doi:10.1006/jmbi.2000.4133 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 303, 489–502

JMB

Insights into the Molecular Recognition of the 5′-GNN-3′ Family of DNA Sequences by Zinc Finger Domains

Birgit Dreier†, David J. Segal† and Carlos F. Barbas III*

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

In order to construct zinc finger domains that recognize all of the possible 64 DNA triplets, it is necessary to understand the mechanisms of protein/DNA interactions on the molecular level. Previously we reported 16 zinc finger domains which had been characterized in detail to bind specifically to the 5′-GNN-3′ family of DNA sequences. Artificial transcription factors constructed from these domains can regulate the expression of endogenous genes. These domains were created by phage-display selection followed by site-directed mutagenesis. A total of 84 mutants of a three-domain zinc finger protein have been analyzed for their DNA-binding specificity. Here, we report the results of this systematic and extensive mutagenesis study. New insights into zinc finger/DNA interactions were obtained by combining specificity data with computer modeling and comparison with known structural data from NMR and crystallographic studies. This analysis suggests that unusual cross-strand and inter-helical contacts are made by some of these proteins, and the general orientation of the recognition helix to the DNA is flexible, even when constrained by flanking zinc finger domains. These findings disfavor the utility of existing simple recognition codes and suggest that highly specific domains cannot be obtained from phage display alone in most cases, but only in combination with rational design. The molecular basis of zinc finger/DNA interaction is complex and its understanding is dependent on the analysis of a large number of proteins. This understanding should enable us to refine rapidly the specificity of other zinc finger domains, as well as polydactyl proteins constructed with these domains to recognize extended DNA sequences.

Keywords: novel DNA-binding proteins; phage display selection; rational design; binding specificity; protein:nucleic acid computer modeling

*Corresponding author

Introduction

Work from our laboratory and others has shown that the malleability of the zinc finger motif may be exploited to create new sequence-specific DNA-binding domains (Segal & Barbas, 2000). We have also shown that these domains can be assembled modularly into polydactyl proteins capable of targeting unique sequences in the human genome with a high level of affinity (Beerli *et al.*, 1998; Segal *et al.*, 1999), thus laying the foundation for the development of applications such as gene-specific transcriptional regulators (Beerli *et al.*, 2000) and novel site-specific endonucleases (Chandrasegaran & Smith, 1999). However, in order for us to develop applications that depend on the reliable and reproducible targeting of any sequence with a high degree of specificity, the underlying principle of DNA recognition within this class of proteins must be further explored. The assembly of polydactyl proteins from modular building blocks requires that each subunit performs its task independently. Therefore, each zinc finger domain must be optimized.

A single zinc finger domain consists of approximately 30 amino acid residues with a simple ββα fold stabilized by hydrophobic interactions and the chelation of a single zinc ion (Brown *et al.*, 1985; Lee *et al.*, 1989; Miller *et al.*, 1985). Presentation of the α-helix of this domain into the major groove of

†These authors contributed equally to this work.

E-mail address of the corresponding author: carlos@scripps.edu

0022-2836/00/040489–14 $35.00/0

DNA allows for sequence-specific base contacts. Among the many thousands of zinc fingers which have been identified, the most studied scaffolds for building proteins of novel specificity have been the murine transcription factor Zif268 and the structurally related human transcription factor Sp1. The structure and binding specificity of both proteins have been well studied (Desjarlais & Berg, 1992; Elrod-Erickson *et al.*, 1996; Narayan *et al.*, 1997; Pavletich & Pabo, 1991; Swirnoff & Milbrandt, 1995). The Zif268-DNA complex structure suggested specific roles for each residue in the recognition helix (Figure 1). With respect to the start of the α-helix, positions −1, 3 and 6 (AA^{-1}, AA^{3}, AA^{6}) were generally observed to contact the 3′, middle, and 5′ nucleotides, respectively of a base triplet. Positions −2, 1 and 5 are often involved in direct or water-mediated contacts to the phosphate backbone. Position 4 is typically a leucine residue that packs in the hydrophobic core of the domain. Position 2 has been shown to interact with other helix residues and with bases depending on the helical protein sequence and operator DNA sequence. Its interaction with DNA, when observed, is almost always a cross-strand contact to a base outside the canonical three-nucleotide site (Elrod-Erickson *et al.*, 1996; Isalan *et al.*, 1997; Pavletich & Pabo, 1991). However, the most distinguishing attributes of Zif268 and Sp1 are their relatively limited inter-domain cooperative binding interactions (that is to say, each domain recognizes predominately its cognate three-nucleotide site) and that all three domains interact with the DNA in essentially the same way. This is true for modified Zif268 or Sp1 domains generated by selection (Elrod-Erickson *et al.*, 1998) or rational design (Kim & Berg, 1996). Crystallographic determination of the structures of these mutant proteins bound to

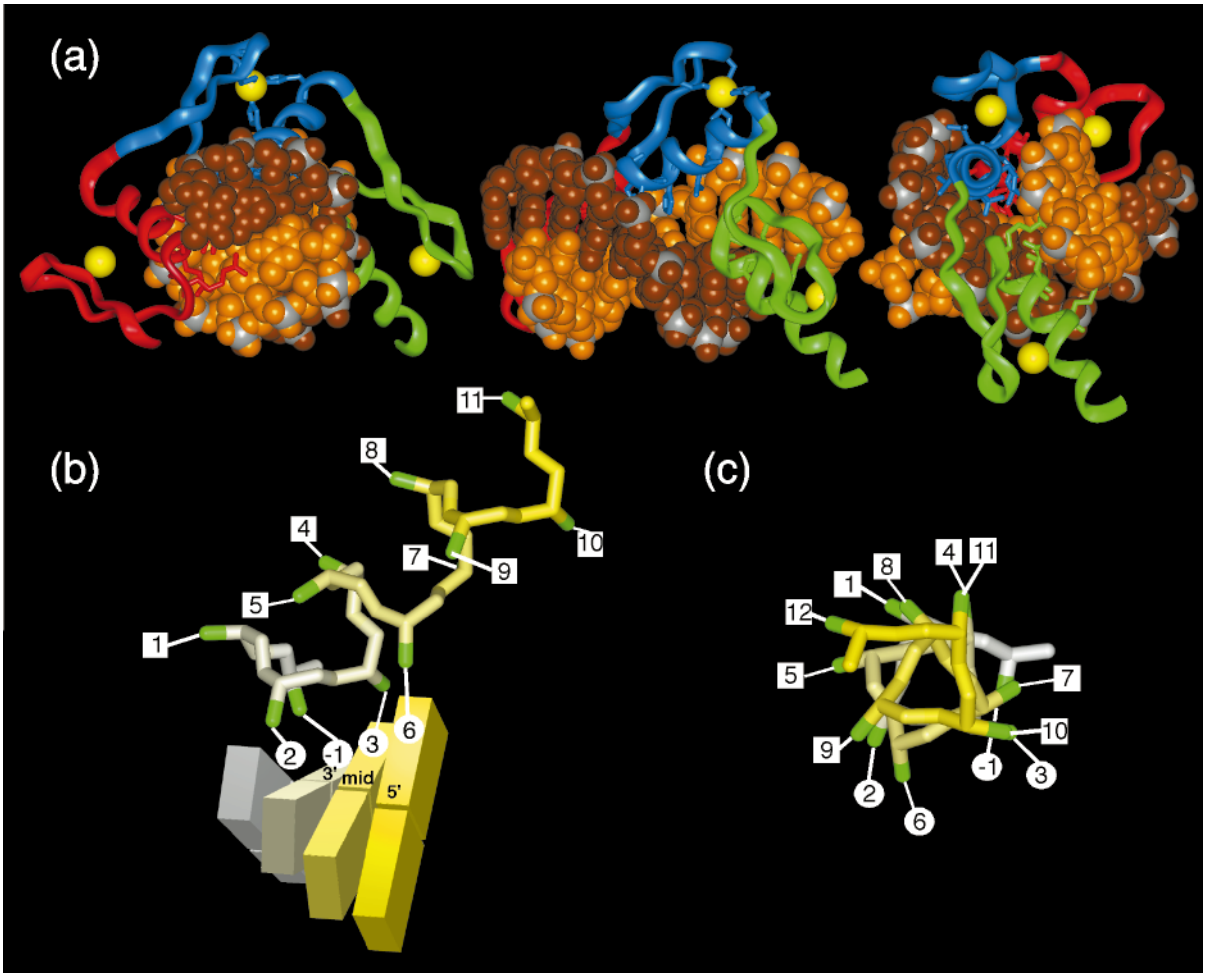

Figure 1. Cys_2-His_2 zinc finger proteins contact DNA with the N terminus of their α-helix. (a) Three rotational views of the three-finger Zif268-DNA complex (Elrod-Erickson *et al.*, 1996). Fingers 1 (red), 2 (blue) and 3 (green) primarily contact one DNA strand (orange) with occasional contacts to the other strand (brown). Phosphate groups (gray) and zinc ions (yellow) are shown. (b) Cartoon of the protein/DNA interface. Amino acid positions −1 through 11 of the α-helix are labeled (β-carbon atoms = green). Residues involved in specific base contacts are highlighted with circular labels. DNA base-pairs are depicted as blocks. The canonical three-nucleotide subsite and one adjacent base-pair are shown, with the 5′, middle and 3′ nucleotides of the heavily contacted strand labeled. (c) An axial view of the α-helix.

their operator DNAs reveals that reorientation of the helix relative to the DNA is sometimes required to achieve the appropriate interactions, but the roles of the amino acid residues are essentially unchanged.

These features have allowed us to construct highly specific polydactyl zinc finger proteins based on the Zif268 and Sp1 protein scaffolds (Beerli *et al*., 1998, 2000). The domains used for their modular assembly were selected by phage display and optimized by rational design (Segal *et al*., 1999). In this study we report on the systematic modifications that were required to optimize the domains that recognize the 5′-GNN-3′ set of DNA sequences. This study attempts to balance the information obtained by crystallographic and NMR studies, in which the elements of specificity for a few helices are investigated in great detail by providing specificity data on 84 closely related helices. Novel interactions, such as cooperativity between positions 2 and 3, are discussed and supported by computer modeling. Overall, our results support the notion that neither selection by phage display nor rational design applied alone are capable of producing domains with specificity sufficient for the practical application of zinc finger technology.

Results

Recognition of the GNG family of DNA sequences

In a previous study, three-finger proteins, in which six residues (helical positions −1, 1, 2, 3, 5, and 6) of finger 2 had been randomized, were displayed on bacteriophage and selected for binding to DNA targets containing all members of the 5′-GNN-3′ family of sequences in the finger-2 recognition subsite (Segal *et al*., 1999). The binding specificity of the new three-finger proteins were then rigorously investigated using a multi-target specificity assay (described previously by Segal *et al*. (1999)). Several finger-2 domains showed high-level specificity, while others showed some degree of cross-specificity. The first to be addressed was the phage-selected helix for 5′-GGG-3′ recognition, RSD-H-LTR (corresponding to helical positions −1, 1, 2, -3-, 4, 5, 6), since this differed from the well-characterized Zif268 finger-2 helix (RSD-H-LTT) by only a change of Thr^6 to Arg^6. Arg^6 restricted 5′-nucleotide recognition from thymine and guanine (Figure 2(a), white bars) to exclusively guanine (Figure 2(b), white bars). However, both domains were still promiscuous for binding adenine or guanine residues in the middle nucleotide position; 5′-G(A/G)G-3′ (Figure 2(a) and (b), black bars).

Rational modification by site-directed mutagenesis was employed to attempt to eliminate the crossreactive binding. The promiscuity for adenine and guanine of the Zif268 finger 2 has been noted by others (Swirnoff & Milbrandt, 1995) and results from the ability of His^3 to make hydrogen bonds to either adenine or guanine (Figure 3(b) and (d)). Other selections with 5′-GGN-3′ targets yielded proteins containing Lys^3 for recognition of middle guanine. A His^3 to Lys^3 substitution was shown to restrict successfully the recognition to only middle guanine, albeit with a 15-fold loss of affinity (Figure 2(c)) (Segal *et al*., 1999). A further Thr^5 to Val^5 substitution was made to assess the ability to use LVR as a common motif to recognize a 5′ guanine. This substitution had no negative effect on binding (Figure 2(d)).

The phage-selected helix for 5′-GAG-3′, RSD-N-LRR, displayed modest crossreactivity with all of the GNG set of targets (Figure 2(e)). Because Asn^3

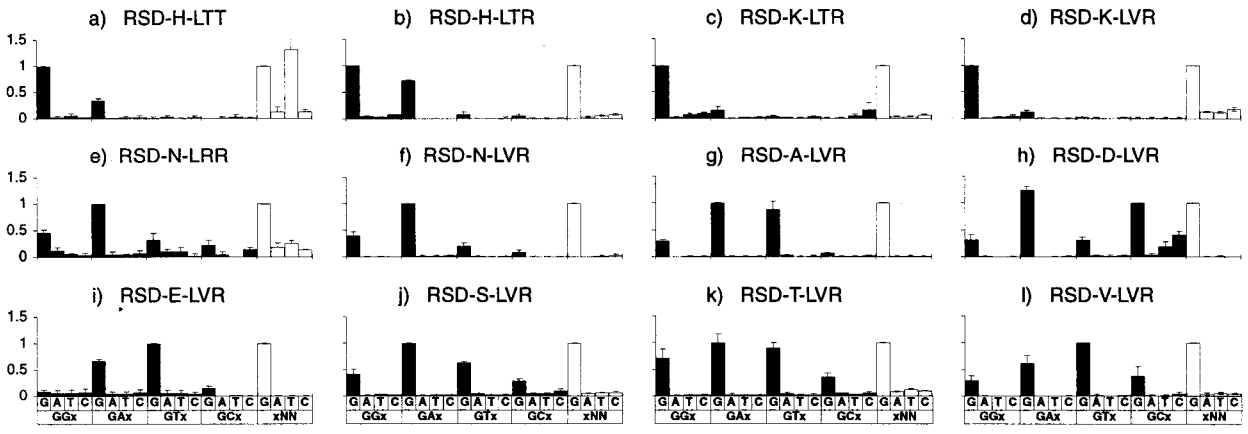

Figure 2. Multi-target specificity assay for helices recognizing 5′-GNG-3′. At the top of each graph is the recognition helix for a finger-2 protein (positions −1 to 6). Position 3 is flanked by dash marks (-) for clarity. Black bars represent target oligonucleotides with different finger-2 subsites: GGG, GGA, GGT, GGC, GAG, GAA, GAT, GAC, GTG, GTA, GTT, GTC, GCG, GCA, GCT, GCC. White bars represent oligonucleotide pools with a unique 5′ nucleotide in their finger-2 subsite: GNN, ANN, TNN, CNN. The height of each bar represents the relative affinity of the protein for each target, averaged over two independent experiments and normalized to the highest signal among the black or white bars. Error bars represent the deviation from the average.

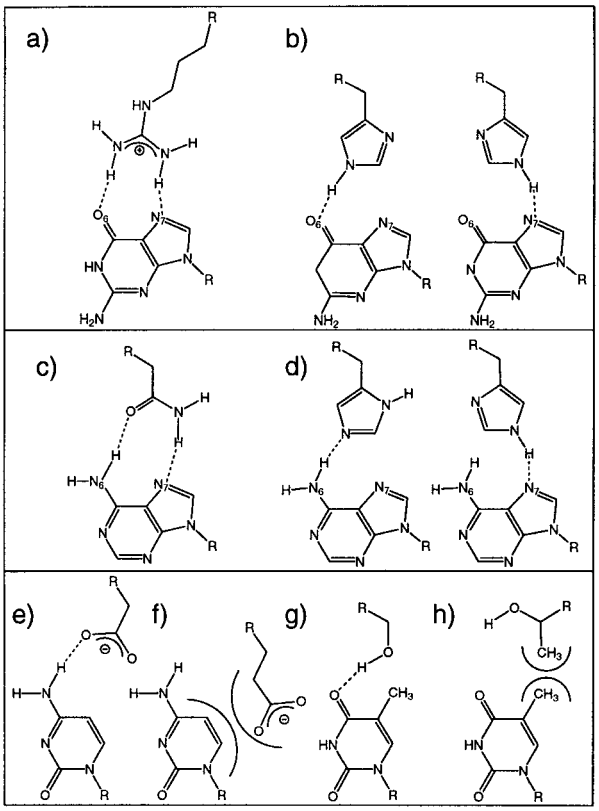

Figure 3. Some amino acid-to-base interactions that have been observed in structural studies. (a) Arg-G, (b) His-G, (c) Asn-A or Gln-A, (d) His-A, (e) Asp-C, (f) Gln-C, (g) Ser-T, (h) Thr-T. Broken lines indicate hydrogen bonding. Curved lines indicate van der Waals interactions. Amino acid residues in (a) and (h) are in an orientation approximating that of position −1 of the recognition helix. Amino acid residues in (b) through (g) are in an orientation approximating that of position 3.

was unanimously selected in all our pannings to recognize a middle adenine residue, no modification was attempted at this position. Instead, we reasoned that Arg^5 might be making non-specific contacts to the phosphate backbone of the DNA, in analogy to Lys^5 in finger 5 of TFIIIA (Nolte *et al.*, 1998). Indeed, changing Arg^5 to Val^5 reduced the background of non-specific binding, although the binding to the 5′-GNG-3′ targets remained essentially unchanged (Figure 2(f)).

The data thus far suggested that 5′-GNG-3′ recognition could be achieved using a common helix, RSD-X-LVR, using the appropriate residue in position 3 to recognize the middle nucleotide. Several different position-3 residues had been selected during phage display against targets containing a middle pyrimidine nucleotide. Unfortunately, specificity analysis revealed strong crossreaction with all 5′-GNG-3′ targets, regardless of whether position 3 was Ala, Asp, Glu, Ser, Thr, or Val (Figure 2(g)-(l)).

The edges of purine bases present two hydrogen bond donors or acceptors in the major groove, whereas those of pyrimidines present only one (Figure 3). One explanation for the strong cross-reactivity is therefore that recognition by these proteins is dominated by the strong bidentate arginine-guanine interactions on the 5′ and 3′ bases and not influenced by pyrimidine interactions at the middle position. This possibility was investigated in two ways. The first was an attempt to disrupt the strong Arg^{-1}/Asp^2 interaction. However, substitutions of Asp^2 to Ala^2 (Figure 4(a)) or Ser^2 (Figure 4(b)), produced proteins of poor specificity (compare with Figure 2(j)).

In a second strategy, Arg^{-1} or Arg^6 was substituted by a lysine residue. Lysine had been selected in these positions to recognize guanine in previous studies (Choo & Klug, 1994b; Isalan *et al.*, 1998; Jamieson *et al.*, 1994; Wu *et al.*, 1995). However, exchanging Arg^{-1} with Lys^{-1} produced fingers of poor specificity, regardless of whether the position 3 residue was Ala, Asp, Ser, Thr, or Val (Figure 4(c)-(g)). Previously, we described a finger (KSA-D-LKR) containing Lys^{-1} that was selected for high-affinity binding to GCG in the finger-1 position (Wu *et al.*, 1995). Surprisingly, this helix bound to 5′-GC(T/C)-3′ in the finger-2 position (Figure 4(h)). The specificity was made moderately worse by changing Asp^3 to Glu^3 (Figure 4(i)), and domains with Val^5 generally produced fingers of very low affinity and specificity, regardless of whether the position-3 residue was Ala, Glu, Ser, or Val (data shown only for Glu^3, Figure 4(j)). In contrast to the effects of Lys^{-1}, Lys^6 produced fingers similar to those containing Arg^6. Lys^6 has been observed to bind 5′ guanine in a number of zinc finger co-crystal structures, including those of TFIIIA (Nolte *et al.*, 1998), GLI1 (Pavletich & Pabo, 1993), and a designed protein (Kim & Berg, 1996). Unfortunately, Lys^6 did not provide improved specificity with Ala, Asp, Glu, Asn, Ser, Thr, or Val in position 3(data shown only for Asp^3 and Glu^3, Figure 4(k) and (l), respectively). These results disfavor the use of Lys^{-1} or Lys^6.

Recognition of the GNA family of DNA sequences

Gln^{-1} was unanimously selected in all our phage-display experiments to recognize a 3′ adenine base (Segal *et al.*, 1999). Our investigation of 5′-GNA-3′ recognition therefore focused on the influence of position 3 in the context of various combinations at positions 1 and 2. We were also interested to know if a common motif could be used for 5′-GNA-3′ recognition, similar to that found for GNG.

The helix that was phage-selected to recognize 5′-GAA-3′, QRS-N-LVR, showed crossreactivity with 5′-GAT-3′ (Figure 5(a)). Using similar reasoning applied to the 5′-GAG-3′ case described above, we mutated the presumptively phosphate-contacting Arg^1 to Ser^1. This resulted in slightly more

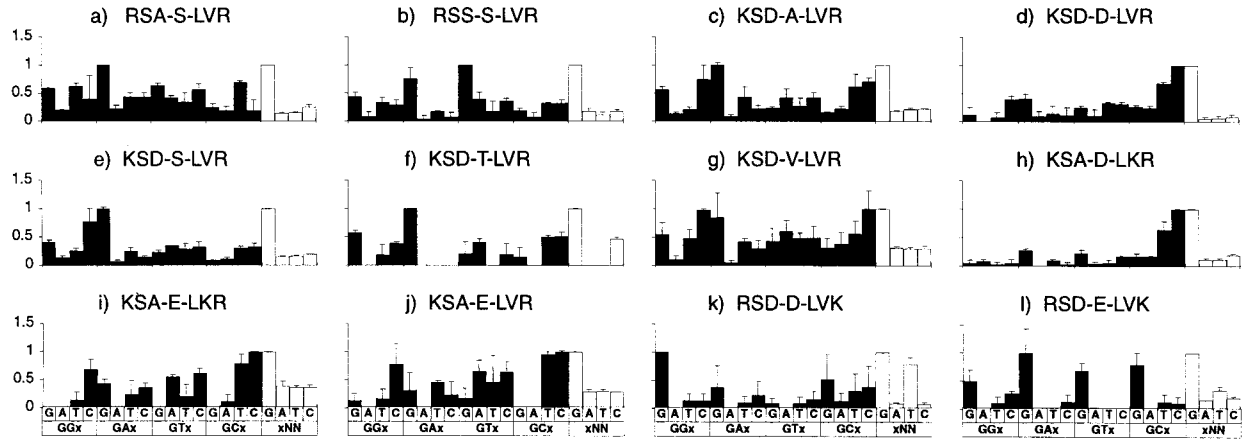

Figure 4. Multi-target specificity assay for helices recognizing 5′-G(T/C)G-3′. Legend as in Figure 2.

specific recognition (Figure 5(b)). The resulting helix, QSS-N-LVR, was similar to the helix which had been phage-selected to recognize 5′-GTA-3′, QSS-S-LVR (Figure 5(c)), suggesting that QSS-X-LVR could be a 5′-GNA-3′ recognition motif. However, the phage-selected helix for 5′-GGA-3′, QRA-H-LER (Figure 5(d)), proved more specific than QSS-H-LVR (Figure 5(e)). The phage-selected helix for 5′-GCA-3′, QSG-D-LRR (Figure 5(f)), was also more specific than QSS-D-LVR (Figure 5(g)), which surprisingly preferred a 3′ thymine, despite the presence of Gln^{-1}. Recognition of 5′-GCA-3′ was problematic. To evaluate the utility of QSG-X-LVR as a 5′-GNA-3′ recognition motif, several position 3 mutants were analyzed. Replacing Arg^{5} of QSG-D-LRR with Val^{5} resulted in slightly reduced 5′-GCA-3′ specificity (Figure 5(h)). QSG-E-LVR (Figure 5(i)) and QSG-T-LRR (Figure 5(j)) preferred recognition of 5′-GTA-3′. Note that the phage selected helix QSS-S-LVR (Figure 5(c)) binds with a specificity profile nearly identical to QSG-E-LVR (Figure 5(i)) despite their very different position-3 residues. QSG-H-LVR (Figure 5(k)) and QSG-N-LVR (Figure 5(l)) were essentially the same as their QSS counterparts.

Recognition of the GNC family of DNA sequences

Asp^{-1}, Glu^{-1}, and Gly^{-1} were selected during phage display using targets with a 3′ cytosine (Segal *et al.*, 1999). One goal was therefore to investigate if a common residue in position −1 could be used to specify a 3′ cytosine residue.

The helix that was phage selected to recognize 5′-GAC-3′, DPG-N-LKR, showed moderate cross-reactivity with 5′-GAT-3′ (Figure 6(a)). Substituting Val^{5} for Lys^{5} reduced crossreactivity (Figure 6(b)). The resulting helix, DPG-N-LVR, was similar to the highly specific phage-selected helix for 5′-GTC-3′, DPG-A-LVR (Figure 6(c)), suggesting that DPG-X-LVR could be a 5′-GNT-3′ recognition motif. DPG-H-LVR proved highly specific for 5′-GGC-3′ (Figure 6(d)). In this case the specificity was better than that of the phage-selected helix, ERS-K-LAR (Figure 6(e)). It is instructive to note that the intermediate steps between these latter two helices

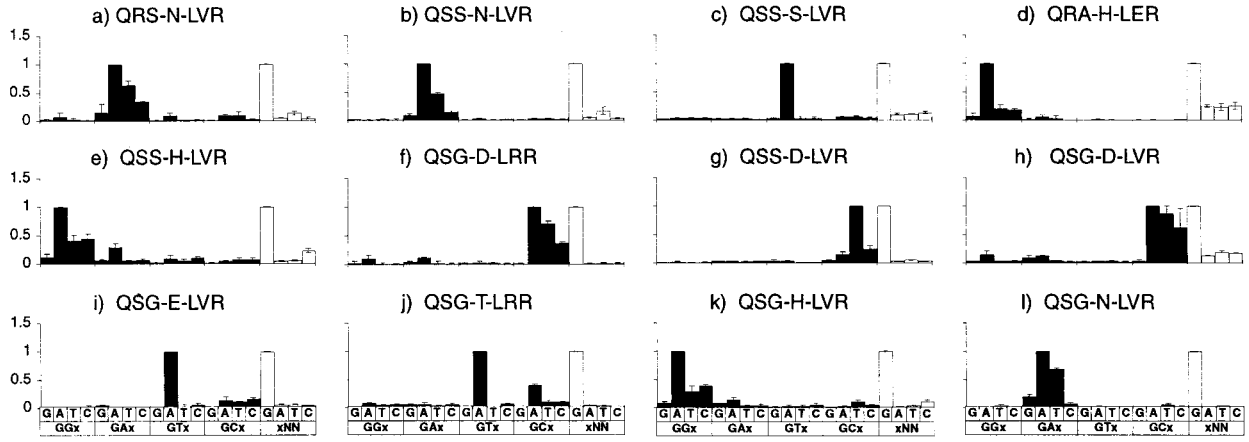

Figure 5. Multi-target specificity assay for helices recognizing 5′-GNA-3′. Legend as in Figure 2.

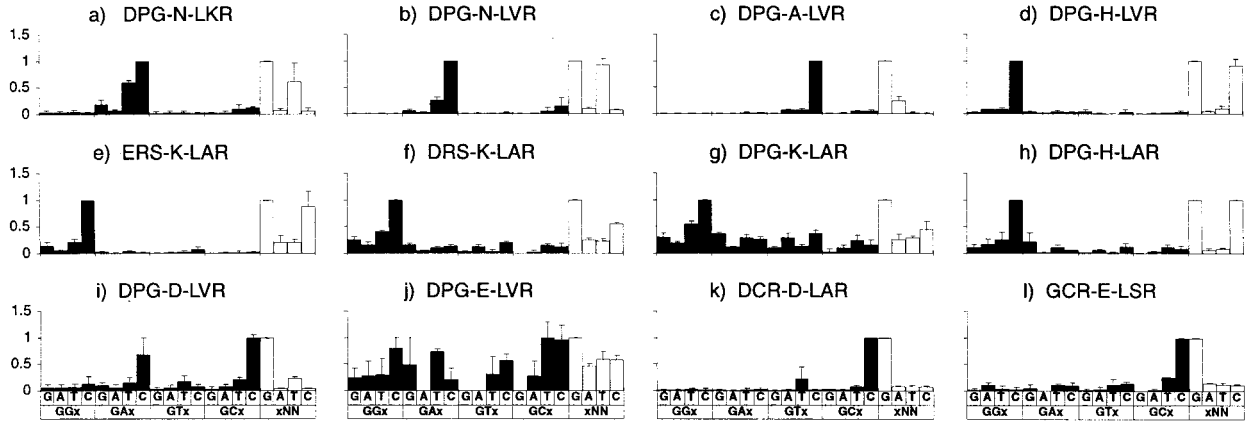

Figure 6. Multi-target specificity assay for helices recognizing 5′-GNC-3′. Legend as in Figure 2. Note that the data regarding 5′ specificity (white bars) are misleading for domains which recognize a 5′-NNC-3′ subsite. The complementary strand of the target oligonucleotide was found to create an alternative binding site, such that domains recognizing 5′-NGC-3′ crossreact with the 5′-CNN-3′ pool, domains recognizing 5′-NAC-3′ crossreact with the 5′-TNN-3′ pool, etc. (Segal *et al.*, 1999).

were all less specific than DPG-H-LVR. For example, simply substituting Asp^{-1} for Glu^{-1} (Figure 6(f)) or even DPG- for ERS- (Figure 6(g)), resulted in proteins of reduced specificity. Even DPG-H-LAR (Figure 6(h)) was less specific than DPG-H-LVR. These examples illustrate the necessity of exploring a large number of sequence combinations and suggest caution in the use of rational design. Unfortunately, DPG-D-LVR (Figure 6(i)) and DPG-E-LVR (Figure 6(j)) were both less specific for 5′-GCC-3′ than the phage-selected helix, DCR-D-LAR (Figure 6(k)). Interestingly, a similar phage-selected helix, GCR-E-LSR, was equally specific for 5′-GCC-3′ but contained Gly^{-1} (Figure 6(l)). Since the 3′ nucleotide can not be specified by Gly^{-1}, this helix must recognize its target site in a fundamentally different way (see Discussion).

Recognition of the GNT family of DNA sequences

Thr^{-1} and Ser^{-1} were most commonly selected during phage display for targets containing a 3′ thymine (Segal *et al.*, 1999). The phage-selected helix, TSG-N-LVR, showed high-level specificity for 5′-GAT-3′ (Figure 7(a)). Because many of the other 5′-GNT-3′ phage-selected helices showed poor specificity (Segal *et al.*, 1999), TSG-X-LVR was tested as a possible 5′-GNT-3′ recognition motif. TSG-H-LVR (Figure 7(b)) was more specific for 5′-GGT-3′ than the phage-selected TAD-K-LSR (Figure 7(c)), which, in fact, preferred 5′-GGG-3′. TSG-E-LVR (Figure 7(d)) seemed modestly more specific for 5′-GCT-3′ than the phage-selected SSQ-T-LTR (Figure 7(e)). The helix QSS-D-LVR (Figure 5(g)) originally constructed to study GCA recognition binds GCT with a specificity approximating TSG-E-LVR (Figure 7(d)). However, recognition of 5′-GTT-3′ proved to be problematic. The phage-selected helix, TSG-S-LTR, preferred other 5′-GNT-3′ sites (Figure 7(f)). Unexpectedly, substituting Val^5 for Thr^5 improved binding to 5′-GTT-3′, although GGT was still the preferred target (Figure 7(g)). TSG-A-LVR (Figure 7(h)) preferred 5′-GGT-3′, while TSG-D-LVR (Figure 7(i)) and TSG-T-LVR (Figure 7(j)) preferred 5′-GCT-3′. Other motifs faired worse. A 5′-GNC-3′-inspired TPG-X-LVR helix was unspecific when position 3 was Ala, Glu, His, Ser, or Thr (data not shown). A 5′-GCT-3′-inspired TSQ-X-LTR failed to bind 5′-GTT-3′ with Ser^3 (Figure 7(k)) or Thr^3 (Figure 7(l)). SSS-S-LVR bound nearly every target (data not shown).

His and Lys in position 3

As discussed previously (Segal *et al.*, 1999), His^3 was promiscuous for adenine or guanine in the middle position in the case of the 5′-GGG-3′-recognition helix RSD-H-LTR (Figures 2(b), 3(b), (d)), but specified only guanine in all other 5′-GGN-3′ helices (Figures 5(d), (e), (k), (d), 6(h), 7(b)). Lys^3 improved specific recognition for middle guanine in RSD-K-LVR (Figure 2(d)), but its application in other 5′-GGN-3′ helices produced unexpected results. A close examination of several mutants suggested that when Asp^2 was not present, Lys^3 resulted in a preference for 5′-GGC-3′. For example, substituting Ala^2 for Asp^2 in the highly specific helix RSD-K-LVR was sufficient to change the specificity from 5′-GGG-3′ toward 5′-GGC-3′ (Figure 8(a) and (b)), although the latter helix preferred 5′-GAG-3′. In addition, preference for 5′-GGC-3′ was observed regardless of whether the residue in position 2 was Ala, Ser, or Gly, and, surprisingly, regardless of whether the residue in position −1 was Q, or T (Figure 8(c)-(h)). In the presence of Asp^2, Lys^3 is observed to only specify middle guanine, as seen in RSD-K-LVR (Figure 2(d)) and TAD-K-LVR (Figure 8(j)). Compare also TAD-H-LVR (Figure 8(i)) with

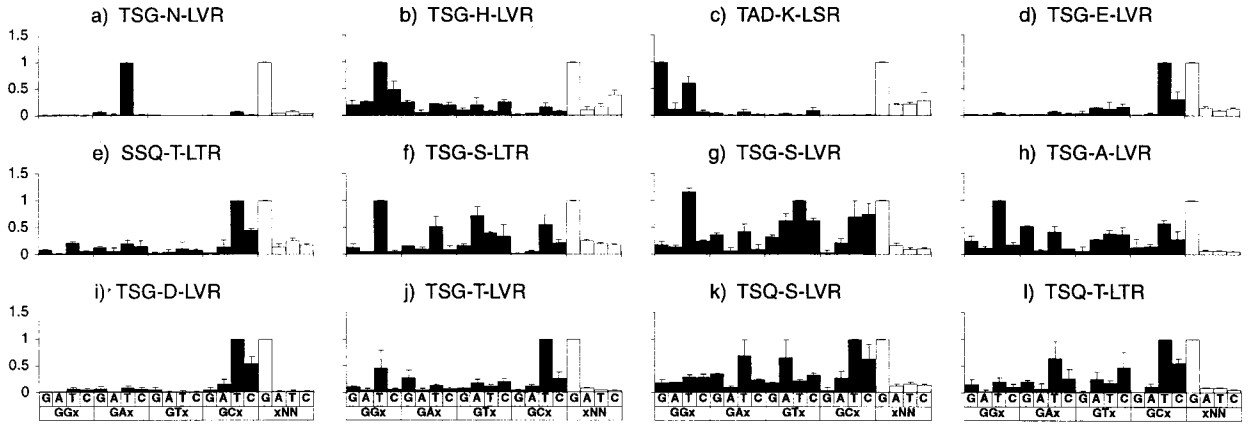

Figure 7. Multi-target specificity assay for helices recognizing 5′-GNT-3′. Legend as in Figure 2.

TAD-K-LVR (Figure 8(j)) and note the middle guanine and adenine binding of TAD-H-LVR.

Discussion

The goal of this study was to investigate zinc finger recognition of the 16-member family of 5′-GNN-3′ DNA sequences in order to understand better zinc finger/DNA-binding mechanisms. Phage-display studies and the rational design of mutant zinc finger domains were used to accomplish this goal. About half of the phage display selected domains showed exquisite binding specificity and therefore did not require optimization. The others domains required optimization to some extent. In the worst case, three finger-2 domains that were selected to recognize one target, actually preferred binding to a different target. Others have reported similar results (Choo & Klug, 1994a; Wolfe *et al.*, 1999), suggesting that inappropriate selection might be a common occurrence in phage display experiments of zinc finger proteins. Therefore, the specificity of novel binding domains created solely by selection, and any ''recognition codes'' derived from such data, must be regarded with caution.

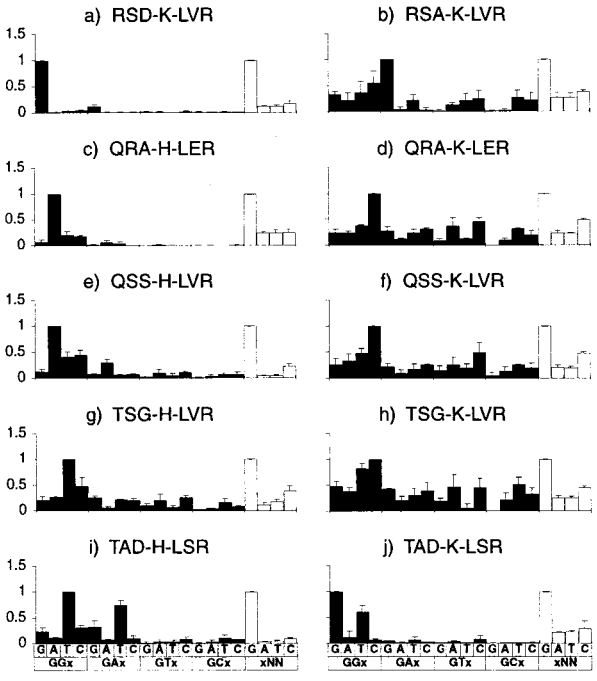

Figure 8. Multi-target specificity assay for helices containing Ser^2/Lys^3. Legend as in Figure 2.

Here, we examined the binding specificity of 84 finger-2 mutants of the 3-finger protein C7 (Wu *et al.*, 1995), a derivative of Zif268. The specificity of each mutant domain was examined for its ability to bind each of the 16 5′-GNN-3′ finger-2 subsites. This approach is arguably more informative than the multiplex target analysis performed by others (Choo & Klug, 1994a; Desjarlais & Berg, 1994), since our method can potentially distinguish, for example, recognition of only 5′-GGC-3′ and 5′-GAG-3′ from 5′-G(G/A)(G/C)-3′ (see Figure 8(b) for example). From these data we have shown that many seemingly benign or conservative substitutions can sometimes have dramatic and unexpected affects on sequence specificity. Consider that the recognition helix QSS-H-LVR (Figure 8(e)) binds 5′-GGA-3′ but QSS-K-LVR (Figure 8(f)) binds 5′-GGC-3′. Note that ''the code'' predicts that both proteins would bind GGA. Helix position 3 was not expected to influence recognition of the 3′ nucleotide, and Gln^{-1}, which had been selected unanimously to recognize a 3′ adenine, failed to specify 3′ adenine in the domain containing Lys^3. As another example, both TSG-D-LVR (Figure 7(i)) and TSG-E-LVR (Figure 7(d)) bind 5′-GCT-3′, but QSG-E-LVR (Figure 5(i)) binds 5′-GTA-3′, while QSG-D-LVR (Figure 5(h)) binds predominantly 5′-GCA-3′. QSS-D-LVR (Figure 5(g)) binds 5′-GCT-3′, also despite the presence of Gln^{-1}. Finally, recognition of GTT was improved by the substitution of valine for threonine in position 5 (Figure 7(f) and (g)), which was not expected to contact any base, and GCR-E-LSR

(Figure 6(l)) is highly specific for a 3′ cytosine, although Gly^{-1} would not be expected to interact specifically with any base.

The fact that these results are at all surprising only highlights the need for a more detailed understanding of zinc finger:DNA interactions. We therefore submit that a major conclusion from our study is that, currently, the best way to make novel domains of high affinity and specificity is by selection, followed by rigorous analysis and optimization involving large sets of mutants. This paradigm is similar to that of the immune system, in which antibodies are selected from the immune repertoire and then optimized through somatic hypermutation. However, our data also suggest some novel interactions, and we offer our speculation on the molecular basis for the observed DNA-binding specificities.

Lessons from the GNG set: interactions of Asp^2 and Lys^3

Lys^3 seemed to specify only middle guanine in the presence of Asp^2 but seemed to direct recognition towards 5′-GGC-3′ when position 2 was not an aspartate residue. An explanation for the restricted activity of Lys^3 in the presence of Asp^2 can be gained from computer modeling. The structure of Zif268 in complex with its operator DNA served as a basis for modeling. A single amino acid residue replacement in finger 2, $Thr^6 \rightarrow$ Arg, and one nucleotide replacement, 5′ thymine → guanine, produced a model for recognition helix RSD-H-LTR (Figure 9(a)). Other amino acid residues and nucleotides were substituted based on this framework to explore potential interactions.

The NMR structure of TFIIIA shows that lysine residues in the recognition helix can undergo dynamic conformational fluctuations (Foster *et al.*, 1997). A reasonable explanation for the specificity of the Asp^2/Lys^3 combination is therefore that Asp^2 draws the ε amino group of Lys^3 towards its carboxyl oxygen atom, allowing Lys^3 to hydrogen bond with both Asp^2 and O6 of the middle guanine base (Figure 9(b)). Since adenine has an amino group in position 6, an Asp^2/Lys^3 interaction would explain the exclusion of middle adenine by the helix RSD-K-LVR (Figure 2(d)). In the absence of Asp^2, Lys^3 may be less restricted. Computer modeling suggests the ability of Lys^3 to interact with O6 of a guanine based-paired to a 3′ cytosine (Figure 9(c)), which might account for the increased recognition of 5′-GGC-3′ by these helices (Figure 8).

Lessons from the GNG set: recognition of middle pyrimidines

Despite considerable effort, no helices were found that could recognize 5′-GTG-3′ or 5′GCG-3′ exclusively. The 5′-GNG-3′ crossreactivity was somewhat surprising because our phage-display experiments showed a strong selection of Ser^3 and Thr^3 for the recognition of 5′-GTG-3′ or 5′-GCG-3′, respectively (Segal *et al.*, 1999). These data also reinforce our position that not all phage-selected sequences are in fact optimal, even when there is a strong consensus sequence in the selected clones.

It is also instructive to note that finger 1 and 3 of Zif268, RSD-E-LTR and RSD-E-RKR, respectively, have been shown to be fairly specific for GCG (Swirnoff & Milbrandt, 1995; Wolfe *et al.*, 1999), and do not significantly crossreact with 5′-GAG-3′. However, in the finger 2 position RSD-E-LVR binds 5′-G(A/T)G-3′ (Figure 2(i)). This observation suggests that the specificity of a finger might change if it is put in a different position. Terminal fingers can cross the major groove in ways the middle finger cannot (Nolte *et al.*, 1998; Wuttke *et al.*, 1997). Local DNA structure may also affect specificity. The extent to which such effects may impact the use of these domains as modular building blocks is currently being investigated by target-site selection assays on novel, multi-finger proteins.

Lessons from GNA: Gln^{-1} flexibility allows multiple interactions

Gln^{-1} has been observed structurally to make bidentate hydrogen bonds with adenine (Elrod-Erickson *et al.*, 1998; Kim & Berg, 1996), similar to the interaction of asparagine shown in Figure 3(c). Although this interaction is obvious chemically, one must consider that a fully extended glutamine residue is approximately 2 Å shorter than an extended arginine residue (Nardelli *et al.*, 1992). Therefore, in comparison with a helix containing Arg^{-1}, the $Glu^{-1}/3'$ adenine interaction requires the reorientation of the α-helix with respect to the DNA. Such a reorientation might influence other interactions along the helix, such as at positions 2 and 3.

It is therefore particularly interesting that, in sharp contrast to the 5′-GNG-3′ situation, QSS-S-LVR (Figure 5(c)) was able to achieve highly specific recognition of a middle thymine, even in the context of two potential double hydrogen bond interactions with the 5′ and 3′ nucleotides. A computer model of this interaction was constructed based on crystallographic data by Pabo and coworkers, in this case QSG-S-LTR bound to 5′-GCA-3′ (Elrod-Erickson *et al.*, 1998). However, our model of QSS-S-LVR bound to 5′-GTA-3′ (Figure 9(d)) does not fully explain the impressive specificity observed here.

Perhaps more surprising was that changing Ser^3 to Asp^3, which produced the anticipated recognition of middle cytosine (Figure 3(e) and (g)), produced the unexpected recognition of 3′ thymine (Figure 5(g)). Recognition of 3′ thymine has been reported frequently for helices containing Gln^{-1} (Desjarlais & Berg, 1994; Nardelli *et al.*, 1992). A clue as to why this might occur derives from the observation that changing Ser^2 to Gly^2 restores recognition to primarily 3′ adenine (Figure 5(f) and

(h)). A potential explanation is therefore that Ser^{2} can, at some frequency, hydrogen bond to the carbonyl oxygen atom of Gln^{-1}. Since the position of Ser2 is fixed by the α-helix, the side-chain of Gln^{-1} would need to move closer to the serine residue, positioning its amino group more towards the center of the major grove and within hydrogen bonding distance of O4 of thymine. Such an interaction can be supported by modeling (Figure 9(e)), and is roughly analogous to the well-characterized Asp^{2}/Arg^{-1} interactions in the fingers of Zif268 (Pavletich & Pabo, 1991).

In the protein TTK, Ser^{2} of finger 1 (HIS-N-FCR) binds 3′ thymine directly (Fairall *et al.*, 1993). This type of interaction could also account for our specificity observations. However, the TTK case is structurally quite different, and no such interactions are anticipated in our helices. More similar to our helices is finger 4 of the protein YY1, QST-N-LKS. In this crystal structure, the hydroxyl group of Thr^{2} actually contacts thymine on the opposite strand, reinforcing recognition of 3′ adenine (Houbaviy *et al.*, 1996). The putative Ser^{2}/3′ thymine contact was the subject of a study by Berg and co-workers who concluded that Ser^{2} does not generally contribute to specificity (Kim & Berg, 1995).

None of the preceding arguments would account for the strong 3′ thymine crossreactivity in helices QSG-D-LVR (Figure 5(h)) or QSG-N-LVR (Figure 5(l)) where Ser^{2} is not present. As Charney and co-workers (Nardelli *et al.*, 1992) proposed previously, ordered water molecules could allow Gln^{-1} to contact the DNA without reorientation of the helix (Figure 9(f)). Finally, the helices QRA-K-LER (Figure 8(d)) and QSS-K-LVR (Figure 8(f)) prefer recognition of 5′-GGC-3′, due perhaps to a dominating Lys^{3} interaction discussed above. Overall, the evidence seems to support that the side-chain of Gln^{-1} is conformationally dynamic and may support multiple interactions, similar to the lysine residues of TFIIIA (Foster *et al.*, 1997).

Lessons from GNC: the roles of unusual residues

Since the 3′ nucleotide cannot be specified by Gly^{-1}, helix GCR-E-LSR must achieve specific recognition of 5′-GCC-3′ (Figure 6(l)) in a fundamentally different way. Two unusual residues were strongly selected by this target during phage display: Cys^{1} and Arg^{2} (Segal *et al.*, 1999). We propose that Arg^{2} donates two hydrogen bonds to the guanine that is base-paired to the 3′ cytosine. This cross-strand contact, which is supported by computer modeling (Figure 9(g)), is reminiscent of the cross-strand contact made by Thr^{2} in the finger 4 of YY1 (Houbaviy *et al.*, 1996). A similar model can be proposed for DCR-D-LAR (Figure 6(k)). The significance of the conservation of Cys^{1} during phage selection is not clear.

Also unclear is the role of the strongly selected Pro^{1} in the 5′-GNC-3′-recognition motif DPG-X-LVR (Figure 6). *A priori*, it might be expected that proline might distort or kink the helix, thus allowing an unusual conformation for Asp^{-1}. However, an alignment of finger 5 from TFIIIA (Nolte *et al.*, 1998) and GLI1 (Pavletich & Pabo, 1993), both of which contain a Pro^{1} (LPS-R-LKR and DPS-S-LRK, respectively) with finger 2 of Zif268 (Pavletich & Pabo, 1991) show no dramatic perturbation of the helix (Figure 10). The 56 atoms of the backbone from positions −1 to 6 align with less than 0.5 Å RMSD, and the α-carbon atoms of the position −1 residues are nearly superimposable. Thus, the role of Pro^{1} may be in fine-tuning interactions or in interacting with other elements of the protein or DNA, rather than dramatically changing the mode of recognition.

The side-chain of Asp^{-1} is shorter than that of Gln^{-1}. The crystal structure of the recognition helix DSS-N-LTR bound to 5′-GAC-3′ (Elrod-Erickson *et al.*, 1998) shows that the N terminus of the α-helix is positioned even closer to the DNA than in the case of QSG-S-LTR. This suggested relatively simple explanations for the specificities and affinities of the helices recognizing 5′-GNC-3′ (Figure 6). For example, Pabo and co-workers observed that both carboxylate oxygen atoms of Asp^{-1} contacted N4 of 3′ cytosine, which might explain the relatively high affinity (3 nM, Segal *et al.*, 1999) of the helix DPG-N-LVR on 5′-GAC-3′ (Figure 9(h)). Also, the side-chain of His^{3} in DPG-H-LVR is relatively long, and since the position of His^{3} would be fixed by the helix the only sterically permissive orientation for His^{3} is to contact O6 of middle guanine (Figure 9(i)), which would explain why this helix does not crossreact with the middle adenine base (Figure 6(d)) (in contrast to RSD-H-LTR (Figure 2(b))). According to the modeling, accommodating the longer side-chain may require middle guanine to be positioned further from the α-helix in a way that is energetically less favorable, accounting for the low affinity of this helix (40 nM).

Lessons from GNT: methyl groups limit interactions

To explore our 5′-GNT-3′ binding helices (Figure 7), computer models were constructed based on finger 3 of protein/DNA structure number 8 in the TFIIIA NMR set (Wuttke *et al.*, 1997). Finger 3 of TFIIIA has the amino acid sequence TKA-N-MKK and binds 5′-GAT-3′. Although the methyl groups of Thr^{-1} and 3′ thymine are 5 Å apart in the corresponding crystal structure (Nolte *et al.*, 1998) they are shown to clearly interact in NMR studies (Figure 3(h)). The recognition helixes TSG-N-LVR (Figure 7(a)) and TSG-H-LVR (Figure 7(b)) were modeled to bind 5′-GAT-3′ and 5′-GGT-3′, respectively, in a manner similar to their DPG-X-LVR counterparts described above (shown for 5′-GAT-3′; Figure 9(j)). Because the helix is slightly further from the DNA, there might be less strain induced at the middle nucleotide, consistent with the affinities of these proteins being somewhat higher (Segal *et al.*, 1999). Several of the other

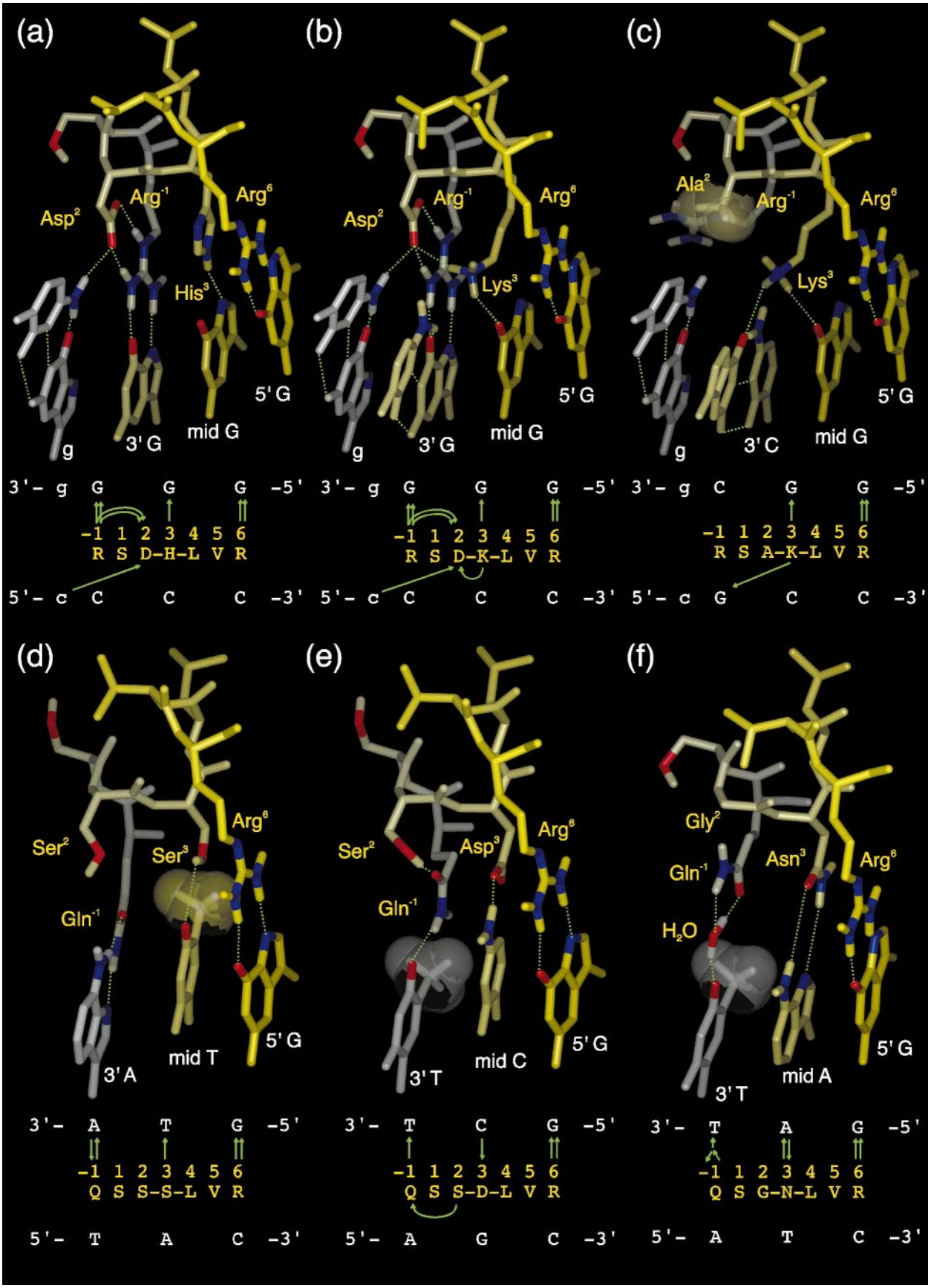

Figure 9 (*legend opposite*)

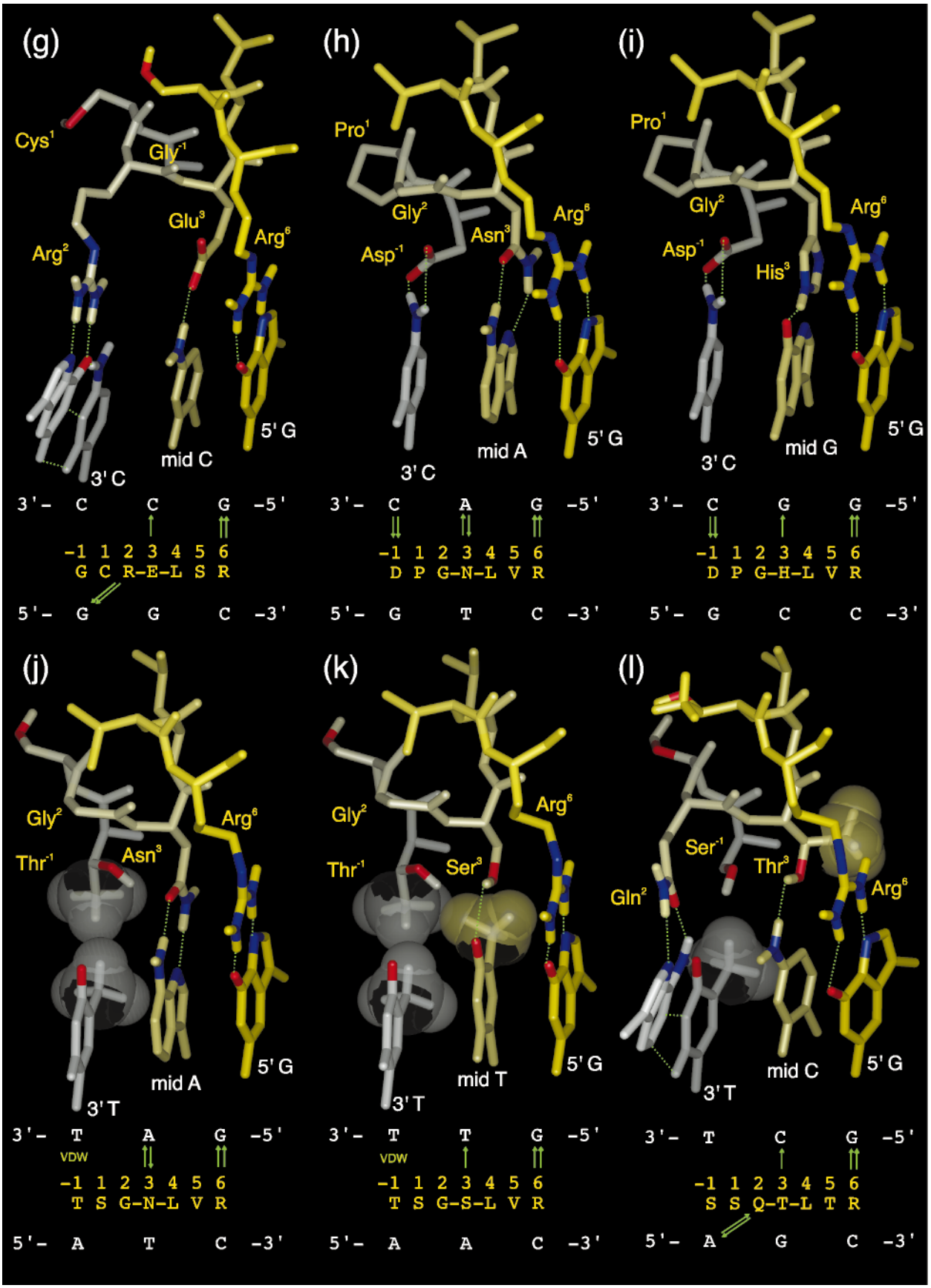

Figure 9. Computer models of finger-2 helices help to explain the molecular features of recognition. Select oxygen (red) and nitrogen (blue) atoms are colored for clarity. Broken lines (green) indicate proposed hydrogen bonds. Shells surrounding methyl groups represent the van der Waals surface of these atoms. The sequence of each helix, the DNA subsite, and the proposed interactions are summarized below each model. Green lines indicate hydrogen bonds. Arrows indicate hydrogen bond acceptors. VDW indicates van der Waals interaction.

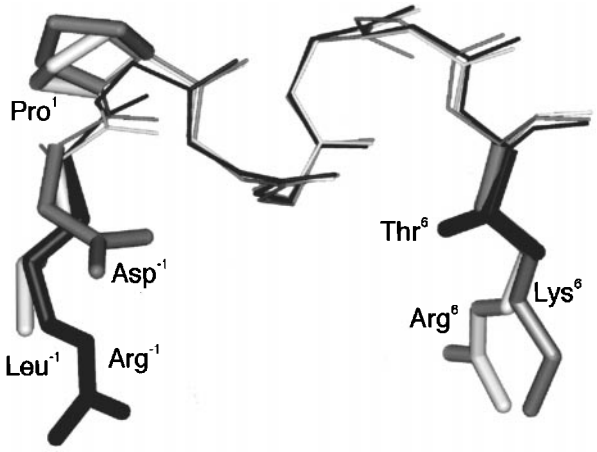

Figure 10. Alignment of helices containing Pro[1]. Shown are finger 5 of TFIIIA (white, Nolte *et al*. (1998)), finger 5 of GLI1 (gray, Pavletich & Pabo (1993)), and finger 2 of Zif268 (black, Elrod-Erickson *et al*. (1996)).

helices shown in Figure 7 show a strong preference for 5′-GCT-3′. Asp[3] of TSG-D-LVR (Figure 7(i)) may hydrogen bond with middle cytosine directly (Figure 3(e)), whereas Glu[3] of TSG-E-LVR (Figure 7(d)) more likely makes van der Waals contacts with the edges of cytosine (Figure 3(f)), as it does in fingers 1 and 3 of Zif268 (Pavletich & Pabo, 1991). This is consistent with the observed low-level affinity of 65 nM for TSG-E-LVR (Segal *et al*., 1999). Based on a model of TSG-T-LVR, the methyl groups of Thr[3] and middle thymine sterically prevent recognition of 5′-GTT-3′. However, the hydroxyl group of Thr[3] might accept a hydrogen bond from N4 of middle cytosine (Figure 7(e) and (j)). Ser[3] in the model of TSG-S-LVR (Figure 9(k)) may be less sterically hindered by middle thymine (Figure 3(g)), consistent with the observed 5′-GTT-3′ recognition (Figure 7(g)). However, structural studies show that serine can interact with any base, which might account for the poor overall specificity of this helix. Modeling suggested Ala[3] of TSG-A-LVR (Figure 7(h)) should provide the best level of specificity for recognition of 5′-GTT-3′. However, two weak 3′ and middle interactions might explain why this helix shows no 5′-GTT-3′ binding. It is not clear how this helix binds 5′-GGT-3′. Modeling also supported the potential role of Gln[2] in helices such as SSQ-T-LTR to hydrogen bond to an adenine base-paired to a 3′ thymine (Figure 9(l)), analogous to the Arg[2] cross-strand contact described earlier. However, none of the Gln[2]-containing helices were specific for 5′-GTT-3′ (Figure 7(e), (k), and (l)). Finally, it is unclear how changes in position 5 can influence specificity (consider Figures 6(a) and (b), or 7(f) and (g)). Potential explanations could include influencing the packing of the helix against the hydrophobic core (since this position often contains a hydrophobic residue) or influencing helical orientation through contacts with DNA or protein (as might occur with Lys[5] or Arg[5]). The hydroxyl group of Thr[5] in finger 2 of the Zif268 crystal structure is more than 5 Å from any part of the DNA, but is within hydrogen bonding distance of the backbone oxygen atom of the position 2 residue. The influence of position 5 seems only important in rare instances, perhaps when the overall affinity is low or when other helix-stabilizing factors are absent.

Conclusions

The modeling presented in this study suggests that the orientation of the recognition helix relative to the DNA is a critical determinant of specificity. Our data supports and extends the concept that the long side-chains of amino acid residues such as arginine, lysine and glutamine permit multiple interactions. The specificity of short chain amino acid residues such as aspartate or threonine seems largely a consequence of helical orientation. However, our ability to understand and describe the molecular aspects of recognition is limited due to the paucity of knowledge concerning the factors that govern helical orientation, sequence-specific DNA structure and the positioning of the nucleotides relative to the helix, and the positions and roles of ordered water molecules. For example, structural studies have shown that the position of the DNA bases changes in different structures. Although clearly affecting specificity, changes to local DNA structure could not be incorporated into our models due to insufficient information regarding the restrictions of base positioning. Specificity also needs to be considered in the context of affinity. Generally, affinity was correlated with the number of hydrogen bonds. Loss of affinity due to the accommodation of unfavorable interactions was proposed to explain the observed low affinity of helices recognizing 5′-GNC-3′ and 5′-GNT-3′. More work will be required to determine if the structural elements proposed by the models and the assertions that underlie them are in fact real. This information will come not only from more structural and specificity studies, but also from the formation and testing of specific hypotheses. Nonetheless, it is clear that zinc finger recognition of DNA is substantially more complex than a simple one to one amino acid to base code.

Methods and Materials

Site-directed mutagenesis

Mutants of the internal finger of a three-finger protein (finger-2 mutants) were constructed by PCR. A helix-2-specific forward primer and a standard back primer were used to amplify fingers 2 and 3 of the three-finger protein C7 (Wu *et al*., 1995) from a modified pMAL-c2 vector (New England Biolabs). The fragment was then used to replace the wild-type fingers 2 and 3 of C7 by cloning into unique *Nsi*I and *Spe*I restriction sites. The

forward primer had the general sequence, 5′-GAGGAGGAGGAGGAGATATGCATGCGTAACTTCAGT $N^{-1}N^{-1}N^{-1}$ $N^1N^1N^1$ $N^2N^2N^2$ $N^3N^3N^3$ CTT $N^5N^5N^5$ $N^6N^6N^6$ CACATCCGCACCCACACAGGC -3′, where $N^{-1}N^{-1}N^{-1}$ is the codon for the −1 position residue, etc. The reverse primer had the sequence, 5′-GTAAAACGACGGCCAGTGCCAAGC -3′.

Multi-target specificity assays

These assays were performed as described (Segal *et al.*, 1999). Essentially, freeze/thaw extracts containing the overexpressed maltose-binding protein zinc finger fusion proteins were prepared from IPTG-induced cultures using the Protein Fusion and Purification System (New England Biolabs) in zinc buffer A (ZBA; 10 mM Tris (pH 7.5), 90 mM KCl, 1 mM $MgCl_2$, 90 μM $ZnCl_2$). Streptavidin (0.2 μg) was applied to a 96-well ELISA plate, followed by the full set of 5′-GNN-3′ DNA targets (0.025 μg). Target hairpin oligonucleotides had the sequence 5′-Biotin-GGACGCN′N′N′CGCGGGTTTTCCCGCGNNNGCGTCC-3′, where NNN was the three-nucleotide finger two-target sequence and N′N′N′ its complement. The plates were blocked with ZBA/3% BSA. Eight twofold serial dilutions of the extracts were applied in 1× binding buffer (ZBA, 1% BSA, 5 mM DTT, 0.12 μg/μl sheared herring sperm DNA), followed by mAb mouse anti-maltose binding protein (Sigma) and mAb goat-anti-mouse IgG conjugated to alkaline phosphatase (Sigma). Alkaline phosphatase substrate (Sigma) was applied, and the A_{405} was quantitated with SOFTmax 2.35 (Molecular Devices). All titration data were background subtracted from ELISA wells containing extract but no oligonucleotide. Since the same initial protein concentration was applied to all targets, the dilution factor required to produce an arbitrarily chosen endpoint A_{405} value of 0.2 was used as a measure of the relative affinity of the protein for each target. Freeze/thaw lysates and purified proteins were found to produce identical results in this assay (data not shown).

Computer modeling

Computer models were generated using InsightII (Molecular Simulations, Inc). Models were based on the coordinates of the co-crystal structures of Zif268-DNA (PDB accession 1AAY), QGSR-GCAC (1A1H), DSNR-GACC (1A1F), TFIIIA-DNA (1TF3 (NMR) and 1TF6 (X-ray)), and GLI-DNA (2GLI). The structures are not energy minimized and are presented only to suggest possible interactions. Hydrogen bonds were considered plausible when the distance between the heavy atoms was 3(±0.3) Å and the angle formed by the heavy atoms and hydrogen was 120° or greater. Plausible van der Waals interactions required a distance between methyl group carbon atoms of 4(±0.3) Å.

Acknowledgments

This study was supported by The Skaggs Institute for Chemical Biology and grant CA86258 from the National Institutes of Health. B. D. was a recipient of a fellowship from the Deutsche Forschungsgemeinschaft. We also thank Roger R. Beerli and Jessica D. Flippin for their contributions to this work.

References

Beerli, R. R., Segal, D. J., Dreier, B. & 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.

Beerli, R. R., Dreier, B. & Barbas, C. F., III (2000). Positive and negative regulation of endogenous genes by designed transcription factors. *Proc. Natl Acad. Sci. USA,* **97**, 1495-1500.

Brown, R. S., Sander, C. & Argos, P. (1985). The primary structure of transcripton factor IIIA has 12 consecutive repeats. *FEBS Letters,* **186**, 271-274.

Chandrasegaran, S. & Smith, J. (1999). Chimeric restriction enzymes: what is next? *Biol. Chem.* **380**, 841-848.

Choo, Y. & Klug, A. (1994a). Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc. Natl Acad. Sci. USA,* **91**, 11168-11172.

Choo, Y. & Klug, A. (1994b). 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.

Desjarlais, J. R. & Berg, J. M. (1992). Redesigning the DNA-binding specificity of a zinc finger protein: a data base-guided approach. *Proteins: Struct. Funct. Genet.* **12**, 101-104.

Desjarlais, J. R. & Berg, J. M. (1994). Length-encoded multiplex binding site determination: application to zinc finger proteins. *Proc. Natl Acad. Sci. USA,* **91**, 11099-11103.

Elrod-Erickson, M., Rould, M. A., Nekludova, L. & Pabo, C. O. (1996). Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure,* **4**, 1171-1180.

Elrod-Erickson, M., Benson, T. E. & Pabo, C. O. (1998). High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure,* **6**, 451-464.

Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T. & Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature,* **366**, 483-487.

Foster, M. P., Wuttke, D. S., Radhakrishnan, I., Case, D. A., Gottesfeld, J. M. & Wright, P. E. (1997). Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nature Struct. Biol.* **4**, 605-608.

Houbaviy, H. B., Usheva, A., Shenk, T. & Burley, S. K. (1996). Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl Acad. Sci. USA,* **93**, 13577-13582.

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.

Isalan, M., Klug, A. & Choo, Y. (1998). Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry,* **37**, 12026-12033.

Jamieson, A. C., Kim, S.-H. & Wells, J. A. (1994). *In vitro* selection of zinc fingers with altered DNA-binding specificity. *Biochemistry,* **33**, 5689-5695.

Kim, C. A. & Berg, J. M. (1995). Serine at position 2 in the DNA recognition helix of a Cys2-His2 zinc fin-

ger peptide is not, in general, responsible for base recognition. *J. Mol. Biol.* **252**, 1-5.

Kim, C. A. & Berg, J. M. (1996). A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. *Nature Struct. Biol.* **3**, 940-945.

Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989). Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science,* **245**, 635-637.

Miller, J., McLachlan, A. D. & Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**, 1609-1614.

Narayan, V. A., Kriwacki, R. W. & Caradonna, J. P. (1997). Structures of zinc finger domains from transcription factor Sp1. Insights into sequence-specific protein-DNA recognition. *J. Biol. Chem.* **272**, 7801-7809.

Nardelli, J., Gibson, T. & Charnay, P. (1992). Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis. *Nucl. Acids Res.* **20**, 4137-4144.

Nolte, R. T., Conlin, R. M., Harrison, S. C. & Brown, R. S. (1998). Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc. Natl Acad. Sci. USA,* **95**, 2938-2943.

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.

Pavletich, N. P. & Pabo, C. O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science,* **261**, 1701-1707.

Segal, D. J. & Barbas, C. F. (2000). Design of novel sequence-specific DNA-binding proteins. *Curr. Opin. Chem. Biol.* **4**, 34-39.

Segal, D. J., Dreier, B., Beerli, R. R. & 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.

Swirnoff, A. H. & Milbrandt, J. (1995). DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Mol. Cell. Biol.* **15**, 2275-2287.

Wolfe, S. A., Greisman, H. A., Ramm, E. I. & 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.

Wu, H., Yang, W.-P. & Barbas, C. F., III (1995). Building zinc fingers by selection: toward a therapeutic application. *Proc. Natl Acad. Sci. USA,* **92**, 344-348.

Wuttke, D. S., Foster, M. P., Case, D. A., Gottesfeld, J. M. & 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.

Edited by M. Yaniv

(Received 8 May 2000; received in revised form 2 August 2000; accepted 22 August 2000)