Zinc fingers in sex determination: only one of the two C.elegans Tra-1 proteins binds DNA in vitro

David Zarkower and Jonathan Hodgkin
Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Received May 10, 1993; Accepted June 28, 1993

ABSTRACT

The tra-1 gene of Caenorhabditis elegans is a major developmental regulator that promotes female development. Two mRNAs are expressed from the tra-1 locus as a result of alternative mRNA processing. One mRNA encodes a protein with five zinc fingers and the other a protein with only the first two zinc fingers. We have derived a preferred in vitro DNA binding site for the five finger protein by selection from random oligonucleotides. The two finger protein does not bind to DNA in vitro. Moreover, removal of the first two fingers from the five finger protein does not eliminate binding and has little effect on its preferred binding site. We find that a protein sequence amino-terminal to the finger domain also appears to play a role in DNA binding.

INTRODUCTION

The tra-1 gene is the last in a pathway of regulatory genes that determines the sex of the nematode Caenorhabditis elegans (1, 2, 3). Animals in which tra-1 is actively expressed develop as females or self-fertile hermaphrodites (females containing some sperm), while animals that do not express tra-1 or in which the gene is missing develop as males (4, 5). The tra-1 gene produces two mRNAs as a result of alternative mRNA processing, both of which encode proteins containing zinc fingers. The shorter mRNA encodes a protein (Tra-1S) of 288 amino acids with two fingers, while the longer mRNA encodes a protein (Tra-1L) of 1110 amino acids with five fingers (the same two fingers plus an additional three; Figure 1A) (6). The zinc fingers of Tra-1L are remarkably similar to those of three other proteins, the products of the Drosophila gene cubitus interuptus Dominant (ciD) (7), and two human genes, GLJ (8) and GLJ3 (9, 10). While the two GLI genes have extended regions of similarity to each other (9), only the fingers are similar in all four proteins (Figure 1B). The tra-1 gene appears to be functionally unrelated to the other three: ciD is a segment polarity gene, GLI is an oncogene, and GLJ3 is involved in limb and craniofacial development.

The similarity of the zinc fingers suggests that the four proteins may bind to very similar DNA sequences, but the extent of conservation is much greater than should be required just to preserve recognition of a particular sequence. This suggests that some additional property has been retained, perhaps a feature of the binding interaction (for example an unusual orientation with the DNA) or possibly a protein/protein interaction involving the zinc fingers. The 'linker' sequences that join the fingers are very similar to the TGKPYX sequence typical of Knüppel-type fingers (11) except between fingers one and two. That linker differs in sequence and is unusually long, particularly in Tra-1.

As the final gene that controls sex determination throughout the animal, tra-1 is likely to regulate a number of genes whose functions are restricted to certain tissues or cell lineages (12). In principle these genes could be identified by genetic screens for mutations that affect only one sex. Indeed a number of such genes have been identified (1). However, tra-1 probably will not directly regulate all of these mutationaly-identified genes. To find the genes whose expression tra-1 controls directly, it is crucial to identify the DNA sequences to which Tra-1 proteins bind. Investigation of DNA binding properties should also clarify the role of the two finger protein. Mutations that truncate only Tra-1L can eliminate most tra-1 function, suggesting that Tra-1S may not bind DNA or otherwise regulate gene expression.

In this report we demonstrate that the zinc fingers of Tra-1L bind DNA in vitro and identify the preferred binding site. We show that the binding site consists of three elements, one recognized by fingers 3-5, the second by finger 1 and the third by a highly conserved part of the amino terminus of the protein. We find that the first two fingers are dispensable for efficient DNA binding, and we discuss the role of Tra-1S.

MATERIALS AND METHODS

Plasmids

All plasmids for in vitro transcription are based on pTTTAG (13), which contains the 5' untranslated region of the human /3 globin gene and an oligonucleotide encoding the c-myc epitope recognized by the monoclonal antibody 9E10 (14). cDNA fragments of Tra-1 were generated by PCR (6) and inserted into the EcoRI site of pTTTAG. The resulting plasmids can be transcribed and translated in vitro to yield proteins with the 9E10 epitope at the amino terminus. Structures of the Tra-1 proteins are diagrammed in Figure 6A. All tra-1 inserts were sequenced for mutations that affect only one sex. Indeed a number of such genes have been identified (1). However, tra-1 probably will not directly regulate all of these mutationaly-identified genes. To find the genes whose expression tra-1 controls directly, it is crucial to identify the DNA sequences to which Tra-1 proteins bind. Investigation of DNA binding properties should also clarify the role of the two finger protein. Mutations that truncate only Tra-1L can eliminate most tra-1 function, suggesting that Tra-1S may not bind DNA or otherwise regulate gene expression.

In this report we demonstrate that the zinc fingers of Tra-1L bind DNA in vitro and identify the preferred binding site. We show that the binding site consists of three elements, one recognized by fingers 3-5, the second by finger 1 and the third by a highly conserved part of the amino terminus of the protein. We find that the first two fingers are dispensable for efficient DNA binding, and we discuss the role of Tra-1S.
pDZ16. The PCR product of primers DZ101 and DZ115 was inserted into pT7TAG. pDZ16 contains Tra-1 sequences from nucleotide 174 to the end of the Tra-1S coding region and encodes protein Tra-1St.

pDZ25. Primers DZ117 and DZ108 were used to generate a cDNA fragment containing fingers 2-5. A DraI site contained in primer DZ117 was used to fuse this fragment to the XmnI site at nucleotide 754, deleting finger 1. pDZ25 encodes protein NF2-5.

pDZ23. An internal AatII fragment was deleted from pDZ7, removing nucleotides 693 to 967, and deleting fingers 1 and 2. pDZ23 encodes protein NF3-5.

pDZ26. The product of primers DZ118 and DZ108 was inserted into pT7TAG. pDZ26 contains Tra-1 sequences from nucleotide 738 to 1304, and encodes protein F1-5. Truncation of pDZ26 at nucleotide 1207 in finger 5 with BamHI allows translation of F1-4.

pTBS1. Oligonucleotides DZ110 and DZ111 were inserted into the Xbal site of pBS2KS+ (Stratagene). The resulting plasmid contains a perfect nonamer consensus (TGGGTGGTGT) and a T-rich sequence. In both pTBS1 and pTBS2, the T-rich sequence is one nucleotide closer to the nonamer than in the consensus derived from selected sites.

pTBS2. Oligonucleotides DZ112 and 113 were inserted in the Xbal site of pBS2SK+. The inserts in pTBS1 and pTBS2 are in the same orientation. pTBS2 contains a deviation from the nonamer consensus (TGGGTAGTGT).

Oligonucleotides

DZ101, 5' CGGAAATTCTACTGAGATCCCGGATACGG 3';
DZ108, 5' CGGAAATTCCGTCAGAGTTGGAATGTGGTC 3';
DZ110, 5' CTAGATACAGTTTCTCTGTGGTGTTCTTATAGGA CT 3';
DZ111, 5' CTAGATGTCCTATAAGACCCACCCACAGCGAAAATCTG TAT 3';
DZ112, 5' CTAGATACAGTTTCTCTGTTGGTAGCTTTATAGG ACT 3';
DZ113, 5' CGAGAGTCTCTATAAGACTACCCACGGGAAACT GTAT 3';
DZ115, 5' CGGAAATTCAATCTCTGATACCCACCAACC 3';
DZ117, 5' GCCGTTTAAGTCCAAATCTACGAGTC 3';
DZ118, 5' CGGAAATTCCGACAGTGGGAATGTGGTC 3';
LMB2, 5' GTAAACCAGCCGCGCTAG 3';
LMB3, 5' CAGGAAACACGCTAGTACG 3';
MDB1, 5' TTTCACACCGGCGGCAGAG 3';
MDB2, 5' GCTATTTTCCGACCGCGAAG 3';

The other oligonucleotides used (R76, Primer F, Primer R) are as described (13).

Antibodies

The mouse monoclonal antibody 9E10 (14) was the generous gift of S. Munro; rabbit anti-mouse IgG was from Sigma.

Binding site selection and gel mobility shift assays

Tra-1 proteins were prepared by in vitro transcription with T7 RNA polymerase (Pharmacia) and in vitro translation in rabbit reticulocyte extract (Promega) according to manufacturers' instructions. RNA was quantitated by measuring incorporation of α[32P]UTP. Proteins were translated in the presence of [35S]methionine and visualized by SDS-PAGE and autoradiography. After translation, ZnSO₄ was added to 50 μM. Each immunoprecipitation or gel-mobility shift reaction contained 1–2 μl of programmed reticulocyte extract. Selection of binding sites and gel mobility shift assays were performed as described (13). Gel mobility-shift reactions were incubated for 20 min. at room temperature before electrophoresis. Under these conditions, complex formation is complete in 5 min. (not shown).

Preparation of probes and competitor oligonucleotides

For gel mobility shift experiments, probes were prepared by PCR amplification as described (13). Unlabelled competitors were prepared by annealing oligonucleotides DZ110 and DZ111 (for TBS1) or DZ112 and DZ113 (for TBS2). Annealing was performed in TA buffer (33 mM Tris acetate pH7.9, 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DTT, 1 mg/ml gelatin) at 94°C for 1 min., then 56°C for 5 min. For the experiment in Figure 5, 200 ng of each oligonucleotide (DZ110/DZ111 or DZ112/DZ113) was annealed and labelled by the addition of 30 μCi α[32P]dCTP, 1mM each dATP, dGTP and dTTP and 2.5 units Klenow fragment DNA polymerase (Cambio) at room temp. for 30 min. Unincorporated nucleotides were removed by passage over a Sephadex G50 (Pharmacia) spin column.

Sequence analysis of selected sites

Oligonucleotides in complexes were recovered from dried mobility-shift gels as described (13) and ligated into pBS2KS+ (Stratagene) between the EcoRI and BamHI sites. Templates were prepared by rescue with helper phage (15) and sequenced by the chain termination method (16) using Sequenase® DNA polymerase (USB), or were prepared using Magic Miniprep® (Promega) and sequenced by linear amplification PCR (17). Consensus sequences were derived from comparison of all sequences. However only those sites in which the primer sequence did not overlap the nonamer sequence were used to compile the final consensus.

RESULTS

Selection of sites bound by five finger and two finger Tra-1 proteins

Several methods have been developed recently for the determination of DNA binding sites (reviewed in (18)). These involve either the isolation of short fragments of genomic DNA containing preferred binding sites or the isolation from a pool of random double stranded oligonucleotides of those containing suitable sequences. We chose the latter method, primarily because of the ease of locating binding sites within the selected DNA fragment. In the initial experiment, the proteins used for selection were the two finger protein Tra-1St (Tra-1 small protein with epitope tag) and the five finger protein NF1-5 (N-terminus and fingers 1-5), which were made by in vitro transcrip-
Figure 1. A. Diagram of the two Tra-1 proteins. Tra-1S is the 288 amino acid predicted product of the shorter tra-l mRNA, and Tra-1L is the 1110 amino acid predicted product of the longer tra-l mRNA. Translation of Tra-1L has been demonstrated genetically (4); the shorter mRNA is cotranscribed for the first 996 nucleotides and is presumed also to be translated. B. Comparison of zinc fingers

| Finger 1: | 
|-------|-------|-------|-------|-------|
| tra-1 | ciD   | GLI   | GLI3  | 
| (N)26TBS1 | HKCTFEGCWGfSRLENLKLPSHTGEKP | HKCTFEGCWGfSRLENLKLPSHTGEKP | HKCTFEGCWGfSRLENLKLPSHTGEKP |
| TBS2   | HKCTFEGCWGfSRLENLKLPSHTGEKP | HKCTFEGCWGfSRLENLKLPSHTGEKP | HKCTFEGCWGfSRLENLKLPSHTGEKP |

Figure 2. Mobility-shift analysis of selected sites. Pools of oligonucleotides from the second, third and fourth rounds of selection were used as probes in gel mobility shift analysis with the in vitro translated Tra-1St and NFl-5 (both proteins are diagrammed in Figure 5). The number of rounds of selection is shown above each lane. Each 16 μl reaction contained 0.5 ng probe and 2 μl rabbit reticulocyte lysate programmed with Tra-1St RNA (lanes 1 -3) or NFl-5 RNA (lanes 5 -7). Lanes 4 and 8: reactions contain unprogrammed lysate and fourth round oligonucleotides selected by Tra-1S (lane 4) or NFl-5 (lane 8). Probes were prepared by amplification with primers F and R (13).
Table 1. Binding sites selected by NF1-5.

<table>
<thead>
<tr>
<th>NF1-5 Binding Sites</th>
<th>4th Round</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Unconstrained Sites</td>
<td></td>
</tr>
<tr>
<td>2. tctctgctTTTATTCTGGCGGTTGCGGTT</td>
<td></td>
</tr>
<tr>
<td>3. tctctgctGATGCCTTTCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>5. tctctgctGCTGCTTTCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>7. tctctgctTCTGTTTTCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>8. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>10. tctctgctTCCGCTTTCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>12. tctctgctGGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>24. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>28. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>29. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>41. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>43. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>44. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>45. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>46. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>49. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>52. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>53. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>54. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>55. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>56. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>57. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>58. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>59. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>60. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>61. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>62. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>63. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>64. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>65. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>66. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>67. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>68. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>69. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>70. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>71. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>72. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>73. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>74. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>75. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>76. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>77. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>78. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>79. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
<tr>
<td>80. tctctgctAGTCTTTCTGCTGCTGCTGCTC</td>
<td></td>
</tr>
</tbody>
</table>

B. Constrained Sites

<table>
<thead>
<tr>
<th>4th Round</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>6. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>9. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>21. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>23. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>25. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>26. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>30. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>50. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>51. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>52. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>53. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>54. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>55. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>56. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>57. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>58. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>59. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
<tr>
<td>60. tctctgctGCGTCTTTCTGCTGCTGCTGCTC</td>
</tr>
</tbody>
</table>

A. Oligonucleotides in which the consensus is not constrained. The selected oligonucleotides in which the nonamer consensus (underlined) does not overlap with the primer sequences. Primer sequences are in lower case and sequences that were random in R76 are in uppercase. Numbers followed by an apostrophe indicate oligonucleotides whose orientation has been reversed to allow alignment. B. Oligonucleotides in which the consensus overlaps with primer sequences. One primer sequence of R76 contains fortuitous similarity to the preferred nonamer. As a result, many of the selected binding sites overlap the primer sequence. The consensus sequences derived from A and B are shown in Figure 6.

500 fmol TBS2 (lane 15). From this and similar experiments (eg Figure 4), we estimate that TBS1 is bound 20-50 fold more efficiently than TBS2 at equilibrium. We see no dramatic difference in the rates of binding of the two oligonucleotides (not shown).

A strong consensus emerges from comparison of the selected oligonucleotides, and single nucleotide changes from this consensus can greatly reduce binding. However, individual oligonucleotides do deviate from the consensus (Table 1A; Table 1B). This might represent genuine flexibility in the sequence requirements for high affinity binding of NF1-5 to its site. Alternatively, such oligonucleotides may be selected in spite of containing relatively poor binding sites. To test the flexibility of NF1-5 binding, we assayed six of the oligonucleotides (Figure 4), including one with only a 5/9 match to the consensus. Each labelled oligonucleotide (5 fmol) was used as a probe in binding reactions containing increasing amounts of unlabelled TBS1 or TBS2 as competitors, as in Figure 3. Four of the oligonucleotides contain a perfect match to the nonamer consensus (A-D). Of these, two also contain the T-rich sequence (A and B) and two do not (C and D). In the remaining two oligonucleotides (E and F), the nonamer differs from the consensus in two or four positions, but the T-rich sequence is present. All six oligonucleotides are bound efficiently by NF1-5 (lanes 2, 10, 18, 26, 34, 42) and are bound detectably in the presence of a 500-fold excess of TBS2 (lanes 8, 16, 24, 32, 40, 48). Thus even a site with four differences from the consensus (F) can bind more efficiently than one with a single difference (TBS2). Since deviations from the consensus can have only minor effects on NF1-5 binding, we conclude that some positions are more important than others for NF1-5 binding. The T-rich sequence appears to contribute little to the binding efficiency of sites that match the nonamer consensus (compare A and B with C and D). However it usually is present in selected sites with a poor match to the nonamer consensus (Table 1), suggesting that it does contribute to the binding efficiency of such sites.
Figure 3. Test of binding site consensus. Two pairs of complementary oligonucleotides were synthesized, one containing a perfect match to the nonamer consensus (TBS1) and the other with the fourth G of the nonamer changed to an A (TBS2). The oligonucleotides were cloned into pBS2SK⁺ (Stratagene) and labelled by PCR with primers F and R. These probes were used in gel-mobility shift assays with in vitro translated NF1-5 (lanes 1–8, TBS1 as probe; lanes 9–16, TBS2 as probe). Each binding reaction contained 5 fmol probe and 2 μl programmed rabbit reticulocyte lysate. Some reactions contained unlabelled TBS1 or TBS2 as competitor (Materials and Methods). Lanes 1 and 9: unprogrammed reticulocyte extract and no competitor. Lanes 2 and 10: NF1-5 programmed extract and no competitor. Lanes 3–8 and 11–16 contained competitor, either TBS1, indicated by the filled wedge (lanes 3 and 11, 50 fmol; lanes 4 and 12, 500 fmol; lanes 5 and 13, 2500 fmol) or TBS2, indicated by the unfilled wedge (lanes 6 and 14, 50 fmol; lanes 7 and 15, 500 fmol; lanes 8 and 16, 2500 fmol). The origin of the extra retarded complex in this experiment is unclear. It has not been seen in other experiments (for example, Figures 2,4,6), but is present in Figure 7, which is part of the same experiment.

The role of fingers one and two

The two finger Tra-1 protein, Tra-1St, did not select oligonucleotides, nor did it bind to those selected by the five finger protein NF1-5 (not shown). Our failure to show binding by Tra-1St could be a technical problem, due for example to using inappropriate binding conditions. Alternatively, Tra-1St may not bind to DNA. To help distinguish between these possibilities, we constructed and analyzed several altered forms of NF1-5. We reasoned that if fingers 1 and 2 are important for binding of NF1-5 and influence the binding sites that it selects, then Tra-1S may bind DNA under some conditions. Conversely, if fingers 1 and 2 are dispensable for binding of NF1-5 and have little or no effect on the selected binding sites, then Tra-1S is unlikely to be capable of binding DNA.

The four proteins we tested are diagrammed in Figure 5A. F1-5 retains all five fingers, but has only 17 amino acids of amino-terminal sequences. NF2-5 lacks finger 1 and 9 amino acids of amino-terminal sequences, while NF3-5 lacks fingers one and two plus 29 amino acids of amino-terminal sequences. The binding of each protein to labelled TBS1 and TBS2 is shown in Figure 5B. NF1-5 binds TBS1 efficiently (lane 1) and TBS2 less efficiently (lane 2). Removing the amino terminal portion of NF1-5 does not reduce the efficiency of binding to TBS1 (F1-5, lane 3). However, it abolishes binding to TBS2 (lane 4). This suggests that the amino terminal portion of NF1-5 can increase the affinity of binding to TBS2 (discussed below). Deletion of finger 1 (NF2-5, lanes 5 and 6) or of both finger 1 and finger 2 (NF3-5, lanes 7 and 8) reduces but does not eliminate binding. Furthermore, the sequence specificity of binding is retained (lanes 6 and 8). Deletion of finger 5 (F1-4, lanes 9 and 10) from F1-5 greatly reduces binding, consistent with finger 5 playing an important role. NF3-5 binds better than NF2-5 in spite of having fewer fingers and being more extensively deleted toward the amino terminus. This may be due to the juxtaposition of interfering sequences with the fingers as a result of the deletion. Both proteins form complexes with lower than expected mobility (lanes 5 and 7), consistent with their having an altered conformation.

Removing the first two fingers of NF1-5 does not abolish DNA binding, but it might change the preferred binding site and thus reveal a role in DNA binding for one or both fingers. We therefore performed five rounds of selection with each of the altered proteins and sequenced 40 clones from each selection. F1-4 failed to select oligonucleotides that could be detected in a gel mobility shift assay, but F1-5, NF2-5, and NF3-5 did select oligonucleotides. Alignment of the selected oligonucleotides (Figure 6) reveals that all of the proteins, including NF3-5, retain a preference for the same nonamer consensus as NF1-5. This
confirms that fingers 3, 4 and 5 are responsible for the interaction with the nonamer sequence.

There are clear differences, however, between the consensus preferred by the different proteins. These are of three types. First, only the proteins which contain the amino-terminal region adjacent to the fingers (NF1-5 and NF2-5) show a preference for the T-rich sequence 5' to the nonamer. The amino terminus may therefore contain a second DNA binding domain. Second, only the proteins that have finger 1 (NF1-5 and F1-5) prefer a C at position −5 relative to the nonamer. Thus finger 1 may interact with the DNA, but with considerably less specificity than fingers 3–5. Position −5 is within the region where finger 1 would interact if each finger spans three base pairs (20). There indication that finger two interacts with DNA. Third, all three of the deleted proteins show additional sequence preferences immediately adjacent to the nonamer. These may be pre-existing preferences that are enhanced to compensate for the interactions eliminated by deletion. Equally, they might reflect new interactions created in some way by the deletions.

In summary, analysis of DNA binding by the altered Tra-1L proteins suggests that finger 1 contributes to the interaction with DNA, but that fingers 1 and 2 are less important than the other three fingers. It also suggests that sequences other than the fingers interact with the DNA. This interaction is dispensible when the nonamer sequence is intact (Figure 5, lane 3), but may increase the affinity of non-consensus sites (lanes 2 and 4). The minor role of fingers 1 and 2 in binding by NF1-5 and its variants suggests that Tra-1S is unlikely to bind DNA with high affinity and specificity.

**Binding of NF1-5 to a putative tra-1 promoter fragment**

Genetic experiments have suggested that tra-1 is subject to negative autoregulation (21). Such regulation could be achieved directly through binding of Tra-1 to a site at the tra-1 promoter. Because both tra-1 mRNAs are trans-spliced, it is not yet known precisely where transcription initiates. However sequencing of the region upstream of the site of trans-splicing has revealed a potential Tra-1L binding site with a 8/9 match to the nonamer consensus and no T-rich sequence (TGTTAGTT; M. de Bono, unpublished). This sequence occurs 400 bp upstream of the site of trans-splicing. We therefore tested binding of NF1-5 to a DNA fragment containing this site (Figure 7). NF1-5 binds to the putative promoter fragment (lane...
efficiently than NF1-5, particularly to a site with a poor nonamer. NF2-5 and NF3-5 bind less binding affinity and specificity. NF2-5 and NF3-5 bind less binding site. In contrast, removing finger 5 almost abolishes DNA elimination and does not alter the preferred nonamer preference. Our data suggest that the preferred binding site consists of three elements, each recognized by a different part of the protein. The nonamer is the same as that bound by the very similar fingers of the GLI protein. Our data suggest that the preferred binding site consists of three elements, each recognized by a different part of the protein. The nonamer is recognized by fingers 3-5, a C at position -5 by finger 1, and a stretch of T residues by a domain amino-terminals to the fingers. Finally, we have shown that a DNA fragment from the putative promoter region of tap-1 is bound in vitro by NF1-5 protein.

The stretch of T residues was preferred only by NF1-5 and NF2-5, proteins that retain most of the amino terminal sequences. In contrast, F1-5 and NF3-5, which lack all or part of the amino terminus do not prefer the T residues. A 45 amino acid sequence in this part of the protein is highly conserved between two Caenorhabditis species (M. de Bono, D.Z., and J.H., in preparation). NF2-5 contains the entire sequence, while NF3-5 is missing part of it and F1-5 lacks it entirely. This sequence may be a new DNA binding domain.

Contrary to expectation, we have found no evidence that Tra-1S binds RNA. Tra-1S neither binds to oligonucleotides selected by NF1-5 nor itself selects oligonucleotides that it can bind. Furthermore, removal of fingers 1 and 2 from NF1-5 does not eliminate DNA binding and does not alter the preferred nonamer binding site. In contrast, removing finger 5 almost abolishes DNA binding and prevents the selection of binding sites. Our data suggest that fingers 1 and 2 do make a minor contribution to binding affinity and specificity. NF2-5 and NF3-5 bind less efficiently than NF1-5, particularly to a site with a poor nonamer. This observation and analysis of the selected sites together suggest that the amino-terminal domain and the first two zinc fingers contribute to the overall binding affinity by interacting with sequences other than the nonamer. However, the only specific preference we observe from fingers 1 and 2 is for a C at position -5. We believe that this is a real preference because it appears in two independent selection experiments. Also, of three genomic DNA fragments selected with GLI, the two that are bound most efficiently both have a C at -5 (22). Contacts with the phosphate backbone might allow fingers 1 and 2 to contribute to the affinity of Tra-1L (20); these contacts would not be apparent in the selected sites. The requirement for only a subset of zinc fingers may be common. In several other proteins fingers have been shown to make unequal contributions to DNA binding (23, 24, 25, 26).

Fingers 1 and 2 have several unusual features that may reflect a non-standard role. First, the spacing between the two fingers is unusually long (12 amino acids between the second conserved histidine and the next cysteine rather than the more usual seven between the other fingers). This might prevent binding to a contiguous block of DNA or otherwise affect binding. As a precedent, a naturally occurring splicing variant of the wt1 gene that inserts 3 amino acids between fingers 3 and 4 of that protein alters the binding specificity (27). Second, finger 2 is unusually long in the region predicted to form a beta sheet. Since contacts between the beta sheet and the phosphate backbone are thought to help orient the finger in the major groove (20), finger 2 might be expected to assume an unusual orientation. Third, none of the amino acids in the predicted nucleotide binding positions of finger 2 have the potential to form hydrogen bonds. As a result, finger 2 probably has little potential to make specific hydrogen bond contacts with the major groove. It might interact with the phosphate backbone, and in fact has four basic amino acids at the carboxyl end of the alpha helix, a region that makes backbone contacts in Zif268.

What is the role of Tra-1S? If Tra-1S does not bind to DNA, does it provide any tra-1 function? Molecular analysis of tra-1 mutations indicates that Tra-1L is essential for tra-1 function, but suggests that Tra-1S also may be important (6). Previously we suggested that Tra-1S might serve as a 'bodyguard' for Tra-1L by titrating negative regulators that otherwise would bind to Tra-1L and block its activity, and speculated that this type of regulation might be widespread (6). A class of tra-1 gain-of-function mutations exists which behave genetically as though they eliminate negative regulation of tra-1 (4). These mutations have been shown to be missense mutations in a short segment of the amino terminal region shared by Tra-1L and Tra-1S (M. de Bono, DZ and JH, in preparation). Thus the gain-of-function mutations are likely to define a site at which interaction with another protein mediates negative regulation of tra-1 (4). This is fully consistent with a 'bodyguard' role for Tra-1S. To test the model, it will be necessary to assay the feminizing activity in vivo of the amino terminal portion of Tra-1S. An important caveat is that Tra-1S might interact with another protein to form a complex capable of binding DNA. Although genetic experiments have not identified any other locus essential for tra-1 function, some transcription factors have been found to bind only in multi-protein complexes, and any second locus could have a lethal or pleiotropic phenotype.

We have not determined the preferred binding site of the full-length Tra-1L protein, but instead have used derivatives that are truncated after the zinc fingers. It is possible that the intact protein might have additional or different sequence preferences. Indeed, we have shown that sequences amino terminal to the fingers can
influence the preferred binding site. However we consider it unlikely that Tra-IL binds to a dramatically different site. This is because full-length GLI and GLI3, although similar to Tra-IL only in the zinc fingers, nevertheless bind to the same nonamer sequence in vitro (9, 22). In addition, nonsense mutations that remove the C terminal two thirds of Tra-IL can retain significant tra-1 function, demonstrating that the C terminus is not essential for function (4, 6).

We have shown the binding sites arbitrarily in an orientation with the G-rich strand of the nonamer as the top strand. Two observations suggest that NF1-5 interacts with this site such that finger 3 binds nearer the 5' end than finger 5 ('parallel' to the G-rich strand). First, finger 1 appears to prefer a C residue 5' to the nonamer. Second, a region amino-terminal to the fingers appears to prefer several T residues 5' to the nonamer. Such an orientation would be different from that of Zif268, which interacts primarily with the G-rich strand in an 'antiparallel' orientation (20). Other zinc finger proteins, however, appear to interact with both strands (eg. (28)), and some bind to an A/T-rich site.

The putative tra-1 promoter fragment does not contain an ideal NF1-5 binding site. The nonamer has a single non-consensus nucleotide and the C at -5 and T's are missing. It is bound less efficiently by NF1-5 than is the TBS1 oligonucleotide (Figures 3 and 5). This is consistent with a role in negative autoregulation of tra-1: lower binding affinity might ensure that regulation occurs only at levels of Tra-1 higher than necessary for binding to higher affinity Tra-IL targets. Otherwise tra-1 might turn itself off before reaching levels sufficient to regulate other targets.

A primary motivation for the experiments presented here is to help identify the targets of tra-1 regulation in vivo. Genetic experiments have identified a number of genes whose activity is restricted mostly or entirely to one sex (1). One gene in particular, mab-3, is an excellent prospect for a direct target of tra-1 regulation: it is required only in males, it functions in more than one tissue of the animal, and its effects are genetically epistatic to those of tra-1 (29). If mab-3 is regulated directly by tra-1, it will be of interest to ascertain whether its promoter contains tra-1 binding sites. Also, rapid progress is being made in the physical mapping and sequencing of the C.elegans genome, with more than 2000 kb of sequence already available for inspection. (30). It should be possible to use the information presented here to identify potential Tra-IL binding sites in the genome and hence to identify new genes subject to tra-1 regulation.

ACKNOWLEDGEMENTS

We thank Jan Fogg and Terry Smith for synthesizing oligonucleotides and Mair Churchill, Mario de Bono and Richard Treisman for providing oligonucleotides and plasmids. We thank many colleagues at the MRC LMB for valuable discussions. We are grateful to Vivian Bardwell, Mario de Bono, Mark Bretscher, Andrew Travers and Richard Treisman for critical reading of the manuscript. D.Z. was supported by postdoctoral fellowships from the Helen Hay Whitney Foundation and the American Cancer Society. J.H. is an International Research Scholar of the Howard Hughes Medical Institute.

REFERENCES
