Volume 12 Number 14 1984 Nucleic Acids Research

An altered DNA sequence encompassing the *ras* gene of Harvey murine sarcoma virus

Shuhei Yasuda*, Masato Furuichi and Eiichi Soeda+

National Institute of Genetics, Mishima 411 and *Technical Department, Fujiya Confectionery Co., Hatano 257, Japan

Received 5 June 1984; Accepted 3 July 1984

ABSTRACT

The DNA fragment encompassing the ras gene of Harvey murine sarcoma virus was sequenced and assigned the coding region of a transforming protein, p21, to the sequence. Examination of nucleotide sequence, taken together with the result of analysis of the *ras* mRNAs (1), has revealed that p21 is encoded from a continuous coding region starting with the 5' proximal initiation codon but not a processed protein. However, there were found several differences between the sequence published by Dhar *et. al.* (2) and ours, including 9 deletions, 7 substitutions and 2 insertions of nucleotides in the published sequence of 997 nucleotides in length. Among these, one of the substitutions occurring in the coding region resulted in amino acid replacement of glycine by alanine at position 122 of p21. The evidences are presented with some of actual gel autoradiographs.

INTRODUCTION

Harvey murine sarcoma virus (Ha-MuSV) was originally isolated from tumours induced by inoculation of Moloney murine leukemia virus (Mo-MuLV) into rats (3). This virus was able to transform fibroblast cells in culture and induce sarcomas in susceptible mice (4,5). In the virus-infected cells, a transformation specific protein with a molecular weight of 21 kilo daltons, p21, was produced (6,7). The closed circular DNA intermediates have been isolated from the infected cells and cloned molecularly (8) and physically characterized (13). Studies on transforming ability of the subgenomic DNA allowed to localize it within the *Sma*I-*Pst*I fragment in the proximal half of the viral genome (9,10).

We have sequenced a fragment spanning the *ras* gene of Ha-MuSV (H-*ras*) by the methods of Maxam and Gilbert (11) and of "dideoxy" sequencing (12) and assigned H-*ras* coding region to the sequence. Comparison of the nucleotide sequence with that reported earlier (2) has showed several discrepancies including insertion, deletion and substitution of nucleotides. One of these changes result in amino acid replacement of glycine by alanine at position 122 of p21 and in frame shift upstream from H-*ras* gene. The

© IRL Press Limited, Oxford, England.

latter suggests that p21 is encoded continuously from the 5' proximal ATG triplet in the single open reading frame. Technical problems will be discuss about DNA sequencing technology.

MATERIALS AND METHODS

[γ - 32 P]ATP and [α - 32 P] dCTP were purchased from RCC Amersham. M13 cloning and sequencing kits were obtained from either RCC Amersham or Takara Shuzo Co., Japan. Restriction endonucleases were gifts from Takara Shuzo Co. or purchased from New England Lab and Bio-Rad. T4 polynucleotide kinase was from Boehringer. Each enzyme and kit was used according to the suppliers specifications. dITP was from Boehringer.

The full length Ha-MuSV DNA cloned in λ gt WES $\cdot$ λ B (8) was generously provided by Dr. M. M. Martin and subsequently subcloned in pBR322 under the P2-EK2 conditions. The viral DNA was cleaved with appropriate restriction endonucleases to generate suitable sizes and separated by electrophoresis on polyacrylamide gels. After elution from the gels, the fragments were purified by the column chromatography on DEAE-sephacel (Pharmacia Co.).

The fragments were sequenced by the methods of Maxam and Gilbert (11) and of Sanger *et. al*. (12).

RESULTS AND DISCUSSION

Coding region of p21

Ha-MuSV was originally isolated by passage of Mo-MuLV in rats (3). The circular duplex DNA intermediates were extracted from the Ha-MuSV infected NIH 3T3 cells and cloned molecularly (8) and the DNA was physically characterized (13). Thus, the structure of Ha-MuSV genome in an integrated form is tripartite (14); i) a 1.2 kilo base (kb) sequence derived from both termini of Mo-MuLV, ii) a 3.2 kb sequence homologous to rat 30S RNA and iii) a 1.2 kb non-homologous sequence derived from the endogeneous transforming gene of rat, designated as c-*ras*. It is known that the *ras* gene family is widely distributed among eucaryotic species and the DNA sequences have been highly conserved since divergence of the species. It has been reported that the variants of human c-*ras* isolated from the bladder carcinoma cells (15,16) and neuroblastoma cells (17) carried a point mutation which resulted in the single amino acid substitution of p21, and were able to induce forcal transformation of NIH 3T3 mouse cells, respectively.

Transfection experiments with the subgenomic DNA (9,10) has circumscribed the coding region of H-*ras* in the *Sma*I-*Pst*I fragment 580 nucleo-

tides downstream from the 5' long terminal repeat sequence (LTR) derived from Mo-MuLV (6). The fragment was isolated by electrophoresis of a polyacrylamide gel and then submitted to sequencing mainly according to the method of Maxam and Gilbert (11).

The AccI-PstI fragment which encompasses the SmaI-PstI fragment was composed of 1042 nucleotide in length (Fig. 1). Nucleotide sequence examination has indicated that a long reading frame was left open from position 178 to 921 of this sequence with the first ATG triplet at position 355. The frame from the first ATG has amino acid sequence information to encode a 21.3 kilo dalton protein, consistent with a molecular weight of p21. In addition, the amino acid sequence predicted from this frame was highly homologous to those from the ras genes of other eucaryotic species published so far. Therefore, the open frame from the ATG was able to be assigned as a coding frame of p21, in agreement with that reported earlier (2) except for two nucleotide substitutions; T for C at position 717 and G for C at position 719 of the present sequence (Fig. 1). The latter change results in amino acid substitution of glycine by alanine at position 122 of p21. In the case of c-H-*ras*, alanine is encoded from the corresponding position (15).

Discrepancies in nucleotide sequences flanked by H-*ras*

Several differences in the sequences were also found in non-coding regions, particularly upstream from the p21 coding region. They included 9 deletions and 5 substitutions and 2 insertions of nucleotides in the published sequences (Fig. 1). Because the viral DNA was originated from the same clone but sequenced in two separate laboratories, these discrepancies in the sequences may be partially caused by technical problems occurring inevitably in sequencing. Thus, most of the regions where many differences were found in the sequences can be characterized by GC rich regions which cause troubles against sequencing due to formation of secondary structures while the gel is running. We tried to sequence them from the both strands by the method of Maxam and Gilbert (11) as well as the "dideoxy" method (12) but failed to do them mainly due to "band compression" emerging in the ladder sequencing patterns. When dITP was added instead of dGTP to "dideoxy" reaction mixtures (18), the band compression disappeared and additional G bands appeared in the ladder patterns with the expected spaces between bands (Fig. 2a).

The active elements for gene expression resided in the 5'LTR approximately 1 kb upstream from H-*ras* gene. It has been presumed that transcrip-

```
      AGACCCCGCTCTAGTGGCAGTGTGTTGGTTGATAGCCAAAGTTAATTTTTAAAA   54

                                                      ▼▼
CATAGTGTTTTGGGGGTTGGGGATTTAGCTCAGTGATAGAGCTCTTGCCTAGCAAGCGCA  114
  ▼                 [CC]                      ▼
AGGCCCTGGGTTCGGTCCCCAGCTCTGAAAAAAAGGAAAGAGAAACAAAACAAAAACATA  174
                                                    ▼
TAGTGTTTTATCTGTGCTTATGCCCGCAGCCCGAGCCGCACCCGCCGCGGACGGAGCCCA  234
      ▼A        ▼  A                        ▼    ▼
TGCGCGGGCCCAGTCGGCGCCCGTCCGCGCCCCCGCCCTGCCCCGGCCCCGGCCCCGGGG  294

GCAGTCGCGCCAGCAAGCGGTGGGGCAAGAGCTCCTGGTTTGGCAGCCCCTGTAGAAGCG  354

ATG ACA GAA TAC AAG CTT GTG GTG GTG GGC GCT AGA GGC GTG GGA   399
Met Thr Glu Tyr Lys Leu Val Val Val Gly Ala Arg Gly Val Gly    15

AAG AGT GCC CTG ACC ATC CAG CTG ATC CAG AAC CAT TTT GTG GAC   444
Lys Ser Ala Leu Thr Ile Gln Leu Ile Gln Asn His Phe Val Asp    30

GAG TAT GAT CCC ACT ATA GAG GAC TCC TAC CGG AAA CAG GTA GTC   489
Glu Tyr Asp Pro Thr Ile Glu Asp Ser Tyr Arg Lys Gln Val Val    45

ATT GAT GGG GAG ACG TGT TTA CTG GAC ATC TTA GAC ACA ACA GGT   534
Ile Asp Gly Glu Thr Cys Leu Leu Asp Ile Leu Asp Thr Thr Gly    60

CAA GAA GAG TAT AGT GCC ATG CGG GAC CAG TAC ATG CGC ACA GGG   579
Gln Glu Glu Tyr Ser Ala Met Arg Asp Gln Tyr Met Arg Thr Gly    75

GAG GGC TTC CTC TGT GTA TTT GCC ATC AAC AAC ACC AAG TCC TTT   624
Glu Gly Phe Leu Cys Val Phe Ala Ile Asn Asn Thr Lys Ser Phe    90

GAA GAC ATC CAT CAG TAC AGG GAG CAG ATC AAG CGG GTG AAA GAT   669
Glu Asp Ile His Gln Tyr Arg Glu Gln Ile Lys Arg Val Lys Asp   105

TCA GAT GAT GTG CCA ATG GTG CTG GTG GGC AAC AAG TGT GAC CTG   714
Ser Asp Asp Val Pro Met Val Leu Val Gly Asn Lys Cys Asp Leu   120
  T  G
GCC GCT CGC ACT GTT GAG TCT CGG CAG GCC CAG GAC CTT GCT CGC   759
Ala Ala Arg Thr Val Glu Ser Arg Gln Ala Gln Asp Leu Ala Arg   135
    Gly
AGC TAT GGC ATC CCC TAC ATT GAA ACA TCA GCC AAG ACC CGG CAG   804
Ser Tyr Gly Ile Pro Tyr Ile Glu Thr Ser Ala Lys Thr Arg Gln   150

GGT GTA GAG GAT GCC TTC TAC ACA CTA GTA CGT GAG ATT CGG CAG   849
Gly Val Glu Asp Ala Phe Tyr Thr Leu Val Arg Glu Ile Arg Gln   165

CAT AAA CTG CGG AAA CTG AAC CCG CCT GAT GAG AGT GGC CCT GGC   894
His Lys Leu Arg Lys Leu Asn Pro Pro Asp Glu Ser Gly Pro Gly   180

TGC ATG AGC TGC AAG TGT GTG CTG TCC TGA CACCAGGTGAGGCAGGGACC  944
Cys Met Ser Cys Lys Cys Val Leu Ser Ter                       189

AGCAAGACATCTGGGGCAGTGGCCTCAGCTAGCCAGATGAACTTCATATCCACTTTGATG 1004
   G   CG
TCCCTGCTCCCCCAATTCTGCCAATCCCCCTGCCTGCA                       1042
```

Fig. 1. The nucleotide sequence of the fragment encompassing H-*ras* gene of Ha-MuSV. The nucleotide sequence of the *Acc*I-*Pst*I fragment were determined by the methods of Maxam and Gilbert (11) and of Sanger *et. al*.(12), and compared with the sequence reported earlier (2). The differences in nucleotide sequences from the published ones are indicated above the sequence line. Arrows(▼) indicate the positions where nucleotides were added to the original sequence. The deletion of dinucleotide CC occurred between position 134 and 135 shown in the box. The predicted amino acid sequence of p21 is also shown below the sequence line with one replacement of glycine by alanine at position 122.

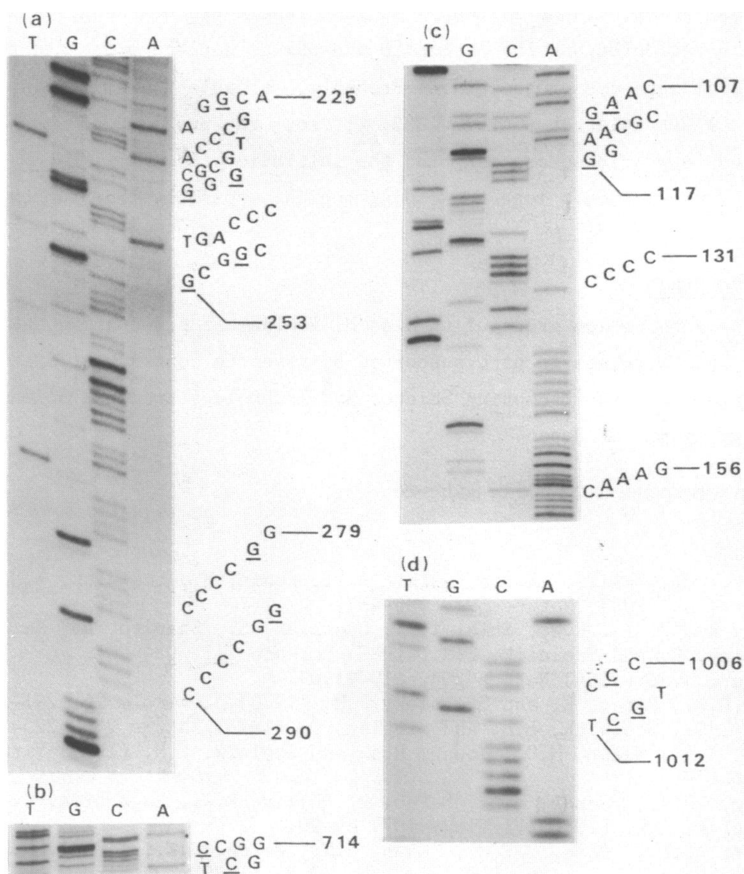

Fig. 2. Autoradiographs of sequencing gels. Nucleotide sequences were determined by the "dideoxy" method of Sanger et. al.(12). dITP was added to "dideoxy" mixture instead of dGTP in order to circumvent band compression occurring successive GC stretches (a). The nucleotide sequences different from the published by Dhar et. al.(2) are indicated with the underlined capitals.

tion of H-*ras* mRNA proceeds exclusively from the upstream LTR, generating the full-genome length of Ha-MuLV. The results of S1 mapping (1) have indicated that the mature genomic and/or subgenomic RNAs were unspliced in this region. In accordance with this, the consensus sequences conserved around the donor and acceptor sites for splicing events were not found in the present sequence. In the sequence upstream from H-*ras* coding gene, all reading frames were interrupted by the presence of several stop codons and the amino acid sequences predicted from such short open frames were not homologous to those predicted from the corresponding regions of the re-

lated ras genes. These evidences suggest that H-ras mRNA carries a long non-coding region beyond the AccI site and p21 is not the processed protein derived from a precursor protein. Probably, translation of p21 initiates from the ATG triplet at position 355. If so, the problems remain to be elucidated how ribosomes can bind the initiation codon of H-ras mRNA, scanning across such a long non-coding region, possibly from the cap site in LTR.

ACKNOWLEDGEMENTS

The authors are grateful to Miss H. Tominaga for preparing the manuscript. This work was in part supported by research contract of Special Coordination Fund for Promoting Science and Technology from the Science and Technology Agency of Japan.

+To whom correspondence should be addressed

REFERENCES

1. Gruss, P., Ellis, R.W., Shih, T.Y., Konig, M., Scolnick, E.M. and Khoury, G. (1981) Nature 293, 486-488.
2. Dhar, R., Ellis, R.W., Shih, T.Y., Oroszlan, S., Shapiro, B., Maizel, J.,Lowy, D. and Scolnick, E.M. (1982) Science 217, 934-937.
3. Harvey, J.J. (1964) Nature 204, 1104-1105.
4. Lowy,D.R., Rands, E. and Scolnick, E.M. (1978) J. Virol. 26, 291-268.
5. Scher, C., Scolnick, E.M. and Siegler, D. (1975) Nature 256, 225-226.
6. Shih, T.Y., Weeks, M.O., Young, H.A. and Scolnik, E.M. (1979) Virology 96, 64-79.
7. Young, H.A., Gonda, M.A., DeFeo, D., Ellis, R.W., Nagashima, K. and Scolnick, E.M. (1980) Virology 107, 89-99.
8. Hager, G.L., Chang, E.H., Chan, H.W., Garon, C.F., Israel, M.A., Martin, M.A., Scolnick, E.M. and Lowy, D.R. (1979) J. Virol. 31, 795-809.
9. Chang, E.H., Ellis, R.W., Scolnick, E.M. and Lowy D.R. (1980) Science 210, 1249-1251.
10. Wei, C.M., Lowy, D.R. and Scolnick, E.M. (1980) Proc. Natl. Acad. Sci. USA. 77, 4674-4678.
11. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acid. Sci. USA 74, 560-564.
12. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acid. Sci. USA 74, 5463-5476.
13. Chan, H.W., Garon, C.F., Chang, E.H., Lowy, D.R., Hager, G.L., Scolnick, E.M., Repaske, R. and Martin M.A. (1980) J. Virol. 33, 845-855.
14. Chang, E.H., Maryak, J.M., Wei, C.-M., Shih, T.Y., Shober, R., Cheung H.L., Ellis, R.W., Hager, G.L., Scolnick, E.M. and Lowy D.R. (1980) J. Virol. 35, 76-92.
15. Capon, D.J., Chen, E.Y., Levinson, A.D., Seeburg, P.H. and Goeddel, D.V. (1983) Narure 302, 33-37.
16. Taparowsky, E., Suard, Y., Fasano, O., Shimizu, K., Goldfarb, M. and Wigler, M. (1982) Nature 300, 762-764.
17. Taparowsky, E., Shimizu, K., Goldfalb, M. and Wigler, M. (1983) Cell 34, 581586.
18. Mills, D.R. and Kramer, F.R. (1979) Proc. Natl. Acid. Sci. USA 76, 2232-2235.