Nucleotide sequence of the transforming early region Elb of adenovirus type 12 DNA: structure and gene organization, and comparison with those of adenovirus type 5 DNA

Takao Kimura¹, Yukiharu Sawada¹, Morikazu Shinawawa², Yasuaki Shimizu¹, Kazuko Shiroki³, Hiroto Shimojo³, Hiroyuki Sugisaki⁴, Mituru Takanami⁴, Yoshiko Uemizu¹ and Kei Fujinaga^{1,5}

¹Cancer Research Inst., Sapporo Medical Coll., Sapporo 060, ²Dep. Veterinary Public Health, Obihiro Univ. Agriculture and Veterinary Medicine, Obihiro, ³Inst. Medical Science, Univ. Tokyo, Minato-ku, Tokyo 108, and ⁴Inst. Chemical Research, Kyoto Univ., Uji, Kyoto 611, Japan

Received 14 September 1981

ABSTRACT

The nucleotide sequence of the entire transforming early region Elb of the highly oncogenic adenovirus type 12 (Ad12) DNA has been determined. The total sequence (3860 base pairs) encompasses the entire transforming early region El of Ad12 DNA. From the sequence for the Elb region of Ad12, and the transcription map of the Elb region (1, 2, 3, and this paper) the structure and gene organization of the early region Elb of Ad12 DNA were analyzed and compared with those of the Elb region in the non-oncogenic Ad5 DNA (4, 5). Most of the sequences in the Elb region of Ad12 was highly homologous to that of Ad5. It is predicted that the Ad12 region Elb codes for polypeptides of 53.9, 19.1, and 8.9 kd. This situation is identical with that of the Ad5 region Elb which codes for polypeptides of 54.9, 20.6, and 8.3 kd. The function of these predicted polypeptides encoded by the Elb regions in cell transformation is discussed.

INTRODUCTION

There are at least 31 well-defined serotypes in human adenoviruses. These human adenovirus serotypes can be divided into 5 subgroups (A-E) according to their DNA homology (6). This classification corresponds well to the tumorigenicity of viruses in newborn hamsters. Thus, members of highly oncogenic subgroup A (Ad12, 18, 31) induce tumors at a high frequency after a short latent period, and those of weakly oncogenic subgroup B (Ad3, 7, 14, 16, 21) do infrequently after a long incubation period, while those of non-oncogenic subgroup C (Ad1, 2, 5, 6), E (Ad4) and D (the remaining serotypes) do not in animals. However, all viruses of subgroup A, B, C and D can transform cells in culture.

In the early stages of viral infection, at least four regions of the genome (El, 2, 3 and 4) are transcribed (7, 8). The early region El, which encompasses the left most 11% on the

linear adenoviral genome, is divided into two subregions : Ela and Elb, with specific promoters (9-11). Cell transformation with restriction fragments of adenoviral DNA has indicated that the fragments containing the leftmost 8% of the genome have the ability to induce complete transformation, but shorter fragments lack some traits typical for virus-transformed cells, and the minimum size fragments capable of transforming cells contain almost all the Ela region (12-14). This observation suggests that the region Ela contains the information for the conversion of normal cells into cells with some transformed phenotypes, but region Elb contains the information necessary for complete transformation (14, 15).

Up to the present, the nucleotide sequences of the termini of three adenoviral serotypes have been determined : the Ela region and a part of the Elb region of Ad12 (subgroup A), the Ela region of Ad7 (subgroup B) and the El region of Ad5 (subgroup C) (1, 16, 4). The comparative studies on the nucleotide sequences of the Ela regions among these adenoviruses were carried out (1, 17). The result showed that the DNA sequences of the Ela regions have been highly conserved among these adenoviruses, and suggests that elucidation of similarity and difference in genetic information encoded in the Elb region is crucial for understanding the molecular mechanisms involved in cell transformation and tumor formation.

In this paper, the nucleotide sequence of the Ad12 Elb region was determined ; the sites for transcription start and stop and for splicing, and the possible coding regions for the Elb polypeptides were assigned on the sequence, based on the data quoted above. We also showed that two basepairs have to be added in the Ad12 DNA sequence reported earlier (1). Since the sequence of the Elb region of Ad5 is available, we compared the structure and gene organization of the Elb region of Ad12 with those of the corresponding region of Ad5.

MATERIALS AND METHODS

Viral DNA and DNA fragments

Ad12 (strain Huie) was grown in KB cells and viral DNA was isolated from purified virions as described by Green and Pina

(18, 19). The HindIII fragments G, I, and F were prepared by cleaving Ad12 DNA with HindIII. Ad12 DNA was also digested with AccI in order to generate the AccI fragments J and F. The restriction fragments were separated by electrophoresis on agarose gels and their extraction from agarose gels were as described previously (1).

Enzymes and other materials

HaeIII, HhaI, HinfI, HindIII and HphI were purified in our laboratories. AccI, AluI, MstI, Sau3A, TaqI and XbaI were purchased from Boehringer Mannheim Corp., New England Biolabs, Inc. and Takara Shuzo Co. Ltd. T4-polynucleotide kinase and bacterial alkaline phosphatase were products of Worthington Biochemicals Corp.

[γ - 32 P] ATP was obtained from New England Nuclear. Acrylamide and urea were obtained from Wako Pure Chemical Industries, Ltd., dimethyl sulfate and hydrazine from Eastman Chemicals, and piperidine from Nakarai Chemicals, Ltd.

Sequencing procedure

The 5'-ends of restriction fragments were labeled by T4-polynucleotide kinase and [γ - 32 P] ATP as described previously (1). The two labeled 5'-ends of each restriction fragment were separated by secondary cleavage with a different restriction enzyme. The fragments labeled at one end were isolated by polyacrylamide gel electrophoresis, chemically degraded (specific for G, T, T+C and A>C), and sequenced according to Maxam and Gilbert (20). The gel systems used were 8%, 10% and 23% polyacrylamide in the presence of 7M urea, and thickness of gels was 0.5 mm.

RESULTS

Nucleotide sequence of the Elb region

The sequence of Ad12 analyzed in this paper runs from map position 4.4 to 11.2 (Fig. 1). We started our sequence determination with the preparation of cleavage maps of Ad12 DNA by various restriction endonucleases (21, 22). When the primary structure of the DNA became known, we found sites for restriction enzymes which we had not used for mapping, and could choose new combinations of enzymes for sequence analysis. HindIII frag-

Fig. 1 Cleavage maps of the left end of Adl2 DNA with *AccI* and *HindIII*. The hatched line (map units 4.4% - 11.2%) denotes the region Elb whose sequence is described in this paper. Letters between cleavage sites (arrows) show the nomenclature of the restriction fragments. The coordinates of the transcription units in this region are indicated by a bottom line.

ments G, I and F, and *AccI* fragments J and F were digested with restriction endonucleases indicated in Fig. 2. The sequences were determined by the method of Maxam and Gilbert (20), and the regions sequenced are indicated on the cleavage maps of the Elb region in Fig. 2.

In the course of sequencing experiments by this method, we observed the situations, in which A sequenced in one strand did not find its counterpart T in the opposite strand. So we decided to reconfirm parts of the *HindIII*-G sequence reported previously (1), where sequences were deduced from only one strand. The result showed A each for position 1,574 in the l-strand and position 2,073 in the r-strand of the genome.

Together with the sequence of the *HindIII*-G fragment which was sequenced earlier (1), the total sequences encompassed the entire Elb region of Adl2. The entire sequence of the l-strand of the Elb region is deduced (Fig. 3).

Assignment of the transcription map and coding regions

The elucidation of the primary structure of the Adl2 Elb region enabled us to determine its coding capacity. Three plausible reading frames are shown in Fig. 4A. In reading frame 3 large termination codon-free stretches are not found. In reading frame 1, there are two large open stretches: the first is between positions 1,843 (TAA) - 3,292 (TGA), and the second is between positions 3,292 (TGA) - 3,805 (TAA). In reading frame 2, we find a large open stretch between positions 1,526 (TAG) - 2,030 (TAA). The transcription map in the Elb region of Adl2

Fig. 2 Fine cleavage maps of Adl2 DNA in the Elb region and sequenced regions. Cleavage sites by restriction endonucleases are illustrated by vertical lines. Arrows indicate the direction of sequencing, and the solid line in each arrow denotes the region in which the sequence was determined. On the bottom of this figure is shown which parts of both strands (l and r) have been sequenced.

has been constructed by the nuclease S1 gel-mapping method as described (2, 3). According to the mapping data, at least three species of mRNA are transcribed from this region (mRNA V - VII). In Fig. 4B we summarize the regions directing these mRNAs and presumed splice sites.

Considerable amounts of the sequence information on the transcription start and stop as well as those on the splice sites are now available ; these are "Hogness box (formulated as TATAAATA)" reported to locate about 30 bp upstream from the mRNA initiation site (23), the sequence AAUAAA found in many characterized eukaryotic mRNAs approximately 10 - 20 nucleotides before the 3' poly (A) tail (24 - 27), and optimal sequences for the donor site (AG₁GTAAGT) and the acceptor site (YYNYAG₁, Y = pyrimidine, N = A,G,T,C) of splicing (28).

Ad5	1663	CTTAAAGGG	1671
Ad12	1487	TTAAACAGG	1495
MetGluAlaTrpGluCysLeuGluAspPheSerAlaValArgAsn			
TATATAATGGCGGTTGGCTTAATCTTGGTTA	CATCTGACCTCATGGAGGCTTGGGAGTGTTCAGAGATTTCTGCTGCGGTAA	1757	
GATATAAAGCTGGGTTGGTGTCTTGAATAGTTCATCTTAG	TATGGAGTTGGAAACTGTGCGGAAAGTTTCAGAGCGGTTGCGCA	1584	
MetGluLeuGluThrValLeuGlnSerPheGlnSerValArgGlu			
LeuLeuGluGlnSerSerAsnSerThrSerTrpPheTrpArgPheLeuTrpGlySerSerGlnAlaLysLeuValCysArgIleLysGlu			
CTTGCTGGAACAGAGCTTAACAGTACTCTTCAATCTCTGTTGGAGGTTCTCTGGGCTCATCCAGGCAGATTAGTCTGCAGAAATTAAGGA	1847		
GCTCTTCAGATATACCTCTAAACACAGTTACAGTTTGGAGGTTATCTGTTGGCTTACCTTAAGGCAAGTGGTAAATAGCGTCAAGA	1674		
LeuLeuGlnTyrThrSerLysAsnThrSerGlyPheTrpArgTyrLeuPheGlySerThrLeuSerLysValValAsnArgValLysGlu			
AspTyrLysTrpGluPheGluGluLeuLysSerCysGlyGluLeuPheAspSerLeuAsnLeuGlyHisGlnAlaLeuPheGlnGlu			
CGATTACAAGTGGAAATTTGAAGAGCTTTTCAATCTCTGTTGGTGGCTTTTGAATCTTGTGAATCTGGCTACCCAGCGCTTTTCCANGA	1937		
AGACTATACAGAGGGAATTTGAAACATATTGGCCGACTGTGCCAGGCGCTTGGCTTACTAGACCTTTCTTACCACCTTGGTCTTTCAGCA	1764		
AspTyrArgGluGluPheGluAsnIleLeuAlaAspCysProGlyLeuLeuAlaSerLeuAspLeuCysTyrHisLeuValPheGlnGlu			
LysValIleLysThrLeuAspPheSerThrProGlyArgAlaAlaAlaValAlaPheLeuSerPheIleLysAspLysTrpSerGlu			
GAAGGTCAATCAAGACTTTGGATTTTTCACACCGGGCGCGCTGGCGCTCTGTGCTTTTTCAGTTTATAACGATAAATGACCGCA	2027		
AAAGTGTTCAGATCTAGATTTTGCATCTGTGGACCGAAGGTTGCTTCTATTTGCTTTTTCGCAACCATATTGGATAATGGACGCA	1854		
LysValValArgSerLeuAspPheSerSerValGlyArgThrValAlaSerIleAlaPheLeuAlaThrIleLeuAspLysTrpSerGlu			
MetGluArg			
ArgAsnProSerGluArgGlyValProAlaGlyPheSerGlyHisAlaSerValGluSerGlyCysGluThrGlnGluSerProAlaThr			
GluThrHisLeuSerLysGlyTyrLeuLeuAspPheLeuAlaMetHisLeuTrpArgAlaValValArgHisLysAsnArgLeuLeuLeu	2117		
AGAAACCCATCTGAGCGGGGGAAGTCTGCTGGATTTCCTGGCAATGCATCTGTGGAGAGGGTTGTGAGACACAAAGATCGCTGCTACT	1944		
GAAATCCCACTGAGTTGGATTACATGCTGGATTACATGTCATGCACCTGTGGAGGGCATGGCTGAGAGGAGGGTTGCAATTTACTC			
LysSerHisLeuSerTrpAspTyrMetLeuAspTyrMetSerMetGlnLeuTrpArgAlaTrpLeuLysArgArgValCysIleTyrSer			
GluIleProProGluLeuGlyLeuHisAlaGlyLeuHisValAsnAlaAlaValGluGlyMetAlaGluGluGlyLeuHisLeuLeu			

[illegible]

2738 Thr Glu Asp Arg Val Ala Phe Arg Cys Ser Met Ile Asn Met Trp Pro Gly Val Leu Gly Met Asp Gly Val Val Ile Met Asn Val Arg
2750 AG AG CA CA GG CG CG CT TT AG AT GT GC CG GG GT CT TC GC AT GC AG GG GT AT TAT TAAT GT AG CG
2760 ACA ACC GCA GCA GG CT GG CT TT AG AT GT GC CG GG GT CT TC GC AT GC AG GG GT AT TAT TAAT GT AG CG
2770 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2780 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2790 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2800 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2810 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2820 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2830 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2840 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2850 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2860 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2870 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2880 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2890 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2900 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2910 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly Val Val Gly Leu Asp Gly Ile Thr Phe Ile Asn Val Arg
2920 Thr Ser Asp Arg Val Ala Phe Arg Cys Arg Met Gln Gly Met Gly Pro Gly

ThrArgCysSerLeuHisLeuGlyAsnArgGlyValPheLeuProTyrGlnCysAsnLeuSerHisThrLysIleLeuLeuGluPro
 ACCGCTGTTTCCTTCATTTCGTTAAGAGGGGGTGTTCCTACCTTACCAATGCAATTTGAGTCAACATAGATATTCCTTGAGCC
 ATCGGTGTACCATACATTAGCTTACGGCGGGGTATGTTAGCTTCCCAATGTAATTTACGCACTCAAACTATATGCTGGACCT
 MetArgCysThrIleHisLeuGlyLeuArgGlyMetPheArgProSerGlnCysAsnPheSerHisSerAsnIleMetLeuGluPro
 3278
 3060
 GlnSerMetSerLysValAsnLeuAsnGlyValPheAspMetThrMetLysIleTrpLysValLeuArgTyrAspGluThrArgThr
 GAGAGCATGTCCAAAGTGAACCTGAACGGGGTGTTCACATGACCATGAAGATCTGAAGGTGCTGAGGTACGATGACACCCACCC
 GAAGTGTTCGTAGAGTGTGTTAAATGGGGTATTTTGAATTCCTCTGGAATTATGTAAGGTTATAGATTAATGATGATACCTGACAT
 GluValPheSerArgValCysLeuAsnGlyValPheAspLeuSerValGluLeuCysLysValIleArgTyrAsnAspAspThrArgHis
 3365
 3150
 ArgCysArgProCysGluCysGlyLysHisIleArgAsnGlnProValMetLeuAspValThrGluGluLeuArgProAspHisLeu
 AGTGGACCTCTCCAGTGTGGGTAAACATATTAGAACCCAGCTGTGATCTGATGTACGAGGAGCTGAGGCCCATCACTTG
 CGTTGGGACAGCTGAGTGTGGTAGCAGTCATCTAGAACTTCGTCCCAATTGCTGTAATGTAACTGAGGAGCTGAGAAGTCAACACTT
 ArgCysArgGlnCysGluCysGlySerSerHisLeuGluLeuArgProIleValLeuAsnValThrGluGluLeuArgSerAspHisLeu
 3455
 3240
 ValLeuAlaCysThrArgAlaGluPheGlySerSerAspGluAspThrAsp
 GTCTGGCTGGACCCGCTGAGTTTGGCTCTAGCGATGAAGATACAGATTGAGGTTA
 ACCCTGCTTGGCTCGGACCTGACTCTGACGATGAAGACGACAACTGAGGTA
 ThrLeuSerCysLeuArgThrAspTyrGluSerSerAspGluAspAsn
 3545
 3319
 GlnProProPro
 MetSer Thr Asn
 AAGAAATATAAAGGTGGGGGTATTATGTAGTATTGTGATCTGTTT
 TATAAAGGTGGAAGTCAACTAAAAATTGTTTGTGTTTAAACAGC
 3620
 3394
 MetAsnGlyThrThrGlnAsn
 AspGluArgAsnTyrSerSerGluGln
 Ser PheAspGlySerIleValSerSerTyrLeuThrThrArgMetProProTyrAlaGlyValArgGlnAsnValMetGly
 TCGTTTGTATGGAAAGCATTTGAGGTTCATATTGAACCGCATCTCCCATGGCCGGGTTCGTGAGAATGTGATGGGT
 ACCGTCCGCTTTTTCATGGAGGGGTTTTAGGCCCTTATTGACTTCAGGTTACCTATTGGGCCGAGTACGTGAGAATGTGTAGAT
 AsnAlaIleLeuPheAspGlyGlyValPheSerProTyrLeuThrSerArgLeuProTyrTyrAlaGlyValArgGlnAsnValValGly
 ArgCysAlaPhe
 3702
 3484

SerSerIleAspGlyArgProValLeuProAlaAsnSerThrThrLeuThrTyrgluThrValSerGlyThrProLeuGluThrAlaAla 3792
 CACAGTGTGATGGTCCGCGCTCTGCGCAAACTGTACTACCTTACGACGCTGCTGAGACGCTGCTGGAACGCGCTTGGAGAGTGCAGCCCT 3574
 GTACAGTGCAGCGGTCCGCTCTGCGCACCTTCCAAATTCATCAAGATTAACCTATGCAGTATTGGAACCTCGCGTTGGATACCGGCGCG
 SerThrValAspGlyArgProValAlaProAlaAsnSerThrLeuThrTyrgluThrAlaAlaThrIleGlyProSerProLeuAspThrAlaAla
 SerAlaAlaAlaSerAlaAlaAlaAlaThrAlaArgGlyIleValThrAspPheAlaPheLeuSerProLeuAlaSerSerAlaAlaSer 3892
 CCGCGCGCTTACGCGCGCTGAGCGCGGAGTGTGACGCTGCTTCTGAGCCGCTTGCAGCAGTGCAGCTTCCC 3864
 CCGGTTCGAGCTTCGAGCGCGCTTACGCGCGAGTATGGCAGCTGATTCAGCTTGTACAATCACTTGGCTTCCAATGCTGTGACAC
 AlaAlaAlaSerAlaAlaAlaSerThrAlaArgSerMetAlaAlaAspPheSerPheTyrgluThrAlaGlnLeuSerGlnLysVal
 ArgSerSerAlaArgAspLysLeuThrAlaLeuLeuAlaGlnLeuAspSerLeuThrArgGluLeuAsnValValSerGlnGlnLeu 3972
 GTTCATCGGCCCGGATGACAACTTGAAGGCTCTTTGGCAAACTGGATCTTTGACCGGGAACCTTAATGTCTGTTCTTCAGCAGCTGT 3754
 GCACCGCAGTTCGACAGGACATTTGACTGTATGCTTGGCAAGCTTGAACCTCTAAGTCTCAGCTGGAAAGGCTATCCCAAGGTTG
 ArgThrAlaValArgGluAspIleLeuThrValMetLeuAlaLysLeuGluThrLeuThrAlaGlnLeuGluGluSerGlnLysVal
 LeuAspLeuArgGlnGlnValSerAlaLeuLysAlaSerSerProProAsnAlaVal 4062
 TGGATCTGGCGCCAGAGGTTTCTGCCCTGAAGGCTTCTCCCTCCCTCCCAATGCGGTTTAAACATTAATAAAACCAGACTCTGTTTGGA 3844
 AGGAATTAGCTGATCTACTACCATACCCAGCGCCAACTGTAAACCAATAAGAAACCTTAATTTGAGATGGTGTATGAATCTTT
 GluGluLeuAlaAspAlaThrThrHisThrProAlaGlnProValThrGln
 TTTGGATCAAGCAAGTGTCTTGTCTTTATT 4095
 ATTGATCTGTTTTT 3860

Fig. 3 The nucleotide sequence of the Elb region of Adl2 DNA, and comparison of the DNA and predicted amino acid sequences in the Adl2 and Ad5 Elb regions. Nucleotides are numbered from the left end of Adl2 DNA, and the sequence of the 1-strand is indicated. Comparison of the DNA and predicted amino acid sequences are aligned by inserting gaps to maximize homology, and common sequences between Adl2 and Ad5 are boxed. Below the sequence, the potential signal sequences for transcription start (Hogness box) and stop (AATAAA), and initiation and termination codons are marked by underlines. Donor and acceptor sequences for splicing are marked by open and filled circles, respectively. Common amino acids, including the pairs of Ser : Thr, Leu : Ile, Arg : Lys, Asp : Glu, and Asn : Gln, are underlined.

In the Elb region, before the mRNA initiation sites suggested by the mRNA mapping data, there are TATAAAG at position 1,498 and TATAAAA at position 3,321. In the vicinity of the presumed common termination site for mRNAs V, VI and VII, AATAAAA is present at position 3,803. In the vicinity of the splice sites suggested by the mRNA mapping data, the corresponding sequences are present ; the possible donor site sequences are AT_↑GTACTT at position 2,060 and AG_↑GTAAGT at position 3,294, and the acceptor site sequence is TAACAG_↓CA at position 3,363, which is preceded by a highly AT-rich region seen for many other intervening sequences (28). These positions of signal sequences for transcription and of splice sites are summarized in Fig. 4C.

On the assumption that the first available ATG (position 1,541) is the only initiation codon in the large termination codon-free stretches (Fig. 4A), mRNAs V and VI with a common 5' end would be translated in the same reading frame. The nonsense triplet TAA at position 2,030, which precedes the possible donor site at position 2,060, functions as a common termination codon in both mRNAs. As a result, both mRNAs V and VI specify identical polypeptides initiating with ATG at position 1,541, terminating with TAA at position 2,030 and consisting of 163 amino acids. The molecular weight (M.W.) computed from the amino acid composition is 19,103. On the other hand, when the ATG at position 1,846 is available in reading frame 1, mRNA V codes for a polypeptide, terminating with TGA at position 3,292 and consisting of 482 amino acids (M.W. 53,870), and mRNA VI codes for a polypeptide (terminating with TGA at position 3,408, 85 amino acids, M.W. 8,924) which have the N-terminal 72 amino acids in common with the mRNA V polypeptide. mRNA VII corresponds to the mRNA coding for the viral protein IX in Ad2 (9, 29). The translation of the mRNA VII would be initiated with ATG at position 3,372 and terminated with TAA position 3,805, resulting in a polypeptide of 144 amino acids (M.W. 15,000).

The assigned coding regions are illustrated in Fig. 4D.

DISCUSSION

There is TTTATT corresponding to AATAAA on the opposite

strand at position 3,842 (Fig. 3). This TTTATT must be associated with the 3' end of the "semi" late mRNA for viral protein IVa2 (9). Consequently, the nucleotide sequence of Ad12 DNA has been determined to encompass the entire early region Elb which plays an important role in the establishment of cell transformation. As can be seen in Fig. 2, a few gaps remain where only one strand has been sequenced. These gaps, however, were situated in the regions of significant sequence homology between Ad12 and Ad5. Therefore, the sequences obtained for the regions were expected to be reliable. We have also assigned the possible regions directing transcripts and coding for the translation products. Of course, we can not make definite conclusions until at least a part of the information on the structures of transcripts and translation products is obtained.

Since the sequences and the transcription maps of the Elb regions of Ad12 (1-3) and Ad5 (4, 5, 30-32) are now elucidated,

Fig. 4 Diagrams of the transcription map and predicted coding regions in the early region El of Ad12.

(A) The occurrence of termination codons in the left to the right direction. The nonsense codons (TAA, TAG, TGA) in the 1-strand of the Elb region of Ad12 DNA are arranged according to each of three possible reading frames, and indicated by vertical lines. In the long stop codon-free stretches inverse triangles indicate the suspected initiation codons (ATG).

(B) The transcription map deduced by the nuclease S1-mapping method (2, 3, 41). mRNA species identified are numbered I to VII, and indicated by lines below the corresponding early regions Ela and Elb. The continuous lines represent the areas found to anneal to mRNA, and the interrupted lines are the intron sequences. Numbers above the lines represent the nucleotide positions of each mRNA. The dots marked R indicate the potential ribosome-binding sites.

(C) Signal sequences in the early region El of Ad12 DNA. Numbers above the line represent nucleotide positions of assigned initiation and termination codons (open and filled triangles), splices and signal sequences for transcription start and stop. Their sequences are also shown below the line. Numbers above the donor and acceptor sequences for splices indicate positions of the end nucleotides of the exons.

(D) Predicted coding regions on each mRNA. The assigned coding regions are indicated by full or half filled lines on the regions corresponding to mRNA sequences. The half filled line means that the mRNA V and the mRNA VI can code for 53.9 kd and/or 19.1 kd and 19.1 kd and/or 8.9 kd polypeptides, respectively. Their estimated molecular weights and their coding frame numbers are also shown near full and half filled lines.

Fig. 5 The coding areas in the Elb regions of Ad12 and Ad5. The thick lines in three plausible frames indicate the predicted coding regions for polypeptides. The Elb mRNAs encoded by Ad12 and Ad5 are represented by the continuous lines. The mRNA splicings are also shown by the broken lines which connect the regions composing mRNAs and coding for polypeptides.

we can compare the organization of the Elb region of Ad12 with that of Ad5. Fig. 5 illustrates the over-all organizations of the Elb regions of Ad12 and Ad5. In the Elb region, their over-all organizations are very homologous. It has been also indicated that the over-all organization of the Ela region of Ad5 is practically identical to that of Ad12 (17). When one realizes that Ad12 and Ad5 have the different tumorigenicity in newborn hamsters, the high homology of the over-all organization of the El region should be noteworthy.

We have assumed that the long termination codon-free

stretch in each reading frame is the translatable region in the transcripts and the first ATG in each long termination codon-free stretch is used. Consequently, we can predict that the mRNA V and the mRNA VI could code for 53.9 kd and/or 19.1 kd and 19.1 kd and/or 8.9 kd polypeptides, respectively. In short, both the mRNA V and the mRNA VI can be translated from two initiation codons located at positions 1,541 and 1,846. In the mRNA V and the mRNA VI, the postulated ribosome-binding sequence was found at positions 1,805-1,818 preceding the ATG(position 1,846), but could not be found upstream of the ATG(position 1,541) (1).

It should be also noted that the distance between the Hogness box and the first ATG on the both mRNA V and mRNA VI is very short. These situations are identical to those of the Ad5 region Elb which encodes for polypeptides of 54.9, 20.6, and 8.3 kd (the postulated ribosome-binding sequence : positions 1,978-1,993, Figs. 3, 5 and 6). It is interesting to know how

Fig. 6 Location of homologous regions in the Elb regions of Adl2 and Ad5. The homologous regions are indicated by filled lines. The assigned initiation (▼/Δ) and termination (▼/▲) codons for each predicted polypeptide are shown above or below the lines representing mRNAs. Signal sites for transcription start (Hogness box ; P) and stop (AATAAA), and donor (d) and acceptor (a) sites for splicing are also represented.

the translation of both polypeptides is regulated on such mRNAs.

Subsequently, the sequences of Ad12 and Ad5 were compared base for base. It was found that the homologous DNA sequence between Ad12 and Ad5 occupied most of the Elb region, separated by one region of divergent sequence. The location of two homologous regions in each genome is shown in Fig. 6, and the sequences in each area are indicated in Fig. 3. They have been aligned by inserting some gaps to maximize homology. In the Elb region, the area comprising the promoter sites for transcription exists in the first homologous region. This region is located within the coding regions as seen in Fig. 6. The next homologous region appears some hundreds of nucleotides downstream from the first, and spans the sequence of about 1,700 bp. This region contains corresponding donor and acceptor site sequences for the smaller splicing, and the intervening regions resemble considerably each other. This region also contains partly the coding region for the viral protein IX. As is the case in Ad5 (4), the potential ribosome-binding site existing on the mRNA VII which probably codes for the viral protein IX in Ad12 is deleted from the Elb mRNA V and VI (Fig. 4). In the non-homologous region between the first and the second homologous regions, Ad12 DNA has a sequence by about 50 nucleotides shorter than Ad5 DNA.

Finally, we compared Elb polypeptides between Ad12 and Ad5 at amino acid level. The over-all amino acid compositions are similar between polypeptides 53.9 kd (Ad12) and 54.9 kd (Ad5), and between polypeptides 19.1 kd (Ad12) and 20.6 kd (Ad5) (Table I). Figs. 3 and 6 reveal the significant similarity in amino acid sequences of the N-terminal 80% between the 19.1 kd and the 20.6 kd polypeptides and of C-terminal 80% between the 53.9 kd and the 54.9 kd polypeptides. The viral proteins IX of Ad12 and Ad5 also resemble each other. However, the amino acid compositions and sequences of the 8.9 kd (Ad12) and the 8.3 kd (Ad5) diverge considerably (Table I, Figs. 3 and 6).

The cell transformation experiments have indicated that the role of the Ela polypeptides seems to be responsible for the conversion of normal cells into incompletely transformed cells, and that the Elb polypeptides also have an important function in

Table 1. Amino acid compositions of the hypothetical early polypeptides encoded for in the Elb regions of Ad12 and Ad5

	Ad12			Ad5		
	8.9kd	19.1kd	53.9kd	8.3kd	20.6kd	54.9kd
Asp	3	7	26	2	7	22
Glu	14	14	38	8	18	36
Arg	4	11	31	5	14	37
Lys	0	9	22	0	9	21
His	6	2	21	1	4	13
Gly	13	4	38	12	7	39
Ser	1	16	29	8	16	31
Thr	0	8	20	3	5	27
Cys	2	3	26	1	3	27
Tyr	1	7	9	0	2	10
Asn	3	4	20	3	5	27
Gln	4	7	13	4	11	12
Ala	14	8	33	13	14	36
Val	5	13	38	6	7	39
Leu	6	23	44	0	23	37
Ile	1	4	19	0	5	21
Pro	3	6	13	13	6	22
Phe	2	7	20	2	10	18
Trp	0	5	6	0	8	6
Met	3	5	16	3	2	15

establishing the complete transformation of cells. All of the regions coding for the middle class Elb polypeptides (the 19.1 kd and the 20.6 kd) and most of the regions coding for the small class Elb polypeptides (the 8.9 kd and the 8.3 kd) are contained in the smallest fragments (7 to 8% of the genome) capable of complete transformation of cells. These observations indicate that most of the regions coding for the large class Elb polypeptides (the 53.9 kd and the 54.9 kd) may not function in cell transformation. Thus, it is suggested that the middle class Elb polypeptides play important roles in establishing the complete transformation of cells.

On the other hand, between Ad12 and Ad5 exists an apparent difference in the tumorigenicity in newborn rodents. Transplantation experiments have also shown the difference in tumor formation in newborn rodents between cells transformed with DNA fragments of Ad12 and Ad5. Whereas Ad12 and Ad5 have a significant similarity in the Ela and the large and the middle class Elb polypeptides, we cannot almost find the similarity between their small class Elb polypeptides, except for a part of the

N-terminal. Thus, there is a possibility that the small class Elb polypeptides are related to the tumorigenic potential in Ad12 and Ad5.

In both in vivo and in vitro translation, it had been reported that the Elb regions of Ad12 and Ad5 code for two polypeptides of 60-65 kd and 19 kd (33-40) corresponding to the large and middle class Elb polypeptides, respectively. Although it is possible that these Elb polypeptides are specifically involved in cell transformation and tumor formation as discussed above, more extensive works will be needed to understand the function of each Elb polypeptide.

ACKNOWLEDGMENTS

We are very grateful to Dr. H. van Ormondt and Dr. J. L. Bos for providing us unpublished sequence data prior to their publication. This work was supported in part by Grants in Aid for Cancer Research and for Scientific Research from the Ministry of Education, Science and Culture, Japan.

⁵To whom reprint requests should be addressed.

REFERENCES

1. Sugisaki, H., Sugimoto, K., Takanami, M., Shiroki, K., Saito, I., Shimojo, H., Sawada, Y., Uemizu, Y., Uesugi, S. and Fujinaga, K. (1980) *Cell* 20, 777-786.
2. Sawada, Y. and Fujinaga, K. (1980) *J. Virol.* 36, 639-651.
3. Saito, I., Sato, J., Handa, H., Shiroki, K. and Shimojo, H. (1981) *Virology* 114, in press.
4. Van Ormondt, H., Maat, J. and Van Beveren, C. P. (1980) *Gene* 11, 299-309.
5. Bos, J. L., Polder, L. J., Bernards, R., Schrier, P. I., Van den Elsen, P. J., Van der Eb, A. J. and Van Ormondt, H. submitted for publication.
6. Green, M., Mackey, J. K., Wold, W. S. M. and Rigden, P. (1979) *Virology* 93, 481-492.
7. Sharp, P. A., Gallimore, P. H. and Flint, J. S. (1974) *Cold Spring Harbor Symp. Quant. Biol.* 39, 457-474.
8. Pettersson, U., Tibbetts, C. and Philipson, L. (1976) *J. Mol. Biol.* 101, 479-501.
9. Chow, L. T., Roberts, J. M., Lewis, J. B. and Broker, T. (1977) *Cell* 11, 819-836.
10. Berk, A. J. and Sharp, P. A. (1978) *Cell* 14, 695-711.
11. Wilson, M. C., Fraser, N. W. and Darnell, Jr. J. E. (1979) *Virology* 94, 175-184.
12. Houwelling, A., Van den Elsen, P. J. and Van der Eb, A. J. (1980) *Virology* 105, 537-550.

13. Dijkema, R., Dekker, B. M. M., Van der Felts, M. J. M. and Van der Eb, A. J. (1979) *J. Virol.* 32, 943-950.
14. Shiroki, K., Shimojo, H., Sawada, Y., Uemizu, Y. and Fujinaga, K. (1979) *Virology* 95, 127-136.
15. Van der Eb, A. J., Van Ormondt, H., Schrier, P. I., Lupker, J. H., Jochemsen, H., Van den Elsen, P. J., Deleys, R. J., Maat, J., Van Beveren, C. P., Dijkema, R. and De Waard, A. (1979) Cold Spring Harbor Symp. Quant. Biol. 44, 383-399.
16. Dijkema, R., Dekker, B. M. M. and Van Ormondt, H. (1980) *Gene* 9, 141-156.
17. Van Ormondt, H., Maat, J. and Dijkema, R. (1980) *Gene* 12, 63-76.
18. Green, M. and Pina, M. (1963) *Virology* 20, 199-210.
19. Green, M. and Pina, M. (1964) *Proc. Nat. Acad. Sci. USA* 51, 1251-1259.
20. Maxam, A. M. and Gilbert, W. (1977) *Proc. Nat. Acad. Sci. USA* 74, 560-564.
21. Kimura, T., Sawada, Y., Uesugi, S. and Fujinaga, K. (1980) *Tumor Res.* 15, 44-55.
22. Sawada, Y., Honda, E., Yano, S. and Fujinaga, K. (1980) *Tumor Res.* 15, 35-43.
23. Proudfoot, N. J. (1979) *Nature* 279, 376.
24. Proudfoot, N. J. and Brownlee, G. G. (1976) *Nature* 263, 211-214.
25. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. and Goodman, H. M. (1977) *Nature* 270, 486-494.
26. Merregaert, J., Van Emmelo, J., Devos, R., Porter, A., Fellner, P. and Fiers, W. (1978) *Eur. J. Biochem.* 82, 55-63.
27. Maat, J. and Van Ormondt, H. (1979) *Gene* 6, 75-90.
28. Seif, I., Khoury, G. and Dhar, R. (1979) *Nucl. Acids Res.* 6, 3387-3398.
29. Pettersson, U. and Mathews, M. B. (1977) *Cell* 12, 741-750.
30. Baker, C. C. and Ziff, E. B. (1979). Cold Spring Harbor Symp. Quant. Biol. 44, 415-428.
31. Aleström, P., Akusjärvi, G., Perricaudet, M., Mathews, M. B., Klessig, D. F. and Pettersson, U. (1980) *Cell* 19, 671-681.
32. Perricaudet, M., Le Moullec, J. M. and Pettersson, U. (1980) *Proc. Nat. Acad. Sci. USA* 77, 3778-3782.
33. Harter, M. L. and Lewis, J. B. (1978) *J. Virol.* 26, 736-749.
34. Halbert, D. N., Spector, D. J. and Raskas, H. J. (1979) *J. Virol.* 31, 621-629.
35. Green, M., Wold, W. S. M., Brackmann, K. H. and Cartas, M. A. (1979) *Virology* 97, 275-286.
36. Schrier, P. I., Van den Elsen, P. J., Hertoghs, J. J. L. and Van der Eb, A. J. (1979) *Virology* 99, 372-385.
37. Ross, S. R., Flint, S. J. and Levine, A. J. (1980) *Virology* 100, 419-432.
38. Spector, D. J., McGrogan, M. and Raskas, H. J. (1979) *J. Mol. Biol.* 126, 395-414.
39. Esche, H., Mathews, M. B. and Lewis, J. B. (1980) *J. Mol. Biol.* 142, 399-417.
40. Jochemsen, H., Daniels, G. S. G., Lupker, J. H. and Van der Eb, A. J. (1980) *Virology* 105, 551-563.
41. Perricaudet, M., Le Moullec, J. M. and Tiollais, P. (1980) *Nature* 288, 174-176.