DNA cloning and amino acid sequence determination of a major constituent protein of mammalian nucleoli

Correspondence of the nucleoplasmin-related protein N038 to mammalian protein B23

Marion S. Schmidt-Zachmann and Werner W. Franke
Division of Membrane Biology and Biochemistry, Institute of Cell and Tumor Biology, German Cancer Research Center, Im Neuenheimer Feld 280, D-6900 Heidelberg, Federal Republic of Germany

Abstract. Using a cDNA probe encoding the nucleolar protein N038 of Xenopus laevis, we have isolated clones that code for the corresponding mammalian protein from cDNA libraries of mouse embryonal carcinoma and fetal liver cells. The murine cDNA-derived amino acid sequence constitutes a polypeptide of 292 amino acids (including the initial methionine) of a total molecular weight of 32 560 and identifies a single ~1.5 kb mRNA on Northern blot hybridization. This polypeptide, which is highly homologous to the Xenopus protein N038, displays an organization in three major domains: (1) an aminoterminal portion of 119 amino acids, which shows a striking homology to nucleoplasmin of Xenopus; (2) a central portion of 68 amino acids that contains two extended acidic domains, a shorter of 13 residues and a longer of 29 residues, separated by an interval enriched in positively charged amino acids; (3) a carboxyterminal portion of 105 amino acids, which is almost identical to the reported partial amino acid sequence of human and rat nucleolar protein termed B23. The sequence comparisons show that the murine protein is the mammalian counterpart to the nucleolar protein N038 of Xenopus and is compatible with the idea that both proteins N038 represent the amphibian and murine equivalents to the human and rat nucleolar phosphoprotein B23. Special sequence features and predicted conformations of this protein are discussed in relation to the specific localization and the possible functions of this major nucleolar protein.

Introduction

The molecular principles governing the compartmentalization of proteins are currently under intensive investigation in many laboratories. In recent years particular progress has been made with respect to proteins accumulating in the nucleus in which relatively short amino acid sequences have been identified that are necessary and sufficient to direct them to the nucleus ("karyophilic signals"); "nuclear localization signals"). Examples include certain viral proteins such as the large T antigens of simian virus SV40 (Kalderon et al. 1984a, b; Lanford and Butel 1984; Goldfarb et al. 1986; Lanford et al. 1986; Richardson et al. 1986) and polyoma virus (Richardson et al. 1986), as well as the E1A protein of adenovirus (Krippl et al. 1985; Lyons et al. 1987; however, see also Richter et al. 1985). Proteins of this group include highly acidic, histone-binding proteins such as nucleoplasmin (Laskey et al. 1978; Krohne and Franke 1980a, b; Mills et al. 1980; Dingwall et al. 1982; Feldherr et al. 1984; Kleinschmidt et al. 1985; Bürghlin and De Robertis 1987; Bürghlin et al. 1987; Dingwall et al. 1987) and protein(s) N1/N2 (Bonner 1977; De Robertis et al. 1978; Dabauvalle and Franke 1982; Kleinschmidt et al. 1986; Kleinschmidt and Seiter 1988). In addition, signal sequences for karyophily have been reported for histone H2B (Moreland et al. 1987), ribosomal protein L3 (Moreland et al. 1985), glucocorticoid receptor (Picard and Yamamoto 1987), certain transcriptional activator proteins (Silver et al. 1984) and have also been proposed for nuclear lamins (Fisher et al. 1986; Krohne et al. 1987).

Another question for nuclear proteins, however, is that of the principles involved in their specific topology, i.e., the reason for the enrichment, if not exclusive localization, of a certain protein in a distinct nuclear substructure such as the lamina, a distinct chromosomal locus, or the nucleolus. Here nucleolar proteins are of special interest because of the homogeneity and enrichment of chromatin elements, i.e., rDNA chromatin, in this structure and the relatively well-defined functions of this structure, i.e., pre-rRNA synthesis, processing, and ribonucleoprotein packaging. Hence, it is perhaps not unexpected that several ribosomal proteins have been localized to the nucleolus (for references see, e.g., Chooi and Leiby 1981; Kalthoff and Richter 1982; Hadjiof 1985; Hügle et al. 1985a; Tsurugi et al. 1988).

In addition, several non-ribosomal proteins have been shown to be enriched in specific nuclear subcompartments. For example, RNA polymerase I has been found primarily in the "fibrillar centers" (Scheer and Rose 1984) whereas a ~M, 100 000 polypeptide variously called "protein C23" or nucleolin (Prestayko et al. 1974; Daskal et al. 1982; Escande et al. 1985), a protein of M, ~180000 (Feldmann et al. 1984), and a M, ~34000 polypeptide termed "fibrillarin" (Lischwe et al. 1985; Ochs et al. 1985b) have been localized to the "dense fibrillar component." On the other hand, in addition to certain ribosomal proteins, several non-ribosomal proteins have been identified in the "granular component" (pars granulosa) of the nucleolus, which is thought to be the compartment of packaging of the pre-ribosomal RNAs with proteins (for review see Hadjiof 1985). Examples reported include a M, ~40000 polypeptide ("ribocharin", Hügle et al. 1985b), a polypeptide of M, ~37000
called "protein B23" that is abundant in nucleoli of mammals and birds (Prestayko et al. 1974; Ochs et al. 1983; Busch 1984; Spector et al. 1984; Morris et al. 1985; Fields et al. 1986; see, however, Daskal et al. 1980) and a Mr ~ 38000 polypeptide ("protein N038") in nucleoli of amphibian and avian cells (Schmidt-Zachmann et al. 1987). All these proteins have also been found in free nucleoplasmic forms of precursor particles for both the large and the small ribosomal subunit.

Surprisingly, comparisons of the amino acid sequence of protein N038 with those of other known nuclear proteins has shown that the aminoterminal portion of the nucleolar protein N038 of the Xenopus laevis is highly homologous to the nucleoplasmic protein, nucleoplasmin, of the same species, whereas the carboxyterminal portion has revealed a striking homology to the partial sequence available for mammalian protein B23 (Schmidt-Zachmann et al. 1987). To clarify the nature of these abundant, and probably functionally important nucleolar proteins and their relationship to nucleoplasmin, we have therefore used the cDNA clone encoding Xenopus protein N038 for the isolation of mRNA and cDNA representing the related mammalian proteins. Here we report the cDNA cloning of murine N038 protein and the deduced amino acid sequence, and we show that this protein is practically identical in its carboxyterminal portion to rat and/or human protein B23, indicating that proteins B23 and N038 are probably the equivalent proteins in amphibia and mammals.

Materials and methods

Selection and characterization of cDNA. Libraries (igt10) constructed from poly(A +)-RNA of cultured murine teratocarcinoma stem cells of line F9 (kindly provided by Dr. A. Alonso, German Cancer Research Center, Heidelberg) or of fetal mouse liver (generously provided by Dr. S. Ruppert (this institute) were used. The Xenopus cDNA-clone JN038-185 (Schmidt-Zachmann et al. 1987), radiolabeled either by nick-translation (Ribgy et al. 1977) or random-priming (Feinberg and Vogelstein 1983), was used for the initial screening of the phage DNA of the F9-teratocarcinoma library by incubation at 42 °C. Non-specific bonding was removed by two washes in 2 × SSC/0.1% SDS at room temperature and then for 30 min in 0.1 × SSC/0.1% SDS at 55 °C. Positive clones were plaque purified, and the phage DNA was isolated. The 0.8 kb EcoRI-insert from the clone selected (gt9-185.3) was excised, subcloned into the bluescript vector and sequenced with the ~ 1.1 kb cDNA clone encoding the amphibian nucleolar protein N038 isolated from a igt11 expression library of Xenopus laevis ovary (Schmidt-Zachmann et al. 1987). Ten positive clones were picked, plaque-purified, and phage-DNA was isolated. As judged from EcoRI-digestion analyses, all clones contained inserts of different lengths. In Southern blot hybridization experiments, only one clone showed strong hybridization with the Xenopus cDNA clone (data not shown). This clone (igt9-185.3), which contained a cDNA insert of ~ 0.8 kb, as judged from its electrophoretic mobility, was subcloned into bluescript vector and sequenced. The sequence analysis (data not shown) revealed that this isolated cDNA has an uninterrupted reading frame of 207 amino acids. Comparison of this partial sequence with the amino acid sequence of the Xenopus cDNA clone encoding protein N038 showed that the clone, in fact, encodes the corresponding mammalian protein N038. Unfortunately, however, the carboxyterminal part of protein N038 and the 3'-non-coding region was missing from this cDNA clone. Therefore, we used this partial mouse cDNA clone for screening a second mouse cDNA library constructed from poly(A +)-RNA of fetal mouse liver. Again several positive clones were isolated, three of which showed, after digestion with EcoRI, inserts of ~ 1.25 kb. One of these (JFML-185.19) was isolated and tested in Northern blot hybridization on poly(A +)-RNA of the mouse myeloma cell line Ag8.653. The insert hybridized to a mRNA band of ~ 1.5 kb (Fig. 1), suggesting that it was a full-length cDNA clone. In vitro translation of the cDNA hybrid-selected mRNA from Ag8.653 poly(A +)-RNA yielded a single polypeptide that migrated on SDS-PAGE (Fig. 2) with a Mr of ~ 38000.

When this cDNA clone was transcribed in the bluescript expression vector with T7-polymerase and the resulting RNA examined by translation in vitro using the reticulocyte lysate, a translation product of the same size (Mr ~ 38000) was obtained, indicating that this cDNA clone indeed contains the entire protein-coding sequence (data not shown; cf., Schmidt-Zachmann et al. 1987).

Nucleotide sequence and amino acid sequence deduced therefrom

Figure 3 presents the nucleotide sequence of clone JFML-185.19 and the amino acid sequence deduced therefrom. The clone, consisting of 1276 nucleotides, contains 78 untranslated nucleotides at the 5'-end and 319 non-coding nucleotides at the 3'-end, with a poly(A +) addition signal 19 nucleotides in front of the poly(A)-tail, of which 16 nu-