JOURNAL OF VIROLOGY, Nov. 1995, p. 7334–7338 Vol. 69, No. 11
0022-538X/95/$04.00+0

Determination of Adeno-Associated Virus Rep68 and Rep78 Binding Sites by Random Sequence Oligonucleotide Selection

JOHN A. CHIORINI, LINDA YANG, BRIAN SAFER, AND ROBERT M. KOTIN*

Molecular Hematology Branch, National Heart, Lung, and Blood Institute, Bethesda, Maryland 20892

Received 26 May 1995/Accepted 4 August 1995

To further define the canonical binding site for the P_5-promoted Rep proteins of the adeno-associated virus, a modified random oligonucleotide selection procedure was performed, using purified recombinant Rep protein. These results may explain the effects of Rep on cellular gene expression.

The P_5-promoted Rep proteins of adeno-associated virus (AAV), Rep68 and Rep78, have been shown to regulate AAV replication (3–5, 14), transcription (13), and integration (11, 12, 18). Each of these diverse activities requires binding of these Rep proteins to tandem repeats of a GAGC motif. Although binding of Rep to the inverted terminal repeat (ITR) structure was previously thought to be absolutely dependent on the presence of a hairpin conformation (1, 10), we have recently shown that Rep can bind to a number of linear DNA substrates with an affinity similar to that of the wild-type (wt) ITR fragment (7). This finding may explain the effect of Rep on expression of heterologous promoters and its cytostatic effects when overexpressed. It is, therefore, important to define a consensus binding site and to determine the contribution of the individual bases at the sites of Rep-DNA interaction.

To define the Rep-DNA interactions, a modified random oligonucleotide selection procedure was performed, using purified MBP-Rep68Δ and MBP-Rep78. Electrophoretic mobility shift assays (EMSA) were used to isolate oligonucleotide probes containing Rep binding sites from a pool of randomly generated oligonucleotides containing over 10^{12} sequence combinations. With this technique, a canonical binding site was identified.

The technique of random selection has many advantages compared with base-by-base mutagenesis studies of wt elements for defining binding sequence requirements (17). The sequence combinations generated allow the determination of a statistically significant canonical binding site. The short oligonucleotides are easily cloned into bacterial vectors and sequenced. Iteration of the selection procedure yields oligonucleotides with higher-affinity binding sites (17). The procedure used for the random selection is similar to one previously described (9) and is outlined in Fig. 1.

The random probes were chemically synthesized and consist of a core of 20 random bases flanked by two constant regions (Fig. 2C). A double-stranded probe was generated by annealing a primer to one of the constant regions and extending it with the Klenow fragment of DNA polymerase and radiolabeled nucleotides. MBP-Rep68Δ is incubated with the pool of oligonucleotide probes under conditions which favor binding of Rep to its specific substrate (6, 7). Probes containing the Rep binding site were separated from nonbinding oligonucleotides by EMSA (7). The random probes were recovered and purified from the shift complex and amplified by using *Taq* polymerase and PCR (cycling conditions [35 cycles]: 95°C for 30 s, 50°C for 30 s, 72°C for 30 s). Aliquots were either selected again to enrich for higher-affinity fragments or cloned into plasmid vectors and sequenced. The band produced by incubating MBP-Rep68Δ with the random oligonucleotide probes (Fig. 3, lane 3) was only present in reaction mixtures containing MBP-Rep68Δ (Fig. 3, lane 1). Furthermore, the addition of wt ITR (Fig. 2A) could inhibit the formation of the shift band, demonstrating the specificity of the complex (Fig. 3, lane 2). Similar results were obtained with MBP-Rep78.

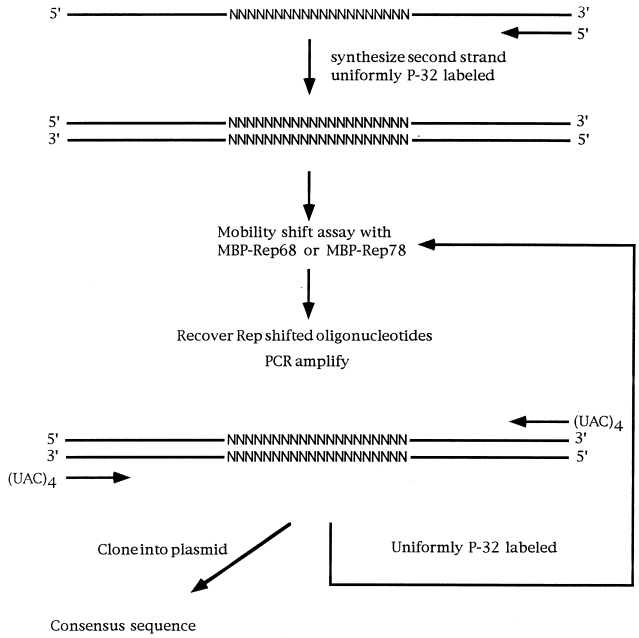

FIG. 1. Random selection procedure. The random probe was chemically synthesized and made double stranded by annealing and extending an oligonucleotide primer. The purified probe was then incubated with recombinant MBP-Rep68Δ or MBP-Rep78, and the complexed DNA was separated from the free DNA by EMSA. The complexed DNA was then isolated from the gel and purified. For cloning, the selected random probes were amplified with *Taq* polymerase by using PCR and oligonucleotide primers which were complementary to the constant regions and which contain $(UAC)_4$ tails complementary to the CloneAmp vector.

* Corresponding author. Mailing address: NIH/NHLBI/MHB, Bldg. 10/7D18, 10 Center Dr., MSC 1654, Bethesda, MD 20892-1654. Phone: (301) 496-1594. Fax: (301) 496-9985.

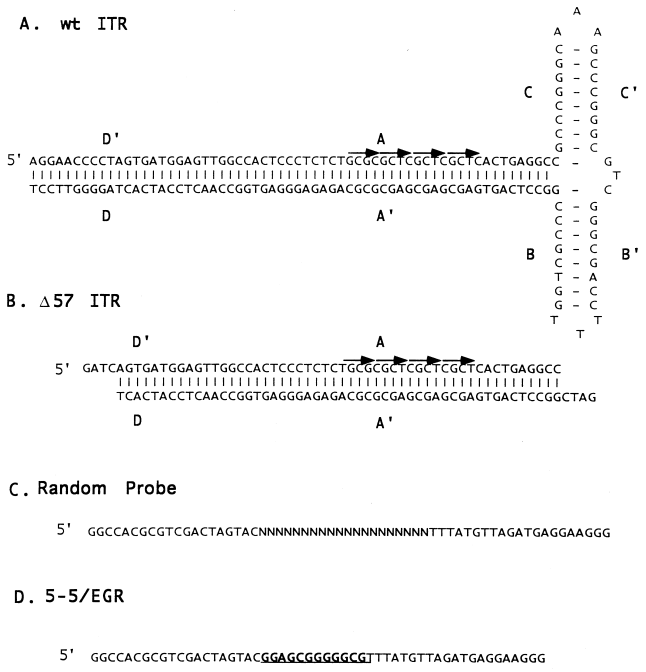

FIG. 2. Sequence of probes. (A) wt ITR was made as previously described (7). (B) Δ57 ITR is an annealed oligonucleotide pair, 57 bases in length, corresponding to all of the A/A′ region and part of the D/D′ region of wt ITR with a GATC 5′ overhang. (C) The random probe was chemically synthesized and consists of a core of 20 random bases (N) flanked on one side by a 20-bp region and on the other by a 21-bp conserved region, both of which are used for annealing primers for amplification and cloning. (D) Probe 5-5/EGR contains the core Rep binding site identified in clone 5-5 flanked by the conserved regions of the random probe. Sequences of the wt ITR and Δ57 ITR are aligned to the GAGC repeat motifs (horizontal arrows).

Of the 1.4×10^8 cpm of probe (4×10^7 cpm/pmol) incubated with 4 pmol of MBP-Rep78 and separated in the first round of EMSA, less than 1% of the probe was bound by Rep and recovered from the gel. These fragments were then incubated a second time with 1 pmol of Rep and EMSA fractionated a second time to enrich for the fragments containing higher-affinity Rep binding sites. Approximately 5% of this material was bound by Rep and recovered from the gel. The fragments isolated from the second round of selection were amplified by PCR. The amplified probes were either relabeled and reselected as described above or cloned, using the CloneAmp system (Bethesda Research Laboratories), and sequenced. This pool of clones is referred to as 78X2. Thirty-two fragments were sequenced, and their random bases were aligned by using a common GAGC motif present in each fragment (Fig. 4A). The position of this element within the probes varied, indicating that the conserved flanking sequence played a less critical role than the random region in the formation of the shift complex.

Log probability ratios, or LOD scores, were calculated for the aligned fragments. Using chi-square analysis, a nucleotide(s) was assigned to each position in the consensus sequence if P was <0.05 and the LOD score at that position was 3 or greater. Alignment of all 32 clones gave a consensus of **GAGC Ga**(g/c)g**G**gg(g/c), with bases in uppercase boldface type indicating statistically significant events (Fig. 4B). While this sequence is similar to the AAV ITR or AAVS1 sequences which contain multiple tandem repeats of GAGC, the consensus developed with random probes after two rounds of selection contained only one full repeat followed by a number of guanine (G) residues. To test the ability of runs of G residues to stabilize the binding of Rep78 to DNA, a probe which contained GAGC GGGGGCG in place of the block of 20 random bases was constructed (Fig. 2D). This sequence was identified in clone 5-5 and resembles the consensus binding site of the early growth response gene, EGR-1 [GCG(G/T)GGGCG] (8). The probe is therefore referred to as 5-5/EGR. The EMSA results indicate that MBP-Rep78 could form a stable interaction with this sequence (Fig. 5, lane 2). However, competition experiments using labeled Δ57 ITR (Fig. 2B) and unlabeled Δ57 ITR or 5-5/EGR (Fig. 2D) indicate that the affinity was approximately 5- to 10-fold lower for 5-5/EGR than for wt Δ57 ITR (Fig. 5, lanes 6 to 8 and 3 to 5, respectively). No competition was observed with a GC-rich probe for the α-pal binding site (9) (Fig. 5, lanes 9 to 11), demonstrating the specificity of the competition.

To identify the higher-affinity sites contained in the pool of selected fragments, a third round of selection was performed with PCR-amplified fragments from the second selection. Approximately 20% of this material was recovered and selected again in the fourth and final round. Nearly 50% of this input material was in the shift complex. The final pool of probes was then amplified by *Taq* polymerase, cloned, and sequenced and is referred to as 78X4. To compare the consensus binding sites for MBP-Rep68Δ and MBP-Rep78, the random selection procedure was performed with MBP-Rep68Δ instead of MBP-Rep78 as described above, and this resulting pool of fragments is referred to as 68X4.

Using either Rep68 or Rep78, the four rounds of selection

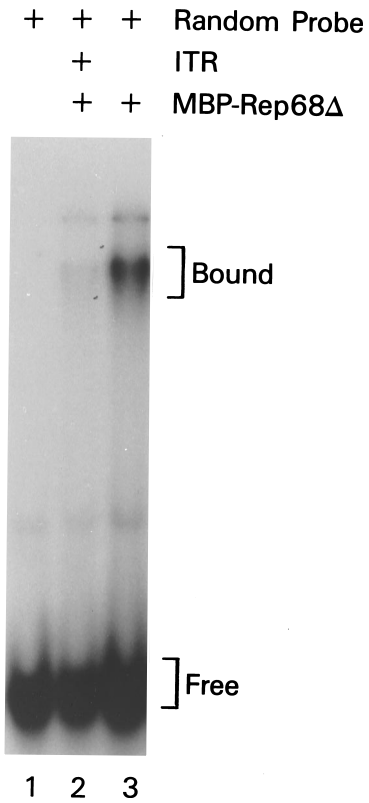

FIG. 3. EMSA of random oligonucleotide probe. A tight cluster of bands was produced by incubating 1 pmol of MBP-Rep68Δ with the random oligonucleotide probes (lane 3). No shift bands were detected in the absence of protein (lane 1). Formation of the lower-mobility complex could be reduced by the addition of wt ITR (lane 2), demonstrating the specificity of the complex.

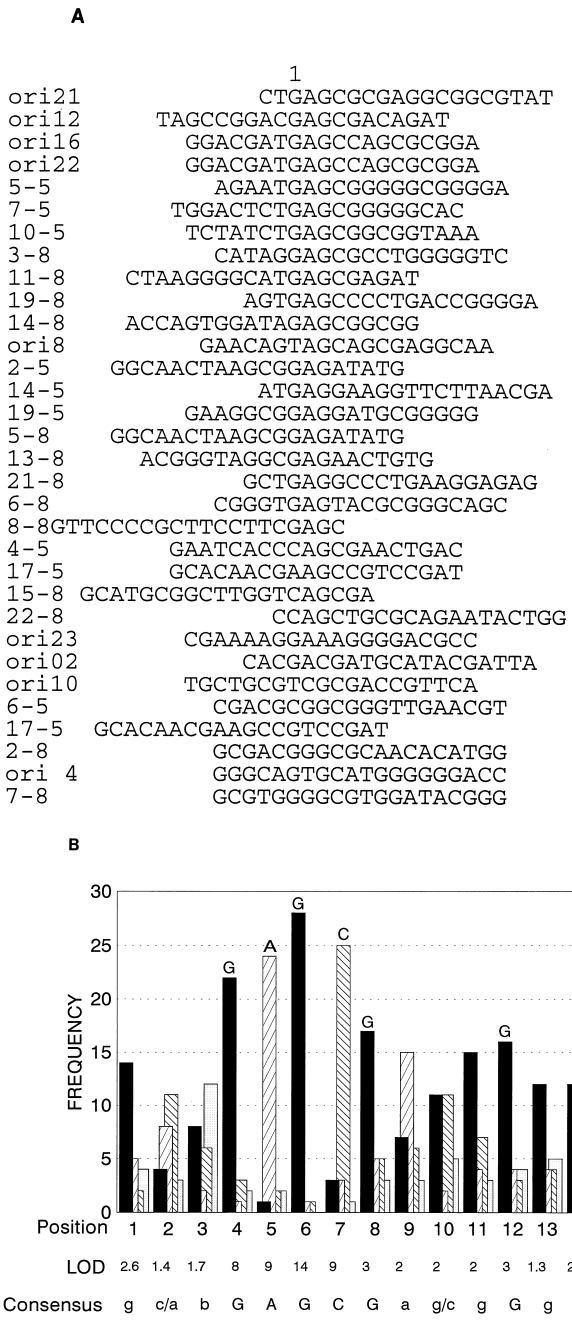

FIG. 4. (A) Sequence alignments of all 32 78X2 affinity-selected oligonucleotides. (B) Consensus sequence and LOD score. Positions of the aligned fragments from panel A are shown, with the frequency of the occurrences of the four bases at each position and the LOD scores for each position. Statistically significant bases are shown above the bars. An LOD score of 3 indicates that the probability of the distribution of the bases at that position occurring by chance is 1 in 1,000. A consensus binding site is also shown. ■, G; ▨, A; ▧, C; ▤, T.

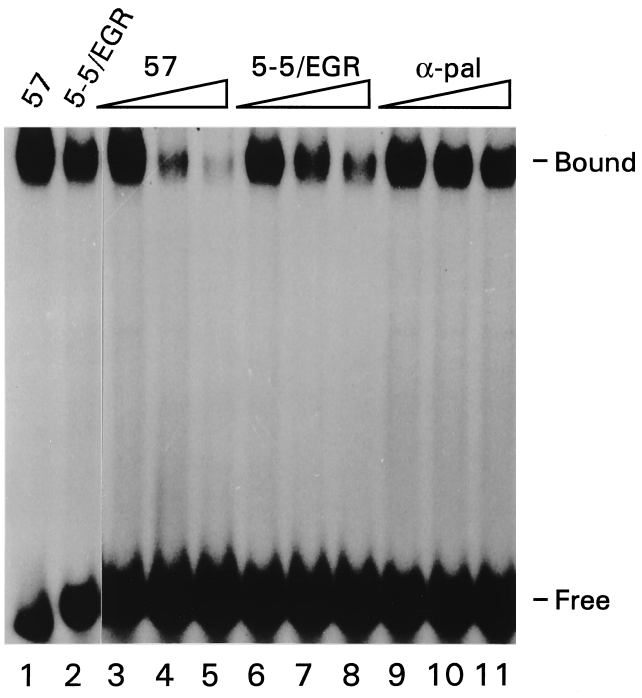

FIG. 5. The binding affinities of recombinant MBP-Rep78 were compared for three probes, Δ57 ITR, 5-5/EGR, and α-pal. Approximately 50,000 cpm of 5′-end-labeled Δ57 ITR was incubated with 100 ng of MBP-Rep78 in the presence of a 5-, 10-, or 15-fold excess of cold competitors: Δ57 ITR (lanes 3 to 5), 5-5/EGR (lanes 6 to 8), or α-pal (lanes 9 to 11). The binding affinities of MBP-Rep78 for Δ57 ITR and 5-5/EGR were also directly compared by incubating MBP-Rep78 with either Δ57 ITR or 5-5/EGR, which had been 5′-end labeled to the same specific activity (lanes 1 and 2, respectively).

resulted in a number of clones containing multiple GAGC repeats. A thymidine-to-cytosine (T-to-C) transition was very common in the 5′-most GAGC repeat motif for both 68X4 and 78X4 but was not as frequently observed in the adjacent repeats. As was seen in the consensus sequence developed for the 78X2 fragments, a number of clones contained repeats of G residues after the tandem GAGC repeats. Sequences of 41 clones from the 78X4 pool (Fig. 6) and those of 46 clones from the 68X4 pool (Fig. 7) were aligned, and consensus sequences of (**A**/**G**)vb**GAGCGAGC** n(**A**/**C**)**G** and (**G**/**A**)nn(**T**/**C**)G AG(**C**/**T**)**GAGCGAG**(c/a)(g/a)**V**, respectively, were identified. These consensus sequences more closely resemble the binding sites identified in the AAV ITR and the AAVS1 sequence (7, 16, 19).

The data presented in this paper suggest that while a preferred site for Rep binding contains tandem repeats of GAGC, stable binding with a lower affinity can be detected with one GAGC motif, provided it is followed by a run of G residues. In addition, a common substitution within the repeats was a T-to-C transition. In the 68X4 pool of fragments, T was as common as C in the first repeat and almost as common as C in the 5′ GAGC element of the 78X4 pool. This T-to-C substitution is also present in the first repeat in the P_5 promoter of AAV, which has been shown to bind Rep (15). The binding of Rep to a fragment containing an EGR-1 consensus sequence is intriguing and warrants further investigation to determine if Rep can effect the expression of EGR-1-responsive genes. The GAGC repeat motif is also similar to the binding site of the *Drosophila* transcription factor Zeste [(C/T)GAG(C/T)G] (2), suggesting that Rep may affect expression of genes regulated by the human homolog of this protein. The identification of a consensus binding site for Rep68 and Rep78, as well as acceptable substitutions within the binding sequence, will help to facilitate the understanding of the effects of Rep68 and Rep78 on the expression of heterologous promoters and their reported cytostatic effects.

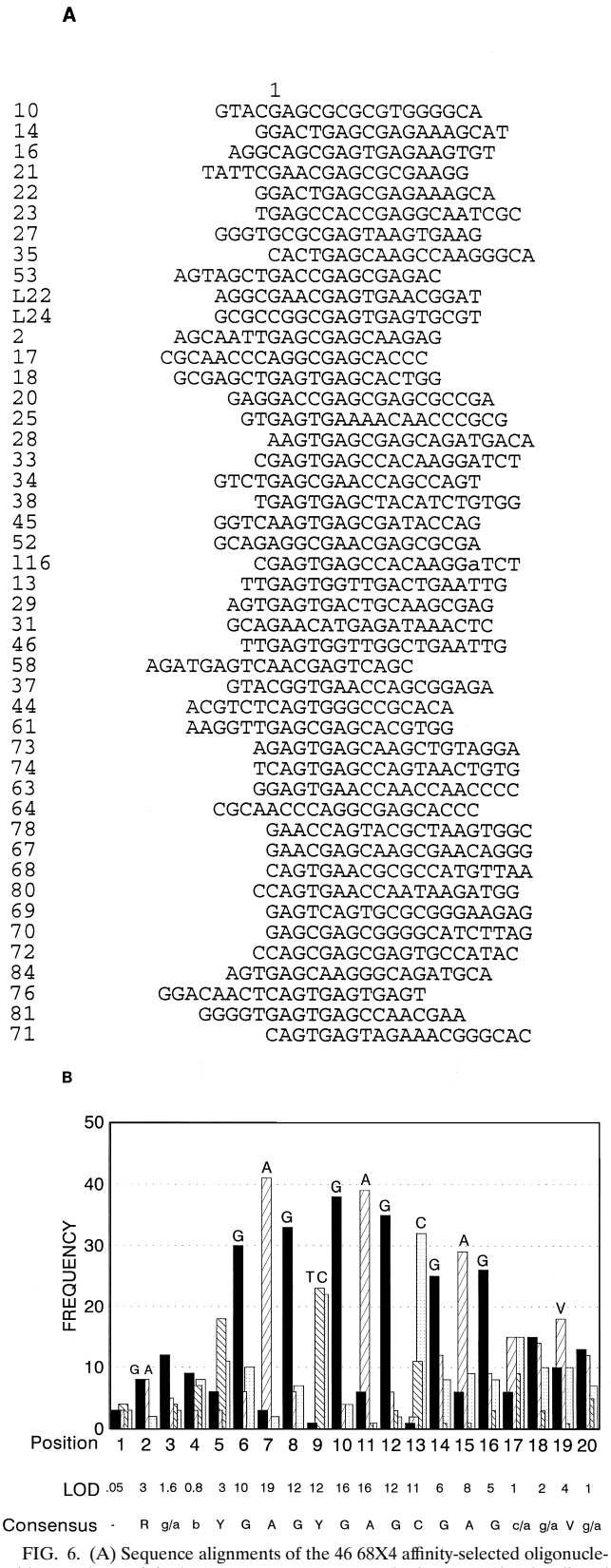

FIG. 6. (A) Sequence alignments of the 46 68X4 affinity-selected oligonucleotides are shown. (B) 68X4 consensus sequence and LOD scores for the aligned fragments in panel 6A are shown, with the sum of the number of occurrences of the four bases at each position and the LOD scores for each position. ■, G; ▨, A; ▧, C; ▤, T.

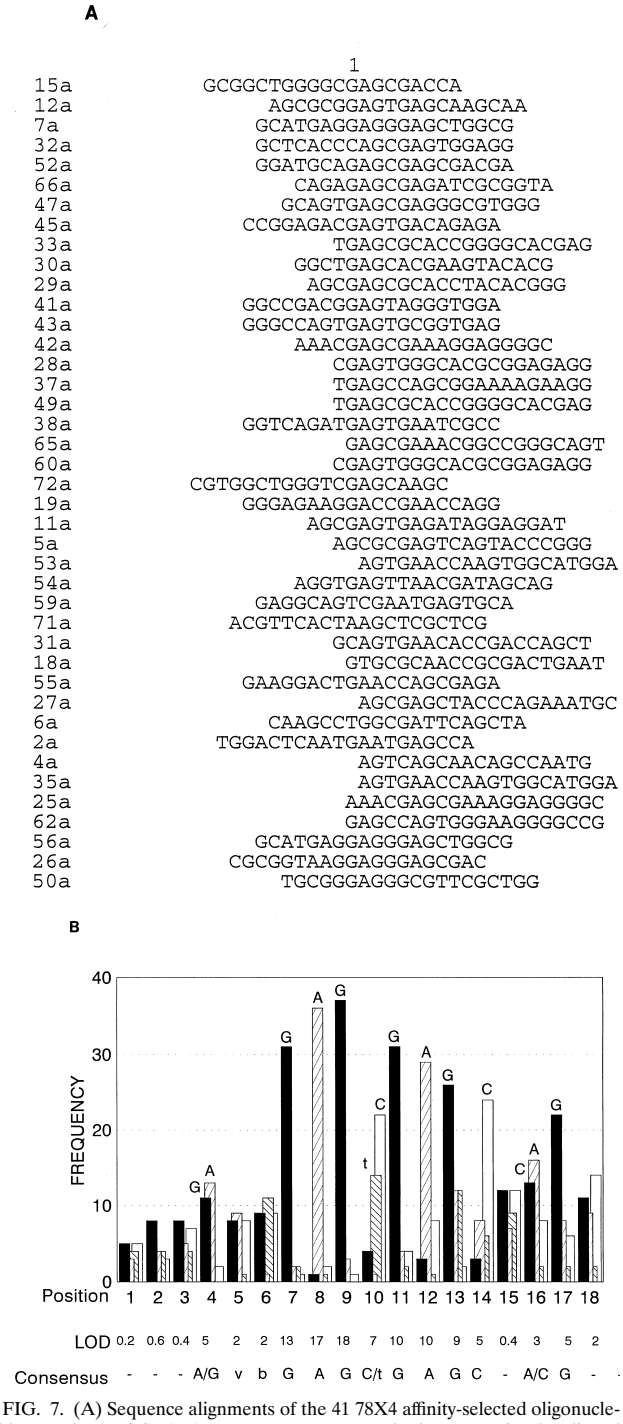

FIG. 7. (A) Sequence alignments of the 41 78X4 affinity-selected oligonucleotides are shown. (B) 78X4 consensus sequence and LOD scores for the aligned fragments in panel A are shown, with the sum of the number of occurrences of the four bases at each position and the LOD scores for each position. ■, G; ▨, A; ▧, C; ▤, T.

We thank Cathy Ciatto for her outstanding administrative skills.
J.A.C. is a Pharmacology Research Associate funded by the NIGMS.

REFERENCES

1. **Ashktorab, H., and A. Srivastava.** 1989. Identification of nuclear proteins that specifically interact with adeno-associated virus type 2 inverted terminal repeat hairpin DNA. J. Virol. **63:**3034–3039.
2. **Biggin, M. D., and R. Tijan.** 1988. Transcription factors that activate the ultrabithorax promoter in developmentally staged extracts. Cell **53:**699–722.
3. **Bohenzky, R. A., and K. I. Berns.** 1989. Interactions between the termini of adeno-associated virus DNA. J. Mol. Biol. **206:**91–100.
4. **Bohenzky, R. A., R. B. LeFebvre, and K. I. Berns.** 1988. Sequence and symmetry requirements within the internal palindromic sequences of the adeno-associated virus terminal repeat. Virology **166:**316–327.
5. **Chejanovsky, N., and B. J. Carter.** 1989. Replication of a human parvovirus nonsense mutant in mammalian cells containing an inducible amber suppressor. Virology **171:**239–247.
6. **Chiorini, J. A., M. D. Weitzman, R. A. Owens, E. Urcelay, B. Safer, and R. M. Kotin.** 1994. Biologically active Rep proteins of adeno-associated virus type 2 produced as fusion proteins in *Escherichia coli*. J. Virol. **68:**797–804.
7. **Chiorini, J. A., S. M. Wiener, R. A. Owens, S. R. M. Kyöstiö, R. M. Kotin, and B. Safer.** 1994. Sequence requirements for stable binding and function of Rep68 on the adeno-associated virus type 2 inverted terminal repeats. J. Virol. **68:**7448–7457.
8. **Christy, B., and D. Nathans.** 1989. DNA binding site of the growth factor-inducible protein Zif268. Proc. Natl. Acad. Sci. USA **86:**8737–8741.
9. **Efiok, B. J. S., J. A. Chiorini, and B. Safer.** 1994. A key transcription factor for eukaryotic initiation factor-2a is strongly homologous to developmental transcription factor and may link metabolic genes to cellular growth and development. J. Biol. Chem. **269:**18921–18930.
10. **Im, D. S., and N. Muzyczka.** 1989. Factors that bind to adeno-associated virus terminal repeats. J. Virol. **63:**3095–3104.
11. **Kotin, R. M., R. M. Linden, and K. I. Berns.** 1992. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. EMBO J. **11:**5071–5078.
12. **Kotin, R. M., M. Siniscalco, R. J. Samulski, X. Zhu, L. Hunter, C. A. Laughlin, S. McLaughlin, N. Muzyczka, M. Rocchi, and K. I. Berns.** 1990. Site-specific integration by adeno-associated virus. Proc. Natl. Acad. Sci. USA **87:**2211–2215.
13. **Labow, M. A., P. L. Hermonat, and K. I. Berns.** 1986. Positive and negative autoregulation of the adeno-associated virus type 2 genome. J. Virol. **60:** 251–258.
14. **LeFebvre, R. B., S. Riva, and K. I. Berns.** 1984. Conformation takes precedence over sequence in adeno-associated virus DNA replication. Mol. Cell. Biol. **4:**1416–1419.
15. **McCarty, D. M., D. J. Pereira, I. Zolotukhin, X. Zhou, J. H. Ryan, and N. Muzyczka.** 1994. Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. J. Virol. **68:**4988–4997.
16. **McCarty, D. M., J. H. Ryan, S. Zolotukhin, X. Zhou, and N. Muzyczka.** 1994. Interaction of the adeno-associated virus Rep protein with a sequence within the A palindrome of the viral terminal repeat. J. Virol. **68:**4998–5006.
17. **Oliphant, A. R., C. J. Brandl, and K. Struhl.** 1989. Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. Mol. Cell. Biol. **9:**2944–2949.
18. **Shelling, A. N., and M. G. Smith.** 1994. Targeted integration of transfected and infected adeno-associated virus vectors containing the neomycin resistance gene. Gene Ther. **2:**1–5.
19. **Weitzman, M. D., S. R. M. Kyöstiö, R. M. Kotin, and R. A. Owens.** 1994. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. Proc. Natl. Acad. Sci. USA **91:**5808–5812.