Activation of the C-myc Gene by Translocation: A Model for Translational Control

Haruo Saito; Adrian C. Hayday; Klas Wiman; William S. Hayward; Susumu Tonegawa


Stable URL:
http://links.jstor.org/sici?sici=0027-8424%2819831215%2980%3A24%3C7476%3AAOTCGB%3E2.0.CO%3B2-J

Proceedings of the National Academy of Sciences of the United States of America is currently published by National Academy of Sciences.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/nas.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
Activation of the c-myc gene by translocation: A model for translational control

(cellular oncogene/non-Hodgkin lymphoma/multiple promoters/DNA sequence)

HARUO SAITO*, ADRIAN C. HAYDAY*, KLAS WIMAN†, WILLIAM S. HAYWARD‡, AND SUSUMU TONEGAWA*

*Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and †Memorial Sloan–Kettering Cancer Center, New York, NY 10021

Communicated by S. E. Luria, September 6, 1983

ABSTRACT We have shown that the human cellular oncogene c-myc is composed of three exons and is transcribed from two initiation sites separated by 175-base-pair DNA in HeLa cells. For both resulting mRNA species, exon 1 comprises the 5′ untranslated region and the initiator methionine is located 16 base pairs downstream from the 5′ splice acceptor of exon 2. In a non-Hodgkin lymphoma, Manca, harboring a t(8;14) translocation, c-myc gene is broken within intron 1, and its exons 2 and 3 are translocated to a site between the heavy chain joining region cluster and Cγ coding DNA segment of the immunoglobulin heavy chain locus.

The translocated c-myc gene is transcribed from points within intron 1 but is apparently still translated from the same methionine codon as the mRNA from the unarranged c-myc gene. The nucleotide sequence of the c-myc gene shows that a region of intron 1 is highly complementary to a region of exon 2. Thus the mRNA from the untranslocated c-myc gene, as opposed to that of the translocated c-myc gene, could form a stable stem-loop structure (ΔG° = −90 kcal/mol) where the intron AUG would be located within the loop. In view of the bind-and-scan model for the initiation of eukaryotic translation, we propose that such a secondary structure will severely hinder the translation. We further propose that the c-myc gene is often activated by translocation through the escape from such a translational suppression.

Avian myelocytomatosis virus, MC29, causes a wide range of tumors in chickens and transforms chicken fibroblasts and macrophages in vitro. The viral oncogene apparently responsible is v-myc. The v-myc protein is a fusion product of part of the major viral structural protein (gag) sequences and sequences transduced from the chicken genome (1, 2). The transduced sequences are derived from a cellular gene, c-myc (3–5).

Activation of chicken c-myc has been suggested as the means by which the nonacutely leukosis virus induces neoplastic disease. By integrating close to the c-myc gene, the avian leukosis provirus can lead to enhanced levels of c-myc transcription (6). In addition, in a human cell line (HL60) derived from an acute promyelocytic leukemia, the human c-myc gene is amplified and there is a concomitant amplification of c-myc mRNA (7, 8). Amplification of the c-myc gene has also been observed in a human colon carcinoma cell line (9).

Nonrandom chromosomal translocations have been observed in a wide variety of vertebrate neoplasms (10–12). These observations, together with the demonstration of c-myc activation by avian leukosis virus integration, have led to the suggestion that c-onc genes might be activated by specific translocation events (6, 11, 12). In support of this idea, recent studies have shown that c-myc is translocated in certain lymphoid neoplasms of both mice and humans. In particular, murine c-myc on chromosome 15 is recombined into the heavy chain locus of mouse immunoglobulin genes (on chromosome 12) in BALB/c plasmacytomas characterized by t(12,15) translocations (13–17). The human c-myc gene has been mapped to a site on chromosome 8 (q24) (16, 17) that corresponds to the break point associated with translocations in Burkitt and other non-Hodgkin lymphomas. In a majority of Burkitt lymphomas characterized by t(8;14) (q24;q32) translocations, the c-myc gene on chromosome 8 is recombined into the immunoglobulin heavy chain locus on chromosome 14 (16, 18, 19). In other Burkitt lymphomas the translocations t(8;22) (q24;q11) and t(2;8) (p12;q24) may recombine the c-myc gene close to the A and κ immunoglobulin light chain loci, respectively. Do these translocations directly activate the oncogenic potential of the c-myc gene? If so, what is the mechanism for the activation?

In this communication we describe the structure and expression of a normal and a translocated human c-myc gene are compared. This comparison led to the hypothesis that the expression of c-myc gene product is ordinarily suppressed at the level of translation and that this suppression is removed as a result of c-myc translocation.

MATERIALS AND METHODS

The Manca cell line (SK-DHL1.2A) was derived from a patient with non-Hodgkin lymphoma. The line is pseudodiploid, carrying a single copy of the characteristic t(8;14) translocation (20). Cells were kindly provided by Bayard Clarkson (Memorial Sloan–Kettering Cancer Center). The isolation of cosmids clones cU2.3 and cAH1, which contain the translocated and untranslocated versions of the human c-myc gene, respectively, was discussed elsewhere. Nucleotide sequences were determined according to the method of Maxam and Gilbert (21). Sequence homologies were analyzed by the computer program SEQ (22). Glyoxal RNA gels, blotting of RNA to nitrocellulose filters, hybridization, and removal of the probes from the nitrocellulose filters were done according to Thomas (23). The nucleotide S1 protection experiments were done by a slightly modified method of Berk and Sharp (24, 25). A 348-base-pair (bp) Hpa I/Sma I fragment and a 171-bp Tag I/Sma I fragment labeled at the 5′ ends by T4 polynucleotide kinase were used as probes after strand separations (see Fig. 1).

RESULTS

Human c-myc Gene Consists of Three Exons. A two-exon structure for both the chicken (3–5) and human (26, 27) c-myc genes has been deduced by comparing restriction maps and nucleotide sequences with those of the v-myc gene of MC29 vi-

Abbreviations: bp, base pair(s); kb, kilobase(s).
rus. However, more recent sequence analysis of a human c-myc cDNA clone (28) suggests the presence of a third exon upstream of these two exons. To establish the exon-intron structure of the human c-myc gene, a series of DNA fragments (Fig. 1, probes 1–5) was dissected from a genomic c-myc gene clone cAH1 (unpublished results) and used as hybridization probes for RNA blotting analysis of the c-myc mRNA from HeLa cells (Fig. 2). Probes 4 and 5, which contain the previously described 5′ and 3′ exons of the c-myc gene, respectively, both hybridized to a mRNA of about 2.3 kilobases (kb). In addition, probe 2, which contains a 550-bp sequence located 1.6 kb upstream of the 5′ exon, hybridized to RNA of the same apparent size. Because probe 3 (a 1.1-kb fragment occurring between the probe 2 and the previously defined 5′ exon) did not hybridize to the 2.3-kb RNA, the RNA sequence detected by probe 2 must have been transcribed from an additional exon—probably the one suggested by Watt et al. (28).

No additional exon was found in the 2.5-kb region immediately upstream of the area covered by probe 2. These results suggest that the human c-myc gene is composed of three exons. Hereafter, these exons will be called exons 1, 2, and 3 from 5′ to 3′.

To confirm that the RNA sequence detected by probe 2 indeed represents a single exon, we determined the DNA sequence of the region covered by this probe and its immediate flanking region, using the strategy indicated in Fig. 1. The sequence is shown in Fig. 3 together with part of the exon 2 sequence previously determined (27). Comparison of the genomic sequence with the cDNA sequence (28) allows the 3′ boundary of exon 1 to be assigned to nucleotide 657. In the region upstream of this splice site the entire 5′ portion of the cDNA sequence of Watt et al. (28) is accounted for by a continuous stretch of the genomic DNA sequence presented here, except for 3 single bp insertions or deletions. This indicates that no additional introns split the c-myc gene, at least not in the region covered by the cDNA clone. The reasons for the discrepancies between the genomic DNA and cDNA sequences are unknown, but at least some may be attributed to human polymorphism.

Two Major Transcription Initiation Sites in the c-myc Gene. To accurately localize the 5′ end of exon 1 on the genomic DNA sequence we carried out nuclease SI protection experiments using probes composed of the two genomic DNA fragments thought to span the 5′ boundary of exon 1 (Fig. 1, probes 7 and 8). The results (Fig. 4) indicate that two alternative sites define the 5′ boundaries of exon 1, one at nucleotide position 104 and the other at position 279 (Fig. 3).

Upstream of each of these sites are "TATA" sequences, characteristic of many eukaryotic promoters (29). No sequence characteristic of splice acceptor sites (30) precedes either of the two boundaries (note that the conserved A-G dinucleotide alone is not a sufficient condition). The sizes of the c-myc mRNAs [2,200 bp and 2,030 bp plus poly(A)] predicted by summing the sizes of the three exons match well with the sizes of the mRNA detected by RNA blotting (2.3 kb). In fact, close inspection of the RNA blot data (Fig. 2) indicates that the 2.3-kb band is a doublet. We thus conclude that the c-myc gene is transcribed from at least two start sites in HeLa cells. The sequences covered by the two probes used in the nuclease SI protection experiment overlap with the 5′ end of the cDNA sequence (28).
Thus, all the above results indicate that no additional introns split the c-myc gene in the 5' region, confirming that this gene is composed by three exons. Because the cDNA reported by Watt *et al.* (28) starts at nucleotide 118, the corresponding mRNA is likely to be transcribed from start site 1.

**Exon 1 Does Not Code for Protein.** The first exon has no ATG codon in any of the three reading frames, while it has termination codons in all of the three reading frames (Fig. 3). These observations strongly indicate that exon 1 of the human c-myc gene has no protein-coding capacity. The first ATG codon appears in the second exon and is followed by a long coding frame showing a strong homology to the v-myc gene of MC29 virus (27).

**The Translocated c-myc Gene Is Often Devoid of Exon 1 and Its Transcription Initiates Within Intron 1.** In several Burkitt lymphomas harboring t(8,14) translocations the break points on the side of the c-myc gene have been mapped within 1–2 kb 5' of exon 2 (19). As shown in Fig. 1, and as will be described in detail elsewhere, the break point of the translocated c-myc gene in a non-Hodgkin lymphoma line (Manca) has been mapped to within intron 1 at a point 291 bp downstream of the 3' end of exon 1. Thus this c-myc gene is devoid of the entire exon 1 as well as the associated transcriptional promoters. To define the intron–exon structure of the translocated c-myc gene RNA blot analyses of the Manca RNA were carried out using exon 3 and intron 1 probes (probes 5 and 6, respectively, in Fig. 1). RNA from HeLa cells was used as a control for the transcription of...
an untranslocated c-myc gene. As shown in Fig. 2 (lanes 6–9),
the exon 3 probe detected the 2.3-kb RNA in HeLa cells and
the 2.6- and 2.2-kb RNA in Manca cells. As expected, the 2.3-
kb HeLa RNA was not detectable with the intron 1 probe. By
contrast, both the 2.6- and 2.2-kb Manca RNAs hybridized with
the intron 1 probe, suggesting that the sequence in this intron
is part of an exon in the translocated c-myc gene. That this ex-
onic sequence is in fact contiguous with the exon 2 sequence
in the mRNA derived from the translocated c-myc gene has been
shown by nucleotide sequence analysis of a cDNA clone iso-
lated from Manca cells (unpublished data). Furthermore, nu-
cleus S1 protection experiments with Manca RNA indicate that
new transcription start sites located within intron 1 are used for
the transcription of the translocated c-myc gene. Despite the
alteration in the mRNA, the translation initiation site is con-
served on translation because no ATG codon is present be-
tween the new transcription start sites and the ATG codon lo-
cated in the 5' region of exon 2 (ref. 27; unpublished results).
Similar features of a translocated c-myc gene in the mouse plas-
macytoma system were recently described by Stanton et al. (31).

**DISCUSSION**

The data presented here indicate that the human c-myc gene
comprises three exons, transcribed from two initiation sites
separated by 175 bp in HeLa cells. For both resultant mRNA
species, exon 1 lacks a methionine codon and is therefore non-
coding. Instead, the initiator methionine is located 16 bp
downstream from the 5' splice acceptor of exon 2. In a non-
Hodgkin lymphoma, Manca, harboring a (8;14) translocation,
the c-myc gene is broken within intron 1, and its exons 2 and
3 are translocated to a site between the Jκ cluster and Cκ-cod-
ing segments of the immunoglobulin heavy chain locus. This
translocated c-myc gene is therefore devoid of exon 1 and is
transcribed from points normally located within intron 1 of the
c-myc gene. Despite this, the mRNA from the translocated c-
muc gene is apparently translated from the same methionine
codon as the mRNA from the unarranged c-myc gene.

Does the lack of expression of exon 1 have an important ef-
fect per se in cells with such c-myc translocations? Examination
of the c-myc sequence (Fig. 3) reveals a region of exon 1 (nu-
cleotides 283–350) to have high complementarity to a region of
exon 2 (nucleotides 2,456–2,526) (Fig. 5A). A stem–loop sec-
ondary structure for the human c-myc RNA may therefore be
proposed. The standard free energy change (ΔGmp) of −90 kcal/
mol (1 cal = 4.184 J) predicted for such a structure (32) would
be sufficient to maintain it under physiological conditions.
The initiator AUG would then be located within the loop (Fig. 5B
and C). According to the “bind-and-scan” model for eukaryotic
translation, the 40S ribosome subunit binds the 5'-terminus of
mRNA and migrates toward the initiator AUG (33). This bind-
ing and migration is severely hindered when secondary struc-
ture is introduced into mRNA; reduction of secondary struc-
ture has the opposite effect (34). The negative effect of base
pairing on the translational efficiency of prokaryotic RNA has
also been demonstrated (35, 36). Therefore, the initiator AUG
in the human c-myc RNA can be adjudged to be inaccessible to
efficient initiation of translation.

By contrast, the lack of exon 1 in the c-myc RNA from the
translocated gene means that such a stem–loop structure cannot
be formed for this RNA. The transcript from the translocated
c-myc gene is therefore quite likely to be translated at a much
higher efficiency than the transcript from the unarranged c-
muc gene. The translocation may therefore result directly in
overproduction of the myc gene product.

To date, precise sites of c-myc rearrangements have been
mapped for only a few Burkitt lymphomas and non-Hodgkin
lymphomas (ref. 19; unpublished results; this study). Among
these, at least three (Lou, W1, and Manca) have rearrangement
sites within exon 1 or intron 1. Similarly, in the three murine
plasmacytomas (J558, M167, and M603) for which myc rear-
rangements have been documented at the nucleotide level, ei-
ther a complete exon 1 or the normal transcriptional promoter is
lost (19, 31). In short, the loss of exon 1 may be a common feature
of human Burkitt lymphomas and murine plasmacytomas in
which c-myc is rearranged. Therefore, the overproduction of
the myc gene product via a translational mechanism may be a
common feature for cells harboring these translocations. This

---

**Fig. 5.** (A) Internal sequence comple-
mentarity seen in exon 1 and exon 2 of the
germ-line (untranslocated) c-myc gene. The
nucleotide numbers are as in Fig. 3. The hy-
dergon bond forming bases are connected by
bars. G-U pairs, which in RNA are allowed to
make weak hydrogen bonds as G-U pairs, are
indicated by dots. Deletions indicated by
dashes are included to maximize the comple-
mentarity. (B) Schematic representations of
the transcription and splicing of the germ-line
(untranslocated) and translocated c-myc genes.
The germ-line c-myc gene is transcribed from
one of the two initiation sites 5' of exon 1; the
translocated c-myc gene is transcribed from
initiation sites within the intron. Only one of
the initiation sites in intron 1 is shown. (C)
Possible secondary structures of c-myc RNA.
The germ-line c-myc RNA can form a stem–
loop structure, with the initiation AUG codon
within the loop. The c-myc mRNA from the
translocated c-myc gene has a different 5'
structure and cannot form the stem–loop
structure.
mechanism may also play a role in avian leukosis virus-induced B-cell lymphomas. The vast majority of proviral integrations in these tumors is located within a region that would correspond to intron 1 of human c-myc (i.e., 0–1 kb upstream of exon 2 of chicken c-myc) (refs. 5, 37–39; unpublished data). Although the precise boundaries of exon 1 in the chicken c-myc gene have not been defined, sequencing data has revealed a region within the putative exon 1 that would form a stable stem–loop structure with sequence in exon 2 (unpublished data), in a manner analogous to that described here for the human c-myc gene.

There may be cases in which c-myc rearrangements occur at some distance from the c-myc gene (e.g., see ref. 19) and in which it is therefore not easy to apply the model for c-myc overproduction presented here. In these cases, overproduction of c-myc may be due to transcriptional effects (14, 40–42). In the case described here, Manca, the rearranged c-myc gene is brought close to sequences that, like their murine counterparts (43, 44), have profound tissue-specific transcription-enhancing activity. The possible effect of these sequences on the transcription of the altered c-myc gene is being examined. It may be possible to find other such enhancers in the neighborhood of other translocated c-myc genes.

In summary, there clearly exist multiple mechanisms by which the c-myc gene can be activated in cancerous cells. These include its amplification in promyelocytic leukemia and its transcriptional activation by nearby viral elements in avian leukosis virus-induced lymphomas. We wish to add to this list its activation by a translational mechanism that occurs as the result of the disruption of the gene during translocation in both mouse and human.

We thank John McMaster, Lena Angnan, and Anne Manwell for excellent technical assistance. A. C. H. was supported by a postdoctoral fellowship from the Imperial Cancer Research Fund (London), and K. W. was supported by a European Molecular Biology Organization long-term fellowship. This work was supported by Grants AI-17879 (S. T.) and CA-14051 (a core grant to S. E. Luria) and CA34502 (W. S. H.) from the National Institutes of Health.