Sequence Discrimination by Alternatively Spliced Isoforms of a DNA Binding Zinc Finger Domain

Joseph A. Gogos, Tien Hsu, Jesse Bolton, Fotis C. Kafatos*

Two major developmentally regulated isoforms of the *Drosophila* chorion transcription factor CF2 differ by an extra zinc finger within the DNA binding domain. The preferred DNA binding sites were determined and are distinguished by an internal duplication of TAT in the site recognized by the isoform with the extra finger. The results are consistent with modular interactions between zinc fingers and trinucleotides and also suggest rules for recognition of AT-rich DNA sites by zinc finger proteins. The results show how modular finger interactions with trinucleotides can be used, in conjunction with alternative splicing, to alter the binding specificity and increase the spectrum of sites recognized by a DNA binding domain. Thus, CF2 may potentially regulate distinct sets of target genes during development.

We have investigated how two forms (CF2-I and CF2-II) of the *Drosophila* zinc finger protein CF2 (*1*) recognize DNA, apparently through the COOH-terminal domain. In this region, form II has four zinc fingers (fingers 3, 4, 5, and 6), whereas form I has an additional finger (fingers 3, 4, 5′, 5, and 6). At least in CF2-II, the NH_2-terminal region also has two additional fingers (fingers 1 and 2) but these, like finger 3, possess some unusual characteristics. These fingers are widely separated, and fingers 1 and 3 contain a proline that is likely to disrupt the α-helical structure of the recognition helices (*2*). In contrast, fingers 4 through 6 have no unusual structural features and are separated from each other by canonical short linkers (Thr-Gly-Glu-Lys-Pro-Phe/Tyr-X) similar to those found in Zif268 (*2*).

We analyzed the recognition properties of full-length CF2 isoform II by using it to select binding sites (*3*) from a pool of random sequence oligonucleotides (*4*, *5*). A modified T7 promoter-driven expression vector (*6*) was used to overproduce the protein in bacteria, and the protein was incubated with a substrate pool of double-stranded oligonucleotides, in which a random 20-bp region was flanked by defined 14-bp primers bearing restriction sites to facilitate cloning. The mixture was subjected to electrophoretic mobility shift assay (EMSA) and the oligonucleotides in the retarded complex (protein-bound fraction) were recovered, amplified by polymerase chain reaction (PCR) with the 14-nucleotide primers, and used as substrate for renewed binding. Five cycles of binding were performed with decreasing protein concentrations for the third, fourth, and fifth cycles to enhance specificity. Cloning and sequencing of the selected oligonucleotides revealed that selection progressively narrowed the spectrum of observed oligonucleotide sequences (Fig. 1, A and B). The 9-bp consensus binding sites G/ATATATA/GTA and GTATATATA were derived from the fourth and fifth cycles of selection, respectively (*7*).

This sequence preference was confirmed as follows: Unlabeled competitor encompassing the sequence GTATATATA effectively blocked CF2-II binding to several of the selected oligonucleotides (*8*). Interference analysis with dimethyl sulfate (DMS) and $KMnO_4$ (*9*) revealed contacts within the region predicted by alignment with the consensus (Fig. 2). Three lines of evidence indicated that the CF2-II recognition properties depend largely on the COOH-terminal DNA binding domain. The recognition properties of full-length CF2-II and CF2-II missing its NH_2-terminal half were comparable: When challenged with the products of the fourth and fifth cycles of selection,

A

GGG	ATATATATA	AGCGCCCC
CCGGAT	GTGTATGTA	CGGCC
GTGGG	GTATATATG	CGGTTG
TGAGCGCGG	GTATATATA	AG
GGCCAG	ATATATATG	TCCCC
CAGT	GTATATGTG	AACGGCC
AGGTGG	ATATATGTA	CGTGC
CGC	ATATATGTA	GACGACCC
ATGCG	CTATATATA	CGTGCC
ACACAAC	GTATATGTA	TGTG
AGGT	CTATATGTG	TGCCGCC
TACGCGGGGAC	CTATATGTG	
GGCAGGGTCAT	GTATATGTG	
GCGT	GTATATATA	ACTAGTG
	GTATATATA	CCACCTGTGCG
GGCCAG	ATATATATA	TCCCC
TCGCGAGAG	ATATATGTG	CC
GGGGCCAC	GTGTATATG	TGC
ACGGCGCGG	GTGTATATC	CG
ACGGGC	GTATATATT	GTGGG
C	GTATATATA	CGCGCCGTGG
GAG	CTATATGTA	GTTCGTGC
TGTCAC	GTATATATA	TTGCC
GAGAGGAC	GTATATGTG	GCC
AG	CTATATGTA	TGCTTTGCCC
GGTCTGA	GTATATGTA	GGGC
GGTCAG	CTATATGTA	TGCCC
AGC	ATATATGTA	CTGGCTGGCC
TGGTGATCA	ATATGTATC	GCC
GGATAC	ATATATGTG	CTGTC
CACC	GTATATATG	ATGGTGC
CGGTG	CTATATGTA	CTTGTC
GAGTCGTAAAC	GTATATGTG	C
	GTATATATG	TAGACCTGGCC
CATGC	ATATATGTG	TGCGC
CGTGAC	GTATCTATA	AGGTG
GGCA	GTATATATG	GTGGGCG
GCTAG	ATATATATC	TTGTGG
GGGC	GTATATGTA	TGCCGTC
GC	GTATATATG	TACTCGGCC
AAGTCCGT	CTATATGTA	TCC
GGCCGC	ATATATGTA	GCGCT
TCGGCT	GTATATATA	ACCAA
CTAC	GTATATCTG	GCACGTC
ACGTGCCTGC	ATATATATG	T
GACCGC	ATATATGTA	GCGCT
AACGGCT	GTATATATA	TCCA
TGGT	GTATATGTA	ATGGCCA
ATGACAAT	CTATATGTA	GCC
CGC	ATATATGTA	CCCCGCAT
GCAAC	ATATATATG	CCATCC
CGAAG	ATATATATA	TACCAC
GG	ATATATGTA	CCCCGTACC
GGGTCGTAC	TTATATCTA	CCC
TGGCGGAA	GTAGATGTA	TGC
CCGTGC	CTATATGCG	GGATG
AGAATTA	ATGTATAAA	AATG
ACAGGAC	GTATATGCA	TGCC
CAGCACGGT	ACATATATA	CG
GGACACGTG	GTATATACG	CC
TGGCAGT	GCATATATA	CCCC
GGGGG	GTAGATATA	TAACTG
	GTACATATA	CATATGCCGCC
GAG	GTATACATA	CCGTGCGT
AGCCA	CTACATATA	GCGCGT
	GTATATACT	CACTGTAAGCC
GGT	ATACATATA	TTGCGCGC
CAGGTAAT	GTATATACA	GCC
	GTATATGCT	TCACTGTAAGC
AAC	ATATATGCA	ACGGGGCC
GGTACA	GTATATTAA	ATTGC
CAGT	CTATACGTA	CAGCCGC
TGGTG	GTACATATA	CCCGTTC
TGAGTAG	ATATATACA	CGGC
ACCA	GTACATATA	GGACACC
GCCCATAG	ATATACACG	CCT
CGCTGTTT	ACATATAGT	GCC
AGAG	ACAGCACCG	CGGGCCC
	GCACGTCCA	GGTTCAATTGT
TC	GTGCGCTGC	CCGCAATGG

B

4th Cycle; 80 sequences

	1	2	3	4	5	6	7	8	9
	G/a	T	A	T	A	T	A/g	T	A
%G	51	0	6	4	4	0	41	2	26
%A	33	0	94	0	94	1	51	2	63
%T	1	94	0	88	0	95	5	83	5
%C	15	6	0	8	2	4	3	13	6

5th Cycle; 70 sequences

	1	2	3	4	5	6	7	8	9
	G	T	A	T	A	T	A	T	A
%G	67	0	6	1	3	0	26	0	16
%A	26	0	94	0	96	0	70	3	81
%T	0	93	0	90	0	96	1	90	3
%C	4	7	0	9	1	4	3	7	0

Fig. 1. Selection of DNA sequences by CF2-II protein from oligonucleotides bearing a 20-bp random region. (**A**) Eighty oligonucleotide sequences cloned after the fourth round of selection. They were aligned as follows: The protein-selected sequences were examined visually, and a $(TR)_4$ motif was detected frequently (R = A or, less frequently, G). Sequences carrying this complete motif were used in a preliminary alignment, which then served as an aid to align the remaining sequences that carried an incomplete set of TR repeats. We obtained optimal alignment by comparing both strands of each sequence (*5*). Only bases included in the degenerate region of the oligonucleotide were tabulated for statistical analysis. Seventy binding-site sequences from the fifth cycle were aligned similarly. (**B**) Analysis of nucleotide usage at each position for the fourth and fifth round sequences; the consensus is also presented.

J. A. Gogos, T. Hsu, J. Bolton, Department of Cellular and Developmental Biology, Harvard University, Cambridge, MA 02138.
F. C. Kafatos, Department of Cellular and Developmental Biology, Harvard University, Cambridge, MA 02138, and also the Institute of Molecular Biology and Biotechnology, FORTH, and Department of Biology, University of Crete, Heraklion 711 10, Crete, Greece.
*To whom correspondence should be addressed.

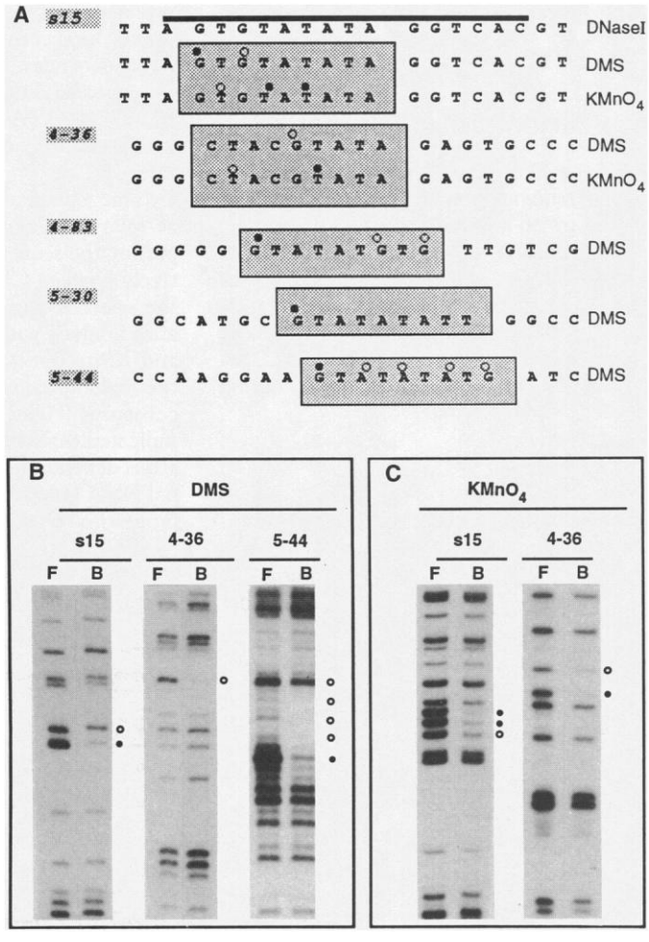

Fig. 2. (**A**) Summary of DNase I footprints and interference patterns by CF2-II on the *s15* promoter binding site, and on typical affinity-selected oligonucleotides. Shaded boxes indicate the binding sites predicted by sequence alignment. Filled and open circles indicate the observed strong and weak interference, respectively, which lie within the predicted site. (**B**) Typical DMS methylation interference patterns in the protein-interacting strand. F, free DNA; B, protein-bound DNA. As in (A), filled and open circles indicate bases whose modification interferes with binding. The labeling was either at the 3′ end (Klenow filling reaction, site 4-36), or at the 5′ end (labeling with T4 polynucleotide kinase, sites *s15* and 5-44). In the overexposed 5-44 pattern, note the interference of modified A's (open circle) and even T's (unmarked) within the binding site. Interference analysis was performed as in (*26*) with the exception that free and bound DNA was transferred to Hybond M-G paper (Amersham), which was processed as described previously (*16*).

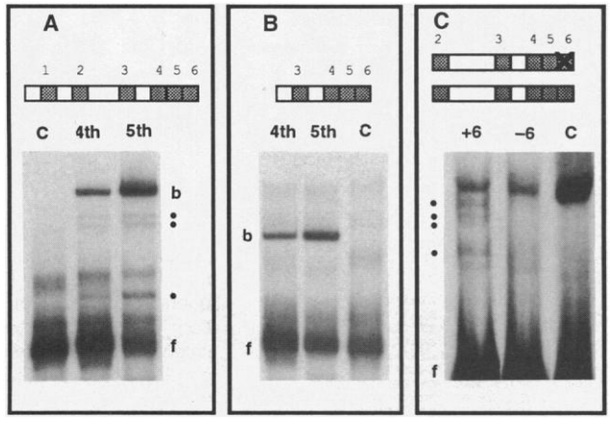

Fig. 3. (**A**) EMSA in which bacterially produced full length CF2-II protein was mixed with the pooled oligonucleotides selected in the fourth and fifth cycles. A negative binding reaction containing only control bacterial extract and the fifth round oligonucleotide pool is indicated by C. The major complex is **b**, free DNA is **f**; dots indicate other retarded complexes, ascribed to CF2 proteolytic products. (**B**) EMSA with a bacterially produced partial CF2-II protein, lacking the NH_2-terminal zinc fingers. (**C**) EMSA with an oligonucleotide encompassing the optimum binding site GTATATATA and reticulocyte cell-free translation products of CF2-II RNA encoding all four fingers of the COOH-terminal DNA binding domain (+6) or missing finger 6 (−6). A control binding reaction containing only unprogrammed reticulocyte extract and the same oligonucleotide is indicated by C. The most intense retarded complex is an endogenous artifact of the reticulocyte extract. Multiple retarded complexes, observed in the presence but not the absence of finger 6 (dots) are CF2 specific since: (i) they are competed by oligonucleotides bearing the CF2 site but not variants bearing mutations in the CF2 site, (ii) they are competed by preincubation of the extract with CF2 antibody, and (iii) similar complexes are formed on fragments of the *s15* chorion gene promoter only if they include the CF2 site. The multiplicity of these specific complexes may be explained by internal initiation of translation.

the protein lacking the NH_2-terminal domain bound to them with progressively increasing affinity (Fig. 3B), as did the full-length protein (Fig. 3A). In contrast, protein lacking the last two COOH-terminal fingers did not detectably interact with these affinity-selected products. Finally, full-length and partial CF2-II yielded similar DNase I (deoxyribonuclease) footprints on the chorion *s15* promoter (*8*, *10*).

According to the crystallographic model of Zif268 complexed with DNA (*2*), each finger recognizes an adjacent DNA triplet on the same strand and interacts with it through one to three amino acid residues at defined positions, between the conserved phenylalanine and histidine of the finger α-helix. The protein/DNA alignment is antiparallel, that is, the first (5′) DNA triplet is recognized by the last (COOH-terminal) protein finger. If we apply this model to the four-finger COOH-terminal domain of CF2-II, it seems surprising that the consensus binding site is 9 bp rather than 12 bp. Because fingers 1 and 2 apparently have no role in sequence recognition and finger 3 has an unusual structure (*11*), we suggest that fingers 4, 5, and 6 are responsible for the interaction of partial or full-length isoform II with the consensus DNA sequence (Fig. 4A). Consistent with this hypothesis, CF2 protein missing finger 6, produced in reticulocyte cell-free extract, fails to interact with a labeled oligonucleotide containing the sequence GTATATATA (Fig. 3C). The proposed arrangement positions finger 5 opposite the middle triplet, TAT. Antiparallel binding with the strand shown requires positioning of finger 6 to the 5′ side of this middle TAT triplet, and finger 4 to the 3′ side to it. Interaction with the strand indicated was suggested by the methylation interference patterns, which showed strong interference with the first guanine of the consensus binding site, and weaker interference with guanines substituted for adenines in the same strand (Fig. 2, A and B). Furthermore, the interaction of CF2 with this strand is supported by the DNase I footprints (*12*). The footprint of CF2-II on the chorion *s15* promoter (*10*) extended asymmetrically 3′ to the GTGTATATA recognition site. This and not a 5′ extension is expected, because the DNA binding domain is at the extreme COOH-terminus: in an antiparallel arrangement the NH_2-terminal bulk of the CF2 protein should extend beyond the 3′ end of the recognition site (Fig. 4A), protecting DNA sequences in that direction.

The inference that finger 5 of CF2-II binds to the middle triplet TAT helped us to determine the DNA binding specificity of CF2-I protein. The extra finger of this latter form, finger 5′, is identical to finger 5 in the three critical residues that contact

DNA according to the Zif268 model (*2*). Assuming that finger specificity is modular, we predicted that the DNA triplet corresponding to finger 5′, like that of finger 5, is TAT. This prediction was confirmed by binding selection from low-degeneracy pools of oligonucleotides (Fig. 4B). First, oligonucleotides containing the sequence GTATATNNN flanked by PCR primer sequences were subjected to three cycles of selection and amplification with a fusion protein, in which the COOH-terminal half of CF2-I was fused to the maltose binding protein (*1*, *13*). The first six fixed nucleotides of this sequence should direct binding of fingers 6 and 5, positioning finger 5′ to select the appropriate triplet from the three degenerate positions. Twenty amplified products of the third selection cycle were cloned and sequenced. In 13, the degenerate positions perfectly matched the TAT expected and the degree of selection at individual positions was higher: the frequency of T in the first position, A in the second, and T in the third were 90, 95, and 70%, respectively (*14*). We further predicted that the cognate triplet for finger 4 should be ATA in form I, just as in form II. This was confirmed by three cycles of selection and amplification with a low-degeneracy oligonucleotide that contained the sequence GTATATTATNNN flanked by sequences of PCR primers. In this case the first nine fixed nucleotides should direct binding of fingers 6, 5, and 5′, positioning finger 4 to select the appropriate triplet from the three degenerate positions. In 20 selected sequences, the base preferences were 70% A, 70% T, and 55% A for the degenerate positions 1, 2, and 3, respectively. Thus, as predicted, the recognition consensus of CF2-I is GTA.TAT.TAT.ATA (dots indicate that individual zinc fingers recognize each nucleotide triplet). However, in comparison with the previous experiment, the degree of selection was clearly lower: only 4 of the 20 sequenced clones contained the optimum subsite sequence, ATA, and the majority were 1- or 2-bp mismatched versions of it. These results suggest that three zinc fingers may provide sufficient affinity and specificity for DNA binding. An intriguing consequence is that a DNA binding domain with four fingers has a degree of redundancy and may be able to recognize disparate sequences by using overlapping sets of three contiguous fingers at a time. Indeed, by means of the two overlapping sets of three fingers each (4, 5, and 5′ or 5, 5′, and 6), CF2-I is able to recognize two seemingly distinct motifs (GTATATTAT and TATTATATA) with reasonable affinities; the full GTATATTATATA sequence that combines these two motifs is recognized most strongly (Fig. 4C).

The distinct binding properties of isoforms I and II are reflected in their recognition of different promoter sequences (*1*). A site in the proximal 5′ flanking DNA of the *CF2* gene is strongly recognized by isoform I but not isoform II. The converse is true for a *cis*-regulatory site of the *s15* chorion promoter. The prediction that these two isoforms regulate distinct promoter elements in vivo can be tested experimentally by genetics and cotransfection experiments in tissue culture cells.

The contribution of each DNA position to specificity is thought to be related to the degree of sequence constraint at this position (*15*). Although the 9-bp recognition site of CF2-II is largely symmetrical, an asymmetry is apparent in terms of the specificity of interactions. DNA positions 2 to 6 and 8 appear to be the most constrained, and the triplet corresponding to finger 5 (positions 4 to 6) seems to be the most important for recognition (Fig. 1B). The interactions of fingers 4 and 6 with DNA appear less constrained, providing some degree of permissiveness in binding and potentially increasing the spectrum of CF2 regulated sites (*16*).

As a first approach to explore the important amino acid–base interactions behind CF2-DNA recognition, we aligned the 9-bp consensus with the amino acid sequences of the CF2-II fingers, using the rules derived for two other zinc finger proteins, Zif268 (*2*) and Krox-20 (*17*). These rules are relatively simple: of the three amino acid residues in each finger that are critical for recognition, the most COOH-terminal may contact the first base in the interacting strand of the respective subsite (5′ X--), the

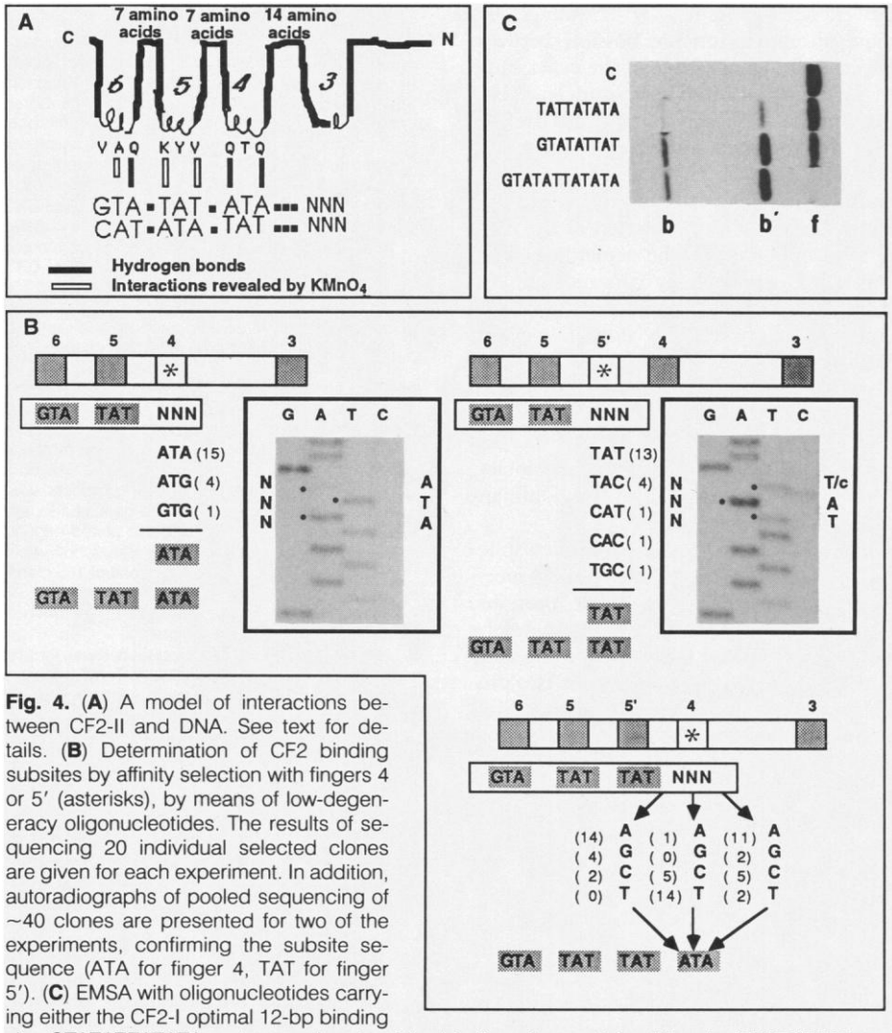

Fig. 4. (**A**) A model of interactions between CF2-II and DNA. See text for details. (**B**) Determination of CF2 binding subsites by affinity selection with fingers 4 or 5′ (asterisks), by means of low-degeneracy oligonucleotides. The results of sequencing 20 individual selected clones are given for each experiment. In addition, autoradiographs of pooled sequencing of ~40 clones are presented for two of the experiments, confirming the subsite sequence (ATA for finger 4, TAT for finger 5′). (**C**) EMSA with oligonucleotides carrying either the CF2-I optimal 12-bp binding site, GTATATTATATA, or two overlapping 9-bp binding sites contained therein (GTATATTAT and TATTATATA). A bacterially produced COOH-terminal half of the CF2-I protein, fused with the maltose binding protein (*1*), was used in this experiment. Note that CF2-I binds to all three oligonucleotides, although most strongly to the full 12-bp site. Lane C is a control reaction containing maltose binding protein only. Free and bound DNA are indicated by **f** and **b**, respectively; **b′** is a binding complex which is presumably caused by proteolysis of CF2-I and behaves similarly to **b**.

middle residue may contact the second base (5′-Y-), and the most NH_2-terminal one may contact the third base (5′--Z). Assuming these rules, we find that two patterns emerged from the alignment (Fig. 4A). First, three out of four adenines interact with glutamines, interactions similar to those described by Seeman *et al.* (*18*). Glutamine residues may be important in zinc fingers for optimal recognition of adenine residues in AT-rich DNA sites. In a similar manner, arginine residues in zinc fingers recognize the guanines of GC-rich DNA sites (*2*, *17*). In addition, because glutamine is both a donor and an acceptor of hydrogen bonds, the constrained permissiveness observed, at positions 7 and 9 of the finger 4 binding subsite (*19*) can be explained. The second observed pattern is correspondence of half the thymines with nonpolar residues (alanine and valine, for the thymines at positions 2 and 6, respectively). Hydrophobic interactions could explain the high degree of constraint observed at these positions. In agreement with this possibility, when we examined $KMnO_4$ interference with binding we detected strong interference for position 6 and weak interference for 2 (Fig. 2, A and C). $KMnO_4$ oxidizes thymine to a 5,6 *cis*-diol, thus adding a disruptive hydroxyl group adjacent to the methyl group that is thought to be involved in hydrophobic interactions; however, other interactions might also be disrupted by $KMnO_4$ (9). Oxidation of the thymine at position 4 also interfered strongly with binding; in this case the cognate amino acid was lysine. In contrast, only a slight $KMnO_4$ interference was detectable at position 8, where thymine is matched with threonine.

Even the least constrained positions of the CF2-II recognition sequence are not indifferent (Fig. 1B). Moreover, all nine positions display clear secondary preferences, which are transitions relative to the primary preference. This suggests a strong bias for alternating purines and pyrimidines in the binding site, perhaps because of conformational reasons. Consistent with this interpretation, long exposures of autoradiograms from DMS experiments indicate that methylation of adenines (at N3 facing the minor groove) interferes with CF2 binding (sequence 5-44, Fig. 2, A and B); a similar conformational interference has been described for the Eco RI–DNA complex (*20*). The crystal structure of the dodecamer CGCATATATGCG, which is related to the CF2-II binding site, is an alternating B helix that is most likely stabilized by improved stacking of bases potentially sensitive to methylation in the minor groove (*21*).

Our binding studies with oligonucleotides of high and low degeneracy have led to a detailed and consistent model for the interactions of the zinc fingers of the two major CF2 isoforms with AT-rich DNA. This model is of general interest because it predicts the modularity of finger-DNA recognition and interactions other than the known hydrogen bonding between arginine or histidine and the guanines.

Two previous attempts using binding-site selection with zinc finger motifs from Sp1 and the Wilms tumor (WT) protein have been largely inconclusive, either because of large protein input, too few selection-amplification cycles, or a small number of sequenced samples; as a result, low affinity sites have been selected and a reliable consensus has not been derived (*22*, *23*).

In the *WT* gene a normal splicing variant inserts 9 bp encoding three amino acids between the third and fourth zinc fingers, thus increasing the length of the linker. This alteration impairs the DNA binding of the encoded protein (*23*). However, because no correlation was possible between fingers and cognate triplets, the exact cause of the impaired binding was unclear. Furthermore, a dominant mutation of the *WT* locus has been identified, consisting of a deletion of 25 bp that includes an exon-intron splice junction and results in a protein with an in-frame deletion of the third of four zinc fingers (the remaining three fingers are separated by canonical linkers and show no unusual characteristics). This mutant protein exhibits only low DNA binding affinity for the consensus oligonucleotide that is recognized by the wild-type protein (*23*). Our results predict that the WT protein with a deleted finger has acquired an altered DNA binding specificity, which might be involved in the dominant mutant phenotype.

One additional case of alternative splicing that changes the DNA binding specificity involves helix-loop-helix proteins. The E47 and E12 proteins are encoded by alternatively spliced mRNAs from the same gene. The spliced mRNAs encode two proteins with differences in the basic and helix-loop-helix regions, and different DNA binding specificities (*24*).

It is now widely appreciated that transcription factors are modular in terms of the arrangement of their activation, DNA binding, and ligand binding domains. Alternative splicing of an activator domain can produce activators and repressors with identical DNA recognition, permitting finely tuned regulation of a defined set of genes (*25*). Conversely, as documented in this and the accompanying report (*1*), alternative splicing within the DNA binding domain might represent a second important level of modularity, producing proteins with related but different specificities, thus increasing the spectrum of genes that are controlled by a single transcriptional regulatory gene.

REFERENCES AND NOTES

1. T. Hsu, J. A. Gogos, S. Kirsh, F. C. Kafatos, *Science* **257**, 1946 (1992).
2. N. P. Pavletich and C. O. Pabo, *ibid.* **252**, 809 (1991).
3. R. Pollock and R. Treisman, *Nucleic Acids Res.* **18**, 6197 (1990); T. K. Blackwell and H. Weintraub, *Science* **250**, 1104 (1990); J. A. Gogos, G. Tzertzinis, F. C. Kafatos, *Nucleic Acids Res.* **19**, 1449 (1991); H. Ueda and S. Hirose, *ibid.*, p. 3689.
4. The degenerate oligonucleotide was annealed to a tenfold molar excess of primer 1, AAGCGGCCGCTCGAGGATCC, a complementary strand was synthesized with Taq DNA polymerase, and the double-stranded product was purified on a 12% polyacrylamide gel. After end-labeling to high specific activity with polynucleotide kinase, protein binding was performed in 20-μl reactions of 4 mM tris-HCl (pH 7.5), 80 mM NaCl, 0.5 mM $ZnSO_4$, 5% glycerol, 0.5 mM DTT (dithiothreitol), 1 mM EDTA, and poly(dI-dC) at 0.15 μg/μl at 23°C for 45 min. Approximately 200 ng of CF2 protein (greater than 50% pure; ~70 ng for cycle 4 and ~20 ng for cycle 5) and 1 μg of the random-sequence oligonucleotide (~100 ng for cycles 2 to 5) were used for binding; bacterial extract was used as a negative control in parallel. Reactions were loaded on a 0.25× TBE [22.5 mM tris-borate (pH 8.3), 0.5 mM EDTA], nondenaturing 6% polyacrylamide gel containing 5% glycerol, which was pre-run and run at 4°C. To isolate selected sequences, we excised a slice approximately 3 to 5 mm wide from the dried gel, including the 3MM (Whatman) paper backing. Slices were incubated at 37°C for 3 hours in 0.5 ml of 0.5 M ammonium acetate, 10 mM $MgCl_2$, 1 mM EDTA, and 0.1% SDS. The eluate was filtered, precipitated with ethanol after addition of 5 μg of glycogen carrier, adjusted to 0.3 M sodium acetate, and reprecipitated with ethanol. Approximately one-fifth of the sample was amplified for 23 cycles in a 100-μl PCR reaction using primers 1 and 2 (primer 2: TGTAAGCTTCCCGGGAATTC), after optimization of the Mg^{2+} concentration. All experiments included a control PCR reaction without template, which did not yield a product. Care was taken to avoid cross-contamination. The products were purified on 12% polyacrylamide gels, eluted, and purified as above. Selected and eluted oligonucleotides were cloned into a Bluescript vector (Stratagene). We sequenced a total of 165 clones with the T_3 or T_7 primer and conventional or ITP double strand dideoxy ^{35}S sequencing (from both strands in case of ambiguities).
5. Alignment of the selected sequences was facilitated by the detailed imprint **G**ta/g.Ta/g**T**.a/gta/g of the CF2-II consensus sequence interaction, detected by $KMnO_4$ and DMS interference [bold letters indicate strong (capital) or weak (lowercase) interference]. Two types of phasing ambiguities were encountered with some sequences. When alternative alignments with the same number of mismatches relative to the consensus were possible, we chose the one conforming best to the secondary preferences that were established by the unambiguously aligned sequences. In agreement with this rule, the observed imprint of sequence 4-36 was gggctac**g**T**a**tagagtgccc, as we expected, rather than gggctac**GtaTa**gagtgccc, as would be predicted by the less favored alternative alignment. In another set of clones, the optimal alignment was unambiguous, but it was obvious that the protein could also interact in a different phase of the sequence; for simplicity, only the optimal alignment was tabulated, although both interactions apparently occurred (the occurrence of both interactions is suggested by the observed imprint of sequence 5-64, acc**gtaTaTaTg**taxgtgcc, which is best explained as an overlay of the two imprints from different binding phases).

6. The modified pT7-7 expression plasmid for the production of full-length CF2-II, carries a 2-bp deletion in the region between the initiation codon and the Shine-Dalgarno sequence, reducing but not eliminating the level of translational initiation. The deletion appears to prevent a basal level of CF2 transcription in the absence of T7 RNA polymerase, which is deleterious and prevents cloning of full-length CF2 in the wild-type vector. The expressed protein formed inclusion bodies that we purified according to F. A. Marston [in *DNA Cloning: A Practical Approach*, D. M. Glover, Ed. (IRL Press, Oxford, 1987), pp. 65 and 80], with slight modifications.
Of the selected sequences, 15 (10/90 from the fourth cycle and 5/75 from the fifth cycle) bore no similarity to the consensus and did not bind CF2 significantly, as verified by gel shift assays. These were not included in the alignment; their occurrence was ascribed to either streaking of the free probe during electrophoresis or to nonspecific binding.
8. J. A. Gogos, J. Bolton, T. Hsu, F. C. Kafatos, unpublished results.
9. M. Truss, G. Chalepakis, M. Beato, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7180 (1990); D. K. Lee, M. Horikoshi, R. Roeder, *Cell* **67**, 1241 (1991).
10. M. J. Shea, D. L. King, M. J. Conboy, B. D. Mariani, F. C. Kafatos, *Genes Dev.* **4**, 1128 (1990).
11. Although finger 6 is clearly important for binding, finger 3 has unusual properties suggesting that either it is structurally variant and unable to interact with DNA, or it recognizes a triplet that is noncontiguous and at a variable distance from the rest. Another possibility is that this finger is not accessible to DNA because of tertiary protein structure.
12. Asymmetric footprints extending 3′ to the consensus were also observed with the maltose binding protein–form I COOH-terminal domain fusion interacting with the CF2 promoter (*1*).
13. C. V. Maina *et al.*, *Gene* **74**, 355 (1988).
14. In a control experiment, the same low-degeneracy oligonucleotide was subjected to three cycles of selection and amplification with a partial CF2-II protein containing only the COOH-terminal zinc finger domain. Twenty amplified products of the third selection cycle were cloned and sequenced; in 15 of them the degenerate positions were ATA, with selection being 95% A in the first position, 100% T in the second, and 75% A in the third (Fig. 4B).
15. S. C. Ekker, K. E. Young, D. P. von Kessler, P. A. Beachy, *EMBO J.* **10**, 1179 (1991).
16. Repeated selection cycles, especially with decreasing protein concentration, underestimate the permissiveness of binding, because they emphasize the highest affinity sequences. The constrained permissiveness was most clearly demonstrated experimentally by means of a simplified selection protocol, with direct DNA sequencing after only one selection cycle, without PCR amplification and cloning. With this approach, we have previously reported [J. A. Gogos *et al.*, *Nucleic Acids Res.* **19**, 1449 (1991)] that an A → T transversion is highly favored at position 8 of the recognized strand, as are T → A/G transversions at position 3. Using two additional low degeneracy oligonucleotides, we further showed that, after only one cycle of selection, T is highly favored over A at position 2, whereas at positions 1 and 7 either purine is acceptable (J. A. Gogos, unpublished data).
17. J. Nardelli, T. J. Gibson, C. Vesque, P. Charnay, *Nature* **349**, 175 (1991).
18. N. C. Seeman, J. M. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 804 (1976).
19. When methylation interference experiments are performed on less-than-ideal binding sites, with guanines replacing the most favored adenine at positions 3, 7, and 9, guanine methylation interferes with binding (Fig. 2, A and B). This suggests that the cognate glutamines are in close contact with the DNA, although alternative interpretations are possible. Note also that substitution of Gln for Glu at position 18 in the second finger of Krox-20 destroys the selectivity of this position as judged by EMSA (*17*).
20. A. L. Lu, W. E. Jack, P. Modrich, *J. Biol. Chem.* **256**, 13200 (1981); J. A. McClarin, *et al.*, *Science* **234**, 1526 (1986).
21. C. Yoon, G. G. Prive, D. S. Goodsell, R. Dickerson, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6332 (1988); W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
22. H. J. Thiesen and C. Bach, *Nucleic Acids Res.* **18**, 3203 (1990).
23. F. J. Rauscher III, J. F. Morris, O. E. Tournay, D. M. Cook, T. Curran, *Science* **250**, 1259 (1990).
24. C. Murre, P. S. McCaw, D. Baltimore, *Cell* **56**, 777 (1989).
25. R. J. Koening *et al.*, *Nature* **337**, 659 (1989); C. Roman, L. Cohn, K. Calame, *Science* **254**, 94 (1991); Y. Nakabeppu and D. Nathans, *Cell* **64**, 751 (1991); N. S. Foulkes, B. Mellstrom, E. Benusiglio, P. Sassane-Corsi, *Nature* **355**, 80 (1992).
26. A. Baldwin, in *Current Protocols in Molecular Biology*, M. Ausubel *et al.*, Eds. (Wiley-Interscience, New York, 1987).
27. We thank A. Khoury for oligonucleotide synthesis and discussions, D. Thanos and M. J. Shea for reagents, T. Maniatis for a critical reading of the manuscript, S. Pedreira for secretarial assistance, and B. Klumpar for photography. Supported by a Gerondelis Foundation predoctoral fellowship (J.A.G.), an NIH postdoctoral fellowship (T.H.), and an NIH grant (F.C.K.).

March 26, 1992; accepted July 28, 1992.

Inactivation of the p34^{cdc2}–Cyclin B Complex by the Human WEE1 Tyrosine Kinase

Laura L. Parker* and Helen Piwnica-Worms*†

Entry into mitosis in *Schizosaccharomyces pombe* is negatively regulated by the *wee1*$^+$ gene, which encodes a protein kinase with serine-, theonine-, and tyrosine-phosphorylating activities. The *wee1*$^+$ kinase negatively regulates mitosis by phosphorylating p34^{cdc2} on tyrosine 15, thereby inactivating the p34^{cdc2}–cyclin B complex. The human homolog of the *wee1*$^+$ gene (*WEE1Hu*) was overproduced in bacteria and assayed in an in vitro system. Unlike its fission yeast homolog, the product of the *WEE1Hu* gene encoded a tyrosine-specific protein kinase. The human WEE1 kinase phosphorylated the p34^{cdc2}–cyclin B complex on tyrosine 15 but not on threonine 14 in vitro and inactivated the p34^{cdc2}–cyclin B kinase. This inhibition was reversed by the human Cdc25C protein, which catalyzed the dephosphorylation of p34^{cdc2}. These results indicate that the product of the *WEE1Hu* gene directly regulates the p34^{cdc2}–cyclin B complex in human cells and that a kinase other than that encoded by *WEE1Hu* phosphorylates p34^{cdc2} on threonine 14.

The mechanisms that regulate progression through the eukaryotic cell cycle are highly conserved. The G_2-M phase transition is universally regulated by p34^{cdc2}, a Ser-Thr protein kinase. The activity of the p34^{cdc2}–cyclin B complex is required for progression of cells into the M phase (*1*). In fission yeast, several mitotic regulators have been identified that are thought to directly regulate the p34^{cdc2}–cyclin B complex. One of these regulators, *wee1*$^+$, encodes a kinase (p107^{wee1}) that has been classified as a dual-specificity kinase on the basis of its ability to autophosphorylate on Ser and Tyr residues (*2–4*). p107^{wee1} phosphorylates the p34^{cdc2}–cyclin B complex on Tyr^{15}, thereby rendering p34^{cdc2} inactive (*2*). A second mitotic regulator, *cdc25*$^+$, encodes a protein phosphatase that dephosphorylates Tyr^{15} and activates the p34^{cdc2}–cyclin B complex (*5–10*). In higher eukaryotes, p34^{cdc2} is negatively regulated by phosphorylation on both Thr^{14} and Tyr^{15} (*11, 12*). The human kinases responsible for these phosphorylations are unknown, although the *cdc25*$^+$ gene product has been implicated in both Thr^{14} and Tyr^{15} dephosphorylation (*6, 8, 10*). On the basis of the conservation of structure and function demonstrated for cell cycle regulators throughout evolution, it was predicted that the human homolog of *wee1*$^+$ would also encode a dual-specificity kinase that would negatively regulate p34^{cdc2} by phosphorylation of Tyr^{15} and Thr^{14} (*13*). A gene (*WEE1Hu*) from a human foreskin fibroblast cDNA library has been cloned by its ability to rescue *wee1*$^+$ mutants in *Schizosaccharomyces pombe* (*14*). *WEE1Hu* shares 29% sequence identity within the kinase domain of *wee1*$^+$ and is predicted to encode a protein kinase of ~49 kD.

To analyze the biochemical activities associated with the human *WEE1* gene product, we expressed it in bacteria as a fusion protein with glutathione-S-transferase (GST) (*15*). A bacterial expression system was chosen on the basis of the apparent lack of endogenous tyrosine kinase activity in bacteria and the success of this expression system for the characterization

Department of Physiology, Tufts University School of Medicine, Boston, MA 02111.

*Present address: Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, and Department of Medicine, Beth Israel Hospital, 330 Brookline Avenue, Boston, MA 02215.
†To whom correspondence should be addressed.