Volume 15 Number 5 1987 Nucleic Acids Research

Characterization of cDNA clones for human myeloperoxidase: predicted amino acid sequence and evidence for multiple mRNA species

Keith R.Johnson, William M.Nauseef[1], Alessandra Care, Margaret J.Wheelock, Sara Shane, Sheila Hudson, H.Phillip Koeffler[2], Michael Selsted[2], Carl Miller[2] and Giovanni Rovera

Wistar Institute of Anatomy and Biology, 36th Street at Spruce, Philadelphia, PA 19104, [1]Department of Medicine, Veterans Administration Medical Center and the University of Iowa, Iowa City, IA 52242 and [2]UCLA School of Medicine, Los Angeles, CA 90024, USA

Received December 3, 1986; Revised and Accepted January 28, 1987

ABSTRACT

Myeloperoxidase is a component of the microbicidal network of polymorphonuclear leukocytes. The enzyme is a tetramer consisting of two heavy and two light subunits. A large proportion of humans demonstrate genetic deficiences in the production of myeloperoxidase. As a first step in analyzing these deficiencies in more detail, we have isolated cDNA clones for myeloperoxidase from an expression library of the HL-60 human promyelocytic leukemia cell line. Two overlapping plasmids (pMPO2 and pMPO62) were identified as myeloperoxidase cDNA clones based on the detection with myeloperoxidase antiserum of 1) 70 kDa protein expressed in pMPO2-containing bacteria and 2) a 75 kDa polypeptide produced by hybridization selection and translation using pMPO62 and HL-60 RNA. Formal identification of the clones was made by matching the predicted amino acid sequences with the amino terminal sequences of the heavy and light subunits. Both subunits are encoded by one mRNA in the following order: pre-pro-sequences -- light subunit -- heavy subunit. The molecular weight of the predicted primary translation product is 83.7 kDa. Northern blots reveal two size classes of hybridizing RNAs (approximately 3.0-3.3 and 3.5-4.0 kilobases) whose expression is restricted to cells of the granulocytic lineage and parallels the changes in enzymatic activity observed during differentiation.

INTRODUCTION

Myeloperoxidase (EC 1.11.1.7) is an enzyme found in the azurophilic granules (or primary lysosomes) of human polymorphonuclear leukocytes (PMN). When stimulated, PMNs produce hydrogen peroxide and concomitantly release the contents of their granules. In the presence of the hydrogen peroxide and the chloride ions in the extracellular milieu, myeloperoxidase catalyzes the production of the potent microbicidal agent hypochlorous acid. This system has been demonstrated to play an important role in human defense against microorganisms (1,2).

When isolated from peripheral blood PMN, myeloperoxidase is a tetramer composed of two 59 kDa and two 12 kDa subunits (3). Studies of the biosynthesis of myeloperoxidase have been aided by the use of HL-60 promyelocytic leukemia cells which express this enzyme (4-7), and when grown in suspension culture,

synthesize and process myeloperoxidase in a way very similar to that of normal bone marrow cells (8).

In vitro translation studies utilizing HL-60 RNA have demonstrated that myeloperoxidase is synthesized as a 75-80 kDa polypeptide (7-11). However, pulse-chase experiments with intact cells have shown that the first product that can be immunoprecipitated migrates at about 90 kDa because of N-linked glycosylation (7,8,11,12). Subsequent processing events produce the two different subunits, modify the oligosaccharide side chains, add the heme-like moieties necessary for catalytic activity and direct the enzyme to the azurophilic granule.

Until recently, hereditary myeloperoxidase deficiency was thought to be rare. However, automated flow cytochemistry analysis of clinical specimens has revealed a partial or total lack of the enzyme in 1 in 2000 people (13-15). PMNs from people with complete myeloperoxidase deficiency lack both the 59 kDa and 12 kDa subunits of the mature enzyme, but like normal PMNs, contain the 90 kDa protein (16). This suggests that precursors of the enzyme can be synthesized in bone marrow from at least some myeloperoxidase-deficient individuals and that the defect(s) lie in post-translational processing steps.

In this paper, we report the isolation of cDNA clones for myeloperoxidase from HL-60 cells. In addition, we report the nucleotide sequence of a 3.2 kb MPO cDNA including the entire protein coding domain. We also demonstrate, by Northern analysis, that at least two RNA species exist for MPO in both HL-60 cells and normal human bone marrow. Our results also indicate that the two RNAs differ primarily in their 3' non-coding sequences.

MATERIALS AND METHODS

Reagents

Guanidine hydrochloride, urea, ^{14}C-molecular weight standards, isopropylthiogalactoside (IPTG), Klenow fragment of DNA polymerase I, sequencing primer, M13mp 18 and 19, and nick-translation kits were from BRL, Gaithersburg, MD; enzymes for cDNA synthesis from Seikagaku America, St. Petersburg, FL and Amersham, Arlington Heights, IL; oligo(dT)-cellulose, Sepharose CL-4B, Protein A-Sepharose CL-4B, deoxy and dideoxynucleotide triphosphates and random hexanucleotides from Pharmacia, Piscataway, NJ; Zetabind from AMF-Cuno, Meridan, CT; nitrocellulose from Schleicher and Schuell, Keene, NH; ^{125}I-Protein A from ICN, Irvine, CA; radiolabeled nucleotides from New England Nuclear, Boston, MA; [^{35}S]methionine from Amersham; restriction enzymes from BRL, New England Biolabs, Beverly, MA or Boehringer-Mannheim, Indianapolis, IN; electrophoresis chemicals from BioRad, Richmond, CA; in vitro translation kit, Promega Biotec,

Madison, WI; E.coli DNA, molecular weight markers and bovine serum albumin (BSA) from Sigma, St. Louis, MO; and 5-bromo-4-chloro-3-indolyl galactoside (X-gal) from Research Organics Inc., Cleveland, OH. Other chemicals were from common laboratory suppliers.

Tissue culture

HL-60 cells were grown in RPMI 1640 supplemented with 100 U penicillin and 100 µg streptomycin per ml plus 10-15% fetal bovine serum. HL-60 cell cultures were induced to differentiate along the granulocytic lineage by addition of 1 µM retinoic acid (17) or along the monocytic lineage by addition of 16 nM 12-tetradecanoylphorbol-13-acetate (TPA, 18).

Construction of cDNA Library

Total RNA was isolated from uninduced HL-60 cells by homogenizing the washed cell pellet in 25 mM EDTA pH 8.0, 10 mM dithiothreitol containing 6 M guanidine hydrochloride (19). Poly(A)$^+$-RNA was isolated by elution from oligo(dT)-cellulose (20). Procedures for the synthesis of double-stranded cDNA, tailing and insertion into the Pst I site of pUC8 were similar to those previously described (21). After overnight growth of the transformants (23), the colonies were scraped from the plates and plasmid DNA isolated from them (22). The library was thus stored as plasmid DNA.

Screening the Library

Aliquots of the library were used to transform (23) E. coli strain JM109 (24) to ampicillin resistance. Approximately two-thirds of the resulting colonies contained inserts. The colonies were transferred to nitrocellulose filters and induced with IPTG for 2-4 h at 37°C. Lysis of induced colonies and blocking the filters with BSA followed established procedures (25). Antiserum specific for human myeloperoxidase (16) was absorbed against boiled extracts of JM109 (pUC8) and used at a final dilution of 1:250 to 1:500. After incubation with the antibody for 2 hrs at 37°C, the filters were washed and positive colonies identified using established procedures (25). Positive colonies were rescreened using the above procedures until all colonies were positive.

Colonies were screened with hybridization probes as described (26). The DNA used to make the probes was purified on glass powder (27) and labeled using nick translation. To reduce the background 0.1-1.0 µg/ml of linearized vector DNA and 0.1-1.0 µg/ml of E. coli DNA was included in the hybridization.

Analysis of Cell Proteins with Antibodies

One clone isolated based on its reactivity with antibody was incubated overnight with or without 0.5 mM IPTG. Cell pellets were lysed as described

(28) and aliquots run on sodium dodecyl sulfate (SDS) gels (29). Electrophoretic transfer of proteins to nitrocellulose and identification of bands recognized by the myeloperoxidase antiserum followed standard procedures (30).

HL-60 cell extracts were prepared by lysing washed cell pellets in SDS solubilization buffer containing 1 mM phenylmethylsulfonyl fluoride and placing the mixture in boiling water.

Nucleic Acid Blotting Procedures

Total cellular RNA was isolated from tissue culture cells as described above. Ten micrograms of RNA were separated on formaldehyde gels (31) and blotted onto Zetabind nylon membranes according to the manufacturer's instructions. Following restriction enzyme digestion, DNA was also blotted to nylon membranes. Both transferred RNA and DNA were covalently bound to the membrane by brief UV irradiation (32). Pre-hybridization, hybridization with nick-translated or oligo-labeled (33) probes and washing followed manufacturer's recommendations. Blots were exposed to either Kodak XAR or Fuji RX film.

Hybrid Selection and Immunoprecipitation

The hybrid selection procedure was similar to that described earlier (34). Briefly, linearized, denatured plasmid DNA was bound to nitrocellulose filters and incubated with 125-250 μg HL-60 RNA. Bound material was eluted with 1 mM EDTA at 100°C and precipitated with carrier tRNA. Translation of the selected messenger RNAs was done in reticulocyte lysate containing [^{35}S]methionine. Products of synthesis were analyzed on SDS gels either directly or following immunoprecipitation.

For immunoprecipitation, the translation mixture was first diluted in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA (NET) containing 1% NP-40 and 1 mg BSA per ml. This solution was incubated first with Sepharose CL-4B to remove non-specifically adherent material and then with myeloperoxidase antiserum and Protein A-Sepharose CL-4B. The Protein A-Sepharose and bound immune complexes were washed 5 times in NET before solubilization for gel electrophoresis. SDS gels containing products of *in vitro* translations were fixed in methanol/acetic acid, embedded with 2,5-diphenyloxazole and exposed to Kodak XAR film (35).

Determination of amino terminal sequences of MPO light and heavy subunits

MPO was purified from peripheral blood PMN as described previously (36). 2 mg MPO was reduced with dithiothreitol and carboxymethylated with iodoacetic acid (37). The light and heavy subunits were purified initially by gel filtration on a TSK-PW 3000 column (Bio-Rad) employing 0.1% trifluoroacetic acid in 36% acetonitrile as the elution solvent. Each subunit was further purified by reversed phase HPLC on a C-4 column (1 X 25 cm, Vydac) employing a linear

gradient of acetonitrile (0-75%, developed at 1% per minute).

Purified subunits were lyophilized and quantitated by amino acid analysis. Sequence determinations were obtained on an Applied Biosystems 470A instrument as described previously (37). Sequence data were derived from two independent determinations of the light subunit and from one analysis of the heavy subunit.

DNA Sequencing

Restriction fragments were subcloned into suitably digested M13 vectors as described (38,39) and sequenced utilizing standard dideoxy chain termination (40,41). Except for the extreme 3' terminus, both strands of the cDNA were completely sequenced. Initially, sequence data was analyzed with SEQPROG (Roger Larsen, University of Minnesota) on an Apple II computer. The complete

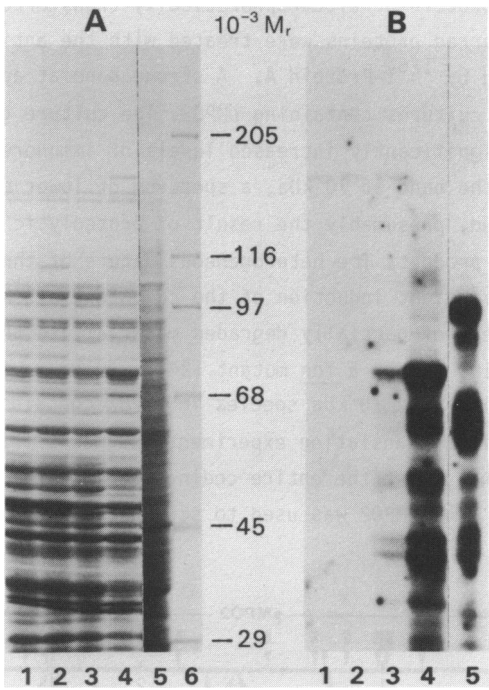

Figure 1. Detection of myeloperoxidase in bacterial lysates and in lysates of HL-60 cells. 100 μg of protein from each bacterial lysate and from an HL-60 cell lysate were run on 7% SDS gels. The gels were either stained with Coomassie R-250 (A) or transferred to nitrocellulose and probed with anti-myeloperoxidase serum and ^{125}I-Protein A (B). The lysates are: JM109 (pUC8) incubated without (lane 1) or with (lane 2) IPTG; JM109 (pMPO2) incubated without (lane 3) or with (lane 4) IPTG; HL-60 cell lysate (lane 5). The molecular weight markers (lane 6) are myosin, β-galactosidase, phosphorylase b, BSA, ovalbumin and carbonic anhydrase. The autoradiogram (B) was exposed for 4 days.

sequence was then compared against the EMBL, NIH Genbank and NBRF protein databases. Further analysis was done with the IntelliGenetics system package.

RESULTS AND DISCUSSION

To identify myeloperoxidase cDNAs, we constructed an expression library of HL-60 cells in pUC8. Following induction with IPTG, colonies of interest were identified with the myeloperoxidase antiserum. Approximately 100,000 colonies with inserts were examined in the primary screen. One clone (pMPO2) remained positive on repeated screenings.

Cultures of JM109 containing both the plasmid vector and pMPO2 were grown overnight in the presence or absence of IPTG. Lysates were prepared from each, and aliquots of approximately 100 μg of protein were run on SDS gels which were either stained for protein or electrophoretically transferred to nitrocellulose (Figure 1). Transferred proteins were treated with the antibody used to screen the library followed by ^{125}I-Protein A. A strong band at approximately 70 kDa was resolved in the cultures containing pMPO2. The culture grown in the presence of IPTG had significantly increased levels of immunoreactive material.

In addition to the band at 70 kDa, a spectrum of lower molecular weight products was observed, presumably the result of proteolytic breakdown of the primary translation product. The heterogeneous nature of these products may be due to the use of overnight induction of the cultures, allowing the accumulation of a wide variety of partially degraded material. In this regard, it may be relevant that JM109 is not a *lon* mutant (24).

The identification of a 70 kDa species in pMPO2-containing *E. coli* and of a 75 kDa species in vitro translation experiments suggests that the insert in pMPO2 (2.2 kilobases) spans almost the entire coding sequence for myeloperoxidase.

When the insert from pMPO2 was used to screen the HL-60 cDNA library,

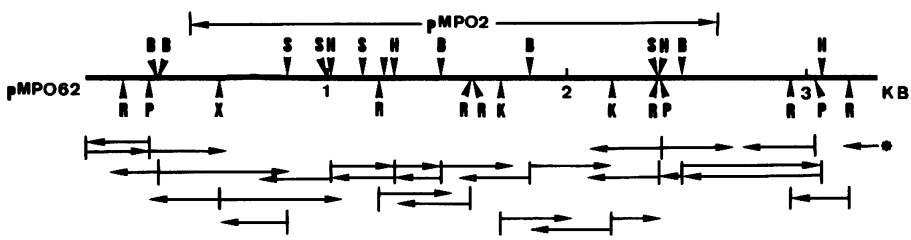

Figure 2. Partial restriction map and sequencing strategy of the insert contained in pMPO62. The following restriction enzyme sites are indicated: *Bal I* (B), *Hinc II* (H), *Kpn I* (K), *Pst I* (P), *Rsa I* (R), *Sma I* (S) and *Xba I* (X). The asterisk (*) indicates the portion of the sequence obtained from a genomic clone. The antibody positive clone pMPO2 extends from 0.4 to 2.6 kb on the map.

several additional clones were identified. Positive clones appeared in this primary screen at a frequency of 0.1%. One such clone (pMP062) was selected for further analysis. A partial restriction map of these two clones is presented in Figure 2.

The plasmid pMP062 was used for hybrid selection experiments with HL-60 RNA. The results of this analysis are shown in Figure 3. The clone pMP062 selected mRNA encoding a polypeptide that migrated at approximately 75,000 daltons on SDS gels, consistent with other reports describing the MPO primary translation product (7-11). This polypeptide was also immune precipitated by the MPO specific antiserum.

The strategy used to sequence the insert from pMP062 is indicated in Figure 2. The sequence obtained is shown in Figure 4. The N-terminal amino acid sequences of the light and heavy subunits of MPO are shown underlined in Figure

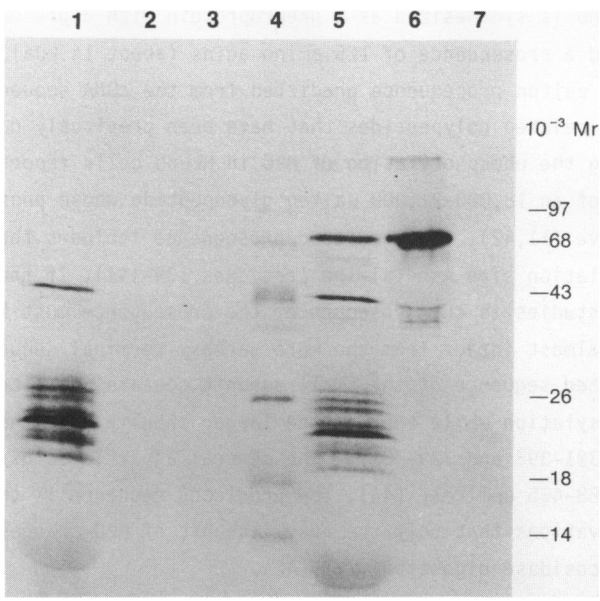

Figure 3. Hybridization-selection with pMP062 and HL-60 RNA. 125 μg RNA from HL-60 cells was incubated with 3 μg Hind III-digested pUC vector or Hind III-digested pMP062 that was bound to nitrocellulose. Selected RNA was translated in vitro in a 40 μl reaction. 3 μl of the reaction was run directly (lanes 1 and 5). One half the remainder was treated either with myeloperoxidase antiserum (lanes 2 and 6) or non-immune serum (lanes 3 and 7) followed by Protein A-Sepharose. Samples were run on 12% SDS gels and processed for fluorography (35). The autoradiogram was exposed for 11 hr. The ^{14}C-labeled molecular weight markers were phosphorylase b, BSA, ovalbumin, α-chymotrypsinogen β-lactoglobulin, and lysozyme.

4 (amino acids 167-178 and 279-301). These protein sequences agree exactly with the amino acids predicted from the nucleic acid sequence. These results prove that our clone is for myeloperoxidase.

The open reading frame in our cDNA clone extends for 745 amino acids totalling 83,747 daltons. If the sequence is broken at the N-termini of the light and heavy subunits, the three polypeptides that result have molecular weights of 18,000, 12,700 and 53,200 daltons. The latter two values agree with those published for the light subunit and the Endo H treated heavy subunit of MPO (7,11,42).

The predicted N-terminus of the primary translation product includes sequences that may function as a signal sequence. A recent algorithm for predicting cleavage sites in presequences (43) identified the Ala-Thr bond at amino acids 41-42 as a likely site. If cleavage does occur here, 125 amino acids remain before the N-terminus of the small subunit is encountered. It thus appears that MPO is synthesized as a preproprotein with a presequence of 41 amino acids and a prosequence of 125 amino acids (about 14 kDa).

The 14,000 dalton prosequence predicted from the cDNA sequence is similar in size to MPO related polypeptides that have been previously described. Two groups studying the phosphorylation of MPO in HL-60 cells reported the immunoprecipitation of an 18,000-21,000 dalton glycopeptide whose phosphorylation was Endo H sensitive (11,42). The predicted prosequence includes the potential N-linked glycosylation site Asn-Val-Thr (residues 139-141). If the glycopeptide seen in these studies is the prosequence, the prosequence must be proteolytically removed almost intact from the more carboxy-terminal sequences of MPO.

The predicted sequence of the small subunit contains no sites of potential N-linked glycosylation while that of the larger subunit contains 4 sites (323-325, 355-357, 391-393 and 729-731). The central aspartic acid residue makes the residues 483-485 unlikely (44). The predicted sequence is thus consistent with the observations that only the heavy subunit of MPO changes mobility following endoglycosidase digestion (7,11,42).

We do not know the carboxy-terminal sequences of either subunit of MPO. As a first approximation, we designated amino acids 278 and 745 as the carboxy-terminal residues of the light and heavy subunits, respectively. With these assumptions, we predicted the amino acid compositions of the light subunit, the heavy subunit and the sum of the two. These predicted values agree closely with amino acid composition data in the literature (Table I).

The 5' untranslated region in the insert of pMPO62 consists of about 25 dG residues followed by 163 nucleotides. Included in this region is a short open

```
                                                                    1
                                                                    AGCTGTGGAGGTGGGGTCCTT
22
GGAAGCTGGATGACAGCAGCTGGCAAGGGGATAAGAGAGCAGTGAGCCCCTCCCTCAAGGAGGTCTGGCTT
93
TATCCATAGACAGGGCCCTCTGAGGTGGGGCTGAGGTACAAAGGGGGATTGAGCAGCCCAGGAGAAGAGAG
164                                                                     18
ATG GGG GTT CCC TTC TTC TCT TCT CTC AGA TGC ATG GTG GAC TTA GGA CCT TGC
MET Gly Val Pro Phe Phe Ser Ser Leu Arg Cys Met Val Asp Leu Gly Pro Cys
218                                                                     36
TGG GCT GGG GGT CTC ACT GCA GAG ATG AAG CTG CTT CTG GCC CTA GCA GGC GTC
Trp Ala Gly Gly Leu Thr Ala Glu Met Lys Leu Leu Leu Ala Leu Ala Gly Val
272                                                                     54
CTG GCC ATT CTG GCC▼ACG CCC CAG CCC TCT GAA GGT GCT GCT CCA GCT GTC CTG
Leu Ala Ile Leu Ala Thr Pro Gln Pro Ser Glu Gly Ala Ala Pro Ala Val Leu
326                                                                     72
GGG GAG GTG GAC ACC TCG TTG GTG CTG AGC TCC ATG GAG GAG GCC AAG CAG CTG
Gly Glu Val Asp Thr Ser Leu Val Leu Ser Ser Met Glu Glu Ala Lys Gln Leu
380                                                                     90
GTG GAC AAG GCC TAC AAG GAG CGG CGG GAA AGC ATC AAG CAG CGG CTT CGC AGC
Val Asp Lys Ala Tyr Lys Glu Arg Arg Glu Ser Ile Lys Gln Arg Leu Arg Ser
434                                                                    108
GGC TCA GCC AGC CCC ATG GAA CTC CTA TCC TAC TTC AAG CAG CCG GTG GCA GCC
Gly Ser Ala Ser Pro Met Glu Leu Leu Ser Tyr Phe Lys Gln Pro Val Ala Ala
488                                                                    126
ACC AGG ACG GCG GTG AGG GCC GCT GAC TAC CTG CAC GTG GCT CTA GAC CTG CTG
Thr Arg Thr Ala Val Arg Ala Ala Asp Tyr Leu His Val Ala Leu Asp Leu Leu
542                                                                    144
GAG AGG AAG CTG CGG TCC CTG TGG CGA AGG CCA TTC AAT GTC ACT GAT GTG CTG
Glu Arg Lys Leu Arg Ser Leu Trp Arg Arg Pro Phe Asn Val Thr Asp Val Leu
596                                                                    162
ACG CCC GCC CAG CTG AAT GTG TTG TCC AAG TCA AGC GGC TGC GCC TAC CAG GAC
Thr Pro Ala Gln Leu Asn Val Leu Ser Lys Ser Ser Gly Cys Ala Tyr Gln Asp
650                                                                    180
GTG GGG GTG ACT TGC CCG GAG CAG GAC AAA TAC CGC ACC ATC ACC GGG ATG TGC
Val Gly Val Thr Cys Pro Glu Gln Asp Lys Tyr Arg Thr Ile Thr Gly Met Cys
704                                                                    198
AAC AAC AGA CGC AGC CCC ACG CTG GGG GCC TCC AAC CGT GCC TTT GTG CGC TGG
Asn Asn Arg Arg Ser Pro Thr Leu Gly Ala Ser Asn Arg Ala Phe Val Arg Trp
758                                                                    216
CTG CCG GCG GAG TAT GAG GAC GGC TTC TCT CTT CCC TAC GGC TGG ACG CCC GGG
Leu Pro Ala Glu Tyr Glu Asp Gly Phe Ser Leu Pro Tyr Gly Trp Thr Pro Gly
812                                                                    234
GTC AAG CGC AAC GGC TTC CCG GTG GCT CTG GCT CGC GCG GTC TCC AAC GAG ATC
Val Lys Arg Asn Gly Phe Pro Val Ala Leu Ala Arg Ala Val Ser Asn Glu Ile
866                                                                    252
GTG CGC TTC CCC ACT GAT CAG CTG ACT CCG GAC CAG GAG CGC TCA CTC ATG TTC
Val Arg Phe Pro Thr Asp Gln Leu Thr Pro Asp Gln Glu Arg Ser Leu Met Phe
920                                                                    270
ATG CAA TGG GGC CAG CTG TTG GAC CAC GAC CTC GAC TTC ACC CCT GAG CCG GCC
Met Gln Trp Gly Gln Leu Leu Asp His Asp Leu Asp Phe Thr Pro Glu Pro Ala
974                                                                    288
GCC CGG GCC TCC TTC GTC ACT GGC GTC AAC TGC GAG ACC AGC TGC GTT CAG CAG
Ala Arg Ala Ser Phe Val Thr Gly Val Asn Cys Glu Thr Ser Cys Val Gln Gln
1028                                                                   306
CCG CCC TGC TTC CCG CTC AAG ATC CCG CCC AAT GAC CCC CGC ATC AAG AAC CAA
Pro Pro Cys Phe Pro Leu Lys Ile Pro Pro Asn Asp Pro Arg Ile Lys Asn Gln
1082                                                                   324
GCC GAC TGC ATC CCG TTC TTC CGC TCC TGC CCG GCT TGC CCC GGG AGC AAC ATC
Ala Asp Cys Ile Pro Phe Phe Arg Ser Cys Pro Ala Cys Pro Gly Ser Asn Ile
1136                                                                   342
ACC ATC CGC AAC CAG ATC AAC GCG CTC ACT TCC TTC GTG GAC GCC AGC ATG GTG
Thr Ile Arg Asn Gln Ile Asn Ala Leu Thr Ser Phe Val Asp Ala Ser Met Val
```

```
1190                                                                    360
TAC GGC AGC GAG GAG CCC CTG GCC AGG AAC CTG CGC AAC ATG TCC AAC CAG CTG
Tyr Gly Ser Glu Glu Pro Leu Ala Arg Asn Leu Arg Asn Met Ser Asn Gln Leu
1244                                                                    378
GGG CTG CTG GCC GTC AAC CAG CGC TTC CAA GAC AAC GGC CGG GCC CTG CTG CCC
Gly Leu Leu Ala Val Asn Gln Arg Phe Gln Asp Asn Gly Arg Ala Leu Leu Pro
1298                                                                    396
TTT GAC AAC CTG CAC GAT GAC CCC TGT CTC CTC ACC AAC CGC TCA GCG CGC ATC
Phe Asp Asn Leu His Asp Asp Pro Cys Leu Leu Thr Asn Arg Ser Ala Arg Ile
1352                                                                    414
CCC TGC TTC CTG GCA GGG GAC ACC CGT TCC AGT GAG ATG CCC GAG CTC ACC TCC
Pro Cys Phe Leu Ala Gly Asp Thr Arg Ser Ser Glu Met Pro Glu Leu Thr Ser
1406                                                                    432
ATG CAC ACC CTC TTA CTT CGG GAG CAC AAC CGG CTG GCC ACA GAG CTC AAG AGC
Met His Thr Leu Leu Leu Arg Glu His Asn Arg Leu Ala Thr Glu Leu Lys Ser
1460                                                                    450
CTG AAC CCT AGG TGG GAT GGG GAG AGG CTC TAC CAG GAA GCC CGG AAG ATC GTG
Leu Asn Pro Arg Trp Asp Gly Glu Arg Leu Tyr Gln Glu Ala Arg Lys Ile Val
1514                                                                    468
GGG GCC ATG GTC CAG ATC ATC ACT TAC CGG GAC TAC CTG CCC CTG GTG CTG GGG
Gly Ala Met Val Gln Ile Ile Thr Tyr Arg Asp Tyr Leu Pro Leu Val Leu Gly
1568                                                                    486
CCA ACG GCC ATG AGG AAG TAC CTG CCC ACG TAC CGT TCC TAC AAT GAC TCA GTG
Pro Thr Ala Met Arg Lys Tyr Leu Pro Thr Tyr Arg Ser Tyr Asn Asp Ser Val
1622                                                                    504
GAC CCA CGC ATC GCC AAC GTC TTC ACC AAT GCC TTC CGC TAC GGC CAC ACC CTC
Asp Pro Arg Ile Ala Asn Val Phe Thr Asn Ala Phe Arg Tyr Gly His Thr Leu
1676                                                                    522
ATC CAA CCC TTC ATG TTC CGC CTG GAC AAT CGG TAC CAG CCC ATG GAA CCC AAC
Ile Gln Pro Phe Met Phe Arg Leu Asp Asn Arg Tyr Gln Pro Met Glu Pro Asn
1730                                                                    540
CCC CGT GTC CCC CTC AGC AGG GTC TTT TTT GCC TCC TGG AGG GTC GTG CTG GAA
Pro Arg Val Pro Leu Ser Arg Val Phe Phe Ala Ser Trp Arg Val Val Leu Glu
1784                                                                    558
GGT GGC ATT GAC CCC ATC CTC CGG GGC CTC ATG GCC ACC CCT GCC AAG CTG AAT
Gly Gly Ile Asp Pro Ile Leu Arg Gly Leu Met Ala Thr Pro Ala Lys Leu Asn
1838                                                                    576
CGT CAG AAC CAA ATT GCA GTG GAT GAG ATC CGG GAG CGA TTG TTT GAG CAG GTC
Arg Gln Asn Gln Ile Ala Val Asp Glu Ile Arg Glu Arg Leu Phe Glu Gln Val
1892                                                                    594
ATG AGG ATT GGG CTG GAC CTG CCT GCT CTG AAC ATG CAG CGC AGC AGG GAC CAC
Met Arg Ile Gly Leu Asp Leu Pro Ala Leu Asn Met Gln Arg Ser Arg Asp His
1946                                                                    612
GGC CTC CCA GGA TAC AAT GCC TGG AGG CGC TTC TGT GGG CTC CCG CAG CCT GAA
Gly Leu Pro Gly Tyr Asn Ala Trp Arg Arg Phe Cys Gly Leu Pro Gln Pro Glu
2000                                                                    630
ACT GTG GGC CAG CTG GGC ACG GTG CTG AGG AAC CTG AAA TTG GCG AGG AAA CTG
Thr Val Gly Gln Leu Gly Thr Val Leu Arg Asn Leu Lys Leu Ala Arg Lys Leu
2054                                                                    648
ATG GAG CAG TAT GGC ACG CCC AAC AAC ATC GAC ATC TGG ATG GGC GGC GTG TCC
Met Glu Gln Tyr Gly Thr Pro Asn Asn Ile Asp Ile Trp Met Gly Gly Val Ser
2108                                                                    666
GAG CCT CTG AAG CGC AAA GGC CGC GTG GGC CCA CTC CTC GCC TGC ATC ATC GGT
Glu Pro Leu Lys Arg Lys Gly Arg Val Gly Pro Leu Leu Ala Cys Ile Ile Gly
2162                                                                    684
ACC CAG TTC AGG AAG CTC CGG GAT GGT GAT CGG TTT TGG TGG GAG AAC GAG GGT
Thr Gln Phe Arg Lys Leu Arg Asp Gly Asp Arg Phe Trp Trp Glu Asn Glu Gly
2216                                                                    702
GTG TTC AGC ATG CAG CAG CGA CAG GCC CTG GCC CAG ATC TCA TTG CCC CGG ATC
Val Phe Ser Met Gln Gln Arg Gln Ala Leu Ala Gln Ile Ser Leu Pro Arg Ile
2270                                                                    720
ATC TGC GAC AAC ACA GGC ATC ACC ACC GTG TCT AAG AAC AAC ATC TTC ATG TCC
Ile Cys Asp Asn Thr Gly Ile Thr Thr Val Ser Lys Asn Asn Ile Phe Met Ser
```

```
2324                                                                738
AAC TCA TAT CCC CGG GAC TTT GTC AAC TGC AGT ACA CTT CCT GCA TTG AAC CTG
Asn Ser Tyr Pro Arg Asp Phe Val Asn Cys Ser Thr Leu Pro Ala Leu Asn Leu
2378                    745
GCT TCC TGG AGG GAA GCC TCC TAGAGGCCAGGTAAGGGGGTGCAGCAGTGAGGGGTATATCTGG
Ala Ser Trp Arg Glu Ala Ser
2442
GCTGGCCAGTTGGAACCACGGAGATCTCCTTGCCCTAGATGAGCCCAGCCCTGTTCTGGGTGCAGCTGAGA
2513
AAATGAGTGACTAGACGTTCATTTGTGTGCTCATGTATGTGCGAAGTATATAAATTGGCTTTTCATGCGTG
2584
TGTGTTGTCTGAACATGGGGAGTGTTTCATGGGTTATGTGTATGTGCCATTTATGTGAGTGTGTGTTTGTG
2655
CTGATGAGAATACTGAGTATGTGGAAGGCAGCAGAGCGGACTGGTGAGGAGCACAGCTCAGGAACTAGACG
2726
TGCCTGGGTTCCAATCCTGGCTCTGTGGCTTGCTAGCTATGTGACCTTGAGCAAATTACCCTCCTTAAACA
2797
AGAGTTTTCTTCCTTGTAAATTACATCTGTCATGGTTTCTTGGAGGGCCCACTTGTATCCTCTGGTTCTTC
2868
ATTTATTGAGCACCTACTACATGCAAGGCACTGTACTAGGCGTGAGAAGCATATAGAGGCAAGAAAGAGAT
2939
ACCAAGATGCCATCTGTGTCCTGGTTAGCAGAGCTGGACCAGTGGTGCCTTGGAGGGATAAGCCAGCTGCA
3010
GCTGGGCTGTGTGGTTGACTTATGGGCCCAGCCAGCCAGGCTCAGGCCATGGCTCCCCTTTTTCTTCCTCA
3081
CCCTGATTTCTTGCTTATTCACTGAAGTTCTCCTGAAGAGGAACTGGGCCTGTTGCCCTTTCTGTACCATT
3152
AGTGCTCCCATGTTTATGATAATAAAGGCACCGTGATGGGGACCTCCACTCTGTCTGTGTCT
```

Figure 4. Sequence of the insert in pMPO62. The restriction fragments were subcloned as outlined in Figure 2 and sequenced using dideoxy termination (40). In some cases, synthetic oligonucleotides were used to prime the synthesis of complementary DNA (61). The N-terminal amino acids of the light and heavy subunits are underlined. The predicted site of signal peptidase cleavage is indicated by an arrow. The short open reading frame in the 5' leader sequence is overlined as are the consensus sequences in the 3' end of the insert. As described in the text, the extreme 3' sequences are from a genomic clone. The length of the insert in the 3' direction is thus not known exactly and is terminated following the TGTGTCT signal (48).

reading frame (nucleotides 31 to 63) that begins with the first ATG in the sequence. The second start codon (beginning with nucleotide 164) is preceded in position -3 and followed in position +4 by purines. Both of the features are characteristic of eukaryotic initiation codons (46). This start codon is thus the one predicted to be used during translation of MPO.

The 3' untranslated region is about 800 nucleotides in length. The insert in pMPO62 contains a poly(A) tail in addition to about 30 dC residues. To overcome sequencing difficulties at this end of the clone, we sequenced a region of a genomic clone that spans the 3' end of the insert. The sequences 3' of the RsaI site at nucleotide 3145 are from the genomic sequence. In this region, there is one consensus polyadenylation signal AATAAA (47) starting at nucleotide 3172. This is followed 30 bpr downstream by the sequence TGTGTCT

Table I. Amino Acid Composition

Amino Acid	Heavy Subunit		Light Subunit		Total	
	Predicted	Olsen and Little (3)	Predicted	Olsen and Little (3)	Predicted	Morita et al. (45)
Ala	28	36	9	8	37	38
Asx*	(56)	51	(12)	10	(68)	69
Cys	12	9	2	3	14	15
Asp	22	--	7	--	29	--
Glu	20	--	6	--	26	--
Phe	20	26	7	6	27	28
Gly	27	35	8	7	35	36
His	5	5	1	1	6	6
Ile	26	32	2	2	28	29
Lys	12	12	2	2	14	17
Leu	52	60	9	9	61	64
Met	15	5	3	1	18	16
Asn	34	--	5	--	39	--
Pro	36	40	10	9	46	49
Gln	23	--	5	--	28	--
Arg	45	47	10	9	55	62
Ser	26	28	6	6	32	31
Thr	22	28	8	7	30	30
Val	23	30	6	5	29	30
Trp	7	3	3	1	10	9
Tyr	12	16	3	2	15	13
Glx*	(43)	49	(11)	10	(54)	56

*Predicted values for Asx and Glx are the sum of the number Asp and Asn or Glu and Gln residues, respectively.

starting at nucleotide 3207. This region is very similar to a 3' consensus sequence that has been found downstream of poly(A) addition sites and may be involved in the formation of 3' termini (48).

No significant homology was found between the sequence for MPO and any other sequence in the data bases.

Northern blots of HL-60 RNA probed with either pMPO2 or pMPO62 showed two regions of hybridization, one at 3.0-3.3 kilobases (kb) and another at 3.5-4 kb

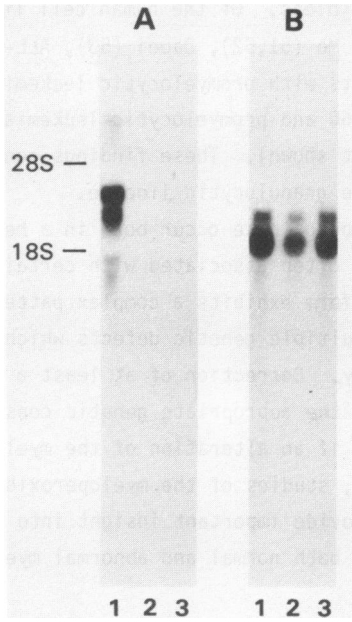

Figure 5. Northern blot analysis of RNA from HL-60 cells. 10 μg of RNA from control HL-60 cells (lane 1), cells induced with TPA (lane 2) and cells induced with retinoic acid (lane 3) were run on a formaldehyde gel and transferred to a nylon membrane. The blot was first probed with oligo-labeled pMPO2 (A) and then with phosphoglycerate kinase cDNA (B, 62). The 28S and 18S ribosomal RNA bands are indicated. The filters were autoradiographed for 6 days (A) and 1 day (B).

(Figure 5). The 3' end of the insert in pMPO2 contains a poly(A) tract in addition to the cloning tail. Preliminary sequencing indicates that the sequence ATATAAA (beginning at nucleotide 2560) may function as a polyadenylation signal with the site of poly(A) addition located about 25 bases downstream. This would result in the two messages differing in length by approximately 600 nucleotides; consistent with the Northern blotting results.

It has been shown that myeloperoxidase mRNA levels decline, as judged by *in vitro* translation, when HL-60 cells are induced to differentiate (7,9,10). In addition, myeloperoxidase activity has been found to decrease as cells are induced to form monocytic cells (6,18). Consistent with these findings, Northern blot analysis of RNA isolated from HL-60 cultures induced to form either granulocytic or monocytic cells demonstrated that the steady-state levels of myeloperoxidase mRNAs approach zero in either differentiated state (Figure 5).

RNAs from several cell lines representing different hematopoietic lineages

were analyzed by Northern blots. Of the human cell lines examined (HL-60 (4), K562 (49), Jurkat E (50), Mo (51,52), Daudi (53), ALL-1 (54), and short-term cell cultures from patients with promyelocytic leukemia or chronic myelomonocytic leukemia), only HL-60 and promyelocytic leukemia cells contained myeloperoxidase mRNAs (not shown). These findings confirm the specificity of the probe for cells of the granulocytic lineage.

Deficiencies in myeloperoxidase occur both in a hereditary form and an acquired form, the latter often associated with certain myeloid leukemias (55-60). The hereditary form exhibits a complex pattern of inheritance, perhaps a reflection of multiple genetic defects which produce the phenotype of myeloperoxidase deficiency. Correction of at least a subset of these deficiencies may be possible with the appropriate genetic constructions and transfection of bone marrow cells if an alteration of the myeloperoxidase gene can be demonstrated. Furthermore, studies of the myeloperoxidase deficiency associated with preleukemia might provide important insight into the molecular mechanisms of differentiation during both normal and abnormal myeloid development.

ACKNOWLEDGEMENTS

This work was supported by grants CA 10816, CA 21124 and CA 26038 from the National Cancer Institute. K.R.J. is a recipient of postdoctoral fellowship HD 06755 from the NIH. H.P.K. and M.E.S. are members of the Jonsson Cancer Center.

We thank A. Palumbo for his help and E. Wang and R. Ricciardi for assistance with the hybrid selection experiments. We thank M. Hoffman for editing and E. Burton for typing the manuscript.

REFERENCES

1. Schultz,J. (1980) in The Reticuloendothelial System (Sbarra, A. J., and Strauss,R.F., eds) Vol. 2, pp. 231-253, Plenum Press, New York.
2. Klebanoff,S.J. (1980) in The Reticuloendothelial System (Sbarra, A. J., and Strauss,R.F., eds) Vol. 2, pp. 279-308, Plenum Press, New York
3. Olsen,R.L. and Little,C. (1984) Biochem. J. **222**, 701-709.
4. Gallagher,R., Collins,S., Trujillo,J., McCredie,K., Ahearn,M., Tsai,S., Metzgar,R., Aulakh,G., Ting,R., Ruscetti,F. and Gallo,R. (1979) Blood **54**, 713-733.
5. Yamada,M., Mori,M. and Sugimura,T. (1981) Biochemistry **20**, 766-771.
6. Yamada,M., Mori,M. and Sugimura,T. (1981) Biochem. Biophys. Res. Commun. **98**, 219-226.
7. Nauseef,W.M. (1986) Blood **67**, 865-872.
8. Olsson,I., Persson,A.-M. and Stromberg,K. (1984) Biochem. J. **223**, 911-920.
9. Yamada,M. and Kurahashi,K. (1984) J. Biol. Chem. **259**, 3021-3025.
10. Koeffler,H.P., Ranyard,J. and Pertcheck,M. (1985) Blood **65**, 484-491.
11. Hasilik,A., Pohlmann,R., Olsen,R.L. and von Figura,K. (1984) EMBO J. **3**, 2671-2676.
12. Yamada,M. (1982) J. Biol. Chem. **257**, 5980-5982.
13. Parry,M.F., Root,R.K., Metcalf,J.A., Delaney,K.K., Kaplow,L.S and

Richar,W.J. (1981) Ann. Intern. Med. **95**, 293-301.
14. Verdot,J.J., Boyle,J., Julan,I., Aillaud,M.F., Guitard,A.M. and Muratore,R. (1981) Sem. Hop. Paris **57**, 450-457.
15. Kitihara,M., Eyre,H.J., Simonian,Y., Atkin,C.L. and Hasstedt,S.J. (1981) Blood **57**, 888-893.
16. Nauseef,W.M., Root,R.K. and Malech,H.L. (1983) J. Clin. Invest. **71**, 1297-1301.
17. Breitman,T.R., Selonick,S.E. and Collins,S.J. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 2936-2940.
18. Rovera,G., Santoli,D. and Damsky,C. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 2779-2783.
19. Woods,D., Crampton,J., Clarke,B. and Williamson,R. (1980) Nucleic Acids Res. **8**, 5157-5168.
20. Aviv,H. and Leder,P. (1972) Proc. Natl. Acad. Sci. U.S.A. **69**, 1408-1412.
21. Gubler,U. and Hoffman,B.J. (1983) Gene (Amst.) **25**, 263-269.
22. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, New York.
23. Hanahan,D. (1983) J. Mol. Biol. **166**, 557-580.
24. Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) Gene (Amst.) **33**, 103-119.
25. Helfman,D.M., Feramisco,J.R., Fiddes,J.C., Thomas,G.P. and Hughes,S.H. (1984) Focus **6** (1), 1-5.
26. Taub,F. and Thompson,E.B. (1982) Anal. Biochem. **126**, 222-230.
27. Vogelstein,B. and Gillespie,D. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 615-619.
28. Cioe,L. and Curtis,P. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 1367-1371.
29. Laemmli,U.K. and Favre,M. (1973) J. Mol. Biol. **80**, 575-599.
30. Towbin,H., Staehelin,T. and Gordon,J. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 4350-4354.
31. Lehrach H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) Biochemistry **16**, 4743-4751.
32. Church,G.M. and Gilbert,W. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 1991-1995.
33. Feinberg,A.P. and Vogelstein,B. (1984) Anal. Biochem. **137**, 266-267.
34. Ricciardi,R.P., Miller,J.S. and Roberts,B.E. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 4927-4931.
35. Bonner,W.M. and Laskey,R.A. (1974) Eur. J. Biochem. **46**, 83-88.
36. Olsson,I., Olofsson,P. and Oderberg,H. (1972) Scand. Journ. Hematol. **9**, 483-491.
37. Selsted,M.E., Harwig,S.L., Ganz,T., Schilling,J.W. and Lehrer,R.I. (1985) Jour. Clin. Invest. **76**, 1436-1439.
38. Struhl,K. (1985) BioTechniques **3**, 452-453.
39. King,P.V. and Blakesley (1986) Focus **8** (1), 1-3.
40. Sanger,F., Nickelen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463-5467.
41. Messing,J. (1983) Meth. Enzymol. **101**, 20-78.
42. Stromberg,K., Person, A.-M. and Olsson,I. (1985) Eur. J. Cell Biol. **398**, 424-431.
43. von Heijne,G. (1986) Nuc. Acids Res. **14**, 4683-4690.
44. Kornfeld,R. and Kornfeld, S. (1985) Ann. Rev. Biochem. **54**, 631-664.
45. Morita,Y., Iwamoto,H. Aibara, S. Kobayashi, T. and Hasegawa, E. (1986) J. Biochem. **99**, 761-770.
46. Kozak,M. (1986) Cell **44**, 283-292.
47. Proudfoot,N.J. and Brownlee,G.G. (1976) Nature **263**, 211-214.
48. McLauchlan,J., Gaffney,D., Whitton,J.L. and Clements,J.B. (1985) Nuc. Acids Res. **13**, 1347-1368.
49. Lozzio,C.B. and Lozzio,B.B. (1975) Blood **45**, 321-334.

50. Morgan,D.A., Ruscetti,F.W. and Gallo,R. (1976) Science (Wash. D.C.) **193**, 1007-1008.
51. Saxon,A., Stevens,R.H., Quan,S.G. and Golde,D.W. (1978) J. Immunol. **120**, 777-782.
52. Kalyanaraman,V.S., Sarngadharan,M.G., Robert-Guroff,M., Miyoshi,I., Blayney,D., Golde,D. and Gallo,R.C. (1982) Science (Wash. D.C.) **218**, 571-573.
53. Benjamin,D., Magrath,I.T., Maguire,R., Janus,C., Todd,H.D. and Parsons,R.G. (1982) J. Immunol. **129**. 1336-1342.
54. Erikson,J., Griffin,C.A., ar-Rushdi,A., Valtieri,M., Hoxie,J., Finan, J., Emmanuel,B.S., Rovera,G., Nowell,P.C. and Croce,C.M. (1986) Proc. Nat. Acad. Sci. U.S.A. **83**, 1807-1811.
55. Davis,T., Brunning,R. and Quie,P.G. (1971) N. Engl. J. Med. **285**, 789-790.
56. Catovsky,D., Galton,D.A.G. and Robinson,J. (1972) Scand. J. Haematol. **9**, 142-148.
57. Breton-Gorius,J., Houssay,D. and Dreyfus,B. (1975) Br. J. Haematol. **30**, 273-278.
58. El Maallem,H. and Fletcher,J. (1977) J. Haematol. **37**, 323-329.
59. Kitahara,M. and Kushner,J.P. (1979) Cancer (Phila.) **44**, 2244-2248.
60. Cech,P., Schneider,P. and Bachmann,F. (1982) Acta Haematol. (Basel) **67**, 180-184.
61. Strauss,E.C., Kobori,J.A., Siu,G. and Hood,L.E. (1986) Anal. Biochem. **154**, 353-360.
62. Michelson,A.M., Markham,A.F. and Orkin,S.A. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 472-476.