Nucleic Acids Research, 2000, Vol. 28, No. 24 **4865–4872**

A novel four zinc-finger protein targeted against p190^{BcrAbl} fusion oncogene cDNA: utilisation of zinc-finger recognition codes

Andrew R. McNamara and Kevin G. Ford*

Department of Molecular Medicine, The Rayne Institute, Guy's, King's and St Thomas' Schools of Medicine and Dentistry, 123 Coldharbour Lane, Camberwell, London SE5 9NU, UK

Received September 1, 2000; Revised and Accepted November 8, 2000

ABSTRACT

A three zinc-finger protein that binds specifically to the cDNA representing the unique fusion gene *BcrAbl*, associated with acute lymphoblastic leukaemia, has previously been characterised. At this breakpoint, a sequence homology of 8/9 bp exists between the *BcrAbl* (fusion) and c-*Abl* (parental) target sequences. We show that the three zinc-finger protein discriminates poorly between the fusion (*BcrAbl*) and parental (*Abl*) sequence (K_ds of 42.8 and 65.1 nM, respectively). In order to improve the discriminatory properties of this protein, and to demonstrate the utility of current zinc-finger databases, we have added a fourth zinc-finger to the original three zinc-finger protein. This fourth finger recognises a 3 bp subsite derived from the *Bcr* portion of the breakpoint and is not present in c-*Abl*. This novel four finger protein, which now recognises a 12 bp sequence, demonstrates improved specific binding to *BcrAbl* (K_d = 17 nM). More significantly we have shown that there is now enhanced discrimination between *BcrAbl* and *Abl* sequences by the four finger protein than the original three finger protein.

INTRODUCTION

Zinc-fingers of the Cys_2/His_2 type are widely distributed in nature and demonstrate sequence-specific DNA binding (reviewed in 1). The structural solutions of a single zinc-finger peptide and latterly of the zinc-finger protein *Zif268* complexed with DNA has revealed the zinc-finger polypeptide backbone as having a well-defined α-helix packed against two β-strands arranged in a hairpin structure (2–4). Key residues of each finger (positions –1, 3 and 6 of the α-helix) are involved in both the recognition of a specific 3 bp DNA subsite, and in stabilising the finger in the major groove, through contacts with the primary DNA phosphate backbone (2,5). Substitutions of these key residues in a zinc-finger have been shown to alter the DNA recognition attributes of the protein, allowing it to bind with greater affinity to the new DNA target than to its original target sequence (6,7).

Phage display technology enabling the expression, selection and isolation of zinc-finger peptides with novel DNA binding specificities, has allowed attempts to elucidate a recognition code, describing the relationships between the sequence of the DNA 3 bp subsite and the critical residues on each zinc-finger's α-helix (8–16). Recently, these recognition codes have been evaluated by looking at how well the predicted amino acid sequence correlates to the actual sequence of fingers selected by phage display for a given DNA target sequence (17). The work concluded that there was indeed a correlation between data obtained by optimising zinc-fingers by phage display and the predicted amino acid sequence. As a by-product of this work, a large body of data now exists that empirically relates a number of zinc-finger peptide sequences to their respective 3 bp DNA subsites.

Great effort is currently being invested in selecting zinc-finger proteins designed to bind biologically relevant targets such as viral promoters, transcription factor binding sites and chromosomal translocations. One such translocation, a cytogenetic hallmark of acute lymphoblastic leukaemia (ALL), is that between the *Bcr* and c-*Abl* sequences, giving rise to a unique gene, p190^{BcrAbl}, that encodes a 190 kDa oncogenic fusion protein (18). This has provided the basis for a model system in which the transcriptional inhibitory properties of specific zinc-finger proteins have been evaluated.

A three finger protein designed to recognise the unique fusion site in the p190^{BcrAbl} cDNA has been described (19). *In vitro* expression of this zinc-finger protein in a murine cell line stabley expressing the p190^{BcrAbl} oncogene resulted in transcriptional repression of this cDNA. The K_d for the anti BcrAbl three zinc-finger protein binding to the p190^{BcrAbl} cDNA fusion site sequence *in vitro* has been reported as $6.2 \pm 0.4 \times 10^{-7}$ M. Close analysis of *BcrAbl* versus *Abl* sequences, however, reveal only a single base difference at the extreme 5′-end of the 9 bp recognition site (Fig. 1a) and we now show that the three zinc-finger protein is able to bind to both sequences with significant affinity, in gel retardation assays.

In an attempt to improve the discriminatory properties of this polypeptide for its eventual use as a gene targeting protein for the loading of effector molecules at specific DNA sequences, we have employed a structural extension strategy, constructing a four zinc-finger protein by single zinc-finger addition. The amino acid sequence of the fourth finger was derived from published database information describing individual zinc-finger peptide sequences and their observed DNA binding subsite. The four finger protein has a 12 bp recognition

*To whom correspondence should be addressed. Tel: +44 207 848 5907; Fax: +44 207 733 3877; Email: kevin.ford@kcl.ac.uk

(a)

Oligodeoxynucleotide name	Sequence
BcrAbl	GATGGCGAGGGCGCCTTCCATGGA**GAC**GCAGAAGCCCTTCAGCGGCCA
3zfBcrAbl	GATGGCGAGGGCGCCTTCCATGGAataGCAGAAGCCCTTCAGCGGCCA
NonSp	GTCGATTGCGAGGGTACCTTCCATGTAACTATTCTCTAACTTCAGGCT
Bcr	GATGGCGAGGGCGCCTTCCATGGA**GAC**GCAGgtgagTTCCTCAGCCCA
Abl	CTTTTTTCTGTTCCCCCCTTTCTCTTcCAGAAGCCCTTCAGCGGCCAG
BcrAbl20	CGCTGACGCAGAAGCCCTCC
Abl20	CGCTCTTCCAGAAGCCCTCC

(b)

```
                    -1  3  6            linker
MAEEKP      FQCRICMRNFSDRSSLTRHTRTH      TGEKP       Finger 1
                                                     (GCC)
             β      β         α

                    -1  3  6
            FQCRICMRNFSQGGNLVRHLRTH      TGEKP       Finger 2
                                                     (GAA)
             β      β         α

                    -1  3  6
            FQCRICMRNFSQAATLQRHLKTH      TGEKp       Finger 3
                                                     (GCA)
             β      β         α

                    -1  3  6
            fqcricmrnfsdrsnlerhtrth      tgeklerphrd Finger 4
                                                     (GAC)
             β      β         α
```

Figure 1. (**a**) The amino acid sequences of pG3zf and pG4zf proteins. pG3zf is shown in upper case. pG4zf is essentially pG3zf with the addition of the fourth finger (lower case). The zinc co-ordinating cysteine and histidine residues are in bold. The β sheets and α helices are underlined. The key 'DNA-recognising' residues are indicated. (**b**) Nucleotide sequences of the oligodeoxynucleotides *BcrAbl*, *3zfBcrAbl*, *NonSp, Bcr* and *Abl* sequences. The 9 bp target sequence of pG3zf is underlined in *BcrAbl*. The fourth finger recognises the immediate upstream subsite GAC (bold) of *Bcr*. Homology of *Abl*, *Bcr* (i.e. the breakpoint regions) and *3zfBcrAbl* with *BcrAbl* is also underlined.

sequence, extending the original recognition site by 3 bp upstream into the *Bcr* region of the breakpoint sequence (Fig. 1a). Here we present the results of binding studies demonstrating improved discriminatory binding properties for the four zinc-finger protein, confirming improved potential for this protein as a targeting tool for the loading of transcriptional effector molecules onto DNA. We therefore also demonstrate that constructional approaches based on simple structural extension, using zinc-finger database information, are sufficient to rapidly generate functional targeting proteins.

MATERIALS AND METHODS

Cloning the three and four zinc-finger genes

The gene for the anti-*BcrAbl* three zinc-finger protein was kindly provided by Yen Choo (LMB, Cambridge, UK) (19).

The gene was amplified by PCR using the following primers with an *Age*I restriction site incorporated into the 3′ primer (underlined): 5′-GCGCAAGCTTCG**CATATG**GCAGAAG-AGAAGCCTTTTCAGTGTCGAATCTGC-3′ and 5′-GCGC-**CTCGAG**TTACTTCTCACCGGTGTGGGTCTTTAGGTG-TCTCTG-3′.

The PCR product was cleaved by *Nde*I and *Xho*I (bold text) and ligated into pET15b to create pET15b3zf. Oligodeoxynucleotides were synthesised coding for the fourth finger (see Fig. 1b) with overhangs compatible for ligation into *Age*I (5′) and *Xho*I (3′) sites. The annealed oligodeoxynucleotide was phosphorylated using T4 polynucleotide kinase (NEB) and ligated into pET15b3zf restricted by *Age*I/*Xho*I, to create pET15b4zf. The three and four zinc-finger genes were recloned into a pGex5X-3 vector with a modified multiple cloning site to create pGex3zf and pGex4zf, allowing recombinant proteins to be expressed as glutathione-*S*-transferase (GST) fusions and to be purified by GST affinity chromatography. The integrity of clones was confirmed by nucleotide sequencing.

Expression and purification of target proteins

Plasmids pGex3zf and pGex4zf were used to transform *Escherichia coli* ER1647 cells [*F- fhuA2 Δ(lacZ)r1 supE44 trp31 mcrA1272::*Tn*10* (Tet^r) *his-1 rpsL104* (Str^r) *xyl-7 mtl-2 metB1 Δ(mrcC-mrr) 102::*Tn*10* (Tet^r)]. Inductions were carried out as follows: 50 ml of LB medium was incubated with a single colony and grown at 37°C in the presence of 50 μg/ml carbenicillin overnight. The cells were harvested by centrifugation (10 000 *g*, 5 min, 4°C) and the cell pellet resuspended in 1 ml of LB and used to inoculate 500 ml of LB medium, supplemented with 100 μM $ZnSO_4$, 50 μg/ml carbenicillin and 0.5% (w/v) glucose. Cells were grown at 37°C with agitation to an OD_{595} of 0.8. Isopropyl-β,D-thiogalactopyranoside was added to a final concentration of 0.5 mM and growth continued for a further 3 h at 30°C. Cells were harvested by centrifugation at 4620 *g* for 10 min at 4°C and resuspended in 10 ml lysis buffer (50 mM Tris–HCl pH 7.5, 750 mM NaCl, 2 mM EGTA, 1 mM EDTA, 1 mM DTT and 20% glycerol). Purification of the protein was performed essentially as described (20) with the following minor modifications. The cell suspension was sonicated and incubated with 250 μl protease inhibitor cocktail (Sigma) and Triton X-100 (to 1%) at 4°C for 30 min with agitation. The cell debris was harvested as before and the supernatant, containing the soluble protein, incubated with GST–agarose beads (pre-equilibrated with lysis buffer) at 4°C for 30 min with agitation. The beads were collected by centrifugation at 1000 *g*, 4°C for 3 min and the pellet washed three times in lysis buffer and once in 50 mM Tris–HCl pH 8.0. The beads were then loaded onto a 2.5 ml disposable plastic column and the recombinant protein eluted by the addition of 10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA and 10 mM reduced glutathione. The protein was dialysed into a solution of 10 mM HEPES pH 7.4, 150 mM NaCl and 3 mM EDTA to remove the glutathione, and glycerol added to 10% (v/v) for storage at –70°C. Protein purity was determined by SDS–PAGE (21) and protein concentration determined in triplicate using the Bio-Rad *DC* Protein Assay (Bio-Rad).

Competition gel retardation assays

Synthetic oligodeoxynucleotides were from either Pharmacia Biotech or MWG Biotech. The sequences are given in Figure 1a. Annealing was promoted by incubation of the complementary single-stranded oligodeoxynucleotides at 94°C for 10 min, with subsequent cooling to 4°C at a rate of 1°C/min in annealing buffer (10 mM Tris–HCl, 10 mM $MgCl_2$, 50 mM NaCl and 1 mM DTT). The oligodeoxynucleotides were 5′-end labelled with [γ-^{32}P]ATP (259 TBq/mmol) (ICN) using T4 polynucleotide kinase (NEB). Unincorporated label was removed using Bio-Rad 6 Spin Columns. Binding reactions typically contained 30 fmol of labelled DNA, 8 pmol protein and 1 μg poly dI–dC/poly dI–dC in 20 mM HEPES pH 7.5, 5 mM DTT, 20 μM $ZnSO_4$, 5 mM $MgCl_2$, 50 mM KCl, 10% glycerol, 0.1% NP-40 and 0.1 mg/ml BSA in a final reaction volume of 20 μl, and incubated for 20 min at room temperature. All competitor DNAs were added at the start of the incubation reaction. The reaction mixture was loaded onto a 4% non-denaturing polyacrylamide (19:1) gel in 0.5× TBE buffer, pre-run for 60 min at 100 V, and electrophoresed at 100 V for 40 min at room temperature. Protein–DNA complexes were visualised after autoradiography with intensifying screens at –70°C overnight.

Dissociation constant determination

To determine the dissociation constant (K_d) for each protein, gel retardations were performed as follows. End-labeled 20mer DNA oligonucleotides (41 fmol) (*BcrAbl20* and *Abl20*, see Fig. 1b) were incubated with 100 ng poly dI–dC/poly dI–dC and increasing concentrations of zinc-finger protein (0.4–5.6 pmol) in 20 mM HEPES pH 7.5, 5 mM DTT, 20 μM $ZnCl_2$, 5 mM $MgCl_2$, 50 mM KCl, 10% glycerol, 0.1% NP-40 and 0.1 mg/ml BSA in a final reaction volume of 25 μl for 20 min at room temperature. The reaction mixture was run on a 4% non-denaturing polyacrylamide gel as described above. The dried gel was scanned using a FLA-2000 PhosphorImager (Fuji), and the subsequent data analysed by Aida densitometry (raytest, Isotopenmeßgeräte, GmgH). Bound protein was determined by quantification of retarded probe signal (assuming a single binding site for each protein on the probe), relative to a control lane probe signal of known concentration. Free protein was calculated by subtraction of bound from total input protein. Each experiment was performed in duplicate. The K_d values for each protein–DNA interaction were calculated as the simple average of the two K_d values derived from Scatchard analysis of each individual data set. For clarity, the data shown in Figure 6 were plotted as bound/free versus bound protein concentration for averaged duplicate data points.

RESULTS

The interactions between the three and four zinc-finger–GST fusion proteins (pG3zf and pG4zf, respectively) with *BcrAbl*, *Abl* and non-specific probes were examined by gel retardation analysis. All experiments were carried out with GST fused to the N-terminus of the zinc-finger proteins. The presence of GST has previously been demonstrated to have no effect on the binding of zinc-fingers to DNA (22). We have also observed that zinc-finger proteins expressed as fusions with poly-histidine behave

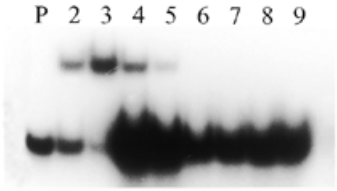

Figure 2. Gel retardation assays of pG3zf and pG4zf binding different radiolabelled oligodeoxynucleotides. Each binding reaction consisted of 0.03 pmol DNA and 8 pmol protein incubated in 10 mM Tris–HCl pH 7.6, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 20 µM $ZnSO_4$, 300 µg/ml BSA, 20 µg/ml poly dI–dC and 5% glycerol in a final volume of 20 µl. Lane P, *BcrAbl* probe alone; lanes 2 and 3, *BcrAbl* oligo with pG3zf and pG4zf, respectively; lanes 4 and 5, *Abl* oligo with pG3zf and pG4zf, respectively; lanes 6 and 7, *Bcr* oligo with pG3zf and pG4zf, respectively; lanes 8 and 9, *NonSp* oligo with pG3zf and pG4zf, respectively.

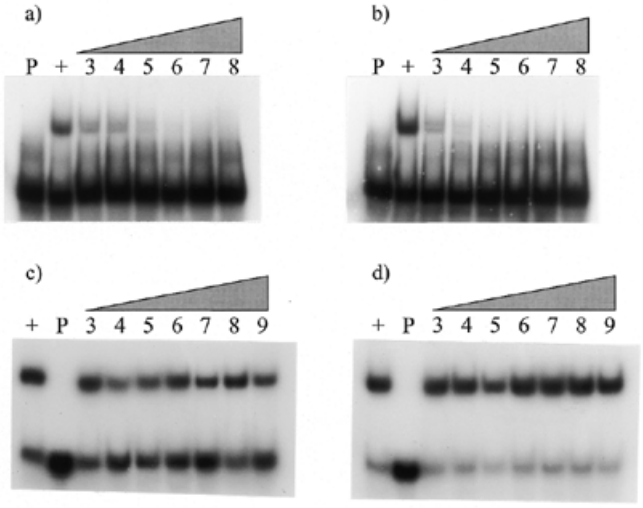

Figure 3. Gel retardation assays of pG3zf and pG4zf binding *BcrAbl* target with increasing amounts of specific and non-specific competitor DNA. For (a) and (b) binding reactions are as described in Materials and Methods. Lane P, probe DNA alone; lane +, probe with zinc-finger protein; lanes 3–8, probe, zinc-finger protein and increasing (50-, 100-, 200-, 500-, 1000- and 2000-fold molar excess) amounts of unlabelled competitor DNA. (**a**) *BcrAbl* probe with pG3zf protein and *BcrAbl* competitor. (**b**) *BcrAbl* probe, pG4zf, *BcrAbl* competitor. (**c** and **d**) Gel retardation assays of pG3zf and pG4zf binding the *BcrAbl* probe in the presence of the *NonSp* competitor. (c) With pG3zf protein. (d) With pG4zf protein. For (c) and (d), assay conditions are as described in Figure 2. Lane +, probe and zinc-finger protein; lane P, probe alone; lanes 3–9, probe, zinc-finger proteins and increasing (1-, 10-, 50-, 100-, 200-, 500- and 1000-fold molar excess) *NonSp* competitor.

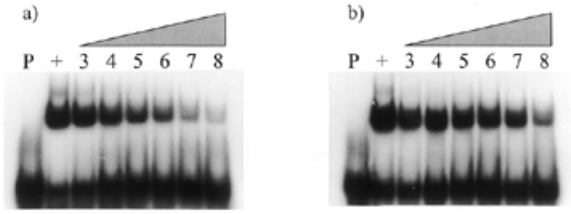

Figure 4. Gel retardation assays of pG3zf and pG4zf binding the *BcrAbl* probe in the presence of the *Abl* competitor. Binding conditions and gel annotation are as shown in Figure 3a and b. (**a**) *BcrAbl* probe with pG3zf protein and *Abl* competitor. (**b**) *BcrAbl* probe with pG4zf protein and *Abl* competitor.

identically to GST fusion zinc-fingers in band shift assays (data not shown). The oligodeoxynucleotide sequences used in the following experiments are given in Figure 1b. Comparative binding of pG3zf protein to *BcrAbl* and *Abl* probes is shown in Figure 2, lanes 2 and 4, respectively. The protein binds to both probes, although much more weakly to the *Abl* probe (note relative probe intensities). Similarly, the pG4zf protein also bound to both probes (Fig. 2, lanes 3 and 5) but bound less weakly to the *Abl* probe. Comparison between the two proteins reveals that pG4zf binds more strongly to the *BcrAbl* probe than pG3zf (Fig. 2, lanes 2 and 3). However, pG4zf binds less strongly than pG3zf to the *Abl* probe (Fig. 2, lanes 4 and 5). These differences in binding interaction become more significant when the relative intensities of the free probes are considered. In this experiment, the *Abl* probe is more intense than the other probes, but the retardation signals seen are weaker, relative to those seen for the *BcrAbl* probe. Figure 2 also reveals that neither pG3zf or pG4zf proteins interact significantly with *Bcr* nor non-specific sequences (Fig. 2, lanes 6–9).

In order to delineate fully the differences in binding characteristics between the three and four finger proteins for their target sequences, competition gel retardation assays were performed. Labelled *BcrAbl* probe was challenged in the binding reaction with unlabelled *BcrAbl* or *NonSp* DNA. These results are shown in Figure 3a–d. The interaction between pG3zf and the probe was competed out by a level of between 200- and 500-fold molar excess of cold (*BcrAbl*) competitor (Fig. 3a). Similarly, the pG4zf complex was competed out, but by a level of only 100–200-fold molar excess of competitor DNA (Fig. 3b). In contrast, no competition was seen using the non-specific oligodeoxynucleotide *NonSp*, even at 1000-fold molar excess of competitor DNA for either protein species (Fig. 3c and d). These results confirm sequence-specific binding of each protein to the *BcrAbl* probe.

The results of experiments designed to evaluate in more detail the binding characteristics of pG3zf and pG4zf to the *Abl* sequence are shown in Figure 4a and b. The *Abl* sequence shares an 8/9 bp homology to the *BcrAbl* sequence recognised by the original pG3zf protein and also completely lacks the fourth finger subsite (see Fig. 1a). Binding of pG3zf to the *BcrAbl* probe was only partially and gradually competed out by the *Abl* competitor, most evident at 1000-fold molar excess (Fig. 4a, lane 7). However, a general trend in terms of reduction of binding was observed from as little as 100-fold molar competitor excess onward, confirming that although pG3zf binds *BcrAbl* sequence specifically, a small but significant interaction occurs between pG3zf and the *Abl* sequence. The same competition experiment repeated with pG4zf is shown in Figure 4b and reveals negligible competition by the *Abl* sequence at 1000-fold excess but some small degree of competition occurring at 2000-fold molar excess. Moreover, the gradual reduction in the binding of the probe seen for pG3zf was not as pronounced in this instance.

To further assess the effect that the fourth finger had on binding we used a competitor that harboured only the original three zinc-finger, 9 bp recognition site (3zf*BcrAbl*, see Fig. 1); i.e. the fourth finger subsite was absent. The results of

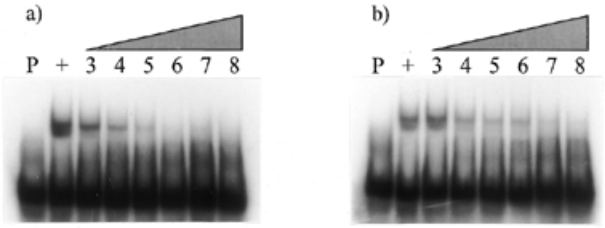

Figure 5. Gel retardation assays of pG3zf and pG4zf binding the *BcrAbl* probe in the presence of the *3zfBcrAbl* competitor. Binding conditions and gel annotation are as shown in Figure 3a and b. (**a**) *BcrAbl* probe with pG3zf protein and *3zfBcrAbl* competitor. (**b**) *BcrAbl* probe with pG4zf protein and *3zfBcrAbl* competitor.

competition binding assays can be seen in Figure 5. In each experiment, labelled *BcrAbl* probe was incubated with either pG3zf (Fig. 5a) or pG4zf (Fig. 5b) in the presence of increasing amounts of *3zfBcrAbl* competitor. Binding of pG3zf to *BcrAbl* was competed out by between a 200- and 500-fold molar excess of *3zfBcrAbl* (Fig. 5a, lane 6), but pG4zf could not be fully competed from *BcrAbl* even at 2000-fold molar excess of competitor, although the intensity of the retarded band was significantly reduced at this end point (Fig. 5b, lane 8). This result is in full agreement with the self-competition experiments described in Figure 3a for pG3zf, and confirms *BcrAbl*, harbouring a 12 bp recognition site, as the target for pG4zf (Fig. 3b).

To fully characterise the binding affinities of pG3zf and pG4zf proteins for the different DNA sequences, gel retardation assays were performed to determine the relative dissociation constants (K_d) of each protein, in which an increasing amount of each zinc-finger protein was incubated with a constant amount of labelled probe. Scatchard plots, for duplicate experiments, were obtained for each zinc-finger protein binding to the target DNA sequences, *BcrAbl20* and *Abl20*. These experiments were performed under conditions that gave the most linear plots (average linear correlation coefficient for data sets = 0.92), and were identical for each protein assayed (Materials and Methods). The averaged Scatchard analyses for these experiments is shown in Figure 6. pG3zf and pG4zf bound to the *BcrAbl* probe with K_ds of 42.8 (±3.3) and 17.05 (±5.7) nM, respectively. This represents an ~2.5-fold increase in affinity for the four finger protein relative to the three finger protein, for the *BcrAbl* sequence. pG3zf and pG4zf bound to the *Abl* probe with K_ds of 65.1 (±11.9) and 121 (±15.7) nM, respectively, confirming competition assay results which suggested a reduced affinity for the *Abl* sequence for pG4zf relative to pG3zf. More significantly, it can be seen that there is a 7-fold difference in binding affinity for pG4zf between *BcrAbl* and *Abl* sequences, compared to only a 1.5-fold difference between these DNA species for pG3zf.

In order to demonstrate the effect of the increased recognition-site size of pG4zf on binding to complex genomes, gel retardation assays were performed as before, except that calf thymus DNA was used as the unlabelled competitor DNA. Calf thymus DNA is comparable to the human genome in both size and sequence complexity and represents a realistic *in vivo* scenario in terms of potential zinc-finger interactions. The results from these experiments can be seen in Figure 7a (pG3zf) and Figure 7b (pG4zf). The difference between the

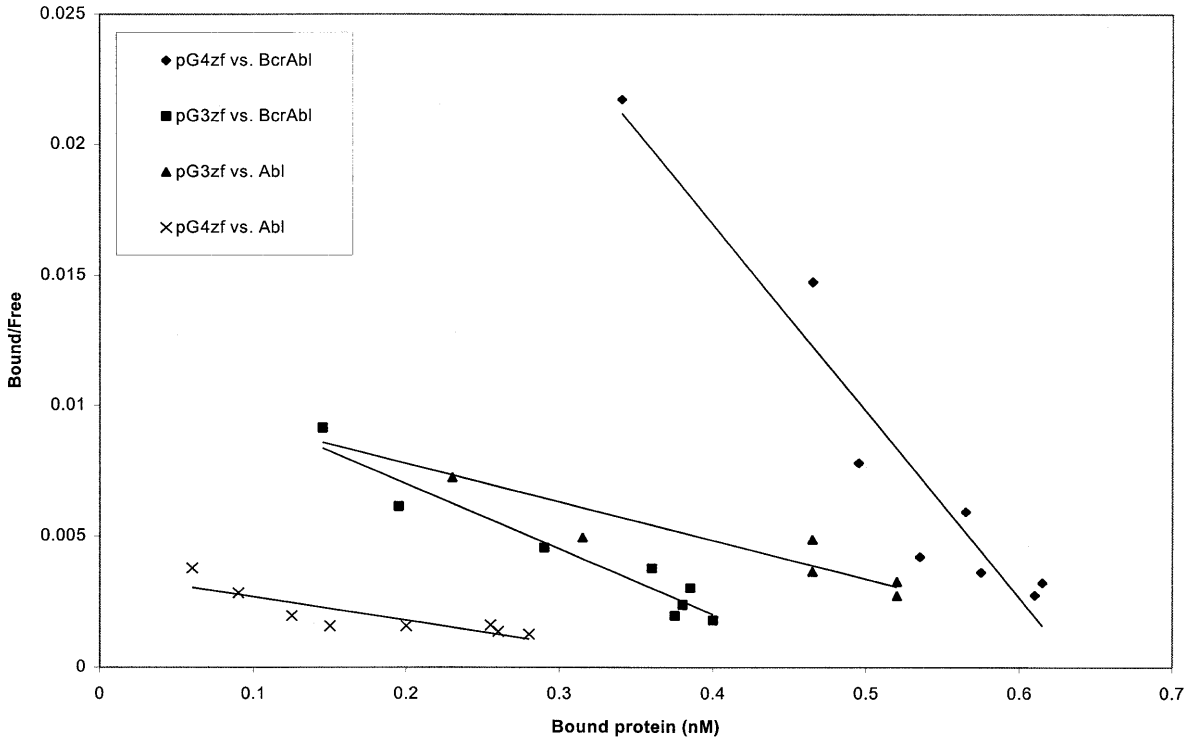

Figure 6. Scatchard analysis for the binding of pG3zf and pG4zf proteins to *BcrAbl20* and *Abl20* oligonucleotides (for details see Materials and Methods).

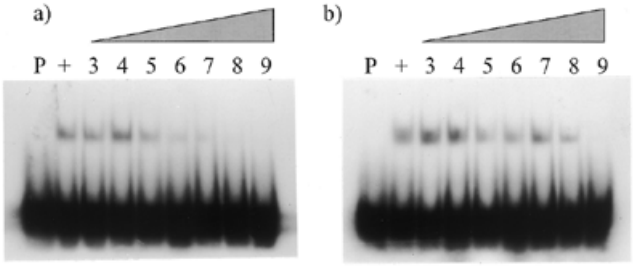

Figure 7. Gel retardation assays of (**a**) pG3zf and (**b**) pG4zf binding to the *BcrAbl* probe in the presence of sonicated calf thymus DNA (Sigma) as competitor. Binding conditions are as shown in Figure 3, except 300 ng of poly dI–dC/poly dI–dC were used per lane. Also, calf thymus DNA was added at 0.1 ng (lane 3), 1 ng (lane 4), 10 ng (lane 5), 50 ng (lane 6), 100 ng (lane 7), 0.5 µg (lane 8) and 1 µg (lane 9). Lane P, probe alone.

two results is immediately apparent. The three finger protein is significantly competed from the labelled *BcrAbl* oligodeoxynucleotide by between 50 and 100 ng of competitor calf thymus DNA. However, the four finger protein remains in complex with the *BcrAbl* target probe up to a level of 500 ng of competitor (Fig. 7b, lane 8). Again these results concur with the observation of pG4zf's increased specificity and discriminatory properties for the target site over pG3zf described in previous experiments.

DISCUSSION

The data presented here describe the characterisation of a synthetic four zinc-finger protein, designed against a specific DNA sequence. The four zinc-finger protein was constructed in order to demonstrate that problems associated with DNA sequence-specific discrimination by a three zinc-finger protein could be solved without the need for labour-intensive phage display methodologies, designed to re-optimise the binding specificity of the protein. Rather, our approach involved addition of a fourth zinc-finger peptide that would recognise the 5′-GAC-3′ 3 bp subsite in the *Bcr* portion of the fusion gene, immediately 5′ to the original three zinc-finger recognition site. We reasoned that since this subsite is not present in the *Abl* sequence, pG4zf would bind this sequence more weakly, but bind the *BcrAbl* sequence with an increased affinity. Consideration of available structural information and the observation that four zinc-finger DNA binding proteins occur naturally [e.g. Wilms' Tumour suppressor (WT1)], led us to assume that addition of a fourth zinc-finger, to a pre-existing three zinc-finger core, would not impose undue torsional/structural constraints on subsequent protein–DNA interactions. In fact, functional proteins have been seen in nature that contain arrays of zinc-fingers up to and including nine zinc-finger units. The amino acid sequence of the fourth finger was selected from a consideration of published data (8), which described a single zinc-finger peptide capable of recognising the 5′-GAC-3′ triplet. The standard zinc-finger linker sequence, TGEKP, was used to join the third and fourth fingers.

Phage display methodologies, where each zinc finger is optimised in the context of a pre-selected neighbouring finger (i.e. iterative methodologies), inherently encompass the binding contributions that arise from the synergy between fingers (23) and 'end effects' (24), where each zinc-finger can interact to some extent with the preceding finger and adjacent subsites. While we concede that such iterative phage display approaches offer optimum prospects for the isolation of high affinity/discriminatory zinc-finger proteins where these synergistic effects are taken into account, our results suggest that a consideration of these potential interactions is not essential for the construction of improved DNA binding proteins through structural extension approaches. The therapeutic potentials of synthetic zinc-finger proteins are numerous and are already being currently exploited through their ability to bind unique sequences within the cell, such as chromosomal translocations, viral sequences and unique elements arising from cellular mutations. The use of zinc-fingers as loading molecules for other effector proteins such as activators/repressors of gene expression (25,26) and endonucleases (27) are also being evaluated.

The data presented are consistent with our hypothesis that the addition of a fourth zinc-finger to the pG3zf protein would enhance its binding and discriminatory properties. We have shown that for the same target (*BcrAbl*), pG4zf binds with higher affinity than pG3zf. It was also shown that when the DNA target at the fourth finger subsite was not complementary to the fourth finger (*3zfBcrAbl* and *Abl* sequences) the affinity of pG4zf was significantly reduced. The reason for this significant decrease in affinity, and that seen when pG4zf tries to bind to *Abl*, is, we believe, due to the inability for appropriate bond formation to occur between residues –1, 3 and 6 of the α helix of the fourth finger and DNA. Furthermore, the inability of the fourth finger to bind the DNA may result in the destabilisation of the rest of the protein, in terms of the interactions of the other three fingers with the DNA.

We have also been able to demonstrate by both competition assay and kinetic analysis, that the single base-pair difference between *Abl* and *BcrAbl* at the breakpoint sequence can affect the binding properties of the zinc-finger protein, i.e. pG3zf shows a greater affinity for 9/9 'correct' bases compared to 8/9 'correct' bases. This difference between *Abl* and *BcrAbl* is a guanine to cytosine substitution. Previous work on the recognition code has shown that cytosine is under-represented at the 5′-end of subsites (8,9). The reason for this remains unclear, but the interaction of adjacent zinc-fingers may play a part (23,24).

The relative order of binding affinities obtained from our study, i.e. pG4zf:*BcrAbl*, pG3zf:*BcrAbl*, pG3zf:*Abl*, pG4zf:*Abl* in descending order of strength of interaction, reflects both the expected increase in affinity of the four finger protein for the target probe and a reduced affinity for the *Abl* probe due to lack of a fourth finger subsite. Omission of poly dI–dC in binding reactions did not significantly alter the K_d values obtained, nor affect the relative order of binding affinities of each protein for each sequence, and suggests that the DNA–protein complexes formed under the conditions of these experiments are due to specific rather than non-specific interactions (data not shown).

The observed K_d for pG4zf binding to the *BcrAbl* sequence (17.05 nM) is in the same range reported for other three zinc-finger proteins, such as *Zif-268* for example (11). Comparing binding affinities to those determined by other groups for similar proteins should however be a task undertaken with caution. For example, the K_d that we determined for the three finger protein against *BcrAbl* is almost 15-fold lower than that

demonstrated by Choo *et al.* (19) This group, however, measured the binding affinities of bacteriophage expressing zinc-finger proteins on their surface, and this, along with variations in protein or DNA concentrations used for example, can affect the apparent binding affinities. It has also been reported that excessive intracellular concentrations of proteins harbouring phage-selected zinc-finger DNA-binding domains results in binding to DNA sequences with up to two base changes from the target site (25). Thus, protein level seems to be crucial in maintaining the binding specificity of zinc-finger proteins and certainly a number of *in vivo* studies relating to zinc-finger protein–DNA interactions may need to be re-evaluated in light of this observation and a degree of caution applied when considering *in vitro* analysis of such proteins.

The increase in binding affinity for the four zinc-finger protein for its target site of only 1.7-fold over the three zinc-finger protein might be considered only a modest improvement, considering that the target site size, and hence the number of possible binding interactions, has increased by 33%. There are a number of possible explanations as to why the observed increase in binding affinity does not reflect the increase in potential DNA–protein contacts. First, as we have already stated, we have not refined the four finger protein, in terms of a mutation and re-selection strategy, so that interaction of the additional finger with its immediately adjacent finger is non-optimal.

Recent work by Segal *et al.* (28) might also suggest that the primary amino acid sequence of the single zinc-finger peptide used in this study may not be optimal. This work looked at zinc-fingers binding to all 16 combinations of the subsite 5′-GNN-3′ and concluded that the 'fundamental' DNA-recognising residues at positions –1, 3 and 6 of the α-helix may not alone be enough to describe truly specific zinc-fingers. In fact, they suggest that the α-helix may act as a series of motifs working in concert rather than as individual amino acids. Some evidence for this conclusion can be seen in the crystal structure of the *Zif268*–DNA complex, where the asparagine at position 2 of the helix is observed to interact with the arginine at position –1 (2). Segal *et al.* also suggest that since the residues at positions 1 and 5 of the α-helix make non-specific contacts with the DNA phosphate, removal of these non-specific contacts would increase the importance of the specific contacts, thereby enhancing specificity (28).

It is interesting to note, for comparative purposes, that the four zinc-finger protein WT1 binds to its 12 bp target site with a K_d of 1.14 nM, while under identical conditions the related three zinc-finger protein EGR1, binds to the first 9 bp of the WT1 target sequence (its consensus binding site) with a K_d of 3.55 nM. Thus, in this instance, the influence of an additional zinc-finger has effectively only increased the binding affinity of WT1 by 3.1-fold (29). In light of such observations, the increase in K_d observed for our own four zinc-finger protein may in fact be within expected parameters. Of more significance, however, is that pG4zf now shows an increased discrimination between the *BcrAbl* and parental *Abl* sequence (K_ds of 17.05 and 121.0 nM, respectively: an ~7-fold difference), with certainly no loss in affinity for its new target site, compared to the original three zinc-finger protein. Therefore, whilst the increase in affinity for the target site may seem modest, when considered in concert with its increased discriminatory properties, pG4zf is more specific for *BcrAbl*, than pG3zf, when presented with related sequences.

The increase in binding site size for the four finger protein, from 9 to 12 bp, results in a recognition site that statistically should occur randomly only once every 16.7 Mb, whereas the original 9 bp recognition site would only be expected to occur once every 250 kb. By using calf thymus DNA as unlabelled competitor, we have tried to recreate this scenario *in vitro*. We subsequently showed that even with these extreme amounts of random, competitor DNA, while pG3zf was still specific for *BcrAbl*, pG4zf was remarkably more resistant to competition, reflected in the K_d values observed. Our interpretation is that addition of the single zinc-finger has indeed resulted in a marked reduction in the number of potential 'random' sites that the protein might specifically interact with. Moreover, if one considers the scenario whereby zinc-fingers are coupled to additional effector proteins, possessing their own intrinsic DNA binding sites, then the combined recognition site would be unlikely to be randomly present within the genome.

In conclusion, we have demonstrated that zinc-finger proteins can be modified by structural extension, based purely on currently available zinc-finger recognition code databases, so that their discriminatory properties and gene targeting attributes are enhanced. We also show that addition of only a single zinc-finger can be sufficient for this purpose. Our results therefore suggest that improved gene targeted effector molecules, with potential therapeutic applications, can be produced rapidly and simply in response to need.

ACKNOWLEDGEMENTS

This work was supported by the Leukaemia Research Fund of the UK. K.G.F. is also supported by The Royal Society.

REFERENCES

1. Kaptein,R. (1991) Zinc fingers. *Curr. Opin. Struct. Biol.*, **1**, 63–70.
2. Pavletich,N.P. and Pabo,C.O.(1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*, **252**, 809–817.
3. Elrod-Erickson,M., Rould,M.A., Nekludova,L. and Pabo,C.O. (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure*, **4**, 1171–1180.
4. Lee,M.S., Gippert,G.P., Soman,K.V., Case,D.A. and Wright,P.E. (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science*, **245**, 635–637.
5. Berg,J.M. (1988) Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc. Natl Acad. Sci. USA*, **85**, 99–102.
6. Thiesen,H.J. and Bach,C. (1991) Determination of DNA binding specificities of mutated zinc finger domains. *FEBS Lett.*, **283**, 23–26.
7. Thiesen,H.J. and Schroder,B. (1991) Amino acid substitutions in the Sp1 zinc finger domain alter the DNA binding affinity to cognate Sp1 target site. *Biochem. Biophys. Res. Commun.*, **175**, 333–338.
8. Choo,Y. and Klug,A. (1994) Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc. Natl Acad. Sci. USA*, **91**, 11163–11167.
9. Choo,Y. and 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. Jamieson,A.C., Kim,S.H. and Wells,J.A.(1994) In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry*, **33**, 5689–5695.
11. Jamieson,A.C., Wang,H. and Kim,S.H. (1996) A zinc finger directory for high-affinity DNA recognition. *Proc. Natl Acad. Sci. USA*, **93**, 12834–12839.
12. Wu,H., Yang,W.P. and Barbas,C.F.,III (1995) Building zinc fingers by selection: toward a therapeutic application. *Proc. Natl Acad. Sci. USA*, **92**, 344–348.

13. Rebar,E.J. and Pabo,C.O. (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science*, **263**, 671–673.
14. Desjarlais,J.R. and Berg,J.M. (1992) Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc. Natl Acad. Sci. USA*, **89**, 7345–7349.
15. Desjarlais,J.R. and Berg,J.M. (1994) Length-encoded multiplex binding site determination: application to zinc finger proteins. *Proc. Natl Acad. Sci. USA*, **91**, 11099–11103.
16. Greisman,H.A. and Pabo,C.O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science*, **275**, 657–661.
17. Wolfe,S.A., Greisman,H.A., Ramm,E.I. and Pabo,C.O. (1999) Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J. Mol. Biol.*, **285**, 1917–1934.
18. Hermans,A., Heisterkamp,N., von Lindern,M., van Baal,S., Meijer,D., van der Plas,D., Weidemann,L.M., Groffen,J., Bootsma,D. and Grosveld,G. (1987) Unique fusion of bcr and c-abl genes in Philadelphia chromosome positive acute lymphoblastic leukaemia. *Cell*, **51**, 33–40.
19. Choo,Y., Sanchez-Garcia,I. and Klug,A. (1994) In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature*, **372**, 642–645.
20. Ford,K., Taylor,C., Connolly,B. and Hornby,D.P. (1993) Effects of co-factor and deoxycytidine substituted oligonucleotides upon sequence-specific interactions between MspI DNA methyltransferase and DNA. *J. Mol. Biol.*, **230**, 779–786.
21. Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
22. Pomerantz,J.L., Sharp,P.A. and Pabo,C.O. (1995) Structure-based design of transcription factors. *Science*, **267**, 93–96.
23. Isalan,M., Choo,Y. and Klug,A. (1997) Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc. Natl Acad. Sci. USA*, **94**, 5617–5621.
24. Choo,Y. (1998) End effects in DNA recognition by zinc finger arrays. *Nucleic Acids Res.*, **26**, 554–557.
25. Choo,Y., Castellanos,A., Garcia-Hernandez,B., Sanchez,G. and Klug,A. (1997) Promoter-specific activation of gene expression directed by bacteriophage-selected zinc fingers. *J. Mol. Biol.*, **273**, 525–532.
26. Kim,J.S., Kim,J., Cepek,K.L., Sharp,P.A. and Pabo,C.O. (1997) Design of TATA box-binding protein/zinc finger fusions for targeted regulation of gene expression. *Proc. Natl Acad. Sci. USA*, **94**, 3616–3620.
27. Nahon,E. and Raveh,D. (1998) Targeting a truncated Ho-endonuclease of yeast to novel DNA sites with foreign zinc fingers. *Nucleic Acids Res.*, **26**, 1233–1239.
28. Segal,D.J., Dreier,B., Beerli,R.R. and Barbas,C.F. (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. *Proc. Natl Acad. Sci. USA*, **96**, 2758–2763.
29. Hamilton,T.B., Borel,F. and Romaniuk,P.J. (1998) Comparison of the DNA binding characteristics of the related zinc finger proteins WT1 and EGR1. *Biochemistry*, **37**, 2051–2058.