Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 2159–2164, March 1996
Genetics

Recognition of diverse sequences by class I zinc fingers: Asymmetries and indirect effects on specificity in the interaction between CF2II and A+T-rich sequence elements

(DNA sequence recognition/asymmetric interactions/mutagenesis/binding site selection)

JOSEPH A. GOGOS*†, JINGMIN JIN*‡, HONG WAN*, MICHAEL KOKKINIDIS§, AND FOTIS C. KAFATOS*¶‖

*The Biological Laboratories, Harvard University, Cambridge, MA 02138; and §Protein Structure Group, Institute for Molecular Biology and Biophysics, Heraklion, Crete, Greece

Contributed by Fotis C. Kafatos, November 16, 1995

ABSTRACT **The *Drosophila* CF2II protein, which contains zinc fingers of the Cys_2His_2 type and recognizes an A+T-rich sequence, behaves in cell culture as an activator of a reporter chloramphenicol acetyltransferase gene. This activity depends on C-terminal but not N-terminal zinc fingers, as does *in vitro* DNA binding. By site-specific mutagenesis and binding site selection, we define the critical amino acid–base interactions. Mutations of single amino acid residues at the leading edge of the recognition helix are rarely neutral: many result in a slight change in affinity for the ideal DNA target site; some cause major loss of affinity; and others change specificity for as many as two bases in the target site. Compared to zinc fingers that recognize G+C-rich DNA, CF2II fingers appear to bind to A+T-rich DNA in a generally similar manner, but with additional flexibility and amino acid–base interactions. The results illustrate how zinc fingers may be evolving to recognize an unusually diverse set of DNA sequences.**

The Cys_2His_2 (class I) zinc finger element (1) is approximately 30 amino acids long and contains pairs of cysteines and histidines at virtually invariant positions. It is a widely prevalent and unusually flexible DNA-binding motif, able to recognize sequences that range from exclusively G+C to exclusively A+T in composition. Zinc fingers are apparently modular in DNA binding, and permutations in the order of the fingers can result in altered target gene regulation. Understanding the rules whereby individual fingers recognize DNA should permit the design of novel and specific transcriptional regulators, with important implications for drug design.

The site-directed mutagenesis studies of Nardelli *et al.* (2) on specificity determinants of krox-20 clearly showed that zinc fingers can recognize discrete triplets of DNA individually. This modularity is consistent with the cocrystal structure of three fingers of krox-24/zif268 bound to their G+C-rich target sequence (3). In the latter case, the fingers form a curved structure, with their α-helices fitting in the major groove of the 9-bp DNA-recognition element. Individual fingers correspond to adjacent DNA triplets. Two bases in each triplet are contacted by amino acids (arginines or histidines), which are spaced three or four residues apart, on the same surface of the α-helix. These residues contact specifically only one strand of DNA in an antiparallel manner (the N- to C-oriented helix contacts a G-rich DNA strand running in the 5′-to-3′ direction). Upon DNA binding, the linker between two adjacent fingers assumes a well-defined backbone structure; it does not contribute to the specific contacts, but because of its flexibility (4) probably plays an important role in establishing the orientation and compact spacing of adjacent fingers.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

We have previously reported on the *Drosophila* CF2 proteins (5, 6), which exist in developmentally regulated, alternatively spliced isoforms with up to seven zinc fingers (numbered 1–6 or F1–F6 from N to C terminus, plus F5′). The CF2II isoform interacts with its (G/A)TATAT(A/G)TA high-affinity target sequence in a manner apparently similar to that of zif268, with the adjacent F6, F5, and F4 modules recognizing the adjacent A+T-rich DNA triplets (G/A)TA, TAT, and (A/G)TA, respectively, in an antiparallel orientation and predominantly in the sense strand. F6 was shown to be essential and F1 and F2 dispensable for binding to the optimal site. The modularity of the interaction was demonstrated by the difference in binding sites selected by CF2II and another isoform, CF2I, which has the extra F5′ finger inserted between F5 and F4.

Our earlier studies (6) did not address the issues of whether CF2 can act as transcription factor *in vivo* and whether F1, F2, and F3 can contribute to binding sequences of lower affinity. Moreover, they left uncertain the exact mechanism of binding and subsite selection, as they did not attempt to identify the F4–F6 residues that are most important for binding. Extending these initial studies, we show here that CF2II can act as transcriptional activator *in vivo*. We present conclusive evidence for nonequivalent functions of the CF2II zinc fingers in DNA binding *in vivo* and *in vitro* and define the critical amino acid–base interactions in detail, by means of site-specific mutagenesis and binding site selection. The results establish CF2 as the prototype of a subfamily of zinc finger proteins with A+T-rich binding sites.

MATERIALS AND METHODS

Protein Production in Bacteria. For protein production in bacteria we used the T7 *Escherichia coli* expression system. An almost full-length fragment from CF2II cDNA clone 18 in pNB40 (7), carrying the coding region and part of the 3′ flanking sequence, was prepared by complete digestion with *Hin*dIII and *Cla* I followed by partial digestion with *Ava* I. The resulting fragment was inserted into the *Nde* I and *Cla* I sites of the pT7-7 expression vector in a three-way ligation using an oligonucleotide linker encoding the first six amino acids of the CF2II protein. Production and partial purification were as previously described (6).

Mutagenesis. For single amino acid replacements the method of Kunkel (8) was used with slight modifications on a

Abbreviations: F1–6, fingers 1–6; EMSA, electrophoretic mobility-shift assay; CAT, chloramphenicol acetyltransferase.

†Present address: Howard Hughes Medical Institute Laboratory, Fred Hutchinson Cancer Research Center, Weintraub Building, Seattle, WA 98104.
‡Present address: New England Biolabs, Beverly, MA 01915.
¶Present address: European Molecular Biology Laboratory, Heidelberg 69117, Germany.
‖To whom reprint requests should be addressed.

Pst I–*Pst* I fragment encompassing the C-terminal zinc finger domain. For the experiments on binding affinity determination, the double-stranded *Pst* I–*Pst* I fragment was subcloned after mutagenesis back to the pNB40-CF2II#24 plasmid, which carries the SP6 promoter and part of the CF2II coding region, not including F1. This truncated CF2II version was preferred to avoid the comigration of the *in vitro* translated full-length CF2II protein–DNA complex with a nonspecific complex in electrophoretic mobility-shift assays (EMSAs). For the construction of the single amino acid replacements in the first two N-terminal fingers a *Hin*dIII–*Kpn* I fragment from pNB40-CF2#18, encompassing F1 and F2, was subcloned into the mp19 vector.

Protein Production in Cell-Free Extracts. *In vitro* transcription reactions were performed according to standard protocols. For the *in vitro* translation experiments home-made or commercially available reticulocyte lysates were used according to the instructions from the supplier. [^{35}S]Methionine was supplied by Amersham.

DNA-Binding Reactions and Electrophoresis. Methods were as described (9). In the EMSA experiments, DNA binding was quantitated by phosphorimager analysis of the shifted ^{32}P-labeled probe after drying of the binding gel and blocking the ^{35}S emission with a piece of film. The protocol for selection by individual fingers was as previously described (6) and optimized (9). To avoid complications from the internal palindromic repeats (TA) of high-affinity binding sites, such as (A/G)TATAT(A/G)TA, we used disfavored sequences to flank the central nonameric target binding site (see Fig. 2 and 3 legends). The results indicated that, at the low concentrations of *in vitro* translated CF2II used in these experiments, the protein was effectively targeted to the highest-affinity available sites, in the top strand and in the desired phase only. We sequenced individual clones rather than pools to ensure that any double base substitutions were on the same molecule.

Cell Transfections and Chloramphenicol Acetyltransferase (CAT) Assays. *Drosophila melanogaster* Schneider line 2 (S2) cells were grown at 25°C in Schneider's *Drosophila* medium M3, supplemented with 10% heat-inactivated fetal calf serum. Cell transfections were carried out as described (10). For quantitation the TLC plate was cut into pieces corresponding to spots on the autoradiogram, and radioactivity was measured by liquid scintillation counting. CAT activity was calculated by the ratio of radioactivity in acetylated species to the sum of radioactivities in acetylated and nonacetylated species.

RESULTS

Asymmetric Roles of CF2 Fingers in DNA Binding and Transcriptional Activation. CF2 proteins contain zinc fingers both in the N-terminal and in the C-terminal domains (5). In its C-terminal domain isoform CF2II has four zinc fingers, F3, F4, F5, and F6. The latter three fingers are tightly clustered, being separated from each other by canonical short linkers (TGEKPF/YX, very similar to those found in zif268); in contrast, a linker of noncanonical length separates F3 from F4. The N-terminal region has two additional, dispersed, fingers (F1 and F2). Our previous studies (6) indicated that the high-affinity GTATATATA target is recognized *in vitro* mainly through F6, F5, and F4. Here we used bacterially produced wild-type CF2II and a mutant protein lacking F6 to address the question of whether, in the absence of F6, CF2II might be capable of binding to DNA sequences other than the known high-affinity site. Pooled products of the third cycle of binding site selection (6), more than 50% of which are not high-affinity sites (data not shown), were recognized by the wild type but not this mutant CF2II (Fig. 1*A*). Similarly, the mutant did not interact with the sequence TATATAN$_6$, corresponding to the F5 (TAT) and F4 (ATA) subsites followed by six degenerate positions (data not shown). These results

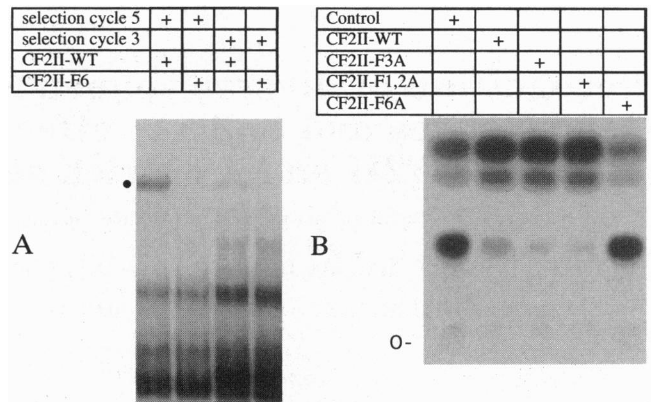

FIG. 1. CF2II F6 is essential and F1, F2, and F3 are dispensable. (*A*) EMSA using wild-type and mutant bacterially expressed CF2II together with pooled products of the fifth and third cycles of CF2II binding selection (6). A single CF2-specific retarded band (dot) was identified by its absence when bacterial extracts were immunodepleted by CF2-specific antibody or prepared from a strain missing the CF2 plasmid (data not shown). This band was observed with wild-type (WT) CF2II but not with mutant CF2II truncated by a stop codon in place of codon Q/s3F6 (CF2II-F6; see text for nomenclature). (*B*) CAT assays of CF2-driven transcription. O, origin. The reporter gene carried a pentamer of ATTA**GTATATATA**GGTC attached to a minimal *Adh* promoter and the CAT gene. The F3A, F1,2A, and F6A mutations have disrupting alanine replacements of the invariant H/h1 residue in the indicated fingers. The F6A protein was also unable to interact with DNA *in vitro* (not shown). Transient cotransfection assays in the *Drosophila* S2 cell line, which expresses endogenous CF2 (as shown by Western blots), were carried out with the expression vector alone as control, or with vectors bearing wild-type or mutant CF2II-coding region. CAT expression is comparably enhanced with wild-type F3A and F1,2A mutant proteins but is reduced to background level when F6 is mutated.

would not be expected if F3, F2, or F1 were able to bind to nearby sequences, in concert with F5 and F4.

It is possible, however, that F1–F3 might contribute to *in vivo* binding or transcriptional regulation in ways not recognizable by *in vitro* binding assays. This possibility was examined by cotransfection experiments in S2 cells, with a multimerized GTATATATA target sequence cloned upstream of a minimal *Adh* promoter and the reporter CAT gene (10). The effector plasmids for these experiments (added at low concentration, 1–2 μg per plate) encoded wild-type CF2II or mutant forms of CF2II which had fingers 3, 1+2, or 6 disrupted by histidine → alanine replacements. Fig. 1*B* shows that under these conditions CF2II can stimulate gene expression, as monitored by CAT activity. The F6 disruption reduces reporter gene expression to the level observed in the absence of any effector plasmid, whereas mutants with disrupted N-terminal fingers are at least as active as wild-type CF2II. Similar results were obtained with the same wild-type and mutant CF2II forms, using monomeric and multimeric high-affinity sites embedded in potential *in vivo* regulatory sequences, such as the s15 chorion gene promoter (11) or two different DNA fragments isolated with CF2 antibody by using a chromatin immunopurification protocol (J.A.G., data not shown). Taken together, these experiments demonstrate that the N-terminal fingers, which have some noncanonical structural features (5, 6), are dispensable for *in vitro* binding and *in vivo* function in cells.

Effects of Alanine Scanning Mutagenesis on Affinity for the Wild-Type Target Site. To determine which residues are involved in binding, we replaced individually with alanine the amino acids at six different positions of F4, F5, and F6 (Fig. 2). Mutated plasmids were transcribed and translated *in vitro*, and the yield was monitored by SDS/PAGE of the ^{35}S-labeled proteins, followed by quantitation in a phosphorimager. Equal amounts of labeled protein were then assayed in EMSA

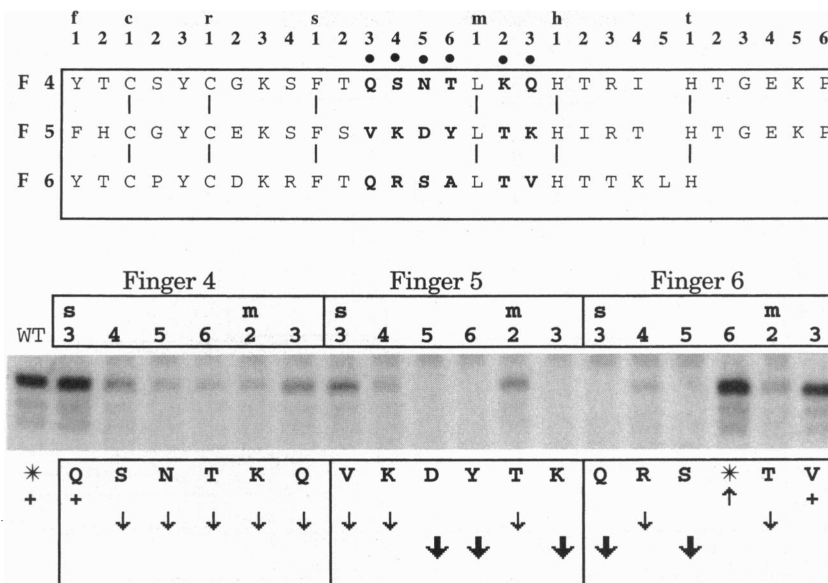

FIG. 2. Mutational analysis of the recognition helix of the CF2II DNA-binding fingers 4, 5, and 6. (*Upper*) Amino acid sequences of these fingers and the intervening linker sequences. Positions are identified according to Jacobs (12); α-helix starts at amino acid s4 and ends at amino acid t2. The s3–m3 residues (boldface letters and dots) were replaced with alanine, except for the essential m1 leucine. (*Lower*) Binding signals from an EMSA experiment using alanine-substituted mutants and ^{32}P-end-labeled double-stranded oligonucleotide including the sequence ATTAG**TATATATA**GGTC. A secondary quantification of three independent experiments shows the amino acids that were replaced by alanine and the resulting changes in binding affinity, estimated as described in the text (+, wild type; ↑, increased and ↓, decreased affinity by ≈1.5- to 3-fold; and **↓**, affinity decreased by at least 15- to 20-fold). The A/s6F6 residue was replaced by serine, resulting in a reproducible increase in the DNA-binding affinity by ≈1.5-fold. Note that drastic reductions in affinity resulted only from some replacements at s3 (helix position −1), s5 (position 2), s6 (position 3), or m3 (position 6), the same positions that make critical contacts in zif268 (3).

experiments for binding to a ^{32}P-end-labeled GTATATATA-containing oligonucleotide, and relative affinities of mutant and wild-type CF2II were estimated from the percentage of retarded oligonucleotide. Amino acid replacements at different positions showed different effects on relative affinity for the wild-type target site (Fig. 2). Results were consistent in three independent experiments.

In choosing the residues to mutate, we made the simplifying assumption that the same general region of the fingers is used for binding of CF2II to A+T-rich DNA as in the well-understood interactions of zif268 and krox-20 fingers with G+C-rich sequences (3, 13). We adopted the convention of Jacobs (12) to refer to positions within the zinc fingers (Fig. 2). In this nomenclature, *s* is a mnemonic for shoulder and *m* for main helix. In the case of zif268, critical contacts are made by certain residues just preceding or within the α-helical region, namely s3, s6, and m3; the s5 position can also play an important role. Thus, we chose to mutate all residues from s3 to m3 in the target CF2II fingers, with the exception of the highly conserved m1 leucine, which stabilizes the framework of the zinc finger (14). For convenience, we refer to these residues by the amino acid symbol followed by their position and finger—e.g., Q/s3F4. The upper part of Fig. 2 shows the sequences of the pertinent CF2II fingers (F4, F5, and F6) and highlights the mutated residues (dotted sites). We chose to replace these residues with alanine to minimize disruption of α-helical structures, since alanine is the most common amino acid in protein α-helices (15). Replacements with alanines should inhibit hydrogen bond or other interactions of DNA with charged and polar amino acid side chains but not interactions with the peptide backbone.

The results shown in Fig. 2 lead to several clear conclusions. First, dramatic (>20-fold) reduction in binding affinities results from some (but not all) of the mutations at four of the six tested positions, the same ones that are most important for zif268 binding to DNA (s3, s5, s6, and m3); this suggests that the A+T-binding fingers of CF2 take on a fold and interact with DNA in a manner similar to that of the G+C-binding fingers of zif268. Second, the critical replacements involve a variety of amino acids: aspartate, tyrosine, lysine, glutamine, and serine, rather than arginine, aspartate and histidine as in the case of zif268. Third, most other alanine replacements in this segment result in a minor but reproducible decrease in affinity (1.5- to 3-fold). Fourth, the three zinc fingers of the C-terminal domain are themselves asymmetrical with respect to their contribution to the overall binding affinity: none of the six tested alanine replacements in F4 resulted in drastic decrease in affinity, even in the case of residues (glutamine) and positions (s3, m3) that are important in the other two fingers. Indeed, even a double alanine replacement, for the glutamines at s3F4 and m3F4, decreased affinity only moderately (4.7-fold, as compared with 1.1-fold increase and 1.9-fold decrease resulting from individual replacements, respectively; data not shown). These results are consistent with the observation that in binding site selection experiments the F4 target triplet is relatively less constrained (ref. 6, and see also below). Fifth, at least minor increases in affinity are possible as a result of mutation: the A → S/s6F6 replacement increased affinity for the optimal site 1.5- to 2-fold.

In addition, we performed a limited number of replacements with amino acids other than alanine. In the case of the s6F6 residue, which is alanine (A/s6F6), we used a comparable but polar residue, serine, as well as bulkier polar residues (aspartate, asparagine) to disrupt potential interactions of alanine with the methyl group of thymine (6). While the A → S/s6F6 replacement slightly increased affinity (1.5- to 2-fold), the A → N/s6F6 and A → D/s6F6 mutations abolished binding to the optimal site. In the case of the m3 position of the same finger, neither V → A/m3F6 nor a more drastic mutation, V → K/m3F6, resulted in detectable change in affinity (data not shown).

Effects of Mutagenesis on DNA Sequence Specificity. To examine the possibility that some mutations in critical residues might alter specificity, permitting binding of the finger to novel DNA sequences, we performed three cycles of PCR-aided binding site selection (6, 9). This experiment used pools of

	POSITION m3 s6 s5 s3	∴	ATC	ATG	GTG	GCA	ACA	TTA	CTA	ATA	GTA	TAA	CAA	GAA	TAC	TAT	CAT	CCT	GCT	GGT	TGT	OTHER	#
			1	1			2	3	1		1	1	1	1		1			2		1	18	34
F6	V A S Q (boxed)	+		*1*	*4*	*2*	*3*	*1*	*12*	*22*	*35*												*80*
	V A S Q (boxed)	+					1	5		9	3												18
	K̲ A S Q	~				1	1	4		10	2												18
	V A A̲ Q	♦					1	4	2	7	3											3	20
	V A S A̲	♦				2		1	2	5	2						1					3	16
	V N̲ S Q	♦									1	14	1	1									17
F5	K Y D V (boxed)	+													*3*	*65*	*5*				*1*	*5*	*80*
	K Y D V (boxed)	+													1	10	9						20
	K A̲ D V	♦													1	9	9						19
	A̲ Y D V	♦															10	9					19
	K Y A̲ V	♦													1		1		5	11	1		19
F4	Q T N Q (boxed)	+	*3*	*9*	*9*	*2*	*2*		*1*	*21*	*20*	*1*							*1*			*11*	*80*

FIG. 3. Binding site selection analysis of CF2II mutant in key amino acids of the recognition helix. Selection was performed as described (6, 9) but was targeted to low-degeneracy double-stranded oligonucleotides, with the top strand primer sequences TGTAAGCTTCCCGGGAATTC and GGATCCTCGAGCGGCCGCTT flanking the 5′ and 3′ ends of the target sequences NNNTATATA (F6 experiment) or GTANNNATA (F5 experiment). In the first column, the amino acid residues in the critical positions are listed in C-terminal to N-terminal orientation, with mutations underlined: the wild-type version is boxed. In the second column (∴) the relative affinities for the GTATATATA site are indicated (+, wild type; ~, approximately equal to wild type; ♦, more than 15- to 20-fold decrease). Subsites selected by binding (listed at the top) are tallied for each wild-type or mutant finger. Control sequences of 34 clones from the starting pools of the F6 and F5 degenerate oligonucleotides showed the triplets in essentially random frequencies, except for an ≈2-fold under-representation of G caused by instability of this nucleotide in the precursor mixture used for oligonucleotide synthesis (data not shown). For comparison, the triplets selected by each wild-type finger from a high-degeneracy oligonucleotide pool under different conditions (figure 1*A* in ref. 6) are shown in italics. Boxes enclose the sets of related triplets that are selected by the various finger versions, and they highlight the shifts in specificity observed with VNSQ/F6, AYDV/F5, and KYAV/F5. A few unrelated sequences observed with the low-affinity F6 mutants VAAQ and VASA could represent slight broadening of the specificity but are probably due to smear contaminants caused by a low signal-to-noise ratio. The last column (#) shows the total number of sequenced clones in each experiment.

low-degeneracy oligonucleotides containing the sequences NNNTATATA and GTANNNATA, for selection by wild-type or mutant F6 and F5, respectively, and was performed with seven selected mutants and the wild-type CF2II protein (Fig. 3). In a convenient shorthand, the proteins are identified in Fig. 3 by a four-letter code, representing the residues at the four critical sites as they face the DNA triplet in antiparallel, C- to N-terminal orientation. Under the conditions of the experiment, each zinc finger selected a limited set of sequence-related triplets out of the 64 possibilities (boxes in Fig. 3; see also figure legend). The wild-type finger 5 (KYDV/F5) selected almost exclusively TAT and CAT. Of its three mutants that were used for site selection, two showed clear evidence of altered specificity. AYDV/F5 still selected CAT, but no longer TAT; instead it selected a novel triplet, CCT. Evidently, this single amino acid mutation partially shifted the spectrum of specificity for the first two nucleotide positions, (T/C)AT → C(A/C)T. The KYAV/F5 mutant changed even more completely the specificity for the same nucleotide positions, from (T/C)AT → G(G/C)T. In contrast, the KADV/F5 mutant showed a normal spectrum of specificity.

Similarly, the wild-type F6 (VASQ/F6) selected a set of related sequences, which were also observed with the KASQ/F6 mutant. The VAAQ/F6 and VASA/F6 mutants appeared to have wild-type specificity, although possibly broadened slightly (see figure legend). In contrast, VNSQ/F6 predominantly selected a novel triplet, TAA.

DISCUSSION

Structural and Functional Asymmetries of Zinc Fingers. The results presented here demonstrate functional asymmetry of the CF2II fingers at two levels. First, only the last three fingers contribute detectably to the affinity and specificity of the interaction with DNA (Fig. 1), in agreement with preliminary evidence presented elsewhere (5, 6). This asymmetry is evident both *in vitro* and in living cells, where we show that CF2II can act as a transcriptional activator without requiring the first three fingers. Potential functions for these fingers include stabilization of a DNA loop within a transcription complex (16) and interactions with RNA (17) or protein (18). Interestingly, a testis-specific alternative splicing event eliminates the DNA-binding fingers of CF2, preserving only the three N-terminal fingers (5).

Functional asymmetries were also observed in the complexes of DNA with transcription factors TFIIIA (19, 20) and GL1 (21). In the latter case, some fingers are involved in specific contacts, some make predominantly or exclusively phosphate contacts, and one does not contact the DNA at all. The structural explanation of these asymmetries is as yet unclear. In the case of CF2, the noncanonical lengths of the linkers flanking each of the first three fingers, and an unusual proline in the helices of F1 and F3 (5, 6), may suppress interaction with DNA.

Second, superimposed on this broad difference between N-terminal and C-terminal CF2II fingers, a finer-scale asymmetry is evident within the latter group. Alanine scanning mutagenesis of F4 results in only minor loss of DNA-binding affinity, in contrast to the dramatic effects of mutagenesis of certain key residues in F6 and, especially, F5. The refractoriness of F4 to alanine mutagenesis is surprising and interesting, as it cannot be explained by postulating that this finger is unimportant for specific DNA binding. F4 selects binding subsites (albeit somewhat less rigorously than F5 and F6; see (Fig. 3), and these subsites are the same whether F4 is adjacent to F5 or F5′ (figure 4 of ref. 6). Reanalysis of published data (figure 1*A* of ref. 6) shows clear preference of F4 for ATA and GTA, and to lesser extent ATG and GTG triplets (Fig. 3).

Moreover, when F4 is replaced by F5′, the selected subsite changes radically, to TAT (6). Refractoriness to alanine mutagenesis cannot be explained by reference to the atypical GL1 fingers, since F4 is a structurally typical triplet-selecting zif268-like finger. Preliminary modeling suggests the possibility that F4 might make multiple contacts: to all three nucleotide positions on one strand and to the terminal position on the other strand (data not shown). It may be that disrupting one or even two of four possible contacts (as by replacement of both F4 glutamines) still permits specific binding. Alternatively, F4 might make only nonspecific contacts with the DNA, but a certain (sequence-dependent) secondary structure of the DNA subsite might be necessary for the establishment of sufficient nonspecific interactions to stabilize binding. Results from minor groove methylation interference assays are consistent with this view (6). Further experiments are necessary to understand the mechanistic basis of these asymmetries in zinc finger binding to DNA.

Mechanisms of Recognition of Diverse Sequences by Zinc Fingers. The results from mutagenesis of CF2II fingers, together with the crystal structures of zif268 (3), GL1 (21), and Tramtrack (22), permit comparison of the mechanisms for recognition between fingers and DNA sequences that vary widely in composition. The salient conclusions can be summarized as follows.

The critical residues of the finger invariably reside in four possible positions, s3, s6, m3, and s5, suggesting a broadly similar mode of recognition by the same helical segment. The first three of these positions are adjacent on the α-helical surface, facing and potentially contacting contiguous bases in one of the DNA strands. In contrast, the s5 residue faces the second strand, where it can contact a base directly. Although an interaction between the s5 and s3 amino acid residues was postulated to explain the importance of s5 in a zif268 finger (3), such an interaction is not possible for CF2II F5, where s3 is a valine. Modeling F5 on the basis of zif268 (data not shown) indicated that the relative positions of the DNA strands and the α-helical segment permit the m3, s6, and s5 residues to make specific contacts with the DNA, forming hydrogen bonds at favorable donor–acceptor distances. The putative m3 and s6 contacts were with the top DNA strand, antiparallel to the protein, and that of s5 was with the bottom DNA strand. Modeling also suggested that direct contact with the DNA is probably responsible for the importance of s5 in F6. Pavletich and Pabo (21) and Fairall *et al.* (22) have reached similar conclusions in the case of the GL1 and Tramtrack zinc fingers, respectively. A statistical analysis (12) confirms that s5 is frequently important for zinc finger function.

The early work in this field pointed to arginines as playing a key role in zinc finger recognition (3, 13). It now is clear that in the same four critical positions a wide variety of amino acids can be used to impart different specificities. In the case of CF2 or its mutants, important residues in critical positions include glutamine, lysine, tyrosine, aspartate, asparagine, and serine.

Interestingly, our mutagenesis studies indicate that even residues at presumed non-hydrogen-bonding positions (s4, m2) contribute weakly to binding affinity. Indirect interactions via water molecules have been suggested for these positions (3, 22). In the present case, a possible clue comes from the observation that the R → A/s4F6 replacement changes slightly the mobility of the complex in EMSA experiments (data not shown): the arginine might contribute to binding affinity through nonspecific interactions that affect DNA conformation. This and other evidence that CF2II induces structural changes upon DNA binding (J.A.G., unpublished data) are consistent with the well-known flexibility of the TA dinucleotide steps (23). When uncomplexed, the ATATAT core of the CF2II site presumably adopts a B-DNA structure, as shown for the dodecamer CGACATATATTGCG (24).

Although specificity typically results from contacts between an individual amino acid and an individual base (3), indirect effects on specificity are also important: it appears that individual amino acid side chains might also contribute to the specificity of the entire finger module, for example by affecting its proximity and relative orientation to the DNA through steric or charge interactions. The importance of indirect effects is supported by the most intriguing result of our mutagenesis study: that drastic changes in specificity, for two different nucleotide positions, can result from a single amino acid change (Fig. 3). Alanine substitutions presumably leave the protein structure unaffected but can perturb specific contacts or indirect effects. For example, this may explain how K → A/m3F5 shifts the preference for the facing base, from T or C to C, even though C cannot be directly selected by alanine. Simultaneously, at the adjacent position of the DNA facing Y/s6, the pyrimidine C is selected, whereas previously only the purine A was acceptable. Part of the explanation may be that when m3 is occupied by the bulky lysine, the adjacent Y/s6 needs a bulky purine for contact, but when m3 is occupied by alanine, tyrosine can reach closer and contact a pyrimidine. In a set of related but less conclusive experiments on the Sp1 middle finger, Desjarlais and Berg (25) observed that the specificity of recognition by N/s6 is reduced by the adjacent mutation Q → R/s3. They argued that while Q/s3 and N/s6 are compatible in length, the longer R/s3 residue requires an amino acid longer than asparagine at s6 for simultaneous contact with DNA. Systematic structural studies and functional analysis using binding site selection will be necessary to define more rigorously the mechanisms of indirect effects on specificity.

The Zinc Finger Recognition Code. The results from this and other studies (2, 12, 25–30) clearly rule out a simple context-independent DNA recognition code involving an invariant "alphabetic" correspondence (29) between an amino acid and the base pair it recognizes. It may still be possible to make "syllabic" (29) predictions about the specificity imposed by a particular residue if one considers structural characteristics of the zinc finger module, such as side-chain lengths. Currently, the most realistic approach is empirical collection of specificity data from powerful genetic selection schemes (28–32). Ultimately this approach should yield specificity information for individual fingers with all acceptable combinations of amino acids at the critical positions. The usefulness of this information will be maximal if individual zinc finger domains select sequences independently of their neighbors, as originally suggested from the results of Nardelli *et al.* (13) and subsequently from the work on CF2 isoforms (5, 6). The zif268 and GL1 cocrystal structures show only minor interdomain interactions (3, 21), but a general assessment of the importance of interdomain context awaits further studies.

Typical zinc fingers are flexible, in the sense that each can recognize a limited set of related DNA sequences (ref. 29 and figure 4 in ref. 6). The tabulation of CF2 subsite selections (Fig. 3) clearly indicates that particular single-base substitutions within a triplet subsite (and sometimes double substitutions) are frequently quite acceptable for recognition by a given finger. This has important evolutionary implications, in that zinc finger recognition is stabilized against sequence drift in target cis-regulatory DNA elements. Conversely, many amino acid replacements, even in the critical positions of the zinc finger, preserve specific DNA recognition, albeit with a change in affinity. A minority of replacements impart radically new DNA specificities. A mixture of conservative and innovative mutations presumably has occurred during evolution in reduplicated zinc finger genes and modules, creating the present highly diverse and versatile family of class 1 zinc finger proteins.

We thank Jesse Bolton for technical assistance; Spyros Triantafyllakis for assistance in modeling; members of the Kafatos laboratory and Daniela Rhodes for thoughtful comments; Bianca Klumpar, Petra

Riedinger, and Alex Sossick for photography; and Susanne Lønstrup and Waltraud Ackermann for manuscript preparation. J.A.G. thanks Maria Karayiorgou for her support and useful discussions during the course of this work. This work was part of J.A.G.'s doctoral thesis (Harvard University, June 1993) and was supported by National Science Foundation and National Institutes of Health grants to F.C.K. and a Gerondelis Foundation Fellowship to J.A.G.

1. Harrison, S. C. (1991) *Nature (London)* **353,** 715–719.
2. Nardelli, J., Gibson, T. & Charnay, P. (1992) *Nucleic Acids Res.* **20,** 4137–4144.
3. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252,** 809–917.
4. Nakaseko, Y., Neuhaus, D., Klug, A. & Rhodes, D. (1992) *J. Mol. Biol.* **228,** 619–636.
5. Hsu, T., Gogos, J. A., Kirsh, S. A. & Kafatos, F. C. (1992) *Science* **257,** 1946–1950.
6. Gogos, J. A., Hsu, T., Bolton, J. & Kafatos, F. C. (1992) *Science* **257,** 1951–1955.
7. Brown, N. H. & Kafatos, F. C. (1988) *J. Mol. Biol.* **203,** 425–437.
8. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 488–492.
9. Gogos, J. A. & Kafatos, F. C. (1994) *Methods Mol. Biol.* **30,** 295–312.
10. Abel, T., Bhatt, R. & Maniatis, T. (1992) *Genes Dev.* **6,** 466–480.
11. Shea, M. J., King, D. L., Conboy, M. J., Mariani, B. D. & Kafatos, F. C. (1990) *Genes Dev.* **4,** 1128–1140.
12. Jacobs, G. H. (1992) *EMBO J.* **11,** 4507–4517.
13. Nardelli, J., Gibson, T. J., Vesque, C. & Charnay, P. (1991) *Nature (London)* **349,** 175–178.
14. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989) *Science* **245,** 635–637.
15. Richardson, J. S. & Richardson, D. C. (1988) *Science* **240,** 1648–1652.
16. Mastrangelo, I. A., Courey, A. J., Wall, J. S., Jackson, S. P. & Hough, P. V. C. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 5670–5674.
17. Joho, K. E., Darby, M. K., Crawford, E. T. & Brown, D. D .(1990) *Cell* **61,** 293–300.
18. Del Rio, S. & Setzer, D. R. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 168–172.
19. Vrana, K. E., Churchill, M. E., Tullius, T. D. & Brown, D. D. (1988) *Mol. Cell. Biol.* **8,** 1684–1696.
20. Churchill, M. E., Tullius, T. D. & Klug, A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 5528–5532.
21. Pavletich, N. P. & Pabo, C. O. (1993) *Science* **261,** 1701–1707.
22. Fairall, L., Schwabe, J. W., Campbell, L., Finch, J. T. & Rhodes, D. (1993) *Nature (London)* **366,** 483–486.
23. Travers, A. (1993) *DNA-Protein Interactions* (Chapman & Hall, London).
24. Yoon, C., Prive, G. G., Goodsell, D. S. & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6332–6336.
25. Desjarlais, J. R. & Berg, J. M. (1992) *Proteins* **12,** 101–104.
26. Desjarlais, J. R. & Berg, J. M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7345–7349.
27. Thurkal, S. K., Morrison, M. L. & Young, E. T. (1992) *Mol. Cell. Biol.* **12,** 2784–2792.
28. Choo, Y. & Klug, A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11163–11167.
29. Choo, Y. & Klug, A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11168–11172.
30. Wu, H., Yang, W. P. & Barbas, C. F. (1994) *Proc. Natl. Acad. Sci. USA* **92,** 344–348.
31. Rebar, E. J. & Pabo, C. O. (1994) *Science* **263,** 671–673.
32. Jamieson, A. C., Kim, S. H. & Wells, J. A. (1994) *Biochemistry* **33,** 5689–5695.