

Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases

Jeffrey G. Mandell¹ and Carlos F. Barbas III^{1,2,*}

¹Department of Molecular Biology and ²Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

Received February 14, 2006; Revised and Accepted March 24, 2006

ABSTRACT

Individual zinc finger (ZF) domains that recognize DNA triplets with high specificity and affinity can be used to create designer transcription factors and nucleases that are specific for nearly any site in the genome. These domains can be treated as modular units and assembled to create a polydactyl protein that recognizes extended DNA sequences. Determination of valid target sites and the subsequent design of ZF proteins (ZFPs) is error-prone and not trivial, however. As a result, the use of ZFPs have been restricted primarily to those labs with the appropriate expertise. To address these limitations, we have created a user-friendly utility called Zinc Finger Tools (ZF Tools) that can be accessed at the URL <http://www.zincfingertools.org>. User-supplied DNA sequences can be searched for target sites appropriate for either gene regulation or nuclease targeting. Using a database of experimentally characterized zinc finger domains, the amino acid sequence for a ZFP expected to bind to any chosen target site can be generated. A reverse engineering utility is provided to predict the binding site for a ZFP of known sequence.

INTRODUCTION

Designer zinc finger proteins (ZFPs) are a promising technology for basic science and clinical applications such as endogenous gene regulation and gene repair [for recent reviews see (1,2)]. ZF domains function as modular units that primarily recognize DNA sites of 3 bp. ZFs of the type Cys₂His₂ are comprised of ~30 amino acids that code for two β -strands and an α -helix that interacts with the DNA and confers specificity. ZFPs provide a DNA-binding domain that can be fused to various effector domains, such as transcriptional activators and repressors, nucleases and integrases (3). Owing to the

specificity of the ZFP, these various activities can be targeted to nearly any desired sequence in the genome.

A goal in the field of zinc finger design is to obtain the full complement of ZF domains that recognize all 64 DNA triplets with high specificity and affinity. Our lab and others have used phage display, rational design and naturally occurring domains to obtain ZFs that specifically recognize many of the 64 triplets (4–10). We demonstrated that when these domains are fused in modular fashion, the resulting proteins are able to recognize DNA sequences of 9–18 bp with exquisite specificity and high affinity (dissociation constants in the low nanomolar range or better).

Historically, only those labs with certain technical expertise could take advantage of this powerful ZFP technology. As ZFPs have gained substantial attention recently (11,12), those without experience designing ZFPs have become increasingly interested in doing so. Unfortunately, successful design of ZFPs is not trivial. First, selection of the appropriate target site depends on knowledge of those domains possessing high specificity and affinity and collecting this information from the literature is not always straightforward. Second, searching the DNA sequence of interest for a target region with the proper length and base composition is highly tedious. Last, assembly of the final protein coding sequence requires knowledge of backbone and linker sequences, and is error-prone since the N-terminus of a ZFP non-intuitively binds to the 3' end of the target site. Here we describe the Zinc Finger Tools web site, a publicly available resource that aims to ameliorate these problems and make ZF technology accessible to more researchers. The web site includes utilities for identifying potential target sites and for designing the amino acid sequence of a ZFP predicted to bind to a certain target site. To aid prediction of ZFP specificity, multitarget ELISA assay results for individual ZF domains may be readily viewed. In addition, we have created a utility for reverse engineering the amino acid sequence of a ZFP, such as that obtained from a library selection, to determine the expected DNA target sequence.

*To whom correspondence should be addressed. Tel: 858 784 9098; Fax: 858 784 2583; Email: carlos@scripps.edu

MATERIALS AND METHODS

Zinc finger domains

The DNA recognition sequence of each ZF usually corresponds to the N-terminal residues of the α -helix (positions -1 to +6 relative to the start of the α -helix). ZF Tools employs a non-redundant set of 49 helices to target as many DNA triplets (4–8). This set consists of ZFs that recognize all 16 GNN DNA triplets, 15 ANN triplets (ATC not represented), 15 CNN triplets (CTC not represented), TGA, TGG and TAG (Supplementary Table S1). In the few cases where a triplet is recognized by more than one ZF, the finger that was found to be most specific by a multitarget ELISA assay (4) was chosen for use in the database (4–6,8). A few less specific ZF helices were included in the database for exclusive use by the ‘Predict Zinc Finger Protein DNA Binding Site’ tool for backwards compatibility with published domains.

Backbone sequence

The ZF sequence outside the recognition helix is the Sp1C consensus framework shown to have enhanced stability toward chelating agents (13,14), and which functions effectively in the context of modular units (15). This framework consists of an N-terminal backbone that constitutes the two β -strands and associated sequence (amino acid sequence YKCPECGKSFS, β -strands underlined) and a C-terminal backbone that is the C-terminal portion of the α -helix (amino acid sequence HQRTH). The fixed cloning sequences on the N-terminus (amino acid sequence LEPGEKP) and C-terminus (amino acid sequence TGKKTS) are based upon a modified Sp1C framework (16). The linker, with amino acid sequence TGEKP, used to fuse the ZFs is a consensus sequence derived from analysis of the Transcription Factor Database [reviewed in (17)].

ZFP generation

To generate the amino acid sequence for a ZFP, each zinc finger repeat is modeled as two invariant sequences (the N- and C-terminal backbones, respectively) surrounding the variant helix whose sequence depends on the triplet. The linker sequence is placed between successive ZF repeats and the entire construct containing all fingers is flanked by the N- and C-terminal fixed sequences (Figure 3B).

Target site overlap (TSO)

Some ZF domains recognize a four-base subsite (15). This effect, termed TSO, results from an aspartate residue located at position 2 of the α -helix making a contact outside of the targeted triplet (15). Although ZF domains recognizing GNN, ANN, CNN and TNN triplets can contain an aspartate residue at position 2 of the α -helix, TSO contacts have been observed only for ZF domains recognizing GNG triplets followed on the 3' side by a G or T base (15,16). Therefore, target sites with unmet TSO preferences were identified as sequences containing a 5'-GNG-3' triplet not followed on the 3' side by either a G or T base. TSO evaluation did not consider sequence 3' of the target site unless such sequence lay in the non-targeted spacer between two half sites.

RESULTS

In the following sections, each of the utilities that make up Zinc Finger Tools is described. The tools are publicly available at the URL <http://www.zincfingertools.org>. Browsers recommended to be used with the site are Internet Explorer for Microsoft Windows and Safari for Mac OS X.

SEARCH DNA SEQUENCE FOR CONTIGUOUS OR SEPARATED TARGET SITES

Description

This tool enables the user to identify target sites within a DNA sequence that are comprised of any of the 49 DNA triplets that can currently be specifically recognized (Materials and Methods) (Figure 1). Either contiguous or separated target sites can be identified. Contiguous sites would be most appropriate for gene regulation. Separated target sites, with the regions recognized by ZFPs divided by a non-targeted spacer, are most suitable for applications such as nuclease design. An identified target site may be further analyzed, or the amino acid sequence for a ZFP expected to recognize the target site obtained, by clicking the appropriate icon next to each target site.

Input

The user must supply a DNA sequence of up to 10 kb to be searched for target sites. If gene regulation is the goal, searching a \sim 1 kb region of the gene's promoter is recommended. Should the promoter be unknown, the several hundred base pairs up and downstream of the transcriptional start site can provide effective regulatory sites (18). Owing to the relatively high bandwidth required to display the results of large search queries, such searches may be slow to display.

Contiguous sites are those that are uninterrupted and are of a minimum, user-specified length measured in base pairs (termed the ‘minimum target size’). The minimum, and not the exact, size is specified to avoid long and redundant target-site lists. Therefore, if a minimum size of 18 bp is specified, a site of 36 bp may be found. To analyze long target sites, the parse feature may be used to find unique subsites within a longer target site (described below). Separated target sites are comprised of two contiguous ‘half sites’ separated by a non-targeted region termed the ‘core.’ Owing to the constrained nature of separated target sites, the exact, rather than minimum, half-site size (in bp) must be specified by the user. The core sequence can be of any length and must be described using either the IUPAC base nomenclature (19) (detailed on the website) or simply a number, where the number refers to the number of IUPAC ‘N’ bases (N matches any base). The relative order in which the half sites occur on the DNA strands determines the juxtaposition of the N- and C-termini of bound ZFPs. The user may choose using the ‘juxtaposition’ radio buttons whether to align N- or C-termini of ZFPs. For nuclease design using the FokI catalytic domain, the C-termini are typically aligned (2,20).

Using the ‘triplets to search’ checkboxes, the user can elect to limit the reported target sites to those comprised of a chosen subset of triplet types (e.g. only GNN and ANN triplets). This feature may be useful when ZF libraries constructed from a

Zinc Finger Tools

See us also at zincfingertools.org.

ZF Tools home
 3/10/06 News
 Intro to ZFPs
 Tools overview
 DNA search tips
 ZFP design tips
 Available triplets
 Synthesize a ZFP
 Effector domains
 Nuclease design
 Citations
 About
 Send feedback

Welcome
 Jeff
 Sign out

Search DNA Sequence for Contiguous or Separated Target Sites

Enter a DNA sequence to be scanned for target sites (10 kb max).

Triplets to search: GNG GN(ACT) ANN CNN TNN

ZF set: Total

Search both strands

Contiguous targets Separated targets

Minimum target size (bp): 18 (for contiguous targets)

Half-site size (bp): 12 (for separated targets)

Core (number of bp or IUPAC bases): NNNNNN (Core, Juxtapose for separated targets)

Juxtapose: C-terms or N-terms

Submit Reset

Figure 1. Home page and the search tool. The interface for the tool for searching a DNA sequence to find potential target sites is shown. The other ZF Tools appear below this on the website and not visible here. On the left are hyperlinks to various information pages. Parameters (described in the text) for searching for either contiguous or separated target sites are shown.

subset of triplets have been screened and the potential target sites are desired or when a particular base composition in the target site is desired (for more details see Discussion). The 'ZF set' option can also be used when searching for targets. Most users should employ the 'total' set comprised of all 49 DNA triplets. The '23–21–19' set is a subset of triplets (Supplementary Table S1) previously used for a library selection (7), and was included to give users of this library an upper estimate of the number of targets they can expect to find.

Output

The search tool identifies all target sites on both strands. Note that a base may belong to as many as three unique target sites on a single strand. A coverage map highlights each base from the searched sequence present in at least one target site (Figure 2). This map permits the rapid identification of those DNA regions containing targets and thus reveals areas of poor or favorable coverage. Because target sites may overlap, the coverage map is potentially redundant and should therefore not be used to identify specific target-site sequences.

Below the coverage map in the output is a table detailing the sequence of each identified target site and other relevant information such its position (Figure 2). Target sites which contain triplets that do not meet TSO requirements (Materials and Methods) are flagged by ZF Tools as having 'TSO issues.' Further, offending GNG triplets are colored in red (Figure 2). Each triplet in a target site is a hyperlink that opens a window displaying the results of a multitarget ELISA specificity assay for the appropriate ZF. This information can be used to assess the specificity of each individual ZF domain, and to predict the

overall specificity of a polydactyl ZFP containing the domains. It should be noted that the specificity of a given ZF domain in the context of an arbitrary ZFP is not well characterized (Discussion). Next to each target site is an arrow icon (Figure 2) that links to the 'Design a Zinc Finger Protein' tool (see below). The design tool generates the amino acid sequence of the ZFP expected to bind the target. Use of this icon to generate an amino acid sequence results in the output of the default backbone and linker sequences. Should sequences other than the defaults be required, direct use of the design tool is necessary. If the target size found is larger than desired (e.g. 36 bp when only 18 bp are preferred), the parse tool, accessed by the double-headed arrow icon next to each target site (Figure 2), should be used. The parse tool (Supplementary Figure S1) moves a window in 3 bp increments along a target site to delineate shorter subsites of a user-specified length. Subsites are listed with triplets that violate TSO rules colored in red. Additionally, pointing the mouse at any triplet in a listed target site results in the immediate display of the appropriate multitarget ELISA graph. Accordingly, a target site may be quickly evaluated to ensure that all corresponding ZF domains are of the desired specificity. An icon is provided next to each target site to obtain the amino acid sequence for the ZFP predicted to bind the subsite.

DESIGN A ZINC FINGER PROTEIN

Description

This tool generates the protein coding sequence for a ZFP expected to recognize the input DNA target sequence. Default backbone and linker sequences are provided that can be readily

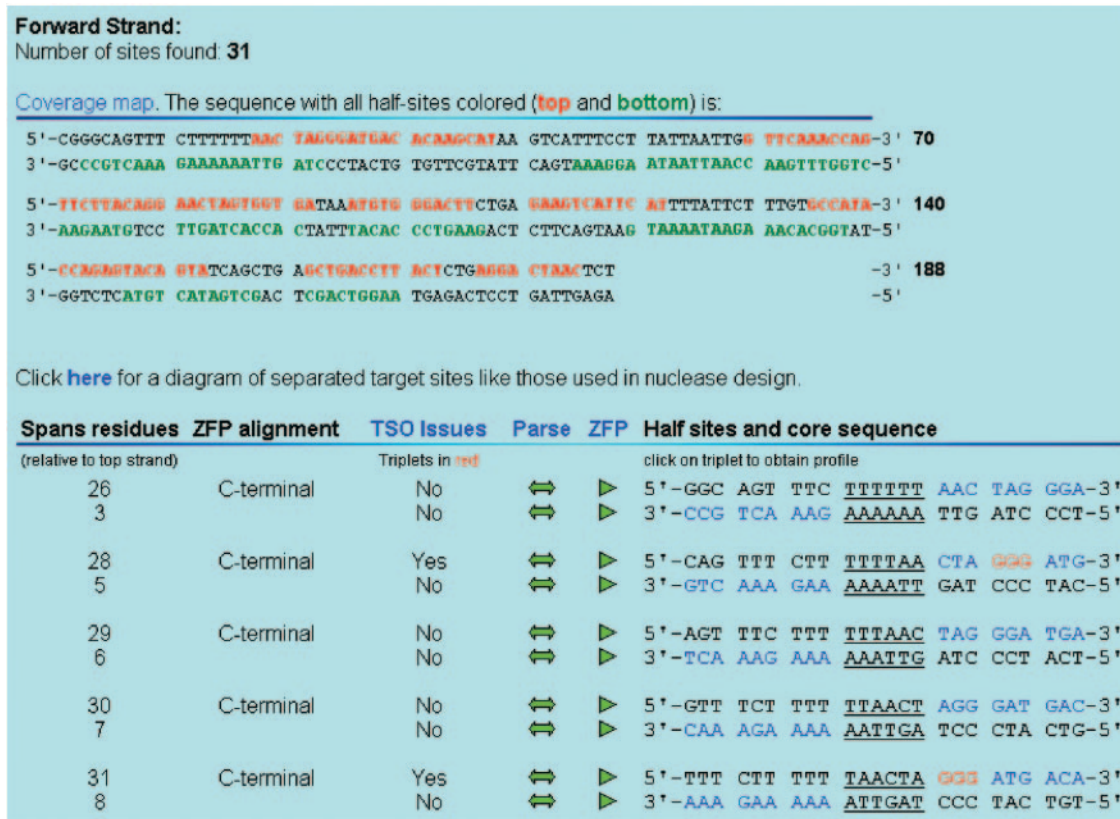


Figure 2. Results of a separated target site search. A search was performed on a DNA sequence 188 bp long for two 9 bp target sites separated by a random 6 bp core. The coverage map highlights those bases that are members of one or more target sites (black bases are in no target sites, red bases are on the top strand, and green bases on the bottom strand). The lower half of the figure displays target sites and non-targeted core sequence (underlined). Each triplet is hyperlinked to the appropriate multitarget ELISA specificity assay. Triplets that do not meet TSO requirements are colored (red) in the list of target sites. Next to each target site are icons either for obtaining the amino acid sequence of a ZFP predicted to bind to the target, or for parsing the target sequence into subsites.

altered by the user. This tool should also be employed when the default backbone or linker sequences provided by the search and parse tools are not desired.

Input

The input for this tool is a DNA target site comprised of valid triplets (Materials and Methods). This target site may have been generated manually or with the aid of the search and parse tools described above. The target site is expected in the 5'-3' orientation. The 'bottom' strands of half sites generated by the search tool are displayed 3'-5' and the orientation must be corrected before manual input with this tool. The default backbone, linker and fixed sequences have been validated in our laboratory and by others and are therefore recommended for most users (Discussion). Nevertheless, the user may edit these sequences through text boxes that display the default values.

Output

The output of the protein design tool is a table that displays each DNA triplet beside the amino acid sequence of the helix expected to recognize it (Figure 3A). Clicking on a hyperlinked triplet yields results of a multitarget ELISA assay for the triplet, thus providing information about its specificity. Note that as per convention, finger number 1 of the ZFP binds to the most 3' triplet. The output also informs the

user whether the ZFP is expected to have TSO issues, and offending triplets are colored in red. The amino acid sequence of the entire ZFP is assembled (Figure 3A and B) and displayed in a text box for convenient copying to other applications or for further manipulation (Discussion).

In order to assess whether a target site is unique within human or mouse genomes, a hyperlink is provided to the NCBI nucleotide-nucleotide (for short, nearly exact matches) BLAST search (<http://www.ncbi.nlm.nih.gov/BLAST/>) (21).

PREDICT ZINC FINGER PROTEIN DNA-BINDING SITE

Description

When supplied with the amino acid sequence of a ZFP, this tool predicts the DNA target sequence. For example, the result of a ZF library selection may yield clones with unknown binding specificity. This tool searches the amino acid sequence and identifies all helices in the ZF Tools database (Materials and Methods). The helices are then queried against the ZF Tools database to determine the corresponding DNA triplets. The algorithm recognizes only helices and ignores all other sequence in an attempt to minimize the impact of poor sequence quality or extraneous sequences. Since only helices are recognized, mutations in a helix sequence will preclude its

A

Design a ZF

N-term backbone: YKCPECGKSFS
 C-term backbone: HQRTH
 ZF linker: TGEKP
 N-term fixed: LEPGEKP
 C-term fixed: TGKKTS

Using 49 triplets.

This target site **has** potential [target site overlap](#) issues (also consider TSO for 3' sequence to this site).
 Triplets with target site overlap issues are in **red**.

Requested target site is:
 5' GGA **GGG** CGA GGA AGG GGG 3'

Perform a short-nucleotide NCBI BLAST with this target sequence against either the [human](#) or [mouse](#) genomes.

Finger	Triplet	Helix
1	GGG	RSDKLVR
2	AGG	RSDHLTN
3	GGA	QRAHLER
4	CGA	QSGHLTE
5	GGG	RSDKLVR
6	GGA	QRAHLER

The amino acid sequence of the ZF protein predicted to bind to the target sequence is:

```
LEPGEKPYKCPECGKSFSRSDDLVRHQRTHTGEKPYKCPECGKSFSRSDDLVTNHQRTHTGEKPYKCPECGKSFSQRAHLE
RHQRTHTGEKPYKCPECGKSFSQSGHLTEHQRTHTGEKPYKCPECGKSFSRSDDLVRHQRTHTGEKPYKCPECGKSFSQR
AHLERHQRTHTGKTS
```

To copy sequence, click mouse in box, select all text (control or ⌘-A), then copy to clipboard (control or ⌘-C), or click on:

(Windows IE users only)

B

LEPGEKP YKCPECGKSFS *Variable* HQRTH (TGEKP) TGKKTS

Figure 3. Generation of the amino acid sequence for a ZFP. (A) The 18 bp target sequence shown was input. Triplets that do not meet TSO requirements are colored (red). Each DNA triplet is shown with the corresponding amino acid sequence of its recognition helix. Clicking on a hyperlinked triplet (blue) opens a window displaying results of a multitarget ELISA assay for the relevant ZF domain. A hyperlink is provided to the NCBI BLAST tool so that the human or mouse genomes can be searched for the target sequence. At the bottom of the screen is the complete amino acid sequence of the ZFP predicted to bind the target. This sequence can be copied to other applications for further manipulation. (B) Order of assembly for fixed, backbone, helix and linker amino acid sequences to generate the sequence of a ZFP. The default amino acid sequence is displayed below each box. The linker is in parenthesis because it is present only between domains and would not be included for a single ZF, as shown here. Sequences between the N-term and C-term fixed sequences repeat with each ZF.

proper identification. This tool can also be used to insure that the intended target site of a designed ZFP is correct; this is especially critical if the amino acid sequence output of the design tool was heavily edited.

Input

The input is the amino acid sequence of a ZFP. If only the DNA sequence is available, it must first be translated and the correct reading frame chosen. The correct reading frame can usually be identified by cursory inspection for backbone sequence.

Output

The predicted target site is displayed, followed by a table listing the location and identity of each identified helix and DNA triplet. Triplets are hyperlinked to multitarget ELISA

specificity assays. Below the table, the input amino acid sequence is shown with any recognized helices highlighted (Supplementary Figure S2). To assess the integrity of ZFPs recovered from a library or another source, it is relatively straightforward for a user to inspect the highlighted output for the expected backbone sequence. A hyperlink is provided to the NCBI BLAST tool so that the human or mouse genomes can be searched for the predicted target sequence.

SEARCH DNA SEQUENCE AND TARGET SITE FOR CLOSE MATCH

Description

This tool permits the user to conduct a search within a long DNA sequence for a shorter sequence that exactly or closely

matches a specified sequence. This tool has two main uses. First, the 'Predict Zinc Finger Protein DNA Binding Site' tool may generate a target site that is not the actual sequence bound by the ZFP. This can occur, for example, if a ZFP binds promiscuously in a library screen against a cloned promoter. In such cases, this tool may be used to search the promoter sequence for a target site that closely matches the predicted site. Second, this tool permits separated target sites (such as nuclease sites) to be identified with more flexibility than the search tool. The search tool is restricted to sequences composed of the 49 triplets for which ZF domains have been validated (Materials and Methods). Greater flexibility may be desired if a user has access to triplets not in the ZF Tools database.

Input

The input DNA sequence can be no longer than 10 kb. The target site to match must be <100 bp and can be described with the IUPAC base nomenclature (19) (provided on the website) to allow undefined or partially defined positions. The user can specify any number of allowed mismatches and these are permitted in any position of the target site. A mismatch contrasts with the IUPAC base 'N' that is restricted to a specific location within the target site. An example of a target-site description using the IUPAC nomenclature to specify a nuclease site with two four-finger proteins separated by a 6 bp core would be NNYNNYNNYNNYNNNNNNNRN-NRNNRNNRNN (where *N* = any base, *R* = G or A, and *Y* = C or T). It is inconvenient and error-prone to generate this type of input so the search tool (described above), that automatically generates the target sequence descriptor, should be used when possible.

Output

The output consists of a list of target sites within the search sequence that match the input criteria.

DISCUSSION

ZF Tools dramatically simplifies the process of zinc finger protein design. The 49 ZF domains that have been experimentally characterized and validated in the context of polydactyl proteins are used for protein design by our laboratory. We chose to populate the ZF Tools database with these zinc fingers since all were selected by similar methods with the intention that they function as independent units and all have been extensively studied. Use of these domains provides a high likelihood of obtaining a functional ZFP. Other zinc finger domains that are minor variants of ours (9), including naturally occurring ones (10), can be incorporated by editing of the ZF Tools amino acid sequence output.

Choosing the right number of fingers and linker sequence for optimal affinity and specificity

Affinity and specificity must be optimized when designing a ZFP. Proteins that bind their target with a dissociation constant (K_d) of 10 nM or better are expected to be productive regulators (18). Naturally occurring three-finger proteins, such as Zif268 and SP1, bind their preferred sequences with K_d values between 10 nM and 10 pM, depending on the assay conditions

(17). ZFPs constructed with the domains included in the ZF Tools database generally have excellent affinities: numerous three-finger proteins targeting the *ErbB-2* gene have K_d values of <10 nM for their preferred targets (22); six-finger proteins 6Fn-642, 6Fn-369 and 6Fn-285 targeting different *ErbB-2* sites have K_d values of 4.8, 4.8 and 9.5 nM, respectively (22); and the K_d values of the six-finger proteins HLTR1, HLTR3 and HLTR6 for their HIV-1 DNA targets are 10, 1 and 6 nM, respectively (23).

Observed DNA affinities tend to improve as the number of fingers increases from one to three, but affinity plateaus beyond three fingers, and only modest improvements in affinity (~70-fold) are seen for six-finger proteins over corresponding three-finger proteins (17). This discrepancy may stem from use of the canonical TGEKP linker. One hypothesis is that the linker results in missing DNA contacts or lost binding energy due to contortion of the DNA (24). Nonetheless, mutagenesis of the linker sequence TGEKP results in substantial loss of DNA-binding affinity (17). In an attempt to ameliorate any defects in the TGEKP linker, use of longer central linkers (9 or 12 residues) between two three-finger proteins containing the canonical TGEKP linker resulted in a 6000-fold tighter interaction with the target site than for the three-finger components (25). In our own studies using these linkers, we have not been able to reproduce this effect. Significantly, it was recently reported that six-finger proteins constructed with only the TGEKP linker performed better for *in vivo* transcriptional regulation than four-finger or five-finger proteins, and better than proteins with 12 residue linkers between two three-finger units or between three two-finger units (26). Our own favorable results are consistent with use of the TGEKP linker between all fingers.

Specificity of a ZFP derives from the number of fingers and the ability of each individual finger to discriminate against other sequences. In general, the domains used by ZF Tools maintain their original high specificity when placed in different positions in a new protein (15). In a survey of 80 three-finger proteins, all bound their intended target and over 90% were highly specific as determined by multitarget ELISA specificity assays (15). However, an 18 bp target site is preferred in order to avoid off-target effects. A 9 bp target site is expected to occur 13 000 times within the human genome, whereas an 18 bp target is expected to occur only once in a genome 20-fold larger than that of humans (15). We previously reported diminished specificity for certain fingers of the E2C protein when moved from a three-finger to a six-finger context (15). Generally, even after factoring in the loss of specificity in certain fingers, the longer recognition sequence of a six-finger protein is expected to confer greater specificity than its three-finger counterpart. Furthermore, if a ZFP binds to both the intended and a related target site, it may not have sufficient affinity (<10 nM) to elicit a biological response at the related target site (15). We find that a six-finger protein designed to recognize an 18 bp target site and constructed using the canonical TGEKP linker is the optimal solution for endogenous gene regulation.

Focusing in on a target site

The list of potential target sites identified by ZF Tools is often long due to the numerous triplets that can now be targeted.

DNase hypersensitivity experiments that indicate accessible DNA regions can be used to assess which of the identified targets are most promising (27). A prudent approach would be to avoid sites that ZF Tools has flagged as having unmet target-site overlap requirements. However, we note that we have obtained high affinity proteins that have unmet target-site overlap requirements. Importantly, ZF Tools does not check for unmet TSO requirements outside of the target-site sequence, unless such sequence falls within the non-targeted spacer between half sites. Users must therefore manually inspect for unmet TSO requirements if the most 3' triplet in a target site is a GNG (Materials and Methods). Lastly, we have anecdotally noted that some of our best performing ZFPs contain many GNG domains; the GNN, ANN, CNN and TNN domains also generally provide high affinity and specificity proteins (data not shown). The search tool provides a means to limit identified target sites to those comprised of a subset of triplets.

From *in silico* to *in vivo*: cloning and effector domains

The DNA coding sequence for a ZFP is typically assembled by standard PCR techniques (6,28). Conveniently, numerous companies have begun recently to offer synthesis of long DNA molecules and the DNA sequence coding for a six-finger protein can be chemically synthesized in good yield. Typically, the DNA encoding the ZFP must be cloned into a vector to create a fusion protein with an effector domain. Activation domains used in the literature include transcriptional activator VP64 (28), a derivative of the Herpes simplex virus protein VP16, and the activation domain of the human p53 protein, a component of the NF- κ B complex (29). Repression domains include the human-derived Krüppel-associated box (KRAB) (30), the Mad mSIN3 interaction domain (SID) (31), the ERF repressor domain (ERD) (32) and a direct fusion with histone deacetylase HDAC1 (33). The vector is then transferred into the cell of interest and the mRNA level of the appropriate gene is monitored, e.g. by RT-PCR. Alternatively, the ZFP response element DNA can be cloned into a minimal promoter upstream of a reporter such as luciferase and activity tested after co-transfection with an expression plasmid containing the ZFP (6).

Future directions

ZF Tools frequently identifies a long list of targets. A desirable feature would be a score for each target that indicates the predicted affinity and specificity for the corresponding ZFP. We have elected not to include a scoring function at this time because much more data are required to provide a reliable rank. In order to establish an ideal scoring system, an in-depth study of the affinity of each domain and its specificity in the context of multiple ZFPs is required. Currently, multitarget ELISA assay information for all triplets is available on the website, thus giving an indication of triplet quality. We are also in the process of selecting for and optimizing the remaining TNN domains.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Roberta Fuller for assistance preparing ELISA graphs, and other members of the Barbas group for helpful discussions. This work was supported by National Institutes of Health Grants CA086258, GM065059 and GM075110. Funding to pay the Open Access publication charges for this article was provided by the National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES

- Blancafort,P., Segal,D.J. and Barbas,C.F., 3rd. (2004) Designing transcription factor architectures for drug discovery. *Mol. Pharmacol.*, **66**, 1361–1371.
- Durai,S., Mani,M., Kandavelou,K., Wu,J., Porteus,M.H. and Chandrasegaran,S. (2005) Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res.*, **33**, 5978–5990.
- Beerli,R.R. and Barbas,C.F.,III (2002) Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.*, **20**, 135–141.
- Segal,D.J., Dreier,B., Beerli,R.R. and Barbas,C.F.,III (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.
- Dreier,B., Segal,D.J. and Barbas,C.F.,III (2000) Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J. Mol. Biol.*, **303**, 489–502.
- Dreier,B., Beerli,R.R., Segal,D.J., Flippin,J.D. and Barbas,C.F.,III (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, **276**, 29466–29478.
- Blancafort,P., Magnenat,L. and Barbas,C.F.,III (2003) Scanning the human genome with combinatorial transcription factor libraries. *Nat. Biotechnol.*, **21**, 269–274.
- Dreier,B., Fuller,R.P., Segal,D.J., Lund,C.V., Blancafort,P., Huber,A., Kokscha,B. and Barbas,C.F.,III (2005) Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, **280**, 35588–35597.
- Liu,Q., Xia,Z., Zhong,X. and Case,C.C. (2002) Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J. Biol. Chem.*, **277**, 3850–3856.
- Bae,K.H., Kwon,Y.D., Shin,H.C., Hwang,M.S., Ryu,E.H., Park,K.S., Yang,H.Y., Lee,D.K., Lee,Y., Park,J. *et al.* (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat. Biotechnol.*, **21**, 275–280.
- Kaiser,J. (2005) Gene therapy. Putting the fingers on gene repair. *Science*, **310**, 1894–1896.
- High,K.A. (2005) Gene therapy: the moving finger. *Nature*, **435**, 577–579.
- Krizek,B.A., Amann,B.T., Kilfoil,V.J., Merkle,D.L. and Berg,J.M. (1991) A consensus zinc finger peptide: design, high-affinity metal binding, a pH-dependent structure, and a His to Cys sequence variant. *J. Am. Chem. Soc.*, **113**, 4518–4523.
- Shi,Y. and Berg,J.M. (1995) A direct comparison of the properties of natural and designed zinc-finger proteins. *Chem. Biol.*, **2**, 83–89.
- Segal,D.J., Beerli,R.R., Blancafort,P., Dreier,B., Effertz,K., Huber,A., Kokscha,B., Lund,C.V., Magnenat,L., Valente,D. *et al.* (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry*, **42**, 2137–2148.
- Beerli,R.R., Segal,D.J., Dreier,B. and Barbas,C.F.,III (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl Acad Sci. USA*, **95**, 14628–14633.
- Wolfe,S.A., Nekludova,L. and Pabo,C.O. (2000) DNA recognition by Cys₂His₂ zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 183–212.

18. Segal,D.J. and Barbas,C.F.,III (2001) Custom DNA-binding proteins come of age: polydactyl zinc-finger proteins. *Curr. Opin. Biotechnol.*, **12**, 632–637.
19. Cornish-Bowden,A. (1985) Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Res.*, **13**, 3021–3030.
20. Kim,Y.G., Cha,J. and Chandrasegaran,S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl Acad Sci. USA*, **93**, 1156–1160.
21. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
22. Lund,C.V., Blancafort,P., Popkov,M. and Barbas,C.F.,III (2004) Promoter-targeted phage display selections with preassembled synthetic zinc finger libraries for endogenous gene regulation. *J. Mol. Biol.*, **340**, 599–613.
23. Segal,D.J., Goncalves,J., Eberhardy,S., Swan,C.H., Torbett,B.E., Li,X. and Barbas,C.F.,III (2004) Attenuation of HIV-1 replication in primary human cells with a designed zinc finger transcription factor. *J. Biol. Chem.*, **279**, 14509–14519.
24. Wuttke,D.S., Foster,M.P., Case,D.A., Gottesfeld,J.M. and Wright,P.E. (1997) Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *J. Mol. Biol.*, **273**, 183–206.
25. Kim,J.-S. and Pabo,C.O. (1998) Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl Acad Sci. USA*, **95**, 2812–2817.
26. Neuteboom,L.W., Lindhout,B.I., Saman,I.L., Hooykaas,P.J. and van der Zaal,B.J. (2006) Effects of different zinc finger transcription factors on genomic targets. *Biochem. Biophys. Res. Commun.*, **339**, 263–270.
27. Liu,P.Q., Rebar,E.J., Zhang,L., Liu,Q., Jamieson,A.C., Liang,Y., Qi,H., Li,P.X., Chen,B., Mendel,M.C. *et al.* (2001) Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J. Biol. Chem.*, **276**, 11323–11334.
28. Beerli,R.R., Dreier,B. and Barbas,C.F.,III (2000) Positive and negative regulation of endogenous genes by designed transcription factors. *Proc. Natl Acad. Sci. USA*, **97**, 1495–1500.
29. Ruben,S.M. and Rosen,C.A. (1990) Suppression of signals required for activation of transcription factor NF-kappa B in cells constitutively expressing the HTLV-I Tax protein. *New Biol.*, **2**, 894–902.
30. Margolin,J.F., Friedman,J.R., Meyer,W.K.-H., Vissing,H., Thiesen,H.-J. and Rauscher,F.J.,III. (1994) Krueppel-associated boxes are potent transcriptional repression domains. *Proc. Natl Acad. Sci. USA*, **91**, 4509–4513.
31. Ayer,D.E., Laherty,C.D., Lawrence,Q.A., Armstrong,A.P. and Eisenman,R.N. (1996) Mad proteins contain a dominant transcription repression domain. *Mol. Cell. Biol.*, **16**, 5772–5781.
32. Sgouras,D.N., Athanasiou,M.A., Beal,G.J., Jr, Fisher,R.J., Blair,D.G. and Mavrothalassitis,G.J. (1995) ERF: an ETS domain protein with strong transcriptional repressor activity, can suppress *ets*-associated tumorigenesis and is regulated by phosphorylation during cell cycle and mitogenic stimulation. *EMBO J.*, **14**, 4781–4793.
33. Ayer,D.E. (1999) Histone deacetylases: transcriptional repression with SINers and NuRDs. *Trends Cell Biol.*, **9**, 193–198.