© 1996 Oxford University Press

Nucleic Acids Research, 1996, Vol. 24, No. 5 **843–849**

Purification and molecular cloning of the scaffold attachment factor B (SAF-B), a novel human nuclear protein that specifically binds to S/MAR-DNA

Andrea Renz and Frank O. Fackelmayer*

Division of Biology, University of Konstanz, 78434 Konstanz, Germany

Received December 2, 1995; Revised and Accepted January 22, 1996

EMBL accession no. L43631

ABSTRACT

We have purified to near homogeneity a novel nuclear protein from HeLa cells, that specifically binds to scaffold or matrix attachment region DNA elements (S/MAR DNA). The protein, designated SAF-B for scaffold attachment factor B, is an abundant component of chromatin, but not of the nuclear matrix and is expressed in all human tissues investigated. Antibodies against the purified protein were raised in rabbit and used to isolate the complete cDNA encoding SAF-B by immunoscreening. As predicted from the cDNA sequence, SAF-B contains 849 amino acids (96 696 Da), without significant homology to any known protein. SAF-B is rich in charged residues, leading to an aberrant migration on SDS gels, and has two putative bipartite nuclear localisation signals.

INTRODUCTION

Eukaryotic chromatin is organised in a higher order structure consisting of thousands of discrete, topologically constrained loop domains. These loops are fixed at their bases to a network composed of proteins and RNA that is generally referred to as the nuclear matrix or scaffold (1,2). Increasing evidence implies that a tight binding of chromatin to the nuclear scaffold is not only important for the compaction of the chromatin fiber, but is also involved in many aspects of nucleic acid metabolism (see 3 for review). It is widely accepted that loop domains are the units of gene expression and replication, and are thus possibly also involved in the formation of nuclear subcompartments (4–6). Attachment of chromatin to the nuclear scaffold seems to occur via specialised DNA elements that have been found in all eukaryotic organisms investigated. Termed SARs or MARs (for scaffold or matrix attachment regions; in the current study we call them S/MARs), these DNA elements are evolutionarily conserved, as shown for example by the fact that mouse S/MARs bind to yeast nuclear scaffolds (7). Consequently, the DNA regions conferring chromatin attachment to the scaffold are of considerable interest and have thoroughly been investigated in the last decade. In many cases, S/MAR elements co-map with boundaries of actively transcribed chromatin domains, and are postulated to protect the transcribed domain from regulatory mechanism from neighboring sequences (8,9). However, the initial interpretation that S/MARs generally form the borders of transcribed regions (and thus delimit units of gene expression) has been questioned by the discovery of intronic S/MARs (e.g. 10–12). These intronic S/MARs are indistinguishable from gene-flanking S/MARs with respect to their nucleotide sequence, their interaction with scaffold preparations and their ability to confer position independent transcriptional activation. It is therefore likely that both types of S/MARs perform the same function *in vivo*, the anchorage of chromatin loops to the nuclear scaffold, and, presumably, regulatory effects on adjacent genes.

Generally, S/MARs are DNA fragments of 300–3000 bp length that contain several A+T rich sequence motifs and sequences resembling topoisomerase II cleavage sites (13,14). However, an interesting result from the comparison of the high number of characterised S/MARs is the fact that no simple consensus sequence seems to exist for nuclear scaffold attachment (13). Although several short A+T rich sequence motifs are clustered in most S/MAR DNA elements, no single one of these sequences is characteristic for all S/MARs. Rather, the binding of S/MARs to the nuclear scaffold is highly dependent on both the A+T-richness and the length of the DNA fragment, indicating that an interaction is involved that is strikingly different from well characterised DNA–protein interactions as, for example, with transcription factors or restriction enzymes. At present, however, the mechanism by which the nuclear scaffold recognises the S/MAR DNA elements is not understood. It has been proposed that the nuclear scaffold contains proteins that specifically recognise unusual DNA structures such as tracts with a narrow minor groove (15), the single-stranded status of 'unwinding elements' (16–18) or DNA bends (19). Further insight into the underlying recognition mechanisms, however, is only possible by identifying and characterising these proteins in molecular detail.

A general assay to screen for proteins with a possible function in the anchorage of chromatin loops is the use of a S/MAR DNA element as molecular affinity probe in Southwestern blot procedures or in direct cDNA screening. Both approaches have been used successfully to identify proteins of the desired specificity (20,11,21). Unfortunately, the few well characterised S/MAR-binding proteins, among those histone H1 (22), topoisomerase II (14), lamins (23,24), SATB1 (21) and SAF-A/hnRNP-U

* To whom correspondence should be addressed

(11,25,26) show no homologies on the level of nucleotide or amino acid sequence. It is therefore not yet possible to say what property of S/MAR DNA is specifically recognised by these proteins. To gain insight into the molecular mechanism of S/MAR DNA recognition, we have set out to identify and characterise proteins that specifically interact with S/MAR DNA elements. In our search, we have previously identified four nuclear proteins from HeLa cells with the desired specificity for S/MAR DNA (11). These proteins were termed scaffold attachment factors A through D (SAF-A to SAF-D), dependent on their relative abundance in nuclear extracts. The protein SAF-A, characterised in our original publication (11), was later shown to be identical to the protein hnRNP-U (25,27–29), and the specific binding of this protein to S/MAR DNA was independently confirmed by others (26,30). In this communication we report on the purification, cloning and characterisation of the second abundant protein SAF-B, the scaffold attachment factor B.

MATERIALS AND METHODS

Purification of SAF-B

All buffers contained 10 mM mercaptoethanol to protect free thiol groups and 10 mM $Na_2S_2O_5$ (sodium metabisulfite, buffered to pH 8.0 with NaOH) as a general protease inhibitor; all purification steps were carried out in the cold. For the purification of SAF-B, nuclear extract was prepared from 1×10^{10} HeLa S3 cells (obtained from Computer Cell Culture Center, Brussels) as previously described (11). Briefly, cells were washed in phosphate buffered saline, allowed to swell for 10 min in hypotonic buffer (10 mM Tris–HCl pH 8.0, 10 mM NaCl, 3 mM $MgCl_2$) and broken by 20 strokes in a loose fitting dounce homogeniser. Nuclei were pelleted (750 *g*, 10 min, 4°C), washed three times in the same buffer and then extracted into 80 ml of extraction buffer (10 mM Tris–HCl pH 8.0, 500 mM NaCl) for 10 min. Nuclear debris is removed by ultracentrifugation (150 000 *g*, 30 min, 4°C) and the cleared extract is directly mixed with 20 ml of pre-equilibrated hydroxylapatite. After 30 min on a rocking platform, the hydroxylapatite is pelleted (160 *g*, 2 min, 4°C), washed twice in 50 ml of binding buffer and twice in 50 ml buffer containing 70 mM potassium phosphate. Bound protein is eluted in 80 ml elution buffer (170 mM potassium phosphate), diluted 2-fold with 10 mM Tris–HCl pH 8.0, 10 mM $Na_2S_2O_5$, and directly loaded on a pre-equilibrated 10 ml Mono Q column (Pharmacia) at a flow rate of 1 ml/ml. Bound protein is eluted in a linear gradient from 100 to 700 mM NaCl in dilution buffer with a total volume of 60 ml. Fractions containing SAF-B are determined in a Southwestern blot assay with the labelled S/MAR element MII and a 1000-fold excess of unlabelled *E.coli* DNA as unspecific competitor. SAF-B elutes at ~300 mM NaCl from Mono Q. Active fractions are pooled, diluted 4-fold and applied to a 1 ml Mono S column (Pharmacia) at a flow rate of 0.5 ml/min. Elution is carried out in a linear gradient from 50 to 700 mM NaCl, where SAF-B elutes between 200 and 350 mM NaCl. For the final purification, the eluate from the Mono S column, containing ~150 µg SAF-B, is diluted 4-fold, mixed with 0.5 mg *E.coli* DNA, incubated on ice for 15 min and then pelleted (8000 *g*, 15 min, 4°C). The pellet was resuspended in high salt buffer (10 mM Tris–HCl pH 8.0, 400 mM NaCl) and passed through a DEAE–Sepharose column to remove the DNA. The flowthrough, containing ~100 µg of nearly homogenous SAF-B, is mixed with the same volume of glycerol and stored at –20°C.

Production of antibodies

Nearly homogeneous SAF-B or the bacterially overexpressed partial protein p43 was further purified by SDS–PAGE and subsequent copper chloride staining to remove remaining impurities. The desired bands, containing 100–200 µg of protein per immunisation, were excised, destained and homogenised by 10 passages through a 25 gauge cannula. The homogenised material was mixed with Freund's complete adjuvant for the first immunisation or RAS adjuvant for the two boost immunisations after 6 weeks each. The obtained antisera were made monospecific by affinity purification with the purified antigen before further use according to (31).

DNA binding assays

DNA binding was assayed by a Southwestern blot procedure as previously described (11) with DNA probes end-labelled radioactively by Klenow polymerase.

Cloning and sequencing

Polyclonal affinity purified serum against the SAF-B protein was used for immunoscreening of a HeLa LambdaZAP II library (Stratagene #937216) as described (32). Several positive clones were isolated and sequenced by the dideoxy method (33) using the USB sequenase kit. A 540 bp probe from the 5′ end of the longest clone was used for conventional hybridisation screening of the same library. Among the 14 positive clones, one contained the complete coding region along with the 5′ and 3′ non-translated regions. This clone, named A5, was sequenced from both strands as described above.

In vitro transcription/translation

The complete cDNA clone A5 was *in vitro* translated with the TNT rabbit reticulocyte lysate kit (Promega) in the presence of [^{35}S]methionine (Amersham) according to the manufacturer's suggested conditions.

Bacterial expression

A fragment of the complete cDNA encoding amino acids 443–718 of SAF-B was introduced in frame into the pRSET prokaryotic expression vector by conventional cloning procedures (32). Positive clones were identified by DNA minipreparations and assayed for protein overexpression after induction with 1 mM IPTG for 2 h. Overexpressed protein, named p43 according to its apparent molecular weight in SDS gels, was purified by metal chelate chromatography over Ni–Agarose and used for the production of antibodies.

Other methods

SDS–PAGE of proteins was performed according to Laemmli (34); the gels were stained with silver (35), with Coomassie brilliant blue (32) or copper chloride (36). Western transfer was performed according to (37) with affinity purified polyclonal serum (31), alkaline phosphatase coupled secondary antibodies (Sigma) and BCIP/NBT substrate. Protein concentrations were determined using the BioRad protein assay reagent with bovine serum albumin as standard. DNA was labelled with [^{32}P]dATP

using Klenow polymerase according to standard protocols (32). Northern blotting was performed on blots obtained from Clonetech according to the conditions suggested by the manufacturer. Cell fractionation experiments were done as described previously (25).

RESULTS

Purification of the scaffold attachment factor B (SAF-B)

The protein SAF-B was first identified by Romig *et al.* (11) as one of four proteins that specifically bind to the S/MAR DNA element MII from the human topoisomerase I gene locus (12). Subsequently, we have developed a purification protocol for this protein. SAF-B was monitored throughout column chromatography by its ability to bind radioactively labelled S/MAR DNA in the presence of a vast excess of unspecific competitor DNA. This assay is highly specific for S/MAR-DNA binding proteins, although several abundant S/MAR binding proteins such as topoisomerase II and HMG I/Y escape detection, most probably because of their inability to refold to an active conformation after denaturing gel electrophoresis. Our purification procedure starts from a 500 mM NaCl nuclear extract prepared from 1×10^{10} HeLa cells. The nuclear extract (Fig. 1, fraction NE) was bound to hydroxylapatite material in a batch procedure. The eluate at 170 mM potassium phosphate (Fig. 1, fraction HAP) was diluted and directly applied to a FPLC Mono-Q column. Unbound proteins were washed off and the column was eluted with a linear gradient from 100 to 700 mM NaCl. Active fractions with the peak of SAF-B at ~300 mM NaCl were combined (Fig. 1, fraction Mono-Q) and loaded to a FPLC Mono-S column. SAF-B eluted from Mono-S between 200 and 350 mM NaCl (Fig. 1, fraction Mono-S). After chromatography on Mono-S, SAF-B is the main protein in the active fractions. However, it still contains the activity of SAF-C, a protein with a molecular weight of 100 kDa. Removal of SAF-C and other contaminating proteins was achieved by exploiting the ability of SAF-B to form aggregates in the presence of nucleic acids. After centrifugation, the SAF-B/DNA aggregates were disrupted in high salt buffer and DNA was removed by a passage over DEAE–Sepharose, resulting in nearly homogeneous SAF-B as judged by silver staining (Fig. 1, fraction aggregation). On average, the purification yields ~100 µg of SAF-B protein from 10^{10} cells, with a recovery of 15–20% as estimated in Southwestern blots by comparing activity of SAF-B in total cell extracts of a known number of cells with a known amount of purified protein. Purified SAF-B migrates as a single band of 150 kDa under denaturing conditions and is a monomeric protein with an apparent sedimentation coefficient of $4.2S_{20,w}$ in non-denaturing glycerol gradient centrifugation. Based on its apparent molecular weight, we speculated that SAF-B could be identical or related to topoisomerase II, a protein with known binding specificity for S/MAR DNA (14). However, purified SAF-B is free of topoisomerase activity and is not recognised by antibodies against purified topoisomerase II in Western blotting experiments. Additionally, the cDNA sequence of SAF-B, reported later in this paper, reveals no homology to topoisomerase sequences.

SAF-B binds specifically to S/MAR-DNA

As described above, SAF-B was purified from nuclear extracts by virtue of its specific binding to the S/MAR DNA element MII in the presence of a 1000-fold excess of unspecific *E.coli* DNA as a competitor. To confirm the binding specificity of purified SAF-B,

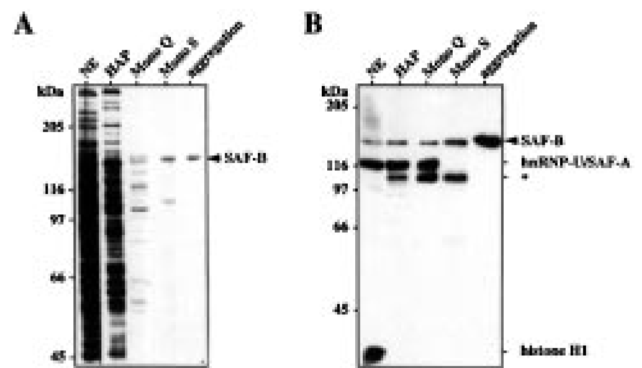

Figure 1. Chromatographic purification of SAF-B. Proteins present in nuclear extract (NE) and in active fractions after chromatography on hydroxylapatite (HAP), Mono Q, Mono S and after aggregation with *E.coli* DNA (aggregation) were separated on 7% SDS–polyacrylamide gels. Proteins were visualised by silver staining (**A**) or by blotting to a nitrocellulose membrane and incubation with the radioactively end-labelled S/MAR DNA MII in the presence of a 1000-fold excess of *E.coli* competitor DNA (**B**). Molecular weight markers, myosin (205 kDa), β-galactosidase (116 kDa), phosphorylase b (97 kDa), bovine serum albumin (66 kDa) and ovalbumin (45 kDa). The positions of the other S/MAR binding proteins histone H1, SAF-A/hnRNP-U and SAF-C (asterisk) are indicated for reference.

we repeated Southwestern blot experiments with the labelled S/MAR DNA element MII (Fig. 2A) and the non-S/MAR DNA pUC18 of similar length (Fig. 2B). We find that SAF-B binds approximately equally well to both DNAs in the absence of competitor DNA, indicating that SAF-B has a general DNA binding activity. However, with increasing amounts of *E.coli* competitor DNA, pUC18 is readily displaced from SAF-B, while ~40% S/MAR DNA remains bound to SAF-B even in the presence of a 2000-fold excess of competitor DNA (Fig. 2C). With even higher amounts of competitor, binding is gradually lost, until no binding is detectable at a 25 000-fold excess of *E.coli* DNA (data not shown). Specific binding was also observed with several other, heterologous S/MAR elements like B1X1 or B4B5 from the chicken lysozyme gene locus (38; a gift of Dr W. Strätling, Hamburg) or fragment IV from the upstream S/MAR of the human interferon-β gene locus (39; a gift of Dr J. Bode, Braunschweig) (data not shown). We conclude therefore, that SAF-B is a DNA binding protein with high specificity for S/MAR DNA elements. In experiments with restriction fragments derived from different S/MARs, we find that SAF-B has no easily defined consensus binding site, but that specific binding is dependent on both A+T richness and length of the subfragment (data not shown). For the purpose of this paper, however, we do not particularly focus on a detailed characterisation of the DNA binding properties of SAF-B.

SAF-B is a novel protein with a unique primary structure

To enable screening for the cDNA encoding SAF-B, we developed a polyclonal antiserum against highly purified SAF-B. The serum obtained recognised SAF-B in both crude nuclear extracts and in its purified form. Specific antibodies were affinity purified by binding to immobilised SAF-B and were used for immunoscreening a Lambda ZAP expression library. Several positive clones were isolated and sequenced. As none of the isolated clones contained the 5′ end of the cDNA, a 540 bp probe from the 5′ end of the longest clone was used to rescreen the same library by DNA

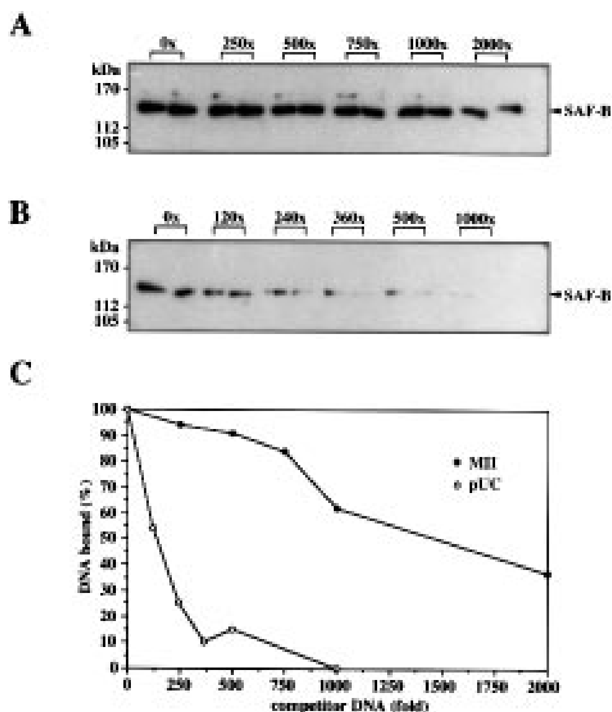

Figure 2. SAF-B binds specifically to S/MAR DNA. Purified SAF-B, 1 µg per lane, was separated on a 7% SDS–polyacrylamide gel and transferred to a nitrocellulose membrane. The membrane was incubated with the end-labelled S/MAR DNA MII (**A**) or the end-labelled non-S/MAR control DNA pUC18 (**B**) in the presence of the indicated excess of *E.coli* competitor DNA. Duplicates of all assays are shown and quantified in (**C**) by determining the percentage of bound DNA by liquid scintillation counting. The amount of DNA bound in the absence of competitor is set as 100%, independently for (A) and (B).

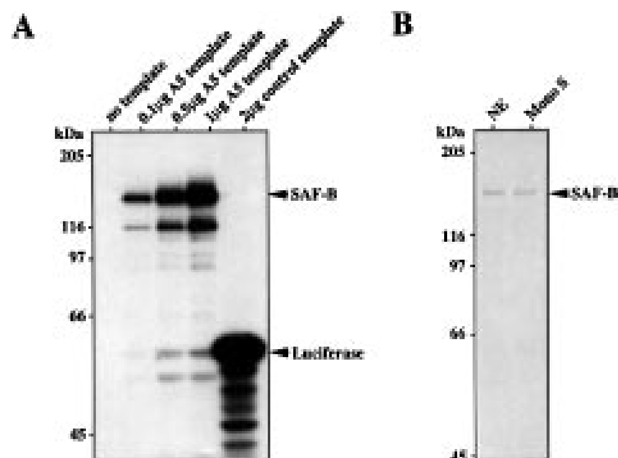

Figure 3. The cDNA clone A5 encodes the full length SAF-B protein. (**A**) Increasing amounts of the cDNA clone A5 were transcribed and translated *in vitro* in the presence of [^{35}S]methionine. Synthesized proteins were separated on a 7% SDS–polyacrylamide gel, along with a positive control (2 µg luciferase cDNA) and a negative control (no template), and visualised by fluorigraphy. Molecular weight markers were as in Figure 1. (**B**) Proteins present in nuclear extract of HeLa cells (NE) and highly enriched SAF-B (Mono S fraction) were separated on a 7% SDS–polyacrylamide gel, transferred to a nitrocellulose membrane and detected with antibodies against the bacterially overproduced partial protein p43 (amino acids 443– 718 of SAF-B).

hybridisation. Fourteen positive clones were isolated, one of which was 2.8 kb in length and contained the complete coding region of SAF-B (sequence deposited to the EMBL, GenBank and DDBJ databases under accession number L43631). To verify that this clone, termed A5, was a cDNA clone encoding the SAF-B protein, we synthesised [^{35}S]methionine-labelled protein in a coupled *in vitro* transcription–translation system. The *in vitro* synthesised protein showed the same electrophoretic mobility as authentic, purified SAF-B from HeLa cells (Fig. 3A). Additionally, antibodies produced to the bacterially overproduced partial protein p43 (amino acids 443–718 of SAF-B) recognise SAF-B in unfractionated HeLa nuclear extracts and in the purified form (Fig. 3B). Taken together, these facts strongly indicate that the A5 cDNA clone encodes the full length SAF-B protein.

The cDNA clone A5 as well as several additional, shorter cDNA clones encoding parts of SAF-B were sequenced by the dideoxy method from both strands (33). Clone A5 is 2825 bp in length and contains an open reading frame encoding a 849 amino acid protein with a calculated mass of 96 696 Da and a pI of 8.8. The proposed initiation codon at nucleotide 48 shows a weak homology to the translation initiation consensus sequence (40) and is preceded by two in-frame stop codons. The clone extends 228 nucleotides 3′ of the translation stop codon at nucleotide 2597, with a consensus polyadenylation signal starting at nucleotide 2752 and a poly(A) tract from nucleotide 2775.

Nucleotide and protein sequence searches of the GenBank and EMBL databases (41) did not reveal significant homology to any known proteins, demonstrating that SAF-B is a novel protein. Sequence identity was found, however, with several expressed sequence tags (e.g. GenBank accession numbers HS54290, HS609163 and HS25591) from human cells.

A schematic representation of the SAF-B protein, as predicted from the cDNA sequence, is given in Figure 4. SAF-B contains two putative bipartite nuclear localisation signals and numerous potential phosphorylation sites. The protein is rich in charged amino acids, with a highly basic N-terminus of 97 aa, an acidic central part of 464 amino acids, and a basic C-terminus starting from amino acid 562. Secondary structure predictions (42,43) show that SAF-B is mostly α-helical and contains only few short β-sheets. The C-terminal 200 amino acids are likely to form an unordered, flexible structure with a high number of predicted turns. The occurrence of differently charged domains of SAF-B is compatible with the aberrant migration of the protein in SDS gels, giving a possible explanation for the high discrepancy between the apparent molecular weight (150 kDa) and the actual calculated molecular weight (96.7 kDa). Phosphorylation at several of the putative phosphorylation sites could also be involved in the aberrant migration. We have tested this possibility by immunoprecipitation of SAF-B from nuclear extracts of HeLa cells labeled with ^{32}P *in vivo*. In these experiments we detected that SAF-B is phosphorylated, but treatment of the purified protein with alkaline phosphatase did not alter the electrophoretic migration of SAF-B (data not shown). Presently we cannot formally rule out that modifications other than phosphorylation are involved in the electrophoretic behavior of SAF-B. However, the finding that the bacterially overproduced protein fragment p43 also show slower migration in SDS gels (43 kDa apparent versus 34 kDa calculated), strongly indicate that the discrepancy

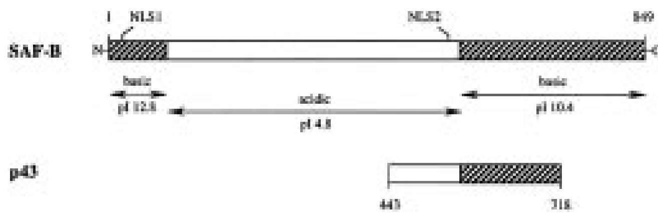

Figure 4. Schematic representation of the SAF-B protein. Charged regions with their calculated isoelectric points and the location of both NLS sequences are indicated. The part of the protein that was bacterially produced (p43) is shown below.

between true and apparent molecular weight is caused by the primary structure of SAF-B alone.

SAF-B is a ubiquitous chromatin protein

We have focused on SAF-B as a DNA binding protein because of its specificity for S/MAR DNA, which makes it a good candidate for a protein located at the attachment point of chromatin loops *in vivo*. Several criteria should be fulfilled by such a protein, and were addressed experimentally. First, a protein with a structural function in the nucleus should be abundant. From our purification that yields ~100 µg of SAF-B from 10^{10} HeLa cells with a recovery of 15–20%, we calculated a copy number of ~10^5 molecules of SAF-B per nucleus. Although this copy number is lower than that of other S/MAR binding proteins (e.g. histone H1 or SAF-A/hnRNP-U), it is compatible with a structural function of SAF-B. Secondly, a general S/MAR binding protein should be ubiquitous, i.e. expressed in all or the majority of cells and tissues. We have performed a Northern blot analysis with a part of the SAF-B cDNA clone A5 as a probe (Fig. 5). The left panel of Figure 5B shows a Northern blot with poly(A)$^+$ RNA from a collection of neoplastic cells of different origins, the right panel is a blot with poly(A)$^+$ RNA from healthy human tissues. A specific mRNA species with a length of 3.4 kb is detected in all cells and tissues at approximately the same level, consistent with SAF-B being a ubiquitous (housekeeping) protein, at least on the level of gene expression.

The subcellular localisation of SAF-B was analysed biochemically and by indirect immunofluorescence microscopy with affinity purified antibodies. Immunofluorescence experiments clearly demonstrate that SAF-B is located in the nucleus, as expected for a protein purified from nuclear extracts (data not shown). This finding is supported by biochemical cell fractionation performed according to the protocol of Fey *et al.* (44,45). This procedure allows the preparation of well defined subcellular and subnuclear fractions with distinct protein compositions under conditions that preserve the non-chromatin structure of the nucleus as well as the spatial organisation of RNA in these structures (6). The partitioning of SAF-B to these fractions was determined by Western blotting experiments and is shown in Figure 6C, along with a schematic representation of the fractionation protocol (Fig. 6A) and a demonstration of the unique protein composition of several of these fractions (Fig. 6B). Lysis of cells in an isotonic buffer containing 0.5% Triton X-100 releases soluble cytosolic proteins, and yields insoluble material consisting of nuclei and cytoskeletal proteins. SAF-B quantitatively partitions to the insoluble material, from which ~50% of SAF-B can be extracted by treatment with 250 mM ammonium sulfate, along with histone H1 and cytoskeletal proteins. The remaining insoluble material, consisting of extracted nuclei that

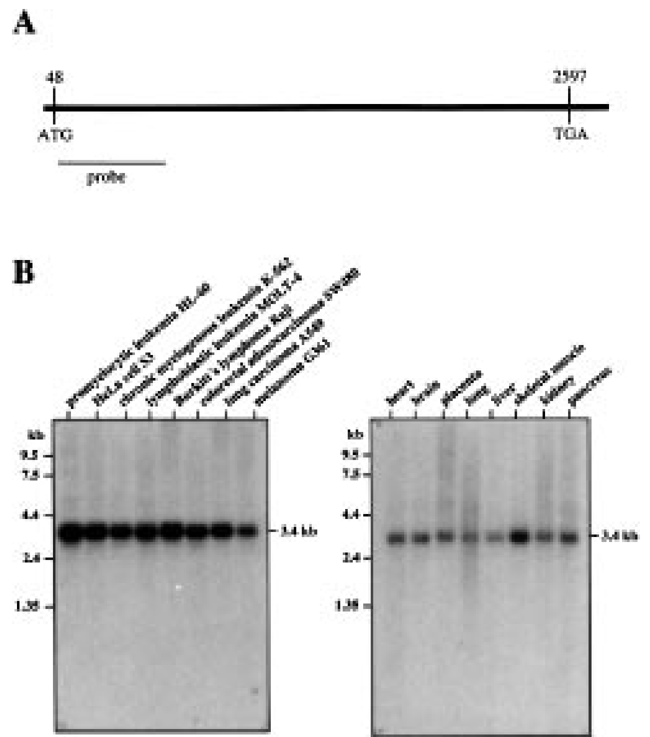

Figure 5. SAF-B is a ubiquitous, housekeeping protein. (**A**) The indicated section of the SAF-B encoding cDNA was used as a probe for (**B**) Northern blotting experiments with 2 µg poly(A)$^+$ RNA each from various human cancer cell lines (left panel) or from different tissues of normal human cells (right panel). The SAF-B specific mRNA runs at 3.4 kb. Exposure times were 24 h for (A) and 10 h for (B).

contain the other half of SAF-B, was digested with DNase I to release proteins tightly bound to DNA ('chromatin proteins', e.g. the core histones). Over 95% of the remaining SAF-B is extracted from the nuclei by this treatment, consistent with the notion that SAF-B is a chromatin protein that is either directly or via other proteins bound to chromosomal DNA *in vivo*. Interestingly, SAF-B is not a component of the 'complete' nuclear matrix that remains after DNase digestion or the 'extracted' nuclear matrix (after extraction of the 'complete' matrix with a buffer containing high salt), as, e.g. SAF-A/hnRNP-U (25).

DISCUSSION

In higher eukaryotes, the genomic DNA of a few meters in length has to be compacted in some way to be confined within the nucleus of only some micrometers in diameter. This impressive compaction is brought about by the formation of chromatin, whose basic architecture—DNA wound around octamers of histone proteins—is well understood. However, many details on higher order structures of chromatin remain obscure. A new area in chromatin research was initiated by the discovery of a proteinaceous nuclear framework (46) and subsequently by the identification of DNA fragments that specifically bind to this framework (47,48). Combining data from electron microscopic examination of chromosomes (49,50) and biochemical work on chromatin structure (51,47), Gasser and Laemmli (1) have proposed a model according to which chromatin is organised in constrained, topologically independent loops attached to a structural entity designated nuclear matrix or scaffold.

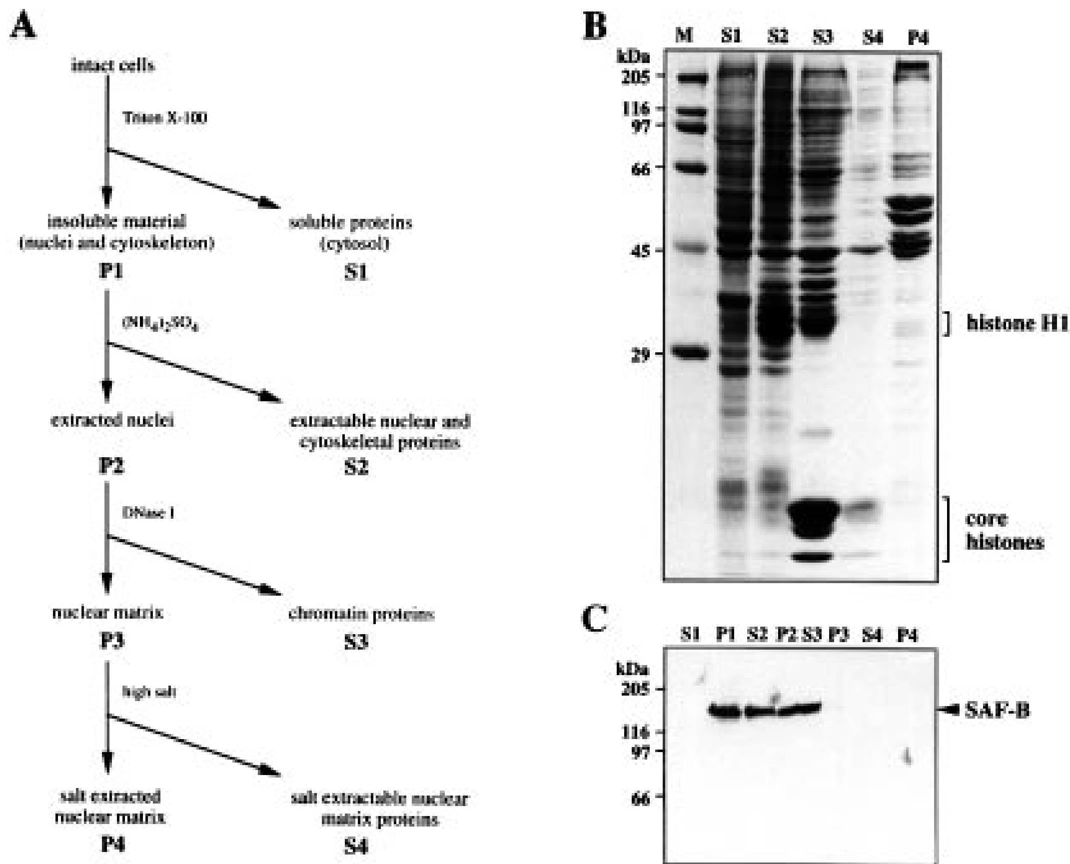

Figure 6. Biochemical fractionation of HeLa cells. HeLa cells were fractionated into subcellular and subnuclear fractions with unique protein compositions according to a standard protocol (44,45). A schematic representation of the procedure is given in (**A**), with the abbreviations for the different fractions indicated. (**B**) Protein composition of the extracted fractions (S1–S4) and the residual, insoluble material (P4) were examined by Coomassie blue staining after resolution on a 12% SDS–polyacrylamide gel. The location of the histones is marked for reference. (**C**) The subcellular distribution of SAF-B was determined by Western blotting of all protein fractions with affinity purified antibodies after electrophoresis on a 10% SDS–polyacrylamide gel.

Although this model is widely accepted today, little is known about the mechanism of attachment of chromatin to nuclear substructures. To gain insight into this functionally important aspect of nuclear structure, DNA elements located at attachment points have been used in several laboratories to identify nuclear proteins that could be involved in the formation of chromatin loop domains (20,11,23,21,26).

In this communication, we describe the purification and cDNA cloning of a novel protein, designated the scaffold attachment factor B (SAF-B). We have identified the SAF-B protein as one of four proteins in HeLa cell nuclear extracts that specifically interact with S/MAR DNA, and consider it a candidate protein for a molecular anchor at the basis of chromatin loops. Biochemical fractionation of cells demonstrate that SAF-B is a chromatin protein, but not a constituent of nuclear matrix preparations. This finding is interesting, as there is a general notion that a protein involved in S/MAR DNA attachment should be part of the insoluble substructure of nuclei. However, at least one other protein with significant binding specificity to S/MAR, namely histone H1, is also not present in nuclear matrix preparations but in negligible amounts (22) and is bound to chromatin even weaker than SAF-B. It is thus possible that two types of S/MAR DNA binding proteins exist, which differ in their partitioning upon biochemical fractionation, but are both involved in S/MAR element function *in vivo*.

As expected for a protein with a general function in chromatin structure, SAF-B is expressed equally in all cell types investigated, suggesting a housekeeping nature of the protein. This behavior is reminiscent of SAF-A/hnRNP-U and histone H1, other known S/MAR binding proteins, but different from SATB1 or topoisomerase II which are differentially expressed in different cell types (21,52).

The complete cDNA for SAF-B, obtained by immunoscreening with antibodies developed against the chromatographically purified protein, revealed that SAF-B is a unique protein with no homology to known proteins. It contains highly charged regions, with both the N- and the C-terminus being basic and the central half being acidic. SAF-B has two putative nuclear localisation signals, compatible with our finding that the protein is located in the nucleus. Although SAF-B is clearly a DNA binding protein, computer comparisons to EMBL and GenBank databases found no significant homologies to any previously identified DNA binding protein. We can therefore not yet define by analogy which part of the protein confers DNA binding. It could be argued that DNA binding of SAF-B occurs due to non-specific electrostatic interaction between basic regions of the protein and the negatively charged phosphate backbone of DNA. However, a simple electrostatic interaction would not be consistent with the specific binding of SAF-B to S/MAR DNA elements.

Although no homology to other cloned S/MAR binding proteins are evident on the level of the amino acid sequence, SAF-B shares the ability for nucleic acid dependent self-aggregation with histone H1 (22), topoisomerase II (14) and SAF-A/hnRNP-U (25). At present, we do not know the molecular basis of the aggregation of these proteins, but it is likely that both a DNA binding domain and a protein–protein interaction domain are involved in this process. It is thus possible that S/MAR DNA binds to protein aggregates due to its intrinsic flexibility brought about by interspersed A+T rich elements. This flexibility could allow the DNA to follow the path of protein aggregates, and could discriminate S/MAR DNA from bulk genomic DNA. Such a binding mode requires strong protein–protein interactions, while the binding of a protein monomer to DNA could be comparatively weak and unspecific. This model could be an explanation why S/MARs have to be of certain length (usually >500 bp) to be specifically bound. Future experiments will focus on the domain structure of SAF-B, with the aim to identify protein regions responsible for specific DNA binding and protein–protein interaction.

NOTE ADDED IN PROOF

After our manuscript was accepted, we learned that Drs. Jean-Pierre Bourquin and Walter Schaffner (Molecular Biology, University of Zurich, Switzerland) had independently cloned the same cDNA. Sequence comparison revealed a critical difference at position 344, resulting in a frameshift concerning the first 100 amino acids of SAF-B. This error has been corrected in the database entry. Additionally, we conclude that the 5′ end of the SAF-B cDNA is missing in our clone A5. We apologize for the error and any confusion it might have caused.

ACKNOWLEDGEMENTS

The authors wish to thank Dr Arndt Richter for encouragement and helpful discussions at the beginning of this work, Annette Hecker for excellent technical assistance and Dr Rolf Knippers for support and critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft through SFB 156.

REFERENCES

1 Gasser,S.M. and Laemmli,U.K. (1987) *TIGS,* **3**, 16–22.
2 Garrard,W.T. (1991). *BioEssays,* **13**, 87–88.
3 De Jong,L., van Driel,R., Stuurman,N., Meijne,A.M.L. and van Renswoude,J. (1990) *Cell Biol. Int. Rep.,* **14**, 1051–1074.
4 Cook,P.R. (1991). *Cell,* **66**, 627–635.
5 Hozak,P., Jackson,D.A. and Cook,P.R. (1994) *J. Cell Sci.,* **107**, 2191–2202.
6 Xing,Y.G. and Lawrence,J.B. (1991) *J. Cell Biol.,* **112,** 1055–1063.
7 Cockerill,P.N. and Garrard,W.T. (1986a) *FEBS Lett.,* **204**, 5–7.
8 Grosveld,F., Greaves,D.R., Kollias,G. and van Assendelft,G.B. (1987) *Cell,* **51**, 975–985.
9 Stief,A., Strätling,W.H., Sippel,A.E. and Winter,D.M. (1989) *Nature,* **341**, 343–345.
10 Käs,E. and Chasin,L.A. (1987) *J. Mol. Biol.,* **198**, 677–692.
11 Romig,H., Fackelmayer,F.O., Renz,A., Ramsperger,U. and Richter,A. (1992) *EMBO J.,* **11**, 3431–3440.
12 Romig,H., Ruff,J., Fackelmayer,F.O., Patil,M.S. and Richter,A. (1994) *Eur. J. Biochem.,* **221**, 411–419.
13 Gasser,S.M., Amati,B.B., Cardenas,M.E. and Hofmann,J.F.X. (1989) *Int. Rev. Cytology,* **119**, 57–96.
14 Adachi,Y., Käs,E. and Laemmli,U.K. (1989) *EMBO J.,* **8**, 3997–4006.
15 Käs,E., Izaurralde,E. and Laemmli,U.K. (1989) *J. Mol. Biol.,* **210**, 587–599.
16 Kohwi-Shigematsu,T. and Kohwi,Y. (1990) *Biochemistry,* **29**, 9551–9560.
17 Bode,J., Kohwi,Y., Dickinson,L., Joh,T., Klehr,D., Mielke,C. and Kohwi-Shigematsu,T. (1992) *Science,* **255**, 195–197
18 Kay,V. and Bode,J. (1994) *Biochemistry,* **33**, 367–374.
19 Homberger,H.P., *Chromosoma*, **98**, 99–104
20 von Kries,J.P., Buhrmester,H. and Strätling,W.H. (1991) *Cell,* **64**, 123–135.
21 Dickinson,L.A., Joh,T., Kohwi,Y. and Kohwi-Shigematsu,T. (1992) *Cell,* **70**, 631–645.
22 Izaurralde,E., Käs,E. and Laemmli,U.K. (1989) *J. Mol. Biol.,* **210**, 573–585.
23 Luderus,M.E.E., de Graaf,A., Mattia,E., den Blaauwen,J.L., Grande, M.A., de Jong,L. and van Driel,R. (1992) *Cell,* **70**, 949–959.
24 Luderus,M.E.E., den Blaauwen,J.L., de Smit,O.J.B., Compton,D.A. and van Driel,R. (1994) *Mol. Cell. Biol.,* **14**, 6297–6305.
25 Fackelmayer,F.O., Dahm,K., Renz,A., Ramsperger,U. and Richter,A. (1994) *Eur. J. Biochem.*, **211**, 749–757
26 Tsutsui,K., Tsutsui,K., Okada,S., Watarai,S., Seki,S., Yasuda,T. and Shohmori, T. (1993) *J. Biol. Chem.,* **268**, 12886–12894.
27 Fackelmayer,F.O. and Richter,A. (1994a) *Biochim. Biophys. Acta,* **1217**, 232–234
28 Fackelmayer,F.O. and Richter,A. (1994b) *Biochemistry,* **33**, 10416–10422
29 Kiledjian,M. and Dreyfuss,G. (1992) *EMBO J.*, **11**, 2655–2664.
30 von Kries,J.P., Buck,F. and Strätling,W.H. (1994) *Nucleic Acids Res.,* **22**, 1215–1220.
31 Harlow,E. and Lane,D.P. (1988) *Antibodies: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
32 Sambrook,J., Maniatis,T. and Fritsch,E.F. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, second edition.
33 Sanger,F., Nicklen,S., and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
34 Laemmli,U.K. (1970) *Nature,* **227**, 680–685.
35 Wray,W., Boulikas,T., Wray,V. and Hancock,R. (1981) *Anal. Biochem.,* **118**, 197203.
36 Lee,C., Levin,A. and Branton,D. (1987) *Anal. Biochem.*, **166**, 308–312.
37 Towbin,H.M., Staehelin,T. and Gordon,J. (1979) *Proc. Natl. Acad. Sci. USA,* **76**, 4350–4354.
38 Phi-Van,L. and Strätling,W.H. (1988) *EMBO J.*, **7**, 655–664.
39 Mielke,C., Kohwi,Y., Kohwi-Shigematsu,T., and Bode,J. (1990) *Biochemistry,* **29,** 7475–7485.
40 Kozak,M. (1983) *Microbiol. Rev.*, **47**, 1–45.
41 Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.
42 Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.,* **120**, 97–120.
43 Chou,P.Y. and Fasman,D.G. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol.,* **74**, 45–148.
44 Fey,E.G., Ornelles,D.A. and Penman,S. (1986) *J. Cell Sci.,* **5**, 99–119.
45 Fey,E.G. and Penman,S. (1988) *Proc. Natl. Acad. Sci. USA,* **85**, 121–125.
46 Berezney,R. and Coffey,D.S. (1974) *Biochem. Biophys. Res. Commun.,* **60**, 1410–1417.
47 Mirkovitch,J., Mirault,M.E. and Laemmli,U.K. (1984) *Cell,* **39**, 223–232.
48 Cockerill,P.N. and Garrard,W.T. (1986b) *Cell,* **44**, 273–282.
49 Cook,P.R., Brazell,I.A. and Jost,E. (1976) *J. Cell Sci.,* **22**, 303–324.
50 Paulson,J.R. and Laemmli,U.K. (1977) *Cell,* **12**, 817–828.
51 Vogelstein,B., Pardoll,D.M. and Coffey,D.S. (1980) *Cell,* **22**, 79–85.
52 Heck,M.M.S. and Earnshaw,W.C. (1986) *J. Cell Biol.,* **103**, 2569–2581.