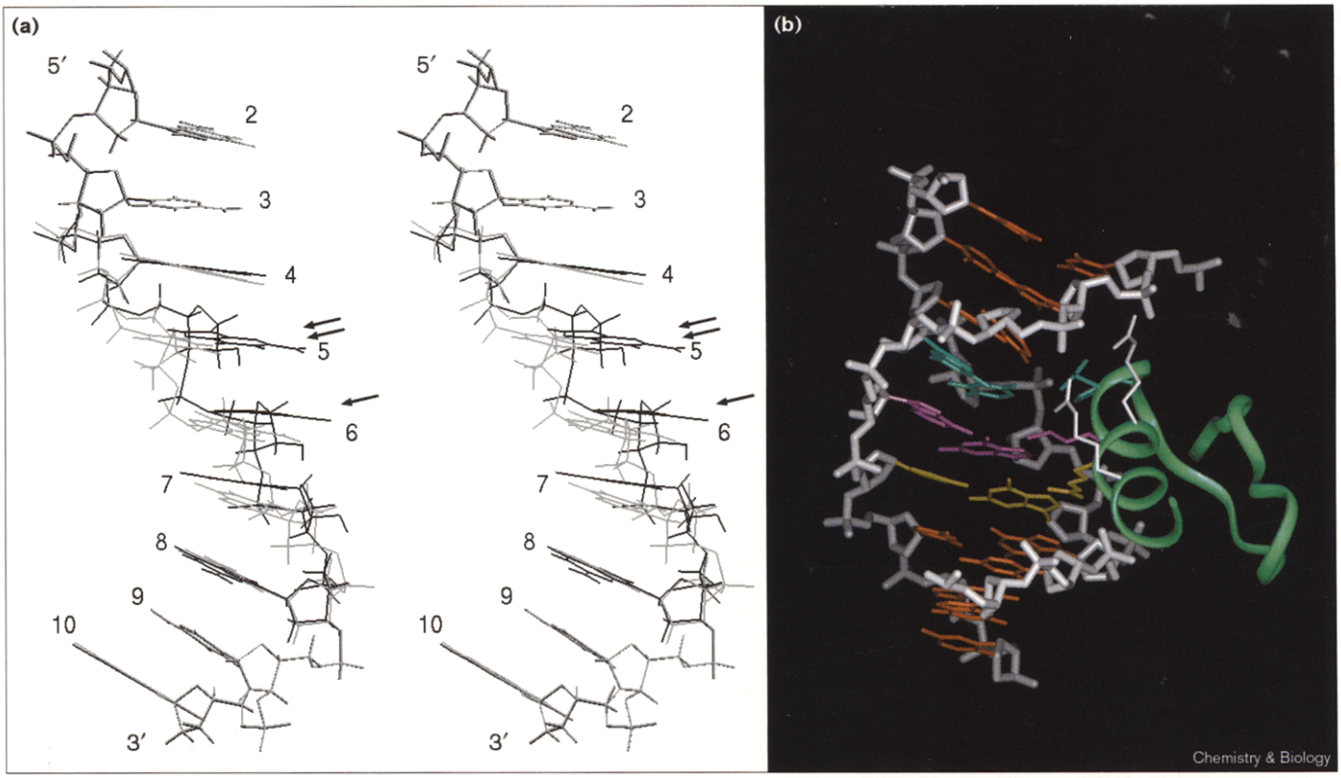
Table 2

Conformations of the phosphates and sugars (riboses or deoxyriboses) of the central triplet in the original DNA and in the modeled rG•A structure.

Original DNA				RNA in G•A			
N	Ph*	Sugar	Angle χ	N	Ph	Sugar	Angle χ
T5	g^-g^-	C1′-*exo*	−109°	G5	g^-g^-	C3′-*endo*	−147°
G6	tg^-	C2′-*endo*	−125°	6	g^-g^-	C3′-*endo*	−160°
G7	g^-g^-	C2′-*endo*	−118°	C7	g^-g^-	C3′-*endo*	−171°
C16	g^-g^-	C2′-*endo*	−120°	G16	g^+g^-	O4′-*exo*	−140°
C17	g^-g^-	O4′-*endo*	−133°	A17	g^-t	C3′-*endo*	−167°
A18	tg^+	O4′-*endo*	−127°	C18	g^-t	C2′-*exo*	−177°

N, nucleotide; Ph, phosphate conformation. *The first letter designates the conformation of the O3′-P bond and the second letter does so for the P-O5′ bond as *trans* (t) or *gauche* (g), and for the gauche conformation as − (−120–0°) or + (0–120°) [3].

Figure 5



Structure of the rG•A model and of its complex with ZF2 of E4.1. **(a)** Superimposition of regions 2–10 in the original DNA (gray) and in rG•A (black) with nucleotide sequences 5′-dGdCdGdTdGdGdG-dCdG-3′ and 5′-dGdCdGrGrGrCdGdCdG-3′, respectively. Nucleotides 5–7 of rG•A contain riboses, all other nucleotides contain deoxyriboses. The flanking nucleotides 2–4 and 8–10 in the two molecules overlap quite well, whereas for nucleotides 5–7 the differences are notable. The main difference occurs around nucleotide 5, where in rG•A a deoxyribonucleotide is followed by a ribonucleotide. G5 in rG•A is shifted about 3 Å toward the minor groove with respect to the position of the corresponding nucleotide T5 in the original DNA (for nucleotides 5–7 the major groove is on the left and the minor groove is on the right). The next nucleotide G6 is also shifted for about 1.5–2.0 Å in the same direction. The positions of G5 and G6 in rG•A are marked by two and one arrow, respectively. Similar differences exist in the opposite chain of the duplex (data not shown). **(b)** The interaction of the middle RNA triplet of rG•A with ZF2 of E4.1. The polypeptide chain of ZF2 folded in two β strands and an α helix is shown as a green ribbon. Only the positions of amino-acid residues His−1 (blue-green) and Asp+2 (blue-green), Lys+3 (purple) and Arg+6 (mustard), which interact directly with the corresponding RNA bases, C7–G16 (blue-green), G6•A17 (purple) and G5–C18 (mustard) are indicated. On the figure, Asp+2 partly eclipses His−1. The sidechains of Arg+5 and Arg+9 (white) are shown interacting with the phosphodiester backbone (white). The flanking DNA triplets are red-orange.

corresponding ZF proteins via nucleotide–amino-acid residue contacts at similar sites. The structural differences identified between rG•A and the original DNA could also affect several aspects of the association.

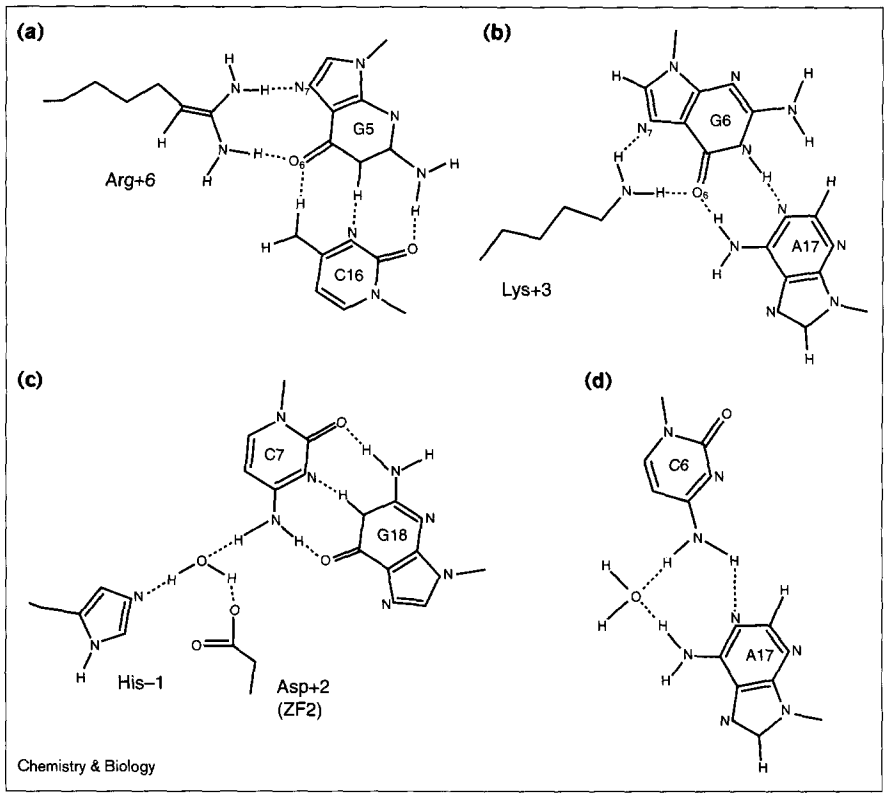
NMR analysis of the rG•A duplex

Out of the four modes of G•A base pairing reported in NMR and X-ray studies of nucleic-acid duplexes, only the sheared base pair gives rise to a characteristic imino proton chemical shift, typically between 10 and 11 ppm [23]. For stable forms of the sheared G•A base pair, the sheared guanine imino resonance is of comparable line width and intensity as other WC base-paired imino protons, observed between 12 and 14 ppm. For the duplex studied here, the absence of a sharp resonance in the 10 to 11 ppm region (Figure 7a) was interpreted as an absence of the sheared type G•A base pair. The chemical shift of the guanine imino proton involved in the head-to-head type of base pair would be expected to coincide with the other WC-paired guanosine imino protons. It was not possible, therefore to make a definitive assignment because of spectral overlap in the 12–14 ppm region.

Proton assignments were obtained for the aromatic and H1′ resonances using a correlated spectroscopy (COSY) and 400 ms nuclear Overhauser effect spectroscopy (NOESY) spectra. Scalar coupling cross peaks between pyrimidine H5 and H6 protons were identified in the COSY spectrum. Sequential and interstrand aromatic to H1′ assignments could be traced on the NOESY spectrum as shown in Figure 7b. The sequential H6/8i-H1′i -H6/8(i+1) pathway

Figure 6

Nucleotide arrangements of the base pairs in the RNA triplet and their interaction with amino-acid residues at positions −1, +2, +3 and +6 of ZF2. **(a)** The Arg+6–G5 interaction; **(b)** The Lys+3–G6 interaction; **(c)** The His−1–Asp+2–C7 interaction; **(d)** Nucleotide arrangement in pair C•A, the best arrangement that is isosteric to the head-to-head G•A pair.



was interrupted between C3 and G6, suggesting localized dynamics for these nucleotides. No unusually intense intranucleotide aromatic to H1′ correlations were detected, ruling out the possibility of *syn* glycosidic conformation for any nucleotides. All other NOESY correlations, notably the sequential H1′–pyrimidine H5 NOEs, were consistent with a B-helical conformation. The sequential A17H2–C18H1′ and cross-strand A17H2–C7H1′ NOEs, and the sequential and intranucleotide NOEs involving the A17H8 proton, confirmed that this nucleotide is stacked in a helical context.

As expected, the six ribose H1′ resonances showed no NOEs with the deoxyribose H2′/H2″ protons (data not shown), thus confirming these crucial assignments. The magnitude of the ribose H1′-H2′ couplings can be used to distinguish between C3′-*endo* and C2′-*endo* ribose puckers. Small couplings, of the order of 3 Hz or less, are indicative of C3′-*endo* sugar conformation. Such small couplings also lead to cancellation of the COSY signal between these two protons. Thus the complete absence of H1′-H2′ COSY signals for the six ribonucleotides (data not shown) suggested that these nucleotides all adopt a C3′-*endo* sugar pucker. In conclusion, our NMR study of rG•A confirmed the main characteristics of the model in both the type of puckering of the riboses and deoxyriboses and the head-to-head arrangement of the G•A pair.

Proposed amino-acid–nucleotide contacts in the ZF–RNA complex

The model of the E4.1–rG•A complex shown in Figure 5 helps rationalize the identity of the residues in E4.1 that were most likely selected because of their interaction with rG•A.

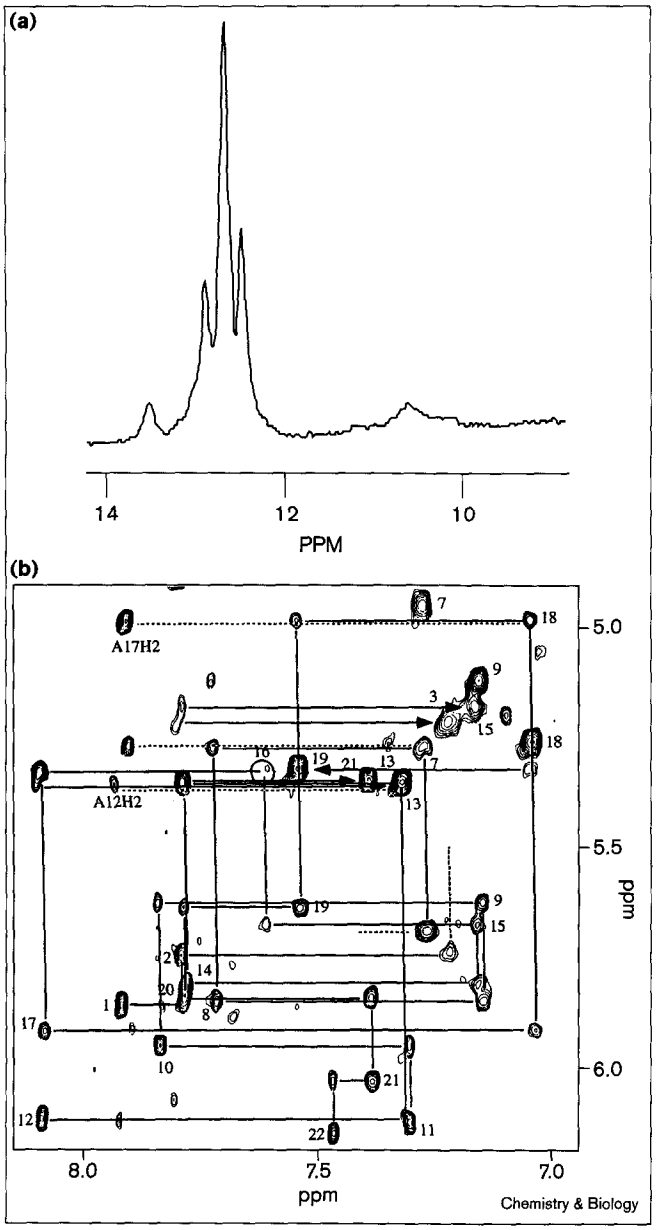
Residues +6 of ZF2 and 76 (+2 of ZF3)

Similar to ZF1 and ZF3 of Zif268, position +6 in ZF2 of E4.1 bears an arginine residue, which could make a bidentate contact with G5 (Figure 6a). The long 'leg' of Arg+6 allows the formation of this interaction, even after the shift of base pair G5–C18 toward the minor groove, as mentioned above. Residue 76 has not changed after selection, and its conservation suggests that it could interact with N4 of C18, exactly as it interacted with N6 of A18 in the Zif268–DNA complex. This Asp76 could also interact with Arg+6 in the same way as the corresponding Asp+2 interacts with Arg+6 in the other ZFs of Zif268 [5].

Residue +3 of ZF2

Position +3 in ZF2 of E4.1 contains a lysine residue, which can form hydrogen bonds with the O6 and N7 groups of G6 (Figure 6b). In some DNA-binding ZFs, including ZF2 of Zif268, position +3 is occupied by a histidine residue, which forms a hydrogen bond with the N7 of the corresponding guanosine. Due to the above-mentioned shift of

Figure 7



NMR spectra of the chimeric duplex. **(a)** The imino proton region at 4°C. The absence of a characteristic sharp peak between 10 and 11 ppm provides evidence against the formation of a sheared G•A base pair (3). **(b)** Aromatic to H1/H5 = region of a NOESY spectrum in D_2O, 400ms mixing time, 25°C. The H5-H6 cytidine cross peaks are numbered. Solid lines trace the $H1 =_{(i-1)} \_ H6/H8_i \_ H1 =_i Y$ pathway, with nucleotide numbering on the H6/8 B H1= cross peaks. There is a discontinuity in this sequential pathway from nucleotides C3 to G6, shown by dotted lines. Sequential and cross strand AH2 B H1 = cross peaks are shown with dotted lines and labeled at the AH2 chemical shift. Arrows indicate sequential $H1 =_i$ to cytidine H5 correlations.

pair G6•A17 in rG•A toward the minor groove, His+3 is too far away from G6 to maintain this interaction. The absence of an additional stabilizing factor, the stacking of residue +3 with the neighbouring nucleotide T5, which existed in the original DNA but not in rG•A, would also discourage the selection of a histidine residue at this position [5]. Correspondingly, mutational analysis of MBP–E4.1 with the replacement Lys+3→His (Figure 8b) abolished the association with rG•A and with rC•A. The replacement Lys+3→Arg in E4.1 (Figure 8a) restored the association, indicating that long sidechain residues are required at this position; this agrees with the predicted shift of the G•A pair toward the minor groove. Lys+3 might also be involved in other contacts, in particular, with O6 of G16.

Residues −1, +1, +2 of ZF2

A serine residue occupies position +1 in ZF2 of E4.1; this same residue is present at position +1 in all three ZFs of Zif268. Although the reasons for such conservation are not clear, it shows that the DNA→RNA replacement of the middle triplet has not changed the requirement for this residue.

Position +2, which was occupied by an aspartate residue, has also not changed after selection. As in the Zif268–DNA complex, this residue can form a hydrogen bond with the NH group of C15. A weaker hydrogen bond is also possible with the NH group of C7.

Histidine in position −1 is too far from the G•A pair to form a direct interaction. Water bridges between this histidine residue and either C7 or Asp+2 are, however, possible (Figure 6c). A replacement of this histidine residue by a negatively charged glutamate (Figure 8d) results in a complete abolition of the association, probably because of the electrostatic repulsion from the Asp+2 and from the phosphates of G6 and C7. These results then, agree with the fact that mutation His–1→Arg (Figure 8c), unlike His–1→Glu, does not affect the association seriously. As in previously described cases [13], this arginine residue can form hydrogen bonds with the phosphates of either G6 or C7. The subsequent replacement C7–G16→G–C in rG•A (Figure 8c) most probably allows Arg–1 to form an optimal bidentate contact with G7. Consistent with this, we found that this substitution increased affinity by two orders of magnitude (Table 1).

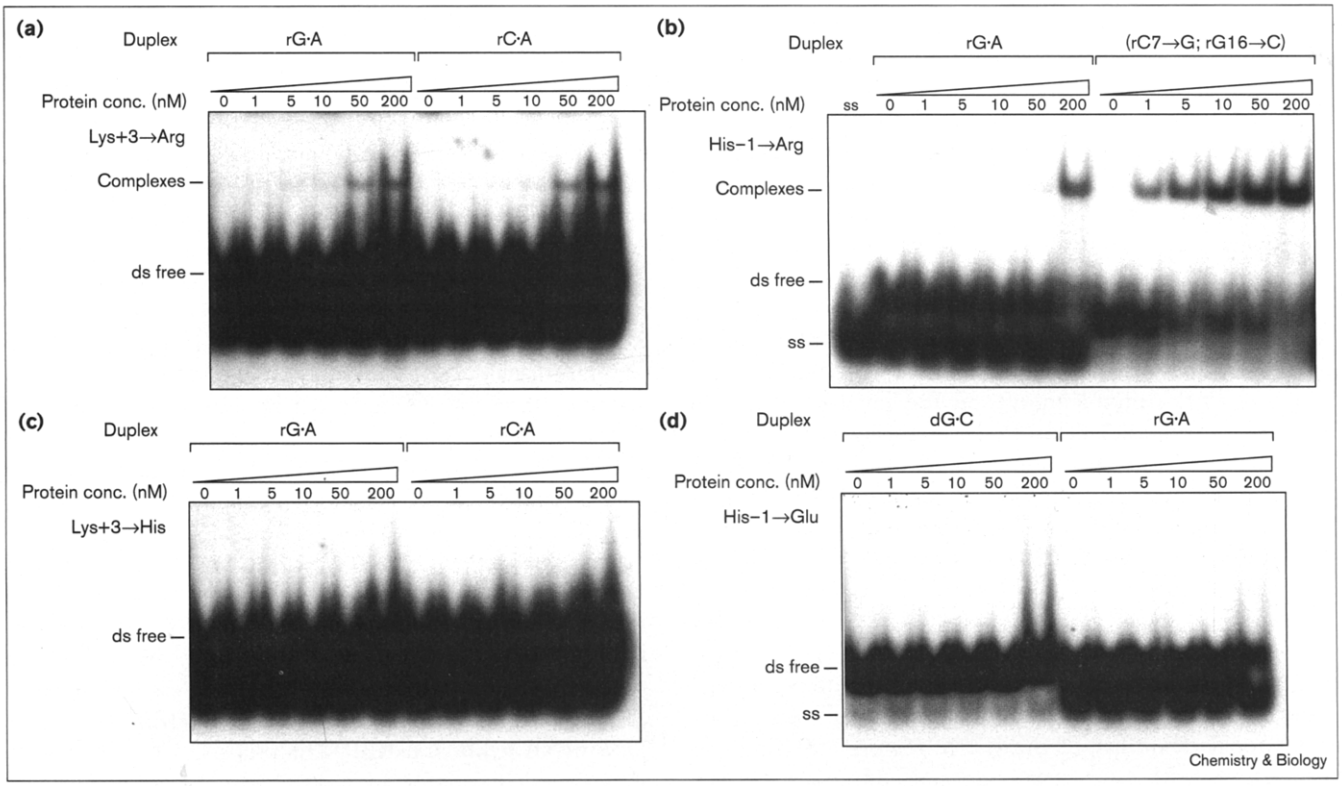
Backbone contacts

Potential contacts with the phosphates include His53, as shown in the Zif268–DNA co-crystal structure [5]. In addition, Arg51 located at position +5 and Arg59, located at position +9, could also be involved in phosphate interactions (Figure 5b). These interactions should not contribute to the specificity but could facilitate the fit of the helix into the major groove, thus stabilizing the interaction with the RNA.

Discussion

We present here a new strategy for the selection of a RNA-binding ZF from a ZF that normally forms a complex with a WC dsDNA. The association between the ZF and RNA

Figure 8



Gel mobility shift experiments assess the interactions of several mutants of the MBP–E4.1 protein with different substrates. Substrates are named as in Figures 2,3.

proved to be specific, as shown by the fact that replacements in the nucleotide sequence of the latter led either to a complete abolition or to a substantial reduction of the affinity. The experimental data suggest that the interaction of the selected ZF with RNA was similar to that of ZF2 in Zif268 with DNA. The identity of the amino-acid residues interacting with the RNA triplet could be easily rationalized within the three-dimensional model of the E4.1–rG·A complex, which was built based on the structure of the Zif268–DNA complex [5]. The similarity of the interactions in E4.1–rG·A and in Zif268–DNA was ensured by replacing only the middle triplet of the original DNA molecule with RNA, whereas the flanking DNA sequences were able to preserve the structural landscape for the ZF2–RNA interaction. The special B_{eg} conformation of the DNA in the Zif268–DNA complex, which shares features of both the B and A forms [14], is another important characteristic that was expected to minimize the perturbing effect of the DNA→RNA replacement.

These factors alone, however, were not sufficient to guarantee a strong RNA–protein interaction. Indeed, all our attempts to select a ZF with WC dsRNA double helices failed. Because our library comprises only a small portion of the ZF sequences possible, we cannot exclude the possibility that such ZFs are missing from the library. However, to be so rare, such ZFs would either require a larger number of critical binding residues than those of the ZF that was selected by the rG·A pair, or be more poorly expressed.

The G·A pair modifies the local geometry of the double helix, and thereby counteracts the effect of the S-puckered riboses on the overall WC RNA geometry. It is well known that purine·purine base pairs are able to widen the major groove of the RNA double helix, and make it more accessible to protein α helices [15]. Another possible role of this base pair is to diminish the effect of the RNA region on the conformation of the whole duplex. Indeed, it is known that the introduction of ribonucleotides into DNA can induce major conformational changes. It has been shown that many such hybrids adopt conformations that are closer to the A than the B form [24]. For example, the presence of even a single ribose in a B-DNA decamer switches it to an A form [25]. From this point of view, the introduction of three RNA base pairs would probably cause the whole molecule to take on the A conformation. The role of the purine·purine base pair G6·A17 could also, therefore, be to change the conformation of the RNA region, such that the propagation of the A form into the

flanking DNA region is prevented. The head-to-head arrangement of the purines in the G•A base pair was suggested based on the requirements for the optimal nucleotide stacking in our model, and confirmed by NMR analysis of rG•A. Although the effect of this pair on the geometry of the RNA double helix is quite specific, another mispair, C•A, was, surprisingly, able to substitute G•A in the E4.1–rG•A complex. Our modeling studies have, therefore, provided clues on the nucleotide arrangement of this pair that makes it isosteric with G•A.

Although the modeled structure of rG•A is fairly close to that of the original DNA, there are some differences, in particular the shift of four nucleotides G5, G6, G16 and A17 toward the minor groove. At the level of E4.1, the shift of the ribonucleotide bases in the direction of the minor groove could be compensated by the substitution of a short His+3 sidechain by a longer one (lysine), which therefore represents a protein determinant that discriminates between DNA and RNA. Some other E4.1 amino-acid residues may also assist in this discrimination. We noticed, that among five amino-acid differences between E4.1 and Zif268, three cases, Thr+5→Arg, Thr+6→Arg and Ile+8→Leu, were C_β-branched substitutions. These replacements could have a cumulative effect on the conformational flexibility of the whole region, affecting the interaction with different targeted duplexes. Arg+5 should also stabilize the association by interacting with the backbone.

The interactions between ZF proteins and dsDNA have attracted a lot of attention in the past decade. Although certain aspects of the interaction are very specific, it is difficult to establish a general 'code' as a dual correspondence between amino-acid residues and nucleotides. As Elrold-Erickson *et al.* [13] showed with different complexes of Zif268 variants with DNA, both the conformation of the ZF and its orientation with respect to the DNA triplet can vary, which makes this amino-acid–nucleotide correspondence even less direct. The introduction of RNA substitutions and nonWC base pair into the WC DNA double helix adds more variables, further complicating the analysis. Despite these difficulties, it has been possible to interpret our experimental results in terms of a simple amino-acid–nucleotide model.

Although we partially or completely randomized eight amino-acid positions in ZF2 and one position in ZF3, the library with which we worked comprised only a fraction of the total complexity possible. However, if we had explored a greater part of the library's total theoretical sequence space, based on our results, we would not expect a great variation of the amino-acid identity in either position +3 or +6. Position –1, on the contrary, could experience some variability, as was demonstrated in the experiments on the post-selection amino-acid replacements in E4.1 (Figure 8).

Recent progress in studying the mechanism of ZF–nucleic-acid interactions indicates that the ZF motif, in particular, that of the Cys2–His2 type, is an extremely versatile scaffold able to adapt easily to binding different surfaces [26]. This versatility of the ZF motif should be attributed to a very simple structure consisting of an α helix packed with two β strands via the coordination of a zinc atom. This arrangement restricts the conformational flexibility of the α helix and fixes the positions of the amino-acid residues located on the outer surface of the helix, enabling them to interact with different molecules in specific ways. The very small size of the ZF, about 30 amino acids, allows the simple, iterative duplication of the domain, and enables several consecutive ZFs to be positioned in close proximity without steric interference; this produces strings that can follow the helical shape of the major groove. Sequential ZF motifs, when separated by short linkers, bind the nucleic acid in a relatively independent manner [5]. This makes it possible for each motif to evolve independently of the others, to adapt to new surfaces, and ultimately to expand its repertoire of biological functions. It is not surprising, therefore, that the Cys2-His2 family of ZFs contains more than 1300 proteins known to specifically bind different types of nucleic acids: RNA (as in the case of TFIIIA [27,28], p43 [29], WT1 [30]) and DNA (Zif268 [5]), or DNA/RNA hybrids (dsRBP–ZFa [31] and YY1 [32]). Some ZFs are evolutionarily specialized to bind either DNA or RNA, and some proteins such as TFIIIA and WT1, contain sets of both DNA and RNA binding ZFs, which enable the same protein to participate in different regulatory pathways, both at the transcriptional and post-transcriptional levels.

The same intrinsic versatility of the ZF domain makes it an excellent scaffold for the *in vitro* evolution of new specificities and/or new protein functions. Recently, Friesen and Darby [33] produced an α-helical library using two RNA-binding fingers from TFIIIA as a scaffold. The authors successfully selected ZFs able to bind with high affinity to either the 5S rRNA or the Rev-binding element (RBE) stem II of HIV-1. This example illustrates that α helices in the ZF context can be used as recognition elements for RNA. In this case, however, it was not known whether the modes of binding of these *in vitro* isolated ZFs are similar to those of the endogenous factors TFIIIA and Rev. In our work, a rational, directed, *in vitro* evolution strategy was developed for converting a ZF from a DNA-binding to a RNA-binding motif. Moreover, our experimental design has allowed us to follow the ZF interaction with the complementary surface of the RNA, and to show that it was similar to the interaction between Zif268 and DNA. Our results demonstrate that the ZF architecture can provide a means for the facile evolution of DNA- or RNA-binding proteins from their ligand counterpart.

In this study we replaced with RNA only one of the three DNA triplets. We recently found that other nonWC RNA

base pairs can also have similar structural properties to that of the G·A pair, and we were able to obtain Zif268 variants that specifically interact with them (P.B., B.P., R.C. and J.K.S., unpublished observations). In the future, based on the current results, ZF selections could be performed against duplexes containing multiple RNA triplets. We believe that manipulating the substrates with different nonWC base pairs will allow us to build a stable RNA duplex that sufficiently resembles the B_{eg}-DNA conformation to allow the successful selection of multi-modular ZF ligands.

Significance

Zinc fingers (ZFs) are versatile motifs that have naturally evolved to bind a wide variety of substrates, including RNA, DNA and RNA/DNA hybrids. Some ZF proteins, such as TFIIIA, contain distinct sets of RNA- and DNA-binding fingers that allow regulation of gene expression both at the transcriptional and post-transcriptional level. Here we presented our data on the directed evolution of a DNA-binding ZF toward RNA-binding. The isolated RNA-binding ZFs were characterized by two levels of specificity. First, they bound exclusively to RNA helices containing noncanonical base pairs. Indeed, incorporation of a G•A base pair within the RNA helix (rG•A duplex) allowed selection of a RNA-binding ZF. In addition, the selected ZF was not able to bind Watson–Crick (WC) RNA duplexes or full-length DNA duplexes. Second, their specificity was both sequence dependent and conformation dependent. The selected ZF bound the rG•A duplex, in which the G•A pair adopted a head-to-head conformation but, significantly, it also bound the C•A-containing RNA helix, which was modeled to be isosteric to the rG•A duplex. In addition, the conformation of the rG•A duplex was shown to be similar to that of the original DNA of the Zif268–DNA complex. This allowed us to rationalize the contacts between the RNA-binding ZF to the RNA helix using the Zif268–DNA complex as a framework. The structural features induced by the G•A pair were successfully used for ZF recognition. We found that a long sidechain at position +3, lysine or arginine, compensated a shift of the G•A pair toward the minor groove.

Interestingly, an analogous sequence-conformation-dependent recognition of a RNA-binding protein such as the case described here has already been observed in nature. The RBE (Rev binding element) of HIV contains two purine•purine pairs, G•A and G•G. The G•G pair has been found to co-vary with an A•A pair, both pairs playing an analogous structural role in widening the RNA major groove, which is required for Rev to bind. In addition, G•A pairs and other noncanonical base pairs are frequently found in natural RNA helices. The strategy presented, therefore, can be used for rational design of ZFs that bind particular RNA structural elements. The designed proteins could then be used to exert specific post-transcriptional regulation. We are currently exploring the possibility of expanding the range of ZF–RNA structural recognition in order to rationally design multi-modular, RNA-binding ZF proteins.

Materials and methods

Phage selection

The gene coding for the mouse transcription factor Zif268 was obtained from the plasmid pMexneol (a generous gift from Dr. D. Skup, Institut du Cancer de Montreal, Montreal, Canada) and cloned into the *Hin*dIII–*Pst*I sites of the filamentous phage vector f88-4 [19], a pVIII-display vector. A linker coding for $(Gly–Ala)_4$ was incorporated at the carboxyl terminus of Zif268, just upstream of the pVIII-coding region. The resulting plasmid, f88-4-ZF, was used to transform *K91* cells, and colonies were grown overnight on LB medium supplemented with tetracycline (20 μg/ml), IPTG (1 mM) and ZnAc (50 μM). Phage expressing recombinant ZF (fd88-4-ZF) were isolated as described in [8]. Phage production was typically $1–5 \times 10^{11}$ phage particles per ml. Control panning experiments with the fd88-4-ZF phage using a ds wild-type DNA operator duplex as ligand were performed as described [8], but 0.1 mg/ml acetylated bovine serum albumin (BSA), prepared as described in [11], was used to block the streptavidin-coated beads (Dynal; Oslo, Norway), and the binding reaction solution was supplemented with 5 mM DTT and 20 units of RNasin (Promega; Madison, WI).

The ZF library was built from two sequential steps of Kunkel mutagenesis using strain *CJ236* for transformation [34]. In the first cycle of mutagenesis, we varied ZF position 76, whereas helix 2 was varied in the second cycle (Figure 1). The original template for mutagenesis was modified to avoid over-representation of the wild-type Zif268 sequence in the library. The modifications consisted of the inclusion of two restriction sites, *Kpn*I and *Sma*I, in helix 2, and a change from aspartate to a leucine residue at position 76. Also, a sequence coding for the epitope tag DVPDYA (using single-letter amino-acid code) was added at the carboxyl terminus of the ZF. After mutagenesis, the fill-in product (10 μg) was digested with *Kpn*I and *Sma*I to eliminate much of the remaining nonmutagenized template. The mutated plasmids were used to transform electrocompetent cells *CM106* cells in ten independent electroporations. The final size of the library after digestion and transformation was 2.5×10^7 TU. Stable expression of independent clones of the library was tested by ELISA using the 17/9 monoclonal antibody that recognizes the linear epitope DVPDYA [35].

The phage selections against the rX•X duplexes (Figure 1b) were performed as described for full-length DNA duplexes [8], but the binding reactions were performed in solution by incubating each of the rX•X-biotinylated duplexes and phage from the library (typically 10^{11} phage) for 1 h at 15°C. Streptavidin-coated beads (150 μg) blocked with acetylated BSA were added to the binding reaction for 10 min at room temperature, and the complexes were washed and eluted as described [8]. The eluted phage were amplified in *Escherichia coli* strain K91. The selection procedure was repeated six times. In the first three rounds of selection, the concentration of the rG:A duplex was 80 nM, but it was decreased thereafter to 8 nM. Single-stranded DNA from ten different phage clones from round four was prepared and sequenced with Sequenase 2.0 according to the instructions provided with the T7 Sequencing Kit (cat. no. 27-1682-01; Pharmacia Biotech, Quebec, Canada). The oligonucleotide sequences of the two strands of rX·X used for panning and gel mobility shift experiments were: 5′-d(TATATAGCG)r(GXC)d(GCGTATATA)-3′ and 5′-biotin-d(TATATACGC)r(GXC)d(CGCTATATAGCG)-3′ where X = A, C, G, U, r = ribonucleotides and d = deoxyribonucleotides. Hybrid oligonucleotides were purchased from Dalton Chemicals (North York, ON, Canada).

Molecular modelling

Three-dimensional models were built using the interactive molecular modelling and energy-minimization facilities in the InsightII/Discover

package (version 97.0; Biosym, San Diego, CA). Calculations and visualizations were performed on a Silicon Graphics Indigo computer. The crystal structure of the Zif268–DNA complex [5] was used as the starting structure. Preliminary models were submitted to partial energy minimization in the AMBER force field. Minimization was performed until a valid stereochemistry was obtained.

NMR spectroscopy

The one-dimensional imino proton spectrum was recorded in 90% H_2O, 10% D_2O and 10 mM sodium phosphate buffer, pH 6.4 at 4°C, using a combination jump-return watergate pulse sequence for water suppression [36,37]. Nonexchangeable proton double quantum filtered COSY and 400 ms mixing time NOESY spectra were recorded in the same buffer in 100% D_2O at 25°C. Spectra were processed using the GIFA program [37]. Typically, the 512(t1) by 2048 (t2) data sets were treated with shifted sinebell window functions, followed by one order of zero-filling and Fourier transformation. Spectra were referenced to the residual water resonance, set to 4.7 ppm.

To determine the GA base-pairing mode, one-dimensional spectra of the imino protons were analyzed. Key resonance assignments obtained from two-dimensional NOESY and COSY spectra of the non-exchangeable protons gave access to critical information concerning the ribonucleotide and helical conformations.

Site-directed mutagenesis, protein expression and purification

Site-directed mutants of the E4.1 ZF were obtained by PCR using two overlapping primers containing the specific mutations in the helix 2 region, as well as two external primers. Inserts were then cloned in pmalc2 (New England Biolabs; Mississauga, ON, Canada) as a *Eco*RI and *Pst*I fragment. The fusion proteins were expressed in *E. coli* strain DHFα and purified using an amylose resin according to the manufacturer's instructions (New England Biolabs). ZF protein appeared more than 90% pure in Coomassie-blue-stained SDS–polyacrylamide gels. Protein concentration was determined using the Bradford assay [38].

Gel mobility shift experiments were performed as described previously [25]. The probe was prepared by annealing a ^{32}P-radiolabeled strand of the duplex, corresponding to the ZF-contacted strand, with the complementary strand. The duplex was purified by electrophoresis on a non-denaturing 5% polyacrylamide gel containing Tris-borate [39], 50 μM ZnAc and 1 mM $MgCl_2$. The concentration of radiolabeled duplex in the binding reaction was 0.1-0.2 nM (ca. 2000 cpm). Binding reactions were incubated for 1 h at 15°C and run on a 5% polyacrylamide gel in Tris-borate [38], 50 μM ZnAc and 1 mM $MgCl_2$ for 3 h at 4°C. Autoradiograms (taken within the linear range of the X-ray film, Fuji medical super RX) were scanned, and bands were quantified by densitometry. The data were then fit to a single-site, saturable binding isotherm according to the equation: %(bound RNA) = [protein] / K_d+[protein]) using Graphpad Prism version 2.01 (Graphpad Software, Inc.) to calculate the K_d.

Acknowledgements

This manuscript is dedicated to the late Robert Cedergren, who, as a world-class experimentalist and mentor, contributed greatly to Canadian science. We thank C.F. Barbas, III and Y. Choo for helpful comments, C.F. Barbas, III for plasmid C7, S. Michnick for critical reading of the manuscript and L. Craig, M. Rashed, M. Zwick, J. Shen, K. Brown, L. Bonnycastle and G. Piñeyro for technical help. R.C. was a Richard Ivey Fellow of the Canadian Institute of Advanced Research. This work was supported by operating grants from the Medical Research Council of Canada (R.C.) and from the Natural Sciences and Engineering Research Council of Canada (J.K.S.).

References

1. Burd, C.G. & Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615-621.
2. Tan, S. & Steitz, T.A. (1988). Eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* **8**, 41-48.
3. Saenger, W. (1984). *Principles of Nucleic Acid Structure*. Springer-Verlag, New-York.
4. Shi, Y. & Berg, J.M. (1995). Specific DNA-RNA by zinc finger proteins. *Science* **268**, 282-284.
5. Pavletich, N.P. & Pabo, C.O. (1991). Zinc finger–DNA recognition: crystal structure of a Zif268–DNA complex at 2.1 Å. *Science* **252**, 809-817.
6. Rebar, E.J. & Pabo, C.O. (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671-673.
7. Jamieson, A.C., Kim, Sung-Hou & Wells, J.A. (1994). *In vitro* selection of zinc fingers with altered DNA-binding specificity. *Biochemistry* **33**, 5689-5695.
8. Choo, Y. & Klug, A. (1994). Toward a code for the interaction of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc. Natl Acad. Sci. USA* **91**, 11163-11167.
9. Choo, Y. & Klug, A. (1994). Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc. Natl Acad. Sci. USA* **91**, 11168-11172.
10. Wu, H., Yang, Wei-Ping & Barbas III, C.F. (1995). Building zinc fingers by selection: toward a therapeutic application. *Proc. Natl Acad. Sci. USA* **92**, 344-348.
11. Jamieson, A., Wang, H. & Kim, S-H. (1996). A zinc finger directory for high-affinity DNA recognition. *Proc. Natl Acad. Sci. USA* **93**, 12834-12839.
12. Greisman, H. & Pabo, C.O. (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657-660.
13. Erold-Erickson, M., Benson, T.B. & Pabo, C.O. (1998). High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger–DNA recognition. *Structure* **6**, 451-464.
14. Nekludova, L. & Pabo, C.O. (1994). Distinctive DNA conformation with enlarged major groove is found in Zn-finger and other protein-DNA complexes. *Proc. Natl Acad. Sci. USA* **91**, 6948-6952.
15. Battiste, J.L., *et al.*, & Williamson, J.R. (1996). α-helix-RNA major groove recognition in an HIV-1 Rev peptide–RRE RNA complex. *Science* **273**, 1547-1551.
16. Puglisi, J.D., Chen, L., Blanchard, S. & Frankel, A.D. (1995). Solution structure of Bovine Immunodeficiency Virus Tat-TAR peptide–RNA complex. *Science* **270**, 1200-1201.
17. Isalan, M., Choo, Y. & Klug, A. (1997). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc. Natl Acad. Sci. USA* **94**, 5617-5621.
18. Isalam, M., Klug, A. & Choo, Y. (1998). Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* **37**, 12026-12033.
19. Zhong, G., Smith, G.P., Berry, J. & Brunham, R.C. (1994). Conformational mimicry of a chlamydial neutralization epitope on filamentous phage. *J. Biol. Chem.* **269**, 24183-24188.
20. Scott, J.K. (1992). Discovering peptide ligands using epitope libraries. *Trends Biochem. Sci.* **17**, 241-245.
21. Smith, G.P. & Scott, J.K. (1993). Libraries of peptides and proteins displayed on filamentous phage. *Methods Enzymol.* **217**, 228-257.
22. Brown, T., Leonard, G.A., Booth, E.D. & Chambers, J. (1989). Crystal structure and stability of a DNA duplex containing A(anti).G(syn) base pairs. *J. Mol. Biol.* **207**, 455-457.
23. Pons, J.L., Malliavin, T.E. & Delsuc, M.A. (1996). Gifa V4: a complete package for NMR data-set processing. *J. Biomol. NMR* **8**, 445-452.
24. Xiong, Y. & Sundaralingam, M. (1998). Crystal structure and conformation of a DNA-RNA hybrid duplex with a polypurine RNA strand: d(TTCTTBr5CTTC)·r(GAAGAAGAA). *Structure* **6**, 1493-1501.
25. Ban, C., Ramakrishnan, B. & Sundaralingam, M. (1994). A single 2′-hydroxyl group converts a B-DNA to A-DNA crystal structure of the DNA–RNA chimeric decamer duplex d(CCGGC)r(G)d)CCGG) with a novel intermolecular G-C paired quadruplet. *J. Mol. Biol.* **236**, 275-285.
26. Berg, J.M. & Shi, Y. (1996). The galvanization of biology: a growing appreciation for the roles of zinc. *Science* **271**, 1081-1085.
27. Clemens, K.R., *et al.*, & Gottesfeld, J.M. (1993). Molecular basis for specific recognition of both RNA and DNA by a zinc finger protein. *Science* **260**, 530-533.
28. McBryant, S.J., *et al.*, & Gottesfeld, J.M. (1995). Interaction of the RNA binding fingers of Xenopus transcription factor IIIA with specific regions of 5 S ribosomal RNA. *J. Mol. Biol.* **248**, 44-57.
29. Zang, Wei-Qing & Romaniuk, P.J. (1995). Characterization of the 5S RNA binding activity of Xenopus zinc finger protein p43. *J. Mol. Biol.* **245**, 549-558.
30. Caricasole, A., *et al.*, & Ward, A. (1996). RNA binding by the Wilms tumour suppressor zinc finger proteins. *Proc. Natl Acad. Sci. USA* **93**, 7562-7566.

31. Finerty, P.J.& Bass, L.B. (1997). A Xenopus zinc finger protein that specifically binds dsRNA and RNA-DNA hybrids. *J. Mol. Biol.* **271**, 195-208.
32. Houbaviy, H.B., Usheva, A. & Burley, S.K. (1996). Co-crystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl Acad. Sci. USA* **93**, 13577-13582.
33. Friesen, J.F. & Darby M.K. (1998). Specific RNA binding proteins constructed from zinc fingers. *Nat. Struct. Biol.* **5**, 543-546.
34. Kunkel, T.A, Bebenek, K. & McClary, J. (1991). Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* **204**, 125-139.
35. Craig, L., *et al.*, & Scott, J.K. (1998). The role of structure in antibody cross-reactivity between peptides and folded proteins. *J. Mol. Biol.* **281**, 183-201.
36. Plateau P. & Gueron, M. (1982). Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J. Am. Chem. Soc.* **104**, 7310-7311.
37. Piotto M., Saudek V. & Sklenar V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, *J. Biomol. NMR* **2**, 661-665.
38. Ausubel, F.M., *et al.*, & Struhl, K. (1992). *Current Protocols in Molecular Biology*. Greene Publishing Associates and Wiley-Interscience, New York.
39. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning* (2nd edn). Cold Spring Harbor Laboratory Press, New York.

Because *Chemistry & Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cmb – for further information, see the explanation on the contents pages.