103), α4 (104 to 111), and α5 (118 to 115). Least squares superpositions were performed with a program originally written by W. A. Hendrickson [*Acta Crystallogr.* **A34**, 713 (1986)].
13. P. V. Luzatti, *ibid.* **5**, 802 (1952).
14. S. R. Sprang, E. J. Goldsmith, R. J. Fletterick, *Science* **237**, 1012 (1987).
15. K. R. Acharya, D. I. Stuart, K. M. Varvill, L. N. Johnson, *Glycogen Phosphorylase b: Description of the Protein Structure* (World Scientific, Singapore, 1991).
16. S. R. Sprang and R. J. Fletterick, *J. Mol. Biol.* **131**, 523 (1979).
17. P. L. Mateo, C. Baron, O. Lopez-Mayorga, J. S. Jimenez, M. Cortijo, *J. Biol. Chem.* **259**, 9384 (1984).
18. H. Buc, *Biochem. Biophys. Res. Commun.* **28**, 59 (1967).
19. W. J. Black and J. H. Wang, *J. Biol. Chem.* **243**, 5892 (1968).
20. J. L. Martin, L. N. Johnson, S. G. Withers, *Biochemistry* **29**, 10745 (1990).
21. N. B. Gusev, J. Hajdu, P. Friedrich, *Biochem. Biophys. Res. Commun.* **90**, 70 (1979).
22. E. J. Goldsmith, S. R. Sprang, R. Hamlin, N.-H. Xuong, R. J. Fletterick, *Science* **245**, 528 (1989).
23. H. E. Morgan and A. Parmeggiani, *J. Biol. Chem.* **239**, 2440 (1964).
24. H. D. Engers, S. Shechosky, N. B. Madsen, *Can. J. Biochem.* **48**, 746 (1970).
25. D. J. Graves, S. A. S. Mann, G. Philip, R. J. Oliveira, *J. Biol. Chem.* **243**, 6090 (1968).
26. C. B. Newgard, K. Nakano, P. K. Hwang, R. J. Fletterick, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8132 (1986).
27. V. L. Rath, C. B. Newgard, S. R. Sprang, E. J. Goldsmith, R. J. Fletterick, *Proteins* **2**, 225 (1987).
28. M. Kobayashi, G. Soman, D. J. Graves, *J. Biol. Chem.* **257**, 14041 (1982).
29. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
30. J. Kraut and L. Jensen, *Acta Crystallogr.* **16**, 79 (1963).
31. We thank L. N. Johnson and D. Barford for making the coordinates of sulfate-activated phosphorylase b available to us (now submitted to the Protein Data Bank as entry 9GPB), to S. Shechosky for excellent technical assistance, to N.-H. Xuong for making the facilities of the Protein Crystallography Research Resource at UCSD available for data collection. Supported in part by NIH research grant. RO1 DK31507 to S.R.S. and RO1 DK26081 to R.J.F. The coordinates of glucose-inhibited phosphosphorylase a and PLPP-GPb have been submitted to the Protein Data Bank.

24 June 1991; accepted 10 September 1991

Characterization of a Zinc Finger Gene Disrupted by the t(15;17) in Acute Promyelocytic Leukemia

AUDREY D. GODDARD, JULIAN BORROW, PAUL S. FREEMONT, ELLEN SOLOMON*

The translocation t(15;17) associated with acute promyelocytic leukemia results in the fusion of the retinoic acid receptor alpha (*RARA*) gene to the *PML* gene. Characterization of *PML* revealed that it is a putative zinc finger protein and potential transcription factor that is commonly expressed, with at least three major transcription products. *PML* breakpoints cluster in two regions on either side of an alternatively spliced exon. Although leukemic cells with translocations characteristically express only one fusion product, both *PML/RARA* (on the 15q+ derivative chromosome) and *RARA/PML* (on the 17q− derivative) are transcribed.

ACUTE PROMYELOCYTIC LEUKEMIA (APL) is characterized cytogenetically by a consistent translocation between chromosomes 15 and 17, t(15;17)(q21;q11.2-12). Both products of the reciprocal translocation are found in these cells (*1*). Investigators have shown that it is the *RARA* gene on chromosome 17 that is disrupted by the translocation (*2, 3*). APL represents an example of a malignancy that is due to a failure of differentiation in which a block in the granulocyte-macrophage pathway leads to the accumulation of abnormal promyelocytes. Treatment of APL promyelocytes with all-trans retinoic acid (RA) in vitro overcomes the block, and this agent has been effective in inducing remission in APL patients (*4*). This observation supports the concept of a role for the RA receptor in the etiology of this disease but leaves open the question of the role of the translocated gene *PML* [formerly *myl* (*3*)] on chromosome 15.

To isolate *PML* cDNA clones, we used a genomic clone (fqp12) (*2*) spanning the t(15;17) breakpoint. With the chromosome 15–specific portion of this clone, we expanded our coverage of the region into 40 kb of genomic DNA (Fig. 1A). Using a genomic fragment (cos15-11/H3.9) as a probe (Fig. 1A), we isolated a total of nine independent overlapping cDNA clones (*5*). Using these cDNAs as probes, we first determined that the normal *PML* locus gives rise to four main transcripts visible on Northern (RNA) blots (Fig. 2). The three largest (4.5, 3.8, and 3.0 kb) are more abundant than the smallest message of 2.1 kb. Northern blots of polyadenylated [poly(A)$^+$] RNA and slot blots of total RNA from a variety of fetal and adult tissues indicated that *PML* was expressed in all tissues examined, including adult brain, gut, liver, lung, muscle, placenta, and testes and fetal brain, gut, liver, and muscle (*6, 7*). The cell lines Daudi (B cell), Molt4 (T cell), HL60 (myeloid), U937 (myeloid) and NB100 (neuroblastoma) were also found to express *PML* (*3, 7*). The four major bands were observed in all RNA samples analyzed on Northern blots. However, some variation in the relative intensity of bands was observed, suggesting some differential regulation is exerted in different tissues.

The nine cDNA clones were sequenced, revealing that three of the four major messages seen on Northern blots were represented in the cDNA clones and that the size differences could be accounted for by alternative 3′-ends. We have designated these *PML-1*, *PML-2*, and *PML-3* (Figs. 1B and 3). The longest of the messages (*PML-1*) is 4463 bp in length and contains a long open-reading frame extending from the initiation codon at position 81 to the first stop codon at position 2261 (*8*). The sequence of the alternative 3′-end (*PML-2*; Fig. 1B) deviates from *PML-1* after codon 548 (they share 1724 bp of cDNA sequence) and represents a transcript 3640 bp in length. The profile of transcripts produced by the *PML* locus is further complicated by the use of alternative splicing. The third class of cDNA clone, *PML-3*, which encodes a 3.0-kb transcript, arises as the result of exclusion of 641 bp from the *PML-2* transcript (base pairs 1725 to 2365) (Fig. 1B) and makes use of an open-reading frame in the *PML-2* 3′-untranslated (UTR) sequences.

The three main classes of transcripts should all encode proteins (Fig. 3) in which the first 548 amino acids are identical. However, 144 bp of alternatively spliced sequence are in this common sequence, and exclusion of this exon from the transcripts results in a loss of 48 amino acids from this part of the putative protein (Figs. 1B and 3A). Beyond residue 548, the three COOH-termini differ. An additional 312 amino acids are encoded in *PML-1*, 41 in *PML-2*, and 254 in *PML-3*. In addition, a third alternatively spliced exon removes 17 amino acids from the *PML-1*–specific COOH-terminal sequences (Fig. 1B).

Whereas the putative *PML* protein sequences do not exhibit a great degree of overall similarity to any known proteins (*9*), the possibility that *PML* protein is involved in transcriptional regulation is suggested by the presence of three clusters of cysteines located towards the NH_2-terminus of all the

A. D. Goddard, J. Borrow, E. Solomon, Somatic Cell Genetics Laboratory, Imperial Cancer Research Fund, London WC2A 3PX, United Kingdom.
P. S. Freemont, Protein Structure Laboratory, Imperial Cancer Research Fund, London WC2A 3PX, United Kingdom.

*To whom correspondence should be addressed.

PML proteins, followed by an amphipathic α helix. Amino acids 35 to 69 appear similar to a family harboring a Zn finger–like motif (*10*) that has been observed towards the NH_2-termini of a number of proteins (Fig. 4). This Zn finger–like motif is found in two transcription factors: the human Rpt-1 protein, which regulates expression of the interleukin-2 receptor and the human immunodeficiency virus–1 genes, and the herpes simplex virus type 1 immediate early protein IE110, which is involved in transcriptional activation of viral late genes (*11*). *MEL-18*, a gene isolated from a murine melanoma, also encodes a protein that contains this motif and has DNA-binding capacity. The presence of the *MEL-18* protein Zn finger is required for the protein to successfully bind DNA (*12*). Furthermore, the ability of this motif to bind Zn and DNA has been demonstrated through a series of biophysical and biochemical studies, including nuclear magnetic resonance, circular dichroism, and optical spectroscopy, that have shown that the cysteine motif of the *RING1* protein (*10*) binds Zn and DNA specifically and that the tertiary structure of the motif is stabilized by Zn binding (*13*). These observations are consistent with the prediction that the proteins that contain this motif are analogous to other Zn finger–containing proteins because they require Zn for DNA binding and suggest that the *PML* protein may bind DNA in a sequence-specific manner.

The second and third cysteine-rich regions (amino acids 107 to 129 and 167 to 205) may form structures more similar to that formed by the C6 cluster characterized by GAL4 and other transcription factors (*14*) than to the Zn finger–like motif of the first cluster. The *PML* proteins do not contain a known nuclear localization signal (*15*).

The *PML* proteins are predicted to contain a long α-helical region with a distribution of amino acids suggestive of an amphipathic helix, starting at amino acid 210 and extending to position 324 (*16*). Amphipathic α helices in which all of the negatively charged residues are exposed on one side of the helix can function as trans-activation domains (*17*), or alternatively this domain could mediate dimerization with other proteins. Trans-activation functions have also been found in association with proline- or glutamine-rich regions (*18*). *PML* protein also has two proline-rich regions. The first 25 amino acids are predominantly proline (36%), and the COOH-terminus of *PML-2* protein (positions 482 to 571) is 16% proline and 20% serine, whereas the analogous regions in *PML-1* protein (amino acids 482 to 569) and *PML-3* protein (482 to 589) are also proline and serine rich.

Using a variety of probes, we have been able to show on Southern (DNA) blots that *PML* is disrupted by the translocations in all APL samples tested (*7*). However, unlike *RARA*, which is principally disrupted in one region, the breakpoints in *PML* cluster in two distinct genomic regions (Fig. 1A). This clustering of breakpoints corresponds to disruption of the *PML* transcripts at two unique points. To demonstrate this directly, we isolated fusion transcripts from two APL patients samples shown to rearrange in either breakpoint cluster region (*19*). The sequence of the fusion transcripts confirmed that the breakpoints occurred at two distinct points in the *PML* cDNA sequence: in sample APL512 between positions 1197 and 1198 and in sample APL511 between base pairs 1671 and 1672 (Figs. 1B and 3A). Presumably, these positions mark exon-exon boundaries, and as in *RARA*, the molecular rearrangement fuses *PML* and *RARA* intronic sequences.

The 5′ breakpoints at position 1198 in the cDNAs contribute 372 amino acids to the *PML/RARA* fusion, and the breakpoints occurring at position 1672 contribute 530 amino acids. The 3′ breakpoints include the alternatively spliced 144-bp exon in the *PML/RARA* product, whereas the 5′ breakpoint cluster translocates this exon to the 17q− derivative chromosome, and the potential for alternative splicing from this exon is restricted to the *RARA/PML* fusion. We found approximately equal numbers of individuals with each breakpoint (Fig. 1A).

A similar situation of two breakpoint re-

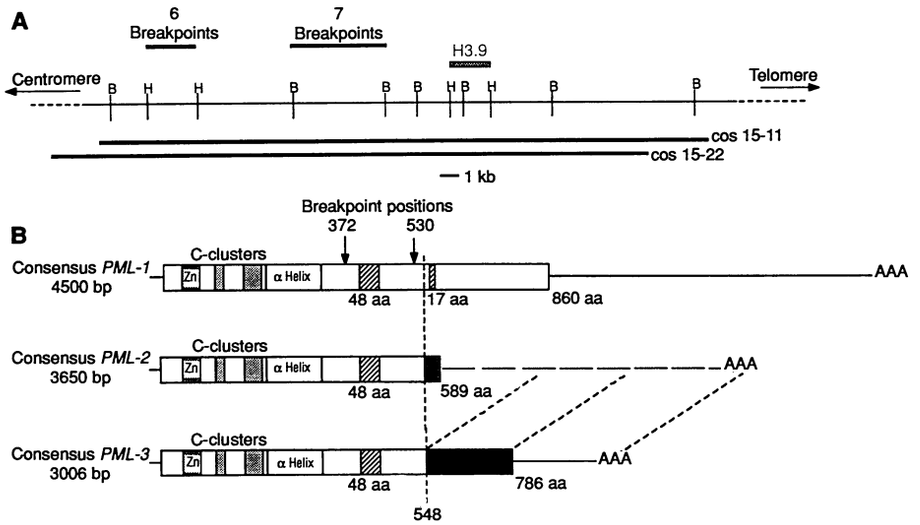

Fig. 1. (**A**) A map of the chromosome 15 genomic region rearranged in APL. The two breakpoint cluster regions and the numbers of breakpoints mapped to each region are indicated above the map. Also indicated are the two cosmids spanning this region and the location of the probe cos15-11/H3.9 we used to isolate cDNA clones from the *PML* locus. The *PML* locus is oriented with its 5′ end centromeric to its 3′ sequences. B, Bam HI; H, Hind III; the map is not complete with respect to Hind III sites. (**B**) The three classes of transcripts and their putative protein products. The location of the three cysteine (C) clusters (stipled boxes with the labeled Zn finger–like domain), the long amphipathic α helix, the breakpoints, and the alternatively spliced exons (striped boxes) are indicated. The COOH-terminal sequences unique to each protein diverge from one another after residue 548, and the lengths of the predicted proteins [in amino acids (aa)] are given at the end of the boxed open-reading frame for each class of transcript. The *PML-3* COOH-terminus is derived from the *PML-2* 3′-UTR. The cDNA sequence from nucleotides 1725 to 2365 is excluded from the *PML-2* consensus to produce the *PML-3* message so that the *PML-3* COOH-terminus is encoded by an open-reading frame from nucleotides 2366 to 3126 in the *PML-2* cDNA sequence.

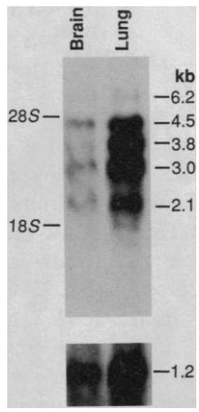

Fig. 2. Northern blot of *PML*-specific transcripts in adult brain and lung poly(A)$^+$ RNA. The upper panel shows the *PML* transcripts that were detected with a cDNA probe (pAGU3). The sizes of the four main transcripts are indicated along with other reference points. The lower panel shows the same samples hybridized to a control probe (pED1) (*22*). Total RNA was extracted by the guanidinium thiocyanate–cesium chloride method (*23*). Poly(A)$^+$ RNA was isolated by the Micro-FastTrack mRNA isolation kit (Invitrogen) or purchased from Cambridge Bioscience (Cambridge, United Kingdom) and size-fractionated through a 1% formaldehyde–agarose gel, followed by transfer to Hybond N membranes (Amersham) according to the manufacturer's instructions. Filters were prehybridized and hybridized to ^{32}P-labeled cDNA probes at 65°C in 0.5 M $NaPO_4$ (pH 7.2), 7% SDS, and 1 mM EDTA (*24*). The final washing was performed in 0.04 M $NaHPO_4$, 1% SDS, and 1 mM EDTA at 65°C.

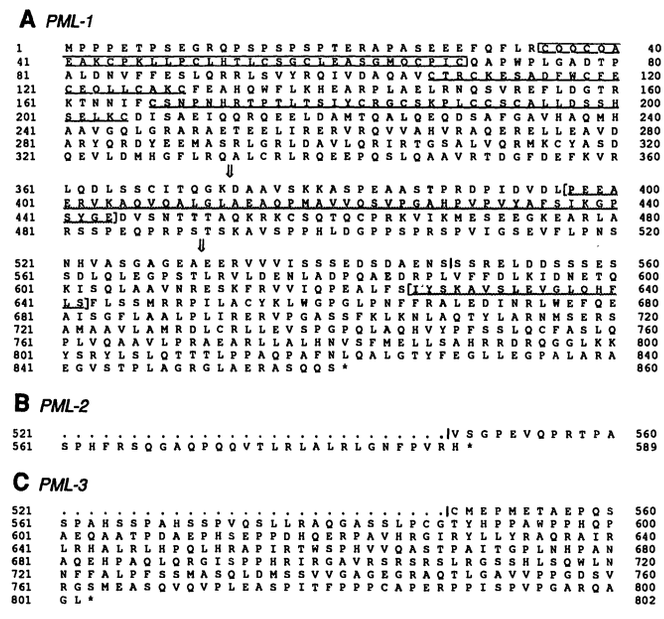

Fig. 3. Predicted amino acid sequence of the *PML* cDNA clones (*25, 26*). Sequencing of *PML* cDNAs and PCR products was performed by the dideoxy-chain termination method with the Sequenase enzyme and kit (United States Biochemical). Primers specific to the vectors (pCDM8 and pBluescript) were used initially, and sequence contigs were completed with insert-specific oligonucleotides. DNA sequence data was assembled, analyzed, and translated with the Intelligenetics Programs (Release 5.4) and the GCG Sequence Package Version 6.2 (*27*). (**A**) Amino acid sequence predicted for the *PML-1* transcript. The Zn finger–like domain is boxed, and the other cysteine-rich regions are underlined with a straight line. The amino acids encoded by the alternatively spliced exons are underlined and enclosed in brackets. The positions of the two breakpoint regions at amino acids 372 and 530 (corresponding to nucleotide positions 1197 and 1671 in the cDNA) are indicated by arrows above the sequence. The point beyond which the different cDNA classes diverge (after nucleotide position 1724, codon 548) is indicated by a vertical line after the last residue in common to all three classes. The termination codon is indicated by an asterix. (**B**) Alternative COOH-terminus predicted for the *PML-2* cDNA. The 41 amino acids unique to this isoform are given after the vertical line. (**C**) Alternative COOH-terminus predicted for the *PML-3* cDNA. The 254 amino acids unique to this isoform are given after the vertical line.

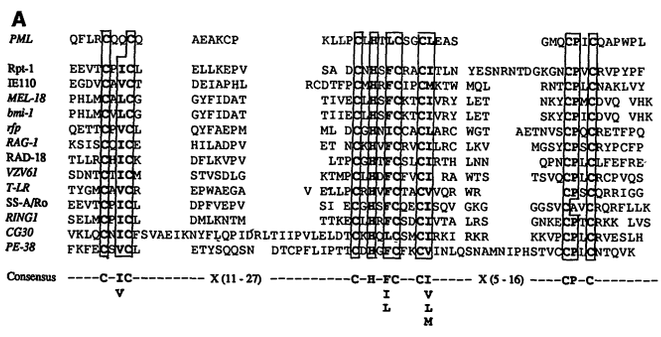

Fig. 4. (**A**) The Zn finger–like motif in the *PML* protein shows similarity to a number of proteins or known or suspected gene products to interact with DNA (*26*). Abbreviations used are as follows: *RAG-1*, the gene product of the human V(D)J recombinase-activating gene *RAG-1*; the *Saccharomyces cerevisiae* postreplication repair protein RAD-18; the varicella zoster virus *VZV61* gene; the baculovirus genes *CG30* and *PE-38*;the trypanosome *T-LR* gene; the human SS-A/Ro 52-kD antigen associated with several rheumatic diseases; *RING1*, a human gene located proximal to the human major histocompatability complex (*28*); two human genes implicated in transformation, *rfp* and *bmi-1* (*29*); and X, any amino acid. Numbers indicate variable numbers of amino acids. (**B**) A representation of the Zn finger–like domain showing the formation of two adjacent Zn fingers from the consensus sequence in (A).

gions exists for the t(9;22) Philadelphia chromosome in which two distinct forms of fusion product are produced, $p210^{bcr/abl}$ and $p190^{bcr/abl}$, as the consequence of the use of the major or minor breakpoint cluster regions in *bcr*. In this case, the two forms are predominantly seen in different leukemias (*20*). Approximately 25% of individuals with APL display features of the variant or microgranular form of APL. It will be interesting to see if use of one breakpoint region or the other will correlate with the diagnosis of variant APL.

Cytogenetically, APL is characterized by the presence of both translocation products (the 15q+ and 17q− derivatives) (*1*), suggesting that both chromosomes are important for leukemogenesis. The 15q+ fusion message contains the *PML* promoter and potential DNA binding and dimerization–trans-activation domains fused to the *RARA* sequences encoding the *RARA* DNA and RA binding and dimerization domains. We have isolated and sequenced a full-length 17q− transcript, in which the 5′ sequences of *RARA*, including the promoter and exons coding for the NH_2-terminal transactivation domain, are fused to the *PML-2* 3′ sequences. Both the *PML/RARA* and the *RARA/PML* fusions are in frame. Both the *PML* and *RARA* promoters are active in hematopoietic cell lines; thus, it is not unexpected that both the *PML/RARA* and *RARA/PML* fusion products are transcribed in APL. In other hematopoietic malignancies associated with the production of fusion proteins, only one of the two derivative chromosomes is always maintained in the leukemic clone. Thus, in chronic myelogenous leukemia only the *bcr/abl* chimera is present, and in pre-B acute lymphocytic leukemia the *E2A/PBX* fusion is observed. The maintenance of both rearranged alleles in APL suggests that both may be important in the etiology of the malignant clone.

The contribution made by *PML* to the oncogenicity of the fusion products is emphasized by three observations. First, no mutations affecting *RARA* alone have been observed in the twenty APLs we have analyzed (*2, 7*). Second, two APLs cytogenetically lacking t(15;17) chromosomes have been shown to carry rearrangements of both *PML* and *RARA* (*7*). Third, *PML*, and not *RARA*, has been molecularly rearranged in a variant APL translocation in which chromosome 15 has been translocated to another chromosome with no visible involvement of chromosome 17 (*21*). These results support a role for the rearrangement of *PML* in the etiology of APL.

Note added in proof: Since the submission of this manuscript, other workers have published similar observations (*21a*).

REFERENCES AND NOTES

1. R. A. Larson *et al.*, *Am. J. Med.* **76**, 827 (1984); D. Sheer *et al.*, *Br. J. Cancer* **52**, 55 (1985).
2. J. Borrow, A. D. Goddard, D. Sheer, E. Solomon, *Science* **249**, 1577 (1990).
3. H. de The *et al.*, *Nature* **347**, 558 (1990).
4. T. R. Breitman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2936 (1980); T. R. Breitman *et al.*, *Blood* **57**, 1000 (1981); M. Huang *et al.*, *ibid.* **72**, 567 (1988); S. Castaigne *et al.*, *ibid.* **76**, 1704 (1990); C. Chomienne *et al.*, *ibid.*, p. 1710; R. P. Warrell *et al.*, *N. Engl. J. Med.* **324**, 1385 (1991).
5. DNA clones for *PML-1*, *PML-2*, and *PML-3*, together with 15q+ and 17q− APL fusion clones, were isolated from cDNA libraries constructed in the plasmid vector pCDM8 [B. Seed, *Nature* **329**, 860 (1987)] from phorbol 12-myristate 13-acetate–stimulated HL60, U937, chronic myelogenous leukemia bone marrow, and APL (APL512) blood cells (from D. Simmons, Imperial Cancer Research Fund, Oxford). Replicates of 5×10^5 colonies were screened with oligolabeled cos15-11/H3.9 or *PML* cDNA clone pAGU3, and the colonies were purified through additional rounds of screening.
6. The slot blots were a gift from Y. Edwards (University College, London) and adult testes total RNA from R. Hawkins (Imperial Cancer Research Fund, London).
7. A. D. Goddard and J. Borrow, unpublished observations.

8. The second ATG is in a more favorable environment with eight out of nine positions (CCACCATGC) matching the Kozak consensus sequence (CC)ACCATGG [M. Kozak, *Cell* **44**, 283 (1986)] as compared to a match at six of the nine positions (GGTCCATGG) for the more 5′ ATG. 5′ rapid amplification of cDNA ends [M. A. Frohman, M. K. Dush, G. R. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998 (1988)] performed on 1 μg of APL and HL60 total RNA suggested that we were lacking less than 50 to 100 bp of 5′ sequence. Also, our estimate of the size of the largest, abundant product on Northern blots is 4.5 kb, close to the 4.46 kb contained in the consensus sequence.
9. We used the *PML* product amino acid sequences to search the OWL10.0 protein database [D. Akrigg *et al.*, *Nature* **335**, 745 (1988)] with the program PROSRCH [J. F. Collins, A. F. W. Coulson, A. Lyall, *Comput. Appl. Biosci.* **4**, 67 (1988)] on a distributed array processor.
10. P. S. Freemont, I. M. Hanson, J. Trowsdale, *Cell* **64**, 483 (1991).
11. L. J. Perry *et al.*, *J. Gen. Virol.* **67**, 2365 (1986); R. Patarca *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2733 (1988).
12. M. Tagawa *et al.*, *J. Biol. Chem.* **265**, 20021 (1990).
13. P. S. Freemont *et al.*, unpublished observations.
14. R. M. Evans and S. M. Hollenberg, *Cell* **52**, 1 (1988).
15. We identified consensus motifs in *PML* by searching with the PROSITE database using the program PROMOT [M. J. E. Sternberg, *Comput. Appl. Biosci.* **7**, 257 (1991)].
16. Predictions of secondary structure were made with the combined algorithm described in E. E. Eliopoulos, A. J. Geddes, M. Bretl, D. J. C. Pappin, and J. B. C. Findley [*Int. J. Biol. Macrmol.* **4**, 263 (1982)]. The α-helical region was predicted by eight separate methods and confirmed by the algorithm produced by A. Lupas, M. Van Dyke, and J. Stock [*Science* **252**, 1162 (1991)].
17. E. Giniger and M. Ptashne, *Nature* **330**, 670 (1987).
18. N. Mermod, E. A. O'Neill, T. J. Kelly, R. Tjian, *Cell* **58**, 741 (1989).
19. A cDNA library was constructed by D. Simmons from poly(A)$^+$ selected RNA from a blood sample from individual APL512. The library was screened with both *RARA*- and *PML*-specific probes, and the fusion cDNA clones purified. A second fusion message was isolated from individual APL511 blood by the combination of reverse transcription from a *RARA*-specific primer and polymerase chain reaction amplification with both *PML* and *RARA* primers, followed with subcloning in pBluescript (Stratagene). Reverse transcription of APL total RNA and subsequent PCR amplification were performed as described in J. Borrow *et al.* (in preparation).
20. E. Fainstein *et al.*, *Nature* **330**, 386 (1987); A. Hermans *et al.*, *Cell* **51**, 33 (1987); S. S. Clark *et al.*, *Science* **239**, 775 (1988); L. Selleri *et al.*, *Blood* **75**, 1146 (1990).
21. The translocation breakpoint in the variant APL translocation was detected with the *PML* cDNA probe pAGU3. No abnormalities were observed in *RARA* when we used the genomic probes LCN4A3/A, /B, and /C (*2*) on Bam HI and Eco RI digests. LCN4A3/D was also used, which detects the Bam HI fragment separating LCN4A3/A and /C, and p124/BE1.4, which hybridizes to the Eco RI and Bam HI fragments at the 5′ end of *RARA* (*2*). These five genomic probes cover the entire region of *RARA* in which APL translocations have been seen to occur.

21a. A. Kakizuka *et al.*, *Cell* **66**, 663 (1991); H. de The *et al.*, *ibid.*, p. 675; P. P. Pandolfi *et al.*, *Oncogene* **6**, 1285 (1991).

22. J. Squire *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6573 (1986).
23. J. M. Chirgwin *et al.*, *Biochemistry* **18**, 5294 (1979).
24. G. M. Church and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1991 (1984).
25. Only the protein sequences encoded by the cDNA clones are provided. The PML-1, PML-2, and PML-3 nucleotide sequences have been deposited with the European Molecular Biology Laboratory–GenBank database (accession numbers M79462, M79463, and M79464, respectively).
26. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
27. J. Devereux, P. Haeberli, O. Smithies, *Nucleic Acids Res.* **12**, 387 (1984).
28. A. J. Davison and J. E. Scott, *J. Gen. Virol.* **67**, 1759 (1986); J. S. Jones *et al.*, *Nucleic Acids Res.* **16**, 7119 (1988); D. G. Schalz *et al.*, *Cell* **59**, 1035 (1989); M. Takahashi and G. M. Cooper, *Mol. Cell. Biol.* **7**, 1378 (1987); M. Takahashi *et al.*, *ibid.* **8**, 1853 (1988); S. M. Thiem and L. K. Miller, *J. Virol.* **63**, 4489 (1989); R. Krappa and D. Knebel-Morsdorf, *ibid.* **65**, 805 (1990); B. L. Smiley *et al.*, *Mol. Cell. Biol.* **10**, 6436 (1990); E. K. L. Chan *et al.*, *J. Clin. Invest.* **87**, 68 (1991).
29. Y. Haupt *et al.*, *Cell* **65**, 753 (1991); M. van Lohuizen *et al.*, *ibid.*, p. 737.
30. We thank I. Goldsmith (Imperial Cancer Research Fund, London) for synthesis of oligonucleotides and J. Cowell (Imperial Cancer Research Fund, London) for donation of tissue for mRNA extraction.

15 July 1991; accepted 19 September 1991

The N-End Rule in Bacteria

JOHN W. TOBIAS, THOMAS E. SHRADER, GABRIELLE ROCAP, ALEXANDER VARSHAVSKY

The N-end rule relates the in vivo half-life of a protein to the identity of its amino-terminal residue. Distinct versions of the N-end rule operate in all eukaryotes examined. It is shown that the bacterium *Escherichia coli* also has the N-end rule pathway. Amino-terminal arginine, lysine, leucine, phenylalanine, tyrosine, and tryptophan confer 2-minute half-lives on a test protein; the other amino-terminal residues confer greater than 10-hour half-lives on the same protein. Amino-terminal arginine and lysine are secondary destabilizing residues in *E. coli* because their activity depends on their conjugation to the primary destabilizing residues leucine or phenylalanine by leucine, phenylalanine–transfer RNA–protein transferase. The adenosine triphosphate–dependent protease Clp (Ti) is required for the degradation of N-end rule substrates in *E. coli*.

AMONG THE FUNCTIONS OF PROtein degradation in living cells are the elimination of abnormal proteins and the temporal control of many cellular processes that involve short-lived regulators (*1*). The metabolic instability of a regulatory protein allows for rapid adjustment of its intracellular concentration through changes in the rate of its synthesis or degradation. Features of a protein that confer metabolic instability are called degradation signals. An essential component of one degradation signal is the protein's N-terminal residue (*2*). The presence of this signal, named the N-degron (*3*), is manifested as the N-end rule, which relates the metabolic stability of a protein to the identity of its N-terminal residue. Distinct versions of the N-end rule operate in all eukaryotes examined, from yeast to mammals (*2–8*). The eukaryotic N-degron comprises two distinct determinants: a destabilizing N-terminal residue and a specific internal lysine residue (or residues) (*5, 7*). The latter is the site of formation of a substrate-attached multiubiquitin chain that is required for degradation of the substrate (*7*).

The N-end rule is organized hierarchically: N-terminal Asp and Glu (and Cys in mammalian reticulocytes) are secondary destabilizing residues in that they are destabilizing through their conjugation, by Arg-tRNA–protein transferase, to Arg, one of the primary destabilizing residues (*2, 6, 8, 9*). N-terminal Asn and Gln are tertiary destabilizing residues in that they are destabilizing through their conversion, by a specific deamidase, into the secondary destabilizing residues Asp and Glu (*6*).

In eukaryotes, linear ubiquitin (Ub) fusions to a test protein such as β-galactosidase (βgal) are deubiquitinated by Ub-specific processing proteases irrespective of the identity of the residue at the Ub-βgal junction (*2*). This finding has made it possible to expose in vivo different residues at the N-termini of otherwise identical test proteins. The Ub fusion technique has so far been inapplicable in bacteria, which lack Ub and Ub-specific enzymes (*10*). The isolation of *UBP1*, a gene for a Ub-specific processing protease of *Saccharomyces cerevisiae* (*10*), has allowed us to bypass this difficulty and to test for the presence of the N-end rule in bacteria.

When expressed in *E. coli* (*11*), the yeast Ubp1 protease efficiently deubiquitinated Ub–X-βgals (X is the junctional residue) both in *E. coli* extracts (*10*) and in vivo (Fig. 1). As expected (*2, 10*), Ub–X-βgals retained their N-terminal Ub moieties in con-

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.